POLYCOPIÉ DE COURS DE STATISTIQUE ET ÉCONOMÉTRIE

MODÈLE LINÉAIRE

M. BONNEU, E. LECONTE

Université des Sciences Sociales
Place Anatole France
31042 Toulouse

Table des matières

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	$\cdots$	x^p
u.s. 1	y_1	x_1^1		x_1^p
$\vdots$	$\vdots$	$\vdots$		$\vdots$
u.s. i	y_i	x_i^1		x_i^p
$\vdots$	$\vdots$	$\vdots$		$\vdots$
u.s. n	y_n	x_n^1		x_n^p

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

$x^1, \ldots, x^p$: variables quantitatives ou qualitatives exogènes.

y : variable quantitative endogène.

Exemple 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 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

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

Exemple 3.

Modèle économique du comportement de production des entreprises :

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

(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.

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

ou encore $$\log Q = \sum_{j=1}^{p} \beta_j \log x^j \text{ , avec } \log x^1 = 1$$

Etude statistique de ce modèle

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

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

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 :

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

Objectif : Estimer les paramètres :
– $\beta_j = \dfrac{\partial \log Q}{\partial \log x^j}$ élasticité.
– $r = \sum \beta_j$ rendement.

Exemple 4.

En Macroéconomie, modèle dynamique explicatif :

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

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

$$\left| \begin{array}{ll} E(e_t) = 0 & \text{“centrées”} \\ V(e_t) = \sigma_t^2 & \text{hétéroscédastiques} \\ E(e_t e_{t-1}) \neq 0 & \text{auto-corrélées} \end{array} \right.$$

OBJECTIF DU COURS

A partir de n observations de $(y, x),\ y \in I\!R$ et $x \in I\!R^p$, on cherche à expliquer la variable y par une combinaison linéaire des variables $x^1, \ldots, x^p$, 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 $x^1, \ldots, x^p$ sont quantitatives ou qualitatives.
- Les données sont des séries temporelles ou des données de panel.

Chapitre 1

Outils algébriques et probabilistes pour le modèle linéaire

1.1 Eléments d'algèbre linéaire

Ce chapitre se propose de rassembler des notations et rappels d'algèbre linéaire ainsi que quelques compléments mathématiques du niveau du premier cycle des Universités.

Dans tout ce qui suit, E et F sont deux espaces vectoriels réels munis respectivement des bases canoniques $\mathcal{E} = \{e_j; j = 1, \ldots, p\}$ et $\mathcal{F} = \{f_i; i = 1, \ldots, n\}$. On note indifféremment un vecteur de E ou de F, un endomorphisme de E, une application linéaire de E dans F et leurs représentations matricielles dans ces bases.

1.1.1 Matrices

Notations

La matrice d'ordre $(n \times p)$ associée à une application linéaire de E dans F est décrite par un tableau :

$$A = \begin{bmatrix} a_1^1 & \cdots & a_1^j & \cdots & a_1^p \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ a_i^1 & \cdots & a_i^j & \cdots & a_i^p \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ a_n^1 & \cdots & a_n^j & \cdots & a_n^p \end{bmatrix}$$

On note par la suite :

$a_i^j = [A]_i^j$ le terme général de la matrice ou a_{ij},

$a_i = [a_i^1, \ldots, a_i^p]'$ un vecteur ligne mis en colonne, où " $'$ " désigne le symbole transposé.

$a^j = [a_1^j, \ldots, a_n^j]'$ un vecteur colonne.

Types de matrices

Une matrice est dite :

- carrée si $n = p$,
- identité (I_p) si $a_i^j = \delta_i^j = \begin{cases} 0 \text{ si } i \neq j \\ 1 \text{ si } i = j \end{cases}$,
- diagonale si $a_i^j = 0$ lorsque $i \neq j$,
- symétrique si $a_i^j = a_j^i$,
- triangulaire supérieure (inférieure) si

$$a_i^j = 0 \text{ lorsque } i > j \ (i < j),$$

- vecteur ligne (colonne) si $n = 1$ $(p = 1)$,
- unité d'ordre p : $\mathbb{1}_p = [1, \ldots, 1]'$,
- scalaire si $n = 1$ et $p = 1$.

Opérations sur les matrices

Somme

$[A + B]_i^j = a_i^j + b_i^j$ pour A et B de même ordre $(n \times p)$.

Multiplication par un scalaire

$[\alpha A]_i^j = \alpha a_i^j$ pour $\alpha \in \mathbb{R}$.

Transposition

$[A']_i^j = a_j^i$, A' est d'ordre $(p \times n)$.
$(A')' = A$; $(A + B)' = A' + B'$; $(AB)' = B'A'$.

Produit scalaire élémentaire

$a'b = \displaystyle\sum_{i=1}^{n} a_i b_i$ où a et b sont des vecteurs colonnes $(n \times 1)$.

Produit

$[AB]_i^j = a_i' b^j$ avec $A_{(n \times p)}, B_{(p \times q)}$.

Propriétés des matrices carrées

La *trace* et le *déterminant* sont des notions intrinsèques qui ne dépendent pas des bases de représentation choisies mais uniquement de l'application linéaire sous-jacente.

Trace

Par définition, si A est une matrice $(p \times p)$:

$$trA = \sum_{i=1}^{p} a_i^i$$

et il est facile de montrer que

$$\begin{aligned}
tr\alpha &= \alpha, \text{ où } \alpha \in I\!R \\
tr(\alpha A) &= \alpha trA, \\
tr(A+B) &= trA + trB, \\
tr(AB) &= tr(BA), \\
tr(CC') &= tr(C'C) = \sum_{i=1}^{p}\sum_{j=1}^{p}(c_i^j)^2.
\end{aligned}$$

Déterminant

On note $\mid A \mid$ (ou det$A$) le *déterminant* de la matrice carrée A. Il vérifie :

$$\begin{aligned}
\mid A \mid &= \prod_{i=1}^{p} a_i^i \text{ si } A \text{ est triangulaire ou diagonale}, \\
\mid \alpha A \mid &= \alpha^p \mid A \mid .
\end{aligned}$$

Inverse

L'*inverse* de A, lorsqu'elle existe, est la matrice unique notée A^{-1} telle que

$$AA^{-1} = A^{-1}A = I_p \ ;$$

elle existe si et seulement si $\mid A \mid \neq 0$.

Quelques propriétés :

$$\begin{aligned}
(A^{-1})' &= (A')^{-1}, \\
(AB)^{-1} &= B^{-1}A^{-1}, \\
\mid A^{-1} \mid &= \frac{1}{\mid A \mid}.
\end{aligned}$$

Définitions

Une matrice *carrée* A est dite :
symétrique si $A' = A$,
singulière si $\mid A \mid = 0$,
régulière si $\mid A \mid \neq 0$,
idempotente si $AA = A$,
orthogonale si $AA' = A'A = I (A' = A^{-1})$.

Exercice :

$$\text{Soit } A = \frac{1}{6} \begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix}$$

Montrer que A est idempotente.
Est-elle régulière ?
Est-elle orthogonale ?

Propriétés des formes quadratiques et matrices définies positives

Définitions

Soit A une matrice $(n \times n)$. La forme quadratique associée à A est l'application de $I\!R^n$ dans $I\!R$ qui à tout vecteur x associe $x'Ax$.

Si $x'Ax > 0$ pour tout $x \neq 0$, on dit que la forme quadratique est *définie positive* et A est dite *matrice définie positive*.

Si $x'Ax \geq 0$ pour tout x, on dit que la forme et la matrice sont *semi-définies positives*. Lorsque les inégalités sont dans l'autre sens, on parle de formes quadratiques et de matrices *définies négatives et semi-définies négatives*. Si une forme quadratique est positive pour certains vecteurs x, et négative pour d'autres, la forme et la matrice sont dites *indéfinies*.

Il est important de disposer de critères permettant de reconnaître les matrices définies positives.

1. Pour qu'une matrice réelle et symétrique A soit définie positive, il faut et il suffit que toutes ses valeurs propres soient positives.
2. Pour qu'une matrice symétrique A soit définie positive, il faut et il suffit que les déterminants de chacune de ses sous-matrices principales soient positifs.

3. Si A est symétrique et définie positive, il existe une matrice non singulière P telle que
$$A = PP'$$
4. Si A, de format $n \times n$, est définie positive, et si P de format $n \times m$, est de rang m, alors $P'AP$ est définie positive.
1. Si A est de format $n \times m$ avec $m < n$ et de rang m, alors $A'A$ est définie positive et AA' est semi-définie positive.
6. Si A et B sont des matrices définies positives et si $A - B$ est aussi définie positive alors $B^{-1} - A^{-1}$ est aussi définie positive.

Propriétés des vecteurs propres et des valeurs propres des matrices réelles symétriques

Dans les problèmes statistiques nous sommes généralement confrontés à des matrices réelles symétriques. Les propriétés énumérées ci-dessous qui sont précédées par un astérisque ne sont valables que pour les matrices réelles symétriques ; celles sans astérisque sont valables pour toutes les matrices réelles.

1. * Les valeurs propres sont réelles.
2. * Les vecteurs propres associés à des valeurs propres différentes sont orthogonaux.
3. * Si une valeur propre est de multiplicité algébrique k (c'est-à-dire, si elle est une racine d'ordre k), il y a k vecteurs propres orthogonaux qui lui sont associés.
4. * La matrice symétrique A d'ordre n possède n valeurs propres λ_1, $\lambda_2, \ldots, \lambda_n$, pas forcément différentes. Les propriétés 2 et 3 montrent qu'il existe un ensemble de n vecteurs propres orthogonaux x_1, $x_2, \ldots, x_n$, tels que
$$x_i'x_j = 0 \qquad i \neq j; i, j = 1, 2, \ldots, n$$
5. * La matrice orthogonale des vecteurs propres permet de diagonaliser A, c'est à dire
$$X'AX = \Lambda = \text{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_n\}.$$
6. La somme des valeurs propres est égale à la somme des éléments diagonaux (la trace) de A.
7. Le produit des valeurs propres d'une matrice A est égale à son déterminant.
8. Le rang de A est égal au nombre de ses valeurs propres non nulles.
9. Les valeurs propres de A^2 sont les carrés des valeurs propres de A et ces deux matrices ont les mêmes vecteurs propres.
10. Les valeurs propres de A^{-1} sont les inverses des valeurs propres de A, mais ces deux matrices ont les mêmes vecteurs propres. Soit :
$$Ax = \lambda x$$
11. Les valeurs propres d'une matrice idempotente ne peuvent être égales qu'à 0 ou à 1.
12. Le rang d'une matrice idempotente est égal à sa trace.

1.1.2 Espaces euclidiens

E est un espace vectoriel réel de dimension p isomorphe à $\mathbb{R}^p$.

Sous-espaces

- Un sous-ensemble E_q de E est un *sous-espace vectoriel* (s.e.v.) de E s'il est non vide et stable :
$$\forall x, y \in E_q,\ \forall \alpha \in \mathbb{R},\ \alpha(x+y) \in E_q.$$
- Le *q*-uple $\{x_1, \ldots, x_q\}$ de E constitue un système *linéairement indépendant* si et seulement si :
$$\sum_{i=1}^{q} \alpha_i x_i = 0 \Rightarrow \alpha_1 = \cdots = \alpha_q = 0.$$
- Un sytème linéairement indépendant $\mathcal{E}_q = \{e_1, \ldots, e_q\}$ qui engendre un s.e.v. $E_q = \text{vect}\{e_1, \ldots, e_q\}$ en constitue une *base* et $\text{dim}(E_q) = \text{card}(\mathcal{E}_q) = q$.

Rang d'une matrice $A_{n\times p}$

Image et noyau

$$\begin{aligned}
\text{Im}(A) &= \text{vect}\,\{a^1, \ldots, a^p\} \text{ est le s.e.v. de } F \text{ image de } A,\\
\text{Ker}(A) &= \{x; Ax = 0\} \text{ est le s.e.v. de } E \text{ noyau de } A,\\
p &= \dim(\text{Im}(A)) + \dim(\text{Ker}(A)).
\end{aligned}$$

Rang

$$\begin{array}{lll}
\text{rang}(A) & = & \dim(\text{Im}(A)),\\
0 \leq & \text{rang}(A) & \leq \min(n,p),\\
\text{rang}(A) & = & \text{rang}(A'),\\
\text{rang}(A+B) & \leq & \text{rang}(A) + \text{rang}(B),\\
\text{rang}(AB) & \leq & \min(\text{rang}(A), \text{rang}(B)),\\
\text{rang}(BAC) & = & \text{rang}(A) \text{ si } B \text{ et } C \text{ sont régulières},\\
\text{rang}(A) & = & \text{rang}(AA') = \text{rang}(A'A).
\end{array}$$

Enfin, si $B(p\times q)$ est de rang q $(q < p)$ et A carrée $(p \times p)$ de rang p alors, la matrice $B'AB$ est de rang q.

Métrique euclidienne

Soit M une matrice carrée $(p \times p)$, symétrique, définie-positive ; M définit sur l'espace E :

– un produit scalaire : $\langle x, y \rangle_M = x'My$,
– une norme : $\| x \|_M = \langle x, x \rangle_M^{1/2}$,
– une distance : $d_M(x, y) = \| x - y \|_M$,
– des angles : $\cos \theta_M(x, y) = \dfrac{\langle x, y \rangle_M}{\| x \|_M \| y \|_M}$,

La matrice M étant donnée, on dit que :
– une matrice A est M-symétrique si $A'M = MA$,
– deux vecteurs x et y sont M-orthogonaux si $\langle x, y \rangle_M = 0$,
– un vecteur x est M-normé si $\| x \|_M = 1$,
– une base $\mathcal{E}_q = \{e_1, \ldots, e_q\}$ est M-orthonormée si

$$\forall (i, j), \langle e_i, e_j \rangle_M = \delta_i^j.$$

Remarque : Si M est la matrice Identité $(M = I_p)$, c'est le cas usuel relatif à la base canonique de $\mathbb{R}^p$.

Projection

Définitions équivalentes

– Soit W un sous-espace de E et $\mathcal{B} = \{b^1, \ldots, b^q\}$ une base de W : $P(p \times p)$ est une matrice de projection M-orthogonale sur W si et seulement si :

$$\forall y \in E, Py \in W \text{ et } \langle Py, y - Py \rangle_M = 0.$$

– Toute matrice idempotente $(P^2 = P)$ et M-symétrique $(P'M = MP)$ est une matrice de projection M-orthogonale.

Propriétés d'une matrice P de projection

– Les valeurs propres de P sont 0 ou 1
$u \in W \quad Pu = u, \lambda = 1$ de multiplicité $\dim(W)$,
$v \in W^\perp \quad Pv = 0, \lambda = 0$ de multiplicité $\dim(W^\perp)$, où $W^\perp$ est le sous-espace orthogonal à W $(W \bigoplus W^\perp = E)$.
– $trP = \dim(W) \ (= rang(P))$,
– $P = B(B'MB)^{-1}B'M$ où $B = [b^1, \ldots, b^q]$,

– dans le cas particulier où les b^j sont M-orthonormés :

$$P = BB'M = \sum_{j=1}^{q} b^j b^{j'} M,$$

– dans le cas particulier où $q = 1$ alors :

$$P = \frac{bb'}{b'Mb} M = \frac{1}{\| b \|_M} bb'M.$$

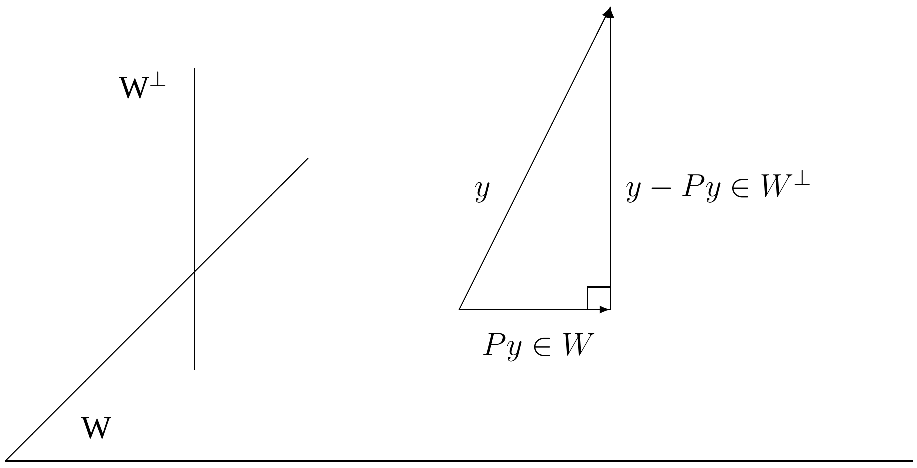
– Si $P_1, \ldots, P_q$ sont des matrices de projection M-orthogonales alors la somme $P_1 + \cdots + P_q$ est une matrice de projection M-orthogonale si et seulement si : $P_k P_j = \delta_k^j P_j$.
– La matrice $I - P$ est la matrice de projection M-orthogonale sur $W^\perp$.

Exemple :

$X = [x^1 \dots x^p]$ matrice $(n \times p)$ des p variables explicatives du modèle linéaire :

$$y_i = \sum_{j=1}^{p} \beta_j x_i^j + e_i \ (i = 1, \dots, n).$$

$W =$ vect$\{x^1, \dots, x^p\}$.



$P = X(X'MX)^{-1}X'M$ est une matrice de projection M-othogonale sur W.
où W est l'espace engendré par les vecteurs colonnes de la matrice X car

$$0 =< Py,\ y - Py >_M = (Py)'M(y - Py)$$

$$Py \in W \text{ car } Py = X\beta \ (\beta \text{ vecteur de composantes } \beta_j)$$

On peut aussi vérifier que : $P^2 = P$ et $P'M = MP$.

1.1.3 Dérivation matricielle

Définition

Soit f une application de $I\!R^k$ dans $I\!R$, on définit le vecteur colonne $(k \times 1)$ des k dérivées partielles de $f(u)$ par rapport à chaque composante u_i :

$$\frac{\partial f(u)}{\partial u} = \begin{pmatrix} \dfrac{\partial f(u)}{\partial u_1} \\ \vdots \\ \dfrac{\partial f(u)}{\partial u_k} \end{pmatrix}$$

Propriétés

(i) Cas des formes linéaires : Si $f(u) = a'u$ où a est un vecteur $(k \times 1)$:

$$\frac{\partial a'u}{\partial u} = a = \frac{\partial u'a}{\partial u}$$

(ii) Cas des formes quadratiques : Si $f(u) = u'Au$ ou A est une matrice carrée d'ordre k, alors :

$$\frac{\partial u'Au}{\partial u} = (A + A')u.$$

De plus, si A est symétrique : $\dfrac{\partial u'Au}{\partial u} = 2Au.$

1.2 Vecteurs aléatoires

1.2.1 Rappels sur les variables aléatoires réelles (v.a.r.)

(cf. tableau de lois, théorèmes sur les combinaisons linéaires indépendantes de v.a.r. et sur les changements de variables donnant les lois usuelles d'échantillonnage : F, T, χ^2).

Formulaire de probabilités

Lois de probabilité usuelles

Lois discrètes

Dénomination X	Loi de probabilité	Moyenne $E(X)$	Variance $V(X)$
Loi binomiale $B(n,p)$ n entier positif ; $0 < p < 1$	$P[X=k] = C_n^k p^k (1-p)^{n-k}$ $k = 0, 1, 2, \ldots, n$	np	$np(1-p)$
Loi multinomiale $n; p_1, \ldots, p_k$, n entier positif ; $p_1 + p_2 + \cdots + p_k = 1$	$P[X_1 = n_1$ et $X_2 = n_2$ et $X_k = n_k]$ $= \frac{n!}{n_1! n_2! \cdots n_k!} p_1^{n_1} \cdot p_2^{n_2} \cdots p_k^{n_k}$ n_i entier positif ; $\sum_{i=1}^{k} n_i = n$.	$E(X_i) = np_i$	var $(X_i) = np_i(1-p_i)$ cov $(X_i, X_j) = -np_i p_j$ pour $i \neq j$
Loi de Poisson de paramètre $\lambda (\lambda > 0)$	$P[X=k] = e^{-\lambda} \frac{\lambda^k}{k!}$ k entier positif ou nul	λ	λ
Loi binomiale négative de paramètres n et p n entier positif - $0 < p < 1$ ($n = 1 \longrightarrow$ loi géométrique)	$P[X=k] = C_{n+k-1}^{n-1} p^n (1-p)^k$ k entier positif ou nul	$\frac{n(1-p)}{p}$	$\frac{n(1-p)}{p^2}$

Lois continues

Dénomination X	Loi de probabilité	Moyenne $E(X)$	Variance $V(X)$
Loi uniforme sur $[a,b]$	$f(x) = \begin{cases} \dfrac{1}{b-a} \text{ pour } a \leq x \leq b \\ 0 \end{cases}$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$
Loi normale $\mathcal{N}(\mu, \sigma^2)$ σ positif	$f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	μ	σ^2
Loi Gamma G(a,p) a et p strictement positifs	$f(x) = \begin{cases} \dfrac{a^p}{\Gamma(p)} x^{p-1} e^{-ax} \text{ si } x > 0 \\ 0 \text{ si } x \leq 0 \end{cases}$ où $\Gamma(p) = \int_0^{+\infty} x^{p-1} e^{-x} dx$	$\dfrac{p}{a}$	$\dfrac{p}{a^2}$
Loi Beta $B(p,q)$ $p > 0 \quad q > 0$	$f(x) = \begin{cases} \dfrac{1}{B(p,q)} x^{p-1}(1-x)^{q-1} \\ \quad \text{si } 0 < x < 1 \\ 0 \text{ sinon} \end{cases}$ où $B(p,q) = \int_0^1 x^{p-1}(1-x)^{q-1} dx$	$\dfrac{p}{p+q}$	$\dfrac{pq}{(p+q)^2(p+q+1)}$
Loi de χ^2 à ν degrés de liberté, ν entier positif $\left(\text{ou encore } G\left(\dfrac{1}{2}, \dfrac{\nu}{2}\right)\right)$	$f(x) = \begin{cases} \dfrac{e^{-\frac{x}{2}} x^{\frac{\nu}{2}-1}}{2^{\frac{\nu}{2}} \Gamma\left(\frac{\nu}{2}\right)} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$	ν	2ν
Loi de Student (ou T) à ν degrés de liberté, ν entier strictement positif	$f(x) = \dfrac{\left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}}{\sqrt{\nu} B\left(\frac{1}{2}, \frac{\nu}{2}\right)}$	0 pour $\nu \geq 2$ non définie pour $\nu = 1$	$\dfrac{\nu}{\nu-2}$ pour $\nu \geq 2$
Loi de Fisher (ou F) de paramètres ν_1 et ν_2 entiers strictement positifs	$f(x) = \begin{cases} \dfrac{\nu_1^{\nu_1/2} \nu_2^{\nu_2/2} x^{\nu_1/2-1}}{B(\frac{\nu_1}{2}, \frac{\nu_2}{2})(\nu_1 x + \nu_2)^{\frac{\nu_1+\nu_2}{2}}} \\ \quad \text{pour } x > 0 \\ 0 \text{ pour } x \leq 0 \end{cases}$	$\dfrac{\nu_2}{\nu_2 - 2}$ pour $\nu_2 \geq 3$ non définie pour $\nu_2 < 3$	$\dfrac{2\nu_2^2(\nu_1+\nu_2-2)}{\nu_1(\nu_2-2)^2(\nu_2-4)}$ si $\nu_2 \geq 5$
Loi de Cauchy	$f(x) = \dfrac{1}{\pi(1+x^2)}$	non définie	non définie
Loi de Laplace de paramètre $\theta > 0$	$f(x) = \dfrac{1}{\theta} \exp\left(-\dfrac{1}{2}\left\|\dfrac{x}{\theta}\right\|\right)$	0	$\dfrac{\theta^2}{2}$

1.2.2 Vecteur aléatoire

Définition : Un vecteur aléatoire X est une application d'un espace probabilisé $(\Omega, \mathcal{A}, P)$ dans un espace vectoriel réel $\mathbb{R}^p$, telle que chaque composante X_i est v.a.r.
On identifie X au p-uple de v.a. formé par ses composantes sur la base canonique de $\mathbb{R}^p$.

