

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA
RECHERCHE SCIENTIFIQUE
UNIVERSITÉ DES SCIENCES ET DE LA TECHNOLOGIE
HOUARI BOUMEDIENE



FACULTÉ DE MATHÉMATIQUES

=====

THESE

présentée pour l'obtention du diplôme de Magister
en Mathématiques
Spécialité : Recherche Opérationnelle

par

MEKHILEF Amal

Thème

Contribution à l'étude des modèles linéaires
multi-échelles

Soutenu publiquement, le 08/03/2008 devant le jury :

A.SEMRI

M. S. MAAMRA

S. BOUROUBI

H. GUERBYENNE

A. AKNOUCHE

Maître de conférences, USTHB

Chargé de cours, USTHB

Maître de conférences, USTHB

Maître de conférences, USTHB

Docteur, USTHB

Président

Directeur de thèse

Examineur

Examineur

Examineur

Remerciements

J'exprime ma profonde reconnaissance à Mr.MAAMRA Mohamed Said pour m'avoir proposé ce thème, pour la confiance qu'il m'a accordé, pour la documentation qu'il a mise à ma disposition, et surtout pour ses précieux conseils et l'aide apportés pour la réalisation de ce travail.

Je remercie également Mr SEMRI Ahmed qui m'a fait l'honneur de présider le jury, ainsi que Mme H.GUERBIENNE, Mr S.BOUROUBI, et Mr A. AKNOUCHE pour avoir accepté de juger ce travail.

SOMMAIRE

Résumé.....1

INTRODUCTION2

CHAPITRE 1

MODÈLES DE RÉGRESSION LINÉAIRE.....4

1. La régression linéaire.....4

2. Le modèle de régression linéaire classique.....9

3. Régression stochastique..... 14

4. Inférence statistique dans un modèle classique de régression linéaire..... 15

5- Modèle de régression linéaire généralisé 27

CHAPITRE 2

MODÈLE MULTI-NIVEAUX DE REGRESSION LINEAIRE....30

Introduction Goldstein, H(1996) 30

1. Principe fondamental de la modélisation multi-niveaux..... 32

2. Description de la modélisation multi-niveaux 34

3- Modèle linéaire à deux niveaux à une seule variable explicative 37

4. Modèle linéaire à deux niveaux à plusieurs variables explicatives..... 41

5. Structure générale d'un modèle linéaire à deux niveaux..... 44

CHAPITRE 3

ESTIMATION D'UN MODÈLE DE RÉGRESSION LINÉAIRE À DEUX NIVEAUX PAR LA MÉTHODE ITÉRATIVE DES MOINDRES CARRÉS GÉNÉRALISÉE.48

Introduction 48

1. Méthode Itérative des Moindres Carrés Généralisée : IMCG 50

2. Généralisation de la méthode IMCG 56

CHAPITRE 4

L'ESTIMATION D'UN MODELE DE REGRESSION LINEAIRE MULTI-NIVEAUX PAR LA METHODE ITERATIVE DE BOOTSTRAP58

Introduction 58

1. L'échantillon de Bootstrap Efron, B and Tibshirani, R. J(1993) 58

2. Estimation d'un modèle de régression linéaire à deux niveaux Goldstein, H(1998) 61

3. l'estimation des intervalles de confiance pour les paramètres fixes et aléatoires 63

Implémentation dans le cas du modèle des composants de la variance à deux niveaux... 73

Conclusion 77

Annexes

- Annexe 1 **Procédure méthode itérative des moindres carrés généralisée**
- Annexe 2 **Programme de la méthode itérative des moindres carrés généralisée**
- Annexe 3 **Programme : estimation par la méthode non paramétrique de Bootstrap**
- Annexe 4 **Programme : estimation par la méthode paramétrique de Bootstrap**

Résumé

Les méthodes d'analyse de données ont fourni dans le cadre des prévisions plusieurs modèles linéaires tels que les modèles de régression linéaire, l'analyse de la variance etc.

Ces modèles appliqués à une réalité à complexité croissante devinrent incapables de traiter tous les cas et d'assurer de bonnes prévisions ; dans cette optique les modèles dits multi-échelles ou multi-niveaux ont fait leur apparition entraînant une plus grande adaptation aux cas réels.

Les modèles linéaires classiques consistent en l'étude de la dépendance d'une variable expliquée d'une ou de plusieurs variables explicatives sous l'hypothèse de la non-variabilité des erreurs, quant aux modèles multi-échelles ils se basent sur la variation de l'erreur selon des classes définies à priori.

Dans la présente étude, chacun des types de modèle est traité, pour ce qui est des modèles classiques on traitera le cas de la régression linéaire multiple ; quant au deuxième type nous avons choisi parmi les modèles multi-niveaux le modèle de régression linéaire à deux niveaux.

Mots clés : structure hiérarchique, modèles multi échelles, multi niveaux, échantillon de Bootstrap.

INTRODUCTION

Certains types de données, notamment les observations recueillies en sciences humaines ou naturelles ont une structure hiérarchique. Par exemple, l'étude de l'hérédité chez les humains ou les animaux prend en considération l'hiérarchie naturelle; cette dernière consiste à grouper les descendants par familles. Les individus descendants des mêmes parents tendent à être plus similaires mentalement et physiquement que les individus choisis aléatoirement d'une population plus large ; les enfants issus d'une même famille tendent tous à avoir la même taille, ceci peut être dû soit à la taille des parents (caractère hérité) soit à leur environnement commun. Certaines expériences créent par elles mêmes une hiérarchie dans les données, dans le cas des essais médicaux faits dans plusieurs centres choisis aléatoirement ou sur différents groupes d'individus, les résultats auront cette structure hiérarchique.

Une hiérarchie consiste à grouper *des unités* par différents *niveaux* : ainsi dans une structure à deux niveaux, où les unités de deuxième niveau sont les familles, les descendants seront les unités de premier niveau; Les données à structure hiérarchique sont devenues une norme, elles sont présentes dans de nombreux domaines d'application; en enseignement, par exemple, on peut citer un cas simple: on enseigne à des élèves ou des étudiants par classes, les classes appartiennent à des écoles ou établissements, ces établissements sont à leur tour sous la tutelle de différentes autorités locales; dans un tel cas les unités s'étendent sur quatre différents niveaux d'hiérarchie, le modèle multi-niveaux de ce système associe les élèves au niveau 1, les classes au niveau 2, les écoles au niveau 3 et enfin les autorités au niveau 4; on dira que les unités de chaque niveau donné sont regroupées en unités d'un niveau immédiatement supérieur.

Il est important de savoir comment la structure hiérarchique d'une population affecte les caractéristiques d'intérêt mesurées au niveau des individus; ainsi si on se propose de mesurer le niveau d'enseignement, on sait que la moyenne variera d'une école à une autre, et ceci veut dire que les élèves aléatoirement choisis dans une même école ont des résultats plus similaires en moyenne que ceux des élèves choisis de différentes écoles.

L'existence d'une hiérarchie dans les données n'est pas accidentelle, on ne peut donc pas l'ignorer. Ceci rend invalides, les techniques traditionnelles de l'analyse statistique des données. L'objet de cette thèse est d'étudier une méthode d'analyse de ces données hiérarchiques en utilisant les modèles de régression linéaire.

Le document se présentera comme suit:

Au premier chapitre, nous présentons les modèles de régression linéaire, et leur estimation par la méthode des moindres carrés ordinaire et la méthode des moindres carrés généralisée.

Au deuxième chapitre nous introduisons la notion de modèles linéaires multi-niveaux, les notations générales dans le cas d'un modèle à deux niveaux, la variance est décomposée en deux variances, l'une due à l'unité du premier niveau, et l'autre à l'unité du deuxième niveau.

Au troisième chapitre, nous introduisons une méthode d'estimation d'un modèle de régression linéaire à deux niveaux, appelée la méthode itérative des moindres carrés généralisée (**IMCG**), elle consiste à utiliser itérativement la méthode des moindres carrés généralisée (**MCG**), à chaque itération nous estimons les variances liées respectivement aux unités de premier et de deuxième niveau. Le critère d'arrêt est donc la convergence de ces variances, nous procéderons par la suite à la comparaison entre cette méthode (**IMCG**) et la méthode des moindres carrés ordinaire (**MCO**) qui ignore la propriété multi-échelles ou la structure hiérarchique des données.

Au quatrième chapitre, nous introduisons l'échantillonnage de Bootstrap. Ce chapitre contient des définitions fondamentales concernant le Bootstrap: qu'est ce qu'un échantillon de Bootstrap? Comment l'obtenir? La méthode paramétrique et la méthode non paramétrique; nous décrivons aussi comment obtenir un estimateur d'un paramètre en utilisant un échantillon de Bootstrap et nous utiliserons par la suite cette méthode pour estimer un modèle de régression linéaire à deux niveaux.

Nous présenterons par la suite une application numérique faite sur des données simulées. Il s'agira d'une étude comparative entre les différentes méthodes: **MCO**, **IMCG** et la méthode d'estimation basée sur l'échantillonnage de Bootstrap. Et nous terminerons par une conclusion où nous aborderons les perspectives de notre étude.

Chapitre 1

MODÈLES DE RÉGRESSION LINÉAIRE

1. La régression linéaire

Dans la régression linéaire simple, on tente d'estimer une relation théorique (dans la population) entre les variables X et Y de la forme :

$$y_i = \beta_1 + \beta_2 x_i + e_i \quad i = 1, n \quad (1.1)$$

Où :

y_i est la $i^{\text{ème}}$ valeur observée de la variable dépendante (ou expliquée) Y

x_1, \dots, x_n , sont n valeurs fixes de la variable explicative X

β_1, β_2 : sont des paramètres inconnus

e_i : est la fluctuation aléatoire non observable et attribuable à un ensemble de facteurs ou de variables non pris en considération dans le modèle.

Dans le tableau ci-dessous on introduit une extension du modèle (1.1), qu'on appelle le modèle de régression linéaire multiple.

Observations	Forme fonctionnelle	Critère d'ajustement
(y_i, x_i)	$y_i = \alpha + \beta x_i + e_i$	$S(\alpha, \beta) = \sum_{i=1}^n (y_i - (\alpha + \beta x_i))^2$
(y_i, x_{i1}, x_{i2})	$y_i = \beta_1 + \beta_2 x_{i1} + \beta_3 x_{i2} + e_i$	$S(\beta_1, \beta_2, \beta_3) = \sum_{i=1}^n (y_i - (\beta_1 + \beta_2 x_{i1} + \beta_3 x_{i2}))^2$
$(y_i, x_{ij})_{j=2, k}$	$y_i = \beta_1 + \sum_{j=2}^k \beta_j x_{ij} + e_i$	$S(\beta_1, \dots, \beta_k) = \sum_{i=1}^n \left(y_i - \left(\beta_1 + \sum_{j=2}^k \beta_j x_{ij} \right) \right)^2$

Pour écrire ce modèle sous la forme matricielle, on définit les vecteurs d'observations de dimension n ,

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad x_1 = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \quad x_2 = \begin{bmatrix} x_{12} \\ \vdots \\ x_{n2} \end{bmatrix}, \dots, x_k = \begin{bmatrix} x_{1k} \\ \vdots \\ x_{nk} \end{bmatrix},$$

le k -vecteur des paramètres et le n -vecteur des erreurs :

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}, \quad e = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}$$

et la (nxk) -matrice :

$$X = \begin{bmatrix} 1 & x_{12} & \dots & \dots & x_{1k} \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ 1 & x_{n2} & \dots & \dots & x_{nk} \end{bmatrix} = [x_1, \dots, x_k]$$

La notation matricielle d'un modèle de régression linéaire multiple peut être écrite comme suit:

$$y = X\beta + e$$

et le critère d'ajustement est donné par :

$$\text{Min } S(\beta) = (y - X\beta)'(y - X\beta).$$

1.1. La méthode des moindres carrés

Cette méthode d'estimation consiste à déterminer le paramètre β qui minimise la fonction $S(\beta)$:

$$\begin{aligned} S(\beta) &= (y - X\beta)'(y - X\beta) \\ &= y'y - y'X\beta - (X\beta)'y + (X\beta)'(X\beta) \\ &= y'y - 2y'X\beta + (X\beta)'(X\beta). \end{aligned}$$

Pour résoudre ce problème de minimisation, on calcule la condition dite de premier ordre :

$$\begin{cases} \frac{\partial S(\beta)}{\partial \beta} = -2X'y + 2X'X\beta \\ \frac{\partial S(\beta)}{\partial \beta} = 0 \end{cases} \quad (1.2)$$

Par identification des deux équations précédentes nous obtenons l'équation :

$$X'y = (X'X)\beta$$

Si $(X'X)^{-1}$ existe, alors le vecteur de paramètres $\hat{\beta}$ est donné par :

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

Sous les hypothèses :

H₁- Les colonnes de X sont linéairement indépendantes, ce qui implique que $(X'X)$ est de plein rang.

H₂- $(X'X)$ est définie positive ce qui implique que $S(\beta)$ est strictement convexe et a un minimum unique.

Ainsi, $\hat{\beta}$ est déterminée de façon unique par les équations normales (1.2)

On définit les résidus par :

$$\hat{e} = y - X\hat{\beta} = y - X(X'X)^{-1} X'y = (I - X(X'X)^{-1} X')y = M_X y,$$

et les valeurs ajustées par :

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

Avec $M_X = I - X(X'X)^{-1} X'$ et $P_X = X(X'X)^{-1} X'$.

M_X et P_X sont deux matrices, aux propriétés particulières suivantes :

- M_X et P_X sont symétriques,
- $M_X + P_X = I$,
- M_X et P_X sont idempotentes $P_X = P_X P_X, M_X = M_X M_X$,
- $P_X M_X = 0, M_X X = 0, P_X X = X$.

1.2 La qualité de l'ajustement (goodness of fit)

La somme totale des carrés des observations de la variable Y ($\sum y_i^2$), satisfaisant à un modèle de régression linéaire multiple notée SST peut être décomposée en deux quantités positives :

- 1- Somme des carrés expliquée par le modèle, notée SSR.
- 2- Somme des carrés des résidus notée SSE, et qui représente la variabilité de y non expliquée par le modèle.

$$y = \hat{y} + \hat{e}$$

La somme des carrés des erreurs est donnée par :

$$\begin{aligned} \text{SSE} &= \hat{e}'\hat{e} = S(\hat{\beta}) = (y - X\hat{\beta})'(y - X\hat{\beta}) = (M_X y)'(M_X y) \\ &= y'M_X' M_X y = y'M_X M_X y = y'M_X y. \end{aligned}$$

La somme des carrés des valeurs ajustées est donnée par

$$\text{SSR} = \hat{y}'\hat{y} = (X\hat{\beta})'(X\hat{\beta})$$

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

Pour avoir une idée sur la contribution des variables explicatives, on prend le modèle de la régression simple qui ne contient aucune variable explicative soit :

$$y_i = \beta_0 + e_i$$

En utilisant la méthode des moindres carrés on obtient $\hat{\beta}_0 = \bar{y}$

Ainsi $\hat{y} = \bar{y}$

D'où :

$$\sum_{i=1}^n y_i^2 = \sum_{i=1}^n \hat{y}_i^2 + \sum_{i=1}^n \hat{e}_i^2 = n\bar{y}^2 + \sum_{i=1}^n \hat{e}_i^2$$

$$n\bar{y}^2 = \sum_{i=1}^n y_i^2 - \sum_{i=1}^n \hat{e}_i^2 = SST - SSE = SSR$$

Ainsi $n\bar{y}^2$ est la somme des carrés expliquée par le paramètre β_0 dit intercept. On décompose SSR en deux quantités :

SSM : La partie de la somme des carrés expliquée par l'intercept,

SSR_m : La partie de la somme des carrés due aux variables explicatives,

où : $SSR = SSM + SSR_m$

Donc : $SST = SSM + SSR_m + SSE$.

Table d'analyse de la variance

Les résultats de la partition de la somme des carrés sont résumés dans un tableau dit tableau de l'analyse de la variance ANOVA.

Source de variation	Degré de liberté.	Somme des carrés	Moyenne
Régresseur	k	$SSR = \hat{\beta}X'y$	$MSR = SSR/k$
Erreur	$n-k$	$SSE = y'y - \hat{\beta}X'y$	$MSE = SSE/(n-k)$
Total	n	$SST = SSR + SSE$	

Source de variation	Degré de liberté.	Somme des carrés	Moyenne
Moyenne	1	$SSM = n\bar{y}^2$	SSM
Régresseur	$k-1$	$SSR_m = SSR - n\bar{y}^2$	$MSR_m = SSR_m / (k-1)$
Erreur	$n-k$	$SSE = y'y - (SSM + SSR_m)$	$MSE = SSE / (n-k)$
Total	n	$SST = SSM + SSR_m + SSE$	

Une autre forme de la table de l'ANOVA est :

Source de variation	Degré de liberté.	Somme des carrés	Moyenne
Moyenne	1	$SSM = n\bar{y}^2$	SSM
Régresseur	$k-1$	$SSR_m = SSR - n\bar{y}^2$	$MSR_m = SSR_m / (k-1)$
Erreur	$n-k$	$SSE = y'y - (SSM + SSR_m)$	$MSE = SSE / (n-k)$
Total	n	$SST_m = SSR_m + SSE$	

Le coefficient de détermination

Soit :

$$R^2 = \begin{cases} \frac{SSR}{SST} & \text{sans intercept} \\ \frac{SSR_m}{SST_m} & \text{avec intercept} \end{cases}$$

où

$$R^2 = \frac{SSR}{SST}$$

représente le pourcentage de variation expliqué par le modèle.

et

$$R^2 = \frac{SSR_m}{SST_m}$$

représente le pourcentage de variation expliqué par les variables explicatives.

et R représente le coefficient de corrélation linéaire entre les variables Y et X .

Le coefficient de détermination R^2 mesure la quantité de variabilité expliquée par le modèle. Il varie entre 0 et 1 où 0 définit une non corrélation et 1 définit une corrélation parfaite.

Intuitivement, on peut déjà prendre SSE comme un facteur possible pour juger de la qualité du modèle, plus la valeur prise par SSE sera petite (SSR sera alors relativement important) plus on pensera que la significativité du modèle est bonne et vice-versa et bien c'est à partir de cette intuition qu'a été défini le coefficient de détermination qui représente en fait le rapport de la somme des carrés expliquée sur la somme totale des carrés.

Remarque

Lorsque l'on rajoute une variable explicative au modèle, le coefficient de détermination ne peut diminuer, en se basant uniquement sur ce dernier, on serait tenté d'accroître le nombre de

variables explicatives afin d'augmenter la proportion de variance expliquée mais l'amélioration, sera totalement artificielle; car celle-ci peut être due à une autre variable.

Coefficient de détermination corrigée (Adjusted R.square)

D'après la remarque faite précédemment, nous avons constaté que R^2 augmente lorsque le nombre de variables explicatives augmente ou reste égale à lui même. Maintenant plutôt que R^2 on définit le coefficient de détermination corrigé \bar{R}^2 dont le rôle est de mesurer la qualité de variabilité expliquée par les variables explicatives (alors que le coefficient de détermination R^2 mesure la qualité de variabilité expliquée par le modèle). Son équation sera similaire à celle de R^2 en remplaçant les deux variances par des estimateurs non biaisés, on obtient alors:

$$\bar{R}^2 = R^2 - \left(\frac{k-2}{n-k-1} \right) (1-R)$$

\bar{R}^2 est donc inférieur à R^2 et peut même être négatif. En se basant sur le coefficient de détermination corrigé, il n'y aura donc pas avantage à introduire des variables explicatives en plus.

2. Le modèle de régression linéaire classique

2.1. Hypothèses

Pour le modèle de régression, linéaire multiple

$$y = X\beta + e \quad (1.3)$$

On commence par le système d'hypothèses suivant :

- A₁ $y = X\beta + e$ $\beta \in \mathbb{R}^k$,
- A₂ X : une (n x k) matrice avec $\text{rang}(X) = k$,
- A₃ $E(e / X) = 0$,
- A₄ $E(ee' / X) = \sigma^2 I$,
- A₅ X est non stochastique .

Remarques

1. L'hypothèse A₁ comprend un grand nombre de formes fonctionnelles, par exemple, on peut modéliser des fonctions exponentielles, en prenant les logarithmes de l'équation suivante :

$$y = e^{\beta_1} x_2^{\beta_2} x_3^{\beta_3} \dots x_k^{\beta_k}$$

2. A₂ a le rôle d'une condition d'identification, dans un modèle à deux dimensions, l'hypothèse indique que X n'est pas constante; il y a assez de variation dans le modèle.

3. A_3

$$E(e / X) = \begin{bmatrix} E(e_1 / X) \\ E(e_2 / X) \\ \vdots \\ E(e_n / X) \end{bmatrix}$$

L'hypothèse de la nullité de l'espérance conditionnelle implique aussi la nullité de l'espérance inconditionnelle puisque :

$$E(e_i) = E_X(E(e_i / X)) = E_X(0) = 0$$

A_3 implique aussi que :

$$E(x_{jl}e_i) = 0 \text{ Pour chaque } i, j = 1, \dots, n, l = 1, \dots, k$$

En effet :

$$E(e_i / x_{jl}) = E(E(e_i / X) / x_{jl}) = 0$$

$$E(x_{jl}e_i) = E(E(x_{jl}e_i) / x_{jl}) = E(x_{jl}E(e_i / x_{jl})) = 0,$$

C'est pourquoi cette hypothèse est aussi appelée condition d'exogénéité stricte. Elle nécessite que les régressions soient orthogonales au terme d'erreur, non seulement celui de la même observation $E(x_{jl}e_i) = 0$, $\forall l = (1, \dots, k)$ mais aussi à tous les termes d'erreurs des autres observations. A_3 implique aussi que :

$$E(y / X) = X\beta,$$

La régression de y sur X est la moyenne conditionnelle.

4. A_4 spécifie d'une manière plus complète la distribution du terme d'erreur.

La condition que $Var(e_i / X) = \sigma^2$ indique des erreurs homoscedastiques et la condition sur les covariances $Cov(e_i e_j / X) = 0$ indique qu'elles ne sont pas auto corrélées.

5. X est non stochastique dans un cadre expérimental, où l'on choisit des variables indépendantes x_i et l'on observe alors "la conséquence" y_i . On note qu'en pratique, cette hypothèse n'est pas très appropriée, et peut être relâchée à moindre coût.

2.2 L'estimation du modèle

A partir de A_1 , on sait que le vrai β existe et on veut l'estimer le mieux que possible.

