

Université de Montréal

Problème des données manquantes dans un plan
d'analyse de variance à mesures répétées.

par

Emmanuel Bernard Lelo

Département de mathématiques et de statistique
Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de
Maître ès sciences (M.Sc.)
en Statistique

août 2003

© Emmanuel Bernard Lelo, 2003



QA
3
U54
2003
v.014

Direction des bibliothèques

AVIS

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.

Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

NOTICE

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.

Université de Montréal

Faculté des études supérieures

Ce mémoire intitulé

**Problème des données manquantes dans un plan
d'analyse de variance à mesures répétées.**

présenté par

Emmanuel Bernard Lelo

a été évalué par un jury composé des personnes suivantes :

Jean-François Angers

(président-rapporteur)

Yves Lepage

(directeur de recherche)

Robert Cléroux

(co-directeur)

Martin Bilodeau

(membre du jury)

Mémoire accepté le:

29 octobre 2003

SOMMAIRE

Ce mémoire se propose d'examiner l'influence des données manquantes dans un plan d'analyse de variance à mesures répétées. Nous avons tout d'abord passé en revue les concepts théoriques des données manquantes en s'inspirant principalement de Rubin(1987), Little et Rubin(1987) et Schafer(1997). Ensuite, nous avons examiné les différentes méthodes d'analyse de variance pour les plans à mesures répétées. Enfin, l'influence des données manquantes complètement aléatoirement dans un plan d'analyse de variance à mesures répétées a été évaluée en se servant des procédures GLM et MIXED de SAS. Nous avons vu qu'en présence des données manquantes, l'approche des modèles linéaires mixtes est la plus appropriée pour l'analyse de variance à mesures répétées.

SUMMARY

This work examines the influence of missing data in the analysis of variance with repeated measurements. We first review the theoretical concepts of missing data as explained in Rubin(1987), Little and Rubin(1987) and Schafer(1997). Then, we examine the various approaches of the analysis of variance with repeated measurements. Lastly, the influence of missing data completely at random in the analysis of variance with repeated measurements was evaluated using SAS GLM and MIXED procedures. We observed that in the presence of missing data, the approach with the mixed linear models is the most appropriate for the analysis of variance with repeated measurements.

REMERCIEMENTS

Je voudrais remercier sincèrement deux personnes qui ont rendu possible la rédaction de ce mémoire grâce à leurs compétences et leurs disponibilités : mon directeur de recherche M. Yves Lepage et mon codirecteur de recherche M. Robert Cléroux.

Encore une fois merci à M. Cléroux pour m'avoir fourni une aide financière durant la rédaction de ce mémoire.

Mes remerciements affectueux vont à ma fiancée Michelle qui m'a enduré tout au long de cette formation.

TABLE DES MATIÈRES

Sommaire	iii
Summary	iv
Remerciements	v
Liste des figures	ix
Liste des tableaux	x
Introduction	1
Chapitre 1. Le problème des données manquantes	2
1.1. Introduction	2
1.2. Modèles pour des mesures répétées	3
1.3. Mécanisme des données manquantes	6
1.4. Méthodes d'analyse des données avec valeurs manquantes	8
1.5. L'imputation multiple	12
1.6. La vraisemblance maximale	15
1.6.1. L'algorithme EM	18
Chapitre 2. Plan d'expérience à deux facteurs dont un comporte des mesures répétées	21

2.1.	Introduction	21
2.2.	Le modèle mathématique.....	24
2.3.	Analyse de variance.....	26
2.4.	Tests des effets des facteurs.	28
2.5.	Vérification des présupposés du modèle à mesures répétées.	30
2.5.1.	La symétrie composée.....	33
2.5.2.	Ajustement des degrés de liberté.....	34
2.5.3.	L'analyse multivariée	35
2.6.	L'approche mixte	39
2.6.1.	Stratégies d'analyse.....	44
Chapitre 3.	Analyse des résultats.....	47
3.1.	Introduction	47
3.2.	Méthodologie de génération des ensembles de données.....	48
3.3.	Résultats issus de la procédure GLM.....	52
3.3.1.	Résultats pour le scénario 1.....	53
3.3.2.	Résultats pour le scénario 2.....	55
3.3.3.	Résultats pour le scénario 3.....	57
3.3.4.	Résultats pour le scénario 4.....	59
3.4.	Résultats issus de la procédure MIXED	61
3.4.1.	Résultats pour le scénario 1.....	62
3.4.2.	Résultats pour le scénario 2.....	63
3.4.3.	Résultats pour le scénario 3.....	66
3.4.4.	Résultats pour le scénario 4.....	68