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix} \text{ où } X_i \; v.a.r. \qquad (1 \leq i \leq p)$$

Loi de probabilité d'un vecteur aléatoire X

Elle est définie par la fonction de répartition F application de $\mathbb{R}^p$ dans $\mathbb{R}$:

$$\begin{aligned} F(x_1, \ldots, x_p) &= P\left(X \in \prod_{1 \leq i \leq p}]-\infty, x_i]\right) \\ &= P(X_1 \leq x_1, \ldots, X_p \leq x_p) \end{aligned}$$

Remarque : La loi de probabilité sera définie de façon équivalente par :

- Soit la densité de X : $f(x_1, \ldots, x_p) = \dfrac{\partial^p F}{\partial x_1 \ldots \partial x_p}$ si la dérivée partielle existe dans le cas de v.a. continues.
- Soit $P(X_1 = x_1, \ldots, X_p = x_p)$ dans le cas de v.a. discrètes.

Application : Montrer que $f(x_1, x_2) = \frac{1}{2\sqrt{x_1 x_2}}$ si $0 < x_2 < x_1 < 1$ est une densité d'un vecteur aléatoire X de $\mathbb{R}^2$.

Espérance d'un vecteur aléatoire X de $\mathbb{R}^p$

$$E(X) = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_p) \end{pmatrix}$$

Remarque : Quand on ne dispose que de la loi de X, on peut calculer $E(X_i)$ par la formule :
$\rightarrow E(X_i) = \int\int_{\boldsymbol{R}^p} x_i f(x_1, \ldots, x_p) dx_1 \cdots dx_p$, dans le cas continu
ou bien $E(X_i) = \int_{\boldsymbol{R}} x_i \underbrace{\left[\int\int_{\boldsymbol{R}^{p-1}} f(x_1, \ldots, x_p) \prod_{j \neq i} dx_j\right]}_{\text{densité marginale de } x_i} dx_i$

Application : dans l'exemple précédent, calculer $E(X_1)$ et $E(X_2)$.

$\rightarrow E(X_i) = \sum_{x_1} \cdots \sum_{x_p} x_i P[X_1 = x_1, \ldots, X_p = x_p]$, dans le cas discret

ou bien $E(X_i) = \sum_{x_i} x_i \sum_{j \neq i} \underbrace{\sum_{x_j} P(X_1 = x_1, \ldots, X_p = x_p)}_{\text{Probabilité marginale } P(X_i = x_i)}$

Propriété :

Soit X un vecteur aléatoire de $\mathbb{R}^p$ et A une matrice non aléatoire $(m \times p)$. On a alors

$$E(AX) = AE(X).$$

Matrice de variance (covariance) de X, v.a. de $\mathbb{R}^p$

Définition : C'est la matrice carrée d'ordre p, notée $\sum_X$ ou $V(X)$ ou var(X), de terme général : cov (X_i, X_j)

$$\text{C'est-à-dire } V(X) = \begin{bmatrix} V(X_1) & \text{cov}\,(X_1, X_2) & \cdots & \text{cov}\,(X_1, X_p) \\ & & & \vdots \\ \text{cov}\,(X_2, X_1) & V(X_2) & & \vdots \\ \vdots & & \ddots & \\ \text{cov}\,(X_p, X_1) & \cdots & \cdots & V(X_p) \end{bmatrix}$$

cf. exemple précédent : déterminer la matrice $V(X)$

1. $V(X) = E\ [\underset{(p \times n)}{(X - E(X))}\ \ \underset{(n \times p)}{(X - E(X))'}]\ = E(XX') - E(X)[E(X)]'$
2. $V(X)$ est symétrique (en exercice)
3. $V(X)$ est définie positive (en exercice)
4. *Réciproquement* si $\sum$ est une matrice carrée d'ordre p symétrique et définie positive, alors $\sum$ est la matrice de variance d'un vecteur aléatoire de $\mathbb{R}^p$.
5. var $(AX + B) = \underset{(m \times p)(p \times p)(p \times m)}{A \cdot V(X) \cdot A'}$ A et B matrices <u>non</u> aléatoires, de dimensions respectives $(m \times p)$ et $(p \times n)$.

Définition :

La matrice de corrélation linéaire de X est la matrice de terme général :

$$\frac{\text{cov}\,(X_i, X_j)}{\sqrt{\text{var}\,(X_i) \cdot \text{var}\,(X_j)}}.$$

Matrice de variance covariance empirique

Soit n observations du vecteur X, définissant la matrice des données A de dimension $(n \times p)$:

$$A = \begin{pmatrix} x_1^1 & \cdots & x_1^p \\ \vdots & & \vdots \\ x_i^1 & \cdots & x_i^p \\ \vdots & & \vdots \\ x_n^1 & \cdots & x_n^p \end{pmatrix}$$

On note $\overline{x}^j$ la moyenne empirique : $\overline{x}^j = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i^j$, pour $j = 1$ à p.

On note $\overline{x}$ le vecteur colonne $(p \times 1)$ de composante $\overline{x}^j$.

On note varx^j, la variance empirique : var$x^j = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i^j - \overline{x}^j)^2$ et cov(x^j, x^k) la covariance empirique :

$$\text{cov}\,(x^j, x^k) = \frac{1}{n}\sum_{i=1}^{n}(x_i^j - \overline{x}^j)(x_i^k - \overline{x}^k)$$

La matrice de variance covariance empirique est la matrice de terme général cov (x^j, x^k), dont l'expression matricielle est :

$$\frac{1}{n}A'A - \overline{x}\,\overline{x}'.$$

Matrice de corrélation empirique

La matrice de corrélation empirique est la matrice de terme général

$$\frac{\text{cov}\,(x^j, x^k)}{\sqrt{\text{var}\,x^j \text{var}\,x^k}}.$$

Exemple (cf. données de l'exemple 2 page 6) :

	CYL	PUIS	LON	LAR	POIDS	VITESSE	PRIX
CYL	1.00000	0.79663	0.70146	0.62976	0.78895	0.66493	0.63858
PUIS	0.79663	1.00000	0.64136	0.52083	0.76529	0.84438	0.79870
LON	0.70146	0.64136	1.00000	0.84927	0.86809	0.47593	0.64376
LAR	0.62976	0.52083	0.84927	1.00000	0.71687	0.47295	0.54665
POIDS	0.78895	0.76529	0.86809	0.71687	1.00000	0.47760	0.75329
VITESSE	0.66493	0.84438	0.47593	0.47295	0.47760	1.00000	0.58176
PRIX	0.63858	0.79870	0.64376	0.54665	0.75329	0.58176	1.00000

1.2.3 Changement de variable

Soit la transformation $\varphi : I\!\!R^p \to I\!\!R^p$ qui à $x = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}$ associe $y = \varphi(x) = \begin{pmatrix} \varphi_1(x) \\ \vdots \\ \varphi_p(x) \end{pmatrix}$.

On suppose que *les applications* φ_i *sont biunivoques* c'est-à-dire φ_i^{-1} existe et $\varphi^{-1}(y) = \begin{pmatrix} \varphi_1^{-1}(y) \\ \vdots \\ \varphi_p^{-1}(y) \end{pmatrix}$.

Propriété de changement de variable

Soit $Y = \varphi(X)$. La densité g de Y s'exprime alors, en fonction de la densité f de X par :

$$g(y_1, \ldots, y_p) = f[\varphi^{-1}(y_1, \ldots, y_p)] \times \mid \det J \mid^{-1},$$

où det J, appelé Jacobien de la transformation, est le déterminant des dérivées partielles premières :

$$\det J = \begin{vmatrix} \dfrac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_p}{\partial x_1} \\ \vdots & & \vdots \\ \dfrac{\partial y_1}{\partial x_p} & \cdots & \frac{\partial y_p}{\partial x_p} \end{vmatrix} \text{ où } y_i = \varphi_i(x)$$

Exercice

Soit $X = (X_1, X_2)$ un couple de deux v.a. indépendantes et de même loi exponentielle de paramètre 1.

Quelle est la densité de $Y = \varphi(X) = \begin{pmatrix} X_1 + X_2 \\ X_2 \end{pmatrix}$?
En déduire que la loi de $X_1 + X_2$ est Gamma(1,2).

Combinaisons linéaires de variables aléatoires usuelles

Hypothèses		Loi de probabilité
X_i binomiale : $\mathcal{B}(n_i, p)$ (X_i) indépendantes ; $Y = \sum_{i=1}^{k} X_i$	$\Longrightarrow$	Y binomiale $\mathcal{B}\left(\sum_{i=1}^{k} n_i, p\right)$
X_i de Poisson : $\mathcal{P}(\lambda_i)$ (X_i) indépendantes ; $Y = \sum_{i=1}^{k} X_i$	$\Longrightarrow$	Y loi de Poisson $\mathcal{P}\left(\sum_{i=1}^{k} \lambda_i\right)$
X_i normale : $\mathcal{N}(\mu_i, \sigma_i^2)$ (X_i) indépendantes ; $Y = \sum_{i=1}^{n} a_i X_i$	$\Longrightarrow$	Y normale $\mathcal{N}\left(\sum_{i=1}^{k} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$
Cas particulier : $\mu_i = \mu$ et $\sigma_i = \sigma$ $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$	$\Longrightarrow$	$\overline{X}$ normale $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$
X_i loi gamma : $G(a, p_i)$ (X_i) indépendantes ; $Y = \sum_{i=1}^{n} X_i$	$\Longrightarrow$	Y loi gamma $G\left(a, \sum_{i=1}^{n} p_i\right)$
Cas particulier : X_i loi de χ^2 à $\gamma_i\ dd\ell$	$\Longrightarrow$	Y loi de χ^2 à $\sum_{i=1}^{n} \gamma_i\ dd\ell$

Changements de variables aléatoires usuels

Hypothèses		Loi de probabilité
X_i normales centrées réduites $\mathcal{N}(0,1)$ et indépendantes	$\Longrightarrow$	$\displaystyle\sum_{i=1}^{n} X_i^2$ loi de χ^2 à n $dd\ell$
X normale $\mathcal{N}(0,1)$, Y loi de χ^2 à n $dd\ell$, X et Y indépendantes	$\Longrightarrow$	$\dfrac{X}{\sqrt{\frac{Y}{n}}}$ loi de Student à n $dd\ell$
X_1 loi de χ^2 à n_1 $dd\ell$, X_2 loi de χ^2 à n_2 $dd\ell$, X_1 et X_2 indépendantes	$\Longrightarrow$	$\dfrac{\frac{X_1}{n_1}}{\frac{X_2}{n_2}}$ loi de Fisher Snedecor à n_1 et n_2 $dd\ell$
T loi de Student à γ $dd\ell$	$\Longrightarrow$	T^2 loi de Fisher $\gamma_1 = 1$ et $\gamma_2 = \gamma$

1.3 Loi normale dans $\mathbb{R}^n$ (Gaussienne)

1.3.1 Diverses caractérisations

Définition 1 : Un vecteur aléatoire X suit une loi gaussienne dans $\mathbb{R}^n$, si

$$\forall a = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}, \; a'X \text{ est une } \textit{v.a. gaussienne} \text{ réelle.}$$

Remarque : X gaussien dans $\mathbb{R}^n \Longrightarrow X_i$ v.a.r. gaussienne $\forall i = 1$ à n.

Exemple : Si X v.a. gaussien dans $\mathbb{R}^n$, A matrice $(m \times n)$ et B un vecteur de $\mathbb{R}^n$ (A et B non aléatoire), alors $y = AX + B$ est un vecteur gaussien de $\mathbb{R}^n$.

Notation :

$$X \sim \mathcal{N}_n(\mu, \Sigma)$$
$$E(X) = \mu \text{ et } V(X) = \Sigma$$

Définition 2. $X \sim \mathcal{N}_n(\mu, \Sigma)$ où Σ est une matrice carrée symétrique et de rang n si la densité de X s'écrit :

$$f_X(x_1, \dots, x_n) = (2\pi)^{-n/2} (\det \Sigma)^{-1/2} \exp\left\{ -\frac{1}{2} \parallel x - \mu \parallel^2_{\Sigma^{-1}} \right\}$$

où $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$

Applications :

1. Expliciter $f_X(x_1, x_2)$ dans le cas $n = 2$
2. Ecrire $f_X(x_1, \dots, x_n)$ si les n v.a. X_i sont indépendantes.

Propriété : Définition 1 $\Longleftrightarrow$ définition 2.
($\Rightarrow$) difficile (utilise les fonctions caractéristiques)
($\Leftarrow$) on utilise un théorème d'algèbre.

Σ symétrique, définie positive, inversible $\Rightarrow \exists$ matrice A d'ordre n, symétrique, inversible telle que

$$\Sigma = AA' = A^2 \Rightarrow \Sigma^{-1} = (A^{-1})^2$$

On considère alors $X = Ay + \mu$, ce qui est équivalent à $y = A^{-1} \cdot (X - \mu)$

Il suffit de vérifier que les composantes y_i sont $\mathcal{N}(0, 1)$ et indépendantes ce qui impliquera que $a'X = a'(Ay + \mu)$ est gaussien pour tout a, car $a'Ay$ est gaussien comme combinaison linéaire de v.a. *indépendantes* gaussiennes.

1.3.2 Non corrélation et indépendance

Théorème 1. Si $X \sim \mathcal{N}_2(\mu, \Sigma)$, alors cov $(X_1, X_2) = 0 \Longleftrightarrow X_1$ et X_2 indépendants.

Démonstration

($\Rightarrow$) toujours vrai dans le cas de lois quelconques.
($\Leftarrow$) cov $(X_1, X_2) = 0 \Rightarrow \Sigma = \text{ diag } (\sigma_1^2, \sigma_2^2)$.

La densité s'écrit :

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2} \exp -\frac{1}{2}\left\{\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right\}$$

c'est-à-dire
$$\underset{\text{densité conjointe}}{f_X(x_1, x_2)} = \underset{\text{produit des densités marginales}}{f_{X_1}(x_1) \times f_{X_2}(x_2)}$$

$\Rightarrow X_1$ et X_2 indépendantes

Théorème 2. Soit $Z = \begin{pmatrix} U \\ W \end{pmatrix} \sim \mathcal{N}_{u+w}(\mu, \Sigma)$ où U v.a. de $\mathbb{R}^u$ et W v.a. de $\mathbb{R}^w$.
La loi de U sachant $W = w$ est une loi gaussienne.

$$\mathcal{N}_u\big(E(U) + \text{cov}\,(U,W)(V(W))^{-1}(W - E(W)), \Omega\big)$$
$$\text{avec} \qquad \Omega = V(U) - \text{cov}\,(U,W)(V(W))^{-1}\text{cov}\,(W,U)$$

Démonstration (en exercice pour $u = 1$ et $w = 1$)

Conséquence du théorème 2

$$\text{cov}\,(U,W) = 0 \quad \Rightarrow \quad U \text{ et } W \text{ indépendants}$$

Théorème 3. Si U_1 et U_2 sont 2 sous-vecteurs d'un v.a. gaussien U, alors :

$$U_1 \text{ et } U_2 \text{ indépendants } \Longleftrightarrow \text{cov}\,(U_1, U_2) = 0$$

Démonstration

Soient U_1 v.a. de $\mathbb{R}^{n_1}$ et U_2 v.a. de $\mathbb{R}^{n_2}$, avec $n_1 + n_2 = n$

$(\Leftarrow)$
$$\text{cov}\,(U_1, U_2) = 0 \quad \Rightarrow \quad V(U) = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix} = \Sigma$$

où $\Sigma_1 = V(U_1)$ et $\Sigma_2 = V(U_2)$

$(\Leftarrow)$
$$[V(U)]^{-1} = \begin{pmatrix} \Sigma_1^{-1} & 0 \\ 0 & \Sigma_2^{-1} \end{pmatrix} = \Sigma^{-1}$$

$(\Leftarrow)$
$$\begin{aligned} f_U(u) &= (2\pi)^{-n/2}(\det V(U))^{-1/2} \exp\left\{-\tfrac{1}{2}(u - E(U))'(V(U))^{-1}(u - E(U))\right\} \\ &= f_{U_1}(u_1) \times f_{U_2}(u_2) \end{aligned}$$

$(\Rightarrow)$ Immédiat d'après la définition de la covariance.

1.3.3 Lois de distribution des formes quadratiques

Théorème 1. Théorème de COCHRAN
Soit $X \sim \mathcal{N}_n(\mu, \Sigma)$
Soit A et B 2 matrices carrées *symétriques* et *idempotentes* (non aléatoires), d'ordre n.
Soit a un vecteur non aléatoire de $I\!R^n$.
Soit les formes quadratiques $Q_1 = X'AX$ et $Q_2 = X'BX$

Alors :
(i) $A\Sigma B = 0 \Rightarrow Q_1$ et Q_2 indépendantes.
(ii) $A\Sigma a = 0 \Rightarrow Q_1$ et $a'X$ indépendantes.

Démonstration

(i) Soit $U = \begin{pmatrix} AX \\ BX \end{pmatrix}$, c'est un vecteur gaussien de dimension $(2n \times 1)$.

$$\text{cov}\,(AX, BX) = AV(X)B' = A\Sigma B$$

$A\Sigma B = 0 \Rightarrow \text{cov}\,(AX, BX) = 0 \Rightarrow AX$ et BX sont des vecteurs indépendants d'après théorème 2 du paragraphe précédent.

$$Q_1 = X'AX = X'AAX \text{ car } A \text{ idempotente}$$

$$\Rightarrow \quad Q_1 = (AX)'AX$$

De même $Q_2 = (BX)'BX$

$\Rightarrow Q_1$ et Q_2 sont indépendantes car elles sont fonctions respectivement de AX et BX, qui sont des variables indépendantes.

(ii) même démonstration, si on considère $U = \begin{pmatrix} AX \\ a'X \end{pmatrix}$,

Théorème 2. Soit $X \sim \mathcal{N}_n(\mu, \Sigma)$
Alors
(i) $\| X - \mu \|^2_{\Sigma^{-1}}$ suit une loi de chi-deux de paramètre n (notée $\chi^2(n)$).
(ii) Pour toute matrice A d'ordre n, symétrique, idempotente, de rang r ($r \leq n$) :
Si $\left| \begin{array}{l} \Sigma = I_n \\ \mu = 0 \end{array} \right.$, alors $X'AX \sim \chi^2(r)$

Démonstration :

(i) On utilise le théorème d'algèbre sur la décomposition d'une matrice symétrique, définie positive et inversible pour Σ.

$\Sigma = A \cdot A$ où A inversible, symétrique. On note $\Sigma^{-1/2} = A^{-1}$

$\Rightarrow \Sigma^{-1} = \Sigma^{-1/2}\Sigma^{-1/2}$

On pose $y = \Sigma^{-1/2}(X - \mu)$
$\Rightarrow y \sim \mathcal{N}_n(0, I_n)$

$$\text{En effet} \quad \begin{aligned} E(y) &= \Sigma^{-1/2}(E(X) - \mu) = 0 \\ V(y) &= \Sigma^{-1/2}V(X)\Sigma^{-1/2} = I_n. \end{aligned}$$

$\Rightarrow$ Les v.a.r. $y_1, \ldots, y_n$ (composants de y) sont indépendantes et de loi $\mathcal{N}(0, 1)$.

$\Rightarrow \displaystyle\sum_{i=1}^{n} y_i^2$ est une v.a.r. qui suit une loi de $\chi^2(n)$, d'après la définition d'une loi de chi-deux.

(ii) La diagonalisation de A donne :

$$A = PLP' \quad \text{où} \quad \left| \begin{array}{l} P \quad \text{matrice des vecteurs propres orthogonale} \\ \\ L \;=\; \mathrm{diag}(\lambda_1, \ldots, \lambda_n) \\ \quad\; = P'AP \text{ matrice des valeurs propres} \end{array} \right.$$

Soit λ une valeur propre de A et u le vecteur propre associé : $Au = \lambda u$

Or A idempotente $\Rightarrow \lambda^2 u = \lambda u$

$\Rightarrow \quad \lambda(\lambda - 1)u = 0$
$\Rightarrow \quad \lambda = 1$ ou $\lambda = 0$

Or rang $A = r \Rightarrow$ il a r valeurs propres non nulles, donc égales à 1.

$$\Rightarrow \quad L = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & 0 & \\ & & 1 & & & \\ & & & 0 & & \\ & 0 & & & \ddots & \\ & & & & & 0 \end{pmatrix}$$

Soit $y = P'X$,

on a

- $E(y) = 0$
- $V(y) = PV(X)P' = I_n$

car $V(X) = I_n$ et $PP' = I_n$.

Donc $y \sim \mathcal{N}_n(0, I_n)$.

$$X'AX = X'PLP'X = (P'X)'L(P'X) = \sum_{i=1}^{r} y_i^r \sim \chi^2_{(r)}$$

Car c'est la somme des carrés de r v.a.r. $\mathcal{N}(0,1)$ *indépendantes*.

Application :

$$\left.\begin{array}{l} X_i \sim \mathcal{N}(\mu, \sigma^2) \\ X_i \text{ indépendants} \end{array}\right\} \Rightarrow \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \overline{X})^2 \sim \chi^2_{(n-1)}$$

Chapitre 2

Modèle linéaire

Cadre général quand les variables exogènes sont non aléatoires.

Autre terminologie :
<u>RLM</u> (plusieurs v. exogènes quantitatives) : Régression Linéaire Multiple

<u>RLS</u> (une seule v. exogène quantitative) : Régression Linéaire Simple

2.1 Formulation du modèle linéaire

Données : Pour $i = 1, \ldots, n$ $\quad \underbrace{y_i}_{\text{endogène(quantitative)}} \quad \underbrace{x_i^1, \ldots, x_i^p}_{\text{exogènes}} \quad n > p$

Modèle linéaire :

(1) $$y_i = \beta_1 x_i^1 + \cdots + \beta_p x_i^p + \underbrace{\varepsilon_i}_{\substack{\text{résidu, bruit,...} \\ \text{(terme d'erreur)}}} \qquad (i = 1, \ldots, n)$$

Remarques :

- 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_i

  ε_i valeur non observable d'une v.a. ε_i

- On admet que $x_i^1 = 1$ pour tout $i = 1, n$ si on considère un terme constant dans (1).
- Les coefficients de régression β_j $(j = 1, p)$ 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 $x^1, \ldots, x^p$ (à une erreur ε près).

2.1.1 Ecriture matricielle

$$\underset{(n\times 1)}{y} = \underset{(n\times p)}{X} \cdot \underset{(p\times 1)}{\beta} + \underset{(n\times 1)}{\varepsilon} \tag{2}$$

2.1.2 Hypothèses du modèle linéaire

Le modèle linéaire relatif aux données $(\underset{(n\times 1)}{y}, \underset{(n\times p)}{X})$ est défini par les 3 hypothèses :

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

H_2 matrice X $\left| \begin{array}{l} X \text{ est non aléatoire} \\ \\ x_i^1 = 1 \text{ pour tout } i = 1, n \\ \\ \text{rg}(X) = p \end{array} \right.$

H_3 Sur les résidus.
ε est un vecteur aléatoire de $I\!\!R^n$ tel que :

$$E(\varepsilon) = 0 \qquad \text{(centré)}$$

$$\underset{(n\times n)}{V(\varepsilon)} = \sigma^2 I_n \quad \text{où } \sigma > 0$$

Remarques

- Hypothèse H_2 :
  $rg(X) = p \Longleftrightarrow$ Les p vecteurs colonnes $x^1, \ldots, x^p$ sont linéairement indépendants.
  Donc $n \geq p$.
- Hypothèse H_3 :
  $E(\varepsilon) = 0$ et $V(\varepsilon) = \sigma^2 I_n$.

$\Longleftrightarrow$ Les résidus $\varepsilon_1, \ldots, \varepsilon_n$ sont des v.a.r :

- centrées $E(\varepsilon_i) = 0$

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

(HOMOSCEDASTICITÉ)

- non corrélées $E(e_i e_j) = 0 \quad$ si $i \neq j$

– Paramètres à estimer :

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

2.2 Estimateur des moindres carrés ordinaires (MCO)

OLSE : “Ordinary Least Square Estimator”

2.2.1 Définition de l’estimateur MCO du paramètre β

Notation

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

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

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

Définition : L’estimateur MCO du paramètre β est le vecteur aléatoire de $I\!R^p$ qui minimise la somme des carrés des écarts $SS(\beta)$.
C’est-à-dire :

$$\hat{\beta}_{MCO} = \arg \min_{\beta \in I\!R^p} \| y - X\beta \|^2_{I_n}$$

Théorème 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 :

$$(1) \qquad \underset{(p\times 1)}{\frac{\partial SS(\hat{\beta})}{\partial \beta}} = 0 \Longleftrightarrow -2X'y + 2X'X\hat{\beta} = 0$$

Justification (cf. Chapitre 1) : $\left| \begin{array}{l} \dfrac{\partial u'a}{\partial u} = a \\ \\ \dfrac{\partial u'Au}{\partial u} = 2Au \text{ si } A \text{ symétrique} \end{array} \right.$

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

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

$$\frac{\partial^2 SS(\beta)}{(\partial\beta)(\partial\beta)'} = 2X'X \text{ matrice définie positive (cf. exercice 1 TD 1)}$$

$\Rightarrow \hat{\beta}$ est un minimum unique.

