
Introduction

En économétrie, quand on est en présence de plusieurs variables, le but n'est pas uniquement de les étudier en tant que variables aléatoires, mais essentiellement de modéliser leurs relations.

Cadre de ce cours : MODELES EXPLICATIFS

Expliquer une variable endogène y quantitative par p variables exogènes, dites explicatives.

Les données se présentent sous la forme du tableau suivant :

Variables	y endogène	x^1	...	x^p
u.s.1	y_1	x_1^1		x_1^p
.				
.				
.				
u.s.i	y_i	x_i^1		x_i^p
.				
.				
.				
u.s.n	y_n	x_n^1		x_n^p

u.s. : unité statistique représentant un individu d'une population ou le temps.

$x^1; \dots; x^p$: variables quantitatives ou qualitatives exogènes.

y : variable quantitative endogène.

Exemple 0..1. *On considère 101 succursales d'une entreprise. L'objectif est d'expliquer le chiffre d'affaire de chaque succursale au moyen de variables telles que : marge commerciale, charge salariale, nombre de vendeurs, . . . Les données sont stockées dans un fichier informatique sous la forme suivante :*

	Chiffre d'affaire		Marge Commerciale		Charges locatives		Charge salariale		
	CA HT	CA TTC	Marge	% CA HT	Taille	Loyer	% CA HT	Nbre Vendeurs	CA / vendeur
Agen	1181201	1441081	592101	50.127	10.001	112201	9.498	2501	472.29
Amboise	700201	954241	355901	50.828	10001	58501	8.354	1501	466.49
...									
...									
Perpignan	1079101	1318501	505201	46.816	95301	129401	11.991	2701	399.519

Exemple 0..2. :

On se propose d'étudier la relation, qu'on suppose linéaire si elle existe, entre le prix et les variables suivantes : cylindrée, puissance, longueur, largeur, poids et vitesse de 18 voitures figurant dans le tableau :

		x^2	x^3	x^4	x^5	x^6	x^7	x^8	y
OBS	NOM	CYL	PUIS	LON	LAR	POIDS	VITESSE	FINITION	PRIX
1	ALFASUD-TI-1350	1350	79	393	161	870	165	B	30570
2	AUDI-100-L	1588	85	468	177	1110	160	TB	39990
3	SIMCA-1307-GLS	1294	68	424	168	1050	152	M	29600
4	CITROEN-CG-CLUB	1222	59	412	161	930	151	M	28250
5	FIAT-132-1600GLS	1585	98	439	164	1105	165	B	34900
6	LANCIA-BETA-1300	1297	82	429	169	1080	160	TB	35480
7	PEUGEOT-504	1796	79	449	169	1160	154	B	32300
8	RENAULT-16-TL	1565	55	424	163	1010	140	B	32000
9	RENAULT-30-TS	2664	128	452	173	1320	180	TB	47700
10	TOYOTA-COROLLA	1166	55	399	157	815	140	M	26540
11	ALFETTA-1.66	1570	109	428	162	1060	175	TB	42395
12	PRINCESS-1800-HL	1798	82	445	172	1160	158	B	33990
13	DATSUN-200L	1998	115	469	169	1370	160	TB	43980
14	TAUNUS-2000-GL	1993	98	438	170	1080	167	B	35010
15	RANCHO	1442	80	431	166	1129	144	TB	39450
16	MAZDA-9295	1769	83	440	165	1095	165	M	27900
17	OPEL-REKORD-L	1979	100	459	173	1120	173	B	32700
18	LADA-1300	1294	68	404	161	955	140	M	22100

$$y_i \beta_1 + \beta_2 x_i^2 + \dots + \beta_8 x_i^8 + e_i$$

Exemple 0..3. :

Modèle économique du comportement de production des entreprises :

$$Q = \alpha \prod_{j=2}^p (x^j)^{\beta_j} \quad (1)$$

(Fonction de Cobb-Douglas)

- var y à expliquer : Q "quantité de production"

- var $x^2 ; \dots ; x^p$ explicatives :

K : Capital,

L : Travail,

E : Energie,

M : Matières premières,

T : Progrès techniques.

s'écrit : $\log Q = \beta_1 + \sum_{j=2}^P \beta_j \log x^j$ (en posant $\log \alpha = \beta_1$)

1. Etude statistique de ce modèle

1. Soit à partir de données statistiques observées pour une entreprise au cours du temps t :

$$\log Q_t = \sum_{j=2}^P \beta_j \log x_t^j + e_t \quad (t = 1, n), \text{ où } n \text{ est le nombre d'unité de temps,}$$

avec des hypothèses sur les v.a. e_t , qui diffèrent selon les 2 cas suivants :

Cas 1. les variables x_t^j ne contiennent pas de valeurs passées de Q_t : modèle explicatif statique.

Cas 2. sinon, modèle explicatif dynamique et dans ce cas les résidus e_t sont des v.a. corrélées.

2. Soit à partir d'un panel de n entreprises dans une période donnée :

$$\log Q_i = \sum_{j=2}^P \beta_j \log x_i^j + e_i \quad (i = 1. \dots, n) \text{ où } n \text{ est le nombre d'entreprises.}$$

Objectif : Estimer les paramètres :

- $\beta_j = \frac{\partial \log Q}{\partial \log x^j}$ élasticité.

- $r = \sum \beta_j$ rendement.

Exemple 1..1. .

En Macroéconomie, modèle dynamique explicatif :

$$\begin{array}{ccccccc} R_t & = & \beta_1 + \beta_2 & M_t & + \beta_3 & I_t & + \beta_4 & U_t + e_t \\ \downarrow & & & \downarrow & & \downarrow & & \downarrow \\ \text{taux} & & & \text{masse} & & \text{taux} & & \text{taux} \\ \text{d'intérêt} & & & \text{monétaire} & & \text{d'inflation} & & \text{de chômage} \\ \text{à long terme} & & & & & & & \end{array}$$

Les résidus e_t sont des v.a. qui seront supposées :

$E(e_t) = 0$ "centrées"

$V(e_t) = \delta_t^2$ hétéroscédastiques

$E(e_t e_{t-1}) \neq 0$ auto-corrélées.

OBJECTIF DU COURS

A partir de n observations de $(y; x)$ $y \in IR$ et $x \in IR^p$, on cherche à expliquer la variable y par une combinaison linéaire des variables $x^1; \dots; x^p$, composantes de x . Ce problème sera abordé différemment selon les cas :

- La variable x est non aléatoire.
- La variable $(x; y)$ est aléatoire
- On considère la loi de $(y; x)$.
- On considère la loi conditionnelle de y sachant $x = x_0$.
- Les variables $x^1; \dots; x^p$ sont quantitatives ou qualitatives.
- Les données sont des séries temporelles ou des données de panel.

linéaire simple et régression linéaire multiple

Cadre général quand les variables exogènes sont non aléatoires. Autre terminologie :

RLM (plusieurs v. exogènes quantitatives) : Régression Linéaire Multiple

RLS (une seule v. exogène quantitative) : Régression Linéaire Simple

1. Formulation du modèle linéaire

Données : Pour $i = 1, \dots, n$

y_i	x_i^1, \dots, x_i^p	$n > p$
⏟	⏟	
endogène(quantitative)	exogènes	

Modèle linéaire :

$$y_i = \beta_1 x_i^1 + \dots + \beta_p x_i^p + \underbrace{\epsilon_i}_{\substack{\text{résidu, bruit,} \dots \\ \text{(terme d'erreur)}}} \quad (i = 1, \dots, n) \quad (1)$$

Remarque 1..1. .

