Résumé du Cours de Modèles de Régression

Yves Tillé

▶ **To cite this version:**

Yves Tillé. Résumé du Cours de Modèles de Régression. Licence. Modèles de Régression, Université de Neuchâtel (Neuchâtel), Suisse. 2015, pp.328. ⟨hal-04952805⟩

HAL Id: hal-04952805

https://hal.science/hal-04952805v1

Submitted on 17 Feb 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons CC BY-NC-ND 4.0 - Attribution - Non-commercial use - No Derivative Works - International License

Résumé du Cours de

Modèles de Régression

Yves Tillé

avec la collaboration de

Erika Antal, Lise Glaus, Caren Hasler, Vittoria Sacco et Audrey-Anne Vallée

14 février 2025

Préface

Ce document contient les notes du cours de "modèle de régression" donné à l'Université de Neuchâtel jusqu'en 2015. Ce cours contient une introduction à la régression multivariée et aux modèles linéaires.

Ce manuel contient aussi un ensemble d'exercices qui ont été réalisés par les collaboratrices qui ont préparé ou animé les séances de travaux pratiques de ce cours : Erika Antal, Lise Glaus, Caren Hasler, Vittoria Sacco et Audrey-Anne Vallée. Un ensemble de questions ayant servi aux différentes évaluations sont aussi présentées. Toutes les solutions se trouvent dans le fascicule. Cette version est donc plus complète que les anciennes versions qui ont été disséminées sur internet.

Après avoir décidé de mettre mon matériel pédagogique en ligne, j'ai reçu un grand nombre de courriers électroniques qui montrent que le cours a été utile pour beaucoup d'étudiants. Un manuel est probablement le meilleur outil pour un étudiant désireux d'apprendre.

Ce fascicule comprend un ensemble de codes en langage R. Pour une introduction bien faite en français, le manuel en libre accès de Paradis (2002) est très bien réalisé. On peut le télécharger facilement depuis le site de développement du langage R.

Yves Tillé, le 14 février 2025

Table des matières

1 Régression bivariée **11**
1.1 Série statistique bivariée . . . 11
1.1.1 Représentation graphique de deux variables . . . 11
1.1.2 Analyse des variables . . . 11
1.1.3 Covariance . . . 12
1.1.4 Corrélation . . . 12
1.1.5 Droite de régression . . . 13
1.1.6 Résidus et valeurs ajustées . . . 15
1.1.7 Sommes de carrés et variances . . . 17
1.1.8 Décomposition de la variance . . . 18
1.2 La régression par l'origine (sans constante) . . . 19
1.3 La régression avec une pente nulle . . . 20

2 Éléments d'algèbre linéaire **21**
2.1 Espace vectoriel . . . 21
2.1.1 Vecteur . . . 21
2.1.2 Multiplication par un scalaire et addition . . . 21
2.1.3 Vecteurs linéairement indépendants . . . 22
2.1.4 Sous-espace vectoriel . . . 22
2.1.5 Système générateur d'un sous-espace vectoriel . . . 22
2.1.6 Base d'un sous-espace vectoriel . . . 22
2.1.7 Base canonique de $\mathbb{R}^n$ . . . 22
2.1.8 Dimension d'un sous-espace vectoriel . . . 22
2.2 Espace euclidien . . . 22
2.2.1 Produit scalaire . . . 22
2.2.2 Norme . . . 23
2.2.3 Distance entre deux vecteurs . . . 23
2.2.4 Vecteurs orthogonaux . . . 23
2.2.5 Orthogonal d'un sous-espace vectoriel . . . 23
2.3 Application linéaire et matrices . . . 23
2.3.1 Application linéaire . . . 23
2.3.2 Matrice . . . 23
2.3.3 Produit d'une matrice et d'un vecteur . . . 24
2.3.4 Produit matriciel . . . 24
2.3.5 Transposition . . . 24
2.3.6 Matrices carrées, symétriques et diagonales . . . 24
2.3.7 Rang d'une matrice . . . 25
2.3.8 Trace d'une matrice . . . 25
2.3.9 Matrices inversibles . . . 25
2.3.10 Inversion par bloc . . . 25
2.3.11 Déterminant . . . 26
2.3.12 Quelques propriétés . . . 26
2.3.13 Matrices orthogonales . . . 26
2.3.14 Valeurs propres et vecteurs propres . . . 27
2.3.15 Formes et applications linéaires, formes quadratiques . . . 27
2.3.16 Image et noyau d'une matrice . . . 28
2.4 Projection et matrice idempotente . . . 28
2.4.1 Projection . . . 28
2.4.2 Projection orthogonale . . . 28
2.4.3 Projection orthogonale dans l'image et le noyau d'une matrice . . . 28

2.4.4 Matrice idempotente . . . 29
2.4.5 Projecteurs obliques . . . 29
2.4.6 Théorème des trois perpendiculaires . . . 30
2.5 Dérivée par rapport à un vecteur . . . 30
2.5.1 Gradient . . . 30
2.5.2 Dérivation d'une forme linéaire . . . 30
2.5.3 Dérivation d'une application linéaire . . . 30
2.5.4 Dérivée d'une forme quadratique . . . 30

3 Régression multivariée **33**
3.1 Représentation matricielle des données . . . 33
3.2 Principe des moindres carrés . . . 33
3.3 Valeurs ajustées et résidus . . . 35
3.4 Variance de régression et variance résiduelle . . . 35
3.5 Coefficient de détermination . . . 35
3.6 Matrice de variance-covariance et matrice de corrélation . . . 36
3.7 Corrélations partielles . . . 37
3.7.1 Cas général . . . 37
3.7.2 Cas où une seule variable x est utilisée . . . 37
3.8 Condition pour que la somme des résidus soit nulle . . . 38
3.9 Décomposition en sommes de carrés . . . 39
3.10 Régression avec les données centrées . . . 39
3.11 Retour au cas bivarié . . . 41
3.12 Méthode 1 . . . 41
3.13 Méthode 2 . . . 42

4 Calcul des probabilités **43**
4.1 Probabilités . . . 43
4.1.1 Événement . . . 43
4.1.2 Opérations sur les événements . . . 43
4.1.3 Relations entre les événements . . . 44
4.1.4 Ensemble des parties d'un ensemble et système complet . . . 44
4.1.5 Axiomatique des Probabilités . . . 45
4.1.6 Probabilités conditionnelles et indépendance . . . 46
4.1.7 Théorème des probabilités totales et théorème de Bayes . . . 47
4.2 Analyse combinatoire . . . 48
4.2.1 Introduction . . . 48
4.2.2 Permutations (sans répétition) . . . 48
4.2.3 Permutations avec répétition . . . 48
4.2.4 Arrangements (sans répétition) . . . 48
4.2.5 Combinaisons . . . 49
4.3 Variables aléatoires . . . 49
4.3.1 Définition . . . 49
4.4 Variables aléatoires discrètes . . . 49
4.4.1 Définition, espérance et variance . . . 49
4.4.2 Variable indicatrice ou bernoullienne . . . 50
4.4.3 Variable binomiale . . . 50
4.4.4 Variable de Poisson . . . 52
4.5 Variable aléatoire continue . . . 53
4.5.1 Définition, espérance et variance . . . 53
4.5.2 Variable uniforme . . . 54
4.5.3 Variable normale . . . 55
4.5.4 Variable normale centrée réduite . . . 56
4.5.5 Distribution exponentielle . . . 57
4.6 Distribution bivariée, cas continu . . . 57
4.6.1 Généralités . . . 57
4.6.2 Indépendance de deux variables aléatoires . . . 59
4.6.3 Propriétés des espérances et des variances . . . 59
4.6.4 Somme de variables aléatoires . . . 60
4.6.5 Loi normale bivariée . . . 60
4.7 Vecteurs aléatoires . . . 64

4.7.1 Généralités . . . 64
4.7.2 Variable normale multivariée . . . 64
4.8 Autres variables aléatoires . . . 66
4.8.1 Variable khi-carrée . . . 66
4.8.2 Variable de Student . . . 66
4.8.3 Variable de Fisher . . . 66
4.9 Inférence statistique . . . 67
4.9.1 Modélisation . . . 67
4.9.2 Intervalle de confiance . . . 67
4.9.3 Tests d'hypothèses . . . 67

5 Le modèle de régression avec une variable explicative et une constante 69
5.1 Le modèle de régression . . . 69
5.2 Estimation des paramètres par les moindres carrés . . . 70
5.3 Valeurs ajustées et résidus . . . 72
5.4 Espérance des sommes de carrés . . . 73
5.5 Le modèle linéaire avec normalité des erreurs . . . 74
5.6 Distribution de probabilités des estimateurs . . . 76
5.7 Intervalles de confiance sur les coefficients de régression . . . 76
5.8 Test d'hypothèses sur un coefficient de régression . . . 77
5.9 Le tableau d'analyse de la variance . . . 77
5.10 Exemple et interprétation . . . 78

6 Le modèle linéaire général 81
6.1 Définition du modèle linéaire général . . . 81
6.2 Régression avec les résidus normaux . . . 81
6.2.1 Hypothèses du modèle linéaire général . . . 81
6.2.2 Données observées et formulation matricielle . . . 81
6.2.3 Autre présentation du modèle linéaire général . . . 82
6.3 Estimation du modèle . . . 82
6.3.1 Estimation par les moindres carrés (ordinaires) . . . 82
6.3.2 Estimateurs du maximum de vraisemblance . . . 84
6.3.3 Propriétés des estimateurs du maximum de vraisemblance . . . 85
6.3.4 Distribution de probabilité des estimateurs . . . 85
6.3.5 Synthèse des résultats . . . 87

7 Inférence dans le modèle linéaire 89
7.1 Intervalle de confiance sur un coefficient de régression . . . 89
7.2 Test d'un seul coefficient de régression . . . 90
7.2.1 Construction du test . . . 90
7.2.2 Modèle linéaire avec uniquement une constante . . . 90
7.3 Tests de Wald sur les coefficients de régression . . . 91
7.3.1 Test général d'une contrainte linéaire . . . 91
7.3.2 Test global des coefficients de régression . . . 92
7.3.3 Test de Fisher sur un coefficient de régression . . . 94
7.4 R-carré et R-carré ajusté . . . 94
7.5 Prévision ponctuelle d'une valeur . . . 95
7.5.1 Cas général . . . 95
7.5.2 Cas bivarié . . . 95
7.6 Un exemple . . . 96

8 Outils de détection en régression 101
8.1 Le qq-plot . . . 101
8.2 Points leviers . . . 103
8.3 Résidus standardisés et studentisés . . . 103
8.4 Distance de Cook . . . 104
8.5 Un premier exemple . . . 104
8.6 Un second exemple . . . 106

9 Analyse de la variance à un facteur 109
9.1 Le problème . . . 109

9.2 Méthode 1 . . . 109
9.3 Méthode 2 . . . 112
9.4 Exemple d'analyse de la variance à un facteur . . . 115
9.4.1 Les données . . . 115
9.4.2 Les résultats . . . 115

I Exercices **117**

10 Exercices : Régression bivariée **119**
10.1 Prix des maisons . . . 119
10.2 Temps et distances . . . 121
10.3 Tailles et poids . . . 123
10.4 Tailles et poids en R . . . 126
10.5 Masse monétaire et produit national brut . . . 128

11 Exercices : Algèbre linéaire **131**
11.1 Vecteurs et matrices . . . 131
11.2 Matrices . . . 133
11.5 Inversion par bloc . . . 136
11.6 Valeurs propres . . . 137
11.7 Décomposition d'une matrice . . . 138
11.8 Vecteurs propres d'une matrice symétrique . . . 140
11.9 Valeurs et vecteurs propres . . . 141
11.10 Puissance de matrice . . . 142
11.11 Projections . . . 143

12 Exercices : Régression multivariée **149**
12.1 Prix, surface et années . . . 149
12.2 Crèmes glacées . . . 152
12.3 Crèmes glacées en R . . . 155

13 Exercices : Calcul des probabilités **157**
13.1 Tennis . . . 157
13.2 Cartes et As . . . 159
13.3 Défauts et probabilités . . . 160
13.4 Sondage . . . 161
13.5 Densité exponentielle . . . 162
13.6 Étude de densité . . . 163
13.7 Probabilités dans une loi normale . . . 164
13.8 Densité jointe . . . 165
13.9 Distribution de Poisson . . . 166
13.10 Densité jointe et conditionnement . . . 167
13.11 Densité jointe normale . . . 169
13.12 Binomiale . . . 171
13.13 Pièces et probabilités . . . 172
13.14 Densité jointe normale et corrélation . . . 173
13.16 Densité uniforme et estimation . . . 178

14 Exercices : Modèle de régression bivariée **181**
14.1 Modèle bivariée . . . 181
14.2 Modèle et sommes de carrés . . . 182
14.3 Prix et poids . . . 183
14.4 Frais d'inscription . . . 186
14.5 Frais de réparation et âge . . . 189
14.6 Revenu et sexe . . . 192

15 Exercices : Modèle linéaire général **195**
15.1 Deux variables explicatives . . . 195

16 Exercices : Inférence dans le modèle linéaire **199**
16.1 Inférence . . . 199

16.2 Inférence et test . . . 201
16.3 Tests de Wald . . . 203
16.4 Relation F et R^2 . . . 206
16.5 Vraisemblance . . . 207
16.6 Estimateurs linéaires . . . 209
16.8 Test de Wald . . . 213
16.9 Test sur les régresseurs . . . 215
16.10 Test sur des contraintes linéaires . . . 216

17 Exercices : Outils de détection en régression **219**
17.1 Détection et villes . . . 219
17.2 Leviers et influence . . . 223

18 Exercices : Analyse de la variance à un facteur **225**
18.1 Traitements et poids . . . 225
18.2 Âge et mémorisation . . . 227
18.3 Âge et traitements . . . 228
18.4 Vente de pains . . . 230
18.5 Comportement maternel de rats . . . 231

II Matériel pour l'évaluation **233**

19 Questions à choix multiples **235**

20 Questions ouvertes **245**

21 Projets en langage R **299**
21.1 Cigarettes . . . 299
21.2 Données macroéconomiques des USA . . . 301
21.3 Accidents de la Route . . . 302
21.4 Résultats aux examens en Californie . . . 304
21.5 Croissance . . . 305

22 Tables statistiques **309**

Liste des tableaux **319**

Liste des figures **321**

Références **323**

Index **325**

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 :

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

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

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

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 :

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

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

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

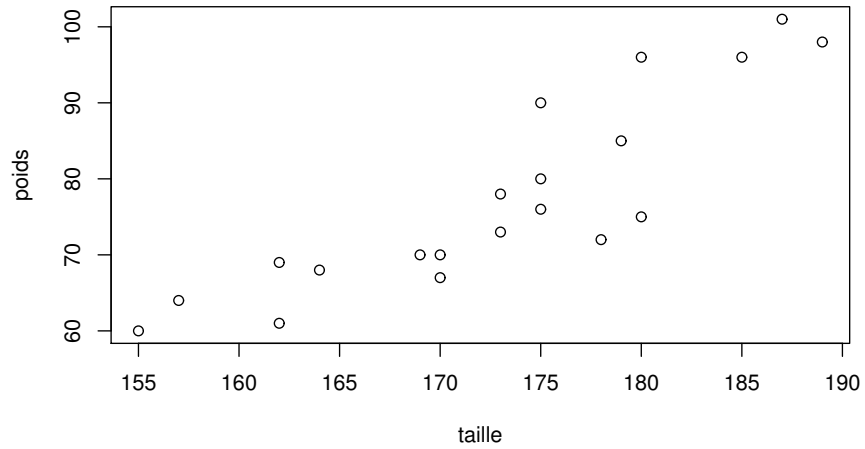


FIGURE 1.1 – Le nuage de points

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 $\mathbb{R}^n$:

$$\tilde{\mathrm{x}} = (x_1 - \bar{x}, \dots, x_i - \bar{x}, \dots, x_n - \bar{x})^\top$$

$$\tilde{\mathrm{y}} = (y_1 - \bar{y}, \dots, y_i - \bar{y}, \dots, y_n - \bar{y})^\top.$$

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

$$\cos(\tilde{\mathrm{x}}, \tilde{\mathrm{y}}) = \frac{< \tilde{\mathrm{x}}, \tilde{\mathrm{y}} >}{||\tilde{\mathrm{x}}|| \times ||\tilde{\mathrm{y}}||},$$

où $< \tilde{\mathrm{x}}, \tilde{\mathrm{y}} >$ est le produit scalaire entre les vecteurs $\tilde{\mathrm{x}}$ et $\tilde{\mathrm{y}}$

$$< \tilde{\mathrm{x}}, \tilde{\mathrm{y}} >= \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}),$$

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

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

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

$$-1 \leq r_{xy} \leq 1. \tag{1.1}$$

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.

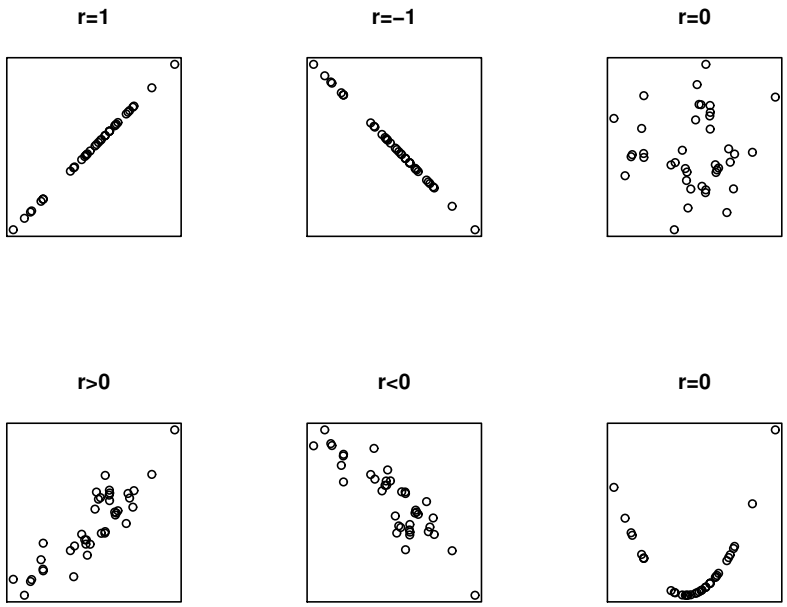


FIGURE 1.2 – Exemples de nuages de points et coefficients de corrélation

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.

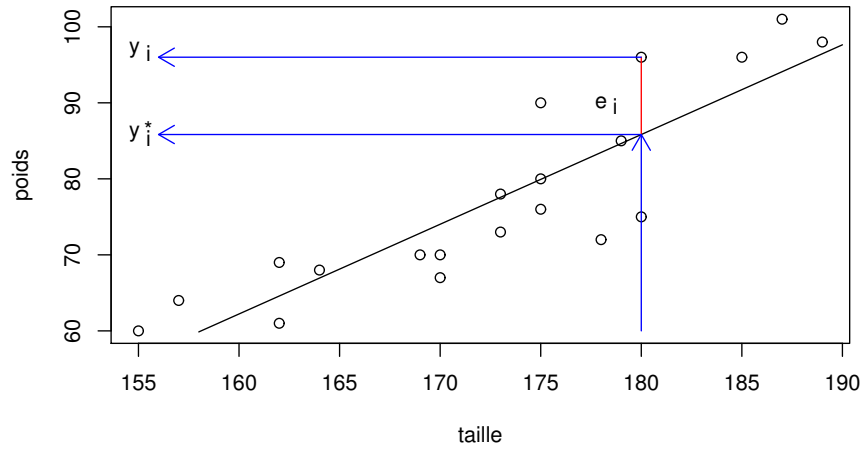


FIGURE 1.3 – Le nuage de points, le résidu

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 :

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

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

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

Démonstration. Le minimum $M(a,b)$ en a, b s'obtient en annulant les dérivées partielles par rapport à a et b.

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

On obtient un système de deux équations à deux inconnues. En divisant les deux équations par $-2n$, on obtient :

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

ou encore

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

ce qui s'écrit aussi

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

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

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

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

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

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 :

$$\mathrm{H} = \begin{pmatrix} \dfrac{\partial^2 M(a,b)}{\partial a^2} & \dfrac{\partial^2 M(a,b)}{\partial a \partial b} \\ \dfrac{\partial^2 M(a,b)}{\partial a \partial b} & \dfrac{\partial^2 M(a,b)}{\partial b^2} \end{pmatrix}.$$

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

$$\mathrm{H} = 2\begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix}.$$

et peut s'écrire $\mathrm{H} = 2\mathrm{X}^\top\mathrm{X}$, où

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

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

La droite 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

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

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,

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

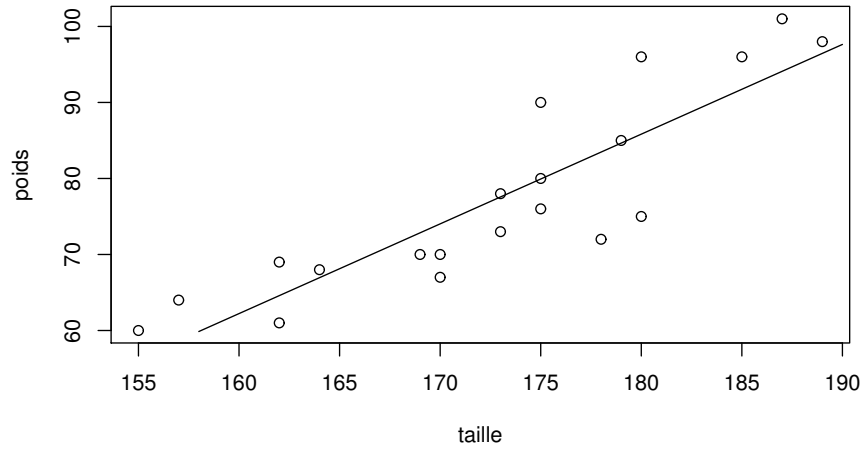


FIGURE 1.4 – La droite de régression

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.

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

Or,

$$\sum_{i=1}^{n} x_i(x_i - \bar{x}) = \sum_{i=1}^{n} x_i^2 - \bar{x}\sum_{i=1}^{n} x_i = n\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}\underbrace{\frac{1}{n}\sum_{i=1}^{n} x_i}_{\bar{x}}\right) = n\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2\right) = ns_x^2. \qquad (1.3)$$

De plus,

$$\sum_{i=1}^{n} x_i(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - \bar{y}\sum_{i=1}^{n} x_i = n\left(\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{y}\underbrace{\frac{1}{n}\sum_{i=1}^{n} x_i}_{\bar{x}}\right) = n\left(\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}\right) = ns_{xy}. \qquad (1.4)$$

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

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

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 $\mathbb{R}^n$:

$$\mathrm{x} = (x_1, \ldots, x_i, \ldots, x_n)^\top \text{ et } \mathrm{y} = (y_1, \ldots, y_i, \ldots, y_n)^\top.$$

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

$$\mathrm{y}^* = (y_1^*, \ldots, y_i^*, \ldots, y_n^*)^\top \text{ et } \mathrm{e} = (e_1, \ldots, e_i, \ldots, e_n)^\top.$$

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

$$1_n = (1, \ldots, 1, \ldots, 1)^\top.$$

La régression permet la décomposition du vecteur y

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

où

$$\mathrm{y}^* = a1_n + b\mathrm{x}.$$

On a donc les relations suivantes
— y^* et e sont orthogonaux, i.e. $< \mathrm{y}^*, \mathrm{e} >= 0$.
— x et e sont orthogonaux, i.e. $< \mathrm{x}, \mathrm{e} >= 0$.
— 1_n et e sont orthogonaux, i.e. $< 1_n, \mathrm{e} >= 0$.
La régression décompose le vecteur y en deux vecteurs orthogonaux y^* et e. Le vecteur e est par ailleurs orthogonal au vecteur contenant la variable explicative et au vecteur 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

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

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

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

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.

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

Démonstration.

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

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

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

□

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.

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

□

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.

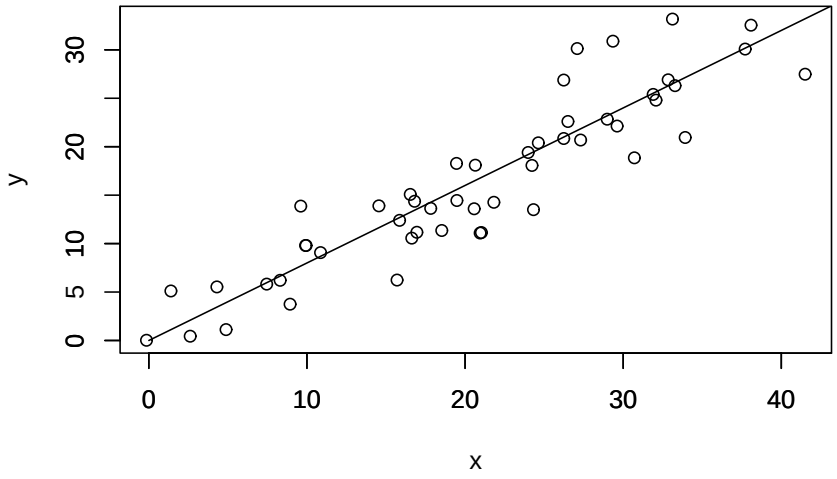


FIGURE 1.5 – Droite passant par l'origine

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 point sont alignés le long d'une droite horizontale comme dans la Figure 1.6.

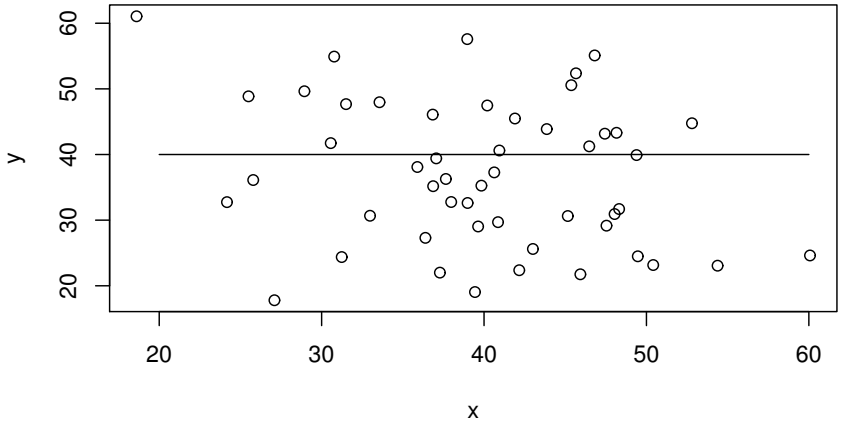


FIGURE 1.6 – Régression avec une pente nulle

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 $\mathrm{a} = [3 \quad 0]$, est un vecteur ligne et le vecteur

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

est un vecteur colonne.

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

Exemple 2.2 Si $\mathrm{a} = (3 \quad 0)$, la transposée de a est

$$\mathrm{a}^\top = \begin{pmatrix} 3 \\ 0 \end{pmatrix}.$$

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 a de $\mathbb{R}^n$, alors

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

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 a et b :

$$c_1\mathrm{a} + c_2\mathrm{b} = c_1 \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} + c_2 \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} c_1a_1 + c_2b_1 \\ \vdots \\ c_1a_n + c_2b_n \end{pmatrix},$$

où $c_1, c_2 \in \mathbb{R}$.

2.1.3 Vecteurs linéairement indépendants

Définition 2.1. *Les vecteurs* $\mathrm{u}_1, \ldots, \mathrm{u}_j, \ldots, \mathrm{u}_J$ *sont dit linéairement indépendants, si*

$$a_1\mathrm{u}_1 + a_2\mathrm{u}_2 + \cdots + a_J\mathrm{u}_J = 0$$

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

2.1.4 Sous-espace vectoriel

Définition 2.2. *Un sous-ensemble non-vide* V *de* $\mathbb{R}^n$ *est un sous-espace vectoriel, si pour tous* $\mathrm{u}, \mathrm{v} \in V$,

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

2.1.5 Système générateur d'un sous-espace vectoriel

Définition 2.3. *Un ensemble de* p *vecteurs* $\mathrm{u}_1, \ldots, \mathrm{u}_p$ *du sous-espace vectoriel* V *forment un système générateur de* V *si et seulement si*

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

2.1.6 Base d'un sous-espace vectoriel

Définition 2.4. *Un ensemble de* p *vecteurs* $\mathrm{u}_1, \ldots, \mathrm{u}_p$ *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 $\mathrm{u}_1, \ldots, \mathrm{u}_p$.

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

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

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

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 a par un vecteur colonne b comme le résultat scalaire :

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

Le produit scalaire de deux vecteurs colonnes u et b de même dimension est noté $< \mathrm{u}, \mathrm{b} >$ et est défini par :

$$< \mathrm{u}, \mathrm{b} >= \mathrm{u}^\top \mathrm{b} = (u_1 \ldots u_n) \times \begin{pmatrix}b_1\\\vdots\\b_n\end{pmatrix} = \sum_{i=1}^{n} u_i b_i.$$

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* u *est*

$$||\mathrm{u}|| = \sqrt{<\mathrm{u}, \mathrm{u}>}.$$

Un vecteur de norme égale à 1 est dit normé.

2.2.3 Distance entre deux vecteurs

Définition 2.8. *La distance entre les vecteurs* u *et* v *de* $\mathbb{R}^n$ *est définie par*

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

Définition 2.9. *La projection d'un vecteur* u *sur un vecteur* v *est définie par*

$$p_v(\mathrm{u}) = \frac{<\mathrm{u}, \mathrm{v}>\mathrm{v}}{||\mathrm{v}||^2}. \tag{2.1}$$

2.2.4 Vecteurs orthogonaux

Définition 2.10. *Deux vecteurs non-nuls* u *et* v *de* $\mathbb{R}^n$ *sont orthogonaux si*

$$<\mathrm{u}, \mathrm{v}>= 0.$$

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

Théorème 2.1. *(de Pythagore) Si* u *et* v *sont orthogonaux, alors*

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

2.2.5 Orthogonal d'un sous-espace vectoriel

Définition 2.11. *Un vecteur* 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*

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

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

2.3 Application linéaire et matrices

2.3.1 Application linéaire

Une application $f(.)$ de $\mathbb{R}^J$ dans $\mathbb{R}^I$ est dite linéaire si pour tous u, v, de $\mathbb{R}^J$ et tout $a \in \mathbb{R}$
— $f(\mathrm{u} + \mathrm{v}) = f(\mathrm{u}) + f(\mathrm{v})$,
— $f(a\mathrm{u}) = af(\mathrm{u})$.

2.3.2 Matrice

Une matrice est un tableau de nombres. Par exemple :

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

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 A de dimension $I \times J$ et un vecteur colonne u de dimension J le produit Au est donné par

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

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

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

où

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

C'est le produit des lignes par les colonnes. La matrice 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

$$\mathrm{A} = \begin{pmatrix} -1 & 2 \\ 4 & 3 \\ -2 & 5 \end{pmatrix} \text{ alors } \mathrm{A}^\top = \begin{pmatrix} -1 & 4 & -2 \\ 2 & 3 & 5 \end{pmatrix}.$$

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

$$(\mathrm{ABC})^\top = \mathrm{C}^\top \mathrm{B}^\top \mathrm{A}^\top.$$

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,

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

est une matrice diagonale.

Définition 2.17. *Une matrice identité* I *est une matrice diagonale dont tous les éléments de la diagonale sont égaux à* 1.

Par exemple,

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

est une matrice identité de dimension 3×3.

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.* $\mathrm{trace}(\mathrm{A} + \mathrm{B}) = \mathrm{trace}(\mathrm{A}) + \mathrm{trace}(\mathrm{B})$.

2. $\mathrm{trace}(\mathrm{AB}) = \mathrm{trace}(\mathrm{BA})$ *mais* $\mathrm{trace}(\mathrm{AB}) \neq \mathrm{trace}(\mathrm{A})\mathrm{trace}(\mathrm{B})$.

2.3.9 Matrices inversibles

Définition 2.21. *Une matrice carrée* A *est dite inversible, s'il existe une matrice* A^{-1} *qui vérifie* $\mathrm{AA}^{-1} = \mathrm{A}^{-1}\mathrm{A} = \mathrm{I}$.

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

2.3.10 Inversion par bloc

Soit une matrice F composée de quatre sous-matrices :

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

Les matrices A et D sont carrées et inversibles.

La technique d'inversion par bloc permet d'obtenir l'inverse de F.

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

où

$$\mathrm{Q} = \left(\mathrm{D} - \mathrm{CA}^{-1}\mathrm{B}\right)^{-1}$$

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

2.3.11 Déterminant

Définition 2.22. *Le déterminant d'une matrice carrée* A *$(J \times J)$ est noté* $|\mathrm{A}|$ *et est défini par*

— *Si* $J = 1$, $|\mathrm{A}| = \mathrm{A}$

— *Si* $J > 1$,

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

pour tout j fixé, où $|\mathrm{M}_{ij}|$ 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 A.

Exemple 2.3 Soit A une matrice (2×2),

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

en prenant $j = 1$, on a

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

On peut aussi calculer le déterminant de A en prenant $j = 2$.

Exemple 2.4 Soit une matrice A de dimension (3×3), le calcul se fait en prenant $j = 1$

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

alors son déterminant vaut

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

Propriété 2.6.

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

2.3.12 Quelques propriétés

Propriété 2.7. *En général, si* A, B *et* C *sont des matrices carrées de même dimension, on a*

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

2.3.13 Matrices orthogonales

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

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

2.3.14 Valeurs propres et vecteurs propres

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

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

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* $\mathrm{u}_i \neq 0$ *est un vecteur propre de* A *associé à la valeur propre* λ_i *si*

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

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

Théorème 2.2. *(de diagonalisation) Soient* A *une matrice symétrique* $(J \times J)$ *et* $\mathrm{u}_i, \lambda_i, i = 1, \ldots, J$, *ses valeurs propres et vecteurs propres associés. Soient la matrice orthogonale* $\mathbf{\Gamma}$ *dont les colonnes sont les* J *vecteurs propres de* A, *et la matrice diagonale* $\mathbf{\Lambda}$ *ayant sur sa diagonale principale les* J *valeurs propres. Alors*
— $\mathbf{\Gamma}^\top \mathrm{A} \mathbf{\Gamma} = \mathbf{\Lambda}$,
— $\mathrm{A} = \mathbf{\Gamma} \mathbf{\Lambda} \mathbf{\Gamma}^\top$.

2.3.15 Formes et applications linéaires, formes quadratiques

Définition 2.26. *Soient* A *une matrice* $(I \times I)$, B *une matrice* $(J \times I)$, a *et* c *deux vecteurs colonnes de dimension* I. *On appelle*
— *forme linéaire définie par le vecteur* $\mathrm{a}^\top$, *l'application de* $\mathbb{R}^I$ *dans* $\mathbb{R}$

$$\mathrm{a}^\top \mathrm{c},$$

— *application linéaire de* $\mathbb{R}^I$ *dans* $\mathbb{R}^J$ *définie par la matrice* B,

$$\mathrm{B}\mathrm{c},$$

— *et forme quadratique définie par la matrice* A, *l'expression*

$$\mathrm{c}^\top \mathrm{A} \mathrm{c}.$$

Définition 2.27. *Une matrice* A *de dimension* $(I \times I)$ *est dite définie positive si*

$$\mathrm{c}^\top \mathrm{A} \mathrm{c} > 0,$$

pour tout $\mathrm{c} \in \mathbb{R}^I \backslash \{0\}$.

Définition 2.28. *Une matrice* A *de dimension* $(I \times I)$ *est dite semi-définie positive si*

$$\mathrm{c}^\top \mathrm{A} \mathrm{c} \geq 0,$$

pour tout $\mathrm{c} \in \mathbb{R}^I$.

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* D, *la matrice* $\mathrm{D}^\top \mathrm{D}$ *est semi-définie positive.*

Démonstration. En posant $\mathrm{a} = \mathrm{D}\mathrm{c}$, la forme quadratique $\mathrm{c}^\top \mathrm{D}^\top \mathrm{D} \mathrm{c}$ peut s'écrire

$$\mathrm{c}^\top \mathrm{D}^\top \mathrm{D} \mathrm{c} = \mathrm{a}^\top \mathrm{a} = \sum_{i=1}^{n} a_i^2 \geq 0.$$

□

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* A *de dimension* $I \times J$ *est le sous-espace de* $\mathbb{R}^J$ *défini par*

$$\mathrm{Ker}(\mathrm{A}) = \left\{\mathrm{u} \in \mathbb{R}^J | \mathrm{Au} = 0\right\}.$$

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

Définition 2.30. *L'image d'une matrice* B *de dimension* $I \times J$ *est le sous-espace de* $\mathbb{R}^I$ *défini par*

$$\mathrm{Im}(\mathrm{B}) = \left\{\mathrm{x} \in \mathbb{R}^I | \textit{ il existe } \mathrm{u} \in \mathbb{R}^J \textit{ tel que } \mathrm{Bu} = \mathrm{x}\right\}.$$

Le sous-espace $\mathrm{Im}(\mathrm{B})$ est l'ensemble des vecteurs qui peuvent s'écrire comme une combinaison linéaire des colonnes de B. L'image de la matrice B est souvent appelé sous-espace engendré par les colonnes de B. La dimension de l'image de B est égale au rang de B.
Remarque 2.2 Le sous-espace $\mathrm{Im}(\mathrm{B})$ est l'orthogonal de $\mathrm{Ker}(\mathrm{B}^\top)$.

Propriété 2.13. *Si* $\mathrm{u} \in \mathrm{Im}(\mathrm{B})$ *et* $\mathrm{v} \in \mathrm{Ker}(\mathrm{B}^\top)$*, alors* u *et* v *sont orthogonaux.*

En statistique, on utilise souvent des matrices X (individus-variables) de dimension $n \times p$ avec $n \geq p$. Le sous-espace engendré par les colonnes de X est l'image de 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* $\mathrm{u} \in \mathbb{R}^n$ *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* u *fait correspondre un vecteur* u^* *tel que* $\mathrm{u} - \mathrm{u}^*$ *soit orthogonal à* V *est appelé projection orthogonale (*$\mathrm{u}^* \in V$*).*

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

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

$$\mathrm{P}_X = \mathrm{X}(\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top.$$

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

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

Remarque 2.3 Si $\mathrm{X} = \mathrm{v}$ est un vecteur, alors le projecteur est

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

et la projection de u sur v

$$p_v(\mathrm{u}) = \mathrm{P}_v\mathrm{u} = \frac{\mathrm{v}}{||\mathrm{v}||^2}\mathrm{v}^\top\mathrm{u} = \mathrm{v}\frac{<\mathrm{v},\mathrm{u}>}{||\mathrm{v}||^2},$$

ce qui correspond à la définition donnée en (2.9).

2.4.4 Matrice idempotente

Définition 2.32. *Une matrice* P *est dite idempotente si* PP = P.

Une matrice de projection est idempotente.

Remarque 2.4 Les matrices P_X et $P_X^\perp$ sont évidemment idempotentes. En effet,

$$\begin{aligned} P_X P_X &= \left\{X(X^\top X)^{-1}X^\top\right\}\left\{X(X^\top X)^{-1}X^\top\right\} \\ &= X(X^\top X)^{-1}\underbrace{X^\top X(X^\top X)^{-1}}_{=I}X^\top \\ &= X(X^\top X)^{-1}X^\top = P_X. \end{aligned}$$

De plus

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

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

$$P_X^\perp = I - X(X^\top X)^{-1}X^\top = I - P_X.$$

Théorème 2.4. *Toutes les valeurs propres d'une matrice idempotente valent 1 ou 0.*

Démonstration. Un vecteur propre non-nul u d'une matrice P doit satisfaire au système d'équation

$$Pu = \lambda u, \tag{2.2}$$

où λ est la valeur propre associée à u. En multipliant (2.2) par P, on obtient

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

et donc,

$$\lambda u = \lambda^2 u.$$

En prémultipliant par $u^\top$ on a

$$u^\top \lambda u = u^\top \lambda^2 u,$$

on obtient donc

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

ce qui n'est vérifié que si λ 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 $X(X^\top X)^{-1}X^\top$ sont égaux au rang de la matrice $(X^\top X)^{-1}$. Cette matrice est supposée de plein rang (sinon $X^\top X$ ne serait pas inversible). Le rang de $(X^\top X)^{-1}$ et donc de $P_X = X(X^\top X)^{-1}X^\top$ est donc égal au nombre de colonnes de X. Le rang de P_X est la dimension du sous-espace sur lequel projette P_X.

2.4.5 Projecteurs obliques

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

$$P_O = X(Z^\top X)^{-1}Z^\top$$

est un projecteur oblique. Il est facile de vérifier que la matrice P_O est idempotente et qu'elle projette sur l'image de 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*

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

2.5 Dérivée par rapport à un vecteur

2.5.1 Gradient

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

$$f(\mathrm{x}) = f(x_1, \ldots, x_j, \ldots, x_p).$$

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

$$\operatorname{grad} f = \frac{\partial f}{\partial \mathrm{x}^\top} = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_j}, \ldots, \frac{\partial f}{\partial x_p} \right).$$

2.5.2 Dérivation d'une forme linéaire

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

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

2.5.3 Dérivation d'une application linéaire

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

$$\mathrm{Ax} = \begin{pmatrix} \sum_{j=1}^p a_{1j} x_j \\ \vdots \\ \sum_{j=1}^p a_{ij} x_j \\ \vdots \\ \sum_{j=1}^p a_{qj} x_j \end{pmatrix}.$$

On a

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

Donc,

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

2.5.4 Dérivée d'une forme quadratique

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

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

Donc,

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

et

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

Si la matrice A est symétrique, on a $\mathrm{A} = \mathrm{A}^\top$ et donc

$$\frac{\partial \mathrm{x}^\top \mathrm{A}\mathrm{x}}{\partial \mathrm{x}} = 2\mathrm{A}\mathrm{x}.$$

Chapitre 3

Régression multivariée

3.1 Représentation matricielle des données

La matrice

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

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 $\mathrm{y} = (y_1 \dots y_i \dots y_n)^\top$ 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 :

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

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 y en X consiste à chercher un vecteur de coefficients de régression :

$$\mathrm{b} = (b_1, \dots, b_j, \dots, b_p)^\top$$

qui permet d'ajuster au mieux les y_i par les x_{ij} au sens où la somme des carrées 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 :

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

La somme des carrés des résidus est le critère à minimiser

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

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

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

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

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

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

$$Q(\mathrm{b}) = \mathrm{y}^\top \mathrm{y} - \mathrm{y}^\top \mathrm{X}\mathrm{b} - \mathrm{b}^\top \mathrm{X}^\top \mathrm{y} + \mathrm{b}^\top \mathrm{X}^\top \mathrm{X}\mathrm{b} = \mathrm{y}^\top \mathrm{y} - 2\mathrm{y}^\top \mathrm{X}\mathrm{b} + \mathrm{b}^\top \mathrm{X}^\top \mathrm{X}\mathrm{b},$$

car $\mathrm{y}^\top \mathrm{X}\mathrm{b} = \mathrm{b}^\top \mathrm{X}^\top \mathrm{y}$.

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

$$\mathrm{b} = \left(\mathrm{X}^\top \mathrm{X}\right)^{-1} \mathrm{X}^\top \mathrm{y},$$

sous réserve que la matrice $\mathrm{X}^\top \mathrm{X}$ *soit inversible.*

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

$$\frac{\partial Q(\mathrm{b})}{\partial \mathrm{b}} = \frac{\partial \mathrm{y}^\top \mathrm{y}}{\partial \mathrm{b}} - \frac{\partial 2\mathrm{y}^\top \mathrm{X}\mathrm{b}}{\partial \mathrm{b}} + \frac{\partial \mathrm{b}^\top \mathrm{X}^\top \mathrm{X}\mathrm{b}}{\partial \mathrm{b}} = 0 - 2\mathrm{X}^\top \mathrm{y} + 2\mathrm{X}^\top \mathrm{X}\mathrm{b} = 2\mathrm{X}^\top \mathrm{X}\mathrm{b} - 2\mathrm{X}^\top \mathrm{y},$$

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

$$\frac{\partial Q(\mathrm{b})}{\partial \mathrm{b}} = 2\mathrm{X}^\top \mathrm{X}\mathrm{b} - 2\mathrm{X}^\top \mathrm{y} = 0,$$

ce qui donne :

$$\mathrm{X}^\top \mathrm{X}\mathrm{b} = \mathrm{X}^\top \mathrm{y}.$$

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

$$\mathrm{b} = \left(\mathrm{X}^\top \mathrm{X}\right)^{-1} \mathrm{X}^\top \mathrm{y}.$$

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

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

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 y au moyen de X et de b, c'est-à-dire

$$\mathrm{y}^* = \mathrm{Xb} = \underbrace{\mathrm{X}(\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top}_{\mathrm{P}_X}\mathrm{y}.$$

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

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

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

$$\mathrm{P}_X = \mathrm{X}(\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top.$$

Le vecteur des résidus est la différence entre y et y^*.

$$\mathrm{e} = \mathrm{y} - \mathrm{y}^* = \mathrm{y} - \mathrm{Xb} = \mathrm{y} - \mathrm{X}(\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top\mathrm{y} = (\mathrm{I} - \mathrm{X}(\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top)\mathrm{y}.$$

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

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

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

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

Propriété 3.1.
- $\mathrm{y} = \mathrm{y}^* + \mathrm{e}$,
- y^* *est une combinaison linéaire des colonnes de* X,
- y^* *et* e *sont orthogonaux,*
- e *est orthogonal avec toutes les colonnes de* X, *c'est-à-dire* $\mathrm{e}^\top\mathrm{X} = 0$.

3.4 Variance de régression et variance résiduelle

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

$$\bar{\mathrm{y}} = (\bar{y}, \dots, \bar{y})^\top.$$

La variance peut être définie simplement par :

$$s_y^2 = \frac{1}{n}(\mathrm{y} - \bar{\mathrm{y}})^\top(\mathrm{y} - \bar{\mathrm{y}}) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2.$$

La variance de régression est la variance des valeurs ajustées :

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

La variance résiduelle est la variance des résidus :

$$s_e^2 = \frac{1}{n}\mathrm{e}^\top\mathrm{e} = \frac{1}{n}(\mathrm{y} - \mathrm{y}^*)^\top(\mathrm{y} - \mathrm{y}^*) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - y_i^*\right)^2 = \frac{1}{n}\sum_{i=1}^{n}e_i^2.$$

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 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

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

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 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 X.

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

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

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

$$\boldsymbol{\Sigma} = \frac{\mathrm{X}_c^\top \mathrm{X}_c}{n}.$$

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

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

La matrice des corrélations :

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

est obtenue par

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

3.7 Corrélations partielles

3.7.1 Cas général

Les corrélations partielles sont des outils importants mettant en jeux deux variables y et z et un groupe de variables explicatives contenues dans la matrice 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 X. Pour ce faire, on calcule le vecteur de coefficients $\mathrm{b}_{y|X}$ de la régression de y en X et le vecteur de coefficients $\mathrm{b}_{z|X}$ de la régression de z en X. On peut ensuite calculer :

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

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

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

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

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

Comme la première colonne de la matrice X est composées de uns, la somme des résidus est nulle, la covariance des résidus s'écrit :

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

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

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

et

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

Le coefficient de corrélation partielle s'écrit :

$$r_{yz|x_2,\ldots,x_p} = \frac{\mathrm{cov}(\mathrm{e}_{y|X}, \mathrm{e}_{z|X})}{\sqrt{\mathrm{var}(\mathrm{e}_{y|X})\mathrm{var}(\mathrm{e}_{z|X})}} = \frac{\mathrm{y}^{\top}\mathrm{P}_{\mathrm{X}}^{\perp}\mathrm{z}}{\sqrt{\mathrm{y}^{\top}\mathrm{P}_{\mathrm{X}}^{\perp}\mathrm{y}\mathrm{z}^{\top}\mathrm{P}_{\mathrm{X}}^{\perp}\mathrm{z}}}.$$

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 X est composée d'une ligne et d'une constante :

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

alors on obtient par le Théorème 1.4, page 19, 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

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

Le coefficient de corrélation partielle vaut donc :

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

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 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 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 $\mathrm{X}\boldsymbol{\lambda} = 1_n = (1 \cdots 1 \cdots 1)^\top$.

Théorème 3.2. *Si la matrice* X *contient une variable constante définie de manière explicite ou implicite, alors la somme des résidus est nulle.*

Démonstration. On a

$$\sum_{i=1}^{n} e_i = 1_n^\top \mathrm{e}$$

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

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

□

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

$$\mathrm{e}^\top \mathrm{X} = 0.$$

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

$$\mathrm{e}^\top \mathrm{X}\boldsymbol{\lambda} = \mathrm{e}^\top 1_n = 0^\top \boldsymbol{\lambda} = 0.$$

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*

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

se décompose donc en une somme de deux termes :

— *la somme des carrés expliquée par la régression,*

$$SC_{REG} = (\mathrm{y}^* - \bar{\mathrm{y}})^\top (\mathrm{y}^* - \bar{\mathrm{y}}) = \sum_{i=1}^{n} (y_i^* - \bar{y})^2,$$

— *la somme des carrés des résidus*

$$SC_{RES} = \mathrm{e}^\top \mathrm{e} = \sum_{i=1}^{n} (y_i - y_i^*)^2 = \sum_{i=1}^{n} e_i^2. \tag{3.3}$$

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

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

Donc,

$$(\mathrm{y} - \bar{\mathrm{y}})^\top (\mathrm{y} - \bar{\mathrm{y}}) = (\mathrm{y}^* - \bar{\mathrm{y}} + \mathrm{e})^\top (\mathrm{y}^* - \bar{\mathrm{y}} + \mathrm{e}) = (\mathrm{y}^* - \bar{\mathrm{y}})^\top (\mathrm{y}^* - \bar{\mathrm{y}}) + \mathrm{e}^\top \mathrm{e} + 2\mathrm{e}^\top (\mathrm{y}^* - \bar{\mathrm{y}})$$

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

$$(\mathrm{y} - \bar{\mathrm{y}})^\top (\mathrm{y} - \bar{\mathrm{y}}) = (\mathrm{y}^* - \bar{\mathrm{y}})^\top (\mathrm{y}^* - \bar{\mathrm{y}}) + \mathrm{e}^\top \mathrm{e}.$$

□

3.10 Régression avec les données centrées

Supposons que la première colonne de la matrice X soit composée de uns :

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

Dans ce cas, la régression multiple s'écrit :

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

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

$$\bar{y} = b_1 + \bar{x}_2 b_2 + \bar{x}_3 b_3 + \cdots + \bar{x}_p b_p, \tag{3.5}$$

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

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

Définissons maintenant

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

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

3. $1 = (1, 1, \ldots, 1)^\top$: le vecteur colonne de n uns,
4. la matrice idempotente qui centre les valeurs :

$$\mathrm{P}_c = \mathrm{I} - \frac{11^\top}{n} = \begin{pmatrix} 1-1/n & -1/n & -1/n & \ldots & -1/n \\ -1/n & 1-1/n & -1/n & \ldots & -1/n \\ -1/n & -1/n & 1-1/n & \ldots & -1/n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1/n & -1/n & -1/n & \ldots & 1-1/n \end{pmatrix}, \tag{3.7}$$

5. $\mathrm{y}_c = \mathrm{P}_c\mathrm{y} = \mathrm{y} - 1\bar{y} = \mathrm{y} - \bar{\mathrm{y}} = (y_1 - \bar{y}, y_2 - \bar{y}, \ldots, y_n - \bar{y})^\top$
6. $\mathrm{X}_c = \mathrm{P}_c\widetilde{\mathrm{X}}$ la matrice $\widetilde{\mathrm{X}}$ centrée

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

La régression multiple peut maintenant s'écrire :

$$\mathrm{y}_c = \mathrm{X}_c\widetilde{\mathrm{b}} + \mathrm{e}.$$

Le vecteur $\widetilde{\mathrm{b}}$ est évidemment défini par

$$\widetilde{\mathrm{b}} = (\mathrm{X}_c^\top\mathrm{X}_c)^{-1}\mathrm{X}_c^\top\mathrm{y}_c = \left(\frac{\mathrm{X}_c^\top\mathrm{X}_c}{n}\right)^{-1}\frac{\mathrm{X}_c^\top\mathrm{y}_c}{n}. \tag{3.8}$$

Cette présentation est intéressante à plus d'un titre. En effet, $(\mathrm{X}_c^\top\mathrm{X}_c)/n$ n'est autre que la matrice variance-covariance $\boldsymbol{\Sigma}$ donnée en (3.2).

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

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

$$\frac{\mathrm{X}_c^\top\mathrm{y}_c}{n} = \begin{pmatrix} s_{2y} \\ \vdots \\ s_{jy} \\ \vdots \\ s_{py} \end{pmatrix},$$

où

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

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

Comme,

$$\mathrm{y}_c = \mathrm{X}_c\widetilde{\mathrm{b}} + \mathrm{e},$$

la décomposition en somme de carrés vient directement :

$$\mathrm{y}_c^\top \mathrm{y}_c = (\mathrm{X}_c\widetilde{\mathrm{b}} + \mathrm{e})^\top(\mathrm{X}_c\widetilde{\mathrm{b}} + \mathrm{e}) = \widetilde{\mathrm{b}}^\top \mathrm{X}_c^\top \mathrm{X}_c\widetilde{\mathrm{b}} + \mathrm{e}^\top\mathrm{e} + \underbrace{2\mathrm{e}^\top \mathrm{X}_c\widetilde{\mathrm{b}}}_{0}.$$

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

$$\mathrm{SC}_{TOT} = \mathrm{SC}_{REG} + \mathrm{SC}_{RES},$$

où

— la somme des carrés totale

$$\mathrm{SC}_{TOT} = \mathrm{y}_c^\top \mathrm{y}_c = \sum_{i=1}^{n}(y_i - \bar{y})^2, \tag{3.9}$$

— la somme des carrés expliquée par la régression,

$$\mathrm{SC}_{REG} = \widetilde{\mathrm{b}}^\top \mathrm{X}_c^\top \mathrm{X}_c\widetilde{\mathrm{b}}, \tag{3.10}$$

car

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

et que donc

$$\widetilde{\mathrm{b}}^\top \mathrm{X}_c^\top \mathrm{X}_c\widetilde{\mathrm{b}} = (\mathrm{y}^* - \bar{\mathrm{y}})^\top(\mathrm{y}^* - \bar{\mathrm{y}}) = \sum_{i=1}^{n}(y_i^* - \bar{y})^2 = \mathrm{SC}_{REG},$$

— la somme des carrés des résidus

$$\mathrm{SC}_{RES} = \mathrm{e}^\top\mathrm{e} = \sum_{i=1}^{n} e_i^2. \tag{3.11}$$

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,

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

On a alors

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

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

où

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

De plus,

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

ce qui permet de calculer b

$$\mathrm{b} = \left(\mathrm{X}^\top\mathrm{X}\right)^{-1}\mathrm{X}^\top\mathrm{y} = \frac{1}{s_x^2}\begin{pmatrix}(s_x^2 + \bar{x}^2)\bar{y} - \bar{x}(s_{xy} + \bar{x}\bar{y})\\ -\bar{x}\bar{y} + (s_{xy} + \bar{x}\bar{y})\end{pmatrix} = \begin{pmatrix}\bar{y} - \bar{x}\dfrac{s_{xy}}{s_x^2}\\ \dfrac{s_{xy}}{s_x^2}\end{pmatrix}.$$

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

$$\mathrm{y}^* = (y_i^*) = \mathrm{Xb},$$

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

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

On obtient

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

Il reste à déduire b_1 de l'équation

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

ce qui donne

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

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 ω un résultat possible de cette expérience aléatoire. L'ensemble de tous les résultats possibles est noté Ω. Par exemple, si on jette deux pièces de monnaie, on peut obtenir les résultats

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

avec F pour "face" et P pour "pile". Un événement est une assertion logique sur une expérience aléatoire comme "avoir deux fois pile" ou "avoir au moins une fois pile". Formellement, un événement est un sous-ensemble de Ω.

— 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 Ω 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 $A \backslash B$ est réalisé quand A est réalisé et que B ne l'est pas.

Le complémentaire

Le complémentaire de l'événement A est l'événement $\Omega \backslash A$. Le complémentaire est noté $\overline{A}$.

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

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

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-vide 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 événement 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 Ω l'ensemble $\mathcal{A}$ de toutes les parties (ou sous-ensembles) de Ω.

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 $A_1, \ldots, A_n$ forment un système complet d'événements, s'ils constituent une partition de Ω, c'est-à-dire si*

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

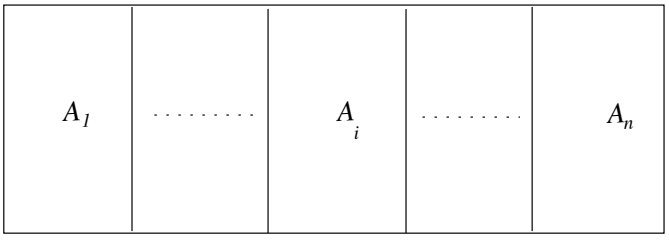


FIGURE 4.1 – Système complet d'événements

4.1.5 Axiomatique des Probabilités

Définition 4.2. *Une probabilité $P(.)$ est une application de $\mathcal{A}$ dans $[0,1]$, telle que :*
- $\Pr(\Omega) = 1$,
- *Pour tout ensemble dénombrable d'événements $A_1, .., A_n$ mutuellement exclusifs (tels que $A_i \cap A_j = \emptyset$, pour tout $i \neq j$,)*

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

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. $\Pr(\overline{A}) = 1 - \Pr(A)$.

Démonstration. On sait que

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

Ainsi, on a que

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

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

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

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

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

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

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

Mais on a que

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

Ainsi, on a

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

Or, une probabilité est à valeur dans [0,1]. Donc, $\Pr(B \cap \overline{A}) \geq 0$. 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

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

et

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

Ainsi

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

Il reste à montrer que

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

Mais

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

avec

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

Donc,

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

□

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

Démonstration. Notons respectivement

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

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

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

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

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

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

□

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

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

Démonstration. Si $A_1, \ldots, A_n$ 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 $i \neq j$

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

On obtient finalement

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

□

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 $A_1, \ldots, A_n$ un système complet d'événements, alors*

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

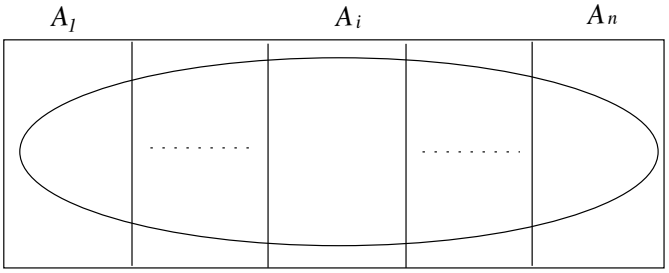


FIGURE 4.2 – Illustration du théorème des probabilités totales

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 $A_1, \ldots, A_n$ 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 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

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

La notation $n!$ se lit factorielle de n (voir Tableau 4.1).

Table 4.1 – 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 $n_1, n_2, \ldots, n_p$, le nombre de permutations avec répétition est

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

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 :

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

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

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

C'est une variable aléatoire discrète dont la distribution de probabilités est présentée en Figure 4.3.

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 :

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

et sa variance

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

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

FIGURE 4.3 – Distribution de "faces" obtenus.

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

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

et la variance vaut

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

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

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

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 :

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

La variance est donnée (sans démonstration) par

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

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,

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

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 4.4.

Exemple 4.10 Supposons que, dans une population d'électeurs, 60% des électeurs s'apprêtent 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'apprêtant 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

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

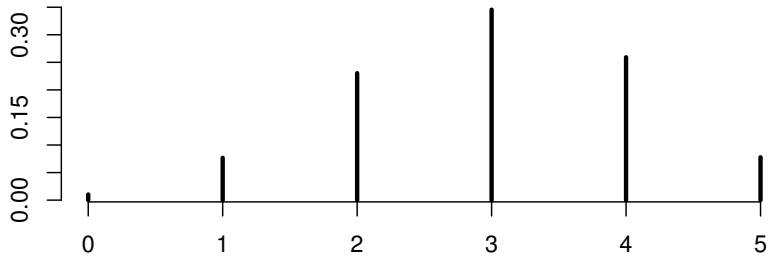


FIGURE 4.4 – Distribution d'une variable aléatoire binomiale avec $n = 5$ et $p = 0.6$.

4.4.4 Variable de Poisson

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

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

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 λ. En effet,

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

En outre, il est possible de montrer que

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

La distribution de probabilités d'une variable de Poisson $\mathcal{P}(\lambda = 1)$ est présentée dans la Figure 4.5.

En langage R

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

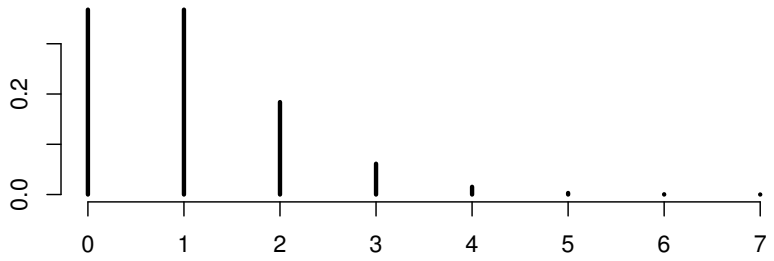


FIGURE 4.5 – Distribution d'une variable de Poisson avec $\lambda = 1$.

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,
— $\lim_{x\to\infty} F(x) = 1$,
— $\lim_{x\to-\infty} F(x) = 0$.

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 4.6, 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

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

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 :

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

et la variance

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

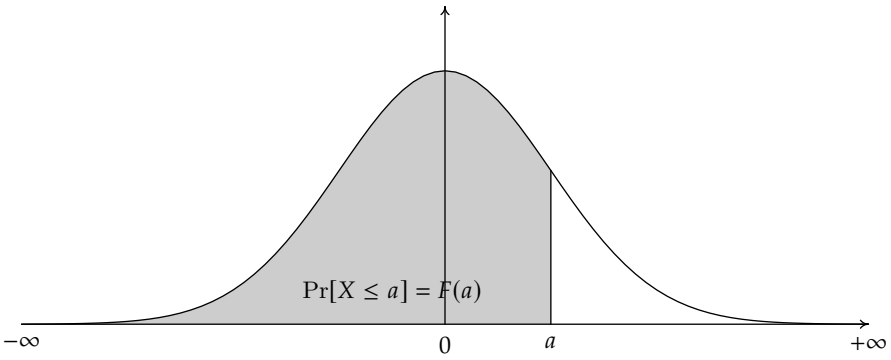


FIGURE 4.6 – Probabilité que la variable aléatoire soit inférieure à a

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.

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

Démonstration.

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

□

Propriété 4.8.

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

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,

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

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 4.7 et 4.8 représentent respectivement les fonctions de densité et de répartition d'une variable uniforme.

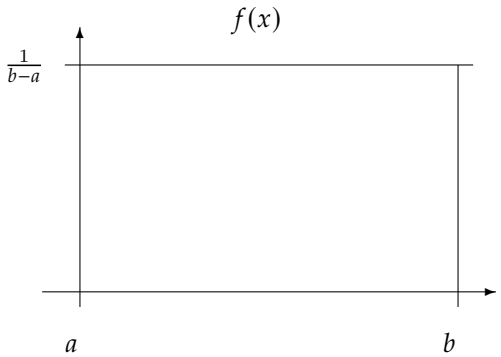


FIGURE 4.7 – Fonction de densité d'une variable uniforme

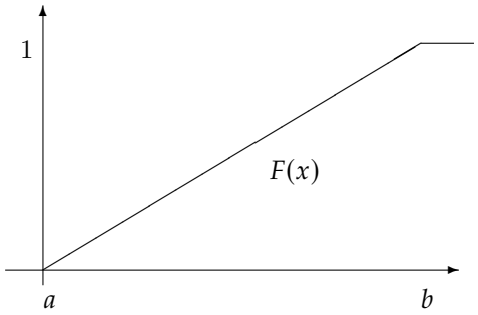


FIGURE 4.8 – Fonction de répartition d'une variable uniforme

4.5.3 Variable normale

Une variable aléatoire X est dite normale si sa densité vaut

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

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

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

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

On peut montrer (sans démonstration) que

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

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

et

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

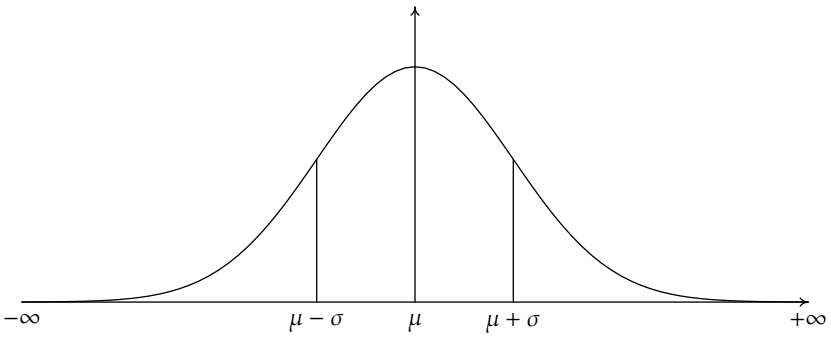


FIGURE 4.9 – Fonction de densité d'une variable normale

La fonction de répartition vaut

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

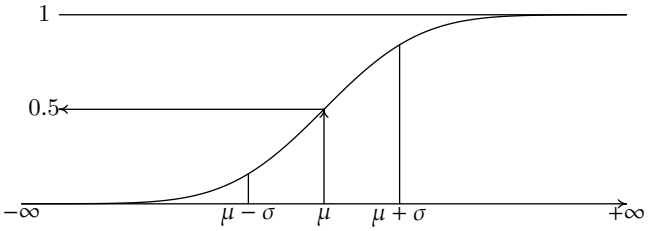


FIGURE 4.10 – Fonction de répartition d'une variable normale

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 4.11.

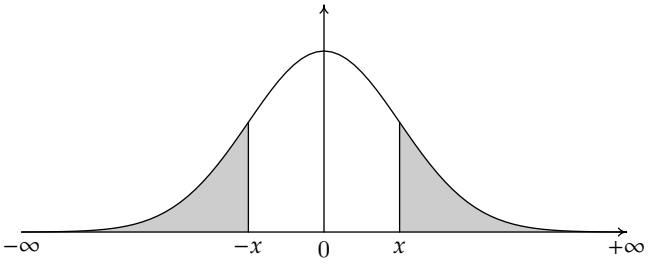


FIGURE 4.11 – Densité d'une normale centrée réduite, symétrie

De plus, le calcul de la répartition d'une variable normale de moyenne μ et de variance σ^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 λ est positif. Quand $x > 0$, sa fonction de répartition vaut :

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

On peut alors calculer la moyenne :

Propriété 4.10. $\mathrm{E}(X) = \dfrac{1}{\lambda}$.

Démonstration.

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

□

Il est également possible de montrer que la variance vaut :

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

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

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

La fonction de répartition jointe est définie par

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

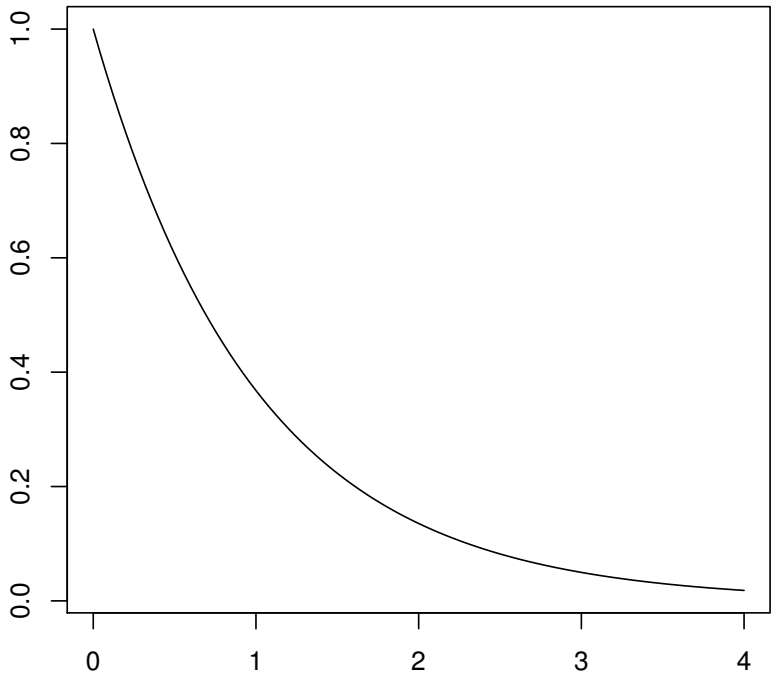


FIGURE 4.12 – Fonction de densité d'une variable exponentielle avec $\lambda = 1$.

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 :

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

$$\sigma_X^2 = \mathrm{var}(X) = \mathrm{E}\left[X - \mathrm{E}(X)\right]^2 = \int_{-\infty}^{\infty} (x-\mu_x)^2 f_X(x)dx,$$

et

$$\sigma_Y^2 = \mathrm{var}(Y) = \mathrm{E}\left[Y - \mathrm{E}(Y)\right]^2 = \int_{-\infty}^{\infty} (y-\mu_y)^2 f_Y(y)dy.$$

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 :

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

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

et

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

Enfin, la covariance entre X et Y est définie par

$$\sigma_{xy} = \mathrm{cov}(X,Y) = \mathrm{E}\left[(X - \mathrm{E}(X))(Y - \mathrm{E}(Y))\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_x)(y-\mu_y) f(x,y)dxdy.$$

Cette covariance peut également s'écrire

$$\sigma_{xy} = \text{cov}(X, Y) = \text{E}(XY) - \text{E}(X)\text{E}(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x, y)dxdy - \mu_x\mu_y.$$

En effet,

$$\begin{aligned}\text{E}\left[(X - \text{E}(X))(Y - \text{E}(Y))\right] &= \text{E}\left[XY - \text{E}(X)Y - X\text{E}(Y) - \text{E}(X)\text{E}(Y)\right] \\ &= \text{E}(XY) - \text{E}(X)\text{E}(Y) - \text{E}(X)\text{E}(Y) + \text{E}(X)\text{E}(Y) = \text{E}(XY) - \text{E}(X)\text{E}(Y).\end{aligned}$$

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 $f_X(.)$ et $f_Y(.)$ 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.

$$\text{E}(a + bX) = a + b\text{E}(X)$$

Démonstration.

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

□

Propriété 4.12.

$$\text{E}(aY + bX) = a\text{E}(Y) + b\text{E}(X)$$

Démonstration.

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

□

Propriété 4.13.

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

Démonstration. Comme la variance de X peut s'écrire $\text{var}(X) = \text{E}\left[X - \text{E}(X)\right]^2$, on a

$$\begin{aligned}\text{var}(a + bX) &= \text{E}\left[a + bX - \text{E}(a + bX)\right]^2 = \text{E}\left[a + bX - a - \text{E}(bX)\right]^2 = \text{E}\left[bX - \text{E}(bX)\right]^2 \\ &= b^2\text{E}\left[X - \text{E}(X)\right]^2 = b^2\text{var}(X).\end{aligned}$$

□

Propriété 4.14.

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

Démonstration.

$$\begin{aligned}\mathrm{var}(X+Y) &= \mathrm{E}\left[X+Y-\mathrm{E}(X+Y)\right]^2 = \mathrm{E}\left[X-\mathrm{E}(X)+Y-\mathrm{E}(Y)\right]^2 \\ &= \mathrm{E}\left[X-\mathrm{E}(X)\right]^2 + \mathrm{E}\left[Y-\mathrm{E}(Y)\right]^2 + 2\mathrm{E}\left[X-\mathrm{E}(X)\right]\left[Y-\mathrm{E}(Y)\right] \\ &= \mathrm{var}(X)+\mathrm{var}(Y)+2\mathrm{cov}(X,Y).\end{aligned}$$

□

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

Démonstration.

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

□

De manière générale, $\mathrm{cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$. Si X et Y sont indépendantes, on a $\mathrm{cov}(X,Y) = \mathrm{E}(X)\mathrm{E}(Y) - \mathrm{E}(X)\mathrm{E}(Y) = 0$, et donc

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

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 :

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

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

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

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 4.13.

La Figure 4.14 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")
```

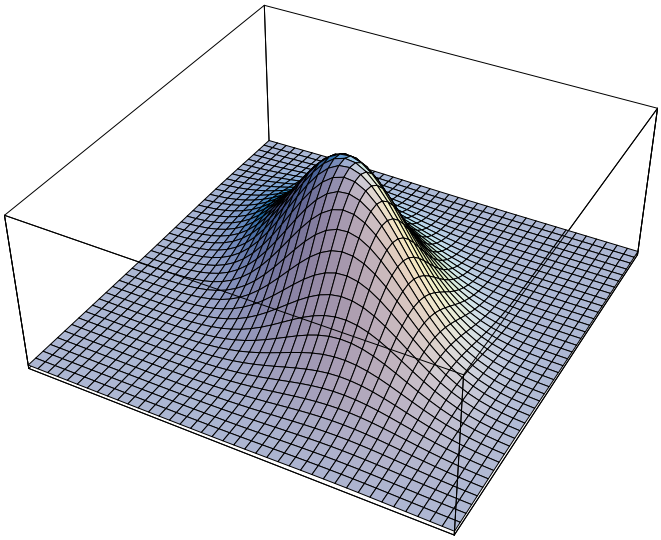


FIGURE 4.13 – Densité d'une normale bivariée

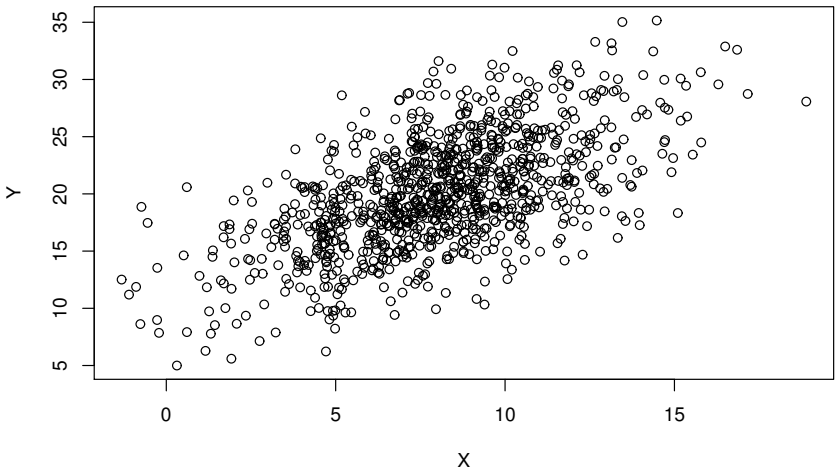


FIGURE 4.14 – Nuage de points de réalisations d'une normale bivariée

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

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

et

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

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

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

□

Le Théorème 4.3 montre que les deux distributions marginales sont normales, que μ_x et μ_y sont les moyennes marginales et que σ_x^2 et σ_x^2 sont les deux variances marginales de la distribution jointe. On peut également montrer à partir du Théorème 4.3 que le volume sous la courbe vaut bien 1. En effet,

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

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 $f(y|x)$)

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

□

Le Theorème 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 :

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

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

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

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

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

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

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

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

Théorème 4.5.

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

Démonstration. La covariance peut également s'écrire

$$\mathrm{cov}(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xyf(x,y)dydx - \mu_x\mu_y.$$

On a :

$$\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xyf(x,y)dxdy &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xyf_X(x)f(y|x)dydx = \int_{-\infty}^{\infty}xf_X(x)\int_{-\infty}^{\infty}yf(y|x)dydx \\
&= \int_{-\infty}^{\infty}xf_X(x)\left[\mu_y + \frac{\sigma_y\rho}{\sigma_x}(x-\mu_x)\right]dx = \mu_y\int_{-\infty}^{\infty}xf_X(x)dx + \frac{\sigma_y\rho}{\sigma_x}\int_{-\infty}^{\infty}xf_X(x)(x-\mu_x)dx \\
&= \mu_y\mu_x + \frac{\sigma_y\rho}{\sigma_x}\sigma_x^2 = \mu_y\mu_x + \sigma_x\sigma_y\rho.
\end{aligned}$$

Donc,

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

□

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

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

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. □

4.7 Vecteurs aléatoires

4.7.1 Généralités

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

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

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

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

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

$$\mathrm{var}(\mathrm{X}) = \mathrm{E}[(\mathrm{X} - \boldsymbol{\nu})(\mathrm{X} - \boldsymbol{\nu})^\top] = \mathrm{E}(\mathrm{X}\mathrm{X}^\top) - \boldsymbol{\nu}\boldsymbol{\nu}^\top = \begin{pmatrix} \mathrm{var}(X_1) & \cdots & \mathrm{cov}(X_1, X_j) & \cdots & \mathrm{cov}(X_1, X_p) \\ \vdots & \ddots & \vdots & & \vdots \\ \mathrm{cov}(X_1, X_j) & \cdots & \mathrm{var}(X_j) & \cdots & \mathrm{cov}(X_j, X_p) \\ \vdots & & \vdots & \ddots & \vdots \\ \mathrm{cov}(X_1, X_p) & \cdots & \mathrm{cov}(X_j, X_p) & \cdots & \mathrm{var}(X_p) \end{pmatrix}.$$

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

$$\mathrm{E}(\mathrm{AX}) = \mathrm{E}\begin{pmatrix} \sum_{j=1}^p a_{1j} X_j \\ \vdots \\ \sum_{j=1}^p a_{ij} X_j \\ \vdots \\ \sum_{j=1}^p a_{qj} X_j \end{pmatrix} = \begin{pmatrix} \mathrm{E}\left(\sum_{j=1}^p a_{1j} X_j\right) \\ \vdots \\ \mathrm{E}\left(\sum_{j=1}^p a_{ij} X_j\right) \\ \vdots \\ \mathrm{E}\left(\sum_{j=1}^p a_{qj} X_j\right) \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^p a_{1j} \mathrm{E}(X_j) \\ \vdots \\ \sum_{j=1}^p a_{ij} \mathrm{E}(X_j) \\ \vdots \\ \sum_{j=1}^p a_{qj} \mathrm{E}(X_j) \end{pmatrix} = \mathrm{AE}(\mathrm{X}) = \mathrm{A}\boldsymbol{\nu},$$

et

$$\begin{aligned} \mathrm{var}(\mathrm{AX}) &= \mathrm{E}\left\{[\mathrm{AX} - \mathrm{E}(\mathrm{AX})][\mathrm{AX} - \mathrm{E}(\mathrm{AX})]^\top\right\} = \mathrm{E}\left\{[\mathrm{AX} - \mathrm{E}(\mathrm{AX})][\mathrm{X}^\top\mathrm{A}^\top - \mathrm{E}(\mathrm{X}^\top\mathrm{A}^\top)]\right\} \\ &= \mathrm{E}\left\{[\mathrm{AX} - \mathrm{AE}(\mathrm{X})][\mathrm{X}^\top\mathrm{A}^\top - \mathrm{E}(\mathrm{X}^\top)\mathrm{A}^\top]\right\} = \mathrm{AE}\left\{(\mathrm{X} - \mathrm{E}(\mathrm{X}))(\mathrm{X}^\top - \mathrm{E}(\mathrm{X}^\top))\right\}\mathrm{A}^\top \\ &= \mathrm{A}\,\mathrm{var}(\mathrm{X})\,\mathrm{A}^\top = \mathrm{A}\boldsymbol{\Sigma}\mathrm{A}^\top. \end{aligned}$$

4.7.2 Variable normale multivariée

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

$$f_X(\mathrm{x}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathrm{x} - \boldsymbol{\nu})^\top \boldsymbol{\Sigma}^{-1}(\mathrm{x} - \boldsymbol{\nu})\right], \tag{4.3}$$

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

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 $\mathbf{\Sigma} = \text{diag}(\sigma_1^2, \ldots, \sigma_p^2)$, toutes les composantes du vecteur X sont non-corrélées. Dans ce cas,

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

où

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

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

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

et on a

$$\mathrm{AX} \sim N\left(\mathrm{A}\boldsymbol{\nu}, \mathrm{A}\mathbf{\Sigma}\mathrm{A}^\top\right).$$

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

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

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

Théorème 4.8. *Soit $X_1, \ldots, X_n$ une suite de variables aléatoires, indépendantes et identiquement distribuées et dont la moyenne μ et la variance σ^2 existent et sont finies, alors si*

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

on a

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

Démonstration.

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

et

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

□

4.8 Autres variables aléatoires

4.8.1 Variable khi-carrée

Soit une suite de variables aléatoires indépendantes, normales, centrées réduites, $X_1, \ldots, X_p$, (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

$$\mathrm{E}(\chi_p^2) = p,$$

et que

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

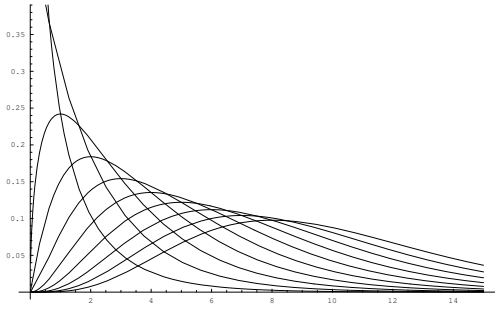


FIGURE 4.15 – Densité d'une variable de chi-carré avec $p = 1, 2, \ldots, 10$

4.8.2 Variable de Student

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

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

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

4.8.3 Variable de Fisher

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

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

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

Remarque 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.

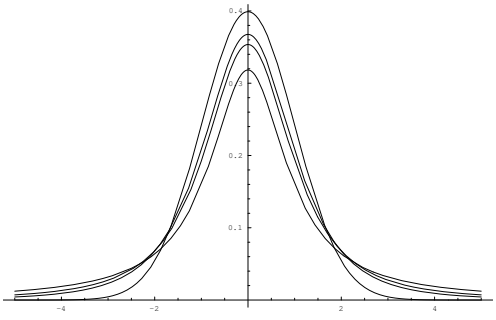


FIGURE 4.16 – Densités de variables de Student avec $p = 1, 2$ et 3 et d'une variable normale

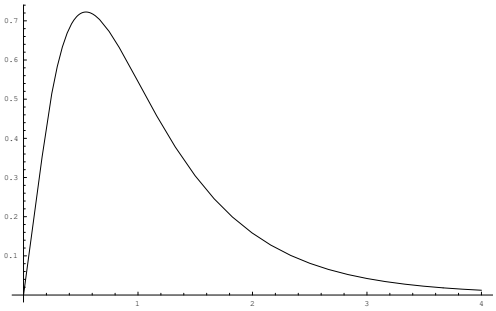


FIGURE 4.17 – Densité d'une variable de Fisher

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 $\widehat{\theta}$ d'un paramètre θ, on préfère produire un intervalle $[L^-, L^+]$ dans lequel pourrait se trouver le paramètre inconnu avec une certaine probabilité que l'on note $1 - \alpha$ (α est une probabilité petite). On relativise ainsi l'information donnée par l'estimateur $\widehat{\theta}$. Pour pouvoir construire un intervalle de confiance, il faut connaître la distribution de probabilité de $\widehat{\theta}$ (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 θ, dont une seule est vraie. Par exemple, on peut tester

— l'hypothèse nulle H_0 que $\theta = \theta_0$,

— l'hypothèse alternative H_1 que $\theta = \theta_1$.

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

— Rejeter H_0 alors que H_0 est vraie, cette erreur est appelée erreur de première espèce.

— Ne pas rejeter H_0 alors que H_0 est fausse, cette erreur est appelée erreur de deuxième espèce.

Table 4.2 – Erreur de première et seconde espèce

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

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

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

Table 4.3 – Probabilité de commettre les erreurs

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

La quantité

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

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

Tests d'hypothèses composites

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

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

Remarque 4.3 L'égalité doit toujours être dans l'hypothèse nulle. Donc, si la question est : "θ est-il strictement plus grand que θ_0 ?" on posera l'hypothèse alternative $H_1 : \theta > \theta_0$ et donc $H_0 : \theta \leq \theta_0$.

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

— On rejette H_0 si $T \notin IA$

— On ne rejette pas H_0 si $T \in IA$

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. α est le coefficient de régression appelé 'constante' (en anglais *intercept*),
3. β est le coefficient de régression appelé 'pente',
4. les ε_i sont les termes d'erreurs du modèle et sont des variables aléatoires telles que
   — $\mathrm{E}(\varepsilon_i) = 0$ pour tout $i = 1, \ldots, n$,
   — $\mathrm{E}(\varepsilon_i \varepsilon_k) = 0$ pour tout $i \neq k = 1, \ldots, n$,
   — $\mathrm{E}(\varepsilon_i^2) = \sigma_\varepsilon^2$ pour tout $i = 1, \ldots, n$.

Dans un premier temps, nous ne faisons pas d'hypothèses sur la distribution de probabilité des ε_i. Les termes d'erreurs sont simplement homoscédastiques (même variance) et non corrélés. Cependant, dans beaucoup d'applications, on supposera que les ε_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.

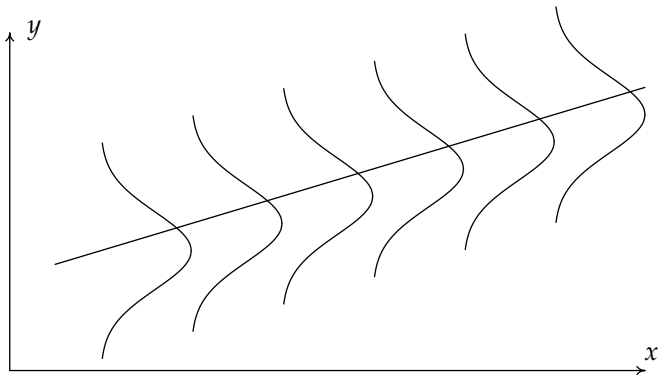


FIGURE 5.1 – Représentation graphique du modèle de régression

Le modèle possède les propriétés suivantes :

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

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

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

5.2 Estimation des paramètres par les moindres carrés

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

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

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

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

$$\widehat{\beta}=\sum_{i=1}^{n}c_i y_i \qquad et \qquad \widehat{\alpha}=\sum_{i=1}^{n}d_i y_i,$$

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

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

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

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

on a

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

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

Démonstration. Comme

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

on a

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

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

$$\mathrm{E}(\widehat{\beta}) = \frac{\beta n s_x^2}{n s_x^2} = \beta.$$

De plus, comme

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

on a

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

□

Théorème 5.3.

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

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

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

où

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

Comme les y_i sont non corrélés, on a directement :

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

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

$$\mathrm{var}(\widehat{\beta}) = \sigma_\varepsilon^2 \sum_{i=1}^{n} \frac{\frac{1}{n^2}(x_i - \bar{x})^2}{s_x^4} = \frac{\sigma_\varepsilon^2}{n}\frac{s_x^2}{s_x^4} = \frac{\sigma_\varepsilon^2}{n s_x^2},$$

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

et

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

□

5.3 Valeurs ajustées et résidus

Les valeurs ajustées se calculent simplement par :

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

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

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

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

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

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

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

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

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

donc $\mathrm{E}(y_i^*) = \mathrm{E}(y_i)$, ce qui implique directement que

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

La variance des y_i^* vaut

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

En utilisant les résultats du Théorème 5.3, on a

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

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

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

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

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

L'Égalité (5.2) nous montre également que

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

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

Théorème 5.4.

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

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

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

est un estimateur sans biais de σ_ε^2.

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

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

et

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

On peut donc construire le Tableau 5.1 qui récapitule les résultats obtenus.

Table 5.1 – Tableau récapitulatif des espérances et variances

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

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 :

$$\mathrm{SC}_{TOT} = \mathrm{SC}_{REG} + \mathrm{SC}_{RES},$$

où

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

Comme

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

et que

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

on a

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

Comme

$$\mathrm{SC}_{REG} = \sum_{i=1}^n (y_i^*-\bar{y})^2 = \sum_{i=1}^n \left[\widehat{\alpha}+\widehat{\beta}x_i - (\widehat{\alpha}+\widehat{\beta}\bar{x})\right]^2 = \widehat{\beta}^2\sum_{i=1}^n (x_i-\bar{x})^2 = \widehat{\beta}^2 n s_x^2,$$

on a

$$\mathrm{E}(\mathrm{SC}_{REG}) = n s_x^2 \mathrm{E}(\widehat{\beta}^2) = n s_x^2\left[\mathrm{var}(\widehat{\beta}) + \mathrm{E}^2(\widehat{\beta})\right] = n s_x^2\left[\frac{\sigma_\varepsilon^2}{ns_x^2} + \beta^2\right] = \sigma_\varepsilon^2 + \beta^2 n s_x^2.$$

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

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

ce qui permet construire le Tableau 5.2.

Table 5.2 – Tableau récapitulatif des espérances des sommes de carrés

	SC_{TOT}	SC_{REG}	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	$(n-1)\sigma_\varepsilon^2+\beta^2 n s_x^2$	$\sigma_\varepsilon^2+\beta^2 n s_x^2$	$(n-2)\sigma_\varepsilon^2$

5.5 Le modèle linéaire avec normalité des erreurs

Si les termes d'erreurs ε_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 $i = 1, \ldots, n$,

— les ε_i sont des variables aléatoires normales indépendantes,

ou encore :

— $y_i \sim \mathcal{N}(\alpha + \beta x_i, \sigma_\varepsilon^2)$, pour tout $i = 1, \ldots, n$,

— 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 linéaires de ε_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 $y_i, \ldots, y_n$ peut s'écrire comme le produit des densités marginales

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

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

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

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 :

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

Théorème 5.5. *Les estimateurs du maximum de vraisemblance des paramètres α, β et σ_ε^2 sont donnés par*

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

Démonstration. En annulant la dérivée de la log-vraisemblance par rapport à α, on obtient :

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

ce qui donne, en multipliant cette équation par σ_ε^2/n

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

En annulant la dérivée de la log-vraisemblance par rapport à β, on obtient :

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

ce qui donne, en multipliant cette équation par σ_ε^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 α par cette valeur dans l'Équation (5.4), on obtient :

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

ce qui donne l'estimateur du maximum de vraisemblance (MV) de β :

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

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

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

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

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

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 α et β par leurs estimateurs du maximum de vraisemblance, on a

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

□

On constate que :
— Les estimateurs du maximum de vraisemblance de α et β sont égaux aux estimateurs des moindres carrés.
— L'estimateur du maximum de vraisemblance de σ_ε^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 linaire 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 :

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

et en centrant et réduisant ces deux variables aléatoires

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

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

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

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

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

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

$$\frac{\dfrac{\sqrt{n} s_x (\widehat{\beta} - \beta)}{\sigma_\varepsilon}}{\sqrt{\dfrac{\dfrac{(n-2)\widehat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2}}{(n-2)}}} \sim t_{n-2} \qquad \text{et} \qquad \frac{\dfrac{\sqrt{n} s_x (\widehat{\alpha} - \alpha)}{\sigma_\varepsilon \sqrt{s_x^2 + \bar{x}^2}}}{\sqrt{\dfrac{\dfrac{(n-2)\widehat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2}}{(n-2)}}} \sim t_{n-2},$$

ce qui s'écrit plus simplement par

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

ou encore

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

Ces deux résultats fondamentaux vont permettre de construire des tests d'hypothèses sur les paramètres α et β.

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 α et β :

$$\mathrm{IC}(0.95) = \left[\widehat{\alpha} - t_{n-2,0.975}\sqrt{\widehat{\mathrm{var}}(\widehat{\alpha})}, \widehat{\alpha} + t_{n-2,0.975}\sqrt{\widehat{\mathrm{var}}(\widehat{\alpha})}\right],$$

$$\mathrm{IC}(0.95) = \left[\widehat{\beta} - t_{n-2,0.975}\sqrt{\widehat{\mathrm{var}}(\widehat{\beta})}, \widehat{\beta} + t_{n-2,0.975}\sqrt{\widehat{\mathrm{var}}(\widehat{\beta})}\right].$$

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 α_0 et β_0.

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

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

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

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

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 :

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

Si cette valeur p est plus petite que l'erreur de première espèce α 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 5.3 – Tableau d'analyse de la variance

Source de variation	Sommes des carrés	Degrés de liberté	Carrés moyens	F_{obs}	p
Régression	SC_{REG}	1	$CM_{reg} = \frac{SC_{REG}}{1}$	$F_{obs} = CM_{reg}/CM_{res}$	$\Pr(F_{1,n-2} > F_{obs} \;\; \|H_0 \text{ vraie})$
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_{REG}. En effet, cette somme de carrés peut s'écrire $SC_{REG} = \widehat{\beta}^2 n s_x^2$ et ne dépend donc que d'une composante aléatoire $\widehat{\beta}$.

— $n-2$ 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 :

$$CM_{tot} = \frac{SC_{TOT}}{n-1}, \; CM_{reg} = \frac{SC_{REG}}{1}, \text{ et } CM_{res} = \frac{SC_{RES}}{n-2}.$$

Le Tableau récapitulatif 5.2 nous permet directement de calculer les espérances des carrés moyens

$$E(CM_{tot}) = \sigma_\varepsilon^2 + \beta^2 s_x^2 \frac{n}{n-1}, \; E(CM_{reg}) = \sigma_\varepsilon^2 + \beta^2 n s_x^2, \text{ et } E(CM_{res}) = \sigma_\varepsilon^2.$$

On constate que si $\beta = 0$, alors les trois carrés moyens sont des estimateurs sans biais de σ_ε^2. Si β 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 :

$$F_{obs} = \frac{CM_{reg}}{CM_{res}}.$$

Il est possible de montrer que, sous H_0, F_{obs} 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

$$p = \Pr(F_{1,n-2} > F_{obs} \;|H_0 \text{ vraie}),$$

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 α (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.

En langage R

```
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 défini 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.