Conclusion	70
Annexe A. Tableaux des données	A-i
Annexe B. Estimateurs des moyennes des groupes dans le temps pour le cas 1 et pour les quatre scénarios	B-i
Annexe C. Tableaux des effets des facteurs résultant de la procédure GLM de SAS	C-i
Annexe D. Tableaux des valeurs des critères d'information AIC et BIC	D-i
Annexe E. Tableaux des valeurs des critères d'information AIC et BIC	E-i
Annexe F. Tableaux des effets des facteurs résultant de la procédure MIXED de SAS	F-i
Annexe G. Suite des tableaux des effets des facteurs résultant de la procédure MIXED de SAS	G-i
Bibliographie	G-i

LISTE DES FIGURES

A.0.1 Les données complètes.....	A-ii
A.0.2 Les données pour le scénario 1, $pvm=0,05$	A-iii
A.0.3 Les données pour le scénario 1, $pvm=0,10$	A-iv
A.0.4 Les données pour le scénario 1, $pvm=0,15$	A-v
A.0.5 Les données pour le scénario 2, $pvm=0,05$	A-vi
A.0.6 Les données pour le scénario 2, $pvm=0,10$	A-vii
A.0.7 Les données pour le scénario 2, $pvm=0,15$	A-viii
A.0.8 Les données pour le scénario 3, $pvm=0,05$	A-ix
A.0.9 Les données pour le scénario 3, $pvm=0,10$	A-x
A.0.10 Les données pour le scénario 3, $pvm=0,15$	A-xi
A.0.11 Les données pour le scénario 4, $pvm=0,05$	A-xii
A.0.12 Les données pour le scénario 4, $pvm=0,10$	A-xiii
A.0.13 Les données pour le scénario 4, $pvm=0,15$	A-xiv

LISTE DES TABLEAUX

2.3.1	Tableau des sommes des carrés des facteurs.	28
2.3.2	Tableau des moyennes des carrés des facteurs.	28
3.2.1	Tableau des moyennes des groupes dans le temps.	50
B.0.1	Tableau des estimateurs des moyennes des groupes dans le temps pour le cas 1 et le scénario 1.....	B-i
B.0.2	Tableau des estimateurs des moyennes des groupes dans le temps pour le cas 1 et le scénario 2.....	B-i
B.0.3	Tableau des estimateurs des moyennes des groupes dans le temps pour le cas 1 et le scénario 3.....	B-ii
B.0.4	Tableau des estimateurs des moyennes des groupes dans le temps pour le cas 1 et le scénario 4.....	B-ii
C.0.1	Résultats des tests avec GLM pour le cas de la matrice de covariances de structure symétrie composée (scénario 1) et des valeurs manquantes générées aléatoirement sans distinction des groupes (contexte 1).	C-ii
C.0.2	Résultats des tests avec GLM pour le cas de la matrice de covariances de structure symétrie composée (scénario 1) et des valeurs manquantes générées aléatoirement à l'intérieur des groupes 1 et 2 séparément(contexte 2).	C-iii

- C.0.3 Résultats des tests avec GLM pour le cas de la matrice de covariances non structurée (scénario 2) et des valeurs manquantes générées aléatoirement sans distinction des groupes (contexte 1).C-iv
- C.0.4 Résultats des tests avec GLM pour le cas de la matrice de covariances non structurée (scénario 2) et des valeurs manquantes générées aléatoirement à l'intérieur des groupes 1 et 2 séparément (contexte 2). C-v
- C.0.5 Résultats des tests avec GLM pour le cas de la matrice de covariances de structure autorégressive d'ordre 1(scénario 3) et des valeurs manquantes générées aléatoirement sans distinction des groupes (contexte 1).C-vi
- C.0.6 Résultats des tests avec GLM pour le cas de la matrice de covariances de structure autorégressive d'ordre 1 (scénario 3) et des valeurs manquantes générées aléatoirement à l'intérieur des groupes 1 et 2 séparément (contexte 2).C-vii
- C.0.7 Résultats des tests avec GLM pour le cas de la matrice de covariances de structure symétrie composée hétérogène (scénario 4) et des valeurs manquantes générées aléatoirement sans distinction des groupes (contexte 1). C-viii
- C.0.8 Résultats des tests avec GLM pour le cas de la matrice de covariances de structure symétrie composée hétérogène (scénario 4) et des valeurs manquantes générées aléatoirement à l'intérieur des groupes 1 et 2 séparément (contexte 2).C-ix
- D.0.1 Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 1 dans le contexte 1 des valeurs manquantes avec $pvm=0,05$ D-i

