L2 MIASHS
Année 2021-22

Statistiques Descriptives 2

TD 2 de Statistiques Descriptives 2

Exercice 1 Notation matricielle

On considère le modèle de régression linéaire simple du Chapitre 1 oû l'on dispose de n observations (x_i, y_i) vérifiant

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

où l'on suppose que les variables $\epsilon_i, i = 1 \dots n$ sont centrées, de variance σ^2 et non-corrélées. On veut retrouver les propriétés du Chapitre 1 à l'aide des notations matricielles du Chapitre 2.

1. Écrire le modèle sous la forme matricielle d'un modèle de régression linéaire multiple.

   Réponse : On peut écrire matriciellement le modèle de la manière suivante :

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

   où

   $$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \quad \text{et} \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

   — Y désigne le vecteur à expliquer de taille $n \times 1$,
   — X la matrice explicative de taille $n \times 2$,
   — ε le vecteur d'erreurs de taille $n \times 1$.

2. Calculer l'estimateur des moindres carrés $\hat{\beta}$ dans le modèle matriciel et retrouver les estimateurs $\hat{\beta}_0$ et $\hat{\beta}_1$ du modèle de régression simple.

   Réponse : comme

   $$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = (X^t X)^{-1} X^t Y$$

   il faut calculer le terme $(X^t X)^{-1} X^t Y$. On commence en calculant $(X^t X)^{-1}$. Pour cela, on utilise la formule pour l'inverse d'une matrice

   $$A^{-1} = \frac{1}{\det A} \operatorname{com}(A)^t$$

où com(A) est la matrice de cofacteurs (ou co-matrice) [1]. Commençons donc par calculer

$$X^t X = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

ce qui donne

$$X^t X = \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix}.$$

En se rappelant que le déterminant de

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

est

$$ad - bc,$$

on obtient que le déterminant de $X^t X$ est

$$\begin{aligned} \det(X^t X) &= n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \\ &= n^2 \left(\overline{x^2} - \bar{x}^2 \right) = n^2 \; s_x^2. \end{aligned}$$

où

$$\overline{x^2} = \frac{1}{n} \sum_{i=1}^n x_i^2.$$

De plus, la matrice des co-facteurs est définie par

$$\mathrm{com}(A)_{i,j} := (-1)^{i+j} \det\left(A_{i,j}\right),$$

où $A_{i,j}$ est la matrice carrée de taille n déduite de A en supprimant la i ème ligne et la $j^{\text{ième}}$ colonne [2]. On a donc

$$\mathrm{com}(A) = \begin{bmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{bmatrix}$$

1. matrice de cofacteurs
2. par exemple, pour

$$A = \begin{pmatrix} 6 & 7 & 1 & 8 & 1 \\ 1 & 0 & 0 & 3 & 2 \\ 0 & -1 & 6 & 3 & 4 \end{pmatrix} \quad \text{on a} \quad A_{23} = \begin{pmatrix} 6 & 7 & 8 & 1 \\ 0 & -1 & 3 & 4 \end{pmatrix}$$

ce qui donne, en divisant par le déterminant,

$$\left(X^t X\right)^{-1} = \frac{1}{n^2\ s_x^2} \begin{bmatrix} \sum_{i=1}^{n} x_i^2 & -\sum_{i=1}^{n} x_i \\ -\sum_{i=1}^{n} x_i & n \end{bmatrix} \tag{1}$$

et donc en prenant un n dans le n^2 du dénominateur $n^2\ s_x^2$ et en s'en servant pour diviser tous les termes de la matrice de droite, on obtient

$$\left(X^t X\right)^{-1} = \frac{1}{n\ s_x^2} \begin{bmatrix} \overline{x^2} & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix} \tag{2}$$

De plus,

$$X^t Y = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

ce qui donne

$$X^t Y = \begin{bmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} Y_i x_i \end{bmatrix}.$$

$$\begin{aligned} (X^t X)^{-1} X^t Y &= \frac{1}{s_x^2} \begin{bmatrix} \overline{x^2} & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{n} \sum_{i=1}^{n} Y_i \\ \frac{1}{n} \sum_{i=1}^{n} x_i Y_i \end{bmatrix}, \\ &= \frac{1}{s_x^2} \begin{bmatrix} \overline{x^2} & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix} \begin{bmatrix} \bar{Y} \\ \overline{Yx} \end{bmatrix}, \\ &= \frac{1}{s_x^2} \begin{bmatrix} \bar{Y}\overline{x^2} - \bar{x}\overline{Yx} \\ -\bar{Y}\bar{x} + \overline{Yx} \end{bmatrix} \end{aligned}$$

On en déduit que

$$\hat{\beta}_1 = \frac{\overline{Yx} - \bar{Y}\bar{x}}{s_x^2}$$

et comme "moyenne des produits moins produit des moyennes" est égal à la covariance, i.e. $\overline{Yx} - \bar{Y}\bar{x} = C_{xY}$, on retrouve bien la formule du cours, i.e.