En langage R

```
> 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*) $\widehat{\alpha} = -126.614$ et la pente $\widehat{\beta} = 1.180$. 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 5.4 des coefficients nous donne les écart-types des coefficients, les statistiques t et la p-valeur.

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,

TABLE 5.4 – Tableau des coefficients de régression

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

le logiciel R affiche l'écart-type estimé des résidus $\widehat{\sigma}_\varepsilon$, 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.

TABLE 5.5 – Tableau d'analyse de la variance

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

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

Chapitre 6

Le modèle linéaire général

6.1 Définition du modèle linéaire général

Le modèle linéaire est une généralisation de la régression au cas où p variables explicatives $x_1, x_2, \ldots, x_p$ sont utilisées pour expliquer la variable dépendante y. La relation la plus simple est une relation linéaire, entre les variables explicatives et la variable dépendante. Le modèle linéaire général s'écrit :

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

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

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

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.

$$\mathrm{y} = \mathrm{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

où

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

Seuls y et X sont observés.

Les hypothèses du modèle linéaire général peuvent être reformulées :

- La matrice X n'est pas aléatoire,

— La matrice X est supposée de plein rang (Dans le cas contraire, on dit qu'il y a multicolinéarité, c'est-à-dire qu'au moins une des colonnes de la matrice peut s'exprimer comme une combinaison linéaire des autres colonnes),
— $\mathrm{E}(\boldsymbol{\varepsilon}) = 0$,
— $\mathrm{var}(\varepsilon_i) = \sigma_\varepsilon^2$ (homoscédasticité).
— $\mathrm{cov}(\varepsilon_i, \varepsilon_j) = 0$ (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 y un vecteur aléatoire de $\mathbb{R}^n$ tel que
— $\mathrm{E}(\mathrm{y}) = \mathrm{X}\boldsymbol{\beta}$, où X est une matrice $n \times p$ et $\boldsymbol{\beta} \in \mathbb{R}^p$,
— $\mathrm{var}(\mathrm{y}) = \mathrm{I}\sigma_\varepsilon^2$, où I est une matrice identité $n \times n$ et σ_ε^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 σ_ε^2. La méthode des moindres carrés consiste à minimiser en $\boldsymbol{\beta}$, l'expression

$$\boldsymbol{\varepsilon}^\top \boldsymbol{\varepsilon} = (\mathrm{y} - \mathrm{X}\boldsymbol{\beta})^\top (\mathrm{y} - \mathrm{X}\boldsymbol{\beta}).$$

La solution (voir Section 3.2) fournit l'estimateur des moindres carrés (ordinaires) $\widehat{\boldsymbol{\beta}}$ de $\boldsymbol{\beta}$, qui se note

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

L'estimateur $\widehat{\boldsymbol{\beta}}$ est une variable aléatoire, car il dépend de 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* $\widehat{\boldsymbol{\beta}} = (\mathrm{X}^\top \mathrm{X})^{-1} \mathrm{X}^\top \mathrm{y}$ *est sans biais.*

Démonstration. Comme

$$\widehat{\boldsymbol{\beta}} = (\mathrm{X}^\top \mathrm{X})^{-1} \mathrm{X}^\top \mathrm{y} = (\mathrm{X}^\top \mathrm{X})^{-1} \mathrm{X}^\top (\mathrm{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = (\mathrm{X}^\top \mathrm{X})^{-1} \mathrm{X}^\top \mathrm{X}\boldsymbol{\beta} + (\mathrm{X}^\top \mathrm{X})^{-1} \mathrm{X}^\top \boldsymbol{\varepsilon} = \boldsymbol{\beta} + (\mathrm{X}^\top \mathrm{X})^{-1} \mathrm{X}^\top \boldsymbol{\varepsilon}.$$

On a

$$\mathrm{E}(\widehat{\boldsymbol{\beta}}) = \mathrm{E}\left\{\boldsymbol{\beta} + (\mathrm{X}^\top \mathrm{X})^{-1} \mathrm{X}^\top \boldsymbol{\varepsilon}\right\} = \boldsymbol{\beta} + (\mathrm{X}^\top \mathrm{X})^{-1} \mathrm{X}^\top \mathrm{E}(\boldsymbol{\varepsilon}) = \boldsymbol{\beta}.$$

□

Théorème 6.2. $\mathrm{var}(\widehat{\boldsymbol{\beta}}) = \sigma_\varepsilon^2 (\mathrm{X}^\top \mathrm{X})^{-1}$.

Démonstration. Comme

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

on a

$$\begin{aligned} \mathrm{var}(\widehat{\boldsymbol{\beta}}) &= \mathrm{var}\left\{(\mathrm{X}^\top \mathrm{X})^{-1} \mathrm{X}^\top \boldsymbol{\varepsilon}\right\} = (\mathrm{X}^\top \mathrm{X})^{-1} \mathrm{X}^\top \mathrm{var}\left\{\boldsymbol{\varepsilon}\right\} \mathrm{X}(\mathrm{X}^\top \mathrm{X})^{-1} = (\mathrm{X}^\top \mathrm{X})^{-1} \mathrm{X}^\top \mathrm{I}\sigma_\varepsilon^2 \mathrm{X}(\mathrm{X}^\top \mathrm{X})^{-1} \\ &= \sigma_\varepsilon^2 \underbrace{(\mathrm{X}^\top \mathrm{X})^{-1} \mathrm{X}^\top \mathrm{X}}_{\mathrm{I}} (\mathrm{X}^\top \mathrm{X})^{-1} = \sigma_\varepsilon^2 (\mathrm{X}^\top \mathrm{X})^{-1}. \end{aligned}$$

□

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

$$\sigma(\widehat{\beta}_j) = \sqrt{\mathrm{var}(\widehat{\beta}_j)} = \sigma_\varepsilon \sqrt{[(\mathrm{X}^\top \mathrm{X})^{-1}]_{jj}},$$

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

Théorème 6.3. *(de Gauss-Markov) L'estimateur* $\widehat{\boldsymbol{\beta}} = (\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top\mathrm{y}$ *est le meilleur (au sens de la plus petite variance) estimateur linéaire en* y *sans biais de* $\boldsymbol{\beta}$.

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

$$\mathrm{E}(\boldsymbol{\beta}^*) = \mathrm{E}\left\{(\mathrm{B} + (\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top)(\mathrm{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})\right\} = (\mathrm{B} + (\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top)\mathrm{X}\boldsymbol{\beta} = \mathrm{BX}\boldsymbol{\beta} + \boldsymbol{\beta},$$

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

$$\mathrm{BX}\boldsymbol{\beta} + \boldsymbol{\beta} = \boldsymbol{\beta},$$

c'est-à-dire que

$$\mathrm{BX}\boldsymbol{\beta} = 0,$$

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

$$\mathrm{BX} = 0. \tag{6.1}$$

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

$$\begin{aligned}
\mathrm{var}(\boldsymbol{\beta}^*) &= (\mathrm{B} + (\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top)\mathrm{var}(\mathrm{y})(\mathrm{B} + (\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top)^\top \\
&= (\mathrm{B} + (\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top)\mathrm{I}\sigma_\varepsilon^2(\mathrm{B} + (\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top)^\top \\
&= \left\{\mathrm{BB}^\top + \underbrace{\mathrm{BX}(\mathrm{X}^\top\mathrm{X})^{-1}}_{0} + \underbrace{(\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top\mathrm{B}^\top}_{0} + (\mathrm{X}^\top\mathrm{X})^{-1}\right\}\sigma_\varepsilon^2.
\end{aligned}$$

Par (6.1), on a finalement

$$\mathrm{var}(\boldsymbol{\beta}^*) = \left\{\mathrm{BB}^\top + (\mathrm{X}^\top\mathrm{X})^{-1}\right\}\sigma_\varepsilon^2. \tag{6.2}$$

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

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

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

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

Lemme 6.1. *Soit un vecteur* u *composé de* n *variables aléatoires d'espérances nulles et tel que* $\mathrm{var}(\mathrm{u}) = \sigma_u^2\mathrm{I}$, *et* A *une matrice symétrique non-aléatoire, alors*

$$\mathrm{E}(\mathrm{u}^\top\mathrm{Au}) = \sigma_u^2\mathrm{trace}(\mathrm{A}).$$

Démonstration.

$$\mathrm{E}(\mathrm{u}^\top\mathrm{Au}) = \sum_{i=1}^n a_{ii}\underbrace{\mathrm{E}(u_i^2)}_{\sigma_u^2} + \sum_{i=1}^n\sum_{\substack{j=1\\ j\neq i}}^n a_{ij}\underbrace{\mathrm{E}(u_iu_j)}_{0}.$$

Or $\mathrm{E}(u_iu_j) = 0$, quand $j \neq i$. Donc,

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

□

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

Théorème 6.4. *Soit* $\mathrm{e} = \mathrm{y} - \mathrm{X}\widehat{\boldsymbol{\beta}}$, *alors*

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

Démonstration. Nous avons vu en Section 3.1 que e peut également s'écrire

$$\mathrm{e} = (\mathrm{I} - \mathrm{P}_X)\,\mathrm{y}, \tag{6.3}$$

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

$$\mathrm{P}_X = \mathrm{X}(\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top.$$

Donc,

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

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

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

On obtient

$$\mathrm{e}^\top\mathrm{e} = \boldsymbol{\varepsilon}^\top(\mathrm{I} - \mathrm{P}_X)^\top(\mathrm{I} - \mathrm{P}_X)\boldsymbol{\varepsilon},$$

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

$$\mathrm{e}^\top\mathrm{e} = \boldsymbol{\varepsilon}^\top(\mathrm{I} - \mathrm{P}_X)\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^\top\mathrm{I}\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^\top\mathrm{P}_X\boldsymbol{\varepsilon}.$$

Par le Lemme 6.1, on obtient

$$\mathrm{E}(\mathrm{e}^\top\mathrm{e}) = \sigma_\varepsilon^2\mathrm{trace}(\mathrm{I}) - \sigma_\varepsilon^2\mathrm{trace}(\mathrm{P}_X).$$

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

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

□

Le Théorème 6.4 nous permet de construire un estimateur sans biais pour σ_ε^2 qui est :

$$\widehat{\sigma}_\varepsilon^2 = \frac{\mathrm{e}^\top\mathrm{e}}{n-p}.$$

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

Table 6.1 – Tableau récapitulatif

Paramètre	Estimateur	Variance	Variance estimée
$\boldsymbol{\beta}$	$\widehat{\boldsymbol{\beta}} = (\mathrm{X}^\top\mathrm{X})^{-1}\,\mathrm{X}^\top\mathrm{y}$	$(\mathrm{X}^\top\mathrm{X})^{-1}\,\sigma_\varepsilon^2$	$(\mathrm{X}^\top\mathrm{X})^{-1}\,\widehat{\sigma}_\varepsilon^2$
σ_ε^2	$\widehat{\sigma}_\varepsilon^2 = \dfrac{(\mathrm{y} - \mathrm{X}\widehat{\boldsymbol{\beta}})^\top(\mathrm{y} - \mathrm{X}\widehat{\boldsymbol{\beta}})}{n-p}$	–	–

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 ε_i sont des variables aléatoires indépendantes ayant des distributions normales de moyennes nulles et de variance σ_ε^2.

On peut donc écrire que le vecteur $\boldsymbol{\varepsilon}$ a une distribution multinormale :

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

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

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

et donc

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

De (4.3), on a

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

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

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

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

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 :

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

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

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

et

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

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

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

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

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

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 $\widehat{\theta}$ d'un paramètre θ est sans biais, si $\mathrm{E}(\widehat{\theta}) = \theta$ pour toute valeur de θ.

— Un estimateur est efficace ou de variance minimum si sa variance est plus petite ou égale que celles de tous les estimateurs du paramètre.

— Un estimateur $\widehat{\theta}$ est convergent, s'il converge en probabilité vers le paramètre à estimer, c'est-à-dire

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

où ε est une quantité arbitrairement petite.

— Une statistique est exhaustive si elle épuise toute l'information relative au paramètre.

La méthode du maximum de vraisemblance fournit des estimateurs ayant les propriétés suivantes :

— S'il existe une statistique exhaustive, alors l'estimateur du maximum de vraisemblance en dépend.

— Si $\widehat{\theta}$ est un estimateur du maximum de vraisemblance de θ alors $f(\widehat{\theta})$ 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 que asymptotiquement sans biais. L'estimateur du maximum de vraisemblance de σ_ε^2 est en effet légèrement biaisé.

6.3.4 Distribution de probabilité des estimateurs

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

$$\widehat{\boldsymbol{\beta}} = \left(\mathrm{X}^\top\mathrm{X}\right)^{-1}\mathrm{X}^\top\mathrm{y} = \left(\mathrm{X}^\top\mathrm{X}\right)^{-1}\mathrm{X}^\top(\mathrm{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \boldsymbol{\beta} + \left(\mathrm{X}^\top\mathrm{X}\right)^{-1}\mathrm{X}^\top\boldsymbol{\varepsilon},$$

Donc, $\widehat{\boldsymbol{\beta}}$ 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,

$$\widehat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{\beta}, \left(\mathrm{X}^\top\mathrm{X}\right)^{-1}\sigma_\varepsilon^2). \tag{6.4}$$

Lemme 6.2. *Soient* u *un vecteur aléatoire de distribution normale de* $\mathbb{R}^n$*, de moyennes nulles et de variance* I*, et* $\mathbf{\Gamma}$ *une matrice orthogonale de dimension* $n \times n$*, alors*

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

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

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

Théorème 6.5. *Soit un vecteur aléatoire* u *de distribution normale, de moyennes nulles et de variance* I*. Si* P *est symétrique, idempotente et de rang* p*, alors* $\mathrm{u}^\top \mathrm{P}\mathrm{u}$ *est une variable* χ^2_p *à* p *degrés de liberté.*

Démonstration. La matrice 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 λ_i de P sur sa diagonale et $\mathbf{\Gamma}$ est une matrice orthogonale contenant les n vecteurs propres de P, alors on peut écrire :

$$\mathrm{P} = \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}^\top.$$

La forme quadratique peut s'écrire

$$\mathrm{u}^\top \mathrm{P}\mathrm{u} = \mathrm{u}^\top \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}^\top \mathrm{u} = \mathrm{v}^\top \mathbf{\Lambda}\mathrm{v},$$

où $\mathrm{v} = \mathbf{\Gamma}^\top \mathrm{u}$. En vertu du Lemme 6.2, $\mathrm{v} \sim \mathcal{N}(0, \mathrm{I})$. En vertu du Théorème 2.4, comme P est idempotente et de rang p, P a p valeurs propres égales à 1 et $n - p$ valeurs propres égales à 0. La forme quadratique

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

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

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

$$(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \frac{\mathrm{X}^\top \mathrm{X}}{\sigma_\varepsilon^2} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim \chi^2_p.$$

En effet,

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left(\mathrm{X}^\top \mathrm{X}\right)^{-1} \mathrm{X}^\top \mathrm{y} - \boldsymbol{\beta} = \left(\mathrm{X}^\top \mathrm{X}\right)^{-1} \mathrm{X}^\top \left(\mathrm{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\right) - \boldsymbol{\beta} = \boldsymbol{\beta} + \left(\mathrm{X}^\top \mathrm{X}\right)^{-1} \mathrm{X}^\top \boldsymbol{\varepsilon} - \boldsymbol{\beta} = \left(\mathrm{X}^\top \mathrm{X}\right)^{-1} \mathrm{X}^\top \boldsymbol{\varepsilon}.$$

Donc,

$$(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \frac{\mathrm{X}^\top \mathrm{X}}{\sigma_\varepsilon^2} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \boldsymbol{\varepsilon}^\top \mathrm{X} \left(\mathrm{X}^\top \mathrm{X}\right)^{-1} \frac{\mathrm{X}^\top \mathrm{X}}{\sigma_\varepsilon^2} \left(\mathrm{X}^\top \mathrm{X}\right)^{-1} \mathrm{X}^\top \boldsymbol{\varepsilon} = \frac{\boldsymbol{\varepsilon}^\top}{\sigma_\varepsilon} \mathrm{X} \left(\mathrm{X}^\top \mathrm{X}\right)^{-1} \mathrm{X}^\top \frac{\boldsymbol{\varepsilon}}{\sigma_\varepsilon}.$$

Comme la matrice $\mathrm{X}\left(\mathrm{X}^\top \mathrm{X}\right)^{-1} \mathrm{X}^\top$ est symétrique idempotente et de rang p et que $\boldsymbol{\varepsilon}^\top / \sigma_\varepsilon$ est un vecteur multinormal non-corrélé, le corollaire s'obtient directement par le Théorème 6.5. □

Corrolaire 6.2. *Dans le modèle linéaire général avec des termes d'erreur normaux,*

$$\frac{\mathrm{e}^\top \mathrm{e}}{\sigma_\varepsilon^2} \sim \chi^2_{n-p}.$$

En effet,

$$\mathrm{e} = \mathrm{y} - \mathrm{X}\widehat{\boldsymbol{\beta}} = \mathrm{y} - \mathrm{X}\left(\mathrm{X}^\top \mathrm{X}\right)^{-1} \mathrm{X}^\top \mathrm{y} = \mathrm{P}_X^\perp \boldsymbol{\varepsilon},$$

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

$$\frac{\mathrm{e}^\top \mathrm{e}}{\sigma_\varepsilon^2} = \frac{\boldsymbol{\varepsilon}^\top}{\sigma_\varepsilon} \mathrm{P}_X^{\perp\top} \mathrm{P}_X^\perp \frac{\boldsymbol{\varepsilon}}{\sigma_\varepsilon} = \frac{\boldsymbol{\varepsilon}^\top}{\sigma_\varepsilon} \mathrm{P}_X^\perp \frac{\boldsymbol{\varepsilon}}{\sigma_\varepsilon} \sim \chi^2_{n-p}.$$

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

Théorème 6.6. *Soient les matrices* B *(*$p \times n$*) et* A *(*$n \times n$*) et un vecteur aléatoire* $\mathrm{u} \sim \mathcal{N}(\boldsymbol{\nu}, \sigma_u^2 \mathrm{I})$*, alors les* p *formes linéaires* $\mathrm{B}\mathrm{u}$ *sont indépendantes de la forme quadratique* $\mathrm{u}^\top \mathrm{A}\mathrm{u}$ *si* $\mathrm{BA} = 0$.

Corrolaire 6.3. *Dans le modèle linéaire avec des termes d'erreur normaux,*

1. $\widehat{\boldsymbol{\beta}}$ *est indépendant de* $\mathrm{e}^\top \mathrm{e}$
2. $\widehat{\boldsymbol{\beta}}$ *est indépendant de* $\widehat{\sigma}_\varepsilon^2 = \frac{\mathrm{e}^\top \mathrm{e}}{n-p}$

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

Théorème 6.7. *Soient deux matrices symétriques* C *(*$n \times n$*) et* A *(*$n \times n$*) et un vecteur aléatoire* $\mathrm{u} \sim \mathcal{N}(\mu, \sigma_u^2 \mathrm{I})$*, alors les deux formes quadratiques* $\mathrm{u}^\top \mathrm{C}\mathrm{u}$ *et* $\mathrm{u}^\top \mathrm{A}\mathrm{u}$ *sont indépendantes si* $\mathrm{CA} = 0$.

6.3.5 Synthèse des résultats

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

— $\widehat{\boldsymbol{\beta}}$ et $\widehat{\sigma}^2_\varepsilon$ sont convergents, exhaustifs, efficaces et sans biais,

— $\widehat{\boldsymbol{\beta}}$ et $\widehat{\sigma}^2_\varepsilon$ sont indépendants,

— $\widehat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{\beta}, (X^\top X)^{-1}\,\sigma^2_\varepsilon)$,

— $\dfrac{(n-p)\widehat{\sigma}^2_\varepsilon}{\sigma^2_\varepsilon} = \dfrac{e^\top e}{\sigma^2_\varepsilon} \sim \chi^2_{n-p}$,

— $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \dfrac{X^\top X}{\sigma^2_\varepsilon}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim \chi^2_p$.

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

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

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

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

On a également que

$$\frac{(n-p)\widehat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} = \frac{\mathrm{e}^\top \mathrm{e}}{\sigma_\varepsilon^2} \sim \chi^2_{n-p}.$$

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

La quantité

$$\frac{\widehat{\beta}_j - \beta_j}{\sqrt{[(\mathrm{X}^\top \mathrm{X})^{-1}]_{jj}\sigma_\varepsilon^2}} \Bigg/ \sqrt{\frac{(n-p)\widehat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2}/(n-p)}$$

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

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

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

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

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

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

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

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

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, $\widehat{\beta}_j \sim \mathcal{N}(\beta_{j0}, \sigma^2(\widehat{\beta}_j))$, où $\sigma(\widehat{\beta}_j)$ est l'erreur-type de $\widehat{\beta}_j$:

$$\sigma^2(\widehat{\beta}_j) = \left[\left(\mathrm{X}^\top \mathrm{X} \right)^{-1} \sigma_\varepsilon^2 \right]_{jj}$$

est simplement la composante correspondante à la $j$ième ligne et la $j$ième colonne de $\mathrm{var}(\widehat{\boldsymbol{\beta}}) = \left(\mathrm{X}^\top \mathrm{X} \right)^{-1} \sigma_\varepsilon^2$. On peut donc estimer simplement $\sigma^2(\widehat{\beta}_j)$ par

$$\widehat{\sigma}^2(\widehat{\beta}_j) = \left[\left(\mathrm{X}^\top \mathrm{X} \right)^{-1} \widehat{\sigma}_\varepsilon^2 \right]_{jj}.$$

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

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

on obtient

$$\frac{(n-p)\widehat{\sigma}^2(\widehat{\beta}_j)}{\sigma^2(\widehat{\beta}_j)} = \frac{(n-p)\left[\left(\mathrm{X}^\top \mathrm{X} \right)^{-1} \widehat{\sigma}_\varepsilon^2 \right]_{jj}}{\left[\left(\mathrm{X}^\top \mathrm{X} \right)^{-1} \sigma_\varepsilon^2 \right]_{jj}} = \frac{(n-p)\widehat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} \sim \chi^2_{n-p}.$$

De plus,

$$\frac{\widehat{\beta}_j - \beta_{j0}}{\sigma_\varepsilon \sqrt{\left[\left(\mathrm{X}^\top \mathrm{X} \right)^{-1} \right]_{jj}}} = \frac{\widehat{\beta}_j - \beta_{j0}}{\sigma(\widehat{\beta}_j)} \sim \mathcal{N}(0,1)$$

Sous H_0, la statistique

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

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 $y_1, \ldots, y_i, \ldots, y_n$ 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

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

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

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

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

$$\widehat{\mu} = \left(1^\top 1 \right)^{-1} 1^\top \mathrm{y} = \frac{1}{n} \sum_{i=1}^{n} y_i = \bar{y},$$

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

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

$$\mathrm{var}(\widehat{\mu}) = \left(1^\top 1\right)^{-1}\sigma^2 = \frac{\sigma^2}{n},$$

$$\widehat{\mathrm{var}}(\widehat{\mu}) = \left(1^\top 1\right)^{-1}\widehat{\sigma}^2 = \frac{\widehat{\sigma}^2}{n}.$$

Par le Corollaire 6.3, $\widehat{\mu}$ et $\widehat{\sigma}^2$ sont indépendants. De plus, on a, par l'Expression (6.4),

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

Donc,

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

En outre, on peut écrire

$$K = \frac{(n-1)\widehat{\sigma}^2}{\sigma^2} = \frac{\boldsymbol{\varepsilon}^\top}{\sigma}\mathrm{P}_c\frac{\boldsymbol{\varepsilon}}{\sigma},$$

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

$$\mathrm{P}_c = \mathrm{I} - \frac{11^\top}{n} = \begin{pmatrix} 1-1/n & -1/n & -1/n & \dots & -1/n \\ -1/n & 1-1/n & -1/n & \dots & -1/n \\ -1/n & -1/n & 1-1/n & \dots & -1/n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1/n & -1/n & -1/n & \dots & 1-1/n \end{pmatrix}. \tag{7.1}$$

Les variables aléatoires d et K sont indépendantes. De plus, par le Théorème 6.5, $K \sim \chi^2_{n-1}$. Donc,

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

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 :

$$H_0 \ : \mathrm{R}\boldsymbol{\beta} = \mathrm{r}, \tag{7.2}$$

contre l'hypothèse alternative

$$H_1 \ : \mathrm{R}\boldsymbol{\beta} \neq \mathrm{r}, \tag{7.3}$$

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

Exemple 7.1

— Le test $H_0 \ : \beta_j = c$ s'obtient en prenant $\mathrm{R} = (0 \cdots 0 \underbrace{1}_{j\text{ième}} 0 \cdots 0)$ et $\mathrm{r} = c$.

— Le test $H_0 \ : \beta_j = 0$ pour tout j s'obtient en prenant $\mathrm{R} = \mathrm{I}_p$ (matrice identité de dimension p) et r est un vecteur de 0 de dimension p.

Sous l'hypothèse H_0,

$$\begin{aligned} \mathrm{R}\widehat{\boldsymbol{\beta}} - \mathrm{r} &= \mathrm{R}\left(\mathrm{X}^\top\mathrm{X}\right)^{-1}\mathrm{X}^\top\mathrm{y} - \mathrm{r} \\ &= \mathrm{R}\left(\mathrm{X}^\top\mathrm{X}\right)^{-1}\mathrm{X}^\top(\mathrm{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) - \mathrm{r} \\ &= \mathrm{R}\boldsymbol{\beta} + \mathrm{R}\left(\mathrm{X}^\top\mathrm{X}\right)^{-1}\mathrm{X}^\top\boldsymbol{\varepsilon} - \mathrm{r} \\ &= \mathrm{R}\left(\mathrm{X}^\top\mathrm{X}\right)^{-1}\mathrm{X}^\top\boldsymbol{\varepsilon}. \end{aligned}$$

De plus,

$$\mathrm{var}(\mathrm{R}\widehat{\boldsymbol{\beta}} - \mathrm{r}) = \mathrm{var}(\mathrm{R}\widehat{\boldsymbol{\beta}}) = \mathrm{R}\mathrm{var}(\widehat{\boldsymbol{\beta}})\mathrm{R}^\top = \sigma_\varepsilon^2\mathrm{R}\left(\mathrm{X}^\top\mathrm{X}\right)^{-1}\mathrm{R}^\top.$$

Examinons maintenant la forme quadratique :

$$(\mathrm{R}\widehat{\boldsymbol{\beta}} - \mathrm{r})^\top\mathrm{var}(\mathrm{R}\widehat{\boldsymbol{\beta}})^{-1}(\mathrm{R}\widehat{\boldsymbol{\beta}} - \mathrm{r}) = \frac{1}{\sigma_\varepsilon^2}\boldsymbol{\varepsilon}^\top\mathrm{W}\boldsymbol{\varepsilon}, \tag{7.4}$$

où

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

On vérifie facilement que W est une matrice idempotente, symétrique de rang q. Par le Théorème 6.5, on obtient donc que

$$\frac{1}{\sigma_\varepsilon^2}\boldsymbol{\varepsilon}^\top\mathrm{W}\boldsymbol{\varepsilon} \sim \chi^2_q,$$

et donc

$$(\mathrm{R}\widehat{\boldsymbol{\beta}} - \mathrm{r})^\top\mathrm{var}(\mathrm{R}\widehat{\boldsymbol{\beta}})^{-1}(\mathrm{R}\widehat{\boldsymbol{\beta}} - \mathrm{r}) = \frac{1}{\sigma_\varepsilon^2}(\mathrm{R}\widehat{\boldsymbol{\beta}} - \mathrm{r})^\top\left\{\mathrm{R}\left(\mathrm{X}^\top\mathrm{X}\right)^{-1}\mathrm{R}^\top\right\}^{-1}(\mathrm{R}\widehat{\boldsymbol{\beta}} - \mathrm{r}) \sim \chi^2_q. \tag{7.5}$$

Si la forme quadratique (7.4) est grande, on soupçonne H_0 d'être faux. Cependant, on ne peut réaliser directement un test χ^2 car l'Expression (7.5) depend de σ_ε^2 qui est inconnu. On sait par ailleurs que

$$\frac{1}{\sigma_\varepsilon^2}\mathrm{e}^\top\mathrm{e} \sim \chi^2_{n-p}.$$

De plus, comme

$$\mathrm{e}^\top\mathrm{e} = \boldsymbol{\varepsilon}^\top(\mathrm{I} - \mathrm{P}_X)\boldsymbol{\varepsilon},$$

et que $(\mathrm{I} - \mathrm{P}_X)\mathrm{W} = 0$, par le Théorème 6.7, on a l'indépendance de $\mathrm{e}^\top\mathrm{e}/\sigma_\varepsilon^2$ et de $\boldsymbol{\varepsilon}^\top\mathrm{W}\boldsymbol{\varepsilon}$.

On peut construire une statistique de test

$$F_{obs} = \frac{(\mathrm{R}\widehat{\boldsymbol{\beta}} - \mathrm{r})^\top\left\{\mathrm{R}\left(\mathrm{X}^\top\mathrm{X}\right)^{-1}\mathrm{R}^\top\right\}^{-1}(\mathrm{R}\widehat{\boldsymbol{\beta}} - \mathrm{r})\frac{1}{q}}{\mathrm{e}^\top\mathrm{e}\dfrac{1}{n-p}}. \tag{7.6}$$

Sous H_0, le numérateur et le dénominateur de F_{obs} sont indépendants et ont, à une constante près, une distribution χ^2. La statistique de test F_{obs} a donc une distribution de Fisher à q et $n-p$ degrés de liberté. Donc, en notant α l'erreur de première espèce, on rejette l'hypothèse 7.2, si

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

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 X est composée de uns, c'est-à-dire que $x_{i1} = 1$ pour tout $i = 1, \ldots, n$. La matrice R est de dimension $(p-1) \times p$ et vaut :

$$\mathrm{R} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

Alors

$$\mathrm{R}\boldsymbol{\beta} = \widetilde{\boldsymbol{\beta}} = (\beta_2 \dots \beta_p)^\top,$$

et

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

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

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

ou encore

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

Théorème 7.1.

$$\left\{\mathrm{R}\left(\mathrm{X}^\top\mathrm{X}\right)^{-1}\mathrm{R}^\top\right\}^{-1} = \widetilde{\mathrm{X}}^\top \mathrm{P}_c \widetilde{\mathrm{X}} = n\boldsymbol{\Sigma}, \tag{7.7}$$

où P_c *est l'opérateur qui centre les données déjà présenté dans l'Expression (3.7) :*

$$\mathrm{P}_c = \mathrm{I} - \frac{11^\top}{n},$$

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

Démonstration. On peut écrire

$$\mathrm{X}^\top\mathrm{X} = \begin{pmatrix} n & \mathrm{u}^\top \\ \mathrm{u} & \mathrm{Z} \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^n x_{i2} & \sum_{i=1}^n x_{i3} & \dots & \sum_{i=1}^n x_{ip} \\ \sum_{i=1}^n x_{i2} & \sum_{i=1}^n x_{i2}^2 & \sum_{i=1}^n x_{i2}x_{i3} & \dots & \sum_{i=1}^n x_{i2}x_{ip} \\ \sum_{i=1}^n x_{i3} & \sum_{i=1}^n x_{i2}x_{i3} & \sum_{i=1}^n x_{i3}^2 & \dots & \sum_{i=1}^n x_{i3}x_{ip} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n x_{ip} & \sum_{i=1}^n x_{i2}x_{ip} & \sum_{i=1}^n x_{i3}x_{ip} & \dots & \sum_{i=1}^n x_{ip}^2 \end{pmatrix},$$

où

$$\mathrm{u} = \begin{pmatrix} \sum_{i=1}^n x_{i2} & \sum_{i=1}^n x_{i3} & \dots & \sum_{i=1}^n x_{ip} \end{pmatrix}^\top,$$

et

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

Par la méthode d'inversion par bloc, on a

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

où

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

De plus,

$$(\mathrm{R}(\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{R}^\top)^{-1} = \mathrm{Q}^{-1} = \mathrm{Z} - \frac{1}{n}\mathrm{u}\mathrm{u}^\top = n\boldsymbol{\Sigma},$$

où $\boldsymbol{\Sigma}$ 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) :

$$(\mathrm{R}\widehat{\boldsymbol{\beta}} - \mathrm{r})^\top \mathrm{var}(\mathrm{R}\widehat{\boldsymbol{\beta}})^{-1}(\mathrm{R}\widehat{\boldsymbol{\beta}} - \mathrm{r}) = \widehat{\widetilde{\boldsymbol{\beta}}}^\top \widetilde{\mathrm{X}}^\top \mathrm{P}_c \widetilde{\mathrm{X}} \widehat{\widetilde{\boldsymbol{\beta}}} = \mathrm{SC}_{REG}.$$

En considérant l'Expression (3.3), la statistique de test (7.6) devient :

$$F_{obs} = \frac{\mathrm{SC}_{REG}/(p-1)}{\mathrm{SC}_{RES}/(n-p)}, \tag{7.8}$$

ce qui peut également s'écrire

$$F_{obs} = \frac{(\mathrm{SC}_{TOT} - \mathrm{SC}_{RES})/(p-1)}{\mathrm{SC}_{RES}/(n-p)}.$$

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_{obs}
Régression	SC_{REG}	$p-1$	$\mathrm{CM}_{reg} = \frac{\mathrm{SC}_{REG}}{p-1}$	$F_{obs} = \mathrm{CM}_{reg}/\mathrm{CM}_{res}$
Résiduelle	SC_{RES}	$n-p$	$\mathrm{CM}_{res} = \frac{\mathrm{SC}_{RES}}{n-p}$	
Totale	SC_{TOT}	$n-1$	$\mathrm{CM}_{tot} = \frac{\mathrm{SC}_{TOT}}{n-1}$	

La règle de décision consiste à rejeter H_0 si $F_{obs} > F_{1-\alpha,p-1,n-p}$, 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
- $q = 1$,
- $\mathrm{R} = (0 \dots \underbrace{1}_{\text{unité j}} \dots 0)$,
- $\mathrm{r} = \beta_{j0}$.

On obtient
- $\mathrm{R}\widehat{\boldsymbol{\beta}} - \mathrm{r} = \widehat{\beta}_j - \beta_{j0}$,
- $\mathrm{R}\,(\mathrm{X}^\top\mathrm{X})^{-1}\,\mathrm{R}^\top = \left[(\mathrm{X}^\top\mathrm{X})^{-1}\right]_{jj}$.

L'Expression (7.6) devient

$$F_{obs} = \frac{(\widehat{\beta}_j - \beta_{j0})^2}{\left[(\mathrm{X}^\top\mathrm{X})^{-1}\right]_{jj}\,\widetilde{\sigma}^2_\varepsilon}.$$

Sous H_0, F_{obs} suit une distribution de Fisher à 1 et $n-p$ degrés de liberté. On rejette donc H_0 si

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

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 le 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 :

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

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

$$\mathrm{x}_j = (x_{j1} \cdots x_{jp}).$$

La prédiction vient simplement et vaut :

$$\widehat{y}_j = (x_{j1} \cdots x_{jp})\widehat{\boldsymbol{\beta}}.$$

La prédiction peut également s'écrire

$$\widehat{y}_j = \mathrm{x}_j\widehat{\boldsymbol{\beta}} = \mathrm{x}_j(\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top\mathrm{y} = \mathrm{x}_j(\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top(\mathrm{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \mathrm{x}_j\boldsymbol{\beta} + \mathrm{x}_j(\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top\boldsymbol{\varepsilon}.$$

Comme la vraie valeur vaut

$$y_j = \mathrm{x}_j\boldsymbol{\beta} + \varepsilon_j,$$

l'erreur de prévision est

$$\widehat{y}_j - y_j = \mathrm{x}_j(\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top\boldsymbol{\varepsilon} - \varepsilon_j.$$

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

$$\mathrm{E}\left(\widehat{y}_j - y_j\right) = \mathrm{E}\left\{\mathrm{x}_j(\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top\boldsymbol{\varepsilon} - \varepsilon_j\right\} = \mathrm{x}_j(\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top\mathrm{E}(\boldsymbol{\varepsilon}) - \mathrm{E}(\varepsilon_j) = 0.$$

Comme la valeur prédite se réfère à une nouvelle observation,

$$\mathrm{E}(\varepsilon_j\boldsymbol{\varepsilon}) = 0,$$

et donc

$$\begin{aligned}
\mathrm{var}\left(\widehat{y}_j - y_j\right) &= \mathrm{var}\left\{\mathrm{x}_j(\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top\boldsymbol{\varepsilon}\right\} + \mathrm{var}\left\{\varepsilon_j\right\} \\
&= \mathrm{x}_j(\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top\sigma^2_\varepsilon\mathrm{X}(\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{x}_j^\top + \sigma^2_\varepsilon \\
&= \sigma^2_\varepsilon\left\{\mathrm{x}_j(\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{x}_j^\top + 1\right\}.
\end{aligned}$$

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 $\widehat{\boldsymbol{\beta}}$ et la seconde partie est due à l'erreur inconnue ε_j.

On estime la variance simplement par

$$\widehat{\mathrm{var}}\left(\widehat{y}_j - y_j\right) = \widehat{\sigma}^2_\varepsilon\left\{\mathrm{x}_j(\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{x}_j^\top + 1\right\},$$

où $\widehat{\sigma}^2_\varepsilon = \mathrm{e}^\top\mathrm{e}/(n-p)$. Enfin, il est possible de construire un intervalle de confiance pour la prévision :

$$\mathrm{IC}(1-\alpha) = \left[\hat{y}_j - t_{1-\alpha/2,n-p}\sqrt{\widehat{\mathrm{var}}\left(\widehat{y}_j - y_j\right)}; \hat{y}_j + t_{1-\alpha/2,n-p}\sqrt{\widehat{\mathrm{var}}\left(\widehat{y}_j - y_j\right)}\right].$$

7.5.2 Cas bivarié

Dans le cas où une seule variable explicative x et une constante sont utilisées, on a

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

$$(\mathrm{X}^\top\mathrm{X})^{-1} = \frac{1}{ns^2_x}\begin{pmatrix} s^2_x + \bar{x}^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}.$$

De plus, on a $\mathrm{x}_j = (1, x_j)$. La variance de l'erreur de prévision devient alors

$$\begin{aligned}
\mathrm{var}\left(\hat{y}_j - y_j\right) &= \sigma_\varepsilon^2 \left\{ \mathrm{x}_j (\mathrm{X}^\top \mathrm{X})^{-1} \mathrm{x}_j^\top + 1 \right\} \\
&= \sigma_\varepsilon^2 \left[\frac{1}{n s_x^2} \left\{ (s_x^2 + \bar{x}^2) \times 1 - \bar{x} x_j - x_j \bar{x} + x_j^2 \right\} + 1 \right] \\
&= \frac{\sigma_\varepsilon^2}{n} \left\{ n + 1 + \frac{(x_j - \bar{x})^2}{s_x^2} \right\}.
\end{aligned}$$

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

$$\widehat{\mathrm{var}}\left(\widehat{y}_j - y_j\right) = \frac{\widehat{\sigma}_\varepsilon^2}{n} \left\{ n + 1 + \frac{(x_j - \bar{x})^2}{s_x^2} \right\},$$

où $\widehat{\sigma}_\varepsilon^2 = \mathrm{e}^\top \mathrm{e}/(n-p)$.

7.6 Un exemple

Les données sur l'épargne comporte 50 lignes (pays) et 5 colonnes : (source : Belsley et al., 2005). Les données contiennent les moyennes de la période 1960-1970. Les données sont présentées dans le tableau 7.2.

— sr taux d'épargne - épargne personnelle divisée par le revenu disponible
— pop15 pourcentage de la population âgée de moins de 15 ans
— pop75 pourcentage de la population âgée de plus de 75 ans
— dpi revenu disponible par habitant en dollars
— ddpi taux de croissance en pourcentage du dpi

Les données se trouvent dans le package faraway en langage R (voir Faraway, 2004).

En langage R

```
#
# On utilise le package faraway qui a été installé préalablement
#
library(faraway)
#
# On utilise les données savings
#
data(savings)
attach(savings)
#
# Description des variables de "savings"
#
plot(savings)
```

En langage R

```
> ll=lm(sr ~ pop15+pop75+dpi+ddpi)
> summary(ll)

Call:
lm(formula = sr ~ pop15 + pop75 + dpi + ddpi)

Residuals:
  Min      1Q  Median      3Q     Max
```

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

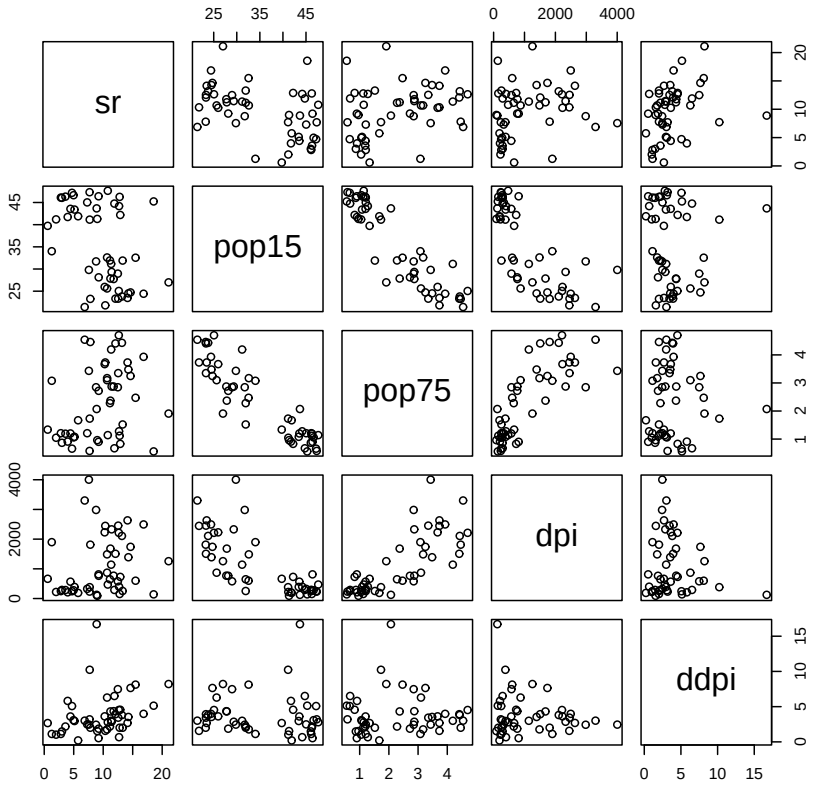


FIGURE 7.1 – Nuages de points de la base de données savings

```
-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.

En langage R

```
> 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.

En langage R

```
> 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 libertés) 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.

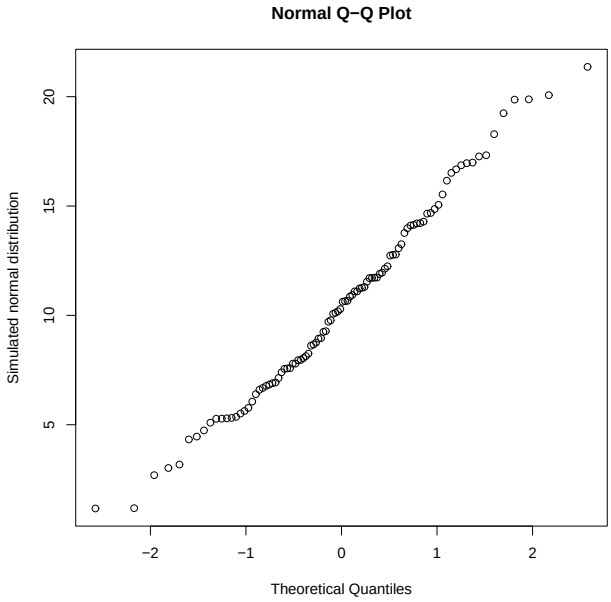


FIGURE 8.1 – qq-plot de 100 variables normales simulées versus une distribution théorique normale

En langage R, on utilise le code suivant :

En langage R

```
# Graphique 1
x=rnorm(100,10,5)
qqnorm(x,ylab = "Simulated normal distribution")
# Graphique 2
x=rexp(100,1)
qqnorm(x,ylab = "Simulated exponential distribution")
# Graphique 3
```

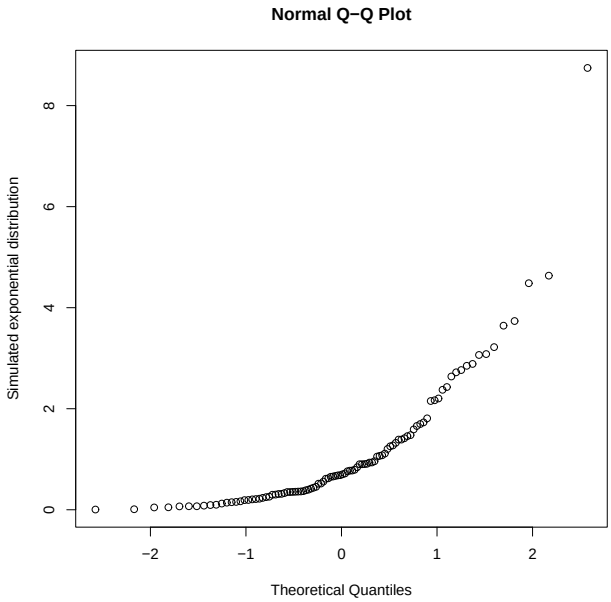


FIGURE 8.2 – qq-plot de 100 variables exponentielles simulées versus une distribution théorique normale

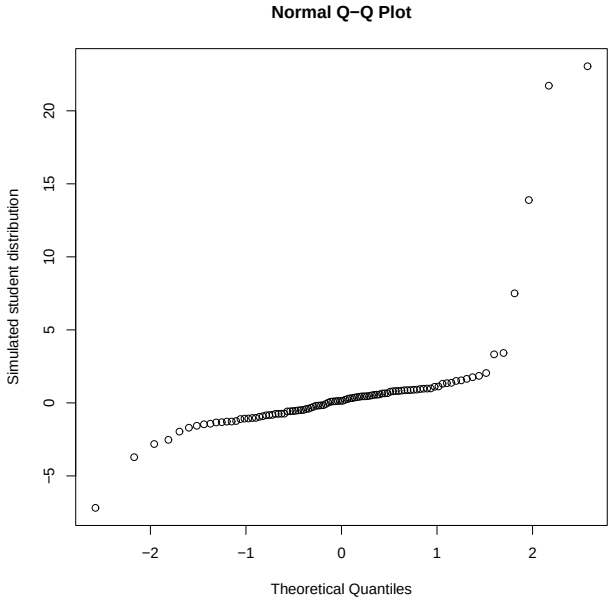


FIGURE 8.3 – qq-plot de 100 variables de Student avec 2 degrés de libertés simulées versus une distribution théorique normale

```
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 y sur le sous-espace engendré par les colonnes de la matrice X, ce qui s'écrit :

$$\mathrm{y}^* = \mathrm{P}_X \mathrm{y}. \tag{8.1}$$

La matrice P_X est une matrice idempotente et est parfois notée H et appelée *hat-matrix* (matrice chapeau) car H est un opérateur qui permet de passer de y aux valeurs ajustées appelées aussi valeurs prédites y^* de y. Le vecteur y^* est parfois noté $\widehat{\mathrm{y}}$. L'opérateur H est celui qui met un chapeau sur y au sens où on peut ré-écrire l'Expression (8.1) sous la forme.

$$\widehat{\mathrm{y}} = \mathrm{H}\mathrm{y}.$$

La variance de y^* est

$$\mathrm{var}(\mathrm{y}^*) = \mathrm{var}(\mathrm{P}_X\mathrm{y}) = \mathrm{P}_X \mathrm{var}(\mathrm{y}) \mathrm{P}_X = \mathrm{P}_X \mathrm{I} \sigma_\varepsilon^2 \mathrm{P}_X = \sigma_\varepsilon^2 \mathrm{P}_X.$$

En notant h_i la composante correspondant à la $i$ème ligne et la $i$ème colonne de P_X i.e. $h_i = [\mathrm{P}_X]_{ii}$, on a

$$\mathrm{var}(y_i^*) = h_i \sigma_\varepsilon^2.$$

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 u, on a $0 \leq \mathrm{u}^\top \mathrm{P}_X \mathrm{u} \leq 1$. En prenant un vecteur v composé de 0 et d'un seul 1 à la $i$ème composante on obtient $\mathrm{v}^\top \mathrm{P}_X \mathrm{v} = h_i$, ce qui implique que

$$0 \leq h_i \leq 1, \text{ pour tout } i = 1, \ldots, n.$$

De plus, comme P_X est une matrice idempotente de rang p, la trace de 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 y sur l'orthogonal du sous-espace engendré par les colonnes de la matrice X, ce qui s'écrit

$$\mathrm{e} = (\mathrm{I} - \mathrm{P}_X)\mathrm{y}.$$

La variance de e est donc

$$\mathrm{var}(\mathrm{e}) = \mathrm{var}[(\mathrm{I} - \mathrm{P}_X)\mathrm{y}] = (\mathrm{I} - \mathrm{P}_X)\mathrm{var}(\mathrm{y})(\mathrm{I} - \mathrm{P}_X)^\top = (\mathrm{I} - \mathrm{P}_X)\mathrm{I}\sigma_\varepsilon^2(\mathrm{I} - \mathrm{P}_X) = \sigma_\varepsilon^2(\mathrm{I} - \mathrm{P}_X).$$

On peut donc écrire

$$\mathrm{var}(e_i) = (1 - h_i)\sigma_\varepsilon^2.$$

Les résidus ne sont donc pas homoscédastiques (de même variance) alors que les termes d'erreurs du modèle sont par définition homoscédastiques. En effet, selon la définition du modèle linéaire, on a

$$\mathrm{var}(\varepsilon_i) = \sigma_\varepsilon^2.$$

Un résidu standardisé est un résidu divisé par son écart-type estimé :

$$r_i = \frac{e_i}{\sqrt{\widehat{\mathrm{var}}(e_i)}} = \frac{e_i}{\widehat{\sigma}_\varepsilon \sqrt{1 - h_i}}.$$

En général, une observation est considérée comme un point aberrnt (*outlier*) si son résidu standardisé est supérieur à 3.

La statistique r_i a cependant un problème : e_i n'est pas indépendant de $\widehat{\sigma}_\varepsilon$. Il n'est donc pas possible de déduire la distribution de probabilité des r_i. On préférera calculer des résidus studentisés défini par

$$t_i = \frac{e_i}{\widehat{\sigma}_{(i)\varepsilon} \sqrt{1 - h_i}},$$

où $\widehat{\sigma}^2_{(i)\varepsilon}$ 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 :

$$D_i = \sum_{j=1}^{n} \frac{[y_j^* - y_j^*(i)]^2}{p\widehat{\sigma}_\varepsilon^2},$$

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. L'observation est alors considérée comme influente.

Il est possible montrer que la distance de Cook peut être réécrite comme une fonction des leviers et des résidus :

$$D_i = \frac{e_i^2}{p \times \widehat{\sigma}_\varepsilon^2} \left[\frac{h_i}{(1-h_i)^2} \right].$$

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 langageR on a généré 7 réalisations de deux variables factices X et Y, avec

$$X_i \sim \mathcal{N}(10, 8), \qquad Y_i = \frac{X}{3} + \varepsilon_i, \qquad \varepsilon_i \sim \mathcal{N}(0, 1).$$

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 valeur particulière et apparait clairement sur dans la Figure 8.4, où sont représentés le nuage de points et la droite de régression. La droite de régression vaut

$$y_i = -4.4932 + 0.9483 x_i + e_i.$$

Le point peut 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 influents

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 8.2, 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 observation ($D_8 = 6.9452521$), ce qui signifie que la droite de régression sera très fortement modifiée si on enlève la 8ème observation.

Le languageR 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 alignées 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 on non des valeurs ajustées. Dans le cas présent,

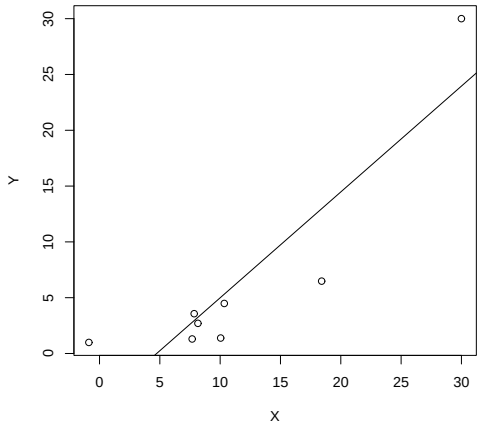


FIGURE 8.4 – Nuage de point et droite de régression : un point extrême joue un rôle de levier sur la droite de régression

il y a une dépendance, qui découle de la présence d'un point particulier. 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 influents, car leur distance de Cook dépasse respectivement 0.5 et 1. On en conclu que la régression devrait être réalisée sans la 8ème 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

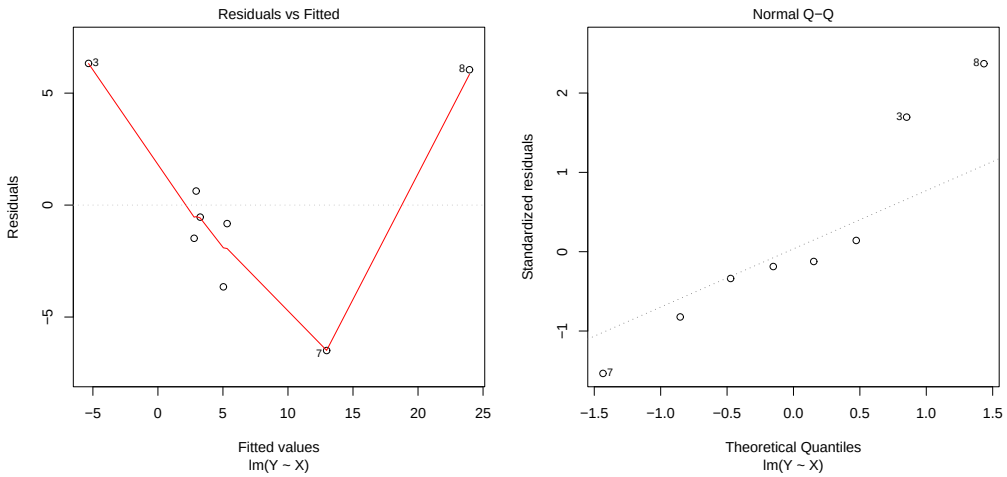


FIGURE 8.5 – Résidus standardisés en fonction des valeurs ajustées et qq-plot des résidus

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)
```

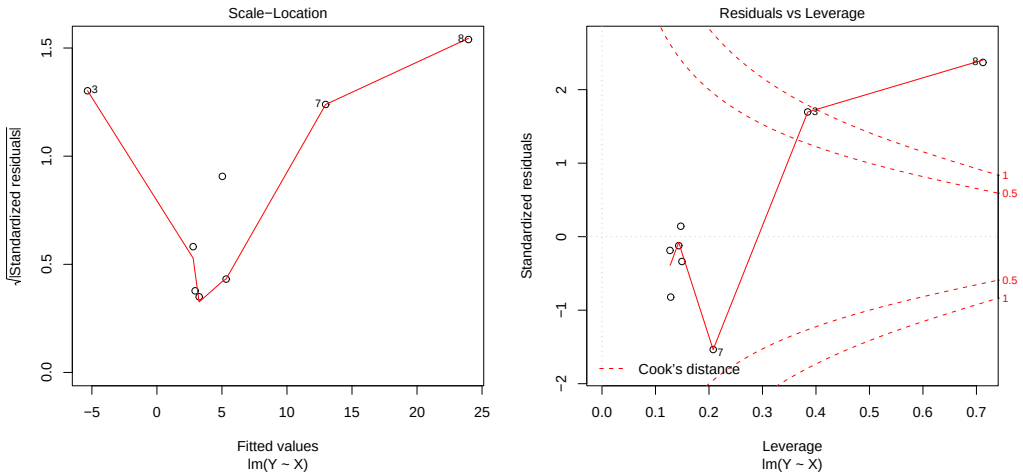


FIGURE 8.6 – Graphique des racines des valeurs absolues des résidus standardisés $\sqrt{|r_i|}$ en fonction des valeurs ajustées et graphique des résidus en fonction des leviers

```
g=lm(Y~X)
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 est un point levier car sa valeur de $h_i > 2p/n = 2 \times 3/50 = 0.12$. Sa distance de Cook relativement importante tout en restant inférieure à 1.

En langage R

```
library(faraway); data(savings); attach(savings)
ll=lm(sr~pop15+ddpi)
summary(ll)
postscript("ll.ps",onefile = TRUE,width=6, height=6,horizontal=FALSE)
plot(ll);dev.off()
```

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.

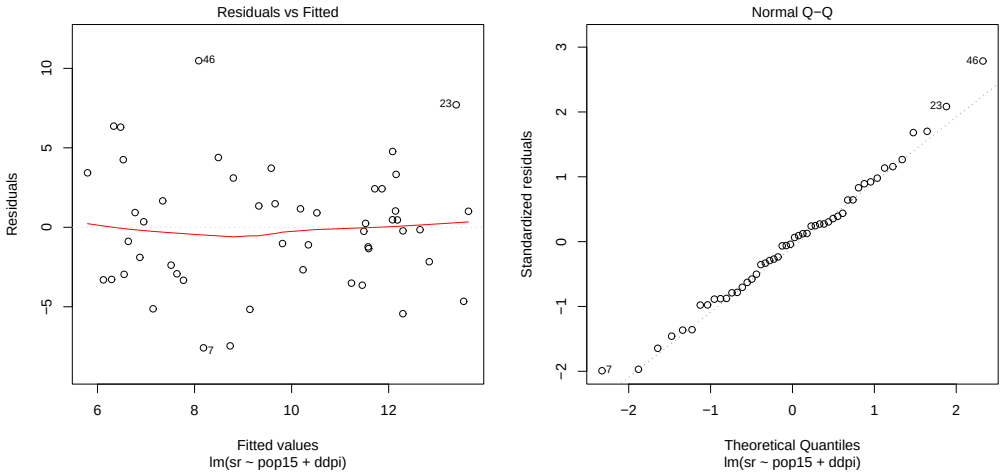


FIGURE 8.7 – Résidus standardisés en fonction des valeurs ajustées et qq-plot des résidus

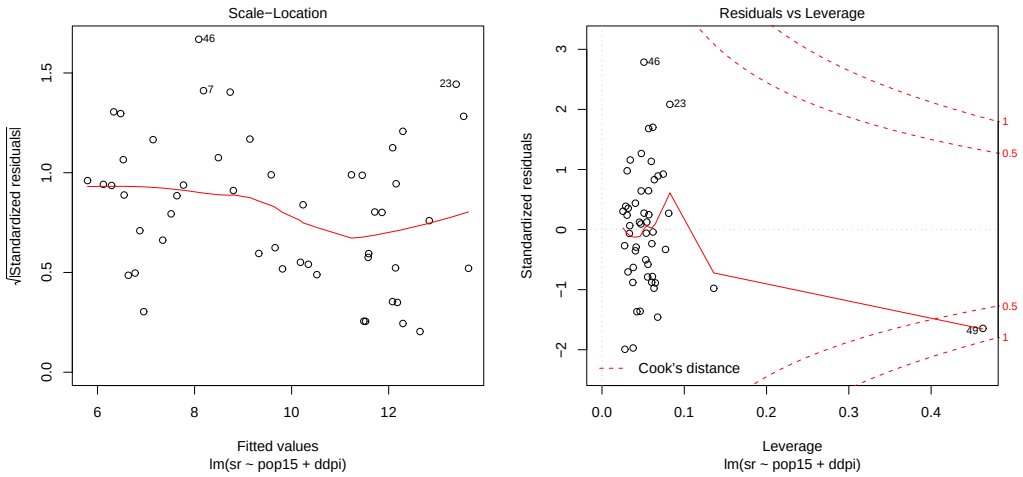


FIGURE 8.8 – Graphique des racines des valeurs absolues des résidus standardisés $\sqrt{|r_i|}$ en fonction des valeurs ajustées et graphique des résidus en fonction des leviers

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 :

$$y_{ih} = \mu_h + \varepsilon_{ih}, \tag{9.1}$$

pour tout $h = 1, \ldots, H$, et $i = 1, \ldots, n_h$, où les μ_h sont H constantes et les ε_i sont des termes d'erreur indépendants, identiquement distribués ayant une distribution normale de moyenne nulle et de variance σ_ε^2. Le modèle (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

$$\begin{cases} H_0 & \mu_1 = \mu_2 = \cdots = \mu_H \\ H_1 & \text{au moins un des } \mu_h \text{ est différent des autres.} \end{cases} \tag{9.2}$$

9.2 Méthode 1

La première méthode consiste à écrire le modèle (9.1) sous la forme d'un modèle linéaire général, où

— y est le vecteur des n observations de y_{ih}
— $\boldsymbol{\beta} = (\mu_1 \ldots \mu_h \ldots \mu_H)^\top$ est le paramètre du modèle,
— $\boldsymbol{\varepsilon}$ est le vecteur des termes d'erreur,
— 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,

$$\mathrm{X} = \left(\begin{array}{cccc} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \hline 0 & 1 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 1 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \hline \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \hline 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 1 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 1 \end{array}\right). \tag{9.3}$$

On peut dès lors écrire le modèle (9.1) sous la forme matricielle habituelle

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

La matrice $\mathrm{X}^\top\mathrm{X}$ est une matrice diagonale qui vaut

$$\mathrm{X}^\top\mathrm{X} = \begin{pmatrix} n_1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & n_2 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ 0 & 0 & \cdots & n_h & \cdots & 0 \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & n_H \end{pmatrix},$$

et son inverse vaut

$$(\mathrm{X}^\top\mathrm{X})^{-1} = \begin{pmatrix} \frac{1}{n_1} & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & \frac{1}{n_h} & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & \frac{1}{n_H} \end{pmatrix}.$$

On a également le produit

$$\mathrm{X}^\top\mathrm{y} = \begin{pmatrix} \sum_{i=1}^{n_1} y_{i1} \\ \vdots \\ \sum_{i=1}^{n_H} y_{iH} \end{pmatrix}.$$

Enfin, l'estimateur de $\boldsymbol{\beta}$ est donné par

$$\widehat{\boldsymbol{\beta}} = \left(\mathrm{X}^\top\mathrm{X}\right)^{-1}\mathrm{X}^\top\mathrm{y} = \begin{pmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_h \\ \vdots \\ \bar{y}_H \end{pmatrix},$$

où $\bar{y}_h$ est la moyenne du groupe h et l'estimateur de μ_h :

$$\widehat{\mu}_h = \bar{y}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} y_{ih},$$

pour $h = 1, \ldots, H$. 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 R de dimension $(H-1)\times H$ suivante :

$$\begin{aligned}
\mathrm{R} &= \begin{pmatrix}
1-\frac{n_1}{n} & -\frac{n_2}{n} & \cdots & -\frac{n_{H-1}}{n} & -\frac{n_H}{n} \\
-\frac{n_1}{n} & 1-\frac{n_2}{n} & \cdots & -\frac{n_{H-1}}{n} & -\frac{n_H}{n} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
-\frac{n_1}{n} & -\frac{n_2}{n} & \cdots & 1-\frac{n_{H-1}}{n} & -\frac{n_H}{n}
\end{pmatrix} \\
&= \begin{pmatrix}
1 & 0 & \cdots & 0 & 0 \\
0 & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 1 & 0
\end{pmatrix} - \begin{pmatrix}
\frac{n_1}{n} & \frac{n_2}{n} & \cdots & \frac{n_{H-1}}{n} & \frac{n_H}{n} \\
\frac{n_1}{n} & \frac{n_2}{n} & \cdots & \frac{n_{H-1}}{n} & \frac{n_H}{n} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\frac{n_1}{n} & \frac{n_2}{n} & \cdots & \frac{n_{H-1}}{n} & \frac{n_H}{n}
\end{pmatrix},
\end{aligned}$$

et r est un vecteur de zéros de dimension $H-1$. On obtient après quelques calculs :

$$\mathrm{R}\boldsymbol{\beta} = \begin{pmatrix}
\mu_1 - \frac{1}{n}\sum_{h=1}^{H} n_h\mu_h \\
\vdots \\
\mu_h - \frac{1}{n}\sum_{h=1}^{H} n_h\mu_h \\
\vdots \\
\mu_{H-1} - \frac{1}{n}\sum_{h=1}^{H} n_h\mu_h
\end{pmatrix} = \begin{pmatrix}
\mu_1 - \mu \\
\vdots \\
\mu_h - \mu \\
\vdots \\
\mu_{H-1} - \mu
\end{pmatrix},$$

et, de la même manière,

$$\mathrm{R}\widehat{\boldsymbol{\beta}} = \begin{pmatrix}
\bar{y}_1 - \bar{y} \\
\vdots \\
\bar{y}_h - \bar{y} \\
\vdots \\
\bar{y}_{H-1} - \bar{y}
\end{pmatrix},$$

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 $\mathrm{R}\boldsymbol{\beta} = \mathrm{r}$ 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 $\mathrm{R}(\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{R}^\top$. Après quelques calculs, on obtient :

$$\mathrm{R}(\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{R}^\top = \frac{1}{n}\begin{pmatrix} \frac{n}{n_1}-1 & -1 & \cdots & -1 \\ -1 & \frac{n}{n_2}-1 & \cdots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & & \frac{n}{n_{H-1}}-1 \end{pmatrix},$$

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

$$\{\mathrm{R}(\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{R}^\top\}^{-1} = \begin{pmatrix} n_1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & n_h & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & n_{H-1} \end{pmatrix} + \frac{\mathrm{n}\mathrm{n}^\top}{n_H},$$

où $\mathrm{n}^\top = (n_1\ n_2 \cdots n_{H-1})$. Enfin, après quelques calculs, on obtient

$$(\mathrm{R}\widehat{\boldsymbol{\beta}} - \mathrm{r})^\top\{\mathrm{R}(\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{R}^\top\}^{-1}(\mathrm{R}\widehat{\boldsymbol{\beta}} - \mathrm{r}) = \sum_{h=1}^{H} n_h(\bar{y}_h - \bar{y})^2,$$

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

$$\mathrm{e}^\top\mathrm{e} = \sum_{h=1}^{H}\sum_{i=1}^{n_h}(y_{ih} - \bar{y}_h)^2,$$

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,

$$\mathrm{SC}_{TOT} = \sum_{h=1}^{H}\sum_{i=1}^{n_h}(y_{ih} - \bar{y})^2,$$

on a la décomposition classique des sommes de carrés

$$\mathrm{SC}_{TOT} = \mathrm{SC}_{INTRA} + \mathrm{SC}_{INTER}.$$

On peut enfin construire la statistique de test de l'Expression (7.6). Comme $q = H - 1$, on a

$$F_{obs} = \frac{\mathrm{SC}_{INTER}/(H-1)}{\mathrm{SC}_{INTRA}/(n-H)}. \tag{9.4}$$

On construit le Tableau 9.1 d'analyse de la variance.

La règle de décision consiste à rejeter H_0 si $F_{obs} > F_{1-\alpha,H-1,n-H}$, 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.

9.3 Méthode 2

Une autre manière d'écrire le modèle (9.1) sous la forme d'un modèle linéaire consiste à poser

$$\alpha_h = \mu_h - \mu, h = 1, \ldots, H,$$

où

$$\mu = \frac{1}{n}\sum_{h=1}^{H} n_h\mu_h.$$

TABLE 9.1 – Tableau d'analyse de la variance à un facteur

Source de variation	Sommes de carrés	Degrés de liberté	Carrés moyens	F_{obs}
INTER	SC_{INTER}	$H-1$	$CM_{INTER} = \frac{SC_{INTER}}{H-1}$	$F_{obs} = \frac{CM_{INTER}}{CM_{INTRA}}$
INTRA	SC_{INTRA}	$n-H$	$CM_{INTRA} = \frac{SC_{INTRA}}{n-H}$	
TOTALE	SC_{TOT}	$n-1$	$CM_{tot} = \frac{SC_{TOT}}{n-1}$	

Le modèle s'écrit alors

$$y_{ih} = \mu + \alpha_h + \varepsilon_{ih}, \tag{9.5}$$

avec la contrainte que

$$\sum_{h=1}^{H} n_h \alpha_h = 0. \tag{9.6}$$

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

$$\alpha_H = -\frac{1}{n_H} \sum_{h=1}^{H-1} n_h \alpha_h, \tag{9.7}$$

ce qui donne

$$\begin{cases} y_{ih} = \mu + \alpha_h + \varepsilon_{ih} & \text{si } 1 \leq h \leq H-1 \\ y_{iH} = \mu - \dfrac{1}{n_H} \displaystyle\sum_{h=1}^{H-1} n_h \alpha_h + \varepsilon_{iH} & \text{sinon.} \end{cases} \tag{9.8}$$

Pour tester l'égalité des moyennes, on peut réaliser le test

$$\begin{cases} H_0 : \alpha_h = 0, \text{ pour tout } h = 1, \ldots, H-1 \\ H_1 : \text{ au moins un des } \alpha_h \text{ est différent de } 0. \end{cases}$$

On remarque qu'un test sur les $H-1$ premiers coefficients α_h suffit, en vertu de l'Expression (9.7). Le modèle (9.5) s'écrit comme un modèle linéaire général

$$\mathrm{y} = \mathrm{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

où

$$\boldsymbol{\beta}^\top = (\mu \;\; \alpha_1 \;\; \alpha_2 \cdots \alpha_{H-1}),$$

et la matrice X est de dimension $n \times H$ et est donnée par

$$\mathrm{X} = \left(\begin{array}{ccccc} 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ \hline 1 & 0 & 1 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 1 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \hline \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ \hline 1 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 1 \\ \hline 1 & -n_1/n_H & -n_2/n_H & \cdots & -n_{H-1}/n_H \\ 1 & -n_1/n_H & -n_2/n_H & \cdots & -n_{H-1}/n_H \\ \vdots & & & & \vdots \\ 1 & -n_1/n_H & -n_2/n_H & \cdots & -n_{H-1}/n_H \\ 1 & -n_1/n_H & -n_2/n_H & \cdots & -n_{H-1}/n_H \end{array}\right). \tag{9.9}$$

La première colonne de la matrice est donc une constante. Comme l'objectif est de tester la nullité des coefficients de régression 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

$$\mathrm{X}^\top\mathrm{X} = \begin{pmatrix} n & 0 & 0 & \cdots & 0 \\ 0 & n_1(1+\frac{n_1}{n_H}) & \dfrac{n_1n_2}{n_H} & \cdots & \dfrac{n_1n_{H-1}}{n_H} \\ 0 & \dfrac{n_1n_2}{n_H} & n_2(1+\frac{n_2}{n_H}) & \cdots & \dfrac{n_2n_{H-1}}{n_H} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dfrac{n_1n_{H-1}}{n_H} & \dfrac{n_2n_{H-1}}{n_H} & \cdots & n_{H-1}(1+\dfrac{n_{H-1}}{n_H}) \end{pmatrix}.$$

Son inverse est

$$(\mathrm{X}^\top\mathrm{X})^{-1} = \begin{pmatrix} 1/n & 0 & 0 & \cdots & 0 \\ 0 & \dfrac{1}{n_1}-\dfrac{1}{n} & -\dfrac{1}{n} & \cdots & -\dfrac{1}{n} \\ 0 & -\dfrac{1}{n} & \dfrac{1}{n_2}-\dfrac{1}{n} & \cdots & -\dfrac{1}{n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & -\dfrac{1}{n} & -\dfrac{1}{n} & \cdots & \dfrac{1}{n_{H-1}}-\dfrac{1}{n} \end{pmatrix}.$$

Le vecteur $\mathrm{X}^\top\mathrm{y}$ vaut

$$\mathrm{X}^\top\mathrm{y} = \begin{pmatrix} n\bar{y} \\ n_1(\bar{y}_1-\bar{y}_H) \\ \vdots \\ n_{H-1}(\bar{y}_{H-1}-\bar{y}_H) \end{pmatrix}.$$

On peut donc calculer l'estimateur de $\boldsymbol{\beta}$.

$$\widehat{\boldsymbol{\beta}} = (\mathrm{X}^\top \mathrm{X})^{-1}\mathrm{X}^\top \mathrm{y} = \begin{pmatrix} \bar{y} \\ \bar{y}_1 - \bar{y} \\ \vdots \\ \bar{y}_{H-1} - \bar{y} \end{pmatrix}.$$

L'estimateur de μ est donc $\bar{y}$ et les estimateurs α_h sont

$$\widehat{\alpha}_h = \bar{y}_h - \bar{y}, h = 1, \ldots, H-1.$$

Les valeurs ajustées valent $\mathrm{X}\widehat{\boldsymbol{\beta}}$, 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

$$e_{ih} = y_{ih} - y^*_{ih} = y_{ih} - \bar{y}_h, h = 1, \ldots, H,$$

On a donc la somme de carrés des résidus qui vaut à nouveau la somme des carrés intra-groupes

$$\mathrm{SC}_{INTRA} = \sum_{i=1}^{n} e_i^2 = \sum_{h=1}^{H}\sum_{i=1}^{n_h}(y_{ih} - \bar{y}_h)^2,$$

et la somme des carrés de la régression qui vaut à nouveau la somme des carrés inter-groupes

$$\mathrm{SC}_{INTER} = \sum_{h=1}^{H}\sum_{i=1}^{n_h}(y^*_{ih} - \bar{y})^2 = \sum_{h=1}^{H} n_h(\bar{y}_h - \bar{y})^2.$$

La statistique de test s'obtient directement à partir de l'Expression (7.8) et vaut

$$F_{obs} = \frac{\mathrm{SC}_{INTER}/(H-1)}{\mathrm{SC}_{INTRA}/(n-H)}$$

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 rejette pas 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			

Première partie

Exercices

Chapitre 10

Exercices : Régression bivariée

Exercice 10.1. Prix des maisons
Le Tableau 10.1 représente le prix (en mille USD) et la taille en sqft (ou pied carré, 1 sqft = 0.092903 m^2) de 14 maisons à vendre à San Diego (en 1990).

Table 10.1 – Prix et taille de 14 maisons à San Diego (1990)

Prix	199.9	228	235	285	239	293	285	365	295	290	385	505	425	415
Surface	1065	1254	1300	1577	1600	1750	1800	1870	1935	1948	2254	2600	2800	3000

Notre but est d'examiner la relation entre ces deux variables et essayer de prévoir le prix d'une maison à 2000 sqft. Pour cela, on cherche la droite qui ajuste au mieux le nuage de ces points au sens des moindres carrés. On doit donc calculer la pente et la constante de la droite de la régression du prix (y) par la surface (x).

1. Pour estimer ces paramètres, calculer
   (a) $\sum_{i=1}^{14} x_i$ et $\sum_{i=1}^{14} y_i$.
   (b) $\bar{x}$ et $\bar{y}$.
   (c) $\sum_{i=1}^{14} x_i^2$, $\sum_{i=1}^{14} y_i^2$ et $\sum_{i=1}^{14} x_i y_i$.
   (d) s_x^2, s_y^2 et s_{xy}, puis r_{xy}.
   Commenter le coefficient de corrélation : que pouvons-nous dire sur la relation linéaire entre x et y ?
2. Donner les paramètres de la droite de la régression :
   (a) b (la pente) et
   (b) a (la constante).
3. D'après l'équation de la régression, donner la prévision du prix d'une maison à 2000 sqft.

Solution

1. Les statistiques sont :
   (a) $\sum_{i=1}^{14} x_i = 26753$ et $\sum_{i=1}^{14} y_i = 4444.9$.
   (b) $\bar{x} = \frac{1}{14}\sum_{i=1}^{14} x_i = \frac{26753}{14} = 1910.929$ et $\bar{y} = 317.493$.
   (c) $\sum_{i=1}^{14} x_i^2 = 55462515$, $\sum_{i=1}^{14} y_i^2 = 1513039,01$ et $\sum_{i=1}^{14} x_i y_i = 9095986$.

(d) $s_x^2 = \frac{1}{14}\sum_{i=1}^{14}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{20} x_i^2 - \bar{x}^2 = 309960.209,\ s_y^2 = 7272.501$

et $s_{xy} = \frac{1}{14}\sum_{i=1}^{14}(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n}\sum_{i=1}^{20} y_i x_i - \bar{y}\bar{x} = 43007.078.$

$r_{xy} = \frac{s_{xy}}{\sqrt{s_y^2 \times s_x^2}} = 0.90583.$

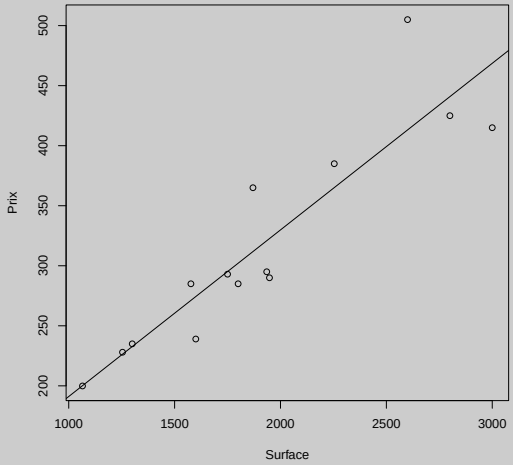
Le coefficient de corrélation est fortement positif. Les points devraient être alignés le long d'une droite croissante.

2. Les paramètres de la droite de la régression

(a)
$$b = \frac{\sum_{i=1}^{14}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{14}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{14} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{14} x_i^2 - n\bar{x}^2} = \frac{ns_{xy}}{ns_x^2} = \frac{602099}{4339443} = 0.139$$

(b)
$$a = \bar{y} - b\bar{x} = 51.874.$$

3. On a pour $x = 2000$:
$$y = a + bx = 51.874 + 0.139 \times 2000 = 329.852.$$



En langage R