- D.0.2 Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 1 dans le contexte 1 des valeurs manquantes avec $pvm=0,10$D-ii
- D.0.3 Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 1 dans le contexte 1 des valeurs manquantes avec $pvm=0,15$D-ii
- D.0.4 Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 1 dans le contexte 2 des valeurs manquantes avec $pvm=0,05$D-ii
- D.0.5 Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 1 dans le contexte 2 des valeurs manquantes avec $pvm=0,10$ D-iii
- D.0.6 Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 1 dans le contexte 2 des valeurs manquantes avec $pvm=0,15$ D-iii
- D.0.7 Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 2 dans le contexte 1 des valeurs manquantes avec $pvm=0,05$ D-iii
- D.0.8 Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 2 dans le contexte 1 des valeurs manquantes avec $pvm=0,10$ D-iv
- D.0.9 Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 2 dans le contexte 1 des valeurs manquantes avec $pvm=0,15$ D-iv

- D.0.10 Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 2 dans le contexte 2 des valeurs manquantes avec $pvm=0,05$ D-iv
- D.0.11 Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 2 dans le contexte 2 des valeurs manquantes avec $pvm=0,10$ D-v
- D.0.12 Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 2 dans le contexte 2 des valeurs manquantes avec $pvm=0,15$ D-v
- D.0.13 Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 3 dans le contexte 1 des valeurs manquantes avec $pvm=0,05$ D-v
- D.0.14 Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 3 dans le contexte 1 des valeurs manquantes avec $pvm=0,10$ D-vi
- D.0.15 Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 3 dans le contexte 1 des valeurs manquantes avec $pvm=0,15$ D-vi
- D.0.16 Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 3 dans le contexte 2 des valeurs manquantes avec $pvm=0,05$ D-vi
- D.0.17 Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 3 dans le contexte 2 des valeurs manquantes avec $pvm=0,10$ D-vii

- D.0.18 Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 3 dans le contexte 2 des valeurs manquantes avec $pvm=0,15$D-vii
- E.0.1 Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 4 dans le contexte 1 des valeurs manquantes avec $pvm=0,05$ E-i
- E.0.2 Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 4 dans le contexte 1 des valeurs manquantes avec $pvm=0,10$ E-ii
- E.0.3 Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 4 dans le contexte 1 des valeurs manquantes avec $pvm=0,15$ E-ii
- E.0.4 Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 4 dans le contexte 2 des valeurs manquantes avec $pvm=0,05$ E-ii
- E.0.5 Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 4 dans le contexte 2 des valeurs manquantes avec $pvm=0,10$ E-iii
- E.0.6 Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 4 dans le contexte 2 des valeurs manquantes avec $pvm=0,15$ E-iii
- F.0.1 Effets de la structure de covariances sur les effets des facteurs pour le scénario 1 dans le contexte 1 des valeurs manquantes avec $pvm=0,05$. F-ii

- F.0.2 Effets de la structure de covariances sur les effets des facteurs pour le scénario 1 dans le contexte 1 des valeurs manquantes avec $pvm=0,10$. F-iii
- F.0.3 Effets de la structure de covariances sur les effets des facteurs pour le scénario 1 dans le contexte 1 des valeurs manquantes avec $pvm=0,15$. F-iv
- F.0.4 Effets de la structure de covariances sur les effets des facteurs pour le scénario 1 dans le contexte 2 des valeurs manquantes avec $pvm=0,05$. F-v
- F.0.5 Effets de la structure de covariances sur les effets des facteurs pour le scénario 1 dans le contexte 2 des valeurs manquantes avec $pvm=0,10$. F-vi
- F.0.6 Effets de la structure de covariances sur les effets des facteurs pour le scénario 1 dans le contexte 2 des valeurs manquantes avec $pvm=0,15$. F-vii
- F.0.7 Effets de la structure de covariances sur les effets des facteurs pour le scénario 2 dans le contexte 1 des valeurs manquantes avec $pvm=0,05$. F-viii
- F.0.8 Effets de la structure de covariances sur les effets des facteurs pour le scénario 2 dans le contexte 1 des valeurs manquantes avec $pvm=0,10$. F-ix
- F.0.9 Effets de la structure de covariances sur les effets des facteurs pour le scénario 2 dans le contexte 1 des valeurs manquantes avec $pvm=0,15$. F-x
- F.0.10 Effets de la structure de covariances sur les effets des facteurs pour le scénario 2 dans le contexte 2 des valeurs manquantes avec $pvm=0,05$. F-xi
- F.0.11 Effets de la structure de covariances sur les effets des facteurs pour le scénario 2 dans le contexte 2 des valeurs manquantes avec $pvm=0,10$. F-xii
- F.0.12 Effets de la structure de covariances sur les effets des facteurs pour le scénario 2 dans le contexte 2 des valeurs manquantes avec $pvm=0,15$. F-xiii
- G.0.1 Effets de la structure de covariances sur les effets des facteurs pour le scénario 3 dans le contexte 1 des valeurs manquantes avec $pvm=0,05$. G-ii