$$\hat{\beta}_1 = \frac{C_{xY}}{s_x^2}.$$

On a aussi

$$\hat{\beta}_0 = \frac{\bar{Y}\overline{x^2} - \bar{x}\overline{Yx}}{s_x^2}$$

Comme $C_{xY} = \overline{Yc} - \bar{Y}\bar{x}$, on a $\overline{Yc} = C_{xY} + \bar{Y}\bar{x}$, et si on remplace $\overline{Yc}$ dans l'expression ci-dessus et on a

$$\begin{aligned}\hat{\beta}_0 &= \frac{\bar{Y}\overline{x^2} - \bar{x}C_{xY} - \bar{x}^2\bar{Y}}{s_x^2} = \frac{\bar{Y}(\overline{x^2} - \bar{x}^2) - \bar{x}C_{xY}}{s_x^2} = \frac{\bar{Y}s_x^2 - \bar{x}C_{xY}}{s_x^2} \\ &= \bar{Y} - \bar{x}\frac{C_{xY}}{s_x^2},\end{aligned}$$

et donc

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

On a donc bien retrouvé les formules de $\hat{\beta}_0$ et $\hat{\beta}_1$.

Exercice 2 Retrouver la formule de l'estimateur $\hat{\beta}$ pour le modèle linéaire multivarié à partir du principe des moindres carrés.

Solution : On veut retrouver la formule

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

L'idée est de minimiser la somme des carrées des résidus pour trouver les coefficients du vecteur de régression. C'est à dire,

$$\min_{b\in\mathbb{R}^{p+1}} \sum_{i=1}^{n} (Y_i - x_i^\top b)^2$$

Appelons

$$f(b) = \sum_{i=1}^{n} (Y_i - x_i^\top b)^2$$

Calculons la dérivée directionnelle de f dans une direction arbitraire $h \in \mathbb{R}^{p+1}$:

$$f'(b;h) = \lim_{t\downarrow 0} \frac{f(b+th) - f(b)}{t}$$

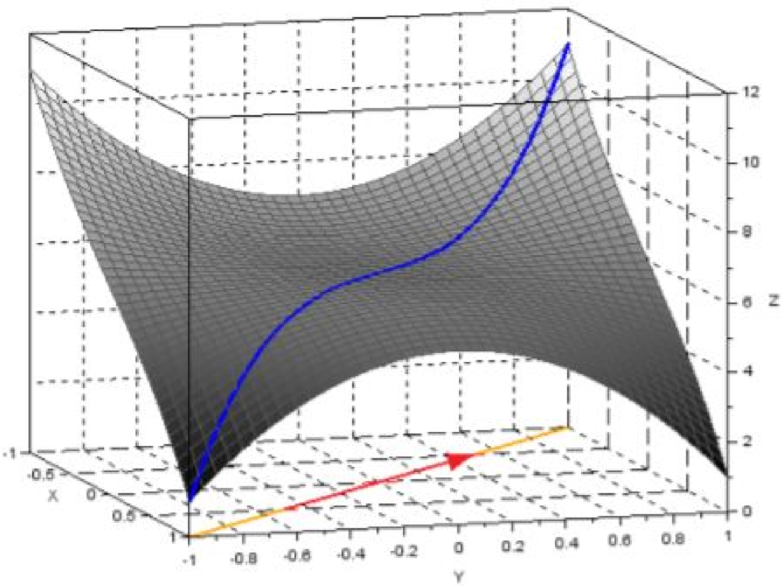

FIGURE 1 – Restriction d'une fonction le long d'une direction (crédit : Mathieu Mansuy).

Comment faire ce calcul ? Il faut écrire f simplement grâce à la formulation matricielle

$$f(b) = \|Y - Xb\|_2^2$$

où $\|\cdot\|_2$ indique la norme euclidienne, $\sqrt{\sum_{i=1}^{2}(\cdot)^2}$. Il faut donc juste calculer

$$\begin{aligned}
f(b+th) &= \|Y - X(b+th)\|_2^2 \\
&= \|Y - Xb - tXh\|_2^2 \\
&= \|Y - Xb\|_2^2 - 2\langle Y - Xb, tXh\rangle + \|tXh\|_2^2 \\
&= \|Y - Xb\|_2^2 - 2t\ \langle Y - Xb, Xh\rangle + t^2\|Xh\|_2^2
\end{aligned}$$

et donc

$$\begin{aligned}
f'(x;h) &= \lim_{t\downarrow 0} \frac{\|Y - Xb\|_2^2 - 2t\ \langle Y - Xb, Xh\rangle + t^2\|Xh\|_2^2 - \|Y - Xb\|_2^2}{t} \\
&= \lim_{t\downarrow 0} \frac{-2t\ \langle Y - Xb, Xh\rangle + t^2\|Xh\|_2^2}{t} \\
&= \lim_{t\downarrow 0}\ -2\ \langle Y - Xb, Xh\rangle + t\|Xh\|_2^2 \\
&= -2\ \langle Y - Xb, Xh\rangle.
\end{aligned}$$

La valeur $\hat{\beta}$ de b où f est minimum est une valeur pour laquelle ou la dérivée sera nulle, pour toutes les directions h, c'est à dire

$$\langle Y - X\hat{\beta}, Xh\rangle = 0$$

quel que soit h. Rappelons que $\langle w, z\rangle = \sum_{i=1}^{d} w_i z_i = w^\top z$. Comme $\langle w, Xz\rangle = w^\top Xz$ et comme $w^\top X = (X^\top w)^\top$, on obtient $\langle w, Xz\rangle = (X^\top w)^\top z = \langle X^\top w, z\rangle$.