```
Prix=c(199.9,228,235,285,239,293,285,365,295,290,385,505,425,415)
Surface=c(1065,1254,1300,1577,1600,1750,1800,1870,1935,1948,2254,2600,2800,3000)
plot(Prix~Surface)
abline(lm(Prix~Surface))
```

Exercice 10.2. Temps et distances
Le Tableau 10.2 représente 15 observations concernant le temps (variable y en minute) et la distance (variable x en km) roulée par des voitures d'une societé de taxi.

TABLE 10.2 – Temps et distance roulée par 15 voitures

minutes (y)	8	19	13	21	11	19	14	19	20	23	47	41	46	48	44
km (x)	3	4	4	6	6	7	8	8	9	12	16	16	20	26	15

Notre but est d'examiner la relation entre ces deux variables et d'essayer de prévoir le temps pour rouler sur une distance de 30 km.

1. Calculer les expressions suivantes :

   (a) $\sum_{i=1}^{15} x_i$ et $\sum_{i=1}^{15} y_i$.

   (b) $\bar{x}$ et $\bar{y}$.

   (c) $\sum_{i=1}^{15} x_i^2$, $\sum_{i=1}^{15} y_i^2$ et $\sum_{i=1}^{15} x_i y_i$.

   (d) s_x^2, s_y^2 et s_{xy}, puis r_{xy}.

2. À l'aide de ces expressions donner la droite de la régression de y en x, puis estimer le temps pour rouler sur une distance de 30 km.

Solution

1. Des statistiques sont :

   (a) $\sum_{i=1}^{15} x_i = 160$ et $\sum_{i=1}^{15} y_i = 393$.

   (b) $\bar{x} = \frac{1}{15}\sum_{i=1}^{15} x_i = \frac{160}{15} = 10.667$ et $\bar{y} = 26.2$.

   (c) $\sum_{i=1}^{15} x_i^2 = 2328$, $\sum_{i=1}^{15} y_i^2 = 13249$ et $\sum_{i=1}^{15} x_i y_i = 5433$.

   (d) $s_x^2 = \frac{1}{15}\sum_{i=1}^{15}(x_i - \bar{x})^2 = 41.422$, $s_y^2 = 196.827$ et $s_{xy} = \frac{1}{15}\sum_{i=1}^{15}(x_i - \bar{x})(y_i - \bar{y}) = 82.733$.
   $r_{xy} = \frac{s_{xy}}{\sqrt{s_y^2 \times s_x^2}} = 0.91627$.

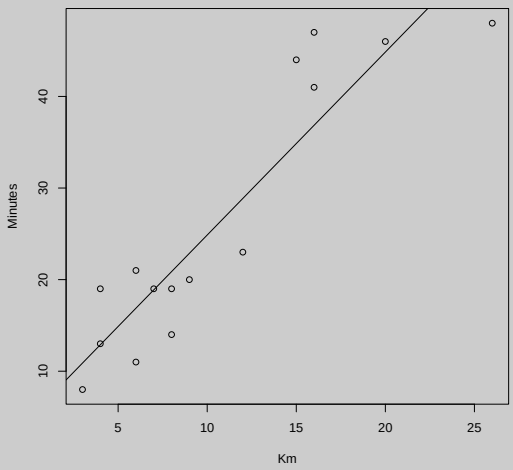
2. Les paramètres de la droite de la régression

$$b = \frac{\sum_{i=1}^{15}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{15}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{15} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{15} x_i^2 - n\bar{x}^2} = \frac{n s_{xy}}{n s_x^2} = \frac{1241}{621} = 1.997,$$

$$a = \bar{y} - b\bar{x} = 4.895.$$

   On a pour $x = 30$:

$$y = a + bx = 4.895 + 1.997 \times 30 = 64.815.$$



En langage R

```
Minutes=c(8,19,13,21,11,19,14,19,20,23,47,41,46,48,44)
Km=c(3,4,4,6,6,7,8,8,9,12,16,16,20,26,15)
plot(Minutes~Km)
abline(lm(Minutes~Km))
```

Exercice 10.3. Tailles et poids
Soit un échantillon de 20 individus dont la taille (x_i) et le poids (y_i) sont donnés dans le Tableau 10.3.

TABLE 10.3 – Taille et poids de 20 individus

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

Soient les totaux suivants :

$$\sum_{i=1}^{20} y_i = 3458, \quad \sum_{i=1}^{20} x_i = 1549, \quad \sum_{i=1}^{20} y_i^2 = 599592, \quad \sum_{i=1}^{20} x_i^2 = 123051 \text{ et } \quad \sum_{i=1}^{20} x_i y_i = 269833.$$

1. Calculer,

   (a) tous les paramètres marginaux (moyennes et variances),

   (b) la covariance,

   (c) le coefficient de corrélation et de détermination, commenter,

   (d) la droite de la régression de y en x,

   (e) les résidus et les valeurs ajustées pour les deux premières observations,

   (f) la somme des carrés totale (SC_{TOT}), la somme des carrés de la régression (SC_{REG}) et la somme des carrés des résidus (ou résiduelle) (SC_{RES}).

2. Vérifier que $SC_{TOT} = SC_{REG} + SC_{RES}$.

Solution

1. Les paramètres sont calculés à partir du tableau complété :

(a) Le Tableau donné ci-dessous permet de calculer les paramètres.

	y_i	x_i	y_i^2	x_i^2	$y_i x_i$
	155	60	24025	3600	9300
	162	61	26244	3721	9882
	157	64	24649	4096	10048
	170	67	28900	4489	11390
	164	68	26896	4624	11152
	162	69	26244	4761	11178
	169	70	28561	4900	11830
	170	70	28900	4900	11900
	178	72	31684	5184	12816
	173	73	29929	5329	12629
	180	75	32400	5625	13500
	175	76	30625	5776	13300
	173	78	29929	6084	13494
	175	80	30625	6400	14000
	179	85	32041	7225	15215
	175	90	30625	8100	15750
	180	96	32400	9216	17280
	185	96	34225	9216	17760
	189	98	35721	9604	18522
	187	101	34969	10201	18887
Total	3458	1549	599592	123051	269833
Moyenne	172.9	77.45	29979.600	6152.550	13491.650

On a donc :

$$\bar{x} = 172.9 \text{ et } \bar{y} = 77.45,$$

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{20} x_i^2 - \bar{x}^2 = 29979.600 - 172.9^2 = 85.190 \text{ et } s_y^2 = 154.048.$$

(b) La covariance :

$$s_{xy} = \frac{1}{n}\sum_{i=1}^{20} y_i x_i - \bar{y}\bar{x} = 13491.650 - 172.9 \times 77.450 = 100.545.$$

(c) Le coefficient de corrélation :

$$r_{xy} = \frac{s_{xy}}{\sqrt{s_y^2 \times s_x^2}} = \frac{100.545}{\sqrt{85.190 \times 154.048}} = 0.87769$$

et le r-carré :

$$r_{xy}^2 = 0.77033.$$

(d) La régression de y en x vaut

$$b = \frac{s_{xy}}{s_x^2} = 1.180, \text{ et } a = \bar{y} - b\bar{x} = -126.614.$$

(e) Les valeurs ajustées et les résidus sont :

x_i	y_i	y_i^*	e_i
155	60	56.324	64.585
162	61	3.676	−3.585

(f) Les sommes de carrés :

$$\text{SC}_{TOT} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = n\, s_y^2 = 20 \times 154.048 = 3080.95,$$

$$\mathrm{SC}_{REG} = \sum_{i=1}^{n} (y_i^* - \bar{y})^2 = n\ s_{y^*}^2 = n\ s_y^2 r^2 = 20 \times 154.048 \times 0.77033 = 2373.353,$$

$$\mathrm{SC}_{RES} = \sum_{i=1}^{n} (y_i - y_i^*)^2 = n\ s_e^2 = n\ s_y^2 (1 - r^2) = 20 \times 154.048 \times (1 - 0.77033) = 707.597.$$

2. On vérifie : $\mathrm{SC}_{REG} + \mathrm{SC}_{RES} = 2373.353 + 707.597 = 3080.95 = \mathrm{SC}_{TOT}$.

Exercice 10.4. Tailles et poids en R
Vérifier les résultats de l'Exercice 10.3 en utilisant le logiciel R.

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

1. Donner les statistiques descriptives demandées à l'exercice 3. (Aide : voir les fonctions "mean", "sum", "var", "cov", "cor").
2. Représenter graphiquement les deux variables pour montrer la relation possible entre elles. (Aide : voir la fonction "plot").

Solution

En langage R

```
taille=c(155,162,157,170,164,162,169,170,178,173,180,175,173,175,179,175,180,185,189,187)
toids=c(60,61,64,67,68,69,70,70,72,73,75,76,78,80,85,90,96,96,98,101)
#
# 2-statistiques
x=taille
y=poids
n=length(y)       # length donne la longueur d'un vecteur
#
xBar1=mean(x)      # moyenne des x
xBar2=sum(x)/n
yBar=mean(y)       # moyenne des y
#
s2x1=var(x)        # variance des x (division par (n-1))
s2x2=sum(x^2)/n-xBar1^2  # variance des x (division par (n))
s2x1*(n-1)/n
s2y=var(y)*(n-1)/n
#
sxy1=cov(x,y)      # covariance des x (division par (n-1))
sxy2=sum(y*x)/n-xBar1*yBar   # covariance des x (division par (n))
sxy1*(n-1)/n
#
r=cor(x,y)        # coefficient de correlation
r2=r^2            # coefficient de determination
#
b=sxy2/s2x2       # pente de la regression
a=yBar-b*xBar1    # constante de la regression
#
vajust=a+b*x      # vecteur des valeurs ajustees
vajust[1]         # valeur ajuste de la premiere observation
vajust[2]         # valeur ajuste de la deuxieme observation
#
residus=y-vajust  # vecteur des residus
residus[1]        # residu de la premiere observation
residus[2]        # residu de la deuxieme observation
#
SCtot1=n*s2y      # SCtot
SCtot2=sum((y-yBar)^2)
SCreg1=n*s2y*r2   # SCreg
SCreg2=sum((vajust-yBar)^2)
SCres1=n*s2y*(1-r2)    # SCres
SCres2=sum((vajust-y)^2)
#
SCreg1+SCres1
# 3- graphique
```

```
plot(poids~taille)# graphique des points
abline(a,b)       # ajouter la droite de regression
```

Exercice 10.5. Masse monétaire et produit national brut
La masse monétaire et le produit national brut (en millions de *snouks*) de la Fictinie (Asie postérieure) sont reproduits dans le Tableau 10.4.

TABLE 10.4 – Masse monétaire et le produit national brut

Année	Masse Monétaire (x)	PNB (y)
1987	2.0	5.0
1988	2.5	5.5
1989	3.2	6.0
1990	3.6	7.0
1991	3.3	7.2
1992	4.0	7.7
1993	4.2	8.4
1994	4.6	9.0
1995	4.8	9.7
1996	5.0	10.0

Notre but est d'examiner la relation entre ces deux variables.

1. Calculer les expressions suivantes :
   (a) $\sum_{i=1}^{10} x_i$ et $\sum_{i=1}^{10} y_i$.
   (b) $\bar{x}$ et $\bar{y}$.
   (c) $\sum_{i=1}^{10} x_i^2$, $\sum_{i=1}^{10} y_i^2$ et $\sum_{i=1}^{10} x_i y_i$.
   (d) s_x^2, s_y^2 et s_{xy}, puis r_{xy}.
2. À l'aide de ces expressions donner la droite de la régression de y en x. Puis, si en tant que ministre des Finances de la Fictinie, vous souhaitez que le PNB soit de 12.0 en 1997. À combien fixeriez-vous la masse monétaire ?

Solution

1. Des statistiques sont :
   (a)
   $$\sum_{i=1}^{10} x_i = 37.2 \text{ et } \sum_{i=1}^{10} y_i = 75.5.$$
   (b)
   $$\bar{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = \frac{37.2}{10} = 3.72 \text{ et } \bar{y} = 7.55.$$
   (c)
   $$\sum_{i=1}^{10} x_i^2 = 147.18, \sum_{i=1}^{10} y_i^2 = 597.03 \text{ et } \sum_{i=1}^{10} x_i y_i = 295.95.$$
   (d)
   $$s_x^2 = \frac{1}{10}\sum_{i=1}^{10}(x_i - \bar{x})^2 = 0.880, s_y^2 = 2.701 \text{ et } s_{xy} = \frac{1}{10}\sum_{i=1}^{10}(x_i - \bar{x})(y_i - \bar{y}) = 1.509.$$
   $$r_{xy} = \frac{s_{xy}}{\sqrt{s_y^2 \times s_x^2}} = 0.979.$$
2. Les paramètres de la droite de la régression

(a)

$$b = \frac{\sum_{i=1}^{10}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{10}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{10} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{10} x_i^2 - n\bar{x}^2} = \frac{ns_{xy}}{ns_x^2} = \frac{15.09}{8.80} = 1.715.$$

(b)

$$a = \bar{y} - b\bar{x} = 1.170.$$

3. Comme

$$y = a + bx,$$

on a

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

On fixe donc x à :

$$x = \frac{12 - 1.170}{1.715} = 6.315.$$

Chapitre 11

Exercices : Algèbre linéaire

Exercice 11.1. Vecteurs et matrices
Soit l'ensemble des vecteurs et matrices suivants :

$$A = \begin{pmatrix} 6 & 3 & 2 \\ 5 & 1 & -5 \\ -8 & 2 & 4 \end{pmatrix}, B = \begin{pmatrix} -4 & 4 & 3 \\ 0 & 2 & 1 \\ -9 & 2 & 1 \end{pmatrix}, a = \begin{pmatrix} 3 & 2 & 6 \end{pmatrix}, b = \begin{pmatrix} -2 \\ 2 \\ 6 \end{pmatrix}, c = \begin{pmatrix} -2 & 4 \end{pmatrix}, d = \begin{pmatrix} 4 \\ 2 \end{pmatrix}.$$

1. Calculer $< a, b >, < b, a >, AB$.
2. Calculer $A^\top$,$|A|$ et $\|a\|$.
3. Calculer B^{-1}.
4. Montrer que c et d sont orthogonaux.

Solution

1. On calcule :

$$< a, b >= \begin{pmatrix} 3 & 2 & 6 \end{pmatrix} \times \begin{pmatrix} -2 \\ 2 \\ 6 \end{pmatrix} = 3(-2) + 2(2) + 6(6) = -6 + 4 + 36 = 34,$$

$$< b, a >= \begin{pmatrix} -2 \\ 2 \\ 6 \end{pmatrix} \times \begin{pmatrix} 3 & 2 & 6 \end{pmatrix} = \begin{pmatrix} -2(3) & -2(2) & -2(6) \\ 2(3) & 2(2) & 2(6) \\ 6(3) & 6(2) & 6(6) \end{pmatrix} = \begin{pmatrix} -6 & -4 & -12 \\ 6 & 4 & 12 \\ 18 & 12 & 36 \end{pmatrix},$$

$$\begin{aligned} AB &= \begin{pmatrix} 6 & 3 & 2 \\ 5 & 1 & -5 \\ -8 & 2 & 4 \end{pmatrix} \times \begin{pmatrix} -4 & 4 & 3 \\ 0 & 2 & 1 \\ -9 & 2 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 6(-4) + 3(0) + 2(-9) & 6(4) + 3(2) + 2(2) & 6(3) + 3(1) + 2(1) \\ 5(-4) + 1(0) + (-5)(-9) & 5(4) + 1(2) + (-5)(2) & 5(3) + 1(1) + (-5)(1) \\ (-8)(-4) + 2(0) + 4(-9) & (-8)(4) + 2(2) + 4(2) & (-8)(3) + 2(1) + 4(1) \end{pmatrix} \\ &= \begin{pmatrix} -42 & 34 & 23 \\ 25 & 12 & 11 \\ -4 & -20 & -18 \end{pmatrix}. \end{aligned}$$

2. On a également :

$$A^\top = \begin{pmatrix} 6 & 5 & -8 \\ 3 & 1 & 2 \\ 2 & -5 & 4 \end{pmatrix},$$

$$\begin{aligned} |A| &= \begin{vmatrix} \mathbf{6} & \mathbf{3} & \mathbf{2} \\ 5 & 1 & -5 \\ -8 & 2 & 4 \end{vmatrix} = 6 \begin{vmatrix} 1 & -5 \\ 2 & 4 \end{vmatrix} - 3 \begin{vmatrix} 5 & -5 \\ -8 & 4 \end{vmatrix} + 2 \begin{vmatrix} 5 & 1 \\ -8 & 2 \end{vmatrix} \\ &= 6(1(4) - (-5)(2)) - 3(5(4) - (-8)(-5)) + 2(5(2) - 1(-8)) = 180, \end{aligned}$$

$$\|a\| = \sqrt{3^2 + 2^2 + 6^2} = \sqrt{49} = 7.$$

3. On utilise la méthode des cofacteurs. On peut écrire :

$$\mathrm{B}^{-1} = \frac{1}{|\mathrm{B}|}\mathrm{B}^{c\top},$$

où B^c est la matrice des cofacteurs. Toutes les matrices ne sont pas inversibles.

(a) On vérifie que le déterminant est non-nul :

$$|\mathrm{B}| = -4\times\begin{vmatrix} 2 & 1 \\ 2 & 1 \end{vmatrix} - 4\times\begin{vmatrix} 0 & 1 \\ -9 & 1 \end{vmatrix} + 3\times\begin{vmatrix} 0 & 2 \\ -9 & 2 \end{vmatrix} = -4\times 0 - 4\times 9 + 3\times 18 = 18 \neq 0.$$

La matrice est inversible et son déterminant vaut 18.

(b) Matrice des cofacteurs :

$$\mathrm{B}^c = \begin{pmatrix} \begin{vmatrix} 2 & 1 \\ 2 & 1 \end{vmatrix} & -\begin{vmatrix} 0 & 1 \\ -9 & 1 \end{vmatrix} & \begin{vmatrix} 0 & 2 \\ -9 & 2 \end{vmatrix} \\ -\begin{vmatrix} 4 & 3 \\ 2 & 1 \end{vmatrix} & \begin{vmatrix} -4 & 3 \\ -9 & 1 \end{vmatrix} & -\begin{vmatrix} -4 & 4 \\ -9 & 2 \end{vmatrix} \\ \begin{vmatrix} 4 & 3 \\ 2 & 1 \end{vmatrix} & -\begin{vmatrix} -4 & 3 \\ 0 & 1 \end{vmatrix} & \begin{vmatrix} -4 & 4 \\ 0 & 2 \end{vmatrix} \end{pmatrix} = \begin{pmatrix} 0 & -9 & 18 \\ 2 & 23 & -28 \\ -2 & 4 & -8 \end{pmatrix}.$$

L'inverse est

$$\mathrm{B}^{-1} = \frac{1}{|\mathrm{B}|}\mathrm{B}^{c\top} = \frac{1}{18}\begin{pmatrix} 0 & 2 & -2 \\ -9 & 23 & 4 \\ 18 & -28 & -8 \end{pmatrix}.$$

4. Comme $< \mathrm{c}, \mathrm{d} > = \begin{pmatrix} -2 & 4 \end{pmatrix} \times \begin{pmatrix} 4 \\ 2 \end{pmatrix} = 0$, c et d sont orthogonaux.

Exercice 11.2. Matrices
Soient les matrices A et B suivantes :

$$\mathrm{A} = \begin{pmatrix} 2/6 & 5/6 \\ 4/6 & 10/6 \end{pmatrix}, \mathrm{B} = \begin{pmatrix} 6 & 8 & 10 \\ 0 & 2 & 4 \\ 12 & 16 & 20 \end{pmatrix}.$$

Calculer

1. leur rang.
2. leur trace.
3. A et B sont-elles idempotentes ?

Solution
Pour la matrice A,

1. On peut écrire $\mathrm{A} = \frac{1}{6}\begin{pmatrix} 2 & 5 \\ 4 & 10 \end{pmatrix} = \frac{1}{6}\mathrm{A}^a$, où $\mathrm{A}^a = \begin{pmatrix} 2 & 5 \\ 4 & 10 \end{pmatrix}$. Donc, le rang(A)=rang (A^a). La seconde ligne de A^a est égale au double de la première ligne. Donc, le rang de A vaut 1.
2. $\mathrm{trace}(\mathrm{A}) = \mathrm{trace}(\begin{pmatrix} 2/6 & 5/6 \\ 4/6 & 10/6 \end{pmatrix} = 2/6 + 10/6 = 12/6 = 2.$
3. On calcule $\mathrm{A}^2 = \mathrm{A}\,\mathrm{A} = \frac{1}{36}\begin{pmatrix} 2 & 5 \\ 4 & 10 \end{pmatrix}\begin{pmatrix} 2 & 5 \\ 4 & 10 \end{pmatrix} = \frac{1}{36}\begin{pmatrix} 24 & 60 \\ 48 & 120 \end{pmatrix} = \frac{12}{36}\begin{pmatrix} 2 & 5 \\ 4 & 10 \end{pmatrix} = \mathrm{A}^2 \neq \mathrm{A}$. Donc, A n'est pas idempotente.

Pour la matrice B,

1. La troisième ligne de la matrice B est égale au double de sa première ligne. La deuxième ligne n'est pas proportionnelle à la première ligne. Donc, il y a seulement deux lignes (lignes 1 et 2 ou ligne 2 et ligne 3) linéairement indépendantes et le rang vaut 2.
2. $\mathrm{trace}(\mathrm{B}) = \begin{pmatrix} 6 & 8 & 10 \\ 0 & 2 & 4 \\ 12 & 16 & 20 \end{pmatrix} = 6 + 2 + 20 = 28.$
3. On calcule $\mathrm{B}^2 = \mathrm{B}\,\mathrm{B} = \begin{pmatrix} 6 & 8 & 10 \\ 0 & 2 & 4 \\ 12 & 16 & 20 \end{pmatrix}\begin{pmatrix} 6 & 8 & 10 \\ 0 & 2 & 4 \\ 12 & 16 & 20 \end{pmatrix} = \begin{pmatrix} 156 & 224 & 292 \\ 48 & 68 & 88 \\ 312 & 448 & 584 \end{pmatrix} \neq \mathrm{B}$. La matrice B n'est pas idempotente.

Exercice 11.3.
Vérifier si les vecteurs suivants sont orthogonaux, ou pas.

1. $(1, 2)$ et $(-2, 1)^\top$
2. $(1, -1, 1)$ et $(-1, 1, -1)^\top$
3. $(a, -b, 1)$ et $(b, a, 0)^\top$

Solution

1. $(1, 2) \times (-2, 1)^\top = 1 \times -2 + 2 \times -1 = 0$ orthogonaux
2. $(1, -1, 1) \times (-1, 1, -1)^\top = 1 \times -1 - 1 \times 1 + 1 \times -1 = -3$ pas orthogonaux
3. $(a, -b, 1) \times (b, a, 0) = a \times b - b \times a + 1 - 0 = 0$ orthogonaux

Exercice 11.4.
Est-il vrai que le produit de matrices symétriques est toujours symétrique ?

Solution
Non. Il suffit de trouver un contre-exemple :

$$\mathbf{A} = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \text{ et } \mathbf{B} = \begin{pmatrix} d & e \\ e & f \end{pmatrix}$$

$$\mathbf{AB} = \begin{pmatrix} ad + be & ae + bf \\ bd + ce & be + cf \end{pmatrix}.$$

Si

$$a = 1, b = 2, c = 3, d = 4, e = 5, f = 6,$$

alors

$$\mathbf{AB} = \begin{pmatrix} 14 & 17 \\ 23 & 28 \end{pmatrix}.$$

Donc, non, le produit des matrices symétriques n'est pas toujours symétrique.

Exercice 11.5. Inversion par bloc
Montrer que

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

où

$$Q = (D - CA^{-1}B)^{-1} \text{ et } F = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

et A, B, C et D sont des sous-matrices de F. Les matrices A et D sont carrées et inversibles.

Solution
Pour montrer cette égalité, on regarde si $FF^{-1} = I$. On calcule :

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} A^{-1} + A^{-1}BQCA^{-1} & -A^{-1}BQ \\ -QCA^{-1} & Q \end{pmatrix} = \begin{pmatrix} AA^{-1} + AA^{-1}BQCA^{-1} - BQCA^{-1} & -AA^{-1}BQ + BQ \\ CA^{-1} + CA^{-1}BQCA^{-1} - DQCA^{-1} & -CA^{-1}BQ + DQ \end{pmatrix}.$$

Par élément :

1. $AA^{-1} + AA^{-1}BQCA^{-1} - BQCA^{-1} = I + IBQCA^{-1} - BQCA^{-1} = I$
2. $-AA^{-1}BQ + BQ = -IBQ + BQ = 0$
3. $CA^{-1} + CA^{-1}BQCA^{-1} - DQCA^{-1} = CA^{-1} - (D - CA^{-1}B)QCA^{-1} = CA^{-1} - Q^{-1}QCA^{-1} = 0.$
4. $-CA^{-1}BQ + DQ = (D - CA^{-1}B)Q = Q^{-1}Q = I.$

Donc,

$$FF^{-1} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} = I.$$

Exercice 11.6. Valeurs propres
Soient deux matrices A et P de dimension $n \times n$. La matrice P est supposée non singulière. Montrer que les valeurs propres de la matrice A sont égales aux valeurs propres de la matrice $P^{-1}AP$.

Solution
Si les valeurs propres de $P^{-1}AP$ et A sont identiques, alors on devrait avoir pour toute valeur propre λ de A l'égalité

$$|A - \lambda I| = |P^{-1}AP - \lambda I|$$

car si ces deux déterminants sont égaux, les valeurs propres de A et de $P^{-1}AP$ sont égales aussi. On a :

$$|P^{-1}AP - \lambda I| = |P^{-1}AP - P^{-1}\lambda IP| = |P^{-1}(A - \lambda I)P| = |P^{-1}||A - \lambda I||P| = \frac{1}{|P|}|A - \lambda I||P| = |A - \lambda I|.$$

Exercice 11.7. Décomposition d'une matrice
Soit

$$\mathrm{A} = \begin{pmatrix} 6 & -14 & 0 \\ 0 & -3 & -6 \\ 0 & -3 & 0 \end{pmatrix}.$$

1. Donner des matrices P et D telles que $\mathrm{A} = \mathrm{PDP}^{-1}$, où D est une matrice diagonale.
2. Vérifier que la trace de la matrice A est égale à la somme de ses valeurs propres réelles ou complexes.
3. Vérifier que le déterminant de la matrice A est égal au produit de ses valeurs propres réelles ou complexes.

Solution

1. On a :

$$\mathrm{P}^{-1}\mathrm{AP} = \mathrm{P}^{-1}\mathrm{PDP}^{-1}\mathrm{P} = \mathrm{D} = \mathrm{diag}(d_1, d_2, \ldots, d_n) = \mathrm{diag}(\lambda_1, \lambda_2, \ldots \lambda_n),$$

où λ_i est une valeur propre de la matrice $\mathrm{D} = \mathrm{P}^{-1}\mathrm{AP}$. Les valeurs propres de $\mathrm{P}^{-1}\mathrm{AP}$ sont les mêmes que les valeurs propres de A. Donc, les éléments diagonaux de D sont les valeurs propres de A. Reste à trouver P. Comme

$$\begin{aligned} \mathrm{AP} &= \mathrm{Pdiag}(\lambda_1, \lambda_2, \ldots \lambda_n) = \begin{pmatrix} p_{11} & p_{12} & \ldots & p_{1n} \\ p_{21} & p_{22} & \ldots & p_{2n} \\ \vdots & & & \\ p_{n1} & p_{n2} & \ldots & p_{nn} \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & 0 & \ldots & 0 \\ 0 & \lambda_2 & 0 & \ldots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \ldots & \lambda_n \end{pmatrix} \\ &= \begin{pmatrix} \lambda_1 p_{11} & \lambda_2 p_{12} & \ldots & \lambda_n p_{1n} \\ \lambda_1 p_{21} & \lambda_2 p_{22} & \ldots & \lambda_n p_{2n} \\ \vdots & & & \\ \lambda_1 p_{n1} & \lambda_2 p_{n2} & \ldots & \lambda_n p_{nn} \end{pmatrix} = (\lambda_1 \mathrm{x}_1, \lambda_2 \mathrm{x}_2, \ldots, \lambda_n \mathrm{x}_n), \end{aligned}$$

avec $\mathrm{x}_k = (p_{1k}, p_{2k}, \ldots, p_{nk})^\top$. Et comme $\mathrm{AP} = (\mathrm{Ax}_1, \mathrm{Ax}_2, \ldots, \mathrm{Ax}_n)$, on obtient $\mathrm{Ax}_k = \lambda_k \mathrm{x}_k$, pour $k = 1, 2, \ldots, n$, ce qui veut dire que $\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n$ sont les vecteurs propres de A. Donc, les colonnes de P sont les vecteurs propres de A et les éléments diagonaux de D sont les valeurs propres de A. On diagonalise donc A. On commence par les valeurs propres :

$$|\mathrm{A} - \lambda \mathrm{I}| = \begin{vmatrix} 6-\lambda & -14 & 0 \\ 0 & -3-\lambda & -6 \\ 0 & -3 & 0-\lambda \end{vmatrix} = (6-\lambda)[(-3-\lambda)(0-\lambda) - 18] = -(\lambda - 6)(\lambda^2 + 3\lambda - 18) = 0.$$

Les valeurs propres sont $\lambda_1 = 6, \lambda_2 = -6, \lambda_3 = 3$. Les vecteurs propres correspondants sont :

(a) Pour $\lambda_1 = 6$,

$$(\mathrm{A} - \lambda_1 \mathrm{I})\mathrm{u} = \begin{pmatrix} 0 & -14 & 0 \\ 0 & -9 & -6 \\ 0 & -3 & -6 \end{pmatrix} \mathrm{u} = \begin{pmatrix} -14u_2 \\ -9u_2 - 6u_3 \\ -3u_2 - 6u_3 \end{pmatrix} = 0.$$

On a donc, $\mathrm{u}_1 = (1, 0, 0)^\top$.

(b) Pour $\lambda_2 = -6$,

$$(\mathrm{A} - \lambda_2 \mathrm{I})\mathrm{u} = \begin{pmatrix} 12 & -14 & 0 \\ 0 & 3 & -6 \\ 0 & -3 & 6 \end{pmatrix} \mathrm{u} = \begin{pmatrix} 12u_1 - 14u_2 \\ 3u_2 - 6u_3 \\ -3u_2 + 6u_3 \end{pmatrix} = 0.$$

On a donc, $u_2 = 2u_3$ et $u_1 = 7u_2/6$. Donc, $\mathrm{u}_2 = (7, 6, 3)^\top$.

(c) Pour $\lambda_3 = 3$,

$$(\mathrm{A} - \lambda_3 \mathrm{I})\mathrm{u} = \begin{pmatrix} 3 & -14 & 0 \\ 0 & -6 & -6 \\ 0 & -3 & -3 \end{pmatrix} \mathrm{u} = \begin{pmatrix} 3u_1 - 14u_2 \\ -6u_2 - 6u_3 \\ -3u_2 - 3u_3 \end{pmatrix} = 0.$$

On a donc, $u_2 = -u_3$ et $u_1 = 14u_2/3$. Donc, $\mathrm{u}_3 = (14, 3, -3)^\top$.

Et finalement,

$$\mathrm{P} = \begin{pmatrix} 1 & 7 & 14 \\ 0 & 6 & 3 \\ 0 & 3 & -3 \end{pmatrix} \text{ et } \mathrm{D} = \mathrm{diag}(6, -6, 3).$$

2. $\text{trace}(\text{A}) = 6 - 3 + 0 = 3 = \sum_{i=1}^{3} \lambda_i = 6 - 6 + 3.$

3. $|\text{A}| = 6 \times \{-3 \times 0 - (-6 \times -3)\} = -108 = \prod_{i=1}^{3} \lambda_i = 6 \times -6 \times 3 = -108.$

Exercice 11.8. Vecteurs propres d'une matrice symétrique
Soit A une matrice symétrique de dimension $n \times n$. Montrer si x et y sont deux vecteurs propres différents de A, alors x et y sont orthogonaux.

Solution
Si x et y sont deux vecteurs propres différents : $\mathrm{Ax} = \lambda \mathrm{x}$ et $\mathrm{Ay} = \mu \mathrm{y}$ où λ et μ sont deux valeurs propres différentes de A. alors

$$\mathrm{y}^\top \mathrm{Ax} = \lambda \mathrm{y}^\top \mathrm{x} \text{ et } \mathrm{x}^\top \mathrm{Ay} = \mu \mathrm{x}^\top \mathrm{y}.$$

Comme A est symétrique

$$\mathrm{x}^\top \mathrm{Ay} = \lambda \mathrm{x}^\top \mathrm{y} \text{ et } \mathrm{y}^t \mathrm{Ax} = \mu \mathrm{y}^\top \mathrm{x}.$$

Donc,

$$\lambda \mathrm{x}^\top \mathrm{y} = \mu \mathrm{x}^\top \mathrm{y}$$

et

$$(\lambda - \mu) \mathrm{x}^\top \mathrm{y} = 0.$$

Comme $\lambda \neq \mu$, $\mathrm{x}^\top \mathrm{y} = 0$, ce qui veut dire qu'ils sont orthogonaux.

Exercice 11.9. Valeurs et vecteurs propres
Soient les matrices C et D suivantes :

$$\mathrm{C} = \begin{pmatrix} 4 & 1 \\ 7 & -2 \end{pmatrix}, \mathrm{D} = \begin{pmatrix} 6 & 3 & -8 \\ 0 & -2 & 0 \\ 1 & 0 & -3 \end{pmatrix}.$$

Calculer

1. Le(s) valeur(s) propre(s) de C et D.
2. Le(s) vecteur(s) propre(s) de C et D.

Solution

— Pour la matrice C, les valeurs propres sont les suivantes :
$|\mathrm{C} - \lambda \mathrm{I}| = \begin{vmatrix} 4-\lambda & 1 \\ 7 & -2-\lambda \end{vmatrix} = (4-\lambda)(-2-\lambda) - 7 = -8 - 4\lambda + 2\lambda + \lambda^2 - 7 = \lambda^2 - 2\lambda - 15 = 0$
$\Rightarrow (\lambda - 5)(\lambda + 3) = 0$
$\Rightarrow \lambda = 5, -3.$
— Si le vecteur propre s'écrit $\mathrm{u} = (a, b)^\top$. On détermine le vecteur propre pour $\lambda = 5$ en résolvant :

$$(\mathrm{C} - \lambda \mathrm{I})\mathrm{u} = 0.$$

Or,

$$(\mathrm{C} - \lambda \mathrm{I})\mathrm{u} = \left\{ \begin{pmatrix} 4 & 1 \\ 7 & -2 \end{pmatrix} - 5 \times \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right\} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ 7 & -7 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a-b \\ 7a-7b \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Il faut donc que $a = b$. On peut prendre le vecteur $(1, 1)$
— Si le vecteur propre s'écrit à nouveau $\mathrm{u} = (a, b)^\top$. On peut trouver le vecteur propre pour $\lambda = -3$ en résolvant

$$(\mathrm{C} - \lambda \mathrm{I})\mathrm{u} = 0.$$

Or,

$$(\mathrm{C} - \lambda \mathrm{I})\mathrm{u} = \left\{ \begin{pmatrix} 4 & 1 \\ 7 & -2 \end{pmatrix} + 3 \times \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right\} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 7 & 1 \\ 7 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 7a+b \\ 7a+b \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Il faut donc que $7a = -b$. On peut prendre le vecteur $(1, -7)$.
— Pour la matrice D, les valeurs propres sont les suivantes :

$$|\mathrm{D} - \lambda \mathrm{I}| = \begin{vmatrix} 6-\lambda & 3 & -8 \\ 0 & -2-\lambda & 0 \\ 1 & 0 & -3-\lambda \end{vmatrix} = (6-\lambda) \begin{vmatrix} -2-\lambda & 0 \\ 0 & -3-\lambda \end{vmatrix} + 1 \begin{vmatrix} 3 & -8 \\ -2-\lambda & 0 \end{vmatrix}$$
$$= (6-\lambda)(-2-\lambda)(-3-\lambda) + 1(0(3) - (-8)(-2-\lambda)) = (6-\lambda)(2+\lambda)(3+\lambda) - 8(2+\lambda))$$
$$= (2+\lambda)(\lambda - 5)(\lambda + 2) = 0.$$

Il y a donc trois valeurs propres -2, -2, et 5.
— Si le vecteur propre s'écrit $\mathrm{u} = (a, b, c)^\top$. On peut trouver le vecteur propre pour $\lambda = 5$ en résolvant

$$(\mathrm{D} - \lambda \mathrm{I})\mathrm{u} = 0.$$

Or,

$$(\mathrm{D} - \lambda \mathrm{I})\mathrm{u} = \left\{ \begin{pmatrix} 6 & 3 & -8 \\ 0 & -2 & 0 \\ 1 & 0 & -3 \end{pmatrix} - 5 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right\} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 1 & 3 & -8 \\ 0 & -7 & 0 \\ 1 & 0 & -8 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} a+3b-8c \\ -7b \\ a-8c \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Il faut donc que $b = 0$ et que $a = 8c$. On peut prendre $\mathrm{u} = (8, 0, 1)^\top$.
— Si le vecteur propre s'écrit à nouveau $\mathrm{u} = (a, b, c)^\top$. On peut chercher les vecteurs propres pour $\lambda = -2$ en résolvant

$$(\mathrm{D} - \lambda \mathrm{I})\mathrm{u} = 0.$$

Or,

$$(\mathrm{D} - \lambda \mathrm{I})\mathrm{u} = \left\{ \begin{pmatrix} 6 & 3 & -8 \\ 0 & -2 & 0 \\ 1 & 0 & -3 \end{pmatrix} + 2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right\} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 8 & 3 & -8 \\ 0 & 0 & 0 \\ 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 8a+3b-8c \\ 0 \\ a-c \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Il faut donc que $a = c$ et que $b = 0$. On peut prendre $\mathrm{u} = (1, 0, 1)^\top$.

Exercice 11.10. Puissance de matrice
Soit

$$A = \frac{1}{2}\begin{pmatrix} -1 & -\sqrt{3} \\ \sqrt{3} & -1 \end{pmatrix}.$$

Montrer que $A^3 = I$. En utilisant ce résultat, calculer A^{-1}.

Solution
On a :

$$A^3 = AAA = AA^2 = A^2A.$$

Donc,

$$A^2 = \frac{1}{2}\begin{pmatrix} -1 & -\sqrt{3} \\ \sqrt{3} & -1 \end{pmatrix} \times \frac{1}{2}\begin{pmatrix} -1 & -\sqrt{3} \\ \sqrt{3} & -1 \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 1-3 & 2\sqrt{3} \\ -2\sqrt{3} & -3+1 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} -1 & \sqrt{3} \\ -\sqrt{3} & -1 \end{pmatrix},$$

et

$$AA^2 = \frac{1}{2}\begin{pmatrix} -1 & -\sqrt{3} \\ \sqrt{3} & -1 \end{pmatrix} \times \frac{1}{2}\begin{pmatrix} -1 & \sqrt{3} \\ -\sqrt{3} & -1 \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 1+3 & 0 \\ 0 & 1+3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I.$$

Comme

$$AA^2 = I = AA^{-1},$$

on obtient

$$A^{-1} = A^2 = \frac{1}{2}\begin{pmatrix} -1 & \sqrt{3} \\ -\sqrt{3} & -1 \end{pmatrix}.$$

Exercice 11.11. Projections

1. Construire des projecteurs orthogonaux P_1, P_2, P_3, sur des sous-espaces engendrés par les colonnes des matrices.
$$X_1 = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, X_2 = \begin{pmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{pmatrix}, X_3 = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_i \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}.$$
$(P_X = X(X^\top X)^{-1}X^\top)$.
2. Construire les trois projecteurs qui projettent sur l'orthogonal des sous-espaces engendré par les colonnes de X_1, X_2, X_3. $(P_X^\perp = I - X(X^\top X)^{-1}X^\top)$
3. Ces 6 projecteurs sont des matrices idempotentes. Vérifier cette propriété pour $P_1, P_2, P_1^\perp$ et $P_2^\perp$.
4. Projeter le vecteur $y = (y_1, \ldots, y_i, \ldots, y_n)^\top$ au moyen des projecteurs $P_1, P_2, P_1^\perp$ et $P_2^\perp$.

Solution

1. Construction des projecteurs orthogonaux P_1, P_2, P_3 sur des sous espaces engendrés par X_1, X_2 et X_3
   (a) Pour calculer
   $$P_1 = X_1(X_1^\top X_1)^{-1}X_1^\top,$$
   on a
   $$X_1^\top X_1 = (1 \cdots 1)\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = n, (X_1^\top X_1)^{-1} = \frac{1}{n},$$
   $$(X_1^\top X_1)^{-1}X_1^\top = \frac{1}{n}(1 \cdots 1) = \left(\frac{1}{n} \cdots \frac{1}{n}\right).$$
   Donc,
   $$P_1 = X_1(X_1^\top X_1)^{-1}X_1^\top = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}\left(\frac{1}{n} \cdots \frac{1}{n}\right) = \begin{pmatrix} 1/n & \cdots & 1/n \\ \vdots & \ddots & \vdots \\ 1/n & \cdots & 1/n \end{pmatrix}.$$
   (b) Pour calculer
   $$P_2 = X_2(X_2^\top X_2)^{-1}X_2^\top,$$
   on a
   $$X_2^\top X_2 = (x_1 \cdots x_n)\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \sum_{i=1}^n x_i^2,$$
   $$(X_2^\top X_2)^{-1} = \frac{1}{\sum_{i=1}^n x_i^2},$$
   $$(X_2^\top X_2)^{-1}X_2^\top = \frac{1}{\sum_{i=1}^n x_i^2}(x_1 \cdots x_n) = \left(\frac{x_1}{\sum_{i=1}^n x_i^2} \cdots \frac{x_n}{\sum_{i=1}^n x_i^2}\right).$$
   Donc,
   $$P_2 = X_2(X_2^\top X_2)^{-1}X_2^\top = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\left(\frac{x_1}{\sum_{i=1}^n x_i^2} \cdots \frac{x_n}{\sum_{i=1}^n x_i^2}\right)$$
   $$= \begin{pmatrix} \frac{x_1^2}{\sum_{i=1}^n x_i^2} & \cdots & \frac{x_1 x_n}{\sum_{i=1}^n x_i^2} \\ \vdots & \ddots & \vdots \\ \frac{x_1 x_n}{\sum_{i=1}^n x_i^2} & \cdots & \frac{x_n^2}{\sum_{i=1}^n x_i^2} \end{pmatrix} = \frac{1}{\sum_{i=1}^n x_i^2}\begin{pmatrix} x_1^2 & \cdots & x_1 x_n \\ \vdots & \ddots & \vdots \\ x_1 x_n & \cdots & x_n^2 \end{pmatrix}.$$

(c) Pour calculer

$$\mathrm{P}_3 = \mathrm{X}_3(\mathrm{X}_3^\top\mathrm{X}_3)^{-1}\mathrm{X}_3^\top,$$

on a

$$\mathrm{X}_3^\top\mathrm{X}_3 = \begin{pmatrix}1 & \cdots & 1\\ x_1 & \cdots & x_n\end{pmatrix}\begin{pmatrix}1 & x_1\\ \vdots & \vdots\\ 1 & x_n\end{pmatrix} = \begin{pmatrix}n & \sum_{i=1}^n x_i\\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2\end{pmatrix}.$$

Comme

$$\begin{pmatrix}a & b\\ c & d\end{pmatrix}^{-1} = \frac{1}{ad-bc}\begin{pmatrix}d & -b\\ -c & a\end{pmatrix},$$

on peut inverser la matrice :

$$\begin{aligned}(\mathrm{X}_3^\top\mathrm{X}_3)^{-1} &= \frac{1}{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}\begin{pmatrix}\sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i\\ -\sum_{i=1}^n x_i & n\end{pmatrix}\\ &= \frac{1}{n^2\left(\frac{1}{n}\sum_{i=1}^n x_i^2 - \bar{x}^2\right)}\begin{pmatrix}\sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i\\ -\sum_{i=1}^n x_i & n\end{pmatrix}\\ &= \frac{1}{n\sum_{i=1}^n (x_i-\bar{x})^2}\begin{pmatrix}\sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i\\ -\sum_{i=1}^n x_i & n\end{pmatrix} = \frac{1}{n^2 s_x^2}\begin{pmatrix}\sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i\\ -\sum_{i=1}^n x_i & n\end{pmatrix}\\ &= \frac{1}{n^2 s_x^2}\begin{pmatrix}ns_x^2 + n\bar{x}^2 & -\bar{x}n\\ -\bar{x}n & n\end{pmatrix} = \frac{1}{ns_x^2}\begin{pmatrix}s_x^2+\bar{x}^2 & -\bar{x}\\ -\bar{x} & 1\end{pmatrix},\end{aligned}$$

où

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

$$(\mathrm{X}_3^\top\mathrm{X}_3)^{-1}\mathrm{X}_3^\top = \frac{1}{ns_x^2}\begin{pmatrix}s_x^2+\bar{x}^2 & -\bar{x}\\ -\bar{x} & 1\end{pmatrix}\begin{pmatrix}1 & \cdots & 1\\ x_1 & \cdots & x_n\end{pmatrix} = \frac{1}{ns_x^2}\begin{pmatrix}s_x^2+\bar{x}^2-x_1\bar{x} & \cdots & s_x^2+\bar{x}^2-x_n\bar{x}\\ -\bar{x}+x_1 & \cdots & -\bar{x}+x_n\end{pmatrix}.$$

$$\begin{aligned}\mathrm{P}_3 &= \mathrm{X}_3(\mathrm{X}_3^\top\mathrm{X}_3)^{-1}\mathrm{X}_3^\top = \begin{pmatrix}1 & x_1\\ \vdots & \vdots\\ 1 & x_n\end{pmatrix}\begin{pmatrix}s_x^2+\bar{x}^2-x_1\bar{x} & \cdots & s_x^2+\bar{x}^2-x_n\bar{x}\\ -\bar{x}+x_1 & \cdots & -\bar{x}+x_n\end{pmatrix}\frac{1}{ns_x^2}\\ &= \frac{1}{ns_x^2}\begin{pmatrix}s_x^2+\bar{x}^2-x_1\bar{x}-x_1\bar{x}+x_1^2 & \cdots & s_x^2+\bar{x}^2-x_n\bar{x}-x_1\bar{x}+x_1x_n\\ \vdots & \ddots & \vdots\\ s_x^2+\bar{x}^2-x_1\bar{x}-x_n\bar{x}+x_nx_1 & \cdots & s_x^2+\bar{x}^2-x_n\bar{x}-x_n\bar{x}+x_n^2\end{pmatrix}\\ &= \frac{1}{ns_x^2}\begin{pmatrix}s_x^2+(x_1-\bar{x})(x_1-\bar{x}) & \cdots & s_x^2+(x_1-\bar{x})(x_n-\bar{x})\\ \vdots & \ddots & \vdots\\ s_x^2+(x_n-\bar{x})(x_1-\bar{x}) & \cdots & s_x^2+(x_n-\bar{x})(x_n-\bar{x})\end{pmatrix}\\ &= \mathrm{P}_1 + \frac{1}{ns_x^2}(\mathrm{I}-\mathrm{P}_1)\mathrm{X}_2\mathrm{X}_2^\top(\mathrm{I}-\mathrm{P}_1).\end{aligned}$$

2. Construction des projecteurs qui projettent sur l'orthogonal des sous-espaces engendrés par les colonnes de X_1, X_2 et X_3 ($\mathrm{P}_\mathrm{X}^\perp = \mathrm{I} - \mathrm{X}(\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top$).

$$\mathrm{P}_1^\perp = \mathrm{I} - \mathrm{X}_1(\mathrm{X}_1^\top\mathrm{X}_1)^{-1}\mathrm{X}_1^\top = \begin{pmatrix}1-1/n & -1/n & \cdots & -1/n\\ -1/n & 1-1/n & \cdots & -1/n\\ \vdots & \vdots & \ddots & \vdots\\ -1/n & -1/n & \cdots & 1-1/n\end{pmatrix}.$$

$$\mathrm{P}_2^\perp = \mathrm{I}-\mathrm{P}_2 = \begin{pmatrix}1-\frac{x_1^2}{\sum_{i=1}^n x_i^2} & \cdots & -\frac{x_1x_n}{\sum_{i=1}^n x_i^2}\\ \vdots & \ddots & \vdots\\ -\frac{x_nx_1}{\sum_{i=1}^n x_i^2} & \cdots & 1-\frac{x_n^2}{\sum_{i=1}^n x_i^2}\end{pmatrix} = \frac{1}{\sum_{i=1}^n x_i^2}\begin{pmatrix}\sum_{i=1}^n x_i^2 - x_1^2 & \cdots & -x_1x_n\\ \vdots & \ddots & \vdots\\ -x_nx_1 & \cdots & \sum_{i=1}^n x_i^2 - x_n^2\end{pmatrix}.$$

$$
\begin{aligned}
\mathrm{P}_3^{\perp} &= \mathrm{I} - \mathrm{P}_3 \\
&= \frac{1}{ns_x^2}\begin{pmatrix} ns_x^2 - (s_x^2 + \bar{x}^2 - x_1\bar{x} - x_1\bar{x} + x_1^2) & \cdots & -(s_x^2 + \bar{x}^2 - x_n\bar{x} - x_1\bar{x} + x_1x_n) \\ \vdots & \ddots & \vdots \\ -(s_x^2 + \bar{x}^2 - x_1\bar{x} - x_n\bar{x} + x_nx_1) & \cdots & ns_x^2 - (s_x^2 + \bar{x}^2 - x_n\bar{x} - x_n\bar{x} + x_n^2) \end{pmatrix}.
\end{aligned}
$$

3. Vérifier que ces 4 projecteurs sont des matrices idempotentes c'est-à-dire : $\mathrm{PP} = \mathrm{P}$

$$
\begin{aligned}
\mathrm{P}_1\mathrm{P}_1 &= \begin{pmatrix} 1/n & \cdots & 1/n \\ \vdots & \ddots & \vdots \\ 1/n & \cdots & 1/n \end{pmatrix}\begin{pmatrix} 1/n & \cdots & 1/n \\ \vdots & \ddots & \vdots \\ 1/n & \cdots & 1/n \end{pmatrix} \\
&= \begin{pmatrix} \underbrace{1/n^2 + 1/n^2 + \cdots + 1/n^2}_{n \text{ termes}} & \cdots & \underbrace{1/n^2 + 1/n^2 + \cdots + 1/n^2}_{n \text{ termes}} \\ \vdots & \ddots & \vdots \\ \underbrace{1/n^2 + 1/n^2 + \cdots + 1/n^2}_{n \text{ termes}} & \cdots & \underbrace{1/n^2 + 1/n^2 + \cdots + 1/n^2}_{n \text{ termes}} \end{pmatrix} \\
&= \begin{pmatrix} \sum_{i=1}^n 1/n^2 & \cdots & \sum_{i=1}^n 1/n^2 \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^n 1/n^2 & \cdots & \sum_{i=1}^n 1/n^2 \end{pmatrix} = \begin{pmatrix} 1/n & \cdots & 1/n \\ \vdots & \ddots & \vdots \\ 1/n & \cdots & 1/n \end{pmatrix} = \mathrm{P}_1.
\end{aligned}
$$

$$
\mathrm{P}_1^{\perp}\mathrm{P}_1^{\perp} = \begin{pmatrix} 1-1/n & -1/n & \cdots & -1/n \\ -1/n & 1-1/n & \cdots & -1/n \\ \vdots & \vdots & \ddots & \vdots \\ -1/n & -1/n & \cdots & 1-1/n \end{pmatrix}\begin{pmatrix} 1-1/n & -1/n & \cdots & -1/n \\ -1/n & 1-1/n & \cdots & -1/n \\ \vdots & \vdots & \ddots & \vdots \\ -1/n & -1/n & \cdots & 1-1/n \end{pmatrix}.
$$

2 cas : 1) Termes diagonaux

$$
\begin{aligned}
d_i &= (1-1/n)^2 - \frac{1}{n}\left(-\frac{1}{n}\right) + \left(-\frac{1}{n}\right)\left(-\frac{1}{n}\right) + \cdots + -\frac{1}{n}\left(-\frac{1}{n}\right) \\
&= \left(\frac{n-1}{n}\right)^2 + \frac{1}{n^2} + \frac{1}{n^2} + \cdots + \frac{1}{n^2} = \left(\frac{n-1}{n}\right)^2 + \frac{n-1}{n^2} \\
&= \left(\frac{n^2 - 2n + 1 + n - 1}{n^2}\right) = \frac{n^2-n}{n^2} = 1 - \frac{1}{n} \text{ pour tout } i = 1, \ldots, n.
\end{aligned}
$$

2) Termes extra-diagonaux : pour tout $i \neq j$

$$
\begin{aligned}
e_{ij} &= 2\left(1 - \frac{1}{n}\right)\left(-\frac{1}{n}\right) + (n-2)\left(-\frac{1}{n}\right)\left(-\frac{1}{n}\right) = -\frac{1}{n}\left(2 - \frac{2}{n} - \frac{n-2}{n}\right) \\
&= -\frac{1}{n}\left(2 - \frac{2}{n} - 1 + \frac{2}{n}\right) = -\frac{1}{n}.
\end{aligned}
$$

Ainsi, $P_1^\perp P_1^\perp = P_1^\perp$.

$$\begin{aligned}
P_2P_2 &= \frac{1}{(\sum_{i=1}^n x_i^2)^2}\begin{pmatrix} x_1^2 & \cdots & x_1x_n \\ \vdots & \ddots & \vdots \\ x_1x_n & \cdots & x_n^2 \end{pmatrix}\begin{pmatrix} x_1^2 & \cdots & x_1x_n \\ \vdots & \ddots & \vdots \\ x_1x_n & \cdots & x_n^2 \end{pmatrix} \\
&= \frac{1}{(\sum_{i=1}^n x_i^2)^2}\begin{pmatrix} x_1^4+\cdots+x_1^2x_n^2 & \cdots & x_1^3x_n+\cdots+x_1x_n^3 \\ \vdots & \ddots & \vdots \\ x_1^3x_n+\cdots+x_1x_n^3 & \cdots & x_n^2x_1^2+\cdots+x_n^4 \end{pmatrix} \\
&= \frac{1}{(\sum_{i=1}^n x_i^2)^2}\begin{pmatrix} x_1^2\sum_{i=1}^n x_i^2 & \cdots & x_1x_n\sum_{i=1}^n x_i^2 \\ \vdots & \ddots & \vdots \\ x_1x_n\sum_{i=1}^n x_i^2 & \cdots & x_n^2\sum_{i=1}^n x_i^2 \end{pmatrix} \\
&= \frac{\sum_{i=1}^n x_i^2}{(\sum_{i=1}^n x_i^2)^2}\begin{pmatrix} x_1^2 & x_1x_2 & \cdots & x_1x_n \\ x_1x_2 & x_2^2 & \cdots & x_2x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_1x_n & x_2x_n & \cdots & x_n^2 \end{pmatrix} \\
&= \frac{1}{\sum_{i=1}^n x_i^2}\begin{pmatrix} x_1^2 & x_1x_2 & \cdots & x_1x_n \\ x_1x_2 & x_2^2 & \cdots & x_2x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_1x_n & x_2x_n & \cdots & x_n^2 \end{pmatrix} = P_2.
\end{aligned}$$

$$\begin{aligned}
P_2^\perp P_2^\perp &= \frac{1}{(\sum_{i=1}^n x_i^2)^2}\begin{pmatrix} \sum_{i=1}^n x_i^2 - x_1^2 & -x_2x_1 & \cdots & -x_1x_n \\ -x_2x_1 & \sum_{i=1}^n x_i^2 - x_2^2 & \cdots & -x_2x_n \\ \vdots & \vdots & \ddots & \vdots \\ -x_nx_1 & -x_nx_2 & \cdots & \sum_{i=1}^n x_i^2 - x_n^2 \end{pmatrix} \\
&\quad\times\begin{pmatrix} \sum_{i=1}^n x_i^2 - x_1^2 & -x_2x_1 & \cdots & -x_1x_n \\ -x_2x_1 & \sum_{i=1}^n x_i^2 - x_2^2 & \cdots & -x_2x_n \\ \vdots & \vdots & \ddots & \vdots \\ -x_nx_1 & -x_nx_2 & \cdots & \sum_{i=1}^n x_i^2 - x_n^2 \end{pmatrix}.
\end{aligned}$$

Le terme général de ce produit est pour la cellule (j,k) avec $j \neq k$:

$$x_jx_k\sum_{i=1}^n x_i^2 - x_jx_k\sum_{i=1}^n x_i^2 - x_kx_j\sum_{i=1}^n x_i^2 = -x_jx_k\sum_{i=1}^n x_i^2.$$

Sur la diagonale, on obtient

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

On a $P_2^\perp P_2^\perp = P_2^\perp$.

4. Projection du vecteur $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ à l'aide des projecteurs calculés ci-dessus :

$$\begin{aligned}
P_1y &= \begin{pmatrix} 1/n & \cdots & 1/n \\ \vdots & \ddots & \vdots \\ 1/n & \cdots & 1/n \end{pmatrix}\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} y_1/n+\cdots+y_n/n \\ y_1/n+\cdots+y_n/n \\ \vdots \\ y_1/n+\cdots+y_n/n \end{pmatrix} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n y_i \\ \vdots \\ \frac{1}{n}\sum_{i=1}^n y_i \end{pmatrix} = \begin{pmatrix} \bar{y} \\ \vdots \\ \bar{y} \end{pmatrix}.
\end{aligned}$$

La meilleure approximation au sens des moindres carrés de y lorsque seule une constante est utilisée

comme régresseur est la moyenne de y.

$$\begin{aligned}
\mathrm{P}_1^{\perp}\mathrm{y} &= \begin{pmatrix} 1-1/n & -1/n & \cdots & -1/n \\ -1/n & 1-1/n & \cdots & -1/n \\ \vdots & \vdots & \ddots & \vdots \\ -1/n & -1/n & \cdots & 1-1/n \end{pmatrix}\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} y_1 - \frac{1}{n}y_1 - \frac{1}{n}y_2 - \cdots - \frac{1}{n}y_n \\ -\frac{1}{n}y_1 + y_2 - \frac{1}{n}y_2 - \frac{1}{n}y_3 - \cdots - \frac{1}{n}y_n \\ \vdots \\ -\frac{1}{n}y_1 - \frac{1}{n}y_2 - \frac{1}{n}y_3 - \cdots - \frac{1}{n}y_n + y_n \end{pmatrix} \\
&= \begin{pmatrix} y_1 - \frac{1}{n}\sum_{i=1}^n y_i \\ y_2 - \frac{1}{n}\sum_{i=1}^n y_i \\ \vdots \\ y_n - \frac{1}{n}\sum_{i=1}^n y_i \end{pmatrix} = \begin{pmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{pmatrix}.
\end{aligned}$$

Cette projection consiste à centrer les données.

$$\begin{aligned}
\mathrm{P}_2\mathrm{y} &= \frac{1}{\sum_{i=1}^n x_i^2}\begin{pmatrix} x_1^2 & x_1x_2 & \cdots & x_1x_n \\ x_1x_2 & x_2^2 & \cdots & x_2x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_nx_1 & x_nx_2 & \cdots & x_n^2 \end{pmatrix}\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \\
&= \frac{1}{\sum_{i=1}^n x_i^2}\begin{pmatrix} x_1^2y_1 + x_1x_2y_2 + \cdots + x_1x_ny_n \\ x_1x_2y_1 + x_2^2y_2 + \cdots + x_2x_ny_n \\ \vdots \\ x_1x_ny_1 + x_2x_ny_2 + \cdots + x_n^2y_n \end{pmatrix} \\
&= \frac{1}{\sum_{i=1}^n x_i^2}\begin{pmatrix} x_1\sum_{i=1}^n x_iy_i \\ x_2\sum_{i=1}^n x_iy_i \\ \vdots \\ x_n\sum_{i=1}^n x_iy_i \end{pmatrix} = \frac{\sum_{i=1}^n x_iy_i}{\sum_{i=1}^n x_i^2}\begin{pmatrix} x_1 \\ \vdots \\ \vdots \\ x_n \end{pmatrix} = \frac{(s_{xy} + \bar{x}\bar{y})}{(s_x^2 + \bar{x}^2)}\mathrm{x}.
\end{aligned}$$

$$\begin{aligned}
\mathrm{P}_2^{\perp}\mathrm{y} &= \frac{1}{\sum_{i=1}^n x_i^2}\begin{pmatrix} \sum_{i=1}^n x_i^2 - x_1^2 & -x_2x_1 & \cdots & -x_1x_n \\ -x_2x_1 & \sum_{i=1}^n x_i^2 - x_2^2 & \cdots & -x_2x_n \\ \vdots & \vdots & \ddots & \vdots \\ -x_nx_1 & -x_nx_2 & \cdots & \sum_{i=1}^n x_i^2 - x_n^2 \end{pmatrix}\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \\
&= \frac{1}{\sum_{i=1}^n x_i^2}\begin{pmatrix} \sum_{i=1}^n x_i^2y_1 - x_1\sum_{i=1}^n x_iy_i \\ \sum_{i=1}^n x_i^2y_2 - x_2\sum_{i=1}^n x_iy_i \\ \vdots \\ \sum_{i=1}^n x_i^2y_n - x_n\sum_{i=1}^n x_iy_i \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} - \frac{\sum_{i=1}^n x_iy_i}{\sum_{i=1}^n x_i^2}\begin{pmatrix} x_1 \\ \vdots \\ \vdots \\ x_n \end{pmatrix}.
\end{aligned}$$

Chapitre 12

Exercices : Régression multivariée

Exercice 12.1. Prix, surface et années
Le Tableau 12.1 contient le prix de vente, la surface habitable et le nombre d'années passées depuis la construction de 20 appartements vendus par une agence immobilière en 1995.

TABLE 12.1 – Prix surface et nombre d'années d'appartements

Numéro	Le prix de vente (en milliers de francs)	Surface habitable (en m^2)	Le nombre d'années passées (en années)
i	y_i	x_{2i}	x_{3i}
1.	174.0	48	2
2.	294.4	55	2
3.	151.6	71	3
4.	298.9	82	0
5.	321.2	100	3
6.	428.3	85	1
7.	214.1	70	8
8.	196.3	73	9
9.	191.8	74	10
10.	174.0	66	7
11.	84.7	35	47
12.	151.6	53	18
13.	169.5	73	18
14.	89.2	39	63
15.	142.7	67	23
16.	98.1	48	64
17.	120.4	51	31
18.	240.9	61	5
19.	111.5	53	40
20.	102.6	54	78
Somme	3755.8	1258.0	432
Moyenne	187.790	62.9	21.6

On connaît également quelques résultats du calcul de régression, comme :

$$\mathrm{X}^\top\mathrm{X} = \begin{pmatrix} 20 & 1258 & 432 \\ 1258 & 84104 & 22424 \\ 432 & 22424 & 20442 \end{pmatrix}, \quad \mathrm{X}^\top\mathrm{y} = 1000 \times \begin{pmatrix} 3.7558 \\ 256.7435 \\ 51.7395 \end{pmatrix}$$

et

$$(\mathrm{X}^\top\mathrm{X})^{-1} = \frac{1}{1000}\begin{pmatrix} 1858.0338 & -24.4835 & -12.4084 \\ -24.4835 & 0.3394 & 0.1451 \\ -12.4084 & 0.1451 & 0.1520 \end{pmatrix},$$

où $\mathrm{x}_1 = (1, \ldots, 1)^\top$.

1. Estimer les coefficients de la régression (b_1, b_2, b_3).

2. Calculer la variance totale de la variable dépendante, la variance de régression et la variance des résidus.
3. Calculer la somme des carrés totale, la somme des carrés de régression et la somme des carrés des résidus.
4. Donner le cofficient de détermination et le coefficient de corrélation multiple.
5. Calculer la matrice variance-covariance et la matrice des corrélations des variables explicatives.

Solution

1. Coefficients de régression :

$$\mathrm{b} = (\mathrm{X}^\top \mathrm{X})^{-1}\mathrm{X}^\top \mathrm{y} = \begin{pmatrix} 50.419452 \\ 2.691016 \\ -1.485583 \end{pmatrix}^\top .$$

2. On peut ajouter les colonnes suivantes au tableau précédent :

	y_i	$(y_i-\bar{y})^2$	x_{i2}	x_{i3}	y_i^*	$y_i^*-\bar{y}$	$(y_i^*-\bar{y})^2$	e_i	e_i^2
	174.0	190.16	48	2	176.86	−10.98	120.53	−2.62	6.85
	294.4	11365.69	55	2	195.74	7.86	61.75	98.95	9790.28
	151.6	1309.72	71	3	237.39	49.43	2443.23	−85.42	7297.40
	298.9	12345.43	82	0	271.53	83.49	6970.08	27.82	773.80
	321.2	17798.23	100	3	315.58	127.47	16248.23	6.14	37.65
	428.3	57845.06	85	1	278.12	90.07	8113.41	150.63	22689.33
	214.1	692.22	70	8	227.23	39.31	1545.29	−12.81	163.99
	196.3	72.42	73	9	233.83	45.90	2106.59	−37.19	1383.35
	191.8	16.08	74	10	235.03	47.10	2218.70	−42.90	1840.31
	174.0	190.16	66	7	217.94	30.03	901.90	−43.63	1903.35
	84.7	10627.55	35	47	74.65	−112.81	12726.81	9.92	98.35
	151.6	1309.72	53	18	166.47	−21.29	453.39	−14.70	216.17
	169.5	334.52	73	18	220.40	32.53	1058.03	−50.62	2562.70
	89.2	9719.99	39	63	61.56	−125.82	15830.27	27.42	752.00
	142.7	2033.11	67	23	196.76	8.95	80.16	−53.85	2899.73
	98.1	8044.30	48	64	84.34	−103.08	10626.49	13.59	184.66
	120.4	4541.41	51	31	141.67	−45.99	2114.86	−21.21	449.79
	240.9	2820.67	61	5	207.44	19.55	382.11	33.76	1139.50
	111.5	5820.16	53	40	133.64	−53.98	2913.39	−22.12	489.29
	102.6	7257.34	54	78	79.62	−107.74	11607.24	22.74	517.16
Somme	3755.8	154333.94	1258	432	3755.80	−0.00	98522.46	3.88	55195.66
Moyenne	187.8	7716.70	63	22	187.79	−0.00	4926.12	0.19	2759.78

où

$$y_i^* = b_1 + b_2 x_{i,2} + b_3 x_{i,3}, \text{ et } e_i^* = y_i - y_i^*.$$

Donc,

$$s_y^2 = 7716.70, s_{y^*}^2 = 4926.12 \text{ et } s_e^2 = 2759.78.$$

3. Les sommes des carrés :

$$\mathrm{SC}_{TOT} = ns_y^2 = 154333.94,\ \mathrm{SC}_{REG} = ns_{y^*}^2 = 98522.46 \text{ et } \mathrm{SC}_{RES} = ns_e^2 = 55195.66.$$

4. Le coefficient de détermination et le coefficient de corrélation multiple :

$$R^2 = \frac{s_{y^*}^2}{s_y^2} = 64.112\% \text{ et } r = \sqrt{R^2} = 0.801$$

5.

$$s_2^2 = \frac{1}{20}\sum_{i=1}^{20}(x_{i,2}-\bar{x}_2)^2 = \frac{1}{20}\sum_{i=1}^{20}x_{i,2}^2 - \bar{x}_2^2 = \frac{84104}{20} - 62.9^2 = 248.790$$

$$s_3^2 = \frac{1}{20}\sum_{i=1}^{20}(x_{i,3}-\bar{x}_3)^2 = \frac{1}{20}\sum_{i=1}^{20}x_{i,3}^2 - \bar{x}_3^2 = \frac{20442}{20} - 21.6^2 = 555.540$$

$$s_{2,3} = s_{3,2} = \frac{1}{20}\sum_{i=1}^{20}(x_{i,2} - \bar{x}_2)(x_{i,3} - \bar{x}_3) = \frac{1}{20}\sum_{i=1}^{20} x_{i,2}x_{i,3} - \bar{x}_2\bar{x}_3 = \frac{22424}{20} - 62.9 * 21.6 = -237.440$$

donc

$$\mathbf{\Sigma} = \begin{pmatrix} s_2^2 & s_{2,3} \\ s_{3,2} & s_3^2 \end{pmatrix} = \begin{pmatrix} 248.790 & -237.440 \\ -237.440 & 555.540 \end{pmatrix}.$$

Comme

$$\mathrm{S} = \begin{pmatrix} s_2 & 0 \\ 0 & s_3 \end{pmatrix} = \begin{pmatrix} 15.7731 & 0 \\ 0 & 23.5699 \end{pmatrix}$$

on obtient

$$\mathrm{R} = \mathrm{S}^{-1}\mathbf{\Sigma}\mathrm{S}^{-1} = \begin{pmatrix} 1 & -0.6387 \\ -0.6387 & 1 \end{pmatrix}.$$

Exercice 12.2. Crèmes glacées
La consommation de crème glacée d'individus a été mesurée pendant 30 périodes. L'objectif est de déterminer si la consommation dépend du revenu et de la température. Les données sont dans le Tableau 12.2.

Table 12.2 – Consommation de crème glacée

consommation y	revenu x_2	température x_3
386	78	41
374	79	56
393	81	63
425	80	68
406	76	69
344	78	65
327	82	61
288	79	47
269	76	32
256	79	24
286	82	28
298	85	26
329	86	32
318	83	40
381	84	55
381	82	63
470	80	72
443	78	72
386	84	67
342	86	60
319	85	44
307	87	40
284	94	32
326	92	27
309	95	28
359	96	33
376	94	41
416	96	52
437	91	64
548	90	71

On sait que

$$\sum_{i=1}^{n} y_i = 10783, \quad \sum_{i=1}^{n} x_{i,2} = 2538, \quad \sum_{i=1}^{n} x_{i,3} = 1473,$$

$$\sum_{i=1}^{n} y_i^2 = 4001293, \quad \sum_{i=1}^{n} x_{i,2}^2 = 215846, \quad \sum_{i=1}^{n} x_{i,3}^2 = 80145,$$

$$\sum_{i=1}^{n} x_{i,2} y_i = 912813, \quad \sum_{i=1}^{n} x_{i,3} y_i = 553747, \quad \sum_{i=1}^{n} x_{i,2} x_{i,3} = 123650,$$

$$\left(\mathrm{X}^\top \mathrm{X}\right)^{-1} = \begin{pmatrix} \sum_{i=1}^{n} x_{i,2}^2 & \sum_{i=1}^{n} x_{i,2} x_{i,3} \\ \sum_{i=1}^{n} x_{i,2} x_{i,3} & \sum_{i=1}^{n} x_{i,3}^2 \end{pmatrix}^{-1} = \begin{pmatrix} 215846 & 123650 \\ 123650 & 80145 \end{pmatrix}^{-1} = \begin{pmatrix} 3.987998 & -6.152797 \\ -6.152797 & 10.74045 \end{pmatrix} \times \frac{1}{10\,0000},$$

et

$$\mathrm{X}^\top \mathrm{y} = \begin{pmatrix} \sum_{i=1}^{n} y_i x_{i,2} \\ \sum_{i=1}^{n} y_i x_{i,3} \end{pmatrix} = \begin{pmatrix} 912813 \\ 553747 \end{pmatrix}.$$

Considérons la relation possible entre x_2, x_3 et y décrite par la formule :

$$y_i = b_2 x_{i,2} + b_3 x_{i,3} + e_i,$$

Attention ! Ce modèle n'a pas de constante.

1. Estimer les coefficients de la régression (b_2, b_3).
2. Calculer la variance totale de la variable dépendante, la variance de régression et la variance des résidus.
3. Calculer la somme des carrés totale, la somme des carrés de régression et la somme des carrés des résidus.
4. Donner le cofficient de détermination et le coefficient de corrélation multiple.
5. Calculer la matrice variance-covariance et la matrice des corrélations des variables explicatives.
6. La somme des résidus de ce modèle est-elle nulle (réponse sans calcul)? Justifier en deux lignes (et toujours sans calcul).

Solution

1. Les coefficients sont :

$$\mathrm{b} = (\mathrm{X}^\top \mathrm{X})^{-1}\mathrm{X}^\top \mathrm{y} = \begin{pmatrix} 3.987998 & -6.152797 \\ -6.152797 & 10.74045 \end{pmatrix} \times \frac{1}{100000}\begin{pmatrix} 912813 \\ 553747 \end{pmatrix} = \begin{pmatrix} 2.332 \\ 3.311 \end{pmatrix}.$$

2. On peut ajouter les colonnes suivantes au Tableau précédent :

$$y_i^{*2} = (b_2 x_{i,2} + b_3 x_{i,3})^2, \text{ et } e_i^{*2} = (y_i - y_i^*)^2$$

et calculer la somme et la moyenne de ces colonnes :

$$\sum_{i=1}^{30} y_i^{*2} = 3961891.750 \text{ et } \frac{1}{30}\sum_{i=1}^{30} y_i^{*2} = 132063.060,$$

$$\sum_{i=1}^{30} e_i^2 = 38912.326 \text{ et } \frac{1}{30}\sum_{i=1}^{30} e_i^2 = 1297.077.$$

Donc,

$$s_y^2 = \frac{1}{30}\sum_{i=1}^{30}(y_i - \bar{y})^2 = \frac{1}{30}\sum_{i=1}^{30} y_i^2 - \bar{y}^2 = 4184.112,$$

$$s_{y^*}^2 = \frac{1}{30}\sum_{i=1}^{30}(y_i^* - \bar{y}^*)^2 = \frac{1}{30}\sum_{i=1}^{30} y_i^{*2} - \bar{y}^2 = 2565.783$$

et

$$s_e^2 = \frac{1}{30}\sum_{i=1}^{30}(y_i - y_i^*)^2 = \frac{1}{30}\sum_{i=1}^{30} e_i^2 = 1297.077.$$

3. Les sommes des carrés :

$$\mathrm{SC}_{TOT} = n s_y^2 = 125523.367,\ \mathrm{SC}_{REG} = n s_{y^*}^2 = 76973.501 \text{ et } \mathrm{SC}_{RES} = n s_e^2 = 38912.326.$$

4. Le coefficient de détermination et le coefficient de corrélation multiple : Comme la régression n'inclut pas une constante R^2 ne peut pas être calculé.
5. Variances et covariances :

$$s_2^2 = \frac{1}{30}\sum_{i=1}^{30}(x_{i,2} - \bar{x}_2)^2 = \frac{1}{30}\sum_{i=1}^{30} x_{i,2}^2 - \bar{x}_2^2 = \frac{215846}{30} - 84.6^2 = 37.707,$$

$$s_3^2 = \frac{1}{30}\sum_{i=1}^{30}(x_{i,3} - \bar{x}_3)^2 = \frac{1}{30}\sum_{i=1}^{30} x_{i,3}^2 - \bar{x}_3^2 = \frac{80145}{30} - 49.1^2 = 260.690,$$

$$s_{2,3} = s_{3,2} = \frac{1}{30}\sum_{i=1}^{30}(x_{i,2} - \bar{x}_2)(x_{i,3} - \bar{x}_3) = \frac{1}{30}\sum_{i=1}^{30} x_{i,2}x_{i,3} - \bar{x}_2\bar{x}_3 = \frac{123650}{30} - 84.6 \times 49.1 = -32.193.$$

Donc,

$$\mathbf{\Sigma} = \begin{pmatrix} s_2^2 & s_{2,3} \\ s_{3,2} & s_3^2 \end{pmatrix} = \begin{pmatrix} 37.707 & -32.193 \\ -32.193 & 260.690 \end{pmatrix}.$$

Comme

$$S = \begin{pmatrix} s_2 & 0 \\ 0 & s_3 \end{pmatrix} = \begin{pmatrix} 6.141 & 0 \\ 0 & 16.146 \end{pmatrix},$$

on obtient

$$R = S^{-1}\Sigma S^{-1} = \begin{pmatrix} 1 & -0.325 \\ -0.325 & 1 \end{pmatrix}.$$

6. Non, car le modèle n'a pas de constante, les résidus ne sont pas orthogonaux au vecteur de constantes.

Exercice 12.3. Crèmes glacées en R
Vérifier vos résultat de l'Exercice 12.2 en utilisant le logiciel R. Aide :
- cbind : fonction pour créer une matrice à l'aide de vecteurs,
- t(A) : pour transposer la matrice A,
- %*% : pour un produit matriciel,
- solve(A) : pour inverser la matrice A,
- matrix : fonction pour créer une matrice à partir d'un vecteur.

Solution

En langage R

```
# données
consommation=c(386.00,374.00,393.00,425.00,406.00,344.00,327.00,288.00,269.00,256.00,
      286.00,298.00,329.00,318.00,381.00,381.00,470.00,443.00,386.00,342.00,319.00,
      307.00,284.00,326.00,309.00,359.00,376.00,416.00,437.00,548.00)
revenu=c(78.00,79.00,81.00,80.00,76.00,78.00,82.00,79.00,76.00,79.00,82.00,85.00,86.00,
    83.00,84.00,82.00,80.00,78.00,84.00,86.00,85.00,87.00,94.00,92.00,95.00,96.00,
    94.00,96.00,91.00,90.00)
temperature=c(41.00,56.00,63.00,68.00,69.00,65.00,61.00,47.00,32.00,24.00,28.00,26.00,
     32.00,40.00,55.00,63.00,72.00,72.00,67.00,60.00,44.00,40.00,32.00,27.00,
     28.00,33.00,41.00,52.00,64.00,71.00)
# 2-statistiques
y=consommation
x2=revenu
x3=temperature
n=length(y)    # length donne la longueur d'un vecteur
#
x2Sum=sum(x2)  # somme des x2
x3Sum=sum(x3)  # somme des x3
ySum=sum(y)    # somme des y
#
x22Sum=sum(x2^2)   # somme des x2^2
x32Sum=sum(x3^2)   # somme des x3^2
y2Sum=sum(y^2)     # somme des y^2
#
x2ySum=sum(x2*y)   # somme des x2*y
x3ySum=sum(x3*y)   # somme des x3*y
x2x3Sum=sum(x2*x3) # somme des x2*x3
#
X=cbind(x2,x3)   # Matrice X
XX=t(X)%*%X      # Produit X'X
XX1=solve(XX)    # Inverse de X'X
Xy=t(X)%*%y      # Produit X'y
# 1- coefficients de la regression
b=XX1%*%Xy
regression<-lm(y~x2+x3-1)
# 2- variances
vajust=X%*%b    # valeurs ajustees y*=X*b
predict(regression)
#
residus=y-vajust   # residus y-y*
residuals(regression)
#
sy2=t(y-mean(y))%*%(y-mean(y))/n   # forme vectorielle: (y-yB)'(y-yB)/n
syE2=t(vajust-mean(y))%*%(vajust-mean(y))/n # forme vectorielle: (y*-yB)'(y*-yB)/n
se2=t(residus)%*%residus/n    # forme vectorielle: e'e/n
# 3- SC
SCtot=n*sy2    # SCtot
```