- G.0.2 Effets de la structure de covariances sur les effets des facteurs pour le scénario 3 dans le contexte 1 des valeurs manquantes avec $pvm=0,10$.G-iii
- G.0.3 Effets de la structure de covariances sur les effets des facteurs pour le scénario 3 dans le contexte 1 des valeurs manquantes avec $pvm=0,15$.G-iv
- G.0.4 Effets de la structure de covariances sur les effets des facteurs pour le scénario 3 dans le contexte 2 des valeurs manquantes avec $pvm=0,05$. G-v
- G.0.5 Effets de la structure de covariances sur les effets des facteurs pour le scénario 3 dans le contexte 2 des valeurs manquantes avec $pvm=0,10$.G-vi
- G.0.6 Effets de la structure de covariances sur les effets des facteurs pour le scénario 3 dans le contexte 2 des valeurs manquantes avec $pvm=0,15$ G-vii
- G.0.7 Effets de la structure de covariances sur les effets des facteurs pour le scénario 4 dans le contexte 1 des valeurs manquantes avec $pvm=0,05$ G-viii
- G.0.8 Effets de la structure de covariances sur les effets des facteurs pour le scénario 4 dans le contexte 1 des valeurs manquantes avec $pvm=0,10$.G-ix
- G.0.9 Effets de la structure de covariances sur les effets des facteurs pour le scénario 4 dans le contexte 1 des valeurs manquantes avec $pvm=0,15$. G-x
- G.0.10 Effets de la structure de covariances sur les effets des facteurs pour le scénario 4 dans le contexte 2 des valeurs manquantes avec $pvm=0,05$.G-xi
- G.0.11 Effets de la structure de covariances sur les effets des facteurs pour le scénario 4 dans le contexte 2 des valeurs manquantes avec $pvm=0,10$ G-xii
- G.0.12 Effets de la structure de covariances sur les effets des facteurs pour le scénario 4 dans le contexte 2 des valeurs manquantes avec $pvm=0,15$ G-xiii

INTRODUCTION

La présence des données manquantes dans un plan d'analyse de variance à mesures répétées est souvent inévitable pour plusieurs raisons et cela suscite un questionnement quant à l'exactitude des résultats de l'analyse effectuée. Il est donc nécessaire de déterminer quelles sont les procédures et méthodes qui sont capables de minimiser l'influence des données manquantes sur les résultats de l'analyse de variance à mesures répétées.

Dans le chapitre 1, nous faisons état de l'aspect théorique des données manquantes, des modèles pour les mesures répétées et de certaines méthodes d'analyse de données avec valeurs manquantes. Nous avons aussi introduit la méthode d'imputation multiple, la méthode d'estimation à vraisemblance maximale et l'algorithme EM.

Dans le chapitre 2, nous présentons l'approche classique, les approches ajustées, l'approche multivariée et l'approche mixte pour l'analyse de variance à deux facteurs dont un est à mesures répétées.

Enfin dans le chapitre 3, à l'aide des procédures GLM et MIXED de SAS, nous avons fait des analyses afin d'évaluer l'influence des données manquantes dans un plan d'analyse de variance à deux facteurs dont un est à mesures répétées en utilisant quatre ensembles de données ayant chacun une structure différente pour la matrice de covariances.

Chapitre 1

LE PROBLÈME DES DONNÉES MANQUANTES

1.1. INTRODUCTION

Les données manquantes sont presque inévitables dans la plupart des études statistiques. Les données manquantes sont particulièrement un problème en analyse des données à mesures répétées comme par exemple dans le cadre des essais cliniques où les données relatives à l'efficacité d'un médicament sont prélevées à plusieurs reprises dans le temps. Les données peuvent être manquantes pour diverses raisons dont le manque de coopération de la part du patient, l'inefficacité du médicament étudié, l'apparition d'effets secondaires, le décès et bien d'autres raisons.

Des plans d'expériences contrôlés sont généralement soigneusement conçus pour permettre que des analyses statistiques soient faites en utilisant des méthodes appropriées. Pour un plan d'expérience équilibré, c'est-à-dire que les combinaisons des niveaux des facteurs sont toutes observées un même nombre de fois, une analyse des moindres carrés permet d'obtenir des estimateurs pour les paramètres et de déduire le tableau d'analyse de variance sous les présupposés usuels.

Étant donné que les niveaux des facteurs dans une expérience contrôlée sont fixés par l'expérimentateur, les données manquantes sont généralement plus fréquentes pour la variable d'intérêt Y que pour les facteurs. L'attention est alors portée sur les données manquantes pour la variable d'intérêt Y . Lorsque ce genre de données manquent, l'équilibre dans le plan d'expérience d'origine est détruit, et par conséquent l'analyse appropriée des moindres carrés devient plus compliquée.