Cela donne

$$\langle X^\top (Y - X\hat{\beta}), h\rangle = 0$$

et donc

$$X^\top (Y - X\hat{\beta}) = 0.$$

Pour le voir, il suffit de dire que si $\langle z, h\rangle = 0$ quel que soit h, alors on peut prendre $h = z$, ce qui donne $\|z\|_2^2 = 0$ et donc $z = 0$ et on fait cela pour $z = X^\top (Y - X\hat{\beta})$. On développe maintenant cette expression et on passe dans l'autre membre pour obtenir

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

et donc

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

Si la fonction f est convexe, alors cela suffit à être sûr que la solution $\hat{\beta}$ est un minimiseur de f. La convexité d'une fonction deux fois différentiable est certifiée à partir du moment où la matrice des dérivées secondes est positive semi-définie (toutes les valeurs propres de la matrices des dérivées secondes sont positives ou nulles.

Exercice 3 Au vu de la représentation de la concentration d'ozone en fonction de la température à midi nous souhaitons modéliser l'ozone par la température via un modèle quadratique.

1. Afficher sous R les données à l'aide des commandes suivantes

```
ozone<-read.table("ozone.txt",header=T)
y = ozone$maxO3
x = ozone$T12
plot(x,y)
res = lm(y~ x)
abline(res$coefficients)
```

2. Ecrire le modèle et estimer les paramètres.

   Réponse : pour un modèle quadratique, on a

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i$$

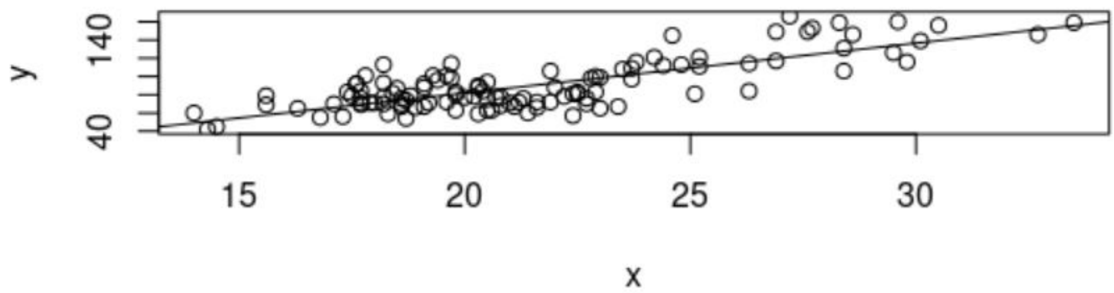

FIGURE 2 – Représentation du nuage de points ozone versus temperature à midi et modèle linéaire pour ces données.

où les ϵ_i sont des variables aléatoires centrées, de variance σ^2 et non-corrélées. Par contre, il est équivalent à un modèle linéaire multivarié car on peut l'écrire

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}.$$

En posant comme d'habitude

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix}$$

on obtient, en utilisant la formule de l'estimation de $\hat{\beta}$ dans le modèle multivarié, on obtient,

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = (X^t X)^{-1} X^t Y.$$

On pourrait essayer de calculer explicitement les formules pour $\hat{\beta}_0$, $\hat{\beta}_1$ et $\hat{\beta}_2$, mais les calculs sont plus longs et compliqués que dans le cas linéaire simple, donc on s'arrête là pour la théorie. Sous R, cela donne :

```
ozone<-read.table("ozone.txt",header=T)
y = ozone$maxO3
x = ozone$T12
plot(x,y)
x2=x**2
res = lm(y~x+x2)
res$coefficients
```

3. Comparer ce modèle au modèle de régression linéaire.

Réponse : ce modèle est plus riche que le modèle linéaire car il le généralise. En effet, lorsqu'on prend $\beta_2 = 0$, on obtient à nouveau le modèle linéaire, mais lorsque $\beta_2 \neq 0$, on a un modèle qui s'en écarte nécéssairement.
On peut comparer les deux modèles à partir du critère BIC vu en cours. Pour cela, il faut retrouver la formule du BIC : on calcule la valeur de -2 le logarithme de la vraisemblance à l'optimum pour les deux modèles et on a ajoute le terme 2 fois le nombre de paramètres à trouver fois $\log(n)$. De plus, la valeur de -2 fois la log-vraisemblance à l'optimum est simple après coup : $n \log(2\pi\hat{\sigma}^2)+n$. Pour le premier modèle (modèle linéaire univarié pur), on obtient grâce à la commande

hatsigma2 =mean(res$residuals**2)

sous R,

$$\hat{\sigma}^2 = 303.1058$$

ce qui donne un BIC égal à

$$BIC = 967.2564$$

et pour le modèle quadratique, on obtient

$$\hat{\sigma}^2 = 286.0124$$

et cela donne un BIC égal à

$$BIC = 965.4736.$$

On peut aussi tester de la même façon le modèle cubique et on trouve

$$BIC = 968.957.$$

Comme décrit dans le cours, le problème principal auquel répond le critère BIC est celui de trouver le bon équilibre entre proposer un modèle plus riche qui risque d'"overfitter" (en français, on parle de "sur-apprentissage") et un modèle pas assez risque qui risque d'avoir un gros biais. Ces situations sont illustrées dans la figure ci-dessous.