CQFD

Autre démonstration : $SS(\beta)$ peut se décomposer (cf. exercice 2 TD 4).

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

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

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

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

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

Théorème 2. 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 : 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 3. (de Gauss Markov). L'estimateur $\hat{\beta}$ est *meilleur* que tout estimateur $\tilde{\beta}$ linéaire et sans biais de β (c'est-à-dire $V(\tilde{\beta}) - V(\hat{\beta})$ matrice semi-définie *positive*).

Démonstration : Soit $\tilde{\beta}$ un estimateur linéaire et sans biais de β
$\Rightarrow$ il existe $\underset{(p\times n)}{A}$ telle que $\tilde{\beta} = Ay$ et $E(\tilde{\beta}) = \beta$

Montrons :

$$\forall u \in I\!R^p\ ,\ u'[V(\tilde{\beta}) - V(\hat{\beta})]u \geq 0$$

$$\begin{aligned} u' \cdot \quad & [V(\tilde{\beta}) - V(\hat{\beta})] \cdot u \\ & = u'[A \underbrace{V(y)}_{\sigma^2 I_n} A' - \sigma^2(X'X)^{-1}] \cdot u \\ & = \sigma^2 \cdot u'A[I_n - X(X'X)^{-1}X']A'u \end{aligned}$$

$$\text{car} \left| \begin{array}{l} E(u'\tilde{\beta}) = u'AE(y) = u'AX\beta \\ \\ E(u'\tilde{\beta}) = u'E(\tilde{\beta}) = u'\beta \end{array} \right\} \Rightarrow (u'AX - u')\beta = 0 \Rightarrow u'AX = u'$$

Or $M = I_n - X(X'X)^{-1}X'$ est une matrice symétrique et idempotente (cf. exercice 2 TD 1).

$$\begin{aligned} \text{Donc :} \quad \sigma^2 u'AMA'u & = \sigma^2 \parallel A'u \parallel_M^2 \\ & = \sigma^2(MA'u)'MA'u = \sigma^2 \parallel MA'u \parallel_{I_n}^2 \geq 0 \end{aligned}$$

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

2.2.3 Estimation de la variance résiduelle σ^2

Théorème : La statistique $\hat{\sigma}^2 = \dfrac{\parallel y - X\hat{\beta} \parallel_{I_n}^2}{n-p}$ est un estimateur sans biais de σ^2.

Démonstration :

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

$$\text{Or} \left| \begin{array}{l} \text{var}\, y = E[(y - X\beta)(y - X\beta)'] = \underset{(n\times n)}{E(yy')} - X \underset{(n\times n)}{\beta\beta' X'} \\ \\ \text{Var}\, y = \sigma^2 I_n \end{array} \right.$$

$$\Rightarrow E(yy') = \sigma^2 I_n + X\beta\beta' X'$$

$$\Rightarrow tr\,(M \cdot E(yy')) = tr(\sigma^2 M) + tr(\underbrace{MX\beta\beta' X'}_{=0}) \text{ car } MX = [I_n - X(X'X)^{-1}X']X = 0$$

$$\begin{aligned}
\Rightarrow E(\hat{\sigma}^2) &= \frac{1}{n-p}\sigma^2 trM \\
\text{Or } trM &= tr(I_n - X(X'X)^{-1}X') \\
&= trI_n - tr(X(X'X)^{-1}X') \\
&= n - tr(\underbrace{X'X(X'X)^{-1}}_{I_p}) = n - p
\end{aligned}$$

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

Remarque : $\hat{\sigma}^2 = \dfrac{\| y - X\hat{\beta} \|^2_{I_n}}{n-p} = \dfrac{RSS}{n-p} = \displaystyle\sum_{i=1}^{n} \hat{\varepsilon}_i^2$ où $\hat{\varepsilon}_i = y_i - \hat{y}_i = y_i - \displaystyle\sum_{j=1}^{p} x_i^j \hat{\beta}_j$.

$\hat{\varepsilon}_i$ est le *résidu estimé* et $\hat{y}_i$ est la valeur ajustée par le modèle de régression.

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

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

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

$\hat{\sigma}_{\hat{\beta}_j} = \sqrt{\hat{\sigma}^2 \cdot [(X'X)^{-1}]_{jj}}$ qui est appelé *erreur standard de* $\hat{\beta}_j$, noté e.s. $(\hat{\beta}_j)$.
Le terme $[(X'X)^{-1}]_{ij}$ représente le terme général de la matrice $(X'X)^{-1}$.

2.3 Interprétation géométrique

2.3.1 Représentation des individus

Chaque individu i $(i = 1, \dots, n)$ est considéré dans l'espace $I\!R^p$ par ses p coordonnées :

$$(y_i, x_i^2, \dots, x_i^p) \in I\!R^p \qquad \text{(on ne considère pas } x_i^1 = 1)$$

L'ensemble des n points de $I\!R^p$ est appelé "nuage de points" dans $I\!R^p$.

$\Rightarrow$ *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 $\Rightarrow$ biplot de y en fonction de x^j.

2.3.2 Espace des variables

Les vecteurs $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ et $x^j = \begin{pmatrix} x_1^j \\ \vdots \\ x_n^j \end{pmatrix}$ pour $j = 1, \dots, p$ sont des vecteurs d'un espace vectoriel $E \subset I\!R^n$.

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

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

Démonstration : (cf. exercice 2 TD 1).

Remarques :

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

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

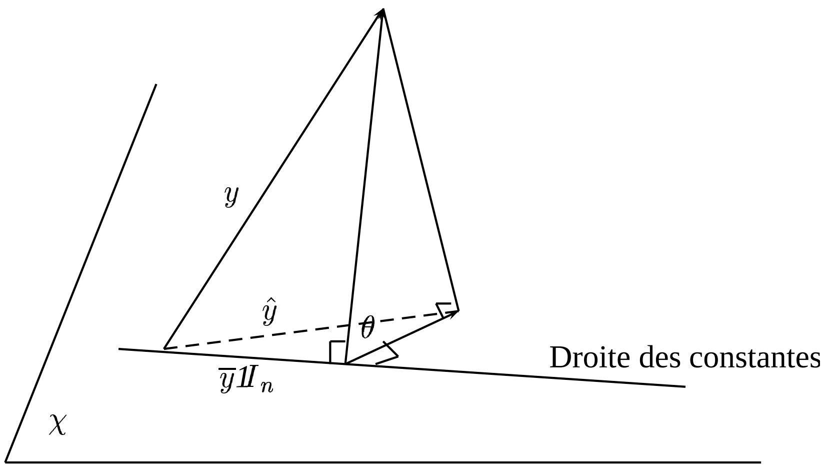
Les coefficients $\hat{\beta}_j$ estimés sont les coefficients de la combinaison linéaire des vecteurs $x^j \quad (j = 1, \dots, p)$.

b. $\displaystyle\sum_{i=1}^{n} \hat{y}_i = \sum_{i=1}^{n} y_i$ (Démonstration exercice 2 TD 4)

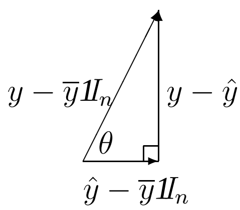
Propriété 2 : Le vecteur colonne unité de $\mathbb{R}^n$ étant noté $\underline{\mathbb{1}_n = x^1}$, la projection de y sur la droite des constantes (sous-espace engendré par x^1) est $\overline{y}\mathbb{1}_n$.

c'est-à-dire $$\mathbb{1}_n(\mathbb{1}_n'\mathbb{1}_n)^{-1}\mathbb{1}_n' y = \overline{y}\mathbb{1}_n$$

Graphique :



$$\cos^2\theta = \frac{\| \hat{y} - \overline{y}\mathbb{1}_n \|^2}{\| y - \overline{y}\mathbb{1}_n \|^2}$$



Propriété : Analyse de la variance de la régression.

$$\| y - \overline{y}\mathbb{1}_n \|^2 = \| y - X\hat{\beta} \|^2 + \| X\hat{\beta} - \overline{y}\mathbb{1}_n \|^2$$

c'est-à-dire $$\underbrace{\sum_{i=1}^{n}(y_i - \overline{y})^2}_{\text{TSS totale}} = \underbrace{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}_{\text{RSS résiduelle}} + \underbrace{\sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2}_{\substack{\text{variation}\\ \text{expliquée par}\\ \text{la régression}\\ \text{des } p \text{ var. } x^j}}$$

Démonstration : (en exercice 2 TD 4)

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

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

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

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

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

Remarque : Si $p = 2$, le coefficient R^2 n'est autre que le carré du coefficient de corrélation linéaire empirique de Pearson (cf. exercice 1 TD 4).

2.3.4 Interprétation et critique de l'utilisation de R^2

a) Exemple 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 page 44).

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

b) Exemple 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			

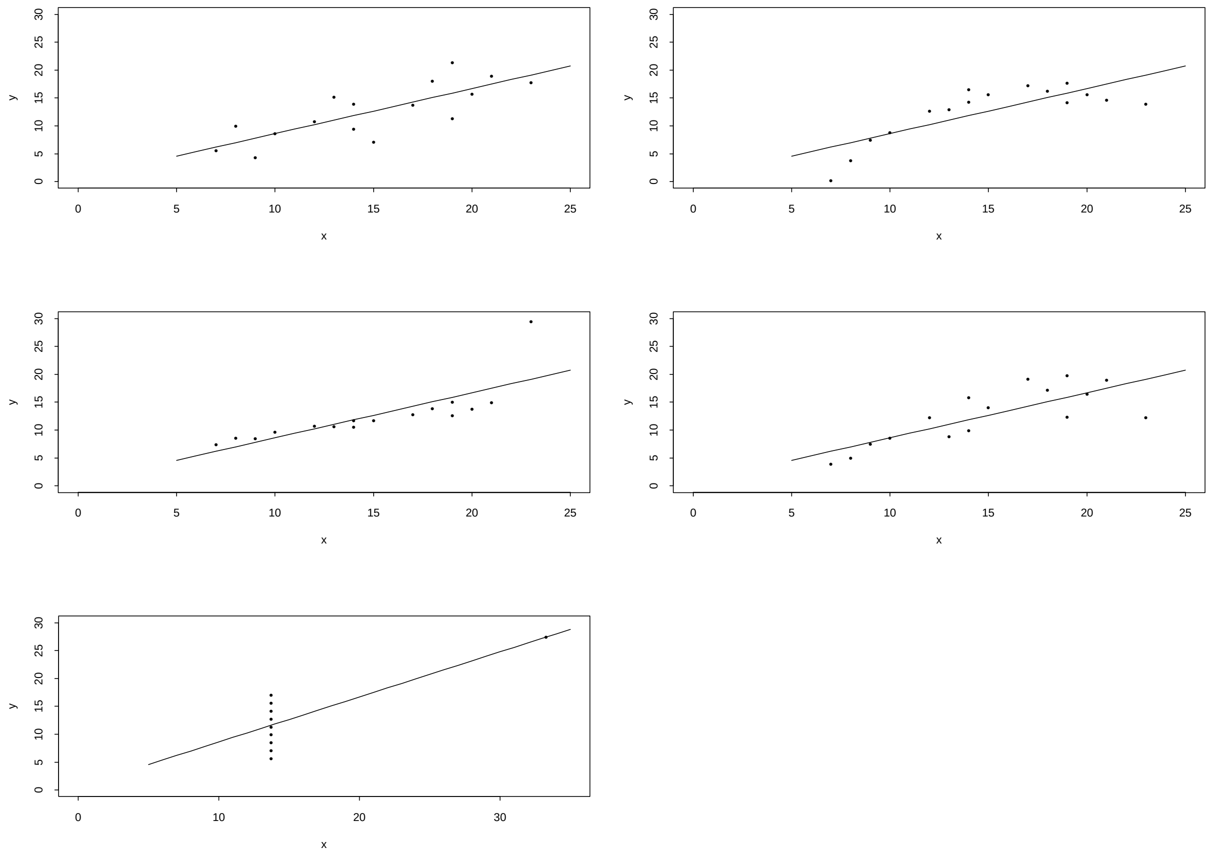


FIG. 2.1 – Graphique de 5 ensembles de données conduisant au même $R^2 = 0.617$. Réf. : TOMASSONE R., LESQUOY E., MILLIER C. (1983), “La régression : nouveaux regards sur une ancienne méthode statistique”, *Masson*.

R^2 = 0,7091 R^2 ajusté = 0,5504 $\hat{\sigma}$ = 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.)

- $R^2_{aj} = \dfrac{(n-1)R^2 - (p-1)}{n-p}$ coefficient de corrélation multiple (empirique) ajusté.
- $F = \dfrac{R^2}{1-R^2} \cdot \dfrac{n-p}{p-1} = \dfrac{\| \hat{y} - \overline{y}1\!\!1_n \|^2 /(p-1)}{\| \hat{y} - y \|^2 /(n-p)}$.
  Statistique de Fisher (cf. Chapitre 3), c'est la statistique du test d'égalité à zéro des $(p-1)$ coefficients $\beta_2 = \beta_3 = \cdots = \beta_p = 0$
- $\hat{\sigma}^2 = \dfrac{\| y - \hat{y} \|^2}{n-p}$ estimateur sans biais de la variance résiduelle σ^2 et qui s'exprime en fonction de R^2_{aj} :

$$\hat{\sigma}^2 = (1 - R^2_{aj}) \frac{\| y - \overline{y}1\!\!1_n \|^2}{n-1} \quad \text{(cf. exercice 2 TD 4).}$$

Chapitre 3

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.

3.1 Modèle linéaire gaussien

3.1.1 Définition

$H_1)$ $y = X\beta + \varepsilon$ (cf. Chapitre 2)

$H_2)$ X matrice non aléatoire de rang p (avec $x_i^1 = 1$)

$H_3')$ $\varepsilon \sim \mathcal{N}_n(0, \sigma^2 I_n)$

Remarque : L'hypothèse H_3' contient les hypothèses H_3 du modèle linéaire :

$$\left| \begin{array}{l} E(\varepsilon) = 0 \\ V(\varepsilon_i) = \sigma^2 \quad \text{homoscédasticité} \\ \text{cov}\,(\varepsilon_i, \varepsilon_j) = 0 \text{ pour } i \neq j \end{array} \right.$$

L'hypothèse $H_3' \Longleftrightarrow$ les v.a.r. $\varepsilon_1, \ldots \varepsilon_n$ sont *i.i.d.* de loi $\mathcal{N}(0, \sigma^2)$

Autre formulation :

$$H_1 \text{ et } H_3' \Longleftrightarrow y \sim \mathcal{N}_n(X\beta, \sigma^2 I_n)$$

structure statistique paramétrique :

$$\left\{ \underbrace{I\!R^n}_{\text{univers}}, \underbrace{\mathcal{B}_{I\!R^n}}_{\text{tribu}}, \underbrace{\mathcal{N}_n(X\beta, \sigma^2 I_n), (\beta, \sigma) \in I\!R^p \times I\!R^+}_{\substack{\text{probabilité} \\ \text{indicée par } (\beta, \sigma^2)}} \right\}$$

3.1.2 Estimateur maximum de vraisemblance EMV ou MLE

Propriété : L'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 $\hat{\beta}_{MCO} = \hat{\beta}_{MV}$.

Remarques :

1. L'estimateur MV du paramètre σ^2 est $\dfrac{\| y - X\hat{\beta} \|^2}{n}$.
   Démonstration : (en exercice)
   C'est un estimateur asymtôtiquement sans biais et convergent en probabilité.
   $\Rightarrow \hat{\sigma}^2 = \dfrac{\| y - X\hat{\beta} \|^2}{n-p}$ estimateur sans biais (cf. Chapitre 2).
2. Les estimateurs $\hat{\beta}$ et $\hat{\sigma}^2$ sont des estimateurs efficaces.

Démonstration : (en exercice)

3.1.3 Loi de probabilité des estimateurs

Propriété 1 : L'estimateur $\hat{\beta} = (X'X)^{-1}X'y$ est un vecteur aléatoire *gaussien* de $I\!R^p$

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

Démonstration : (cf. Chapitre 1, III), Combinaison linéaire de v.a. indépendantes gaussiennes.

Propriété 2 : La v.a.r. $\dfrac{\| y - X\hat{\beta} \|^2}{\sigma^2}$ suit une loi de chi-deux à $(n-p)$ d.d.ℓ.
c'est-à-dire :

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

Démonstration :

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

où $\quad M = I_n - X(X'X)^{-1}X' \qquad$ (rang $M = n - p$)

où $\quad \| y - X\hat{\beta} \|^2 = \varepsilon' M \varepsilon$ car $\hat{\varepsilon} = M\varepsilon$

D'après le théorème de COCHRAN (chapitre 1)

$$\left.\begin{array}{l} \bullet \quad \dfrac{1}{\sigma} \cdot \varepsilon \sim \mathcal{N}_n(0, I_n) \\[2ex] \bullet \quad M \text{ matrice symétrique idempotente de rang } n-p \end{array}\right\} \Rightarrow \frac{1}{\sigma^2}\varepsilon' M \varepsilon \sim \chi^2_{(n-p)}$$

Donc $\dfrac{1}{\sigma^2} \| y - X\hat{\beta} \|^2 \sim \chi^2_{(n-p)}$

Remarques :

• On retrouve

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

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

en effet $\quad E(Z) = \nu$ si $Z \sim \chi^2_{(\nu)}$

• On calcule $V(\hat{\sigma}^2) = \dfrac{2\sigma^2}{n-p}$

En effet $V(Z) = 2\nu$ si $Z \sim \chi^2_{(\nu)}$

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

Propriété 3 :

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

Démonstration : On considère les vecteurs gaussiens

$$\begin{array}{lll} & y - X\hat{\beta} = My = M\varepsilon & \text{de } I\!\!R^n \\ \text{et} & & \\ & (X'X)^{-1}X'\varepsilon & \text{de } I\!\!R^p \end{array}$$

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

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

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

$$\left.\begin{array}{l} \text{Donc cov}\,(u,w) = 0 \\ u \text{ et } w \text{ vecteurs gaussiens} \end{array}\right\} \Rightarrow u \text{ et } w \text{ indépendants (cf. chapitre 1)}$$

$$\begin{array}{lllll} \Rightarrow & y - X\hat{\beta} & \text{et} & (X'X)^{-1}X'\varepsilon & \text{indépendants} \\ \Rightarrow & y - X\hat{\beta} & \text{et} & \hat{\beta} & \text{indépendants} \\ \Rightarrow & \parallel y - X\hat{\beta} \parallel^2 & \text{et} & \hat{\beta} & \text{indépendants} \\ \Rightarrow & \hat{\sigma}^2 & \text{et} & \hat{\beta} & \text{indépendants} \end{array}$$

Propriété 4 :

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

$$T_j = \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}\sqrt{\gamma_{jj}}} \text{ suit une loi de Student de paramètre } n - p$$

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

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

Propriété 5 :

Soit L une matrice $q \times p$ de rang q $(q < p)$. La v.a.r. $\dfrac{1}{\sigma^2}[L(\hat{\beta} - \beta)]' \cdot [L(X'X)^{-1}L']^{-1} \cdot [L(\hat{\beta} - \beta)]$ suit une loi de chi-deux de paramètre $q \cdot (\chi^2_{(q)})$.

Démonstration :
- rang $(L) = q \Rightarrow$ rang$[L(X'X)^{-1}L'] = q$ (cf. chapitre 1).
- $L\hat{\beta}$ est un vecteur aléatoire gaussien de $I\!R^q$
- Var $(L\hat{\beta}) = L$ Var $\hat{\beta}L' = \sigma^2 L(X'X)^{-1}L'$ qui est inversible car définie positive de rang q.

$$\Rightarrow \quad L\hat{\beta} \sim \mathcal{N}_n(L\beta, \sigma^2 L(X'X)^{-1}L')$$

$$\Rightarrow \quad [L\hat{\beta} - L\beta]' \cdot \tfrac{1}{\sigma^2}[L(X'X)^{-1}L']^{-1} \cdot [L\hat{\beta} - L\beta] \sim \chi^2_{(q)}$$

d'après le théorème de COCHRAN (chapitre 1).

3.1.4 Estimation par intervalle de confiance (IC)

Propriété 1 : Un IC au coefficient de sécurité $\gamma = 1-\alpha$, d'un coefficient β_j (pour $j = 1, \ldots, p$) est :

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

où

γ_{jj}	$j$ème élément diagonal de $(X'X)^{-1}$
$t_{n-p};\dfrac{\alpha}{2}$	valeur d'une v.a. de Student de paramètre $n-p$ qui est dépassée avec une probabilité $\dfrac{\alpha}{2}$

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 $\sqrt{}$ d'une v.a. de χ^2.

Propriété 2 : Région de confiance (R.C) pour q paramètres β_j $(q \leq p)$. Une R.C au coefficient de sécurité $\gamma = 1 - \alpha$, de q coefficients $\beta_{j_1}, \dots, \beta_{j_q}$ est :

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

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 un probabilité α.

Démonstration : résulte de la propriété 5 du paragraphe précédent et de la propriété de changement de variable (chapitre 1)

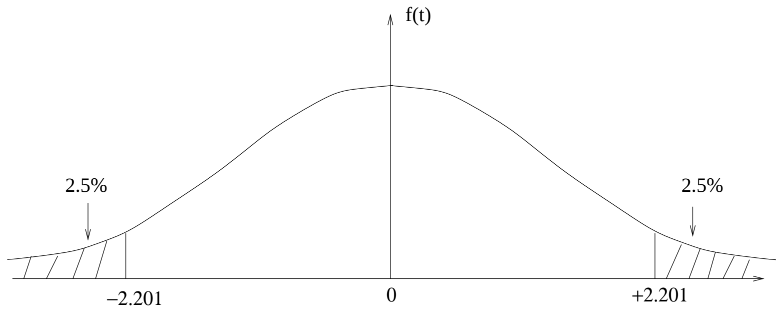
Application sur l'exemple 2 du chapitre 2.

Au coefficient de sécurité $\gamma = 95\%$, 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 $\quad n = 18 \quad p = 7.$

$$\left[\underbrace{\hat{\beta}_2}_{-3.505} - \underbrace{t_{n-p;\frac{\alpha}{2}}}_{t_{11;2.5\%}=2.201} \times \underbrace{\hat{\sigma}\sqrt{\gamma_{jj}}}_{\substack{\text{erreur standard} \\ 5.55}} , \hat{\beta}_2 + t_{n-p;\frac{\alpha}{2}} \times \hat{\sigma}\sqrt{\gamma_{jj}} \right]$$



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

$$[-15.7 \quad , \quad +8.7]$$

Remarque : $0 \in [-15.7 \quad , \quad +8.7]$
La valeur 0 est une valeur possible du paramètre β_j.