La plupart des méthodes statistiques usuelles présument la présence des données pour chaque individu pour toutes les variables incluses dans l'analyse. Une solution simple utilisée par défaut par plusieurs logiciels statistiques consiste à exclure de l'analyse tout individu ayant au moins une donnée manquante. Ceci a pour conséquence de contribuer à une perte substantielle d'information lorsque plusieurs individus ont au moins une donnée manquante et surtout de biaiser les résultats de l'analyse lorsque le mécanisme qui produit les données manquantes dépend des valeurs de Y .

Une façon simple de résoudre ce problème est de remplacer les données manquantes par des estimateurs non biaisés avant de procéder à l'analyse en se servant d'un plan d'expérience rendu équilibré. Le problème qui se pose alors est d'appliquer une méthode adéquate qui tient compte du mécanisme des données manquantes et de leur tendance pour déterminer les estimateurs non biaisés.

1.2. MODÈLES POUR DES MESURES RÉPÉTÉES

Pour des études avec des mesures répétées, les données pour la variable d'intérêt Y sont mesurées sur des individus à plusieurs reprises dans le temps. Par conséquent les observations sur chaque individu dans le temps ne sont pas indépendantes. Dans le contexte du modèle linéaire général, la dépendance des mesures de Y sur les variables explicatives et les corrélations entre les mesures de Y doivent être prises en considération. De plus, les ensembles de données à mesures répétées sont souvent non équilibrés et pour cette raison, ils ne peuvent

pas être analysés en utilisant des techniques usuelles de régression multivariée (voir Hand et Taylor, 1987).

Une alternative naturelle survient en observant que le profil des mesures répétées spécifique à chaque sujet peut être bien approximé par des fonctions de régression linéaire. Ainsi le vecteur des mesures répétées pour chaque sujet peut être réduit à un vecteur d'un nombre restreint de coefficients de régression estimés spécifiques à l'individu et des techniques de régression multivariée peuvent être utilisées par la suite pour relier ces estimateurs aux covariables.

Laird et Ware(1982) ont introduit un modèle pour les mesures répétées lorsque les données suivent une loi normale. Soit

$$Y_i = X_i\beta + Z_ib_i + \varepsilon_i \quad (1.2.1)$$

où Y_i est le vecteur $(m_i \times 1)$ des résultats à mesures répétées pour le sujet i , $1 \leq i \leq N$, N est le nombre de sujets, $X_i : (m_i \times p)$ est une matrice des p covariables dépendantes du temps avec coefficients de régression β , $Z_i : (m_i \times q)$ est une matrice des q covariables dépendantes du temps avec coefficients de régression aléatoires b_i , $b_i : (q \times 1)$ est le vecteur de coefficients individuels qui varient pour chaque sujet et $\varepsilon_i : (m_i \times 1)$ est le vecteur de l'erreur pour chaque sujet i , β est considéré fixe. Les vecteurs ε_i sont $N_{m_i}(0_{m_i}, \Sigma_i)$ et indépendants entre les sujets, les vecteurs b_i sont $N_q(0_q, D)$ et indépendants des ε_i , D est une matrice de covariances $(q \times q)$ et Σ_i est une matrice de covariances $(m_i \times m_i)$ qui dépend de i seulement à travers sa dimension m_i .

Le modèle (1.2.1) a stimulé plusieurs approches concernant les données à mesures répétées. Ainsi, Diggle *et al.* (1994) traite trois modèles distincts : le modèle marginal, le modèle à effets aléatoires et le modèle de transition. La première approche modélise séparément la moyenne marginale et la covariance. Une espérance mathématique non conditionnelle de Y_i de l'équation (1.2.1) donne le modèle marginal suivant :

$$E(Y_i) = X_i\beta,$$

$i = 1, \dots, N$. La corrélation entre les Y_i est une fonction connue avec quelques paramètres inconnus. Les coefficients de régression et les paramètres de la matrice de corrélation doivent être estimés. L'objectif primaire est de modéliser les effets des covariables sur la distribution marginale de Y_i . Liang et Zeger (1986) propose des équations d'estimations généralisées ("generalized estimating equations") pour la distribution marginale de Y_i . Cette approche évite une hypothèse explicite de la distribution et traite les corrélations comme des paramètres de nuisance. Sous certaines hypothèses concernant la dépendance par rapport au temps, elle donne des estimations convergentes.