```
SCreg=n*syE2   # SCreg
SCres=n*se2    # SCres

# 5- matrice variance-covariance et matrice correlation multiple
s2=var(x2)*(n-1)/n     # variance x2
s23=cov(x2,x3)*(n-1)/n    # covariance x2,x3
s32=cov(x3,x2)*(n-1)/n    # covariance x3,x2
s3=var(x3)*(n-1)/n     # variance x2
sigma=matrix(c(s2,s23,s32,s3),ncol=2)     # matrice var-cov
s=matrix(c(sqrt(s2),0,0,sqrt(s3)),ncol=2) # matrice s
R=solve(s)%*%sigma%*%solve(s)        # matrice corr multiple
cor(x2,x3)
```

Chapitre 13

Exercices : Calcul des probabilités

Exercice 13.1. Tennis
Trois amis A, B et C font un tournoi de tennis. Chacun joue contre chacun. On admet que $\Pr(\text{A bat B}) = 0.7$, $\Pr(\text{B bat C}) = 0.6$ et $\Pr(\text{C bat A}) = 0.2$. Ces événements sont supposés être indépendants.

1. Calculer la probabilité que :
   (a) A gagne ses deux parties.
   (b) A gagne exactement une partie.
   (c) A ne gagne aucune partie.
   (d) chacun gagne et perd une partie (match nul).
   (e) B gagne sa partie contre A en sachant que la rencontre se solde par un match nul.
2. Les joueurs décident de parier de l'argent (A : 70 francs, B : 50 francs, C : 30 francs) et le gagnant de chaque partie reçoit 50 francs. Le jeu est-il financièrement intéressant pour A ?

Solution

1. (a) Probabilité que A gagne ses deux parties :

$$\Pr(\text{A bat B et A bat C}) = \Pr(\text{A bat B} \cap \text{A bat C}) = \Pr(\text{A bat B})\Pr(\text{A bat C}) = 0.7 \times 0.8 = 0.56$$

   (b) Probabilité que A gagne exactement une partie :

$$\begin{aligned}&\Pr(\text{A bat B mais bat pas C}) + \Pr(\text{A bat C mais pas B}) = \\ &= \Pr(\text{A bat B})\Pr(\text{A bat pas C}) + \Pr(\text{A bat C})\Pr(\text{A bat pas B}) \\ &= 0.7 \times 0.2 + 0.8 \times 0.3 = 0.38\end{aligned}$$

   (c) Probabilité que A ne gagne aucune partie :

$$\Pr(\text{B bat A et C bat A}) = \Pr(\text{B bat A} \cap \text{C bat A}) = \Pr(\text{B bat A})\Pr(\text{C bat A}) = 0.3 \times 0.2 = 0.06$$

   (d) Probabilité que chacun gagne et perd une partie (match nul) :

$$\begin{aligned}\Pr(\text{match nul}) &= \Pr(\text{A bat B, B bat C et C bat A}) + \Pr(\text{A bat C, C bat B et B bat A}) \\ &= \Pr(\text{A bat B})\Pr(\text{B bat C})\Pr(\text{C bat A}) + \Pr(\text{A bat C})\Pr(\text{C bat B})\Pr(\text{B bat A}) \\ &= 0.7 \times 0.6 \times 0.2 + 0.8 \times 0.4 \times 0.3 = 0.18\end{aligned}$$

   (e) Probabilité B que gagne sa partie contre A en sachant que la rencontre se solde par un match nul :

$$\begin{aligned}\Pr(\text{B bat A}|\text{match nul}) &= \frac{\Pr(\text{B bat A et match nul})}{\Pr(\text{match nul})} = \frac{\Pr(\text{B bat A} \cap \text{match nul})}{\Pr(\text{match nul})} \\ &= \frac{\Pr(\text{A bat C, C bat B et B bat A})}{\Pr(\text{match nul})} = \frac{0.8 \times 0.4 \times 0.3}{0.18} = \frac{0.096}{0.18} = \frac{96}{180} = \frac{8}{15}.\end{aligned}$$

2. Soit la variable aléatoire X = “gain de A” :

$$\Pr(X = 30) = \Pr(\text{A bat B et C}) = 0.56.$$

$$\Pr(X = -20) = \Pr(\text{A gagne une seule partie}) = 0.38.$$

$$\Pr(X = -70) = \Pr(\text{A ne gagne aucune partie}) = 0.06.$$

$$\mathrm{E}(X) = 30 \times 0.56 + (-20) \times 0.38 + (-70) \times 0.06 = 5 > 0.$$

L’espérance est positive donc le jeu est financièrement intéressant pour A.

Exercice 13.2. Cartes et As
On a un jeu de 52 cartes. On tire 2 cartes au hasard.

1. Quelle est la probabilité de tirer 2 As ?
2. Quelle est la probabilité d'obtenir au moins une main ayant 2 As sur 10 parties ?
3. Combien de parties faut-il jouer pour avoir plus d'une chance sur deux d'obtenir une main de 2 As ?

Solution

1. Probabilité de tirer 2 As :

$$\Pr(2As) = \frac{\text{nombre de possibilités de tirer deux As parmi 4}}{\text{divisé par le nombre de possibilités de tirer deux cartes parmi 52}}.$$

$$\Pr(2As) = \frac{C_2^4}{C_2^{52}} = \frac{\binom{4}{2}}{\binom{52}{2}} = 0.004525.$$

2. Probabilité d'obtenir au moins une main ayant 2 As sur 10 parties :
On a 10 essais identiques et indépendants. À chaque essai, on a deux résultats différents : succès ou échec. Chaque essai a la même probabilité de succès, $p = 0.004525$.
Soit la variable aléatoire X = "nombre de mains ayant 2 As en 10 parties". La variable X suit une loi binomiale $B(10, 0.004525)$. La probabilité d'obtenir au moins une fois deux As sur 10 parties correspond à :

$$\Pr(X > 0) = 1 - \Pr(X = 0) = 1 - (1 - 0.004525)^{10} = 0.0443.$$

Même sur 10 parties il y a peu de chance d'obtenir une main de 2 As.

3. Nombre de parties à jouer pour avoir plus d'une chance sur deux d'obtenir une main de 2 As :
La variable aléatoire X = "le nombre de main ayant deux As sur n parties". n est inconnue. La variable X suit une variable aléatoire binomiale $B(n, p = 0.004525)$. La probabilité d'obtenir au moins une fois deux As, sur n parties correspond à :

$$\Pr(X > 0) = 1 - \Pr(X = 0) = 1 - (1 - 0.004525)^n.$$

On veut que :

$$\begin{aligned} 1 - (1 - 0.004525)^n &> 0.5 \\ 0.5 &> (1 - 0.004525)^n \\ \log(0.5) &> n \times \log(1 - 0.004525) \\ n &> \frac{\log(0.5)}{\log(1 - 0.004525)} \\ n &> 152.8. \end{aligned}$$

Il faudra jouer au moins 153 parties pour avoir plus d'une chance sur deux d'obtenir une main de 2 As.

Exercice 13.3. Défauts et probabilités
Une menuiserie possède 100 fenêtres en stock. Par expérience on sait que parmi les premières 50 fenêtres achetées à un commerçant A, il y a 60% de produits sans défauts et 40 % avec des petits défauts. La qualité des 50 autres fenêtres, provenant d'un commerçant B, est un peu meilleure, 80 % sont sans défauts. On choisit au hasard une caisse de 10 fenêtres. (Dans une caisse, les fenêtres sont toutes du même commerçant.)

1. Quelle est la probabilité que parmi ces 10 fenêtres, elles soient toutes sans défaut ?
2. On regarde les dix fenêtres choisies et on constate que toutes sont sans défaut. Quelle est la probabilité que cette caisse de dix fenêtres ait été achetée au commerçant A ?

Solution

1. Soit X, le nombre de fenêtres défectueuses dans l'échantillon et soit $Y = 0$ si la caisse choisie était achetée à ce commerçant A et soit $Y = 1$ sinon.
   Comme la moitié des caisses étaient achetées au commerçant A on peut dire

   $$\Pr(Y = 0) = \Pr(Y = 1) = 0.5$$

   La question est la probabilité, $\Pr(X= 0)$. Selon le théorème des probabilités totales :

   $$\Pr(X = 0) = \Pr(X = 0|Y = 0)\Pr(Y = 0) + \Pr(X = 0|Y = 1)\Pr(Y = 1) = 0.5 \times \frac{\binom{30}{10} + \binom{40}{10}}{\binom{50}{10}} = 0.0427.$$

2. En effet, si on sait que dans la caisse choisie toutes les fenêtres sont sans défauts la question est : cette caisse a-t-elle été achetée au commerçant A, qui produit 60% de fenêtres sans défauts, ou à quelqu'un d'autre mais dans ce cas les 80% des fenêtres sont sans défauts ? Soient X et Y les variables aléatoires définies au point précédent. Alors ona :

   $$\Pr(X = 0|Y = 0) = \frac{\binom{30}{10}}{\binom{50}{10}},$$

   $$\Pr(X = 0|Y = 1) = \frac{\binom{40}{10}}{\binom{50}{10}}.$$

   On obtient donc :

   $$\begin{aligned} \Pr(Y = 0|X = 0) &= \frac{\Pr(X = 0|Y = 0)\Pr(Y = 0)}{\Pr(X = 0|Y = 0)\Pr(Y = 0) + \Pr(X = 0|Y = 1)\Pr(Y = 1)} \\ &= \frac{0.5 \times \frac{\binom{30}{10}}{\binom{50}{10}}}{0.5 \times \frac{\binom{30}{10}}{\binom{50}{10}} + 0.5 \times \frac{\binom{40}{10}}{\binom{50}{10}}} = 0.034. \end{aligned}$$

Exercice 13.4. Sondage
Dans une ville, on évalue à 20% le nombre d'individus qui approuvent la politique économique du président, les 80% restant s'y opposent.

1. Considérons le cas d'un échantillon choisi avec remise :
   (a) Quelle est la probabilité que parmi 8 personnes choisies au hasard, 3 exactement approuvent la politique économique ?
   (b) Quelle est la probabilité que parmi 8 personnes choisies au hasard, un nombre inférieur ou égal à 3 personnes approuvent la politique économique ?
   (c) Donner les expressions de l'espérance et de la variance de la loi de probabilité utilisée.
   (d) Calculer les valeurs de l'espérance et de la variance.
2. Une rencontre organisée par les opposants a réuni 10% des opposants et 1% des individus favorables. Déterminer la probabilité qu'un participant à la rencontre, choisi au hasard, soit un opposant.

Solution

1. (a) X : nombre de personnes approuvant la politique économique du président. X est une variable aléatoire qui suit la loi binomiale : $X \sim \mathcal{B}(8, 0.2)$ avec

$$n = 8,$$

et

$$p = \frac{20}{100} = 0.2.$$

Donc,

$$q = 1 - p = 0.8.$$

On peut calculer :

$$\Pr(X = 3) = \binom{8}{3}(0.2)^3(0.8)^5 = 0.1468.$$

(b) On a :

$$\begin{aligned}
\Pr(X \leq 3) &= \Pr(X = 0) + \Pr(X = 1) + \Pr(X = 2) + \Pr(X = 3) \\
&= \binom{8}{0}(0.2)^0(0.8)^8 + \binom{8}{1}(0.2)^1(0.8)^7 + \binom{8}{2}(0.2)^2(0.8)^6 + \binom{8}{3}(0.2)^3(0.8)^5 \\
&= 0.167 + 0.1468 + 0.2936 + 0.3355 \\
&= 0.944.
\end{aligned}$$

(c) La loi de probabilité utilisée est la loi binomiale. Donc,

$$\mathrm{E}(X) = \sum_{x=1}^{n} x\Pr(x) = np,$$

$$\mathrm{var}(X) = \mathrm{E}\left[(X - \mathrm{E}(X))^2\right] = np(1-p) = npq.$$

(d)

$$\mathrm{E}(X) = np = 8 \times 0.2 = 1.6.$$

$$\mathrm{var}(X) = np(1-p) = 8 \times 0.2 \times 0.8 = 1.28.$$

2. Un meeting a réuni 10% des opposants et 1% des individus favorables.
Soient les événements suivants : M : participer au meeting, F : favorable, O : Opposant. On a

$$\Pr(M|F) = 0.01, \Pr(M|O) = 0.1.$$

Donc,

$$\Pr(O|M) = \frac{\Pr(O \cap M)}{\Pr(M)} = \frac{\Pr(M|O)\Pr(O)}{\Pr(M|F)\Pr(F) + \Pr(M|O)\Pr(O)} = \frac{0.1 \times 0.8}{0.01 \times 0.2 + 0.1 \times 0.8} = 0.97.$$

En effet,

$$\Pr(M) = \Pr(M|F)\Pr(F) + \Pr(M|O)\Pr(O).$$

Exercice 13.5. Densité exponentielle
Soit une variable aléatoire x dont la fonction de densité est donnée par l'équation suivante (distribution exponentielle), où λ est positif :

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0. \end{cases}$$

1. Calculer $F(a) = \Pr(X \leq a)$.
2. Montrer que $f(x)$ est une fonction de densité.
3. Calculer $\mathrm{E}(X)$.
4. Calculer $\mathrm{var}(X)$.

Solution

1. Pour $a > 0$,

$$\Pr(X \leq a) = F(a) = \int_{-\infty}^{a} f(x)dx = \int_{0}^{a} \lambda e^{-\lambda x}dx = \left[-e^{-\lambda x}\right]_0^a = -e^{-\lambda a} - (-e^{0\lambda}) = 1 - e^{-a\lambda}.$$

2. C'est une fonction de densité car $\displaystyle\int_{-\infty}^{\infty} f(x)dx = [-e^{-\lambda x}]_0^{\infty} = 1$ et $f(x) \geq 0$.
3. Calcul de l'espérance : Nous allons utiliser l'intégration par partie. Pour rappel :

$$\int_a^b u(x)v'(x)\,dx = \Big[u(x)v(x)\Big]_a^b - \int_a^b u'(x)v(x)\,dx = u(b)v(b) - u(a)v(a) - \int_a^b u'(x)v(x)\,dx.$$

On doit calculer :

$$\mathrm{E}(X) = \int_0^{\infty} x\lambda e^{-\lambda x}dx.$$

On pose $u(x) = x$ et $v(x) = -e^{-\lambda x}$. Donc, $u'(x) = 1$ et $v'(x) = f(x) = \lambda e^{-\lambda x}$. On a donc :

$$\mathrm{E}(X) = \int_0^{\infty} x\lambda e^{-\lambda x}dx = [u(x)v(x)]_0^{\infty} - \int_0^{\infty} u'(x)v(x)\,dx = \left[-xe^{-\lambda x}\right]_0^{\infty} + \int_0^{\infty} e^{-\lambda x}\,dx = 0 + \left[\frac{e^{-\lambda x}}{\lambda}\right]_0^{\infty} = \frac{1}{\lambda}.$$

4. Calcul de la variance. Nous allons d'abord calculer $\mathrm{E}(X^2)$. À nouveau, nous allons utiliser l'intégration par partie. On doit calculer :

$$\mathrm{E}(X) = \int_0^{\infty} x^2\lambda e^{-\lambda x}dx.$$

On pose $u(x) = x^2$ et $v(x) = -e^{-\lambda x}$. Donc, $u'(x) = 2x$ et $v'(x) = f(x) = \lambda e^{-\lambda x}$. On a donc :

$$\begin{aligned}\mathrm{E}(X) &= \int_0^{\infty} x^2\lambda e^{-\lambda x}dx = [u(x)v(x)]_0^{\infty} - \int_0^{\infty} u'(x)v(x)\,dx = \left[-x^2e^{-\lambda x}\right]_0^{\infty} + \int_0^{\infty} 2xe^{-\lambda x}\,dx \\ &= 0 + \frac{2}{\lambda}\int_0^{\infty} \lambda x e^{-\lambda x}\,dx = \frac{2}{\lambda^2}.\end{aligned}$$

Donc,

$$\mathrm{var}(X) = \mathrm{E}(X^2) - \mathrm{E}^2(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

Exercice 13.6. Étude de densité
Soit une variable aléatoire x dont la fonction de densité est donnée par l'équation suivante,

$$f(x) = \begin{cases} c(1-x^2) & \text{si } -1 < x < 1 \\ 0 & \text{si } x \notin (-1,1). \end{cases}$$

1. Calculer c.
2. Calculer $F(a) = \Pr(X \leq a)$.

Solution

1. Comme

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-1}^{1} c(1-x^2)dx = \left[cx - \frac{cx^3}{3}\right]_{-1}^{1} = c - \frac{c}{3} - (-c + \frac{c}{3}) = \frac{4c}{3} = 1.$$

   On doit prendre $c = 3/4$.
2. Pour $-1 < a < 1$,

$$\Pr(X \leq a) = \int_{-\infty}^{a} c(1-x^2)dx = \left[cx - \frac{cx^3}{3}\right]_{-1}^{a} = \frac{3a}{4} - \frac{a^3}{4} + \frac{1}{2}.$$

Exercice 13.7. Probabilités dans une loi normale
Utiliser le Tableau 22.2 à la page 310.

1. Supposons que $X \sim \mathcal{N}(0,1)$, trouver $\Pr(-1 < X < 1)$.
2. Supposons que $X \sim \mathcal{N}(5, \sigma^2 = 10)$, trouver $\Pr(X < 15)$. Noter que si $X \sim \mathcal{N}(\mu, \sigma^2)$, alors $\frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$.

Solution

1. $\Pr(-1 < X < 1) = \Pr(X < 1) - \Pr(X < -1) = 0.6826$.
2. Si, $Z = \dfrac{X-5}{\sqrt{10}} \sim \mathcal{N}(0,1)$, $\Pr(X < 15) = \Pr\left(Z < \dfrac{15-5}{\sqrt{10}}\right) = \Pr\left(Z < \sqrt{10}\right) = \Pr(Z < 3.162278) = 0.9992$.

Exercice 13.8. Densité jointe
Soit la distribution jointe de deux variables aléatoires (X et Y) qui est définie par :

$$f(x,y) = \begin{cases} \frac{6}{5}(x+y^2) & \text{si } 0 < x < 1 \text{ et } 0 < y < 1 \\ 0 & \text{sinon.} \end{cases} \tag{13.1}$$

1. Donner les distributions marginales.
2. Donner le coefficient de corrélation.

Solution

1. Les distributions marginales :

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

(a)

$$f(x) = \int_{-\infty}^{\infty} \frac{6}{5}(x+y^2)dy = \int_0^1 \frac{6}{5}(x+y^2)dy = \frac{6}{5}\left(\int_0^1 x dy + \int_0^1 y^2 dy\right) = \frac{6}{5}\left[xy+\frac{y^3}{3}\right]_0^1 = \frac{6}{5}\left(x+\frac{1}{3}\right),$$

si $0 < x < 1$,

(b) et

$$f(y) = \int_0^1 \frac{6}{5}(x+y^2)dx = \frac{6}{5}\left(y^2+\frac{1}{2}\right),$$

si $0 < y < 1$.

2. La corrélation vaut :

$$\rho_{xy} = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\text{var}(Y)}} = \frac{\text{E}(XY) - \text{E}(X)\text{E}(Y)}{\sqrt{[\text{E}(X^2) - \text{E}^2(X)][\text{E}(Y^2) - \text{E}^2(Y)]}}.$$

Donc, on a besoin de calculer : $\text{E}(X), \text{E}(Y), \text{E}(X^2), \text{E}(Y^2),$ et $\text{E}(XY)$.

$$\text{E}(X) = \int_{-\infty}^{\infty} xf(x)dx = \frac{3}{5}, \quad \text{E}(Y) = \int_{-\infty}^{\infty} yf(y)dy = \frac{3}{5},$$

$$\text{E}(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \frac{13}{30}, \quad \text{E}(Y^2) = \int_{-\infty}^{\infty} y^2 f(y)dy = \frac{11}{25}, \quad \text{E}(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dxdy = \frac{7}{20}.$$

Ainsi,

$$\rho_{xy} = \frac{\frac{7}{20} - \frac{3}{5}\frac{3}{5}}{\sqrt{\left(\frac{13}{30} - \frac{9}{25}\right)\left(\frac{11}{25} - \frac{9}{25}\right)}} = -0.1305.$$

Exercice 13.9. Distribution de Poisson
Dans une usine pour fabriquer un produit on utilise des plaques en inox. Ces plaques sont découpées en 25 parties (pièces) identiques et le déchet est négligable. Mais sur les plaques il y a des petits défauts, dont le nombre est une variable aléatoire qui suit une distribution de Poisson avec $\lambda = 3.5$. Combien de plaques l'usine doit-elle commander si elle veut fabriquer 500 000 pièces sans défauts ?

Solution
Comme une plaque est coupée en 25 parties identiques et qu'il n'y a pas de déchets, le nombre de défauts par partie est

$$\lambda = \frac{3.5}{25} = 0.14$$

en moyenne.
La probabilité qu'une partie soit sans défauts est :

$$\Pr(x = 0) = e^{-0.14} = 0.8694,$$

alors le nombre de pièces sans défauts est de 87%Àinsi, pour que l'usine puisse fabriquer 500 000 pièces sans défauts, il faut commander à peu près 23 000 plaques.

Exercice 13.10. Densité jointe et conditionnement
Soit la distribution jointe de deux variables aléatoires donnée par

$$f(x,y) = \begin{cases} xe^{-x(1+y)} & \text{si } x \geq 0 \text{ et } y \geq 0 \\ 0 & \text{sinon.} \end{cases}$$

1. Donner les densités marginales $f_X(x)$ et $f_Y(y)$.
2. Donner $\mathrm{E}(X)$ et $\mathrm{var}(X)$.
3. Donner la densité conditionnelle $f(y|x)$.
4. Donner l'espérance conditionnelle $\mathrm{E}(Y|X=x)$ et la variance conditionnelle $\mathrm{var}(Y|X=x)$.
5. X et Y sont-elles indépendantes ?

Solution

1. Les distributions marginales :

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

   (a) On doit calculer :

$$f_Y(y) = \int_{-\infty}^{\infty} x \times e^{-x(1+y)}dx.$$

   Pour réaliser une intégration par bloc, on pose :

$$u(x) = x, v'(x) = -\frac{1}{1+y}e^{-x(1+y)}, u'(x) = 1, v(x) = e^{-x(1+y)}.$$

   On a :

$$\begin{aligned} f_Y(y) &= \left[-\frac{x}{1+y}e^{-x(1+y)}\right]_{x=0}^{x=\infty} + \frac{1}{1+y}\int_0^{\infty} e^{-x(1+y)}dx \\ &= (0-0) + \frac{-1}{(1+y)^2} \times \left[e^{-x(1+y)}\right]_{x=0}^{x=\infty} = \frac{1}{(1+y)^2}. \end{aligned}$$

   Donc,

$$f_Y(y) = \begin{cases} \frac{1}{(1+y)^2} & \text{si } y \geq 0 \\ 0 & \text{sinon.} \end{cases}$$

   (b) La densité marginale de X :

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} x \times e^{-x(1+y)}dy = x\int_0^{\infty} e^{-x(1+y)}dy = x\int_0^{\infty} e^{-x} \times e^{-x\times y}dy \\ &= xe^{-x}\left[-\frac{1}{x}e^{-x\times y}\right]_{y=0}^{y=\infty} = xe^{-x}\left(0+\frac{1}{x}\right) = e^{-x}. \end{aligned}$$

   Donc,

$$f_X(x) = \begin{cases} e^{-x} & \text{si } x \geq 0 \\ 0 & \text{sinon.} \end{cases}$$

   La variable X a une distribution exponentielle de paramètre $\lambda = 1$.

2. On a :

$$\mathrm{E}(X) = 1 \text{ et } \mathrm{var}(X) = 1,$$

   car $X \sim \exp(1)$. En effet, si $X \sim \exp(\lambda)$ alors $\mathrm{E}(X) = \frac{1}{\lambda}$ et $\mathrm{var}(X) = \frac{1}{\lambda^2}$. On peut aussi calculer :

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} xf_X(x)dx = \int_0^{\infty} xe^{-x}dx.$$

   Pour réaliser une intégration par bloc, on pose :

$$u(x) = x, v'(x) = e^{-x}, u'(x) = 1, v(x) = -e^{-x}.$$

Donc,

$$\mathrm{E}(X) = [-xe^{-x}]_0^\infty + \int_0^\infty e^{-x}dx = 0 + [-e^{-x}]_0^\infty = 1.$$

Calcul de la variance :

$$\mathrm{var}(X) \quad = \quad \mathrm{E}(X^2) - \mathrm{E}(X)^2 = \int_{-\infty}^{\infty} x^2 f_X(x)dx - 1^2 = \int_0^\infty x^2 e^{-x}dx - 1.$$

Pour réaliser une intégration par bloc, on pose :

$$u(x) = x^2, v'(x) = e^{-x}, u'(x) = 2x, v(x) = -e^{-x}.$$

Donc,

$$\mathrm{var}(X) \quad = \quad [-2xe^{-x}]_0^\infty + 2\int_0^\infty xe^{-x}dx - 1 = 2 - 1 = 1.$$

3. La densité conditionnelle :

$$f(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{xe^{-x}e^{-xy}}{e^{-x}} = xe^{-xy}, \text{ pour } x, y \geq 0.$$

$$f(y|x) = \begin{cases} xe^{-xy} & \text{si } x, y \geq 0 \\ 0 & \text{sinon.} \end{cases}$$

4. On peut maintenant calculer :

$$\mathrm{E}(Y|X = x) \quad = \quad \int_{-\infty}^{\infty} yf(y|x)dy = x\int_0^\infty ye^{-xy}dy.$$

Pour réaliser une intégration par bloc, on pose :

$$u(y) = y, v'(y) = e^{-xy}, u'(y) = 1, v(y) = \frac{-1}{x}e^{-xy}.$$

On a :

$$\mathrm{E}(Y|X = x) = x\left(\left[\frac{-y}{x}e^{-xy}\right]_{y=0}^{y=\infty} + \frac{1}{x}\int_0^\infty e^{-xy}dy\right) = x\left(\frac{1}{x}\left[\frac{-1}{x}e^{-xy}\right]_{y=0}^{y=\infty}\right) = x\frac{1}{x^2} = \frac{1}{x}.$$

Donc,

$$\mathrm{E}(Y|X = x) = \begin{cases} 1/x & \text{si } x > 0 \\ 0 & \text{sinon.} \end{cases}$$

Pour la variance, on a :

$$\mathrm{var}(Y|X = x) = \mathrm{E}(Y^2|X = x) - \mathrm{E}(Y|X = x)^2 = \int_0^\infty y^2 e^{-xy}dy - \frac{1}{x^2}.$$

Pour réaliser une intégration par bloc, on pose :

$$u(y) = y^2, v'(y) = e^{-xy}, u'(y) = 2y, v(y) = \frac{-1}{x}e^{-xy}.$$

$$\mathrm{var}(Y|X = x) = \left[\frac{-y^2}{x}e^{-xy}\right]_{y=0}^{y=\infty} + \frac{2}{x}\int_0^\infty ye^{-xy}dy - \frac{1}{x^2} = 0 + \frac{2}{x}\frac{1}{x^2} - \frac{1}{x^2} = \frac{2}{x^3} - \frac{1}{x^2}, \text{ pour } x > 0.$$

Donc,

$$\mathrm{var}(Y|X = x) = \begin{cases} \frac{2}{x^3} - \frac{1}{x^2} & \text{si} x > 0 \\ 0 & \text{sinon.} \end{cases}$$

5. Les variables X et Y ne sont pas indépendantes. En effet, $f_X(x)f_Y(y) \neq f(x,y)$ pour tout x, y.

Exercice 13.11. Densité jointe normale
Soit la distribution jointe de deux variables aléatoires (X et Y) donnée par :

$$f(x,y) = \frac{1}{\pi\sqrt{3}} \exp\left\{-\frac{2}{3}(x^2 - x(y-1) + (y-1)^2)\right\}.$$

1. Donner la densité marginale $f_Y(y)$. On trouve pour la densité marginale de X,
$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp -\frac{1}{2}x^2.$$
2. Donner $\mathrm{E}(X)$, $\mathrm{var}(X)$, $\mathrm{E}(Y)$ et $\mathrm{var}(Y)$.
3. Donner la densité conditionnelle $f(x|y)$.
4. Calculer $\mathrm{E}(X|Y=y)$ et $\mathrm{var}(X|Y=y)$.
5. On sait que le coefficient de corrélation des variables vaut $1/2$. Montrer que (X,Y) a une distribution normale bivariée.
6. Calculer $\mathrm{E}(X-3Y)$, et $\mathrm{var}(X-3Y)$. Aide : on sait que $\mathrm{cov}(X,Y) = 1/2$.

Rappelons que :

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

Solution

1. La distribution marginale de $f_Y(y)$ est :
$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f(x,y)dx = \frac{1}{\pi\sqrt{3}} \int_{-\infty}^{\infty} \exp\left\{-\frac{2}{3}(x^2 - x(y-1) + (y-1)^2)\right\} dx \\ &= \frac{1}{\pi\sqrt{3}} \int_{-\infty}^{\infty} \exp\left\{-\frac{2}{3}\left[x^2 - x(y-1) + \frac{(y-1)^2}{4}\right]\right\} \exp\left\{-\frac{2}{3}\frac{3(y-1)^2}{4}\right\} dx \\ &= \frac{1}{\pi\sqrt{3}} \exp -\frac{1}{2}(y-1)^2 \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2\frac{3}{4}}[x-(y-1)]^2\right\} dx \\ &= \frac{1}{\pi\sqrt{3}} \exp -\frac{1}{2}(y-1)^2 \sqrt{2\pi\frac{3}{4}} \int_{-\infty}^{\infty} dx = \frac{1}{\sqrt{2\pi}} \exp -\frac{1}{2}(y-1)^2. \end{aligned}$$
2. On a :
$$X \sim \mathcal{N}(0,1) \text{ donc } \mathrm{E}(X) = 0 \text{ et } \mathrm{var}(X) = 1,$$
$$Y \sim \mathcal{N}(1,1) \text{ donc } \mathrm{E}(Y) = 1 \text{ et } \mathrm{var}(Y) = 1.$$
3. La densité conditionnelle :
$$\begin{aligned} f(x|y) &= \frac{f(x,y)}{f_Y(y)} = \frac{\frac{1}{\pi\sqrt{3}} \exp -\frac{2}{3}\left[x^2 - x(y-1) + (y-1)^2\right]}{\frac{1}{\sqrt{2\pi}} \exp -\frac{1}{2}(y-1)^2} \\ &= \frac{1}{\sqrt{2\pi\frac{3}{4}}} \exp -\frac{2}{3}\left[x^2 - x(y-1) + \frac{(y-1)^2}{4}\right] = \frac{1}{\sqrt{2\pi\frac{3}{4}}} \exp -\frac{1}{2\frac{3}{4}}\left[x - \frac{(y-1)}{2}\right]^2, \end{aligned}$$
pour $x, y \geq 0$, ce qui est la densité d'une normale $\mathcal{N}((y-1)/2, 3/4)$.
4. Donc, $\mathrm{E}(X|Y=y) = (y-1)/2$ et $\mathrm{var}(X|Y=y) = 3/4$.
5. On a :
$$X \sim \mathcal{N}(\mu_x, \sigma_x^2) = \mathcal{N}(0,1) \text{ donc } \sigma_x = 1,$$
$$Y \sim \mathcal{N}(\mu_y, \sigma_y^2) = \mathcal{N}(1,1) \text{ donc } \sigma_y = 1.$$
Les variables X et Y suivent une loi normale bivariée si leur densité jointe est :
$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} - \frac{2\rho(x-\mu_x)(y-\mu_y}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right]\right\}.$$

Dans notre cas, $\sigma_x = \sigma_y = 1, \mu_x = 0, \mu_y = 1$ et $\rho = \frac{1}{2}$ on a :

$$f(y,x) = \frac{1}{2\pi\sqrt{\frac{3}{4}}} \exp -\frac{1}{2\frac{3}{4}}(x^2 - x(y-1) + y^2) = \frac{1}{\pi\sqrt{3}} \exp -\frac{2}{3}(x^2 - x(y-1) + y^2).$$

Ce qui correspond à la densité jointe de nos variables. Ainsi, X et Y suivent une loi normale bivariée.

6. Soit : $\mathrm{Z} = \begin{pmatrix} X & Y \end{pmatrix}^\top$. On sait que, pour toute matrice non aléatoire A et vecteur aléatoire Z de dimension égale aux nombre de colonnes de A,

$$\mathrm{E}(\mathrm{AZ}) = \mathrm{AE}(\mathrm{Z}) \text{ et } \mathrm{var}(\mathrm{AZ}) = \mathrm{A}\boldsymbol{\Sigma}\mathrm{A}^\top,$$

où $\boldsymbol{\Sigma}$ est la matrice de variances/covariances. On s'intéresse ici à : $\mathrm{Z} = (X,Y)^\top$ et $\mathrm{A} = (1,-3)$.

$$\mathrm{AZ} = (1,-3)\begin{pmatrix} X \\ Y \end{pmatrix} = X - 3Y.$$

Donc,

$$\mathrm{E}(X - 3Y) = \mathrm{AE}\begin{pmatrix} X \\ Y \end{pmatrix} = \mathrm{A}\begin{pmatrix} \mathrm{E}(X) \\ \mathrm{E}(Y) \end{pmatrix} = \begin{pmatrix} 1 & -3 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = 0 - 3 = -3.$$

De plus, on a,

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}.$$

Donc,

$$\mathrm{var}(X - 3Y) = \mathrm{var}(\mathrm{AZ}) = \mathrm{A}\boldsymbol{\Sigma}\mathrm{A}^\top = \begin{pmatrix} 1 & -3 \end{pmatrix}\begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}\begin{pmatrix} 1 \\ -3 \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} & -\frac{5}{2} \end{pmatrix}\begin{pmatrix} 1 \\ -3 \end{pmatrix} = -\frac{1}{2} + \frac{15}{2} = \frac{14}{2} = 7.$$

Exercice 13.12. Binomiale
Dans un pays donné on sait qu'il y a en moyenne 516 garçons et 484 filles parmi 1000 nouveau-nés. Quelle est la probabilité que dans une famille de 6 enfants il y a au moins autant de garçons que de filles ?

Solution
Soit X représentant le nombre de garçons dans une famille de 6 enfants. Donc, X suit une distribution binomiale avec des paramètres $n = 6$ et $p = 0.516$.

$$\Pr(X \geq 3) = \sum_{i=3}^{6} \binom{6}{k} 0.516^k \times 0.484^{6-k} = 0.6858.$$

Exercice 13.13. Pièces et probabilités
On a deux pièces de monnaie. Une a, comme d'habitude, deux côté differents, pile et face. L'autre pièce est un peu particulière car elle a les deux côtés faces. On prend une de ces deux pièces de monnaie et on la jette. Le résultat est face. Si l'on jette cette pièce de nouveau, quelle est la probabilité que le résultat soit face de nouveau ?

Solution
On sait que si l'on a choisi la pièce qui a les deux côtés faces on va toujours jeter face, alors dans ce cas la probabilité de jeter face est 1. Si l'on a choisi la pièce régulière, on a 50% de chances de jeter face. La première question est : laquelle des deux pièces a-t-on choisie ? On a choisi par hasard avec une probabilité 50% - 50%Mais on sait aussi qu'après avoir lancé cette pièce choisie, le résultat était face. En sachant que l'on a obtenu face avec la pièce choisie, quelle est la probabilité que l'on ait choisi la pièce normale et quelle est la probabilité que l'on ait choisi l'autre ? Soient A, B et F les évenements suivants :

— A : On choisi la pièce qui a les deux côtés faces.
— B : On choisi la pièce qui a deux côtés differents.
— F : On jette face pour la première fois.

Donc,

$$\Pr(A) = \frac{1}{2}, \Pr(B) = \frac{1}{2}, \Pr(F|A) = 1, \Pr(F|B) = \frac{1}{2},$$

$$\Pr(A|F) = \frac{\Pr(F|A)\Pr(A)}{\Pr(F)} = \frac{2}{3}, \text{ et } \Pr(B|F) = \frac{\Pr(F|B)\Pr(B)}{\Pr(F)} = \frac{1}{3}.$$

On a

$$\Pr(G) = \Pr(F|A)\Pr(A|F) + \Pr(F|B)\Pr(B|F) = 1 \times \frac{2}{3} + \frac{1}{2} \times \frac{1}{3} = \frac{5}{6},$$

où $\Pr(G)$ est la probabilité que le résultat soit face de nouveau en utilisant la même pièce.

Exercice 13.14. Densité jointe normale et corrélation
Soit la distribution jointe de deux variables aléatoires (X et Y) est définie par :

$$f(x,y) = \frac{1}{\pi\sqrt{3}} \exp\left\{-\frac{2}{3}(x^2 - xy + y^2)\right\}.$$

1. Donner les distributions marginales, $f_X(x)$ et $f_Y(y)$.
2. Calculer $\mathrm{E}(X)$, $\mathrm{var}(X)$, $\mathrm{E}(Y)$ et $\mathrm{var}(Y)$.
3. Donner les distributions conditionnelles, $f(x|y)$ et $f(y|x)$.
4. Calculer $\mathrm{E}(X|Y=y)$, $\mathrm{var}(X|Y=y)$, $\mathrm{E}(Y|X=x)$ et $\mathrm{var}(Y|X=x)$.
5. Est-ce que X et Y sont indépendantes ? Si non, calculer le coefficient de corrélation.
6. Montrer que (X,Y) a une distribution normale bivariée.
7. Calculer $\mathrm{E}(X+Y)$, $\mathrm{E}(X-Y)$, $\mathrm{var}(X+Y)$ et $\mathrm{var}(X-Y)$.

Rappelons que :

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} dx = 1$$

Solution

1. Les distributions marginales :

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

   (a) On a :

$$\begin{aligned}
f_Y(y) &= \int_{-\infty}^{\infty} \frac{1}{\pi\sqrt{3}} \exp\left\{-\frac{2}{3}(x^2 - xy + y^2)\right\} dx \\
&= \frac{1}{\pi\sqrt{3}} \int_{-\infty}^{\infty} \exp\left\{-\frac{2}{3}(x^2 - xy + y^2)\right\} dx \\
&= \frac{1}{\pi\sqrt{3}} \int_{-\infty}^{\infty} \exp\left\{-\frac{2}{3}[(x - \frac{1}{2}y)^2 - \frac{1}{4}y^2 + y^2]\right\} dx \\
&= \frac{1}{\pi\sqrt{3}} \int_{-\infty}^{\infty} \exp\left\{-\frac{2}{3}(x - \frac{1}{2}y)^2 - \frac{2}{3}\frac{3}{4}y^2\right\} dx \\
&= \frac{1}{\pi\sqrt{3}} \exp\left\{-\frac{1}{2}y^2\right\} \int_{-\infty}^{\infty} \exp\left\{-\frac{2}{3}(x - \frac{1}{2}y)^2\right\} dx \\
&= \frac{1}{\pi\sqrt{3}} \exp\left\{-\frac{1}{2}y^2\right\} \sqrt{2\pi\sigma^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\{x - \mu(y)\}^2\right\} dx,
\end{aligned}$$

   où

$$\mu(y) = \frac{y}{2} \text{ et } \sigma^2 = \frac{3}{4}.$$

   Comme

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\{x - \mu(y)\}^2\right\} dx = 1,$$

   on a

$$f_Y(y) = \frac{1}{\pi\sqrt{3}} \exp\left\{-\frac{1}{2}y^2\right\} \sqrt{2\pi\sigma^2} = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}y^2\right\}.$$

   (b) Comme X et Y sont symétriques

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x^2\right\}.$$

2. Donc, X et Y sont des variables aléatoire normales centrées réduites. Ainsi :

$$\mathrm{E}(X) = 0, \mathrm{var}(X) = 1, \mathrm{E}(Y) = 0 \text{ et } \mathrm{var}(Y) = 1.$$

3. Les densités conditionnelles sont :

$$\begin{aligned} f(x|y) &= \frac{f(x,y)}{f_Y(y)} = \frac{\frac{1}{\pi\sqrt{3}}e^{-\frac{2}{3}(x^2-xy+y^2)}}{\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2}} = \frac{1}{\sqrt{\frac{3}{2}\pi}}e^{-\frac{2}{3}(x^2-xy+y^2)+\frac{1}{2}y^2} \\ &= \frac{1}{\sqrt{2\pi\frac{3}{4}}}e^{-\frac{2}{3}(x-\frac{1}{2}y)^2-\frac{1}{2}y^2+\frac{1}{2}y^2} = \frac{1}{\sqrt{2\pi\frac{3}{4}}}e^{-\frac{2}{3}(x-\frac{1}{2}y)^2} = \frac{1}{\sqrt{2\pi\frac{3}{4}}}e^{-\frac{1}{2\frac{3}{4}}(x-\frac{1}{2}y)^2} \end{aligned}$$

qui est une fonction d'une distribution normale avec $(\mu,\sigma) = \left(\frac{1}{2}y, \sqrt{\frac{3}{4}}\right)$.
Par symétrie,

$$f(y|x) = \frac{f(x,y)}{f(x)} = \frac{1}{\sqrt{2\pi\frac{3}{4}}}e^{-\frac{1}{2\frac{3}{4}}(y-\frac{1}{2}x)^2},$$

qui est une fonction d'une distribution normale avec $(\mu,\sigma) = \left(\frac{1}{2}y, \sqrt{\frac{3}{4}}\right)$.

4. Donc,

$$\mathrm{E}(X|Y=y) = \frac{y}{2}, \mathrm{var}(X|Y=y) = \frac{3}{4}, \mathrm{E}(Y|X=x) = \frac{x}{2} \text{ et } \mathrm{var}(Y|X=x) = \frac{3}{4}.$$

5. Les deux variables ne sont pas indépendantes, car

$$f(x,y) = \frac{1}{\pi\sqrt{3}}e^{-\frac{2}{3}(x^2-xy+y^2)} \neq f_X(x)f_Y(y) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2} \times \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2} = \frac{1}{2\pi}e^{-\frac{1}{2}(x^2+y^2)}.$$

La corrélarion est :

$$\rho = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)\mathrm{var}(Y)}} = \frac{\mathrm{E}(XY)-\mathrm{E}(X)\mathrm{E}(Y)}{\sqrt{[\mathrm{E}(X^2)-\mathrm{E}^2(X)][\mathrm{E}(Y^2)-\mathrm{E}^2(Y)]}}$$

Il faut donc calculer : $\mathrm{E}(X)$, $\mathrm{E}(Y)$, $\mathrm{E}(X^2)$, $\mathrm{E}(Y^2)$ et $\mathrm{E}(XY)$. Comme X et Y sont des variables aléatoires normales centrées réduites, leur espérances sont de 0 et leur variances sont de 1. Ainsi,

$$\mathrm{E}(X) = 0, \mathrm{E}(Y) = 0, \mathrm{E}(X^2) = \mathrm{var}(X) + \mathrm{E}^2(X) = 1 \text{ et } \mathrm{E}(Y^2) = \mathrm{var}(Y) + \mathrm{E}^2(Y) = 1.$$

Nous devons calculer

$$\mathrm{E}(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dxdy = \frac{1}{\pi\sqrt{3}}\int_{-\infty}^{\infty} ye^{-\frac{1}{2}y^2}\sqrt{2\pi\frac{3}{4}}\int_{-\infty}^{\infty} x\frac{1}{\sqrt{2\pi\frac{3}{4}}}e^{-\frac{1}{2\frac{3}{4}}(x-\frac{1}{2}y)^2}dxdy.$$

Comme

$$\int_{-\infty}^{\infty} x\frac{1}{\sqrt{2\pi\frac{3}{4}}}e^{-\frac{1}{2\frac{3}{4}}(x-\frac{1}{2}y)^2}dx$$

est l'espérance d'une variable normale $(\mu,\sigma) = \left(\frac{1}{2}y, \sqrt{\frac{3}{4}}\right)$

$$\mathrm{E}(XY) = \frac{1}{\pi\sqrt{3}}\int_{-\infty}^{\infty} ye^{-\frac{1}{2}y^2}\sqrt{2\pi\frac{3}{4}}\frac{1}{2}ydy = \sqrt{\frac{2\pi\frac{3}{4}}{3\pi^2}}\frac{1}{2}\int_{-\infty}^{\infty} y^2e^{-\frac{1}{2}y^2}dy.$$

Puis comme

$$\int_{-\infty}^{\infty} y^2\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2}dy$$

est le deuxième moment centré ($\mathrm{E}(Y^2)$) d'une variable normale centrée réduite, elle vaut 1. ($\mathrm{E}(Y^2) = \mathrm{var}(Y) + \mathrm{E}^2(Y) = 1$). Il reste donc

$$\mathrm{E}(XY) = \sqrt{\frac{2\pi\frac{3}{4}2\pi}{3\pi^2}} \times \frac{1}{2} = \frac{1}{2}.$$

Finalement,

$$\rho = \frac{1}{2}.$$

6. Il faut montrer qu'en utilisant les espérances et les variances de ces deux variables aléatoire X et Y ainsi que leur coefficient de corrélation, trouvés dans les points précédents, la distribution jointe peut s'écrire sous la forme :

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

En utilisant nos données :

$$\mu_x = \mathrm{E}(X) = 0, \mu_y = \mathrm{E}(Y) = 0,\ \sigma_x = \sqrt{\mathrm{var}(X)} = 1 \sigma_y = \sqrt{\mathrm{var}(Y)} = 1 \text{ et } r = R = \frac{1}{2}$$

$$f(x,y) = \frac{1}{2\pi\sqrt{1-\frac{1}{4}}} e^{-\frac{1}{2(1-\frac{1}{4})}[x^2 - 2\frac{1}{2}xy + y^2]} = \frac{1}{\pi\sqrt{3}} e^{-\frac{2}{3}[x^2 - xy + y^2]}.$$

7. On peut maintenant calculer :

$$\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y) = 0 + 0 = 0, \mathrm{E}(X-Y) = \mathrm{E}(X) - \mathrm{E}(Y) = 0 - 0 = 0,$$

$$\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y) + 2\mathrm{cov}(X,Y) = 1+1+2(\mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)) = 2+2\left(\frac{1}{2} - 0\times 0\right) = 2+1 = 3,$$

et

$$\mathrm{var}(X-Y) = \mathrm{var}(X) + \mathrm{var}(Y) - 2\mathrm{cov}(X,Y) = 1 + 1 - 2(\mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)) = 2 - 1 = 1.$$

Exercice 13.15.
Soit une suite de variables aléatoires $X_1, \ldots, X_n$ indépendantes et identiquement distribuées dont la distribution de probabilité, la fonction de densité, ou la fonction de répartition d'un X_i est donnée par

1. Distribution de Bernoulli (discrète) :
$$p_{X_i}(x) = \begin{cases} \theta^x(1-\theta)^{1-x} & \text{si } x = 0, 1; 0 \leq \theta \leq 1 \\ 0 & \text{sinon.} \end{cases}$$
2. Distribution binomiale (discrète)
$$p_{X_i}(x) = \begin{cases} \binom{m}{x}\theta^x(1-\theta)^{n-x} & \text{si } x = 0, \ldots, m; 0 \leq \theta \leq 1 \\ 0 & \text{sinon.} \end{cases}$$
3. Distribution géometrique (discrète)
$$p_{X_i}(x) = \begin{cases} \theta(1-\theta)^x & \text{si } x = 0, 1, \ldots, 0 \leq \theta \leq 1 \\ 0 & \text{sinon.} \end{cases}$$
4. Distribution binomiale négative (discrète) :
$$p_{X_i}(x) = \begin{cases} \binom{r+x-1}{x}\theta^r(1-\theta)^x & \text{si } x = 0, 1, \ldots, 0 \leq \theta \leq 1 \\ 0 & \text{sinon.} \end{cases}$$
5. Distribution Poisson (discrète) :
$$p_{X_i}(x) = \begin{cases} \dfrac{e^{-\theta}\theta^x}{x!} & \text{si } x = 0, 1, \ldots, \theta > 0 \\ 0 & \text{sinon.} \end{cases}$$
6. Distribution exponentielle (continue)
$$f_{X_i}(x) = \begin{cases} \theta e^{-\theta x} & \text{si } x \geq 0, \theta > 0 \\ 0 & \text{sinon.} \end{cases}$$

Donner la fonction de vraisemblance et son logarithme. Dériver la fonction de vraisemblance. Estimer le paramètre au moyen de la méthode du maximum de vraisemblance.

Solution

1. Bernoulli
$$L(\theta) = \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}$$
$$\ell(\theta) = \sum_{i=1}^n x_i \log(\theta) + \left(n - \sum_{i=1}^n x_i\right)\log(1-\theta)$$
$$\frac{d\ell(\theta)}{d\theta} = \frac{1}{\theta}\sum_{i=1}^n x_i - \frac{1}{1-\theta}\left(n - \sum_{i=1}^n x_i\right) = \frac{n}{\theta(1-\theta)}\{(1-\theta)\bar{x} - \theta(1-\bar{x}))\} = \frac{n}{\theta(1-\theta)}\{\bar{x} - \theta\}$$
$$\widehat{\theta} = \frac{\sum_{i=1}^n x_i}{n} = \bar{x}$$
2. Binomiale
$$L(\theta) = \theta^{\sum_{i=1}^n x_i}(1-\theta)^{mm-\sum_{i=1}^n x_i}\prod_{i=1}^n \binom{m}{x_i}$$
$$\ell(\theta) = \sum_{i=1}^n x_i \log(\theta) + \left(nm - \sum_{i=1}^n x_i\right)\log(1-\theta) + \log\left\{\prod_{i=1}^n \binom{m}{x_i}\right\}$$
$$\frac{d\ell(\theta)}{d\theta} = \frac{1}{\theta}\sum_{i=1}^n x_i - \frac{1}{1-\theta}\left(nm - \sum_{i=1}^n x_i\right) = \frac{n}{\theta(1-\theta)}\{(1-\theta)\bar{x} - \theta(m-\bar{x})\} = \frac{n}{\theta(1-\theta)}(\bar{x} - \theta m)$$
$$\widehat{\theta} = \frac{\bar{x}}{m}$$

3. Géométrique

$$L(\theta) = \theta^n (1-\theta)^{\sum_{i=1}^n x_i}$$

$$\ell(\theta) = n \log(\theta) + \sum_{i=1}^n x_i \log(1-\theta)$$

$$\frac{d\ell(\theta)}{d\theta} = \frac{n}{\theta} - \frac{\sum_{i=1}^n x_i}{1-\theta} = \frac{n}{\theta(1-\theta)} \{(1-\theta) - \theta\bar{x}\}$$

$$\widehat{\theta} = \frac{1}{1+\bar{x}}$$

4. Binomiale négative

$$L(\theta) = \theta^{nr} (1-\theta)^{\sum_{i=1}^n x_i} \prod_{i=1}^n \binom{r+x_i+1}{x_i}$$

$$\ell(\theta) = nr \log(\theta) + \sum_{i=1}^n x_i \log(1-\theta) + \log\left\{\prod_{i=1}^n \binom{r+x_i+1}{x_i}\right\}$$

$$\frac{d\ell(\theta)}{d\theta} = \frac{nr}{\theta} - \frac{1}{1-\theta} \sum_{i=1}^n x_i = \frac{n}{\theta(1-\theta)} \{r(1-\theta) - \theta\bar{x}\}$$

$$\widehat{\theta} = \frac{r}{\bar{x}+r}$$

5. Poisson

$$L(\theta) = e^{-n\theta} \theta^{\sum_{i=1}^n x_i} \prod_{i=1}^n \frac{1}{x_i!}$$

$$\ell(\theta) = -n\theta + \sum_{i=1}^n x_i \log(\theta) + \log\left\{\prod_{i=1}^n \frac{1}{x_i!}\right\}$$

$$\frac{d\ell(\theta)}{d\theta} = -n + \frac{1}{\theta} \sum_{i=1}^n x_i = \frac{n}{\theta} (\bar{x} - \theta)$$

$$\widehat{\theta} = \frac{\sum_{i=1}^n x_i}{n} = \bar{x}$$

6. Exponentielle

$$L(\theta) = \theta^n e^{-\theta \sum_{i=1}^n x_i}$$

$$\ell(\theta) = n \log(\theta) - \theta \sum_{i=1}^n x_i$$

$$\frac{d\ell(\theta)}{d\theta} = \frac{n}{\theta} - \sum_{i=1}^n x_i = n\left(\frac{1}{\theta} - \bar{x}\right)$$

$$\widehat{\theta} = \frac{1}{\bar{x}}$$

Exercice 13.16. Densité uniforme et estimation
Soit une suite de variables aléatoires $X_1, \ldots, X_n$ indépendantes et identiquement distribuées dont les densités sont données par

$$f_{x_i}(x_i) = \begin{cases} \frac{1}{\theta} & \text{si } 0 \leq x_i \leq \theta, \\ 0 & \text{sinon.} \end{cases} \tag{13.2}$$

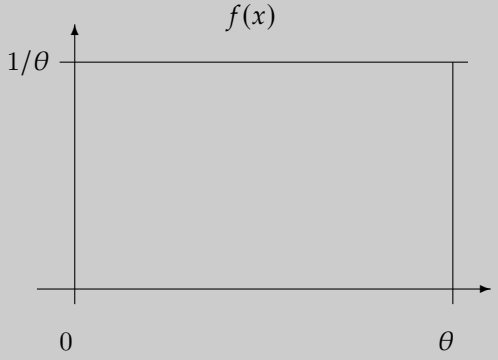
1. Dessiner la fonction de densité et la fonction de répartition de X_i.
2. Quelle est la densité jointe du vecteur $X = (X_1, \ldots, X_i, \ldots, X_n)$?
3. Donner la fonction de vraisemblance.
4. Estimer θ par maximum de vraisemblance.
5. Donner les fonctions de densité et de répartition de l'estimateur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

Solution

1. (a) La fonction de densité vaut

$$f(x) = \begin{cases} 1/\theta & \text{si } x \in [0, \theta] \\ 0 & \text{sinon.} \end{cases}$$

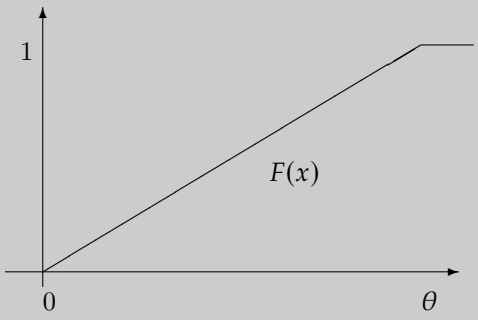
Fonction de densité



(b) La fonction de répartition vaut

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

Fonction de répartition



2. La répartition jointe vaut :

$$F(x_1, \ldots, x_n) = \Pr[X_1 \leq x_1, \ldots, X_n \leq x_n] = \Pr\left[\bigcap_{i=1}^{n}(X_i \leq x_i)\right] = \prod_{i=1}^{n} \Pr(X_i \leq x_i) = \prod_{i=1}^{n} \frac{x_i}{\theta},$$

où $0 \leq x_i \leq \theta$, pour tout, $i = 1, \ldots, n$.
La densité jointe vaut :

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \frac{I\left[0 \leq x_i \leq \theta\right]}{\theta},$$

où $I[A]$ vaut 1 si A est vrai et 0 sinon.

3. La fonction de vraisemblance est donc :

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\theta} I\left[0 \leq x_i \leq \theta\right] = \frac{I\left[x_{\max} \leq \theta\right]}{\theta^n}.$$

4. L'estimateur de maximum de vraisemblance vaut : $\widehat{\theta}_{mv} = x_{\max}$.
5. La répartition de $x_{\max}$ est :

$$F_{\max}(x) = \Pr\left[x_{\max} \leq x\right] = \Pr\left[\bigcap (X_i \leq x)\right] = \prod_{i=1}^{n} \Pr[X_i \leq x] = \left(\frac{x}{\theta}\right)^n = \frac{x^n}{\theta^n}.$$

Donc,

$$f_{\max}(x) = \frac{1}{\theta^n} n x^{n-1}.$$

6. L'espérance de $x_{\max}$ est :

$$\mu\max = \int_0^\theta x \frac{1}{\theta^n} n x^{n-1} dx = \frac{n}{\theta^n} \int_0^\theta x^n dx = \frac{n}{\theta^n} \left[\frac{x^{n+1}}{n+1}\right]_0^\theta = \theta \frac{n}{n+1}.$$

7. On obtient donc un estimateur sans biais

$$\widehat{\theta}_{sb} = x_{\max} \frac{n+1}{n}.$$

8. Calculons maintenant la variance de $x_{\max}$

$$\begin{aligned} \sigma^2_{max} &= \int_0^\theta x^2 \frac{1}{\theta^n} n x^{n-1} dx - \mu^2_{\max} = \frac{n}{\theta^n} \int_0^\theta x^{n+1} dx - \mu^2_{\max} \\ &= \frac{n}{\theta^n} \left[\frac{x^{n+2}}{n+2}\right]_0^\theta - \mu^2_{\max} = \theta^2 \frac{n}{n+2} - \left(\theta \frac{n}{n+1}\right)^2 = \theta^2 \frac{n}{(n+2)(n+1)^2}. \end{aligned}$$

Donc,

$$\operatorname{var}\left(\widehat{\theta}_{sb}\right) = \operatorname{var}\left(x_{\max} \frac{n+1}{n}\right) = \left(\frac{n+1}{n}\right)^2 \frac{n}{(n+2)(n+1)^2} \theta^2 = \theta^2 \frac{1}{n(n+2)}$$

et

$$\operatorname{var}\left(\frac{\widehat{\theta}_{sb}}{2}\right) = \theta^2 \frac{1}{4n(n+2)}.$$

Si on calcule maintenant la moyenne des observations x_i, on sait que

$$\operatorname{var}(\bar{x}) = \frac{\sigma^2}{n}.$$

De plus, on peut calculer

$$\begin{aligned} \sigma^2 &= \operatorname{var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2 = \int_0^\theta x^2 \frac{1}{\theta} dx - \left[\int_0^\theta x \frac{1}{\theta} dx\right]^2 \\ &= \frac{1}{\theta} \left[\frac{x^3}{3}\right]_0^\theta - \left\{\frac{1}{\theta} \left[\frac{x^2}{2}\right]_0^\theta\right\}^2 = \frac{\theta^2}{3} - \frac{\theta^2}{4} = \frac{\theta^2}{12}. \end{aligned}$$

Donc,

$$\mathrm{var}(\bar{x}) = \frac{\theta^2}{12n}.$$

On constate que

$$\mathrm{var}(\bar{x}) - \mathrm{var}\left(\frac{\widehat{\theta}_{sb}}{2}\right) = \frac{\theta^2}{4n}\left[\frac{1}{3} - \frac{1}{n+2}\right] = \theta^2 \frac{n-1}{12n(n+2)} \geq 0.$$

L'estimateur $\widehat{\theta}_{sb}/2$ est donc plus efficace.

Chapitre 14

Exercices : Modèle de régression bivariée

Exercice 14.1. Modèle bivariée
Soit le modèle linéaire :

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

Montrer que :

$$\text{SC}_{RES} = ns_y^2 - \widehat{\beta}^2 \times ns_x^2.$$

Solution
On a

$$\text{SC}_{RES} = \sum_{i=1}^{n}(y_i - \widehat{y}_i)^2 = \sum_{i=1}^{n}(y_i - (\widehat{\alpha} + \widehat{\beta}x_i))^2.$$

Comme

$$\bar{y} - \widehat{\beta}\bar{x} - \widehat{\alpha} = 0,$$

$$\begin{aligned}\text{SC}_{RES} &= \sum_{i=1}^{n}(y_i - \widehat{y}_i)^2 = \sum_{i=1}^{n}(y_i - \bar{y} + \widehat{\beta}\bar{x} + \widehat{\alpha} - (\widehat{\alpha} + \widehat{\beta}x_i))^2 = \sum_{i=1}^{n}(y_i - \bar{y} - \widehat{\beta}(x_i - \bar{x}))^2 \\ &= \sum_{i=1}^{n}(y_i - \bar{y})^2 + \widehat{\beta}^2\sum_{i=1}^{n}(x_i - \bar{x})^2 - 2\widehat{\beta}\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x}) = ns_y^2 + n\widehat{\beta}^2 s_x^2 - 2n\widehat{\beta}s_{xy} = ns_y^2 - n\widehat{\beta}^2 s_x^2,\end{aligned}$$

car $\widehat{\beta}^2 s_x^2 = n\widehat{\beta}s_{xy}$.

Exercice 14.2. Modèle et sommes de carrés
Soit le modèle linéaire :

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

On suppose que chaque x_i est remplacé par cx_i, où $c \neq 0$ est une constante. Quelles valeurs prendront les estimateurs $\widehat{\alpha}$ et $\widehat{\beta}$ et le coefficient de détermination R^2, ainsi que les statistiques suivantes :

$$\widehat{\sigma}^2_\varepsilon = \frac{\text{SC}_{RES}}{n-2}, \quad \widehat{\text{var}}(\widehat{\beta}) = \frac{\widehat{\sigma}^2_\varepsilon}{ns^2_x} \quad \text{et} \quad t_\beta = \frac{\widehat{\beta}}{\sqrt{\widehat{\text{var}}(\widehat{\beta})}}.$$

Solution
On note $z_i = cx_i$

$$ns^2_z = \sum_{i=1}^n (cx_i - c\bar{x})^2 = c^2 \sum_{i=1}^n (x_i - \bar{x})^2 = c^2 \sum_{i=1}^n x_i^2 - c^2 n\bar{x}^2 = c^2 ns^2_x,$$

$$ns_{zy} = \sum_{i=1}^n (cx_i - c\bar{x})(y_i - \bar{y}) = \sum_{i=1}^n c(x_i - \bar{x})(y_i - \bar{y}) = c \sum_{i=1}^n x_i y_i - cn\bar{x}\bar{y} = cns_{xy},$$

$$ns^2_y = \sum_{i=1}^n y_i^2 - n\bar{y}^2 = ns^2_y.$$

Donc,

$$\widehat{\beta}_{y|z} = \frac{cs_{xy}}{c^2 s^2_x} = \frac{1}{c}\widehat{\beta}_{y|x}$$

et

$$\widehat{\alpha}_{y|z} = \bar{y} - \widehat{\beta}_{y|z} c\bar{x} = \bar{y} - \frac{1}{c}\widehat{\beta} c\bar{x} = \widehat{\alpha}_{y|x},$$

car

$$\bar{z} = c\bar{x}.$$

On a

$$\text{SC}_{RES,y|z} = ns^2_y - \widehat{\beta}^2_{y|z} ns^2_z = ns^2_y - \frac{1}{c^2}\widehat{\beta}^2_{y|x} c^2 ns^2_x = \text{SC}_{RES,y|x},$$

$$\widehat{\sigma}^2_{\varepsilon,y|z} = \frac{\text{SC}_{RES,y|z}}{n-2} = \widehat{\sigma}^2_{\varepsilon,y|x},$$

et

$$R^2_{y|z} = \frac{\text{SC}_{reg,y_z}}{\text{SC}_{TOT}} = 1 - \frac{\text{SC}_{RES,y|z}}{\text{SC}_{tot,y|z}} = 1 - \frac{\text{SC}_{RES,y|z}}{ns^2_y} = 1 - \frac{\text{SC}_{RES,y|x}}{ns^2_y} = R^2_{y|x},$$

$$\widehat{\text{var}}(\widehat{\beta}_{y|z}) = \frac{\widehat{\sigma}^2_{\varepsilon,y|z}}{ns^2_z} = \frac{\widehat{\sigma}^2_{\varepsilon,y|x}}{c^2 ns^2_x} = \frac{1}{c^2}\widehat{\text{var}}(\widehat{\beta}_{y|x}),$$

et

$$t_{\beta,y|z} = \frac{\widehat{\beta}_{y|z}}{\sqrt{\widehat{\text{var}}(\widehat{\beta}_{y|z})}} = \frac{\frac{1}{c}\widehat{\beta}_{y|x}}{\sqrt{\frac{1}{c^2}\widehat{\text{var}}(\widehat{\beta}_{y|x})}} = \frac{\frac{1}{c}\widehat{\beta}_{y|x}}{\frac{1}{c}\sqrt{\widehat{\text{var}}(\widehat{\beta}_{y|x})}}. = t_{\beta,y|x}.$$

Exercice 14.3. Prix et poids
À partir des données (x_i, y_i), où Y : prix (CHF) et X : poids (gr) concernant un certain produit alimentaire, on a calculé les sommes suivantes :

$$\sum_{i=1}^{8} x_i = 1235\ ,\ \sum_{i=1}^{8} y_i = 32.3\ ,\ \sum_{i=1}^{8} x_i^2 = 208405\ ,\ \sum_{i=1}^{8} y_i^2 = 139.89\ \text{ et }\ \sum_{i=1}^{8} y_i x_i = 5320.1$$

Soit le modèle linéaire :

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

1. Estimer les paramètres du modèle par la méthode des moindres carrés et écrire l'équation.
2. Utiliser l'équation de la régression pour prévoir :
   — le prix d'un aliment pesant 200gr,
   — le prix d'un aliment pesant 300gr,
   — l'augmentation du prix pour 50gr supplémentaires.
3. Donner un estimateur sans biais de la variance des erreurs.
4. Donner la loi de distribution de $\widehat{\beta}$ et celle de $\widehat{\alpha}$.
5. Tester l'hypothèse H_0 : $\beta = 0$ vs $\beta \neq 0$ au seuil de 5%.
6. Tester l'hypothèse H_0 : $\alpha = 0$ vs $\alpha \neq 0$ au seuil de 5%.
7. Calculer les expressions suivantes : SC_{TOT}, SC_{REG} et SC_{RES}, et donner les expressions des espérances.
8. Compléter le Tableau de l'analyse de la variance.
9. Effectuer le F-test et commenter.
10. Calculer le coefficient de détermination R^2 et commenter.

Source	d.l.	SC	CM	F_{obs}
Régression	$\cdots$	$\cdots$	$\cdots$	$\cdots$
Résiduelle	$\cdots$	$\cdots$	$\cdots$	
Totale	$\cdots$	$\cdots$		

Solution

1. Les paramètres sont :

$$\widehat{\beta} = \frac{n s_{xy}}{n s_x^2} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} = \frac{333.788}{17751.875} = 0.019,$$

   et

$$\widehat{\alpha} = \bar{y} - \widehat{\beta}\bar{x} = 4.038 - 0.019 \times 154.375 = 1.105,$$

   l'équation est :

$$y_i = \widehat{\alpha} + \widehat{\beta} x_i = 1.105 + 0.019 \times x_i.$$

2. Prix d'un aliment pesant 200gr

$$y_j = 1.105 + 0.019 \times 200 = 4.199.$$

   Prix d'un aliment pesant 200gr

$$y_k = 1.105 + 0.019 \times 300 = 6.299.$$

   L'augmentation du prix :

$$\Delta y = \alpha + \beta(x + 50) - (\alpha + \beta x) = 50\beta = 50 \times 0.019 = 0.95.$$

3. L'estimateur sans biais de la variance des erreurs :

$$\widehat{\sigma}_\varepsilon^2 = \frac{\text{SC}_{RES}}{n - p} = \frac{n s_y^2 - \widehat{\beta}^2 n s_x^2}{n - 2} = \frac{\sum_{i=1}^{n} y_i^2 - n\bar{y}^2 - \widehat{\beta}^2 \times 17751.875}{n - 2} = \frac{3.038}{6} = 0.506.$$

   $(\widehat{\sigma}_\varepsilon = 0.711)$.

4. La distribution de $\widehat{\beta}$ est :

$$\widehat{\beta} \sim \mathcal{N}\left(\mathrm{E}(\widehat{\beta}), \mathrm{var}(\widehat{\beta})\right) \quad = \quad \mathcal{N}\left(\beta, \frac{\sigma_\varepsilon^2}{n s_x^2}\right).$$

En effet,

$$\frac{\widehat{\beta} - \mathrm{E}(\widehat{\beta})}{\sqrt{\mathrm{var}(\widehat{\beta})}} = \frac{\sqrt{n} s_x (\widehat{\beta} - \beta)}{\sigma_\varepsilon} \sim N(0,1)$$

Si la variance est estimée, la distribution suit une loi de Student :

$$\frac{\widehat{\beta} - \mathrm{E}(\widehat{\beta})}{\sqrt{\widehat{\mathrm{var}}(\widehat{\beta})}} \sim t_{n-p}.$$

La distribution de $\widehat{\alpha}$:

$$\widehat{\alpha} \sim \mathcal{N}\left(\mathrm{E}(\widehat{\alpha}), \mathrm{var}(\widehat{\alpha})\right) \quad = \quad \mathcal{N}\left(\alpha, \frac{\sigma_\varepsilon^2}{n s_x^2}(s_x^2 + \bar{x}^2)\right).$$

En effet,

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

Si la variance est estimée, la distribution suit une loi de Student :

$$\frac{\widehat{\alpha} - \mathrm{E}(\widehat{\alpha})}{\sqrt{\widehat{\mathrm{var}}(\widehat{\alpha})}} \sim t_{n-p}.$$

5. Comme

$$\begin{aligned} t_\beta &= \frac{\sqrt{n} s_x (\widehat{\beta} - \beta^0)}{\widehat{\sigma}_\varepsilon} = \frac{\sqrt{8}\sqrt{\frac{1}{n}\sum_{i=1}^8 x_i^2 - \bar{x}^2} \times (\widehat{\beta} - \beta^0)}{0.711} \\ &= \frac{\sqrt{8}\left[\sqrt{\frac{1}{8} \times 208405 - 154.375^2}\right] \times 0.019}{0.711} = \frac{2.531}{0.711} = 3.560, \\ & t_{hyp} = t_{0.975,6} = 2.447, \end{aligned}$$

on rejette l'hypothèse : $H_0 \ : \ \beta = 0$.

6. Comme

$$t_\alpha = \frac{\sqrt{n} s_x (\widehat{\alpha} - \alpha^0)}{\widehat{\sigma}_\varepsilon \sqrt{s_x^2 + \bar{x}^2}} = \frac{\widehat{\alpha} - \alpha^0}{\sqrt{\widehat{\mathrm{var}}(\widehat{\alpha})}} = \frac{1.105}{\sqrt{1.045}} = 1.081,$$

on ne rejette pas l'hypothèse : $H_0 \ : \ \alpha = 0$.

7. Les sommes de carrés :

$$\mathrm{SC}_{TOT} = \sum_{i=1}^8 (y_i - \bar{y})^2 = \sum_{i=1}^8 y_i^2 - n\bar{y}^2 = 139.890 - 8 \times 4.038^2 = 9.446,$$

$$\begin{aligned} \mathrm{SC}_{RES} &= n s_y^2 - \widehat{\beta}^2 \times n s_x^2 = \sum_{i=1}^8 y_i^2 - n\bar{y}^2 - \widehat{\beta}^2 \times \left(\sum_{i=1}^8 x_i^2 - n\bar{x}^2\right) \\ &= 139.890 - 8 \times 4.038^2 - 0.019^2 (208405 - 8 \times 154.375) = 3.038 \end{aligned}$$

et

$$\mathrm{SC}_{REG} = \mathrm{SC}_{TOT} - \mathrm{SC}_{RES} = 6.408.$$

Leurs espérances sont :

$$\mathrm{E}(\mathrm{SC}_{RES}) = (n-2)\sigma_\varepsilon^2,$$
$$\mathrm{E}(\mathrm{SC}_{REG}) = \sigma_\varepsilon^2 + \beta^2 n s_x^2$$

et

$$\mathrm{E}(\mathrm{SC}_{TOT}) = \mathrm{E}(\mathrm{SC}_{REG} + \mathrm{SC}_{RES}) = \mathrm{E}(\mathrm{SC}_{REG}) + \mathrm{E}(\mathrm{SC}_{RES}) = (n-2)\sigma_\varepsilon^2 + \sigma_\varepsilon^2 + \beta^2 n s_x^2 = (n-1)\sigma_\varepsilon^2 + \beta^2 n s_x^2.$$

8. Le tableau de l'analyse de la variance :

Source	d.l.	SC	CM	F_{obs}
Régression	1	6.408	6.408	12.664
Résiduelle	6	3.038	0.506	
Totale	7	9.446		

9. Comme
$$F_{hyp} = F_{0.95,1,6} = 5.987 < F_{test} = 11.758,$$
on rejette l'hypothèse $H_0 \ : \ \beta = 0$.
10. Le R-carré :
$$R^2 = \frac{\text{SC}_{REG}}{\text{SC}_{TOT}} = \frac{6.408}{9.446} = 0.678.$$
68% de la variance de y est expliquée par la régression.

Exercice 14.4. Frais d'inscription
Dans le Tableau 14.1, on trouve les frais d'inscription (pour une année) des 25 meilleurs cours MBA aux Etats-Unis en 1996 et l'augmentation moyenne des salaires (de 1995 à 1996) dans ces écoles.

TABLE 14.1 – Frais d'inscription et augmentation moyenne des salaires

	inc.salaries (x)	tuition fee (y)
	55	23.608
	42.9	23.18
	50	23.025
	55.64	23.84
	45	19.627
	44.85	23.83
	54	23.1
	42	24.655
	57.75	23.9
	60	23.7
	33	23.69
	48	18.963
	41	18.788
	36	22.5
	32	15.613
	28.75	20.1
	44.18	22.2
	33.25	22.45
	37.25	15.332
	33	11.854
	30	21.24
	43.05	23.13
	32.35	20.524
	29	21.8
	27	18.95
somme	1035	529
moyenne	41.399	21.144

Afin d'examiner si une relation existe entre l'augmentation moyenne des salaires et les frais d'inscription, on voudrait ajuster un modèle linéaire sur les données. On voudrait savoir si l'augmentation des salaires pourrait influencer les frais d'inscription, ainsi la variable indépendante est l'augmentation moyenne des salaires et la variable dépendante est les frais d'inscription.

1. Estimer les paramètres du modèle par la méthode des moindres carrés et écrire l'équation.
2. Donner un estimateur sans biais de la variance des erreurs.
3. Tester l'hypothèse $H_0 : \beta = 0 \quad \text{vs} \quad \beta \neq 0$ au seuil 5%.
4. Tester l'hypothèse $H_0 : \alpha = 0 \quad \text{vs} \quad \alpha \neq 0$ au seuil 5%.
5. Donner des intervalles de confiance de la pente et de la constante de la droite de régression.
6. Dresser le tableau de l'analyse de la variance.
7. Effectuer le F-test et commenter.
8. Calculer le coefficient de détermination R^2 et commenter.
9. Vérifier vos résultat en utilisant le logiciel R.

TABLE 14.2 – Tableau de l'analyse de la variance

Source	d.l.	SC	CM	F_{obs}
Régression	$\cdots$	$\cdots$	$\cdots$	$\cdots$
Résiduelle	$\cdots$	$\cdots$	$\cdots$	
Totale	$\cdots$	$\cdots$		

Solution

1. Les paramètres sont :

$$\widehat{\beta} = \frac{ns_{xy}}{ns_x^2} = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} = \frac{367.179}{2390.673} = 0.154,$$

et

$$\widehat{\alpha} = \bar{y} - \widehat{\beta}\bar{x} = 21.144 - 0.154 \times 41.399 = 14.786,$$

l'équation est :

$$y_i = \widehat{\alpha} + \widehat{\beta} x_i = 14.786 + 0.154 \times x_i.$$

2. L'estimateur sans biais de la variance des erreurs :

$$\widehat{\sigma}_\varepsilon^2 = \frac{\text{SC}_{RES}}{n-p} = \frac{ns_y^2 - \widehat{\beta}^2 ns_x^2}{n-2} = \frac{\sum_{i=1}^n y_i^2 - n\bar{y}^2 - \widehat{\beta}^2 \times 2390.673}{n-2} = \frac{199.847}{23} = 8.689.$$

($\widehat{\sigma}_\varepsilon = 2.948$).

3. Comme

$$t_\beta = \frac{\sqrt{n} s_x (\widehat{\beta} - \beta^0)}{\widehat{\sigma}_\varepsilon} = \frac{\sqrt{25}\sqrt{\frac{1}{n}\sum_{i=1}^{25} x_i^2 - \bar{x}^2} \times (\widehat{\beta} - \beta^0)}{2.948} = \frac{\sqrt{25}\sqrt{\frac{1}{25} \times 890.022 - 41.399^2} \times 0.154}{2.948}$$

$$= \frac{\widehat{\beta} - \beta^0}{\sqrt{\widehat{\text{var}}(\widehat{\beta})}} = \frac{0.154}{\sqrt{0.00363}} = 2.548$$

$$t_{hyp} = t_{0.975,23} = 2.070,$$

on rejette l'hypothèse : $H_0 \ : \ \beta = 0$.

4. Comme

$$t_\alpha = \frac{\sqrt{n} s_x (\widehat{\alpha} - \alpha^0)}{\widehat{\sigma}_\varepsilon \sqrt{s_x^2 + \bar{x}^2}} = \frac{\widehat{\alpha} - \alpha^0}{\sqrt{\widehat{\text{var}}(\widehat{\alpha})}} = \frac{14.786}{\sqrt{6.577}} = 5.765$$

on rejette l'hypothèse : $H_0 \ : \ \alpha = 0$.

5. Intervalles de confiance autour des coefficients :

$$\text{IC} = \left[\widehat{\beta} - t_{1-\alpha/2,n-2} se(\widehat{\beta}), \widehat{\beta} + t_{1-\alpha/2,n-2} se(\widehat{\beta})\right]$$

et respectivement

$$\text{IC} = \left[\widehat{\alpha} - t_{1-\alpha/2,n-2} se(\widehat{\alpha}), \widehat{\alpha} + t_{1-\alpha/2,n-2} se(\widehat{\alpha})\right],$$

où

$$se(\widehat{\beta}) = \sqrt{\text{var}(\widehat{\beta})} \text{ et } se(\widehat{\alpha}) = \sqrt{\text{var}(\widehat{\alpha})}.$$

Dans notre cas,

$$\text{IC}_{\widehat{\beta}} = (0.029, 0.278)$$

et

$$\text{IC}_{\widehat{\alpha}} = (9.477, 20.094)$$

La tableau de l'analyse de la variance :

6.

Source	d.l.	SC	CM	F_{obs}
Régression	1	56.394	56.394	6.490
Résiduelle	23	199.847	8.689	
Totale	24	256.241		

7. Le test :

$$F_{hyp} = F_{0.95,1,23} = 4.28 < F_{obs} = 6.479,$$

donc on rejette l'hypothèse $H_0 \ : \ \beta = 0$.

8. Le R^2 :

$$R^2 = \frac{\text{SC}_{REG}}{\text{SC}_{TOT}} = \frac{56.394}{256.241} = 0.220083 = 22.0083\%.$$

9. **En langage R**

```
tuition=c(23.608,23.18,23.025,23.84,19.627,23.83,23.1,24.655,23.9,23.7,23.69,18.963,
18.788,22.5,15.613,20.1,22.2,22.45,14.332,11.854,21.24,23.13,20.524,21.8,18.95)
slrygain=c(55,42.9,50,55.64,45,44.85,54,42,57.75,60,33,48,41,36,32,28.75,44.18,
    33.25,37.25,33,30,43.05,32.35,29,27)
y=tuition
x=slrygain
model1=lm(y~x)
summary(model1)
anova(model1)
plot(y~x)
abline(model1)     ### droite
confint(model1)    ### IC
```

Exercice 14.5. Frais de réparation et âge
Le Tableau 14.3 nous montre une partie des 57 observations concernant les frais de réparation cumulés (en USD) et l'âge (en semaine) d'un véhicule.

TABLE 14.3 – Frais cumulés et âge des véhicules

	coût (y)	âge (x)
	11	5
	16	12
	55	30
	⋮	⋮
	3401	526
	3412	527
	3425	538
somme	81318	15932
moyenne	1426.632	279.509

Afin d'examiner si une relation existe entre les frais cumulés et l'âge de ce véhicule, on voudrait ajuster un modèle sur les données. Notre variable indépendante (x) est l'âge, la variable dépendante est les frais cumulés. On connaît également quelques résultats :

$$\sum_{i=1}^{57} x_i^2 = 5692060.500, \sum_{i=1}^{57} x_i y_i = 31827126.0, \sum_{i=1}^{57} y_i^2 = 190223678.0.$$

1. Estimer les paramètres du modèle par la méthode des moindres carrés et écrire l'équation.
2. Donner un estimateur sans biais de la variance des erreurs.
3. Tester l'hypothèse $H_0 : \beta = 0$ vs $\beta \neq 0$ au seuil 5%.
4. Tester l'hypothèse $H_0 : \alpha = 0$ vs $\alpha \neq 0$ au seuil 5%.
5. Donner des intervalles de confiance de la pente et de la constante de la droite de la régression.
6. Construire un intervalle de prédiction pour la valeur actuelle de y si on connaît $x = 624$.
7. Dresser le tableau de l'analyse de la variance.
8. Effectuer le F-test et commenter.
9. Calculer le coefficient de détermination R^2 et commenter.

TABLE 14.4 – Tableau de l'analyse de la variance

Source	d.l.	SC	CM	F_{obs}
Régression	⋯	⋯	⋯	⋯
Résiduelle	⋯	⋯	⋯	
Totale	⋯	⋯		

10. Montrer que
$$t_\beta^2 = \left(\frac{\widehat{\beta} - 0}{\sqrt{\widehat{\text{var}}(\widehat{\beta})}} \right)^2 \text{ et } F_{obs} = \frac{\text{SC}_{REG}/1}{\text{SC}_{RES}/(n-2)} \text{ sont reliés par } F = t^2.$$
11. Vérifier vos résultat en utilisant le logiciel R.

Solution

1. Les paramètres sont :
$$\widehat{\beta} = \frac{n s_{xy}}{n s_x^2} = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} = \frac{9098031.684}{1238926.746} = 7.343,$$

et

$$\widehat{\alpha} = \bar{y} - \widehat{\beta}\bar{x} = 1426.632 - 7.343 \times 279.509 = -625.935,$$

l'équation est :

$$y_i = \widehat{\alpha} + \widehat{\beta}x_i = -625.935 + 7.343 \times x_i.$$

2. L'estimateur sans biais de la variance des erreurs :

$$\widehat{\sigma}_\varepsilon^2 = \frac{\text{SC}_{RES}}{n-p} = \frac{ns_y^2 - \widehat{\beta}^2 ns_x^2}{n-2} = \frac{\sum_{i=1}^n y_i^2 - n\bar{y}^2 - \widehat{\beta}^2 \times 1238926.746}{n-2} = \frac{7401653}{55} = 134575.508.$$

($\widehat{\sigma}_\varepsilon = 366.845$).

3. Comme

$$t_\beta = \frac{\widehat{\beta} - \beta^0}{\sqrt{\widehat{\text{var}}(\widehat{\beta})}} = \frac{7.343}{\sqrt{0.10862}} = 22.281$$

$$t_{hyp} = t_{0.975,55} = 2.000,$$

on rejette l'hypothèse : $H_0 \ : \ \beta = 0$.

4. Comme

$$t_\alpha = \frac{\sqrt{n}s_x(\widehat{\alpha} - \alpha^0)}{\widehat{\sigma}_\varepsilon\sqrt{s_x^2 + \bar{x}^2}} = \frac{\widehat{\alpha} - \alpha^0}{\widehat{\text{var}}(\widehat{\alpha})} = \frac{-625.935}{\sqrt{10847.135}} = -6.010$$

on rejette l'hypothèse : $H_0 \ : \ \alpha = 0$.

5. Donner des intervalles de confiance autour des coefficients.

$$\text{IC} = \left[\widehat{\beta} - t_{1-\alpha/2,n-2}se(\widehat{\beta}), \widehat{\beta} + t_{1-\alpha/2,n-2}se(\widehat{\beta})\right]$$

et respectivement

$$\text{IC} = \left[\widehat{\alpha} - t_{1-\alpha/2,n-2}se(\widehat{\alpha}), \widehat{\alpha} + t_{1-\alpha/2,n-2}se(\widehat{\alpha})\right]$$

ou

$$se(\widehat{\beta}) = \sqrt{\text{var}(\widehat{\beta})} \text{ et } se(\widehat{\alpha}) = \sqrt{\text{var}(\widehat{\alpha})}$$

dans notre cas :

$$\text{IC}_{\widehat{\beta}} = (6.684, 8.003)$$

et

$$\text{IC}_{\widehat{\alpha}} = (-834.234, -417.636)$$

6. L'intervalle de prédiction pour la valeur actuelle de y si on connaît $x = 624$:

$$\text{IC}_{\widehat{y}^*} = \left[\widehat{y}^* - t_{1-\alpha/2,n-2}\widehat{\sigma}_\varepsilon\sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{ns_x^2}}; \widehat{y}^* + t_{1-\alpha/2,n-2}\widehat{\sigma}_\varepsilon\sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{ns_x^2}}\right],$$

où $\widehat{y}^* = \widehat{\beta}_0 + \widehat{\beta}_1 x^* = 3956.395$. Donc,

$$\text{IP}_{\widehat{y}^*} = (3182.245; 4730.546)$$

7. Tableau de l'analyse de la variance :

Source	d.l.	SC	CM	F_{obs}
Régression	1	66811198.338	6681198.338	496.459
Résiduelle	55	7401652.926	134575.508	
Totale	56	74212851.263		

8.

$$F_{hyp} = F_{0.95,1,55} = F_{hyp} = F_{0.95,1,60} = 4.000 < F_{obs} = 496.459,$$

donc on rejette l'hypothèse $H_0 \ : \ \beta = 0$.

9.

$$R^2 = \frac{\text{SC}_{REG}}{\text{SC}_{TOT}} = \frac{66811198.338}{74212851.263} = 0.900265 = 90.0265\%.$$

10. Le tableau de l'analyse de la variance est :

Source de la variation	Degrés de liberté	Sommes des carrées	Carrés moyens	F_{obs}
Régression	1	SC_{REG}	$SC_{REG}/1$	$F = \frac{SC_{REG}/1}{SC_{RES}/(n-2)}$
Résiduelle	$n-2$	SC_{RES}	$SC_{RES}/(n-2)$	
Totale	$n-1$	SC_{TOT}		

$$t_{\hat{\beta}}^2 = 22.281^2 = 496.443 \approx F_{obs} = 496.459.$$

11. **En langage R**