On choisit un estimateur parmi la classe des estimateurs linéaires :

$$\tilde{\beta} = \underset{[k \times n]}{D} y + \underset{[k \times 1]}{d}.$$

Sous la condition que cet estimateur ne soit pas biaisé

$$E(\tilde{\beta}) = \beta, \quad \forall \beta \in R^k.$$

De ceux là, on choisit le meilleur estimateur, celui qui a une variance minimale.

$$Var(\tilde{\beta}) = E[(\tilde{\beta} - E(\tilde{\beta}))(\tilde{\beta} - E(\tilde{\beta}))'] \text{ est minimale}$$

Remarque

Le concept de plus petite variance est le suivant :

Soit Σ une (nxn)-matrice de variances-covariances, par conséquent elle est symétrique et définie positive et

$$\Sigma_1 \prec \Sigma_2 \Leftrightarrow x'(\Sigma_2 - \Sigma_1)x \geq 0, \quad \forall x \in R^n.$$

Lemme 1 [Weber, A (2003)]

Sous les hypothèses A_1, A_3, A_5 , l'estimateur linéaire $\tilde{\beta} = Dy + d$ est non biaisé si et seulement si $DX = I$ et $d = 0$

Lemme 2 [Weber, A (2003)]

Soit y une variable aléatoire de dimension n , pour laquelle le premier et le deuxième moment existe, et soit $z = Cy + c$ une variable aléatoire de dimension k alors :

$$Var(z) = CVar(y)C'$$

Une conséquence de ces lemmes on note :

$$Var(y) = Var(e).$$

$\tilde{\beta} = Dy$ est un estimateur linéaire sans biais, avec une matrice de variance covariance.

$$Var(\tilde{\beta}) = DVar(y)D' = DVar(e)D' = \sigma^2 DD'$$

Lemme 3 Weber, A (2003)

Soit $X^+ = (X'X)^{-1} X'$, X est de plein rang et soit $DX = L$, alors

$$DD' = (LX^+)(LX^+)' + (D-LX^+)(D-LX^+)'.$$

On applique ce lemme avec $L = I$ et on obtient :

$$DX = I$$

$$DD' = \underbrace{X^+(X^+)'}_{\text{Constante}} + \underbrace{(D-X^+)(D-X^+)'}_{= 0} \Leftrightarrow D = X^+$$

Avec ces étapes, on vient de prouver le théorème de Gauss Markov

Théorème 4 Théorème de Gauss-Markov [Weber, A (2003)]

Soit C un k vecteur colonne, sous les hypothèses A_1, A_2, A_3, A_4, A_5 , le meilleur estimateur linéaire sans biais (BLUE) de la fonction linéaire $C'\beta$ existe et est donné par $C'\hat{\beta}$ où $\hat{\beta}$ est l'unique estimateur des moindres carrés ordinaire de β .

Preuve

La preuve se déduit par les propriétés exigées de l'estimateur :

1. l'estimateur est linéaire donc il est de la forme $\lambda'Y$

2. l'estimateur linéaire de $C'\beta$ est sans biais alors :

$$E(\lambda'Y) = \lambda'E(Y) = \lambda'X\beta = C'\beta \quad , \forall \beta \Leftrightarrow \lambda'X = C'$$

3. l'estimateur de $C'\beta$ est de variance minimale $Var(\lambda'Y) = \lambda'Var(Y)\lambda$ et en utilisant la technique de Lagrange, on trouve $\lambda = X(X'X)^{-1}C$. L'estimateur

$$\lambda'Y = C'(X'X)^{-1}X'Y = C'\hat{\beta}.$$

Remarque

En général, parmi les estimateurs sans biais, il existe de meilleurs estimateurs non linéaires de β . On montrera que si les termes d'erreurs e suivent la loi normale $N(0, \sigma^2 I)$, $\hat{\beta}$ est le meilleur estimateur sans biais (BLUE) qui veut dire que β est aussi un estimateur efficace.

Estimation de la variance de $\hat{\beta}$

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

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

$$Var(\hat{\beta}) = E(\hat{\beta} - \beta)(\hat{\beta} - \beta)' = E(X'X)^{-1}X'ee'X(X'X)^{-1}$$

$$= (X'X)^{-1}X'X(X'X)^{-1}\sigma^2 = (X'X)^{-1}\sigma^2$$

On note que $Var(\hat{\beta})$ dépend de :

- σ^2 variance de l'erreur e_i
- $(X'X)^{-1}$, si $(X'X)^{-1}$ est presque singulière on parlera alors de problème de multi colinéarité (entre les colonnes de X).

Transformations linéaires de β

Corollaire 5 (corollaire du théorème de Gauss Markov) [Weber, A (2003)]

Sous les hypothèses A_1, A_2, A_3, A_4, A_5 l'estimateur $\hat{\gamma} = L\hat{\beta}$ (L est une $(s \times k)$ -matrice) est BLUE pour $\gamma = L\beta$.

Ce corollaire a d'importantes conséquences :

L'estimateur BLUE pour y est donné par :

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

De variance

$$\begin{aligned} Var(\hat{y}) &= Var(P_X y) = P_X' Var(y) P_X = P_X' Var(e) P_X = \sigma^2 P_X' P_X \\ &= \sigma^2 P_X. \end{aligned}$$

Estimation des erreurs e

Comme ci-dessus, on veut obtenir un estimateur BLUE:

- Estimateur linéaire $\tilde{e} = Cy + c$
- Estimateur sans biais c'est-à-dire $E(\tilde{e}) = E(e)$.

On définit l'erreur d'estimation δ comme : $\delta = e - \tilde{e}, E(\delta) = 0$

$$\begin{aligned} E(\delta) &= E(e - Cy - c) = E(e) - E(Cy) - E(c) \\ &= E(e) - E(C(X\beta + e)) - E(c) \\ &= (I - C)E(e) - E(CX\beta) - E(c) = -CX\beta - c = 0 \end{aligned}$$

Alors les conditions pour que δ soit sans biais sont

$$\begin{aligned} CX\beta &= -c \quad \forall \beta \in R^k \\ \Rightarrow CX &= 0 \quad \text{et} \quad c = 0 \end{aligned}$$

et le problème de minimisation à résoudre est :

$$\begin{aligned} \min \text{Var}(\delta) &= E(\delta\delta') \quad \text{tel que} \quad CX = 0 \\ E(\delta\delta') &= E((I - C)e - CX\beta)((I - C)e - CX\beta)' \\ &= E\left((I - C)ee'(I - C)'\right) = \sigma^2(I - C)(I - C)' \end{aligned}$$

En appliquant le lemme 3 avec $L = X$ on obtient

$$\begin{aligned} E(\delta\delta') &= \left(XX^+ + ((I - C) - XX^+)\right)\sigma^2 \\ (I - C) - XX^+ &= 0 \quad \text{si} \quad I - C = XX^+ \\ \text{c'est à dire} \quad C &= I - XX^+ = I - P_X \end{aligned}$$

Ainsi, on obtient un estimateur BLUE pour e :

$$\begin{aligned} \hat{e} &= (I - P_X)y = y - P_X y = y - X(X'X)^{-1}X'y \\ &= y - X\hat{\beta} \end{aligned}$$

avec

$$\begin{aligned} \text{Var}(\hat{e}) &= \sigma^2 M_X = \sigma^2 (I - XX^+) \\ E(\hat{e}) &= 0 \end{aligned}$$

Estimation de σ^2

Soit l'expression $\frac{1}{n}e'e$ et supposons que les e_i ont été observés, alors on obtient :

$$E\left(\frac{1}{n}e'e\right) = \frac{1}{n}E(e'e) = \frac{1}{n}E\left(\sum_{i=1}^n e_i^2\right) = \frac{1}{n}\sum_{i=1}^n E(e_i^2) = \sigma^2.$$

Ceci suggère qu'on peut prendre $\tilde{\sigma}^2 = \frac{1}{n}\hat{e}'\hat{e}$ comme un estimateur de σ^2 , mais par la construction des estimateurs des moindres carrés on a :

$$\hat{e}'\hat{e}/e'e$$

Ainsi $\hat{\sigma}^2 = \frac{1}{n} \hat{e}'\hat{e}$ est biaisé.

On peut obtenir un estimateur sans biais pour σ^2 par

$$\hat{\sigma}^2 = \frac{\hat{e}'\hat{e}}{n-k}$$

ce qui permet d'obtenir un estimateur pour la variance de $\hat{\beta}$

$$Var(\hat{\beta}) = \hat{\sigma}^2 (X'X)^{-1} = \frac{\hat{e}'\hat{e}}{n-k} (X'X)^{-1}$$

Remarques

- L'écart type $\hat{\sigma}$ est aussi appelé l'écart-type de la régression (*standard error*)
- La variance d'un seul paramètre $\hat{\beta}_j$ est donné par $\hat{\sigma}^2 (X'X)^{-1}_{jj}$

3. Régression stochastique

Il est parfois nécessaire de considérer que les variables explicatives sont aléatoires, une méthode pour obtenir les propriétés statistiques de $\hat{\beta}$ est de

- 1- Obtenir des résultats sur les propriétés statistiques conditionnées sur X (équivalent au cas où les régresseurs ne sont pas stochastiques).
- 2- Trouver des résultats inconditionnels en moyennant (intégrant sur) les lois conditionnelles.

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

Alors :

L'espérance conditionnelle $E(\hat{\beta} / X) = \beta + (X'X)^{-1} X'E(e / X) = \beta$

$$E(\hat{\beta}) = E_X(E(\hat{\beta} / X)) = E_X(\beta) = \beta$$

La variance conditionnelle $Var(\hat{\beta} / X) = (X'X)^{-1} \sigma^2$.

$$\text{Et } Var(\hat{\beta}) = E_X(Var(\hat{\beta} / X)) + Var_X(E(\hat{\beta} / X))$$

$$\text{On a } E(\hat{\beta} / X) = \beta \quad \text{et} \quad Var_X(\beta) = 0$$

$$\text{d'où } Var(\hat{\beta}) = E_X(\sigma^2 (X'X)^{-1}) = \sigma^2 E_X(X'X)^{-1}$$

avec le théorème de Gauss Markov précédent on a montré que :

$$Var(\hat{\beta} / X) \leq Var(\beta^* / X), \quad \forall \beta^* \neq \hat{\beta}.$$

Si cette inégalité est vraie pour chaque X , elle reste vraie pour les valeurs moyennes de X .

$$Var(\hat{\beta}) = E(Var(\hat{\beta} / X))$$

Théorème de Gauss Markov (suite)[Weber, A (2003)]

Dans un modèle classique de régression linéaire, l'estimateur des moindres carrés : $\hat{\beta} = (X'X)^{-1} X'y$ est un estimateur sans biais de variance minimale de β dans le cas où X est stochastique ou non.

4. Inférence statistique dans un modèle classique de régression linéaire

Après avoir estimé le modèle, il reste quand même des questions sans réponses comme :

- Quelles sont les variables qui expliquent la variabilité de y plus que les autres ?
- Quelle est la distribution des résidus ?

On a besoin de tester le modèle et de faire des hypothèses sur la distribution exacte des erreurs.

On fait alors l'hypothèse

$$A_6 \quad e/X \sim N(0, \sigma^2 I)$$

Cette hypothèse présente de nombreux avantages :

- Les transformations linéaires préservent la normalité.
- Les formes quadratiques des normales donnent des variables aléatoires distribuées selon χ^2 Chi-deux et F Fisher-Snedecor.
- Si e suit la normale, $\hat{\beta}$ l'estimateur des moindres carrés ordinaires est aussi l'estimateur du maximum de vraisemblance.

Définition

- Une variable aléatoire e de dimension n , est normalement distribué si sa fonction de densité a la forme suivante :

$$f(e/\mu, \Sigma) = \frac{1}{(\sqrt{2\pi})^n \sqrt{\det(\Sigma)}} \exp\left\{-\frac{1}{2}(e-\mu)' \Sigma^{-1}(e-\mu)\right\}$$

Avec $\mu \in R^n$ et Σ une matrice symétrique définie positive,

$$E(e) = \mu$$

$$Var(e) = \Sigma$$

Corollaire 7 [Weber, A (2003)]

Soit x une variable aléatoire $n \times 1$ avec $x \sim N(\mu, \Sigma)$; Si $y = Cx + c$, où C est une matrice $(s \times n)$ et $\text{rang}(c) = s$ et $1 \leq s \leq n$ alors :

$$y \sim N(C\mu + c, C\Sigma C')$$

En appliquant ce corollaire au modèle de régression linéaire on obtient :

$$e/X \sim N(0, \sigma^2 I),$$

$$y/X \sim N(X\beta, \sigma^2 I),$$

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

4.1. Le principe du maximum de vraisemblance

Supposons y_1, \dots, y_n un échantillon d'observations donné, de fonction de densité $f_y(y_1, \dots, y_n / \theta)$. On veut estimer le vecteur paramètre θ dépendant des observations noté $\hat{\theta}(y_1, \dots, y_n)$.

On définit la fonction de vraisemblance comme :

$$L(\theta / y_1, \dots, y_n) = f_y(y_1, \dots, y_n / \theta).$$

Le principe de maximum de vraisemblance déclare qu'un estimateur de θ désigné par $\hat{\theta}(y_1, \dots, y_n)$ est donné par le maximum de la fonction de vraisemblance $L(\theta / y_1, \dots, y_n)$.

i.e. $\hat{\theta}(y_1, \dots, y_n)$ est tel que $L(\hat{\theta} / y_1, \dots, y_n) > L(\theta / y_1, \dots, y_n) \quad \forall \theta \neq \hat{\theta}$

Dans notre cas, on veut obtenir un estimateur du maximum de vraisemblance (EMV) pour le modèle de régression linéaire (1.3).

Écrivons d'abord la fonction de vraisemblance pour y/X

$$f_y(y_1, \dots, y_n / X\beta, \sigma^2 I) = \frac{1}{(\sqrt{2\pi\sigma^2})^{n/2}} \exp\left\{-\frac{1}{2\sigma^2} (y - X\beta)' (y - X\beta)\right\}$$

$$= L(\beta, \sigma^2 / y_1, \dots, y_n)$$

En prenant les logarithmes :

$$\ln L = \frac{-n}{2} (\ln 2\pi + \ln \sigma^2) - \frac{1}{2\sigma^2} (y - X\beta)' (y - X\beta)$$

On calcule les dérivées par rapport aux paramètres :

$$\frac{\partial \ln L}{\partial \beta} = \frac{-1}{2\sigma^2} 2X'(y - X\beta)$$

$$\frac{\partial \ln L}{\partial \sigma^2} = \frac{-n}{2\sigma^2} + \frac{1}{2\sigma^4} (y - X\beta)' (y - X\beta)$$

En annulant ces deux dérivées

$$\frac{\partial \ln L}{\partial \beta} = 0 \quad \frac{\partial \ln L}{\partial \sigma^2} = 0$$

On obtient les EMV de β et σ^2 suivants :

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

$$\hat{\sigma}_{MV}^2 = \frac{1}{n} (y - X\hat{\beta})' (y - X\hat{\beta}) = \frac{1}{n} \hat{e}'\hat{e}.$$

L'estimateur du maximum de vraisemblance de β est égal à l'estimateur de MCO et l'estimateur du maximum de vraisemblance de σ^2 est donné par l'estimateur biaisé de la variance.

Théorème 8 [Weber, A (2003)]

Dans le modèle classique de régression linéaire et sous l'hypothèse de normalité des erreurs l'estimateur des moindres carrés ordinaire $\hat{\beta}$ est de variance minimale sur l'ensemble de tous les estimateurs sans biais.

Remarque

Pour des erreurs qui ne suivent pas la loi normale, l'estimateur du maximum de vraisemblance $\hat{\beta}_{MV}$ a souvent une variance plus petite que celle de l'estimateur des moindres carrés ordinaire $\hat{\beta}$. Ainsi, dans le cas où la normalité n'est pas satisfaite, il est préférable d'utiliser l'estimateur du maximum de vraisemblance que l'estimateur des moindres carrés.

4.2. Les tests d'hypothèses

Soit Θ un espace de paramètre.

L'hypothèse nulle : $H_0 \quad \theta \in \Theta_0 \quad \Theta_0 \subset \Theta$

L'alternative : $H_1 \quad \theta \in \Theta_1 \quad \Theta_1 \subset \Theta \setminus \Theta_0$

Exemple

$$\Theta = R^k$$

$$\Theta_0 = R^{k-1} \times \{\theta^*\}$$

Un test statistique est une règle de décision basée sur un échantillon. La règle de décision détermine si H_0 est acceptée ou rejetée.

Pour une procédure de test, les résultats possibles sont les suivants :

	Accepter H_0	Rejeter H_0
H_0 vrai	Ok	Erreur de type 1
H_0 fausse	Erreur de type 2	Ok

α = la probabilité (rejeter H_0/H_0 est vrai) = Probabilité (erreur de type 1)

α est appelée la taille du test

β = Probabilité (erreur de type 2)

$1 - \beta$ est appelée la puissance du test

Les tests sont comparés en termes de taille et de puissance.

Les principes de test d'hypothèse

On présente trois principes de test basés sur l'estimation par le maximum de vraisemblance d'un paramètre θ . Soit $c(\cdot)$ une fonction, le test d'hypothèse est le suivant :

$$H_0 : c(\theta) = 0$$

Test du rapport de vraisemblance

Si la restriction $c(\theta) = 0$ est valide. L'imposer ne va pas simplifier la fonction de vraisemblance. Par conséquent, le test est basé sur la différence $\ln L - \ln L_R$, où L est la vraisemblance pour une valeur de θ non restreinte et L_R est la valeur de la vraisemblance quand l'estimateur est restreint. Soit $\hat{\theta}_{MV}$ l'estimateur obtenu du maximum de vraisemblance de θ et $\hat{\theta}_R$ l'estimateur obtenu en maximisant la fonction de vraisemblance sous la contrainte $c(\theta) = 0$

Si \hat{L}_{MV} et \hat{L}_R sont les valeurs de la fonction de vraisemblance calculées en ces deux estimateurs, alors le rapport de vraisemblance est :

$$\lambda = \frac{\hat{L}_R}{\hat{L}_{MV}}$$

Cette fonction doit être comprise entre 0 et 1. Si λ est trop petite on rejette l'hypothèse nulle.

Test de WALD

Si la restriction $c(\theta) = 0$ est valide, $c(\hat{\theta}_{MV})$ doit être très proche de zéro, car l'estimateur du maximum de vraisemblance est convergent sous certaines hypothèses. Par conséquent, le test est basé sur $c(\hat{\theta}_{MV})$, on rejette l'hypothèse nulle si elle est significativement différente de zéro.

Test de multiplicateur de Lagrange

Si la restriction $c(\theta) = 0$ est vraie, l'estimateur $\hat{\theta}_R$ doit être proche du point qui maximise le log de la vraisemblance. Aussi la pente de la fonction de vraisemblance doit être proche de zéro à l'estimateur restreint. Le test est basé sur la pente du log de la vraisemblance au point où la fonction est maximisée sous la restriction. La dérivée de la vraisemblance par rapport au paramètre est appelée le score.

$$s(\theta) = \frac{\partial \ln L(\theta)}{\partial \theta}$$

Le test est basé sur $s(\hat{\theta}_R)$ et l'hypothèse nulle est rejetée s'il est significativement différent de zéro.

Ces trois tests ont un comportement asymptotique équivalent, mais différent dans le cas des petits échantillons.

Parmi les trois principes, le choix est fait selon les considérations de calcul.

Test de Wald : nécessite l'estimation du modèle non restreint.

Test de multiplicateur de Lagrange : nécessite l'estimation du modèle restreint .

Test du rapport de vraisemblance : nécessite l'estimation des deux modèles.

Test d'hypothèse concernant un seul paramètre β_j

$$H_0 : \beta_j = \beta_j^*$$

Pour obtenir une règle de décision, on doit trouver le test statistique qui convient et dont on connaît la loi. Rappelons que sous l'hypothèse A_6 .

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

$$z_j = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\sigma^2 (X'X)^{-1}_{jj}}} \sim N(0,1).$$

Comme on ne connaît pas la valeur de σ^2 , on la remplace par son estimateur $\hat{\sigma}^2$.

$$t_j = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 (X'X)^{-1}_{jj}}} = \frac{\hat{\beta}_j - \beta_j}{Se(\hat{\beta}_j)},$$

suit la loi de Student avec $n - k$ degrés de liberté notée t_{n-k} .

Soit le test d'hypothèse

$$H_0 : \beta_j = 0$$

Cette hypothèse indique que la variable x_j n'a pas un effet sur y après avoir considéré les variables $x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_k$ dans le modèles.

Si l'hypothèse nulle n'est pas rejetée, x_j peut être éliminée de l'équation de régression.

Dans ce cas, la t-statistique est :

$$t_{\hat{\beta}_j} = \frac{\hat{\beta}_j}{Se(\hat{\beta}_j)}$$

$t_{\hat{\beta}_j}$ est petit si $\hat{\beta}_j$ est presque nul ou $Se(\hat{\beta}_j)$ est grand.

Il existe deux possibilités pour définir l'hypothèse alternative. On commence par :

$$H_0 : \beta_j = 0,$$

$$H_1 : \beta_j > 0 .$$

Et on choisit le niveau de signification ou une taille $\alpha = 0.05$; pour construire une règle de décision, on doit trouver une valeur de $t_{\hat{\beta}_j}$ suffisamment grande dans le but de rejeter H_0 :

$\beta_j = 0$, on rejette l'hypothèse nulle si : avec $c = 95$ percentile de la loi t_{n-k} .

Selon cette décision, parmi 100 échantillons aléatoires dans lesquels H_0 est vrai, on rejettera cette dernière dans 5 de ces échantillons (erreur de type 1).

La 2^{ème} possibilité est de définir une hypothèse alternative :

$$H_0 : \beta_j = 0 ,$$

$$H_1 : \beta_j \neq 0 .$$

Dans ce cas on rejette H_0 si $\left| t_{\hat{\beta}_j} \right| > c$.

Test des valeurs du paramètre β

Définition (la loi χ^2)

Soit $x = (x_1, \dots, x_n)$ un vecteur aléatoire de dimension n , tel que $x \sim N(0, I)$, alors :

$Q = x'x = \sum_{i=1}^n x_i^2$ suit une loi de fonction de densité :

$$f(Q, n) = \begin{cases} \frac{1}{c(n)} Q^{\frac{n}{2}-1} e^{-\frac{1}{2}Q} , & \text{si } Q \geq 0 \\ 0 & , \quad \text{sinon} \end{cases}$$

$$\text{où : } c(n) = 2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right) ,$$

et on dit que Q suit une loi χ^2 centrée avec n degrés de liberté.

