ÉCRITURE MATRICIELLE
D'UN MODÈLE LINÉAIRE

Plan du chapitre 3

1. Ecriture matricielle d'un modèle linéaire
2. Projection orthogonale : la méthode des moindres carrés
3. Comprendre la table d'analyse de variance
4. L'estimation de θ pose-t-elle problème ?
5. Méthodes de construction d'une base de $\mathcal{P}$
6. Calcul de l'estimateur de θ
7. Calcul de l'estimateur de la variance résiduelle
8. Fonctions linéaires estimables

Écriture Matricielle d'un modèle linéaire

Exemple : Analyse de Variance à deux facteurs sans interaction

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

$$A \rightarrow \alpha_1 \quad \alpha_2 \qquad B \rightarrow \beta_1 \quad \beta_2 \quad \beta_3$$

modèle additif :

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

$$
\begin{array}{cccccccccc}
\mathbf{Y} & = \mu\ \mathbb{1} & +\alpha_1\,\mathbf{A_1} & +\alpha_2\ \mathbf{A_2} & +\beta_1\ \mathbf{B_1} & +\beta_2\ \mathbf{B_2} & +\beta_3\ \mathbf{B_3} & + & \varepsilon
\end{array}
$$

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

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

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

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

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

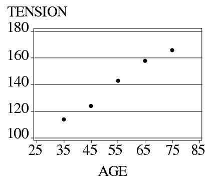
est structuré en

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

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

Relation entre tension artérielle et âge

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



modèle :

$$\mathrm{Y_n} = \alpha + \beta \mathrm{Z_n} + \varepsilon_\mathrm{n}, \qquad \mathrm{n} = 1 \cdots 5$$

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

$$\begin{pmatrix} \mathrm{Y}_1 \\ \mathrm{Y}_2 \\ \mathrm{Y}_3 \\ \mathrm{Y}_4 \\ \mathrm{Y}_5 \end{pmatrix} = \alpha \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \beta \begin{pmatrix} 35 \\ 45 \\ 55 \\ 65 \\ 75 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{pmatrix}$$

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

$$\begin{pmatrix} \mathrm{Y}_1 \\ \mathrm{Y}_2 \\ \mathrm{Y}_3 \\ \mathrm{Y}_4 \\ \mathrm{Y}_5 \end{pmatrix} = \begin{pmatrix} 1 & 35 \\ 1 & 45 \\ 1 & 55 \\ 1 & 65 \\ 1 & 75 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{pmatrix}$$

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 :

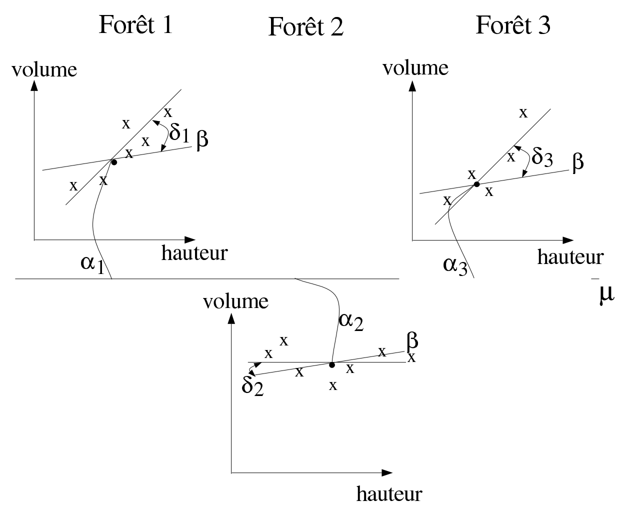
$$Y_{ir} = \mu + \alpha_i + \beta Z_{ir} + \delta_i Z_{ir} + \varepsilon_{ir}$$

$$Y \quad = \mu \quad \mathbb{1} \quad +\alpha_1 \quad A_1 \quad +\alpha_2 \quad A_2 \quad +\alpha_3 \quad A_3 \quad +\beta \quad Z \quad +\delta_1 \quad Z_1 \quad +\delta_2 \quad Z_2 \quad +\delta_3 \quad Z_3 \quad + \quad \varepsilon$$

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

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

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



<u>En résumé</u>

Tous les modèles linéaires s'écrivent :