2. R.C. de β_2 et β_3

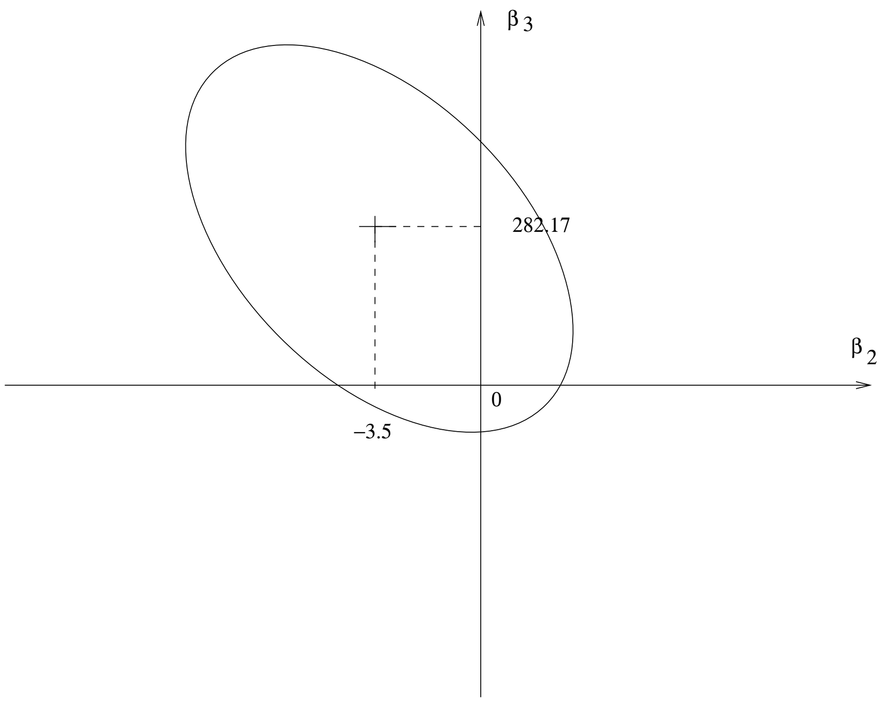
$$L_2 = \begin{pmatrix} 0 & 1 & 0 & & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \end{pmatrix} \Rightarrow L_2(\hat{\beta}_2 - \beta) = \begin{pmatrix} \hat{\beta}_2 & - & \beta_2 \\ \hat{\beta}_3 & - & \beta_3 \end{pmatrix}$$

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

où $\Big|$ Ω est l'inverse de $L_2(X'X)^{-1}L_2' = \begin{pmatrix} \gamma_{22} & \gamma_{23} \\ \gamma_{32} & \gamma_{33} \end{pmatrix}$ avec $(\gamma_{ij}) = (X'X)^{-1}$.

$\hat{\beta}_2 = -3.5 \quad \hat{\beta}_3 = 282.17 \quad \hat{\sigma}^2 \gamma_{22} = 5.55^2 \quad \hat{\sigma}^2 \gamma_{33} = 174.88^2 \quad \hat{\sigma}^2 \gamma_{23} = \hat{\sigma}^2 \gamma_{32} = 42.1$
$F_{2;11;5\%} = 3.98$

$J_{2;\alpha}(\beta_2, \beta_3) = \left\{ (\beta_2, \beta_3) / (3.5 + \beta_2)^2 \omega_1 + (\beta_3 - 282.17)^2 \omega_2 + \omega_3(-3.5 - \beta_2)(282.17 - \beta_3) \leq \omega_4 \right\}$
où $\omega_1, \omega_2, \omega_3, \omega_4$ sont déterminés par le calcul.



Ellipsoïde de confiance

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

La valeur prédite (ou ajustée) est :

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

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

$$[x_0'\hat{\beta} - t_{n-p;\frac{\alpha}{2}} \times \hat{\sigma}\sqrt{\varepsilon_0}; x_0'\hat{\beta} + t_{n-p;\frac{\alpha}{2}}\hat{\sigma}\sqrt{\varepsilon_0}]$$

où $\varepsilon_0 = 1 + x_0'(X'X)^{-1}x_0$ qui se déduit de la variance de y_0 :

$$\begin{aligned} \text{var}\,(y_0) &= \text{var}\,(x_0'\hat{\beta}) + \text{var}\,e_0 \\ &= x_0'\text{var}\,\hat{\beta}x_0 + \sigma^2 \\ &= \sigma^2 x_0'(X'X)^{-1}x_0 + \sigma^2 = \sigma^2\varepsilon_0 \end{aligned}$$

3.2 Tests d'hypothèses

3.2.1 Estimation sous contrainte linéaire

On souhaite estimer le paramètre β sous contrainte linéaire.

Exemples :

(1) $$\beta_1 = 0 \xrightarrow{\text{Ecriture matricielle}} (1\ 0\ \cdots\ 0)\beta = 0$$

(2) $$\beta_1 = \beta_2 = \beta_3 = 0 \longrightarrow \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & \cdots & \cdots \\ 0 & 0 & 1 & 0 & \cdots & 0 \end{pmatrix} \beta = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

(3) $$\left.\begin{array}{r} \beta_1 + \beta_2 = 6 \\ 2\beta_2 - \beta_3 = 9 \end{array}\right\} \longrightarrow \begin{pmatrix} 1 & 1 & 0 & \cdots & \cdots \\ 0 & 2 & -1 & \cdots & \cdots \end{pmatrix} \beta = \begin{pmatrix} 6 \\ 9 \end{pmatrix}$$

Contrainte linéaire :

$$R\beta = r$$

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 coefficients β_j.

Propriété : 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 :

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

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.

3.2.2 Test de student de signification d'un coefficient β_j

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 $R = (0\ 0\ 0 \underbrace{\ 1\ }_{\text{jème}} 0 \cdots 0)$ et $r = 0$.

Règle de décision, au niveau α :

$$\text{on rejette } H_0 \text{ si } t = \left| \frac{\hat{\beta}_j}{\hat{\sigma}_{\hat{\beta}_j}} \right| > t_{n-p;\frac{\alpha}{2}}$$

Démonstration : découle des propriétés du paragraphe 1.

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

3.2.3 Test de Fisher de signification de plusieurs coefficients

Soit q coefficients $\beta_{j1}, \ldots, \beta_{jq}$ $\quad (q \leq p)$.
On souhaite tester :

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

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

$$\text{où } \underset{(q\times p)}{R} = \begin{pmatrix} 0 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 0 & 1 & 0 & \cdots \\ \vdots & & & & & \\ 0 & \cdots & \cdots & \cdots & 1 & 0 \end{pmatrix} \text{ et } \underset{(q\times 1)}{r} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

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 :

$$R\hat{\beta} \sim \mathcal{N}_q(\underbrace{R\beta}_{\substack{R\beta = 0 \\ \text{si } H_0 \text{ vraie}}}, \underbrace{\sigma^2 R(X'X)^{-1}R'}_{\substack{\text{inversible} \\ \text{(déf. positive de rang } q)}})$$

Règle de décision, au niveau α.

$$\text{on rejette } H_0 \text{ si } F_q = \frac{[R\hat{\beta}]' \cdot [R(X'X)^{-1}R']^{-1}R\hat{\beta}}{q\hat{\sigma}^2} > \underbrace{F_{q;n-p;\alpha}}_{\substack{\text{F de Fisher,} \\ \text{dépassé avec une} \\ \text{probabilité } \alpha.}}$$

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 :
   – $R\hat{\beta} = \hat{\beta}_j, \qquad q = 1$
   – $[R(X'X)^{-1}R']^{-1} = (\gamma_{jj})^{-1}$ où $\gamma_{jj} = ((X'X)^{-1})_{jj}$
   – $F_1 = \dfrac{\hat{\beta}_j^2}{\hat{\sigma}^2 \cdot \gamma_{jj}} = \dfrac{\hat{\beta}_j^2}{e.s.(\hat{\beta}_j)}$

– $F_{1;n-p;\alpha} = t^2_{n-p;\alpha}$ (cf. propriété chapitre 1 : T Student $\Rightarrow T^2$ 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) : $R\beta = r = 0$.

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

Dans le modèle sans contrainte

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

Propriété : (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 :

$$\text{on rejette} H_0 \text{ si } F_q = \frac{(RSS_c - RSS)/q}{RSS/(n-p)} > F_{q;n-p;\alpha}$$

Remarque : Dans le cas $q = p-1$ et $\{j_1, \ldots, j_q\} = \{2, \ldots, p\}$ 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 2).

3.2.4 Test de Fisher d'une hypothèse linéaire générale

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

$$H_0 : R\beta = r \qquad R \text{ matrice } (q \times p) \text{ de rang } q$$

$$\text{contre} \quad H_1 : R\beta \neq r$$

La règle de décision de rejet de H_0 au niveau α est :

$$\text{On rejette } H_0 \text{ si } F_q = \frac{(R\hat{\beta} - r)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r)/q}{\| y - X\hat{\beta} \|^2 /(n-p)} > F_{q;n-p,\alpha}$$

Remarque : F_q peut encore s'écrire :

$$F_q = \frac{(\| y - X\hat{\beta}_c \|^2 - \| y - X\hat{\beta} \|^2)/q}{\| y - X\hat{\beta} \|^2 /(n-p)} = \frac{(RSS_c - RSS)/q}{\hat{\sigma}^2}$$

où $\hat{\beta}_c$ est l'estimateur MCO contraint par $H_0 : R\beta = r$

c'est-à-dire $\hat{\beta}_c = \hat{\beta} + (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(r - R\hat{\beta})$

3.3 Critiques du modèle linéaire gaussien

- Hypothèse de linéarité ?
- Hypothèse sur les résidus :
  - Normalité ?
  - Même variance ?
  - Non corrélation ?
- Quel sous-ensemble de variables exogènes faut-il retenir ?

3.3.1 Test de linéarité

On peut tester l'hypothèse de linéarité, dans le cas où on dispose de mesures répétées de la variable à expliquer :

pour chaque $\ell = 1, L$ $\quad x_\ell = (x_\ell^1, \ldots, x_\ell^p) \rightarrow \quad \underbrace{y_{\ell_1}, \ldots, y_{\ell_{n\ell}}}_{n_\ell \text{ mesures}}$

répétées $(n = \sum_{\ell=1}^{L} n_\ell)$

$\underline{(n > L > p)}$

Modèle linéaire gaussien :

$$y_{\ell_r} = (\beta_1 x_\ell^1 + \cdots + \beta_p x_\ell^p) + \underbrace{e_{\ell_r}}_{\text{i.i.d}}$$

pour $\ell = 1, L$ et $r = 1, n_\ell$

Ecriture matricielle :

$$y = X\beta + e$$

$$\text{où} \quad y = \begin{pmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{L1} \\ \vdots \\ y_{Ln_L} \end{pmatrix} \qquad X_{(n\times p)} = \begin{pmatrix} x_1^1 & x_1^2 & \cdots & x_1^p \\ \vdots & & & \\ x_1^1 & x_1^2 & \cdots & x_1^p \\ x_2^1 & x_2^2 & \cdots & x_2^p \\ \vdots & & & \\ x_2^1 & x_2^2 & \cdots & x_2^p \\ \vdots & & & \\ x_L^1 & x_L^2 & \cdots & x_L^p \\ \vdots & & & \\ x_L^1 & x_L^2 & \cdots & x_L^p \end{pmatrix} \begin{array}{l} \left.\vphantom{\begin{matrix}a\\a\\a\end{matrix}}\right\} n_1 \text{ fois} \\ \left.\vphantom{\begin{matrix}a\\a\\a\end{matrix}}\right\} n_2 \text{ fois} \\ \vdots \\ \left.\vphantom{\begin{matrix}a\\a\\a\end{matrix}}\right\} n_L \text{ fois} \end{array}$$

On veut éprouver les 2 hypothèses :
$H_0 : y = X\beta + e$ (la liaison entre y et x est *linéaire*),
contre $H_1 : y_{\ell_r} = m(x_\ell^1, \dots, x_\ell^p) + e_{\ell_r}$ $\begin{pmatrix} \ell = 1, L \\ r = 1, n_\ell \end{pmatrix}$ (la liaison entre y_ℓ et x_ℓ est traduite à l'aide d'une fonction m quelconque de $\mathbb{R}^p$ dans $\mathbb{R}$ qui n'est pas une application linéaire).

Remarque : On estime les paramètres par :
- sous H_0 : $\hat{\beta}$ estimateur MCO : $\hat{\beta} = (X'X)^{-1}X'y$
- sous H_1 : $\hat{m}(x_\ell) = \overline{y}_\ell = \dfrac{1}{n_\ell}\displaystyle\sum_{r=1}^{n_\ell} y_{\ell_r}$

Règle de décision, au niveau α

$$\text{On rejette } H_0 \text{ si } \frac{\displaystyle\sum_{\ell=1}^{L} n_\ell(\hat{y}_\ell - \overline{y}_\ell)^2/(L-p)}{\displaystyle\sum_{\ell=1}^{L}\sum_{j=1}^{n_\ell}(y_{\ell j} - \overline{y}_\ell)^2/(n-L)} > F_{L-p;n-L;\alpha}$$

où $\hat{y}_\ell = \hat{\beta}_1 x_\ell^1 + \cdots + \hat{\beta}_p x_\ell^p$

Remarque :

$$\underbrace{\sum_{\ell=1}^{L}\sum_{j=1}^{n_\ell}(y_{\ell j} - \hat{y}_\ell)^2}_{\|y - X\hat{\beta}\|^2} = \sum_{\ell=1}^{L}\sum_{j=1}^{n_\ell}(y_{\ell j} - \overline{y}_\ell)^2 + \sum_{\ell=1}^{L} n_\ell(\hat{y}_\ell - \overline{y}_\ell)^2$$

(à démontrer en exercice).

3.3.2 Analyse des résidus

Les résidus théoriques e_i sont estimés par $\hat{e}_i = y_i - \hat{y}_i$ qui ne sont pas de variance constante et qui sont corrélés.

$$E(\hat{e}_i) = 0$$

$$\text{var}\,(\hat{e}) = \sigma^2(I_n - \underbrace{X(X'X)^{-1}X'}_{P}) \Rightarrow \left| \begin{array}{ll} \text{cov}\,(\hat{e}_i, \hat{e}_j) = -\sigma^2 p_{ij} & \text{si } i \neq j \\ \\ \text{var}\,\hat{e}_i = \sigma^2(1 - p_{ii}) & \end{array} \right.$$

avec $P = (P_{ij})$.

Définitions :

1. On appelle résidus standardisés :

$$r_i = \frac{y_i - \hat{y}_i}{\hat{\sigma}} \qquad i = 1, \ldots, n.$$

2. On appelle *résidus studentisés* :

$$T_i = \frac{y_i - \hat{y}_i}{\hat{\sigma}\sqrt{1 - p_{ii}}} \qquad i = 1, \ldots, n$$

où $\hat{\sigma}^2 = \dfrac{\| y - X\hat{\beta} \|^2}{n - p}$ et p_{ii} élément diagonal de $P = X(X'X)^{-1}X'$

3. On appelle *résidus studentisés par validation croisée*

$$T_i^* = \frac{y_i - \hat{y}_i}{\hat{\sigma}_{(i)}\sqrt{1 - p_{ii}}} \qquad i = 1, \ldots, n$$

où $\hat{\sigma}_{(i)}$ est l'estimation de σ dans le modèle linéaire gaussien obtenu en supprimant l'observation i. Ce modèle est noté $(\underset{(n-1)\times 1}{y_{(i)}}, X_{(i)}\beta, \sigma^2 I_{n-1})$

Remarque : $\hat{y}_{(i)}$ est la $i$ème composante de

$$X_{(i)}(X'_{(i)}X_{(i)})^{-1}X'_{(i)}y_{(i)} \qquad \text{cf. exercice}$$

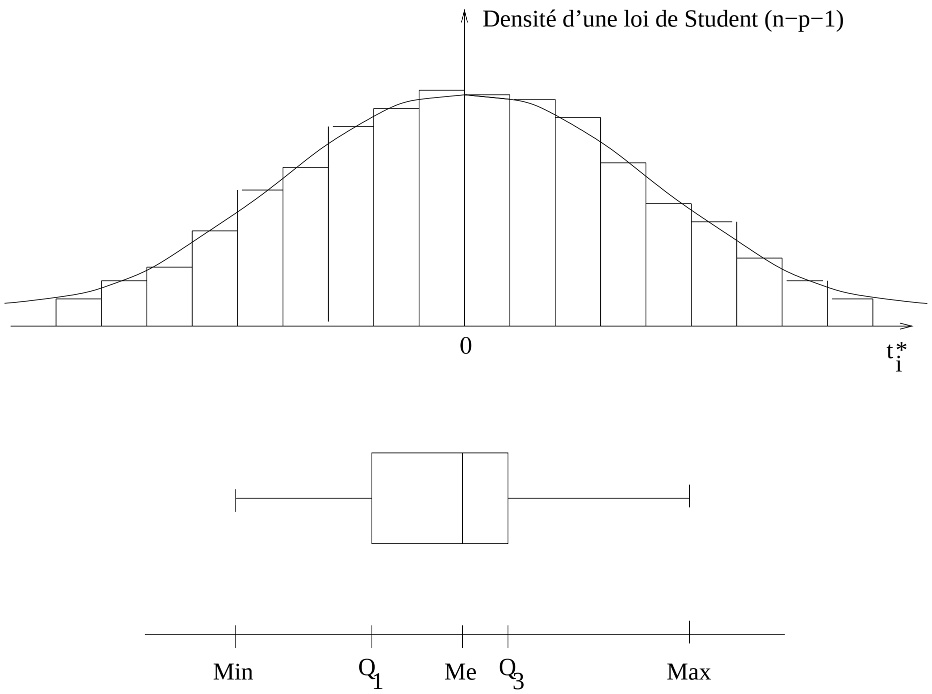
Propriété : Sous l'hypothèse $y \sim \mathcal{N}_n(X\beta, \sigma^2 I_n)$ les v.a. T_i^* (*résidus par validation croisée*) sont identiquement distribuées de loi de student à $n - p - 1$ d.d.ℓ.

Démonstration : cf "Régression non linéaire et applications", Antoniadis et al.

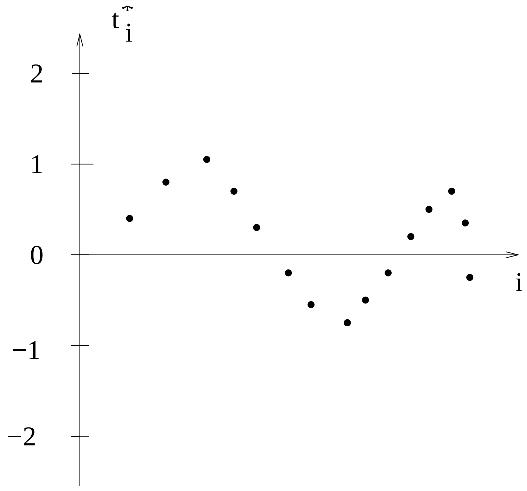
Dans la pratique, on utilisera les résidus studentisés T_i de la même façon que les résidus T_i^* par validation croisée, pour contrôler les hypothèses faites dans le modèle gaussien (les v.a. T_i sont considérées comme des v.a. de Student *quand* n *est grand*)

L'analyse des résidus consiste alors à réaliser des graphiques sur les résidus T_i^* ou T_i, pour vérifier l'hypothèse

$$y \sim \mathcal{N}(X\beta, \sigma^2 I_n)$$



Graphique 1
Histogramme (lissé) et boîte à moustaches
de la série empirique des résidus t_i^*

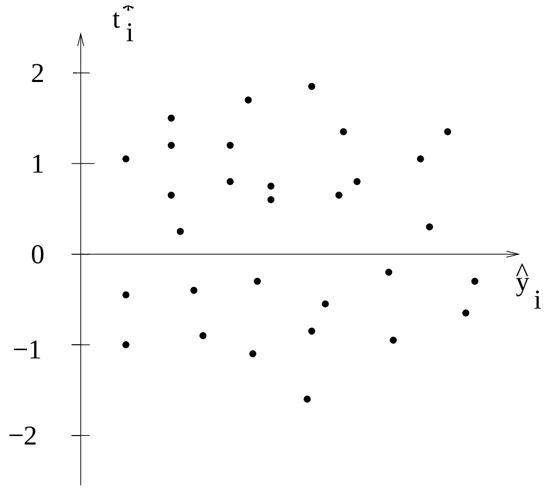


Graphique 2
Tracé chronologique des t_i^*
(uniquement si l'ordre est connu, cas des séries temporelles)

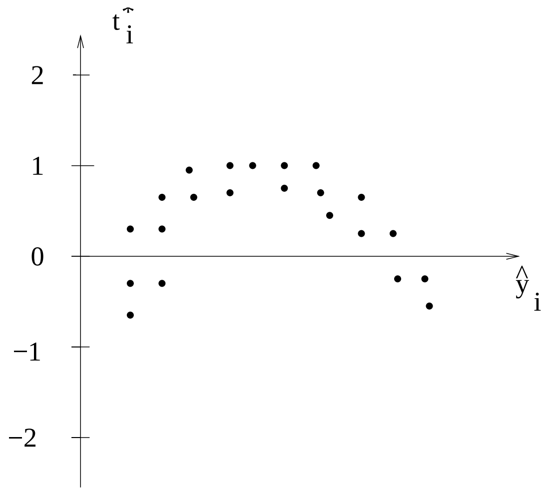
Dans l'exemple, amas de résidus positifs, puis négatifs $\Rightarrow$ hypothèse de non corrélation des résidus incorrecte.

Graphique 3

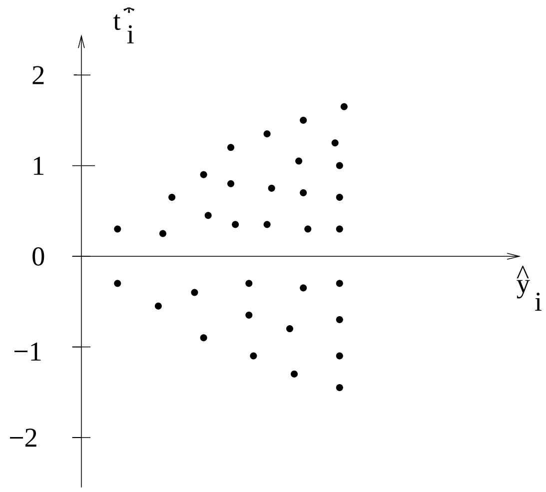
Bi-plot des résidus t_i^* en fonction de $\hat{y}_i$



Ajustement satisfaisant



Hypothèse de linéarité
ou de non corrélation *incorrecte*



Hypothèse d'homogénéité
des variances des résidus *incorrecte*

3.3.3 Détection de valeurs aberrantes ou influentes

Un exemple de valeur aberrante et de valeur influente a été présenté dans le chapitre 2 (cf. exemple 1). La définition de valeur aberrante varie d'un auteur à l'autre ; dans le cadre de ce cours, nous choisirons la définition suivante :

Définition : Etant donné un niveau α, une donnée aberrante est un point (y_i, x_i) pour lequel la valeur du résidu réduit t_i^* n'appartient pas à l'intervalle suivant :

$$[-t_{n-p-1;\frac{\alpha}{2}}, +t_{n-p-1;\frac{\alpha}{2}}]$$

Remarque : Un graphique bi-plot des résidus t_i^* (ou t_i) en fonction de y_i (ou de $\hat{y}_i$) permet de détecter les données aberrantes.

Donnée influente : Une donnée sera dite influente si le terme p_{ii} est "grand".

Exercice : Montrer que $p_{ii} = x_i'(X'X)^{-1}x_i$ où $x_i = \begin{pmatrix} x_i^1 \\ \vdots \\ x_i^p \end{pmatrix}$ et $\sum_{i=1}^{n} p_{ii} = p$.

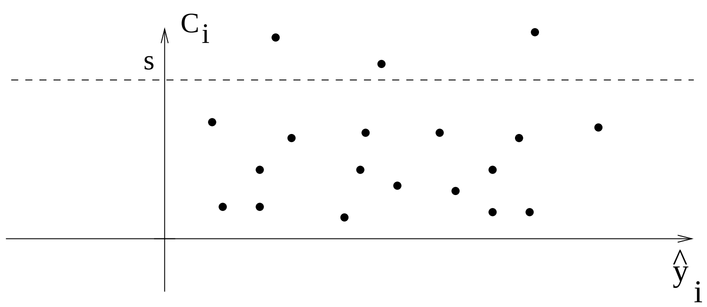
Une donnée influente intervient fortement dans les résultats de l'ajustement car :

$$\hat{y} = Py \ ; \ \text{var}\, \hat{\varepsilon}_i = \sigma^2(1 - p_{ii}) \ ; \ T_i = \frac{y_i - \hat{y}_i}{\hat{\sigma}\sqrt{1 - p_{ii}}}$$

Il existe plusieurs règles pour repérer si une donnée (y_i, x_i) est influente et pour préciser "p_{ii} grand". A titre d'exemple, on donne la distance de COOK très utilisée dans la pratique :

$$C_i = \frac{1}{p} \times \frac{p_{ii}}{1 - p_{ii}} \cdot T_i^2$$

Une donnée (y_i, x_i) sera influente si $C_i > s$ où s est un seuil, au choix du statisticien.