```
cost=c(11,16,55,66,76,83,135,160,163,211,258,322,374,408,478,489,536,590,
       604,704,985,1021,1030,1096,1114,1134,1157,1176,1182,1182,1231,1244,
       1257,1260,1342,1356,1467,1518,1557,1565,1583,1609,2825,2893,2918,
       3011,3077,3095,3154,3162,3217,3274,3320,3329,3401,3412,3425)
age=c(5,12,30,40,42,53,66,73,79,101,114,129,150,180,195,196,204,212,
      224,227,232,235,239,249,260,271,272,273.5,276,279,281,313,326,
      328,329,336.5,338,342.5,344.5,351,366,384,388,402,432,433,436,
      446,456,463.5,465,478,485,498.5,526,527,538)
y=cost
x=age
n=length(y)
model2=lm(y~x)  # Modele lineaire
summary(model2) # Coefficients de la regression
        # Student d'alpha et Student de beta
        # R^2
anova(model2)   # estimateur de variance des residus
        # ANOVA
        # Fisher
plot(y~x)     # graphique des points et de la droite
abline(model2$coefficients[1],model2$coefficients[2])
confint(model2)               # IC des parametres
new=data.frame(x=624)
predict(model2,new,se.model2=TRUE)     # prediction pour x=624
predict(model2,new,interval="prediction") # IC de prediction pour x=624
```

Exercice 14.6. Revenu et sexe
Afin d'examiner l'utilisation d'une variable "dummy" (ou variable indicatrice) comme variable indépendante de la régression simple ($y_i = \alpha + \beta x_i + \varepsilon_i$), voir Tableau 14.5 (une partie de 49 observations). La variable "dummy" est la variable "gender" (sexe) (0 pour une femme et 1 pour un homme) et la variable dépendante est le "wage" (salaire).

TABLE 14.5 – Sexe et salaire des observations

	wage (y)	gender (x)
	1345	0
	2435	1
	1715	1
	⋮	⋮
	1839	1
	1288	1
	1288	0
somme	89190	26
moyenne	1820.204	0.531

Les données sont dans le code R suivant :

En langage R

```
WAGE=c(1345,2435,1715,1461,1639,1345,1602,1144,1566,1496,1234,1345,
       1345,3389,1839,981,1345,1566,1187,1345,1345,2167,1402,2115,
       2218,3575,1972,1234,1926,2165,2365,1345,1839,2613,2533,1602,
       1839,2218,1529,1461,3307,3833,1839,1461,1433,2115,1839,1288,1288)
GENDER=c(0,1,1,1,1,0,0,0,1,1,0,0,0,1,1,1,0,0,0,0,0,1,1,1,
        1,1,1,0,1,0,0,0,0,1,1,0,0,1,0,1,1,1,1,0,1,0,1,1,0)
```

1. Ajuster un modèle de régression linéaire simple sur les données :

$$\text{wage} = \alpha + \beta \times \text{gender},$$

si

$$\sum_{i=1}^{49} x_i y_i = 54260.$$

2. Importer les données en R.
   — Représenter les données graphiquement.
   — Donner les boxplots du salaire des femmes et des hommes.
   — Ajuster un modèle de régression linéaire simple et vérifier les résultats du point 1.
   — Dessiner la droite de la régression.
3. Tester si $\beta > 0$. (Cela veut dire, tester, si la moyenne des salaires des hommes est significativement plus grande que celle des femmes)($\widehat{\sigma}_\varepsilon^2 = 345354$).
4. Estimer ces moyennes.
5. Donner un intervalle de confiance pour β.

Solution

1. La droite de régression :

$$\widehat{\beta} = \frac{\sum_{i=1}^{49}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{49}(x_i - \bar{x})^2} = \frac{s_{xy} \times (n-1)}{s_x^2 \times (n-1)} = \frac{\sum_{i=1}^{49} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{49} x_i^2 - n\bar{x}^2} = \frac{n s_{xy}}{n s_x^2} = \frac{6934.694}{12.204} = 568.227$$

et

$$\widehat{\beta}_0 = \bar{y} - \widehat{\beta}\bar{x} = 1518.696.$$

Le modèle linéaire est :

$$\widehat{y} = 1518.696 + 568.227\ x.$$

2. **En langage R**

```
y=WAGE
x=GENDER
boxplot(y[x==0],y[x==1])# boxplot des hommes et des femmes
model3=lm(y~x)  # Modele lineaire
summary(model3) # Coefficients de la regression
       # student d'alpha et student de beta
    # R^2
plot(WAGE~GENDER)
abline(model3)
```

3. Tester si $\beta > 0$. Dans ce cas, comme test alternatif, on doit utiliser un test unilatéral :

$$H_0\ : \beta = 0 \text{ contre } H_A\ : \beta > 0.$$

la statistique de test est :

$$t_\beta = \frac{\widehat{\beta} - 0}{se(\widehat{\beta})} \sim t_{n-2} \text{ quand } H_0 \text{ est vraie .}$$

Comme

$$se(\widehat{\beta}) = \frac{\widehat{\sigma}_\varepsilon^2}{\sqrt{n s_x^2}} = \sqrt{\widehat{\text{var}}(\widehat{\beta})} = \sqrt{28298.22127} = 168.221,$$

$$t_\beta = 3.378.$$

Comme notre test alternatif est un test unilatéral, on doit comparer cette valeur avec la valeur du quantile : $t_{1-\alpha,n-2} = t_{0.95,47}$. Dans la table, on trouve $t_{0.95,40} = 1.68$ et $t_{0.95,50} = 1.67$. On peut voir qu'à ce niveau de confiance, la statistique de test est plus grande ces deux t-valeurs des quantiles. Donc, on rejette l'hypothèse. On conclut que β est plus grand que 0, cela veut dire que la moyenne des salaires des hommes est significativement plus grande que celle des femmes.

4. La moyenne estimée des salaires des femmes :

$$\widehat{\overline{y}}|_{x=0} = 1518.696 + 568.227 \times 0 = 1518.696$$

et celle des hommes :

$$\widehat{\overline{y}}|_{x=1} = 1518.696 + 568.227 \times 1 = 2086.923.$$

5. L'intervalle de confiance de β_1 :

$$\text{IC}_\beta = \left[\widehat{\beta} - t_{1-\alpha/2,n-2} se(\widehat{\beta}), \widehat{\beta} + t_{1-\alpha/2,n-2} se(\widehat{\beta})\right],$$

ce qui vaut dans notre cas : $\text{IC}_\beta = [230.104, 906.351]$.

Chapitre 15

Exercices : Modèle linéaire général

Exercice 15.1. Deux variables explicatives
On dispose des données suivantes :

x_2	x_3	y
-2	4	18
1	3	12
0	-6	10
-1	-5	16
2	3	11
0	-3	9
-1	6	11
1	-1	8
1	0	7
-1	-1	12

Soit le modèle linéaire :

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i.$$

1. Estimer les paramètres du modèle par la méthode des moindres carrés et écrire l'équation. (Utiliser les calculs matriciels)
2. Donner un estimateur sans biais de la variance des erreurs.
3. Tester, à l'aide des statistiques de Student, les hypothèses :
   $H_0 : \beta_1 = 0$ vs $\beta_1 \neq 0$ au seuil 5%.
   $H_0 : \beta_2 = 0$ vs $\beta_2 \neq 0$ au seuil 5%.
   $H_0 : \beta_3 = 0$ vs $\beta_3 \neq 0$ au seuil 5%.
4. Donner des intervalles de confiance des coefficients de la droite de la régression.(β_1, β_2 et β_3).
5. Dresser le tableau de l'analyse de la variance (ANOVA).
6. Effectuer le F-test ($\alpha = 5\%$) et commenter.
7. Calculer le coefficient de détermination R^2 et commenter.
8. Vérifier vos résultats en utilisant le logiciel R.

Table 15.1 – Tableau de l'analyse de la variance

Source	d.l.	SC	CM	F_{obs}
Régression	$\cdots$	$\cdots$	$\cdots$	$\cdots$
Résiduelle	$\cdots$	$\cdots$	$\cdots$	
Totale	$\cdots$	$\cdots$		

Solution

1. On a $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$ et $n = 10$

$$\mathrm{y}^\top \mathrm{y} = \sum_{i=1}^{n} y_i^2 = 1404, \ \sum_{i=1}^{n} y_i = 114, \ \text{d'où } \bar{y} = 11.4, \ \sum_{i=1}^{n} y_i x_{i2} = -26, \ \sum_{i=1}^{n} y_i x_{i3} = 20,$$

$$\sum_{i=1}^{n} x_{i2}^2 = 14,\ \sum_{i=1}^{n} x_{i3}^2 = 142,\ \sum_{i=1}^{n} x_{i2} = \sum_{i=1}^{n} x_{i3} = \sum_{i=1}^{n} x_{i2}x_{i3} = 0.$$

$$\mathrm{X}^\top \mathrm{y} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n y_i x_{i2} \\ \sum_{i=1}^n y_i x_{i3} \end{pmatrix} = \begin{pmatrix} 114 \\ -26 \\ 20 \end{pmatrix}, \mathrm{X}^\top \mathrm{X} = \begin{pmatrix} n & \sum_{i=1}^n x_{i2} & \sum_{i=1}^n x_{i3} \\ \sum_{i=1}^n x_{i2} & \sum_{i=1}^n x_{i2}^2 & \sum_{i=1}^n x_{i2}x_{i3} \\ \sum_{i=1}^n x_{i3} & \sum_{i=1}^n x_{i2}x_{i3} & \sum_{i=1}^n x_{i3}^2 \end{pmatrix} = \begin{pmatrix} 10 & 0 & 0 \\ 0 & 14 & 0 \\ 0 & 0 & 142 \end{pmatrix},$$

$$\left(\mathrm{X}^\top \mathrm{X}\right)^{-1} = \begin{pmatrix} \frac{1}{10} & 0 & 0 \\ 0 & \frac{1}{14} & 0 \\ 0 & 0 & \frac{1}{142} \end{pmatrix}.$$

Donc,

$$\widehat{\boldsymbol{\beta}} = (\mathrm{X}^\top \mathrm{X})^{-1} \mathrm{X}^\top \mathrm{y} = \begin{pmatrix} \frac{1}{10} & 0 & 0 \\ 0 & \frac{1}{14} & 0 \\ 0 & 0 & \frac{1}{142} \end{pmatrix} \begin{pmatrix} 114 \\ -26 \\ 20 \end{pmatrix} = \begin{pmatrix} 11.4 \\ -1.857 \\ 0.141 \end{pmatrix}.$$

L'équation de la régression est

$$y_i = 11.4 - 1.857 x_{i2} + 0.141 x_{i3} + e_i.$$

2. L'estimateur sans biais de la variance des erreurs :

$$\widehat{\sigma}_\varepsilon^2 = \frac{\mathrm{SC}_{RES}}{n-p} = \frac{\mathrm{SC}_{RES}}{10-3} = \frac{\mathrm{SC}_{RES}}{7} = \frac{\mathrm{y}^\top \mathrm{y} - \boldsymbol{\beta}^\top \mathrm{X}^\top \mathrm{y}}{7},$$

où

$$\mathrm{y}^\top \mathrm{y} = \sum_{i=1}^{n} y_i^2 = 1404$$

et

$$\boldsymbol{\beta}^\top \mathrm{X}^\top \mathrm{y} = \begin{pmatrix} 11.4 & -1.857 & 0.141 \end{pmatrix} \begin{pmatrix} 114 \\ -26 \\ 20 \end{pmatrix} = 1350.70262.$$

Donc,

$$\widehat{\sigma}_\varepsilon^2 = \frac{53.297}{7} = 7.614.$$

3. Les tests : $H_0 : \beta_j = 0 \quad \text{vs} \quad \beta_j \neq 0$. Donc,

$$t_{\beta_j} = \frac{\widehat{\beta}_j - 0}{se(\widehat{\beta}_j)},$$

où

$$se(\widehat{\beta}_j) = \sqrt{\widehat{\beta}_j} = \sqrt{\widehat{\sigma}_\varepsilon^2 (\mathrm{X}^\top \mathrm{X})_{j,j}^{-1}}.$$

On peut construire le tableau :

	Coef	se	t_{obs}	t_{hyp}
Constante	11.400	0.873	13.058	2.36
x_2	−1.857	0.738	−2.516	2.36
x_3	0.141	0.232	0.608	2.36

Donc, on rejette les hypothèses :
$H_0 : \beta_1 = 0 \quad \text{vs} \quad \beta_1 \neq 0$ au seuil 5%,
et
$H_0 : \beta_2 = 0 \quad \text{vs} \quad \beta_2 \neq 0$ au seuil 5%,
car $|t_{obs}| > t_{hyp}$. D'autre part, on ne rejette pas l'hypothèse
$H_0 : \beta_3 = 0 \quad \text{vs} \quad \beta_3 \neq 0$ au seuil 5%.
On conclut que la constante et le coefficient de la variable x_2 sont significativement différents de 0. Cependant, la variable x_3 semble être inutile dans le modèle.

4. L'intervalle de confiance de β_j (à un seuil de α) est l'intervalle autour du coéfficient $\widehat{\beta}_j$ définie par :

$$\text{IC} = (\widehat{\beta}_j - t_{1-\alpha/2,n-p} se(\widehat{\beta}_j), \widehat{\beta}_j + t_{1-\alpha/2,n-p} se(\widehat{\beta}_j)).$$

Donc,

$$\text{IC}_{\beta_1} = (9.335, 13.465), \text{IC}_{\beta_2} = (-3.602, -0.112) \text{ et } \text{IC}_{\beta_3} = (-0.408, 0.690).$$

5. Les sommes de carrés sont :

$$\text{SC}_{TOT} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 = \quad 1404 - 10 \times 11.4^2 = 104.4$$

$$\text{SC}_{RES} = 53.297$$

et

$$\text{SC}_{REG} = \text{SC}_{TOT} - \text{SC}_{RES} = 104.4 - 53.297 = 51.103.$$

Source	d.l.	SC	CM	F_{obs}
Régression	$p - 1 = 3 - 1 = 2$	51.103	25.551	3.356
Résiduelle	$n - 1 - (p - 1) = n - p = 7$	53.297	7.614	
Totale	$n - 1 = 9$	104.400		

6. Comme

$$F_{obs} = 3.356 < 4.74 = F_{hyp} = F_{2,7,0.95} = F_{p-1,n-p,1-\alpha},$$

on ne rejette pas l'hypothèse que tous les coefficients (sauf la constante) de la droite de la régression sont différents de 0. On avait également vu dans le point précédent que β_3 n'est pas significativement différent de 0.

7. Le R^2 vaut :

$$R^2 = \frac{\text{SC}_{REG}}{\text{SC}_{TOT}} = \frac{51.103}{104.400} = 0.48949 = 48.949\%.$$

Le modèle n'explique que 48.949% de la variance de la variable dépendante (y).

8. **En langage R**

```
y=c(18,12,10,16,11,9,11,8,7,12)
x2=c(-2,1,0,-1,2,0,-1,1,1,-1);x3=c(4,3,-6,-5,3,-3,6,-1,0,-1)
n=length(y)
model=lm(y~x2+x3)
summary(model) # 1- coefficients
        # 3- tests sur les betas
        # 6- F-statistic
        # 7- Multiple R-square
anova(model)   # 2- sigma chapeau
        # 5- SCres=53.297 SCtot=48.286+2.817+53.297=104.4
        #  SCreg=48.286+2.817=51.103
confint(model) # 4- intervalles de confiances des betas
```

Chapitre 16

Exercices : Inférence dans le modèle linéaire

Exercice 16.1. Inférence
Voici des estimations d'une régression linéaire : $\widehat{\boldsymbol{\beta}} = \begin{pmatrix} 116.8 & 0.54 & -1.67 \end{pmatrix}^\top$.

$$se(\widehat{\beta}_1) = \sqrt{\widehat{\sigma}^2_\varepsilon [(\mathrm{X}^\top \mathrm{X})^{-1}]_{11}} = 452, se(\widehat{\beta}_2) = \sqrt{\widehat{\sigma}^2_\varepsilon [(\mathrm{X}^\top \mathrm{X})^{-1}]_{22}} = 0.21, se(\widehat{\beta}_3) = \sqrt{\widehat{\sigma}^2_\varepsilon [(\mathrm{X}^\top \mathrm{X})^{-1}]_{33}} = 1.79.$$

Le Tableau 16.1 contient d'analyse de la variance.

Table 16.1 – Tableau de l'analyse de la variance

Source	d.l.	SC	CM	F_{obs}
Régression	$\cdots$	19848749	$\cdots$	$\cdots$
Résiduelle	$\cdots$	$\cdots$	$\cdots$	
Totale	$\cdots$	81003467		

On sait que n=30.

1. Donner les éléments manquants du tableau d'analyse de la variance et faire le test de F à 5%.
2. Donner les p-valeurs des hypothèses suivantes : $H_0 \ : \ \beta_j = 0 \quad$ vs $\quad \beta_j \neq 0,$ pour $j = 1, 2, 3$.

Solution

1. Le tableau complété :

Source	d.l.	SC	CM	F_{obs}
Régression	$p - 1 = 2$	19848749	9924374.500	4.382
Résiduelle	$n - p = 27$	61154718	2.264989.556	
Totale	$n - 1 = 29$	81003467		

$$F_{obs} = 4.382 > 3.35 = F_{hyp} = F_{2,7,0.95} = F_{p-1,n-p,1-\alpha},$$

donc on rejette l'hypothèse :

$$H_0 \ : \ \beta_j = 0 \quad \text{pour tout les } j \neq 1 \text{ vs } H_1 \ : \ \text{au moins un des } \beta_j \neq 0.$$

2. On calcule les statistiques de test :

$$t_{\beta_j} = \frac{\widehat{\beta}_j - \beta_{j0}}{\widehat{\sigma}(\widehat{\beta}_j)},$$

$$t_{\beta_1} = \frac{116.8 - 0}{452} = 0.2584, t_{\beta_2} = \frac{0.54 - 0}{0.21} = 2.5714, t_{\beta_3} = \frac{-1.67 - 0}{1.79} = -0.9330.$$

La p-valeur associée à l'observation d'une statistique de test est le seuil auquel on rejetterait l'hypothèse nulle compte tenu de cette observation. Donc,

$$p\text{-valeur} = \alpha_{min}, \text{ où } t_{1-\frac{\alpha}{2},n-p} > t_{obs}.$$

Autrement : $\Pr(|t_{obs}| < t_{1-\frac{\alpha}{2},n-p})$, la probabilité d'observer une statistique aussi grande, si les paramètres de la Student sont définis selon h_0. (En R : 2*pt(-0.2584,27), pour β_1).

$$p\text{-valeurs} = \begin{pmatrix} 0.79 \\ 0.01 \\ 0.35 \end{pmatrix}.$$

Exercice 16.2. Inférence et test
On considère le modèle :

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i, i = 1, \ldots, 10.$$

Les données de l'échantillon sont résumées de la façon suivante :

$$\sum_{i=1}^{n} y_i^2 = 177, \ \sum_{i=1}^{n} y_i = 10, \ \sum_{i=1}^{n} y_i x_{i2} = 20, \ \sum_{i=1}^{n} y_i x_{i3} = 40,$$

$$\sum_{i=1}^{n} x_{i2}^2 = 5, \ \sum_{i=1}^{n} x_{i3}^2 = 20, \ \sum_{i=1}^{n} x_{i2} = \sum_{i=1}^{n} x_{i3} = \sum_{i=1}^{n} x_{i2} x_{i3} = 0.$$

1. Construire le tableau d'analyse de la variance,
2. Calculer le R^2,
3. Tester, au seuil de signification $\alpha = 0.05$, les hypothèses suivantes :
   (a) $H_0 : \beta_j = 0$ pour tout les j $\neq 1$ vs H_1 : au moins un des $\beta_j \neq 0$,
   (b) $H_0 : \beta_3 = 3$ vs $\beta_3 \neq 3$,
   (c) $H_0 : \beta_3 \leq 1.5$ vs $\beta_3 > 1.5$,
   (d) $H_0 : \beta_2 + \beta_3 = 8$ vs $\beta_2 + \beta_3 \neq 8$.

Solution
On a $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$ et $n = 10$

$$\mathrm{y}^\top \mathrm{y} = \sum_{i=1}^{n} y_i^2 = 177, \ \sum_{i=1}^{n} y_i = 10, \ \text{d'où } \bar{y} = 1$$

$$\sum_{i=1}^{n} y_i x_{i2} = 20, \ \sum_{i=1}^{n} y_i x_{i3} = 40,$$

$$\sum_{i=1}^{n} x_{i2}^2 = 5, \ \sum_{i=1}^{n} x_{i3}^2 = 20, \ \sum_{i=1}^{n} x_{i2} = \sum_{i=1}^{n} x_{i3} = \sum_{i=1}^{n} x_{i2} x_{i3} = 0,$$

$$\mathrm{X}^\top \mathrm{y} = \begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} y_i x_{i2} \\ \sum_{i=1}^{n} y_i x_{i3} \end{pmatrix} = \begin{pmatrix} 10 \\ 20 \\ 40 \end{pmatrix}$$

$$\mathrm{X}^\top \mathrm{X} = \begin{pmatrix} n & \sum_{i=1}^{n} x_{i2} & \sum_{i=1}^{n} x_{i3} \\ \sum_{i=1}^{n} x_{i2} & \sum_{i=1}^{n} x_{i2}^2 & \sum_{i=1}^{n} x_{i2} x_{i3} \\ \sum_{i=1}^{n} x_{i3} & \sum_{i=1}^{n} x_{2i} x_{i3} & \sum_{i=1}^{n} x_{i3}^2 \end{pmatrix} = \begin{pmatrix} 10 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 20 \end{pmatrix}$$

$$\left(\mathrm{X}^\top \mathrm{X}\right)^{-1} = \begin{pmatrix} \frac{1}{10} & 0 & 0 \\ 0 & \frac{1}{5} & 0 \\ 0 & 0 & \frac{1}{20} \end{pmatrix}$$

$$\widehat{\boldsymbol{\beta}} = \begin{pmatrix} \frac{1}{10} & 0 & 0 \\ 0 & \frac{1}{5} & 0 \\ 0 & 0 & \frac{1}{20} \end{pmatrix} \begin{pmatrix} 10 \\ 20 \\ 40 \end{pmatrix} = \begin{pmatrix} 1 \\ 4 \\ 2 \end{pmatrix}$$

1. Les sommes de carrés :

$$\mathrm{SC}_{TOT} = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 = 177 - 10 \times 1^2 = 167$$

$$\mathrm{SC}_{RES} = \mathrm{e}^\top \mathrm{e} = \mathrm{y}^\top \mathrm{y} - \boldsymbol{\beta}^\top \mathrm{X}^\top \mathrm{y} = 177 - \begin{pmatrix} 1 & 4 & 2 \end{pmatrix} \begin{pmatrix} 10 \\ 20 \\ 40 \end{pmatrix} = 177 - 170 = 7.$$

$$\mathrm{SC}_{REG} = \mathrm{SC}_{TOT} - \mathrm{SC}_{RES} = 167 - 7 = 160.$$

Source	d.l.	SC	CM	F_{obs}
Régression	$p-1=3-1=2$	160	80	80
Résiduelle	$n-1-(p-1)=n-p=7$	7	1	
Totale	$n-1=9$	167		

2. Le R^2 :
$$R^2 = \frac{\text{SC}_{REG}}{\text{SC}_{TOT}} = 1 - \frac{\text{SC}_{RES}}{\text{SC}_{TOT}} = 1 - \frac{7}{167} = 0.96.$$

3. Test d'hypothèses avec $\alpha = 0.05$:

   (a) Test de Wald pour le test global des coefficients :
   $$F_{obs} = 80 > 4.74 = F_{hyp} = F_{2,7,0.95} = F_{p-1,n-p,1-\alpha}.$$
   On rejette l'hypothèse : $H_0 \ : \ \beta_j = 0 \quad$ pour tout les j $\neq 1$ vs $H_1 \ :$ au moins un des $\beta_j \neq 0$. Au moins une des variables est significative dans notre modèle.

   (b) Test de Student sur un coefficient : $H_0 \ : \ \beta_3 = 3 \quad$ vs $\quad \beta_3 \neq 3$,
   $$t_{obs} = \frac{\widehat{\beta}_3 - \beta_{30}}{se(\widehat{\beta}_3)} = \frac{2-3}{\sqrt{\frac{1}{20}\left(\frac{7}{7}\right)}} = -4.472,$$
   $$|t_{obs}| > t_{0.975,10-3} = 2.36.$$
   Donc, on rejette H_0.

   (c) Test de Student sur un coefficient : $H_0 \ : \ \beta_3 \leq 1.5 \quad$ vs $\quad \beta_3 > 1.5$. C'est un test unilatéral.
   $$t_{obs} = \frac{2-1.5}{\sqrt{\frac{1}{20}}} = 2.236 > t_{0.95,10-3} = 1.9.$$
   Donc, on rejette H_0. (remarque : on met le signe $\leq$ ou $\geq$ dans l'hypothèse H_0.)

   (d) Test de Wald avec matrice R : $H_0 \ : \ \beta_2 + \beta_3 = 8 \quad$ vs $\quad \beta_2 + \beta_3 \neq 8$, $H_0 \ : \text{R}\boldsymbol{\beta} = r$, où $r = 8$ et $\text{R} = \begin{pmatrix} 0 & 1 & 1 \end{pmatrix}$.
   Donc,
   $$\text{R}\widehat{\boldsymbol{\beta}} - r = 4 + 2 - 8 = -2,$$
   $$\text{R}\left(\text{X}^\top\text{X}\right)^{-1}\text{R}^\top = \begin{pmatrix} 0 & \frac{1}{5} & \frac{1}{20} \end{pmatrix}\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = \frac{5}{20} = \frac{1}{4}.$$
   La statistique de test :
   $$t_{obs} = \frac{\text{R}\widehat{\boldsymbol{\beta}} - r}{\widehat{\sigma}\sqrt{\text{R}\left(\text{X}^\top\text{X}\right)^{-1}\text{R}^\top}} = \frac{-2}{1\sqrt{\frac{1}{4}}} = -4,$$
   $$|t_{obs}| = 4 > t_{0.975,10-3} = 2.365.$$
   On rejette H_0, car
   $$F_{obs} = \frac{(\text{R}\widehat{\beta} - r)^2}{\widehat{\sigma_\varepsilon}\left[\text{R}\left(\text{X}^\top\text{X}\right)^{-1}\text{R}^\top\right]} = \frac{4}{1/4} = 16 > F_{1-\alpha,1,10-3} = 5.59.$$

Exercice 16.3. Tests de Wald
On dispose d'un échantillon de 30 observations et on connaît les résultats suivants :

$$\sum_{i=1}^{30}(y_i-\bar{y})^2=150,\ \sum_{i=1}^{30}(x_{i2}-\bar{x}_2)^2=45,\ \sum_{i=1}^{30}(x_{i3}-\bar{x}_3)^2=90,$$

$$\sum_{i=1}^{30}(y_i-\bar{y})(x_{i2}-\bar{x}_2)=45,\ \sum_{i=1}^{30}(y_i-\bar{y})(x_{i3}-\bar{x}_3)=30,\ \sum_{i=1}^{30}(x_{i2}-\bar{x}_2)(x_{i3}-\bar{x}_3)=60,$$

$$\bar{y}=15,\bar{x}_2=3,\bar{x}_3=3.$$

On considère le modèle suivant :

$$y_i=\beta_1+\beta_2x_{i2}+\beta_3x_{i3}+\varepsilon_i,$$

où y_i, x_{i2} et x_{i3} sont les variables non-centrées de l'échantillon. On suppose que les erreurs, ε_i sont homoscédastiques, de variance σ_ε^2, non-corrélées, d'espérance nulle et ont une distribution normale.

1. Écrire le modèle sous forme matricielle $\mathrm{y}=\mathrm{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$ (en précisant le contenu et les dimensions de chaque vecteur et matrice).
2. Estimer les 3 paramètres du modèle en utilisant la méthode des MCO (suggestion : utiliser les données centrées).
3. Donner l'estimateur sans biais ($\widehat{\sigma_\varepsilon^2}$) de la variance des erreurs (σ_ε^2).
4. Dresser le tableau d'analyse de la variance. Calculer la statistique de Fisher associée.
5. Tester (avec $\alpha=0.05$) $H_0\ :\beta_3=0$ contre $H_1\ :\beta_3\neq0$.
6. Tester (avec $\alpha=0.05$) $H_0\ :\beta_2=\beta_3=0$ contre $H_1\ :$ au moins un de ces coefficients est non nul.
7. Sous l'hypothèse de normalité des erreurs, tester avec $\alpha=0.05$ l'hypothèse selon laquelle $\beta_2+\beta_3=1.8$.
8. Calculer et interpréter le coefficient de détermination R^2.
9. Donner une prévision optimale de y_{31} sachant qu'on connaît $x_{31,2}=10$ et $x_{31,3}=61$.

Solution

1. La forme matricielle du modèle est :

$$\begin{pmatrix}y_1\\ \vdots\\ y_{30}\end{pmatrix}=\begin{pmatrix}1 & x_{1,2} & x_{1,3}\\ \vdots & \vdots & \vdots\\ 1 & x_{30,2} & x_{30,3}\end{pmatrix}\begin{pmatrix}\beta_1\\ \beta_2\\ \beta_3\end{pmatrix}+\begin{pmatrix}\varepsilon_1\\ \vdots\\ \varepsilon_{30}\end{pmatrix},$$

$$\underset{(30\times1)}{\mathrm{y}}=\underset{(30\times3)}{\mathrm{X}}\underset{(3\times1)}{\boldsymbol{\beta}}+\underset{(30\times1)}{\boldsymbol{\varepsilon}},$$

$$\mathrm{X}^\top\mathrm{X}=\begin{pmatrix}1 & \cdots & 1\\ x_{1,2} & \cdots & x_{30,2}\\ x_{1,3} & \cdots & x_{30,3}\end{pmatrix}\begin{pmatrix}1 & x_{1,2} & x_{1,3}\\ \vdots & \vdots & \vdots\\ 1 & x_{30,2} & x_{30,3}\end{pmatrix}=\begin{pmatrix}n & \sum_{i=1}^n x_{i2} & \sum_{i=1}^n x_{i3}\\ \sum_{i=1}^n x_{i2} & \sum_{i=1}^n x_{i2}^2 & \sum_{i=1}^n x_{i2}x_{i3}\\ \sum_{i=1}^n x_{i3} & \sum_{i=1}^n x_{i2}x_{i2} & \sum_{i=1}^n x_{i3}^2\end{pmatrix},$$

$$\mathrm{X}^\top\mathrm{y}=\begin{pmatrix}1 & \cdots & 1\\ x_{1,2} & \cdots & x_{30,2}\\ x_{1,3} & \cdots & x_{30,3}\end{pmatrix}\begin{pmatrix}y_1\\ \vdots\\ y_{30}\end{pmatrix}=\begin{pmatrix}\sum_{i=1}^n y_i\\ \sum_{i=1}^n y_ix_{i2}\\ \sum_{i=1}^n y_ix_{i3}\end{pmatrix}.$$

2. Dans notre cas :

$$\widetilde{\beta}=(\mathrm{X}_c^\top\mathrm{X}_c)^{-1}\mathrm{X}_c^\top\mathrm{y}_c,$$

où

$$\mathrm{X}_c^\top\mathrm{X}_c=\begin{pmatrix}ns_{x_2}^2=\sum_{i=1}^n(x_{i2}-\bar{x}_2)^2 & ns_{x_2x_3}=\sum_{i=1}^n(x_{i2}-\bar{x}_2)(x_{i3}-\bar{x}_3)\\ ns_{x_2x_3}=\sum_{i=1}^n(x_{i2}-\bar{x}_2)(x_{i3}-\bar{x}_3) & ns_{x_3}^2=\sum_{i=1}^n(x_{i3}-\bar{x}_3)^2\end{pmatrix}$$

et

$$\mathrm{X}_c^\top\mathrm{y}_c=\begin{pmatrix}\sum_{i=1}^n(y_i-\bar{y})(x_{i2}-\bar{x}_2)\\ \sum_{i=1}^n(y_i-\bar{y})(x_{i3}-\bar{x}_3)\end{pmatrix}=\begin{pmatrix}ns_{yx_2}\\ ns_{yx_3}\end{pmatrix}.$$

Donc,

$$\mathrm{X}_c^\top \mathrm{X}_c = \begin{pmatrix} 45 & 60 \\ 60 & 90 \end{pmatrix},$$

$$(\mathrm{X}_c^\top \mathrm{X}_c)^{-1} = \begin{pmatrix} 0.200 & -0.133 \\ -0.133 & 0.100 \end{pmatrix}$$

et

$$\mathrm{X}_c^\top \mathrm{y}_c = \begin{pmatrix} 45 \\ 30 \end{pmatrix}.$$

Ainsi,

$$\widehat{\widetilde{\boldsymbol{\beta}}} = (\mathrm{X}_c^\top \mathrm{X}_c)^{-1} \mathrm{X}_c^\top \mathrm{y}_c = \begin{pmatrix} 0.200 & -0.133 \\ -0.133 & 0.100 \end{pmatrix} \begin{pmatrix} 45 \\ 30 \end{pmatrix} = \begin{pmatrix} 5 \\ -3 \end{pmatrix} = \begin{pmatrix} \widehat{\beta}_2 \\ \widehat{\beta}_3 \end{pmatrix}.$$

On détermine $\widehat{\beta}_1$ en utilisant :

$$\widehat{\beta}_1 = \bar{y} - \widehat{\beta}_2 \bar{x}_2 - \widehat{\beta}_3 \bar{x}_3 = 15 - 5 \times 3 - -3 \times 3 = 9.$$

3. La somme des carrés de la régression est :

$$\mathrm{SC}_{REG} = \widetilde{\beta}^\top \mathrm{X}_c^\top \mathrm{X}_c \widetilde{\beta} = \begin{pmatrix} 5 & -3 \end{pmatrix} \begin{pmatrix} 45 & 60 \\ 60 & 90 \end{pmatrix} \begin{pmatrix} 5 \\ -3 \end{pmatrix} = 135.$$

Comme

$$\mathrm{SC}_{TOT} = 150,$$

on obtient la somme des carrés des résidus par différence.

$$\mathrm{SC}_{RES} = 150 - 135 = 15$$

et les carrés moyens :

$$\mathrm{CM}_{res} = \frac{15}{27} = 0.556 = \widehat{\sigma}_\varepsilon^2.$$

4. Tableau de l'analyse de la variance :

Source	d.l.	SC	CM	F_{obs}
Régression	$p - 1 = 3 - 1 = 2$	135	67.500	121.500
Résiduelle	$n - 1 - (p - 1) = n - p = 27$	15	0.556	
Totale	$n - 1 = 29$	150		

5. Test de Student avec $\alpha = 0.05$:
$H_0 : \beta_3 = 0$ contre $H_1 : \beta_3 \neq 0$. On a si $[(\mathrm{X}^\top \mathrm{X})^{-1}]_{33} = [(\mathrm{X}_c^\top \mathrm{X}_c)^{-1}]_{22} = 0.1$. Donc,

$$t_{obs} = \frac{\widehat{\beta}_3 - 0}{se(\widehat{\beta}_3)} = \frac{-3}{\sqrt{0.1 \times 0.556}} = -12.728.$$

Comme $|t_{obs}| > t_{1-\alpha/2,n-p} = t_{0.975,27} = 2.052$, on rejette H_0 et on conclut que β_3 n'est pas égal à 0.

6. Test de Wald : test global des coefficients : On a

$$F_{obs} = 121.500 > F_{2,27,0.95} = 3.354.$$

Donc, on rejette H_0.

7. Test de Wald $H_0 : \beta_2 + \beta_3 = 1.8$.

$$\begin{cases} H_0 : & \beta_2 + \beta_3 = 1.8 \\ H_1 : & \beta_2 + \beta_3 \neq 1.8. \end{cases}$$

On pose $Z = \beta_2 + \beta_3$. Donc,

$$\begin{cases} H_0 : & Z = 1.8 \\ H_1 : & Z \neq 1.8. \end{cases}$$

Comme

$$\widehat{\mathrm{var}}(Z) = \widehat{\mathrm{var}}(\widehat{\beta}_2 + \widehat{\beta}_3) = \widehat{\mathrm{var}}(\widehat{\beta}_2) + \widehat{\mathrm{var}}(\widehat{\beta}_3) + 2\widehat{\mathrm{cov}}(\widehat{\beta}_2, \widehat{\beta}_3),$$

on a :

$$\begin{aligned} |t| &= \left| \frac{\widehat{Z} - 1.8}{\sqrt{\widehat{\text{var}}(\widehat{Z})}} \right| = \left| \frac{\widehat{\beta}_2 + \widehat{\beta}_3 - 1.8}{\sqrt{\widehat{\text{var}}(\widehat{\beta}_2) + \widehat{\text{var}}(\widehat{\beta}_3) + 2\widehat{\text{cov}}(\widehat{\beta}_2, \widehat{\beta}_3)}} \right| \\ &= \left| \frac{0.2}{\sqrt{0.111 + 0.056 + 2 \times -0.074}} \right| = |1.470| < 2.052. \end{aligned}$$

On ne rejette pas H_0.

8. Calcul du R^2 :

$$R^2 = \frac{135}{150} = 0.9.$$

9. Donner une prévision optimale de y_{31} sachant qu'on connaît $x_{31,2} = 10$ et $x_{31,3} = 61$:

$$\widehat{y}_{31} = 9 + 5 \times 10 + -3 \times 61 = -124.$$

Exercice 16.4. Relation F et R^2
Montrer que :

$$R^2 = \frac{\nu_1 F}{\nu_1 F + \nu_2},$$

où F est la statistique F_{obs} et ν_1 et ν_2 correspondant au nombre de degrés de liberté de la régression et de degrés de liberté des résidus, respectivement.

Solution

$$R^2 = \frac{\text{SC}_{REG}}{\text{SC}_{TOT}} = \frac{\text{SC}_{REG}}{\text{SC}_{REG} + \text{SC}_{RES}} = \frac{\nu_1 \text{CM}_{reg}}{\nu_1 \text{CM}_{reg} + \nu_1 \text{CM}_{res}} = \frac{\nu_1 \text{CM}_{reg}/\text{CM}_{res}}{\nu_1 \text{CM}_{reg}/\text{CM}_{res} + \nu_1} = \frac{\nu_1 F}{\nu_1 F + \nu_1}$$

Exercice 16.5. Vraisemblance
Soit $x_1, \ldots, x_i, \ldots, x_n$ une suite de variables aléatoires indépendantes et identiquement distribuées (i.i.d.) de loi $\mathcal{N}(\mu, \sigma^2)$.

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

2. On considère que μ est connue. Estimez σ^2 par la méthode du maximum de vraisemblance.

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

Solution

1. On considère que σ^2 est connue. Estimons μ par la méthode du maximum de vraisemblance.

$$f_{x_i}(u_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(u_i-\mu)^2}{2\sigma^2}\right), u_i \in \mathbb{R}.$$

Par l'independance, on a :

$$f_X(u_1, \ldots, u_n) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(u_i-\mu)^2}{2\sigma^2}\right) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(\sum_{i=1}^{n} -\frac{(u_i-\mu)^2}{2\sigma^2}\right).$$

La fonction de vraisemblance dépend de μ :

$$L(\mu \mid x_1, \ldots, x_n) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right).$$

Le log de la fonction (log-likelihood) :

$$\ell(\mu) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2.$$

On annule la dérivée de ℓ :

$$\frac{\partial \ell(\mu)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = 0.$$

Donc,

$$\sum_{i=1}^{n} x_i - n\mu = 0,$$

et

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

2. On considère maintenant que μ est connue. Estimons σ^2 par la méthode du maximum de vraisemblance. On a, en reprenant les calculs précédents :

$$\ell(x_i \cdots x_n \mid \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2,$$

$$\frac{\partial \ell(\sigma^2)}{\partial \sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} + \sum_{i=1}^{n}(x_i-\mu)^2\frac{1}{2}\frac{1}{\sigma^4} = 0.$$

Donc,

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2.$$

3. On considère enfin que μ et σ^2 sont inconnues. Estimons μ et σ^2 par la méthode du maximum de vraisemblance. On a :

$$\ell(\sigma^2, \mu) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log \sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2.$$

On annule la dérivée partielle :

$$\frac{\partial \ell(\sigma^2, \mu)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0.$$

Donc,

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

De plus,

$$\frac{\partial \ell(\sigma^2, \mu)}{\partial \sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} + \sum_{i=1}^{n}(x_i - \mu)^2 \frac{1}{2}\frac{1}{\sigma^4} = 0.$$

Donc,

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2.$$

Mais on doit remplacer μ par $\widehat{\mu}$ pour ne pas contredire la définition d'un estimateur. Donc,

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \widehat{\mu})^2.$$

Cet estimateur est légèrement biaisé. En effet,

$$\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \widehat{\mu})^2$$

est sans biais de σ^2.

Exercice 16.6. Estimateurs linéaires
On considère le modèle linéaire sans constante suivant :

$$y_i = \beta x_i + \varepsilon_i, \quad \text{avec } i = 1, \dots n \quad \text{et } \varepsilon_i \sim N(0, \sigma^2). \tag{16.1}$$

Les ε_i sont non-indépendants.

1. Recherche du meilleur estimateur de β :
   (a) On pose $\beta^* = \sum_{i=1}^{n} c_i y_i$ un estimateur linéaire de β. Les c_i sont des constantes. Sous quelles conditions β^* est-il sans biais ?
   (b) Donner l'expression de la variance de β^*.
   (c) Le problème est donc de déterminer les c_i tels que la variance de β^* soit minimale. Écrire le programme d'optimisation sous contrainte et déterminer l'expression de β^*.
2. Estimation de l'équation (16.1) par MCO : déterminer l'expression de $\widehat{\beta}$ l'estimateur de β par moindre carrés ordinaires.
3. Estimation par la méthode du maximum de vraisemblance :
   (a) Écrire la fonction de vraisemblance de l'échantillon $(y_1, \dots, y_n)$.
   (b) Estimer β et σ^2 par la méthode du maximum de vraisemblance.
4. Montrer que σ^2_{MV} est biaisé et calculer son biais.

Solution

1. (a) Soit $\beta^* = \sum_{i=1}^{n} c_i y_i$ un estimateur linéaire de β. Donc,

$$\mathrm{E}(\beta^*) = \mathrm{E}\left(\sum_{i=1}^{n} c_i y_i\right) = \sum_{i=1}^{n} c_i \mathrm{E}(y_i).$$

Comme $\mathrm{E}(y_i) = \beta x_i$, on a

$$\mathrm{E}(\beta^*) = \sum_{i=1}^{n} c_i \beta x_i = \beta \sum_{i=1}^{n} c_i x_i$$

La condition est $\sum_{i=1}^{n} c_i x_i = 1$.

De plus

$$\mathrm{var}(y_i) = \mathrm{var}(\beta x_i + \varepsilon_i) = \mathrm{var}(\varepsilon_i) = \sigma^2.$$

(b)

$$\begin{aligned} \mathrm{var}(\beta^*) = \mathrm{var}\left(\sum_{i=1}^{n} c_i y_i\right) &= \sum_{i=1}^{n} \mathrm{var}(c_i y_i) + \underbrace{2 \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{cov}(c_i y_i, c_j y_j)}_{0} \\ &= \sum_{i=1}^{n} c_i^2 \mathrm{var}(y_i) = \sigma^2 \sum_{i=1}^{n} c_i^2. \end{aligned}$$

(c) Le programme

$$\begin{cases} \min \ \mathrm{var}(\beta^*) \\ \text{sous la contrainte} \quad \sum_{i=1}^{n} c_i x_i = 1, \end{cases}$$

peut aussi s'écrire

$$\begin{cases} \min \sum_{i=1}^{n} c_i^2 \\ \text{sous la contrainte} \quad 2 \sum_{i=1}^{n} c_i x_i = 2. \end{cases}$$

La fonction lagrangienne est

$$\mathcal{L}_{c_i,\lambda} = \sum_{i=1}^{n} c_i^2 - 2\lambda\left(\sum_{i=1}^{n} c_i x_i - 1\right).$$

En annulant les dérivées partielles, on obtient le système de $n+1$ équations :

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial c_i} = 2c_i - 2\lambda x_i = 0, i = 1, \ldots, n \\ \frac{\partial \mathcal{L}}{\partial \lambda} = -2\left(\sum_{i=1}^{n} c_i x_i - 1\right) = 0, \end{cases}$$

ce qui donne

$$\begin{cases} c_i = \lambda_i x_i, i = 1, \ldots, n \\ \sum_{i=1}^{n} c_i x_i = 1. \end{cases}$$

Comme $c_i x_i = \lambda x_i^2$ on a

$$\sum_{i=1}^{n} c_i x_i = \lambda \sum_{i=1}^{n} x_i^2,$$

et donc

$$\lambda = \frac{1}{\sum_{i=1}^{n} x_i^2}, \quad c_i = \frac{x_i}{\sum_{i=1}^{n} x_i^2}, \quad \beta^* = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}.$$

2. Estimation par MCO. La matrice $\mathrm{X} = \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix}$. On a, de manière générale,

$$\widehat{\beta} = \left(\mathrm{X}^\top \mathrm{X}\right)^{-1} \mathrm{X}^\top \mathrm{y}.$$

Donc,

$$\mathrm{X}^\top \mathrm{X} = \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \sum_{i=1}^{n} x_i^2, \quad \left(\mathrm{X}^\top \mathrm{X}\right)^{-1} = \frac{1}{\sum_{i=1}^{n} x_i^2}, \quad \mathrm{X}^\top \mathrm{y} = \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \sum_{i=1}^{n} x_i y_i.$$

Ce qui donne

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

Comme $\widehat{\beta} = \beta^*$, l'estimateur $\widehat{\beta}$ est efficace.

3. (a) Méthode du maximum de vraisemblance :

$$\mathcal{L}(\beta, \sigma_\varepsilon^2) = \frac{1}{(2\pi\sigma_\varepsilon^2)^{\frac{n}{2}}} \exp - \frac{(\mathrm{y} - \mathrm{X}\beta)^\top (\mathrm{y} - \mathrm{X}\beta)}{2\sigma_\varepsilon^2} = \frac{1}{(2\pi\sigma_\varepsilon^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma_\varepsilon^2} \sum_{i=1}^{n} (y_i - \beta x_i)^2\right).$$

(b) Le logarithme de la fonction de vraisemblance :

$$\begin{aligned} \ell(\beta, \sigma_\varepsilon^2) &= \log \mathcal{L}(\beta, \sigma_\varepsilon^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma_\varepsilon^2) - \frac{\sum_{i=1}^{n}(y_i - \beta x_i)^2}{2\sigma_\varepsilon^2} \\ &= \log \mathcal{L}(\beta, \sigma_\varepsilon^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma_\varepsilon^2) - \frac{1}{2\sigma_\varepsilon^2}\left(\sum_{i=1}^{n} y_i^2 - 2\sum_{i=1}^{n} \beta y_i x_i + \beta^2 \sum_{i=1}^{n} x_i^2\right). \end{aligned}$$

On annule la dérivée :

$$\frac{\partial \ell}{\partial \beta} = \frac{1}{\sigma_\varepsilon^2}\left(\sum_{i=1}^{n} x_i y_i - \beta \sum_{i=1}^{n} x_i^2\right) = 0.$$

Donc,

$$\widehat{\beta}_{MV} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}.$$

De plus,

$$\frac{\partial \ell}{\partial \sigma_\varepsilon^2} = -\frac{n}{2\sigma_\varepsilon^2} + \frac{1}{2\sigma_\varepsilon^4}\sum_{i=1}^{n}(y_i - \beta x_i)^2 = \frac{1}{2\widehat{\sigma}_\varepsilon^2}\left(-n + \frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^{n}(y_i - \beta x_i)^2\right) = 0.$$

Donc,

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

$$\widehat{\sigma}_{\varepsilon MV}^2 = \frac{\mathrm{e}^{\top}\mathrm{e}}{n},$$

4. L'estimateur $\widehat{\sigma}_{\varepsilon}^2 = \frac{\mathrm{e}^{\top}\mathrm{e}}{n-1}$ est sans biais de σ_{ε}^2. Donc,

$$\widehat{\sigma}_{\varepsilon MV}^2 = \frac{n-1}{n}\widehat{\sigma}_{\varepsilon}^2$$

et

$$\mathrm{E}(\widehat{\sigma}_{\varepsilon MV}^2) = \frac{n-1}{n}\sigma_{\varepsilon}^2 \neq \sigma_{\varepsilon}^2$$

Donc, $\widehat{\sigma}_{\varepsilon MV}^2$ est biaisé.

$$\mathrm{E}(\widehat{\sigma}_{\varepsilon MV}^2) - \sigma_{\varepsilon}^2 = \frac{n-1}{n}\sigma_{\varepsilon}^2 - \sigma_{\varepsilon}^2 = \frac{-\sigma_{\varepsilon}^2}{n}.$$

Exercice 16.7.
En considérant le même modèle que dans l'Exercice 12.2, on veut tester l'hypothèse que $2\beta_1 = \beta_2$.

1. Donner une matrice de contrainte R et le vecteur r à utiliser pour construire ce test. (La notation est celle utilisée au cours).
2. Donner l'expression théorique et simplifiée de la statistique de test.
3. Faire le test. Peut-on admettre au niveau de 95% l'hypothèse que $2\beta_1 = \beta_2$?

Solution

1. $\mathrm{R} = (2, -1)$ et $\mathrm{r} = 0$.
2. $$F_c = \frac{(\mathrm{R}\widehat{\boldsymbol{\beta}} - \mathrm{r})^\top \left\{\mathrm{R}\left(\mathrm{X}^\top\mathrm{X}\right)^{-1}\mathrm{R}^\top\right\}^{-1}(\mathrm{R}\widehat{\boldsymbol{\beta}} - \mathrm{r})\frac{1}{q}}{\mathrm{e}^\top\mathrm{e}\frac{1}{n-p}} = \frac{(2\widehat{\beta}_1 - \widehat{\beta}_2)^2\left\{4a_{11} + a_{22} - 2a_{12} - 2a_{21}\right\}^{-1}\frac{1}{1}}{\widehat{\sigma}^2_\varepsilon}$$
$$= \frac{(2 \times 2.332 - 3.311)^2\left\{[4 \times 3.987998 + 10.740450 - 4 \times (-6.152797)]/100\,000\right\}^{-1}}{1389.725} = 2.567548336,$$
où a_{ij} est l'élément général de la matrice $(\mathrm{X}^\top\mathrm{X})^{-1}$.
3. Le quantile d'ordre 0.95 d'une variable de Fisher à 1 et 28 degrés de liberté est 4.20. Donc, on ne rejette pas cette hypothèse.

Exercice 16.8. Test de Wald
Donnez le test général de Wald pour tester un ensemble de contraintes linéaires sur les coefficients de régression. Examiner les cas particuliers :

1. $H_0 : \beta_j = 0$ (uniquement pour un coefficient),
2. $H_0 : \beta_j = 0$, pour tout $j = 2, \ldots, p$ (c'est-à-dire pour tous les coefficients sauf la constante),
3. $H_0 : \beta_i = \beta_j$ pour deux coefficients i et j donnés,
4. $H_0 : \mathrm{R}\boldsymbol{\beta} = \mathrm{r}$ (test sur une combinaison linéaire des coefficients).

Donner la forme de R pour chaque test.

Solution
On veut tester une hypothèse sur les coefficients de régression de type $H_0 : \mathrm{R}\boldsymbol{\beta} = \mathrm{r}$ contre $H_1 : \mathrm{R}\boldsymbol{\beta} \neq \mathrm{r}$. On peut construire une statistique de test

$$F_{obs} = \frac{\left(\mathrm{R}\widehat{\boldsymbol{\beta}} - \mathrm{r}\right)^\top \left(\mathrm{R}\left(\mathrm{X}^\top\mathrm{X}\right)^{-1}\mathrm{R}^\top\right)^{-1}\left(\mathrm{R}\widehat{\boldsymbol{\beta}} - \mathrm{r}\right)\frac{1}{q}}{\mathrm{e}^\top\mathrm{e}\frac{1}{n-p}}.$$

Sous H_0, $F_{obs} \sim F_{q,n-p}$. Ainsi, en prenant $\alpha = 0.05$ comme erreur de première espèce, on rejette l'hypothèse H_0 si $F_{obs} > F_{1-\alpha,q,n-p}$. Cependant, F_{obs} est assez lourd à calculer. Nous allons maintenant examiner des tests $\mathrm{R}\boldsymbol{\beta} = \mathrm{r}$ particuliers et nous allons dans chaque cas essayer de simplifier au maximum l'expression F_{obs}.

1. Premier cas : test sur un coefficient individuel $H_0 : \beta_i = \beta_{i0}$ contre $H_1 : \beta_i \neq \beta_{i0}$.
   On obtient $\mathrm{R}\boldsymbol{\beta} = \mathrm{r}$ où $\mathrm{R} = \begin{pmatrix}0 & \cdots & 0 & 1 & 0 & \cdots & 0\end{pmatrix}$ et $\mathrm{r} = \beta_{i0}$. Donc, on a

$$F_{obs} = \frac{(\widehat{\beta}_i - \beta_{i0})^\top \left[(\mathrm{X}^\top\mathrm{X})^{-1}\right]_{ii}^{-1}(\widehat{\beta}_i - \beta_{i0})}{\widehat{\sigma}^2_\varepsilon} = \frac{(\widehat{\beta}_i - \beta_{i0})^2}{\widehat{\sigma}^2_\varepsilon\left[(\mathrm{X}^\top\mathrm{X})^{-1}\right]_{ii}} = \frac{(\widehat{\beta}_i - \beta_{i0})^2}{\widehat{\sigma}^2(\widehat{\beta}_i)} = \left(\frac{\widehat{\beta}_i - \beta_{i0}}{\widehat{\sigma}(\widehat{\beta}_i)}\right)^2 = t^2_{obs}.$$

   On rejette donc H_0 si $F_{obs} > F_{1-\alpha,1,n-p,}$ ou si $\mid t_{obs} \mid > t_{1-\alpha/2,n-p}$.
   Si $H_0 : \beta_i = 0$ alors

$$F_{obs} = \left(\frac{\widehat{\beta}_i}{\widehat{\sigma}(\widehat{\beta}_i)}\right)^2.$$

2. Deuxième cas : test de la nullité de tous les coefficients sauf la constante, autrement dit $H_0 : \beta_j = 0$ pour tout $j = 2, \ldots, p$. Or

$$F_{obs} = \frac{(\mathrm{SC}_{TOT} - \mathrm{SC}_{RES})/(p-1)}{\mathrm{SC}_{RES}/(n-p)} = \frac{\left(1 - \frac{\mathrm{SC}_{RES}}{\mathrm{SC}_{TOT}}\right)/(p-1)}{\frac{\mathrm{SC}_{RES}}{\mathrm{SC}_{TOT}}/(n-p)} = \frac{R^2/(p-1)}{(1-R^2)/(n-p)}.$$

   Plus le R^2 est grand, plus proportion de variance expliquée par la régression est élevée. Les valeurs élevées de R^2 conduisent à des valeurs élevées de F_{obs}, ce sont donc bien les valeurs élevées de R^2 qui nous conduisent à rejeter l'hypothèse H_0 selon laquelle y ne dépend pas de variables explicatives $x_2, \ldots, x_p$.
3. On veut tester l'égalité de 2 coefficients i.e. $H_0 : \beta_i = \beta_j$ pour deux coefficients i et j donnés contre $H_1 : \beta_i \neq \beta_j$.
   Donc, on a :

$$\mathrm{R} = \begin{pmatrix}0 & \cdots & 0 & \underbrace{1}_{i^e} & 0 & \cdots & 0 & \underbrace{-1}_{j^e} & 0 & \cdots\end{pmatrix}, \quad \mathrm{r} = 0,$$

$$\begin{aligned}\mathrm{R}\left(\mathrm{X}^\top\mathrm{X}\right)^{-1}\mathrm{R}^\top &= \left[\left(\mathrm{X}^\top\mathrm{X}\right)^{-1}\right]_{ii} - \left[\left(\mathrm{X}^\top\mathrm{X}\right)^{-1}\right]_{ji} - \left[\left(\mathrm{X}^\top\mathrm{X}\right)^{-1}\right]_{ij} + \left[(\mathrm{X}^\top\mathrm{X})^{-1}\right]_{jj} \\ &= \left[\left(\mathrm{X}^\top\mathrm{X}\right)^{-1}\right]_{ii} + \left[\left(\mathrm{X}^\top\mathrm{X}\right)^{-1}\right]_{jj} - 2\left[(\mathrm{X}^\top\mathrm{X})^{-1}\right]_{ij},\end{aligned}$$

$$\mathrm{R}\widehat{\boldsymbol{\beta}} - \mathrm{r} = \widehat{\beta}_i - \widehat{\beta}_j,$$

et

$$\begin{aligned} F_{obs} &= \frac{\left(\widehat{\beta}_i - \widehat{\beta}_j\right)^\top \left[\left[(\mathrm{X}^\top\mathrm{X})^{-1}\right]_{ii} + \left[(\mathrm{X}^\top\mathrm{X})^{-1}\right]_{jj} - 2\left[(\mathrm{X}^\top\mathrm{X})^{-1}\right]_{ij}\right]^{-1}\left(\widehat{\beta}_i - \widehat{\beta}_j\right)\frac{1}{1}}{\widehat{\sigma}^2_\varepsilon} \\ &= \frac{\left(\widehat{\beta}_i - \widehat{\beta}_j\right)^2}{\widehat{\sigma}^2_\varepsilon\left[\left[(\mathrm{X}^\top\mathrm{X})^{-1}\right]_{ii} + (\mathrm{X}^\top\mathrm{X})^{-1}_{jj} - 2\left[(\mathrm{X}^\top\mathrm{X})^{-1}\right]_{ij}\right]} = \frac{\left(\widehat{\beta}_i - \widehat{\beta}_j\right)^2}{\widehat{\sigma}^2_\varepsilon(\mathrm{X}^\top\mathrm{X})^{-1}_{ii} + \widehat{\sigma}^2_\varepsilon\left[(\mathrm{X}^\top\mathrm{X})^{-1}\right]_{jj} - 2\widehat{\sigma}^2_\varepsilon\left[(\mathrm{X}^\top\mathrm{X})^{-1}\right]_{ij}} \\ &= \frac{\left(\widehat{\beta}_i - \widehat{\beta}_j\right)^2}{\widehat{\mathrm{var}}\left(\widehat{\beta}_i\right) + \widehat{\mathrm{var}}(\widehat{\beta}_j) - 2\widehat{\mathrm{cov}}\left(\widehat{\beta}_i, \widehat{\beta}_j\right)} = \frac{\left(\widehat{\beta}_i - \widehat{\beta}_j\right)^2}{\widehat{\mathrm{var}}\left(\widehat{\beta}_i - \widehat{\beta}_j\right)} = \left(\frac{\left(\widehat{\beta}_i - \widehat{\beta}_j\right) - 0}{\widehat{\sigma}\left(\widehat{\beta}_i - \widehat{\beta}_j\right)}\right)^2 \sim F_{1,n-p} = t^2_{n-p}. \end{aligned}$$

Donc,

$$\left|\frac{\widehat{\beta}_i - \widehat{\beta}_j}{\widehat{\sigma}\left(\widehat{\beta}_i - \widehat{\beta}_j\right)}\right| \sim t_{n-p}.$$

4. Test sur une seule combinaison linéaire des coefficients.
H_0 : $\mathrm{R}\boldsymbol{\beta} = \mathrm{r}$ contre H_1 : $\mathrm{R}\boldsymbol{\beta} \neq \mathrm{r}$, où r est un scalaire.
Par exemple, si on prend $\mathrm{R} = (1\ 1\ 0\cdots 0)$ et $\mathrm{r} = 1$, on a $\beta_1 + \beta_2 = 1$.

$$F_{obs} = \frac{\left(\mathrm{R}\widehat{\boldsymbol{\beta}} - \mathrm{r}\right)^\top \left[\mathrm{R}\left(\mathrm{X}^\top\mathrm{X}\right)^{-1}\mathrm{R}^\top\right]^{-1}\left(\mathrm{R}\widehat{\boldsymbol{\beta}} - \mathrm{r}\right)/q}{\mathrm{e}^\top\mathrm{e}/(n-p)}$$

Comme $q = 1$, car le vecteur r est une constante, donc sa dimension est 1, on a

$$F_{obs} = \frac{\left(\mathrm{R}\widehat{\boldsymbol{\beta}} - \mathrm{r}\right)^2}{\widehat{\sigma}^2_\varepsilon\left(\mathrm{R}\left(\mathrm{X}^\top\mathrm{X}\right)^{-1}\mathrm{R}^\top\right)}.$$

On rejette H_0 si $F_{obs} > F_{1,n-p,1-\alpha}$.
Si on veut faire un test uniquement sur le $i$ème coefficient de régression, on prend

$$\mathrm{R} = (0\cdots 0 \underbrace{1}_{i\text{ème composante}} 0\cdots 0)^\top$$

et $\mathrm{r} = \beta_{i0}$. On obtient alors

$$F_{obs} = \frac{\left(\widehat{\beta}_i - \beta_{i0}\right)^2}{(\mathrm{X}^\top\mathrm{X})^{-1}_{ii}\widehat{\sigma}^2_\varepsilon}.$$

Ce test peut aussi être fait au moyen d'un test de Student. Il suffit de prendre la racine carrée des deux parties de cette égalité :

$$t_{obs} = \frac{\widehat{\beta}_i - \beta_{i0}}{\sqrt{(\mathrm{X}^\top\mathrm{X})^{-1}_{ii}\widehat{\sigma}^2_\varepsilon}}.$$

On rejette H_0 si $t_{obs} > t_{1-\frac{\alpha}{2},n-p}$. Le test est le même car $t^2_{1-\frac{\alpha}{2},n-p} = F_{1,n-p,1-\alpha}$.

Exercice 16.9. Test sur les régresseurs
On considère le modèle suivant :

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t,$$

où $t = 1, \ldots, 11$, et $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$. On a déjà réalisé les calculs suivants :

$$\bar{x} = 10.2, \sum_{t=1}^{11}(x_t - \bar{x})^2 = 2641.82, \widehat{\beta}_0 = -1.54, \widehat{\beta}_1 = 2.5.$$

et le tableau partiel d'analyse de la variance suivant :

Table 16.2 – Tableau d'analyse de la variance : ANOVA

Source de variation	Somme des Carrés	Degrés de libertés	Carrés Moyens	Statistique F
Régression	?	?	?	?
Résidu	65.32	?	?	
Total	2578	?	?	

1. Écrire le modèle sous forme matricielle $\mathrm{y} = \mathrm{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ (en précisant le contenu et les dimensions de chaque vecteur et matrice).
2. Compléter le tableau de l'analyse de la variance.
3. Tester (avec $\alpha = 0.05$) $H_0 : \beta_1 = 0$ contre $H_1 : \beta_1 \neq 0$.
4. Donner un intervalle de confiance à 95% pour β_1.
5. Donner une prévision de y sachant qu'on connaît $x = 2.15$.

Solution

1. La forme matricielle du modèle est :

$$\begin{pmatrix} y_1 \\ \vdots \\ y_{11} \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_{11} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_{11} \end{pmatrix}$$

$$\underset{(11,1)}{\mathrm{y}} = \underset{(11,2)}{\mathrm{X}} \underset{(2,1)}{\boldsymbol{\beta}} + \underset{(11,1)}{\boldsymbol{\varepsilon}}$$

2. Tableau d'analyse de la variance : ANOVA

Source de variation	Somme des Carrés	Degrés de libertés	Carrés Moyens	Statistique F
Régression	2512.68	1	2512.68	346.2422
Résidu	65.32	9	7.257	
Total	2578	10	257.8	

3. $F_{1,9,0.95} = 5.117 < 346.2422 = F_{obs}$, ne pas rejeter H_0.
4. Intervalle de confiance :

$$\mathrm{IC}(\beta_1) = \left[\widehat{\beta}_1 \pm t_{n-p,1-\frac{\alpha}{2}} \times \sqrt{\frac{\widehat{\sigma}_\varepsilon^2}{\sum_t (x_t - \bar{x})^2}} \right].$$

   Avec $t_{9,0.975} = 2.262$, $\widehat{\sigma}_\varepsilon^2 = 7.257$, l'intervalle de confiance est $[2.381445, 2.61855]$.
5. $y^*(2.15) = -1.54 + 2.15 \times 2.5 = 3.835$.

Exercice 16.10. Test sur des contraintes linéaires
On considère le modèle linéaire suivant :

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \varepsilon_t,$$

où $t = 1, \ldots, 10$,

$$\mathrm{E}(\varepsilon_t) = 0, \text{ pour tout } t,$$

et

$$\mathrm{E}(\varepsilon_t \varepsilon_{t'}) = \begin{cases} 0 & \text{si } t \neq t' \\ \sigma^2 & \text{si } t = t'. \end{cases}$$

1. Écrire le modèle sous forme matricielle : $\mathrm{y} = \mathrm{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ (en précisant le contenu et les dimensions de chaque vecteur et matrice) puis calculer analytiquement $\mathrm{X}^\top\mathrm{X}$ et $\mathrm{X}^\top\mathrm{y}$.
2. Donner une expression simplifée pour l'estimation de $\boldsymbol{\beta}$ sachant que

$$\sum_{t=1}^{10} x_{t1} = \sum_{t=1}^{10} x_{t2} = \sum_{t=1}^{10} x_{t1}x_{t2} = 0.$$

3. Donner une expression analytique simplifée de l'estimateur sans biais de $\widehat{\sigma}^2$ de la variance des erreurs σ^2.
4. À l'aide des informations suivantes :

$$\sum_{t=1}^{10} y_t = 25, \sum_{t=1}^{10} x_{t1}^2 = 60, \sum_{t=1}^{10} x_{t2}^2 = 80, \sum_{t=1}^{10} y_t^2 = 1940, \sum_{t=1}^{10} x_{t1}y_t = 30 \text{ et } \sum_{t=1}^{10} x_{t2}y_t = 90,$$

   calculer numériquement $\widehat{\boldsymbol{\beta}}_{MCO}$, $\widehat{\sigma}^2$ puis la matrice variance-covariance estimée de $\widehat{\boldsymbol{\beta}}_{MCO}$.
5. Calculer et interpréter le coefficient de détermination.
6. Sous l'hypothèse de normalité des erreurs, tester avec $\alpha = 5\%$ l'hypothèse selon laquelle $\beta_1 + \beta_2 = 1.5$.
7. Donner une prévision optimale de y_{11} sachant qu'on connaît $x_{11,1} = 3$ et $x_{11,2} = 4$.

Solution

1. La forme matricielle du modèle est :

$$\begin{pmatrix} y_1 \\ \vdots \\ y_{10} \end{pmatrix} = \begin{pmatrix} 1 & x_{1,1} & x_{1,2} \\ \vdots & \vdots & \vdots \\ 1 & x_{10,1} & x_{10,2} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_{10} \end{pmatrix}$$

$$\underset{(10,1)}{\mathrm{y}} = \underset{(10,3)}{\mathrm{X}} \underset{(3,1)}{\boldsymbol{\beta}} + \underset{(10,1)}{\boldsymbol{\varepsilon}},$$

   où

$$\mathrm{X}^\top\mathrm{X} = \begin{pmatrix} 1 & \cdots & 1 \\ x_{1,1} & \cdots & x_{10,1} \\ x_{1,2} & \cdots & x_{10,2} \end{pmatrix} \begin{pmatrix} 1 & x_{11} & x_{12} \\ \vdots & \vdots & \vdots \\ 1 & x_{10,1} & x_{10,2} \end{pmatrix} = \begin{pmatrix} T & \sum_{t=1}^T x_{t1} & \sum_{t=1}^T x_{t2} \\ \sum_{t=1}^T x_{t1} & \sum_{t=1}^T x_{t1}^2 & \sum_{t=1}^T x_{t1}x_{t2} \\ \sum_{t=1}^T x_{t2} & \sum_{t=1}^T x_{t2}x_{t1} & \sum_{t=1}^T x_{t2}^2 \end{pmatrix},$$

   et $T = 10$. Donc,

$$\mathrm{X}^\top\mathrm{y} = \begin{pmatrix} 1 & \cdots & 1 \\ x_{1,1} & \cdots & x_{10,1} \\ x_{1,2} & \cdots & x_{10,2} \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_{10} \end{pmatrix} = \begin{pmatrix} \sum_{t=1}^T y_t \\ \sum_{t=1}^T y_t x_{t1} \\ \sum_{t=1}^T y_t x_{t2} \end{pmatrix}.$$

2. Puisque les erreurs sont homoscedastiques et non corrélées, la meilleure estimation de $\boldsymbol{\beta}$ sera celle donnée par les moindres carrés ordinaires : $\widehat{\boldsymbol{\beta}}_{MCO} = (\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top\mathrm{y}$.
   x_1 est centré car $\bar{\mathrm{X}}_1 = 0$, $\sum_{t=1}^T x_{t1} = 0$,
   x_2 est centré car $\bar{\mathrm{X}}_2 = 0$, $\sum_{t=1}^T x_{t2} = 0$,
   x_1 est orthogonal à x_2 car $\sum_{t=1}^T x_{t1}x_{t2} = 0$. Donc,

$$\mathrm{X}^\top\mathrm{X} = \begin{pmatrix} T & 0 & 0 \\ 0 & \sum_{t=1}^T x_{t1}^2 & 0 \\ 0 & 0 & \sum_{t=1}^T x_{t2}^2 \end{pmatrix}.$$

et

$$(\mathrm{X}^\top\mathrm{X})^{-1} = \begin{pmatrix} \frac{1}{T} & 0 & 0 \\ 0 & \frac{1}{\sum_{t=1}^T x_{t1}^2} & 0 \\ 0 & 0 & \frac{1}{\sum_{t=1}^T x_{t2}^2} \end{pmatrix}.$$

On obtient :

$$\widehat{\boldsymbol{\beta}}_{MCO} = \begin{pmatrix} \frac{1}{T} & 0 & 0 \\ 0 & \frac{1}{\sum_{t=1}^T x_{t1}^2} & 0 \\ 0 & 0 & \frac{1}{\sum_{t=1}^T x_{t2}^2} \end{pmatrix} \begin{pmatrix} \sum_{t=1}^T y_t \\ \sum_{t=1}^T y_t x_{t1} \\ \sum_{t=1}^T y_t x_{t2} \end{pmatrix} = \begin{pmatrix} \sum_{t=1}^T \frac{y_t}{T} \\ \sum_{t=1}^T \frac{y_t x_{t1}}{\sum_{t=1}^T x_{t1}^2} \\ \sum_{t=1}^T \frac{y_t x_{t2}}{\sum_{t=1}^T x_{t2}^2} \end{pmatrix} = \begin{pmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{pmatrix}.$$

3. On estime la variance par :

$$\widehat{\sigma}_\varepsilon^2 = \frac{\mathrm{e}^\top\mathrm{e}}{T-3} = \frac{(\mathrm{y}-\mathrm{X}\boldsymbol{\beta})^\top(\mathrm{y}-\mathrm{X}\boldsymbol{\beta})}{T-3}.$$

4. Les estimations sont :

$$\widehat{\boldsymbol{\beta}} = \begin{pmatrix} \sum_{t=1}^T \frac{y_t}{T} \\ \sum_{t=1}^T \frac{y_t x_{t1}}{\sum_{t=1}^T x_{t1}^2} \\ \sum_{t=1}^T \frac{y_t x_{t2}}{\sum_{t=1}^T x_{t2}^2} \end{pmatrix} = \begin{pmatrix} 2.5 \\ 0.5 \\ 1.125 \end{pmatrix} = \begin{pmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{pmatrix}$$

$$\widehat{\sigma}_\varepsilon^2 = 245.1786$$

$$\begin{aligned} \widehat{\sigma}(\widehat{\boldsymbol{\beta}}) &= \widehat{\sigma}_\varepsilon^2(\mathrm{X}^\top\mathrm{X})^{-1} \\ &= 141.857 \begin{pmatrix} \frac{1}{10} & 0 & 0 \\ 0 & \frac{1}{60} & 0 \\ 0 & 0 & \frac{1}{80} \end{pmatrix} = \begin{pmatrix} 24.5179 & 0 & 0 \\ 0 & 4.0863 & 0 \\ 0 & 0 & 3.0647 \end{pmatrix} \end{aligned}$$

5.

$$R^2 = 1 - \frac{e^\top e}{\sum_{t=1}^T y_t^2 - T\bar{y}^2} = 1 - \frac{1716.25}{1940 - 62.5} = 0.08588.$$

$R^2 \to 0$: très mauvaise qualité d'ajustement linéaire

6. Le test :

$$\begin{cases} H_0 : & \beta_1 + \beta_2 = 1.5 \\ H_1 : & \beta_1 + \beta_2 \neq 1.5, \end{cases}$$

On pose $Z = \beta_1 + \beta_2$. Donc,

$$\begin{cases} H_0 : & Z = 1.5 \\ H_1 : & Z \neq 1.5, \end{cases}$$

La variance vaut :

$$\widehat{\mathrm{var}}(Z) = \widehat{\mathrm{var}}(\widehat{\beta}_1 + \widehat{\beta}_2) = \quad \widehat{\mathrm{var}}(\widehat{\beta}_1) + \widehat{\mathrm{var}}(\widehat{\beta}_2) + 2\widehat{\mathrm{cov}}(\widehat{\beta}_1, \widehat{\beta}_2) = \widehat{\mathrm{var}}(\widehat{\beta}_1) + \widehat{\mathrm{var}}(\widehat{\beta}_2).$$

$$T = \frac{\widehat{Z} - 1.5}{\sqrt{\widehat{\mathrm{var}}(\widehat{Z})}} = \frac{\widehat{\beta}_1 + \widehat{\beta}_2 - 1.5}{\sqrt{\widehat{\mathrm{var}}(\widehat{\beta}_1) + \widehat{\mathrm{var}}(\widehat{\beta}_2)}} = 0.0467.$$

Comme $|Z| < t_{7,0.975} = 2.36$, on ne rejette pas H_0.

7. La prédiction :

$$\hat{y}_t = 2.5 + 0.5x_{t1} + 1.125x_{t2},$$

$$\hat{y}_{11} = 2.5 + 0.5x_{11,1} + 1.125x_{11,2} = 8.5.$$

Chapitre 17

Exercices : Outils de détection en régression

Exercice 17.1. Détection et villes
Le Tableau 17.1 donne les médailles d'or pour le saut en longeur pendant les jeux olympiques de 1900 à 2004. Les performances sont mesurées en centimètres.

Table 17.1 – Médailles d'or aux jeux olympiques pour le saut en longueur

Ville	saut en longueur(Y)	année (X)
Paris	719	0
St Louis	734	4
London	748	8
Stockholm	760	12
Antwerpen	715	20
Paris	745	24
Amsterdam	774	28
Los Angeles	764	32
Berlin	806	36
London	782	48
Helsinki	757	52
Melbourne	783	56
Rome	812	60
Tokyo	807	64
Mexico City	890	68
München	824	72
Montreal	834	76
Moscow	854	80
Los Angeles	854	84
Seoul	872	88
Barcelona	867	92
Atlanta	850	96
Sydney	855	100
Athen	859	104
somme	19184	1304
moyenne	799.333	54.333

On voudrait donner une estimation de la performance pour les prochains jeux olympiques, pour cela on ajuste un modèle linéaire sur les données :

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

On connaît également les expressions suivantes :

$$\sum_{i=1}^{24} y_i^2 = 15390858, \quad \sum_{i=1}^{24} x_i^2 = 95424, \text{ et } \sum_{i=1}^{24} y_i x_i = 1077188.$$

1. Donner la matrice $\mathrm{X}^\top\mathrm{X}$ et le vecteur $\mathrm{X}^\top\mathrm{y}$.
2. Donner les valeurs des coefficients $\widehat{\beta}_1$ et $\widehat{\beta}_2$.
3. Donner un estimateur sans biais de la variance des erreurs du modéle ($\widehat{\sigma}^2_\varepsilon$).
4. Tester, séparément, à 95% si les deux coefficients sont significativement différent de 0.
5. Donner les intervalles de confiance à 95% pour les deux coefficients.
6. Donner l'intervalle de confiance pour la prévision de la performance en saut en longeur pour les jeux olympiques de 2012 ($x_{26} = 112$).
7. Construire le tableau d'analyse de la variance et donner le R^2.
8. Vérifier vos résultats en utilisant le logiciel R.
9. À partir du Tableau 17.2, donner les points leviers et déterminer l'observation qui a la plus grande influence sur le modèle (utiliser la distance de Cook).

Table 17.2 – Médailles d'or aux jeux olympiques pour le saut en longueur

Ville	saut en longueur (Y)	année (X)	h_{ii}	st.res	D_{Cook}
Paris	719	0	0.162	0.200	0.004
St Louis	734	4	0.145	-0.368	0.011
London	748	8	0.129	-0.864	0.055
Stockholm	760	12	0.115	-1.234	0.099
Antwerpen	715	20	0.090	2.093	0.216
Paris	745	24	0.079	0.660	0.019
Amsterdam	774	28	0.070	-0.699	0.018
Los Angeles	764	32	0.062	0.211	0.001
Berlin	806	36	0.055	-1.884	0.104
London	782	48	0.043	0.478	0.005
Helsinki	757	52	0.042	2.235	0.109
Melbourne	783	56	0.042	1.071	0.025
Rome	812	60	0.043	-0.265	0.002
Tokyo	807	64	0.045	0.347	0.003
Mexico City	890	68	0.049	0.559	0.008
München	824	72	0.054	0.023	0.000
Montreal	834	76	0.061	-0.227	0.002
Moscow	854	80	0.068	-1.060	0.041
Los Angeles	854	84	0.077	-0.734	0.023
Seoul	872	88	0.088	-1.462	0.103
Barcelona	867	92	0.099	-0.841	0.039
Atlanta	850	96	0.112	0.502	0.016
Sydney	855	100	0.127	0.546	0.022
Athen	859	104	0.142	0.653	0.035
somme	19184	1304	2.000	-0.061	0.960
moyenne	799.333	54.333	0.083	-0.003	0.040

Solution

1. La matrice $\mathrm{X}^\top\mathrm{X}$ et le vecteur $\mathrm{X}^\top\mathrm{y}$ sont :

$$\mathrm{X}^\top\mathrm{X} = \begin{pmatrix} 24 & 1304 \\ 1304 & 95424 \end{pmatrix} \text{ et } \mathrm{X}^\top\mathrm{y} = \begin{pmatrix} 19184 \\ 1077188 \end{pmatrix}.$$

2. Comme

$$ns_x^2 = 24574.203,\ ns_{xy} = 34864.163 \text{ et } ns_y^2 = 56460.123,$$

on a

$$\widehat{\beta}_1 = 722.234 \text{ et } \widehat{\beta}_2 = 1.419.$$

3. L'estimateur de variances :

$$\widehat{\sigma}_\varepsilon^2 = \frac{\text{SC}_{RES}}{n-p} = \frac{ns_y^2 - \widehat{\beta}_2^2 ns_x^2}{n-p} = \frac{6978.466}{22} = 317.203$$

4. Pour les tests, les statistiques à calculer sont :

$$t_{\beta_2} = \frac{\widehat{\beta}_2 - \beta_{2,0}}{se(\widehat{\beta}_2)}$$

et

$$t_{\beta_1} = \frac{\widehat{\beta}_1 - \beta_{1,0}}{se(\widehat{\beta}_1)},$$

où

$$se(\widehat{\beta}_2) = \frac{\widehat{\sigma}_\varepsilon^2}{\sqrt{ns_x^2}} = \frac{17.810}{\sqrt{24574.203}} = 0.114$$

et

$$se(\widehat{\beta}_1) = \widehat{\sigma}_\varepsilon \sqrt{1/n + \bar{x}^2/ns_x^2} = 17.810\sqrt{1/24 + 54.333^2/24574.203} = 7.164.$$

Donc,

$$t_{\beta_2} = \frac{1.419 - 0}{0.114} = 12.447$$

et

$$t_{\beta_1} = \frac{722.234 - 0}{7.164} = 100.814.$$

Dans chaque cas on doit comparer ces statistiques avec la valeur t hypothètique, $t_{0.975,n-2} = t_{0.975,22} = 2.070$. Les valeurs absolues de ces statistiques sont plus grandes que les quantiles. Donc, on rejette les hypothèses et on conclut que β_1 et β_2 sont significativement différents de 0.

5. Construire des intervalles de confiance pour la pente et pour la constante.

$$\text{IC}_{\beta_2} = (\widehat{\beta}_2 - t_{1-\beta_1/2,n-2} se(\widehat{\beta}_2), \widehat{\beta}_2 + t_{1-\alpha/2,n-2} se(\widehat{\beta}_2))$$

et respectivement

$$\text{IC}_{\beta_1} = (\widehat{\beta}_1 - t_{1-\beta_1/2,n-2} se(\widehat{\beta}_1), \widehat{\beta}_1 + t_{1-\alpha/2,n-2} se(\widehat{\beta}_1))$$

dans notre cas :

$$\text{IC}_{\beta_2} = (1.183; 1.655)$$

et

$$\text{IC}_{\beta_1} = (707.405; 737.063).$$

6. L'intervalle de prédiction pour la valeur actuelle de y si on connaît $x = 112$:

$$\text{IP}_{\widehat{y}^*} = \left[\widehat{y}^* - t_{1-\alpha/2,n-2}\widehat{\sigma}_\varepsilon \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{ns_x^2}}; \widehat{y}^* + t_{1-\alpha/2,n-2}\widehat{\sigma}_\varepsilon \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{ns_x^2}}\right],$$

où $\widehat{y}^* = \widehat{\beta}_1 + \widehat{\beta}_2 x^* = 881.162$. Donc, $\text{IP}_{\widehat{y}^*} = (824.455; 926.517)$.

7. Tableau d'analyse de la variance :

Source	d.l.	SC	CM	F_{obs}
Régression	1	49481.657	49481.657	155.994
Résiduelle	21	6978.466	317.203	
Totale	23	56460.123		

$\widehat{\sigma}_\varepsilon^2 = 317.203$, $\widehat{\sigma}_\varepsilon = 17.810$, $R^2 = 0.876$ et $F_{obs} = 155.994$.

8. **En langage R**

```
LJ=c(719,734,748,760,715,745,774,764,806,782,757,783,812,807,809,824,834,854,854,
   872,867,850,855,859)
Années=c(0,4,8,12,20,24,28,32,36,48,52,56,60,64,68,72,76,80,84,88,92,96,100,104)
#
y=LJ
x=Années
fit=lm(y~x)
summary(fit)
anova(fit)
#
xnew=data.frame(x=112)
predict(fit,xnew,interval="confidence")
predict(fit,xnew,interval="prediction")
#
influence(fit)
hii=influence(fit)$hat
leverages=hii[hii>2*2/24]
leverages
Sigma2=sum((fit$residuals)^2)/(24-2)
st.res=fit$residuals/(sqrt(Sigma2*(1-hii)))
st.res
cooksd=st.res^2*hii/(2*(1-hii))
cooksd
outliers=st.res[abs(st.res)>3]
outliers
influ=cooksd[cooksd>2*3/(24-2)]
influ
plot(x,y)
abline(fit)
plot(fit)
plot(st.res~fit$fitted)
```

9. On a $p = 2$ et $n = 24$. Donc, $2p/n = 4/24 = 1/6 = 0.1666667$. Auncun des h_{ii} ne dépasse 0.1666666. Il n'y a donc pas de point levier. Aucune distance de Cook n'est plus grande que 1. Il n'y a donc pas d'observations très influentes. L'observation la plus influente est Anvers avec une distance de Cook de 0.216.

Exercice 17.2. Leviers et influence
Dans le tableau ci-dessous on voit les résidus et les termes diagonaux de la matrice H pour quelques observations d'un modèle à 3 variables (constante + 2 autres variables).

	résidus	h_{ii}
1	0.366	0.067
2	−0.872	0.042
3	6.965	0.043
4	0.840	0.079
5	5.169	0.033
6	−1.474	0.031
7	−11.967	0.149
8	−14.039	0.075
9	46.825	0.288
10	−8.151	0.064
⋮	⋮	⋮
42	2.106	0.046
43	−3.199	0.060
44	−2.664	0.060
45	4.684	0.072
46	−0.269	0.032
47	−10.408	0.073
48	−2.749	0.077
49	−2.530	0.032
50	−0.080	0.031
51	−0.936	0.031

$$\widehat{\sigma}^2_\varepsilon = 75.256$$

1. Donner les points leviers.
2. Calculer les résidus standardisés et les distances de Cook, puis donner les outliers (points "aberrants") et les points qui ont une grande influence sur la droite de la régression (selon leur distance de Cook).
3. Utiliser le logiciel R pour ajuster un modèle linéaire de la variable "mr" par une constante, "south" et "pcy". Vérifier vos résultats des points précédents.

En langage R