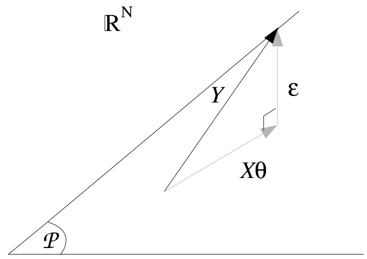
$$\underset{(\mathrm{N},1)}{\mathrm{Y}} = \underbrace{\underset{(\mathrm{N},\mathrm{C})}{\mathrm{X}} \qquad \underset{(\mathrm{C},1)}{\theta}}_{\substack{\Downarrow \\ E(\mathrm{Y})}} + \underset{\substack{\Downarrow \\ \mathrm{Var}(\mathrm{Y})}}{\underset{(\mathrm{N},1)}{\varepsilon}}$$

• 4 postulats

la méthode des moindres carrés

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

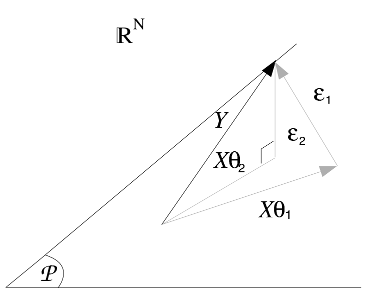
$X\theta$: combinaison linéaire des C vecteurs colonne de X.

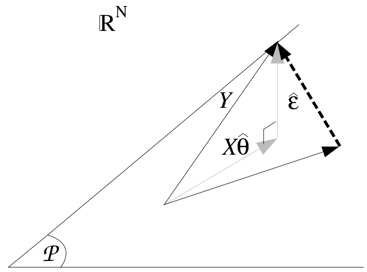


$\mathcal{P}$: sous-espace vectoriel de $\mathbb{R}^N$ engendré par les C vecteurs colonne de X. Ce sous-espace correspond au modèle linéaire choisi pour prédire le vecteur Y des observations $(Y \in \mathbb{R}^N)$

Il existe une infinité de θ qui vérifient la décomposition

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

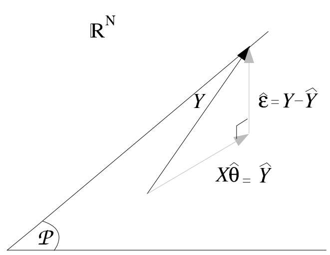




Une valeur de θ notée $\widehat{\theta}$ qui rend $\|\ \mathrm{Y} - \mathrm{X}\theta\ \|^2$ minimale.

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

Projeter orthogonalement Y sur $\mathcal{P}$



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

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

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

Projection orthogonale $\Leftrightarrow$

minimiser le critère des moindres carrés

d'analyse de la variance

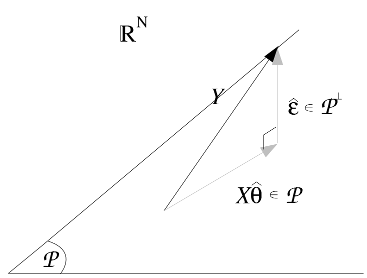
<u>du modèle $\mathbf{M_1}$ contre le modèle $\mathbf{M_0}$ le plus simple</u>

$$\begin{array}{lcr} \text{modèle} & & \text{espace} \\ M_1 & \longrightarrow & \mathcal{P} \\ | & & \\ M_0 & \longrightarrow & \text{droite engendrée par } \mathbb{1} \\ E(Y_n) = \mu & & \mathbb{1} \subset \mathcal{P} \end{array}$$

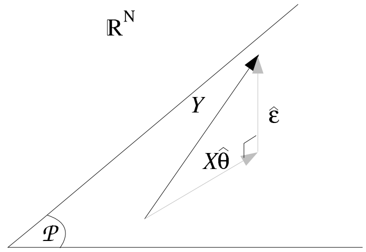
$\boxed{\text{On veut quantifier l'utilité de } M_1 \text{ par rapport à } M_0}$

En choisissant un modèle linéaire M_1,

on a décomposé $\mathbb{R}^N$ en deux sous-espaces orthogonaux : $\mathcal{P}$ et $\mathcal{P}^\perp$



$$\begin{array}{ccccc} Y & = & \hat{Y} & + & \hat{\varepsilon} \\ \mathbb{R}^N & = & \mathcal{P} & + & \mathcal{P}^\perp \end{array}$$



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

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.