- On notera y indifféremment la variable aléatoire y ou la valeur observée y .
- La relation (1) est stochastique :
 y_i observation d'une v.a. y
 ϵ_i valeur non observable d'une v.a. ϵ
- On admet que $x_i^1 = 1$ pour tout $i = 1, \dots, n$ si on considère un terme constant dans (1).
- Les coefficients de régression

$\beta_j (j = 1, \dots, p)$ sont les p paramètres inconnus, à estimer.

– La relation (1) équivaut à considérer que la variable y peut être approximée par une fonction affine des p variables x_1, \dots, x_p (à une erreur ϵ près).

1.1 Écriture matricielle

$$\begin{array}{ccccccc} y & = & X & \beta & + & \epsilon & (2) \\ (n \times 1) & & (n \times p) & (p \times 1) & & (n \times 1) & \end{array}$$

1.2 Hypothèses du modèle linéaire

Le modèle linéaire relatif aux donnée $y = X$ est défini par les 3 hypothèses :

H_1 $y = X\beta + \epsilon$ (relation (2) de linéarité)

H_2 Sur la matrice X :

X est non aléatoire

$x_i^1 = 1$ pour tout $i = 1, \dots, n$

$rg(X) = p$

H_3 Sur les résidus :

ϵ est un vecteur aléatoire de R^n tel que :

$$\begin{aligned} E(\epsilon) &= 0 \quad (\text{centré}) \\ V(\epsilon)_{(n \times n)} &= \sigma^2 I_n \text{ Ou } \sigma > 0 \end{aligned}$$

Remarque 1..2. :

- Hypothèse H_2 :

$rg(X) = p \iff$ Les p vecteurs colonnes x^1, \dots, x^p sont linéairement indépendants. Donc $n > p$.

- Hypothèse H_3 :

$E(\epsilon) = 0$ et $V(\epsilon) = \sigma^2 I_n$

\leftrightarrow Les résidus $\epsilon_1, \dots, \epsilon_n$ sont des v.a.r :

- centrées $E(\epsilon_i) = 0$

- de même variance $V(\epsilon_i) = \sigma^2$

(HOMOSCEDASTICITÉ)

- Paramètres à estimer :

$$\beta(p \times 1) = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} \text{ et } \sigma^2$$

1.3 Modélisation de la régression linéaire simple

Exemple

Pour des raisons de santé publique, on s'intéresse à la concentration d'ozone O₃ dans l'air (en microgrammes par millilitre). En particulier, on cherche à savoir s'il est possible d'expliquer le taux maximal d'ozone de la journée par la température T12 à midi. Les données sont :

Température à 12h	23.8	16.3	27.2	7.1	25.1	27.5	19.4	19.8	32.2	20.7
O ₃ max	115.4	76.8	113.8	81.6	115.4	125	83.6	75.2	136.8	102.8

FIGURE 1.1 – 10 données journalières de température et d'ozone.

D'un point de vue pratique, le but de cette régression est double :

- ajuster un modèle pour expliquer O₃ en fonction de T12 ;
- prédire les valeurs d'O₃ pour de nouvelles valeurs de T12.

Avant toute analyse, il est intéressant de représenter les données, comme sur la figure suivante

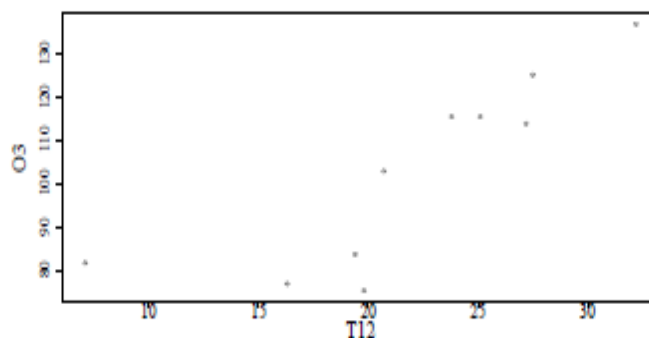


FIGURE 1.1 – 10 données journalières de température et d'ozone.

Dans de nombreuses situations, en première approche, une idée naturelle est de supposer que la variable à expliquer y est une fonction affine de la variable explicative x , c'est-à-dire de

chercher f dans l'ensemble \mathcal{F} des fonctions affines de R dans R . C'est le principe de la régression linéaire simple. On suppose dans la suite disposer d'un échantillon de n points (x_i, y_i) du plan.

Définition 1..1. *Un modèle de régression linéaire simple est défini par une équation de la forme :*

$$y_i = \beta_1 + \beta_2 x_i + \epsilon_i \quad (i = 1, \dots, n) \quad (2)$$

Les quantités ϵ_i viennent du fait que les points ne sont jamais parfaitement alignés sur une droite. On les appelle les erreurs (ou bruits) et elles sont supposées aléatoires. Sous les hypothèses $H3$, les erreurs sont donc supposées centrées, de même variance (homoscédasticité) et non corrélées entre elles.

Notons que le modèle de régression linéaire simple de la définition 1.1 peut encore s'écrire de façon vectorielle :

$$Y = \beta_1 \mathbf{1} + \beta_2 X + \epsilon,$$

où :

- le vecteur $Y = [y_1, \dots, y_n]'$ est aléatoire de dimension n ,
- le vecteur $\mathbf{1} = [1, \dots, 1]'$ est le vecteur de R^n dont les n composantes valent toutes 1,
- le vecteur $X = [x_1, \dots, x_n]'$ est un vecteur de dimension n donné (non aléatoire),
- les coefficients β_1 et β_2 sont les paramètres inconnus (mais non aléatoires) du modèle,
- le vecteur $\epsilon = [\epsilon_1, \dots, \epsilon_n]'$ est aléatoire de dimension n .

2. ESTIMATEUR DES MOINDRES CARRÉS ORDINAIRES (MCO)

OLSE : "Ordinary Least Square Estimator"

2.1 Définition de l'estimateur MCO du paramètre β

Notation

$$SS(\beta) = \| \beta \|_{I_n}^2 = \epsilon' \epsilon = \sum_{i=1}^n \epsilon_i^2$$

La somme des carrés des résidus $SS(\beta)$ s'écrit encore :

$$SS(\beta) = \| y - X\beta \|_{I_n}^2 = \sum_{i=1}^n [y_i - \sum_{j=1}^p x_{ij}\beta_j]^2$$

Définition 2..1. L'estimateur MCO du paramètre β

est le vecteur aléatoire de IR^p qui minimise la somme des carrés des écarts $SS(\beta)$.

C'est-à-dire :

$$\hat{\beta}_{MCO} = \arg_{\beta \in R^p} \min SS(\beta) = \|y - X\beta\|_{I_n}^2$$

Théorème 2..1. Sous les hypothèses H_1 et H_2 l'estimateur MCO, noté $\hat{\beta}$ ou $\hat{\beta}_{MCO}$ est :

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

Démonstration : La fonction à minimiser $SS(\beta)$ s'écrit :

$$SS(\beta) = y'y - 2\beta'X'y + \beta'X'X\beta$$

Le vecteur $\hat{\beta}$ qui minimise $SS(\beta)$ est solution de :

$$\frac{\partial SS(\hat{\beta})}{\partial \beta} = 0 \iff -2X'y + 2X'X\hat{\beta} = 0 \quad (3)$$

Justification (cf. Chapitre 0) :

$$\begin{aligned} \frac{\partial u'a}{\partial u} &= a \\ \frac{\partial u'Au}{\partial u} &= 2Au \end{aligned}$$

Or $rg(X) = p \implies X'X$ inversible car de rang p .

Donc (3) $\iff X'X\hat{\beta} = X'y \iff \hat{\beta} = (X'X)^{-1}y$

$$\frac{\partial^2 SS(\beta)}{(\partial \beta)(\partial \beta')} = 2X'X \text{ matrice définie positive (cf. exercice)}$$

$\implies \beta$ est un minimum unique.

