ÉCRITURE MATRICIELLE D'UN MODÈLE LINÉAIRE

Plan du chapitre 3

FPSTAT INRA 1997 III-Écriture matricielle d'un modèle linéaire-1

Écriture Matricielle d'un modèle linéaire

<u>Exemple : Analyse de Variance à deux facteurs sans interaction</u>

Étude de la teneur en huile de populations de tournesol

Testeur	Origine	Teneur en Huile
T_1	Afrique	43.54
		45.30
	Hongrie	44.25
		42.55
	Maroc	47.28
		49.40
T_2	Afrique	47.21
		47.73
	Hongrie	44.34
		46.49
	Maroc	47.75
		49.47

A = Testeur

B = Origine

$$\begin{array}{ccc} & A & \\ \swarrow & & \searrow \\ \alpha_1 & & \alpha_2 \end{array} \qquad \begin{array}{ccc} & B & \\ \swarrow & \downarrow & \searrow \\ \beta_1 & \beta_2 & \beta_3 \end{array}$$

modèle additif :

$$Y_{ijr} = \mu + \alpha_i + \beta_j + \epsilon_{ijr}$$

On suppose les postulats vérifiés
On cherche le vecteur θ de R^C, tel que $Y = X\theta + \varepsilon$
C étant le nombre de paramètres déclarés dans le modèle, soit $1(\mu) + 2(I = 2) + 3(J = 3)$

Exemple d'une ANOVA à deux facteurs à deux et trois niveaux, avec deux répétitions

$\Rightarrow$ 12 données : $Y_{111}, Y_{112}, \cdots, Y_{231}, Y_{232}$

On choisit un modèle additif :

- i : indice du facteur A : testeur
- j : indice du facteur B : origine
- r : indice de répétition
- μ : effet moyen
- α_i : effet du niveau i de A
- β_j : effet du niveau j de B
- ε_{ijr} : erreurs

$$
\begin{array}{ccccccccccccccccc}
\mathbf{Y} & = \mu & \mathbb{1} & +\alpha_1 & \mathbf{A_1} & +\alpha_2 & \mathbf{A_2} & +\beta_1 & \mathbf{B_1} & +\beta_2 & \mathbf{B_2} & +\beta_3 & \mathbf{B_3} & + & \varepsilon \\
\begin{pmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ \vdots \\ \vdots \\ \\ \\ \\ \\ \\ \\ Y_{232} \end{pmatrix} & = \mu &
\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} & +\alpha_1 &
\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} & +\alpha_2 &
\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} & +\beta_1 &
\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} & +\beta_2 &
\begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} & +\beta_3 &
\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} & + &
\begin{pmatrix} \varepsilon_{111} \\ \varepsilon_{112} \\ \varepsilon_{121} \\ \vdots \\ \vdots \\ \\ \\ \\ \\ \\ \\ \varepsilon_{232} \end{pmatrix}
\end{array}
$$

$$
\begin{pmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ \vdots \\ \vdots \\ Y_{232} \end{pmatrix} =
\begin{pmatrix}
\mu + 1 \times \alpha_1 + 0 \times \alpha_2 + 1 \times \beta_1 + 0 \times \beta_2 + 0 \times \beta_3 \\
\mu + 1 \times \alpha_1 + 0 \times \alpha_2 + 0 \times \beta_1 + 1 \times \beta_2 + 0 \times \beta_3 \\
\vdots \\
\vdots \\
\mu + 0 \times \alpha_1 + 1 \times \alpha_2 + 0 \times \beta_1 + 0 \times \beta_2 + 1 \times \beta_3
\end{pmatrix} +
\begin{pmatrix} \varepsilon_{111} \\ \varepsilon_{112} \\ \varepsilon_{121} \\ \vdots \\ \vdots \\ \varepsilon_{232} \end{pmatrix}
$$

$$
\begin{array}{ccccccc}
\mathbf{Y} & = & \mathbf{X} & \theta & + & \varepsilon \\
\begin{pmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ \vdots \\ \vdots \\ \\ \\ \\ \\ \\ \\ Y_{232} \end{pmatrix} & = &
\begin{pmatrix}
1 & 1 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 \\
1 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 1 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 0 & 1
\end{pmatrix} &
\begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} & + &
\begin{pmatrix} \varepsilon_{111} \\ \varepsilon_{112} \\ \varepsilon_{121} \\ \vdots \\ \vdots \\ \\ \\ \\ \\ \\ \\ \varepsilon_{232} \end{pmatrix}
\end{array}
$$

— X : Matrice du plan d'expérience (N,C) avec N=12 et C=6

— Y : vecteur de R^N

— ε vecteur de R^N

— $X\theta$: vecteur R^N combinaison linéaire des C vecteurs colonne de X : $\mathbb{1}$, A_1, A_2, B_1, B_2, B_3

$$Y_n = \mu \mathbb{1} + \alpha_1 A_1 + \alpha_2 A_2 + \beta_1 B_1 + \beta_2 B_2 + \beta_3 B_3 + \varepsilon$$

est structuré en

$$\begin{array}{cccccc}
 & \multicolumn{2}{c}{\text{facteur A}} & \multicolumn{3}{c}{\text{facteur B}} \\
\mu & \alpha_1 & \alpha_2 & \beta_1 & \beta_2 & \beta_3 \\
\downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
\mathbb{1} & A_1 & A_2 & B_1 & B_2 & B_3
\end{array}$$

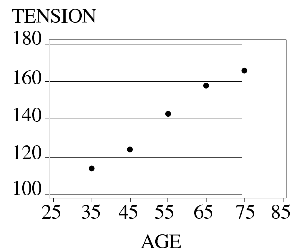
$$X = \begin{pmatrix}
1 & 1 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1 & 0 & 1 & 0 & 0 & 1
\end{pmatrix}$$

FPSTAT INRA 1997 III-Écriture matricielle d'un modèle linéaire-4

Exemple : Régression linéaire simple

Relation entre tension artérielle et âge

Données	
(Z) Âge	(Y) Tension
35	114
45	124
55	143
65	158
75	166



modèle :

$$Y_n = \alpha + \beta Z_n + \varepsilon_n, \qquad n = 1 \cdots 5$$

N = 5 données dans le modèle

α = ordonnée à l'origine

β = pente de la droite

ε_n = erreurs

FPSTAT INRA 1997 III-Écriture matricielle d'un modèle linéaire-5

$$Y = \alpha \; \mathbb{1} + \beta \; Z + \varepsilon$$

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{pmatrix} = \alpha \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \beta \begin{pmatrix} 35 \\ 45 \\ 55 \\ 65 \\ 75 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{pmatrix}$$

$$Y = X \; \theta + \varepsilon$$

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{pmatrix} = \begin{pmatrix} 1 & 35 \\ 1 & 45 \\ 1 & 55 \\ 1 & 65 \\ 1 & 75 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{pmatrix}$$

— X matrice (N,C) avec N = 5 et C = 2

— Y vecteur de $\mathbb{R}^N$

— ε vecteur de $\mathbb{R}^N$

— $X\theta$ vecteur de $\mathbb{R}^N$ combinaison linéaire des C vecteurs colonne de X : $\mathbb{1}, Z$

Exemple : une généralisation de l'analyse de covariance

Relation entre volume de bois et hauteur d'arbre par forêt d'origine

Forêt 1		Forêt 2		Forêt 3	
Hauteur	Volume	Hauteur	Volume	Hauteur	Volume
23.4	0.26	18.9	0.26	22.5	0.25
24.4	0.28	21.1	0.28	22.9	0.24
24.6	0.27	21.1	0.27	29.7	0.29
24.9	0.28	22.1	0.27	24.0	0.27
25.0	0.30	22.5	0.29	24.0	0.27
26.2	0.31	23.5	0.32		
		24.5	0.30		
Z_{1r}	Y_{1r}	Z_{2r}	Y_{2r}	Z_{3r}	Y_{3r}

modèle :

$$Y_{ir} = \mu + \alpha_i + \beta Z_{ir} + \delta_i Z_{ir} + \varepsilon_{ir}$$

On émet l'hypothèse que les hauteurs diffèrent entre les forêts et qu'il existe des relations hauteur/volume, différentes selon les forêts.

$N_1 + N_2 + N_3 = 6 + 7 + 5$ données dans le modèle.

$i =$ numéro de la forêt i=1..3
$r =$ numéro de l'arbre r=1..n_i
$\mu =$ moyenne générale (ordonnée à l'origine)
$\alpha_i =$ effet de la forêt i sur l'ordonnée à l'origine
$\beta =$ pente "moyenne"
$\delta_i =$ effet de la forêt i sur la pente
$\varepsilon_{ir} =$ erreurs