En supposant que les coefficients de régression varient pour chaque individu, la deuxième approche modélise l'espérance mathématique conditionnelle des mesures de Y_i étant donné le coefficient individuel b_i . Dans l'équation (1.2.1), étant donné b_i , l'espérance mathématique conditionnelle de Y_i a la forme du modèle linéaire général :

$$E(Y_i|b_i) = X_i\beta + Z_ib_i,$$

$i = 1, \dots, N$. Ceci est le modèle à effets aléatoires. Ce modèle est très utile lorsque l'objectif est de faire de l'inférence sur des individus plutôt que sur la moyenne de la population comme dans le cas du modèle marginal. Ce modèle suppose aussi qu'étant donné b_i , les réponses Y_i sont indépendantes. Le coefficient individuel b_i a une forme spécifique de distribution avec quelques paramètres connus ou inconnus. Les coefficients de régression sont communs aux réponses pour un individu donné mais varient entre les individus, induisant ainsi la corrélation. Traditionnellement, une distribution multinormale est supposée avec un vecteur moyen 0 et une matrice de covariance indexée par certains paramètres. Soient $\Sigma_b = \sigma_b^2 I$ et $\Sigma_\varepsilon = \sigma^2 I$, alors la variance de Y est $\Sigma_Y = \sigma_b^2 Z Z' + \sigma^2 I$. Lorsque $Z = 0$ ou $\sigma_b^2 = 0$, le modèle à effets aléatoires se réduit au modèle linéaire général. L'estimation des paramètres dans un modèle à effets aléatoires peut s'effectuer par la méthode des moindres carrés généralisés ou la méthode de vraisemblance maximale.

La méthode des moindres carrés généralisés estime β en minimisant $(Y - X\beta)' \Sigma_Y (Y - X\beta)$. Lorsque Σ_Y est inconnue ou lorsque les estimations de σ^2 et de σ_b^2 ne sont pas disponibles, cette méthode ne peut pas être utilisée. Les méthodes basées sur la vraisemblance supposent que b et ε sont distribués selon une loi multinormale. Une alternative est la vraisemblance maximale restreinte qui est déduite en utilisant une formulation bayésienne (voir Harville (1974) et Dempster *et al.*, 1977). L'algorithme EM ("Expectation Maximisation" dans la littérature) peut être utilisé pour obtenir les estimations.

Comme dans le cas du modèle à effets aléatoires, le modèle de transition considère l'objectif de régression et la corrélation intra sujet simultanément. Ce modèle combine les équations de régression et l'hypothèse de corrélation en une seule équation. On spécifie un modèle de régression pour l'espérance conditionnelle de la réponse courante étant donné des réponses précédentes, comme une fonction des covariables et des réponses passées :

$$E(Y_t | Y_{t-1}, \dots, Y_1) = X\beta + \sum_{j=1}^n f_j(Y_{t-1}, \dots, Y_1),$$

où les f_j sont des fonctions connues. Notons que pour des réponses normalement distribuées avec une liaison identité, ces trois approches sont équivalentes et donnent les mêmes estimations (voir Diggle *et al.*, 1994).

1.3. MÉCANISME DES DONNÉES MANQUANTES.

De façon pratique, les mesures répétées sur un sujet peuvent ne pas être disponibles pour tous les temps. Soit $Y_i = (Y_{i1}, \dots, Y_{ib})$ le vecteur complet de mesures répétées pour le i^e sujet, $1 \leq i \leq N$, N est le nombre de sujets. Pour la k^e mesure, la réponse ou la non-réponse de Y_{ik} est indiquée par la variable aléatoire $M_{ik} = 1$ si Y_{ik} est manquante, et $M_{ik} = 0$ si Y_{ik} est observée, où $k = 1, \dots, b$.

La distribution de $M_i = (M_{i1}, \dots, M_{ib})$ peut dépendre de Y_i et des autres variables. La distribution conditionnelle de M_{ik} étant donné Y_{i1}, \dots, Y_{ik-1} est le

processus ou le mécanisme des données manquantes. Ce processus est décrit par la probabilité des données manquantes :

$$Prob(M_{ik} = 1) = Prob(Y_{ik} \text{ est manquante} | Y_{i1} = y_{i1}, \dots, Y_{i(k-1)} = y_{i(k-1)}).$$

La question importante est de déterminer si le processus des données manquantes est dépendant du processus des mesures de Y_i . Le mécanisme qui génère les données manquantes est un élément clé pour ce qui est du choix de la méthode d'analyse appropriée et de l'interprétation des résultats. Little et Rubin (1987) donnent un traitement général de l'analyse statistique avec données manquantes et une classification utile du processus des données manquantes. Les données manquantes sont dites :

(a) manquantes complètement aléatoirement ("Missing Completely at Random"(MCAR)) si le processus des données manquantes est indépendant des mesures observées et des mesures non observées ;

(b) manquantes aléatoirement ("Missing at Random"(MAR)) si le processus des données manquantes ne dépend pas des mesures non observées ;

(c) non ignorables ou informatives si le processus des données manquantes dépend des mesures non observées.