Autre démonstration : $SS(\beta)$ peut se décomposer (cf. exercice).

$$SS(\beta) = \underbrace{y'y - y'X(X'X)^{-1}X'y}_{\text{ne dépend pas de } \beta} + \underbrace{[\beta - (X'X)^{-1}X'y]'(X'X)[\beta - (X'X)^{-1}X'y]}_{\substack{\text{de la forme } u'(X'X)u \text{ qui s'annule} \\ \text{ssi } u = 0 \text{ (car } X'X \text{ définie positive)}}$$

Donc le minimum de $SS(\beta)$ est atteint pour $u = 0 \iff \hat{\beta} = (X'X)^{-1}X'y$.

Notation : On appelle somme des carrés résiduelle (ou Residual Sum Squares) le minimum de la fonction $SS(\beta)$ qui est noté SCE ou RSS.

$$\begin{aligned} RSS &= SS(\hat{\beta}) = \|y - X\hat{\beta}\|_{I_n}^2 \\ &= y'y - y'X(X'X)^{-1}X'y \end{aligned}$$

2.2 Estimation dans le modèle simple

Définition 2..2. On appelle estimateurs des Moindres Carrés Ordinaires (en abrégé MCO) $\hat{\beta}_1$ et $\hat{\beta}_2$ les valeurs minimisant la quantité $SS(\beta_1, \beta_2) = \sum_{i=1}^n (y_i - \beta_1 - \beta_2 x_i)^2$.

Autrement dit, la droite des moindres carrés minimise la somme des carrés des distances verticales des points (x_i, y_i) du nuage à la droite ajustée $y = \hat{\beta}_1 + \hat{\beta}_2 x$.

La fonction de deux variables SS est une fonction quadratique et sa minimisation ne pose aucun problème, comme nous allons le voir maintenant.

Proposition 2..2. Les estimateurs des MCO ont pour expressions :

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$$

avec,

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

L'expression de $\hat{\beta}_2$ suppose que le dénominateur $\sum_{i=1}^n (x_i - \bar{x})^2$ est non nul. Or ceci ne peut arriver que si tous les x_i sont égaux, situation sans intérêt pour notre problème et que nous excluons donc a priori dans toute la suite.

Démonstration : ...

Exemple (suite)

- Expliquer le taux maximal d'ozone O_3max de la journée par la température T_{12} à midi par la méthode de MCO.

2.3 Propriétés de l'estimateur MCO $\hat{\beta}$

Sous les hypothèses H_1 , H_2 et H_3 , voici quelques propriétés statistiques de l'estimateur $\hat{\beta}$.

Théorème 2..3. *L'estimateur $\hat{\beta} = (X'X)^{-1}X'y$ est un estimateur sans biais de β , de variance $V(\hat{\beta}) = \sigma^2(X'X)^{-1}$.*

Démonstration : ...

Remarque 2..1. : *La propriété d'optimalité pour un estimateur sans biais est d'avoir une variance minimum. Cette propriété est énoncée pour $\hat{\beta}$ dans le théorème suivant.*

Théorème 2..4. (de Gauss Markov).

L'estimateur $\hat{\beta}$ est meilleur que tout estimateur $\bar{\beta}$ linéaire et sans biais de β (c'est-à-dire $V(\bar{\beta}) - V(\hat{\beta})$ matrice semi-définie positive).

Démonstration : Soit $\bar{\beta}$ un estimateur linéaire et sans biais de $\beta \implies A$ telle que $\bar{\beta} = Ay$ et $E(\bar{\beta}) = \beta$

$$\begin{aligned} \forall u \in R^p, u'[V(\bar{\beta}) - V(\beta)]u &\geq 0 u'[V(\bar{\beta}) - V(\beta)].u \\ &= u'[AVV(y)A' - \sigma^2(X'X)^{-1}].u \\ &= \sigma^2.u'A[I_n - X(X'X)^{-1}X']AU' \end{aligned}$$

car :

$$\begin{aligned} E(u'\bar{\beta}) &= u'AE(y) = u'AX\beta \\ E(u'\bar{\beta}) &= u'E(\beta) = u'\beta \implies (u'AX - u')\beta = 0 \implies u'AX = u' \text{ Or } M = I_n - X(X'X)^{-1}X' \text{ est} \\ &\text{une matrice symétrique et idempotente (cf. exercice).} \end{aligned}$$

Donc :

$$\begin{aligned} \sigma^2 &= 2u'AMA'u = \sigma^2 \|A'u\|_{I_n}^2 \\ &= \sigma^2(MA'u)'MA'u = \sigma^2 \|MA'u\|_{I_n}^2 \geq 0 \end{aligned}$$

Donc $V(\bar{\beta}) - V(\beta)$ est une matrice semi-définie positive. Donc $\hat{\beta}$ a une variance minimum dans la classe des estimateurs linéaires et sans biais de β .

2.4 Estimation de la variance résiduelle σ^2

Théorème 2..5. La statistique $\hat{\sigma}^2 = \frac{\|y - X\hat{\beta}\|^2}{n-p}$ est un estimateur sans biais de σ^2

Démonstration :

$$\begin{aligned} E(\hat{\sigma}^2) &= \frac{1}{n-p} \cdot E(\|y - X\hat{\beta}\|_{I_n}^2) \\ &= \frac{1}{n-p} \cdot E(y'My) \text{ ou } M = I_n - X(X'X)^{-1}X' \\ &= \frac{1}{n-p} \cdot E[\text{tr}(My y')] = \frac{1}{n-p} \text{tr}(M \cdot E(yy')) \end{aligned}$$

$$\begin{aligned} \text{or : } \begin{cases} \text{var } y = E[(y - X\beta)(y - X\beta)'] = E(yy') - X\beta\beta'X' \\ \text{Var } y = \sigma^2 I_n \end{cases} &\implies E(yy') = \sigma^2 I_n + X\beta\beta'X' \\ \text{tr}(M \cdot E(yy')) &= \text{tr}(\sigma^2 M) + \text{tr}(MX\beta\beta'X') \text{ car } MX = [I_n - X(X'X)^{-1}X']X = 0 \end{aligned}$$

$$\begin{aligned} \implies E(\hat{\sigma}^2) &= \frac{1}{n-p} \sigma^2 \text{tr } M \\ \text{or } \text{tr } M &= \text{tr}(I_n - X(X'X)^{-1}X') \\ &= \text{tr} I_n - \text{tr}(X(X'X)^{-1}X') \\ &= n - \text{tr}(X'X(X'X)^{-1}) = n - p \end{aligned}$$

$\implies E(\hat{\sigma}^2) = \sigma^2$. Donc $\hat{\sigma}^2$ est un estimateur sans biais de σ^2

Remarque 2..2. : $\hat{\sigma}^2 = \frac{\|y - X\hat{\beta}\|_{I_n}^2}{n-p} = \frac{RSS}{n-p} = \sum_{i=1}^n \hat{\epsilon}_i^2$ ou $\hat{\epsilon}_i = y_i - \hat{y}_i = y_i - \sum_{j=1}^p x_i^j \hat{\beta}_j$
 $\hat{\epsilon}_i$ est le résidu estimé et \hat{y}_i est la valeur ajustée par le modèle de régression.

Estimateur de $V(\hat{\beta})$

$$\widehat{V(\hat{\beta})} = \hat{\sigma}^2 (X'X)^{-1} = \frac{\|y - X\hat{\beta}\|^2}{n-p}$$

En particulier :

L'écart-type estimé de l'estimateur $\hat{\beta}_j$ de chaque coefficient de régression β_j est :

$$\hat{\sigma}_{\hat{\beta}_j} = \sqrt{\hat{\sigma}^2 \cdot [(X'X)^{-1}]_{jj}} \text{ qui est appelé erreur standard de } \hat{\beta}_j, \text{ noté e.s. } (\hat{\beta}_j).$$

Le terme $[(X'X)^{-1}]_{ij}$ représente le terme général de la matrice $(X'X)^{-1}$.