$$
\begin{array}{ccccccccccccccccccc}
Y & =\mu & 1\!\!1 & +\alpha_1 & A_1 & +\alpha_2 & A_2 & +\alpha_3 & A_3 & +\beta & Z & +\delta_1 & Z_1 & +\delta_2 & Z_2 & +\delta_3 & Z_3 & + & \varepsilon \\
\begin{pmatrix} Y_{11}\\ Y_{12}\\ \vdots\\ Y_{16}\\ Y_{21}\\ Y_{22}\\ \vdots\\ Y_{27}\\ Y_{31}\\ Y_{32}\\ \vdots\\ Y_{35} \end{pmatrix}
& =\mu &
\begin{pmatrix} 1\\ 1\\ \vdots\\ 1\\ 1\\ 1\\ \vdots\\ 1\\ 1\\ 1\\ \vdots\\ 1 \end{pmatrix}
& +\alpha_1 &
\begin{pmatrix} 1\\ 1\\ \vdots\\ 1\\ 0\\ 0\\ \vdots\\ 0\\ 0\\ 0\\ \vdots\\ 0 \end{pmatrix}
& +\alpha_2 &
\begin{pmatrix} 0\\ 0\\ \vdots\\ 0\\ 0\\ 1\\ \vdots\\ 1\\ 0\\ 0\\ \vdots\\ 0 \end{pmatrix}
& +\alpha_3 &
\begin{pmatrix} 0\\ 0\\ \vdots\\ 0\\ 0\\ 0\\ \vdots\\ 0\\ 1\\ 1\\ \vdots\\ 1 \end{pmatrix}
& +\beta &
\begin{pmatrix} 23.4\\ 24.4\\ \vdots\\ 26.2\\ 18.9\\ 21.1\\ \vdots\\ 24.5\\ 22.5\\ 22.9\\ \vdots\\ 24.0 \end{pmatrix}
& +\delta_1 &
\begin{pmatrix} 23.4\\ 24.4\\ \vdots\\ 26.2\\ 0\\ 0\\ \vdots\\ 0\\ 0\\ 0\\ \vdots\\ 0 \end{pmatrix}
& +\delta_2 &
\begin{pmatrix} 0\\ 0\\ \vdots\\ 0\\ 18.9\\ 21.1\\ \vdots\\ 24.5\\ 0\\ 0\\ \vdots\\ 0 \end{pmatrix}
& +\delta_3 &
\begin{pmatrix} 0\\ 0\\ \vdots\\ 0\\ 0\\ 0\\ \vdots\\ 0\\ 22.5\\ 22.9\\ \vdots\\ 24.0 \end{pmatrix}
& + &
\begin{pmatrix} \varepsilon_{11}\\ \varepsilon_{12}\\ \vdots\\ \varepsilon_{16}\\ \varepsilon_{21}\\ \varepsilon_{22}\\ \vdots\\ \varepsilon_{27}\\ \varepsilon_{31}\\ \varepsilon_{32}\\ \vdots\\ \varepsilon_{35} \end{pmatrix}
\end{array}
$$

$$
\begin{array}{ccccc}
Y & = & X & \theta & + \quad \varepsilon
\end{array}
$$

$$
\begin{pmatrix} Y_{11}\\ Y_{12}\\ \vdots\\ Y_{16}\\ Y_{21}\\ Y_{22}\\ \vdots\\ Y_{27}\\ Y_{31}\\ Y_{32}\\ \vdots\\ Y_{35} \end{pmatrix}
=
\begin{pmatrix}
1 & 1 & 0 & 0 & 23.4 & 23.4 & 0 & 0\\
1 & 1 & 0 & 0 & 24.4 & 24.4 & 0 & 0\\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\
1 & 1 & 0 & 0 & & 26.2 & 0 & 0\\
1 & 0 & 1 & 0 & & 0 & 18.9 & 0\\
1 & 0 & 1 & 0 & & 0 & 21.1 & 0\\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\
1 & 0 & 1 & 0 & & 0 & 24.5 & 0\\
1 & 0 & 0 & 1 & & 0 & 0 & 22.5\\
1 & 0 & 0 & 1 & & 0 & 0 & 22.9\\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\
1 & 0 & 0 & 1 & 24.0 & 0 & 0 & 24.0
\end{pmatrix}
\begin{pmatrix} \mu\\ \alpha_1\\ \alpha_2\\ \alpha_3\\ \beta\\ \delta_1\\ \delta_2\\ \delta_3 \end{pmatrix}
+
\begin{pmatrix} \varepsilon_{11}\\ \varepsilon_{12}\\ \vdots\\ \varepsilon_{16}\\ \varepsilon_{21}\\ \varepsilon_{22}\\ \vdots\\ \varepsilon_{27}\\ \varepsilon_{31}\\ \varepsilon_{32}\\ \vdots\\ \varepsilon_{35} \end{pmatrix}
$$

— X matrice (N,C) avec N = 18 et C = 8

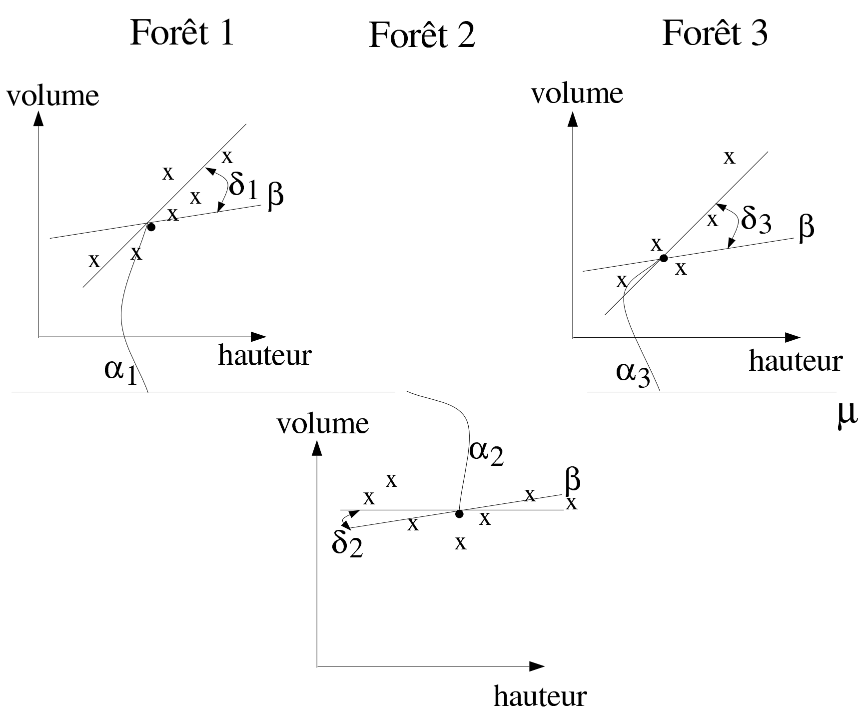
— Y vecteur de $\mathbb{R}^N$

— ε vecteur de $\mathbb{R}^N$

— $X\theta$ vecteur de $\mathbb{R}^N$ combinaison linéaire des C vecteurs colonne de X :

$1\!\!1, A_1, A_2, A_3, Z, Z_1, Z_2, Z_3$

Cette décomposition en vecteurs révèle bien que selon les paramètres on s'intéresse soit à l'ensemble des données (μ, β), soit aux données de chaque forêt $(\alpha_1, \delta_1, \cdots, \alpha_3, \delta_3)$.



FPSTAT INRA 1997
III-Écriture matricielle d'un modèle linéaire-9

En résumé

Tous les modèles linéaires s'écrivent :

$$\begin{array}{ccccccc} Y = & X & & \theta & + & \varepsilon \\ (N,1) & \underbrace{(N,C) \quad (C,1)}_{\Downarrow} & & & & (N,1) \\ & E(Y) & & & & \Downarrow \\ & & & & & \mathrm{Var}(Y) \end{array}$$

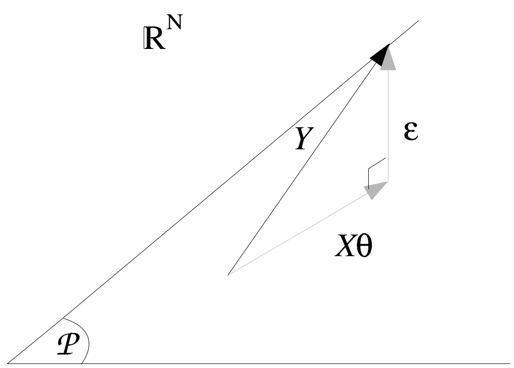
• 4 postulats

FPSTAT INRA 1997 III-Écriture matricielle d'un modèle linéaire-10

Projection orthogonale : la méthode des moindres carrés

$$Y = X\theta + \varepsilon$$

$X\theta$: combinaison linéaire des C vecteurs colonne de X.



$\mathcal{P}$: sous-espace vectoriel de $\mathbb{R}^N$ engendré par les C vecteurs colonne de X. Ce sous-espace correspond au modèle linéaire choisi pour prédire le vecteur Y des observations $(Y \in \mathbb{R}^N)$

Soit $\mathcal{P}$ le sous-espace vectoriel de $\mathbb{R}^N$, engendré par les C vecteurs colonne de X.

Le vecteur Y est décomposé en :

— un vecteur $X\theta$ de $\mathcal{P}$ $\rightarrow$ modèle choisi $X\theta = \mu_n = E(Y)$

— un vecteur ε de $\mathbb{R}^N$ $\rightarrow$ erreur du modèle $\varepsilon = Y - X\theta = Y - \mu_n = Y - E(Y)$

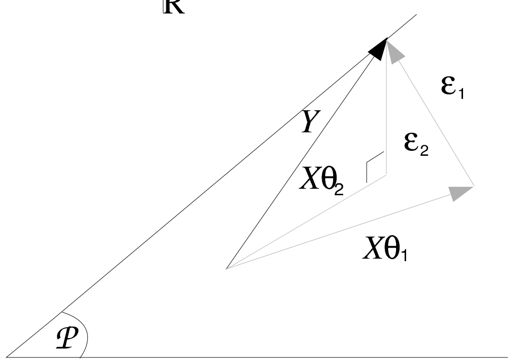
$X\theta \in \mathcal{P} \subset \mathbb{R}^N$

Voici une décomposition possible de Y en $X\theta$ et ε

Il existe une infinité de θ qui vérifient la décomposition

$$\mathrm{Y} = \mathrm{X}\theta + \varepsilon$$

$\mathbb{R}^N$



Il existe une infinité de θ qui vérifient la décomposition

$$\mathrm{Y} = \mathrm{X}\theta + \varepsilon$$

Il est donc nécessaire de disposer d'un critère de choix des θ.