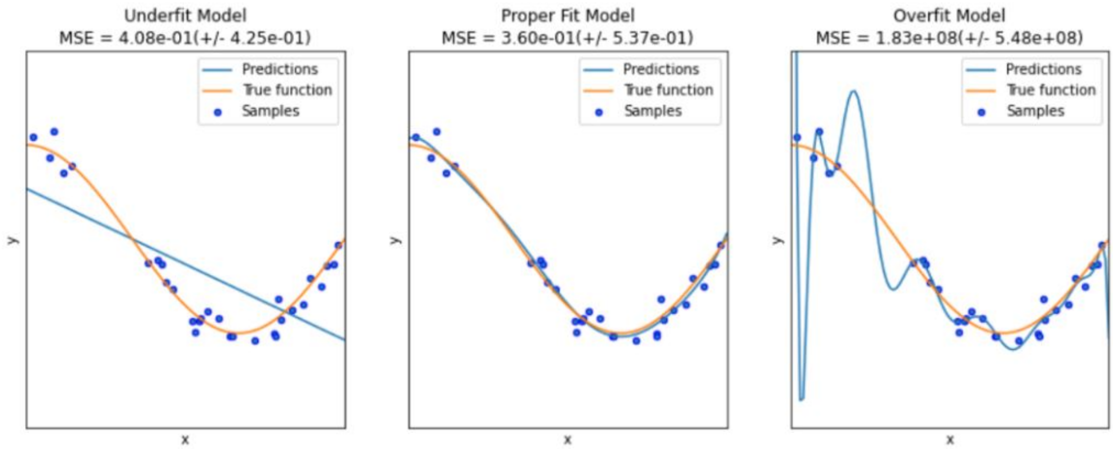

FIGURE 3 – Les divers types de modèle : les trop pauvres, les bien équilibrés et les trop riches qui "overfit". Cette figure a été capturée sur internet et il manque les crédits.

Exercice 4 1. Montrer que pour un modèle linéaire multiple

$$Y_i = \beta_0 + \beta^t x_i + \epsilon_i$$

on peut modifier les données X_i en $X_{i,*}$ de telle maniere que

$$Y_i = \beta_*^t x_{i,*} + \epsilon_i$$

pour un certain vecteur β_*. On pourra donc supposer que tous les modèles linéaires seront de ce type. On notera $\beta = \beta_*$ dans la suite.

Réponse : On peut prendre

$$X_{i,*} = \begin{bmatrix} 1 \\ x_{i,1} \\ \vdots \\ x_{i,p} \end{bmatrix}$$

où $x_{i,1}, \ldots, x_{i,p}$ sont les coordonnées[3] de x_i. On peut aussi définir

$$\beta_* = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}.$$

Maintenant, quand on calcule $\beta_*^t x_{i,*}$, on trouve bien $\beta_0 + \beta^t x_i$ comme prévu.

3. on a noté le nombre de coordonnées par p mais on aurait pu prendre une autre lettre. La lettre p est quand même très souvent utilisée.

2. On rappelle que la solution du problème d'estimation pour modèle linéaire Gaussien est donnée par

$$\hat{\beta} = (X^t X)^{-1} X^t Y \tag{3}$$

— Montrer que $\mathbb{E}\left[\hat{\beta}\right] = \beta$.

Réponse : On a

$$\begin{aligned}\mathbb{E}[\hat{\beta}] &= \mathbb{E}\left[(X^t X)^{-1} X^t Y\right] \\ &= (X^t X)^{-1} X^t \mathbb{E}\left[Y\right]\end{aligned}$$

D'un autre coté, on a

$$Y = X\beta + \epsilon$$

et donc

$$\begin{aligned}\mathbb{E}[Y] &= \mathbb{E}\left[X\beta + \epsilon\right], \qquad (4) \\ &= X\beta + \mathbb{E}[\epsilon], \qquad (5)\end{aligned}$$

mais $\mathbb{E}[\epsilon] = 0$ car ϵ est un bruit centré. Plus précisément ϵ est un vecteur dont les composantes sont des variables aléatoires[4] d'espérance nulle. L'espérance d'un vecteur est par définition le vecteur des espérances des composantes de ce vecteur, et c'est pour cela que $\mathbb{E}[\epsilon] = 0$.