3. Interprétation géométrique

3.1 Représentation des individus

Chaque individu $i (i = 1, \dots, n)$ est considéré dans l'espace IR^p par ses p coordonnées : $(y_i, x_i^2, \dots, x_i^p) \in IR^p$ (on ne considère pas $x_i^1 = 1$)

L'ensemble des n points de IR^p est appelé “nuage de points” dans IR^p .

\implies Représentation graphique possible si $p = 2$ (RLS).

Elle permet :

- de visualiser le type de liaison entre y et x^2
 - d'apprécier les distances entre individus
- \rightarrow Si $p > 3$, représentation graphique difficile à utiliser) biplot de y en fonction de x^j .

3.2 Espace des variables

Les vecteurs

$$y = \begin{pmatrix} y_1 \\ . \\ . \\ . \\ y_n \end{pmatrix} \quad \text{et} \quad x^j = \begin{pmatrix} x_1^j \\ . \\ . \\ . \\ x_n^j \end{pmatrix}$$

pour $j = 1, \dots, p$ sont des vecteurs d'un espace vectoriel $E \in IR^n$.

Notation : Soit $\chi = \text{vect}\{x^1, \dots, x^p\}$ l'espace engendré par les vecteurs x^1, \dots, x^p

Proposition 3..1. La matrice $P = X(X'X)^{-1}X'$ est un opérateur de projection I_n -orthogonal sur χ .

Démonstration : (exercice).

Remarque 3..1. :

a. L'estimateur MCO $\hat{\beta} = (X'X)^{-1}X'y$ correspond à la projection de y sur χ :

$$\hat{y} = Py = X(X'X)^{-1}X'y = X\hat{\beta} = \sum_{j=1}^p \hat{\beta}_j x^j$$

Les coefficients $\hat{\beta}_j$ estimés sont les coefficients de la combinaison linéaire des vecteurs $x^j (j = 1, \dots, p)$.

b. $\sum_{i=1}^n \hat{y}_i = \sum_{i=1}^n y_i$ (exercice)

Proposition 3..2. Le vecteur colonne unité de \mathbb{R}^n étant noté $I_n = x^1$ la projection de y sur la droite des constantes (sous-espace engendré par x^1) est $\bar{y}I_n$.

c'est-à-dire :

$$I_n(I_n' I_n)^{-1} I_n' y = \bar{y} I_n$$

Graphique :

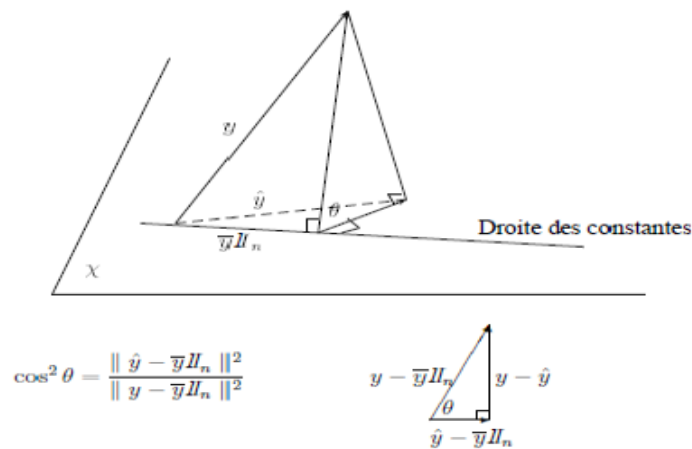


FIGURE 1.2 – Représentation de la projection dans l'espace des variables $p > 2$

Proposition 3..3. Analyse de la variance de la régression

$$\|y - \bar{y}I_n\|^2 = \|y - X\hat{\beta}\|^2 + \|X\hat{\beta} - \bar{y}I_n\|^2$$

c'est-à-dire :

$$\underbrace{\sum_{i=1}^n (y_i - \bar{y})^2}_{TSS \text{ totale}} = \underbrace{\sum_{i=1}^n (y_i - \hat{y}_i)^2}_{RSS \text{ résiduelle}} + \underbrace{\sum_{i=1}^n \hat{y}_i - \bar{y}}_{\text{variation expliquée par la régression}}$$

3.3 Coefficient de corrélation multiple (empirique) ou coefficient de détermination

Définition 3..1. : C'est le nombre défini par $R^2 = \cos^2 \theta$

$$R^2 = \frac{\| \hat{y}_i - \bar{y} I_n \|^2}{\| y_i - \bar{y} I_n \|^2} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Interprétation : C'est le rapport de la variation expliquée par la régression, sur la variation totale.

- Si $R^2 = 1 \iff \theta = 0 \iff y \in \chi \iff y = X\beta$
- Si $R^2 \simeq 0 \iff \theta \simeq 90$ résidus très élevés \iff modèle de régression linéaire inadapté.

Remarque 3..2. Si $p = 2$, le coefficient R^2 n'est autre que le carré du coefficient de corrélation linéaire empirique de Pearson (exercice).

Exemple 3..1. Pour $p = 2$, cinq ensembles de données ($n = 16$) ont la même valeur $R^2 = 0.617$. Et pourtant le modèle linéaire n'est pas toujours satisfaisant (voir figure 1.3).

Conclusion : Le critère R^2 d'ajustement à un modèle linéaire est insuffisant.

Exemple 3..2. Réf. : SAPORTA G. (1990), "Probabilités, analyse des données et statistique", Editions Technip.

A. Tableau d'analyse de variance de la régression

	d.d.l.	Somme de carrés	Carré moyen	F	Sig F
Régression	6	520591932,37	86765322,06	4,4691	0,0156
Résiduelle	11	213563857,91	19414896,17		
Totale	17	734155790,28			

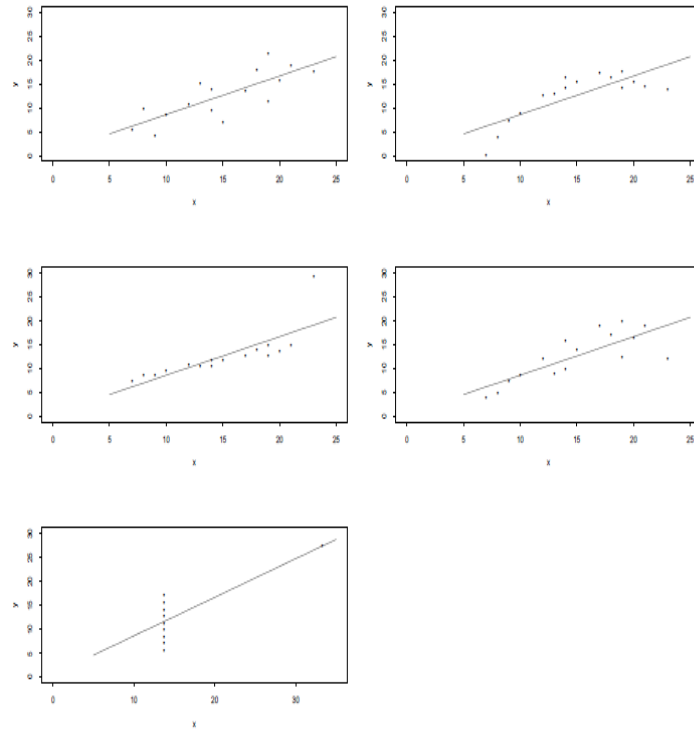


FIGURE 1.3 – Graphique de 5 ensembles de données conduisant au même $R^2 = 0.617$. Réf. : TOMASSONE R., LESQUOY E., MILLIER C. (1983), “La régression : nouveaux regards sur une ancienne méthode statistique”, Masson..