Soit $x \sim N(\mu, I)$, alors $Q = x'x$ suit un loi χ^2 non centrée, avec n degré de liberté et de moyenne $\lambda = \mu'\mu$.

Remarque

Si $x \sim N(\mu, \Sigma)$, $X^2 = (x - \mu)' \Sigma^{-1} (x - \mu)$ suit une loi χ^2 centrée avec n degrés de liberté et

$X^2 = x' \Sigma^{-1} x$ suit la loi χ^2 non centrée de paramètres n et $\lambda = \mu' \Sigma^{-1} \mu$.

Définition (Loi Fisher-Snedecor)

Soit χ_1^2 et χ_2^2 deux variables aléatoires indépendantes suivant une χ^2 à m et n degrés de liberté respectivement, la variable aléatoire :

$$F = \frac{\chi_1^2/m}{\chi_2^2/n}$$

suit une loi F centrée, sa fonction de densité est donnée par :

$$f(F, m, n) = \begin{cases} c(m, n) \frac{F^{\frac{m}{2}-1}}{\left(1 - \frac{m}{n}\right) F^{\frac{m+n}{2}}}, & \text{si } F \geq 0 \\ 0, & \text{sinon} \end{cases}$$

$$\text{où } c(m, n) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{m/2}.$$

Soit χ_1^2 et χ_2^2 deux variables aléatoires indépendantes, χ_1^2 suit une χ^2 non centrée, χ_2^2 une χ^2 centrée, alors le rapport des deux suit la loi F non centrée.

Définition (Loi de Student)

La loi de Student t est un cas particulier de la loi F, si $m=1$ et la variable aléatoire F suit une Fisher centrée alors la variable aléatoire $t = \sqrt{F}$ suit une loi de Student centrée à n degré de libertés si F suit une Fisher non centrée, t suit une Student non centrée.

Exemple

Le rapport d'une variable aléatoire normale standard et la racine carrée d'une variable aléatoire indépendante χ^2 est une variable aléatoire qui suit la loi t de Student.

Distributions de quelques formes quadratiques

Théorème 9 [Weber, A (2003)]

Soit x un vecteur aléatoire de dimension n , tel que $x \sim N(\mu, \sigma^2 I)$ alors $Q = \frac{x'Ax}{\sigma^2}$ suit une loi

$\chi_{r,\lambda}^2$, $\lambda = \frac{\mu' A \mu}{\sigma^2}$ si et seulement si A est idempotente, symétrique et de rang r .

Théorème 10 [Weber, A (2003)]

Soit A une matrice idempotente et de rang r , soit B une matrice $(m \times n)$ telle que $BA = 0$, et soit $x \sim N(\mu, \sigma^2 I)$ un vecteur aléatoire $(n \times 1)$, alors le vecteur $z = Bx$ et la variable aléatoire $Q = x'Ax$ sont indépendamment distribués.

Exemple

Dans un modèle de régression classique, où les erreurs suivent une loi normale, $\hat{\beta}$ et $\hat{\sigma}^2$ sont indépendants.

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

$$\hat{\sigma}^2 = \frac{1}{n-k} \hat{e}'\hat{e} = \frac{1}{n-k} y'(I - X(X'X)^{-1} X')y,$$

$$y \sim N(X\beta, \sigma^2 I).$$

Dans ce cas : $A = I - X(X'X)^{-1} X'$ et $B = (X'X)^{-1} X'$.

$$BA = (X'X)^{-1} X'(I - X(X'X)^{-1} X') = 0.$$

Théorème 11 [Weber, A (2003)]

Soit A une matrice $(n \times n)$ symétrique et idempotente de rang r , soit B une matrice $(n \times n)$ symétrique et $x \sim N(\mu, \sigma^2 I)$

Si $BA = 0$ alors les formes quadratiques $Q_1 = x'Ax$ et $Q_2 = x'Bx$ sont indépendantes.

Théorème 12 [Weber, A (2003)]

Sous les hypothèses A_1, A_2, A_3, A_4, A_6 la loi de la forme quadratique $\hat{e}'\hat{e} / \sigma^2$ est une χ_{n-k}^2 à $n-k$ degrés de libertés.

Preuve On a

$$\hat{e} = (I - X(X'X)^{-1} X')y = M_X y. \text{ Donc}$$

$$\frac{\hat{e}'\hat{e}}{\sigma^2} = \frac{y'M_X y}{\sigma^2},$$

où M_X est symétrique, idempotente et de rang $n-k$; d'après le théorème 8, $\frac{\hat{e}'\hat{e}}{\sigma^2} \sim \chi_{n-k}^2$

c.q.f.d

Théorème 13 [Weber, A (2003)]

Soit $\beta^* \in R^k$ fixe; sous les hypothèses A_1, A_2, A_3, A_4, A_6 la forme quadratique :

$$(\hat{\beta} - \beta^*)X'X(\hat{\beta} - \beta^*) \frac{1}{\sigma^2} \sim \chi_{k,\lambda}^2$$

$$\text{avec : } \lambda = (\beta - \beta^*)X'X(\beta - \beta^*) \frac{1}{\sigma^2}$$

est indépendante de $\frac{\hat{e}'\hat{e}}{\sigma^2}$.

Preuve

$$\begin{aligned}\hat{\beta} &= \beta + (X'X)^{-1} X'e \\ \hat{\beta} - \beta^* &= (X'X)^{-1} X'e + (\beta - \beta^*) = (X'X)^{-1} X'(e + X(\beta - \beta^*)) \\ (\hat{\beta} - \beta^*)' X'X(\hat{\beta} - \beta^*) &= (e + X(\beta - \beta^*))' X^+ X'X X^+ (e + X(\beta - \beta^*)) \text{ où } X^+ = (X'X)^{-1} X' \\ &= (e + X(\beta - \beta^*))' (X X^+)' (X X^+) (e + X(\beta - \beta^*)), \\ &= (e + X(\beta - \beta^*))' P_X (e + X(\beta - \beta^*)), \quad P_X = (X X^+)' \end{aligned}$$

On a $e + X(\beta - \beta^*) \sim N(X(\beta - \beta^*), \sigma^2 I)$

Comme P_X est idempotente, symétrique ainsi la distribution χ^2 résulte du théorème 8. aussi

$M_X P_X = 0$, des théorèmes 10 et 11 on déduit l'indépendance. c.q.f.d

Les théorèmes précédents constituent les outils nécessaires pour produire les tests statistiques.

Notons d'abord une équation qui nous sera utile, et qui peut facilement être vérifiée.

$$(y - X\hat{\beta})'(y - X\hat{\beta}) = (y - X\hat{\beta})'(y - X\hat{\beta}) + (\hat{\beta} - \beta^*)' X'X(\hat{\beta} - \beta^*) \quad (1.4)$$

L'hypothèse I $H_0 : \beta = \beta^*$

La première hypothèse est faite sur tout le vecteur de paramètre, on obtient le F test statistique à partir du test du rapport de vraisemblance. Rappelons que ce dernier est donné par :

$$RV = \frac{\text{Max}_{\theta \in \Theta_0} L(\theta)}{\text{Max}_{\theta \in \Theta} L(\theta)},$$

à partir de :

$$\ln L = \frac{-n}{2} \ln (2\pi \sigma^2) - \frac{1}{2\sigma^2} (y - X\beta)'(y - X\beta),$$

nous avons obtenu les EMV des paramètres β et σ^2

$$\hat{\theta}_{MV} = \begin{pmatrix} \hat{\beta} \\ \hat{\sigma}_{MV}^2 \end{pmatrix} = \begin{pmatrix} (X'X)^{-1} X'y \\ \frac{1}{n} y' M_X y \end{pmatrix}.$$

Dans le modèle restreint, on a $\hat{\beta} = \beta^*$ et $\hat{\sigma}^2 = \frac{1}{n} (y - X\hat{\beta})'(y - X\hat{\beta})$

Les maximums des fonctions de vraisemblance dans le modèle restreint et le modèle non restreint sont :

$$\underset{\theta \in \Theta_0}{\text{Max}} \ln L(\theta) = \frac{-n}{2} \ln \left(2\pi \frac{1}{n} (y - X\beta^*)' (y - X\beta^*) - \frac{1}{\frac{2}{n} (y - X\beta^*)' (y - X\beta^*)} (y - X\beta^*)' (y - X\beta^*) \right)$$

$$\underset{\theta \in \Theta}{\text{Max}} \ln L(\theta) = \frac{-n}{2} \ln \left(2\pi \frac{1}{n} (y - X\hat{\beta})' (y - X\hat{\beta}) \right) - \frac{n}{2}$$

On obtient donc le test statistique

$$RV = \frac{\left(\frac{1}{n} (y - X\hat{\beta})' (y - X\hat{\beta}) \right)^{n/2}}{\left(\frac{1}{n} (y - X\beta^*)' (y - X\beta^*) \right)^{n/2}} = \left(\frac{1}{1 + F \frac{k}{n-k}} \right)^{n/2}. \quad (1.5)$$

La dernière égalité est obtenue en utilisant l'équation (1.4) et

$$F = \frac{(\hat{\beta} - \beta^*)' X'X (\hat{\beta} - \beta^*)}{\hat{e}'\hat{e}} \frac{n-k}{k}.$$

Sous l'hypothèse H_0 , F suit la distribution F centrée de paramètres $(k, n-k)$.

Sous l'hypothèse H_1 , F suit la distribution F non centrée de paramètres $(k, n-k, \lambda)$,

et $\lambda = (\beta - \beta^*)' X'X (\beta - \beta^*) / \sigma^2$.

Hypothèse II $H_0 : \beta_2 = \beta_2^*$

Cette hypothèse est faite sur une partie du vecteur paramètre, on partitionne β en deux parties selon l'hypothèse :

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

où β_1 est un vecteur $((k-s) \times 1)$ et β_2 un vecteur $(s \times 1)$, ceci veut dire qu'on va tester s restrictions.

$$H_0 : \beta_2 = \beta_2^*$$

Le modèle restreint est donné par :

$$y = (X_1, X_2) \begin{pmatrix} \beta_1 \\ \beta_2^* \end{pmatrix} + e,$$

$$z = y - X_2\beta_2^* = X_1\beta_1 + e,$$

$$\hat{\beta}_1 = (X_1'X_1)^{-1} X_1'(y - X_2\beta_2^*).$$

On définit M_{X_1} par :

$$M_{X_1} = I - X_1(X_1'X_1)^{-1} X_1',$$

Et $H = X_2' M_{X_1} X_2$.

De cette régression on peut exprimer les résidus par :

$$\hat{e} = M_{X_1} z = z - X_1 \hat{\beta}_1$$

On note une équation auxiliaire analogue à l'équation (1.4)

$$(y - X_2 \beta_2^*)' M_{X_1} (y - X_2 \beta_2^*) = (y - X \hat{\beta})' (y - X \hat{\beta}) + (\hat{\beta}_2 - \beta_2^*)' H (\hat{\beta}_2 - \beta_2^*) \quad (1.6)$$

On obtient :

$$RV = \frac{\text{Max}_{\theta \in \Theta_0} L(\theta)}{\text{Max}_{\theta \in \Theta} L(\theta)} = \frac{\left[\frac{1}{n} (y - X \hat{\beta})' (y - X \hat{\beta}) \right]^{n/2}}{\left[\frac{1}{n} (z - X_1 \hat{\beta}_1)' (z - X_1 \hat{\beta}_1) \right]^{n/2}} = \left[\frac{\hat{e}' \hat{e}}{(y - X_2 \beta_2^*)' M_{X_1} (y - X_2 \beta_2^*)} \right]^{n/2}.$$

En appliquant l'équation (1.6)

$$RV = \left[\frac{1}{1 + F \frac{s}{n-k}} \right]^{n/2},$$

$$\text{Avec } F = \frac{(\hat{\beta}_2 - \beta_2^*)' H (\hat{\beta}_2 - \beta_2^*) (n-k)}{\hat{e}' \hat{e} s}.$$

Sous l'hypothèse nulle, F suit une distribution F centrée de paramètres $(s, n-k)$. Sous H_1 elle suit une distribution F centrée de paramètres $(s, n-k)$ et $\lambda = (\hat{\beta}_2 - \beta_2^*)' H (\hat{\beta}_2 - \beta_2^*) / \sigma^2$.

L'hypothèse III $H_0 : \beta_k = \beta_k^*$

Cette troisième hypothèse concerne un seul élément du vecteur paramètre, c'est un cas particulier de la deuxième hypothèse avec $s = 1$

$$H_0 : \beta_k = \beta_k^*$$

$$H_1 : \beta_k \neq \beta_k^*$$

Comme c'est un cas particulier de l'hypothèse II avec $s = 1$

$$H = X_k' M_{X_1} X_k = \frac{n}{\frac{1}{n} (X'X)_{kk}^{-1}}.$$

Ainsi

$$F = \frac{(\hat{\beta}_k - \beta_k^*)' (\hat{\beta}_k - \beta_k^*)}{(X'X)_{kk}^{-1} \hat{e}'\hat{e}} \frac{n-k}{k-(k-1)} = \frac{(\hat{\beta}_k - \beta_k^*)^2}{(X'X)_{kk}^{-1} \hat{e}'\hat{e}} (n-k).$$

On peut obtenir le test de Student t par :

$$t = \sqrt{F} = \frac{\hat{\beta}_k - \beta_k^*}{\sqrt{\widehat{Var}(\hat{\beta}_k)}} \sim t_{n-k}.$$

Sous H_0 , t a une distribution t_{n-k} centrée, sous H_1 t suit la distribution t_{n-k} non centrée.

4.3 Intervalles de confiance

Pour un paramètre β_k , on construit l'intervalle de confiance, en utilisant la statistique t .

$$t = \frac{\hat{\beta}_k - \beta_k}{\sqrt{\widehat{Var}(\hat{\beta}_k)}} = \frac{\hat{\beta}_k - \beta_k}{Se(\hat{\beta}_k)} \sim t_{n-k}.$$

un intervalle de confiance pour β_k est donné par :

$$P \left[\hat{\beta}_k - t_{\frac{\alpha}{2}} Se(\hat{\beta}_k) < \beta_k < \hat{\beta}_k + t_{\frac{\alpha}{2}} Se(\hat{\beta}_k) \right] = 1 - \alpha.$$

Exemple Pour $\alpha = 0.05$, l'intervalle de confiance est :

$$[\hat{\beta}_k - c.Se(\hat{\beta}_k), \hat{\beta}_k + c.Se(\hat{\beta}_k)].$$

où c est 97,5 le quantile de la distribution t_{n-k}

Pour le paramètre σ^2 , on rappelle que :

$$\frac{\hat{e}'\hat{e}}{\sigma^2} \sim \chi_{n-k}^2,$$

$$\hat{\sigma}^2 = \frac{1}{n-k} \hat{e}'\hat{e}.$$

On construit l'intervalle de confiance par :

$$P \left[\frac{(n-k)\hat{\sigma}^2}{\chi_{1-\frac{\alpha}{2}}^2} \leq \sigma^2 \leq \frac{(n-k)\hat{\sigma}^2}{\chi_{\frac{\alpha}{2}}^2} \right] = 1 - \alpha.$$

5- Modèle de régression linéaire généralisé

5-1 Estimation par la méthode des moindres carrés généralisée

Dans notre système d'hypothèses on modifie A_4 :

$$A_1 : y = X\beta + e ,$$

$$A_2 : \dim(X) = (n \times k); \quad \text{rang}(X) = k ,$$

$$A_3 : E(e/X) = 0 ,$$

$$A_4^* : E(ee'/X) = \sigma^2 \Omega \quad \Omega = \Omega' \text{ et } \det(\Omega) > 0 ,$$

$A_5 : X$ est non stochastique ;

Si Ω est inconnue, on a $\frac{n(n+1)}{2}$ autres paramètres dans notre modèle. Il y aura plus de paramètres à estimer que d'observations.

On suppose d'abord que Ω est connue et que seulement σ^2 est inconnue. Comme condition de normalisation, on prend $\text{tr}(\Omega) = n$.

La matrice symétrique définie positive Ω et son inverse peuvent être décomposées selon la décomposition de Cholesky en :

$$\Omega = RR' ,$$

$$\Omega^{-1} = P'P ,$$

$$\text{avec } (PR)'(PR) = I \quad \text{et} \quad (PR)'(PR) = I .$$

En multipliant le modèle de régression par la matrice P

$$Py = PX\beta + Pe .$$

et on obtient un nouveau modèle transformé

$$y^* = X^*\beta + e^* . \tag{1.7}$$

Ce modèle a les propriétés suivantes :

- 1- X^* n'est pas stochastique,
- 2- $X^{*'}X^* = X'P'PX$ est non singulière,
- 3- $E(e^*) = E(Pe) = 0$,
- 4- $E(e^*e^{*'}) = E(Pee'P') = \sigma^2 P\Omega P' = \sigma^2 I$.

C'est-à-dire que dans le modèle transformé (1.7), les hypothèses de la régression linéaire classique sont vérifiées, on peut appliquer l'estimateur des moindres carrés sur ce modèle et obtenir le résultat suivant :

Théorème 14 [Weber, A (2003)]

Sous les hypothèses A_1, A_2, A_3, A_4^* et A_5 l'estimateur BLUE de β est donné par :

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

Preuve

$$(X^{*'}X^*)^{-1}X^{*'}y^* = (X'P'PX)^{-1}X'P'Py = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y.$$

$\tilde{\beta}$ est appelé l'estimateur des Moindres Carrés Généralisés (MCG). c.q.f.d

Matrice de covariance de l'estimateur MCG

$$\begin{aligned} Var(\tilde{\beta}) &= E(\beta - \tilde{\beta})(\beta - \tilde{\beta})' \\ &= E(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}ee'\Omega^{-1}X(X'\Omega^{-1}X)^{-1} \\ &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}E(ee')\Omega^{-1}X(X'\Omega^{-1}X)^{-1} \\ &= \sigma^2(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\Omega\Omega^{-1}X(X'\Omega^{-1}X)^{-1} \\ &= \sigma^2(X'\Omega^{-1}X)^{-1} \end{aligned}$$

$$\tilde{e} = y - X\tilde{\beta} \quad \text{et} \quad \tilde{\sigma}^2 = \frac{1}{n-2} \tilde{e}'\Omega^{-1}\tilde{e}$$

5-2 Propriétés des estimateurs MCO dans le modèle de régression généralisé

- 1- L'estimateur MCO est sans biais.
- 2- En général $\hat{\beta}$ est moins efficace que $\tilde{\beta}$.
- 3- L'estimateur MCO de σ^2 est biaisé.

5-3 Propriétés asymptotiques du MCG

Théorème 14 [Weber, A (2003)]

Sous les hypothèses A_1, A_2, A_3, A_4^* et A_5 et

$$\lim_{n \rightarrow \infty} \frac{1}{n} X'\Omega^{-1}X = Q$$

si Q est définie positive, alors l'estimateur MCG $\tilde{\beta}$ est convergent.

Preuve

$$\tilde{\beta} = \beta + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}e$$

$$E(\tilde{\beta}) = \beta$$

$$Var(\tilde{\beta}) = E(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)' = \sigma^2(X'\Omega^{-1}X)^{-1}$$

$$= \frac{\sigma^2}{n} \left(\frac{1}{n} X'\Omega^{-1}X \right)^{-1} \rightarrow 0$$

Ainsi $\tilde{\beta} \xrightarrow{M} \beta$ (converge en moyenne) ce qui implique que $\tilde{\beta} \xrightarrow{P} \beta$.
c.q.f.d.

Remarque

L'hypothèse $\lim_{n \rightarrow \infty} \frac{1}{n} X' \Omega^{-1} X = Q$ avec Q définie positive implique que

$$\lim_{n \rightarrow \infty} \left(\frac{1}{n} X' \Omega^{-1} X \right)^{-1} = Q^{-1}.$$

Pour $\Omega = I, k = 1$ on peut donner l'interprétation de cette hypothèse.

$$\lim_{n \rightarrow +\infty} \left(\frac{1}{n} \sum_{i=1}^n x_i^2 \right)^{-1} = Q^{-1}$$

Avec Q définie positive, c'est-à-dire $\frac{1}{n} \sum x_i^2$ ne doit pas converger vers 0. Il doit y avoir assez de variation dans les données. Plus la taille de l'échantillon est grande plus la variation augmente.

Chapitre 2

MODELE MULTI-NIVEAUX DE REGRESSION LINEAIRE

Introduction Goldstein, H(1996)

L'objet de ce chapitre est d'introduire le modèle hiérarchique de régression linéaire et les notations de base que nous utiliserons par la suite.

Dans un modèle hiérarchique de régression linéaire, les observations sont groupées par unités, chaque ensemble d'unités correspond à un niveau. Ainsi dans un modèle à deux niveaux, on note les observations y_{ij} telles que y_{ij} est l'observation qui correspond à la i ème unité de premier niveau appartenant à la j ème unité de deuxième niveau.

Le système éducatif présente un exemple évident de la structure hiérarchique, avec les élèves groupés par école, les écoles groupées par région etc. ..., y_{ij} peut être par exemple la note de lecture du i ème élève de la j ème école. L'ensemble des élèves correspond à l'ensemble des unités de niveau 1, de même l'ensemble des écoles correspond à l'ensemble des unités de niveau 2.

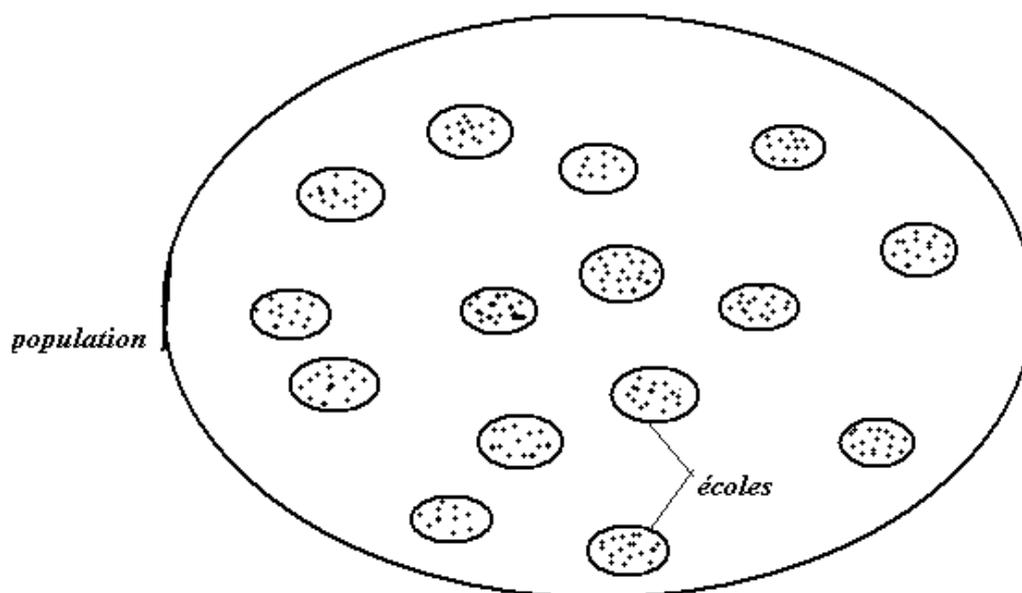


Fig-2.1 Représentation d'un modèle à deux niveaux

En modélisation multi-niveaux ou multi-échelles, le modèle statistique utilisé doit explicitement reconnaître la structure hiérarchique des données, l'utilisation d'un modèle qui néglige cette hiérarchie ne sera pas sans conséquences.