On cherche θ tel que :

— $\mathrm{X}\theta$ ressemble le plus possible à Y

c'est à dire tel que :

— l'erreur ε soit la plus petite

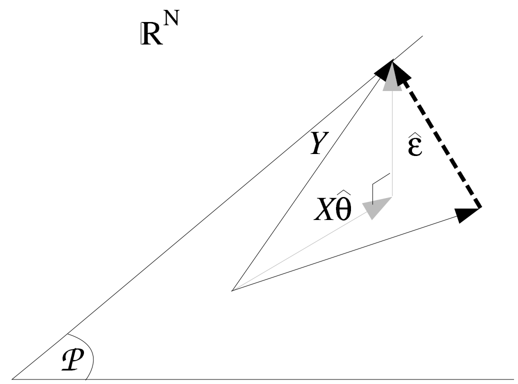
c'est à dire

— $\| \varepsilon \|^2$ soit minimale

équivalent

— $\| \mathrm{Y} - \mathrm{X}\theta \|^2$ soit minimale

Comment choisir un $\boldsymbol{\theta}$ qui rend $\| \mathbf{Y} - \mathbf{X}\boldsymbol{\theta} \|^2$ minimale ?



Une valeur de θ notée $\widehat{\theta}$ qui rend $\| Y - X\theta \|^2$ minimale.

est telle que : $X\widehat{\theta} = \mathrm{Proj}^{\perp}_{\mathcal{P}}(Y)$

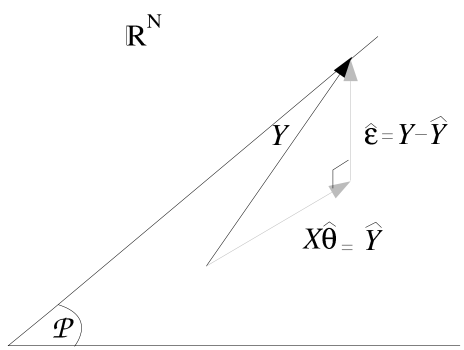
$\| Y - X\theta \|^2 = \| \varepsilon \|^2$ est minimale lorsque $X\theta \perp \varepsilon$

$\widehat{\theta}$ est donc tel que :

$X\widehat{\theta} = \mathrm{Proj}^{\perp}_{\mathcal{P}}(Y)$

$\mathrm{Proj}^{\perp}_{\mathcal{P}}(Y)$: projection orthogonale sur $\mathcal{P}$ de Y

Projeter **orthogonalement** Y sur $\mathcal{P}$



$$\|\widehat{\varepsilon}\|^2 \quad \text{minimale}$$

$$\Leftrightarrow \left\| \mathrm{Y} - \widehat{\mathrm{Y}} \right\|^2 \quad \text{minimale}$$

$$\Leftrightarrow \sum_{n=1}^{N} \left(\mathrm{Y}_n - \widehat{\mathrm{Y}}_n \right)^2 \text{minimale}$$

Projection orthogonale $\Leftrightarrow$

minimiser le critère des moindres carrés

Projeter orthogonalement Y sur $\mathcal{P}$,

— c'est rendre $\| \varepsilon \|^2$ minimale,

— c'est donc rendre la somme des carrés des écarts entre valeurs observées et valeurs ajustées minimale,

— c'est minimiser le critère des moindres carrés.

conclusion :
La meilleur estimation de Y, au sens des moindres carrés, est la projection orthogonale de Y sur $\mathcal{P}$ notée $\widehat{\mathrm{Y}}$.

$\mathcal{P}$ est le sous-espace vectoriel de $\mathbb{R}^N$ engendré par les C vecteurs colonne de X.

Comprendre la table d'analyse de la variance

<u>du modèle M_1 contre le modèle M_0 le plus simple</u>

modèle — espace

$$
\begin{tikzcd}
M_1 \arrow[r] \arrow[d, dash] & \mathcal{P} \\
M_0 \arrow[r] & \text{droite engendrée par } \mathbb{1}
\end{tikzcd}
$$

$E(Y_n) = \mu$ $\qquad$ $\mathbb{1} \subset \mathcal{P}$

$$\boxed{\text{On veut quantifier l'utilité de } M_1 \text{ par rapport à } M_0}$$

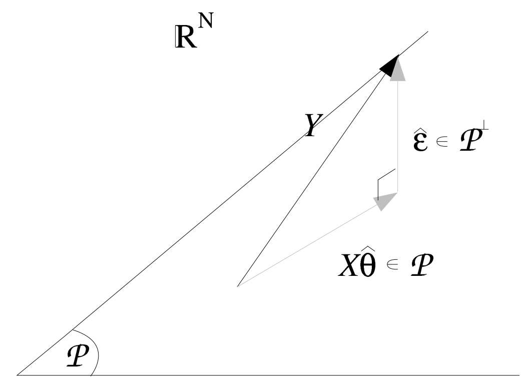
M_0 étant emboîté dans M_1 la droite engendrée par le vecteur $\mathbb{1}$ est incluse dans l'espace $\mathcal{P}$

FPSTAT INRA 1997 III-Écriture matricielle d'un modèle linéaire-15

En choisissant un modèle linéaire M_1,

on a décomposé $\mathbb{R}^{\mathrm{N}}$ en deux sous-espaces orthogonaux :

$\mathcal{P}$ et $\mathcal{P}^{\perp}$



$$\mathrm{Y} \quad = \quad \widehat{\mathrm{Y}} \quad + \quad \widehat{\varepsilon}$$

$$\mathbb{R}^{\mathrm{N}} \quad = \quad \mathcal{P} \quad + \quad \mathcal{P}^{\perp}$$

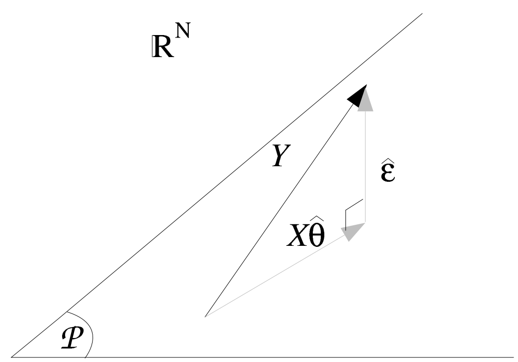
Vient de Pythagore.

En effet par construction

$$\widehat{\varepsilon} \perp \mathrm{X}\widehat{\theta}$$

FPSTAT INRA 1997 III-Écriture matricielle d'un modèle linéaire-16

Degré de liberté associé à la somme des carrés



$$\begin{array}{ccccc} Y & = & X\widehat{\theta} & + & \widehat{\varepsilon} \\ \mathbb{R}^N & = & \mathcal{P} & + & \mathcal{P}^{\perp} \\ \| Y \|^2 & = & \| X\widehat{\theta} \|^2 & + & \| \widehat{\epsilon} \|^2 \end{array}$$

Par définition le degré de liberté associé à la somme des carrés d'un vecteur est la dimension de l'espace auquel le vecteur appartient.

$$\begin{array}{ccccc} \mathrm{ddl}(\| Y \|^2) & = & \mathrm{ddl}\left(\| X\widehat{\theta} \|^2\right) & + & \mathrm{ddl}(\| \widehat{\varepsilon} \|^2) \\ \dim \mathbb{R}^N & = & \dim \mathcal{P} & + & \dim \mathcal{P}^{\perp} \\ N & = & P & + & N - P \end{array}$$

Remarque : On a toujours $P \leq C$

$\mathrm{ddl}(\| Y \|^2)$ est une notation que l'on lira **degré de liberté associé à la norme au carré de Y**

La dimension de $\mathcal{P}$ (sous espace vectoriel engendré par les C vecteurs colonne de X) = nombre de vecteurs colonne de X linéairement indépendants = P

$P \leq C$, par définition de P

Par définition, les dimensions des espaces auquels appartiennent Y, $X\widehat{\theta}$ et $\widehat{\varepsilon}$ sont appelées les degrés de liberté, c'est aussi le nombre de statistiques indépendantes de chacun des vecteurs.

Synthèse

En choisissant un modèle linéaire M_1, on a décomposé $\mathbb{R}^{\mathrm{N}}$ en deux sous-espaces orthogonaux $\mathcal{P}$ et $\mathcal{P}^\perp$

$$\begin{array}{ccccc}
\mathrm{Y} & = & \mathrm{X}\widehat{\theta}+\widehat{\varepsilon} & = & \widehat{\mathrm{Y}}+\widehat{\varepsilon} \\
\|\mathrm{Y}\|^2 & = & \|\mathrm{X}\widehat{\theta}\|^2 & + & \|\widehat{\epsilon}\|^2 \\
\mathbb{R}^{\mathrm{N}} & = & \mathcal{P} & + & \mathcal{P}^\perp \\
\dim(\mathbb{R}^{\mathrm{N}}) & = & \dim(\mathcal{P}) & + & \dim(\mathcal{P}^\perp) \\
\mathrm{ddl}(\|\mathrm{Y}\|^2) & = & \mathrm{ddl}\Big(\|\mathrm{X}\widehat{\theta}\|^2\Big) & + & \mathrm{ddl}(\|\widehat{\epsilon}\|^2) \\
\mathrm{N} & = & \mathrm{P} & + & \mathrm{N}-\mathrm{P}
\end{array}$$

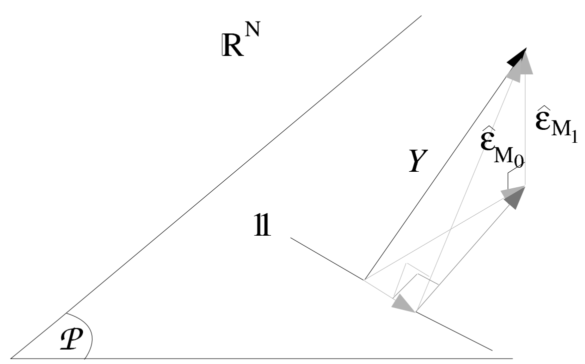
Sous les 4 postulats, et sous H_0

— $\|\mathrm{Y}\|^2 \sim \sigma^2 \quad \chi^2(\mathrm{N})$

— $\|\mathrm{X}\widehat{\theta}\|^2 \sim \sigma^2 \quad \chi^2(\mathrm{P})$

— $\|\widehat{\varepsilon}\|^2 \sim \sigma^2 \quad \chi^2(\mathrm{N}-\mathrm{P})$