$$R^2 = 0.7091 R^2_{\text{ajusté}} = 0.5504 \quad \hat{\sigma} = 4406.23.$$

B Estimation des paramètres

Variable	Coefficient estimé	Ecart- type	T si H_0 : coeff. = 0	Sig T
Constante	-8239,362677	42718,42313	-0,193	0,8506
Cylindrée	-3,505182	5,550595	-0,631	0,5406
Largeur	208,693773	412,047884	0,506	0,6225
Longueur	-15,037660	129,747488	-0,116	0,9098
Poids	12,574678	24,622191	0,511	0,6197
Puissance	282,168803	174,882966	1,613	0,1349
Vitesse	-111,113551	222,256575	-0,500	0,6270

C Etude des résidus estimés

		Prix	Prix estimé
1	ALFASUD-TI-1350	30570	29616,1
2	AUDI-100-L	39990	36259,7
3	SIMCA-1307-GLS	29600	31411,1
4	CITROEN-CG-CLUB	28250	26445,8
5	FIAT-132-1600GLS	34900	37043,0
6	LANCIA-BETA-1300	35480	34972,8
7	PEUGEOT-504	32300	33749,1
8	RENAULT-16-TL	32000	26580,0
9	RENAULT-30-TS	47700	44445,6
10	TOYOTA-COROLLA	26540	24650,2
11	ALFETTA-1,66	42395	38270,5
12	PRINCESS-1800-HL	33990	34830,4
13	DATSUN-200L	43980	44872,4
14	TAUNUS-2000-GL	35010	36343,5
15	RANCHO	39450	35638,1
16	MAZDA-9295	27900	32233,4
17	OPEL-REKORD-L	32700	37103,5
18	LADA-1300	22100	30389,8

D Calcul du R^2 pour différents modèles

k	Modèle	R^2	$\hat{\sigma}$
1	Puis.	0.638	4076.0
2	Puis., Poids	0.686	3916.4
3	Cyl., Puis., Poids	0.699	3974.4
4	Cyl., Puis., Larg., Poids	0.702	4103.7
5	Cyl., Puis., Larg., Poids, Vitesse	0.709	4221.2
6	Complet	0.709	4406.2

c) **Autres critères :** (que l'on justifiera dans les chapitres suivants.)

- $R_{aj}^2 = \frac{(n-1)R^2 - (p-1)}{n-p}$ coefficient de corrélation multiple (empirique) ajusté.

- $F = \frac{R^2}{1-R^2} \cdot \frac{n-p}{p-1} = \frac{\|y_i - \bar{y}I_n\|^2 / (p-1)}{\|\hat{y}_i - y\|^2 / (n-p)}$ Statistique de Fisher (cf. Chapitre 2), c'est la statistique du test d'égalité à zéro des $(p-1)$ coefficients $\beta_2 = \beta_3 = \dots = \beta_p = 0$

- $\hat{\sigma}^2 = \frac{\|y - \hat{y}\|^2}{n-p}$ estimateur sans biais de la variance résiduelle σ^2 et qui s'exprime en fonction de R_{aj}^2 :

$$\hat{\sigma}^2 = (1 - R_{aj}^2) \frac{\|y - \bar{y}I_n\|^2}{n-1} \quad (\text{exercice}).$$

Induction statistique dans le modèle linéaire

Si on ajoute aux hypothèses du modèle linéaire une hypothèse de normalité des résidus, il est possible de réaliser des tests et des estimations par intervalle de confiance.

1. Modèle linéaire gaussien

Définition 1..1. : Rappelons le contexte du chapitre précédent. Nous avons supposé un modèle de la forme :

$$\begin{matrix} y & = & X & \beta & + & \epsilon & (2) \\ (n \times 1) & & (n \times p) & (p \times 1) & & (n \times 1) \end{matrix}$$

avec les hypothèses H_1), H_2) et H'_3). Dans chapitre, comme ce fut le cas en fin de Chapitre 1, nous allons faire une hypothèse plus forte, à savoir celle de gaussianité des résidus. Nous supposons donc désormais :

$$H_1) \quad y = X\beta + \epsilon \quad (\text{cf. Chapitre 2})$$

$$H_2) \quad X \text{ matrice non aléatoire de rang } p \text{ (avec } x_i^1 = 1)$$

$$H'_3) \quad \epsilon \sim N_n(0, \sigma^2 I_n)$$

Remarque 1..1. : L'hypothèse H'_3 contient les hypothèses H_3 du modèle linéaire :

$$E(\epsilon) = 0$$

$$V(\epsilon_i = \delta^2) \text{ homoscédasticité}$$

$$\text{cov}(\epsilon_i, \epsilon_j) = 0 \text{ pour } i \neq j$$

L'hypothèse $H'_3 \iff$ les v.a.r. $\epsilon_1, \dots, \epsilon_n$ sont i.i.d. de loi $(N_n(0, \sigma^2))$

Autre formulation :

$$H_1 \text{ et } H_3' \iff y \sim N_n(X\beta, \sigma^2 I_n)$$

structure statistique paramétrique :

$$\underbrace{\{IR^n, B_{IR^n}, N_n(X\beta, \sigma^2 I_n), (\beta, \sigma) \in IR^p \times IR^+\}}_{\text{Univers} \quad \text{Tribu} \quad \text{Probabilité indexée par } (\beta, \sigma)}$$

1.1 Estimateur maximum de vraisemblance EMV ou MLE

Nous allons commencer par faire le lien entre l'estimateur du maximum de vraisemblance et l'estimateur des moindres carrés vu au chapitre précédent.

Proposition 1..1. : "estimateur MV du paramètre β
est égal à l'estimateur MCO :

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

Démonstration : ...

Notation : Dans tout le chapitre $\hat{\beta}$ désigne $\hat{\beta}_{MCO} = \hat{\beta}_{MV}$.

Remarque 1..2.

1) L'estimateur MV du paramètre σ^2 est $\frac{\|y - X\hat{\beta}\|^2}{n}$

Démonstration : (en exercice)

C'est un estimateur asymptotiquement sans biais et convergent en probabilité.

$\Rightarrow \hat{\sigma}^2 = \frac{\|y - X\hat{\beta}\|^2}{n-p}$ estimateur sans biais (cf. Chapitre 1).

2) Les estimateurs $\hat{\beta}$ et $\hat{\sigma}^2$ sont des estimateurs efficaces.

Démonstration : (en exercice)

1.2 Loi de probabilité des estimateurs

Proposition 1..2. : L'estimateur $\hat{\beta} = (X'X)^{-1}X'y$ est un vecteur aléatoire gaussien de IR^p

$$\hat{\beta} \sim \mathcal{N}_p(\beta, \sigma^2(X'X)^{-1})$$

Démonstration : (cf. Chapitre 0, III), Combinaison linéaire de v.a. indépendantes gaussiennes.