Gardons toujours le système éducatif pour discuter le cas où on ignore la structure hiérarchique des données et aussi pour illustrer des notions élémentaires des modèles multi-niveaux; dans ces deux objectifs nous nous basons sur un cas réel étudié à **English Local Education Authorities** qui est le suivant: on s'intéresse à la relation entre la variable dépendante *notes d'examens* des élèves de 16ans et une variable indépendante (explicative) : *notes de lecture* de ces mêmes élèves quand ils avaient 11ans (c'est-à-dire juste avant qu'ils n'entrent en école secondaire). On fera référence à la variable explicative par "*notes LRT*" où *LRT* est la diminution de London Reading Test. Par le passé, il a été question de décider si l'analyse sera réalisée au niveau des écoles ou au niveau des élèves, mais dans les deux cas proposés (à un niveau), les résultats d'analyse ne sont pas satisfaisants. Au niveau école, la moyenne des notes d'examens et celle des notes LRT sont d'abord calculées pour chaque école, une régression ordinaire est alors utilisée pour estimer la relation entre elles, le problème majeur dans ce cas est qu'on ne peut pas interpréter clairement la relation trouvée, une interprétation causale doit inclure les élèves d'une manière individuelle. Au niveau élève la relation entre les notes a été estimée en utilisant les données de tous les 4059 élèves. Dans ce cas, on a la possibilité de modéliser la variation entre

l'ensemble des écoles en incorporant des termes séparés pour l'ensemble des élèves de chaque école mais cette procédure reste inefficace du moment qu'elle nécessite l'estimation d'un grand nombre de coefficients et aussi qu'elle ne considère pas les écoles comme un échantillon aléatoire.

1. Principe fondamental de la modélisation multi-niveaux

On utilisera l'exemple précédent pour illustrer le Principe fondamental de la modélisation multi-niveau: existence de plusieurs niveaux de variation. Dans ce but, on s'intéresse aux écoles.

On commence par observer les données d'une seule école; dans la figure 2.2 les notes d'examen d'un échantillon de 73 élèves d'une même école sont contre les notes LRT de ces mêmes élèves; la relation se résume par une droite de régression.

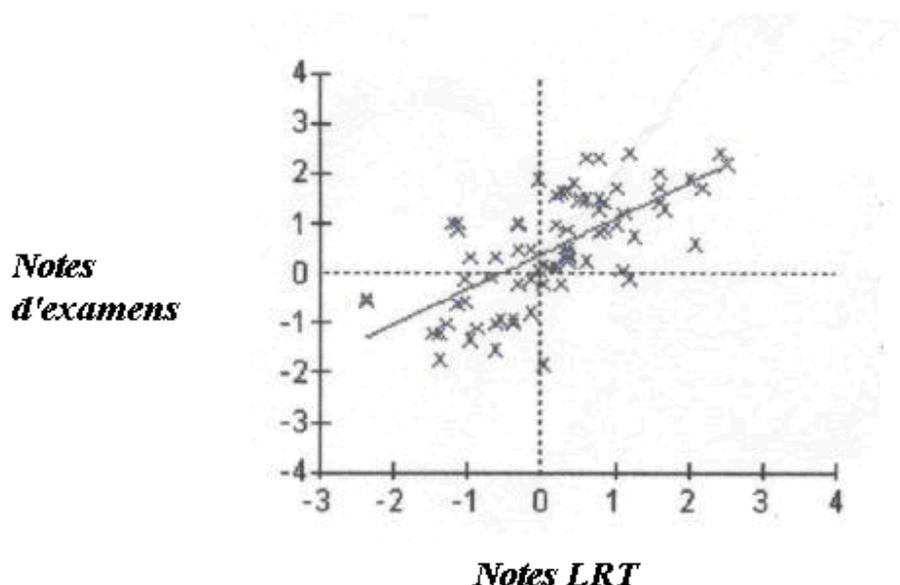


Fig2.2 Variation au niveau 1

Chaque point de ce graphe représente une paire de valeurs pour un élève; la droite de régression est un résumé; au niveau école de cette relation; qui peut varier d'une école à une autre. Dans la figure 2.3, quatre droites de régression sont tracées pour quatre écoles en contraignant les lignes à être parallèles

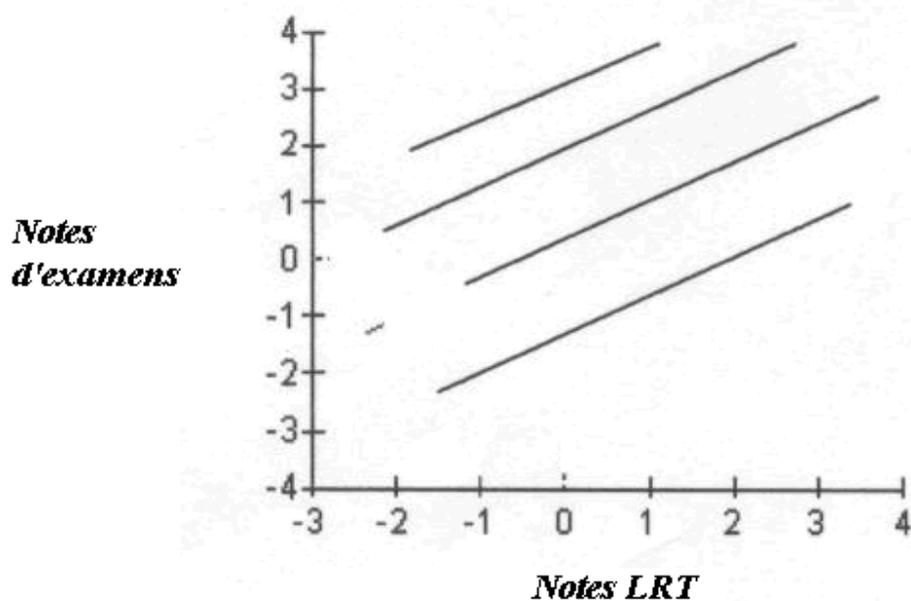


Fig 2.3 Variation au niveau 2

Les droites dans la figure 2.3 ont des intercepts différents; la variation entre ces intercepts est appelée variation de deuxième niveau puisque dans cet exemple les écoles sont les unités de deuxième niveau. L'ensemble des écoles forme un échantillon aléatoire d'une large population d'écoles. On ne s'intéresse ni aux élèves ni aux écoles en tant qu'individus; mais on s'intéresse aux inférences qu'on peut faire sur la variation entre toutes les écoles de la population, en utilisant l'échantillon d'écoles.

Si on considère qu'en utilisant les droites de la Fig. 2.3 pour prédire la note d'examen pour une note LRT donnée, alors il est clair que les différences entre les écoles est constante sur l'ensemble des notes LRT. Ce type de variation est appelé la variation simple de deuxième niveau.

L'exemple d'une variation plus complexe de deuxième niveau peut être obtenu si on permet au slopes des droites de varier. Comme dans la Fig2.4. C'est le cas où la différence entre les écoles dépend des notes LRT des élèves.

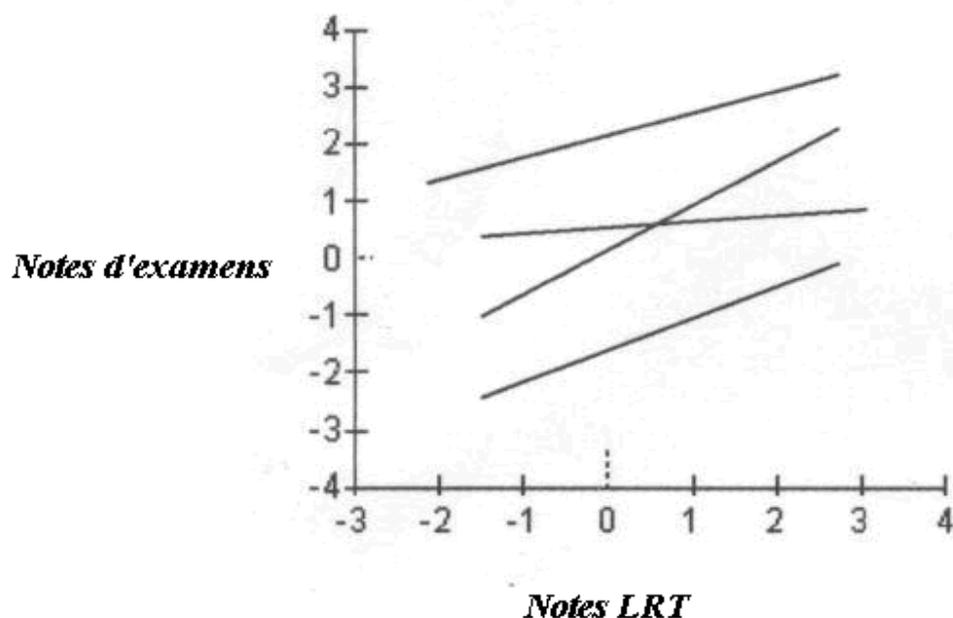


Fig 2.4 variation complexe au niveau 2

L'objet principal de l'analyse multi-niveaux ne concerne pas les écoles en tant qu'individus mais l'estimation de la structure de la variation (fluctuation) dans toute la population d'écoles. Une fois que cet objectif est atteint, il sera possible d'essayer d'expliquer cette variation en termes de caractéristiques générales des écoles en incorporant d'autres variables au modèle.

2. Description de la modélisation multi-niveaux

La fig. 2.2 fournit une illustration de la variation au premier niveau dans une seule école, ainsi qu'une droite de régression qui résume la relation entre les notes d'examens et les notes LRT des élèves de cette école. Cette droite est évidemment obtenue en utilisant la méthode classique des moindres carrés ordinaires.

Cependant on s'intéresse à la variation entre les écoles de notre échantillon comme un outil indispensable aux inférences possibles concernant cette même variation dans l'ensemble de toute la population d'écoles.

Soit une relation de régression linéaire pour une seule école; elle peut être exprimée par :

$$y_i = a + bx_i + e_i \quad (2.1)$$

où l'indice i varie de 1 au nombre d'élèves de l'école; dans notre exemple: y_i et x_i sont respectivement les notes d'examens et les notes LRT du i ème élève ; l'expression $a + bx_i$ est appelée partie fixe du modèle. La relation peut aussi être exprimée par:

$$\hat{y}_i = \hat{a} + \hat{b}x_i \quad (2.2)$$

où \hat{y}_i est la note d'examen du i ème élève prédite par la droite de régression. L'intercept \hat{a} est le point où cette droite croise l'axe vertical et \hat{b} son slope.

Dans l'équation (2.1), e_i est l'écart entre la note d'examen du i ème élève et sa note expliquée, on l'appellera l'erreur. L'erreur est la partie de la note y_i qui n'est pas expliquée par la partie fixe de la relation de régression de l'équation (2.2), avec une seule école, la variation du niveau 1 est la variance des e_i .

Observons maintenant le cas multi-niveaux à plusieurs écoles qui sont considérées comme un échantillon aléatoire d'une population d'écoles. On suppose une relation de régression pour chaque école:

$$y_{i1} = a_1 + bx_{i1}$$

$$y_{i2} = a_2 + bx_{i2}$$

.....

où les slopes sont parallèles, et de façon plus concise on peut l'écrire:

$$y_{ij} = a_j + bx_{ij} \quad (2.3)$$

L'indice j prend ses valeurs de 1 vers le nombre d'écoles dans l'échantillon. On peut écrire le modèle complet de la façon suivante:

$$y_{ij} = a_j + bx_{ij} + e_{ij} \quad (2.4)$$

Remarque

Si un terme est à deux indices ij alors il varie d'un élève à un autre dans une même école. Quand il est à un indice j , il varie seulement selon les écoles mais a la même valeur pour les élèves d'une même école; et quand il ne possède aucun indice; ceci voudra dire que c'est une constante sur l'ensemble des élèves et des écoles.

En analyse multi-niveaux; les groupes du deuxième niveau, dans notre cas les écoles, sont considérés comme un échantillon aléatoire d'une population; pour cette raison, on peut re exprimer l'équation (2.4) comme suit:

$$a_j = a + u_j$$

$$y_{ij} = a + bx_{ij} + e_{ij}$$

où a , sans aucun indice est une constante et u_j l'écart entre l'intercept relatif à la jème école et la valeur totale; on appellera u_j le résidu de deuxième niveau; il a la même valeur pour l'ensemble des élèves d'une même école.

Le modèle (2.4) peut maintenant être exprimé de la façon suivante:

$$y_{ij} = a + bx_{ij} + u_j + e_{ij} \quad (2.5)$$

Dans cette équation u_j et e_{ij} sont des quantités aléatoires de moyenne nulle, elles forment la partie aléatoire du modèle (2.5). Etant à plusieurs niveaux; on suppose que ces variables sont non corrélées puis on fera l'hypothèse standard qu'elles suivent la loi normale; il suffira alors d'estimer leurs variances notées σ_u^2 et σ_e^2 respectivement; les quantités: a ; l'intercept moyen et b le slope sont fixes et qu'on doit aussi estimer.

C'est la présence de deux paramètres aléatoires u_j et e_{ij} dans l'équation (2.5) qui fait de celle-ci un modèle multi-niveaux et plus précisément un modèle à deux niveaux.

Ainsi les variances σ_u^2 et σ_e^2 sont les paramètres aléatoires de notre modèle et les quantités a et b en sont les paramètres fixes. et comme les seules paramètres aléatoires de ce modèle sont les variances des résidus de chaque niveau, on se trouve dans le cas le plus simple appelé **modèle des composants de la variance**.

Si les slopes ne sont pas parallèles, la relation de régression pour chaque école sera :

$$y_1 = a_1 + b_1x_{i1}$$

$$y_2 = a_2 + b_2x_{i2}$$

.....

qu'on peut écrire sous la forme générale :

$$y_{ij} = a_j + b_jx_{ij}$$

Le modèle complet sera donc :

$$y_{ij} = a_j + b_j x_{ij} + e_{ij} \quad (2.6)$$

Supposons qu'on peut écrire a_j et b_j respectivement sous les formes suivantes :

$$a_j = a + u_{0j}$$

$$b_j = b + u_{1j}$$

où a et b sont des constantes et u_{0j} et u_{1j} sont des variables aléatoires de paramètres :

$$\begin{cases} E(u_{0j}) = E(u_{1j}) = 0 \\ \text{Var}(u_{0j}) = \sigma_{u0}^2, \text{Var}(u_{1j}) = \sigma_{u1}^2 \\ \text{Cov}(u_{0j}, u_{1j}) = \sigma_{u01} \end{cases}$$

On écrit alors le modèle comme suit :

$$y_{ij} = a + b x_{ij} + u_{1j} x_{ij} + u_{0j} + e_{ij} \quad (2.7)$$

Dans cette équation, y_{ij} est exprimé comme la somme d'une partie fixe $a + b x_{ij}$ et une partie aléatoire $u_{1j} x_{ij} + u_{0j} + e_{ij}$

3- Modèle linéaire à deux niveaux à une seule variable explicative

Notations

Dans le but d'exprimer plus généralement un modèle à deux niveaux, on adopte la notation du modèle (2.5), on introduit une variable explicative x_0 , qui prend la valeur 1 sur l'ensemble des élèves, on utilise les notations β_0, β_1, \dots etc. pour les paramètres fixes, les indices 0 et 1... etc. indiquent les variables explicatives auxquelles ils sont associés (β_0 est associé à la variable x_0) ; enfin et de façon similaire on incorpore l'indice 0 aux variables aléatoires (e_{ij} sera notée e_{0ij} et u_j sera notée u_{0j}). Ce qui permet d'obtenir :

$$y_{ij} = \beta_0 x_0 + \beta_1 x_{1ij} + u_{0j} x_0 + e_{0ij} x_0 \quad (2.8)$$

Remarques

L'indice 0 ajouté aux variables aléatoires des deux niveaux indique qu'elles sont associées à x_0
On écrit généralement la partie fixe sous la forme matricielle suivante :

$$E(y) = X\beta \quad \text{avec } y = \{y_{ij}\}$$

$$E(y_{ij}) = x_{ij}\beta = (X\beta)_{ij} \quad X = \{x_{ij}\}, \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

Où $\{ \}$ dénote une matrice.

Structure de variance de la variable dépendante

Le modèle (2.7) montre la différence essentielle qui existe entre les modèles multi-niveaux et les modèles simples à 1 niveau.

En effet l'estimation du modèle revient à estimer les deux coefficients fixes et les quatre paramètres $\sigma_{u0}^2, \sigma_{u1}^2, \sigma_{u01},$ et σ_{e0}^2 ; les variances et covariances sont les paramètres aléatoires du modèle.

On considère le modèle à deux niveaux, le plus simple qui comprend deux paramètres aléatoires $\sigma_{u0}^2, \sigma_{e0}^2$ qui seront notés respectivement par σ_u^2, σ_e^2 .

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_j + e_{ij} \quad (2.9)$$

Sous les hypothèses suivantes :

$$SH_1 \begin{cases} E(u) = E(e) = 0 \\ Var(u_j) = \sigma_u^2, \quad Var(e_{ij}) = \sigma_e^2 \\ Cov(u_j, e_{ik}) = 0 \quad \forall i, \forall j, \forall k \end{cases}$$

Remarque

Dans chaque unité du deuxième niveau j , i varie de 1 à n_j (n_j est le nombre d'unités de premier niveau dans la j ème unité de deuxième niveau).

Sous le système d'hypothèse (SH1), le modèle (2.9) est appelé **le modèle des composants de la variance** ; car la variance de la variable dépendante sachant les composants fixes est :

$$Var(y_{ij} / \beta_0, \beta_1, x_{ij}) = \sigma_u^2 + \sigma_e^2 \quad (2.10)$$

En effet

$$\begin{aligned} Var(\beta_0 + \beta_1 x_{ij} + u_j + e_{ij} / \beta_0, \beta_1, x_{ij}) &= Var(u_j + e_{ij}) \\ &= Var(u_j) + Var(e_{ij}) + 2cov(u_j, e_{ij}) \\ &= \sigma_u^2 + \sigma_e^2 \end{aligned}$$

Car u_j et e_{ij} sont non corrélés par hypothèses.

Cette variance est appelée le prédicteur fixe, c'est la somme des variances de premier et deuxième niveau. L'équation (2.10) implique que chaque groupe d'information de premier niveau a la même variance résiduelle et que la covariance entre deux unités de premier niveau (notée i_1, i_2) dans une même unité de niveau deux (notée j) est donnée par :

$$\text{cov}(y_{i_1j}, y_{i_2j}) = \text{cov}(u_j + e_{i_1j}, u_j + e_{i_2j}) = \text{cov}(u_j, u_j) = \sigma_u^2$$

En effet

$$\begin{aligned} \text{cov}(u_j + e_{i_1j}, u_j + e_{i_2j}) &= \frac{1}{2} (\text{Var}(2u_j + e_{i_1j} + e_{i_2j}) - \text{Var}(u_j + e_{i_1j}) - \text{Var}(u_j + e_{i_2j})) \\ &= \frac{1}{2} (4\text{Var}(u_j) + \text{Var}(e_{i_1j}) + \text{Var}(e_{i_2j}) - \text{Var}(u_j) - \text{Var}(e_{i_1j}) - \text{Var}(u_j) - \text{Var}(e_{i_2j})) \\ &= \frac{1}{2} (2\text{Var}(u_j)) = \text{Var}(u_j) = \sigma_u^2 \end{aligned}$$

Car u_j, e_{ij} sont non corrélés et e_{ij} sont indépendants par hypothèse.

La corrélation

Comme les résidus du premier niveau sont indépendants par hypothèse, la corrélation entre ces deux unités est :

$$\rho = \text{Cor}(u_j + e_{i_1j}, u_j + e_{i_2j}) = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}.$$

En effet :

$$\begin{aligned} \text{Cor}(u_j + e_{i_1j}, u_j + e_{i_2j}) &= \frac{\text{Cov}(u_j + e_{i_1j}, u_j + e_{i_2j})}{\sqrt{\text{Var}(u_j + e_{i_1j})\text{Var}(u_j + e_{i_2j})}} \\ &= \frac{\sigma_u^2}{\sqrt{(\sigma_u^2 + \sigma_e^2)(\sigma_u^2 + \sigma_e^2)}} \\ &= \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}. \end{aligned}$$

On peut appeler ρ corrélation intra unité, cette corrélation mesure la proportion de la variance totale entre les unités de premier niveau due à la variation entre les unités de deuxième niveau.

En modèle multi-niveaux, on doit séparer les variances relatives aux différents niveaux ; logiquement en citant les noms des unités. Par exemple dans un modèle à trois niveaux avec élèves, classes et écoles, on aura deux corrélations :

1. une corrélation intra école : mesure la corrélation entre les élèves des différentes classes d'une même école.
2. une corrélation intra classe : mesure la corrélation entre les élèves d'une même classe.

Ainsi, deux observations appartenant à la même unité de deuxième niveau sont corrélées à moins que $\sigma_u^2 = 0$; cas auquel correspond un modèle simple à un seul niveau.

L'existence d'une corrélation intra unité non nulle, résulte du fait qu'il y a plus d'un terme résiduel dans le modèle. Ceci implique que les procédures d'estimation classiques telles que la méthode des moindres carrés ordinaire utilisée par exemple en régression multiple ne sont pas applicables.

Regardons avec plus de détails la structure d'un groupe de deuxième niveau, en s'intéressant à la structure de la covariance.