— $\|\mathrm{X}\widehat{\theta}\|^2$ et $\|\widehat{\varepsilon}\|^2$ sont des χ^2 **indépendants** par construction



$$Y = \mathbb{1}Y_{.} + \hat{\varepsilon}_{M_0} \qquad \text{modèle } M_0$$

$$Y = X\hat{\theta} + \hat{\varepsilon}_{M_1} \qquad \text{modèle } M_1$$

donc

$$\mathbb{1}Y_{.} + \hat{\varepsilon}_{M_0} = X\hat{\theta} + \hat{\varepsilon}_{M_1}$$

$$\hat{\varepsilon}_{M_0} = \underbrace{\left(X\hat{\theta} - \mathbb{1}Y_{.}\right)}_{=} + \hat{\varepsilon}_{M_1}$$

$$\mathbb{R}^{N-1} = \mathbb{R}^{P-1} \quad \mathbb{R}^{N-P}$$

$\boxed{X\hat{\theta} - \mathbb{1}\, Y_{.} \text{ mesure l'apport de la projection sur } \mathcal{P} \text{ après la projection sur } \mathbb{1}}$

— S'intéresser à M_0, c'est projeter sur la droite engendrée par $\mathbb{1}$.

$\text{proj}^{\perp}_{\mathbb{1}}(Y) = \mathbb{1}\, Y_{.} \qquad \text{proj}^{\perp}_{\mathbb{1}^{\perp}}(Y) = \hat{\varepsilon}_{M_0}$

— S'intéresser à M_1, c'est projeter sur $\mathcal{P}$.

$\text{proj}^{\perp}_{\mathcal{P}}(Y) = X\hat{\theta} \qquad \text{proj}^{\perp}_{\mathcal{P}^{\perp}}(Y) = \hat{\varepsilon}_{M_1}$

— $\mathbb{1} \subset \mathcal{P}$ donc $\mathcal{P}^{\perp} \subset \mathbb{1}^{\perp}$

cela implique que $\underset{\hat{\varepsilon}_{M_1}}{\underset{=}{\text{proj}^{\perp}_{\mathcal{P}^{\perp}}(Y)}} = \text{proj}^{\perp}_{\mathcal{P}^{\perp}}\Big(\underset{\hat{\varepsilon}_{M_0}}{\underset{=}{\text{proj}^{\perp}_{\mathbb{1}^{\perp}}(Y)}}\Big)$

donc si on projette $\hat{\varepsilon}_{M_0}$ sur $\mathcal{P}$

On trouvera $\hat{\varepsilon}_{M_0} = \text{proj}^{\perp}_{\mathcal{P}}\Big(\text{proj}^{\perp}_{\mathbb{1}^{\perp}}(Y)\Big) + \hat{\varepsilon}_{M_1}$

Mais comme on a aussi $\hat{\varepsilon}_{M_0} = X\hat{\theta} - \mathbb{1}Y_{.} + \hat{\varepsilon}_{M_1}$

On en déduit que $X\hat{\theta} - \mathbb{1}Y_{.} \perp \hat{\varepsilon}_{M_1}$

Et que $X\hat{\theta} - \mathbb{1}Y_{.}$

mesure bien l'apport de la projection sur $\mathcal{P}$ après la projection sur $\mathbb{1}$

Table d'analyse de la variance

Source de variation	SCE	ddl
Modèle	$\mathrm{SCE}_{M_0} - \mathrm{SCE}_{M_1}$	P-1
Résiduelle	SCE_{M_1}	N-P
Totale du M_0	SCE_{M_0}	N-1

Source de variation	SCE	ddl=dim de
Apport de M_1 par rapport à M_0	$\lVert \widehat{\varepsilon}_{M_0} - \widehat{\varepsilon}_{M_1} \rVert^2 = \lVert X\widehat{\theta} - \mathbb{1}Y. \rVert^2$	$\mathbb{R}^{P-1}$
Erreur du modèle M_1	$\lVert \widehat{\varepsilon}_{M_1} \rVert^2$	$\mathbb{R}^{N-P}$
Erreur du modèle M_0	$\lVert \widehat{\varepsilon}_{M_0} \rVert^2$	$\mathbb{R}^{N-1}$

Le test de M_1 par rapport à M_0 fournit une table d'analyse de la variance

Avec les 4 postulats, sous H_0

$$\lVert X\widehat{\theta} - \mathbb{1}Y. \rVert^2 \sim \sigma^2 \quad \chi^2(P-1)$$

$$\lVert \widehat{\varepsilon}_{M_1} \rVert^2 \sim \sigma^2 \quad \chi^2(N-P)$$

Nous avons vu que

$$X\widehat{\theta} - \mathbb{1}Y. \perp \widehat{\varepsilon}_{M_1}$$

$\Rightarrow \lVert X\widehat{\theta} - \mathbb{1}Y. \rVert^2$ et $\lVert \widehat{\varepsilon}_{M_1} \rVert^2$ sont des χ^2 **indépendants**

d'où

$$\frac{\lVert X\widehat{\theta} - \mathbb{1}Y. \rVert^2/(P-1)}{\lVert \widehat{\varepsilon}_{M_1} \rVert^2/(N-P)} \underset{\text{sous } H_0}{\sim} F(P-1,\ N-P)$$

L'estimation de θ pose-t-elle problème ?

$\widehat{\theta}$ n'est pas forcément unique

Par construction $\widehat{\mathrm{Y}}$ est unique et s'écrit sous la forme $\widehat{\mathrm{Y}} = \mathrm{X}\widehat{\theta}$

Problème :

Existe-t-il un unique vecteur $\widehat{\theta}$ tel que $\widehat{\mathrm{Y}} = \mathrm{X}\widehat{\theta}$?

— oui si les vecteurs colonne de X forment une base de $\mathcal{P}$

— non dans le cas contraire

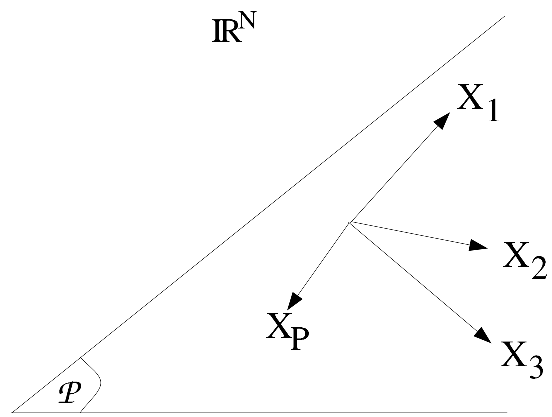
$\widehat{\mathrm{Y}}$ est unique car c'est la projection orthogonale de Y sur $\mathcal{P}$

Si les C vecteurs de X forment une base de X :

$$\begin{cases} \text{X est de plein rang} \\ \text{Le modèle est écrit sous une forme irréductible si } \mathrm{P} = \mathrm{C} \end{cases}$$

FPSTAT INRA 1997 III-Écriture matricielle d'un modèle linéaire-21

Les P vecteurs colonne de X forment ils une base de $\mathcal{P}$?



$\mathcal{P}$: engendré par les P vecteurs colonne de X : $X_1, X_2, \cdots, X_P$

Rappels :

Des vecteurs $x_1, \cdots, x_n$ forment une base d'un espace vectoriel $\mathcal{E}$, si tout vecteur x de $\mathcal{E}$ peut s'écrire, de manière unique, sous forme d'une combinaison linéaire des vecteurs $x_1, \cdots, x_n$ ou encore :

$x_1, \cdots, x_n$ forment une base si ces n vecteurs engendrent l'espace $\mathcal{E}$, et si $x_1, \cdots, x_n$ sont linéairement indépendants (aucun des vecteurs $x_1, \cdots, x_n$ ne peut s'écrire comme combinaison linéaire des autres).

Les C vecteurs colonne de X forment une base de $\mathcal{P}$

$\Leftrightarrow$ ils sont linéairement indépendants

$\Leftrightarrow$ aucun vecteur colonne de X ne s'exprime comme combinaison linéaire des autres vecteurs de X

FPSTAT INRA 1997 III-Écriture matricielle d'un modèle linéaire-22

Exemple : Analyse de variance à 2 facteurs sans interaction

$$X = \begin{array}{cccccc} \mathbb{1} & A_1 & A_2 & B_1 & B_2 & B_3 \\ \end{array}$$
$$X = \left(\begin{array}{cccccc}
1 & 1 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 \\
1 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 1 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 0 & 1
\end{array}\right)$$

Les vecteurs colonne de X forment ils une base de $\mathcal{P}$?

Non, en effet

$$\begin{cases} A_1 + A_2 & = \mathbb{1} \\ B_1 + B_2 + B_3 & = \mathbb{1} \end{cases}$$

Seuls C–2 = 5–2 vecteurs sont linéairement indépendants.

Existe t-il des vecteurs colonne combinaisons linéaires des autres vecteurs colonne ?

Il existe deux relations liant les vecteurs colonne entre eux.

$\Rightarrow$ les C vecteurs colonne sont liés entre eux.

Exemple : Régression linéaire simple

$$X = \begin{pmatrix} 1 & Z_1 \\ \vdots & \vdots \\ \vdots & \vdots \\ 1 & Z_N \end{pmatrix}$$

Les vecteurs colonne de X forment ils une base $\mathcal{P}$?

Oui si les (Z_n) pour $n = 1..N$ ne sont pas tous égaux.

Dans ce cas, les C = 2 vecteurs colonne de X sont linéairement indépendants.