Exemple : $s = F_{p;n-p;\alpha}$ (v.a.r. de Fisher à p et $n-p$ d.d.ℓ. qui est dépassée avec une probabilité α)

Remarques :

1. $C_i = \frac{1}{p}\frac{\| \hat{y} - \hat{y}_{(i)} \|^2}{\hat{\sigma}^2}$
2. Dans les logiciels, des règles équivalentes à celles de COOK sont utilisées avec des seuils s qui sont fonction du caractère *très influent* ou *anormalement influent*.

3.3.4 Sélection des variables exogènes

Pour avoir un modèle plus stable et ayant un bon pouvoir prédictif, l'objectif est d'éliminer les variables redondantes.

a) Recherche exhaustive :

$$2^p - 1 \text{ modèles linéaires possibles}$$

Si on envisage tous les sous-ensembles de k variables exogènes parmi les p régresseurs :

Critères de choix de modèle

1. *Test* d'un modèle M_0 contre un modèle M_1 (avec $M_0 \subset M_1$)
2. Si l'objectif est de minimiser l'erreur de prévision (moyenne quadratique) pour une observation future $y_0 = x_0'\beta + e_0$ définie par : $MSEP(y_0) = E[(y_0 - x_0'\hat{\beta})^2]$. On utilise le critère PRESS qui est un estimateur sans biais de $MSEP(y) = \frac{1}{n}\sum_{i=1}^{n} E[(y_i - \hat{y}_i)^2]$, qui s'écrit :
$$PRESS = \frac{1}{n}\sum_{i=1}^{n}\frac{(y_i - \hat{y}_i)^2}{(1-p_{ii})^2}$$
$$\text{où} \quad p_{ii} = x_i'(XX)^{-1}x_i \text{ et } \hat{y} = X\hat{\beta} = X(X'X)^{-1}X'y$$
Remarques :
   - La v.a. PRESS est un estimateur *sans biais* de $MSEP(y)$.
   - Parmi tous les modèles linéaires correspondant aux sous-ensembles de k variables exogènes $(k \leq p)$, on choisira le modèle correspondant à la valeur PRESS minimum.
   - $$PRESS = \frac{\hat{\sigma}^2}{n}\sum_{i=1}^{n} T_i^2$$
$$\text{où} \quad T_i = \frac{y_i - \hat{y}_i}{\hat{\sigma}\sqrt{1-p_{ii}}} \qquad \text{résidu studentisé}$$

3. Autres critères de choix de modèles linéaires
   - Le coefficient R^2 de détermination *uniquement* si les modèles linéaires ont le *même nombre* k de paramètres non liés

$$R^2 = \frac{\Sigma(\hat{y}_i - \overline{y})^2}{\Sigma(y_i - \overline{y})^2}$$

   - R^2 ajusté maximum ou $\hat{\sigma}^2$ minimum (cf. Chapitre 2).
   - C_p Mallows minimum

$$C_p = \frac{\Sigma(y_i - \hat{y}_i)^2}{\tilde{\sigma}^2} + 2p$$

     où $\tilde{\sigma}^2$ est un estimateur de σ^2.
   - Critère AIC (d'Akaike) minimum.

$$AIC(M_k) = \log \underbrace{\hat{\sigma}^2_{M_k}}_{} + \frac{2k}{n}$$

$\downarrow$

Estimation de la variance
résiduelle dans le modèle M_k
contenant k coefficients de régression.

b) Recherche par méthode pas à pas : (*step technique*)

- Méthode ascendante : On introduit la variable explicative la plus corrélée (linéairement) avec y.
  On ajoute ensuite la variable qui fait progresser le plus le coefficient R^2.
  A chaque étape on teste (test de Student) la signification de chaque coefficient de régression.

- Méthode descendante : Dans le modèle linéaire contenant les p variables explicatives, on élimine la variable correspondant au t de Student le moins significatif.
  On recalcule les estimations des coefficients dans le modèle linéaire ayant $p-1$ variables.
  On recommence jusqu'à ce que tous les coefficients soient significatifs.

Application : Exemple 2, analyse des résultats obtenus par le logiciel SAS.

A. Ajustement du modèle complet

Variable	Coefficient estimé	Ecart-type	T si H_0 : coeff. = 0	Sig T	Facteur d'inflation*
Constante	-8239,362677	42718,42313	-0,193	0,8506	0
Cylindrée	-3,505182	5,550595	-0,631	0,5406	3,77201352
Largeur	208,693773	412,047884	0,506	0,6225	11,11881980
Longueur	-15,037660	129,747488	-0,116	0,9098	7,20419514
Poids	12,574678	24,622191	0,511	0,6197	4,19760475
Puissance	282,168803	174,882966	1,613	0,1349	9,95728271
Vitesse	-111,113551	222,256575	-0,500	0,6270	6,37511213

* facteur d'inflation : $\frac{1}{1-R_j^2}$ | R_j^2 carré du coefficient de corrélation multiple de la variable x^j et des $(p-2)$ autres.

Mais : le test de $H_0 : \beta_2 = \beta_3 = \cdots = \beta_7 = 0$ est significatif à 5 % ($P(F > 4.489) = 0.016$).

Résidus

	PRIX	PRIX ESTIME	E-TYPE PREDICT	LIMITE INF 95%	LIMITE SUP 95%	RESIDU	RESIDU STUDENT.	DISTANCE DE COOK
ALFASUD-TI-1350	30570	29616,1	2914,0	17989,1	41243,1	953,891	0,2886	0,009
AUDI-100-L	39990	36259,7	3572,5	23774,5	48744,8	3730,345	1,4463	0,573
SIMCA-1307-GLS	29600	31411,1	2486,0	20276,0	42546,3	-1811,149	-0,4978	0,017
CITROEN-CG-CLUB	28250	26445,8	2259,2	15547,2	37344,3	1804,249	0,4769	0,012
FIAT-132-1600GLS	34900	37043,0	2160,8	26241,5	47844,5	-2142,997	-0,5581	0,014
LANCIA-BETA-1300	35480	34972,8	2707,1	23590,7	46355,0	507,166	0,1459	0,002
PEUGEOT-504	32300	33749,1	1945,4	23147,9	44350,3	-1449,145	-0,3665	0,005
RENAULT-16-TL	32000	26580,0	2760,8	15135,5	38024,5	5420,043	1,5783	0,230
RENAULT-30-TS	47700	44445,6	3683,5	31805,2	57086,0	3254,423	1,3459	0,600
TOYOTA-COROLLA	26540	24650,2	3039,9	12868,1	36432,4	1889,759	0,5925	0,046
ALFETTA-1.66	42395	38270,5	3006,8	26529,5	50011,4	4124,538	1,2806	0,204
PRINCESS-1800-HL	33990	34830,4	2018,2	24163,5	45497,4	-840,418	-0,2146	0,002
DATSUN-200L	43980	44872,4	3343,6	32698,3	57046,6	-892,423	-0,3110	0,019
TAUNUS-2000-GL	35010	36343,5	2320,9	25382,4	47304,6	-1333,489	-0,3560	0,007
RANCHO	39450	35638,1	2453,2	24538,2	46737,9	3811,935	1,0415	0,070
MAZDA-9295	27900	32233,4	2726,5	20828,9	43638,0	-4333,420	-1,2519	0,139
OPEL-REKORD-L	32700	37103,5	2535,7	25914,2	48292,8	-4403,495	-1,2220	0,106
LADA-1300	22100	30389,8	2755,1	18952,0	41827,6	-8289,814	-2,4108	0,533

B. Recherche exhaustive

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. Modèle le meilleur (pour le critère $\hat{\sigma}$ ou $PRESS$) :

$$\text{Prix}_i = \beta_1 + \beta_2 \text{ Puis}_i + \beta_3 \text{ Poids}_i + e_i \qquad i = 1, \ldots, 18.$$

Résidus

	PRIX	PRIX ESTIME	LIMITE INF 95%	LIMITE SUP 95%	RESIDU	RESIDU STUDENT.	DISTANCE DE COOK
ALFASUD-TI-1350	30570	29752,5	20216,0	39289,0	817,5	0,2504	0,009
AUDI-100-L	39990	34738,6	26136,2	43341,0	5251,4	1,3845	0,042
SIMCA-1307-GLS	29600	30811,1	21981,3	39640,9	-1211,1	-0,3294	0,005
CITROEN-CG-CLUB	28250	27280,2	18325,9	36234,6	969,8	0,2687	0,004
FIAT-132-1600GLS	34900	36904,9	28170,9	45638,9	-2004,9	-0,5381	0,010
LANCIA-BETA-1300	35480	33726,2	25139,5	42312,8	1753,8	0,4614	0,004
PEUGEOT-504	32300	34523,4	25565,3	43481,4	-2223,4	-0,6164	0,023
RENAULT-16-TL	32000	27904,5	18637,2	37171,8	4095,5	1,1937	0,144
RENAULT-30-TS	47700	45630,9	36023,3	55238,6	2069,1	0,6429	0,066
TOYOTA-COROLLA	26540	24696,5	15274,9	34118,1	1843,5	0,5524	0,038
ALFETTA-1.66	42395	38067,3	28559,2	47575,3	4327,7	1,3183	0,245
PRINCESS-1800-HL	33990	35042,3	26191,4	43893,1	-1052,3	-0,2871	0,004
DATSUN-200L	43980	44204,9	34599,8	53810,1	-224,9	-0,0699	0,001
TAUNUS-2000-GL	35010	36493,6	27676,7	45310,6	-1483,6	-0,4028	0,007
RANCHO	39450	34186,3	25431,9	42940,8	5263,7	1,4166	0,074
MAZDA-9295	27900	34145,9	25549,8	42742,0	-6245,9	-1,6453	0,058
OPEL-REKORD-L	32700	37497,6	28742,6	46252,6	-4797,6	-1,2913	0,062
LADA-1300	22100	29248,2	20470,3	38026,2	-7148,2	-1,9302	0,147

Paramètres estimés

Variable	paramètre	écart-type	T	Sig T
POIDS	16,451161	10,774488	1,527	0,1476
PUIS	172,967225	72,419998	2,388	0,0305
(Constant)	1775,601201	8030,952416	0,221	0,8280

D. Régression pas à pas ascendante

Les analyses ont été réalisées avec le logiciel SPSS.

```
          * * * *   M U L T I P L E   R E G R E S S I O N   * * * *

Equation Number 1    Dependent Variable..   PRIX

Block Number  1.  Method:  Stepwise    Criteria   PIN 0,1500  POUT 0,2000
   CYL      LAR      LON      POIDS    PUIS     VITESSE

Variable(s) Entered on Step Number
   1..    PUIS

Multiple R          0,79870
R Square            0,63792
Adjusted R Square   0,61529
Standard Error   4076,00771

Analysis of Variance
                   DF      Sum of Squares      Mean Square
Regression         1      468334369,05604  468334369,05604
Residual           16     265821421,22173   16613838,82636

F =      28,18941       Signif F = 0,0001

------------------ Variables in the Equation ------------------

Variable               B           SE B      Beta     T    Sig T

PUIS          257,589788      48,516071  0,798700   5,309 0,0001
(Constant)  12363,652921    4215,922818             2,933 0,0098


------------- Variables not in the Equation -------------

Variable     Beta In    Partial  Min Toler        T     Sig T

CYL         0,006334   0,006363   0,365384     0,025    0,9807
LAR         0,179298   0,254366   0,728734     1,019    0,3245
LON         0,223392   0,284837   0,588654     1,151    0,2678
POIDS       0,342858   0,366762   0,414327     1,527    0,1476
```

```
VITESSE    -0,322784  -0,287389   0,287023    -1,162    0,2634

Variable(s) Entered on Step Number
   2..    POIDS

Multiple R          0,82863
R Square            0,68663
Adjusted R Square   0,64484
Standard Error   3916,33023

Analysis of Variance
                 DF      Sum of Squares      Mean Square
Regression        2     504091153,79101     252045576,89550
Residual         15     230064636,48677      15337642,43245

F =      16,43314       Signif F = 0,0002

------------------ Variables in the Equation ------------------

Variable                  B             SE B       Beta      T       Sig T

POIDS              16,451161       10,774488   0,342858     1,527   0,1476
PUIS              172,967225       72,419998   0,536314     2,388   0,0305
(Constant)       1775,601201     8030,952416                0,221   0,8280

------------- Variables not in the Equation -------------

Variable        Beta In     Partial    Min Toler         T     Sig T

CYL           -0,205562   -0,196994     0,287794    -0,752    0,4646
LAR            0,044471    0,055281     0,275311     0,207    0,8389
LON            0,008787    0,007772     0,172546     0,029    0,9772
VITESSE       -0,159514   -0,133170     0,117236    -0,503    0,6230

End Block Number   1   PIN =    0,150 Limits reached.
```

Remarque : Au niveau 15 %, on rejette l'hypothèse H_0 : "le modèle M_1 est bon" pour considérer que le modèle M_2 est meilleur que le modèle M_1.

Test de Fisher de $H_0 : \beta_2 = 0$ contre $\beta_2 \neq 0$

$$F = \frac{(RSS_1 - RSS_2)/1}{RSS_2/15} = 2.33 > F_{1;15;0.15}$$

Chapitre 4

Modèles linéaires gaussiens particuliers

On considère le cas où les variables exogènes peuvent être des variables qualitatives, c'est-à-dire des variables indicatrices dites "auxiliaires".

Exemple : Considérons le modèle linéaire simple relatif à une série chronologique comportant une tendance linéaire et une composante saisonnière trimestrielle (par exemple le volume d'importations ou le nombre de chômeurs ou le prix des biens alimentaires,...) :

$$y_{ts} = a + bt + \alpha_s + e_{ts} \qquad \begin{array}{ll} t = 1, \dots, n & \text{indice du mois} \\ s = 1, \dots, 4 & \text{indice du trimestre} \end{array}$$

où les paramètres α_s représentant l'effet de la saison s vérifient la contrainte $\sum_{s=1}^{4} \alpha_s = 0$.

Ecriture matricielle : Expliciter X dans le modèle linéaire suivant :

$$y = X\beta + e$$

$$\text{où} \qquad \underset{(n\times 1)}{y} = \begin{bmatrix} y_{11} \\ y_{21} \\ y_{31} \\ y_{41} \\ y_{52} \\ \vdots \\ y_{n4} \end{bmatrix} \qquad \underset{(5\times 1)}{\beta} = \begin{bmatrix} a \\ b \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} \qquad e = \begin{bmatrix} e_{11} \\ e_{21} \\ e_{31} \\ e_{41} \\ e_{52} \\ \vdots \\ e_{n4} \end{bmatrix}$$

Remarque : Si on n'impose pas la contrainte $\sum_{s=1}^{4} \alpha_s = 0$, pour que la matrice X soit de rang 5, il faut contraindre par exemple le terme constant : $a = 0$.

Dans ce cas le modèle s'écrit :

$$y = X\beta + e$$

$$\text{où } X = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 & 0 \\ 4 & 1 & 0 & 0 & 0 \\ 5 & 0 & 1 & 0 & 0 \\ 6 & 0 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} = [x^1 \quad \underbrace{x^2\ x^3\ x^4\ x^5}_{\substack{\text{variables indicatrices} \\ \text{de la saison “trimestre”}}} \quad] \text{ et } \beta = \begin{bmatrix} b \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix}$$

4.1 Analyse de variance à un facteur

La variable endogène est expliquée par une variable qualitative ayant I modalités, appelée facteur explicatif ou cause contrôlée ou variable indicatrice auxiliaire.

4.1.1 Exemple

On considère une étude sur les salaires de responsables de ventes auprès de 6 entreprises. On a obtenu le tableau des salaires annuels bruts en centaine de FF :

Entreprise i	1	2	3	4	5	6
	1602	1472	1548	1435	1493	1585
	1615	1477	1555	1438	1498	1592
	1624	1485	1559	1448	1509	1598
	1631	1493	1563	1449	1516	1604
		1496	1575	1454	1521	1609
		1504		1458	1523	1612
		1510		1467		
				1475		

Question : Peut-on considérer que les salaires moyens d'un responsable de ventes différent suivant le type d'entreprise ?

On souhaite donc étudier l'influence du type d'entreprise sur la variable salaire. Le type d'entreprise est un facteur ayant 6 niveaux.

4.1.2 Approche intuitive de l'analyse de variance

Notations : n : nombre total d'individus.

y_{ij} : observation de la variable aléatoire correspondant au salaire du $j$ème individu dans le niveau i $(i = 1, \ldots, I; j = 1, \ldots, n_i)$.

n_i : nombre d'individus du niveau i $(n = \sum_{i=1}^{I} n_i)$.

$\overline{y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}; \overline{y} = \frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i} y_{ij} = \frac{1}{n}\sum_{i=1}^{I}\overline{y}_i n_i.$

Equation d'analyse de variance

$$(E) \qquad \underbrace{\sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{ij}-\overline{y})^2}_{\text{VARIATION TOTALE}} = \underbrace{\sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{ij}-\overline{y}_i)^2}_{\substack{\text{VARIATION INTRA} \\ \text{(effet du hasard)}}} + \underbrace{\sum_{i=1}^{I} n_i(\overline{y}_i-\overline{y})^2}_{\substack{\text{VARIATION INTER} \\ \text{(effet du facteur } x)}}$$

Si le facteur x a un effet sur la variable exogène y, la variation INTER sera importante par rapport à la variation INTRA.

Modélisation statistique

$$(1) \qquad y_{ij} = \mu_i + e_{ij} \qquad (i = 1, \ldots, I; j = 1, \ldots, n_i)$$

Le paramètre μ_i représente l'effet du niveau i du facteur x. Le terme e_{ij} est la réalisation d'une variable aléatoire $\mathcal{N}(0, \sigma^2)$ et les n v.a.r. e_{ij} sont indépendantes.

Ecriture matricielle du modèle linéaire gaussien (1)

$$\boxed{\underset{(n\times 1)}{y} = \underset{(n\times I)(I\times 1)}{X\beta} + \underset{(n\times 1)}{e}}$$

avec $\beta = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_I \end{bmatrix} \qquad y = \begin{bmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ \vdots \\ y_{I1} \\ \vdots \\ y_{In_I} \end{bmatrix} \qquad X = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 1 & 0 & \cdots & \vdots \\ 0 & 1 & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & 1 & \cdots & \vdots \\ \vdots & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & 1 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \begin{array}{l} \left.\begin{array}{l} \\ \\ \\ \end{array}\right\} n_1 \text{ lignes} \\ \left.\begin{array}{l} \\ \\ \\ \end{array}\right\} n_2 \text{ lignes} \\ \vdots \\ \left.\begin{array}{l} \\ \\ \\ \end{array}\right\} n_I \text{ lignes} \end{array}$

La matrice X a pour ligne i : $(x_i^1, \ldots, x_i^I)$. Les vecteurs colonnes de X correspondant à $x = (x^1, \ldots, x^I)$ ne sont autres que les variables indicatrices des I niveaux du facteur x.

Comme dans les hypothèses du modèle linéaire gaussien, on suppose de plus :
$e \sim \mathcal{N}_n(0, \sigma^2 I_n)$ et on vérifie d'autre part : X matrice non aléatoire de rang I.

Autre paramétrisation

Si l'on décompose l'effet μ_i du niveau i en un effet moyen général $\mu = \frac{1}{n}\sum_{i=1}^{I} n_i \mu_i$ et un effet différentiel $\alpha_i = -\mu + \mu_i$, on peut reparamétriser le modèle (1) de façon équivalente pour obtenir la modélisation (2) :

$$y_{ij} = \mu + \alpha_i + e_{ij} \text{ avec la contrainte : } \sum_{i=1}^{I} n_i \alpha_i = 0 \qquad (i = 1, \ldots, I \quad j = 1, \ldots, n_i)$$

Ecriture matricielle du modèle linéaire gaussien (2) $\qquad y = X\beta + e$

$$\text{avec} \quad \underset{(I\times 1)}{\beta} = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{I-1} \end{bmatrix} \quad \text{et} \quad \underset{(n\times I)}{X} = \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 \\ 1 & \vdots & \vdots & \cdots & \vdots \\ \vdots & 1 & 0 & \cdots & \vdots \\ \vdots & 0 & 1 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & 1 & \cdots & \vdots \\ \vdots & \vdots & 0 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & 1 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \vdots & 0 & 0 & \cdots & 1 \\ \vdots & -\frac{n_1}{n_I} & -\frac{n_2}{n_I} & \cdots & -\frac{n_{I-1}}{n_I} \\ 1 & \vdots & \vdots & \cdots & \vdots \\ 1 & -\frac{n_1}{n_I} & -\frac{n_2}{n_I} & \cdots & -\frac{n_{I-1}}{n_I} \end{bmatrix} \begin{array}{l} \left.\begin{array}{l} \\ \\ \\ \end{array}\right\} n_1 \text{ lignes} \\ \left.\begin{array}{l} \\ \\ \\ \end{array}\right\} n_2 \text{ lignes} \\ \vdots \\ \left.\begin{array}{l} \\ \\ \\ \end{array}\right\} n_{I-1} \text{ lignes} \\ \left.\begin{array}{l} \\ \\ \\ \end{array}\right\} n_I \text{ lignes} \end{array}$$

et $e \sim \mathcal{N}_n(0, \sigma^2 I_n)$

4.1.3 Estimation des paramètres

On peut considérer les résultats obtenus dans le chapitre 3 pour le modèle linéaire gaussien, que l'on applique par exemple au modèle (1) correspondant à la matrice X des variables indicatrices des I niveaux du facteur x.

Propriété : l'EMCO (ou MV) de $\beta = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_I \end{bmatrix}$ est $\hat{\beta} = \begin{bmatrix} \overline{y}_1 \\ \vdots \\ \overline{y}_I \end{bmatrix}$ le vecteur des I moyennes des valeurs de y_{ij} pour chaque niveau i.

Démonstration :

Il suffit de calculer $\hat{\beta} = (X'X)^{-1}X'y$ en remarquant que $X'y$ a pour $i$ème composante $\sum_{j=1}^{n_i} y_{ij}$ et que $X'X = \text{diag}\,(n_1, n_2, \ldots, n_I)$.

Une deuxième démonstration possible est de minimiser $\| y - X\beta \|^2 = \sum_{i=1}^{i} \sum_{j=1}^{n_i} (y_{ij} - \mu_i)^2$, expression obtenue également en écrivant la vraisemblance de y.

Remarque : On vérifiera que $\hat{\mu} = \overline{y}$ et $\hat{\alpha}_i = \overline{y}_i - \overline{y}$, estimations des paramètres du modèle (2).

Estimation de la variance résiduelle σ^2 :

$$\hat{\sigma}^2 = \frac{\| y - X\hat{\beta} \|^2}{n - I} = \frac{\sum_{i=1}^{I} \sum_{j=1}^{n_i} (y_{ij} - \overline{y}_i)^2}{n - I}$$

d'après les propriétés démontrées dans le chapitre 3 : $\hat{\sigma}^2$ estimateur sans biais et Var $(\hat{\sigma}^2) = \frac{2\sigma^4}{n - I}$.

On retrouve au numérateur de $\hat{\sigma}^2$ la variation INTRA qui est la variation résiduelle due au hasard.

Intervalle de confiance, <u>au coefficient de confiance $\gamma = 1 - \alpha$</u>

a) d'un paramètre μ_i :

$$[\overline{y}_i - t_{n-I;\frac{\alpha}{2}} \times \frac{\hat{\sigma}}{\sqrt{n_i}}; \overline{y}_i + t_{n-I;\frac{\alpha}{2}} \times \frac{\hat{\sigma}}{\sqrt{n_i}}]$$

b) de q paramètres $\mu_{i_1}, \ldots, \mu_{i_q} \quad (q \leq I)$:

$$\mu_{i_j} \in \left[\overline{y}_{i_j} - \sqrt{b_{i_j}}, \overline{y}_{i_j} + \sqrt{b_{i_j}}\right] \qquad \text{pour } j = 1 \text{ à } q.$$

où $\quad b_{i_j} = \frac{q}{n_{i_j}} \times \hat{\sigma}^2 \times F_{q;n-I;\alpha}$

c) des différences $\mu_i - \mu_{i'}$:
On ordonne suivant les valeurs croissantes les estimations $\overline{y}_i$ des paramètres μ_i. On obtient alors $\hat{\mu}_{i_1} \leq \hat{\mu}_{i_2} \leq \cdots \leq \hat{\mu}_{i_I}$.
On considère alors les $(I - 1)$ différences $\mu_i - \mu_{i'}$ pour deux indices i et i' successifs. L'intervalle de confiance des $(I - 1)$ différences est alors :

$$\left[\overline{y}_i - \overline{y}_{i'} - \sqrt{b_{ii'}}, \overline{y}_i - \overline{y}_{i'} + \sqrt{b_{ii'}}\right] \text{ avec } b_{ii'} = (I-1)\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)\hat{\sigma}^2 f_{I-1;n-I;\alpha}.$$

Dans la pratique, on regarde si 0 appartient à l'intervalle en commençant par $\mu_{i_1} - \mu_{i_2}$ et on regroupe les 2 niveaux i_1 et i_2 si c'est le cas. On continue alors avec $\mu_{i_2} - \mu_{i_3}, \ldots$.