— Montrer que la matrice de covariance de $\hat{\beta}$

$$\mathrm{Var}(\hat{\beta}) =: \mathbb{E}\left[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^t\right] \tag{6}$$

est donnée par

$$\mathrm{Var}(\hat{\beta}) = \sigma^2 (X^t X)^{-1}. \tag{7}$$

On a

$$\begin{aligned}\hat{\beta} - \beta &= (X^t X)^{-1} X^t Y - \beta \\ &= (X^t X)^{-1} X^t (X\beta + \epsilon) - \beta \\ &= {\color{red}(X^t X)^{-1} (X^t X)}\beta + (X^t X)^{-1} X^t \epsilon - {\color{green}\beta}\end{aligned}$$

ce qui donne, la quantité en rouge disparaissant, donnant β, qui s'annule lui même avec la quantité en vert,

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

4. indépendantes, mais cela n'est pas utile ici

On en déduit que, comme la transposée de $(X^tX)^{-1}X^t\epsilon$ est $\epsilon^tX(X^tX)^{-1}$,

$$\mathbb{E}\left[(\hat\beta-\beta)(\hat\beta-\beta)^t\right]=\mathbb{E}\left[(X^tX)^{-1}X^t\epsilon\epsilon^tX(X^TX)^{-1}\right]$$

ce qui donne

$$\mathbb{E}\left[(\hat\beta-\beta)(\hat\beta-\beta)^t\right]=(X^tX)^{-1}X^t\mathbb{E}\left[\epsilon\epsilon^t\right]X(X^TX)^{-1} \tag{8}$$

C'est maintenant le moment d'utiliser le fait que les composantes de ϵ sont centrées et indépendantes et de variance σ^2, car on obtient alors que

$$\mathbb{E}\left[\epsilon\epsilon^t\right]=\sigma^2 I$$

En réinjectant ce résultat dans (8), on obtient

$$\mathbb{E}\left[(\hat\beta-\beta)(\hat\beta-\beta)^t\right]=\sigma^2\ (X^tX)^{-1}X^tX(X^TX)^{-1} \tag{9}$$

et donc,

$$\mathbb{E}\left[(\hat\beta-\beta)(\hat\beta-\beta)^t\right]=\sigma^2\ (X^TX)^{-1} \tag{10}$$

— Retrouver alors les formules (11) et (12) suivantes :

$$\hat\beta_1\sim\mathcal{N}\left(\beta_1,\frac{\sigma^2}{\sum_{i=1}^n(x_i-\bar x)^2}\right) \tag{11}$$

et

$$\hat\beta_0\sim\mathcal{N}\left(\beta_0,\frac{\sigma^2\ \sum_{i=1}^n\ x_i^2}{n\ \sum_{i=1}^n\ (x_i-\bar x)^2}\right) \tag{12}$$

Réponse : Il faut calculer $(X^tX)^{-1}$ pour le modèle linéaire simple. On a déjà fait ce calcul pour l'exercice 1 et on a trouvé

$$\left(X^tX\right)^{-1}=\frac{1}{n^2\ \left(\overline{x^2}-\bar x^2\right)}\begin{bmatrix}\sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i\\ -\sum_{i=1}^n x_i & n\end{bmatrix}. \tag{13}$$

Comme $s_x^2=\overline{x^2}-\bar x^2=\dfrac{1}{n}\displaystyle\sum_{i=1}^n(x_i-\bar x)^2$, on obtient

$$\left(X^tX\right)^{-1}=\frac{1}{n\ \sum_{i=1}^n(x_i-\bar x)^2}\begin{bmatrix}\sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i\\ -\sum_{i=1}^n x_i & n\end{bmatrix}. \tag{14}$$

La variance de β_0 est la composante (1,1) de la matrice de covariance, c'est à dire

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

et

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

comme annoncé.

Exercice 5 On considère les données suivantes :

y	x_1	x_2	x_3	x_4
125	13	18	25	11
158	39	18	59	30
207	52	50	62	53
182	29	43	50	29
196	50	37	65	56
175	64	19	79	49
145	11	27	17	14
144	22	23	31	17
160	30	18	34	22
175	51	11	58	40
151	27	15	29	31
161	41	22	53	39
200	51	52	75	36
173	37	36	44	27
175	23	48	27	20
162	43	15	65	36
155	38	19	62	37
230	62	56	75	50
162	28	30	36	20
153	30	25	41	33

où
— Y - est une mesure du succès à un test de langue.
— x_1 - est une mesure du score moyen du candidat à un test de compétance en éthnologie.
— x_2 - est une mesure de l'étique de travail du candidat.
— x_3 - est une mesure du score moyen à un test d'allemand
— x_4 - est une mesure du score moyen à un test de mathématiques.

1. Ecrire le modèle sous forme matricielle en supposant les erreurs d'observation Gaussiennes.

Réponse : Le modèle est naturellement

$$\underbrace{\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}}_{Y} = \underbrace{\begin{bmatrix} 1 & x_{1,1} & x_{1,2} & x_{1,3} & x_{1,4} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n,1} & x_{n,2} & x_{n,3} & x_{n,4} \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}}_{\beta} + \underbrace{\begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}}_{\epsilon}$$

avec $\epsilon_1, \ldots, \epsilon_n$ i.i.d. $\mathcal{N}(0, \sigma^2)$.