Proposition 1..3. La v.a.r. $\frac{\|y - X\hat{\beta}\|^2}{\sigma^2}$ suit une loi de chi-deux à $(n - p)$ d.d.l.

c'est-à-dire :

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

Démonstration :

On a ;

$$\|y - X\hat{\beta}\|^2 = (y - X\beta)'M(y - X\beta)$$

$$\text{avec } M = I_n - X(X'X)^{-1}X' \text{ (rang } M = n-p)$$

$$\text{d'où } \|y - X\hat{\beta}\|^2 = \epsilon' M \epsilon \text{ Car } \hat{\epsilon} = M\epsilon$$

D'après le théorème de COCHRAN (chapitre 0)

$$\begin{cases} \frac{1}{\sigma} \cdot \epsilon \sim \mathcal{N}_n(0, I_n) \\ M \text{ matrice symétrique idempotente de rang } n - p \end{cases} \implies \frac{1}{\sigma^2} \epsilon' M \epsilon \sim \chi_{(n-p)}^2$$

$$\text{Donc } \frac{1}{\sigma^2} \|y - X\hat{\beta}\|^2 \sim \chi_{(n-p)}^2$$

Remarque 1..3. :

- On retrouve

$$E(\hat{\sigma}^2) = E\left(\frac{\|y - X\hat{\beta}\|^2}{n - p}\right) = \frac{\sigma^2}{n - p} \times \underbrace{E\left(\frac{1}{\sigma^2} \|y - X\hat{\beta}\|^2\right)}_{n-p}$$

$$E(\hat{\sigma}^2) = \sigma^2$$

$$\text{en effet } E(Z) = v \text{ si } Z \sim \chi_v^2$$

- On calcule $V(\hat{\sigma}^2) = \frac{2\sigma^2}{n-p}$ En effet $V(Z) = 2v$ si $Z \sim \chi_v^2$

$$\begin{aligned} \Rightarrow V(\hat{\sigma}^2) &= V\left(\frac{\sigma^2}{n - p} \times \frac{\|y - X\hat{\beta}\|^2}{\sigma^2}\right) \\ &= \frac{\sigma^4}{(n - p)^2} \times V\left(\frac{\|y - X\hat{\beta}\|^2}{\sigma^2}\right) \\ &= \frac{\sigma^4}{(n - p)^2} \times 2(n - p) \\ &= \frac{2\sigma^4}{n - p}. \end{aligned}$$

Proposition 1..4. :

Les v.a. $\hat{\beta}$ et $\hat{\sigma}^2$ sont indépendantes.

Démonstration : On considère les vecteurs gaussiens

$$y - X\hat{\beta} = My = M\epsilon \text{ de } R^n \text{ et } (X'X)^{-1}X'\epsilon \text{ de } R^n$$

Ils sont centrés car $E(\epsilon) = 0$.

Calcul de la matrice de covariance de ces 2 vecteurs (de dimension $n \times p$.)

$$\begin{aligned} \text{cov}(y - X\hat{\beta}, (X'X)^{-1}X'\epsilon) &= \text{cov}(M\epsilon, (X'X)^{-1}X'\epsilon) \\ &= E[M\epsilon((X'X)^{-1}X'\epsilon)'] - 0 \\ &= E[M\epsilon\epsilon'X(X'X)^{-1}] \\ &= ME(\epsilon\epsilon')X(X'X)^{-1} = M\sigma^2 I_d X(X'X)^{-1} \\ &= \sigma^2 MX(X'X)^{-1} = 0 \text{ car } MX = 0 \end{aligned}$$

$$\left\{ \begin{array}{l} \text{Donc } \text{cov}(u, w) = 0 \\ u \text{ et } w \text{ vecteurs gaussiens} \end{array} \right. \implies u \text{ et } w \text{ indépendants (cf. chapitre 0)}$$

$$\begin{aligned} &\implies y - X\hat{\beta} \text{ et } (X'X)^{-1}X'\epsilon \text{ indépendants} \\ &\implies y - X\hat{\beta} \text{ et } \hat{\beta} \text{ indépendants} \\ &\implies \|y - X\hat{\beta}\| \text{ et } \hat{\beta} \text{ indépendants} \\ &\implies \hat{\sigma}^2 \text{ et } \hat{\beta} \text{ indépendants} \end{aligned}$$

Proposition 1..5. Pour $j = 1, \dots, p$ la v.a.r

$$T_j = \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma} \sqrt{\gamma_{jj}}} \sim \mathcal{T}_{n-p}$$

où γ_{jj} jème élément diagonal de $(X'X)^{-1}$

Remarque 1..4. $\hat{\sigma} \sqrt{\gamma_{jj}} = e.s(\hat{\beta}_j) = \hat{\sigma}_{\hat{\beta}_j}$

Démonstration :

D'après la proposition précédente, on sait d'une part que :

$$\hat{\beta} \sim \mathcal{N}_P(\beta, \sigma^2(X'X)^{-1})$$

et d'autre part que

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

et enfin que Les v.a. $\hat{\beta}$ et $\hat{\sigma}^2$ sont indépendantes. Il reste alors à écrire T_j sous la forme :

$$T_j = \frac{\frac{\hat{\beta}_j - \beta_j}{\sigma \sqrt{\gamma_{jj}}}}{\frac{\hat{\sigma}}{\sigma}}$$

pour reconnaître une loi de Student \mathcal{T}_{n-p}

Proposition 1..6. :

Soit L une matrice $q \times p$ de rang q ($q < p$)

i) La v.a.r.

$$\frac{1}{\sigma^2} [L(\hat{\beta}) - \beta]' \cdot [L(X'X)^{-1}L']^{-1} \cdot [L(\hat{\beta}) - \beta] \sim \chi_{(q)}^2$$

.

ii) La v.a.r.

$$\frac{1}{q\hat{\sigma}^2} [L(\hat{\beta}) - \beta]' \cdot [L(X'X)^{-1}L']^{-1} \cdot [L(\hat{\beta}) - \beta] \sim \mathcal{F}_{q,n-p}$$

.

Démonstration :

i) On a :

- $\text{rang}(L) = q \implies \text{rang}[L(X'X)^{-1}L'] = q$ (Chapitre 0).
- $L\hat{\beta}$ est un vecteur aléatoire gaussien de IR^q
- $\text{Var}(L\hat{\beta}) = L\text{Var}\hat{\beta}L' = \sigma^2 L(X'X)^{-1}L'$ qui est inversible car définie positive de rang q .

$$\begin{aligned} \implies L\hat{\beta} &\sim N_n(L\beta, \sigma^2 L(X'X)^{-1}L') \\ \implies [L\hat{\beta} - L\beta] \cdot \frac{1}{\sigma^2} [L(X'X)^{-1}L']^{-1} \cdot [L\hat{\beta} - L\beta] &\sim q \cdot (\chi^2_{(q)}) \end{aligned}$$

d'après le théorème de COCHRAN (Chapitre 0).

ii) (exercice)

1.3 Estimation par intervalle de confiance et Région de confiance (RC)

De ces résultats vont découler les régions de confiance de la section suivante. Auparavant, donnons un exemple illustrant le second point que l'on vient d'établir.

Proposition 1..7. : *Un IC au niveau de confiance $\gamma = 1 - \alpha$ d'un coefficient β_j (pour $j = 1, \dots, p$) est :*

$$[\hat{\beta}_j - t_{(n-p, \frac{\alpha}{2})} \hat{\sigma} \sqrt{\gamma_{jj}}; \hat{\beta}_j + t_{(n-p, \frac{\alpha}{2})} \hat{\sigma} \sqrt{\gamma_{jj}}]$$

Où γ_{jj} jème élément diagonal de $(X'X)^{-1}$

$t_{(n-p, \frac{\alpha}{2})}$ valeur d'une v.a. de Student de paramètre $(n - p)$ qui est dépassée avec une probabilité $\frac{\alpha}{2}$.

Démonstration : Découle de la propriété 4 du paragraphe précédent et de la propriété (chapitre 1) de changement de variable : rapport d'un v.a. normale et de la $\sqrt{\cdot}$ d'une v.a. de χ^2 .