```
\begin{verbatim}
assa=data.frame(mr=c(11.60,9,8.60,10.20,13.10,5.80,6.30,5,78.50,8.90,11.40,
     3.80,2.90,11.40,7.50,2.30,6.40,6.60,20.30,1.60,12.70,3.90,9.80,3.40,
     13.50,11.30,3,11.30,1.70,3.90,10.40,2,5.30,8,13.30,6,8.40,4.60,
     6.80,3.90,10.30,3.40,10.20,11.90,3.10,3.60,8.30,6.90,5.20,4.40,3.40),
south=c(1,0,0,1,0,0,0,1,1,1,1,0,0,0,0,0,0,1,1,0,1,0,0,0,1,0,0,1,0,0,0,0,0,0,0,0,0,
     1,0,0,0,1,0,1,1,0,0,1,1,0,0,0),
pcy=c(13.53,18.22,14.29,12.63,17.30,16.98,22.24,17.26,23.30,16.31,15.21,18.57,
    13.83,17.82,15.18,14.44,15.68,13.34,13.12,14.83,18.89,19.28,16.26,16.57,11.65,
    15.45,13.73,14.75,13.49,15.54,18.08,17.66,21.23,12.91,19.61,15.56,13.45,15.35,
    16.80,16.78,13.32,14.12,14.57,15.12,12.75,15.35,17.10,12.77,17.20,15.65,15.58))
```

Solution

1. Comme
$$\frac{2p}{n} = \frac{2 \times 3}{51} = 0.118,$$
les observations 7 et 9 ont sont des points levier car leur valeurs h_{ii} sont supérieures à 0.118. On peut calculer
$$\text{st.res}_i = \frac{\text{res}_i}{\sqrt{\widehat{\sigma}^2_\varepsilon(1 - h_{ii})}}$$

et

$$D_{i.cook} = \frac{e_i^2}{p\widehat{\sigma}_\varepsilon^2}\frac{h_{ii}}{(1-h_{ii})^2} = \frac{(\text{st.res})^2 \times h_{ii}}{2(1-h_{ii})},$$

afin d'obtenir le tableau :

	résidus	h_{ii}	st.res	D_{cook}
1	0.366	0.067	0.044	0.000
2	−0.872	0.042	−0.102	0.000
3	6.965	0.043	0.819	0.015
4	0.840	0.079	0.101	0.000
5	5.169	0.033	0.604	0.006
6	−1.474	0.031	−0.172	0.000
7	−11.967	0.149	−1.492	0.195
8	−14.039	0.075	−1.678	0.114
9	46.825	0.288	6.381	8.234
10	−8.151	0.064	−0.969	0.032
⋮	⋮	⋮	⋮	⋮
42	2.106	0.046	0.248	0.001
43	−3.199	0.060	−0.379	0.005
44	−2.664	0.060	−0.316	0.003
45	4.684	0.072	0.559	0.012
46	−0.269	0.032	−0.031	0.000
47	−10.408	0.073	−1.243	0.061
48	−2.749	0.077	−0.329	0.005
49	−2.530	0.032	−0.296	0.001
50	−0.080	0.031	−0.009	0.000
51	−0.936	0.031	−0.109	0.000

L'observation 9 est outlier car son résidu standardisé est beaucoup plus grand que 3. Enfin, l'observation 9 est un point influent car sa distance de Cook est plus grande que 1.

2.
```
attach(assa)
fit=lm(mr~south+pcy)
summary(fit)
anova(fit)
influence(fit)
hii=influence(fit)$hat
leverages=hii[hii>2*3/51]
leverages
Sigma2=sum((fit$residuals)^2)/(51-3)
st.res=fit$residuals/(sqrt(Sigma2*(1-hii)))
st.res
cooksd=st.res^2*hii/(2*(1-hii))
cooksd
outliers=st.res[abs(st.res)>3]
outliers
influ=cooksd[cooksd>1]
influ
par(mfrow=c(3,2))
plot(fit)
plot(st.res~fit$fitted)
```

Chapitre 18

Exercices : Analyse de la variance à un facteur

Exercice 18.1. Traitements et poids
Un groupe de recherche médicale voudrait comparer 5 traitements de régimes différents. Pour cela, ils demandent à 25 personnes en surpoids (5-5 personnes dans chaques groupes) de suivre ces traitements. (On suppose que la distribution de la perte des poids suit une loi normale et qu'il n'y a pas de différences entre les variances). Les résultats sont dans le Tableau 18.1 :

TABLE 18.1 – Les résultats (perte de poids en kg) des 5 traitements

A	13	16	16	15	15
B	7	4	7	8	9
C	12	8	6	9	10
D	6	10	5	7	7
E	9	11	13	11	11

Tester au seuil de signification $\alpha = 0.05$ s'il y a différence entre les traitements du point de vue de l'efficacité.

Solution
On calcule les moyennes et les sommes de carrés chaque groupe.

Les moyennes et les variances des 5 traitements.

Traitements	moyennes	sommes de carrés
A	15.0	6
B	7.0	14
C	9.0	20
D	7.0	14
E	11.0	8
Total	9.8	286

Donc,

$$\begin{aligned} SC_{intra} &= 6 + 14 + 20 + 4 + 8 = 62 \\ SC_{TOT} &= 286 \\ SC_{inter} &= SC_{TOT} - SC_{intra} = 286 - 62 = 224 \end{aligned}$$

Tableau de l'analyse de la variance

Source	d.l.	SC	MC	F_{obs}
Inter groupe	4	224	56	18.06
Intra groupe	20	62	3.1	
Totale	24	286		

Comme $F_{0.95,4,20} = 2.87$, on rejette l'hypothèse qu'il n'y a pas de différences entre les moyennes des groupes. Donc, les traitements differents ont des efficacités differentes.

En langage R

```
X=rbind(A=c(13,16,16,15,15),
B=c(7,4,7,8,9),
C=c(12,8,6,9,10),
D=c(6,10,5,7,7),
E=c(9,11,13,11,11))
xb=rowMeans(X)
SCG=rowSums((X-xb)^2)
xbt=mean(X)
SCT=sum((X-xbt)^2)
SGI=sum(SCG)
XX=c(X)
F=as.factor(rep(1:5,5))
anova(lm(XX~F))
qf(0.95,4,20)
```

Exercice 18.2. Âge et mémorisation
Une partie d'une étude d'Eysenck (1974) compare les sujets plus jeunes et plus âgés quand à leur aptitude à se rappeler le matériel alors qu'on les avait prévenus qu'ils devaient mémoriser les données de manière à s'en souvenir ultérieurement (cette tâche exigeait vraisemblablement un niveau élevé de traitement.) Les données figurent dans le Tableau 18.2. (la variable dépendante étant le nombre d'éléments rappelés).

TABLE 18.2 – Nombre d'éléments rappelés selon l'âge

Sujets plus jeunes :	21	19	17	15	22	16	22	22	18	21
Sujets plus âgés :	10	19	14	5	10	11	14	15	11	11

Effectuer une analyse de variance afin de comparer les moyennes de ces deux groupes.

Solution
On calcule les moyennes et les sommes de carrés :

	Moyenne	Somme de Carrés
Jeunes	19.30	64.10
Âgés	12.00	126.00
Total	15.65	456.55

Tableau de l'analyse de la variance

Source	d.l.	SC	MC	F_{obs}
Inter groupe	1	266.45	266.450	25.229
Intra groupe	18	190.10	10.561	
Totale	19	456.55		

Le quantile $F_{0.95,1,18} = 4.413873 < 25.229 = F_{obs}$. On rejette donc l'hypothèse de l'égalité des moyennes.

En langage R

```
J=c(21,19,17,15,22,16,22,22,18,21)
A=c(10,19,14,5,10,11,14,15,11,11)
X=rbind(J,A)
xb=rowMeans(X)
SCG=rowSums((X-xb)^2)
xbt=mean(X)
SCT=sum((X-xbt)^2)
SGI=sum(SCG)
XX=c(X)
F=factor(rep(1:2,10))
anova(lm(XX~F))
qf(0.95, 1,18)
```

Exercice 18.3. Âge et traitements
Une partie d'une étude d'Eysenck (1974) consiste à comparer quatre groupes de sujets. L'un des groupes se composait de jeunes sujets à qui l'on présentait les mots dans une condition qui suscitait un niveau peu élevé de traitement. Un deuxième groupe se composait des sujets jeunes à qui l'on donnait des tâches requérant un niveau de traitement plus élevé. Les deux autres groupes comprenaient des sujets plus âgés à qui l'on donnait des tâches requérant un niveau de traitement soit peu élevé, soit élevé. Les données sont dans le Tableau 18.3 :

Table 18.3 – Nombre d'éléments rappelés selon l'âge et le niveau

Jeunes/Peu élevé	8	6	4	6	7	6	5	7	9	7
Jeunes/Elevé	21	19	17	15	22	16	22	22	18	21
Agés/Peu élevé	9	8	6	8	10	4	6	5	7	7
Agés/Elevé	10	19	14	5	10	11	14	15	11	11

1. Effectuer une analyse de variance à un critère de classification sur ces données.
2. Effectuer à présent une analyse de variance à un critère de classification en opposant les traitements 1 et 3 combinés ($H = 2$) aux traitements 2 et 4 combinés. À quelle question répondez-vous ?

Solution

1. Test pour quatre groupes :
   Calcul des moyennes et des sommes de carrés pour quatre groupes.

	Moyennes	SC
Jeunes/Peu élevé	6.50	18.50
Jeunes/Élevé	19.30	64.10
Agés/Peu élevé	7.00	30.00
Agés/Elevé	12.00	126.00
Total	11.20	1298.40

   Le tableau d'analyse de la variance :

Tableau de l'analyse de la variance (4 groupes)

Source	d.l.	SC	MC	F_{obs}
Inter groupe	3	1059.8	353.27	53.30
Intra groupe	36	238.6	6.63	
Totale	39	1298.4		

   On rejette donc l'hypothèse de l'égalité des moyennes car $F_{0.95,3,36} = 2.866266$.
2. Test pour deux groupes :
   Calcul des moyennes et des sommes de carrés pour deux groupes :

	Moyennes	SC
Peu élevé	6.75	49.75
Élevé	15.65	456.55
Total 3	11.20	1298.40

   Le tableau d'analyse de la variance :

Tableau de l'analyse de la variance (2 groupes)

Source	d.l.	SC	MC	F_{obs}
Inter groupe	1	792.1	792.1	59.45
Intra groupe	38	506.3	13.32	
Totale	39	1298.4		

   On rejette donc l'hypothèse de l'égalité des moyennes en répondant à la question de savoir si les moyennes sont différentes selon le niveau de traitement des tâches qui sont confiées car $F_{0.95,1,38} = 4.098172$.

En langage R

```
AP=c(9,8,6,8,10,4,6,5,7,7),
AE=c(10,19,14,5,10,11,14,15,11,11))
# 4 groupes
xb=rowMeans(X)
SCG=rowSums((X-xb)^2)
xbt=mean(X)
SCT=sum((X-xbt)^2)
SGI=sum(SCG)
XX=c(X)
F=factor(rep(1:nrow(X),ncol(X)))
anova(lm(XX~F))
qf(0.95, nrow(X)-1,length(X)-nrow(X))
xtable(data.frame(Moyennes=c(xb,xbt),SC=c(SCG,SCT)))
#2 groupes
X1=rbind( c(X[1,],X[3,]), c(X[2,],X[4,]) )
xb=rowMeans(X1)
SCG=rowSums((X1-xb)^2)
xbt=mean(X1)
SCT=sum((X1-xbt)^2)
SGI=sum(SCG)
XX=c(X1)
F=factor(rep(1:nrow(X1),ncol(X1)))
anova(lm(XX~F))
qf(0.95, nrow(X1)-1,length(X1)-nrow(X1))
xtable(data.frame(Moyennes=c(xb,xbt),SC=c(SCG,SCT)))
```

Exercice 18.4. Vente de pains
Le propriétaire d'une petite boulangerie suppose que la vente du pain les lundis et les samedis sont differents de celle des autres jours de la semaine. Afin de pouvoir mieux gérer la commande, il note le nombre des pains de 1 kg vendus lors de 6 lundis, 6 samedis et 10 autres jours choisis au hasard et indépendamment les uns de des autre. Les résultats sont dans le Tableau 18.4 :

Table 18.4 – Les nombres des pains vendus

le jours	nbr. des jours	nbr. des pains vendus	la moyenne	somme de carrés
lundi	6	30, 40, 54, 34, 44, 50	42.00	424
autre jour	10	49, 43, 30, 59, 35, 46, 42, 35, 36, 43	41.80	633.6
samedi	6	52, 58, 57, 70, 54, 53	57.33	219.36
Totale			46.09	2319.82

On suppose que le nombre des pains vendus pendant les jours différents de la semaine suivent une loi normale, avec la même variance. Tester l'hypothèse que les espérances de la quantité des pains vendus pendants les jours différents de la semaine sont égales.

Solution
Donc,

$$\begin{aligned} SC_{intra} &= 424 + 633.6 + 219.33 = 1276.93 \\ SC_{TOT} &= 2319.82 \\ SC_{inter} &= SC_{TOT} - SC_{intra} = 2319.82 - 1276.93 = 1042.88 \end{aligned}$$

Le tableau d'analyse de la variance :

Tableau de l'analyse de la variance

Source	d.l.	SC	MC	F_{obs}
Inter groupe	2	1042.88	521.44	7.76
Intra groupe	19	1276.93	67.21	
Totale	21	2319.82		

Comme $F_{0.95,2,19} = 3.52$, on rejette l'hypothèse qu'il n'y a pas de différences entre les moyennes des groupes. Donc, les ventes de pains sont différentes pendant les différents jours de la semaine.

Exercice 18.5. Comportement maternel de rats
Pour étudier le comportement maternel de rats de laboratoire, nous éloignons le bébé rat de sa mère d'une distance fixée et enregistrons le temps nécessaire à la mère (en secondes) pour ramener son bébé au nid. Nous réalisons cette expérience avec des bébés rats de 5, 20 et 35 jours. Les données figurent ci-dessous pour six bébés par groupe.

Table 18.5 – Temps selon les groupes

5 jours	15	10	25	15	20	18
20 jours	30	15	20	25	23	20
35 jours	40	35	50	43	45	40

1. Peut-on dire au niveau de probabilité 0.05 que le temps nécessaire pour ramener le bébé dépend de l'âge ? Justifier votre réponse.
2. Donner le quantile d'ordre 0.95 de la variable de Fisher correspondant à l'analyse de la variance ?

Solution

1. On donne le tableau d'analyse de la variance suivant :

Tableau d'analyse de la variance

	Sommes des carrés	degrés de liberté	Carrés moyens	F	Sig.
Inter-groupes	2100.0	2	1050.000	40.127	0.000
Intra-groupe	392.5	15	26.167		
Total	2492.5	17			

   Oui, le temps nécessaire pour ramener le bébé dépend de l'âge car nous pouvons voir que la valeur de F est "grande" et suffisante à 0.05% pour affirmer que les temps diffèrent significativement d'un groupe à l'autre.
2. La valeur du quantile d'ordre 0.95 d'une Fisher à 2 et 15 degrés de liberté vaut 3.682.

Deuxième partie

Matériel pour l'évaluation

Chapitre 19

Questions à choix multiples

Question 1.
On jette deux dés équilibrés. Quelle est la probabilité qu'au moins l'un d'entre eux montre 6, sachant que les résultats des deux dés sont différents ?

A) $\frac{1}{2}$ B) $\frac{1}{3}$ C) $\frac{1}{5}$ D) $\frac{1}{4}$ E) Aucune de ces réponses

Solution

$$\frac{\frac{1}{6}\frac{5}{6}+\frac{5}{6}\frac{1}{6}}{\frac{30}{36}}=\frac{\frac{10}{36}}{\frac{5}{6}}=\frac{1}{3}.$$

Réponse : B

Question 2.
Les variables aléatoires X et Y ont la densité conjointe suivante :

$$f(x,y)=\begin{cases}2xy+\frac{3}{2}y^2 & \text{si } 0\leq x\leq 1 \text{ et } 0\leq y\leq 1\\ 0 & \text{sinon.}\end{cases}$$

Quelle est la densité marginale de $f_X(x)$ quand $0\leq x\leq 1$?
A) $f_X(x)=x+\frac{1}{2}$ B) $f_X(x)=y+\frac{3}{2}y^2$ C) $f_X(x)=2x^2+\frac{3}{2}x$ D) Aucune des réponses préalables.

Solution

$$\int_0^1\left(2xy+\frac{3}{2}y^2\right)dy=2x\int_0^1 ydy+\frac{3}{2}\int_0^1 y^2dy=2x\left[\frac{y^2}{2}\right]_0^1+\frac{3}{2}\left[\frac{y^3}{3}\right]_0^1=x\left[y^2\right]_0^1+\frac{1}{2}\left[y^3\right]_0^1=x+\frac{1}{2}$$

Réponse : A

Question 3.
S'il existe une très forte dépendance linéaire entre deux variables, il s'ensuit que la corrélation est :
A) Proche de zéro, B) Proche de 1 ou de -1, C) Proche de -1, D) Proche de 1. E) Aucune de ces réponses.

Solution
Réponse : B

Question 4.
Le changement d'unité de mesure de la variable indépendante x dans une régression linéaire simple $y_i=\alpha+\beta x_i+\varepsilon_i$ affecte dans la régression :
A) L'estimation de la pente du modèle. B) L'estimation de la constante du modèle.
C) L'estimation de la somme des carrés des résidus. D) Aucune des réponses préalables.

Solution
Réponse : A

Question 5.
Quelle affirmation est vraie ?
A) La moyenne des valeurs ajustées est toujours égale à la moyenne des valeurs observées.
B) La moyenne des valeurs ajustées est égale à la moyenne des valeurs observées, si la régression est sans constante.
C) La moyenne des valeurs ajustées n'est jamais égale à la moyenne des valeurs observées.
D) La moyenne des valeurs ajustées est égale à la moyenne des valeurs observées, si la somme des résidus est nulle.
E) Aucune des affirmations précédentes n'est correcte.

Solution
Réponse : D

Question 6.
À partir de données sur les variables x et y, on a calculé les coefficients de la droite de régression de y en x. On a alors trouvé la droite de régression suivante : $y = -4 + 5x$. Laquelle des valeurs suivantes pourrait être plausible comme coefficient de corrélation de x et y ?
A) $r_{xy} = -5$ B) $r_{xy} = -4$ C) $r_{xy} = -0.8$
D) $r_{xy} = 0.8$ E) $r_{xy} = 4$ F) $r_{xy} = 5$

Solution
Réponse : D

Question 7.
Nous avons effectué une régression avec une variable explicative x et une constante, nous recevons une nouvelle observation x_{n+1} et nous calculons la prédiction correspondante $\widehat{y}_{n+1}$. La variance de la valeur prédite est minimale si x_{n+1} est tel que :
A) $x_{n+1} = 0$ B) $x_{n+1} = \bar{x}$ C) $x_{n+1} = \min_i x_i$ D) $x_{n+1} = \max_i x_i$ E) On ne peut pas savoir.

Solution
Réponse : B

Question 8.
Dans une régression linéaire, le vecteur y est-il orthogonal au vecteur des résidus e ?
A) Oui B) Non C) Seulement si la constante fait partie du modèle de régression.

Solution
Réponse : B

Question 9.
Nous avons effectué une régression multiple, une des variables explicatives est la constante. La somme des résidus calculés vaut :
A) 0 B) Approximativement 0 C) Parfois 0 C) 1 D) -1 ou 1

Solution
Réponse : A

Question 10.
Le vecteur y^* est-il orthogonal au vecteur des termes d'erreurs $\boldsymbol{\varepsilon}$?
A) Oui B) Non C) Seulement si 1 fait partie des variables explicatives.

Solution
Réponse : B

Question 11.
Supposons que $X \sim \mathcal{N}(2, 4)$. Quelle est la probabilité que $\Pr(X < 7)$?
A) 0.8944 B) 0.6103 C) 0.9938

Solution
Réponse : C

Question 12.
La droite des moindres carrés de la régression définie par $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$ passe-t-elle par le point $(\bar{x}, \bar{y})$:
A) Toujours B) Jamais C) Parfois

Solution
Réponse : A

Question 13.
La fonction de densité d'une variable aléatoire X est définie par :

$$f(x) = \begin{cases} 3\exp^{-3x} & \text{si } x > 0 \\ 0 & \text{sinon.} \end{cases}$$

Quelle est sa variance ?
A) $\frac{1}{3}$ B) $\frac{1}{9}$ C) $\frac{1}{6}$ D) Aucune des réponses préalables.

Solution
La fonction $f(x)$ la densité d'une variable exponentielle avec $\lambda = 3$. Donc, $\text{var}(X) = 1/\lambda^2 = 1/3^2 = 1/9$.
Réponse : B

Question 14.
On examine l'évolution d'une variable y_i en fonction de deux variables explicatives $x_{i,2}$ et $x_{i,3}$ avec le modèle $y_i = \beta_1 + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \varepsilon_i$. Nous avons obtenus les résultats suivants :

$$\mathrm{X}^\top \mathrm{X} = \begin{pmatrix} 10 & 0 & 0 \\ 0 & 1010 & 65.6 \\ 0 & 65.6 & 132.78 \end{pmatrix}$$

La taille n de l'echantillon vaut :
A) 14 B) 25 C) 10 D) Aucune des réponses proposées

Solution
Réponse : C

Question 15.
Pour la matrice

$$\mathrm{X}^\top \mathrm{X} = \begin{pmatrix} 10 & 0 & 0 \\ 0 & 1010 & 65.6 \\ 0 & 65.6 & 132.78 \end{pmatrix}$$

d'un modèle avec deux variables explicatives et une constante, le coefficient de corrélation linéaire empirique ente $x_{i,2}$ et $x_{i,3}$ vaut :
A) 0.5 B) 0.1797 C) 0 D) 1 E) Aucune des réponses proposées

Solution
Les variables x_2 et x_3 sont centrées. Donc,

$$r_{23} = \frac{\sum_{i=1}^{n} x_{i,2}x_{i,3}}{\sqrt{\sum_{i=1}^{n} x_{i,2}^2 \times \sum_{i=1}^{n} x_{i,3}^2}} = \frac{65.6}{\sqrt{1010 \times 132.78}} = 0.1797.$$

Réponse : B

Question 16.
Soit la droite de régression suivante : $y_i = a + bx_i + e_i$. On remarque que la pente est nulle dans ce modèle ($b = 0$). Que vaut la corrélation entre la variable x et la variable y ?
A) 1 B) 0 C) 0.5 D) -1 E) Aucune de ces réponses

Solution
Réponse : B

Question 17.
Soit une fonction de densité avec la forme suivante :

$$f(x) = \begin{cases} 6e^{-6x} & \text{si } x > 0 \\ 0 & \text{sinon.} \end{cases}$$

$\mathrm{E}(X)$ vaut
A) $\frac{1}{6}$ B) $\frac{2}{5}$ C) $\frac{1}{5}$ D) $\frac{4}{6}$ E) Aucune de ces réponses

Solution
C'est une variable exponentielle de paramètre $\lambda = 6$. Donc, $\mathrm{E}(X) = 1/\lambda = 1/6$.
Réponse : A

Question 18.
Dans l'analyse d'une régression simple avec x comme variable indépendante et y comme variable dépendante, si la valeur de la constante de la régression est négative :
A) les valeurs prédites de y sont toujours négatives,
B) la corrélation entre x et y peut être soit positive, négative ou nulle,
C) la corrélation entre x et y est négative,
D) la corrélation entre x et y est positive.

Solution
Réponse : B

Question 19.
Dans le modèle linéaire simple suivant : $y_i = \alpha + \beta x_i + \varepsilon_i$, où les ε_i sont les termes d'erreurs et iid tels que $\mathrm{E}(\varepsilon_i) = 0$ et $\mathrm{var}(\varepsilon_i) = \sigma_\varepsilon^2$, la moyenne $\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i$ de la distribution des termes d'erreurs est toujours égale à :
A) 0 B) 1 C) -1 D) β, la pente de la régression E) aucune des réponses ci-dessus.

Solution
Réponse : E

Question 20.
À partir de données sur les variables y (poids en kg) et x (taille en centimètres), on a calculé les coefficients de

la droite de régression de y en x. Si on change les unités de la variable x en mètres, alors :
A) la pente de la droite de régression est multipliée par 100
B) la pente de la droite est divisée par 100
C) la constante de la droite de régression est multipliée par 100
D) la constante de la droite de régression est divisée par 100
E) Aucune des réponses ci-dessus n'est correcte.

Solution
Réponse : A

Question 21.
Dans un modèle de régression avec une constante et une variable explicative :
A) $\beta^2 = \mathrm{E}(\widehat{\beta}^2)$ B) $\beta^2 = \mathrm{E}(\widehat{\beta}^2) - \frac{\sigma_\varepsilon^2}{ns_x^2}$ C) $\beta^2 = \mathrm{E}(\widehat{\beta}^2) + \frac{\sigma_\varepsilon^2}{ns_x^2}$ D) $\mathrm{E}(\beta^2) = \widehat{\beta}^2$ E) $\mathrm{E}(\beta^2) = \frac{\sigma_\varepsilon^2}{ns_x^2} + \widehat{\beta}^2$

Solution
Comme $\mathrm{E}(\widehat{\beta}^2) = \mathrm{var}(\widehat{\beta}) + \mathrm{E}^2(\widehat{\beta}) = \frac{\sigma_\varepsilon^2}{ns_x^2} + \beta^2$, on a $\beta^2 = \mathrm{E}(\widehat{\beta}^2) - \frac{\sigma_\varepsilon^2}{ns_x^2}$.
Réponse : B

Question 22.
Dans une régression linéaire :
A) La somme des résidus n'est jamais toujours égale à 0.
B) La somme des résidus est égale à 0, si la régression est sans constante.
C) Si la somme des résidus est égale à 0, la constante de la régression est égale à zéro.
D) Si la somme des carrés des résidus est égale à 0, les valeurs ajustées sont égales aux valeurs observées.
E) La moyenne des valeurs ajustées n'est jamais égale à la moyenne des valeurs observées.
F) Aucune des affirmations précédentes n'est correcte.

Solution
Réponse : D

Question 23.
Dans un modèle linéaire général :
A) L'estimateur du maximum de vraisemblance des paramètres : $\widehat{\beta}_{MV}$ surestime β.
B) L'estimateur du maximum de vraisemblance des paramètres : $\widehat{\beta}_{MV}$ sous-estime β.
C) L'estimateur du maximum de vraisemblance : $\widehat{\sigma}^2_{\varepsilon MV}$ surestime la valeur de la variance des erreurs (σ_ε^2).
D) L'estimateur du maximum de vraisemblance : $\widehat{\sigma}^2_{\varepsilon MV}$ sous-estime la valeur de la variance des erreurs (σ_ε^2).
E) L'estimateur du maximum de vraisemblance : $\widehat{\sigma}^2_{\varepsilon MV}$ est un estimateur sans biais de la variance des erreurs (σ_ε^2).
F) Aucune des affirmations précédentes n'est correcte.

Solution
Réponse : D

Question 24.
Quelle affirmation est vraie ?
A) Si la corrélation entre deux variable (r) est égale à 0, cela veut dire qu'il n'existe aucune relation entre elles.
B) La somme des erreurs (ε_i) est toujours égale à 0.
C) Dans le cadre de la régression simple (une constante et une variable explicative) le test de Fisher à partir du tableau d'analyse de la variance et le test de Student pour la nullité du coefficient de la variable explicative, aboutissent exactement à la même décision.
D) Le R-carré est toujours inférieur au R-carré ajusté.
E) Plus la p-valeur d'un test est grande plus le test est significatif.
F) Aucune des affirmations précédentes n'est correcte.

Solution
Réponse : D

Question 25.
On considère le modèle de régression : $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$. On suppose que les termes d'erreurs sont des variables aléatoires indépendantes normalement distribuées d'espérance nulle et de variance $\sigma^2 > 0$ ($\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$). Soient les assertions suivantes :

I) $\mathrm{E}(y_i) = \beta_0 + \beta_1 x_i$, II) $\mathrm{var}(\beta_0 + \beta_1 x_i) = \mathrm{var}(-\varepsilon_i)$, III) $\mathrm{var}(y_i) = \mathrm{var}(\varepsilon_i)$, IV) $\mathrm{var}(y_i) = \sigma^2$,

Quelle affirmation est correcte :
A) I et II seulement sont vraies,
B) I, II et III seulement sont vraies,
C) I, II et IV seulement sont vraies,
D) I, III et IV seulement sont vraies,
E) II et III seulement sont vraies,
F) II, III et IV seulement sont vraies,
G) II et IV seulement sont vraies,
H) III et IV seulement sont vraies,
I) I, II, III et IV sont vraies.

Solution
Réponse : D

Question 26.
Quelle affirmation est vraie ?
A) La somme des erreurs (ε_i) est toujours égale à 0.
B) Dans un modèle de régression avec une seule variable explicative et la constante : si la corrélation entre les deux variables est négative, le coefficient ($\boldsymbol{\beta}_1$) de la variable explicative doit être également négatif.
C) Dans un modèle linéaire général Y (la variable dépendante) et X(les variables explicatives) sont des variables aléatoires.
D) La somme des carrés des résidus est toujours égale à 0.
E) Aucune des affirmations précédentes n'est correcte.

Solution
Réponse : B

Question 27.
On examine l'évolution d'une variable y_i en fonction de deux variables explicatives $x_{i,2}$ et $x_{i,3}$ avec le modèle $y_i = \beta_1 + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \varepsilon_i$. Nous avons obtenu les résultats suivants :

$$\left(\mathrm{X}^\top \mathrm{X}\right)^{-1} = \begin{pmatrix} 1/20 & 0 & 0 \\ 0 & 1/80 & 0 \\ 0 & 0 & 1/142 \end{pmatrix}.$$

La taille n de l'échantillon vaut :
A) 45 B) 20 C) 10 D) Aucune des réponses proposées
E) On ne posséde pas assez d'information pour déterminer n.

Solution
Réponse : B

Question 28.
On considère le modèle $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$. Soient alors $\widehat{\beta}_1$ et $\widehat{\beta}_2$ les estimateurs des paramètres (estimés par la méthode des moindres carrés). On peut alors affirmer :
A) $\mathrm{E}(\beta_1) = \widehat{\beta}_1$ B) $\mathrm{E}(\widehat{\beta}_1) = \beta_1$ C) $\mathrm{E}(y_i) = \beta_1 + \beta_2$ D) $\mathrm{E}(\widehat{\beta}_1) = \mathrm{E}(\widehat{\beta}_2)$
E) Aucune des réponses ci-dessus n'est correcte.
Remarque : E(.) est l'espérance.

Solution
Réponse : B

Question 29.
Soit le modèle linéaire général avec des termes d'erreur normaux (formulation matricielle) suivant

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

Soit $\widehat{\boldsymbol{\beta}}$ l'estimateur des moindres carrés de $\boldsymbol{\beta}$. On peut alors affirmer que
A) Puisque $\widehat{\boldsymbol{\beta}}$ est un estimateur sans biais, sa variance est nulle,
B) $\widehat{\boldsymbol{\beta}}$ est un estimateur biaisé,
C) $\widehat{\boldsymbol{\beta}}$ est normalement distribué, d'espérance $\boldsymbol{\beta}$,
D) $\widehat{\boldsymbol{\beta}}$ est normalement distribué, d'espérance nulle,
E) Aucune des réponses ci-dessus n'est correcte.

Solution
Réponse : C

Question 30.
À partir de données sur les variables x et y, on a calculé les coefficients de la droite de régression de y en x. On a alors trouvé la droite de régression suivante : $y = -4 + 5x$. Laquelle des valeurs suivantes pourrait être le coefficient de corrélation de x et y ?
A) $r_{xy} = -4$ B) $r_{xy} = -0.8$ C) $r_{xy} = 0.8$ D) $r_{xy} = 5$
E) Il n'est pas possible de répondre à cette question.

Solution
Réponse : C

Question 31.
Afin d'ajuster un modèle linéaire $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$, on dispose de la matrice de données et du vecteur de la variable dépendante :

$$\mathbf{X} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & -1 \\ 1 & 1 \end{pmatrix} \text{ et } \quad \mathbf{y} = \begin{pmatrix} 1 \\ 3 \\ 4 \\ 2 \end{pmatrix}.$$

L'estimateur de $\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$ obtenu à l'aide des moindres carrés est

A) $\widehat{\boldsymbol{\beta}} = \begin{pmatrix} 2.5 \\ -1 \end{pmatrix}$ B) $\widehat{\boldsymbol{\beta}} = \begin{pmatrix} 10 \\ -1 \end{pmatrix}$ C) $\widehat{\boldsymbol{\beta}} = \begin{pmatrix} 4 \\ 0 \end{pmatrix}$ D) $\widehat{\boldsymbol{\beta}} = \begin{pmatrix} 0 \\ 0 \\ -1 \\ 1 \end{pmatrix}$
E) Aucune des réponses ci-dessus n'est correcte

Solution

$$\mathrm{X}^\top\mathrm{X} = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}, (\mathrm{X}^\top\mathrm{X})^{-1} = \begin{pmatrix} 1/4 & 0 \\ 0 & 1/2 \end{pmatrix}, \mathrm{X}^\top\mathrm{y} = \begin{pmatrix} 10 \\ -2 \end{pmatrix}, \widehat{\boldsymbol{\beta}} = (\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top\mathrm{y} = \begin{pmatrix} 2.5 \\ -1 \end{pmatrix}.$$

Réponse : A

Question 32.
La Figure 19.1 représente des données ainsi que le droite de régression de y en x ajustée. Notons respectivement $\widehat{\beta}_1$ et $\widehat{\beta}_2$ la constante et la pente de cette droite.

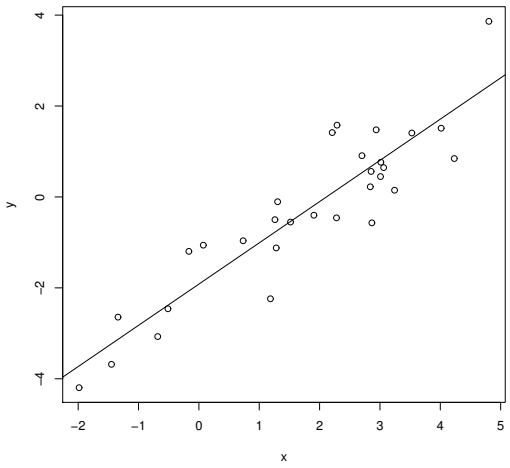


FIGURE 19.1 – Nuage de points et droite de régression

Au vu de ce graphique, on peut assurer que :
A) $\widehat{\beta}_1 < 0$ et $\widehat{\beta}_2 < 0$, B) $\widehat{\beta}_1 < 0$ et $\widehat{\beta}_2 > 0$ C) $\widehat{\beta}_1 > 0$ et $\widehat{\beta}_2 < 0$ D) $\widehat{\beta}_1 > 0$ et $\widehat{\beta}_2 > 0$
E) Il n'est pas possible de répondre à cette question.

Solution
Réponse : B

Question 33.
On souhaite calculer les coefficients de la droite de régression de y en x passant par l'origine. On utilise les résultats suivants :

$$\frac{1}{20}\sum_{i=1}^{20} x_i^2 = 2, \quad \frac{1}{20}\sum_{i=1}^{20} y_i^2 = 7, \quad \frac{1}{20}\sum_{i=1}^{20} x_i y_i = 12, \quad \bar{x} = 2, \quad \bar{y} = 3.$$

Soient $\widehat{\beta}_1$ et $\widehat{\beta}_2$ respectivement la constante et la pente de la droite de régression. On trouve alors :
A) $\widehat{\beta}_1 = 9$ et $\widehat{\beta}_2 = -3$ B) $\widehat{\beta}_1 = 0$ et $\widehat{\beta}_2 = -3$ C) $\widehat{\beta}_1 = 9$ et $\widehat{\beta}_2 = 6$ D) $\widehat{\beta}_1 = 0$ et $\widehat{\beta}_2 = 6$
E) Aucune des réponses ci-dessus n'est correcte.

Solution

$$\widehat{\beta}_2 = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} = \frac{12}{2} = 6 \text{ et } \widehat{\beta}_1 = 0.$$

Réponse : D

Question 34.
On dispose des résultats d'une régression linéaire simple (avec une variable explicative et une constante), avec 25 individus, du logiciel R suivants.

```
Analysis of Variance Table

Response : y
          Df  Sum Sq Mean Sq F value  Pr(>F)
x          1  56.394  56.394  6.4903 0.01799 *
Residuals~?      ~?        ~?       ~?       ~?
---
Signif. codes~:  0 "***" 0.001 "**" 0.01 "*" 0.05 "." 0.1 " " 1
```

Malheureusement, les résultats de la ligne Residuals ont été effacés. Quels résultats pourrait-on trouver sur cette ligne ?

	Df	Sum Sq	Mean Sq	F value	Pr(> F)
A)	1	56.394	56.394		
B)	23	199.846	8.689		
C)	25	162.2575	6.4903	3.2452	0.05
D)	24	208.536	8.689	6.4903	0.01799
E)	24	208.536	8.689		
F) Aucune des réponses ci-dessus n'est correcte.					

Solution

```
Analysis of Variance Table
Response~: y
     Df  Sum Sq Mean Sq F value  Pr(>F)
x      1  56.394  56.394  6.4903 0.01799 *
Residuals 23 199.847   8.689
---
Signif. codes~:  0 "***" 0.001 "**" 0.01 "*" 0.05 "." 0.1 " " 1
```

Réponse : B

Question 35.
Le Tableau 19.1 présente 24 médaillés d'or en saut en longueur aux Jeux Olympiques. Le modèle est le suivant :

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i.$$

À partir du Tableau 19.1, donner la (ou les) observation(s) étant un (ou des) point(s) levier(s). Noter que h_i est le levier de l'observation i.

A) Paris(1)
B) Paris(1), St Louis, Athen, London(1), Stockholm
C) Helsinki, Melbourne, Rome, London(2)
D) Helsinki, Melbourne
E) Aucune des réponses ci-dessus n'est correcte

Solution
Un observation est point levier si $h_i > 2p/n = 2 \times 2/24 = 0.16666$.
Réponse : E

Question 36.
Soient le modèle de régression $\mathrm{y} = \mathrm{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, le vecteur des valeurs ajustées y^*, le vecteur des termes d'erreur e tel que $\mathrm{y} = \mathrm{X}\widehat{\boldsymbol{\beta}} + \mathrm{e}$, le projecteur P_X dans l'image de la matrice X et le projecteur $\mathrm{P}_X^\perp$ dans le noyau de la matrice $\mathrm{X}^\top$. Quel énoncé est faux ? (Si tous les énoncés sont vrais, inscrire F).

A) Le vecteur y^* est orthogonal au vecteur e.
B) Le vecteur e est orthogonal aux colonnes de la matrice X.
C) Le vecteur e est orthogonal aux colonnes de P_X.
D) Le vecteur e est orthogonal aux colonnes de $\mathrm{P}_X^\perp$.
E) Le vecteur y^* est orthogonal aux colonnes $\mathrm{P}_X^\perp$.
F) Tous les énoncés qui précèdent sont vrais.

Solution
Réponse : D

Question 37.
Soient le modèle de régression $\mathrm{y} = \mathrm{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ et le projecteur orthogonal P_X dans l'image de la matrice X. Parmi les estimateurs suivants, lequel n'exprime pas l'estimateur des moindres carrés ordinaires du vecteur $\boldsymbol{\beta}$?

A) $\widehat{\boldsymbol{\beta}} = (\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top\mathrm{P}_X\mathrm{y}$
B) $\widehat{\boldsymbol{\beta}} = (\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top\mathrm{y}$
C) $\widehat{\boldsymbol{\beta}} = (\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top(\mathrm{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})$
D) $\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top\boldsymbol{\varepsilon}$
E) $\widehat{\boldsymbol{\beta}} = \mathrm{X}^\top\mathrm{P}_X\mathrm{y}$
F) Aucune des réponses ci-dessus n'est correcte.

Table 19.1 – Médailles d'or aux jeux olympiques pour le saut en longueur

Ville	saut en longueur (Y)	année (X)	h_i
Paris(1)	719	0	0.162
St Louis	734	4	0.145
London(1)	748	8	0.129
Stockholm	760	12	0.115
Antwerpen	715	20	0.090
Paris(2)	745	24	0.079
Amsterdam	774	28	0.070
Los Angeles(1)	764	32	0.062
Berlin	806	36	0.055
London(2)	782	48	0.043
Helsinki	757	52	0.042
Melbourne	783	56	0.042
Rome	812	60	0.043
Tokyo	807	64	0.045
Mexico City	890	68	0.049
München	824	72	0.054
Montreal	834	76	0.061
Moscow	854	80	0.068
Los Angeles(2)	854	84	0.077
Seoul	872	88	0.088
Barcelona	867	92	0.099
Atlanta	850	96	0.112
Sydney	855	100	0.127
Athen	859	104	0.142
somme	19184	1304	2.000
moyenne	799.333	54.333	0.083

Solution
Réponse : E

Question 38.
Dans un problème de régression simple $y = \alpha + \beta X + \varepsilon$, on note respectivement $\widehat{\alpha}$, $\widehat{\beta}$ et r_{xy} les estimateurs de la droite de la régression et la corrélation entre X et Y. Pour chaque combinaison ci-dessous, dire quelle combinaison est impossible.

A) $\widehat{\alpha} = 1.06$ et $\widehat{\beta} = 0.67$. B) $\widehat{\alpha} = 0.67$ et $\widehat{\beta} = 1.06$. C) $\widehat{\beta} = 0.67$ et $r_{xy} = 0.9$.
D) $\widehat{\beta} = 0.41$ et $r_{xy} = -0.5$. E) $\widehat{\beta} = 0.76$ et $r_{xy} = 0.76$.

Solution
$\widehat{\beta}$ et r_{xy} doivent avoir le même signe.
Réponse : D

Chapitre 20

Questions ouvertes

Exercice 1.
En considérant que $\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top\boldsymbol{\varepsilon}$:

1. Donner une expression simplifiée de $\widehat{\boldsymbol{\beta}}^\top\mathrm{X}^\top\mathrm{X}\widehat{\boldsymbol{\beta}}$ en fonction de X, $\boldsymbol{\beta}$ et $\boldsymbol{\varepsilon}$.
2. Calculer l'espérance de $\widehat{\boldsymbol{\beta}}^\top\mathrm{X}^\top\mathrm{X}\widehat{\boldsymbol{\beta}}$.
3. Est-ce que $\widehat{\boldsymbol{\beta}}^\top\mathrm{X}^\top\mathrm{X}\widehat{\boldsymbol{\beta}}$ est un estimateur sans biais de $\boldsymbol{\beta}^\top\mathrm{X}^\top\mathrm{X}\boldsymbol{\beta}$?

Solution

1. Version simplifiée :

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}^\top\mathrm{X}^\top\mathrm{X}\widehat{\boldsymbol{\beta}} &= (\boldsymbol{\beta} + (\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top\boldsymbol{\varepsilon})^\top\mathrm{X}^\top\mathrm{X}(\boldsymbol{\beta} + (\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top\boldsymbol{\varepsilon}) \\
&= \boldsymbol{\beta}^\top\mathrm{X}^\top\mathrm{X}\boldsymbol{\beta} + 2\boldsymbol{\beta}^\top\mathrm{X}^\top\mathrm{X}(\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}^\top\mathrm{X}(\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top\mathrm{X}(\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top\boldsymbol{\varepsilon} \\
&= \boldsymbol{\beta}^\top\mathrm{X}^\top\mathrm{X}\boldsymbol{\beta} + 2\boldsymbol{\beta}^\top\mathrm{X}^\top\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}^\top\mathrm{X}(\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top\boldsymbol{\varepsilon} \\
&= \boldsymbol{\beta}^\top\mathrm{X}^\top\mathrm{X}\boldsymbol{\beta} + 2\boldsymbol{\beta}^\top\mathrm{X}^\top\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}^\top\mathrm{P}_X\boldsymbol{\varepsilon}.
\end{aligned}$$

2. Espérance :

$$\mathrm{E}(\widehat{\boldsymbol{\beta}}^\top\mathrm{X}^\top\mathrm{X}\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}^\top\mathrm{X}^\top\mathrm{X}\boldsymbol{\beta} + \sigma_\varepsilon^2 \times \mathrm{trace}(\mathrm{P}_X) = \boldsymbol{\beta}^\top\mathrm{X}^\top\mathrm{X}\boldsymbol{\beta} + \sigma_\varepsilon^2 \times p.$$

3. Non, car $\mathrm{E}(\widehat{\boldsymbol{\beta}}^\top\mathrm{X}^\top\mathrm{X}\widehat{\boldsymbol{\beta}}) \neq \boldsymbol{\beta}^\top\mathrm{X}^\top\mathrm{X}\boldsymbol{\beta}$.

Exercice 2.
Soit le modèle linéaire : $\mathrm{y} = \mathrm{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, où $\mathrm{var}(\boldsymbol{\varepsilon}) = \mathrm{I}\sigma_\varepsilon^2$. On définit $\widehat{\boldsymbol{\beta}}_2 = (\mathrm{Z}^\top\mathrm{X})^{-1}\mathrm{Z}^\top\mathrm{y}$ un estimateur de $\boldsymbol{\beta}$, où Z est une matrice $n \times p$ de plein rang.

1. Est-ce que $\widehat{\boldsymbol{\beta}}_2$ est sans biais ?
2. Cet estimateur a-t-il une plus petite variance que l'estimateur $\widehat{\boldsymbol{\beta}} = (\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top\mathrm{y}$?

Justifier vos réponses en deux lignes.

Solution

1. Comme $\mathrm{y} = \mathrm{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, on peut écrire :

$$\widehat{\boldsymbol{\beta}}_2 = (\mathrm{Z}^\top\mathrm{X})^{-1}\mathrm{Z}^\top\mathrm{y} = (\mathrm{Z}^\top\mathrm{X})^{-1}\mathrm{Z}^\top(\mathrm{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}).$$

   Pour voir si cet estimateur est sans biais ou non, on doit calculer son espérance :

$$\begin{aligned} \mathrm{E}(\widehat{\boldsymbol{\beta}}_2) &= \mathrm{E}((\mathrm{Z}^\top\mathrm{X})^{-1}\mathrm{Z}^\top\mathrm{y}) = \mathrm{E}((\mathrm{Z}^\top\mathrm{X})^{-1}\mathrm{Z}^\top(\mathrm{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})) = \mathrm{E}((\mathrm{Z}^\top\mathrm{X})^{-1}\mathrm{Z}^\top\mathrm{X})\boldsymbol{\beta} + (\mathrm{Z}^\top\mathrm{X})^{-1}\mathrm{Z}^\top\boldsymbol{\varepsilon}) \\ &= \mathrm{E}(\boldsymbol{\beta}) + \mathrm{E}(\mathrm{Z}^\top\mathrm{X})^{-1}\mathrm{Z}^\top\boldsymbol{\varepsilon} = \boldsymbol{\beta} + (\mathrm{Z}^\top\mathrm{X})^{-1}\mathrm{Z}^\top\mathrm{E}(\boldsymbol{\varepsilon}) = \boldsymbol{\beta} + (\mathrm{Z}^\top\mathrm{X})^{-1}\mathrm{Z}^\top 0 = \boldsymbol{\beta}. \end{aligned}$$

   Donc, l'estimateur est sans biais.
2. Le Theorème de Gauss-Markov montre que $\widehat{\boldsymbol{\beta}}$ est le meilleur estimateur linéaire sans biais. Cependant on peut aussi calculer la variance de $\widehat{\boldsymbol{\beta}}_2$:

$$\mathrm{var}(\widehat{\boldsymbol{\beta}}_2) = \sigma_\varepsilon^2(\mathrm{Z}^\top\mathrm{X})^{-1}\mathrm{Z}^\top\mathrm{Z}(\mathrm{X}^\top\mathrm{Z})^{-1} = \sigma_\varepsilon^2\left\{(\mathrm{X}^\top\mathrm{Z})(\mathrm{Z}^\top\mathrm{Z})^{-1}(\mathrm{Z}^\top\mathrm{X})\right\}^{-1} = \sigma_\varepsilon^2\left\{\mathrm{X}^\top\mathrm{P}_Z\mathrm{X})\right\}^{-1},$$

   où

$$\mathrm{P}_Z = \mathrm{Z}(\mathrm{Z}^\top\mathrm{Z})^{-1}\mathrm{Z}^\top.$$

   Pour montrer que $\widehat{\boldsymbol{\beta}}$ est plus efficace que $\widehat{\boldsymbol{\beta}}_2$, on doit montrer que la matrice $\mathrm{var}(\widehat{\boldsymbol{\beta}}_2) - \mathrm{var}(\widehat{\boldsymbol{\beta}})$ est semi definie positive. Il est équivalent de montrer que $\mathrm{var}(\widehat{\boldsymbol{\beta}})^{-1} - \mathrm{var}(\widehat{\boldsymbol{\beta}}_2)^{-1}$ est semi definie positive. Or

$$\mathrm{var}(\widehat{\boldsymbol{\beta}})^{-1} - \mathrm{var}(\widehat{\boldsymbol{\beta}}_2)^{-1} = \frac{1}{\sigma_\varepsilon^2}\left\{\mathrm{X}^\top\mathrm{X} - \mathrm{X}^\top\mathrm{P}_Z\mathrm{X})\right\} = \frac{1}{\sigma_\varepsilon^2}\left\{\mathrm{X}^\top(\mathrm{I} - \mathrm{P}_Z)\mathrm{X})\right\},$$

   qui est semi définie positive, car $(\mathrm{I} - \mathrm{P}_Z)$ est une matrice de projection.

Exercice 3.
On considère le modèle linéaire suivant : $y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \varepsilon_i$, où $i = 1, \ldots, 10$. On voudrait estimer les coefficients de ce modèle (β_0, β_1 et β_2) pour cela, on avait déjà calculé les expressions suivantes :

$$\sum_{i=1}^{10} y_i = 25, \quad \sum_{i=1}^{10} y_i^2 = 245, \quad \sum_{i=1}^{10} x_{1,i} = \sum_{i=1}^{10} x_{2,i} = \sum_{i=1}^{10} x_{1,i}x_{2,i} = 0,$$

$$\sum_{i=1}^{10} x_{1,i}^2 = 60 \quad \sum_{i=1}^{10} x_{2,i}^2 = 80, \quad \sum_{i=1}^{10} x_{1,i}y_i = 30 \quad \text{et} \quad \sum_{i=1}^{10} x_{2,i}y_i = 90.$$

1. Écrire le modèle sous forme matricielle $\mathrm{y} = \mathrm{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, en précisant le contenu et les dimensions de chaque vecteur et matrice.
2. Calculer analytiquement $\mathrm{X}^\top\mathrm{X}$, $(\mathrm{X}^\top\mathrm{X})^{-1}$ et $\mathrm{X}^\top\mathrm{y}$.
3. Donner les estimations des coefficients, les erreurs standard et les t-valeurs (pour le test générale H_0 : $\beta_j = 0$). Donc, remplir le tableau suivant :

j	$\widehat{\beta_j}$	$se(\widehat{\beta_j})$	t-valeurs
0	...	0.973	...
1	...	...	...
2	...	...	...

4. Construire le tableau d'ANOVA.
5. Donner la valeur de S^2, la valeur du R^2 puis interpreter le R^2.

Solution

1. La forme matricielle du modèle est :

$$\begin{pmatrix} y_1 \\ \vdots \\ y_{10} \end{pmatrix} = \begin{pmatrix} 1 & x_{1,1} & x_{2,1} \\ \vdots & \vdots & \vdots \\ 1 & x_{1,10} & x_{2,10} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_{10} \end{pmatrix},$$

$$\underset{(10,1)}{\mathrm{y}} = \underset{(10,3)}{\mathrm{X}} \underset{(3,1)}{\boldsymbol{\beta}} + \underset{(10,1)}{\boldsymbol{\varepsilon}},$$

$$\mathrm{X}^\top\mathrm{X} = \begin{pmatrix} 1 & \cdots & 1 \\ x_{1,1} & \cdots & x_{1,10} \\ x_{2,1} & \cdots & x_{2,10} \end{pmatrix} \begin{pmatrix} 1 & x_{1,1} & x_{2,1} \\ \vdots & \vdots & \vdots \\ 1 & x_{1,10} & x_{2,10} \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^n x_{1,i} & \sum_{i=1}^n x_{2,i} \\ \sum_{i=1}^n x_{1,i} & \sum_{i=1}^n x_{1,i}^2 & \sum_{i=1}^n x_{1,i}x_{2,i} \\ \sum_{i=1}^n x_{2,i} & \sum_{i=1}^n x_{1,i}x_{2,i} & \sum_{i=1}^n x_{2,i}^2 \end{pmatrix},$$

$$\mathrm{X}^\top\mathrm{y} = \begin{pmatrix} 1 & \cdots & 1 \\ x_{1,1} & \cdots & x_{1,10} \\ x_{2,1} & \cdots & x_{2,10} \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_{10} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{1,i}y_i \\ \sum_{i=1}^n x_{2,i}y_i \end{pmatrix}.$$

2. On calcule les matrices :

$$\mathrm{X}^\top\mathrm{X} = \begin{pmatrix} 10 & 0 & 0 \\ 0 & 60 & 0 \\ 0 & 0 & 80 \end{pmatrix}, \qquad (\mathrm{X}^\top\mathrm{X})^{-1} = \begin{pmatrix} \frac{1}{10} & 0 & 0 \\ 0 & \frac{1}{60} & 0 \\ 0 & 0 & \frac{1}{80} \end{pmatrix}.$$

3. L'estimateur de $\boldsymbol{\beta}$:

$$\widehat{\boldsymbol{\beta}} = \begin{pmatrix} \frac{1}{10} & 0 & 0 \\ 0 & \frac{1}{60} & 0 \\ 0 & 0 & \frac{1}{80} \end{pmatrix} \begin{pmatrix} 25 \\ 30 \\ 90 \end{pmatrix} = \begin{pmatrix} 2.500 \\ 0.500 \\ 1.125 \end{pmatrix} = \begin{pmatrix} \widehat{\beta_0} \\ \widehat{\beta_1} \\ \widehat{\beta_2} \end{pmatrix}.$$

$$\mathrm{SC}_{TOT} = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - n\bar{y}^2 = 245 - 10 \times (25/10)^2 = 182.500.$$

$$\text{se}(\widehat{\beta}_0) = \widehat{\sigma}_\varepsilon \sqrt{(\mathrm{X}^\top \mathrm{X})^{-1}_{11}} = \sqrt{\widehat{\sigma}^2_\varepsilon \frac{1}{10}} = 0.973.$$

Donc, $\widehat{\sigma}_\varepsilon = 3.077$. On a

$$\text{se}(\widehat{\beta}_1) = \widehat{\sigma}_\varepsilon \sqrt{(\mathrm{X}^\top \mathrm{X})^{-1}_{22}} = \sqrt{3.077^2 \times \frac{1}{60}} = 0.397,$$

$$\text{se}(\widehat{\beta}_2) = \widehat{\sigma}_\varepsilon \sqrt{(\mathrm{X}^\top \mathrm{X})^{-1}_{33}} = \sqrt{3.077^2 \times \frac{1}{80}} = 0.344.$$

$\widehat{\beta}_j$	$se(\widehat{\beta}_j)$	$t-$ valeurs
2.500	0.973	2.570
0.500	0.397	1.259
1.125	0.344	3.271

Le quantile est $t_{0.95,7} = 2.36$ donc $\widehat{\beta}_0$ et $\widehat{\beta}_2$ sont significativement differents de 0. Cependant, on ne rejette pas l'hypothèse, que $\widehat{\beta}_1$ est egal à 0.

4. Le tableau d'analyse de la variance :

Source	d.l.	SC	CM	F
Régression	2	116.224	58.112	6.138
Résidus	7	66.276	9.468	
Total	9	182.500		

Le quantile est $F_{0.95,2,7} = 4.74$. Donc, le modèle est significatif.

$$\widehat{\sigma}^2_\varepsilon = 9.468 \quad \widehat{\sigma}_\varepsilon = 3.077 \text{ et } \; R^2 = 63.684\%.$$

Le modèle explique 63.684% de la variabilité de la variable dépendante.

Exercice 4.
On dispose des loyers moyens par m^2 selon le nombre de pièces de 13 appartements. Ces données se trouve dans le Tableau 20.1.

TABLE 20.1 – Loyers moyens par m^2 selon le nombre de pièces

Nombre de pièces (X)	1	1	2	2	3	3	3	4	4	4	5	5	6
Loyers par m^2 (Y)	19	18.5	15.7	15.2	13.7	13.6	13.2	13.1	13.1	13	12.9	12.2	11.7

On voudrait donner une estimation du prix par m^2 pour un appartement de 7 pièces. Pour cela, on ajuste un modèle de régression linéaire sur nos données : $Y = \beta_0 + \beta_1 X + \varepsilon$. Les expressions suivantes sont connues :

$$\sum_{i=1}^{13} y_i^2 = 2692.030 \quad \sum_{i=1}^{13} x_i^2 = 171 \text{ et } \quad \sum_{i=1}^{13} y_i x_i = 573.300$$

1. Donner le matrice $\mathrm{X}^\top\mathrm{X}$ et le vecteur $\mathrm{X}^\top\mathrm{y}$.
2. Donner les valeurs estimées des coefficients $\widehat{\beta}_0$ et $\widehat{\beta}_1$.
3. Donner un estimateur sans biais de la variance des erreurs du modéle ($\widehat{\sigma}_\varepsilon^2$).
4. Tester au niveau 5% si ces deux coefficients sont significativement différents de 0.
5. Donner des intervalles de confiance à 95% pour β_0 et pour β_1.
6. Donner des intervalles de confiance et des intervalles de prediction pour un appartement de 7 pièces.
7. Construire le tableau d'ANOVA et donner les valeurs de
$$\mathrm{SC}_{RES}, \quad \widehat{\sigma}_\varepsilon^2, \quad R^2 \text{ et } F.$$
8. À partir du Tableau 20.2, donner des "leverage points", des outliers et des observations qui ont une grande influence sur le modèle, basée de leur distance de Cook.

TABLE 20.2 – Loyers moyens par m^2 selon le nombre de pièces, leviers, résidus standardisés et distances de Cook

Nombre de pièces (X)	Loyers par m^2 (Y)	h_{ii}	st.res	D_{Cook}
1	19.0	0.262	-1.966	0.686
1	18.5	0.262	-1.389	0.343
2	15.7	0.136	0.281	0.006
2	15.2	0.136	0.814	0.052
2	13.7	0.080	0.963	0.040
3	13.6	0.080	1.066	0.050
3	13.2	0.080	1.479	0.095
4	13.1	0.094	0.210	0.002
4	13.1	0.094	0.210	0.002
4	13.0	0.094	0.314	0.005
5	12.9	0.176	-0.014	0.110
5	12.2	0.176	-0.250	0.007
6	11.7	0.329	-1.282	0.403

Solution

1. On calcule :
$$\mathrm{X}^\top\mathrm{X} = \begin{pmatrix} 13 & 43 \\ 43 & 171 \end{pmatrix} \text{ et } \mathrm{X}^\top\mathrm{y} = \begin{pmatrix} 184.900 \\ 573.300 \end{pmatrix}.$$
2. Les estimateurs des paramètres sont : $\widehat{\beta}_0 = 18.626$ et $\widehat{\beta}_1 = -1.331$, car
$$ns_x^2 = \sum_{i=1}^{n}(x_1 - \bar{x})^2 = 28.769, \quad ns_{xy} = \sum_{i=1}^{n}(x_1 - \bar{x})(y_1 - \bar{y}) = -38.292 \text{ et } \mathrm{SC}_{TOT} = \sum_{i=1}^{n}(y_1 - \bar{y})^2 = 62.183.$$

3. L'estimateur de la variance des erreurs :

$$\widehat{\sigma}_\varepsilon^2 = \frac{\sum_{i=1}^n e_i^2}{n-2} = \frac{\text{SC}_{TOT} - \widehat{\beta}_1^2 \sum_{i=1}^n (x_1 - \bar{x})^2}{n-2} = \frac{11.217}{11} = 1.020.$$

4. Pour les tests, les statistiques sont :

$$t_{\beta_1} = \frac{\widehat{\beta}_1 - \beta_1^0}{se(\widehat{\beta}_1)} \text{ et } t_{\beta_0} = \frac{\widehat{\beta}_0 - \beta_0^0}{se(\widehat{\beta}_0)},$$

où

$$se(\widehat{\beta}_1) = \frac{\widehat{\sigma}_\varepsilon}{\sqrt{ns_x^2}} = \frac{1.010}{\sqrt{28.769}} = 0.188$$

et

$$se(\widehat{\beta}_0) = \widehat{\sigma}_\varepsilon \sqrt{1/n + \bar{x}^2/ns_x^2} = 1.010\sqrt{1/13 + 3.308^2/28.769} = 0.683.$$

Donc,

$$t_{\beta_1} = \frac{-1.331 - 0}{0.188} = -7.068 \text{ et } t_{\beta_0} = \frac{18.626 - 0}{0.683} = 27.273.$$

Pour chaque cas, on doit comparer ces statistiques avec la t-valeur hypothètique, $t_{0.975,n-2} = t_{0.975,11} = 2.200$. Les valeurs absolues de ces statistiques sont plus grandes que les valeurs hypothètique. Donc, on rejette les hypothèses nulles et on conclut que β_0 et β_1 sont significativement différents de 0.

5. Construire des intervalles de confiance pour la pente et pour la constante.

$$\text{IC}_{\beta_1} = \left[\widehat{\beta}_1 - t_{1-\alpha/2,n-2} se(\widehat{\beta}_1), \widehat{\beta}_1 + t_{1-\alpha/2,n-2} se(\widehat{\beta}_1)\right]$$

et respectivement

$$\text{IC}_{\beta_0} = \left[\widehat{\beta}_0 - t_{1-\alpha/2,n-2} se(\widehat{\beta}_0), \widehat{\beta}_0 + t_{1-\alpha/2,n-2} se(\widehat{\beta}_0)\right].$$

Dans notre cas :

$$\text{IC}_{\beta_1} = [-1.745, -0.917] \text{ et } \text{IC}_{\beta_0} = [17.123, 20.129].$$

6. L'intervalle de confiance pour la moyenne de Y, si on connaît $X = 7$:

$$\text{IC}_{\widehat{y}^*} = \left[\widehat{y}^* - t_{1-\alpha/2,n-2}\widehat{\sigma}_\varepsilon \sqrt{\left\{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{ns_x^2}\right\}}, \widehat{y}^* + t_{1-\alpha/2,n-2}\widehat{\sigma}_\varepsilon \sqrt{\left\{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{ns_x^2}\right\}}\right],$$

où $\widehat{y}^* = \widehat{\beta}_0 + \widehat{\beta}_1 x^* = 9.309$. Donc, $\text{IC}_{\widehat{y}^*} = (6.356, 12.262)$.

7. L'intervalle de prédiction de Y si on connaît $X = 7$:

$$\text{IP}_{\widehat{y}^*} = \left[\widehat{y}^* - t_{1-\alpha/2,n-2}\widehat{\sigma}_\varepsilon \sqrt{\left\{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{ns_x^2}\right\}}), \widehat{y}^* + t_{1-\alpha/2,n-2}\widehat{\sigma}_\varepsilon \sqrt{\left\{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{ns_x^2}\right\}}\right],$$

où $\widehat{y}^* = \widehat{\beta}_0 + \widehat{\beta}_1 x^* = 9.309$. Donc, $\text{IP}_{\widehat{y}^*} = (4.134, 14.484)$.

8. Le tableau d'analyse de la variance :

Source	d.l.	SC	CM	F
Régression	1	50.966	50.966	49.967
Résiduelle	11	11.217	1.020	
Total	12	62.183		

On a donc $\text{SC}_{RES} = 11.217$, $\widehat{\sigma}_\varepsilon^2 = 1.020$, $R^2 = 0.820$ et $F = 49.967$.

9. La dernière observation a une valeur $h_{13,13}$ supérieure à $2\frac{p}{n} = 2 \times \frac{2}{13} = 0.308$. Cette observation est un point levier. Tous les résidus standardisés sont inférieurs à 3. Aucune observation n'a une distance de Cook supérieure à 1. Il n'y a donc ni outliers ni observations influentes.

Exercice 5.
Le Tableau 20.3 la concentration (mg/ml) d'un composant chimique dans l'eau et la transparence de l'eau lue grâce à un colorimètre.

Table 20.3 – Composant chimique dans l'eau et la transparence de l'eau

Concentration (mg/ml) (X)	4	4	5	5	6	6	7	8	8	9
Mesure lue sur le colorimètre (Y)	69	72	175	180	265	272	335	490	492	415

On voudrait donner une estimation de la mesure de transparence (Y) dans le cas où la concentration du ce composant chimique dans l'eau est de 10 (mg/ml). On ajuste un modèle de régression linéaire sur nos données : $y_i = \alpha + \beta x_i + \varepsilon_i$. Pour le calcul des estimations des paramètres, les expressions suivantes sont connues :

$$\sum_{i=1}^{10} y_i^2 = 983793 \quad \sum_{i=1}^{10} x_i^2 = 412 \text{ et } \quad \sum_{i=1}^{10} y_i x_i = 19497.$$

1. Donner la matrice $\mathrm{X}^\top\mathrm{X}$ et la matrice $\mathrm{X}^\top\mathrm{y}$.
2. Donner les valeurs estimées des coefficients α et β.
3. Donner un estimateur sans biais de la variance des erreurs du modéle ($\widehat{\sigma}_\varepsilon^2$).
4. Tester si ces deux coefficients sont significativement différents de 0 ($\alpha = 0.05$).
5. Donner des intervalles de confiance à 95% pour α et pour β.
6. Donner l'intervalle de confiance et l'intervalle de prediction de la mesure si la concentration est de 10 (mg/ml).
7. Construire le tableau d'ANOVA et donner les valeurs de SC_{RES}, $\widehat{\sigma}_\varepsilon^2$, R^2 et F.
8. À partir du Tableau 20.4 donner des "leverage points".

Table 20.4 – Composant chimique et transparence, leviers, résidus standardisés et distance de Cook

Concentration (mg/ml) (X)	Mesure lue sur le colorimètre (Y)	h_{ii}	st.res	D_{Cook}
4	69	0.275	-0.485	0.045
4	72	0.275	-0.412	0.032
5	175	0.152	0.019	0.000
5	180	0.152	0.132	0.002
6	265	...	0.122	0.001
6	272	...	0.276	0.004
7	335	...	-0.216	0.003
8	490	0.217	1.410	0.276
8	492	0.217	1.457	0.295
9	415	0.384	-2.658	2.203

9. Peut-on considérer que les observations 5, 6 et 7 sont des points "leviers" ? Expliquer brièvement pourquoi.

Solution

1. On a :
$$\mathrm{X}^\top\mathrm{X} = \begin{pmatrix} 10 & 62 \\ 62 & 412 \end{pmatrix} \text{ et } \mathrm{X}^\top\mathrm{y} = \begin{pmatrix} 2765 \\ 19497 \end{pmatrix}.$$
2. Les estimateurs :
$$\widehat{\alpha} = -252.297 \text{ et } \widehat{\beta} = 85.290,$$
car
$$ns_x^2 = 72.6, \quad \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = 2354 \text{ et } \mathrm{SC}_{TOT} = 219270.5$$

3. L'estimateur de la variance :

$$\widehat{\sigma}_\varepsilon^2 = \frac{\text{SC}_{RES}}{n-p} = \frac{\text{SC}_{TOT} - \widehat{\beta}^2 n s_x^2}{n-2} = \frac{18497.499}{8} = 2312.187$$

4. Pour les tests, les statistiques sont :

$$t_\beta = \frac{\widehat{\beta} - \beta^0}{se(\widehat{\beta})} \text{ et } t_\alpha = \frac{\widehat{\alpha} - \alpha^0}{se(\widehat{\alpha})},$$

où

$$se(\widehat{\beta}) = \frac{\widehat{\sigma}_\varepsilon}{\sqrt{n s_x^2}} = \frac{48.085}{\sqrt{27.600}} = 9.153$$

et

$$se(\widehat{\alpha}) = \widehat{\sigma}_\varepsilon \sqrt{1/n + \bar{x}^2/n s_x^2} = 48.085\sqrt{1/10 + 6.2^2/2354} = 58.750.$$

Donc,

$$t_\beta = \frac{85.290 - 0}{9.153} = 9.318 \text{ et } t_\alpha = \frac{-252.297 - 0}{58.750} = -4.294.$$

Pour chaque cas, on doit comparer ces statistiques avec le quantile $t_{0.975,n-2} = t_{0.975,8} = 2.310$. Les valeurs absolues de ces statistiques sont plus grandes que les quantiles. Donc, on rejette les hypothèses et on conclut que α et β sont significativement différents de 0.

5. Construction des intervalles de confiance pour la pente et pour la constante.

$$\text{IC}_\beta = \left[\widehat{\beta} - t_{1-\alpha/2,n-2} se(\widehat{\beta}), \widehat{\beta} + t_{1-\alpha/2,n-2} se(\widehat{\beta})\right]$$

et respectivement

$$\text{IC}_\alpha = \left[\widehat{\alpha} - t_{1-\alpha/2,n-2} se(\widehat{\alpha}), \widehat{\alpha} + t_{1-\alpha/2,n-2} se(\widehat{\alpha})\right].$$

Dans notre cas,

$$\text{IC}_\beta = [64.147, 106.433] \text{ et } \text{IC}_\alpha = [-388.010, -116.585].$$

6. L'intervalle de confiance pour la moyenne de Y, si on connaît $X = 10$:

$$\text{IC}_{\widehat{y}^*} = \left[\widehat{y}^* - t_{1-\alpha/2,n-2}\widehat{\sigma}_\varepsilon \sqrt{\left\{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{n s_x^2}\right\}}, \widehat{y}^* + t_{1-\alpha/2,n-2}\widehat{\sigma}_\varepsilon \sqrt{\left\{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{n s_x^2}\right\}}\right],$$

où $\widehat{y}^* = \widehat{\alpha} + \widehat{\beta} x^* = 600.603$. Donc,

$$\text{IC}_{\widehat{y}^*} = (531.382, 669.824)$$

7. L'intervalle de prédiction de Y si on connaît $X = 10$:

$$\text{PI}_{\widehat{y}^*} = \left[\widehat{y}^* - t_{1-\alpha/2,n-2}\widehat{\sigma}_\varepsilon \sqrt{\left\{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{n s_x^2}\right\}}, \widehat{y}^* + t_{1-\alpha/2,n-2}\widehat{\sigma}_\varepsilon \sqrt{\left\{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{n s_x^2}\right\}}\right],$$

où $\widehat{y}^* = \widehat{\alpha} + \widehat{\beta} x^* = 600.603$. Donc,

$$\text{PI}_{\widehat{y}^*} = [420.305, 780.901].$$

8. Le tableau d'analyse de la variance :

Source	d.l.	SC	CM	F
Régression	1	200773.001	200773.001	86.833
Résiduelle	8	18497.499	2312.187	
Total	9	219270.5		

$$\text{SC}_{RES} = 18497.499, \quad \widehat{\sigma}_{\varepsilon}^2 = 2312.187, \quad R^2 = 0.916 \text{ et } F = 86.833.$$

9. Une observation est un point levier si son $h_{ii} > 2\frac{p}{n} = 2 \times \frac{2}{n} = 0.4$. La somme des h_{ii} doit être égale à $p = 2$. La somme des valeurs des h_{ii} données est déjà 1.67. Donc, la somme des 3 valeur manquante est 0.33. Donc, aucune de ces trois valeurs ne peut pas être plus grande que 0.4. Donc, aucune de ces trois point n'est pas un point levier. Il n'y a donc pas de points leviers.

Exercice 6.
On dispose des prix trimestriels d'un titre pendant les années 2006 à 2008. Dans la suite de l'exercice, X et Y représenteront respectivement le temps (en trimestre) et le prix du titre (en CHF) en question. Ces données se trouvent dans le tableau ci-dessous. On souhaite effectuer un ajustement (régression) linéaire.

X (temps)	1	2	3	4	5	6	7	8	9	10	11	12
Y (prix)	128	130	140	112	100	110	120	116	80	90	100	80

On a déjà les résultats suivants :

$$\sum_{i=1}^{12} x_i = 78 \qquad \sum_{i=1}^{12} y_i = 1306 \qquad \sum_{i=1}^{12} x_i^2 = 650 \qquad \sum_{i=1}^{12} y_i^2 = 146284 \qquad r_{xy} = -0.811$$

1. Calculer les variances marginales de X et Y.
2. Calculer la covariance entre X et Y.
3. (a) Déterminer l'équation de la droite de régression de Y en fonction de X.
   (b) Expliquer en trois lignes au maximum pourquoi on fait ici une régression de Y en X plutôt qu'une régression de X en Y.
4. Déterminer le coefficients de détermination entre les variables X et Y et ensuite déterminer la qualité de l'ajustement.
5. Donner la valeur ajustée et le résidu de la dernière observation.
6. Estimer à l'aide du modèle le prix du titre le premier trimestre de 2009 (c'est-à-dire pour $t = 13$).

Solution

1. Les variances :
$$s_x^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = \frac{1}{12}650 - \left(\frac{78}{12}\right)^2 = 11.917$$
et
$$s_y^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \bar{y}^2 = \frac{1}{12}146284 - \left(\frac{1306}{12}\right)^2 = 345.639.$$
2. La covariance :
$$s_{xy} = r_{xy}s_x s_y = -0.811\sqrt{11.917}\sqrt{345.639} = -52.049.$$
3. (a) La droite de régression :
$$b = \frac{s_{xy}}{s_x^2} = \frac{-52.049}{11.917} = -4.368,$$
$$a = \bar{y} - b\bar{x} = \frac{1306}{12} + 4.368\frac{78}{12} = 137.225$$
et
$$y = a + bx = 137.225 - 4.368x.$$
   (b) Car cela a un sens de supposer que le prix peut être une fonction du temps mais cela n'a pas de sens de supposer que le prix peut expliquer le temps.
4. La corrélation : $r_{xy}^2 = 0.658$. Dons, 65.8% de la variance est expliquée par le modèle.
5. Les valeurs prédites :
$$y_{12}^* = a + bx_{12} = 137.225 - 4.368 \times 12 = 84.809$$
et
$$e_{12} = y_{12} - y_{12}^* = 80 - 84.809 = -4.809.$$
6. Le prix du titre le premier trimestre de 2009 :
$$y_{13}^* = a + 13b = 137.225 - 13 \times 4.368 = 80.441.$$

Exercice 7.
On dispose des données suivantes :

y	x_2	x_3
32.4	−6	−8
21.6	3	−6
18.0	0	12
28.8	−3	10
19.8	6	−6
16.2	0	6
19.8	−3	−12
14.4	3	2
12.6	3	0
21.6	−3	2

Soit le modèle linéaire : $y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i$. On voudrait estimer les coefficients de ce modèle (β_1, β_2 et β_3). Pour ce faire, on a déjà calculé les expressions suivantes :

$$\sum_{i=1}^{10} y_i^2 = 4549, \sum_{i=1}^{10} x_{2,i}x_{3,i} = 0, \sum_{i=1}^{10} x_{2,i}^2 = 126,$$

$$\sum_{i=1}^{10} x_{3,i}^2 = 568, \sum_{i=1}^{10} x_{2,i}y_i = -140.4 \text{ et } \sum_{i=1}^{10} x_{3,i}y_i = -72.$$

1. Écrire le modèle sous forme matricielle : $\mathrm{y} = \mathrm{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. (en précisant le contenu et les dimensions de chaque vecteur et matrice)
2. Calculer analytiquement, puis numériquement $\mathrm{X}^\top\mathrm{X}$, $(\mathrm{X}^\top\mathrm{X})^{-1}$ et $\mathrm{X}^\top\mathrm{y}$.
3. Donner les estimations des coefficients, les erreurs standard et les t-valeurs (pour le test général H_0 : $\beta_j = 0$). (Donc, remplir le tableau suivant :)

j	$\widehat{\beta}_j$	$se(\widehat{\beta}_j)$	$t-$valeurs
1	...	1.571	...
2	...	...	...
3	...	...	...

4. Construire le tableau d'ANOVA.
5. Donner la valeur de $\widehat{\sigma}_\varepsilon^2$, la valeur du R^2 puis interpréter le R^2.
6. Tester, au seuil de signification $\alpha = 0.05$, les hypothèses suivantes :
   (a) $H_0 : \beta_j = 0$ pour tout les $j \neq 1$ vs H_1 : au moins un des $\beta_j \neq 0$, $j \neq 1$.
   (b) $H_0 : \beta_2 = -1$ vs $\beta_2 \neq -1$.
   (c) $H_0 : \beta_2 \geq -1$ vs $\beta_2 < -1$.
   (d) $H_0 : \beta_1 + \beta_2 = 18$ vs $\beta_1 + \beta_2 \neq 18$.

Solution

1. La forme matricielle du modèle est :

$$\begin{pmatrix} y_1 \\ \vdots \\ y_{10} \end{pmatrix} = \begin{pmatrix} 1 & x_{2,1} & x_{3,1} \\ \vdots & \vdots & \vdots \\ 1 & x_{2,10} & x_{3,10} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_{10} \end{pmatrix},$$

où

$$\underset{(10,1)}{\mathrm{y}} = \underset{(10,3)}{\mathrm{X}} \underset{(3,1)}{\boldsymbol{\beta}} + \underset{(10,1)}{\boldsymbol{\varepsilon}}$$

On calcule :

$$\mathrm{X}^\top\mathrm{X} = \begin{pmatrix} 1 & \cdots & 1 \\ x_{2,1} & \cdots & x_{2,10} \\ x_{3,1} & \cdots & x_{3,10} \end{pmatrix} \begin{pmatrix} 1 & x_{1,1} & x_{2,1} \\ \vdots & \vdots & \vdots \\ 1 & x_{2,10} & x_{3,10} \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^n x_{2,i} & \sum_{i=1}^n x_{3,i} \\ \sum_{i=1}^n x_{2,i} & \sum_{i=1}^n x_{2,i}^2 & \sum_{i=1}^n x_{2,i}x_{3,i} \\ \sum_{i=1}^n x_{3,i} & \sum_{i=1}^n x_{2,i}x_{3,i} & \sum_{i=1}^n x_{3,i}^2 \end{pmatrix},$$

$$\mathrm{X}^\top \mathrm{y} = \begin{pmatrix} 1 & \cdots & 1 \\ x_{2,1} & \cdots & x_{2,10} \\ x_{3,1} & \cdots & x_{3,10} \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_{10} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{2,i} y_i \\ \sum_{i=1}^n x_{3,i} y_i \end{pmatrix}.$$

2. L'estimateur de $\boldsymbol{\beta}$:

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

Comme

$$\mathrm{X}^\top \mathrm{X} = \begin{pmatrix} 10 & 0 & 0 \\ 0 & 126 & 0 \\ 0 & 0 & 568 \end{pmatrix},$$

on a

$$\mathrm{X}^\top \mathrm{X})^{-1} = \begin{pmatrix} \frac{1}{10} & 0 & 0 \\ 0 & \frac{1}{126} & 0 \\ 0 & 0 & \frac{1}{568} \end{pmatrix}.$$

3. Finalement,

$$\widehat{\boldsymbol{\beta}} = \begin{pmatrix} \frac{1}{10} & 0 & 0 \\ 0 & \frac{1}{126} & 0 \\ 0 & 0 & \frac{1}{568} \end{pmatrix} \begin{pmatrix} 205.2 \\ -140.4 \\ -72 \end{pmatrix} = \begin{pmatrix} 20.520 \\ -1.114 \\ -0.127 \end{pmatrix} = \begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \widehat{\beta}_3 \end{pmatrix}.$$

La somme des carrés totale :

$$\mathrm{SC}_{TOT} = n \times s_x^2 = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - n\bar{y}^2 = 4549 - 10 \times (205.2/10)^2 = 338.256.$$

On a :

$$\mathrm{se}(\widehat{\beta}_1) = \widehat{\sigma}_\varepsilon \sqrt{(\mathrm{X}^\top \mathrm{X})^{-1}_{11}} = \widehat{\sigma}_\varepsilon \sqrt{\frac{1}{10}} = 1.571,$$

Donc, $\widehat{\sigma}_\varepsilon = 4.967$. De plus,

$$\mathrm{se}(\widehat{\beta}_2) = \widehat{\sigma}_\varepsilon \sqrt{(\mathrm{X}^\top \mathrm{X})^{-1}_{22}} = 4.967 \sqrt{\frac{1}{126}} = 0.442,$$

$$\mathrm{se}(\widehat{\beta}_3) = \widehat{\sigma}_\varepsilon \sqrt{(\mathrm{X}^\top \mathrm{X})^{-1}_{33}} = 4.967 \sqrt{\frac{1}{568}} = 0.208.$$

$\widehat{\beta}_j$	$se(\widehat{\beta}_j)$	$t-$valeurs
20.520	1.571	13.065
−1.114	0.442	−2.518
−0.127	0.208	−0.608

Le quantile est $t_{0.95,7} = 2.36$ donc $\widehat{\beta}_1$ et $\widehat{\beta}_2$ sont significativement differents de 0. Par contre, on ne rejette pas l'hypothèse, que $\widehat{\beta}_3$ est égale a 0.

4. Le tableau d'analyse de la variance :

Source	d.l.	SC	CM	F
Régression	2	165.572	82.786	3.356
Résidus	7	172.684	24.669	
Total	9	338.256		

Le quantile est $F_{0.95,2,7} = 4.74$. Donc, le test n'est pas significatif.

5. $\widehat{\sigma}_\varepsilon^2 = 24.669$ $\widehat{\sigma}_\varepsilon = 4.967$ et $R^2 = 48.949\%$. Le modèle explique les 48.949% de la variabilité de la variable dépendante.

6. Tests :

(a) On teste : $H_0 : \beta_j = 0$ pour tout les $j \neq 1$ vs H_1 : au moins un des $\beta_j \neq 0$. La statistique de test :

$$F_{obs} = 3.356 < 4.74 = F_{hyp} = F_{2,7,0.95} = F_{p-1,n-p,1-\alpha},$$

On ne rejette pas H_0.

(b) On teste : $H_0 \ : \ \beta_2 = -1 \quad$ vs $\quad \beta_2 \neq -1$,

$$t_{obs} = \frac{\widehat{\beta}_2 - \beta_{20}}{se(\widehat{\beta}_2)} = \frac{-1.114 - -1}{0.442} = -0.258,$$

$$|t_{obs}| < t_{0.975,10-3} = 2.36.$$

Donc, on ne rejette pas H_0.

(c) On teste : $H_0 \ : \ \beta_2 \geq -1 \quad$ vs $\quad \beta_2 < -1$. C'est un test unilatéral.

$$t_{obs} = \frac{-1.114 - -1}{0.442} = -0.258 < t_{0.95,10-3} = 1.9.$$

Donc, on ne rejette pas H_0. (noter : on met le signe $\leq$ ou $\geq$ dans l'hypothèse H_0.)

(d) On teste : $H_0 \ : \ \beta_1 + \beta_2 = 18 \quad$ vs $\quad \beta_1 + \beta_2 \neq 10$. $H_0 \ : \mathrm{R}\boldsymbol{\beta} = r$, où $r = 18$ et $\mathrm{R} = \begin{pmatrix} 1 & 1 & 0 \end{pmatrix}$. Donc,

$$\mathrm{R}\widehat{\boldsymbol{\beta}} - r = 20.520 - 1.114 - 18 = 1.406,$$

$$\mathrm{R}\left(\mathrm{X}^\top \mathrm{X}\right)^{-1} \mathrm{R}^\top = \begin{pmatrix} \frac{1}{10} & \frac{1}{126} & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \frac{69}{630} = 0.110.$$

Donc,

$$t_{obs} = \frac{\mathrm{R}\widehat{\boldsymbol{\beta}} - r}{\widehat{\sigma}\sqrt{\mathrm{R}\left(\mathrm{X}^\top \mathrm{X}\right)^{-1} \mathrm{R}^\top}} = \frac{1.406}{4.967\sqrt{\frac{69}{630}}} = 0.855,$$

$$|t_{obs}| = 0.855 < t_{0.975,10-3} = 2.365.$$

Donc, on ne rejette pas H_0.

Exercice 8.
Montrer que dans un modèle linéaire général : $SC_{RES} = y^\top y - \widehat{\boldsymbol{\beta}}^\top X^\top y$.

Solution

$$\begin{aligned} SC_{RES} &= e\top e = (y - X\widehat{\boldsymbol{\beta}})^\top (y - X\widehat{\boldsymbol{\beta}}) = (y^\top - \widehat{\boldsymbol{\beta}}^\top X^\top)(y - X\widehat{\boldsymbol{\beta}}) \\ &= y^\top y - \widehat{\boldsymbol{\beta}}^\top X^\top y - y^\top X\widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{\beta}}^\top X^\top X\widehat{\boldsymbol{\beta}} = y^\top y - 2\widehat{\boldsymbol{\beta}}^\top X^\top y + \widehat{\boldsymbol{\beta}}^\top X^\top X(X^\top X)^{-1} X^\top y \\ &= y^\top y - 2\widehat{\boldsymbol{\beta}}^\top X^\top y + \widehat{\boldsymbol{\beta}}^\top X^\top y = y^\top y - \widehat{\boldsymbol{\beta}}^\top X^\top y. \end{aligned}$$

Exercice 9.
Les données que nous allons utiliser concernent 463 cours des années académiques 2000-2002 de l'University du Texas de Austin. Les variables nous intéressant sont les suivantes :

y_i :	note d'enseignement sur une échelle de 1 (enseignement très insatisfaisant) à 5 (enseignement excellent).
x_{i2} :	âge du professeur
x_{i3} :	variable binaire valant $\begin{cases} 1, & \text{si le professeur est une femme,} \\ 0, & \text{sinon.} \end{cases}$

On ajuste le modèle : $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$. Pour ce faire, on a utilisé

$$\sum_{i=1}^{n} y_i = 1851.2, \sum_{i=1}^{n} x_{i2} = 22393, \sum_{i=1}^{n} x_{i3} = 195, \sum_{i=1}^{n} y_i^2 = 7543.84,$$

$$\sum_{i=1}^{n} x_{i2}^2 = 1127433, \sum_{i=1}^{n} x_{i3}^2 = 195, \sum_{i=1}^{n} x_{i2}x_{i3} = 8793, n = 463.$$

On a alors trouvé les valeurs des estimateurs suivants : $\widehat{\beta}_1 = 4.364$, $\widehat{\beta}_2 = -0.006$, $\widehat{\beta}_3 = -0.201$ et $\widehat{\sigma}_\varepsilon^2 = 0.299$.

1. Le modèle s'écrit matriciellement $\mathrm{y} = \mathrm{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. Donner la matrice $\mathrm{X}^\top\mathrm{X}$ dans le cas considéré ici.
2. (a) À l'aide du modèle, donner une prédiction de la note obtenue par un professeur de sexe féminin âgé de 36 ans.
   (b) Dans nos données, on trouve un individu ayant ces caractéristiques (professeur de sexe féminin et âgé de 36 ans) dont la note d'enseignement est de 4.3. Expliquer (en moins de 2 lignes) pourquoi on trouve une note différente de celle prédite.
3. À l'aide du modèle (20.1) estimer la différence de note d'enseignement selon que le professeur est un homme ou une femme de même âge.
4. Construire le tableau d'analyse de la variance.
5. Tester : $\begin{cases} H_0 : & \beta_2 = \beta_3 = 0 \\ H_1 : & \text{au moins un des } \beta_j \neq 0 \ (j = 2, 3) \end{cases}$ au seuil 95%.
6. Donner le R^2 et commenter (en une ligne).

Solution

1. On calcule
$$\mathrm{X}^\top\mathrm{X} = \begin{pmatrix} n & \sum_{i=1}^n x_{i2} & \sum_{i=1}^n x_{i3} \\ \sum_{i=1}^n x_{i2} & \sum_{i=1}^n x_{i2}^2 & \sum_{i=1}^n x_{i2}x_{i3} \\ \sum_{i=1}^n x_{i3} & \sum_{i=1}^n x_{i2}x_{i3} & \sum_{i=1}^n x_{i3}^2 \end{pmatrix} = \begin{pmatrix} 463 & 22393 & 195 \\ 22393 & 1127433 & 8793 \\ 195 & 8793 & 195 \end{pmatrix}.$$
2. (a) $\widehat{y} = 4.464 - 0.006 \times 36 - 0.201 \times 1 = 3.947$
   (b) La note ci-dessus est une prédiction faite à partir du modèle. Elle n'est pas égale à la valeur observée car les coefficients de régression sont estimés et le terme d'erreur est inconnu.
3. Cette différence vaut $\widehat{\beta}_3 = -0.201$.
4. Les sonmmes de carrés :
$$\begin{aligned} \mathrm{SC}_{RES} &= (n-3)\widehat{\sigma}_\varepsilon^2 = 460 \times 0.299 = 137.54, \\ \mathrm{SC}_{TOT} &= \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{2} y_i^2 - n\bar{y}^2 \\ &= 7543.84 - 463 \times \left(\frac{1}{463} \times 1851.2\right)^2 = 7543.84 - 463 \times 15.986 = 142.239, \\ \mathrm{SC}_{REG} &= \mathrm{SC}_{TOT} - \mathrm{SC}_{RES} = 142.239 - 137.54 = 4.699. \end{aligned}$$

   Tableau de l'analyse de la variance :

Source de variation	Somme des carrés	Degrés de liberté	Carrés moyens	F_c
Régression	4.699	2	2.350	$F_{obs} = \frac{2.350}{0.299} = 7.860$
Résiduelle	137.54	460	0.299	
Totale	142.239	462	0.308	

5. On teste $\begin{cases} H_0 : & \beta_2 = \beta_3 = 0 \\ H_1 : & \text{au moins un des } \beta_j \neq 0 \ (j = 2, 3). \end{cases}$
   Avec le test de Fisher, on a $F_{obs} = 7.860 > F_{0.95,2,460} = F_{0.95,2,\infty} = 2.996$. On rejette ainsi H_0.
6. Le R^2 vaut
$$R^2 = \frac{\text{SC}_{REG}}{\text{SC}_{RES}} = \frac{4.699}{142.239} = 0.033.$$
   Donc, 3.3% de la variance de y est expliquée pas le modèle.

Exercice 10.
Nous allons ajuster et tester un modèle : $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$, sur les 10 données de la Figure 20.1. Dans ce but, nous avons déjà calculé

$$\left(\mathrm{X}^\top \mathrm{X}\right)^{-1} = \begin{pmatrix} 4.75 & -0.91 \\ -0.91 & 0.18 \end{pmatrix}, \left(\mathrm{X}^\top \mathrm{X}\right)^{-1} \mathrm{X}^\top \mathrm{y} = \begin{pmatrix} 4.32 \\ 3.01 \end{pmatrix}, \widehat{\sigma}^2_\varepsilon = 24.71,$$

ainsi que

i	1	2	3	4	5	6	7	8	9	10
h_i	0.24	0.18	0.13	0.13	0.22	0.12	0.11	0.54	0.10	0.23

où $h_i = [P_\mathrm{X}]_{ii}$.

1. Donner les coefficients de la régression.
2. (a) Tester $\begin{cases} H_0 : & \beta_2 = 4 \\ H_1 : & \beta_2 \neq 4 \end{cases}$ au seuil 95%.
   (b) Donner un intervalle de confiance à 95% pour β_2.
   (c) Donner un intervalle de confiance à 90% pour β_2.
   (d) On remarque que l'un des intervalles est contenu dans l'autre.
   Expliquer (en moins de 4 lignes) ce phénomène.
3. Tester $\begin{cases} H_0 : & \beta_1 + 2\beta_2 = 0 \\ H_1 : & \beta_1 + 2\beta_2 \neq 0 \end{cases}$ au seuil 95%.
4. Donner un intervalle de confiance à 95% pour la prévision correspondant à $x_i = 4$.
5. On observe sur la Figure 20.1 que certains points pourraient avoir un levier important. Déterminer par calculs quels points ont en effet un levier important.

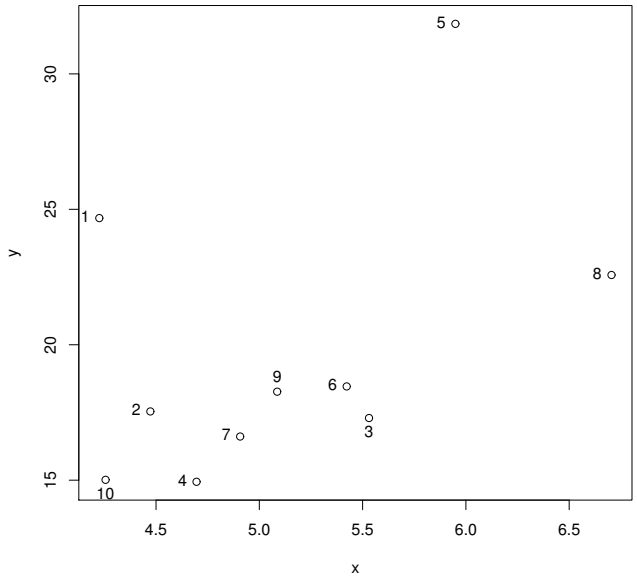


FIGURE 20.1 – Nuage de points

6. Parmi les observations 5 et 8, laquelle à une plus grande influence sur la régression ?
   Justifier au moyen de la distance de Cook.
   Aide : $(x_5, y_5) = (6, 32)$ et $(x_8, y_8) = (6.7, 22.5)$.

Solution

1. Les coefficients sont : $\widehat{\beta}_1 = 4.32, \widehat{\beta}_2 = 3.01$.
2. (a) Il y a deux tests différents possibles.

— Test de Student :

$$t_{obs} = \frac{\widehat{\beta}_2 - 4}{\sqrt{\widehat{\sigma}_\varepsilon^2 \left[(\mathrm{X}^\top \mathrm{X})^{-1}\right]_{22}}} = \frac{3.01 - 4}{\sqrt{24.71 \times 0.18}} = -0.469.$$

Comme $|t_{obs}| < t_{0.975,n-p} = t_{0.975,8} = 2.306$, on ne rejette pas H_0.

— Test de Fisher :

$$F_c = \frac{\left(\widehat{\beta}_2 - 4\right)^2}{\widehat{\sigma}_\varepsilon^2 \left[(\mathrm{X}^\top \mathrm{X})^{-1}\right]_{22}} = t_{obs}^2 = 0.220,$$

Comme $F_{obs} < F_{0.95,n-p} = F_{0.95,8} = 5.318$, on ne rejette pas H_0.

(b) Intervalle de confiance à 95% :

$$\begin{aligned} \mathrm{IC}(95\%)_{\beta_2} &= \left[\widehat{\beta}_2 - t_{0.975,8} \times \sqrt{\widehat{\sigma}_\varepsilon^2 \left[(\mathrm{X}^\top \mathrm{X})^{-1}\right]_{22}}, \widehat{\beta}_2 + t_{0.975,8} \times \sqrt{\widehat{\sigma}_\varepsilon^2 \left[(\mathrm{X}^\top \mathrm{X})^{-1}\right]_{22}}\right] \\ &= \left[3.01 - 2.306 \times \sqrt{24.71 \times 0.18}, 3.01 + 2.306 \times \sqrt{24.71 \times 0.18}\right] \\ &= [-1.853, 7.873] \end{aligned}$$

(c) Intervalle de confiance à 90% :

$$\begin{aligned} \mathrm{IC}(90\%)_{\beta_2} &= \left[\widehat{\beta}_2 - t_{0.95,8} \times \sqrt{\widehat{\sigma}_\varepsilon^2 \left[(\mathrm{X}^\top \mathrm{X})^{-1}\right]_{22}}, \widehat{\beta}_2 + t_{0.95,8} \times \sqrt{\widehat{\sigma}_\varepsilon^2 \left[(\mathrm{X}^\top \mathrm{X})^{-1}\right]_{22}}\right] \\ &= \left[3.01 - 1.860 \times \sqrt{24.71 \times 0.18}, 3.01 + 1.860 \times \sqrt{24.71 \times 0.18}\right] \\ &= [-0.913, 6.933] \end{aligned}$$

(d) $\mathrm{IC}(90\%)_{\beta_2} \subset \mathrm{IC}(95\%)_{\beta_2}$. Plus le niveau de confiance de l'intervalle est grand plus la longueur de l'intervalle de confiance augmente.

3. Tester $H_0 : \beta_1 + 2\beta_2 = 0$ revient à tester $\mathrm{R}\boldsymbol{\beta} = \mathrm{r}$, avec $\mathrm{R} = (1, 2)$ et $\mathrm{r} = 0$. On peut donc faire un test de Wald :

$$F_{obs} = \frac{\left(\mathrm{R}\widehat{\boldsymbol{\beta}} - \mathrm{r}\right) \top \left\{\mathrm{r}(\mathrm{X}^\top \mathrm{X})^{-1} \mathrm{r}^\top\right\}^{-1} \left(\mathrm{R}\widehat{\boldsymbol{\beta}} - \mathrm{r}\right) \frac{1}{q}}{\widehat{\sigma}_\varepsilon^2} = \frac{(10.34)\{1.83\}^{-1}(10.34)\frac{1}{q}}{24.71} = 2.364$$

Comme $F_{obs} < F_{1-\alpha,q,n-p} = F_{0.95,1,8} = 5.318$, on ne rejette pas H_0.

4. Intervalle sur une prédiction :

$$\mathrm{IC}(95\%) = \left[\widehat{y}_i - t_{0.95,8} \times \sqrt{\widehat{\mathrm{var}}\left(\widehat{y}_i - y_i\right)}, \widehat{y}_i + t_{0.95,8} \times \sqrt{\widehat{\mathrm{var}}\left(\widehat{y}_i - y_i\right)}\right]$$

On a de plus

$$\begin{aligned} (\mathrm{X}^\top \mathrm{X})_{22}^{-1} &= \frac{1}{n S_x^2} \text{ donc } S_x^2 = \frac{1}{n \times 0.18} = 0.556 \\ (\mathrm{X}^\top \mathrm{X})_{12}^{-1} &= \frac{-\bar{x}}{n S_x^2} = -0.91 \text{ donc } \bar{x} = \frac{0.91}{0.18} = 5.056 \end{aligned}$$

et on peut alors calculer

$$\begin{aligned} \widehat{\mathrm{var}}\left(\widehat{y}_i - y_i\right) &= \frac{\widehat{\sigma}_\varepsilon^2}{n}\left\{n + 1 + \frac{(x_j - \bar{x})^2}{S_x^2}\right\} = \frac{24.71}{10}\left\{10 + 1 + \frac{(4 - 5.056)^2}{0.556}\right\} \\ &= \frac{24.71}{10}\{10 + 1 + 2.006\} = 32.137 \end{aligned}$$

Remarque : Il est également possible de calculer $\widehat{\text{var}}\left(\widehat{y}_i - y_i\right)$ avec la formule du cas général.
De plus,

$$\widehat{y}_i = \widehat{\beta}_1 + \widehat{\beta}_2 \times 4 = 4.32 + 3.01 \times 4 = 16.36.$$

On peut alors donner l'intervalle de confiance :

$$\text{IC}(95\%) = \left[16.36 - 2.306 \times \sqrt{32.137}, 16.36 + 2.306 \times \sqrt{32.137}\right] = [3.287, 29.433]\,.$$

5. L'observation i a un levier important si $h_i > \frac{p}{n} = \frac{2}{10}$. Les observations 1, 5, 8 et 10 ont ainsi un levier important.
6. La distance de Cook est donnée par

$$D_i = \frac{e_i^2}{p \times \widehat{\sigma}_\varepsilon^2}\left[\frac{h_i}{(1-h_i)^2}\right].$$

On trouve, pour les observations 5 et 8

$$\begin{aligned}
e_5 &= y_5 - \widehat{y}_5 = 32 - (4.32 + 3.01 \times 6) = 9.62,\\
D_5 &= \frac{9.62^2}{2 \times 24.71}\left(\frac{0.22}{(1-0.22)^2}\right) = 0.677,\\
e_8 &= y_8 - \widehat{y}_8 = 22.5 - (4.32 + 3.01 \times 6.7) = -1.987,\\
D_8 &= \frac{-1.987^2}{2 \times 24.71}\left(\frac{0.54}{(1-0.54)^2}\right) = 0.2.
\end{aligned}$$

Comme $D_5 > D_8$, l'observation 5 a une plus grande influence sur la régression.

Exercice 11.
Soit le modèle linéaire général $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots + \beta_p x_{ip} + \varepsilon_i$. Montrer que

$$\mathrm{SC}_{TOT} = \frac{(n-p)\widehat{\sigma}_\varepsilon^2}{1-R^2},$$

où SC_{TOT} est la somme des carrés totale, n est le nombre d'observations et $\widehat{\sigma}_\varepsilon^2$ est l'estimateur sans biais de $\sigma_\varepsilon^2 = \mathrm{var}(\varepsilon_i), i = 1, \ldots, n$.

Solution

$$\frac{(n-p)\widehat{\sigma}_\varepsilon^2}{1-R^2} = \frac{\mathrm{SC}_{RES}}{1-R^2} = \frac{\mathrm{SC}_{RES}}{1-\frac{\mathrm{SC}_{REG}}{\mathrm{SC}_{TOT}}} = \frac{\mathrm{SC}_{RES}}{\frac{\mathrm{SC}_{RES}}{\mathrm{SC}_{TOT}}} = \mathrm{SC}_{TOT}.$$

Exercice 12.
On dispose des données suivantes :

x_i	-5	-4	-3	-2	-1	0	1	2	3	4	5
y_i	1	5	4	7	10	8	9	13	14	13	18

On considère le modèle : $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$. On a calculé

$$\sum_{i=1}^{11} x_i = 0, \sum_{i=1}^{11} x_i^2 = 110, \sum_{i=1}^{11} x_i y_i = 173, \sum_{i=1}^{11} y_i = 102, \sum_{i=1}^{11} y_i^2 = 1311, \widehat{\sigma}_\varepsilon^2 = 4.05.$$

1. Calculer $\widehat{\beta}_1$ et $\widehat{\beta}_2$, les estimateurs des moindres carrés de β_1 et β_2.
   Donner l'expression fonctionnelle de la droite de régression.
2. (a) Donner un intervalle de confiance à 95% pour β_1.
   (b) En se basant sur l'intervalle de confiance donné au point précédent, dire si on rejette ou non l'hypothèse $\begin{cases} H_0 : & \beta_1 = 0 \\ H_1 : & \beta_1 \neq 0 \end{cases}$ au seuil 95%. Justifier.
   (c) En se basant sur les deux points précédents et sur la relation d'inclusion qu'il existe entre un intervalle de confiance à 95% et un intervalle de confiance à 90%, dire si on rejette ou non l'hypothèse $\begin{cases} H_0 : & \beta_1 = 0 \\ H_1 : & \beta_1 \neq 0 \end{cases}$ au seuil 90%. Justifier.
3. Tester $\begin{cases} H_0 : & \beta_2 = 2 \\ H_1 : & \beta_2 \neq 2 \end{cases}$ au seuil 90%.
4. Donner un intervalle de confiance à 95% pour la prévision de y lorsque $x = 3$.

Solution

1. Les paramètres et le modèle :

$$\begin{aligned} \widehat{\beta}_2 &= \frac{s_{xy}}{s_x^2} = \frac{\frac{1}{n}\sum_{i=1}^n x_i y_i - \bar{x} \times \bar{y}}{\frac{1}{n}\sum_{i=1}^n x_i^2 - \bar{x}^2} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} = \frac{173}{110} = 1.573, \\ \widehat{\beta}_1 &= \bar{y} - \widehat{\beta}_2 \bar{x} = \bar{y} = \frac{102}{11} = 9.273, \\ \widehat{y} &= 9.273 + 1.573x. \end{aligned}$$

2. (a) Intervalle pour β_1 :

$$\begin{aligned} \widehat{\text{var}}(\widehat{\beta}_1) &= \frac{\widehat{\sigma}_\varepsilon^2}{n s_x^2}(s_x^2 + \bar{x}^2) = \frac{\widehat{\sigma}_\varepsilon^2}{n} = \frac{4.05}{11} = 0.368, \\ \text{IC}(95\%) &= \widehat{\beta}_1 \pm t_{0.975,9} \times \sqrt{\widehat{\text{var}}(\widehat{\beta}_1)} = 9.273 \pm 2.262 \times \sqrt{0.368} = [7.901, 10.645]. \end{aligned}$$

   (b) Comme $0 \notin \text{IC}(95\%)$, on rejette $H_0 : \beta_1 = 0$ à 95%.
   (c) Comme $\text{IC}(90\%) \subseteq \text{IC}(95\%)$, on a $0 \notin \text{IC}(90\%)$, on rejette $H_0 : \beta_1 = 0$ à 90%.
3. Test pour β_2 :

$$\begin{aligned} \widehat{\text{var}}(\widehat{\beta}_2) &= \frac{\widehat{\sigma}_\varepsilon^2}{n s_x^2} = \frac{4.05}{11\frac{110}{11}} = 0.037, \\ t_{\beta_2} &= \frac{\widehat{\beta}_2 - \beta_{0,2}}{\sqrt{\widehat{\text{var}}(\widehat{\beta}_2)}} = \frac{1.573 - 2}{\sqrt{0.037}} = -2.220. \end{aligned}$$

   Comme $t_{0.95,9} = 1.833$, $|t_\beta| > 1.833$ donc on rejette H_0.
4. Intervalle de confiance sur la prédiction :

$$\begin{aligned} \widehat{\text{var}}(\widehat{y}_j - y_j) &= \frac{\widehat{\sigma}_\varepsilon^2}{n}\left(n + 1 + \frac{(x_j - \bar{x})^2}{s_x^2}\right) = \frac{4.05}{11}\left(12 + \frac{3^2}{\frac{110}{11}}\right) = \frac{4.05}{11} \times 12.9 = 4.750, \\ \widehat{y}_j &= \widehat{\beta}_1 + \widehat{\beta}_2 \times 3 = 9.273 + 1.573 \times 3 = 13.992, \\ \text{IC}(95\%) &= 13.992 \pm 2.262\sqrt{4.75} = [9.062, 18.992]. \end{aligned}$$

Exercice 13.
On dispose des données suivantes :

	x_{i2}	x_{i3}	y_i
	-1	-1	7
	-1	0	8
	0	0	10
	1	0	12
	1	1	13
Somme	0	0	50
Somme des carrés	4	2	526

Nous allons ajuster et tester un modèle : $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$. Dans ce but, on a calculé :

$$\begin{pmatrix} 5 & 0 & 0 \\ 0 & 4 & 2 \\ 0 & 2 & 2 \end{pmatrix}^{-1} = \begin{pmatrix} 0.2 & 0 & 0 \\ 0 & 0.5 & -0.5 \\ 0 & -0.5 & 1 \end{pmatrix} \text{ et } \widehat{\sigma}^2_\varepsilon = 0.01.$$

1. Le modèle considéré ici s'écrit matriciellement $\mathrm{y} = \mathrm{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. Donner la matrice X.
2. Donner $\widehat{\boldsymbol{\beta}}$, l'estimateur de $\boldsymbol{\beta}$.
3. Donner une prévision pour y si $x_{i2} = -1$ et x_{i3}=1.
4. À l'aide du modèle, estimer l'augmentation de y liée à une augmentation de une unité de x_{i2}, x_{i3} restant constant. On s'intéresse à un cas général, on ne choisira pas de valeur particulière pour x_{i3}.
5. Tester $\begin{cases} H_0 : & \beta_3 = 0 \\ H_1 : & \beta_3 \neq 0 \end{cases}$ au seuil 95%.
6. Tester $\begin{cases} H_0 : & \beta_1 - 8\beta_3 = 0 \\ H_1 : & \beta_1 - 8\beta_3 \neq 0 \end{cases}$ au seuil 95%.
7. Construire le tableau d'analyse de la variance.
8. Tester $\begin{cases} H_0 : & \beta_2 = \beta_3 = 0 \\ H_1 : & \text{au moins un des } \beta_j \neq 0 \ (j = 2, 3). \end{cases}$ au seuil 95%.
9. Donner le R^2 et commenter (en une ligne).

Solution

1. La matrice X :
$$\mathrm{X} = \begin{pmatrix} 1 & -1 & -1 \\ 1 & -1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}.$$
2. On a :
$$\mathrm{X}^\top\mathrm{X} = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 4 & 2 \\ 0 & 2 & 2 \end{pmatrix}, \left(\mathrm{X}^\top\mathrm{X}\right)^{-1} = \begin{pmatrix} 0.2 & 0 & 0 \\ 0 & 0.5 & -0.5 \\ 0 & -0.5 & 1 \end{pmatrix},$$
$$\mathrm{X}^\top\mathrm{y} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & 0 & 1 & 1 \\ -1 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 7 \\ 8 \\ 10 \\ 12 \\ 13 \end{pmatrix} = \begin{pmatrix} 50 \\ 10 \\ 6 \end{pmatrix}$$
et
$$\widehat{\boldsymbol{\beta}} = \left(\mathrm{X}^\top\mathrm{X}\right)^{-1}\mathrm{X}^\top\mathrm{y} = \begin{pmatrix} 0.2 & 0 & 0 \\ 0 & 0.5 & -0.5 \\ 0 & -0.5 & 1 \end{pmatrix} \begin{pmatrix} 50 \\ 10 \\ 6 \end{pmatrix} = \begin{pmatrix} 10 \\ 2 \\ 1 \end{pmatrix}.$$

3. $\widehat{y} = 10 + 2 \times (-1) + 1 \times 1 = 9$.
4. $\widehat{\beta}_2 = 2$.
5. La statistique de test :

$$t_{\beta_3} = \frac{\widehat{\beta}_3 - \beta_{3,0}}{\sqrt{\left[(\mathrm{X}^\top \mathrm{X})^{-1} \widehat{\sigma}_\varepsilon^2\right]_{33}}} = \frac{1}{\sqrt{1 \times 0.01}} = \frac{1}{0.1} = 10.$$

   Comme $|t_{\beta_3}| = 10 > t_{0.975,2(n-p)} = 4.303$, on rejette H_0.
6. On peut faire un test de Wald avec $q = 1$, $\mathrm{R} = (1, 0, -8)$ et $\mathrm{R} = 0$. On calcule :

$$\mathrm{R}\widehat{\boldsymbol{\beta}} - \mathrm{r} = 1 \times 10 + 0 \times 2 - 8 \times 1 = 2,$$

$$\mathrm{R}\left(\mathrm{X}^\top \mathrm{X}\right)^{-1} \mathrm{R}^\top = \begin{pmatrix} 1 & 0 & -8 \end{pmatrix} \begin{pmatrix} 0.2 & 0 & 0 \\ 0 & 0.5 & -0.5 \\ 0 & -0.5 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ -8 \end{pmatrix} = 64.2$$

   et

$$F_{obs} = \frac{2 \times 64.2^{-1} \times 2}{0.01} = 0.0006.$$

   Comme $F_{obs} < F_{1-\alpha,q,n-p} = F_{0.95,1,2} = 18.5$, on ne rejette pas H_0.
7. Les sommes de carrés :

$$\begin{aligned} \mathrm{SC}_{RES} &= (n-p)\widehat{\sigma}_\varepsilon^2 = 2 \times 0.01 = 0.02, \\ \mathrm{SC}_{TOT} &= \sum_{i=1}^{n} y_i^2 - n \times \bar{y}^2 = 526 - 5 \times 10^2 = 26, \\ \mathrm{SC}_{REG} &= 26 - 0.02 = 25.98. \end{aligned}$$

   Le tableau de l'analyse de la variance

Source de variation	Somme des carrés	Degrés de liberté	Carrés moyens	F_c
Régression	25.98	2	12.99	$F_{obs} = \frac{12.99}{0.01} = 1299$
Résiduelle	0.02	2	0.01	
Totale	26	4	6.5	

8. Comme $F_{obs} = 1299 > F_{0.95,2,2} = 19$, on rejette H_0.
9. $R^2 = \frac{\mathrm{SC}_{REG}}{\mathrm{SC}_{TOT}} = \frac{25.98}{26} = 0.999$. Donc, 99% de la variance de la variable dépendante (c'est-à-dire de y) est expliquée par le modèle.

Exercice 14.
Soit le modèle linéaire

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i.$$

Le but de cet exercice est de montrer que le R-carré est égal au carré du coefficient de corrélation. Pour rappel, le coefficient de corrélation est

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

1. Montrer que la somme des carrés de la régression satisfait

$$\mathrm{SC}_{REG} = \frac{s_{xy}^2}{s_x^4} \sum_{i=1}^{n} (x_i - \bar{x})^2 .$$

2. Utiliser l'égalité précédente pour montrer que $R^2 = r_{xy}^2$.

Solution

1. On a :

$$\begin{aligned} \mathrm{SC}_{REG} &= \sum_{i=1}^{n} \left(\hat{y}_i - \bar{y}\right)^2 = \sum_{i=1}^{n} \left(\widehat{\beta}_1 + \widehat{\beta}_2 x_i - \bar{y}\right)^2 = \sum_{i=1}^{n} \left(\bar{y} - \widehat{\beta}_2 \bar{x} + \widehat{\beta}_2 x_i - \bar{y}\right)^2 \\ &= \widehat{\beta}_2^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{s_{xy}^2}{s_x^4} \sum_{i=1}^{n} (x_i - \bar{x})^2 . \end{aligned}$$

2. Donc,

$$R^2 = \frac{\mathrm{SC}_{REG}}{\mathrm{SC}_{TOT}} = \frac{\frac{s_{xy}^2}{s_x^4} n \times s_x^2}{n \times s_y^2} = \frac{s_{xy}^2}{s_x^2 s_y^2} = r_{xy}^2.$$

Exercice 15.
Soit le modèle linéaire : $y_i = \alpha + \beta x_i + \varepsilon_i$. On suppose que chaque y_i est remplacé par cy_i, où $c \neq 0$ est une constante. Quelles valeurs prendront les estimateurs $\widehat{\alpha}$ et $\widehat{\beta}$, le coefficient de détermination R^2 et la variance estimée de $\widehat{\beta}$?

Solution
On pose les modèles

$$y_i = \alpha_{y|x} + \beta x_i \alpha_{y|x} + \varepsilon_i.$$

et

$$z_i = \alpha_{z|x} + \beta z_i \alpha_{z|x} + \eta_i,$$

où $z_i = cy_i$ pour tout $i, =, \ldots, n$. On a $\bar{z} = c\bar{y}$,

$$ns_z^2 = \sum_{i=1}^{n}(z - \bar{z})^2 = \sum_{i=1}^{n}(cy_i - c\bar{y})^2 = \sum_{i=1}^{n} c^2(y_i - \bar{y})^2 = c^2 \sum_{i=1}^{n} y_i^2 - c^2 n\bar{y}^2 = c^2 ns_y^2,$$

$$ns_{xz} = \sum_{i=1}^{n}(x_i - \bar{x})(z_i - \bar{z}) = \sum_{i=1}^{n}(x_i - \bar{x})(cy_i - c\bar{y}) = \sum_{i=1}^{n} c(x_i - \bar{x})(y_i - \bar{y}) = c \sum_{i=1}^{n} x_i y_i - cn\bar{x}\bar{y} = cns_{xy}.$$

Donc,

$$\widehat{\beta}_{z|x} = \frac{s_{xz}}{s_x^2} = \frac{cs_{xy}}{s_x^2} = c\widehat{\beta}_{y|x} \text{ et } \widehat{\alpha}_{z|x} = \bar{z} - \widehat{\beta}_{z|x}\bar{x} = c\bar{y} - c\widehat{\beta}_{y|x}\bar{x} = c\widehat{\alpha}_{y|x}.$$

Finalement,

$$\mathrm{SC}_{RES,z|x} = ns_z^2 - \widehat{\beta}_{z|x}^2 ns_x^2 = c^2 ns_y^2 - c^2\widehat{\beta}_{y|x}^2 ns_x^2 = c^2 \mathrm{SC}_{RES,y|x},$$

$$R_{z|x}^2 = \frac{\mathrm{SC}_{reg,z|x}}{\mathrm{SC}_{tot,z|x}} = 1 - \frac{\mathrm{SC}_{RES,z|x}}{\mathrm{SC}_{tot,z|x}} = 1 - \frac{c^2\mathrm{SC}_{RES,y|x}}{c^2 ns_y^2} = 1 - \frac{\mathrm{SC}_{RES,y|x}}{ns_y^2} = R_{y|x}^2.$$

et

$$\widehat{\mathrm{var}}(\widehat{\beta}_{z|x}) = \frac{SC_{RES,z|x}}{n(n-2)s_x^2} = \frac{c^2 SC_{RES,y|x}}{n(n-2)s_x^2} = c^2 \frac{SC_{RES,y|x}}{n(n-2)s_x^2} = c^2 \widehat{\mathrm{var}}(\widehat{\beta}_{y|x}).$$

Exercice 16.
Les 22 observations sont tirées de "the High School and Beyond survey" conduit par le Département d'éducation américain en 1980 (Owings, 1985). On cherche à savoir si des tests de rendement (i.e. variable bytest) sont expliqués par les années de scolarités effectuées (i.e. variable yresd) et la distance du collège en 4ème année (i.e. variable dist). Pour cela, on dispose des résultats d'une régression effectuée à l'aide du logiciel R :

En langage R