Dans le cas (a), les données observées sont un sous échantillon aléatoire de l'échantillon des données complètes, qui représentent l'ensemble des données qui seraient recueillies dans le cas où il n'y aurait pas de données manquantes. La non réponse de la k^e mesure pour le i^e sujet est indépendante de Y_{ik} et des mesures précédentes $Y_{i1}, \dots, Y_{i(k-1)}$.

Dans le cas (b), les données observées ne sont plus un sous échantillon aléatoire de l'échantillon des données complètes, mais elles sont encore un échantillon aléatoire dans chaque sous classe définie par les valeurs de $Y_{i1}, \dots, Y_{i(k-1)}$. La non-réponse de la k^e mesure pour le i^e sujet est indépendante de Y_{ik} , mais elle peut dépendre des mesures précédentes $Y_{i1}, \dots, Y_{i(k-1)}$.

Dans le cas (c), la non réponse de la k^e mesure pour le i^e sujet dépend de Y_{ik} et elle peut aussi dépendre des mesures précédentes $Y_{i1}, \dots, Y_{i(k-1)}$.

1.4. MÉTHODES D'ANALYSE DES DONNÉES AVEC VALEURS MANQUANTES.

Il y a deux approches simples pour analyser les ensembles de données avec valeurs manquantes. L'approche la plus radicale consiste à restreindre l'analyse aux individus dont toutes les données sont observées à toutes les périodes, c'est-à-dire exclure de l'analyse tout individu ayant au moins une valeur manquante à une période quelconque. C'est l'analyse des cas complets ("complete case analysis" ou "listwise deletion method"). Toutefois cette méthode peut exclure de l'analyse une quantité importante des données. Les sujets dont toutes les données sont observées ne sont pas un échantillon aléatoire de l'échantillon d'origine. Ainsi les données exclues ne sont pas manquantes complètement aléatoirement ("MCAR") même si les données non observées sont manquantes complètement aléatoirement. Les estimateurs qui en résultent pourront être biaisés.

Une approche moins radicale consiste à ne pas exclure de l'analyse tous les sujets dont certaines valeurs manquent, mais plutôt de porter l'analyse sur toutes les données observées et de considérer l'ensemble des données comme étant non équilibré, c'est-à-dire que les combinaisons des niveaux des facteurs ne sont pas toutes observées un même nombre de fois. Ceci est l'analyse des cas disponibles ("available case analysis" ou "pairwise deletion method"). Pour un échantillon basé sur N individus pour lesquels les informations sont prélevées à b périodes, la k^e composante μ_k du vecteur moyen et la k^e variance de la matrice de covariances sont estimées en utilisant toutes les données observées à la k^e période, $k = 1, \dots, b$. Le $(k, k')^e$ élément ($k \neq k', k = 1, \dots, b$ et $k' = 1, \dots, b$) de la matrice des variances-covariances est calculé en utilisant toutes les données observées simultanément pour les deux périodes k et k' . Cette méthode est plus efficace que la méthode

des cas complets dans la mesure où elle utilise plus d'information et les résultats issus de cette méthode sont non biaisés lorsque les données sont manquantes complètement aléatoirement (voir Little et Rubin, 1987).

Des méthodes statistiques usuelles peuvent être utilisées pour l'analyse des cas complets et l'analyse des cas disponibles, mais seule l'analyse des cas disponibles donne des estimateurs non biaisés avec des données manquantes ignorables (voir Little et Rubin, 1987).

Il y a deux classes de procédures disponibles pour des données manquantes : les procédures basées sur l'imputation et les procédures basées sur des modèles. En utilisant les données observées, l'objectif de la technique d'imputation est de "rectangulariser" l'information contenue dans les sujets observés pour que l'application de la technique de l'analyse des cas complets sur les données rectangularisées donne des réponses statistiquement valides. Avec cette procédure, les données manquantes sont remplacées et les données complètes ainsi obtenues sont analysées par des méthodes usuelles. Little et Rubin (1987) suggèrent plusieurs procédures d'imputation.

L'imputation "hot deck" est une méthode par laquelle les sujets sont classifiés dans des cellules basées sur certaines mesures de proximité en rapport avec le fait que la mesure Y_{ik} du sujet i soit observée ou manquante à la période k . Un sujet avec valeur observée est alors choisi aléatoirement pour imputer les caractéristiques non observées.

L'imputation non conditionnelle par la moyenne ("mean imputation") est une forme particulièrement simple. Les valeurs manquantes sont substituées par la moyenne des données observées pour les autres individus et pour la même période (voir Little et Rubin, 1987).

Comme une extension de l'imputation moyenne, Buck (1960) proposa l'imputation par la régression où les valeurs manquantes sont estimées par les valeurs prédites de la régression sur les données connues.