Exercice :

Pourquoi quand les Z_n sont égaux, n'y a t-il pas indépendance linéaire ?

Si les Z_n sont égaux $\Rightarrow Z_n = z = z\mathbb{1}$

Z_n et $\mathbb{1}$ non linéairement indépendants

<u>Exemple</u> <u>:</u> <u>Généralisation</u> <u>de</u> <u>l'analyse</u> <u>de</u> <u>covariance</u>

$$X = \begin{array}{c} \begin{array}{cccccccc} \mathbb{1} & A_1 & A_2 & A_3 & Z & Z_1 & Z_2 & Z_3 \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & 0 & 23.4 & 23.4 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & 24.4 & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & 1 & 0 & \vdots & \vdots & 26.2 & 0 & \vdots \\ \vdots & 0 & 1 & \vdots & \vdots & 0 & 18.9 & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & 1 & 0 & \vdots & \vdots & 24.5 & 0 \\ \vdots & \vdots & 0 & 1 & \vdots & \vdots & 0 & 22.5 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 1 & 24.0 & 0 & 0 & 24.0 \end{pmatrix} \end{array}$$

Les C vecteurs colonne de X forment-ils une base de $\mathcal{P}$?

Non en effet il existe des vecteurs colonne qui s'expriment comme combinaison linéaire des autres :

$$\begin{cases} \mathbb{1} & = A_1 + A_2 + A_3 \\ Z & = Z_1 + Z_2 + Z_3 \end{cases}$$

Seuls C–2 = 8–2 vecteurs colonne sont linéairement indépendants

FPSTAT INRA 1997 III-Écriture matricielle d'un modèle linéaire-25

$\widehat{\theta}$ n'est pas uniquement déterminé lorsque $\mathbf{P} = \dim(\mathcal{P}) < \mathbf{C}$

Le modèle n'est pas écrit sous une forme irréductible.

Que faire ?

$\boxed{\text{Construire une base de } \mathcal{P}}$

Vocabulaire : on dit que le modèle est écrit sous forme irréductible si P = C

FPSTAT INRA 1997 III-Écriture matricielle d'un modèle linéaire-26

Méthodes de construction d'une base de $\mathcal{P}$

$\Leftrightarrow$ Construire à partir des C vecteurs colonne de X, P vecteurs linéairement indépendants

$\Leftrightarrow \left\{\begin{matrix}\text{choisir}\\ \text{imposer}\end{matrix}\right\} \mathrm{C}-\mathrm{P}$ contraintes linéaires sur le vecteur des paramètres θ

FPSTAT INRA 1997 III-Écriture matricielle d'un modèle linéaire-27

Exemple : Analyse de variance à 2 facteurs sans interaction

$$X = \begin{array}{c} \begin{array}{cccccc} \mathbb{1} & A_1 & A_2 & B_1 & B_2 & B_3 \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

C = 6 P = 4

Les vecteurs $\mathbb{1}, A_2, B_2, B_3$ sont linéairement indépendants.

D'autres quadruplets forment une base de $\mathcal{P}$, exemples :

$(\mathbb{1},\ A_1,\ B_2,\ B_3)$ ou $(A_1,\ A_2,\ B_2,\ B_3)$ ou ...

Méthode 1 : construire P vecteurs linéairement indépendants

On a déja vu que

$$\begin{cases} A_1 + A_2 = \mathbb{1} \\ B_1 + B_2 + B_3 = \mathbb{1} \end{cases}$$

$\mathbb{1}$ est combinaison linéaire de A_1, A_2

$\mathbb{1}$ est combinaison linéaire de B_1, B_2, B_3

Pour construire P = 4 vecteurs linéairement indépendants, on choisit arbitrairement de supprimer A_1 et B_1
Les P vecteurs $\mathbb{1}, A_2, B_2, B_3$ forment une base de $\mathcal{P}$

Exercice : Faire chercher les autres quadruplets de vecteurs indépendants qui forment une base de $\mathcal{P}$

Correction :

$(\mathbb{1},\ A_2,\ B_1,\ B_2)$; $(\mathbb{1},\ A_2,\ B_1,\ B_3)$;
$(A_1,\ A_2,\ B_1,\ B_2)$; $(A_1,\ A_2,\ B_1,\ B_3)$;
$(\mathbb{1},\ A_1,\ B_1,\ B_2)$; $(\mathbb{1},\ A_1,\ B_1,\ B_3)$;
$(A_1,\ B_1,\ B_2,\ B_3)$; $(A_2,\ B_1,\ B_2,\ B_3)$.

Le modèle écrit sous la forme

$\mathrm{Y} = \mathrm{X}\theta + \varepsilon$ devient $\mathrm{Y} = \widetilde{\mathrm{X}}\widetilde{\theta} + \varepsilon$

$$\widetilde{\mathrm{X}}\widetilde{\theta} = \underbrace{\begin{array}{c} \begin{array}{cccccc} \mathbb{1} & \mathrm{A}_1 & \mathrm{A}_2 & \mathrm{B}_1 & \mathrm{B}_2 & \mathrm{B}_3 \end{array} \\ \begin{pmatrix} 1 & . & 0 & . & 0 & 0 \\ 1 & . & 0 & . & 0 & 0 \\ 1 & . & 0 & . & 1 & 0 \\ 1 & . & 0 & . & 1 & 0 \\ 1 & . & 0 & . & 0 & 1 \\ 1 & . & 0 & . & 0 & 1 \\ 1 & . & 1 & . & 0 & 0 \\ 1 & . & 1 & . & 0 & 0 \\ 1 & . & 1 & . & 1 & 0 \\ 1 & . & 1 & . & 1 & 0 \\ 1 & . & 1 & . & 0 & 1 \\ 1 & . & 1 & . & 0 & 1 \end{pmatrix} \end{array}}_{\widetilde{X}} \cdot \underbrace{\begin{pmatrix} \mu \\ \alpha_2 \\ \beta_2 \\ \beta_3 \end{pmatrix}}_{\widetilde{\theta}}$$

La transformation $\mathrm{X} \to \widetilde{\mathrm{X}} \Leftrightarrow \theta \to \widetilde{\theta}$

$\Leftrightarrow$ ajouter les contraintes $\begin{cases} \alpha_1 = 0 \\ \beta_1 = 0 \end{cases}$

Transformer X en $\widetilde{\mathrm{X}}$ revient à :

— transformer θ en $\widetilde{\theta}$

— fixer la paramétrisation en ajoutant **les contraintes** $\begin{cases} \alpha_1 = 0 \\ \beta_1 = 0 \end{cases}$

Exemple : Analyse de variance à 2 facteurs sans interaction

Une autre méthode C = 6, P = 4 $\Rightarrow$ choisir C-P=2 contraintes indépendantes sur le vecteur θ

$$\begin{cases} \alpha_1 + \alpha_2 & = 0 \\ \beta_1 + \beta_2 + \beta_3 & = 0 \end{cases}$$

$$\Leftrightarrow \mathrm{X}\theta = \mu.\mathbb{1} - \alpha_2.\mathrm{A}_1 + \alpha_2.\mathrm{A}_2 - (\beta_2 + \beta_3).\mathrm{B}_1 + \beta_2.\mathrm{B}_2 + \beta_3.\mathrm{B}_3$$

$$\mathrm{X}\theta = \widetilde{\mathrm{X}}\widetilde{\theta} = \mu.\mathbb{1} + \alpha_2.(\mathrm{A}_2 - \mathrm{A}_1) + \beta_2.(\mathrm{B}_2 - \mathrm{B}_1) + \beta_3.(\mathrm{B}_3 - \mathrm{B}_1)$$

$$\text{avec } \widetilde{\mathrm{X}} = \begin{array}{c} \begin{pmatrix} 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 0 \\ 1 & -1 & 1 & 0 \\ 1 & -1 & 0 & 1 \\ 1 & -1 & 0 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \end{pmatrix} \\ \begin{array}{cccc} \downarrow & \downarrow & \downarrow & \downarrow \\ \mathbb{1} & \mathrm{A}_2 - \mathrm{A}_1 & \mathrm{B}_2 - \mathrm{B}_1 & \mathrm{B}_3 - \mathrm{B}_1 \end{array} \end{array} \text{ et } \widetilde{\theta} = \begin{pmatrix} \mu \\ \alpha_2 \\ \beta_2 \\ \beta_3 \end{pmatrix}$$

Autre solution :

Utiliser un autre système de contraintes, plus naturel :

$$\begin{cases} \alpha_1 + \alpha_2 & = 0 \\ \beta_1 + \beta_2 + \beta_3 & = 0 \end{cases}$$

On introduit les contraintes dans θ

Utiliser ces contraintes revient à transformer X en $\widetilde{X}$, matrice à 4 colonnes en remplaçant les colonnes :

$\mathbb{1}$ par $\mathbb{1}$

A_2 par $\mathrm{A}_2 - \mathrm{A}_1$

B_2 par $\mathrm{B}_2 - \mathrm{B}_1$

B_3 par $\mathrm{B}_3 - \mathrm{B}_1$

et en supprimant les colonnes A_1 et B_1

Cela s'appelle recentrer la matrice X

<u>Exercice</u> :

Reprendre les données de l'exemple sur la généralisation de l'analyse de covariance de ce chapitre (page III-7)

Le modèle est il écrit sous une forme irréductible ?