Pour une unité de deuxième niveau j , supposons n_j le nombre d'unités de niveau 1 qu'elle contient, on peut former la $(n_j \times n_j)$ matrice V_j qui contient les variances des résidus $\sigma_u^2 + \sigma_e^2$ sur la diagonale et la covariance σ_u^2 par tout ailleurs :

$$V_j(i_1, i_2) = Cov(y_{i_1 j}, y_{i_2 j}) = \begin{cases} \sigma_e^2 + \sigma_u^2 & \text{si } i_1 = i_2 \\ \sigma_u^2 & \text{sinon} \end{cases}$$

$$V_j = \begin{bmatrix} \sigma_u^2 + \sigma_e^2 & & & & \\ & \ddots & & & \\ & & \sigma_u^2 & & \\ & & & \ddots & \\ & \sigma_u^2 & & & \\ & & & & \sigma_u^2 + \sigma_e^2 \end{bmatrix}$$

Comme les corrélations entre les unités du premier niveau des différentes unités de deuxième niveau sont nulles (c'est-à-dire $Cov(e_{ik_1}, e_{ik_2}) = 0$) la $(n \times n)$ matrice de variance covariance pour l'échantillon total est :

$$V = \bigoplus_{j=1}^m V_j$$

Où m est le nombre total d'unités de deuxième niveau, n le nombre total d'unités de premier niveau, $n = \sum_{j=1}^m n_j$ et \bigoplus est l'opérateur somme directe des matrices.

D'où V peut être écrite dans la forme suivante :

$$V = \begin{bmatrix} V_1 & & & \\ & V_2 & & O \\ & & \ddots & \\ & O & & \ddots \\ & & & & V_m \end{bmatrix} \quad (2.11)$$

Exemple Supposons un modèle à deux niveaux avec élèves et écoles. Pour deux écoles; la première à 3 élèves et la deuxième à 2 élèves. La matrice des covariances totale est :

$$\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$$

$$\text{où : } A = \begin{bmatrix} \sigma_e^2 + \sigma_u^2 & \sigma_u^2 & \sigma_u^2 \\ \sigma_u^2 & \sigma_e^2 + \sigma_u^2 & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2 & \sigma_e^2 + \sigma_u^2 \end{bmatrix}$$

et

$$B = \begin{bmatrix} \sigma_e^2 + \sigma_u^2 & \sigma_u^2 \\ \sigma_u^2 & \sigma_e^2 + \sigma_u^2 \end{bmatrix}$$

Cette structure reflète le fait que la covariance entre les élèves des différentes écoles est nulle et on écrit plus généralement :

$$V = \begin{bmatrix} \sigma_u^2 J(3) + \sigma_e^2 I(3) & 0 \\ 0 & \sigma_u^2 J(2) + \sigma_e^2 I(2) \end{bmatrix}$$

Où $I_{(k)}$ est la $(k \times k)$ -matrice unité et $J_{(k)}$ est la $(k \times k)$ -matrice de 1

Dans le cas classique de la régression linéaire à un niveau σ_u^2 est nulle ce qui réduit cette matrice à la forme standard $\sigma^2 I$ où σ^2 est la seule variance résiduelle et I la matrice unité de dimension n , n étant le nombre total d'unités de niveau 1.

4. Modèle linéaire à deux niveaux à plusieurs variables explicatives

La partie fixe du modèle (2.8) peut contenir d'autres variables explicatives :

$$y_{ij} = \beta_0 x_0 + \beta_1 x_{1ij} + \sum_{h=2}^p \beta_h x_{hij} + (u_{0j} + u_{1j} x_{1ij} + e_{0ij})$$

Ou bien :

$$y_{ij} = X_{ij}\beta + \sum_{h=0}^1 u_{hj}z_{hij} + e_{0ij}z_{0ij} \quad (2.12)$$

où on utilise de nouvelles variables explicatives dans la partie aléatoire du modèle et on écrit plus généralement :

$$\begin{aligned} Z &= \{Z_0, Z_1\} \\ Z_0 &= \{1\} \quad \text{C'est-à-dire un vecteur de } 1 \\ Z_1 &= \{x_{1ij}\} \end{aligned}$$

Remarques

1. Dans la partie aléatoire du modèle, les variables explicatives sont souvent un sous ensemble de celles de la partie fixe.
2. Dans le modèle (2.12), la variable possède un coefficient aléatoire au niveau 2. Notons qu'il faut distinguer entre la matrice de variance covariance de la variable dépendante donnée en (2.11) et la matrice de variance covariance des coefficients aléatoires.

Soit Ω_1 la matrice de covariance pour l'ensemble des coefficients aléatoires de niveau 1, dans le cas de notre modèle $\Omega_1 = \sigma_{e0}^2$.

Soit Ω_2 la matrice de covariance de l'intercept u_{0j} et le slope u_{1j} qui sont aléatoire au niveau 2 :

$$\Omega_2 = \begin{bmatrix} \sigma_{u0}^2 & \sigma_{u01} \\ \sigma_{u01} & \sigma_{u1}^2 \end{bmatrix}$$

On notera l'ensemble de matrices de covariance des coefficients aléatoires $\Omega = \{\Omega_i\}$.

Cas particulier

Dans un modèle à deux niveaux et pour une unité j de deuxième niveau qui contient deux unités de niveau 1, la matrice de covariance de la variable dépendante y est :

$$\begin{bmatrix} A & B \\ B & C \end{bmatrix}$$

Où:

$$A = \sigma_{u0}^2 + 2\sigma_{u01}x_{1j} + \sigma_{u1}^2x_{1j}^2 + \sigma_{e0}^2$$

$$B = \sigma_{u0}^2 + \sigma_{u01}(x_{1j} + x_{2j}) + \sigma_{u1}^2x_{1j}x_{2j}$$

$$C = \sigma_{u0}^2 + 2\sigma_{u01}x_{2j} + \sigma_{u1}^2x_{2j}^2 + \sigma_{e0}^2$$

Preuve

$$\begin{aligned} A &= \text{Var}(y_{1j} / \beta) = \text{Var}(u_{0j} + u_{1j}x_{1j} + e_{01j}) \\ &= \text{Var}(u_{0j}) + \text{Var}(u_{1j}x_{1j}) + \text{Var}(e_{01j}) + 2\text{Cov}(u_{0j}, u_{1j}x_{1j}) \\ &= \sigma_{u0}^2 + 2\sigma_{u01}x_{1j} + \sigma_{u1}^2x_{1j}^2 + \sigma_{e0}^2 \end{aligned}$$

$$\begin{aligned} B &= \text{Cov}(y_{1j}, y_{2j}) = \frac{1}{2}(\text{Var}(y_{1j} + y_{2j}) - \text{Var}(y_{1j}) - \text{Var}(y_{2j})) \\ &= \frac{1}{2}(T_1 - A - C) \end{aligned}$$

$$\begin{aligned} T_1 &= \text{Var}(y_{1j} + y_{2j}) = \text{Var}(2u_{0j} + u_{1j}(x_{1j} + x_{2j}) + e_{01j} + e_{02j}) \\ &= 4\sigma_{u0}^2 + (x_{1j} + x_{2j})^2 \sigma_{u1}^2 + \sigma_{e0}^2 + \sigma_{e0}^2 + 2\text{Cov}(2u_{0j}, u_{1j}(x_{1j} + x_{2j})) \\ &= 4\sigma_{u0}^2 + (x_{1j} + x_{2j})^2 \sigma_{u1}^2 + 2\sigma_{e0}^2 + 2.2(x_{1j} + x_{2j})\sigma_{u01} \end{aligned}$$

$$\begin{aligned} C &= \text{Var}(y_{2j} / \beta) = \text{Var}(u_{0j} + u_{1j}x_{2j} + e_{02j}) \\ &= \text{Var}(u_{0j}) + \text{Var}(u_{1j}x_{2j}) + \text{Var}(e_{02j}) + 2\text{Cov}(u_{0j}, u_{1j}x_{2j}) \\ &= \sigma_{u0}^2 + 2\sigma_{u01}x_{2j} + \sigma_{u1}^2x_{2j}^2 + \sigma_{e0}^2 \end{aligned}$$

D'où on peut calculer B :

$$\begin{aligned} B &= \frac{1}{2} \left[4\sigma_{u0}^2 + (x_{1j} + x_{2j})^2 \sigma_{u1}^2 + 2\sigma_{e0}^2 + 4(x_{1j} + x_{2j})\sigma_{u01} - \sigma_{u0}^2 - 2\sigma_{u01}x_{1j} \right. \\ &\quad \left. - \sigma_{u1}^2x_{1j}^2 - \sigma_{e0}^2 - \sigma_{u0}^2 - 2\sigma_{u01}x_{1j} - 2\sigma_{u01}x_{1j} - \sigma_{u1}^2x_{2j}^2 - \sigma_{e0}^2 \right] \\ &= \frac{1}{2} \left[2\sigma_{u0}^2 + 2(x_{1j} + x_{2j})\sigma_{u01} + 2x_{1j}x_{2j} \sigma_{u1}^2 \right] \\ &= \sigma_{u0}^2 + \sigma_{u01}(x_{1j} + x_{2j}) + \sigma_{u1}^2x_{1j}x_{2j} \end{aligned}$$

Étant donné :

$$\begin{bmatrix} A & B \\ B & C \end{bmatrix} = X_j \Omega_2 X_j^t + \begin{bmatrix} \Omega_1 & \\ & \Omega_1 \end{bmatrix}$$

Tel que :

$$X_j = \begin{bmatrix} 1 & x_{1j} \\ 1 & x_{2j} \end{bmatrix} \quad \Omega_2 = \begin{bmatrix} \sigma_{u0}^2 & \sigma_{u01} \\ \sigma_{u01} & \sigma_{u0}^2 \end{bmatrix} \quad \text{et} \quad \Omega_1 = \sigma_{e0}^2$$

5. Structure générale d'un modèle linéaire à deux niveaux

Soit la structure générale d'un modèle à deux niveaux suivante :

$$Y = X\beta + E \quad (2.13)$$

Où :

$$Y = \{y_{ij}\}$$

$$X = \{X_{ij}\} \quad X_{ij} = \{x_{0ij}, x_{1ij}, \dots, x_{pij}\}$$

j : L'indice de l'unité de niveau deux $j = 1, \dots, n$.

i : L'indice de l'unité de premier niveau, pour chaque valeur de j , i varie de 1 à n_j (n_j étant le nombre d'unités de premier niveau dans la j ème unité de niveau deux).

p : Le nombre de variables explicatives dans la partie fixe du modèle.

E est la partie aléatoire du modèle qu'on peut exprimer par :

$$E = E_1 + E_2 = \{e_{ij}\}, \quad e_{ij} = e_{ij}^{(1)} + e_j^{(2)}$$

E_1 et E_2 : sont deux matrices de résidus respectivement de premier et deuxième niveau.

Où :

$$e_{ij}^{(1)} = \sum_{h=0}^{q_1} Z_{hij}^{(1)} e_{hij}^{(1)} \quad e_j^{(2)} = \sum_{h=0}^{q_2} Z_{hij}^{(2)} e_{hj}^{(2)}$$

$e_{ij}^{(1)}$ représente la partie aléatoire au premier niveau; q_1 représente le nombre de variables explicatives à coefficients aléatoires au premier niveau.

$e_j^{(2)}$ représente la partie aléatoire au deuxième niveau; q_2 représente le nombre de variables explicatives à coefficients aléatoires au deuxième niveau.

On re note $e_{ij}^{(1)}$ par e_{ij} et $e_j^{(2)}$ par u_j , ce qui permet d'écrire la partie aléatoire du modèle sous la forme matricielle suivante :

$$E = Z^{(2)}u + Z^{(1)}e$$

D'où la forme matricielle du modèle (2.13)

$$Y = X\beta + Z^{(2)}u + Z^{(1)}e \quad (2.14)$$

Sous les hypothèses :

$$\begin{cases} E(E_1) = 0, & E(E_2) = 0, \\ \text{Var}(E_1) = V_{2(1)}, & \text{Var}(E_2) = V_{2(2)} \\ E(E_1 E_2^t) = 0 & V_2 = V_{2(1)} + V_{2(2)} \end{cases}$$

Remarques :

Dans le modèle (2.14) :

X est la matrice des variables explicatives de la partie fixe.

$Z^{(2)}$ est la matrice des variables explicatives de la partie aléatoire ayant des coefficients aléatoires au deuxième niveau.

$Z^{(1)}$ est la matrice des variables explicatives de la partie aléatoire ayant des coefficients aléatoires au premier niveau.

$E(E_1 E_2^t) = 0$ indique que les coefficients aléatoires du premier niveau sont indépendants de ceux du deuxième niveau.

Structure de la variance de la variable dépendante

-Les résidus de premier niveau sont supposés indépendants, $V_{2(1)}$ est donc une matrice diagonale où son élément ij est :

$$\text{Var}(e_{ij}) = \sigma_{e_{ij}}^2 = Z_{ij}^{(1)t} \Omega_e Z_{ij}^{(1)}, \quad \Omega_e = \text{Cov}\left(e_{(h)}^{(1)}\right).$$

-Les résidus du deuxième niveau sont indépendants sur l'ensemble des unités de deuxième niveau, leur matrice de covariance $V_{2(2)}$ est une matrice diagonale par bloc dont le j ème bloc est défini par :

$$V_{2(2)j} = Z_j^{(2)t} \Omega_u Z_j^{(2)}, \quad \Omega_2 = \text{Cov}\left(e_{(h)}^{(2)}\right).$$

Le j ème bloc de la matrice est alors donné par :

$$V_{2j} = \oplus_i \sigma_{e_{ij}}^2 + V_{2(2)j}.$$

Où \oplus est l'opérateur somme directe

Remarque

D'après les notations utilisées dans la définition du modèle (2.12) ; la matrice $V_{2(2)j}$ est une matrices $(n_j \times n_j)$.

Chapitre 3

ESTIMATION D'UN MODÈLE DE RÉGRESSION LINÉAIRE À DEUX NIVEAUX PAR LA MÉTHODE ITÉRATIVE DES MOINDRES CARRÉS GÉNÉRALISÉE.

Introduction

Ce chapitre sera consacré à l'étude de la méthode d'estimation d'un modèle de régression linéaire à deux niveaux, cette méthode, dite méthode itérative des moindres carrés généralisée a le principe suivant :

À chaque itération, on estime d'abord les paramètres fixes c'est-à-dire qu'on va considérer qu'on n'a qu'un seul terme résiduel ; une fois ce terme estimé, on construit un modèle de régression linéaire en prenant les résidus estimés comme les observations d'une variable dépendante. Estimer le modèle ainsi obtenu donnera les estimateurs de variances et des covariances des erreurs des deux niveaux, on estime les paramètres fixes en utilisant ces variances et ces covariances ; ce qui produit un cycle itératif de la méthode.

Il est à noter que chaque modèle sera estimé par la méthode des Moindres Carrés Généralisée (MCG), sauf pour la première itération où on estimera les paramètres fixes de la méthode des Moindres Carrés Ordinaires (MCO), car on ne dispose pas des variances. Avant de voir plus en détails cette méthode sur un cas particulier, voici quelques définitions utiles (Intrilligator, et al.)

Définition Produit de Kronecker

Le produit de Kronecker d'une matrice par une matrice, multiplie chaque élément de la matrice à gauche par toute la matrice de droite. Ainsi :

Soit A une matrice $(m \times n)$ et B une matrice $(p \times q)$; le produit de Kronecker est donné par:

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \dots & a_{1n}B \\ a_{21}B & a_{22}B & \dots & a_{2n}B \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1}B & a_{m2}B & \dots & a_{mn}B \end{bmatrix}$$

Le résultat est une matrice $(mp \times nq)$.

Propriété du produit de Kronecker de deux matrices

- $(A \otimes B)' = A' \otimes B'$
- $(A + B) \otimes C = (A \otimes C) + (B \otimes C)$
- $A \otimes (B + C) = (A \otimes B) + (A \otimes C)$
- $(A \otimes B)(C \otimes D) = AC \otimes BD$
- $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ Si A et B sont carrées et non singulières.

Définition L'opérateur Vec

Soit $A = (a_1, a_2, \dots, a_p)$ une matrice $(n \times p)$ (c'est-à-dire a_i est un n vecteur colonne $i = 1, \dots, p$).

On définit un vecteur de dimension np :

$$Vec(A) = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}$$

Ainsi, $y = \{y_{ij}\}$ peut être noté $y = (y_1, y_2, \dots, y_m)$;

où y_j est un vecteur colonne de dimension n_j

$$y_j = (y_{1j}, y_{2j}, \dots, y_{n_j j})$$

On définit $Y = Vec(y) = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$

Y est alors un vecteur colonne de dimension $N = \sum_{j=1}^m n_j$ de matrice de variance covariance V .

De façon analogue, pour $x = \{x_{ij}\}$ on définit le N vecteur colonne :

$$Vec(x) = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$$

où x_j est un vecteur colonne de dimension n_j ; $j = 1, \dots, m$.

Soit 1_N un N vecteur colonne de 1 et soit $X = (1_N, Vec(x))$ une matrice $(N \times 2)$.

Cette dernière matrice X ; la matrice plan d'expérience, et le vecteur (variable dépendante) Y seront utilisés pour estimer les paramètres fixes du modèle.

1. Méthode Itérative des Moindres Carrés Généralisée : IMCG

Soit le modèle de régression linéaire simple à deux niveaux :

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_j + e_{ij} \quad (3.1)$$

Où i et j indiquent l'unité de 1^{er} et 2^{ème} niveau respectivement.

Sous les hypothèses suivantes :

$$\begin{cases} e_{ij} \sim N(0, \sigma_e^2) \\ Cov(e_{ij}, e_{ik}) = 0 \text{ si } j \neq k \\ u_j \sim N(0, \sigma_u^2) \\ Cov(e_{ij}, u_j) = 0 \end{cases}$$

Un tel modèle est appelé modèle des composants de la variance.

Soit m le nombre total d'unités de niveau 2 et n_j le nombre d'unités de niveau 1 dans la $j^{\text{ème}}$ unité de niveau 2 ; le nombre total d'unités de niveau 1 est $N = \sum_{j=1}^m n_j$

La matrice de variance covariance de la variable dépendante est donné par :

$$V = \begin{bmatrix} V_1 & & & 0 \\ & V_2 & & \\ & & \ddots & \\ 0 & & & V_m \end{bmatrix}$$

où chaque bloc $V_j = \sigma_e^2 J_j + \sigma_u^2 I_j$

J_j est la $(n_j \times n_j)$ matrice de 1

I_j est la $(n_j \times n_j)$ matrice unité

La matrice V est de dimension $(N \times N)$.

Estimation des paramètres

Supposons qu'on connaît les variances σ_e^2 et σ_u^2 , on peut alors construire V et appliquer la méthode MCG pour estimer les paramètres fixes du modèle (3.1) $\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}$.

On suppose que dans (3.1), on a un seul terme d'erreur R , ce qui permet de l'écrire sous la forme :

$$y = X\beta + R$$

Connaissant V , l'estimateur du MCG du paramètre β est :

$$\hat{\beta} = (X'V^{-1}X)^{-1} X'V^{-1}y \quad (3.2)$$

À partir de cet estimateur, on calcule les résidus :

$$\tilde{y}_{ij} = y_{ij} - \hat{\beta}_0 - \hat{\beta}_1 x_{ij}$$

Remarque

Chaque \tilde{y}_{ij} est un estimateur de $e_{ij} + u_j$. Le N vecteur des résidus est noté $\tilde{y} = \{\tilde{y}_{ij}\}$, si on forme la matrice $\tilde{y}\tilde{y}'$ on voit simplement que c'est un estimateur de V .

Cas particulier

Reprenons le cas particulier du chapitre II où l'on a

$$Vec(y) = \begin{bmatrix} y_{11} \\ y_{21} \\ y_{31} \\ y_{12} \\ y_{22} \end{bmatrix}$$

$$\tilde{y} = (\tilde{y}_1, \tilde{y}_2)$$

tel que :

$$\tilde{y}_1 = (\tilde{y}_{11}, \tilde{y}_{21}, \tilde{y}_{31})'$$

$$\tilde{y}_2 = (\tilde{y}_{12}, \tilde{y}_{22})'$$

On a

$$\tilde{y}_1 \tilde{y}_1' = \begin{bmatrix} \tilde{y}_{11}^2 & \tilde{y}_{11} \tilde{y}_{21} & \tilde{y}_{11} \tilde{y}_{31} \\ \tilde{y}_{21} \tilde{y}_{11} & \tilde{y}_{21}^2 & \tilde{y}_{21} \tilde{y}_{31} \\ \tilde{y}_{31} \tilde{y}_{11} & \tilde{y}_{31} \tilde{y}_{21} & \tilde{y}_{31}^2 \end{bmatrix}$$

qui est estimateur de V_1 et

$$\tilde{y}_2 \tilde{y}_2' = \begin{bmatrix} \tilde{y}_{12}^2 & \tilde{y}_{12} \tilde{y}_{22} \\ \tilde{y}_{22} \tilde{y}_{12} & \tilde{y}_{22}^2 \end{bmatrix}$$

qui est estimateur de V_2

Rappelons que

$$V = \begin{bmatrix} V_1 & 0 \\ 0 & V_2 \end{bmatrix}$$

Si on construit $Vec(\tilde{y}\tilde{y}')$, on obtient un vecteur colonne de dimension $3^2 + 2^2 = 13$ qui est le suivant :

$$Vec(\tilde{y}\tilde{y}') = \begin{bmatrix} Vec(\tilde{y}_1 \tilde{y}_1') \\ Vec(\tilde{y}_2 \tilde{y}_2') \end{bmatrix}$$

De même $Vec(V) = \begin{bmatrix} Vec(V_1) \\ Vec(V_2) \end{bmatrix}$ qui est aussi un vecteur colonne de dimension $3^2 + 2^2 = 13$ et

son estimateur est $Vec(\tilde{y}\tilde{y}')$.

Dans le cas général, la relation entre ces deux vecteurs peut être exprimée par le modèle de régression linéaire suivant :

$$\begin{bmatrix} \tilde{y}_{11}^2 \\ \tilde{y}_{11}\tilde{y}_{21} \\ \vdots \\ \vdots \\ \tilde{y}_{n_m}^2 \end{bmatrix} = \begin{bmatrix} \sigma_e^2 + \sigma_u^2 \\ \sigma_u^2 \\ \vdots \\ \sigma_e^2 + \sigma_u^2 \end{bmatrix} + R = \sigma_u^2 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ \vdots \\ 1 \end{bmatrix} + \sigma_e^2 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ \vdots \\ 1 \end{bmatrix} + R \quad (3.3)$$

Où R est le vecteur des résidus.

