MATH 60604
Modélisation statistique
§ 6d - Ordonnée à l'origine aléatoire

HEC Montréal
Département de sciences de la décision

Modèles avec ordonnée à l'origine aléatoire

Le modèle de régression linéaire mixte suivant n'a qu'une seule composante aléatoire (une ordonnée à l'origine par groupe). L'équation de la moyenne est

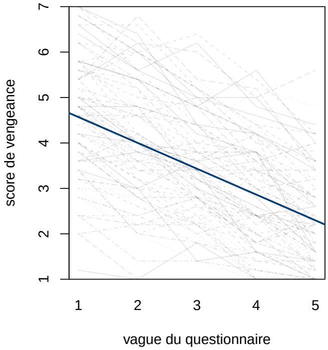
$$Y_{ij} = (\beta_0 + b_i) + \beta_1 \mathrm{X}_{ij1} + \cdots + \beta_p \mathrm{X}_{ijp} + \varepsilon_{ij}, \quad \boldsymbol{\varepsilon}_i \sim \mathrm{No}(\mathbf{0}, \mathbf{R}_i)$$

pour $i = 1, \ldots, m$ et $j = 1, \ldots, n_i$ et où Y_{ij} est l'observation j du groupe i. L'**ordonnée à l'origine** du groupe i est désormais $\beta_0 + b_i$. Elle est composée

- d'un effet fixe commun à tous les groupes, β_0;
- d'un effet aléatoire spécifique au groupe i, b_i.

Illustration de l'effet aléatoire

On montre l'impact de l'effet aléatoire pour les données `vengeance` avec covariance AR(1) et `t` comme variable explicative continue.



Modèle sans effet aléatoire (gauche) et avec effet aléatoire pour `id` (droite).

Modèle avec ordonnée à l'origine aléatoire

Soit l'équation du modèle,

$$Y_{ij} = (\beta_0 + b_i) + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \cdots + \beta_p X_{ijp} + \varepsilon_{ij}$$

- Les effets aléatoires $b_1, \ldots, b_m$ sont supposées indépendants des termes d'aléas ε et des variables explicatives
- On postule pour le moment
  - $b_i \stackrel{\text{iid}}{\sim} \text{No}(0, \sigma_b^2)$ $(i = 1, \ldots, m)$.
  - $\varepsilon_{ij} \stackrel{\text{iid}}{\sim} \text{No}(0, \sigma^2)$ $(i = 1, \ldots, m; j = 1, \ldots, n_i)$.

Modèles avec effets aléatoires: covariance

Comme il est aléatoire, le terme b_i introduit de la corrélation intra-groupe dans le modèle. Puisque ε_{ij} est indépendant de b_i pour tout i, j, la variance (conditionelle) d'une observation est

$$\mathrm{Var}\left(Y_{ij} \mid \mathbf{X}_i\right) = \mathrm{Var}\left(b_i\right) + \mathrm{Var}\left(\varepsilon_{ij}\right) = \sigma_b^2 + \sigma^2$$

La covariance entre deux individus d'un même groupe est

$$\mathrm{Cov}\left(Y_{ij}, Y_{ik} \mid \mathbf{X}_i\right) = \sigma_b^2, \qquad j \neq k.$$

Par conséquent, la corrélation entre deux individus d'un même groupe est

$$\mathrm{Corr}\left(Y_{ij}, Y_{ik} \mid \mathbf{X}_i\right) = \frac{\sigma_b^2}{\sigma^2 + \sigma_b^2}, \qquad j \neq k.$$

Cette quantité est habituellement appelée corrélation intra-groupe.

Aparté mathématique

Les coefficients β_j et les covariables ne sont pas aléatoires, ce qui fait que

$$\begin{aligned}
\mathsf{Cov}\left(Y_{ij}, Y_{ik} \mid \mathbf{X}_i\right) &= \mathsf{Co}\left(\beta_0 + b_i + \beta_1 \mathrm{X}_{ij1} + \cdots + \varepsilon_{ij},\right. \\
&\qquad\qquad \left.\beta_0 + b_i + \beta_1 \mathrm{X}_{ik1} + \cdots + \varepsilon_{ik} \mid \mathbf{X}_i\right) \\
&= \mathsf{Cov}\left(b_i + \varepsilon_{ij}, b_i + \varepsilon_{ik}\right) \\
&= \mathsf{Var}\left(b_i\right) + \mathsf{Cov}\left(\varepsilon_{ij}, \varepsilon_{ik}\right) \\
&= \sigma_b^2 + \sigma^2 \mathbf{1}_{j=k}.
\end{aligned}$$

où la dernière étape vient de l'indépendance entre b_i et $\boldsymbol{\varepsilon}$ et puisque $\mathsf{Cov}\left(Y_{ij}, Y_{ij}\right) = \mathsf{Var}\left(Y_{ij}\right)$.