— calculer C et P

Si le modèle n'est pas écrit sous une forme irréductible

— proposer une solution

FPSTAT INRA 1997 III-Écriture matricielle d'un modèle linéaire-31

Exercice correction :

Le modèle n'est pas écrit sous une forme irréductible en effet

P = 6 et C = 8

Deux solutions équivalentes :

— Construire P = 6 vecteurs linéairement indépendants à partir des C = 8 vecteurs colonne de X

— Imposer C - P contraintes linéaires sur le vecteur θ

FPSTAT INRA 1997 III-Écriture matricielle d'un modèle linéaire-32

Exercice correction : construire P vecteurs linéairement indépendants

$$X = \begin{array}{c} \begin{array}{cccccccc} \mathbb{1} & A_1 & A_2 & A_3 & Z & Z_1 & Z_2 & Z_3 \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & 0 & 23.4 & 23.4 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & 24.4 & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & 1 & 0 & \vdots & \vdots & 26.2 & 0 & \vdots \\ \vdots & 0 & 1 & \vdots & \vdots & 0 & 18.9 & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & 1 & 0 & \vdots & \vdots & 24.5 & 0 \\ \vdots & \vdots & 0 & 1 & \vdots & \vdots & 0 & 22.5 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 1 & 24.0 & 0 & 0 & 24.0 \end{pmatrix} \end{array}$$

C = 8   P = 6

Les vecteurs $\mathbb{1}, A_2, A_3, Z, Z_2, Z_3$ sont linéairement indépendants

On a déja vu que

$$\begin{cases} \mathbb{1} &= A_1 + A_2 + A_3 \\ Z &= Z_1 + Z_2 + Z_3 \end{cases}$$

$\mathbb{1}$ est combinaison linéaire de A_1, A_2, A_3

Z est combinaison linéaire de Z_1, Z_2, Z_3

Pour construire P = 6 vecteurs linéairement indépendants, on choisit de supprimer A_1 et Z_1

Exercice correction :

Le modèle écrit sous la forme

$Y = X\theta + \varepsilon$ devient $Y = \widetilde{X}\widetilde{\theta} + \varepsilon$

$$\widetilde{X}\widetilde{\theta} = \begin{array}{c} \begin{array}{cccccccc} \mathbb{1} & A_1 & A_2 & A_3 & Z & Z_1 & Z_2 & Z_3 \end{array} \\ \begin{pmatrix} 1 & \vdots & 0 & 0 & 23.4 & \vdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & 0 & \vdots & \vdots & \vdots & 0 & \vdots \\ \vdots & \vdots & 1 & \vdots & \vdots & \vdots & 18.9 & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & 1 & 0 & \vdots & \vdots & 24.5 & 0 \\ \vdots & \vdots & 0 & 1 & \vdots & \vdots & 0 & 22.5 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \vdots & 0 & 1 & 24.0 & \vdots & 0 & 24.0 \end{pmatrix} \end{array} \begin{pmatrix} \mu \\ \alpha_2 \\ \alpha_3 \\ \beta \\ \delta_2 \\ \delta_3 \end{pmatrix}$$

La transformation $X \to \widetilde{X} \Leftrightarrow \theta \to \widetilde{\theta}$

$\Leftrightarrow$ ajouter les contraintes $\begin{cases} \alpha_1 = 0 \\ \delta_1 = 0 \end{cases}$

Transformer X en $\widetilde{X}$ revient à :

— transformer θ en $\widetilde{\theta}$

— fixer la paramétrisation en ajoutant **les contraintes** $\begin{cases} \alpha_1 = 0 \\ \delta_1 = 0 \end{cases}$

Exercice correction : une autre méthode

On choisit C−P = 2 contraintes linéaires sur le paramètre θ

Par exemple $\begin{cases} \alpha_1 + \alpha_2 + \alpha_3 = 0 \\ \delta_1 + \delta_2 + \delta_3 = 0 \end{cases}$

On obtient alors

$$X\theta = \mu\mathbb{1} - (\alpha_2 + \alpha_3)A_1 + \alpha_2 A_2 + \alpha_3 A_3 + \beta Z - (\delta_2 + \delta_3)Z_1 + \delta_2 Z_2 + \delta_3 Z_3$$

$$X\theta = \widetilde{X}\widetilde{\theta} = \mu\mathbb{1} + \alpha_2(A_2 - A_1) + \alpha_3(A_3 - A_1) + \beta Z + \delta_2(Z_2 - Z_1) + \delta_3(Z_3 - Z_1)$$

$$\text{avec } \widetilde{X} = \begin{pmatrix} 1 & -1 & -1 & 23.4 & -23.4 & -23.4 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & -1 & -1 & \vdots & -26.2 & -26.2 \\ \vdots & 1 & 0 & \vdots & 18.9 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & 1 & 0 & \vdots & 24.5 & 0 \\ \vdots & 0 & 1 & \vdots & 0 & 22.5 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & 24.0 & 0 & 24.0 \end{pmatrix} \text{ et } \widetilde{\theta} = \begin{pmatrix} \mu \\ \alpha_2 \\ \alpha_3 \\ \beta \\ \delta_2 \\ \delta_3 \end{pmatrix}$$

Utiliser ces contraintes revient à transformer X en $\widetilde{X}$

en remplaçant les colonnes

$\mathbb{1}$ par $\mathbb{1}$

A_2 par $A_2 - A_1$

A_3 par $A_3 - A_1$

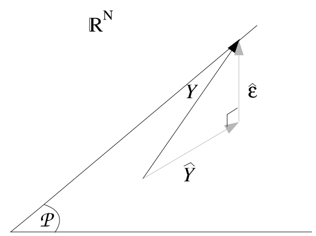
Z par Z

Z_2 par $Z_2 - Z_1$

Z_3 par $Z_3 - Z_1$

et en supprimant les colonnes A_1 et Z_1

Calcul de l'estimateur θ : $\widehat{\widetilde{\theta}}$



$\widehat{\widetilde{\theta}}$?

Egalité successive	Utilisation de
$\widehat{Y} = \widetilde{X}\,\widehat{\widetilde{\theta}}$	la multiplication par $\widetilde{X}'$ à gauche et à droite de l'égalité
$\widetilde{X}'\widehat{Y} = \widetilde{X}'\widetilde{X}\,\widehat{\widetilde{\theta}}$	$\widetilde{X}'\widehat{\varepsilon} = 0$
$\widetilde{X}'\left(\widehat{Y} + \widehat{\varepsilon}\right) = \widetilde{X}'\widetilde{X}\,\widehat{\widetilde{\theta}}$	$Y = \widehat{Y} + \widehat{\varepsilon}$
$\widetilde{X}'Y = \widetilde{X}'\widetilde{X}\,\widehat{\widetilde{\theta}}$	$\widetilde{X}'\widetilde{X}$ est carrée et inversible
$\widehat{\widetilde{\theta}} = \left(\widetilde{X}'\widetilde{X}\right)^{-1}\widetilde{X}\ Y$	

$\widetilde{X}$ est de plein rang

$\Rightarrow \widetilde{X}'\widetilde{X}$ est inversible

Objectif
On part de $\widehat{Y}$ car il est unique et l'on cherche à exprimer $\widehat{\widetilde{\theta}}$ comme une fonction de $\widetilde{X}$ et de Y car ce sont des quantités connues

— Peut-on écrire $\widehat{\widetilde{\theta}} = \widetilde{X}^{-1}\widehat{Y}$? NON, $\widetilde{X}$ de dimension (N, P) n'est pas carrée.

— mais on peut écrire $\widetilde{X}'\widehat{Y} = \widetilde{X}'\widetilde{X}\widehat{\widetilde{\theta}}$

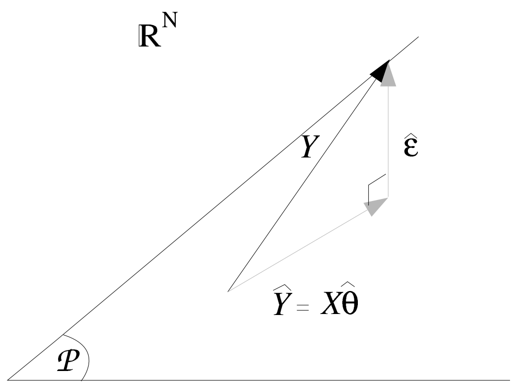
— de plus $\widehat{\varepsilon} \in P^{\perp}$ d'où $\widetilde{X}'\widehat{\varepsilon} = 0$

donc, $\widetilde{X}'(Y + \widehat{\varepsilon}) = \widetilde{X}'\widetilde{X}\widehat{\widetilde{\theta}}$

or $Y = \widehat{Y} + \widehat{\varepsilon}$

et comme $\widetilde{X}'\widetilde{X}$ de dimension (P, P) est carrée et inversible, on obtient le résultat.

Calcul de l'estimateur de la variance résiduelle



$\dfrac{\| \hat{\varepsilon} \|^2}{\mathrm{N} - \mathrm{P}}$ est un estimateur **sans biais** de σ^2 noté $\hat{\sigma}^2$

Il est aussi appelé

"carré moyen résiduel" ou

"variance résiduelle"

Sous les quatres postulats

— $\| \hat{\varepsilon} \|^2 \sim \sigma^2 \quad \chi^2(\mathrm{N} - \mathrm{P})$

donc

— $E(\| \hat{\varepsilon} \|^2) = (\mathrm{N} - \mathrm{P})\ \sigma^2$, car $E(\chi^2(\nu)) = \nu$

c'est pourquoi $\frac{\|\hat{\varepsilon}\|^2}{\mathrm{N}-\mathrm{P}}$ est un estimateur sans biais de σ^2

Il est de plus optimal, c'est à dire de variance minimale.

Fonctions linéaires estimables

— Une fonction linéaire des paramètres est estimable si son estimation ne dépend pas **des contraintes ou de la base choisie**

— Un contraste est une fonction linéaire estimable dont la somme des coefficients est nulle

<u>Exemple :</u>

$\alpha_1\ \theta_1 + \alpha_2\ \theta_2 + . + \alpha_c\ \theta_c$ est un contraste :

si son estimation ne dépend pas des contraintes choisies

et si $\alpha_1 + \cdots + \alpha_c = 0$ ou encore $\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_c \end{pmatrix} \perp \mathbb{1}$

Pour se dégager de l'arbitraire sur le choix des contraintes ou de la base de $\mathcal{P}$, le statisticien s'intéresse uniquement aux fonctions linéaires estimables, SAS ne donne d'estimation que pour ces fonctions

Remarque : le vecteur $\widehat{Y}$ est unique, il ne dépend pas des contraintes ou de la base choisie ; chacune de ses composantes est l'estimation d'une fonction linéaire estimable.