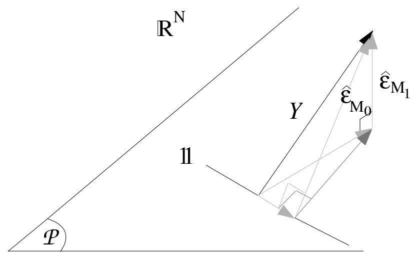
$$
\begin{array}{ccccc}
\mathrm{ddl}(\|\,\mathrm{Y}\,\|^2) & = & \mathrm{ddl}\left(\|\,\mathrm{X}\widehat{\theta}\,\|^2\right) & + & \mathrm{ddl}(\|\,\widehat{\varepsilon}\,\|^2) \\
\dim \mathbb{R}^{\mathrm{N}} & = & \dim \mathcal{P} & + & \dim \mathcal{P}^{\perp} \\
\mathrm{N} & = & \mathrm{P} & + &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

Remarque : On a toujours $\mathrm{P} \leq \mathrm{C}$

Synthèse

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

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



$$Y = \mathbb{1}Y. + \hat{\varepsilon}_{M_0} \quad \text{modèle } M_0$$
$$Y = X\hat{\theta} + \hat{\varepsilon}_{M_1} \quad \text{modèle } M_1$$

donc

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

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

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

$X\hat{\theta} - \mathbb{1}\, Y.$ mesure 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	$SCE_{M_0} - SCE_{M_1}$	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	$\Vert \hat{\varepsilon}_{M_0} - \hat{\varepsilon}_{M_1} \Vert^2 = \Vert X\hat{\theta} - \mathbb{1}Y. \Vert^2$	$\mathbb{R}^{P-1}$
Erreur du modèle M_1	$\Vert \hat{\varepsilon}_{M_1} \Vert^2$	$\mathbb{R}^{N-P}$
Erreur du modèle M_0	$\Vert \hat{\varepsilon}_{M_0} \Vert^2$	$\mathbb{R}^{N-1}$

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 $\widehat{\mathrm{Y}} = \mathrm{X}\widehat{\theta}$

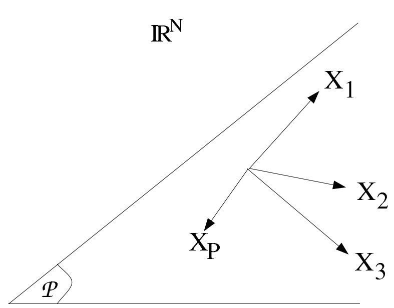
Problème :

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

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



$\mathcal{P}$: engendré par les P vecteurs colonne de X : $\mathrm{X}_1, \mathrm{X}_2, \cdots, \mathrm{X}_\mathrm{P}$

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

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.

Exemple : Régression linéaire simple

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

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.

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

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

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

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

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

Exemple : Analyse de variance à 2 facteurs sans interaction

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

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 $\ldots$

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

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

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

$\Leftrightarrow$ ajouter 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

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

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

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

$$
\text{avec } \widetilde{\mathrm{X}} = \begin{pmatrix}
1 & -1 & -1 & -1 \\
1 & -1 & -1 & -1 \\
1 & -1 & 1 & 0 \\
1 & -1 & 1 & 0 \\
1 & -1 & 0 & 1 \\
1 & -1 & 0 & 1 \\
1 & 1 & -1 & -1 \\
1 & 1 & -1 & -1 \\
1 & 1 & 1 & 0 \\
1 & 1 & 1 & 0 \\
1 & 1 & 0 & 1 \\
1 & 1 & 0 & 1
\end{pmatrix}
\text{ et } \widetilde{\theta} = \begin{pmatrix} \mu \\ \alpha_2 \\ \beta_2 \\ \beta_3 \end{pmatrix}
$$

$$
\begin{array}{cccc}
\downarrow & \downarrow & \downarrow & \downarrow \\
\mathbb{1} & \mathrm{A}_2 - \mathrm{A}_1 & \mathrm{B}_2 - \mathrm{B}_1 & \mathrm{B}_3 - \mathrm{B}_1
\end{array}
$$

