République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITÉ MOHAMED KHIDER BISKRA

Faculté des Sciences Exactes et Sciences de la Nature et de la Vie
Département de Mathématiques

Première Année Master

Notes de Cours

Analyse de Données

Chapitre 2 : Régression linéaire multiple
(Séance 4)

Auteur des notes :

Dr. Sana BENAMEUR

Année universitaire : 2021-2022

Chapitre 2

Régression Linéaire Multiple

Le modèle de régression linéaire multiple constitue une généralisation de la régression linéaire simple, et qui sert à la mise en œuvre de l'étude des données multidimensionnelle (de dimension ≥ 2)

2.1 Modèle

Considérons une variable aléatoire (va) dépendante Y, qui l'on veut modéliser par $(X_1, X_2, \cdots, X_p)$ l'ensemble de p variables déterministes (indépendantes). Les données sont supposées d'un échantillon de taille n.

Définition 2.1 (Modèle de régression linéaire multiple)**.**
Le modèle de régression linéaire multiple se base sur l'écriture de chaque observation y_i, $i = \overline{1, n}$ $(n > p)$*par une équation de la forme :*

$$y_i = b_0 + \sum_{j=1}^{p} b_j x_{i,j} + \varepsilon_i, \tag{2.1}$$

où :

$x_{i,j}$ *sont des variables déterministes,*

les paramètres du modèle b_j $\left(j = \overline{0, p}\right)$ *sont des constantes inconnus,*

ε_i *sont des termes d'erreur d'une variable aléatoire* $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^t$.

Le modèle (2.1) prend la forme matricielle suivante :

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

avec : Y est un vecteur aléatoire de dimension n,

X est une matrice de dimension $(n, p+1) \in \mathbb{R}^n \times \mathbb{R}^{p+1}$, contient l'ensemble des observations sur les variables explicatives et la première colonne formée par la valeur

1 indique la constante b_0 dans l'équation, $\beta = (b_0, b_1, \cdots, b_p)^t$ est le vecteur de dimension $p+1$ des paramètres du modèle et ε est le vecteur des erreurs de dimension n. D'où

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\ 1 & x_{2,1} & x_{2,2} & \cdots & x_{2,p} \\ \vdots & \vdots & & \ddots & \vdots \\ 1 & x_{n,1} & x_{n,2} & \cdots & x_{n,p} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}. \tag{2.2}$$

Les hypothèses concernant le modèle sont :

(H_1) La matrice X est de plein rang, c-à-d.

$$rang(X) = p + 1.$$

(H_2) Les erreurs sont centrées, de même variance, et non corrélées entre elles, i.e

$$\mathbb{E}[\varepsilon] = 0_n, \; Var(\varepsilon) = \sigma_\varepsilon^2 I_n,$$

avec I_n la matrice identité d'ordre n.

$(H3)$ Les erreurs sont indépendantes des X_j $\left(j = \overline{1,p}\right)$.

2.2 Estimation des Paramètres

A partir de la connaissance des valeurs X_j, on estime les paramètres inconnues du modèle (le vecteur β et la variance σ_ε^2) par minimisation des moindres carrés.

2.2.1 Calcul des Estimateurs

On appelle estimateur des moindres carrés $\hat{\beta}$ de β la valeur suivante :

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} \varepsilon_i^2$$

On cherche alors la statistique $\hat{\beta} = \left(\hat{b}_0, \hat{b}_1, \cdots, \hat{b}_p\right)^t$ qui minimise

$$S(\beta) = S(b_0, b_1, \cdots, b_p) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left(y_i - b_0 - \sum_{j=1}^{p} b_j x_{i,j}\right)^2$$