Remarque : dans un modèle sous une forme irréductible, tous les paramètres entrant dans l'écriture du modèle sont des fonctions linéaires estimables. Ce n'est plus vrai, pour les paramètres d'un modèle sous une forme réductible.

Un exemple numérique :

On utilise les données des hauteurs d'arbres en fonction de la forêt d'origine avec le modèle

modèle : $\mathrm{Y}_{\mathrm{ir}} = \mu + \alpha_{\mathrm{i}} + \varepsilon_{\mathrm{ir}}$

$$\theta = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}$$

θ n'est pas estimable, car X n'est pas de plein rang et qu'il faut mettre des contraintes

FPSTAT INRA 1997 III-Écriture matricielle d'un modèle linéaire-39

Fin de l'exemple

$$\widehat{\theta} = \begin{pmatrix} \widehat{\mu} \\ \widehat{\alpha}_1 \\ \widehat{\alpha}_2 \\ \widehat{\alpha}_3 \end{pmatrix}$$

3 contraintes usuelles

$\alpha_1 = 0$	$\mu = 0$	$\sum_{i=1}^{3} n_i \, \alpha_i = 0$
$\downarrow$	$\downarrow$	$\downarrow$
$\widehat{\theta} = \begin{pmatrix} 24.75 \\ 0. \\ -2.79 \\ -1.33 \end{pmatrix}$	$\widehat{\theta} = \begin{pmatrix} 0. \\ 24.75 \\ 21.96 \\ 23.42 \end{pmatrix}$	$\widehat{\theta} = \begin{pmatrix} 23.29 \\ 1.46 \\ -1.33 \\ 0.13 \end{pmatrix}$

Quelle fonction des paramètres est estimée identiquement dans les trois cas ?

$$\mu + \alpha_i$$

$$\alpha_i - \alpha_{i'} \Leftarrow \text{contraste}$$

Exercice :

Les estimations de μ, α_i dépendent de la contrainte choisie. Par contre, $(\mu + \alpha_i)$ et $(\alpha_i - \alpha_{i'})$ sont estimés identiquement quelle que soit la contrainte.

Faire vérifier par les stagiaires que $\mu + \alpha_i$ et $\alpha_i - \alpha_{i'}$ sont estimés identiquement dans les trois cas

Correction partielle de l'exercice

$\widehat{\mu} + \widehat{\alpha}_1 = 24.75 + 0 = 0 + 24.75 = 23.29 + 1.46$

$\widehat{\alpha}_1 - \widehat{\alpha}_2 = 0 + 2.79 = 24.75 - 21.96 = 1.46 + 1.33$

$\alpha_i - \alpha_{i'}$ est un contraste en effet :

Exemple pour le contraste $\alpha_1 - \alpha_2$
$= 0 \times \mu + 1 \times \alpha_1 - 1 \times \alpha_2 + 0 \times \alpha_3$
$= (0, 1, -1, 0).\theta$
on a bien $0 + 1 - 1 + 0 = 0$
ou encore $(0, 1, -1, 0) \perp \mathbb{1}$

Propriétés des estimateurs

Propriétés de $\widehat{\widetilde{\theta}}$ sous les 4 postulats du ML :

— $\widehat{\widetilde{\theta}}$ est un estimateur sans biais de $\widetilde{\theta}$: $E\left(\widehat{\widetilde{\theta}}\right) = \widetilde{\theta}$

— $\mathrm{Var}\left(\widehat{\widetilde{\theta}}\right) = \sigma^2\left(\widetilde{\mathrm{X}}'\widetilde{\mathrm{X}}\right)^{-1}$

— $\widehat{\widetilde{\theta}}$ est normalement distribué

En résumé

$$\widehat{\widetilde{\theta}} \sim \mathcal{N}\left(\widetilde{\theta}\ ;\ \sigma^2\left(\widetilde{\mathrm{X}}'\widetilde{\mathrm{X}}\right)^{-1}\right)$$

FPSTAT INRA 1997 III-Écriture matricielle d'un modèle linéaire-41

Propriétés des estimateurs

Propriétés de $\widehat{\sigma}^2$ sous les 4 postulats du ML :

— $\widehat{\sigma}^2 \sim \frac{\sigma^2 \chi^2 \,(\mathrm{N-P})}{\mathrm{N-P}}$

— $\widehat{\sigma}^2$ est un estimateur sans biais de σ^2 : $E\left(\widehat{\sigma}^2\right) = \sigma^2$

— $\widehat{\sigma}^2$ et $\widehat{\widehat{\theta}}$ sont des variables aléatoires indépendantes

Rappel :

$$\widehat{\sigma}^2 = \frac{\|\,\widehat{\varepsilon}\,\|^2}{\mathrm{N-P}} = \frac{\|\,\mathrm{Y} - \widehat{\mathrm{Y}}\,\|^2}{\mathrm{N-P}} = \frac{\|\,\mathrm{Y} - \widetilde{\mathrm{X}}\widehat{\widehat{\theta}}\,\|^2}{\mathrm{N-P}}$$

$E\left(\widehat{\sigma}^2\right) = \sigma^2$ car $E\left(\chi^2(\nu) = \nu\right)$

On a vu que $\widehat{\varepsilon} \in \mathcal{P}^{\perp}$ $\mathrm{X}\widehat{\widehat{\theta}} \in \mathcal{P}$ comme $\widehat{\sigma}^2$ est $\widehat{\widehat{\theta}}$ sont construits avec des variables aléatoires appartenant à des espaces orthogonaux, $\widehat{\sigma}^2$ et $\widehat{\theta}$ sont des variables aléatoires indépendantes

Démonstration

$$\mathrm{E}\left(\widehat{\widetilde{\theta}}\right) = \widetilde{\theta}$$

$$\mathrm{E}\left(\widehat{\widetilde{\theta}}\right) = \mathrm{E}\left[\left(\widetilde{\mathrm{X}}'\widetilde{\mathrm{X}}\right)^{-1}\widetilde{\mathrm{X}}'\mathrm{Y}\right]$$

On utilise que : $\mathrm{E}(\mathrm{A\ Y}) = \mathrm{AE}(\mathrm{Y})$

D'où : $\mathrm{E}\left(\widehat{\widetilde{\theta}}\right) = \left[\left(\widetilde{\mathrm{X}}'\widetilde{\mathrm{X}}\right)^{-1}\widetilde{\mathrm{X}}'\right]\mathrm{E}(\mathrm{Y})$

$$\mathrm{E}(\mathrm{Y}) = \mathrm{E}\left(\widetilde{\mathrm{X}}\,\widetilde{\theta} + \varepsilon\right) = \mathrm{E}\left(\widetilde{\mathrm{X}}\,\widetilde{\theta}\right) + 0 = \widetilde{\mathrm{X}}\,\widetilde{\theta}$$

$$\mathrm{E}\left(\widehat{\widetilde{\theta}}\right) = \left[\left(\widetilde{\mathrm{X}}'\widetilde{\mathrm{X}}\right)^{-1}\widetilde{\mathrm{X}}'\right]\left[\widetilde{\mathrm{X}}\,\widetilde{\theta}\right] = \left[\underbrace{\left(\widetilde{\mathrm{X}}'\widetilde{\mathrm{X}}\right)^{-1}\left(\widetilde{\mathrm{X}}'\widetilde{\mathrm{X}}\right)}_{\mathrm{Id}}\right]\widetilde{\theta}$$

Démonstration

$$\mathrm{Var}\,\widehat{\widetilde{\theta}} = \sigma^2\left(\widetilde{\mathrm{X}}'\widetilde{\mathrm{X}}\right)^{-1}$$

On utilise que pour une matrice A :

$$\mathrm{Var}\,(\mathrm{A\ Y}) = \mathrm{A\ Var}\,(\mathrm{Y})\mathrm{A}'$$

D'où :

$$\begin{aligned}\mathrm{Var}\left(\widehat{\widetilde{\theta}}\right) &= \mathrm{Var}\left[\left(\widetilde{\mathrm{X}}'\widetilde{\mathrm{X}}\right)^{-1}\widetilde{\mathrm{X}}'\mathrm{Y}\right] \\ &= \mathrm{Var}\left[\left(\left(\widetilde{\mathrm{X}}'\widetilde{\mathrm{X}}\right)^{-1}\widetilde{\mathrm{X}}'\right)\mathrm{Y}\right] \\ &= \left[\left(\widetilde{\mathrm{X}}'\widetilde{\mathrm{X}}\right)^{-1}\widetilde{\mathrm{X}}'\right]\mathrm{Var}(\mathrm{Y})\ \left[\left(\widetilde{\mathrm{X}}'\widetilde{\mathrm{X}}\right)^{-1}\widetilde{\mathrm{X}}'\right]'\end{aligned}$$

$$\mathrm{Var}\,(\mathrm{Y}) = \mathrm{Var}(\varepsilon) = \sigma^2\begin{pmatrix}1 & . & . & . & 0\\ . & 1 & . & . & .\\ . & . & 1 & . & .\\ . & . & . & 1 & .\\ 0 & . & . & . & 1\end{pmatrix} = \sigma^2\mathrm{Id}$$

Démonstration suite

$$\mathrm{Var}\left(\widehat{\widetilde{\theta}}\right) = \sigma^2\left[\left(\widetilde{\mathrm{X}}'\widetilde{\mathrm{X}}\right)^{-1}\widetilde{\mathrm{X}}'\right]\left[\left(\widetilde{\mathrm{X}}'\widetilde{\mathrm{X}}\right)^{-1}\widetilde{\mathrm{X}}'\right]'$$