2. Estimer le vecteur β puis donner l'équation de la droite des moindres carrés. Pour cela, on donne (n'oubliez pas la colonne de 1 dans la matrice X !)

$$\left(X^t X\right)^{-1} = \begin{bmatrix} 0.5464 & 0.0050 & -0.0056 & -0.0075 & -0.0045 \\ 0.0050 & 0.0032 & -0.0000 & -0.0016 & -0.0014 \\ -0.0056 & -0.0000 & 0.0003 & -0.0000 & -0.0001 \\ -0.0075 & -0.0016 & -0.0000 & 0.0013 & 0.0000 \\ -0.0045 & -0.0014 & -0.0001 & 0.0000 & 0.0017 \end{bmatrix}$$

Réponse : On utilise la formule

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

Comme on a les valeurs de la matrice $(X^t X)^{-1}$, on n'a plus qu'a calculer

$$X^t Y = \begin{bmatrix} 3389 \\ 130863 \\ 103654 \\ 173209 \\ 114600 \end{bmatrix}$$

et multiplier $(X^t X)^{-1}$ avec $X^t Y$ et on obtient

$$\hat{\beta} = \begin{bmatrix} 102.7439 \\ 1.2540 \\ 1.0643 \\ -0.3714 \\ 0.2339 \end{bmatrix}$$

et on obtient la droite

$$Y \approx 102.7439 + 1.2540 x_1 + 1.0643 x_2 - .3714 x_3 + 0.2339 x_4.$$

3. Calculer les estimations de σ^2 et $\mathrm{Var}(\hat{\beta})$ en prenant la formule (7) de l'exercice précédant, c'est à dire

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

Réponse : On trouve dans la section 3.4.2 p.6 du polycopié sur ma page web

$$s^2 = \frac{\sum_{i=1}^n (Y_i - \hat{y}_i)^2}{n-4-1} \tag{15}$$

avec

$$\begin{bmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_n \end{bmatrix} = X\hat{\beta}.$$

Le calcul donne les valeurs :

Valeur prédite	
$\hat{y}_1$	131.4917,
$\hat{y}_2$	155.9131,
$\hat{y}_3$	210.5381,
$\hat{y}_4$	173.0888,
$\hat{y}_5$	193.7818,
$\hat{y}_6$	185.3442,
$\hat{y}_7$	142.2350,
$\hat{y}_8$	147.2744,
$\hat{y}_9$	152.0404,
$\hat{y}_{10}$	166.2216,
$\hat{y}_{11}$	149.0474,
$\hat{y}_{12}$	167.0117,
$\hat{y}_{13}$	202.6085,
$\hat{y}_{14}$	177.4314,
$\hat{y}_{15}$	177.3229,
$\hat{y}_{16}$	156.9114,
$\hat{y}_{17}$	156.2465,
$\hat{y}_{18}$	223.9344,
$\hat{y}_{19}$	161.0933,
$\hat{y}_{20}$	159.4636.

De plus,

$$s^2 = \frac{\sum_{i=1}^n (Y_i - \hat{y}_i)^2}{n-4-1}. \tag{16}$$

En utilisant ces valeurs, la formule (16) donne

$$s^2 = 39.1840.$$

Les résidus sont toujours importants à visualiser. Ici, ils sont remarquablement gentils pour notre modèle car la moyennes des résidus est très petites et les signes s'alternent presque harmonieusement sur les données !

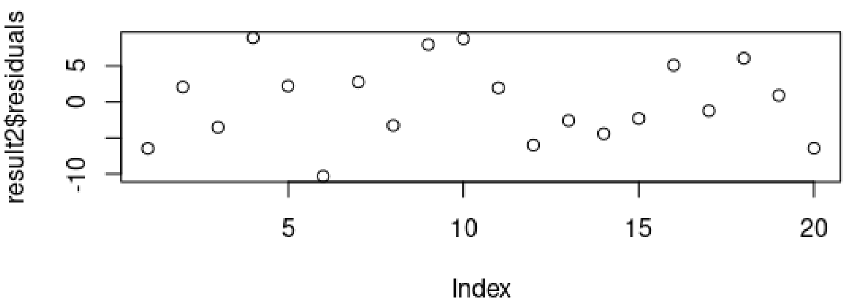

FIGURE 4 – Les résidus pour ce modèle linéaire multiple

Tournons-nous maintenant vers la matrice de variance-covariance de $\hat{\beta}$: on prend comme estimateur

$$s^2(X^\top X)^{-1}$$

où on a remplacé σ^2 par son estimateur s^2. Cela donne

$$39.1840\,\left(X^tX\right)^{-1} = 39.1840 \times \begin{bmatrix} 0.5464 & 0.0050 & -0.0056 & -0.0075 & -0.0045 \\ 0.0050 & 0.0032 & -0.0000 & -0.0016 & -0.0014 \\ -0.0056 & -0.0000 & 0.0003 & -0.0000 & -0.0001 \\ -0.0075 & -0.0016 & -0.0000 & 0.0013 & 0.0000 \\ -0.0045 & -0.0014 & -0.0001 & 0.0000 & 0.0017 \end{bmatrix}$$

$$= \begin{bmatrix} 21.4102077 & 0.1956510354 & -0.2178526911 & -0.2954384348 & -0.177860934 \\ 0.1956510 & 0.1240232098 & -0.0005394934 & -0.0612044395 & -0.053986850 \\ -0.2178527 & -0.0005394934 & 0.0112474326 & -0.0002625021 & -0.002353997 \\ -0.2954384 & -0.0612044395 & -0.0002625021 & 0.0513147016 & 0.001179114 \\ -0.1778609 & -0.0539868495 & -0.0023539966 & 0.0011791140 & 0.067334946 \end{bmatrix}$$