En (3.3), le côté gauche est évidemment le vecteur dépendant du modèle et le côté droit deux variables explicatives avec les deux coefficients σ_e^2 et σ_u^2 qui sont à estimer.

Remarque

Le vecteur $Vec(\tilde{y}\tilde{y}')$ qui correspond au modèle (3.1) peut être écrit sous la forme :

$$\tilde{y} = Vec(\tilde{y}\tilde{y}') = \begin{bmatrix} Vec(\tilde{y}_1\tilde{y}'_1) \\ Vec(\tilde{y}_2\tilde{y}'_2) \\ \vdots \\ Vec(\tilde{y}_m\tilde{y}'_m) \end{bmatrix}$$

Il est de dimension $D = \sum_{j=1}^m n_j^2$

De même $Vec(V)$ peut aussi être écrit sous la forme :

$$Vec(V) = \begin{bmatrix} Vec(V_1) \\ Vec(V_2) \\ \vdots \\ \vdots \\ Vec(V_m) \end{bmatrix}$$

Il est aussi de dimension $D = \sum_{j=1}^m n_j^2$

Cette notation nous permet d'expliquer comment obtenir les deux vecteurs variables explicatives du modèle (3.3) ; en effet le premier vecteur qui a σ_u^2 comme coefficient est obtenu par :

$$J = \begin{bmatrix} \text{Vec}(J_{(1)}) \\ \text{Vec}(J_{(2)}) \\ \vdots \\ \text{Vec}(J_{(m)}) \end{bmatrix},$$

Il est de dimension D .

Le deuxième vecteur est aussi de dimension D , il est obtenu en superposant les vec des matrices identité $I_{(1)}, I_{(2)}, \dots, I_{(m)}$ c'est-à-dire

$$I = \begin{bmatrix} \text{Vec}(I_{(1)}) \\ \text{Vec}(I_{(2)}) \\ \vdots \\ \text{Vec}(I_{(m)}) \end{bmatrix}.$$

L'estimation du modèle (3.3) nécessite l'application de la méthode MCG en utilisant la matrice des covariances estimée du vecteur $\text{Vec}(\tilde{y}\tilde{y}')$ qui sous hypothèse de normalité est $\Sigma = 2(V^{-1} \otimes V^{-1})$ [Goldstein, H (1996)]

Le modèle (3.3) peut être écrit sous la forme suivante :

$$\begin{aligned} \tilde{y} &= J\sigma_u^2 + I\sigma_e^2 + R \\ &= Z^*\theta + R \end{aligned}$$

Où $Z^* = (J, I)$ une matrice $(2 \times D)$ et $\theta = \begin{pmatrix} \sigma_u^2 \\ \sigma_e^2 \end{pmatrix}$.

L'estimateur de MCG du paramètre θ est :

$$\hat{\theta} = \left(Z^* \hat{\Sigma}^{-1} Z^{*'} \right)^{-1} Z^* \hat{\Sigma}^{-1} \tilde{y}. \quad (3.4)$$

En utilisant ces estimateurs, on peut construire la matrice \hat{V} et reprendre (3.2) pour calculer de nouveaux estimateurs des paramètres fixes ; on alterne alors entre l'estimation des paramètres fixes en (3.2) et les paramètres aléatoires (3.4) jusqu'à ce que la procédure converge. C'est-à-dire que les estimateurs de tous les paramètres ne changent plus de valeurs d'une itération à une autre.

Remarque La procédure d'estimation est itérative, on doit commencer par le calcul de $\hat{\beta}$ par (3.2), et comme on ne dispose pas de la matrice V à la 1^{ère} itération on commence usuellement par des estimateurs raisonnables des paramètres fixes, ceux-ci sont obtenus par la méthode des MCO.

Estimation des résidus

Dans un modèle à un niveau, l'estimateur du terme résiduel est justement $\tilde{y}_i = y_i - \hat{y}_i$, cependant dans un modèle multi-niveaux on a plusieurs résidus à plusieurs niveaux. Dans ce qui suit on ne considérera que l'estimation des résidus individuels (C'est-à-dire que chaque résidu sera estimé individuellement).

Étant donné les estimateurs des paramètres d'un modèle des composants de la variance à deux niveaux, pour chaque unité de deuxième niveau on estime u_j par :

$$\hat{u}_j = E(u_j / y, \hat{\beta}, \hat{\Omega}). \quad (3.5)$$

Si on néglige la variation des estimateurs de (3.5), c'est-à-dire qu'on considère qu'ils sont fixes et non aléatoires on obtient :

$$\begin{aligned} Cov(\tilde{y}_{ij}, u_j) &= Var(u_j) = \sigma_u^2, \\ Cov(\tilde{y}_{ij}, e_{ij}) &= Var(e_{ij}) = \sigma_e^2, \\ Var(\tilde{y}_{ij}) &= \sigma_u^2 + \sigma_e^2. \end{aligned} \quad (3.6)$$

On considère (3.5) comme une régression linéaire de u_j sur l'ensemble des $\{y_{ij}\}$ (pour la j ème unité de 2^{ème} niveau), et (3.6) nous permet d'estimer les coefficients de la régression et de là \hat{u}_j . on obtient :

$$\begin{aligned}\hat{u}_j &= \frac{n_j \sigma_u^2}{n_j \sigma_u^2 + \sigma_e^2} \tilde{y}_j, \\ \hat{e}_{ij} &= \tilde{y}_{ij} - \hat{u}_j, \\ \tilde{y}_j &= \frac{\sum_{i=1}^{n_j} \tilde{y}_{ij}}{n_j}.\end{aligned}\tag{3.7}$$

2. Généralisation de la méthode IMCG

Pour un modèle linéaire à deux niveaux :

$$y = X\beta + Z^{(2)}u + Z^{(1)}e$$

Etape N°1 Calcul des estimateurs des paramètres fixes.

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y\tag{3.8}$$

Etape N°2

On calcule le vecteur $y^* = \tilde{y}\tilde{y}'$, $\tilde{y} = y - X\hat{\beta}$

On sait que $E(y^*) = V$

On construit $y^{**} = \text{Vec}(y^*)$

Etape N°3

y^{**} peut être utilisée comme variable dépendante dans un modèle de régression linéaire invoquant les paramètres aléatoires c'est-à-dire les éléments de Ω_u et Ω_e et on obtient le modèle suivant :

$$E(y^{**}) = Z^* \theta \quad \text{Var}(y^{**}) = V \otimes V$$

Où Z^* est la matrice plan d'expérience, et θ le vecteur contenant les éléments de Ω_u et Ω_e à estimer.

Remarque

Un exemple d'une telle matrice plan d'expérience Z^* est donné par celle concernant le modèle (3.3) à savoir $Z^* = (J, I)$.

L'estimateur des MCG pour θ est

$$\hat{\theta} = \left(Z^{*'} V^{*-1} Z^* \right)^{-1} Z^{*'} V^{*-1} y^{**} \quad \text{où } V^* = V \otimes V\tag{3.9}$$

La procédure itère entre (3.9) et (3.8) [Goldstein, H (1996)]. À la première itération on considère les estimateurs des moindres carrés ordinaires des paramètres fixes.

La matrice de covariance de $\hat{\theta}$ est :

$$Cov(\hat{\theta}) = 2 \left(Z^{*'} V^{*-1} Z^* \right)^{-1}$$

En effet

$$\begin{aligned} Cov(\hat{\theta}) &= Cov \left(\left(Z^{*'} V^{*-1} Z^* \right)^{-1} Z^{*'} V^{*-1} y^{**} \right) \\ &= \left(Z^{*'} V^{*-1} Z^* \right)^{-1} Z^{*'} V^{*-1} Cov(y^{**}) V^{*-1} Z^* \left(Z^{*'} V^{*-1} Z^* \right)^{-1} \end{aligned}$$

On a

$$y^{**} = Vec(\tilde{y}\tilde{y}') = \tilde{y} \otimes \tilde{y}$$

D'après Searle et al (1992); on a $Cov(\tilde{y} \otimes \tilde{y}) = (V \otimes V)(I + S_N)$ où $(V \otimes V) = V^*$ et S_N est le Vec de la matrice de permutation Goldstein, H et Rasbash, J (1992) :

Soit une matrice A symétrique et $Z^* = Vec(A)$:

$$V^{*-1} Z^* = (V \otimes V)^{-1} Vec(A) = (V^{-1} \otimes V^{-1}) Vec(A) = Vec(V^{-1} A V^{-1})$$

$V^{-1} A V^{-1}$ est symétrique; et $S_N V^{*-1} Z^* = V^{*-1} Z^*$

En remplaçant dans l'expression de $Cov(\hat{\theta})$ et on obtient $Cov(\hat{\theta}) = 2 \left(Z^{*'} V^{*-1} Z^* \right)^{-1}$ **c.q.f.d**

Chapitre 4

L'ESTIMATION D'UN MODELE DE REGRESSION LINEAIRE MULTI-NIVEAUX PAR LA METHODE ITERATIVE DE BOOTSTRAP

Introduction

La méthode de Bootstrap est introduite en 1979 comme une méthode basée sur le calcul machine. Elle a l'avantage d'être complètement automatique, l'estimation d'un paramètre θ par la méthode de Bootstrap ne nécessite aucun calcul théorique et est applicable quelque soit la complexité du calcul de θ . Les méthodes de Bootstrap dépendent de la notion d'échantillonnage de Bootstrap.

1. L'échantillon de Bootstrap Efron, B and Tibshirani, R. J(1993)

A partir d'un échantillon aléatoire $X = (x_1, x_2, \dots, x_n)$ issu d'une fonction de distribution F inconnue on veut estimer un paramètre $\theta = t(F)$ sur la base de X , soit $\hat{\theta} = s(X)$ cet estimateur.

1.1 Définition

Soit \hat{F} la distribution empirique donnant une probabilité $1/n$ pour chaque valeur observée x_i , $i = 1, \dots, n$. Un échantillon de Bootstrap X^* est un échantillon aléatoire de taille n obtenue à partir de \hat{F} tel que:

$$\begin{aligned} X^* &= (x_1^*, x_2^*, \dots, x_n^*) \\ \hat{F} &\rightarrow (x_1^*, x_2^*, \dots, x_n^*) \end{aligned} \quad (4.1)$$

Le symbole * indique que X^* n'est pas l'ensemble des données X mais une version randomisée de X .

Il y a une autre façon de voir l'échantillon (4.1); les valeurs $(x_1^*, x_2^*, \dots, x_n^*)$ forment un échantillon aléatoire de taille n , obtenu, avec remplacement à partir d'une population de n objets (x_1, x_2, \dots, x_n) ainsi on peut avoir: $x_1^* = x_4, x_2^* = x_1, x_3^* = x_{12}, \dots, x_n^* = x_1$.

Un échantillon de Bootstrap $(x_1^*, x_2^*, \dots, x_n^*)$ consiste en des éléments de l'ensemble de données initial (x_1, x_2, \dots, x_n) ; certains de ces éléments n'apparaissent aucune fois, d'autres apparaissent une fois, d'autres deux fois etc....

1.2 Comment obtenir un échantillon de Bootstrap?

Simuler des entiers i_1, i_2, \dots, i_n chacun d'eux est compris entre 1 et n avec une probabilité $1/n$, l'échantillon de Bootstrap est l'ensemble de valeurs de l'ensemble des données X tel que:

$$x_1^* = x_{i_1}, x_2^* = x_{i_2}, x_3^* = x_{i_3}, \dots, x_n^* = x_{i_n} \quad (4.2)$$

Une méthode d'estimation par Bootstrap utilise plusieurs échantillons indépendants obtenus de cette façon.

1.3 Application L'estimation de l'erreur standard par la méthode de Bootstrap

A partir d'un échantillon aléatoire $X = (x_1, x_2, \dots, x_n)$ d'une fonction de distribution F inconnue on veut estimer un paramètre $\theta = t(F)$ sur la base de X , dans ce but on calcule l'estimateur $\hat{\theta} = s(X)$. Par la méthode de Bootstrap, on peut estimer l'erreur standard ie $se_F(\hat{\theta})$ ce qui permet de mesurer la précision de $\hat{\theta}$

On note par $\hat{\theta}^*$, la réplique de Bootstrap (*Bootstrap replicate*) de $\hat{\theta}$ associée à l'échantillon X^* , on a:

$$\hat{\theta}^* = s(X^*) \quad (4.3)$$

$s(X^*)$ Est obtenue en appliquant la même fonction $s(\cdot)$ à X^* comme appliquée à X .

Exemple

Si $s(X)$ est la moyenne de l'échantillon alors $s(X^*)$ est la moyenne des données de Bootstrap

$$s(X^*) = \bar{X}^* = \sum_{i=1}^n x_i^* / n$$

L'estimation par la méthode de Bootstrap de $se_F(\hat{\theta})$: l'erreur standard de la statistique $\hat{\theta}$, utilise la fonction de distribution empirique \hat{F} au lieu de la distribution inconnue F ; l'estimateur obtenue par cette méthode est noté: $se_{\hat{F}}(\hat{\theta}^*)$

En d'autres termes; c'est l'erreur standard de $\hat{\theta}$ pour les échantillons aléatoires de taille n obtenus en utilisant \hat{F} . $se_{\hat{F}}(\hat{\theta}^*)$ est appelée l'estimateur idéal de Bootstrap de l'erreur type de la statistique $\hat{\theta}$.

1.3.1 Algorithme de Bootstrap pour estimer l'erreur standard

Cet algorithme permet d'obtenir une bonne approximation de la valeur numérique de $se_{\hat{F}}(\hat{\theta}^*)$:

Etape N° 1

Obtenir B échantillons indépendants de Bootstrap $X^{*1}, X^{*2}, \dots, X^{*B}$, chacun est un ensemble de valeurs tirées des n valeurs de X avec remplacement, comme en (4.1) ou (4.2)

Etape N° 2

Calculer la réplication correspondant à chaque échantillon :

$$\hat{\theta}^*(b) = s(X^{*b}) \quad b = 1, 2, \dots, B \quad (4.4)$$

Etape N° 3

Estimer l'erreur standard $se_F(\hat{\theta})$ par l'écart type de l'échantillon des B réplifications

$$s\hat{e}_B = \left\{ \sum_{b=1}^B [\hat{\theta}^*(b) - \hat{\theta}^*(.)]^2 / (B-1) \right\}^{1/2} \quad (4.5)$$

$$\text{ou } \hat{\theta}^*(.) = \sum_{b=1}^B \hat{\theta}^*(b) / B$$

Remarque

La limite de $s\hat{e}_B$ quand B tend vers l'infini est l'estimateur idéal de Bootstrap de $se_F(\hat{\theta})$,

$$\lim_{B \rightarrow \infty} se_B = se_{\hat{F}} = se_{\hat{F}}(\hat{\theta}^*) \quad (4.6)$$

Ceci revient à dire que l'écart type empirique approche l'écart type de la population autant que le nombre de répliques B est grand; la population ici est la population des B valeurs $\hat{\theta}^* = s(X^*)$ ou $\hat{F} \rightarrow (x_1^*, x_2^*, \dots, x_n^*) = X^*$.

1.3.2 La méthode paramétrique

L'estimateur idéal de Bootstrap $se_{\hat{F}}(\hat{\theta}^*)$ et son approximation $s\hat{e}_B$ sont parfois appelés des estimateurs non paramétriques car ils sont basés sur \hat{F} l'estimateur non paramétrique de F .

L'estimateur paramétrique de l'erreur standard est noté: $se_{\hat{F}_{par}}(\hat{\theta}^*)$

où \hat{F}_{par} est un estimateur de F tiré d'un modèle paramétrique; un exemple pour illustrer cette idée : au lieu d'estimer F par la distribution empirique on peut supposer que F suit une loi normale notée \hat{F}_{Norm} , l'estimateur paramétrique de Bootstrap de l'erreur standard de $\hat{\theta}$ est $s\hat{e}_{\hat{F}_{Norm}}(\hat{\theta}^*)$ comme dans le cas non paramétrique. Par conséquent, pour approximer $se_{\hat{F}_{par}}(\hat{\theta}^*)$ on simule B échantillons de taille n à partir de l'estimateur paramétrique \hat{F}_{par} ,

$$\hat{F}_{par} \rightarrow (x_1^*, x_2^*, \dots, x_n^*).$$

au lieu d'utiliser des échantillons générés, avec remplacement, à partir de l'ensemble des données initial. Par la suite on procède de la même façon étape2/ étape3/de l'algorithme de Bootstrap.

2. Estimation d'un modèle de régression linéaire à deux niveaux Goldstein, H(1998)

2.1 Principe de la méthode

Soit un modèle de régression linéaire à deux niveaux, ayant ajusté (estimé) ce modèle et obtenu des estimateurs de ses paramètres; le principe de cette méthode est d'utiliser le modèle estimé pour simuler un ensemble de données, utiliser ces dernières pour réajuster (ré estimer) le modèle et obtenir par la suites de nouveaux estimateurs des paramètres appelés les estimateurs de Bootstrap.

On considère le modèle de composants de la variance à deux niveaux

$$y_{ij} = (X\beta)_{ij} + u_i + e_{ij} \quad (4.7)$$

Considérons d'abord la méthode d'échantillonnage, le processus de simulation d'un échantillon de Bootstrap, correspond au mécanisme qui est supposé avoir généré les données. Ceci peut être modélisé comme une sélection d'un échantillon aléatoire de résidus des unités de deuxième niveau

selon la densité $f(u)$ et dans chacune de ces unités, un échantillon de résidus des unités de premier niveau.

2.2 Cas non paramétrique (Comment obtenir une unité complète non paramétrique de Bootstrap?)

2.2.1 Utiliser les données complètes (y_{ij}, x_{ij})

On génère, avec remplacement, un échantillon aléatoire d'unités de deuxième niveau, qu'on numérote $1, 2, \dots, B$; dans chacune de ces unités on génère, avec remplacement un échantillon d'unités de premier niveau, dans les différents échantillons obtenus de cette façon, le nombre total d'unités de premier niveau (noté N) ne sera pas le même. Cependant cette procédure permet de maintenir la structure des données d'origine.

Une autre possibilité est de prendre directement un échantillon d'unités de premier niveau, de la taille nécessaire et les trier dans leurs unités de deuxième niveau, ceci mène à un nombre variable d'unités de deuxième niveau, et dans celles-ci un nombre variable d'unités de premier niveau, mais le nombre total d'unités de premier niveau sera le même.

2.2.2 Utiliser les résidus estimés (\hat{U}, \hat{e})

Par la méthode non paramétrique, on utilise les résidus estimés après les avoir centré pour assurer une moyenne nulle; la procédure suivante conserve la structure de l'échantillon:

1. Générer avec remplacement les résidus de deuxième niveau :

A partir de $\hat{U} = (\hat{u}_1, \hat{u}_2, \dots, \hat{u}_J)$, on obtient l'échantillon de Bootstrap $U^* = (u_1^*, u_2^*, \dots, u_J^*) = (\hat{u}_{l_1}, \hat{u}_{l_2}, \dots, \hat{u}_{l_J})$ $l_k = 1, \dots, J; k = 1, \dots, J$.

2. Pour chaque unité de deuxième niveau, on génère avec remplacement le nombre nécessaire de résidus de premier niveau, ces derniers doivent correspondre à l'unité de deuxième niveau de Bootstrap et le nombre nécessaire est le nombre d'unités de premier niveau dans cette unité de deuxième niveau.

On utilise $\{e_{ij}\}_{i=1, \dots, n_j; j=1, \dots, J}$ pour générer $\{e_{ij}^*\}$

Pour chaque unité de deuxième niveau j , on génère:

$$e_j^* = (\hat{e}_{i_1 l_j}, \hat{e}_{i_2 l_j}, \dots, \hat{e}_{i_{n_j} l_j}) = (e_{1j}^*, e_{2j}^*, \dots, e_{n_j j}^*)$$

Notons que dans certains cas, ceci voudra dire générer plus de résidus de premier niveau qu'il n'y a actuellement de résidus associés à l'unité de deuxième niveau choisie.

2.2.3 Algorithme itératif de Bootstrap pour l'estimation du modèle

1. Ajuster le modèle (4.7), calculer l'ensemble des résidus: $\{\hat{e}_{ij}\}_{i=1,\dots,n_j; j=1,\dots,J}$ et $\{\hat{u}_j\}_{j=1,\dots,J}$
2. Générer avec remplacement à partir de ces deux ensembles les deux nouveaux ensembles $\{e_{ij}^*\}_{i=1,\dots,I_j; j=1,\dots,J}$ et $\{u_j^*\}_{j=1,\dots,J}$.
3. L'ensemble des données de Bootstrap est alors (y_{ij}^*, x_{ij}) où

$$y_{ij}^* = \hat{\beta}_0 + \hat{\beta}_1 x_{ij} + u_j^* + e_{ij}^*$$

Un grand nombre B d'ensembles de données de Bootstrap est généré, le modèle supposé est ajusté à chacun de ces ensembles. Ainsi on obtient B estimateurs de Bootstrap pour chaque paramètre, qu'on peut utiliser pour estimer le biais, l'erreur standard, ou les intervalles de confiances.

2.3 Cas paramétrique

Algorithme itératif de Bootstrap pour l'estimation du modèle

Supposons qu'on a ajusté le modèle (4.7) et estimé ses paramètres, la méthode paramétrique de Bootstrap simule:

1. $e_{ij}^* \sim N(0, \hat{\sigma}_e^2)$ ($i = 1, 2, \dots, I_j; j = 1, \dots, J$) ou $\hat{\sigma}_e^2$ est l'estimateur de $\sigma_e^2 = \text{Var}(e_{ij})$ obtenu à partir des données initiales.
2. $u_j^* \sim N(0, \hat{\sigma}_u^2)$ ($j = 1, \dots, J$) ou $\hat{\sigma}_u^2$ est l'estimateur de $\sigma_u^2 = \text{Var}(u_j)$ obtenu à partir des données initiales.
3. L'ensemble des données de Bootstrap est alors $\{(y_{ij}^*, x_{ij})\}_{i=1,\dots,I_j; j=1,\dots,J}$ où:

$$y_{ij}^* = \hat{\beta}_0 + \hat{\beta}_1 x_{ij} + u_j^* + e_{ij}^*$$

3. l'estimation des intervalles de confiance pour les paramètres fixes et aléatoires

On estime les intervalles de confiance pour les paramètres fixes et aléatoires du modèle des composants de la variance (1.3).

Comme dans les modèles à un niveau, on peut utiliser les résidus estimés pour vérifier les hypothèses du modèle. Les deux hypothèses particulières qui peuvent être étudiées sont celles de normalité et que les variances dans le modèle sont constants.