On utilise que pour deux matrices A et B :

$$(\mathrm{AB})' = \mathrm{B}'\mathrm{A}' \text{ et } (\mathrm{A}')^{-1} = \left(\mathrm{A}^{-1}\right)'$$

$$\begin{aligned}\mathrm{Var}\left(\widehat{\widetilde{\theta}}\right) &= \sigma^2\left[\left(\widetilde{\mathrm{X}}'\widetilde{\mathrm{X}}\right)^{-1}\widetilde{\mathrm{X}}'\right]\left[\widetilde{\mathrm{X}}\left[\left(\widetilde{\mathrm{X}}'\widetilde{\mathrm{X}}\right)^{-1}\right]'\right] \\ &= \sigma^2\underbrace{\left[\left(\widetilde{\mathrm{X}}'\widetilde{\mathrm{X}}\right)^{-1}\widetilde{\mathrm{X}}'\widetilde{\mathrm{X}}\right]}_{\mathrm{Id}}\underbrace{\left[\left(\widetilde{\mathrm{X}}'\widetilde{\mathrm{X}}\right)^{-1}\right]'}_{\text{matrice symétrique}} \\ &= \sigma^2\left(\widetilde{\mathrm{X}}'\widetilde{\mathrm{X}}\right)^{-1}\end{aligned}$$

FPSTAT INRA 1997 III-Écriture matricielle d'un modèle linéaire-42

Exercice

Reprendre l'exemple des forêts

$Y_{ir} = \mu + \alpha_i + \varepsilon_{ir}$

— Donner un intervalle de confiance de $\alpha_1 - \alpha_2$

— Expliquer quel intérêt il y a d'utiliser les données de la forêt 3 pour estimer la différence moyenne des volumes de bois produits dans les forêts 1 et 2

Rappel :

Soit Y une variable aléatoire (de $\mathbb{R}^k$) distribuée selon une loi $\mathcal{N}(\mu ; V)$ (V est une matrice k $\times$ k définie positive)

Soit L la matrice d'une application linéaire de $\mathbb{R}^k$ dans $\mathbb{R}^l$ avec rang (L) = l (L est une matrice $l \times$ k)

alors

$LY \sim \mathcal{N}(L\,\mu\,; L\,V\,L')$

Correction de l'exercice

On choisit la contrainte $\alpha_3 = 0$ le modèle devient

$$Y = \widetilde{X}\,\widetilde{\theta} + \varepsilon$$

$$\text{avec } \widetilde{\theta} = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{pmatrix}$$

On cherche un intervalle de confiance de $\alpha_1 - \alpha_2$

Notons L = (0, 1, – 1)

$\alpha_1 - \alpha_2 = 0 \times \mu + 1 \times \alpha_1 - 1 \times \alpha_2 = L\widetilde{\theta}$

On cherche donc un intervalle de confiance de $L\widetilde{\theta}$

$$\Updownarrow$$

C'est à dire encadrer $L\widetilde{\theta}$ par des valeurs dépendants de

$\widehat{\widetilde{\theta}}$ et de la distribution de $\widehat{\widetilde{\theta}}$

FPSTAT INRA 1997 III-Écriture matricielle d'un modèle linéaire-44

— Estimation le $\mathrm{L}\widetilde{\theta}$ par un estimateur

$$\mathrm{L}\widehat{\widetilde{\theta}} \ \Rightarrow \ \widehat{\alpha}_1 - \widehat{\alpha}_2$$

— Calcul de la variance de $\mathrm{L}\widehat{\widetilde{\theta}}$

$$\mathrm{Var}\left(\mathrm{L}\widehat{\widetilde{\theta}}\right) = \sigma^2 \, \mathrm{L} \left(\widetilde{\mathrm{X}}'\widetilde{\mathrm{X}}\right)^{-1} \mathrm{L}'$$

— Alors $\mathrm{U} = \frac{\mathrm{L}\widehat{\widetilde{\theta}} - \mathrm{L}\widetilde{\theta}}{\sqrt{\mathrm{Var}\left(\mathrm{L}\widehat{\widetilde{\theta}}\right)}} \sim \mathcal{N}(0,1)$

Mais dans U on ne connait pas σ^2

— Estimation de σ^2 par $\widehat{\sigma}^2$

— Un estimateur sans biais de $\mathrm{L}\theta$ est donné par $\widehat{\mathrm{L}\widetilde{\theta}} = \mathrm{L}\widehat{\widetilde{\theta}}$

— $\mathrm{L}\widehat{\widetilde{\theta}}$ est une variable aléatoire normalement distribuée comme combinaison linéaire de variables aléatoires normalement distribuées

Soit :

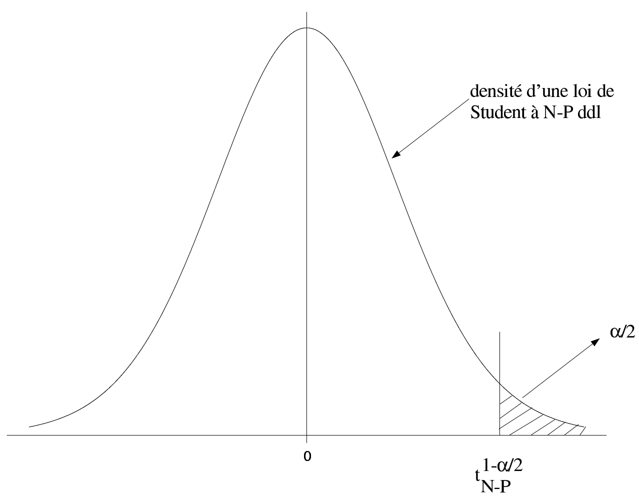
$$\mathrm{T} = \frac{\mathrm{U}}{\sqrt{\frac{\widehat{\sigma}^2}{\sigma^2}}} \sim \frac{\mathcal{N}(0\ ,\ 1)}{\sqrt{\frac{\chi^2(\mathrm{N-P})}{\mathrm{N-P}}}}$$

T suit une loi de Student à N-P degrés de liberté

D'où :

$$\mathrm{p}\Big(\mid \mathrm{T} \mid \leq \mathrm{t}_{\mathrm{N-P}}^{1-\alpha/2}\Big) = 1 - \alpha$$

où $\mathrm{t}_{\mathrm{N-P}}^{1-\alpha}$ est tel que



— $\widehat{\sigma}^2 \sim \frac{\sigma^2\chi^2(\mathrm{N-P})}{\mathrm{N-P}}$

— $\widehat{\sigma}^2$ et L $\widehat{\widehat{\theta}}$ sont indépendantes

$\Rightarrow \widehat{\sigma}^2$ et U sont indépendantes

Rappel :

(1) $\sum_{\mathrm{i}=1}^{\nu} \mathrm{Z}_{\mathrm{i}}^2 \sim \chi^2(\nu)$ si les Z_{i} sont des $\mathcal{N}(0,1)$ indépendantes

(2) $\frac{\mathcal{N}(0,1)}{\sqrt{\frac{\chi^2(\nu)}{\nu}}} \sim \mathrm{t}(\nu)$ si la $\mathcal{N}(0,1)$ et le $\chi^2(\nu)$ sont indépendants

(3) $\frac{\chi^2(\nu_1)/\nu_1}{\chi^2(\nu_2)/\nu_2} \sim \mathrm{F}(\nu_1,\nu_2)$ si les χ^2 sont indépendants

(2+3) $\Rightarrow \mathrm{t}^2(\nu) = \mathrm{F}(1,\nu)$

Correction de l'exercice : fin

En simplifiant par σ^2 on obtient comme un intervalle de confiance

$$\mathrm{L}\widehat{\widetilde{\theta}} - \mathrm{t}\,^{1-\alpha/2}_{\mathrm{N-P}} \sqrt{\widehat{\sigma}^2 \mathrm{L}\left(\widetilde{\mathrm{X}}'\widetilde{\mathrm{X}}\right)^{-1}\mathrm{L}'} \leq \mathrm{L}\widetilde{\theta} \leq \mathrm{L}\widehat{\widetilde{\theta}} + \mathrm{t}\,^{1-\alpha/2}_{\mathrm{N-P}} \sqrt{\widehat{\sigma}^2 \mathrm{L}\left(\widetilde{\mathrm{X}}'\widetilde{\mathrm{X}}\right)^{-1}\mathrm{L}'}$$

$$\widehat{\alpha}_1 - \widehat{\alpha}_2 - \mathrm{t}\,^{1-\alpha/2}_{\mathrm{N-P}} \sqrt{\widehat{\sigma}^2 \mathrm{L}\left(\widetilde{\mathrm{X}}'\widetilde{\mathrm{X}}\right)^{-1}\mathrm{L}'} < \alpha_1 - \alpha_2 \leq \widehat{\alpha}_1 - \widehat{\alpha}_2 + \underbrace{\mathrm{t}\,^{1-\alpha/2}_{\mathrm{N-P}} \sqrt{\widehat{\sigma}^2 \mathrm{L}\left(\widetilde{\mathrm{X}}'\widetilde{\mathrm{X}}\right)^{-1}\mathrm{L}'}}$$

— L'intérêt d'utiliser toutes les forêts repose sur la qualité d'estimation de la variance résiduelle (qui repose sur le postulat d'homoscédasticité).

Le degré de liberté de la loi de student est N-P et, par conséquent, le seuil de rejet $\mathrm{t}^{1-\alpha/2}_{\mathrm{N-P}}$ est inférieur à celui qui aurait été obtenue si seuls les arbres des forêts 1 et 2 avaient été utilisés dans le calcul. (Ce qui traduit que l'on estime mieux un paramètre (variance, moyenne) lorsque l'effectif observé est plus grand).