4.1.4 Test d'hypothèses

(On applique la théorie des tests vue au chapitre 3).

On souhaite tester l'hypothèse $H_0 : \mu_1 = \cdots = \mu_I = \mu$, c'est-à-dire la variable auxiliaire x n'a pas d'effet sur la variable à expliquer y.

Modèle quand H_0 est vraie

$$y_{ij} = \mu + e_{ij} \text{ avec } e_{ij} \; v.a. \; i.i.d. \text{ de loi } \mathcal{N}(0, \sigma_0^2)$$

Les paramètres μ et σ_0^2 sont estimés par : $\hat{\mu} = \overline{y}$ et $\hat{\sigma}_0^2 = \frac{1}{n-1}\sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{ij} - \overline{y})^2$.

Tableau d'analyse de variance

Source de variation	Somme des carrés RSS (SCR)	Degré de liberté d.f. (d.d.ℓ)	Somme des carrés moyens	Statistique F de Fisher
INTER (due au facteur x)	$\sum_{i=1}^{I} n_i(\overline{y}_i - \overline{y})^2$	$I-1$	$\frac{\sum_{i=1}^{I} n_i(\overline{y}_i - \overline{y})^2}{I-1}$	$F_0 = \frac{\sum_{i=1}^{I} n_i(\overline{y}_i - \overline{y})^2}{(I-1)\hat{\sigma}^2}$
INTRA (due au hasard)	$\sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{ij} - \overline{y}_i)^2$	$n-I$	$\hat{\sigma}^2 = \frac{\sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{ij} - \overline{y}_i)^2}{n-I}$	
TOTALE	$\sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{ij} - \overline{y})^2$	$n-1$	$\hat{\sigma}_0^2 = \frac{\sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{ij} - y)^2}{n-1}$	

La statistique de Fisher, compte tenu de l'équation (E) s'écrit aussi :

$$F_0 = \frac{\left[\sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{ij} - \overline{y})^2 - \sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{ij} - \overline{y}_i)^2\right]/(I-1)}{\sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{ij} - \overline{y}_i)^2/(n-I)}.$$

Règle de décision : au niveau α de l'hypothèse H_0 contre H_1 : non H_0

On rejette H_0 si $F_0 > F_{I-1;n-I;\alpha}$ où $f_{I-1;n-I;\alpha}$ est déterminée d'après la v.a. $F_{I-1;n-I}$ de Fisher par :

$$P[F_{I-1;n-I} > F_{I-1;n-I;\alpha}] = \alpha.$$

Remarque : Dans les sorties de logiciel on dispose de la "tail probability" :
$P(F_{I-1;n-I} > F_0)$, et on rejette alors H_0 si $P(F_{I-1;n-I} > F_0) \leq \alpha$.

4.2 Analyse de variance à deux facteurs

La variable endogène est expliquée par 2 variables qualitatives ayant respectivement I et J modalités ou niveaux.

4.2.1 Exemple

Le tableau ci-dessous donne les rendements laitiers de 40 vaches pour des modes d'alimentation différents. On se propose d'étudier l'influence des 2 variables Dose et Produit de base sur le rendement laitier.

Produit de base Dose	Paille	Foin	Herbe	Aliments ensilés
Dose faible	8 11 11 10 7	12 13 14 11 10	10 12 12 13 14	17 13 17 14 13
Dose forte	8 9 8 10 9	10 7 10 12 11	11 9 11 11 12	17 19 17 16 21

Notation :

n : Nombre total d'observations.

n_{ij} : Nombre d'observations relatives au niveau i du facteur F_1 et au niveau j du facteur F_2 $(i = 1, \dots, I$ et $j = 1, \dots, J)$

y_{ij_k} : Observation de la variable y pour le $k$ième individu dans la cellule $(i, j) \in I \times J$.

X : Matrice $(n \times IJ)$ dont les vecteurs colonnes sont les indicatrices x^{ij} de la cellule $(i, j) \in \{1, \dots, I\} \times \{1, \dots, J\}$

$$\overline{y} = \frac{1}{n} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{n_{ij}} y_{ij_k} \ ; \ \overline{y}_{ij} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} y_{ij_k} \quad \text{pour } (i, j) \text{ tel que } n_{ij} \geq 1.$$

Définition : Le plan d'expérience est *complet* si $n_{ij} \geq 1$ pour tout couple (i, j).
Le plan est dit *équilibré* si $n_{ij} = \frac{n}{IJ}$ pour tout couple (i, j).

4.2.2 Modélisation statistique et estimation

(On considère un plan complet).

L'observation y_{ij_k} se décompose additivement en un effet fixe μ_{ij} des niveaux i et j des facteurs F_1 et F_2 respectivement, et en un effet aléatoire e_{ij_k} :

$$y_{ij_k} = \mu_{ij} + e_{ij_k} \qquad \text{où} \qquad e_{ij_k} \text{ i.i.d. } \mathcal{N}(0, \sigma^2)$$

C'est donc un modèle linéaire gaussien qui s'écrit matriciellement :
(3)

$$\underset{(n\times 1)}{y} = X\beta + e \qquad \text{avec} \quad \underset{(IJ\times 1)}{\beta} = \begin{bmatrix} \mu_{11} \\ \vdots \\ \mu_{1J} \\ \mu_{21} \\ \vdots \\ \mu_{2J} \\ \vdots \\ \mu_{I1} \\ \vdots \\ \mu_{IJ} \end{bmatrix} \quad \underset{(n\times IJ)}{X} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 1 & 0 & \cdots & \vdots \\ 0 & 1 & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & 1 & \cdots & \vdots \\ \vdots & 0 & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & 0 \\ \vdots & \vdots & \cdots & 1 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \begin{array}{l} \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} n_{11} \text{ lignes} \\ \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} n_{12} \text{ lignes} \\ \quad \vdots \\ \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} n_{IJ} \text{ lignes} \end{array}$$

$$\text{où } e \sim \mathcal{N}_n(0, \sigma^2 I_n)$$

La variable x est la variable indicatrice des IJ niveaux des 2 facteurs F_1 et F_2.

Autre paramétrisation : (Voir section 4.2.4, pour la signification des interactions γ_{ij}^{12}).

$$(4) \qquad y_{ij_k} = \mu + \alpha_i^1 + \beta_j^2 + \gamma_{ij}^{12} + e_{ij_k} \qquad \text{où } e_{ij_k} \text{ i.i.d. } \mathcal{N}(0, \sigma^2)$$

avec les contraintes :

$$\sum_{i=1}^{I} \alpha_i^1 = \sum_{j=1}^{J} \beta_j^2 = \sum_{i=1}^{I} \gamma_{ij}^{12} = \sum_{j=1}^{J} \gamma_{ij}^{12} = 0.$$

Exercice : Dans le cas où $I = 2$, $J = 3$ et $n_{ij} = 2$, écrire matriciellement le modèle d'analyse de variance selon les paramétrisations (3) et (4). Combien de paramètres non liés doit-on estimer dans la modélisation (4) ?

Estimation des paramètres, pour la modélisation (3)

$$\hat{\mu}_{ij} = \overline{y}_{ij} \text{ pour tout couple } (i,j) \in \{1, \ldots, I\} \times \{1, \ldots, J\}$$
$$\text{(composante de } \hat{\beta} = (X'X)^{-1}X'y)$$
$$\hat{\sigma}^2 = \frac{\| y - X\hat{\beta} \|^2}{n - IJ} = \frac{\displaystyle\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{n_{ij}} (y_{ij_k} - \overline{y}_{ij})^2}{n - IJ}$$

Démonstration : Soit en calculant $\hat{\beta}$ (cf Chapitre 2), soit en maximisant la vraisemblance.

Estimation par intervalle de confiance, de q paramètres μ_{ij} conjointement au coefficient de confiance $\gamma = 1 - \alpha$:

$$\left[\overline{y}_{ij} - \sqrt{b_{ij}}\ ;\ \overline{y}_{ij} + \sqrt{b_{ij}}\right] \qquad \text{avec } b_{ij} = \frac{q}{n_{ij}} \cdot \hat{\sigma}^2 \cdot f_{q,n-IJ;\alpha}$$

4.2.3 Tests d'hypothèses

On peut s'intéresser aux 3 hypothèses nulles suivantes :

H_0^1 : $\alpha_i^1 = 0$ pour tout $i = 1, \ldots, I$ "le facteur F_1 n'a pas d'effet moyen sur y".

H_0^2 : $\beta_j^2 = 0$ pour tout $j = 1, \ldots, J$ "le facteur F_2 n'a pas d'effet moyen sur y''.

H_0^2 : $\gamma_{ij}^{12} = 0$ pour tout $(i,j) \in \{1, \ldots, I\} \times \{1, \ldots, J\}$ "les facteurs F_1 et F_2 n'ont pas d'effet d'interaction sur y.

1. Cas général d'un plan complet déséquilibré (c'est-à-dire n_{ij} différents).

Pour chacune des hypothèses H_0^1, H_0^2 ou H_0^3, on considère les matrices correspondant aux contraintes linéaires : R^1, R^2 ou R^3.
On utilise alors le test de Fisher d'une hypothèse linéaire (cf. II.4 Chapitre 3).
La statistique du test de Fisher ne s'explicite pas simplement dans le cas d'un plan complet déséquilibré.

2. Cas d'un plan complet équilibré (c'est-à-dire $n_{ij} = \frac{n}{IJ}$ que l'on notera K.)

On peut établir dans ce cas l'équation d'analyse de variance qui donne la décomposition de la variation totale de façon orthogonale :

$$(E) \qquad \begin{aligned} \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(y_{ij_k} - \overline{y})^2 &= \sum_i\sum_j\sum_k(y_{ij_k} - \overline{y}_{ij})^2 + JK\sum_i(\overline{y}_{i\cdot} - \overline{y})^2 \\ &+ IK\sum_j(\overline{y}_{\cdot j} - \overline{y})^2 + K\sum_i\sum_j(\overline{y}_{ij} - \overline{y}_{i\cdot} - \overline{y}_{\cdot j} + \overline{y})^2 \end{aligned}$$

$$\text{avec } \overline{y} = \frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} y_{ij_k} \ ; \ \overline{y}_{i\cdot} = \frac{1}{J}\sum_{j=1}^{J}\overline{y}_{ij} \ ; \ \overline{y}_{\cdot j} = \frac{1}{I}\sum_{i=1}^{I}\overline{y}_{ij}$$

On en déduit, dans le cas d'un plan complet équilibré, le tableau d'analyse de variance :

Source de variation	Somme des carrés	d.d.ℓ.	Statistique du test de Fisher	Valeur théorique	Hyp. nulle
Due au facteur F_1 $RSS_{23} - RSS$	$JK\sum_{i=1}^{I}(\overline{y}_{i.} - \overline{y})^2$	$I-1$	$\dfrac{JK\sum_{i=1}^{I}(\overline{y}_{i.} - \overline{y})^2}{\hat{\sigma}^2(I-1)}$	$F_{I-1;n-IJ;\alpha}$	H_0^1
Due au facteur F_2 $RSS_{13} - RSS$	$IK\sum_{j=1}^{J}(\overline{y}_{.j} - \overline{y})^2$	$J-1$	$\dfrac{IK\sum_{j}(\overline{y}_{.j} - \overline{y})^2}{\hat{\sigma}^2(J-1)}$	$F_{J-1;n-IJ;\alpha}$	H_0^2
Due à l'interaction $F_1 \times F_2$ des 2 facteurs $RSS_{12} - RSS$	$K\sum_{i}^{I}\sum_{j}^{J}(\overline{y}_{ij} - \overline{y}_{i.} - \overline{y}_{.j} + \overline{y})^2$	$(I-1) \times (J-1)$	$\dfrac{\sum_{i}\sum_{j}(\overline{y}_{ij} - \overline{y}_{i.} - \overline{y}_{.j} + \overline{y})^2}{(I-1)(J-1)\hat{\sigma}^2}$	$F_{(I-1)(J-1);n-IJ;\alpha}$	H_0^3
Résiduelle RSS	$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(y_{ij_k} - \overline{y}_{ij})^2$	$n-IJ$			
Totale TSS	$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(y_{ij_k} - \overline{y})^2$	$n-1$	$\dfrac{R^2}{1-R^2} \times \dfrac{n-IJ}{IJ-1}$	$F_{n-1;n-IJ;\alpha}$	H_0^4

$$\text{où } \hat{\sigma}^2 = \frac{\sum_i \sum_j \sum_k (y_{ij_k} - \overline{y}_{ij})^2}{n - IJ}$$

L'hypothèse nulle H_0^4 correspond à la nullité des $IJ-1$ paramètres $\alpha_i^1, \beta_j^1, \gamma_i^1$ ce qui revient encore à tester tous les paramètres μ_{ij} égaux à μ.

La statistique du test de l'hypothèse H_0^4 s'explicite alors en fonction du coefficient de détermination : (cf Chapitre 2)

$$R^2 = \frac{\displaystyle\sum_i \sum_j \sum_k (\overline{y}_{ij} - \overline{y})^2}{\displaystyle\sum_i \sum_j \sum_k (y_{ij_k} - \overline{y})^2} = \frac{\| \hat{y} - \overline{y}\mathbb{1}_n \|^2}{\| y - \overline{y}\mathbb{1}_n \|^2}$$

Notation : RSS_{23} (resp. RSS_{13} et RSS_{12}) désigne la somme des carrés des résidus estimés sous l'hypothèse $H_2 \cap H_3$ (resp. $H_1 \cap H_3$ et $H_1 \cap H_2$)

Application : Pour l'exemple II.1, construire le tableau d'analyse de variance.
Au niveau 5%, peut-on conclure s'il y a un effet moyen de la dose sur le rendement laitier ? s'il

y a un effet moyen du produit de base ? s'il y a une interaction des facteurs dose et produit ? Expliquer la signification de l'effet de l'interaction, en précisant quel est l'effet de la dose sur le rendement laitier.

4.2.4 Interprétation de l'interaction de deux facteurs

On considère le modèle linéaire gaussien (3) :

$$y_{ij_k} = \mu_{ij} + e_{ij_k} \quad \begin{array}{l} i = 1, \ldots, I \\ j = 1, \ldots, J \\ k = 1, \ldots, n_{ij} \end{array}$$

Le paramètre μ_{ij} peut se décomposer :

$$\mu_{ij} = \overline{\mu}_{..} + (\overline{\mu}_{i.} - \overline{\mu}_{..}) + (\overline{\mu}_{.j} - \overline{\mu}_{..}) + (\mu_{ij} - \overline{\mu}_{i.} - \overline{\mu}_{.j} + \overline{\mu}_{..})$$

avec

$$\overline{\mu}_{..} = \frac{1}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} \mu_{ij}$$ représentant *l'effet moyen général* commun à tous les traitements.

$$\overline{\mu}_{i.} = \frac{1}{J} \sum_{j=1}^{J} \mu_{ij}$$ représentant *l'effet principal* du niveau i du facteur F_1.

$$\overline{\mu}_{.j} = \frac{1}{I} \sum_{j=1}^{J} \mu_{ij}$$ représentant *l'effet principal* du niveau j du facteur F_2.

Définitions :

- $\alpha_i^1 = \overline{\mu}_{i.} - \overline{\mu}_{..}$ est *l'effet différentiel du niveau* i du facteur F_1 $(i = 1, \ldots, I)$.
- $\beta_j^2 = \overline{\mu}_{.j} - \overline{\mu}_{..}$ est *l'effet différentiel du niveau* j du facteur F_2 $(j = 1, \ldots, J)$.
- $\gamma_{ij}^{12} = \mu_{ij} - \overline{\mu}_{i.} - \overline{\mu}_{.j} + \overline{\mu}_{..}$ est *l'interaction* entre le niveau i du facteur F_1 et le niveau j du facteur F_2.

Remarque : On vérifie que

$$\sum_{i=1}^{I} \alpha_i^1 = \sum_{j=1}^{J} \beta_j^2 = \sum_{i=1}^{I} \gamma_{ij}^{12} = \sum_{j=1}^{J} \gamma_{ij}^{12} = 0$$

On déduit de ces définitions une reparamétrisation équivalente du modèle (3) :

$$(4) \qquad \begin{array}{l} y_{ij_k} = \overline{\mu}_{..} + \alpha_i^1 + \beta_j^2 + \gamma_{ij}^{12} + e_{ij_k} \\ \text{avec} \quad \sum_{i=1}^{I} \alpha_i^1 = \sum_{j=1}^{J} \beta_j^2 = \sum_{i=1}^{I} \gamma_{ij}^{12} = \sum_{j=1}^{J} \gamma_{ij}^{12} = 0 \end{array} \left\{ \begin{array}{l} i = 1, \ldots, I \\ j = 1, \ldots, J \\ k = 1, \ldots, n_{ij} \end{array} \right.$$

e_{ij_k} réalisation d'une v.a. $E_{ij_k} \sim \mathcal{N}(0, \sigma^2)$ et les v.a. E_{ij_k} indépendantes.

L'écriture matricielle est dans ce cas :

$$y = \underset{(n \times IJ)(IJ \times 1)}{X\beta} + e$$

où β est le vecteur des IJ paramètres inconus :

$$\beta' = (\overline{\mu}_{..}, \alpha_1^1, \ldots, \alpha_{I-1}^1, \beta_1^2, \ldots, \beta_{J-1}^2, \gamma_{11}^{12}, \ldots, \gamma_{(I-1)(J-1)}^{12})$$

et X est une matrice dont les éléments sont 0,-1 ou $+1$, qui s'écrit dans l'exemple : $I = 2, J = 3$ et $n_{ij} = K = 2$:

$$\begin{array}{ccccccc} y & = & X & \times & \beta & + & e \\ \begin{pmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{131} \\ y_{132} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \\ y_{231} \\ y_{232} \end{pmatrix} & = & \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 0 & -1 & 0 \\ 1 & -1 & 1 & 0 & -1 & 0 \\ 1 & -1 & 0 & 1 & 0 & -1 \\ 1 & -1 & 0 & 1 & 0 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 & 1 & 1 \end{pmatrix} & \cdot & \begin{pmatrix} \overline{\mu}_{..} \\ \alpha_1^1 \\ \beta_1^2 \\ \beta_2^2 \\ \gamma_{11}^{12} \\ \gamma_{12}^{12} \end{pmatrix} & + & \begin{pmatrix} e_{111} \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ e_{232} \end{pmatrix} \end{array}$$

Interprétation et signification des interactions

Pour expliquer la signification des interactions, nous allons considérer *deux exemples fictifs* donnant le tableau des valeurs des paramètres μ_{ij}.

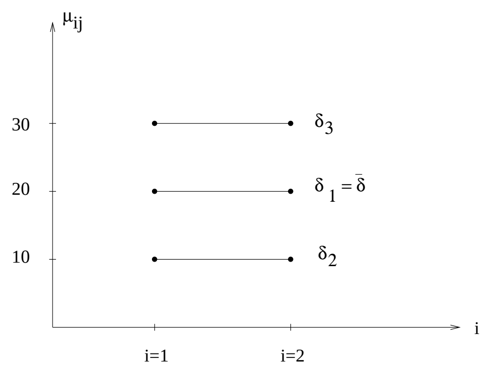
1er exemple : Facteur F_1 à 2 niveaux et F_2 à 3 niveaux.

$F_1 \backslash F_2$	$j = 1$	$j = 2$	$j = 3$
$i = 1$	$\mu_{11} = 20$	$\mu_{12} = 10$	$\mu_{13} = 30$
$i = 2$	$\mu_{21} = 20$	$\mu_{22} = 10$	$\mu_{23} = 30$

On calcule :
$\overline{\mu}_{1.} = \overline{\mu}_{2.} = \overline{\mu}_{..} = 20$
$\overline{\mu}_{.1} = 20; \overline{\mu}_{.2} = 10; \overline{\mu}_{.3} = 30$

Ce qui entraine :
$\alpha_1^1 = \alpha_2^1 = 0 \quad \Longleftrightarrow$ le facteur F_1 n'a pas d'effet différentiel.
$\gamma_{ij}^{12} = 0 \quad \forall(i,j) \quad \Longleftrightarrow$ les facteurs F_1 et F_2 sont sans interaction.

Si on représente dans un repère, les points de coordonnées (i, μ_{ij}), on obtient :



- $\star$ δ_j ligne polygonale des points (i, μ_{ij}).
- $\star$ $\overline{\delta}$ ligne polygonale des points $(i, \mu_{i.})$.

Remarque :

- Le parallélisme entre $\overline{\delta}$ et l'axe des abscisses des niveaux du facteur F_1, caractérise *l'absence d'effet moyen* du facteur F_2.
- Le parallélisme des lignes polygonales entre elles caractérise *l'absence d'interaction* entre les facteurs F_1 et F_2.

2ème exemple : Facteur F_1 à 3 niveaux et F_2 à 4 niveaux.

F_2 \ F_1	$j = 1$	$j = 2$	$j = 3$	$j = 4$
$i = 1$	16	6	7	11
$i = 2$	8	10	17	5
$i = 3$	6	14	6	14

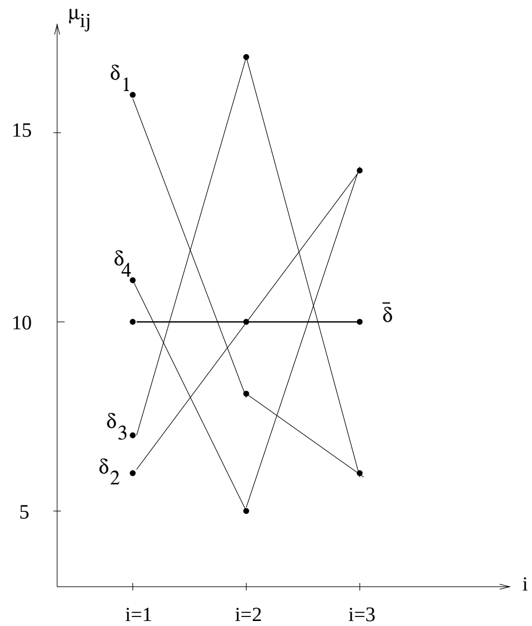
Tableau des paramètres μ_{ij}

On calcule :
$\overline{\mu}_{1\cdot} = \overline{\mu}_{2\cdot} = \overline{\mu}_{3\cdot} = \overline{\mu}_{\cdot\cdot} = 10$
$\overline{\mu}_{\cdot 1} = \overline{\mu}_{\cdot 2} = \overline{\mu}_{\cdot 3} = \overline{\mu}_{\cdot 4} = \overline{\mu}_{\cdot\cdot} = 10$

Ce qui entraine :
$\alpha_i^1 = 0 \quad \forall i = 1, 2, 3$ $\Longleftrightarrow$ *Pas d'effet moyen de* F_1.
$\beta_j^2 = 0 \quad \forall j = 1, 2, 3, 4$ $\Longleftrightarrow$ *Pas d'effet moyen de* F_2.

De plus les paramètres μ_{ij} étant différents, on a obligatoirement *existence d'interaction* entre F_1 et F_2.

On calcule $\gamma_{11}^{12} = 6, \gamma_{12}^{12} = -4; \gamma_{13}^{12} = -3; \gamma_{14}^{12} = +1; \gamma_{21}^{12} = -2; \ldots$.



- ⋆ Parallélisme entre $\overline{\delta}$ et l'axe des abscisses ⇒ pas d'effet moyen de F_1.
- ⋆ Les lignes polygonales δ_i ne sont pas parallèlles et recoupent $\overline{\delta}$ souvent ⇒ les interactions sont fortes.