Quand on construit un intervalle de confiance ou un test d'hypothèse pour chaque paramètre, il convient d'utiliser les estimateurs de ces derniers ainsi que les erreurs standards. Cependant dans bien des cas, on s'intéresse à des combinaisons de paramètres, pour les tests d'hypothèse, ce cas se présente pour les variables explicatives groupées : quand n effets sont définis en terme de $(n-1)$ effets et on souhaite tester simultanément si les effets sont nuls.

3.1. Les intervalles de confiance

L'estimation de l'écart type (standard error) est souvent utilisée pour obtenir un intervalle de confiance approximé pour un paramètre d'intérêt θ .

Supposons un échantillon aléatoire $X = (x_1, x_2, \dots, x_n)$ d'une fonction de distribution F inconnue, et soit $\hat{\theta} = t(\hat{F})$ un estimateur d'un paramètre d'intérêt $\theta = t(F)$ et soit $s\hat{e}$ l'estimateur de l'écart type de $\hat{\theta}$, basé sur les calculs de Bootstrap (par exemple), quand la taille de l'échantillon n augmente, la loi de $\hat{\theta}$ tend vers la loi normale:

$$\hat{\theta} \xrightarrow[n \rightarrow \infty]{} N(\theta, s\hat{e}^2)$$

Est de façon équivalente:

$$\frac{\theta - \hat{\theta}}{s\hat{e}} \xrightarrow[n \rightarrow \infty]{} N(0,1) \quad (4.8)$$

Soit $Z^{(\alpha)}$ le 100. α ième percentile de la $N(0,1)$, à partir de la table de la normale standard $Z^{(.025)} = -1.960$ et $Z^{(.975)} = 1.960$.

Si on considère le résultat asymptotique (4.8) alors :

$$prob_F \left\{ Z^{(\alpha)} \leq \frac{\hat{\theta} - \theta}{s\hat{e}} \leq Z^{(1-\alpha)} \right\} = 1 - 2\alpha \quad \text{qui peut aussi être écrite comme:}$$

$$prob_F \left\{ \theta \in \left[\hat{\theta} - Z^{(1-\alpha)} \cdot s\hat{e}, \hat{\theta} + Z^{(\alpha)} \cdot s\hat{e} \right] \right\} = 1 - 2\alpha ,$$

On dit que l'intervalle $\left[\hat{\theta} - Z^{(1-\alpha)} \cdot s\hat{e}, \hat{\theta} + Z^{(\alpha)} \cdot s\hat{e} \right]$ est l'intervalle de confiance de niveau (coverage probability) $1 - 2\alpha$ et comme $Z^{(\alpha)} = -Z^{(1-\alpha)}$ on peut écrire l'intervalle de confiance sous la forme:

$$\hat{\theta} \pm Z^{(1-\alpha)} \cdot s\hat{e} .$$

3.2. Les intervalles de confiance basés sur le Bootstrap Efron, B and Tibshirani, R. J(1993)

3.2.1 IC basé sur la "table" de Bootstrap

Par la méthode de Bootstrap on peut obtenir des intervalles de confiance précis sans avoir à considérer la tendance asymptotique vers la loi normale. Le principe de la procédure est d'estimer la répartition de Z directement à partir des données: on construit une table telle que celle de la normale standard ou de Student, mais qui n'est valable que pour notre ensemble de données. Pour construire cette table, appelée table de Bootstrap, on génère B échantillons de Bootstrap et pour chacun de ces B échantillons on calcule la version Bootstrap de Z la table de Bootstrap consiste en les percentiles de ces B valeurs.

Ainsi on génère B échantillons indépendants de Bootstrap $X^{*1}, X^{*2}, \dots, X^{*B}$, chacun est un ensemble de valeurs tirés des n valeurs de X avec remplacement, comme en (1) ou (2).

Pour chaque échantillon X^{*b} on calcule:

$$Z^*(b) = \frac{\hat{\theta}^*(b) - \hat{\theta}}{s\hat{e}^*(b)}$$

Où $\hat{\theta}^*(b)$ est la valeur de $\hat{\theta}$ basé sur l'échantillon X^{*b} et $s\hat{e}^*(b)$ l'estimateur de son erreur standard. Le percentile α de $Z^*(b)$ est estimé par la valeur $\hat{t}^{(\alpha)}$ tel que:

$$nb \{ Z^*(b) \leq \hat{t}^{(\alpha)} \} / B = \alpha.$$

Par exemple si $B=1000$, le point 5% estimé est la 50^{ième} plus grande valeur des $Z^*(b)$ et le point 95% estimé est la 950^{ième} plus grande valeur des $Z^*(b)$ enfin l'intervalle de confiance de t_Bootstrap est :

$$\left(\hat{\theta} - \hat{t}^{(1-\alpha)} \cdot s\hat{e}, \hat{\theta} - \hat{t}^{(\alpha)} \cdot s\hat{e} \right) \quad (4.9)$$

Remarques

1. Si $B\alpha$ n'est pas un entier, la procédure suivante peut être utilisée: En supposant $\alpha < .5$, soit $K = [(B+1)\alpha]$ le plus grand entier $\leq (B+1)\alpha$; les quantiles empiriques α et $1-\alpha$ sont respectivement la K ^{ième} et la $(B+1+K)$ ^{ième} plus larges valeurs des $Z^*(b)$. Cette procédure est

basée sur l'hypothèse que la répartition de la quantité $Z = (\hat{\theta} - \theta) / \hat{s\hat{e}}$ est approximativement la même pour chaque valeur de θ ; c'est la propriété qui nous permet de construire l'intervalle (4.9).

2. les IC basés sur les tables de la normale et de Student sont symétriques, ce qui n'est pas le cas pour ce de t_Bootstrap. ces derniers sont donc plus longs sur la gauche ou sur la droite. cette asymétrie représente une part importante de l'amélioration du niveau de confiance.

3. cette procédure est très utile et constitue une généralisation intéressante de la méthode usuelle de la table de Student. Elle est particulièrement applicable aux: moyenne de l'échantillon, médiane, le percentile (point de pourcentage).

4. la méthode t_Bootstrap, au moins sous sa forme simple n'est pas très appropriée dans le cas des problèmes plus généraux, comme l'obtention d'un IC pour le coefficient de corrélation.

3.2.2 IC basé sur les points de pourcentage

Définitions

1. Le point de pourcentage 100α d'une variable aléatoire x est la valeur z_α telle que : $P(x \geq z_\alpha) = 1 - F(z_\alpha) = \alpha$ où F est la fonction de répartition de la va x .

2. La fonction de distribution cumulée G d'une va x est :

$$G(x) = \frac{\text{le nb de valeurs } \leq x}{\text{le nb total de valeurs}}$$

3. Etant donné B réplifications de $\hat{\theta}^*$ et \hat{G} la fonction de répartition cumulée de $\hat{\theta}^*$, l'intervalle de confiance de niveau $(1 - 2\alpha)\%$ est défini par les points de pourcentage α et $(1 - \alpha)$ de \hat{G} :

$$\left[\hat{G}^{-1}(\alpha), \hat{G}^{-1}(1 - \alpha) \right] = \left[\hat{\theta}^{*(\alpha)}, \hat{\theta}^{*(1-\alpha)} \right]$$

Par définition $\hat{G}^{-1}(\alpha) = \hat{\theta}^{*(\alpha)}$ le point de pourcentage 100α de la distribution de Bootstrap.

Comment obtenir cet intervalle de confiance?

1. Générer B ensembles de données de Bootstrap indépendants: $X^{*1}, X^{*2}, \dots, X^{*B}$.

2. Pour chaque ensemble X^{*b} $b = 1, 2, \dots, B$ calculer la réplification

$$\hat{\theta}^*(b) = s(X^{*b}).$$

Soit $\hat{\theta}_B^{*(\alpha)}$ le point de pourcentage empirique d'ordre 100α des valeurs $\hat{\theta}^*(b)$, $b = 1, 2, \dots, B$; c'est la valeur n° $B\alpha$ de la liste ordonnée des B réplifications de $\hat{\theta}^*$; si $B = 2000$ et $\alpha = 0.05$; $\hat{\theta}_B^{*(\alpha)}$ est la 100^{ième} valeur des réplifications.

De même, soit $\hat{\theta}_B^{*(1-\alpha)}$ le point de pourcentage empirique d'ordre $100.(1-\alpha)$ des valeurs $\hat{\theta}^*(b)$, $b=1,2,\dots,B$.

L'intervalle de confiance de niveau $(1-2\alpha)\%$ est :

$$\left[\hat{\theta}_B^{*(\alpha)} \quad \hat{\theta}_B^{*(1-\alpha)} \right] \quad (4.10)$$

si la distribution de Bootstrap des $\hat{\theta}$ approche la normale, les intervalles de confiance basés sur le point de pourcentage et ceux basés sur la loi normale sont presque les mêmes. Le théorème central limit "dit que" quand n tend vers l'infini, l'histogramme de Bootstrap aura une forme normale mais quand l'échantillon est de petite taille; l'histogramme est très différent de la normale.

Remarques

1. Si $B.\alpha$ n'est pas un entier, on utilise la même procédure que dans le cas des intervalles de confiance basés sur la table de Bootstrap.

2. Dans le cas des petits échantillons, les intervalles de confiance basés sur le point de pourcentage sont préférés aux intervalles standards; ces derniers nécessitent une transformation appropriée du paramètre d'intérêt θ .

3. La difficulté dans l'estimation des intervalles de confiance standards réside dans le fait qu'on doit connaître une transformation différente pour chaque paramètre. La méthode du point de pourcentage peut être vu comme un algorithme qui incorpore automatiquement de telles transformations.

Le résultat suivant exprime le fait que la méthode du point de pourcentage connaît toujours la transformation appropriée.

Lemme

Supposons la transformation $\hat{\phi} = m(\hat{\theta})$ qui normalise parfaitement la distribution de $\hat{\theta}$:

$$\hat{\phi} \rightarrow N(\phi, c^2) \quad (4.11)$$

Alors l'intervalle de confiance du paramètre basé sur le point de pourcentage θ est:

$$\left[m^{-1}(\hat{\phi} - Z^{(1-\alpha)}.c) \quad m^{-1}(\hat{\phi} - Z^{(\alpha)}.c) \right]$$

Sous l'hypothèse que $\hat{\phi} = m(\hat{\theta})$ et $\phi = m(\theta)$ satisfont (4.11); ce lemme montre que la méthode de point de pourcentage transforment correctement les bornes de l'intervalle de confiance.

la méthode de point de pourcentage utilise $\hat{G}^{-1}(\alpha)$ comme borne inférieure et $\hat{G}^{-1}(1-\alpha)$ comme borne supérieure de l'intervalle de confiance. L'approche t_Bootstrap estime la distribution d'un pivot $Z = (\hat{\theta} - \theta) / \hat{s}\hat{e}$ puis inverse ce pivot pour obtenir un intervalle de confiance. Si on utilise cette approche pour obtenir un intervalle de confiance basé sur $\hat{\theta} - \theta$, ainsi le dénominateur du pivot =1, on obtiendra l'intervalle :

$$\left[2\hat{\theta} - \hat{G}^{-1}(1-\alpha), \left[2\hat{\theta} - \hat{G}^{-1}(\alpha) \right] \right]$$

il y a d'autres façons par lesquelles les intervalles standards peuvent échouer, en outre, la non normalité, par exemple $\hat{\theta}$ peut être biaisé, mais suit la normale, ainsi :

$$\hat{\theta} \rightarrow N(\theta + \text{biais}, \hat{s}\hat{e}^2) \quad (4.12)$$

dans quel cas, aucune autre transformation ne peut être appropriée, on introduit dans ce qui suit une extension de cette méthode de point de pourcentage qui prend en considération le biais et les transformations et une autre extension permettant à l'erreur standard de (4.12) de varier avec θ , cette extension a un avantage théorique très important.

Un des principaux objectifs de la théorie de Bootstrap, est de produire automatiquement de "bons" intervalles de confiance, c à d des intervalles de confiance qui approchent le plus les intervalles de confiance exacts dans le cas où la théorie statistique mène à une réponse exacte, et donnent des niveaux de confiance (*coverage probability*) précis dans tous les cas ; cependant, aucune des deux méthodes introduites ne satisfait ces critères.

On introduit une version améliorée de la méthode de point de pourcentage appelée BCa abréviation de (*bias corrected and accelerated*) . Les intervalles BCa sont une amélioration substantielle de la méthode du point de pourcentage en théorie et en pratique.

3.2.3 La méthode BCa

Soit $\hat{\theta}^{*(\alpha)}$ le point de pourcentage d'ordre $100.\alpha$ des B répliquions de Bootstrap $\hat{\theta}^*(1), \hat{\theta}^*(2), \dots, \hat{\theta}^*(B)$. L'intervalle de point de pourcentage de niveau de confiance $1-2\alpha$ est directement obtenu :

$$I_{pp} = \left(\hat{\theta}^{*(\alpha)}, \hat{\theta}^{*(1-\alpha)} \right)$$

Exemple

Si $B=2000$ et $\alpha=.05$ l'intervalle de confiance est : $I_{pp} = (\hat{\theta}^{*(.05)}, \hat{\theta}^{*(.95)})$ où $\hat{\theta}^{*(.05)}$ est la 100^{ième} valeur de la liste ordonnée des $\hat{\theta}^*(b)$, $b=1, \dots, 2000$ et $\hat{\theta}^{*(.95)}$ la 1900^{ième} valeur dans cette même liste.

Les bornes d'un intervalle de confiance BCa sont aussi données par les points de pourcentage, ceux-ci dépendent de deux nombres:

- \hat{a} appelé accélération
- \hat{z}_0 appelé correction du biais.

Définition d'un intervalle de confiance BCa

Un intervalle BCa de niveau de confiance $1-2\alpha$ est donné par:

$$BCa = (\hat{\theta}^{*(\alpha_1)}, \hat{\theta}^{*(\alpha_2)})$$

Où:

$$\alpha_1 = \Phi \left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(\alpha)}}{1 - \hat{a}(\hat{z}_0 + z^{(\alpha)})} \right)$$

$$\alpha_2 = \Phi \left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(1-\alpha)}}{1 - \hat{a}(\hat{z}_0 + z^{(1-\alpha)})} \right)$$

où $\Phi(\cdot)$ est la distribution cumulée de la normale standard et $z^{(\alpha)}$ est le point de pourcentage 100. α de la distribution de la $N(0, 1)$.

Remarque

Le cas où $\hat{a}=0$ et $\hat{z}_0=0$ nous ramène à intervalle de point de pourcentage.

Comment calculer \hat{z}_0 et \hat{a} ?

La correction du biais est directement obtenue à partir de la proportion des répliques de Bootstrap inférieures à l'estimateur d'origine $\hat{\theta}$,

$$\hat{z}_0 = \Phi^{-1} \left(\frac{\#\{\hat{\theta}^*(b) < \hat{\theta}\}}{B} \right) \quad (4.13)$$

Où $\Phi^{-1}(\cdot)$ est la fonction inverse de la fonction de distribution de la $N(0, 1)$.

Remarque

On obtient $\hat{z}_0=0$ si exactement la moitié des valeurs $\hat{\theta}^*(b)$ sont inférieures ou égales à $\hat{\theta}$.

La façon la plus facile à expliquer pour calculer \hat{a} est donnée en termes des valeurs de Jackknife d'une statistique $\hat{\theta} = s(X)$;

Soit $X_{(i)}$, l'échantillon d'origine X en omettant le point x_i ,

Soit $\hat{\theta}_{(i)} = s(X_{(i)})$

Et on définit $\hat{\theta}_{(.)} = \sum_{i=1}^n \hat{\theta}_{(i)} / n$;

Une simple expression pour l'accélération est:

$$\hat{a} = \frac{\sum_{i=1}^n (\hat{\theta}_{(.)} - \hat{\theta}_{(i)})^3}{6 \left\{ \sum_{i=1}^n (\hat{\theta}_{(.)} - \hat{\theta}_{(i)})^2 \right\}^{3/2}} \quad (4.13)$$

3.3. Les intervalles de confiance pour les paramètres d'un modèle multi-niveaux

3.3.1 Les paramètres fixes Goldstein,H (1996)

Soit C la matrice des contraintes ($r \times p$), elle est utilisée pour former les fonctions linéairement indépendantes des p paramètres fixes du modèle. On aura $f = C\beta$ telle que chaque ligne de C définit une fonction linéaire particulière. Les paramètres qui ne sont pas concernés par le test ont l'ensemble des éléments correspondants nuls.

Supposons qu'on a à tester l'hypothèse $H_0 : f=k$, on construit alors

$$R = (\hat{f} - k)' [C(X^t \hat{V}^{-1} X) C^t] (\hat{f} - k) \quad \hat{f} = C\hat{\beta}$$

Sous l'hypothèse nulle $R \longrightarrow \chi_r^2$.

Notons que $X^t \hat{V}^{-1} X$ est l'estimateur de la matrice de covariance des coefficients fixes. Le test est alors fait en comparant R calculé aux valeurs critiques de $\chi_{r,\alpha}^2$.

On peut aussi obtenir une région de confiance des paramètres fixes :

$$\hat{R} = (f - \hat{f})' [C(X^t \hat{V}^{-1} X) C^t] (f - \hat{f})$$

une région de taille $\alpha\%$ de loi χ_r^2

Ceci produira une fonction quadratique des coefficients estimés, formant une région ellipsoïdale de dimension r . si on s'intéresse à des intervalles de confiance séparés pour toutes les fonctions linéaires possibles invoquant q paramètres ou q fonctions linéaires indépendantes des paramètres. Pour un intervalle de $(1-\alpha)\%$, soit C_i la i ème ligne de C ; l'intervalle simultané pour $C_i\beta$ est donné par :

$$\left(C_i\hat{\beta} - d_i, C_i\hat{\beta} + d_i \right) \text{ ou } d_i = \left[C_i \left(X' \hat{V}^{-1} X \right)^{-1} C_i' \chi_{q,\alpha}^2 \right]^{1/2}$$

3.3.2 Les paramètres aléatoires Goldstein,H (1996)

On utilisera les hypothèses initiales concernant les lois, à partir des lois supposées normales ; on simule des valeurs pour obtenir des ensembles d'estimateurs de Bootstrap.

En effet, pour générer un échantillon par la méthode de Bootstrap, on procède comme suit :

A partir de $N(0, \sigma_u^2)$, on génère un ensemble de valeurs u_j^* , et à partir de $N(0, \sigma_e^2)$, on génère un ensemble de valeurs e_{ij}^* , et qu'on additionne à $(X\beta)_{ij}$; ce qui donnera un ensemble de pseudo valeurs y_{ij}^* qui sera traité comme un ensemble de variables dépendantes et à partir desquelles on peut obtenir un nouvel ensemble de valeurs $\hat{\beta}^*, \hat{\sigma}_u^{2*}, \hat{\sigma}_e^{2*}$

Une fois l'ensemble des valeurs de Bootstrap est obtenu, on peut l'utiliser pour estimer les matrices de variances et de covariance des paramètres ou les erreurs standards, et sous l'hypothèses de normalités on peut aussi obtenir des intervalles de confiance pour ces estimateurs ou pour des fonctions de ces derniers

La méthode non paramétrique consiste à utiliser les percentiles (point de pourcentage) de l'ensemble des valeurs empiriques de Bootstrap ; quand la valeur de la médiane d'un paramètre ou d'une fonction de paramètres dévie de l'estimateur du paramètre, une procédure de correction doit être utilisée.

3.3.3. Les résidus

Ces résidus peuvent avoir deux rôles, leur interprétation de base est comme des variables aléatoires avec une distribution dont les valeurs des paramètres expliquent la variation au niveau des unités de deuxième niveau et qui produit des estimateurs efficaces pour les coefficients fixes .La seconde interprétation est comme des estimateurs individuels pour chaque unité de niveau 2 ou on utilise l'hypothèse qu'elles appartiennent à une population d'unités. En particulier, pour les

unités qui ont un petit nombre d'unités de niveau 1 on peut obtenir des estimateurs plus précis que si on ignore l'appartenance à une population et on utilise seulement l'information tirée de ces unités

Ceci devient important pour les estimateurs des résidus et des coefficients aléatoires et dans le cas extrême d'une seule unité de niveau 1 dans chaque unité de niveau 2, on manque alors d'information pour calculer un estimateur indépendant.

Pour obtenir des estimateurs de Bootstrap pour les résidus de 2^{ème} niveau, on procède comme suit :

Pour chaque échantillon de Bootstrap on estime aussi les résidus \hat{u}_j^* ; ainsi pour estimer la variance « comparative » des résidus pour chaque unité de niveau 2, on utilise $\tilde{u}_j^* = \hat{u}_j^* - u_j^*$ pour estimer la variance ou la matrice de covariance dans le cas où il existe plusieurs paramètres aléatoires. On peut aussi les utiliser pour construire des intervalles de confiance non paramétrique.

Implémentation dans le cas du modèle des composants de la variance à deux niveaux

On utilisera comme illustration le modèle des composants de la variance à deux niveaux et l'ensemble des données Aitkin et Longford (1986);

$$y_{ij} = (X\beta)_{ij} + u_j + e_{ij}$$

L'implémentation informatique des méthodes d'estimation du modèle des composants de la variance à deux niveaux :

- La méthode itérative des moindres carrés généralisée IMCG
- La méthode paramétrique de Bootstrap
- La méthode non paramétrique de Bootstrap

a été réalisée en langage MATLAB 7.0 sous WINDOWS XP Edition familiale Version2002, le matériel utilisé est un TOSHIBA Guenuine Intel CPU à 1.66GHz, 0.99Go de RAM .