Proposition 1..8. : *Région de confiance (R.C) pour q paramètres $\beta_j (q < p)$.*

Une R.C au niveau de sécurité $\gamma = 1 - \alpha$, de q coefficients $\beta_{j_1}, \dots, \beta_{j_q}$ est :

$$J_{q,\alpha}(L_q\beta) = \{L_q\beta \in R^q / \frac{[L_q(\hat{\beta} - \beta)]' \cdot [L_q(X'X)^{-1}L_q']^{-1} \cdot [L_q(\hat{\beta} - \beta)]}{q\hat{\sigma}^2} \leq F_{q,n-p,\alpha}\}$$

Où :

L_q est la matrice $(q \times p)$ telle que les éléments : $(L_q)_{ij_i} = 1$ et les autres sont égaux à zéro.

$F_{q,n-p,\alpha}$ est la valeur d'une v.a. de Fisher $(q; n - p)$ qui est dépassée avec une probabilité α .

Démonstration : Résulte de la propriété 5 du paragraphe précédent et de la propriété de changement de variable (chapitre 0)

Exemple 1..1. Considérons le cas $p = q = 2$ et la matrice $R = I_2$, de sorte que

$$L_2(\hat{\beta} - \beta) = \begin{pmatrix} \hat{\beta}_1 & - & \beta_1 \\ \hat{\beta}_2 & - & \beta_2 \end{pmatrix}$$

Si la constante fait partie du modèle, X est la matrice $n \times 2$ dont la première colonne est uniquement composée de 1 et la seconde est composée des x_i , si bien que

$$X'X = \frac{1}{\sum (x_i - \bar{x})^2} \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} = \frac{1}{\sum (x_i - \bar{x})^2} \begin{pmatrix} n & n\bar{x} \\ n\bar{x} & \sum x_i^2 \end{pmatrix}$$

et le point (ii) s'écrit

$$\frac{1}{2\hat{\sigma}^2} (n(\hat{\beta}_1 - \beta_1)^2 + 2n\bar{x}(\hat{\beta}_1 - \beta_1)(\hat{\beta}_2 - \beta_2) + \sum x_i^2(\hat{\beta}_2 - \beta_2)^2) \sim \mathcal{F}_{2,n-2}$$

. permettant de construire une ellipse de confiance pour $\beta = (\beta_1, \beta_2)$.

Exemple 1..2. Application sur l'exemple 2 du chapitre 1.

Au niveau $\gamma = 95\%$, construire :

1. I.C. de β_2 (coefficient de la v.a.r. cylindrée).
2. R.C. de β_2, β_3 (coefficients des v.a.r. cylindrée et puissance).
1. I.C. de β_2 $n = 18$ $p = 7$

$$\left\{ \underbrace{\hat{\beta}_2}_{-3.505} - \underbrace{t_{n-p, \frac{\alpha}{2}}}_{t_{11, 2.5\%} = 2.201} \times \underbrace{\hat{\sigma}\sqrt{\gamma_{jj}}}_{\text{erreur standard}}, \hat{\beta}_2 + t_{n-p, \frac{\alpha}{2}} \times \hat{\sigma}\sqrt{\gamma_{jj}} \right\}$$

I.C. de β_2 au coefficient de sécurité 95% :

$$[-15.7, +8.7]$$

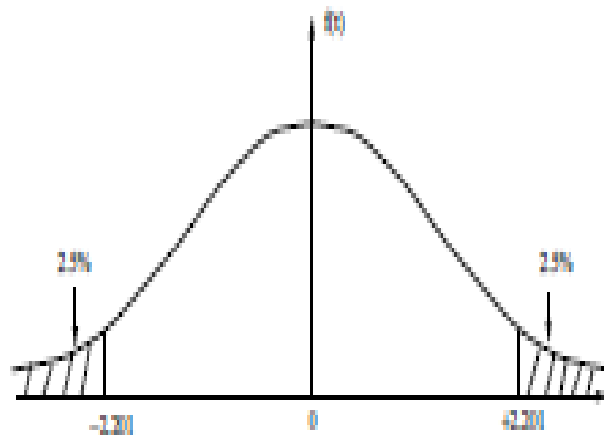
Remarque 1..5. : $0 \in [-15 : 7; +8 : 7]$

La valeur 0 est une valeur possible du paramètre β_j .

2. R.C. de β_2 et β_3

$$L_2 = \begin{pmatrix} 0 & 1 & 0 & . & . & . & 0 \\ 0 & 0 & 1 & . & . & . & 0 \end{pmatrix} \Rightarrow L_2(\hat{\beta} - \beta) \begin{pmatrix} \hat{\beta}_2 & - & \beta_2 \\ \hat{\beta}_3 & - & \beta_3 \end{pmatrix}$$

$$J_{2,\alpha}(\beta_2) = \{(\beta_2, \beta_3) / \frac{1}{2\hat{\sigma}^2} (\hat{\beta}_2 - \beta_2 \quad \hat{\beta}_3 - \beta_3) \Omega \begin{pmatrix} \hat{\beta}_2 & - & \beta_2 \\ \hat{\beta}_3 & - & \beta_3 \end{pmatrix} \leq F_{(2;11;5\%)}\}$$



Où Ω est l'inverse de

$$L_2(X'X)^{-1}L_2' = \begin{pmatrix} \gamma_{22} & \gamma_{23} \\ \gamma_{32} & \gamma_{33} \end{pmatrix} \text{ avec } (\gamma_{ij}) = (X'X)^{-1}$$

$$\hat{\beta}_2 = -3.5; \hat{\beta}_3 = 282.17; \hat{\sigma}^2\gamma_{22} = 5.55^2; \hat{\sigma}^2\gamma_{33} = 174.88^2; \hat{\sigma}^2\gamma_{23} = \hat{\sigma}^2\gamma_{32} = 42.1;$$

$$F_{(2,11;5\%)} = 3.98$$

$$J_{2,\alpha}(\beta_2, \beta_3) = \{(\beta_2, \beta_3) / (3.5 + \beta_2)^2\omega_1 + (\beta_3 - 282.17)^2\omega_2 + \omega_3(-3.5 - \beta_2)(282.17 - \beta_3) \leq \omega_4\}$$

où $\omega_1, \omega_2, \omega_3, \omega_4$ sont déterminés par le calcul.

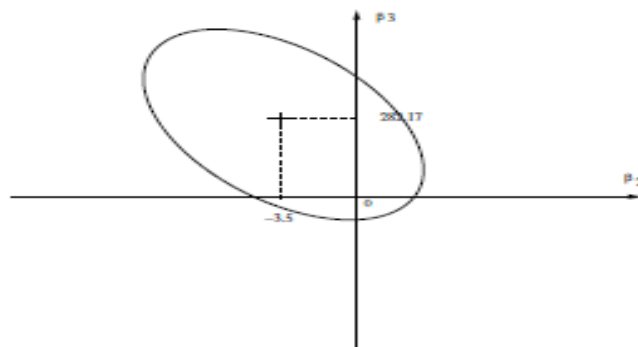


FIGURE 2.1 – Ellipsoïde de confiance

1.4 Prévision pour une valeur future $x_0 = (x_0^1, \dots, x_0^p)'$

La valeur prédite (ou ajustée) est :

$$\hat{y}_0 = x_0' \hat{\beta} \text{ ou } x_0 = \begin{pmatrix} x_0^1 \\ \vdots \\ x_0^p \end{pmatrix}$$

L'intervalle de prévision de $y_0 = x_0' \hat{\beta} + e_0$ est alors (au coefficient de sécurité $(\gamma = 1 - \alpha)$) :

$$[x_0' \hat{\beta} - t_{n-p, \frac{\alpha}{2}} \times \hat{\sigma} \sqrt{\epsilon_0}; x_0' \hat{\beta} + t_{n-p, \frac{\alpha}{2}} \sigma \sqrt{\epsilon_0}]$$

où $\epsilon_0 = 1 + x_0'(X'X)^{-1}x_0$ qui se déduit de la variance de y_0 :

$$\begin{aligned} \text{var}(y_0) &= \text{var}(x_0' \hat{\beta}) + \text{var}e_0 \\ &= x_0' \text{var} \hat{\beta} x_0 + \sigma^2 \\ &= \sigma^2 x_0'(X'X)^{-1}x_0 + \sigma^2 = \sigma^2 \epsilon_0 \end{aligned}$$

2. Tests d'hypothèses

2.1 Estimation sous contrainte linéaire

On souhaite estimer le paramètre β sous contrainte linéaire.