Cette matrice nous indique que certaines composantes sont très peu dépendantes les unes des autres sous l'hypothèse Gaussienne. Par exemple, la covariance entre $\hat{\beta}_1$ et $\hat{\beta}_2$ est assez petite, i.e. -0.0005394934. Pour en savoir plus, il faut calculer les coefficients de corrélation, c'est à dire diviser les covariances par les racines carrées des variances respectives des deux termes, i.e. des deux "beta chapeau" pris en compte.

4. Tester l'hypothese nulle H_0 : "$\beta_j = 0$" contre l'alternative H_1 : "$\beta_j \neq 0$" pour $j = 0, 1, 2$. Pour cette question, il faut regarder la section 3.6 du polycopié sur ma page web.

Réponse : Pour β_0, on calcule la variance de $\hat{\beta}_0$, qui est donnée par

$$\mathrm{Var}(\hat{\beta}_0) = \sigma^2 (X^t X)^{-1}_{1,1}$$

c'est à dire la composante (1,1) de la matrice $\sigma^2(X^tX)^{-1}$, qui est égale à $0.5464\ \sigma^2$. On en déduit

$$\frac{\hat{\beta}_0 - \beta_0}{\sqrt{s^2 \ \cdot \ 0.5464}} \sim t_{n-5}$$

Sous l'hypothèse H_0, on a $\beta_0 = 0$ et donc

$$\frac{\hat{\beta}_0}{\sqrt{s^2 \ \cdot \ 0.5464}} \sim t_{n-5}$$

ce qui implique en regardant la table de la loi de Student avec 15 degrés de liberté, que

$$\frac{\hat{\beta}_0}{\sqrt{s^2 \ \cdot \ 0.5464}} \in [-2.13; 2.13]$$

avec probabilité .95. On remplace par les nombres obtenus pour $\hat{\beta}_1$, s^2 et n et on obtient

$$\frac{102.7439}{\sqrt{39.1840 \ \cdot \ 0.5464}} = 22.20 \notin [-2.13; 2.13].$$

On en déduit que $\beta_0 \neq 0$. Pour β_1, on recommence la même routine : on a

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{s^2 \ \cdot \ 0.0032}} \sim t_{n-5}$$

et comme sous (H_0) on a $\beta_1 = 0$, on en déduit

$$\frac{\hat{\beta}_1}{\sqrt{s^2 \ \cdot \ 0.0032}} \sim t_{n-5}$$

et on calcule

$$\frac{\hat{\beta}_1}{\sqrt{s^2 \ \cdot \ 0.0032}} = 3.560826 \notin [-2.13; 2.13],$$

de quoi on déduit que $\beta_1 \neq 0$. On continue avec β_2 : on a

$$\frac{\hat{\beta}_2 - \beta_2}{\sqrt{s^2 \ \cdot \ 0.0003}} \sim t_{n-5}$$

et comme sous (H_0) on a $\beta_1 = 0$, on en déduit

$$\frac{\hat{\beta}_2}{\sqrt{s^2 \ \cdot \ 0.0003}} \sim t_{n-5}$$

et on calcule

$$\frac{\hat{\beta}_2}{\sqrt{s^2 \cdot 0.0003}} = 10.03547 \notin [-2.13; 2.13],$$

de quoi on déduit que $\beta_2 \neq 0$. Et ainsi de suite ...
Pour β_3, on a la valeur de la statistique

$$\frac{\hat{\beta}_3}{\sqrt{s^2 \cdot 0.0013}} = -1.639 \in [-2.13; 2.13],$$

et donc on conclut que $\beta_3 = 0$. De même pour β_4, on a la valeur de la statistique

$$\frac{\hat{\beta}_4}{\sqrt{s^2 \cdot 0.0017}} = 0.901 \in [-2.13; 2.13],$$

et donc on conclut que $\beta_4 = 0$.

Exercice 6 Jeux videos

Les données que nous étudions le score y de à un certain jeu vidéo en fonction de divers facteurs, ou covariables :

— x_1 est le taux d'une certaine hormone,
— x_2 est le score à un test de logique
— x_3 est le nombre d'amis avec lesquels on communique activement a propos de ce jeux.

x_1	x_2	x_3	y
124	33	8	81
49	31	6	55
181	38	8	80
4	17	2	24
152	39	6	52
75	30	7	88
54	29	7	45
43	35	6	50
41	31	5	69
17	23	4	66
22	21	3	45
16	8	3	24
10	23	3	43
63	37	6	38
170	40	8	72
15	38	6	41
15	25	4	38
221	39	7	52
171	33	7	52
97	38	6	66
254	39	8	89

(17)

Rappelons que (voir section 3.5.2 du polycopié sur ma page web)

$$s^2 = \frac{\sum_{i=1}^{n} (Y_i - \hat{y}_i)^2}{n - p - 1} \tag{18}$$

et (voir section 3.5.3 du polycopié sur ma page web)