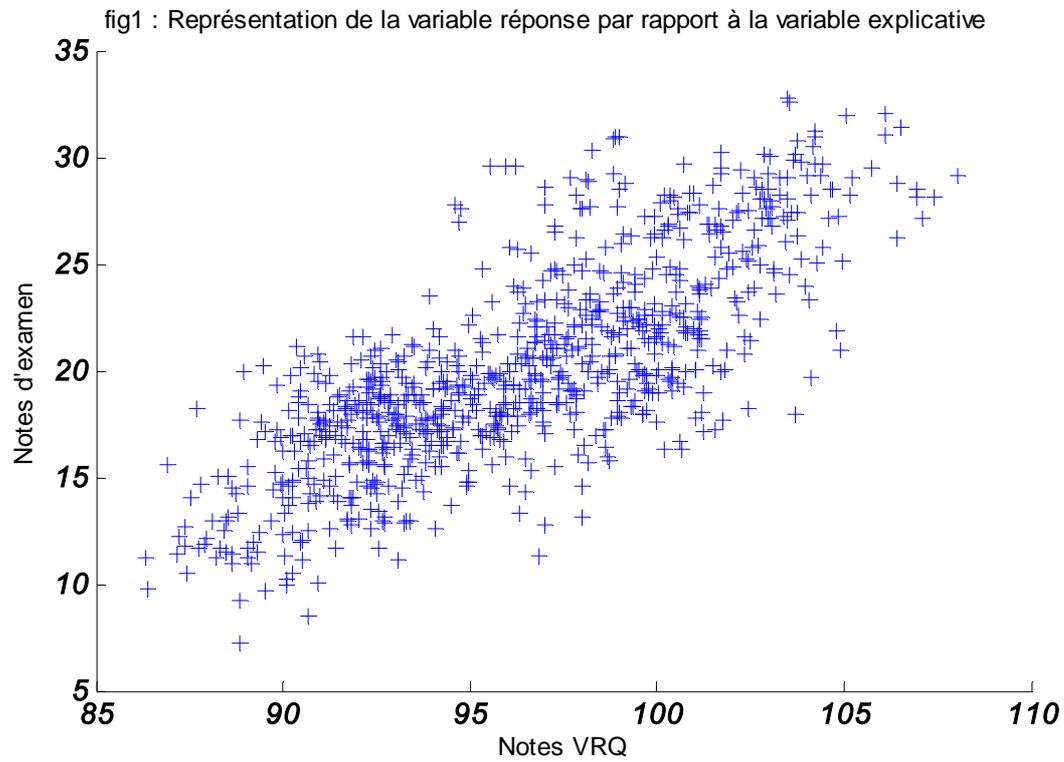
Les données sont des notes obtenues par 864 étudiants de 16 établissements polyvalents d'enseignement secondaires (Secondary comprehensive school); ces 16 établissements sont sous une seule tutelle appelée English local education authority (LEA). La variable réponse est une note d'examen de chaque étudiant; chaque étudiant a des notes pour chaque épreuve de l'examen de certificat général d'éducation ou de certificat de l'examen secondaire d'éducation passé à l'âge de 16 ans ; et la variable explicative est une note obtenue par l'étudiant dans un test de raisonnement verbal (Verbal reasoning test), et cela à l'âge de 11ans; c'est-à-dire juste avant qu'il n'entre à l'école secondaire; on appellera cette variable la VRQ (Verbal Reasoning Quotient).

La table 1 montre les moyennes et les écart-types des notes d'examens et des VRQ pour les 16 écoles;

Table 1 les moyennes et les écart-types des notes d'examens et des VRQ pour les 16 établissements					
ECOLE	Nombre d'étudiants	variable dépendante notes d'examen		variable indépendante notes de VRQ	
		Moyenne	ecart-type	Moyenne	ecart-type
1	65	26,9	1,98	102,1	1,48
2	79	18,7	1,54	92,6	1,33
3	48	28,5	1,93	104,5	1,50
4	47	21,1	2,07	101,4	1,82
5	66	17,5	1,83	92,2	1,39
6	41	11,5	1,89	89,0	1,55
7	52	14,3	1,64	91,8	1,84
8	67	21,2	1,62	99,1	1,47
9	49	16,3	2,13	91,7	1,94
10	47	26,5	1,95	98,6	1,85
11	50	20,3	1,71	98,8	1,88
12	41	22,6	3,01	98,1	1,95
13	49	16,8	1,99	93,7	2,01
14	29	22,6	2,92	99,1	2,25
15	72	18,0	1,52	95,1	1,50
16	62	19,2	1,64	96,3	1,60

A partir de cette table on voit facilement qu'il y a de grandes différences entre les notes d'examens (variable réponse) et qu'elles sont liées aux VRQ. Les données utilisées sont simulées à partir de cette table.

La figure 1 représente les notes d'examens contre les VRQ montrant l'existence d'une relation linéaire



Les résultats de l'estimation par la méthode IMCG et la méthode des Moindres Carrés Ordinaires sont présentés dans la table 2

Table 2 : Estimateurs des paramètres du modèle des composants de la variance à deux niveaux		
Paramètres	IMCG	MCO
fixes		
Intercept β_0	16,607	16.769
Note VRQ β_1	0,0352	0.0336
aléatoires		
Niveau 2 σ_u^2	3,7373	
Niveau 1 σ_e^2	18,732	
Corrélation intra école ρ	0,1663	

Prendre en considération la structure hiérarchique des données permet de décomposer la variance ce qui permet une estimation de la corrélation intra-école, celle-ci mesure la proportion de la variance totale entre les unités de premier niveau (les étudiants) due à la variation entre les unités de niveau 2 (les écoles).

La table 3 montre les résultats de la méthode paramétrique de Bootstrap, ainsi que ceux de la méthode paramétrique ; dans le cas non paramétrique, la simulation est faite à partir des paramètres aléatoires du modèle.

Table 3 : Estimateurs des paramètres du modèle des composants de la variance à deux niveaux par la méthode paramétrique de Bootstrap et la méthode non paramétrique			
Paramètres	Modèle estimé	paramétrique	Non paramétrique
fixes			
Intercept β_0	16,607	16.560	13,074
Note VRQ β_1	0,0352	0.0357	0,0736
aléatoires			
Niveau 2 σ_u^2	3,7373	3.7331	3,8287
Niveau 1 σ_e^2	18,732	17.457	18,394
Corrélation intra école ρ	0,1663	0.1761	0,1722

Les estimations non paramétriques sont moins précises que celles paramétriques, ceci est dû aux échantillons de Bootstrap utilisés pour chaque cas ; dans le cas paramétrique les échantillons générés à partir de la loi normale ont toujours la même structure, par contre dans le cas non paramétrique les échantillons obtenus ne sont pas très représentatifs par rapport à l'ensemble de données initial ; dans certains cas on génère plus d'unités de niveau 1 qu'il n'y a actuellement d'unités dans l'unité de deuxième niveau choisie, même si les échantillons ainsi obtenus respectent la structure hiérarchique des données initiales mais il contiennent souvent des mesures répétées.

Conclusion

Une analyse qui modélise la structure hiérarchique des données présente de nombreux avantages; elle permet d'obtenir des estimateurs efficaces des coefficients de régressions; en considérant leur structure on a des erreurs standards et intervalles de confiance plus corrects que ceux obtenus en ignorant cette structure; aussi en utilisant les covariances à chaque niveau de la hiérarchie, cette analyse permet d'explorer les différences existantes entre les valeurs de la variable réponse et quelles sont les variables explicatives qui y participent .

Des chercheurs dans le domaine de l'éducation et de l'enseignement se sont intéressés à faire des comparaisons entre les écoles et entre d'autres institutions, très souvent en utilisant les résultats obtenus par les étudiants : ces études peuvent avoir pour objectif la comparaison entre les écoles publiques et autres (Goldstein 1992) mais les chercheurs se sont surtout intéressés à savoir quels sont les facteurs (variables indépendantes) qui expliquent les différences entre les écoles.

Nous avons présenté une méthode d'estimation pour des modèles hiérarchiques à deux niveaux qui est la méthode itérative des moindres carrés généralisée ainsi que deux autres alternatives :

- Méthode non paramétrique de Bootstrap
- Méthode paramétrique de Bootstrap

La méthode non paramétrique présente l'inconvénient qu'il existe plusieurs façons d'obtenir les échantillons certaines détruisent la structure de l'ensemble des données mais permet de réduire la taille des intervalles de confiance. La méthode paramétrique n'est pas très appropriée quand les données ne sont pas vraiment normalement distribuées ou quand la taille de l'échantillon n'est pas assez grande pour utiliser le théorème Central Limit.

Une difficulté commune à ces trois méthodes est le calcul de l'inverse de matrices de très grande taille; connaissant la structure de ces matrices, certains résultats permettent de réduire ce calcul au calcul de l'inverse de plusieurs matrices de plus petite taille.

Ce travail est une introduction aux modèles linéaires multi-niveaux qui peut avoir plusieurs extensions telles que l'étude des valeurs aberrantes (outliers), les mesures répétées c'est-à-dire qu'une caractéristique d'un même individu ou unité est considérée plus d'une fois dans l'ensemble des données; aussi le cas où les données sont des variables discrètes. La structure multi-échelles est aussi étudiée dans le cas non linéaires et /ou multivarié.

Bibliographie

1. **Aitkin, M and Longford, N. (1986)** *Statistical modelling in school effectiveness studies* (with discussion). **J. Royal statist. Soc., A**, 149, 1-43.
2. **Carpenter, J and Goldstein, H and Rasbash, J** *A non-parametric Bootstrap for multilevel models* **Institute of Education**
www.ioe.ac.uk/multilevel
3. **Efron, B and Tibshirani, R. J (1993)***An introduction to the Bootstrap* **Chapman & Hall**
4. **Goldstein,H and Rasbash, J (1992).** *Efficient computational procedures for the estimation for parameters in multilevel models based on iterative generalised least squares.* **Computational Statistics and Data Analysis**, 13, 63-71.
5. **Goldstein, H** *Bootstrapping for multilevel models* **Multilevel Models Project Working paper** (december 1998)
www.ioe.ac.uk/multilevel
6. **Goldstein, H (1986)** *Multilevel mixed linear model analysis using iterative generalised least squares.* **Biometrika**, 73, 43-56
7. **Goldstein, H** *multilevel models*
www.ioe.ac.uk/multilevel
8. **Goldstein, H (1992)** *Statistical information and the measurement of education outcomes* (editorial). **J.Roy.Statist.Soc.,A** 155; 313-315
9. **Intrilligator, M, Bodkin.R and Hsiao.C** *Econometric Models techniques and applications* **Prentice Hall** Second Edition
10. **Redfern.D and Campbell.C (1998)** *The MATLAB 5 Handbook* **Springer**
11. **Searle, S. R, Casella, G. and McCulloch, C. E. (1992)** *Variance Components*. New York, **Wiley**.
12. **Weber, A** *Lecture Notes: Econometrics I* **Institute of Advanced studies** (December 2003)

Annexe 1

Nom du fichier : imcg.m

Procédure méthode itérative des moindres carrés généralisée

```
% "imcg.m(function file
%Méthode itérative des moindres carrés généralisée
%-----
function [b,sig,e,u,sg,tlb,tlb,tlb]=imcg(nc,nu,ntu,xs,ys)
%estimation des bi
b=inv(xs'*xs)*xs'*ys; tlb=b;
% calcul des erreurs
ye=ys-xs*b;sse=ye'*ye;
sg(1)=sse/(ntu-2);
%construction de la matrice z utilisée pour estimer les e et u
une=eye(nu(1),nu(1));
z= une(:,1);
for l=2 : nu(1)
    z=[z;une(:,l)];
end;
for c =2:nc
    une=eye(nu(c),nu(c));
    z1= une(:,1);
    for l=2 : nu(c)
        z1=[z1;une(:,l)];
    end;
    z=[z;z1];
end;
[sz, qq]=size(z);
uns=ones(sz,1);
z=[z,uns];
% calcule de l'estimateur des sigmas
yy=ones(nu(1),1);
for l=1 :nu(1)
    yy(l)=ye(l);
end;vs=ones(nu(1));
vs=yy*yy';uns=ones(nu(1));
for p=1:nu(1)
    for q=1 : nu(1)
        if vs(p,q)<0
            vs(p,q)=0;
        end;
    end;
end;
end;

s1=inv(vs)*inv(vs);s2=inv(vs)*uns*inv(vs);
s3=inv(vs)*vs*inv(vs);yz=s3(:,1);
vz=[s1(:,1),s2(:,1)];
for li=2 :nu(1)
    vz=[vz;s1(:,li),s2(:,li)];
    yz=[yz;s3(:,li)];
```

```
end;
for c= 2: nc
    %re initialisation
    yy=ones(nu(c),1);
    for ll=1 : nu(c)
        yy(ll)=ye(ll+1);
    end;
    %re initialisation
    vs=ones(nu(c),nu(c));
    vs=yy*yy';
    for p=1:nu(c)
        for q=1 : nu(c)
            if vs(p,q)<0
                vs(p,q)=0;
            end;
        end;
    end;
    uns=ones(nu(c));

    %re ininitialisation
    s1=ones(nu(c));
    s2=ones(nu(c));
    s3=ones(nu(c));
    s1=inv(vs)*inv(vs);
    s2=inv(vs)*uns*inv(vs);
    s3=inv(vs)*vs*inv(vs);
    vv=[s1(:,1),s2(:,1)];vy=s3(:,1);
    for li=2 :nu(c)
        vv=[vv;[s1(:,li),s2(:,li)]];
        vy=[vy;s3(:,li)];
    end;
    vz=[vz;vv];yz=[yz;vy];
    l=1+nu(c);
end;
sig=inv(z'*vz)*z'*yz; tlsig=sig;

% 1. construcion de la première matrice de variance covariance
v2=ones(nu(1));
for l= 1 : nu(1)
    for c= 1 : nu(1)
        if l==c
            v2(l,c)=sig(1)+sig(2);
        else v2(l,c)=sig(2);
        end;
    end;
end;

for n=2 : nc
    v=ones(nu(n));
```

```
for l= 1 :nu(n)
  for c= 1 : nu(n)
    if l==c
      v(l,c)=sig(1)+sig(2);
    else v(l,c)=sig(2);

    end;
  end;
end;
v2=blkdiag(v2,v);
end;
%répétition de la methode
conv=0; iter=2;
while conv==0
  %best=lscov(xs,ys,v2);
  best=inv(xs'*inv(v2)*xs)*xs'*inv(v2)*ys;tlb=[tlb, best];
  % calcul des erreurs
  ye=ys-xs*best;sse1=ye'*ye;
  sg(iter)=sse1/(ntu-2);
  yy=ones(nu(1),1);
  for l=1 :nu(1)
    yy(l)=ye(l);
  end;vs=ones(nu(1));vn=ones(nu(1));
  vs=yy*yy';uns=ones(nu(1));
  for l= 1 : nu(1)
    for c= 1 : nu(1)
      if l==c
        vn(l,c)=sig(1)+sig(2);
      else vn(l,c)=sig(2);
      end;
    end;
  end;
end;

s1=ones(nu(1));s1=s2;s3=s2;
s1=inv(vn)*inv(vn);s2=inv(vn)*uns*inv(vn);
s3=inv(vn)*vs*inv(vn);yz=s3(:,1);
vz=[s1(:,1),s2(:,1)];
for li=2 :nu(1)
  vz=[vz;[s1(:,li),s2(:,li)]];
  yz=[yz;s3(:,li)];
end;
for c= 2: nc
  %re initialisation
  yy=ones(nu(c),1);
  for ll=1 : nu(c)
    yy(ll)=ye(ll+1);
  end;
  %re initialisation
  vs=ones(nu(c));vn=ones(nu(c));
```

```

vs=yy*yy';
for p=1:nu(c)
    for q=1 : nu(c)
        if p==q
            vn(p,q)=sig(1)+sig(2);
        else vn(p,q)=sig(2);
        end;
    end;
end;
uns=ones(nu(c));
%re innitialisation
s1=ones(nu(c));
s2=ones(nu(c));
s3=ones(nu(c));
s1=inv(vn)*inv(vn);
s2=inv(vn)*uns*inv(vn);
s3=inv(vn)*vs*inv(vn);
vv=[s1(:,1),s2(:,1)];vy=s3(:,1);
for li=2 :nu(c)
    vv=[vv;[s1(:,li),s2(:,li)]];
    vy=[vy;s3(:,li)];
end;
vz=[vz;vv];yz=[yz;vy];
l=l+nu(c);
end;
sig1=inv(z'*vz)*z'*yz; tlsig=[tlsig,sig1];
dif=sig1-sig;diff=dif(1)^2+dif(2)^2;
dif1=best-b;diff1=dif1(1)^2+dif1(2)^2;
if (diff<.01) & (diff1<.01)
    conv=1;
else %construire la nouvelle matrice de variance covariance
    v2=ones(nu(1));
    for l= 1 : nu(1)
        for c= 1 : nu(1)
            if l==c
                v2(l,c)=sig1(1)+sig1(2);
            else v2(l,c)=sig1(2);
            end;
        end;
    end;
end;

for n=2 : nc
    v=ones(nu(n));
    for l= 1 :nu(n)
        for c= 1 : nu(n)
            if l==c
                v(l,c)=sig1(1)+sig1(2);
            else v(l,c)=sig(2);
            end;
        end;
    end;
end;

```

```
        end;
    end;
    v2=blkdiag(v2,v);
    end;
    sig=sig1;b=best;iter=iter+1;
end;
end;

%estimation des résidus
l=0;
for c= 1: nc
    %re initialisation
    yy=ones(nu(c),1);
    for ll=1 : nu(c)
        yy(ll)=ye(ll+1);
    end;
    moy=mean(yy);u(c)=(nu(c)*sig(2))*moy/((nu(c)*sig(2))+sig(1));
    for i=1: nu(c)
        e(i,c)=yy(i)-u(c);
    end;
    l=l+nu(c);
end;
```

Annexe 2

Nom du fichier : Etape1.m

Programme de la méthode itérative des moindres carrés généralisée

```
clear all; tic
%nc: nombre d'unités de niveau 2
nc=xlsread('classeur1','feuille1','A18')
%les nu nombres d'unités de niveau 1 dans chaque unité de niveau 2
nu=xlsread('classeur1','feuille1','B3:B18');
mey=xlsread('classeur1','feuille1','C3:C18');
sy=xlsread('classeur1','feuille1','D3:D18');
mex=xlsread('classeur1','feuille1','E3:E18');
sx=xlsread('classeur1','feuille1','F3:F18');
vd=cell(nc,1); vi=cell(nc,1);
for i=1 : nc
    vd{i}=sy(i)*randn(nu(i),1)+mey(i);
    vi{i}=sx(i)*randn(nu(i),1)+mex(i);
end
% nombre total d'unités
ntu=sum(nu);
ys=cell2mat(vd);
xx=cell2mat(vi);
xs=[ones(ntu,1) xx];
%pour une representation graphique
subplot(2,1,1)
y1=vd{1,1}; x1=vi{1,1};xr=[ones(nu(1),1) x1];
b=inv(xr*xr)*xr*y1;
l1=min(x1);l2=max(x1);
xabs= l1:.2:l2;yabs=b(1)+b(2)*xabs;
yest=xr*b;
plot( x1, y1, '+',xabs,yabs,'-'); xlabel('la variable indépendante');ylabel('la variable dépendante');title('FIG1')
subplot(2,1,2)
y1=ones(nu(10),1); x1=y1;
y1=vd{10,1}; x1=vi{10,1};xr=[ones(nu(10),1) x1];
b=inv(xr*xr)*xr*y1;
l1=min(x1);l2=max(x1);
xabs= l1:.2:l2;yabs=b(1)+b(2)*xabs;
yest=xr*b;
plot( x1, y1, '+',xabs,yabs,'-'); xlabel('la variable indépendante');ylabel('la variable dépendante');title('FIG2')

[b,sig,e,u,sg,tlb,tlsg]=imcg(nc,nu,ntu,xs,ys);
sig
sig(2)/(sig(1)+sig(2))
sig(1)/(sig(1)+sig(2))
tps=toc;
```

Annexe 3

Nom du fichier : Bnp1.m

Programme : estimation par la méthode non paramétrique de Bootstrap

```
%estimateur de Bootstrap non paramétrique utilisant (U,e)
nrb=500;tic
sigma=sig; bbn=b;
for repb=1 :nrb
    ub=ones(nc,1);
    for i=1 : nc
        tb=round((nc*rand));
        if tb==0 tb=1;
        elseif tb>nc tb=nc;
        end;
        ub(i,1)=u(1,tb);
        %à partir de l'unité tb on genere nu(1)
        ss=nu(tb);
        for j=1 :nu(i)
            sb=round((ss*rand)+1);
            if sb==0 sb=1;
            elseif sb>ss sb=ss;
            end;
            eb(j,i)=e(sb,tb);
        end;
        %centrer les eb
        eb(1:nu(i),i)=eb(1:nu(i),i)-(mean(eb(1:nu(i),i))*ones(nu(i),1));
        %corriger sa variance
        %j1=std(e(1:nu(i),tb));
        j1=sqrt(sig(1));j2=std(eb(1:nu(i),i));eb(1:nu(i),i)=(j1/j2)*eb(1:nu(i),i);
    end;
    %centrer les ub
    ub=ub-mean(ub)*ones(nc,1);
    %corriger sa variance
    %j1=std(u);
    j1=sqrt(sig(2));j2=std(ub);ub=(j1/j2)*ub;
    %costruire le nouvel ensemble ysb
    uy=ones(nu(1),1);erb=ones(nu(1),1);
    erb=eb(1:nu(1),1)+ub(1,1)*uy;
    for i=2 : nc
        erb1=ones(nu(i),1);
        uy=ones(nu(i),1);
        erb1=eb(1:nu(i),i)+ub(i,1)*uy;
        erb=[erb;erb1];
    end;
    ys=xs*b;
    %mys=mean(ys);sys=std(ys);
    %for i=1:ntu
    % ys(i)=(ys(i)-mys)/sys;
    %end;
```

```
ysb=ys+erb;
%ysb=ys+erb;
[bb,sigb,eb,ub,sg,tlb,tlsig]=imcg(nc,nu,ntu,xs,ysb);
sigma=[sigma,sigb];bbn=[bbn,bb];
end;
%mue=mean(sigma(1,:))
%mub=mean(sigma(2,:))
%bb1=mean(bbn(1,:))
%bb2=mean(bbn(2,:))
tpsbnp1=toc
```

Annexe 4

Nom du fichier : Bp2.m

Programme : estimation par la méthode paramétrique de Bootstrap

%estimateur de Bootstrap paramétrique

nrb=500;tic

for repb=1 :nrb

 %simuler les u

 ssu=sqrt(sig(2));

 ub=ssu*randn(nc,1);sse=sqrt(sig(1));

 for c=1:nc

 eb(1:nu(c),c)=sse*randn(nu(c),1) ;

 end;

 %costruire le nouvel ensemble ysb

 uy=ones(nu(1),1);erb=ones(nu(1),1);

 erb=eb(1:nu(1),1)+ub(1,1)*uy;

 for i=2 : nc

 erb1=ones(nu(i),1);

 uy=ones(nu(i),1);

 erb1=eb(1:nu(i),i)+ub(i,1)*uy;

 erb=[erb;erb1];

 end;

 ysb=xs*b+erb;

 [b,sigb,e,u]=imcg(nc,nu,ntu,xs,ysb);

 repb

 sigb

end;toc