Exemple 2..1. :

$$\beta_1 = 0 \quad \underline{\text{Ecriture matricielle}} \quad (1 \ 0 \ \dots \ 0)\beta = 0$$

$$\beta_1 = \beta_2 = \beta_3 = 0 \longrightarrow \begin{pmatrix} 1 & 0 & \dots & \dots & \dots & 0 \\ 0 & 1 & 0 & \dots & \dots & \dots \\ 0 & 0 & 1 & 0 & \dots & 0 \end{pmatrix} \beta = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{cases} \beta_1 + \beta_2 = 6 \\ 2\beta_2 - \beta_3 = 9 \end{cases} \longrightarrow \begin{pmatrix} 1 & 1 & 0 & \dots & \dots \\ 0 & 2 & -1 & \dots & \dots \end{pmatrix} \beta = \begin{pmatrix} 6 \\ 9 \end{pmatrix}$$

Contrainte linéaire :

où :

R est une matrice $q \times p$ de rang q

r est un vecteur $q \times 1$

R et r sont connus

Cette écriture matricielle signifie que l'on impose un ensemble de q contraintes linéaires sur les β_j

Proposition 2..1. *Si la matrice $R(q \times p)$ est de rang q , l'E.M.C.O. contraint par $R\beta = r$, que l'on notera $\hat{\beta}^c$, est égal à :*

$$\hat{\beta}^c = \hat{\beta} + (X'X)^{-1}R'[R(X'X)^{-1}R' - 1]^{-1}(r - R\hat{\beta}) \quad \hat{\beta} = (X'X)^{-1}X'y \text{ est l'E.M.C.O. (non contraint)}$$

Démonstration : (en exercice)

idée : on minimisera $\|y - X\beta\|^2 - 2\lambda'(R\beta - r)$

où λ est un vecteur de q multiplicateurs de Lagrange.

2.2 Test de student de signification d'un coefficient

On souhaite tester :

$$H_0 = \beta_j = 0 \text{ contre } H_1 : \beta_j \neq 0$$

qui est un test bilatéral de signification de β_j . Dans ce cas la contrainte $R\beta = r$ peut s'écrire pour $R = (0 \ 0 \ 0 \ \underbrace{1}_{\text{à la } j\text{-ième}} \ 0 \ \dots \ 0)$ et $r = 0$.

Règle de décision, au niveau α :

$$\text{on rejette : } H_0 \text{ Si } t = \left| \frac{\hat{\beta}_j}{\delta_{\beta_j}} \right| > t_{n-p, \frac{\alpha}{2}}$$

Démonstration : découle des propriétés du paragraphe 1.

Remarque : On pourra construire un test unilatéral (à titre d'exercice) pour l'hypothèse alternative $H_1 : \beta_j > 0$.

2.3 Test de Fisher de signification de plusieurs coefficients

Soit q coefficients $\beta_{j_1}, \dots, \beta_{j_q}$ ($q \leq p$)

On souhaite tester :

$$H_0 = \beta_{j_1} = \beta_{j_2} = \dots = \beta_{j_q} = 0 \text{ contre } H_1 : \exists j \in \{j_1, \dots, j_q\} \text{ tel que } \beta_j \neq 0$$

L'hypothèse H_0 s'écrit matriciellement $R\beta = r$, où

$$R_{(q \times p)} = \begin{pmatrix} 0 & 1 & 0 & \dots & \dots & 0 \\ 0 & 0 & 0 & 1 & 0 & \dots \\ 0 & \dots & \dots & \dots & \dots & 1 & 0 \end{pmatrix} \text{ et } r_{(q \times 1)} = \begin{pmatrix} 0 \\ \dots \\ 0 \end{pmatrix}$$

Remarque : Si H_0 est vraie, alors $R\beta = 0$ et donc $R\hat{\beta}$ doit être “proche” de 0.

Donc la première idée intuitive du test de H_0 c'est de rejeter H_0 si la norme de $R\hat{\beta}$ est trop grande. On rappelle

$$R\hat{\beta} \sim N_q\left(\underbrace{R\beta}_{R\beta=0 \text{ si } H_0 \text{ vraie}}, \underbrace{\sigma^2 R(X'X)^{-1}R'}\right)$$

Règle de décision, au niveau α

$$\text{on rejette } H_0 \text{ si } F_q = \frac{[R\hat{\beta}]' \cdot [R(X'X)^{-1}R\hat{\beta}]}{q\hat{\sigma}^2} > F_{q,n-p,\alpha}$$

Remarques :

1. Ce résultat se déduit de la propriété 5 paragraphe 1 et de la propriété du rapport de 2 lois de chi-deux.
2. On retrouve le cas particulier pour 1 coefficient : $H_0 : \beta_j = 0$.
En effet :
 - $R\hat{\beta} = \hat{\beta}_j, q = 1$
 - $[R(X'X)^{-1}R']^{-1} = (\gamma_{jj})^{-1}$ où $\gamma_{jj} = ((X'X)^{-1})_{jj}$
 - $F_1 = \frac{\hat{\beta}_j^2}{\sigma^2 \cdot \gamma_{jj}} = \frac{\hat{\beta}_j^2}{e.s(\hat{\beta}_j)}$
 - $F_{1,n-p,\alpha} = t_{n-p,\alpha}^2$ (cf. propriété chapitre 0 : $T \text{ Student} \implies T^2 \text{ Fisher}$)

Ce test peut être justifié par une deuxième idée intuitive. On doit rejeter H_0 si l'écart entre la “ RSS_c ” dans le modèle contraint et la “ RSS ” dans le modèle non contraint, est trop grand (ce qui revient à l'écart entre $X\hat{\beta}^c$ et $X\hat{\beta}$ trop grand). Dans le modèle contraint (si H_0 est vraie) : $R\beta = r = 0$.

$$RSS_c = \|y - X\hat{\beta}_c\|_{I_n}^2 \text{ où } \hat{\beta}_c = \hat{\beta} - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R\hat{\beta}$$

Dans le modèle sans contrainte

$$RSS = \|y - X\hat{\beta}\|_{I_n}^2 = \hat{\sigma}^2 \times (n - p)$$

Proposition 2..2. : (démonstration en exercice).

$$RSS_c - RSS = \|X\hat{\beta} - X\hat{\beta}_c\|^2 = (R\hat{\beta})'[R(X'X)^{-1}R']^{-1}R\hat{\beta}$$

Formulation du test de Fisher équivalente :

on rejette H_0 si $F_1 = \frac{RSS_c - RSS}{RSS} \cdot \frac{q}{n-p} > F_{1,n-p,\alpha}$

Remarque : Dans le cas $q = p - 1$ et $\{j_1, \dots, j_q\} = \{2, \dots, p\}$ la statistique de Fisher s'écrit :

$$F_{p-1} = \frac{R^2}{1-R^2} \cdot \frac{n-p}{p-1} \frac{\sum ((\hat{y}_i - y_i)^2 / (p-1))}{\sum ((\hat{y}_i - y_i)^2 / (n-p))}$$

où R^2 est le coefficient de corrélation multiple empirique (cf. Chapitre 1).

2.4 Test de Fisher d'une hypothèse linéaire générale

Le raisonnement est le même que dans le paragraphe précédent mais pour l'hypothèse :