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (Y_i - \hat{y}_i)^2}{n} \tag{19}$$

1. Régresser Y sur x_1 et tester la signification de cette régression.
2. Trouver l'équation de la régression multiple de Y sur x_1 et x_2. On donne la matrice $(X^tX)^{-1}$ dans ce cas[5] :

$$(X^tX)^{-1} = \begin{bmatrix} 0.4919 & -0.0074 & -0.0058 \\ -0.0074 & 0.0003 & -0.0001 \\ -0.0058 & -0.0001 & 0.0003 \end{bmatrix}.$$

3. Construire la régression multiple de Y sur x_1, x_2 et x_3. On donne la matrice $(X^tX)^{-1}$ dans ce cas :

$$(X^tX)^{-1} = \begin{bmatrix} 0.5344 & 0.0014 & -0.0057 & -0.0075 \\ 0.0014 & 0.0021 & -0.0001 & -0.0015 \\ -0.0057 & -0.0001 & 0.0003 & -0.0000 \\ -0.0075 & -0.0015 & -0.0000 & 0.0013 \end{bmatrix}$$

Réponse : encore, à nouveau, comme d'habitude, on calcule la valeur du vecteur

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = (X^tX)^{-1}X^tY$$

et on trouve

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} 19.9633 \\ 0.0314 \\ -0.2597 \\ 7.1873 \end{bmatrix}$$

La variance après calcul est donnée par 213.9229

4. Quel modèle est le plus pertinent au vu du critère AIC ? Il faut se référer à la section 3.8.2 du polycopié sur ma page web.

5. n'oubliez pas la colonne de "1" dans la matrice X !

5. Même question au vu du critère BIC (même section du poly de cours).

Exercice 7 Refaire les deux derniers exercices en utilisant les fonctions *R* appropriées.

Solution de l'exercice 5 :

```
x1 <- c(13,39,52,29,50,64,11,22,30, 51,27,41,51,37,23,43,38,62,28,30)
x2 <- c(18,18,50,43,37,19,27,23,18,11,15,22,52,36,48,15,19,56,30,25)
x3 <- c(25,59,62,50,65,79,17,31,34,58,29,53,75,44,27,65,62,75,36,41)
x4 <- c(11,30,53,29,56,49,14,17,22,40,31,39,36,27,20,36,37,50,20,33)
c1 <- rep(1,20) data <- cbind.data.frame(c1,x1,x2,x3,x4)
X <- as.matrix(data)
Y <- as.matrix(c(125,158,207,182,196,175,145,144,160,175,151,161,200,173,175,162,155,230,162,153))

XtX_inv <- solve((t(X)%*%X))
Beta_hat <- XtX_inv%*%t(X)%*%Y
pred <- c() for (i in 1 :nrow(X)){ pred <- append(pred, X[i,]%*%Beta_hat) } predbis = X%*%Beta_hat
numerateur <- 0
for (i in 1 :nrow(X)){ numerateur <- numerateur + (Y[i]-pred[i])^2}
s_carre <- numerateur/(20-4-1)
s_carrebis = sum(result$residuals**2)/(20-4-1) VarBetha_hat = s_carre*XtX_inv
result = lm(Y~ x1+x2+x3+x4)
```

Solution pour l'exercice 6 :

```
XY = rbind(c(124 , 33 , 8 , 81 ),
c( 49 , 31 , 6 , 55 ),
c(181 , 38 , 8 , 80 ),
c(4 , 17 , 2 , 24 ),
c(152 , 39 , 6 , 52 ),
c(75 , 30 , 7 , 88 ),
c(54 , 29 , 7 , 45 ),
c(43 , 35 , 6 , 50 ),
c(41 , 31 , 5 , 69 ),
c(17 , 23 , 4 , 66 ),
c(22 , 21 , 3 , 45 ),
c(16 , 8 , 3 , 24 ),
c(10 , 23 , 3 , 43 ),
c(63 , 37 , 6 , 38 ),
c(170 , 40 , 8 , 72 ),
c(15 , 38 , 6 , 41 ),
c(15 , 25 , 4 , 38 ),
c(221 , 39 , 7 , 52 ),
c(171 , 33 , 7 , 52 ),
c(97 , 38 , 6 , 66 ),
c(254 , 39 , 8 , 89 ))
X=XY[,c(1,2,3)]
Y=XY[,4]
x1=X[,1]
```

x2=X[,2]
x3=X[,3]

modele1 = lm(Y x3)
modele2 = lm(Y x1+x2)
modele3 = lm(Y x1+x2+x3)

summary(modele1)
summary(modele2)
summary(modele3)
n = 21
AICx3 = n*log(2*pi*var(modele1$residuals))+n+2*2 BICx3 = n*log(2*pi*var(modele1$residuals))+n+lo
AICx1x2 = n*log(2*pi*var(modele2$residuals))+n+2*3 BICx1x2 = n*log(2*pi*var(modele2$residuals))+r
AICx1x2x3 = n*log(2*pi*var(modele3$residuals))+n+2*3 BICx1x2x3 = n*log(2*pi*var(modele3$residual