MATH 60604
Modélisation statistique
§ 6e - Pente aléatoire

Léo Belzile

HEC Montréal
Département de sciences de la décision

Formulation du modèle

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

$$\boldsymbol{Y}_i \mid \boldsymbol{\mathcal{B}}_i = \boldsymbol{b}_i \sim \mathsf{No}_5\left(\mathbf{X}_i\beta + \mathbf{Z}_i\boldsymbol{b}_i, \sigma^2\mathbf{I}_5\right)$$
$$\boldsymbol{\mathcal{B}}_i \sim \mathsf{No}_2(\mathbf{0}_2, \boldsymbol{\Omega})$$

où $\mathbf{Z}_i = [\mathbf{1}_5, \mathtt{t}_i]$ est une matrice 5×2 pour les effets aléatoires et $\boldsymbol{\Omega} = \left(\begin{smallmatrix}\omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22}\end{smallmatrix}\right)$.
Les colonnes de $\mathbf{Z}_i$ incluent d'ordinaire les variables suivantes:

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

Effet aléatoire sur une variable explicative

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

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

- L'effet conditionel 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.

Covariance de la réponse

- La matrice de covariance de $\boldsymbol{Y}_i$ 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

$$\mathsf{Var}\left(Y_{ij} \mid \mathbf{X}_i\right) = \omega_{11} + \mathrm{X}_{ij1}^2\omega_{22} + 2\mathrm{X}_{ij1}\omega_{12} + \sigma_\varepsilon^2.$$

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

$$\mathsf{Cov}\left(Y_{ij}, Y_{ik} \mid \mathbf{X}_i\right) = \omega_{11} + \mathrm{X}_{ij1}\mathrm{X}_{1ik}\omega_{22} + (\mathrm{X}_{ij1} + \mathrm{X}_{1ik})\omega_{12}.$$

- Il peut être difficile d'estimer les paramètres si la structure de covariance de $\boldsymbol{Y}_i$ est complexe (en plus des coûts computationnels).

Code SAS pour ajuste un modèle avec pente aléatoire

```
proc mixed data=modstat.vengeance;
model vengeance = sexe age vc wom t
    / ddfm=kenwardroger 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 $-0{,}72$.

Tester la corrélation entre effets aléatoires

- On peut tester si $\mathscr{H}_0 : \omega_{12} = 0$ versus $\mathscr{H}_a : \omega_{12} \neq 0$ en ajustant un modèle avec matrice de covariance de $\boldsymbol{b}_i$ 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^2_1 (problème régulier, la covariance peut être négative)
  - la valeur-p est $0{,}002$:
  - la corrélation entre effets aléatoires est fortement significative.

Comparaison de modèles

- On pourrait comparer avec le modèle qui inclut uniquement un effet aléatoire pour l'ordonnée à l'origine aléatoire
  - ce qui revient à tester $\mathscr{H}_0 : \omega_{22} = \omega_{12} = 0$.
  - La loi nulle asymptotique est $\frac{1}{2}\chi^2_1 + \frac{1}{2}\chi^2_2$,
  - mais l'approximation est mauvaise en échantillon fini…mieux vaut utiliser un critère d'information.