

MATH 60604

Modélisation statistique

§ 6e - Pente aléatoire

HEC Montréal
Département de sciences de la décision

On considère un modèle linéaire mixte avec ordonnée à l'origine et pente aléatoire pour les données vengeance,

$$Y_i | \mathbf{B}_i = b_i \sim \text{No}_5 \left(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i b_i, \sigma^2 \mathbf{I}_5 \right)$$
$$\mathbf{B}_i \sim \text{No}_2(\mathbf{0}_2, \boldsymbol{\Omega})$$

où $\mathbf{Z}_i = [\mathbf{1}_5, \mathbf{t}_i]$ est une matrice 5×2 pour les effets aléatoires et $\boldsymbol{\Omega} = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{pmatrix}$.

Les colonnes de \mathbf{Z}_i incluent d'ordinaire les variables suivantes:

- temps
- indicateurs de variables binaires/catégorielles (effet de groupe).

Soit la matrice $\mathbf{Z}_i = [\mathbf{1}_{n_i}, \mathbf{X}_{1i}]$.

$$Y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})X_{ij1} + \beta_2 X_{ij2} + \cdots + \beta_p X_{ijp} + \varepsilon_{ij}.$$

- L'effet conditionnel de la variable X_1 pour le groupe i est $\beta_1 + b_{1i}$
- Le paramètre β_1 est la « pente » X_1 moyenne pour la population.
- $\beta_1 + b_{1i}$ est l'effet de X_1 spécifique au groupe i .

- La matrice de covariance de Y_i dépend des variables explicatives de \mathbf{Z}_i qui induisent un effet aléatoire.
- Par exemple, si $\mathbf{Z}_i = [\mathbf{1}_{n_i}, \mathbf{X}_{1i}]$, la variance marginale de Y_{ij} est

$$\text{Var}(Y_{ij} \mid \mathbf{X}_i) = \omega_{11} + X_{ij1}^2 \omega_{22} + 2X_{ij1} \omega_{12} + \sigma_\varepsilon^2.$$

- Si les aléas sont indépendants, la covariance entre deux observations d'un même groupe est

$$\text{Cov}(Y_{ij}, Y_{ik} \mid \mathbf{X}_i) = \omega_{11} + X_{ij1} X_{1ik} \omega_{22} + (X_{ij1} + X_{1ik}) \omega_{12}.$$

- Il peut être difficile d'estimer les paramètres si la structure de covariance de Y_i est complexe (en plus des coûts computationnels).

Code SAS pour ajuste un modèle avec pente aléatoire

```
proc mixed data=modstat.vengeance;  
class id;  
model vengeance = sexe age vc wom t / solution;  
random intercept t / subject=id type=un v=1 vcorr=1;  
run;
```

La sortie inclut des information sur le nombre de paramètres de covariance, le nombre d'effets aléatoires, etc.

Dimensions	
Paramètres de covariance	4
Colonnes dans X	6
Colonnes dans Z par sujet	2
Sujets	80
Max. obs. par sujet	5

Matrice de covariance de la réponse

Valeur estimée du paramètre de covariance		
Param. de cov.	Sujet	Estimation
UN(1,1)	id	0.3064
UN(2,1)	id	-0.05268
UN(2,2)	id	0.01730
Residual		0.2055

Matrice V estimée pour Subject 1					
Ligne	Col1	Col2	Col3	Col4	Col5
1	0.4239	0.1830	0.1476	0.1122	0.07682
2	0.1830	0.3704	0.1468	0.1287	0.1106
3	0.1476	0.1468	0.3515	0.1452	0.1444
4	0.1122	0.1287	0.1452	0.3672	0.1782
5	0.07682	0.1106	0.1444	0.1782	0.4175

- La variance de l'effet aléatoire sur l'ordonnée à l'origine est $\omega_{11} = 0,306$
- La variance de l'effet aléatoire sur la pente est $\omega_{22} = 0,017$
- La corrélation entre les deux effets aléatoires est $-0,72$.

- On peut tester si $\mathcal{H}_0 : \omega_{12} = 0$ versus $\mathcal{H}_a : \omega_{12} \neq 0$ en ajustant un modèle avec matrice de covariance de b_i diagonale et en faisant un test de rapport de vraisemblance (REML, car les effets fixes sont les même)
 - dans SAS, le modèle de covariance `type=vc` (option par défaut pour effets aléatoires).
 - la statistique de test est $R = 8,98$
 - sa loi nulle asymptotique est χ_1^2 (problème régulier, la covariance peut être négative)
 - la valeur- p est 0,002:
 - la corrélation entre effets aléatoires est fortement significative.

- On pourrait comparer avec le modèle qui inclut uniquement un effet aléatoire pour l'ordonnée à l'origine aléatoire
 - ce qui revient à tester $\mathcal{H}_0 : \omega_{22} = 0$, et la loi nulle de la statistique du rapport de vraisemblance est $\frac{1}{2}\chi_1^2$ si les effets aléatoires sont indépendants.
 - si les erreurs sont corrélés, la variance nulle force $\omega_{12} = 0$ et on « perd » un autre paramètre...
 - la loi nulle du test est compliquée, voir
Andrews, D.W. (2001), Testing when a parameter is on the boundary of the maintained hypothesis, Econometrica, 69 (3)
- L'approximation n'est pas très bonne ...la plupart des gens se rabattent sur les critères d'information.