4.3 Analyse de covariance

La variable endogène est expliquée par une variable quantitative z et par les I indicatrices correspondant à une variable qualitative F ayant I modalités ou niveaux.

4.3.1 Modélisation avec interaction

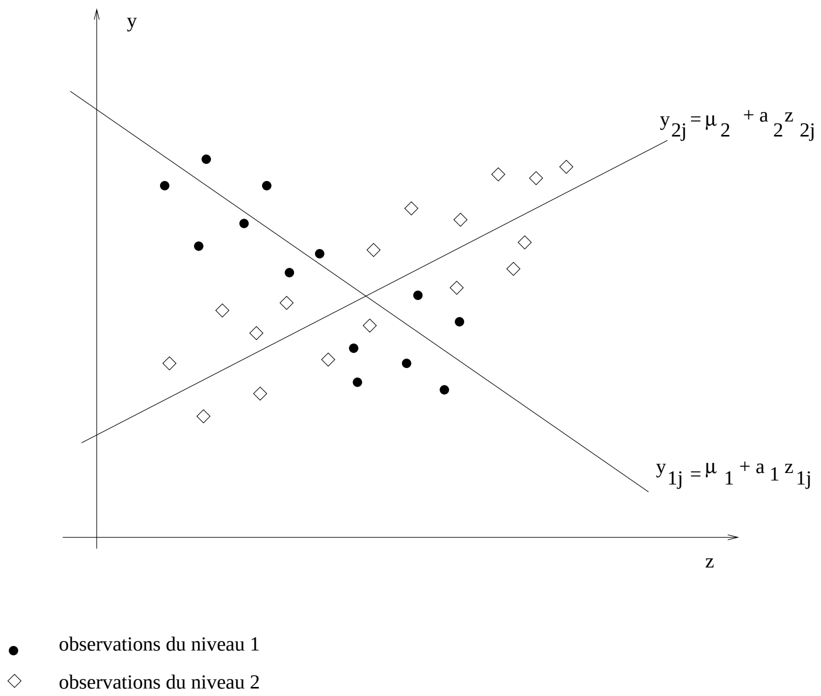
$$y_{ij} = \mu_i + a_i z_{ij} + e_{ij} \qquad \text{où} \qquad e_{ij} \text{ i.i.d. } \mathcal{N}(0, \sigma^2)$$

$$\text{pour} \quad i = 1, \dots, I \qquad \text{et} \qquad j = 1, \dots, n_i \text{ et } n = \sum_{i=1}^{I} n_i \text{ tel que } n > n_i > 1$$

Le coefficient μ_i représente l'effet du niveau i du facteur F.

Le coefficient a_i est le coefficient de régression de z correspondant aux observations du niveau i du facteur F. Il représente l'interaction entre la variable quantitative z et le facteur F.

Exemple : $I = 2, n_1 = 13, n_2 = 17$.



Le modèle linéaire s'écrit matriciellement :

$$\underset{(n\times 1)}{y} = \underset{(n\times 2I)(2I\times 1)}{X\beta} + e$$

$$\text{où } \beta = \begin{bmatrix} \mu_1 \\ a_1 \\ \mu_2 \\ a_2 \\ \vdots \\ \mu_I \\ a_I \end{bmatrix}, \; X = \begin{array}{c} \left[\begin{array}{ccccccccc} 1 & z_{11} & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ 1 & z_{1n_1} & 0 & 0 & \vdots & \cdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & z_{21} & \vdots & \cdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & 1 & z_{2n_2} & 0 & \cdots & 0 & \vdots & \vdots \\ \vdots & \vdots & 0 & 0 & \cdots & \cdots & \cdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \cdots & \cdots & \cdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \cdots & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & 0 & \cdots & 0 & 1 & z_{I1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1 & z_{In_I} \end{array}\right] \end{array} \begin{array}{l} \left.\begin{array}{c} \\ \\ \\ \end{array}\right\} n_1 \text{ lignes} \\ \left.\begin{array}{c} \\ \\ \\ \end{array}\right\} n_2 \text{ lignes} \\ \quad \vdots \\ \\ \left.\begin{array}{c} \\ \\ \\ \end{array}\right\} n_I \text{ lignes} \end{array}$$

Les $(2i-1)$ème et $(2i)$ème vecteurs colonnes de X étant orthogonaux à tous les autres vecteurs colonnes, on peut déterminer les estimateurs de μ_i et a_i pour tout $i = 1, \ldots, I$, en considérant I modèles linéaires successifs :

$$y_i = X_i \cdot \begin{pmatrix} \mu_i \\ a_i \end{pmatrix} + e_i \qquad \text{avec} \qquad \underset{(n_i\times 1)}{y_{(i)}} = \begin{bmatrix} y_{i1} \\ \vdots \\ y_{in_i} \end{bmatrix} \underset{(n_i\times 2)}{X} = \begin{bmatrix} 1 & z_{i1} \\ \vdots & z_{i2} \\ \vdots & \vdots \\ 1 & z_{in_i} \end{bmatrix}$$

Estimateurs (cf. RLS). Pour tout $i = 1, \ldots, I$:

$$\left| \begin{array}{lll} \hat{\mu}_i = \overline{y}_i - \hat{a}_i \overline{z}_i & \text{où} & \overline{y}_i = \frac{1}{n_i} \displaystyle\sum_{j=1}^{n_i} y_{ij} \\ \hat{a}_i = \dfrac{\displaystyle\sum_{j=1}^{n_i} (y_{ij} - \overline{y}_i)(z_{ij} - \overline{z}_i)}{\displaystyle\sum_{j=1}^{n_i} (z_{ij} - \overline{z}_i)^2} & & \overline{z}_i = \frac{1}{n_i} \displaystyle\sum_{j=1}^{n_i} z_{ij} \end{array} \right.$$

La variance des résidus est alors estimée par :

$$\hat{\sigma}^2 = \frac{\| y - X\hat{\beta} \|^2}{n-2I} = \frac{\displaystyle\sum_{i=1}^{I}\sum_{j=1}^{n_i}[y_{ij} - \overline{y}_i - \hat{a}_i(z_{ij} - \overline{z}_i)]^2}{n-2I} \qquad \text{si } n > 2I$$

4.3.2 Tests d'hypothèses

On peut s'intéresser aux hypothèses suivantes :

H_0^1 : $a_i = a$ pour tout $i = 1, I$.

H_0^2 : $a_i = 0$ pour tout $i = 1, I$.

H_0^3 : $\mu_i = \mu$ pour tout $i = 1, I$.

Test de l'hypothèse H_0^1, relatif au parallélisme des droites de régression.

La règle de décision, de $H_0^1 : a_i = a$ pour tout $i = 1, I$ contre $\overline{H}_0^1 : \exists i \neq j, a_i \neq a_j$, est au niveau α :

$$\text{on rejette } H_0^1 \text{ si } \frac{\displaystyle\sum_{i=1}^{I}\sum_{j=1}^{n_i}(\hat{a}_i^2 - \hat{a}^2)\cdot(z_{ij} - \overline{z}_i)^2}{\hat{\sigma}^2 \cdot (I-1)} > f_{I-1;n-2I;\alpha}$$

On montrera, *à titre d'exercice*, ce résultat (cf. test de Fisher Chapitre 3) en précisant l'estimateur $\hat{a}$ qui est l'estimateur de a sous l'hypothèse H_0^1.

Test de l'hypothèse H_0^2 contre l'hypothèse H_0^1, pour μ_i quelconque.

La règle de décision, au niveau α est :

$$\text{on rejette } H_0^2 \text{ si } \frac{\left[\displaystyle\sum_i\sum_j(y_{ij} - \overline{y}_i)^2 - \sum_i\sum_j(y_{ij} - \overline{y}_i - \hat{a}(z_{ij} - z_i))^2\right]}{\displaystyle\sum_i\sum_j(y_{ij} - \overline{y}_i - \hat{a}(z_{ij} - \overline{z}_i))^2/(n-I-1)} > f_{1;n-I-1;\alpha}$$

A titre d'exercice, on montrera que la statistique de ce test s'écrit de façon équivalente en remplaçant le numérateur par : $\hat{a}^2\displaystyle\sum_i\sum_j(z_{ij} - \overline{z}_i)^2$.

Exercice : Montrer que l'estimation de a est différente dans les 2 modèles suivants :

$$1) \qquad y_{ij} = az_{ij} + \mu_i + e_{ij}$$

$$2) \qquad y_{ij} = az_{ij} + e_{ij}$$

Chapitre 5

Généralisation du modèle linéaire gaussien

Le modèle linéaire gaussien, s'il n'est pas adapté au problème considéré, peut être étendu dans les cas des différentes hypothèses :

– *Hypothèse de linéarité*

$$y = X\beta + e \qquad \rightarrow \qquad y = f(x, \beta) + e.$$

MODÈLE DE RÉGRESSION
NON LINÉAIRE

– *Hypothèse sur le vecteur x des variables exogènes.*

X de rang p	$\rightarrow$	X de rang inférieur à p (MULTICOLINÉARITÉ) MODÈLE CONTRAINT : $R\beta = 0$
x vecteur *non* aléatoire	$\rightarrow$	x vecteur aléatoire de $I\!R^p$ – Loi du $p+1$-uple (y, x). – Loi de y sachant x. MODÈLE CONDITIONNEL – Variable x, corrélée avec les résidus. MODÈLES À ERREUR sur les variables exogènes (variables instrumentales).

– *Hypothèse sur les résidus*

Var $(e_i) = \sigma^2$ (Résidus homogènes)	$\longrightarrow$	var$(e_i) = \sigma_i^2$ (Résidus hétérogènes) MOINDRES CARRÉS PONDÉRÉS
$E(e_ie_j) = 0$ si $i \neq j$ (Résidus non corrélés)	$\longrightarrow$	$E(e_ie_j) \neq 0$ si $i \neq j$ (Résidus autocorrélés) MOINDRES CARRÉS GÉNÉRALISÉS
e_i v.a. gaussienne (Normalité des résidus)	$\longrightarrow$	Loi de probabilité de y_i paramétrique : – Binomiale – Poisson – Gamma $\vdots$ VRAISEMBLANCE $\mathcal{L}(y_1, \ldots, y_n; \beta)$, (Si x aléatoire, vraisemblance conditionnelle). $\rightarrow$ Loi de probabilité de (y, x) sans spécification de lois particulières paramétriques. (STATISTIQUE NON PARAMÉTRIQUE).

5.1 Moindres carrés généralisés

5.1.1 Estimateur des moindres carrés généralisés

On le note MCG ou GLS (Generalized Least Squares).

On se place dans le cas où l'hypothèse H_3 des perturbations sphériques : $E(ee') = \sigma^2 I_n$ du modèle linéaire, n'est plus vraie.

Définition du modèle linéaire (Lorsque les perturbations ne sont pas sphériques).

H_1 $\quad \underset{(n\times 1)}{y} = \underset{(n\times p)(p\times 1)}{X\beta} + \underset{(n\times 1)}{e}$

H_2 $\quad X$ matrice non aléatoire de rang p et la première colonne de X est $1\!\!1_n$.

H_3'' $\quad E(e) = 0 \quad$ et $\quad E(ee') = \sigma^2\Omega \quad$ où Ω matrice d'ordre n, symétrique définie positive, de rang n.

Définition

L'estimateur MCG (ou estimateur d'Aitken) est :

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

Propriété de Gauss-Markov : Sous les hypothèses H_1, H_2 et H_3'', l'estimateur $\hat{\beta}_{MCG}$ est *sans biais*, de variance : $\sigma^2(X'\Omega^{-1}X)^{-1}$ et *meilleur que tout estimateur linéaire et sans biais*.

Démonstration : On vérifie :

$$E(\hat{\beta}_{MCG}) = \beta \text{ et Var } (\hat{\beta}_{MCG}) = \sigma^2(X'\Omega^{-1}X)^{-1}$$

On se ramène au cas du théorème de Gauss-Markov énoncé pour le modèle linéaire classique.

On utilise la propriété (cf. Chapitre I) :
Ω matrice symétrique définie positive (d'ordre n).

$\Rightarrow$ Il existe une matrice P d'ordre n, *inversible* telle que :

$$\Omega = P\,P'$$

On pose :

$$z = P^{-1}y$$

- L'hypothèse $H_1 : y = X\beta + e$ s'écrit :

$$P^{-1}y = P^{-1}X\beta + P^{-1}e$$

  C'est-à-dire $\quad z = A\beta + u \quad$ avec $\begin{array}{l} A = P^{-1}X \\ u = P^{-1}e \end{array}$.

- L'hypothèse H_3'' sur les résidus e, s'écrit pour les *résidus* u :

$$\begin{array}{ll} \star & E(u) = 0 \\ \star & E(uu') = \sigma^2 I_n \end{array}$$

- Dans le modèle linéaire : $z = A\beta + u$, l'estimateur des moindres carrés s'écrit :

$$(A'A)^{-1}A'z = (X' \underbrace{P'^{-1}P^{-1}}_{\Omega^{-1}} X)^{-1}X' \underbrace{P'^{-1}P^{-1}}_{\Omega^{-1}} y = \hat{\beta}_{MCG}$$

  $\Rightarrow$ Théorème de Gauss-Markov (Chapitre 2).

Propriétés :

1. Sous l'hypothèse H_3'', l'estimateur MCO $\hat{\beta}_{MCO} = (X'X)^{-1}X'y$, qui est linéaire et toujours sans biais, n'est plus de variance minimale.
   Exercice : Calculer Var $(\hat{\beta}_{MCO})$.

2. L'estimateur MCG $\quad \hat{\beta}_{MCG} = (X'\Omega'X)^{-1}X'\Omega^{-1}y$ est l'estimateur qui :
   – MINIMISE $\| y - X\beta \|^2_{\Omega^{-1}}$
   – MAXIMISE LA VRAISEMBLANCE DU MODÈLE LINÉAIRE GAUSSIEN $(y, X\beta, \sigma^2\Omega)$ défini par H_1, H_2, H_3'' et l'hypothèse de normalité des résidus e.
   $$\mathcal{L}(y,\beta) = \sigma^{-n} \cdot (2\pi)^{-n/2} \cdot (\det \Omega)^{-1/2} \exp\{-\frac{1}{2\sigma^2} \| y - X\beta \|^2_{\Omega^{-1}}\}$$
   Exercice : Montrer que $\hat{\beta}_{MCG} = \arg\min_\beta \| y - X\beta \|^2_{\Omega^{-1}} = \hat{\beta}_{MV}$

Remarques :

1. Si la matrice Ω est connue :
   – $\hat{\beta}_{MCG} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$ est un estimateur sans biais de β.
   – $\hat{\sigma}^2 = \dfrac{\| y - X\hat{\beta}_{MCG} \|^2_{\Omega^{-1}}}{n-p}$ est un estimateur sans biais de σ^2. (à montrer à titre d'exercice).
2. Si la matrice Ω est inconnue :
   – Il y a $\dfrac{n(n-1)}{2}$ paramètres Ω_{ij} de la matrice Ω à estimer ainsi que n paramètres Ω_{ii}.
   $\Rightarrow$ Problème insoluble dans le cas général.
   – Estimation de Ω possible dans des cas simples particuliers :
     * Ω diagonale
     * Ω a une expression particulière, par exemple :

$$\Omega = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & & & \\ \vdots & & 1 & & \\ \vdots & & & 1 & \\ \rho^{n-1} & & & & 1 \end{pmatrix}$$

Processus autorégressif du premier ordre : $e_t = \rho e_{t-1} + u_t$

Test dans le modèle $(y, X\beta, \sigma^2, \Omega)$ *gaussien où* Ω *connue* : Test de Fisher. (cf. Chapitre 3.)

– Hypothèse nulle :

$$H_0 : R\beta = r$$

$$\text{contre} \quad R\beta \neq r \qquad \text{avec } R(q \times p) \text{ de rang } q \text{ et } r(q \times 1)$$

– Règle de décision au niveau α.

On rejette H_0 si $F > F_{q,n-p,\alpha}$ où la statistique de Fisher s'écrit de façon analogue au chapitre 3.

$$F = \frac{\frac{1}{q} \parallel r - R\hat{\beta}_{MCG} \parallel^2_{[R(X'\Omega^{-1}X)^{-1}R']^{-1}}}{\frac{1}{n-p} \parallel y - X\hat{\beta}_{MCG} \parallel^2_{\Omega^{-1}}}$$

Exercice : Par analogie aux résultats du chapitre 3, écrire :

– L'expression de l'estimateur contraint par $R\beta = r$.
– Le test de student de signification d'un coefficient β_j.
– Le test de Fisher de signification de q coefficients.
– Le numérateur de F sous la forme $\frac{1}{q}(RSS_C - RSS)$.

5.1.2 Estimateur des moindres carrés pondérés

On le note MCP ou WLS (Weighted Least Squares).

Hétéroscédasticité et non corrélation des résidus

Exemple 1 : Cas de données groupées.

$$y_{ij} = \beta_1 + \beta_2 x^1_{ij} + e_{ij} \text{ avec } \begin{cases} i = 1, \ldots, I; j = 1, \ldots, n_i \\ e_{ij} \text{ indépendantes} \\ E(e_{ij}) = 0 \text{ Var } (e_{ij}) = \sigma^2 \end{cases}$$

$$\Longleftrightarrow \qquad \overline{y}_i = \beta_1 + \beta_2 \overline{x}^1_i + u_i \text{ avec } \begin{cases} u_i \text{ indépendants} \\ E(u_i) = 0 \quad \text{Var } (u_i) = \dfrac{\sigma^2}{n_i} \end{cases}$$

$$\text{Var } u = \sigma^2\Omega = \sigma^2 \begin{pmatrix} \frac{1}{n_1} & & 0 \\ 0 & \vdots & \frac{1}{n_I} \end{pmatrix}$$

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{I} n_i(\overline{y}_i - \beta_1 - \beta_2 \overline{x}^1_i)^2$$

Exemple 2 : Hétéroscédasticité due à une (ou plusieurs) variable exogène x^j.

$$y = X\beta + e \text{ avec } \begin{cases} E(e) = 0 \\ \\ \text{Var } (e) = \sigma^2\Omega \end{cases} \quad \text{où} \quad \Omega = \text{ diag } ((x_1^j)^2, \ldots, (x_n^j)^2)$$

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} (y_i - \sum_{j=1}^{p} x_i^j \beta_j)^2 w_i$$

$$\text{où la pondération } w_i = \frac{1}{(x_i^j)^2},$$

Estimateur MCP

C'est l'estimateur MCG dans le modèle particulier $(y, X\beta, \sigma^2\Omega)$ *où* Ω *est une matrice diagonale* : $\Omega =$ diag $(\frac{1}{w_1}, \ldots, \frac{1}{w_n})$ définie par :

$$\begin{aligned} \hat{\beta}_{MCP} &= \arg\min_{\beta} \sum_{i=1}^{n} w_i (y_i - \sum_{j=1}^{p} x_i^j \beta_j)^2 \\ &= \arg\min_{\beta} \parallel y - X\beta \parallel^2_{\Omega^{-1}} \end{aligned}$$

et égal à :

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

5.1.3 Autocorrélation des résidus

On considère le modèle linéaire dans le cas de *résidus corrélés*, c'est-à-dire : $y = X\beta + e$ avec $E(e) = 0$ et $E(ee') = \Sigma$ où Σ est une matrice symétrique, définie positive, non diagonale, inversible.

On traitera le cas de corrélation temporelle, très classique dans les séries chronologiques :

Autocorrélation du 1er ordre dans les processus autorégressifs :

Définition :

Les résidus e_t suivent un *processus autorégressif d'ordre 1* (AR(1)) si :

$$e_t = \rho e_{t-1} + u_t \qquad \text{(pour } t = 2, \dots, n) \text{ avec } \left| \begin{array}{l} u_t \text{ indépendant de } e_{t-1} \\ E(u_t) = 0 \\ \text{Var}\,(u_t) = \sigma^2 \\ E(u_t u'_t) = 0 \text{ si } t \neq t' \end{array} \right.$$

c'est-à-dire le résidu e_t dépend du résidu e_{t-1} :

$$E(e_t / e_{t-1}) = \rho e_{t-1}$$

Le paramètre ρ est tel que : $\mid \rho \mid < 1$.

Propriétés : Si les résidus e_t du modèle linéaire $(y, X\beta, \Sigma)$ suivent un processus $AR(1)$, alors :

1. La matrice Σ s'écrit :

$$\sum_{(n\times n)} = \frac{\sigma^2}{1-\rho^2} \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{pmatrix} = \sigma^2 \Omega$$

$$\Omega^{-1} = \begin{pmatrix} 1 & \rho & 0 & \cdots & 0 \\ -\rho & 1+\rho^2 & -\rho & \cdots & \vdots \\ 0 & -\rho & 1+\rho^2 & \cdots & \vdots \\ \vdots & \vdots & -\rho & \ddots & 0 \\ \vdots & \vdots & \cdots & 1+\rho^2 & -\rho \\ 0 & \cdots & 0 & -\rho & 1 \end{pmatrix} \text{ et } \det \Omega = \frac{1}{1-\rho^2}$$

2. Les paramètres β et ρ sont estimés :
   – Soit par une méthode en 2 étapes :

Etape 1 : ⋆ Calcul de $\hat{\beta}_{MCO} = (X'X)^{-1}X'y$.

⋆ Calcul de $\hat{e} = y - X\hat{\beta}_{MCO}$.

⋆ Calcul de $\hat{\rho}_{MCO}$ dans le modèle linéaire :

$$\hat{e}_t = \rho \hat{e}_{t-1} + u_t \qquad t = 2, \dots, n$$

$$\Rightarrow \qquad \hat{\rho}_{MCO} = \frac{\sum_{t=2}^{n} \hat{e}_t \hat{e}_{t-1}}{\sum_{t=2}^{n} (\hat{e}_{t-1}^2)}$$

Etape 2 : ⋆ Calcul de $\hat{\Omega}$, estimation de Ω obtenue en remplaçant ρ par $\hat{\rho}_{MCO}$
⋆ Calcul de l'estimateur MCG :

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

– Soit par MV (difficile)

3. Un test de l'autocorrélation pour l'hypothèse nulle H_0 : $\rho = 0$ (c'est-à-dire $\sigma^2\Omega = \sigma^2 I_n$) contre $\rho \neq 0$ (ou $\rho > 0$), peut être réalisé à partir de la statistique de DURBIN-WATSON :

$$DW = \frac{\sum_{t=2}^{n} (\hat{e}_t - \hat{e}_{t-1})^2}{\sum_{t=1}^{n} (\hat{e}_t^2)} \text{ où } \hat{e} = y - X(X'X)^{-1}X'y$$

Des tables de Durbin-Watson permettent de déterminer la procédure du test suivant.

Si $DW < d_{inf}$, on rejette H_0.

Si $DW > d_{sup}$, on ne rejette pas H_0.

Si $d_{inf} < DW < d_{sup}$, on ne peut pas conclure.

Conditions d'utilisation du test :
- $x_i^1 = 1$
- X non aléatoire
- les variables exogènes ne contiennent pas la variable endogène retardée.

5.2 Régression stochastique

On a supposé jusqu'à présent que les variables exogènes $x_i = (x_i^1, \dots, x_i^p)$ étaient déterministes et que les erreurs e_i étaient aléatoires (ce qui entraîne que la variable endogène y_i est aléatoire).

En fait, dans beaucoup d'exemples considérés, les variables exogènes sont aléatoires.

On suppose donc que $x_i = (x_i^1, \dots, x_i^p)$ est un vecteur aléatoire de $I\!R^p$, et donc X est une matrice aléatoire.

Remarque : Dans ce cas on ne considère plus l'hypothèse $x_i^1 = 1$, et si le modèle comprend un terme constant (où des variables exogènes déterministes) il conviendra de le faire intervenir additionnellement à X.

On se bornera à donner quelques idées générales suivant les 2 cas :
- On considère la loi conditionnelle de y_i sachant x_i.
- On considère la loi du *couple* (y_i, x_i), dans le cas $y_i \in I\!R$ et $x_i \in I\!R^p$.

5.2.1 Modèles linéaires conditionnels

Définitions : Un modèle linéaire conditionnel à $X = \tilde{X}$ (X matrice $(n \times p)$ aléatoire) est défini par les hypothèses :

A_1 $y = X\beta + e$.