Pour calculer $\hat{\beta}$, nous proposons l'écriture matricielle suivante :

$$\begin{aligned} S(\beta) &= \varepsilon^t \varepsilon \\ &= \|Y - X\beta\|^2 \\ &= (Y - X\beta)^t (Y - X\beta) \\ &= \left(Y^t - \beta^t X^t\right)(Y - X\beta) \\ &= Y^t Y - Y^t X\beta - \beta^t X^t Y + \beta^t X^t X\beta \end{aligned}$$

et puisque $Y^t X\beta = \beta^t X^t Y \in \mathbb{R}$, on obtient donc

$$S(\beta) = Y^t Y - 2\beta^t X^t Y + \beta^t X^t X\beta.$$

En dérivant $S(\beta)$ par rapport à chacun des paramètres $b_0, b_1, \cdots, b_p$, on obtient le système des équations :

$$\begin{aligned} -2X^t Y + 2\ X^t X\hat{\beta} &= 0 \\ &\Rightarrow X^t Y - X^t X\hat{\beta} = 0 \\ &\Rightarrow X^t Y = X^t X\hat{\beta} \\ &\Rightarrow \left(X^t X\right)^{-1} X^t Y = \left(X^t X\right)^{-1}\left(X^t X\right)\hat{\beta} \end{aligned}$$

d'où on tire finalement :

$$\hat{\beta} = \left(X^t X\right)^{-1} X^t Y. \tag{2.3}$$

2.2.2 Propriétés des Estimateurs

Proposition 2.1.

$\hat{\beta}$ est un estimateur sans biais de β, de matrice de variance-covariance

$$Var\left(\hat{\beta}\right) = \sigma_\varepsilon^2 \left(X^t X\right)^{-1}.$$

Démonstration.

Pour le biais il suffit d'écrire

$$\begin{aligned} \hat{\beta} &= \left(X^t X\right)^{-1} X^t Y \\ &= \left(X^t X\right)^{-1} X^t (X\beta + \varepsilon) \\ &= \left(X^t X\right)^{-1} X^t X\beta + \left(X^t X\right)^{-1} X^t \varepsilon \\ &= \beta + \left(X^t X\right)^{-1} X^t \varepsilon \end{aligned}$$

$$\mathbb{E}\left[\hat{\beta}\right] = \beta + \left(X^t X\right)^{-1} X^t \mathbb{E}\left(\varepsilon\right).$$

et puisque $\mathbb{E}(\varepsilon) = 0_n$, il vient

$$\mathbb{E}\left[\hat{\beta}\right] = \beta$$

Pour la variance, on procède de même

$$\begin{aligned}
Var\left(\hat{\beta}\right) &= \mathbb{E}\left[\left(\hat{\beta} - \mathbb{E}\left(\hat{\beta}\right)\right)\left(\hat{\beta} - \mathbb{E}\left(\hat{\beta}\right)\right)^t\right] \\
&= \mathbb{E}\left[\left(\hat{\beta} - \beta\right)\left(\hat{\beta} - \beta\right)^t\right] \\
&= \mathbb{E}\left[\left(X^t X\right)^{-1} X^t \varepsilon \left(\left(X^t X\right)^{-1} X^t \varepsilon\right)^t\right] \\
&= \mathbb{E}\left[\left(X^t X\right)^{-1} X^t \varepsilon \varepsilon^t X \left(X^t X\right)^{-1}\right] \\
&= \left(X^t X\right)^{-1} X^t \mathbb{E}\left(\varepsilon \varepsilon^t\right) X \left(X^t X\right)^{-1} \\
&= \left(X^t X\right)^{-1} X^t \sigma_\varepsilon^2 I_n X \left(X^t X\right)^{-1} \\
&= \sigma_\varepsilon^2 \left(X^t X\right)^{-1} X^t X \left(X^t X\right)^{-1} \\
&= \sigma_\varepsilon^2 \left(X^t X\right)^{-1}
\end{aligned}$$

□

Remarque 2.1.

1) *La matrice $X^t X$ est carrée d'ordre $p+1$, symétrique et inversible car X est de rang $p+1$.*

2) *$X^t X$ est définie positive.*

3) *$\hat{Y} = X\hat{\beta}$ est la valeur ajusté de Y.*

4) *Le vecteur des résidus $\hat{\varepsilon}$, tel que*

$$\hat{\varepsilon} = Y - \hat{Y} = Y - X\hat{\beta} = Y - X\left(X^t X\right)^{-1} X^t Y = \left[I_n - X\left(X^t X\right)^{-1} X^t\right] Y,$$

on pose $H = X\left(X^t X\right)^{-1} X^t$ et $P = I_n - H$, d'où

$$\hat{\varepsilon} = \left(I_n - H\right) Y = PY$$