```
Call:
lm(formula = bytest ~ yrsed + dist)

Residuals:
   Min     1Q   Median     3Q     Max
-12.1027  -2.2825  -0.8178   3.7111  10.8907

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  31.5863    14.1386   2.234 0.037695 *
yrsed     1.5768     0.9936   1.587 0.129014
dist      2.4883     0.5926   4.199 0.000486 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.754 on 19 degrees of freedom
Multiple R-squared: 0.7637,     Adjusted R-squared: 0.7388
F-statistic:  30.7 on 2 and 19 DF,  p-value: 1.117e-06
```

1. Préciser le contenu et les dimensions de chaque vecteur et matrice du modèle : $\mathrm{y} = \mathrm{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.
2. Construire le tableau d'analyse de la variance en sachant que la $\mathrm{SC}_{TOT} = 2540.754$. La regréssion est-elle significative ? Justifier en utilisant le test F et le R^2.
3. En sachant que $\mathrm{se}(\widehat{\beta}_2) = 0.994$, donner les intervalles de confiance à 90% et 95% pour le coefficient de la variable yrsed.
4. Donner la prédiction du résultat du test sachant que l'individu a effectué 15 ans d'études (yrsed=15) et que la distance du collège est de 0.1 km (dist=0.1). Puis pour cette nouvelle valeur, donner l'intervalle de confiance pour la prévision à 95% tout en sachant que :

$$(\mathrm{X}^\top\mathrm{X})^{-1} = \begin{pmatrix} 6.0384097 & -0.42232159 & 0.17305116 \\ -0.4223216 & 0.02982082 & -0.01277483 \\ 0.1730512 & -0.01277483 & 0.01060741 \end{pmatrix}$$

Solution

1. Dimensions du modèle :
$$\underset{(22,1)}{\mathrm{y}} = \underset{(22,3)}{\mathrm{X}}\ \underset{(3,1)}{\boldsymbol{\beta}} + \underset{(22,1)}{\boldsymbol{\varepsilon}}.$$
2. Tableau d'analyse de la variance :

Source	d.l.	SC	CM	F_{obs}
Régression	2	1911.683	955.842	30.7
Résidus	19	629.062	33.109	
Totale	21	2540.745		

   Le quantile est $F_{0.95,2,19} = 3.522$. Donc, le test est significatif. De plus, le R^2 est supérieur à 0.5.
3. On sait que :
$$\mathrm{IC}_{\widehat{\beta}_2} = \left[\widehat{\beta}_2 - t_{1-\frac{\alpha}{2},n-p} \times \mathrm{se}(\widehat{\beta}_2), \widehat{\beta}_2 + t_{1-\frac{\alpha}{2},n-p} \times \mathrm{se}(\widehat{\beta}_2)\right].$$

Intervalle de confiance à 90% :

$$\mathrm{IC}_{\widehat{\beta}_2} = [1.576825 - 1.729 \times 0.994, 1.576825 + 1.729 \times 0.994] = [-0.1410859, 3.2947362]\,.$$

Intervalle de confiance à 95% :

$$\mathrm{IC}_{\widehat{\beta}_2} = [1.576825 - 2.093 \times 0.994, 1.576825 + 2.093 \times 0.994] = [-0.5027513, 3.6564017]\,.$$

4. La nouvelle valeur est :

$$\widehat{y}_j = 31.5863 + 1.576825 \times 15 + 2.488318 \times 0.1 = 55.48751.$$

L'intervalle de confiance pour la prévision est :

$$\mathrm{IC}(1-\alpha) = \left[\hat{y}_j - t_{1-\alpha/2,n-p}\sqrt{\widehat{\mathrm{var}}\left(\widehat{y}_j - y_j\right)}, \hat{y}_j + t_{1-\alpha/2,n-p}\sqrt{\widehat{\mathrm{var}}\left(\widehat{y}_j - y_j,\right)}\right],$$

où

$$\widehat{\mathrm{var}}\left(\widehat{y}_j - y_j\right) = \widehat{\sigma}^2_\varepsilon\{\mathrm{x}_j(\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{x}_j^\top + 1\}.$$

En remplaçant on trouve :

$$\begin{aligned}\widehat{\mathrm{var}}\left(\widehat{y}_j - y_j\right) &= 5.754^2 \times \left[\begin{pmatrix}1 & 15 & 0.1\end{pmatrix} \times \begin{pmatrix}6.0384097 & -0.42232159 & 0.17305116\\ -0.4223216 & 0.02982082 & -0.01277483\\ 0.1730512 & -0.01277483 & 0.01060741\end{pmatrix} \times \begin{pmatrix}1\\15\\0.1\end{pmatrix} + 1\right]\\ &= 35.58236.\end{aligned}$$

Finalement,

$$\mathrm{IC}(1-\alpha) = \left[55.48751 - 2.093 \times \sqrt{35.58236}, 55.48751 + 2.093 \times \sqrt{35.58236}\right] = [43.00257, 67.97245]\,.$$

Exercice 17.
Le tableau ci-dessus présente les résultats de l'effet de la température (x_2) et de la pression (x_3) sur le rendement d'un procédé chimique :

y	1	5	4	7	10	8	9	13	14	13	18
x_2	−5	−4	−3	−2	−1	0	1	2	3	4	5
x_3	40	50	−60	70	−80	−40	−80	70	−60	50	40

Soit le modèle linéaire : $y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i$. On voudrait estimer les coefficients de ce modèle (β_1, β_2 et β_3) pour cela, on a déjà calculé les expressions suivantes :

$$\sum_{i=1}^{11} y_i^2 = 1194, \sum_{i=1}^{11} x_{2,i}x_{3,i} = 0, \sum_{i=1}^{11} x_{2,i}^2 = 110, \sum_{i=1}^{11} x_{3,i}^2 = 39600, \sum_{i=1}^{11} x_{2,i}y_i = 158 \text{ et } \sum_{i=1}^{11} x_{3,i}y_i = 140.$$

1. Préciser le contenu et les dimensions de chaque vecteur et matrices du modéle : $\mathrm{y} = \mathrm{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.
2. Calculer analytiquement, puis numériquement $\mathrm{X}^\top\mathrm{X}$, $(\mathrm{X}^\top\mathrm{X})^{-1}$ et $\mathrm{X}^\top\mathrm{y}$.
3. Donner les estimations des coefficients, les erreurs standards et les t-valeurs (pour le test général H_0 : $\beta_j = 0$)(Donc, remplir le tableau suivant).

j	$\widehat{\beta}_j$	$se(\widehat{\beta}_j)$	$t-$valeurs
1	...	0.485	...
2	...	...	...
3	...	...	...

4. Construire le tableau d'ANOVA.
5. Donner la valeur de $\widehat{\sigma}_\varepsilon^2$ et la valeur du R^2 puis interpréter le R^2.
6. Tester, au seuil de signification $\alpha = 0.05$, les hypothèses suivantes :
   (a) $H_0 : \beta_j = 0$ pour tout les $j \neq 1$ vs H_1 : au moins un des $\beta_j \neq 0$,
   (b) $H_0 : \beta_2 = -2$ vs $\beta_2 \neq -2$,
   (c) $H_0 : \beta_2 \leq 2$ vs $\beta_2 > 2$,
   (d) $H_0 : \beta_1 + \beta_2 = 10$ vs $\beta_1 + \beta_2 \neq 10$.

Solution

1. La forme matricielle du modèle est :

$$\begin{pmatrix} y_1 \\ \vdots \\ y_{11} \end{pmatrix} = \begin{pmatrix} 1 & x_{2,1} & x_{3,1} \\ \vdots & \vdots & \vdots \\ 1 & x_{2,11} & x_{3,11} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_{11} \end{pmatrix},$$

ou

$$\underset{(11,1)}{\mathrm{y}} = \underset{(11,3)}{\mathrm{X}} \underset{(3,1)}{\boldsymbol{\beta}} + \underset{(11,1)}{\boldsymbol{\varepsilon}}.$$

2. On a :

$$\mathrm{X}^\top\mathrm{X} = \begin{pmatrix} 1 & \cdots & 1 \\ x_{2,1} & \cdots & x_{2,11} \\ x_{3,1} & \cdots & x_{3,11} \end{pmatrix} \begin{pmatrix} 1 & x_{2,1} & x_{3,1} \\ \vdots & \vdots & \vdots \\ 1 & x_{2,11} & x_{3,11} \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^n x_{2,i} & \sum_{i=1}^n x_{3,i} \\ \sum_{i=1}^n x_{2,i} & \sum_{i=1}^n x_{2,i}^2 & \sum_{i=1}^n x_{2,i}x_{3,i} \\ \sum_{i=1}^n x_{3,i} & \sum_{i=1}^n x_{2,i}x_{3,i} & \sum_{i=1}^n x_{3,i}^2 \end{pmatrix},$$

$$\mathrm{X}^\top\mathrm{y} = \begin{pmatrix} 1 & \cdots & 1 \\ x_{2,1} & \cdots & x_{2,11} \\ x_{3,1} & \cdots & x_{3,11} \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_{11} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{2,i}y_i \\ \sum_{i=1}^n x_{3,i}y_i \end{pmatrix},$$

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

On calcule :

$$\mathrm{X}^\top\mathrm{X} = \begin{pmatrix} 11 & 0 & 0 \\ 0 & 110 & 0 \\ 0 & 0 & 39600 \end{pmatrix}, (\mathrm{X}^\top\mathrm{X})^{-1} = \begin{pmatrix} \frac{1}{11} & 0 & 0 \\ 0 & \frac{1}{110} & 0 \\ 0 & 0 & \frac{1}{39600} \end{pmatrix}, \mathrm{X}^\top\mathrm{y} = \begin{pmatrix} 102 \\ 158 \\ 140 \end{pmatrix}.$$

Donc,

$$\widehat{\boldsymbol{\beta}} = \begin{pmatrix} \frac{1}{11} & 0 & 0 \\ 0 & \frac{1}{110} & 0 \\ 0 & 0 & \frac{1}{39600} \end{pmatrix} \begin{pmatrix} 102 \\ 158 \\ 140 \end{pmatrix} = \begin{pmatrix} 9.273 \\ 1.437 \\ 0.004 \end{pmatrix} = \begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \widehat{\beta}_3 \end{pmatrix}.$$

3. La variance des erreurs :

$$\widehat{\sigma}^2_\varepsilon = \frac{\mathrm{SC}_{RES}}{n-p} = \frac{\mathrm{y}^\top \mathrm{y} - \boldsymbol{\beta}^\top \mathrm{X}^\top \mathrm{y}}{11-3} = \frac{1194 - 1173.259}{8} = 2.593,$$

où

$$\boldsymbol{\beta}^\top \mathrm{X}^\top \mathrm{y} = \begin{pmatrix} 9.273 & 1.436 & 0.004 \end{pmatrix} \begin{pmatrix} 102 \\ 158 \\ 140 \end{pmatrix} = 1173.259.$$

Donc,

$$\widehat{\sigma}_\varepsilon = 1.610,$$

$$\mathrm{se}(\widehat{\beta}_1) = \widehat{\sigma}_\varepsilon \sqrt{(\mathrm{X}^\top \mathrm{X})^{-1}_{11}} = \widehat{\sigma}_\varepsilon \sqrt{\frac{1}{11}} = 0.485,$$

$$\mathrm{se}(\widehat{\beta}_2) = \widehat{\sigma}_\varepsilon \sqrt{(\mathrm{X}^\top \mathrm{X})^{-1}_{22}} = 1.610 \sqrt{\frac{1}{110}} = 0.154,$$

$$\mathrm{se}(\widehat{\beta}_3) = \widehat{\sigma}_\varepsilon \sqrt{(\mathrm{X}^\top \mathrm{X})^{-1}_{33}} = 1.610 \sqrt{\frac{1}{39600}} = 0.008.$$

$\widehat{\beta}_j$	$se(\widehat{\beta}_j)$	$t-$ valeurs
9.273	0.485	19.000
1.436	0.154	9.356
0.004	0.008	0.437

Le quantile est $t_{0.95,8} = 2.306$ donc $\widehat{\beta}_1$ et $\widehat{\beta}_2$ sont significativement differents de 0. Par contre, on ne rejette pas l'hypothèse, que $\widehat{\beta}_3$ est égale a 0.

4. Le tableau d'analyse de la variance :

Source	d.l.	SC	CM	F_{obs}
Régression	2	227.440	113.720	43.862
Résidus	8	20.741	2.593	
Totale	10	248.182		

Le quantile est $F_{0.95,2,8} = 4.459$. Donc, notre modèle est significatif.

5. La variance des erreurs et le R^2 :

$$\widehat{\sigma}^2_\varepsilon = 2.593 \text{ et } R^2 = \frac{\mathrm{SC}_{REG}}{\mathrm{SC}_{TOT}} = 91.6\%.$$

Le modèle explique 91.6% de la variabilité de la variable dépendante. C'est un bon modèle.

6. Le test global :
   (a) Comme
   $$F_{obs} = 43.862 > 4.459 = F_{hyp} = F_{2,8,0.95} = F_{p-1,n-p,1-\alpha},$$
   on rejette l'hypothèse H_0, où
   $$H_0 \ : \ \beta_j = 0 \quad \text{pour tout les } j \neq 1 \ \text{ vs } \ H_1 \ : \ \text{au moins un des } \ \beta_j \neq 0.$$
   (b) Le test bilatéral : $H_0 \ : \ \beta_2 = -2 \quad \text{vs} \quad \beta_2 \neq -2.$
   $$t_{obs} = \frac{\widehat{\beta}_2 - \beta_{20}}{se(\widehat{\beta}_2)} = \frac{-1.436 - (-2)}{0.154} = 22.312.$$
   Comme $|t_{obs}| > t_{0.975,11-3} = 2.306$, on rejette H_0.

(c) Le test unilatéral : $H_0 \ : \ \beta_2 \leq 2 \quad$ vs $\quad \beta_2 > 2$. C'est un test unilatéral.

$$t_{obs} = \frac{1.436 - 2}{0.154} = -3.66.$$

Comme $|t_{obs}| > t_{0.95,11-3} = 1.860$, on rejette H_0 (noter : on met le signe $\leq$ ou $\geq$ dans l'hypothèse H_0).

(d) Test de Wald : $H_0 \ : \ \beta_1 + \beta_2 = 10 \quad$ vs $\quad \beta_1 + \beta_2 \neq 10$. ou

$$H_0 \ : \mathrm{R}\boldsymbol{\beta} = \mathrm{r}, \text{ où } \mathrm{r} = 10 \text{ et } \mathrm{R} = (1, 1, 0).$$

Donc,

$$\mathrm{R}\boldsymbol{\beta} - \mathrm{x} = 9.273 + 1.436 - 10 = 0.709,$$

$$\mathrm{R}\left(\mathrm{X}^\top \mathrm{X}\right)^{-1} \mathrm{R}^\top = \begin{pmatrix} 0.091 & 0.009 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = 0.1.$$

Donc,

$$t_{obs} = \frac{\mathrm{R}\widehat{\boldsymbol{\beta}} - \mathrm{r}}{\widehat{\sigma}\sqrt{\mathrm{R}\left(\mathrm{X}^\top \mathrm{X}\right)^{-1} \mathrm{R}^\top}} = \frac{0.709}{1.610\sqrt{0.1}} = 1.393.$$

Comme $|t_{obs}| = 1.393 < t_{0.975,11-3} = 2.306$, on ne rejette pas H_0.

Exercice 18.
On dispose de données annuelles dans 24 états des USA durant les années 1985 à 1995 (source : Stock and Watson, 2003). On cherche à savoir si la quantité de cigarettes consommées (i.e. variable packpc) dépend du revenu des états (i.e. variable income) et de la moyenne des taxes locales, fédérales et étatiques (i.e. variable tax). Pour cela, on dispose des résultats d'une régression effectuée à l'aide du logiciel R :

En langage R

```
Call~:
lm(formula = packpc ~ income + tax)

Residuals~:
Min      1Q  Median      3Q     Max
-31.121  -8.025   0.459   8.906  45.074

Coefficients~:
Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.690e+02  2.086e+01   8.104  6.7e-08 ***
income      -4.872e-08  3.739e-08  -1.303   0.2068
tax         -1.367e+00  6.370e-01  -2.146   0.0437 *
---
Signif. codes~:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error~: 16.2 on 21 degrees of freedom
Multiple R-squared~: 0.2167,     Adjusted R-squared~: 0.1421
F-statistic~: 2.904 on 2 and 21 DF,  p-value~: 0.077
```

1. Préciser le contenu et les dimensions de chaque vecteur et matrice du modèle : $\mathrm{y} = \mathrm{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.
2. Écrire l'équation de la droite de régression.
3. Donner le seuil de signification des coefficients de la droite de régression. Interpréter.
4. Déterminer la qualité de l'ajustement au moyen du coefficient de détermination.
5. Construire le tableau d'analyse de la variance. La regréssion est-elle significative ? Justifier.
6. En sachant que $\mathrm{se}(\widehat{\beta}_3) = 0.637$, donner les intervalles de confiance à 90% et 95% pour le coefficient de régression de la variable tax.
7. Donner la prédiction de la quantité de cigarettes consomées en sachant que le revenu des États income=52856936 et que la moyenne des taxes locales tax=36. Puis pour cette nouvelle valeur, donner l'intervalle de confiance pour la prévision à 95% tout en sachant que : $\widehat{\mathrm{var}}\left(\widehat{y}_j - y_j\right) = 281.546$.

Solution

1. Dimension des vecteurs et matrices :
$$\underset{(24,1)}{\mathrm{y}} = \underset{(24,3)}{\mathrm{X}}\ \underset{(3,1)}{\boldsymbol{\beta}} + \underset{(24,1)}{\boldsymbol{\varepsilon}}$$
2. Le modèle : $y_i = 169 - 4.782 \times 10^{-8} \times x_{i,2} - 1.367 \times x_{i,3}$.
3. Dans notre cas, la constante possède un niveau de signification élevé (***). La variable tax a un niveau de signification acceptable, car la p-valeur est inférieure à 0.05 (*). Par contre, la variable income semble pas être nécessaire dans le modèle et nous aider à expliquer la variable dépendante.
4. La valeur du R^2 est de 0.2167. Le modèle explique les 21.67% de la variabilité de la variable dépendante. Ce n'est pas un très bon modèle.
5. Le tableau d'analyse de la variance :

Source	d.l.	SC	CM	F_{obs}
Régression	2	1524.252	762.126	2.904
Résidus	21	5511.24	262.44	
Totale	23	7035.492		

Le quantile est $F_{0.95,2,21} = 3.467$. Donc, notre modèle n'est pas significatif.

6. On sait que : $\text{IC}_{\widehat{\beta}_3} = \left[\widehat{\beta}_3 - t_{1-\frac{\alpha}{2},n-p} \times \text{se}(\widehat{\beta}_3), \widehat{\beta}_3 + t_{1-\frac{\alpha}{3},n-p} \times \text{se}(\widehat{\beta}_3)\right]$.
   Intervalle de confiance à 90% : $\text{IC}_{\widehat{\beta}_3} = [-1.367 - 1.721 \times 0.637, -1.367 + 1.721 \times 0.637] = [-2.463, -0.271]$.
   Intervalle de confiance à 95% : $\text{IC}_{\widehat{\beta}_2} = [-1.367 - 2.080 \times 0.637, -1.367 + 2.080 \times 0.637] = [-2.692, -0.042]$.
7. La nouvelle valeur est : $\widehat{y_j} = 169 - 4.782 \times 10^{-8} \times 52856936 - 1.367 \times 36 = 117.228$.
   L'intervalle de confiance pour la prévision est :

$$\text{IC}(1-\alpha) = \left[\hat{y}_j - t_{1-\alpha/2,n-p}\sqrt{\widehat{\text{var}}\left(\widehat{y}_j - y_j\right)}, \hat{y}_j + t_{1-\alpha/2,n-p}\sqrt{\widehat{\text{var}}\left(\widehat{y}_j - y_j\right)}\right].$$

   D'où

$$\text{IC}(1-\alpha) = \left[117.228 - 2.080 \times \sqrt{281.546}, 117.228 + 2.080 \times \sqrt{281.546}\right] = [82.327, 152.129].$$

Exercice 19.
On dispose d'un modèle de régression simple $y_i = \alpha + \beta x_i + \varepsilon_i$. Montrer que

$$F_{obs} = \frac{\text{SC}_{REG}/1}{\text{SC}_{RES}/(n-2)} \text{ et } t_\beta = \frac{\widehat{\beta} - 0}{\sqrt{\widehat{\text{var}}(\widehat{\beta})}}$$

sont reliés par $F_{obs} = t_\beta^2$. Pour ce faire, élever t_β au carré, puis faites apparaître les sommes des carrés.

Solution
On sait que :

$$\widehat{\beta} = \frac{s_{xy}}{s_x^2} \text{ et } \widehat{\text{var}}(\widehat{\beta}) = \frac{\widehat{\sigma}_\varepsilon^2}{n s_x^2}.$$

On a donc :

$$t_\beta^2 = \frac{s_{xy}^2}{s_x^4} \times \frac{n s_x^2}{\widehat{\sigma}_\varepsilon^2} = \frac{s_{xy}^2 \times n}{s_x^2 \times \widehat{\sigma}_\varepsilon^2} = \frac{s_{xy}^2 \times n}{s_x^2 \times \text{SC}_{RES}/(n-2)} = \frac{n s_y^2 \frac{s_{xy}^2}{s_x^2 s_y^2}}{\text{SC}_{RES}/(n-2)} = \frac{n\ s_y^2\ r_{xy}^2}{\text{SC}_{RES}/(n-2)} = \frac{\text{SC}_{REG}/1}{\text{SC}_{RES}/(n-2)} = F_{obs}.$$

Exercice 20.
Les données que nous allons utiliser concernent les rémunérations dans le championnat de football américain. Les variables nous intéressant sont les suivantes :

y_i :	salaire (en million de dollars) pour un quarterback régulier
x_{i2} :	total des salaires de l'équipe
x_{i3} :	variable binaire valant $\begin{cases} 1 & \text{s'il s'agit de la "National Football Conference"} \\ 0 & \text{s'il s'agit de l'"American Football Conference"} \end{cases}$

Le terme "quarterback" désigne le nom d'un joueur d'une équipe de football américain. Il s'agit d'un poste offensif dans l'équipe. On ajuste le modèle suivant

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i. \tag{20.1}$$

Pour ce faire, on a utilisé

$$\sum_{i=1}^{n} y_i = 40.142, \quad \sum_{i=1}^{n} x_{i2} = 633.096, \quad \sum_{i=1}^{n} x_{i3} = 14, \quad \sum_{i=1}^{n} y_i^2 = 70.579,$$
$$\sum_{i=1}^{n} x_{i2}^2 = 14551.82, \quad \sum_{i=1}^{n} x_{i3}^2 = 14, \quad \sum_{i=1}^{n} x_{i2}x_{i3} = 316.834, \quad n = 28.$$

On a alors trouvé les valeurs des estimateurs suivants : $\widehat{\beta}_1 = -1.674$, $\widehat{\beta}_2 = 0.1363$, $\widehat{\beta}_3 = 0.05215$ et $\widehat{\sigma}_\varepsilon^2 = 0.344$.

1. Le modèle (2) s'écrit matriciellement $\mathrm{y} = \mathrm{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. Donner la matrice $\mathrm{X}^\top\mathrm{X}$ dans le cas considéré ici.
2. (a) À l'aide du modèle (2), donner une prédiction du salaire d'un quarterback jouant dans la National Football Conference et dont le total des salaires de l'équipe se monte à 25.642 dollars.
   (b) Dans nos données, on trouve un joueur ayant ces caractéristiques (joueur évoluant dans la National Football Conference et dans une équipe ayant des salaires équivalents à 25.642) dont le salaire s'élève à 2.250 dollars. Expliquer (en moins de 2 lignes) pourquoi on trouve un salaire différent de celui prédit.
3. À l'aide du modèle (2) estimer la différence de salaire d'un quarterback selon qu'il joue dans la "National Football Conference" ou dans l'"American Football Conference".
4. Construire le tableau d'analyse de la variance.

Table 20.5 – Tableau de l'analyse de la variance

Source de variation	Somme des carrés	Degrés de liberté	Carrés moyens	F_c
Régression				
Résiduelle				
Totale				

5. Tester $\begin{cases} H_0 : & \beta_2 = \beta_3 = 0 \\ H_1 : & \text{au moins un des } \beta_j \neq 0 \ (j = 2, 3) \end{cases}$, au seuil 95%.
6. Donner le R^2 et commenter (en une ligne).

Solution

1. La matrice $\mathrm{X}^\top\mathrm{X}$:
$$\mathrm{X}^\top\mathrm{X} = \begin{pmatrix} n & \sum_{i=1}^n x_{i2} & \sum_{i=1}^n x_{i3} \\ \sum_{i=1}^n x_{i2} & \sum_{i=1}^n x_{i2}^2 & \sum_{i=1}^n x_{i2}x_{i3} \\ \sum_{i=1}^n x_{i3} & \sum_{i=1}^n x_{i2}x_{i3} & \sum_{i=1}^n x_{i3}^2 \end{pmatrix} = \begin{pmatrix} 28 & 633.096 & 14 \\ 633.096 & 14551.82 & 316.834 \\ 14 & 316.834 & 14 \end{pmatrix}.$$
2. (a) $\widehat{y} = -1.674 + 0.1363 \times 25.642 + 0.05215 \times 1 = 1.873$.
   (b) Le salaire ci-dessus est une prédiction faite à partir du modèle, il n'est pas égal à la valeur observée, car les paramètres sont estimés et il manque le terme d'erreur inconnu ε_i.

3. Cette différence de salaire correspond au coefficient $\widehat{\beta}_3 = 0.05215$.
4. Le tableau d'analyse de la variance :

$$\begin{aligned} \text{SC}_{RES} &= (n-3)\widehat{\sigma}_\varepsilon^2 = 25 \times 0.344 = 8.6, \\ \text{SC}_{TOT} &= \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 \\ &= 70.579 - 28 \times \left(\frac{1}{28} \times 40.142\right)^2 = 70.579 - 28 \times 2.055 = 13.0297, \\ \text{SC}_{REG} &= \text{SC}_{TOT} - \text{SC}_{RES} = 13.0297 - 8.6 = 4.4297. \end{aligned}$$

Source de variation	Somme des carrés	Degrés de liberté	Carrés moyens	F_c
Régression	4.4297	2	2.21485	$F_{obs} = \frac{2.21485}{0.344} = 6.4385$
Résiduelle	8.6	25	0.344	
Totale	13.0297	27	0.4826	

5. On teste

$$\begin{cases} H_0 : & \beta_2 = \beta_3 = 0 \\ H_1 : & \text{au moins un des } \beta_j \neq 0 \ (j = 2, 3), \end{cases}$$

au moyen du test de Fisher. On a $F_{obs} = 6.4385, F_{0.95,2,28} = 3.340$. Donc, $F_{obs} > F_{0.95,2,28}$. On rejette H_0.

6. Le R^2 vaut :

$$R^2 = \frac{\text{SC}_{REG}}{\text{SC}_{TOT}} = \frac{4.4297}{13.0297} = 0.3399.$$

Donc, 34% de la variance de y est expliquée pas le modèle.

Exercice 21.
Nous allons ajuster et tester un modèle $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$ sur les 10 données de la Figure 20.2.

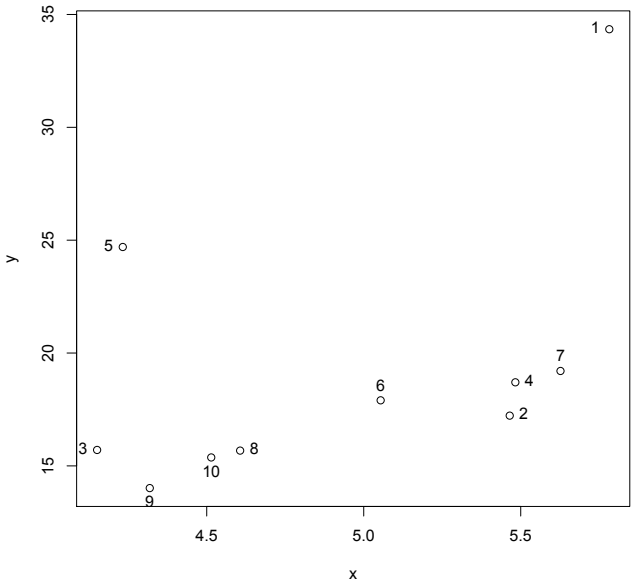


FIGURE 20.2 – Nuage de points

Dans ce but, nous avons déjà calculé

$$(\mathrm{X}^\top\mathrm{X})^{-1} = \begin{pmatrix} 6.90 & -1.38 \\ -1.38 & 0.28 \end{pmatrix}, (\mathrm{X}^\top\mathrm{X})^{-1}\,\mathrm{X}^\top\mathrm{y} = \begin{pmatrix} -3.118 \\ 4.550 \end{pmatrix}, \widehat{\sigma}^2_\varepsilon = 32.217,$$

ainsi que

i	1	2	3	4	5	6	7	8	9	10
h_i	0.31	0.18	0.27	0.19	0.23	0.10	0.24	0.13	0.20	0.15

où $h_i = [\mathrm{P_X}]_{ii}$ et $\mathrm{P_X} = \mathrm{X}(\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{x}^\top$.

1. Donner les coefficients de la régression.
2. (a) Tester $\begin{cases} H_0 : & \beta_2 = 4 \\ H_1 : & \beta_2 \neq 4 \end{cases}$ au seuil 95%.
   (b) Donner un intervalle de confiance à 95% pour β_2.
   (c) Donner un intervalle de confiance à 90% pour β_2.
   (d) On remarque que l'un des intervalles est contenu dans l'autre.
   Expliquer (en moins de 4 lignes) ce phénomène.
3. On observe sur la Figure 20.2 que certains points pourraient avoir un levier important. Déterminer par calculs quels points ont en effet un levier important.
4. Parmi les observations 5 et 8, laquelle à une plus grande influence sur la régression ?
   Justifier au moyen de la distance de Cook.
   Aide : $(x_5, y_5) = (4.23, 24.69)$ et $(x_8, y_8) = (4.61, 15.67)$.

Solution

1. $\widehat{\beta}_1 = -3.118, \widehat{\beta}_2 = 4.55$
2. (a) Il y a deux tests différents possibles.

— Test de Student :

$$t_{obs} = \frac{\widehat{\beta}_2 - 4}{\sqrt{\widehat{\sigma}^2_\varepsilon \left[(\mathrm{X}^\top \mathrm{X})^{-1}\right]_{22}}} = \frac{4.55 - 4}{\sqrt{32.217 \times 0.28}} = 0.183$$

Comme $|t_{obs}| < t_{0.975,n-p} = t_{0.975,8} = 2.306$, on ne rejette pas H_0.

— Test de Fisher :

$$F_c = \frac{\left(\widehat{\beta}_2 - 4\right)^2}{\widehat{\sigma}^2_\varepsilon \left[(\mathrm{X}^\top \mathrm{X})^{-1}\right]_{22}} = t^2_{obs} = 0.0335.$$

Comme $F_{obs} < F_{0.95,n-p} = F_{0.95,8} = 5.318$, on ne rejette pas H_0.

(b) Intervalle de confiance à 95%.

$$\begin{aligned} \mathrm{IC}(95\%)_{\beta_2} &= \left[\widehat{\beta}_2 - t_{0.975,8} \times \sqrt{\widehat{\sigma}^2_\varepsilon \left[(\mathrm{X}^\top \mathrm{X})^{-1}\right]_{22}}, \widehat{\beta}_2 + t_{0.975,8} \times \sqrt{\widehat{\sigma}^2_\varepsilon \left[(\mathrm{X}^\top \mathrm{X})^{-1}\right]_{22}}\right] \\ &= \left[4.55 - 2.306 \times \sqrt{32.217 \times 0.28}, 4.55 + 2.306 \times \sqrt{32.217 \times 0.28}\right] \\ &= [-2.376, 11.476] \end{aligned}$$

(c) Intervalle de confiance à 90%.

$$\begin{aligned} \mathrm{IC}(90\%)_{\beta_2} &= \left[\widehat{\beta}_2 - t_{0.95,8} \times \sqrt{\widehat{\sigma}^2_\varepsilon \left[(\mathrm{X}^\top \mathrm{X})^{-1}\right]_{22}}, \widehat{\beta}_2 + t_{0.95,8} \times \sqrt{\widehat{\sigma}^2_\varepsilon \left[(\mathrm{X}^\top \mathrm{X})^{-1}\right]_{22}}\right] \\ &= \left[4.55 - 1.860 \times \sqrt{32.217 \times 0.28}, 3.01 + 1.860 \times \sqrt{32.217 \times 0.28}\right] \\ &= [-1.036, 10.136] \end{aligned}$$

(d) $\mathrm{IC}(90\%)_{\beta_2} \subset \mathrm{IC}(95\%)_{\beta_2}$. La longueur de l'intervalle augmente avec le niveau de confiance.

3. L'observation i a un levier important si $h_i > \frac{p}{n} = \frac{2}{10}$. Les observations 1, 3, 5 et 7 ont ainsi un levier important.

4. La distance de Cook est donnée par

$$D_i = \frac{e_i^2}{p \times \widehat{\sigma}^2_\varepsilon} \left[\frac{h_i}{(1-h_i)^2}\right].$$

On trouve, pour les observations 5 et 8

$$\begin{aligned} e_5 &= y_5 - \widehat{y}_5 = 24.69 - (-3.118 + 4.55 \times 4.23) = 8.5615, \\ D_5 &= \frac{8.5612^2}{2 \times 32.217} \left(\frac{0.23}{(1-0.23)^2}\right) = 0.441, \\ e_8 &= y_8 - \widehat{y}_8 = 15.67 - (-3.118 + 4.55 \times 4.61) = -2.1875, \\ D_8 &= \frac{-2.1875^2}{2 \times 32.217} \left(\frac{0.13}{(1-0.13)^2}\right) = 0.018. \end{aligned}$$

Comme $D_5 > D_8$, l'observation 5 a une plus grande influence sur la régression.

Exercice 22.
Soit le modèle suivant : $y_i = \alpha + \beta x_i + \varepsilon_i$, où $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$. Montrer que le rapport $\text{SC}_{REG}/\text{SC}_{TOT}$ est égal au carré du coefficient de corrélation entre x et y, noté r_{xy}, c'est-à-dire que :

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

est égal à

$$\frac{\text{SC}_{REG}}{\text{SC}_{TOT}} = \frac{\sum_{i=1}^n (y_i^* - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})}.$$

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

$$\hat{\alpha} = \bar{y} - \widehat{\beta}x$$

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

Solution
On a :

$$\frac{\text{SC}_{REG}}{\text{SC}_{TOT}} = \frac{\sum_{i=1}^n (y_i^* - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})} = \frac{\sum_{i=1}^n (\hat{\alpha} + \widehat{\beta}x_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}.$$

Comme $\hat{\alpha} = \bar{y} - \widehat{\beta}x$,

$$\frac{\text{SC}_{REG}}{\text{SC}_{TOT}} = \frac{\sum_{i=1}^n (\bar{y} - \widehat{\beta}\bar{x} + \widehat{\beta}x_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{\widehat{\beta}^2 \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2}{\frac{1}{n}\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{s_{xy}^2 s_x^2}{s_x^4 s_y^2} = \frac{s_{xy}^2}{s_x^2 s_y^2} = r_{xy}^2.$$

Exercice 23.
On possède des données concernant les ventes d'un magasin en milliers de dollars et le nombre de ses clients. On cherche à déterminer s'il existe un lien entre le nombre de clients et le volume des ventes. On dispose des données suivantes :

Ventes hebdomadaires (y_i)	Nombre de clients (x_i)
11.20	907
11.05	926
6.84	506
9.21	741
9.42	789
10.08	889
9.45	874
6.73	510
7.24	529
6.12	420
11.00	550
7.63	679
9.43	872
9.46	924
7.64	607
6.92	452
8.95	729
9.33	794
10.23	844
11.77	1010

On considère le modèle : $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$. On a calculé :

$$\sum_{i=1}^{20} x_i = 14552, \quad \sum_{i=1}^{20} x_i^2 = 11223068, \quad \sum_{i=1}^{20} x_i y_i = 135576.3, \quad \sum_{i=1}^{20} y_i = 179.7, \quad \sum_{i=1}^{20} y_i^2 = 1668.189, \quad \widehat{\sigma}_\varepsilon^2 = 0.939.$$

1. Calculer $\widehat{\beta}_1$ et $\widehat{\beta}_2$, les estimateurs des moindres carrés de β_1 et β_2.
   Donner l'expression fonctionnelle de la droite de régression.
2. (a) Donner un intervalle de confiance à 95% pour β_1.
   (b) En se basant sur l'intervalle de confiance donné au point précédent, dire si on rejette ou non l'hypothèse $\begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{cases}$ au seuil 95%. Justifier.
   (c) En se basant sur les deux points précédents et sur la relation d'inclusion qui existe entre un intervalle de confiance à 95% et un intervalle de confiance à 90%, dire si on rejette ou non l'hypothèse $\begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{cases}$ au seuil 90%. Justifier.
3. Tester $\begin{cases} H_0 : \beta_2 = 2 \\ H_1 : \beta_2 \neq 2 \end{cases}$ au seuil 90%.
4. Donner un intervalle de confiance à 95% pour la prévision de y lorsque $x = 600$.
5. En sachant que $\sum_{i=1}^{20} h_i = 2$, calculer pour les points $(550, 11)$ et $(741, 9.21)$ les h_i ainsi que la distance de Cook. Interpréter.

Solution

1. Estimation des paramètres :

$$\begin{aligned} \widehat{\beta}_2 &= \frac{s_{xy}}{s_x^2} = \frac{\frac{1}{n}\sum_{i=1}^n x_i y_i - \bar{x} \times \bar{y}}{\frac{1}{n}\sum_{i=1}^n x_i^2 - \bar{x}^2} = \frac{241.329}{31751.64} = 0.0076, \\ \widehat{\beta}_1 &= \bar{y} - \widehat{\beta}_2 \bar{x} = 8.985 - 0.0076 \times 727.6 = 3.455. \end{aligned}$$

   La droite est : $\widehat{y} = 3.455 + 0.0076x$.

2. (a) La variance de $\widehat{\beta}_1$:

$$\widehat{\text{var}}(\widehat{\beta}_1) = \frac{\widehat{\sigma}_\varepsilon^2}{n s_x^2}(s_x^2 + \bar{x}^2) = \frac{\widehat{\sigma}_\varepsilon^2}{n} = \frac{0.939}{20} = 0.04695.$$

L'intervalle de confiance :

$$\text{IC}(95\%) = \widehat{\beta}_1 \pm t_{0.975,20} \times \sqrt{\widehat{\text{var}}(\widehat{\beta}_1)} = 3.455 \pm 2.086 \times \sqrt{0.04695} = [3.003, 3.907]$$

(b) Comme $0 \notin \text{IC}(95\%)$, on rejette $H_0 : \beta_1 = 0$ à 95%.

(c) Comme $\text{IC}(90\%) \subseteq \text{IC}(95\%)$, on a $0 \notin \text{IC}(90\%)$, on rejette $H_0 : \beta_1 = 0$ à 90%.

3. La variance de $\widehat{\beta}_2$:

$$\widehat{\text{var}}(\widehat{\beta}_2) = \frac{\widehat{\sigma}_\varepsilon^2}{n s_x^2} = \frac{0.939}{20 \times 31751.64} = 0.00000148$$

$$t_{\beta_2} = \frac{\widehat{\beta}_2 - \beta_{0,2}}{\sqrt{\widehat{\text{var}}(\widehat{\beta}_2)}} = \frac{0.0076 - 2}{\sqrt{0.00000148}} = -1637.743$$

Comme $t_{0.95,30} = 1.725$, $|t_\beta| > 1.725$ donc on rejette H_0.

4. L'erreur de prédiction :

$$\widehat{\text{var}}(\widehat{y}_j - y_j) = \frac{\widehat{\sigma}_\varepsilon^2}{n}\left(n + 1 + \frac{(x_j - \bar{x})^2}{s_x^2}\right) = \frac{0.939}{20}\left(21 + \frac{-127.6^2}{31751.64}\right) = 1.01$$

$$\widehat{y}_j = \widehat{\beta}_1 + \widehat{\beta}_2 \times 600 = 3.455 + 0.0076 \times 600 = 8.015$$

$$\text{IC}(95\%) = 8.015 \pm 2.086\sqrt{1.01} = [5.919, 10.111]$$

5. Recherche des points leviers :

$$h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{n S_x^2}$$

$$h_{550} = \frac{1}{20} + \frac{(550 - 727.6)^2}{635032.8} = 0.0997$$

$$h_{741} = \frac{1}{20} + \frac{(741 - 727.6)^2}{635032.8} = 0.0503$$

Si l'on calcule la valeur $\frac{2p}{n} = \frac{4}{20} = 0.2$, on voit bien qu'aucun des h_i est supérieur à cette valeur et l'on peut donc conclure qu'il n'y a pas de points leviers.

$$D_i = \frac{e_i^2}{p \times \widehat{\sigma}_\varepsilon^2} \times \frac{h_i}{(1 - h_i)^2}$$

$$D_{550} = \frac{(11 - (3.455 + 0.0076 \times 550))^2}{2 \times 0.939} \times \frac{0.0997}{(1 - 0.0997)^2} = 0.742$$

$$D_{741} = \frac{(9.21 - (3.455 + 0.0076 \times 741))^2}{2 \times 0.939} \times \frac{0.0503}{(1 - 0.0503)^2} = 0.000489$$

On considère en général que la distance de Cook est importante si elle est supérieur à 1. Aucune des observations n'a donc une grande influence sur la régression.

Exercice 24.
On dispose des données suivantes :

y_i	15	11	8	17	12	8	11	10	8	11
x_{i2}	4	3	−6	−5	3	−3	6	−1	0	−1
x_{i3}	−2	1	0	−1	2	0	−1	1	1	−1

Soit le modèle linéaire : $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$. On voudrait estimer les coefficients de ce modèle (β_1, β_2 et β_3) pour cela, on a déjà calculé les expressions suivantes :

$$\sum_{i=1}^{10} y_i^2 = 1313, \quad \sum_{i=1}^{10} y_i = 111, \quad \sum_{i=1}^{10} x_{i2}x_{i3} = 0,$$

$$\sum_{i=1}^{10} x_{i2}^2 = 142, \quad \sum_{i=1}^{10} x_{i3}^2 = 14, \quad \sum_{i=1}^{10} x_{i2}y_i = 17 \quad \text{et} \quad \sum_{i=1}^{10} x_{i3}y_i = -16.$$

1. Écrire le modèle sous forme matricielle : $\mathrm{y} = \mathrm{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. (en précisant le contenu et les dimensions de chaque vecteur et matrice).
2. Calculer analytiquement, puis numériquement $\mathrm{X}^\top\mathrm{X}$, $(\mathrm{X}^\top\mathrm{X})^{-1}$ et $\mathrm{X}^\top\mathrm{y}$.
3. Donner les estimations des coefficients, les erreurs standard et les t-valeurs (pour le test général H_0 : $\beta_j = 0$). Remplir le tableau suivant :

j	$\widehat{\beta}_j$	$se(\widehat{\beta}_j)$	$t-$ valeurs
1	...	0.9303	...
2	...	...	...
3	...	...	...

4. Construire le tableau d'ANOVA.
5. Donner la valeur de $\widehat{\sigma}_\varepsilon^2$, la valeur du R^2 puis interpréter le R^2.
6. Tester, au seuil de signification $\alpha = 0.05$, les hypothèses suivantes :
   (a) $H_0 : \beta_j = 0$ pour tous les $j \neq 1$ vs H_1 : au moins un des $\beta_j \neq 0$,
   (b) $H_0 : \beta_2 = 1$ vs $\beta_2 \neq 1$,
   (c) $H_0 : \beta_1 + \beta_2 = 10$ vs $\beta_1 + \beta_2 \neq 10$.

Solution

1. La forme matricielle du modèle est :

$$\begin{pmatrix} y_1 \\ \vdots \\ y_{10} \end{pmatrix} = \begin{pmatrix} 1 & x_{2,1} & x_{3,1} \\ \vdots & \vdots & \vdots \\ 1 & x_{2,10} & x_{3,10} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_{10} \end{pmatrix}$$

$$\underset{(10,1)}{\mathrm{y}} = \underset{(10,3)}{\mathrm{X}} \underset{(3,1)}{\boldsymbol{\beta}} + \underset{(10,1)}{\boldsymbol{\varepsilon}}$$

2. On a :

$$\mathrm{X}^\top\mathrm{X} = \begin{pmatrix} 1 & \cdots & 1 \\ x_{2,1} & \cdots & x_{2,10} \\ x_{3,1} & \cdots & x_{3,10} \end{pmatrix} \begin{pmatrix} 1 & x_{1,1} & x_{2,1} \\ \vdots & \vdots & \vdots \\ 1 & x_{2,10} & x_{3,10} \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^n x_{2,i} & \sum_{i=1}^n x_{3,i} \\ \sum_{i=1}^n x_{2,i} & \sum_{i=1}^n x_{2,i}^2 & \sum_{i=1}^n x_{2,i}x_{3,i} \\ \sum_{i=1}^n x_{3,i} & \sum_{i=1}^n x_{2,i}x_{3,i} & \sum_{i=1}^n x_{3,i}^2 \end{pmatrix},$$

$$\mathrm{X}^\top\mathrm{y} = \begin{pmatrix} 1 & \cdots & 1 \\ x_{2,1} & \cdots & x_{2,10} \\ x_{3,1} & \cdots & x_{3,10} \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_{10} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{2,i}y_i \\ \sum_{i=1}^n x_{3,i}y_i \end{pmatrix}$$

$$\widehat{\boldsymbol{\beta}} = (\mathrm{X}^\top\mathrm{X})^{-1}\mathrm{X}^\top\mathrm{y},$$

$$X^\top X = \begin{pmatrix} 10 & 0 & 0 \\ 0 & 142 & 0 \\ 0 & 0 & 14 \end{pmatrix},$$

et

$$(X^\top X)^{-1} = \begin{pmatrix} \frac{1}{10} & 0 & 0 \\ 0 & \frac{1}{142} & 0 \\ 0 & 0 & \frac{1}{14} \end{pmatrix}.$$

3. L'estimateur de $\boldsymbol{\beta}$:

$$\widehat{\boldsymbol{\beta}} = \begin{pmatrix} \frac{1}{10} & 0 & 0 \\ 0 & \frac{1}{142} & 0 \\ 0 & 0 & \frac{1}{14} \end{pmatrix} \begin{pmatrix} 111 \\ 17 \\ -16 \end{pmatrix} = \begin{pmatrix} 11.1 \\ 0.1197 \\ -1.143 \end{pmatrix} = \begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \widehat{\beta}_3 \end{pmatrix}.$$

La somme des carrés totale :

$$SC_{TOT} = n * s_y^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 = 1313 - 10 \times (111/10)^2 = 80.9.$$

On a :

$$\text{se}(\widehat{\beta}_1) = \widehat{\sigma}_\varepsilon \sqrt{(X^\top X)^{-1}_{11}} = \widehat{\sigma}_\varepsilon \sqrt{\frac{1}{10}} = 0.9303.$$

Donc, $\widehat{\sigma}_\varepsilon = 2.942$

$$\text{se}(\widehat{\beta}_2) = \widehat{\sigma}_\varepsilon \sqrt{(X^\top X)^{-1}_{22}} = 2.942 \sqrt{\frac{1}{142}} = 0.247,$$

$$\text{se}(\widehat{\beta}_3) = \widehat{\sigma}_\varepsilon \sqrt{(X^\top X)^{-1}_{33}} = 2.942 \sqrt{\frac{1}{14}} = 0.786,$$

ce qui donne

$\widehat{\beta}_j$	$se(\widehat{\beta}_j)$	$t-$ valeurs
11.1	0.9303	11.932
0.1197	0.247	0.485
−1.143	0.786	−1.454

Le quantile est $t_{0.95,7} = 2.36$ donc $\widehat{\beta}_1$ est significativement differente de 0. Par contre, on ne rejette pas l'hypothèse, que $\widehat{\beta}_2$ est égale à 0 et que $\widehat{\beta}_3$ est égal à 0.

4. Le tableau d'analyse de la variance :

Source	d.l.	SC	CM	F
Régression	2	20.312	10.156	1.173
Résidus	7	60.588	8.656	
Total	9	80.9		

Le quantile est $F_{0.95,2,7} = 4.74$. Donc, le test n'est pas significatif.

5.

$$\widehat{\sigma}_\varepsilon^2 = 8.654 \quad \widehat{\sigma}_\varepsilon = 2.942 \text{ et } R^2 = 0.251\%.$$

Le modèle explique les 25.1% de la variabilité de la variable dépendante.

6. (a) $F_{obs} = 1.173 < 4.74 = F_{hyp} = F_{2,7,0.95} = F_{p-1,n-p,1-\alpha}$, On ne rejette pas l'hypothèse : $H_0 : \beta_j = 0$ pour tout les $j \neq 1$ vs $H_1 :$ au moins un des $\beta_j \neq 0$.

   (b) On teste :

$$H_0 : \beta_2 = 1 \quad \text{vs} \quad \beta_2 \neq 1,$$

$$t_{obs} = \frac{\widehat{\beta}_2 - \beta_{20}}{se(\widehat{\beta}_2)} = \frac{0.1197 - 1}{0.247} = -4.0001,$$

$$|t_{obs}| > t_{0.975,10-3} = 2.36.$$

Donc, on rejette H_0.

(c) On teste :

$$H_0 \ : \ \beta_1 + \beta_2 = 10 \quad \text{vs} \quad \beta_1 + \beta_2 \neq 10,$$

H_0 : $\mathrm{R}\boldsymbol{\beta} = \mathrm{r}$, où $\mathrm{r} = 10$ et $\mathrm{R} = \begin{pmatrix} 1 & 1 & 0 \end{pmatrix}$. Donc,

$$\mathrm{R}\widehat{\boldsymbol{\beta}} - r = 11.1 + 0.1197 - 10 = 1.2197,$$

$$\mathrm{R}\left(\mathrm{X}^\top \mathrm{X}\right)^{-1} \mathrm{R}^\top = \begin{pmatrix} \frac{1}{10} & \frac{1}{142} & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = 0.107.$$

Donc,

$$t_{obs} = \frac{\mathrm{R}\widehat{\boldsymbol{\beta}} - r}{\widehat{\sigma}\sqrt{\mathrm{R}\left(\mathrm{X}^\top \mathrm{X}\right)^{-1} \mathrm{R}^\top}} = \frac{-8.7803}{2.942\sqrt{0.107}} = -9.124,$$

$$|t_{obs}| = 9.124 > t_{0.975,10-3} = 2.365.$$

Donc, on rejette H_0.

Exercice 25.
On dispose des données suivantes :

	x_{i2}	x_{i3}	y_i
	-1	-1	7
	-1	0	8
	0	0	10
	1	0	12
	1	1	13
Somme	0	0	50
Somme des carrés	4	2	526

Considérons le modèle : $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$. On a :

$$\begin{pmatrix} 5 & 0 & 0 \\ 0 & 4 & 2 \\ 0 & 2 & 2 \end{pmatrix}^{-1} = \begin{pmatrix} 0.2 & 0 & 0 \\ 0 & 0.5 & -0.5 \\ 0 & -0.5 & 1 \end{pmatrix}$$

et que $\widehat{\sigma}^2_\varepsilon = 0.01$.

1. Le modèle considéré ici s'écrit matriciellement $\mathrm{y} = \mathrm{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. Donner théoriquement $\mathrm{X}^\top\mathrm{X}$ et $\mathrm{X}^\top\mathrm{y}$.
2. Soit $\widehat{\boldsymbol{\beta}}$, l'estimateur de $\boldsymbol{\beta}$. Donner les valeurs de X et $\widehat{\boldsymbol{\beta}}$.
3. Donner une prévision pour y si $x_{i2} = -1$ et x_{i3}=1.
4. À l'aide du modèle, estimer l'augmentation de y liée à une augmentation de une unité de x_{i2}, x_{i3} restant constant. On s'intéresse à un cas général, on ne choisira pas de valeur particulière pour x_{i3}.
5. Tester $\left\{ \begin{array}{lcl} H_0 & : & \beta_3 = 0 \\ H_1 & : & \beta_3 \neq 0 \end{array} \right.$ au seuil 95%.
6. Tester $\left\{ \begin{array}{lcl} H_0 & : & \beta_1 - 8\beta_3 = 0 \\ H_1 & : & \beta_1 - 8\beta_3 \neq 0 \end{array} \right.$ au seuil 95%.
7. Construire le tableau d'analyse de la variance.
8. Tester $\left\{ \begin{array}{lcl} H_0 & : & \beta_2 = \beta_3 = 0 \\ H_1 & : & \text{au moins un des } \beta_j \neq 0 \ (j = 2, 3) \end{array} \right.$ au seuil 95%.
9. Donner le R^2 et commenter (en une ligne).

Solution

1. On a :
$$\mathrm{X}^\top\mathrm{X} = \begin{pmatrix} n & \sum_{i=1}^n x_{i2} & \sum_{i=1}^n x_{i3} \\ \sum_{i=1}^n x_{i2} & \sum_{i=1}^n x_{i2}^2 & \sum_{i=1}^n x_{i2}x_{i3} \\ \sum_{i=1}^n x_{i3} & \sum_{i=1}^n x_{i2}x_{i3} & \sum_{i=1}^n x_{i3}^2 \end{pmatrix} \text{ et } \mathrm{X}^\top\mathrm{y} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n y_i x_{i2} \\ \sum_{i=1}^n y_i x_{i3} \end{pmatrix}$$
2. La matrice X :
$$\mathrm{X} = \begin{pmatrix} 1 & -1 & -1 \\ 1 & -1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$
et l'estimation de $\boldsymbol{\beta}$:
$$\widehat{\boldsymbol{\beta}} = \left(\mathrm{X}^\top\mathrm{X}\right)^{-1}\mathrm{X}^\top\mathrm{y} = \begin{pmatrix} 0.2 & 0 & 0 \\ 0 & 0.5 & -0.5 \\ 0 & -0.5 & 1 \end{pmatrix}\begin{pmatrix} 50 \\ 10 \\ 6 \end{pmatrix} = \begin{pmatrix} 10 \\ 2 \\ 1 \end{pmatrix}$$
3. $\widehat{y} = 10 + 2 \times (-1) + 1 \times 1 = 9$
4. $\widehat{\beta}_2 = 2$.

5. Test sur β_3 :
$$t_{\beta_3} = \frac{\beta_3 - \beta_{3,0}}{\sqrt{\left[(\mathrm{X}^\top \mathrm{X})^{-1}\, \widehat{\sigma}_\varepsilon^2\right]_{33}}} = \frac{1}{\sqrt{1 \times 0.01}} = \frac{1}{0.1} = 10.$$
Comme $|t_{\beta_3}| > t_{0.975,2(n-p)} = 4.303$, on rejette H_0.
6. Test de Wald.
H_0 : $\mathrm{R}\boldsymbol{\beta} = \mathrm{r}$, où $\underbrace{\mathrm{R} = \begin{pmatrix} 1 & 0 & -8 \end{pmatrix}}_{1\times 3 = q \times p}$ et $\underbrace{\mathrm{r} = 0}_{1\times 1 = q \times 1}$. On a :
$$\mathrm{R}\widehat{\boldsymbol{\beta}} - r = 1 \times 10 + 0 \times 2 - 8 \times 1 = 2,$$
$$\mathrm{R}\left(\mathrm{X}^\top \mathrm{X}\right)^{-1} \mathrm{R}^\top = \begin{pmatrix} 1 & 0 & -8 \end{pmatrix} \begin{pmatrix} 0.2 & 0 & 0 \\ 0 & 0.5 & -0.5 \\ 0 & -0.5 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ -8 \end{pmatrix} = 64.2$$
et
$$F_{obs} = \frac{(\mathrm{R}\widehat{\boldsymbol{\beta}} - r)^2}{\widehat{\sigma}_\varepsilon^2 \mathrm{R}\left(\mathrm{X}^\top \mathrm{X}\right)^{-1} \mathrm{R}^\top} = \frac{2 \times 64.2^{-1} \times 2}{0.01} = 6.23.$$
Comme $F_{obs} < F_{1-\alpha,q,n-p} = F_{0.95,1,2} = 18.51$, on ne rejette pas H_0.
7. Les sommes de carrés :
$$\begin{aligned} \mathrm{SC}_{RES} &= (n-p)\widehat{\sigma}_\varepsilon^2 = 2 \times 0.01 = 0.02 \\ \mathrm{SC}_{TOT} &= \sum_{i=1}^{n} y_i^2 - n \times \bar{y}^2 = 526 - 5 \times 10^2 = 26 \\ \mathrm{SC}_{REG} &= 26 - 0.02 = 25.98 \end{aligned}$$
Tableau de l'analyse de la variance :

Source de variation	Somme des carrés	Degrés de liberté	Carrés moyens	F_c
Régression	25.98	2	12.99	$F_{obs} = \frac{12.99}{0.01} = 1299$
Résiduelle	0.02	2	0.01	
Totale	26	4		

8. Comme $F_{obs} = 1299 > F_{0.95,2,2} = 19$, on rejette H_0.
9. $R^2 = \frac{\mathrm{SC}_{REG}}{\mathrm{SC}_{TOT}} = \frac{25.98}{26} = 0.999$, ce qui signifie que 99% de la variance de la variable dépendante (c'est-à-dire de y) est expliquée par le modèle.

Exercice 26.
Soient le modèle $y = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ et le vecteur des valeurs ajustées de la régression linéaire par les moindres carrés y^*. Montrer que $\text{var}(y^*) = \sigma_\varepsilon^2 P_X$.

Solution
On peut calculer :

$$\text{var}(y^*) = \text{var}(X^\top \widehat{\boldsymbol{\beta}}) = X^\top \text{var}(\widehat{\boldsymbol{\beta}}) X = X^\top \sigma_\varepsilon^2 (X^\top X)^{-1} X = \sigma_\varepsilon^2 P_X$$

ou

$$y^* = P_X y \text{ donc } \text{var}(y^*) = P_X \text{var}(y) P_X^\top = \sigma_\varepsilon^2 P_X.$$

Exercice 27.
Démontrer que lorsqu'on se trouve dans le cas d'un modèle linéaire général : $SC_{RES} = y^\top y - \widehat{\boldsymbol{\beta}}^\top X^\top y$ Pour commencer, vous pouvez utiliser le fait que : $SC_{RES} = e^\top e$.

Solution
On a :

$$\begin{aligned} SC_{RES} &= e^\top e = (y - X\widehat{\boldsymbol{\beta}})^\top (y - X\widehat{\boldsymbol{\beta}}) = (y^\top - (X\widehat{\boldsymbol{\beta}})^\top)(y - X\widehat{\boldsymbol{\beta}}) = (y^\top - \widehat{\boldsymbol{\beta}}^\top X^\top)(y - X\widehat{\boldsymbol{\beta}}) \\ &= y^\top y - y^\top X\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^\top X^\top y + \widehat{\boldsymbol{\beta}}^\top X^\top X\widehat{\boldsymbol{\beta}} = y^\top y - 2\widehat{\boldsymbol{\beta}}^\top X^\top y + \widehat{\boldsymbol{\beta}}^\top X^\top X\widehat{\boldsymbol{\beta}}. \end{aligned}$$

Or,

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

Donc,

$$SC_{RES} = y^\top y - \widehat{\boldsymbol{\beta}}^\top X^\top y.$$

Exercice 28.
On s'intéresse aux résultats de "saut en hauteur" (Y), donnés dans le Tableau 20.6 et étant donné que :

$$\sum_{i=1}^{22} x_i = 1100, \ , \ \sum_{i=1}^{22} y_i = 46.41, \ , \ \sum_{i=1}^{22} x_i^2 = 74608, \ , \ \sum_{i=1}^{22} y_i^2 = 98.6217, \text{ et } \sum_{i=1}^{22} y_i x_i = 2435.04.$$

TABLE 20.6 – Médailles d'or aux jeux olympiques pour le "saut en hauteur" et le "saut en longueur"

Ville	saut en hauteur (Y)	saut en longueur (Z)	année (X)
Paris	1.90	7.19	0
St Louis	1.80	7.34	4
London	1.91	7.48	8
Stockholm	1.93	7.60	12
Antwerp	1.94	7.15	20
Paris	1.98	7.45	24
Amsterdam	1.94	7.74	28
Los Angeles	1.97	7.64	32
Berlin	2.03	8.06	36
London	1.98	7.82	48
Helsinki	2.04	7.57	52
Melbourne	2.11	7.83	56
Rome	2.16	8.12	60
Tokyo	2.18	8.07	64
Mexico City	2.24	8.90	68
Munich	2.23	8.24	72
Montreal	2.25	8.34	76
Moscow	2.36	8.54	80
Los Angeles	2.35	8.54	84
Seoul	2.38	8.72	88
Barcelona	2.34	8.67	92
Atlanta	2.39	8.50	96

1. Estimez les paramètres du modèle $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ par la méthode des moindres carrés.
2. Testez l'hypothèse $H_0 \ : \ \beta_1 = 0$ versus $H_1 \ : \ \beta_1 > 0$ au seuil 0.05 et interprétez le résultat.
3. Combien vaut en moyenne l'augmentation de la performance en "saut en hauteur" d'une olympiade à l'autre.
4. Sachant que la somme des carrés des résidus du modèle de la question 1 vaut $\text{CM}_{res} = 0.00243$ calculez le R^2.
5. Calculez une prévision pour $x = 100$. Construisez alors l'intervalle de confiance à 95% pour la prévision obtenue.
6. À Sydney en 2000, la médaille d'or du saut en hauteur a été gagnée par l'athlète russe Sergey Kliugin avec $2.35\,m$. Sa performance fait-elle partie de votre intervalle de confiance ? Interprétez.
7. Pourriez-vous donner une estimation de la médaille d'or pour les jeux olympiques d'Athènes ? (L'année 2004 selon l'échelle du Tableau 20.6 prend la valeur 104).

Solution

1. $\widehat{y_i} = 1.81747 + 0.00584 \times x_i$.
2. $\dfrac{\text{E}(\beta_1)}{\sqrt{\text{var}(\beta_1)}} = \dfrac{0.00584}{0.00035} = 16.59 \ >> \ 2.086 = t_{0.025,20}$. On rejette H_0.
3. Les jeux olympiques se déroulent tous les quatre ans. $\widehat{\beta_1} \approx 0.00584$ est l'augmentation par année. Alors, à chaque olympiade, on attend en moyenne une augmentation de $4 \times 0.00584 = 0.02336$m sur le saut en hauteur.

4. $R^2 = \dfrac{\text{SC}_{REG}}{\text{SC}_{TOT}} = 1 - \dfrac{\text{SC}_{RES}}{\text{SC}_{TOT}}$.

   $\text{SC}_{TOT} = \displaystyle\sum_{i=1}^{22} y_i^2 - n \times \bar{y}^2 = 98.6217 - 22 \times 4.45018 = 0.7177.$

   $R^2 = 1 - \dfrac{0.00243}{0.7177} = 0.932.$

5. $\hat{y}(x = 100) = 1.81747 + 0.00584 \times 100 = 2.4014,$

   $\text{var}(\hat{y}(x = 100)) = \sigma_\varepsilon^2 \left\{ \dfrac{1}{22} + \dfrac{(100 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \right\}$, où $\bar{x} = 50$ et σ_ε^2 estimé par la $\sigma_\varepsilon^2 = \text{CM}_{res}$.

   Alors, $\text{var}(\hat{y}(x = 100)) = 0.00243 \left(0.04545 + \dfrac{50}{19608} \right) = 0.0001166.$

   Comme $t_{0.025,20} = 2.086$, on a : $\text{IP}(95\%) = [2.4014 \pm 2.086 \times \sqrt{0.0001166}] = [2.378875\ ,\ 2.423925]$.

6. $2.35 \notin \text{IP}$.

7. $\hat{y}(x = 104) = 1.81747 + 0.00584 \times 104 = 2.42483.$

Exercice 29.
En reprenant les données du Tableau 20.6, on utilise la régression linéaire simple afin de modéliser l'évolution des performances pour le "saut en longueur" (Z). On a effectué l'analyse de régression et on a obtenu le Tableau 20.7.

TABLE 20.7 – Résultats de l'analyse de régression

	Coef	Écart Type Coef	t
Constante	···	0.0975	73.84
année	0.0155	···	···
S = 0.2346	R^2=···		

Source	d.l.	SC	MC	F
Régression	···	···	···	85.12
Résidu	···	···	···	
Total	···	5.7862		

1. Complétez les valeurs manquantes du Tableau 20.7 (indiquées par "···").
2. Donnez le modèle $\hat{z}_i = \widehat{\beta}_0 + \widehat{\beta}_1 \, x_i$ estimé par la méthode des moindres carrés.
3. Donnez un estimateur sans biais de la variance des erreurs du modèle.
4. Donnez un intervalle de confiance à 95% pour β_1.
5. Selon le modèle du Tableau 20.6, donnez une estimation de la performance de la médaille d'or pour les jeux olympiques d'Athènes. (L'année 2004 selon l'échelle de Tableau 20.6 prend la valeur 104.)
6. L'observation numéro 15 a été détectée comme observation aberrante. Pourquoi ? Commentez. Comment détecte-t-on les valeurs aberrantes ? Pour $i = 15$ on a : $h_{ii} = 0.06197$.

Solution

1. La tableau complété :

	Coef	SE Coef	t	p
Constante	7.20483	0.09757	73.84	0.000
année	0.01545	0.00167	9.23	0.000
S = 0.2346	$R^2 = 81.0\%$			

Analyse de la variance

Source	d.l.	SC	MS	F
Régression	1	4.6853	4.6853	85.12
Résidu	20	1.1009	0.0550	
Total	21	5.7862		

2. $\hat{z}_i = 7.20483 + 0.015458 \, x_i$.
3. $\mathrm{CM}_{res} = 0.0550$.
4. $\mathrm{E}(\beta_1) = 0.015458$ et
$$\widehat{\mathrm{var}}(\widehat{\beta}_1) = \frac{\widehat{\sigma}_\varepsilon^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{0.0550}{19608} = 0.000002805.$$
De plus, $t_{0.025,20} = 2.086$.
L'intervalle de confiance à 95% est $[0.015458 \pm 2.086 \times \sqrt{0.000002805}] = [0.01196283, 0.01895317]$.
5. $z(x_i = 104) = 7.20483 + 0.015458 \times 104 = 8.8064$.
6. Le plus grand résidu standardisé est celui de l'observation 15 et vaut 2.83. Ce n'est pas à proprement parler un *outlier*.

Exercice 30.
À partir d'une étude économique, on a mesuré le salaire moyen (Y en 1000 CHF) ainsi que l'expérience professionelle (X en termes d'années de travail) pour un ensemble de 150 employés. Étant donné que :

$$\sum_{i=1}^{150} x_i = 1410\ ,\ \sum_{i=1}^{150} y_i = 8695\ ,\ \sum_{i=1}^{150} x_i^2 = 16458\ ,\ \sum_{i=1}^{150} y_i^2 = 555625\ \text{ et }\ \sum_{i=1}^{150} y_i x_i = 93958.$$

1. Estimez les paramètres du modèle $y_i = \beta_0 + \beta_1\, x_i + \varepsilon_i$ par la méthode des moindres carrés.
2. Testez l'hypothèse $H_0\ :\ \beta_1 = 0$ versus $H_1\ :\ \beta_1 \neq 0$ au seuil 0.05 et interprétez le résultat.
3. Donnez un intervalle de confiance à 95% pour β_0.
4. Estimez sans biais la variance des erreurs du modèle.

Solution

1. Les paramètres sont : $\widehat{\beta}_0 = 22.1006$ et $\widehat{\beta}_1 = 3.8155$.
2. On a :

$$\begin{aligned} r_{xy} &= \frac{\sum_{i=1}^n x_i y_i - \frac{1}{n}\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{\left(\sum_{i=1}^n x_i^2 - \frac{1}{n}(\sum_{i=1}^n x_i)^2\right)\left(\sum_{i=1}^n y_i^2 - \frac{1}{n}(\sum_{i=1}^n y_i)^2\right)}} \\ &= \frac{93958 - \frac{1410\times 8695}{150}}{\sqrt{\left(16458 - \frac{1410^2}{150}\right)\left(5555625 - \frac{8695^2}{150}\right)}} = 0.9507305. \end{aligned}$$

$$r_{xy}^2 = 0.9038885.$$

$$s_y^2 = \frac{1}{n}\left\{\sum_{i=1}^n y_i^2 - \frac{1}{n}\left(\sum_{i=1}^n y_i\right)^2\right\} = \frac{1}{150}\left(555625 - \frac{8695^2}{150}\right) = 344.0322.$$

$$s_x^2 = \frac{1}{n}\left\{\sum_{i=1}^n x_i^2 - \frac{1}{n}\left(\sum_{i=1}^n x_i\right)^2\right\} = \frac{1}{150}\left(16458 - \frac{1410^2}{150}\right) = 21.36.$$

$$\text{SC}_{TOT} = n s_y^2 = 150 \times 344.0322 = 51604.83.$$

$$\text{SC}_{RES} = (1 - r_{xy}^2)\text{SC}_{TOT} = (1 - 0.9038885) \times 51604.83 = 4959.819.$$

$$\widehat{\text{var}}(\widehat{\beta}_1) = \frac{\widehat{\sigma}_\varepsilon^2}{n s_x^2} = \frac{\text{SC}_{RES}}{(n-2) n s_x^2} = \frac{4959.819}{148 \times 150 \times 21.36} = 0.01045952.$$

$$\sqrt{\widehat{\text{var}}}(\widehat{\beta}_1) = 0.1022718$$

Puisque

$$t_c = \frac{3.8155}{0.1022718} = 37.31 >> 1.96 \approx t_{0.025,149},$$

l'hypothèse H_0 est rejetée.

3. $\text{IC}_{\beta_0}(95\%) = [21.209,\ 24.2005]$.
4. Un estimateur sans biais de la variance des erreurs du modèle est calculé ainsi :

$$\text{CM}_{res} = \frac{\text{SC}_{RES}}{n-2} = \frac{4959.819}{148} = 33.51229.$$

Exercice 31.
Soient les données suivantes :

x_i	3	8	9	13	3	7	11	21
y_i	29	42	49	57	33	28	44	56

1. Calculez les statistiques suivantes :

$$\bar{x}\ ,\ \bar{y}\ ,\ S_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \text{ ainsi que } S_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2.$$

2. Estimez le coefficient de corrélation ρ_{xy}.
3. Testez au seuil de signification $\alpha = 5\%$ l'hypothèse $H_0 : \rho_{xy} = 0$ vs $H_1 : \rho_{xy} \neq 0$, étant donné que les variables X and Y, desquelles proviennent les x_i and y_i, sont conjointement normalement distribuées.

Solution

1. $\bar{x} = 9.375,\ \bar{y} = 42.25,\ S_x^2 = 34.268$ et $S_y^2 = 131.3571$.
2. On a : $r_{xy} = 0.833$.
3. On utilise la statistique $\frac{\sqrt{n-2}\times r_{xy}}{\sqrt{1-r_{xy}^2}} \sim t_{\alpha/2,n-2}$.
   On trouve $t_c = 3.69 > 2.447 = t_{0.025,6}$, et alors on rejette H_0 au seuil 5%.

Exercice 32.
Le Tableau 20.8 donne les valeurs de : $Y = CO_1$, X_1 = taux bas de goudron (mg), X_2 = taux bas de nicotine (mg), X_3 = Poids (gr), enregistrées pour de cigarettes *light* (c'est-à-dire des cigarettes de taux bas de goudron et de nicotine).

TABLE 20.8 – Taux de goudron (mg), taux de nicotine (mg), poids (gr), et monoxide de carbone (mg) pour 25 firmes de cigarettes, (source Mendenhall et al., 2003).

Firme	Taux de goudron (mg)	Taux de nicotine (mg)	Poids (gr)	CO_1 (mg)
Alpine	14.1	0.86	0.9853	13.6
Benson & Hedges	16.0	1.06	1.0938	16.6
Bull Durham	29.8	2.03	1.1650	23.5
Camel Lights	8.0	0.67	0.9280	10.2
Carlton	4.1	0.40	0.9462	5.4
Chesterfield	15.0	1.04	0.8885	15.0
Golden Lights	8.8	0.76	1.0267	9.0
Kent	12.4	0.95	0.9225	12.3
Kool	16.6	1.12	0.9372	16.3
L & M	14.9	1.02	0.8858	15.4
Lark Lights	13.7	1.01	0.9643	13.0
Marlboro	15.1	0.90	0.9316	14.4
Merit	7.8	0.57	0.9705	10.0
Multi Filter	11.4	0.78	1.1240	10.2
Newport Lights	9.0	0.74	0.8517	9.5
Now	1.0	0.13	0.7851	1.5
Old Gold	17.0	1.26	0.9186	18.5
Pall Mall Light	12.8	1.08	1.0395	12.6
Raleigh	15.8	0.96	0.9573	17.5
Salem Ultra	4.5	0.42	0.9106	4.9
Tareyton	14.5	1.01	1.0070	15.9
VRAI	7.3	0.61	0.9806	8.5
Viceroy Rich Light	8.6	0.69	0.9693	10.6
Virginia Slims	15.2	1.02	0.9496	13.9
Winston Lights	12.0	0.82	1.1184	14.9

À partir du Tableau 20.8, on a fait l'analyse de régression linéaire et on a obtenu l'équation :

$$\widehat{y}_i = 3.202 + 0.9626\, x_{1i} - 2.632\, x_{2i} - 0.130\, x_{3i},$$

ainsi que le tableau des coefficients :

	$\widehat{\beta}_j$	$s(\widehat{\beta})_j)$
Constante	3.202	3.462
X_1	0.962	0.242
X_2	−2.632	3.901
X_3	−0.130	3.885
$s = 1.446$	$R^2 = 91.90\%$	

1. Quels sont les paramètres significatifs du modèle dessus ?
2. Sur la base du modèle obtenu, combien vaut en moyenne le monoxide du carbone émis d'une cigarette pesant 1gr, de taux de nicotine 0.8mg et de taux de goudron 10mg ?
3. Un expert nous a expliqué que le monoxide du carbone ne dépend que du taux de goudron. On a refait l'analyse et on a obtenu le modèle suivant : $\widehat{y}_i = 2.74 + 0.801\, x_{1i}$, ainsi que le tableau des coefficients :

	$\widehat{\beta}_j$	$s(\widehat{\beta})_j)$
Constante	2.7433	0.6752
X_1	0.801	0.0503
$s = 1.397$	$R^2 = 91.89\%$	

Quels sont les paramètres significatifs dans ce modèle ?

4. Selon vous, l'expert a-t-il raison ? (en termes statistiques la question peut se poser comme "le modèle simple est-il mieux que le modèle à trois variables explicatives ?").
5. Combien vaut, cette fois-ci, en moyenne le monoxide du carbone émis d'une cigarette de taux de goudron 10 ? Comparez à 3. et commentez.
6. Le même expert nous a suggéré de faire attention aux firmes "Bull Durham" (observation 3) et "Now" (observation 16). Selon lui, les deux firmes peuvent tromper les estimateurs du modèle final, c'est-à-dire le modèle $\widehat{y_i} = 2.74 + 0.801\, x_{1i}$.
Sachant que : $h_{3,3} = 0.4413$ et $h_{16,16} = 0.2032$, contrôlez si les observations 3 et 16 sont des valeurs aberrantes ou des points leviers (on vous rappelle que les $h_{i,i}$ correspondent aux éléments diagonaux de la matrice H).

Solution

1. Seul le taux de goudron est significatif au seuil 5%.

	Coef	SE Coef	t_c	P-value
Constante	3.202	3.462	0.93	0.365
X_1	0.9626	0.2422	3.97	0.001
X_2	-2.632	3.901	-0.67	0.507
X_3	-0.130	3.885	-0.03	0.974

2. À partir du modèle obtenu on a : $y_k^* = 3.20 + 0.963 \times 10 - 2.63 \times 0.8 - 0.130 \times 1 = 10.5960$.
3. Tous les les paramètres sont significatifs dans ce modèle.

	Coef	SE Coef	t_c	P-value
Constant	2.7433	0.6752	4.06	0.000
X_1	0.8009	0.0503	15.92	0.000

4. Ce modèle est plus parcimonieux et toutes les variables sont significatives. Le R^2 est presque le même que celui du premier modèle.
5. Cette fois on a obtenu : $y_k^* = 2.7433 + 0.8009 \times 10 = 10.7523$.
6. Une observation i est un point levier si $h_{ii} > 2p/n = 4/25 = 0.16$.
$st.res_3 = -2.98128$ et $h_{3,3} = 0.441329$. L'unité 3 est presque un outlier et est un point levier.
$st.res_{16} = -1.63973$ et $h_{16,16} = 0.203283$. L'unité 16 n'est pas un outlier mais est un levier.

Chapitre 21

Projets en langage R

21.1 Cigarettes

On dispose de données annuelles dans 48 états aux USA durant les années 1985 à 1995 (source : Stock and Watson, 2003; Croissant and Graves, 2022). On cherche à savoir si la quantité de cigarettes consommées dépend des variables qui se trouvent dans le Tableau 21.1. La variable packpc est donc la variable dépendante.

Table 21.1 – Variables à utiliser dans la régression

cpi	Indices des prix à la consommation (U.S.).
packpc	nombre de paquets par habitant.
income	revenu des états (total, valeur nominale).
tax	moyenne des taxes locales, fédérales et étatiques.
avgprs	prix moyen pendant l'exercice, y compris les taxes de ventes.
taxs	taxes d'accise moyennes pour l'année financière y compris les taxes de ventes.

Les données sont accessibles grâce au code suivant :

En langage R