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

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

vecteurs linéairement indépendants

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

$C = 8 \quad P = 6$

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

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 $X \rightarrow \widetilde{X} \Leftrightarrow \theta \rightarrow \widetilde{\theta}$

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

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

Par exemple $\left\{\begin{array}{l}\alpha_1+\alpha_2+\alpha_3=0\\ \delta_1+\delta_2+\delta_3=0\end{array}\right.$

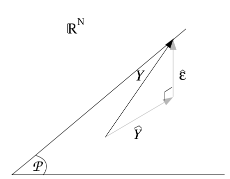
On obtient alors

$$\mathrm{X}\theta=\mu\mathbb{1}-(\alpha_2+\alpha_3)\mathrm{A}_1+\alpha_2\mathrm{A}_2+\alpha_3\mathrm{A}_3+\beta\mathrm{Z}-(\delta_2+\delta_3)\mathrm{Z}_1+\delta_2\mathrm{Z}_2+\delta_3\mathrm{Z}_3$$

$$\mathrm{X}\theta=\widetilde{\mathrm{X}}\widetilde{\theta}=\mu\mathbb{1}+\alpha_2(\mathrm{A}_2-\mathrm{A}_1)+\alpha_3(\mathrm{A}_3-\mathrm{A}_1)+\beta\mathrm{Z}+\delta_2(\mathrm{Z}_2-\mathrm{Z}_1)+\delta_3(\mathrm{Z}_3-\mathrm{Z}_1)$$

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

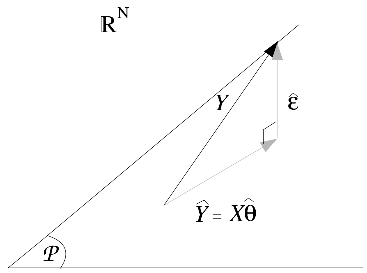
Calcul de l'estimateur θ : $\widehat{\widehat{\theta}}$



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

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

de la variance résiduelle



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

Il est aussi appelé

"carré moyen résiduel" ou

"variance résiduelle"

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

Exemple :

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

si son estimation ne dépend pas des contraintes choisies

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

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

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

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

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

Fin de l'exemple

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

3 contraintes usuelles

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

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

$$\mu + \alpha_i$$

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

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

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

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 $\widehat{\widetilde{\theta}}$ sont des variables aléatoires indépendantes

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

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

— 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 $\widehat{\sigma}^2$

Soit :

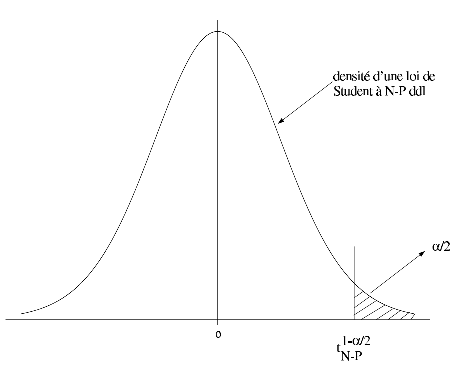
$$\mathrm{T} = \frac{\mathrm{U}}{\sqrt{\frac{\widehat{\sigma}^2}{\sigma^2}}} \sim \frac{\mathcal{N}(0\ ,\ 1)}{\sqrt{\frac{\chi^2(\mathrm{N-P})}{\mathrm{N-P}}}}$$

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

D'où :

$$\mathrm{p}\left(|\,\mathrm{T}\,| \leq \mathrm{t}_{\mathrm{N-P}}^{1-\alpha/2}\right) = 1 - \alpha$$

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



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

$$L\widehat{\widetilde{\theta}} - t_{N-P}^{1-\alpha/2}\sqrt{\widehat{\sigma}^2 L\left(\widetilde{X}'\widetilde{X}\right)^{-1}L'} \leq L\widetilde{\theta} \leq L\widehat{\widetilde{\theta}} + t_{N-P}^{1-\alpha/2}\sqrt{\widehat{\sigma}^2 L\left(\widetilde{X}'\widetilde{X}\right)^{-1}L'}$$

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

— 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 $t_{N-P}^{1-\alpha/2}$ 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).