Master MIASHS - Master 1

V2DMI5

Analyse de données de panels

Évaluation 1- février 2019

Durée : 2 heures

Exercice 1 Soit le modèle théorique suivant :

$$y_{ijk} = \mu + \alpha_i + \omega_{ij} + \gamma\ x_{ijk} + \varepsilon_{ijk} \quad i = 1, 2, 3, 4, 5,\ j = 1, 2, 3,\ k = 1, 2, 3, 4$$

avec $\alpha_i \sim \mathcal{N}(0, \sigma^2_\alpha),\ \omega_{ij} \sim \mathcal{N}(0, \sigma^2_\omega)$ et $\varepsilon_{ijk} \sim \mathcal{N}(0, \sigma^2_\varepsilon)$, les α_i, ω_{ij} et ε_{ijk} sont indépendants les uns des autres.

Dans tout l'exercice et pour plus de facilité on rangera Y de la façon suivante

$$(y_{111}, y_{112}, y_{113}, y_{114}, y_{121}, \ldots).$$

1. Donner l'expression de l'espérance mathématique marginale de Y_{ijk} .
2. Identifier les paramètres associés à la partie fixe du modèle et identifier les variables intervenant dans cette partie fixe.
3. Donner la matrice de design X associée à cette partie fixe.
4. Identifier les facteurs à effet aléatoire de ce modèle en précisant le nombre de niveaux de chacun. Préciser aussi le sens de la modélisation de ces effets.
5. Donner l'expression de la variance marginale de Y_{ijk} .
6. Donner la matrice de design Z associée aux effets aléatoires en précisant ses dimensions.
7. Dans ce modèle, écrire les covariances suivantes :
   (a) $\text{cov}(y_{113}, y_{133})$
   (b) $\text{cov}(y_{211}, y_{112})$
   (c) $\text{cov}(y_{123}, y_{124})$
   (d) $\text{cov}(y_{312}, y_{312})$
8. En déduire la matrice de variance-covariance marginale de Y selon le modèle théorique.
9. Identifier les paramètres à estimer de ce modèle.

Exercice 2 Une étude visant à analyser l'effet de l'âge et de l'école sur un score de performance scolaire a été faite. Pour cela 27 enfants ont été suivis à 8, 10, 12 et 14 ans dans 3 écoles différentes. L'objectif de cet exercice est de retrouver les codes R et d'analyser les résultats proposés à partir des sorties R. Les analyses ont été réalisées avec la fonction lmer du package lme4.

1. Un premier modèle a été estimé et on a obtenu les résultats suivants (table 1):

Model 1	Sum Sq	Mean Sq	NumDF	DenDF	F.value	Pr(>F)
Age	235.36	235.36	1.00	78.00	121.28	0.0000
Ecole	3.30	1.65	2.00	101.13	0.85	0.4308
Age:Ecole	12.59	6.29	2.00	78.00	3.24	0.0443

Model 1	Groups	Name	Std.Dev.
	Subject	(Intercept)	1.98
	Residual		1.39

Table 1: Résultats numériques de l'analyse 1

a) Déduire la nature (variables quantitatives/qualitatives, effets fixes/alétoires) des différentes variables.
b) Écrire le code R ayant permis d'estimer ce modèle.
c) Interpréter les résultats. Que concluez-vous ?

2. Deux analyses supplémentaires ont été faites dont les résultats sont présentés ci-dessous (table 2 et table 3) :

Model 2	Sum Sq	Mean Sq	NumDF	DenDF	F.value	Pr(>F)
Age	235.36	235.36	1.00	79.17	121.61	0.0000
Age:ecole	19.32	9.66	2.00	44.28	4.99	0.0111

Model 2	Groups	Name	Std.Dev.
	Subject	(Intercept)	1.99
	Residual		1.39

Table 2: Résultats numériques de l'analyse 2

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
Model 2		445.44	461.54	-216.72	433.44			
Model 1		447.66	469.11	-215.83	431.66	1.79	2	0.4093

Table 3: Résultats numériques de l'analyse 3

a) Qu'est-ce qui a changé entre le premier modèle (model 1) et le second (model 2) ? Ce nouveau modèle (model 2) est-il emboîté dans le modèle précédent (model 1) ?

b) Qu'est-ce qui a amené le responsable de l'étude à s'intéresser à ce modèle 2 ?

c) Les résultats de la table 3 sont issus de la fonction *anova*. Interpréter ces résultats en vous appuyant sur plusieurs critères.

3. Une dernier modèle a été estimé. Les résultats sont présentés ci-dessous (table 4) :

Model 3	Sum Sq	Mean Sq	NumDF	DenDF	F.value	Pr(>F)
Age	166.72	166.72	1.00	25.17	97.14	0.0000
Age:Ecole	12.67	6.33	2.00	24.00	3.69	0.0400

Model 3	Groups	Name	Std.Dev.	Corr
	Subject	(Intercept)	2.33	
	Subject	Age	0.19	-0.59
	Residual		1.31	

Table 4: Résultats numériques de l'analyse 4

a) Donner la commande R qui a permis d'estimer ce modèle.

b) Que représente le paramètre de la colonne Corr estimé à -0.59 ?

c) Proposer une modification de la commande R pour que ce paramètre ne soit plus estimé.

d) Donner la structure de la matrice de variance-covariance marginale en précisant l'ordre que vous donnez aux observations dans Y.