```
install.packages("Ecdat")
library(Ecdat)
summary(Cigarette)
```

1. Ouvrir le logiciel R et charger les données.
2. (a) Effectuer le graphique sous-forme de nuages de points.
   (b) Commenter le graphique.
3. Effectuer un ajustement linéaire.
4. Donner les coefficients de la régression et leur seuil de signification.
   Interpréter.
5. Déterminer la qualité de l'ajustement au moyen du coefficient de détermination.
6. Donner
   (a) les résidus,
   (b) les valeurs ajustées,
   (c) les intervalles de confiance à 90% et à 95% pour le coefficient de la variable avgprs.
7. (a) Donner la prédiction du nombre de paquets par habitant sachant que l'indice des prix à la consommation est égale à 1.5 (cpi=1.5), que le revenu des états est de 50 000 000 (income=50 000 000), que la moyenne des taxes locales, fédérales et étatiques s'élève à 39 (tax=39), que le prix moyen pendant l'exercice est de 150 (avgprs=150) et que les taxes d'accise moyennes pour l'année financière sont de 46 (taxs=46).
   (b) Donner, pour cette nouvelle valeur, l'intervalle de confiance et l'intervalle de confiance pour la prévision (à 95%).

8. (a) Enlever une à une les variables les moins significatives jusqu'à ce que toutes les variables restantes aient une p-valeur inférieure à 0.1% (Attention, la constante reste dans le modèle).
9. Choisir la variable explicative qui vous semble le mieux expliquer la variable dépendante. Justifier votre choix.
10. (a) Effectuer une régression linéaire simple (une variable dépendante, une variable explicative et une constante) de la variable packpc en la variable avgprs.
    (b) Donner les intervalles de confiance pour la prévision pour la variable explicative avgprs lorsqu'elle prend les valeurs 100 et 200.
    Comparer la longueur de ces deux intervalles, commenter et expliquer.
    (c) Créer un graphique contenant :
    — Les données (en abscisse, la variable explicative et en ordonnée la variable dépendante),
    — la droite de régression,
    — la courbe des limites inférieures et celle des limites supérieures des intervalles de confiance pour la prévision pour chaque avgprs observé.

21.2 Données macroéconomiques des USA

On dispose de données concernant des séries temporelles macroéconomiques des USA recueillies et publiées par différentes agences gouvernementales (source : Stock and Watson, 2003; Croissant and Graves, 2022). On cherche à savoir si l'index des prix à la consommation américain est lié directement ou indirectement par les autres variables qui se trouvent dans le tableau ci-dessous. Ces données sont accessibles avec le code R suivant :

En langage R

```
library(Ecdat)
summary(Macrodat)
```

On considérera les variables se trouvant dans le Tableau 21.2. La variable PUNEW sera donc la variable dépendante.

Table 21.2 – Variables à utiliser dans la régression

LHUR	Taux de chômage.
PUNEW	Indice des prix à la consommation.
FYFF	Taux d'intérêt des fonds fédéraux.
FYGM3	Taux d'intérêt des bons du Trésor (trimestriel).
FYGT1	Taux d'intérêt des bons du Trésor (annuel).
EXRUK	Taux de change dollar-livre.
GDP_JP	GDP du Japon.

1. Ouvrir le logiciel R et charger les données.
2. (a) Effectuer le graphique sous-forme de nuages de points.
   (b) Commenter le graphique.
3. Effectuer un ajustement linéaire.
4. Donner les coefficients de la régression et leur seuil de signification. Interpréter.
5. Déterminer la qualité de l'ajustement au moyen du coefficient de détermination.
6. Donner
   (a) les résidus,
   (b) les valeurs ajustées,
   (c) les intervalles de confiance à 90% et à 95% pour le coefficient de la variable GDP_JP.
7. (a) Donner la prédiction de l'indice des prix à la consommation sachant que le taux de chômage est égal à 5.4 (LHUR=5.4), que le taux d'intérêt des fonds fédéraux est de 5.8 (FYFF=5.8), que celui des bons du Trésor trimestriel est de 5.7 (FYGM3=5.7), que celui des bons du Trésor annuel s'élève à 6.5 (FYGT1=6.5), que le taux de change dollar-livre est de 160 (EXRUK=160) et que le GDP du Japon s'élève à 492390 (GDP_JP=492390).
   (b) Donner, pour cette nouvelle valeur, l'intervalle de confiance et l'intervalle de confiance pour la prévision (à 95%).
8. (a) Enlever une à une les variables les moins significatives jusqu'à ce que toutes les variables restantes aient un seuil de signification inférieur à 0.1% (Attention, la constante reste dans le modèle).
   (b) Recommencer les point 4 et 5 avec ce nouveau modèle.
9. Choisir la variable explicative qui vous semble le mieux expliquer la variable dépendante. Justifier votre choix.
10. (a) Effectuer une régression linéaire simple (une variable dépendante, une variable explicative et une constante) de la variable PUNEW en la variable GDP_JP.
    (b) Donner les intervalles de confiance pour la prévision de la variable GDP_JP prenant les valeurs 10150 et 490000. Comparer la longueur de ces deux intervalles, commenter et expliquer.
    (c) Créer un graphique contenant :
    — Les données (en abscisse, la variable explicative et en ordonnée la variable dépendante)
    — la droite de régression
    — la courbe des limites inférieures et celle des limites supérieures des intervalles de confiance pour la prévision pour chaque GDP_JP observé.

21.3 Accidents de la Route

On dispose de données concernant les accidents de la route dans 50 états aux USA, plus le district de Colombie et de variables qui pourraient expliquer ceux-ci durant les années 1983 à 1997 (source : Cohen and Einav, 2003; Stock and Watson, 2003; Kleiber and Zeileis, 2008). Ces données sont accessibles avec le code R suivant :

En langage R

```
install.packages("AER")
library(AER)
data("USSeatBelts")
```

On considérera les variables se trouvant dans le Tableau 21.3. La variable fatality est la variable dépendante.

Table 21.3 – Variables à utiliser dans la régression

sb_useage	Taux d'utilisation de la ceinture de sécurité.
fatality	Nombre de décès.
speed65	Si la limite de vitesse de 65 miles par heure est respectée = 1 sinon = 0.
speed70	Si la limite de vitesse de 70 miles par heure est respectée = 0 sinon = 1.
ba08	Si le taux d'alcoolémie maximal est inférieur ou égal à 0.8 % = 0 sinon = 1.
drinkage21	Si l'âge pour la consommation d'alcool (21 ans) est respecté = 0 sinon = 1.
income	Revenu par habitant.
age	Âge moyen du conducteur.
primary	Si l'application 1ère des lois sur la ceinture de sécurité est respectée = 0 sinon = 1.
secondary	Si l'application 2ème des lois sur la ceinture de sécurité est respectée = 0 sinon = 1.

1. Ouvrir le logiciel R et charger les données.
2. (a) Effectuer le graphique sous-forme de nuages de points.
   (b) Commenter le graphique.
3. Effectuer un ajustement linéaire.
4. Donner les coefficients de la régression et leur seuil de signification.
   Interpréter.
5. Déterminer la qualité de l'ajustement au moyen du coefficient de détermination.
6. Donner
   (a) les résidus,
   (b) les valeurs ajustées,
   (c) les intervalles de confiance à 90% et à 95% pour le coefficient de la variable primary.
7. (a) Donner la prédiction du nombre d'accidents de la route sachant que le taux d'utilisation de la ceinture de sécurité s'élève à 0.5 (sb_useage=0.5), que la limite de vitesse de 65 miles par heure n'est pas respectée (speed65=0), mais celle de 70 miles par heure l'est (speed70=0), que le taux maximal d'alcoolémie est supérieur à 0.8% (ba08=1), que le conducteur est âgé de 35 ans (age=35), que l'âge pour la consommation d'alcool selon la loi est respecté (drinkage21=0), que le revenu est de 21000 (income=21000) et que l'application des lois primaires et secondaires sur la ceinture de sécurité n'est pas respectée (primary=1 & secondary=1).
   (b) Donner, pour cette nouvelle valeur, l'intervalle de confiance et l'intervalle de confiance pour la prévision (à 95%).
8. (a) Enlever une à une les variables les moins significatives jusqu'à ce que toutes les variables restantes aient une p-valeur inférieure à 0.1% (Attention, la constante reste dans le modèle).
9. Choisir la variable explicative qui vous semble le mieux expliquer la variable dépendante. Justifier votre choix.
10. (a) Effectuer une régression linéaire simple (une variable dépendante, une variable explicative et une constante) de la variable fatality en la variable income.

(b) Donner les intervalles de confiance pour la prévision de la variable explicative income prenant les valeurs 15 000 et 22 900.
Comparer la longueur de ces deux intervalles, commenter et expliquer.

(c) Créer un graphique contenant :
— Les données (en abscisse, la variable explicative et en ordonnée la variable dépendante),
— la droite de régression,
— la courbe des limites inférieures et celle des limites supérieures des intervalles de confiance pour la prévision pour chaque income observé.

21.4 Résultats aux examens en Californie

On dispose de données concernant la moyenne des résultats aux examens en Californie et de variables qui pourraient expliquer ceux-ci durant les années 1998 à 1999 (source : Stock and Watson, 2003; Croissant and Graves, 2022). Ces données sont accessibles avec le code R suivant :

En langage R

```
library(Ecdat)
summary(Caschool)
```

On considérera les variables se trouvant dans le Tableau 21.4. La variable testscr est la variable dépendante. Utiliser le logiciel R.

Table 21.4 – Variables à utiliser dans la régression

testscr	moyenne des notes,
meal_pct	pourcentage d'élèves ayant des prix préférentiels au repas,
comp_stu	nombre d'ordinateurs par élèves,
str	ratio du nombre de professeurs par rapport au nombre d'élèves,
avginc	revenu moyen par district,
el_pct	pourcentage du nombre d'élèves apprenant l'anglais.

1. Ouvrir le logiciel et charger les données.
2. (a) Effectuer le graphique sous-forme de nuages de points
   (b) Commenter le graphique
3. Effectuer un ajustement linéaire.
4. Donner les coefficients de la régression et leur seuil de signification. Interpréter
5. Donner
   (a) les résidus,
   (b) les valeurs ajustées,
   (c) les intervalles de confiance à 90% et à 95% pour le coefficient de la variable meal_pct.
6. (a) Donner la prédiction de la moyenne des notes sachant que 14.4% des élèves ont des prix préférentiels pour les repas (meal_pct=14.4), qu'il y a 1 ordinateur pour 5 élèves (comp_stu=0.2), qu'il y a 1 professeur pour 17.655 élèves (str=17.655), que le revenu moyen par district est de 14.177 (avginc=14.177) et que tous les élèves savent l'anglais (el_pct=0).
   (b) Donner, pour cette nouvelle valeur, l'intervalle de confiance et l'intervalle de prédiction (à 95%).
7. (a) Enlever une à une les variables les moins significatives jusqu'à ce que toutes les variables restantes aient un seuil de signification inférieur à 0.1% (Attention, la constante reste dans le modèle).
   (b) Recommencer les point 4 et 5 avec ce nouveau modèle.
8. Choisir la variable explicative qui vous semble le mieux expliquer la variable dépendante. Justifier votre choix.
9. (a) Effectuer une régression linéare simple (une variable dépendante, une variable explicative et une constante) de la variable testscr en la variable meal_pct.
   (b) Donner les intervalles de prédiction pour les valeurs de la variable explicative meal_pct 5 et 44. Comparer la longueur de ces deux intervalles, commenter et expliquer.
   (c) Créer un graphique contenant :
   — Les données (en abscisse, la variable explicative et en ordonnée la variable dépendante
   — la droite de régression
   — la courbe des limites inférieures et celle des limites supérieures des intervalles de prédiction pour chaque meal_pct observé.

21.5 Croissance

On dispose de données concernant le taux de croissance dans 65 pays de 1960 à 1995 ainsi que des variables qui sont potentiellement liées à la croissance. Ces données ont été récoltées par Ross Levine de l'Université de Brown et ont été utilisées dans l'article de Beck et al. (2000). Ces données se trouvent ci-dessous. On considérera les variables se trouvant dans le Tableau 21.5. La variable growth est la variable dépendante.

TABLE 21.5 – Variables à utiliser dans la régression

growth	Croissance annuelle moyenne du PIB de 1960 à 1995.
rgdp60	Valeur du PIB par habitant en 1960 convertie en dollars américain (cours de 1960).
tradehare	Part moyenne des échanges de 1960 à 1995.
yearsshcool	Années de scolarité des adultes en 1960.
rev_coups	Moyenne annuelle de révolutions, d'insurrections et de coups d'état de 1960 à 1995.
assasinations	Moyenne annuelle d'assassinats politiques de 1960 à 1995 (par million d'habitants).

1. Ouvrir le logiciel R et charger les données.
2. (a) Effectuer le graphique sous-forme de nuages de points.
   (b) Commenter le graphique.
3. Effectuer un ajustement linéaire.
4. Donner les coefficients de la régression et leur seuil de signification.
   Interpréter.
5. Déterminer la qualité de l'ajustement au moyen du coefficient de détermination.
6. Donner
   (a) les résidus,
   (b) les valeurs ajustées,
   (c) les intervalles de confiance à 90% et à 95% pour le coefficient de la variable yearsshcool.
7. (a) Donner la prédiction du taux de croissance annuel sachant que la valeur du PIB par habitant en 1960 s'élève à 3000 dollars (rgdp60=3000), que la part moyenne des échanges est de 0.33 (tradehare=0.33), que les années de scolarité sont égal à 5 ans (yearsshcool=5), que le nombre annuel moyen de révolutions, d'insurrections et de coups d'état est de 0.1 (rev_coups=0.1) et que le nombre annuel moyen d'assassinats politiques s'élève à 0.16 (assasinations=0.16).
   (b) Donner, pour cette nouvelle valeur, l'intervalle de confiance et l'intervalle de confiance pour la prévision (à 95%).
8. (a) Enlever une à une les variables les moins significatives jusqu'à ce que toutes les variables restantes aient une p-valeur inférieure à 0.1% (Attention, la constante reste dans le modèle).
9. Choisir la variable explicative qui vous semble le mieux expliquer la variable dépendante. Justifier votre choix.

10. (a) Effectuer une régression linéaire simple (une variable dépendante, une variable explicative et une constante) de la variable growth en la variable yearsshcool.
    (b) Donner les intervalles de confiance pour la prévision de la variable explicative yearsshcool prenant les valeurs 4 et 9.
    Comparer la longueur de ces deux intervalles, commenter et expliquer.
    (c) Créer un graphique contenant :
    — Les données (en abscisse, la variable explicative et en ordonnée la variable dépendante),
    — la droite de régression,
    — la courbe des limites inférieures et celle des limites supérieures des intervalles de confiance pour la prévision pour chaque yearsschool observé.

En langage R

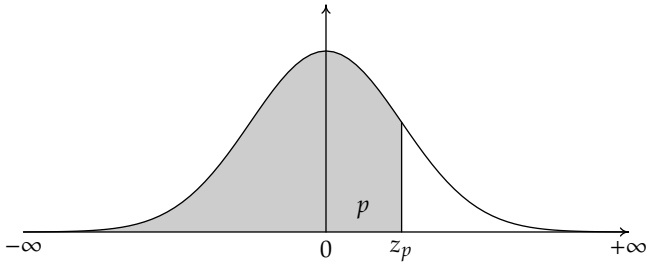
```
Growth=read.table(file=stdin(),header = TRUE,sep=",")
growth,rgdp60,tradeshare,yearsschool,rev_coups,assasinations
1.915168,765.9998,0.140502,1.45,0.1333333,0.8666667
0.6176451,4462.001,0.156623,4.99,0.9333333,1.933333
4.304759,2954,0.1577032,6.71,0,0.2
2.930097,1784,0.1604051,2.89,0.1,0.1
1.712265,9895.004,0.160815,8.66,0,0.4333333
0.7082631,951.9998,0.2214584,0.79,0.3064815,0.175
2.880327,3123.001,0.2994059,3.8,0.0666667,1.433333
2.227014,1684,0.3130729,2.97,0.1,0.7666667
0.060206,2019,0.3246127,3.02,0.2666667,0.5666667
-0.6579341,923.9999,0.3247456,0.7,0.3740741,0.2
1.975147,7782.002,0.3294792,9.03,0,0.0666667
2.932981,4564,0.3300217,4.56,0.0333333,1.2
3.22405,2093,0.337879,4.37,0.1666667,0.1666667
2.431281,5823,0.3397062,4.65,0,0.3
-2.811944,488.9999,0.3523176,0.54,0.1481481,0.0555556
1.025309,3968,0.3588567,5.07,0,0.1666667
1.973942,2836,0.3932509,2.41,0,0.1666667
2.698163,638,0.4176039,0.63,0.2666667,0.2666667
-2.751478,531.9999,0.4258372,0.2,0.1333333,0
0.3550578,1148,0.4355793,4.68,0.9703704,0.2
2.453688,6569.999,0.4421429,7.64,0.0666667,0.2333333
2.385988,7257.997,0.4443746,8.07,0,0.0666667
1.962219,6822.999,0.4505467,7.67,0.2666667,0.3333333
1.124132,7960.001,0.4557693,9.55,0,0
1.158741,1133,0.4683466,3.77,0.7333333,1.033333
2.798494,5291.002,0.4732549,7.51,0,0
-0.8835508,6337.998,0.4846008,2.53,0.2666667,0.1
7.156855,904.0001,0.4889496,3.23,0.3333333,0.1
0.9292306,1660,0.5095903,1.42,0.5407407,2.466667
0.5977848,1039,0.5176646,1.69,0.2666667,0.0666667
-0.6075561,1427,0.5277733,1.7,0.6333333,1.733333
1.446995,2885,0.5346847,4.99,0.1666667,0.4666667
4.876695,943,0.5433371,3.45,0.4,0.0333333
1.888134,7592.002,0.5442435,7.67,0,0
-0.4378241,1047,0.5508126,1.62,0.0333333,0.0666667
1.120786,5627.002,0.5527223,4.31,0.0333333,0
2.388093,1461,0.5606425,2.95,0.5333334,0
2.17937,6759.999,0.5607497,10.07,0,0
1.421865,9408.997,0.567131,6.87,0,0
2.889185,5143.001,0.5752748,3.65,0,0
0.8381555,988.9999,0.5828143,1.54,0.2,0.2333333
2.381932,1177,0.6030353,3.35,0.0666667,0.0333333
1.613697,2096,0.6050313,3.92,0,0
3.64731,1869,0.6136169,1.94,0.2333333,0
0.4627753,366.9999,0.6260797,0.27,0.1666667,0.0333333
3.012389,4963.999,0.6394842,5.63,0,0
2.810969,3476.999,0.6469135,6.99,0.0666667,0.2
0.3920211,2191,0.6512036,4.06,0.1,0.3666667
3.182494,5610,0.6815552,5.56,0,0
-0.3398342,877.9999,0.6994329,0.53,0.4740741,0.0333333
2.498769,1195,0.7048199,2.38,0.3,0.2
-0.9631622,894,0.7292323,0.69,0.4370371,0.1
2.704598,1259,0.7429716,3.43,0.2,0.2
6.624734,1256,0.7454978,3.32,0.0333333,0.0666667
2.027189,1575,0.7798471,4.26,0.1333333,0.0666667
1.120371,1235,0.7983758,1.13,0.0333333,0
1.962509,659.0001,0.81471,1.2,0.0333333,0.1444445
3.254494,3310.999,0.8229564,6.45,0,0.0666667
0.4177902,1773,0.8302383,2.46,0,0.1333333
2.200577,6076.997,0.8342042,5.27,0,0
5.384184,2037,0.9793554,4.29,0.1,0.1666667
4.114544,1420,1.105364,2.34,0.0333333,0.0333333
2.651335,5495.002,1.115917,7.46,0,0
3.024178,2861.999,1.127937,2.44,0,0
6.652838,1374,1.992616,5.64,0,0

summary(Growth)
```

Chapitre 22

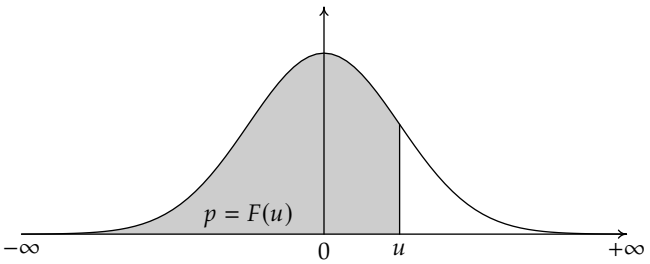
Tables statistiques

TABLE 22.1 – Table des quantiles d'une variable normale centrée réduite



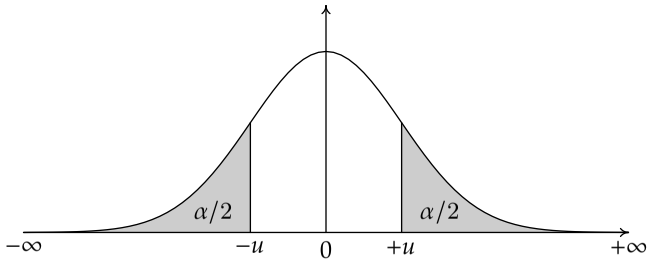
Ordre du quantile (p)	Quantile (z_p)	Ordre du quantile (p)	Quantile (z_p)
0.500	0.0000	0.975	1.9600
0.550	0.1257	0.976	1.9774
0.600	0.2533	0.977	1.9954
0.650	0.3853	0.978	2.0141
0.700	0.5244	0.979	2.0335
0.750	0.6745	0.990	2.3263
0.800	0.8416	0.991	2.3656
0.850	1.0364	0.992	2.4089
0.900	1.2816	0.993	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 22.2 – 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 22.3 – Quantiles de la loi normale centrée réduite
(*u : valeur ayant la probabilité α d'être dépassé en valeur absolue*)



α	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	∞	2.5758	2.3263	2.1701	2.0537	1.9600	1.8808	1.8119	1.7507	1.6954
0.1	1.6449	1.5982	1.5548	1.5141	1.4758	1.4395	1.4051	1.3722	1.3408	1.3106
0.2	1.2816	1.2536	1.2265	1.2004	1.1750	1.1503	1.1264	1.1031	1.0803	1.0581
0.3	1.0364	1.0152	0.9945	0.9741	0.9542	0.9346	0.9154	0.8965	0.8779	0.8596
0.4	0.8416	0.8239	0.8064	0.7892	0.7722	0.7554	0.7388	0.7225	0.7063	0.6903
0.5	0.6745	0.6588	0.6433	0.6280	0.6128	0.5978	0.5828	0.5681	0.5534	0.5388
0.6	0.5244	0.5101	0.4958	0.4817	0.4677	0.4538	0.4399	0.4261	0.4125	0.3989
0.7	0.3853	0.3719	0.3585	0.3451	0.3319	0.3186	0.3055	0.2924	0.2793	0.2663
0.8	0.2533	0.2404	0.2275	0.2147	0.2019	0.1891	0.1764	0.1637	0.1510	0.1383
0.9	0.1257	0.1130	0.1004	0.0878	0.0753	0.0627	0.0502	0.0376	0.0251	0.0125

TABLE 22.4 – Table des quantiles d'une variable χ^2 à n degrés de liberté

	ordre du quantile					
	0.01	0.025	0.05	0.95	0.975	0.99
n=1	0.000157	0.000982	0.003932	3.841	5.024	6.635
2	0.02010	0.05064	0.103	5.991	7.378	9.210
3	0.115	0.216	0.352	7.815	9.348	11.34
4	0.297	0.484	0.711	9.488	11.14	13.28
5	0.554	0.831	1.145	11.07	12.83	15.09
6	0.872	1.237	1.635	12.59	14.45	16.81
7	1.239	1.690	2.167	14.07	16.01	18.48
8	1.646	2.180	2.733	15.51	17.53	20.09
9	2.088	2.700	3.325	16.92	19.02	21.67
10	2.558	3.247	3.940	18.31	20.48	23.21
11	3.053	3.816	4.575	19.68	21.92	24.72
12	3.571	4.404	5.226	21.03	23.34	26.22
13	4.107	5.009	5.892	22.36	24.74	27.69
14	4.660	5.629	6.571	23.68	26.12	29.14
15	5.229	6.262	7.261	25.00	27.49	30.58
16	5.812	6.908	7.962	26.30	28.85	32.00
17	6.408	7.564	8.672	27.59	30.19	33.41
18	7.015	8.231	9.390	28.87	31.53	34.81
19	7.633	8.907	10.12	30.14	32.85	36.19
20	8.260	9.591	10.85	31.41	34.17	37.57
21	8.897	10.28	11.59	32.67	35.48	38.93
22	9.542	10.98	12.34	33.92	36.78	40.29
23	10.20	11.69	13.09	35.17	38.08	41.64
24	10.86	12.40	13.85	36.42	39.36	42.98
25	11.52	13.12	14.61	37.65	40.65	44.31
26	12.20	13.84	15.38	38.89	41.92	45.64
27	12.88	14.57	16.15	40.11	43.19	46.96
28	13.56	15.31	16.93	41.34	44.46	48.28
29	14.26	16.05	17.71	42.56	45.72	49.59
30	14.95	16.79	18.49	43.77	46.98	50.89
31	15.66	17.54	19.28	44.99	48.23	52.19
32	16.36	18.29	20.07	46.19	49.48	53.49
33	17.07	19.05	20.87	47.40	50.73	54.78
34	17.79	19.81	21.66	48.60	51.97	56.06
35	18.51	20.57	22.47	49.80	53.20	57.34
36	19.23	21.34	23.27	51.00	54.44	58.62
37	19.96	22.11	24.07	52.19	55.67	59.89
38	20.69	22.88	24.88	53.38	56.90	61.16
39	21.43	23.65	25.70	54.57	58.12	62.43
40	22.16	24.43	26.51	55.76	59.34	63.69
42	23.65	26.00	28.14	58.12	61.78	66.21
44	25.15	27.57	29.79	60.48	64.20	68.71
46	26.66	29.16	31.44	62.83	66.62	71.20
48	28.18	30.75	33.10	65.17	69.02	73.68
50	29.71	32.36	34.76	67.50	71.42	76.15
60	37.48	40.48	43.19	79.08	83.30	88.38
70	45.44	48.76	51.74	90.53	95.02	100.43
80	53.54	57.15	60.39	101.88	106.63	112.33
90	61.75	65.65	69.13	113.15	118.14	124.12
100	70.06	74.22	77.93	124.34	129.56	135.81
110	78.46	82.87	86.79	135.48	140.92	147.41
120	86.92	91.57	95.70	146.57	152.21	158.95

TABLE 22.5 – Table des quantiles d'une variable de Student à n degrés de liberté

	ordre du quantile			
	0.95	0.975	0.99	0.995
n=1	6.314	12.71	31.82	63.66
2	2.920	4.303	6.965	9.925
3	2.353	3.182	4.541	5.841
4	2.132	2.776	3.747	4.604
5	2.015	2.571	3.365	4.032
6	1.943	2.447	3.143	3.707
7	1.895	2.365	2.998	3.499
8	1.860	2.306	2.896	3.355
9	1.833	2.262	2.821	3.250
10	1.812	2.228	2.764	3.169
11	1.796	2.201	2.718	3.106
12	1.782	2.179	2.681	3.055
13	1.771	2.160	2.650	3.012
14	1.761	2.145	2.624	2.977
15	1.753	2.131	2.602	2.947
16	1.746	2.120	2.583	2.921
17	1.740	2.110	2.567	2.898
18	1.734	2.101	2.552	2.878
19	1.729	2.093	2.539	2.861
20	1.725	2.086	2.528	2.845
21	1.721	2.080	2.518	2.831
22	1.717	2.074	2.508	2.819
23	1.714	2.069	2.500	2.807
24	1.711	2.064	2.492	2.797
25	1.708	2.060	2.485	2.787
26	1.706	2.056	2.479	2.779
27	1.703	2.052	2.473	2.771
28	1.701	2.048	2.467	2.763
29	1.699	2.045	2.462	2.756
30	1.697	2.042	2.457	2.750
31	1.696	2.040	2.453	2.744
32	1.694	2.037	2.449	2.738
33	1.692	2.035	2.445	2.733
34	1.691	2.032	2.441	2.728
35	1.690	2.030	2.438	2.724
36	1.688	2.028	2.434	2.719
37	1.687	2.026	2.431	2.715
38	1.686	2.024	2.429	2.712
39	1.685	2.023	2.426	2.708
40	1.684	2.021	2.423	2.704
50	1.676	2.009	2.403	2.678
60	1.671	2.000	2.390	2.660
70	1.667	1.994	2.381	2.648
80	1.664	1.990	2.374	2.639
90	1.662	1.987	2.368	2.632
100	1.660	1.984	2.364	2.626
120	1.658	1.980	2.358	2.617
∞	1.645	1.960	2.327	2.576

TABLE 22.6 – Table des quantiles d'ordre 0.95 d'une variable de Fisher à n_1 et n_2 degrés de liberté

	n_1=1	2	3	4	5	6	7	8	9	10	12	14	16	20	30	∞
n_2=1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.4	246.5	248.0	250.1	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.42	19.43	19.45	19.46	19.50
3	10.13	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.786	8.745	8.715	8.692	8.660	8.617	8.526
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964	5.912	5.873	5.844	5.803	5.746	5.628
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735	4.678	4.636	4.604	4.558	4.496	4.365
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060	4.000	3.956	3.922	3.874	3.808	3.669
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637	3.575	3.529	3.494	3.445	3.376	3.230
8	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438	3.388	3.347	3.284	3.237	3.202	3.150	3.079	2.928
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137	3.073	3.025	2.989	2.936	2.864	2.707
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978	2.913	2.865	2.828	2.774	2.700	2.538
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854	2.788	2.739	2.701	2.646	2.570	2.404
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753	2.687	2.637	2.599	2.544	2.466	2.296
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671	2.604	2.554	2.515	2.459	2.380	2.206
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602	2.534	2.484	2.445	2.388	2.308	2.131
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544	2.475	2.424	2.385	2.328	2.247	2.066
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494	2.425	2.373	2.333	2.276	2.194	2.010
17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450	2.381	2.329	2.289	2.230	2.148	1.960
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412	2.342	2.290	2.250	2.191	2.107	1.917
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378	2.308	2.256	2.215	2.155	2.071	1.878
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348	2.278	2.225	2.184	2.124	2.039	1.843
21	4.325	3.467	3.072	2.840	2.685	2.573	2.488	2.420	2.366	2.321	2.250	2.197	2.156	2.096	2.010	1.812
22	4.301	3.443	3.049	2.817	2.661	2.549	2.464	2.397	2.342	2.297	2.226	2.173	2.131	2.071	1.984	1.783
23	4.279	3.422	3.028	2.796	2.640	2.528	2.442	2.375	2.320	2.275	2.204	2.150	2.109	2.048	1.961	1.757
24	4.260	3.403	3.009	2.776	2.621	2.508	2.423	2.355	2.300	2.255	2.183	2.130	2.088	2.027	1.939	1.733
25	4.242	3.385	2.991	2.759	2.603	2.490	2.405	2.337	2.282	2.236	2.165	2.111	2.069	2.007	1.919	1.711
26	4.225	3.369	2.975	2.743	2.587	2.474	2.388	2.321	2.265	2.220	2.148	2.094	2.052	1.990	1.901	1.691
27	4.210	3.354	2.960	2.728	2.572	2.459	2.373	2.305	2.250	2.204	2.132	2.078	2.036	1.974	1.884	1.672
28	4.196	3.340	2.947	2.714	2.558	2.445	2.359	2.291	2.236	2.190	2.118	2.064	2.021	1.959	1.869	1.654
29	4.183	3.328	2.934	2.701	2.545	2.432	2.346	2.278	2.223	2.177	2.104	2.050	2.007	1.945	1.854	1.638
30	4.171	3.316	2.922	2.690	2.534	2.421	2.334	2.266	2.211	2.165	2.092	2.037	1.995	1.932	1.841	1.622
32	4.149	3.295	2.901	2.668	2.512	2.399	2.313	2.244	2.189	2.142	2.070	2.015	1.972	1.908	1.817	1.594
34	4.130	3.276	2.883	2.650	2.494	2.380	2.294	2.225	2.170	2.123	2.050	1.995	1.952	1.888	1.795	1.569
36	4.113	3.259	2.866	2.634	2.477	2.364	2.277	2.209	2.153	2.106	2.033	1.977	1.934	1.870	1.776	1.547
38	4.098	3.245	2.852	2.619	2.463	2.349	2.262	2.194	2.138	2.091	2.017	1.962	1.918	1.853	1.760	1.527
40	4.085	3.232	2.839	2.606	2.449	2.336	2.249	2.180	2.124	2.077	2.003	1.948	1.904	1.839	1.744	1.509
50	4.034	3.183	2.790	2.557	2.400	2.286	2.199	2.130	2.073	2.026	1.952	1.895	1.850	1.784	1.687	1.438
60	4.001	3.150	2.758	2.525	2.368	2.254	2.167	2.097	2.040	1.993	1.917	1.860	1.815	1.748	1.649	1.389
120	3.920	3.072	2.680	2.447	2.290	2.175	2.087	2.016	1.959	1.910	1.834	1.775	1.728	1.659	1.554	1.254
∞	3.841	2.996	2.605	2.372	2.214	2.099	2.010	1.938	1.880	1.831	1.752	1.692	1.644	1.571	1.459	1.000

TABLE 22.7 – Table des quantiles d'ordre 0.99 d'une variable de Fisher à n_1 et n_2 degrés de liberté

	n_1=1	2	3	4	5	6	7	8	9	10	12	14	16	20	30	∞
n_2=1	4052	5000	5403	5625	5764	5859	5928	5981	6022	6056	6106	6143	6170	6209	6261	6366
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.44	99.45	99.47	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.92	26.83	26.69	26.51	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.25	14.15	14.02	13.84	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.888	9.770	9.680	9.553	9.379	9.020
6	13.75	10.93	9.780	9.148	8.746	8.466	8.260	8.102	7.976	7.874	7.718	7.605	7.519	7.396	7.229	6.880
7	12.25	9.547	8.451	7.847	7.460	7.191	6.993	6.840	6.719	6.620	6.469	6.359	6.275	6.155	5.992	5.650
8	11.26	8.649	7.591	7.006	6.632	6.371	6.178	6.029	5.911	5.814	5.667	5.559	5.477	5.359	5.198	4.859
9	10.56	8.022	6.992	6.422	6.057	5.802	5.613	5.467	5.351	5.257	5.111	5.005	4.924	4.808	4.649	4.311
10	10.04	7.559	6.552	5.994	5.636	5.386	5.200	5.057	4.942	4.849	4.706	4.601	4.520	4.405	4.247	3.909
11	9.646	7.206	6.217	5.668	5.316	5.069	4.886	4.744	4.632	4.539	4.397	4.293	4.213	4.099	3.941	3.602
12	9.330	6.927	5.953	5.412	5.064	4.821	4.640	4.499	4.388	4.296	4.155	4.052	3.972	3.858	3.701	3.361
13	9.074	6.701	5.739	5.205	4.862	4.620	4.441	4.302	4.191	4.100	3.960	3.857	3.778	3.665	3.507	3.165
14	8.862	6.515	5.564	5.035	4.695	4.456	4.278	4.140	4.030	3.939	3.800	3.698	3.619	3.505	3.348	3.004
15	8.683	6.359	5.417	4.893	4.556	4.318	4.142	4.004	3.895	3.805	3.666	3.564	3.485	3.372	3.214	2.868
16	8.531	6.226	5.292	4.773	4.437	4.202	4.026	3.890	3.780	3.691	3.553	3.451	3.372	3.259	3.101	2.753
17	8.400	6.112	5.185	4.669	4.336	4.102	3.927	3.791	3.682	3.593	3.455	3.353	3.275	3.162	3.003	2.653
18	8.285	6.013	5.092	4.579	4.248	4.015	3.841	3.705	3.597	3.508	3.371	3.269	3.190	3.077	2.919	2.566
19	8.185	5.926	5.010	4.500	4.171	3.939	3.765	3.631	3.523	3.434	3.297	3.195	3.116	3.003	2.844	2.489
20	8.096	5.849	4.938	4.431	4.103	3.871	3.699	3.564	3.457	3.368	3.231	3.130	3.051	2.938	2.778	2.421
21	8.017	5.780	4.874	4.369	4.042	3.812	3.640	3.506	3.398	3.310	3.173	3.072	2.993	2.880	2.720	2.360
22	7.945	5.719	4.817	4.313	3.988	3.758	3.587	3.453	3.346	3.258	3.121	3.019	2.941	2.827	2.667	2.305
23	7.881	5.664	4.765	4.264	3.939	3.710	3.539	3.406	3.299	3.211	3.074	2.973	2.894	2.781	2.620	2.256
24	7.823	5.614	4.718	4.218	3.895	3.667	3.496	3.363	3.256	3.168	3.032	2.930	2.852	2.738	2.577	2.211
25	7.770	5.568	4.675	4.177	3.855	3.627	3.457	3.324	3.217	3.129	2.993	2.892	2.813	2.699	2.538	2.169
26	7.721	5.526	4.637	4.140	3.818	3.591	3.421	3.288	3.182	3.094	2.958	2.857	2.778	2.664	2.503	2.131
27	7.677	5.488	4.601	4.106	3.785	3.558	3.388	3.256	3.149	3.062	2.926	2.824	2.746	2.632	2.470	2.097
28	7.636	5.453	4.568	4.074	3.754	3.528	3.358	3.226	3.120	3.032	2.896	2.795	2.716	2.602	2.440	2.064
29	7.598	5.420	4.538	4.045	3.725	3.499	3.330	3.198	3.092	3.005	2.868	2.767	2.689	2.574	2.412	2.034
30	7.562	5.390	4.510	4.018	3.699	3.473	3.304	3.173	3.067	2.979	2.843	2.742	2.663	2.549	2.386	2.006
32	7.499	5.336	4.459	3.969	3.652	3.427	3.258	3.127	3.021	2.934	2.798	2.696	2.618	2.503	2.340	1.956
34	7.444	5.289	4.416	3.927	3.611	3.386	3.218	3.087	2.981	2.894	2.758	2.657	2.578	2.463	2.299	1.911
36	7.396	5.248	4.377	3.890	3.574	3.351	3.183	3.052	2.946	2.859	2.723	2.622	2.543	2.428	2.263	1.872
38	7.353	5.211	4.343	3.858	3.542	3.319	3.152	3.021	2.915	2.828	2.692	2.591	2.512	2.397	2.232	1.837
40	7.314	5.179	4.313	3.828	3.514	3.291	3.124	2.993	2.888	2.801	2.665	2.563	2.484	2.369	2.203	1.805
50	7.171	5.057	4.199	3.720	3.408	3.186	3.020	2.890	2.785	2.698	2.562	2.461	2.382	2.265	2.098	1.683
60	7.077	4.977	4.126	3.649	3.339	3.119	2.953	2.823	2.718	2.632	2.496	2.394	2.315	2.198	2.028	1.601
120	6.851	4.787	3.949	3.480	3.174	2.956	2.792	2.663	2.559	2.472	2.336	2.234	2.154	2.035	1.860	1.381
∞	6.635	4.605	3.782	3.319	3.017	2.802	2.639	2.511	2.407	2.321	2.185	2.082	2.000	1.878	1.696	1.000

TABLE 22.8 – Valeur critique du test de Durbin-Watson au seuil de 5%

n	$k=1$		$k=2$		$k=3$		$k=4$		$k=5$	
ine2-11	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 22.9 – 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 . . . 11

4.1 Factorielle des nombres de 1 à 10 . . . 48
4.2 Erreur de première et seconde espèce . . . 68
4.3 Probabilité de commettre les erreurs . . . 68

5.1 Tableau récapitulatif des espérances et variances . . . 73
5.2 Tableau récapitulatif des espérances des sommes de carrés . . . 74
5.3 Tableau d'analyse de la variance . . . 77
5.4 Tableau des coefficients de régression . . . 80
5.5 Tableau d'analyse de la variance . . . 80

6.1 Tableau récapitulatif . . . 84

7.1 Tableau d'analyse de la variance . . . 94
7.2 Data savings . . . 97

8.1 Variables générées : la 8ième observation est un point influents . . . 104
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 . . . 105

9.1 Tableau d'analyse de la variance à un facteur . . . 113
9.2 Nombre de mots selon les groupes . . . 116
9.3 Moyennes selon les groupes . . . 116
9.4 Tableau d'analyse de la variance . . . 116

10.1 Prix et taille de 14 maisons à San Diego (1990) . . . 119
10.2 Temps et distance roulée par 15 voitures . . . 121
10.3 Taille et poids de 20 individus . . . 123
10.4 Masse monétaire et le produit national brut . . . 128

12.1 Prix surface et nombre d'années d'appartements . . . 149
12.2 Consommation de crème glacée . . . 152

14.1 Frais d'inscription et augmentation moyenne des salaires . . . 186
14.2 Tableau de l'analyse de la variance . . . 186
14.3 Frais cumulés et âge des véhicules . . . 189
14.4 Tableau de l'analyse de la variance . . . 189
14.5 Sexe et salaire des observations . . . 192

15.1 Tableau de l'analyse de la variance . . . 195

16.1 Tableau de l'analyse de la variance . . . 199
16.2 Tableau d'analyse de la variance : ANOVA . . . 215

17.1 Médailles d'or aux jeux olympiques pour le saut en longueur . . . 219
17.2 Médailles d'or aux jeux olympiques pour le saut en longueur . . . 220

18.1 Les résultats (perte de poids en kg) des 5 traitements . . . 225
18.2 Nombre d'éléments rappelés selon l'âge . . . 227
18.3 Nombre d'éléments rappelés selon l'âge et le niveau . . . 228
18.4 Les nombres des pains vendus . . . 230
18.5 Temps selon les groupes . . . 231

19.1 Médailles d'or aux jeux olympiques pour le saut en longueur . . . 244

20.1 Loyers moyens par m^2 selon le nombre de pièces . . . 249
20.2 Loyers moyens par m^2 selon le nombre de pièces, leviers, résidus standardisés et distances de Cook . . . 249
20.3 Composant chimique dans l'eau et la transparence de l'eau . . . 251
20.4 Composant chimique et transparence, leviers, résidus standardisés et distance de Cook . . . 251
20.5 Tableau de l'analyse de la variance . . . 278
20.6 Médailles d'or aux jeux olympiques pour le "saut en hauteur" et le "saut en longueur" . . . 292
20.7 Résultats de l'analyse de régression . . . 294
20.8 Taux de goudron (mg), taux de nicotine (mg), poids (gr), et monoxide de carbone (mg) pour 25 firmes de cigarettes, (source Mendenhall et al., 2003). . . . 297

21.1 Variables à utiliser dans la régression . . . 299
21.2 Variables à utiliser dans la régression . . . 301
21.3 Variables à utiliser dans la régression . . . 302
21.4 Variables à utiliser dans la régression . . . 304
21.5 Variables à utiliser dans la régression . . . 305

22.1 Table des quantiles d'une variable normale centrée réduite . . . 309
22.2 Fonction de répartition de la loi normale centrée réduite . . . 310
22.3 Quantiles de la loi normale centrée réduite . . . 311
22.4 Table des quantiles d'une variable χ^2 à n degrés de liberté . . . 312
22.5 Table des quantiles d'une variable de Student à n degrés de liberté . . . 313
22.6 Table des quantiles d'ordre 0.95 d'une variable de Fisher à n_1 et n_2 degrés de liberté . . . 314
22.7 Table des quantiles d'ordre 0.99 d'une variable de Fisher à n_1 et n_2 degrés de liberté . . . 315
22.8 Valeur critique du test de Durbin-Watson au seuil de 5% . . . 316
22.9 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 . . . 317

Table des figures

1.1 Le nuage de points . . . 12
1.2 Exemples de nuages de points et coefficients de corrélation . . . 13
1.3 Le nuage de points, le résidu . . . 14
1.4 La droite de régression . . . 16
1.5 Droite passant par l'origine . . . 19
1.6 Régression avec une pente nulle . . . 20

4.1 Système complet d'événements . . . 44
4.2 Illustration du théorème des probabilités totales . . . 47
4.3 Distribution de "faces" obtenus. . . . 50
4.4 Distribution d'une variable aléatoire binomiale avec $n = 5$ et $p = 0.6$. . . . 52
4.5 Distribution d'une variable de Poisson avec $\lambda = 1$. . . . 53
4.6 Probabilité que la variable aléatoire soit inférieure à a . . . 54
4.7 Fonction de densité d'une variable uniforme . . . 55
4.8 Fonction de répartition d'une variable uniforme . . . 55
4.9 Fonction de densité d'une variable normale . . . 56
4.10 Fonction de répartition d'une variable normale . . . 56
4.11 Densité d'une normale centrée réduite, symétrie . . . 56
4.12 Fonction de densité d'une variable exponentielle avec $\lambda = 1$. . . . 58
4.13 Densité d'une normale bivariée . . . 61
4.14 Nuage de points de réalisations d'une normale bivariée . . . 61
4.15 Densité d'une variable de chi-carré avec $p = 1, 2, \ldots, 10$ . . . 66
4.16 Densités de variables de Student avec $p = 1, 2$ et 3 et d'une variable normale . . . 67
4.17 Densité d'une variable de Fisher . . . 67

5.1 Représentation graphique du modèle de régression . . . 69

7.1 Nuages de points de la base de données savings . . . 98

8.1 qq-plot de 100 variables normales simulées versus une distribution théorique normale . . . 101
8.2 qq-plot de 100 variables exponentielles simulées versus une distribution théorique normale . . . 102
8.3 qq-plot de 100 variables de Student avec 2 degrés de libertés simulées versus une distribution théorique normale . . . 102
8.4 Nuage de point et droite de régression : un point extrême joue un rôle de levier sur la droite de régression . . . 105
8.5 Résidus standardisés en fonction des valeurs ajustées et qq-plot des résidus . . . 105
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 . . . 106
8.7 Résidus standardisés en fonction des valeurs ajustées et qq-plot des résidus . . . 107
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 . . . 107

19.1 Nuage de points et droite de régression . . . 242

20.1 Nuage de points . . . 261
20.2 Nuage de points . . . 280

Bibliographie

Beck, T., Levine, R., and Loayza, N. (2000). Finance and the sources of growth. *Journal of financial economics*, 58(1-2) :261–300.

Belsley, D. A., Kuh, E., and Welsch, R. E. (2005). *Regression diagnostics : Identifying influential data and sources of collinearity*. John Wiley & Sons.

Cohen, A. and Einav, L. (2003). The effects of mandatory seat belt laws on driving behavior and traffic fatalities. *Review of Economics and Statistics*, 85(4) :828–843.

Croissant, Y. and Graves, S. (2022). *Ecdat : Data Sets for Econometrics*. R package version 0.4-2.

Eysenck, M. W. (1974). Age differences in incidental learning. *Developmental Psychology*, 10(6) :936–941.

Faraway, J. J. (2004). *Linear models with R*. Chapman and Hall/CRC.

Kleiber, C. and Zeileis, A. (2008). *Applied Econometrics with R*. Springer-Verlag, New York. ISBN 978-0-387-77316-2.

Mendenhall, W., Sincich, T., and Boudreau, N. S. (2003). *A second course in statistics : regression analysis*, volume 6. Prentice Hall New York.

Owings, J. A. (1985). *High School and Beyond, a National Longitudinal Study for the 1980's : Characteristics of High School Students who Identify Themselves as Handicapped*. US Department of Education, Office of Educational Research and Improvement.

Paradis, E. (2002). R pour les débutants. Montpellier (F) : University of Montpellier https://cran.r-project.org/doc/contrib/Paradis-rdebuts_fr.pdf.

Stock, J. H. and Watson, M. W. (2003). *Introduction to Econometrics*, volume 104. Addison Wesley Boston.

Index

aberrant, 103
analyse, 11
    combinatoire, 48
    de la variance, 77, 80
    de la variance à un facteur, 109
application linéaire, 23, 27
arrangement, 48
axiomatique, 45

base
    canonique, 22
    d'un sous-espace vectoriel, 22
Bernoulli, 50
bernoullienne, 50
binôme de Newton, 50

cas bivarié, 41
CM carrés moyens, 77, 78, 94, 183, 185–187, 189, 190, 195, 197, 199, 202, 204, 206, 221, 248, 250, 252, 256, 270, 273, 275, 286, 292–295
coefficient
    de corrélation, 13, 37
      multiple, 35
      partielle, 37, 38
    de détermination, 13, 19, 35
    de régression, 69, 76
combinaison, 49
    linéaire
      de deux vecteurs, 21
combinaison linéaire
    de deux matrices, 24
complémentaire, 43
corrélation, 12
    coefficient, 12
    partielle, 37
    simple, 38
covariance, 12, 58, 63, 72

degré de liberté, 76, 78, 80, 84, 89, 90
densité
    jointe, 74
    marginale, 63, 74
différence, 43
dimension d'un sous-espace vectoriel, 22
distance
    de Cook, 104
    entre deux vecteurs, 23
distribution
    binomiale, 50, 51
    bivariée, 57
    bivarée, 62
    conditionnelle, 58
    d'une variable de Poisson, 53
    de Fisher, 78
    de probabilité, 49, 51, 67, 84
    de probabilité des estimateurs, 76, 85
    de Student, 90
    exponentielle, 57
    marginale, 58, 61, 62
    multinormale, 84
    normale bivariée, 61, 63
    normale multivariée, 64
données centrées, 42
droite
    de régression, 15
droite de régression, 13
dérivation
    d'une application linéaire, 30
    d'une forme linéaire, 30
dérivation d'une forme
    quadratique, 30
dérivée
    partielle, 14, 30, 85
déterminant, 26

écart-type, 82
    de la variable de régression, 55
    des coefficients, 79
    marginal, 11
ensemble
    parties d'un ensemble, 44
    système complet, 44
erreur
    de deuxième espèce, 67
    de première espèce, 67
erreur-type, 82
espace
    euclidien, 22
      norme, 23
    vectoriel, 21
espérance, 49, 59
    conditionnelle, 63
    d'une variable
      binomiale, 51
      indicatrice, 50
    des sommes des carrés, 73
    propriétés, 59
estimateur
    convergent, 85
    du maximum de vraisemblance, 75, 84, 85
    efficace, 85
    linéaire, 70, 83
    sans biais, 84, 85
estimation

par les moindres carrés (ordinaires), 82
événements, 43
indépendants, 47
mutuellement exclusifs, 44
exercice
Binomiale, 171
Cartes et As, 159
Comportement maternel de rats, 231
Crèmes glacées en R, 155
Crèmes glacées, 152
Densité exponentielle, 162
Densité jointe et conditionnement, 167
Densité jointe normale et corrélation, 173
Densité jointe normale, 169
Densité jointe, 165
Densité uniforme et estimation, 178
Deux variables explicatives, 195
Distribution de Poisson, 166
Décomposition d'une matrice, 138
Défauts et probabilités, 160
Détection et villes, 219
Estimateurs linéaires, 209
Frais d'inscription, 186
Frais de réparation et âge, 189
Inférence et test, 201
Inférence, 199
Inversion par bloc, 136
Leviers et influence, 223
Masse monétaire et produit national brut, 128
Matrices, 133
Modèle bivariée, 181
Modèle et sommes de carrés, 182
Pièces et probabilités, 172
Prix des maisons, 119
Prix et poids, 183
Prix, surface et années, 149
Probabilités dans une loi normale, 164
Projections, 143
Puissance de matrice, 142
Relation F et R^2, 206
Revenu et sexe, 192
Sondage, 161
Tailles et poids en R, 126
Tailles et poids, 123
Temps et distances, 121
Tennis, 157
Test de Wald, 213
Test sur des contraintes linéaires, 216
Test sur les régresseurs, 215
Tests de Wald, 203
Traitements et poids, 225
Valeurs et vecteurs propres, 141
Valeurs propres, 137
Vecteurs et matrices, 131
Vecteurs propres d'une matrice symétrique, 140
Vente de pains, 230
Vraisemblance, 207
Âge et mémorisation, 227
Âge et traitements, 228
Étude de densité, 163
expérience aléatoire, 43, 49

fonction, 53
de densité, 56, 74
conditionnelle, 58
d'une variable aléatoire continue, 53
d'une variable exponentielle, 58
d'une variable normale multivariée, 64
d'une variable uniforme, 55
marginale, 58
de log-vraisemblance, 74
de répartition jointe, 57
de vraisemblance, 74, 85
forme
linéaire, 27
quadratique, 27, 86
formulation matricielle, 81

gradient, 30

homoscédasticité, 81
homoscédastique, 63, 69

image d'une matrice, 28
indépendance, 59
influent, 104
inférence, 89
inférence statistique, 43, 67
intersection, 43
intervalle
d'acceptation, 68
de confiance, 67, 76
sur un coefficient de régression, 89
inversion par bloc, 25

log-vraisemblance, 75
loi
binomiale, 50
de Poisson, 52
normale bivariée, 60

matrice, 23
carrée, 24, 26
carée symétrique, 27
de corrélation, 36
de plein rang, 82
de projection, 29
de variance-covariance, 36
des corrélations, 36
des variables explicatives, 109
diagonale, 25, 86
définie positive, 27
déterminant, 26
hessienne, 15
idempotente, 28, 29, 40, 83, 86
trace, 29
identité, 25, 26
image, 28
inversible, 25
noyau, 28
orthogonale, 26, 86
produit d'une matrice et d'un vecteur, 24
propriétés, 26
rang, 25

semi-définie positive, 27
sous-espace, 28
symétrique, 24, 86
trace, 25
transposition, 24
valeurs propres, 27
variance-covariance, 40
vecteurs propres, 27
modèle
de régression, 69
estimé, 79
linéaire, 74
avec une constante, 90
linéaire général, 81, 86
définition, 81
hypothèses, 81
paramétrique, 74
modélisation, 67
moindres carrés, 13, 14, 70, 85
principe, 34
moyenne, 11
conditionnelle, 58
marginale, 11, 58, 60, 62
multicolinéarité, 82
multiplication par un scalaire, 21

normale bivariée, 60
normalité des erreurs, 74
norme, 23
noyau d'une matrice, 28

orthogonal d'un sous-espace vectoriel, 23
outlier, 103

paramètres marginaux, 11
pente, 15
nulle, 20
permutation
avec répétition, 48
sans répétition, 48
point levier, 103
principe des moindres carrés, 33
probabilité, 43, 45
conditionnelle et indépendance, 46
théorème des probabilités totales, 47
produit
d'une matrice et d'un vecteur, 24
matriciel, 24
scalaire, 12, 22
projecteur oblique, 29
projecteur orthogonal, 28, 29
projection, 23, 28
orthogonale, 28
dans l'image, 28
dans le noyau, 28
propriétés, 65
propriétés des espérances et des variances, 59
prévision ponctuelle d'une valeur, 95
puissance d'un test, 68

qq-plot, 101

rang d'une matrice, 25
régression, 13
bivariée, 11, 69
données centrées, 39
linéaire, 78
multiple, 39
multivariée, 33, 77
représentation
graphique de deux variables, 11
matricielle des données, 33
résidus, 15, 16, 72
normaux, 81
standardisés et studentisés, 103

SC somme de carrés, 17, 18, 39, 41, 73, 74, 77, 78, 93, 94, 112, 113, 115, 123–125, 150, 153, 181–187, 189–191, 195–197, 199, 201, 202, 204, 206, 213, 221, 225, 230, 247–253, 256, 258–260, 264, 267–270, 273, 275, 277, 279, 282, 286, 289, 291, 293–295
scalaire, 21
somme
des carrés, 39, 41
de la régression, 17
des résidus, 14, 18, 39, 80
expliquée par la régression, 39
inter-groupes, 112, 115
intra-groupes, 112, 115
totale, 17
totale des écarts à la moyenne, 39
des carrés des résidus, 76
des résidus, 38
des variables aléatoires, 60
sous ensemble, 43
sous-espace vectoriel, 22, 28
base, 22
base canonique de R, 22
dimension, 22
orthogonal, 23
système générateur, 22
statistique exhaustive, 85
système
complet d'événements, 44, 46, 47
générateur d'un sous-espace vectoriel, 22
série statistique bivariée, 11

terme d'erreur, 69, 72, 81–83, 85–87, 109
test
d'hypothèses, 76, 77
composites, 68
simples, 67
d'un seul coefficient de régression, 90
de Fisher sur un coefficient de régression, 94
de Wald sur les coefficients de régression, 91
global sur les coefficients de régression, 92
théorie des ensembles, 43
théorème
de Bayes, 47
de diagonalisation, 27
de Gauss-Markov, 83
de Pythagore, 23
des probabilités totales, 47

des trois perpendiculaires, 30
trace d'une matrice, 25
idempotente, 29
transposition, 21

union, 43

valeur
ajustée, 15, 72
propre, 27, 86
d'une matrice idempotente, 29
valeur propre, 29
variable
aléatoire, 43, 49, 54, 55, 57, 59, 64, 70, 72, 76
continue, 53, 57
de Fisher, 78
de Student, 89
discrète, 49
indépendante, 59
normale indépendante, 74
uniforme, 55
binomiale, 50
constante, 38
de Fisher, 66
de Poisson, 52
de Student, 66
de student, 76
dépendante, 40
espérance, 49
explicative, 38, 40, 41, 80
indicatrice, 50
indépendante, 37
khi-carrée, 66
normale, 55
centrée réduite, 56
multivariée, 64
uniforme, 54, 55
variance, 11, 17, 18, 49–59, 64, 66, 72, 83
conditionnelle, 58, 63
d'une variable
binomiale, 51
indicatrice, 50
de régression, 18, 19
régression multivariée, 35
des résidus, 72
marginale, 11, 19, 58, 60, 62
minimum, 85
propriétés, 59
résiduelle, 18
régression multivariée, 35
vecteur, 21
aléatoire, 64, 82
colonne, 21, 22, 40
des résidus, 35, 37, 38
des valeurs ajustées, 35, 42
ligne, 21, 22
linéairement indépendant, 22
multinormal, 65
multinormal non-corrélé, 86
orthogonal, 23
projection, 23
propre, 27, 86