Variabilité des Y

De manière équivalente, on peut écrire

$$
\begin{aligned}
\mathsf{Var}\left(Y_i \mid \mathbf{X}_i\right) &= \mathsf{Var}\left(b_i \mathbf{1}_{n_i}\right) + \mathsf{Var}\left(\boldsymbol{\varepsilon}_i\right) \\
&= \sigma_b^2 \mathbf{1}_{n_i} \mathbf{1}_{n_i}^\top + \sigma^2 \mathbf{I}_{n_i} \\
&= \begin{pmatrix} \sigma_b^2 & \sigma_b^2 & \cdots & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 & \cdots & \sigma_b^2 \\ \vdots & \ddots & \ddots & \vdots \\ \sigma_b^2 & \sigma_b^2 & \cdots & \sigma_b^2 \end{pmatrix} + \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix}.
\end{aligned}
$$

Structure d'équicorrélation avec ordonnée à l'origine aléatoire

- Quand les aléas ε_i sont indépendants et homoscédastiques, soit $\mathsf{Var}\left(\varepsilon_i\right) = \sigma^2 \mathbf{I}_{n_i}$, introduire un effet aléatoire pour l'ordonnée à l'origine fait en sorte que la corrélation intra-groupe est constante.
- Dans ce cas précis, la matrice de covariance conditionnelle de Y_i est donc la même que si on considérait une régression linéaire sans effet aléatoire avec une structure d'équicorrélation pour $\mathsf{Var}\left(\varepsilon_i\right)$.
- La différence ici est que la corrélation doit être non-négative. Ce n'est généralement pas une limitation, car les corrélations intra-groupe ont tendance à être positives.

Ajout d'un effet aléatoire avec la commande `random`

- La commande `repeated` nous permettait de spécifier la structure de covariance des erreurs avec `proc mixed`.
- Si on utilise pas la commande `repeated`, les erreurs sont postulées indépendantes.

Code SAS pour ajuster un modèle avec ordonnée à l'origine aléatoire avec erreurs indépendantes

```
proc mixed data=modstat.mobilisation;
class idunite;
model mobilisation = sexe anciennete agegest nunite / solution;
random intercept / subject=idunite v=1 vcorr=1;
run;
```

L'inclusion d'une ordonnée à l'origine aléatoire induit une structure d'équicorrélation pour les réponses. Il n'est donc pas nécessaire de spécifier une structure de covariance pour les erreurs.

Matrice de covariance du modèle avec ordonnée à l'origine aléatoire

Matrice V estimée pour idunite 1

Ligne	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9
1	1.3709	0.2448	0.2448	0.2448	0.2448	0.2448	0.2448	0.2448	0.2448
2	0.2448	1.3709	0.2448	0.2448	0.2448	0.2448	0.2448	0.2448	0.2448
3	0.2448	0.2448	1.3709	0.2448	0.2448	0.2448	0.2448	0.2448	0.2448
4	0.2448	0.2448	0.2448	1.3709	0.2448	0.2448	0.2448	0.2448	0.2448
5	0.2448	0.2448	0.2448	0.2448	1.3709	0.2448	0.2448	0.2448	0.2448
6	0.2448	0.2448	0.2448	0.2448	0.2448	1.3709	0.2448	0.2448	0.2448
7	0.2448	0.2448	0.2448	0.2448	0.2448	0.2448	1.3709	0.2448	0.2448
8	0.2448	0.2448	0.2448	0.2448	0.2448	0.2448	0.2448	1.3709	0.2448
9	0.2448	0.2448	0.2448	0.2448	0.2448	0.2448	0.2448	0.2448	1.3709

Estimés des paramètres de covariance

Valeur estimée du paramètre de covariance

Param. de cov.	Sujet	Estimation
Intercept	idunite	0.2448
Residual		1.1261

- L'estimé de la variance de l'effet aléatoire est $\widehat{\sigma}_b^2 = 0{,}2448$, tandis que l'estimé de la variance des erreurs est $\widehat{\sigma}^2 = 1{,}1261$.
- Conséquemment, la corrélation intra-unité est

$$\widehat{\rho} = \frac{\widehat{\sigma}_b^2}{\widehat{\sigma}_b^2 + \sigma^2} = 0{,}1785.$$

- C'est exactement la même corrélation que celle rapportée pour le modèle d'équicorrélation intra-unité des erreurs (commande `repeated`).

Estimés des effets fixes

Solution pour effets fixes

Effet	Estimation	Erreur type	DDL	Valeur du test t	Pr > |t|
Intercept	13.7633	0.3955	97	34.80	<.0001
sexe	0.5622	0.06835	914	8.23	<.0001
anciennete	-0.4722	0.006015	914	-78.50	<.0001
agegest	0.01929	0.006801	97	2.84	0.0056
nunite	0.006470	0.02019	97	0.32	0.7493

L'effet des variables explicatives (et leurs erreurs-types) sont exactement les même que pour le modèle d'équicorrélation des erreurs — les deux modèles sont équivalents pour la réponse si la corrélation intra-unité est positive.