A_2 La matrice X est aléatoire et sa réalisation $\tilde{X}$ est de rang p (la loi de probabilité de x ne dépend pas de β.)

A_3 $E(e/X = \tilde{X}) = 0; \text{Var}\,(e/X = \tilde{X}) = \sigma^2\Omega$.
(Ω matrice de rang n, symétrique et définie positive, *connue*).

Remarques :

1. L'hypothèse A_3 est équivalente à
$$E(y/X = \tilde{X}) = \tilde{X}\beta; \text{Var}\,(y/X = \tilde{X}) = \sigma^2\Omega$$
2. Si, de plus, on suppose la normalité pour la loi conditionnelle de y sachant X (cas du modèle linéaire gaussien conditionnel), on a :
$$\text{la loi de } y \text{ sachant } X = \tilde{X} \text{ est } \mathcal{N}_n(\tilde{X}\beta; \sigma^2\Omega)$$
3. Si le modèle contient un terme constant, on écrira :
$$y = \beta_0 + X\beta + e$$
avec les mêmes hypothèses A_2 et A_3.
4. Tous les calculs et propriétés démontrés dans les chapitres précédents sont valables, pour des raisonnements évidemment *conditionnellement à* X.

5.2.2 Approximation d'une v.a. y par un vecteur aléatoire x

Dans cette section, on considère le vecteur aléatoire de $I\!R^{p+1}$:

$$z = (y, x) \text{ où } y \text{ v.a.r. et } x = (x^1, \ldots, x^p) \in I\!R^p$$

Régression de y en x

Problème : Peut-on trouver une fonction f de $\mathbb{R}^p$ dans $\mathbb{R}$ telle que :

$$y = \underbrace{f(x)}_{\substack{\text{expliquée} \\ \text{par } x}} + \underbrace{e}_{\substack{\text{reste} \\ \text{indépendant} \\ \text{de } x}}$$

en considérant comme critère *à minimiser* “l’erreur moyenne quadratique” :

$$E[(y - f(x))^2]$$

où l’espérance E est relative à la loi de $z = (y, x)$

Propriété 1 : $E(y)$ est le nombre le plus proche de y en moyenne quadratique, c’est-à-dire :

$$E(y) = \arg\min_k E[(y - k)^2]$$

Démonstration : On décompose :

$$E[(y - k)^2] = V(y) + (E(y) - k)^2$$

$$\text{Car} \quad E[(y - E(y))(E(y) - k)] = 0$$

Propriété 2 : La fonction f qui minimise $E[(y - f(x))^2]$ est celle qui à $\tilde{x}$ associe l’espérance conditionnelle de y sachant $x = \tilde{x}$, notée $E(y/x = \tilde{x})$ ou $\mathcal{R}(\tilde{x})$ et appelée *régression de* y *en* x.

Démonstration : On rappelle :

$$E(U) = E[E(U/x = \tilde{x})] \text{ si } U \text{ est une fonction de } x \text{ et de } y.$$

On décompose :

$$E[(y - f(x))^2] = E\left([y - E(y/x = \tilde{x})]^2\right) + E\left([E(y/x = \tilde{x}) - f(\tilde{x})]^2\right)$$

car le double produit est nul.

La fonction f qui minimise $E[(y - f(x))^2]$ est donc

$$f(\tilde{x}) = E(y/x = \tilde{x}) \text{ pour toute valeur } \tilde{x} \text{ de la variable aléatoire } x.$$

Remarque : $\mathcal{R}(\tilde{x})$ s’explicite :

– Dans le cas où (y, x) a une densité g.

$$\mathcal{R}(\tilde{x}) = \int y \times \frac{g(y, \tilde{x}^1, \ldots, \tilde{x}^p)}{\int g(y, \tilde{x}^1, \ldots, \tilde{x}^p) dy} dy$$

– Dans le cas où (y, x) est un vecteur aléatoire discret :

$$\mathcal{R}(\tilde{x}) = \sum_{\tilde{y}} \tilde{y} \cdot \frac{P(y = \tilde{y}, x = \tilde{x})}{P(x = \tilde{x})}$$

Remarque : Si y et x indépendants, alors

$$\mathcal{R}(\tilde{x}) = E(y) \text{ pour tout } \tilde{x}$$

Exemples : Sachant que $x = \tilde{x}$, que prévoir pour y ?

(1) (y, x) de densité $g(y, x) = \dfrac{1}{2\sqrt{xy}}$ si $0 < y < x < 1$

(2) (y, x) discrète de loi :

x \ y	0	1
0	25/174	4/174
1	120/174	25/174

Propriété 3 : Analyse de variance :

$$V(y) = V(y - \mathcal{R}(x)) + V(\mathcal{R}(x))$$

Régression linéaire de y en x.

Problème : Peut-on trouver une fonction affine f de $\mathbb{R}^p$ dans $\mathbb{R}$ (c'est-à-dire $f(x) = \sum_{j=1}^{p} a_j x^j + a_0$) telle que : $y = f(x) + e$ en considérant comme critère le minimum de $E[(y - f(x))^2]$?

$$\text{c'est-à-dire } E\left[(y - \sum_{j=1}^{p} a_j x^j - a_0)^2\right] \text{ minimum}$$

C'est la méthode des moindres carrés probabiliste.

Propriété 4 : La fonction affine $f(x) = \sum_{j=1}^{p} a_j x^j + a_0$ qui minimise $E[(y - f(x))^2]$ est la *fonction notée* $EL(y/x)$ (appelée *Régression linéaire de* y *par* x*)* :

$$EL(y/x = \tilde{x}) = E(y) + \underset{(1\times p)}{\text{cov}\,(y,x)}\underset{(p\times p)}{[V(x)]^{-1}}\underset{(p\times 1)}{(\tilde{x} - E(x))}$$

Démonstration : On pose $\underset{(1\times p)(p\times 1)}{a'x} = \sum_{j=1}^{p} a_j x^j =< a, x >$.

On décompose :

$$E[(y - a'x - a_0)^2] = E[(y^* - a'x^*)^2] + (E(y) - a'E(x) - a_0)^2$$

car le double produit est nul

avec $y^* = y - E(y)$ et $x^* = x - E(x)$

On décompose :

$$\begin{array}{rcl} E[(y^* - a'x^*)^2] & = & V(y) - \text{cov}\,(y,x)(V(x))^{-1}\text{cov}\,(x,y) \\ & + & (a - (V(x))^{-1}\text{cov}\,(x,y))'V(x)(a - (V(x))^{-1}\text{cov}\,(x,y)) \end{array}$$

$\Rightarrow E[(y - a'x - a_0)^2]$ minimum si :

$$\left| \begin{array}{rcl} a & = & (V(x))^{-1}\text{cov}\,(x,y) \\ a_0 & = & E(Y) - a'E(x) \end{array} \right.$$

$\Rightarrow f(x) = a'x + a_0$ qui minimise $E[(y - a'x - a_0)^2]$ est :

$$EL(y \mid x) = (\text{cov}\,(x,y))'(V(x))^{-1}(x - E(x)) + E(y)$$

CQFD

Aspect statistique

Quand on ne connait pas la loi de (y,x) mais quand on dispose d'un échantillon $(y_i,x_i)_{i=1,n}$ de taille n, on peut estimer $EL(y/x)$.

Propriété : Soit (y,x) centrée ($E(y)=0$ et $E(x)=0$)

$$\text{Soit } \underline{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \text{ et } \underset{(n\times p)}{X} = \begin{pmatrix} x_1^1 & \cdots & x_1^p \\ \vdots & \ddots & \vdots \\ x_n^1 & \cdots & x_n^p \end{pmatrix} \text{ les observations de } (y,x)$$

Alors la régréssion linéaire de y par x, $EL(y/x)$ est estimée à l'aide des moments empiriques, par :

$$EL\widehat{(y_i}/x_i) = x_i'(X'X)^{-1}X'\underline{y} \text{ où } x_i = \begin{pmatrix} x_i^1 \\ \vdots \\ x_i^p \end{pmatrix}$$

Remarque : On retrouve les formules du modèle linéaire (Chapitre 2).

Démonstration : On a vu que $a=(V(x))^{-1}\text{cov}\,(x,y)$ comme y et x sont centrées :

$$a = \left(E(xx')\right)^{-1} E(xy)$$

- $\underset{(p\times p)}{E(xx')}$ est estimée par la matrice de variance covariance empirique $\dfrac{1}{n}X'X$ de terme général : $\frac{1}{n}\sum_{t=1}^{n} x_t^j x_t^{j'}$.

$$\text{En effet, } E(\frac{1}{n}\sum_{t=1}^{n} x_t^j x_t^{j'}) = E(x^j x^{j'}) \text{ terme général de } E(xx').$$

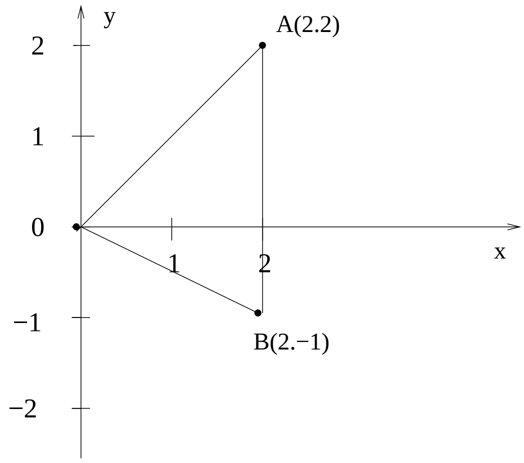
- $\underset{(p\times 1)}{E(xy)}$ est estimée par le vecteur de covariance empirique $\dfrac{1}{n}X'\underline{y}$, de composantes : $\frac{1}{n}\sum_{t=1}^{n} y_t x_t^i$.

$$\text{En effet, } E\left(\frac{1}{n}\sum_{t=1}^{n} y_t x_t^j\right) = E(y_t x_t^j) = E(yx^j)$$

$$j\text{ème composante de } E(xy) \qquad (j=1,\ldots,p)$$

- Donc $\hat{a}=(X'X)^{-1}X'\underline{y}$ et $\hat{y}_i = x_i'\hat{a}$.

Application : Au concours de tir à l'arc de la forêt de Brocéliande, les archers visent une cible triangulaire. Merlin l'Enchanteur, en avance sur son temps, a repéré les points de cette cible dans un système d'axes orthonormés, selon le schéma :



La cible a atteindre est OAB et le résultat d'un tir est un couple (x, y) représentant les coordonnées du point atteint. On suppose que (x, y) suit une loi uniforme sur le triangle OAB, nulle en dehors.

- Calculer la régression de y par x.
- Calculer la régression linéaire de y par x.

Bibliographie

ANTONIADIS A., BERRUYER J., CARMONA R. (1992), “Régression non linéaire et applications”, *Economica*.

BOURBONNAIS R. (1998), “Econométrie : manuel et exercices corrigés”, *Dunod*.

MADDALA G.S. (1977), “Econometrics”, *International Student Edition.*

FOURGEAUD C., LENCLUD B. (1978), “Econométrie”, *PUF*.

GREENE W. H. (1993), “Econometric Analysis”, *Prentice Hall*.

GIRAUD R., CHAIX N. (1989), “Econométrie”, *PUF*.

MONFORT A. (1982), “Cours de statistique mathématique”, *Economica*, ENSAE.

RAFFIN C. [1993), “ Econométrie”, Flash U, *Armand Colin*.

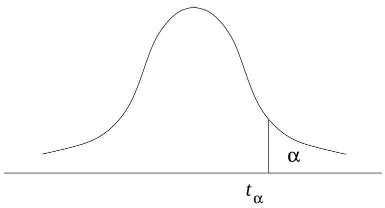
TOMASSONE R., LESQUOY E., MILLIER C. (1983), “La régression : nouveaux regards sur une ancienne méthode statistique”, *Masson*.

JOHNSTON J. (1985), “Méthodes économétriques : tomes 1 et 2”, *Economica*.

SAPORTA G. (1990), “Probabilités, analyse des données et statistique”, *Technip*.

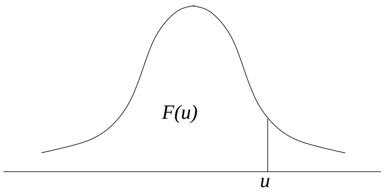
Tables statistiques

Loi de Student



d.d.l. α	0.10	0.05	0.025	0.01	0.005
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
∞	1.282	1.645	1.960	2.326	2.576

Loi normale



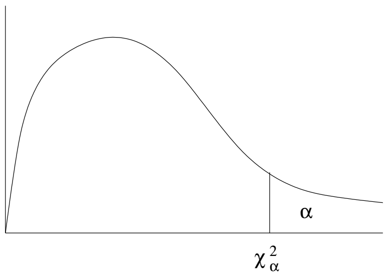
u	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

Table pour les grandes valeurs de u :

u	3.0	3.1	3.2	3.3	3.4	3.5
$F(u)$	0.998650	0.999032	0.999313	0.999517	0.999663	0.999767

u	3.6	3.7	3.8	3.9	4.0	4.5
$F(u)$	0.999841	0.999892	0.999928	0.999952	0.999968	0.999997

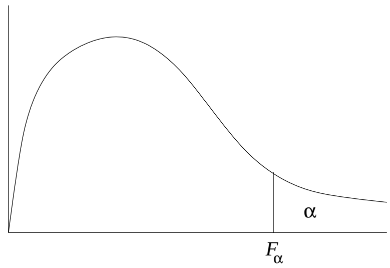
Loi du khi-deux



d.d.l. α	0.95	0.90	0.50	0.30	0.20	0.10	0.05	0.02	0.01	0.001
1	0.004	0.016	0.455	1.074	1.642	2.706	3.841	5.412	6.635	10.828
2	0.103	0.211	1.386	2.408	3.219	4.605	5.991	7.824	9.210	13.816
3	0.352	0.584	2.366	3.665	4.642	6.251	7.815	9.837	11.345	16.266
4	0.711	1.064	3.357	4.878	5.989	7.779	9.488	11.668	13.277	18.467
5	1.145	1.610	4.351	6.064	7.289	9.236	11.07	13.388	15.086	20.515
6	1.635	2.204	5.348	7.231	8.558	10.645	12.592	15.033	16.812	22.458
7	2.167	2.833	6.346	8.383	9.803	12.017	14.067	16.622	18.475	24.322
8	2.733	3.490	7.344	9.524	11.030	13.362	15.507	18.168	20.090	26.124
9	3.325	4.168	8.343	10.656	12.242	14.684	16.919	19.679	21.666	27.877
10	3.940	4.865	9.342	11.781	13.442	15.987	18.307	21.161	23.209	29.588
11	4.575	5.578	10.341	12.899	14.631	17.275	19.675	22.618	24.725	31.264
12	5.226	6.304	11.340	14.011	15.812	18.549	21.026	24.054	26.217	32.909
13	5.892	7.042	12.340	15.119	16.985	19.812	22.362	25.471	27.688	34.528
14	6.571	7.790	13.339	16.222	18.151	21.064	23.685	26.873	29.141	36.123
15	7.261	8.547	14.339	17.322	19.311	22.307	24.996	28.259	30.578	37.697
16	7.962	9.312	15.338	18.418	20.465	23.542	26.296	29.633	32.000	39.252
17	8.672	10.085	16.338	19.511	21.615	24.769	27.587	30.995	33.409	40.790
18	9.390	10.865	17.338	20.601	22.760	25.989	28.869	32.346	34.805	42.312
19	10.117	11.651	18.338	21.689	23.900	27.204	30.144	33.687	36.191	43.820
20	10.851	12.443	19.337	22.775	25.037	28.412	31.410	35.020	37.566	45.315
21	11.591	13.240	20.337	23.858	26.171	29.615	32.671	36.343	38.932	46.797
22	12.338	14.041	21.337	24.939	27.301	30.813	33.924	37.659	40.289	48.268
23	13.091	14.848	22.337	26.018	28.429	32.007	35.172	38.968	41.638	49.728
24	13.848	15.659	23.337	27.096	29.553	33.196	36.415	40.270	42.980	51.179
25	14.611	16.473	24.337	28.172	30.675	34.382	37.652	41.566	44.314	52.620
26	15.379	17.292	25.336	29.246	31.795	35.563	38.885	42.856	45.642	54.052
27	16.151	18.114	26.336	30.319	32.912	36.741	40.113	44.140	46.963	55.476
28	16.928	18.939	27.336	31.391	34.027	37.916	41.337	45.419	48.278	56.892
29	17.708	19.768	28.336	32.461	35.139	39.087	42.557	46.693	49.588	58.301
30	18.493	20.599	29.336	33.530	36.250	40.256	43.773	47.962	50.892	59.703
40	26.509	29.051	39.335	44.165	47.269	51.805	55.758	60.436	63.691	73.402
50	34.764	37.689	49.335	54.723	58.164	63.167	67.505	72.613	76.154	86.661
60	43.188	46.459	59.335	65.227	68.972	74.397	79.082	84.580	88.379	99.607
70	51.739	55.329	69.334	75.689	79.715	85.527	90.531	96.388	100.425	112.317
80	60.391	64.278	79.334	86.120	90.405	96.578	101.879	108.069	112.329	124.839
90	69.126	73.291	89.334	96.524	101.054	107.565	113.145	119.648	124.116	137.208
100	77.929	82.358	99.334	106.906	111.667	118.498	124.342	131.142	135.807	149.449

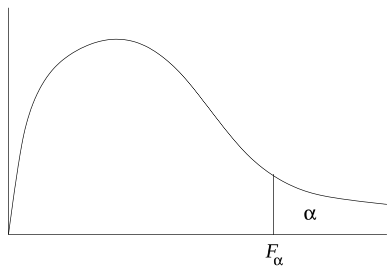
Quand le nombre de degrés de liberté est élevé, $\sqrt{2\chi_2}$ est à peu près distribué normalement autour de $\sqrt{2d.d.l. - 1}$ avec une variance égale à 1.

Loi de Fisher : $\alpha = 0.10$



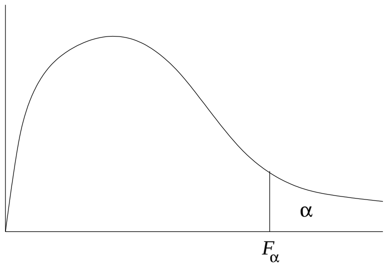
Degrés de liberté du dénominateur : ν_2	Degrés de liberté du numérateur : ν_1								
	1	2	3	4	5	6	7	8	9
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63

Loi de Fisher : $\alpha = 0.10$ (suite)



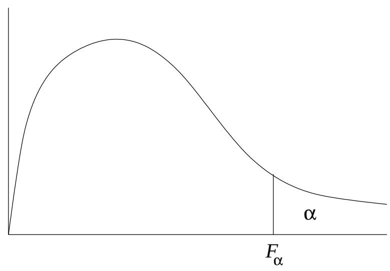
Degrés de liberté du dénominateur : ν_2	Degrés de liberté du numérateur : ν_1 10	12	15	20	24	30	40	60	120	∞
1	60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33
2	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
3	5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13
4	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76
5	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10
6	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
7	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
8	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
9	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
10	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
11	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97
12	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
13	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
14	2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80
15	2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76
16	2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72
17	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69
18	1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66
19	1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63
20	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
21	1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59
22	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57
23	1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55
24	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53
25	1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52
26	1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50
27	1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49
28	1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48
29	1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47
30	1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46
40	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38
60	1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29
120	1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19
∞	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00

Loi de Fisher : $\alpha = 0.05$



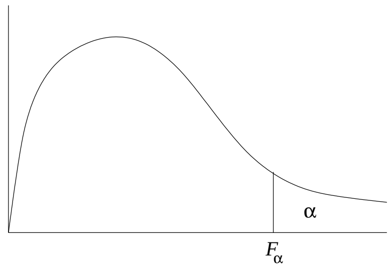
Degrés de liberté du dénominateur : ν_2	Degrés de liberté du numérateur : ν_1								
	1	2	3	4	5	6	7	8	9
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88

Loi de Fisher : $\alpha = 0.05$ (suite)



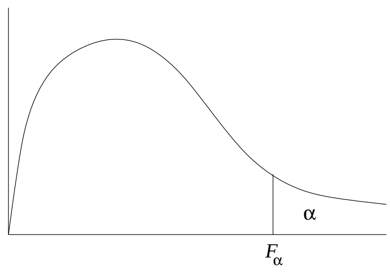
	Degrés de liberté du numérateur : ν_1									
Degrés de liberté du dénominateur : ν_2	10	12	15	20	24	30	40	60	120	∞
1	241.88	243.91	245.95	248.01	249.05	250.10	251.14	252.20	253.25	254.30
2	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

Loi de Fisher : $\alpha = 0.025$



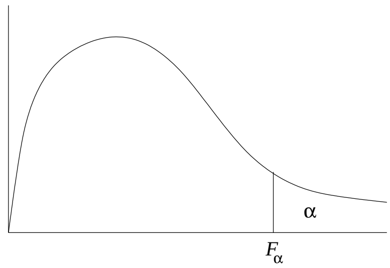
	Degrés de liberté du numérateur : ν_1								
Degrés de liberté du dénominateur : ν_2	1	2	3	4	5	6	7	8	9
1	647.79	799.50	864.16	899.58	921.85	937.11	948.22	956.66	963.28
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11

Loi de Fisher : $\alpha = 0.025$ (suite)



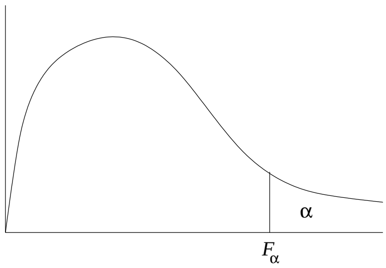
	Degrés de liberté du numérateur : ν_1									
Degrés de liberté du dénominateur : ν_2	10	12	15	20	24	30	40	60	120	∞
1	968.63	976.71	984.87	993.10	997.25	1001.41	1005.60	1009.80	1014.02	1018
2	39.40	39.41	39.43	39.45	39.46	39.46	39.47	39.48	39.49	39.50
3	14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	13.90
4	8.84	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26
5	6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02
6	5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85
7	4.76	4.67	4.57	4.47	4.41	4.36	4.31	4.25	4.20	4.14
8	4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67
9	3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33
10	3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08
11	3.53	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88
12	3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72
13	3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60
14	3.15	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49
15	3.06	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.46	2.40
16	2.99	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	2.32
17	2.92	2.82	2.72	2.62	2.56	2.50	2.44	2.38	2.32	2.25
18	2.87	2.77	2.67	2.56	2.50	2.44	2.38	2.32	2.26	2.19
19	2.82	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	2.13
20	2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09
21	2.73	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	2.04
22	2.70	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	2.00
23	2.67	2.57	2.47	2.36	2.30	2.24	2.18	2.11	2.04	1.97
24	2.64	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94
25	2.61	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	1.91
26	2.59	2.49	2.39	2.28	2.22	2.16	2.09	2.03	1.95	1.88
27	2.57	2.47	2.36	2.25	2.19	2.13	2.07	2.00	1.93	1.85
28	2.55	2.45	2.34	2.23	2.17	2.11	2.05	1.98	1.91	1.83
29	2.53	2.43	2.32	2.21	2.15	2.09	2.03	1.96	1.89	1.81
30	2.51	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79
40	2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64
60	2.27	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48
120	2.16	2.05	1.94	1.82	1.76	1.69	1.61	1.53	1.43	1.31
∞	2.05	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.00

Loi de Fisher : $\alpha = 0.01$



Degrés de liberté du dénominateur : ν_2	Degrés de liberté du numérateur : ν_1								
	1	2	3	4	5	6	7	8	9
1	4052.18	4999.50	5403.35	5624.58	5763.65	5858.99	5928.36	5981.07	6022.47
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41

Loi de Fisher : $\alpha = 0.01$ (suite)



Degrés de liberté du dénominateur : ν_2	Degrés de liberté du numérateur : ν_1									
	10	12	15	20	24	30	40	60	120	∞
1	6055.85	6106.32	6157.28	6208.73	6234.63	6260.65	6286.78	6313.03	6339.39	6366
2	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
4	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

Test de Durbin et Watson

Valeurs critiques au seuil $\alpha = 0.05$ pour $H_0 : \rho = 0$

p : nombre de variables explicatives (constante exclue)
n : nombre d'observations

	$p=1$		$p=2$		$p=3$		$p=4$		$p=5$	
n	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2
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