Pour un échantillon normal multivarié, la première étape consiste à déterminer le vecteur moyen μ et la matrice des variances-covariances Σ par la méthode des cas complets qui consiste à exclure de l'analyse tous les individus ayant une ou des valeurs manquantes pour certaines périodes. On suppose que Y est $N(\mu, \Sigma)$, c'est-à-dire que l'ensemble des données obtenu à partir de la méthode des cas complets est un sous-échantillon aléatoire des données complètes formé des données observées $Y^{(o)}$ et non observées $Y^{(m)}$. Pour un individu comportant des valeurs manquantes, la densité conditionnelle des composantes avec valeurs manquantes $Y_i^{(m)}$ sur les composantes avec valeurs observées $Y_i^{(o)}$ donne que la loi conditionnelle de $Y_i^{(m)}|Y_i^{(o)}$ est

$$N(\mu^{(m)} + \Sigma^{(mo)}(\Sigma^{(oo)})^{-1}(Y_i^{(o)} - \mu_i^{(o)}), \quad \Sigma^{(mm)} - \Sigma^{(mo)}(\Sigma^{(oo)})^{-1}\Sigma^{(om)})$$

où les indices (o) et (m) réfèrent aux valeurs "observées" et "manquantes" respectivement.

La seconde étape consiste à calculer la moyenne conditionnelle à partir des résultats de la densité conditionnelle et de la substituer aux valeurs manquantes. Buck (1960) a montré que cette méthode est valide lorsque le mécanisme des données manquantes est MCAR (manquante complètement aléatoirement). Little et Rubin (1987) affirme que cette méthode est valide sous certains types de mécanisme MAR (manquante aléatoirement).

Avec l'imputation stochastique par la régression, une valeur manquante est remplacée par la valeur prédite par la régression plus une erreur tirée aléatoirement pour refléter le hasard dans la valeur prédite. L'imputation simple, c'est-à-dire imputer une valeur pour chaque valeur manquante, nécessite un grand échantillon pour avoir des bonnes estimations. Elle traite les données imputées sans ajustement et sans considérer la variabilité. Une sous estimation des erreurs standards peut se produire même si le modèle utilisé pour générer les imputations est correct. Rubin (1987) proposa la méthode d'imputation multiple pour surmonter ces désavantages. Au lieu d'avoir un ensemble de données imputées,

plusieurs ensembles d'imputation sont tirés indépendamment à plusieurs reprises sous un modèle pour les données manquantes. Chaque ensemble des valeurs imputées et des valeurs observées donne un ensemble complet de données. Les mêmes inférences sur des données complètes sont alors exécutées pour chaque ensemble de données. L'inférence combinée reflète le hasard dans les données manquantes. Cette technique donne des estimateurs non biaisés aussi bien pour les données manquantes ignorables que pour les données manquantes non ignorables (Rubin, 1987).

Pour des procédures basées sur des modèles, on distingue deux approches pour formuler les modèles pour les données manquantes : le modèle de sélection ("selection model") et le modèle de mélange ("mixture model"). Dans le modèle de sélection, la fonction de densité conjointe de M_i et Y_i étant donné X_i est :

$$f(Y_i, M_i|X_i) = f(Y_i|X_i)f(M_i|Y_i, X_i) \quad (1.4.1)$$

où la deuxième composante modélise l'incidence des données manquantes comme fonction de Y_i et de X_i . Alternativement le modèle de mélange donne la fonction de densité conjointe de M_i et de Y_i étant donné X_i comme suit :

$$f(Y_i, M_i|X_i) = f(Y_i|X_i, M_i)f(M_i|X_i) \quad (1.4.2)$$

où la distribution de Y_i est caractérisée conditionnellement à X_i et M_i , et l'incidence des données manquantes est une fonction de X_i seulement. Pour le modèle de sélection, la méthode d'estimation à vraisemblance maximale peut être utilisée. Dans la plupart des cas des techniques itératives sont nécessaires pour maximiser la vraisemblance. Pour le modèle de mélange, la fonction de densité conditionnelle $f(Y_i|X_i, M_i)$ se rapporte aux données manquantes, ainsi il n'y a pas de données pour l'estimer.

Il y a deux types de données manquantes pour les données à mesures répétées : les "drop outs" et les données manquantes intermittentes. On parle des "drop-outs" lorsque des séquences de mesures répétées sur des sujets se terminent de façon

prématurées, c'est-à-dire si Y_{ik} est manquante, $Y_{ik'}$ est aussi manquante pour $k' > k$. Dans le cas contraire les données manquantes sont dites intermittentes. Ce mémoire traite le cas des données manquantes intermittentes où le mécanisme des données manquantes est connu.

1.5. L'IMPUTATION MULTIPLE.

L'imputation multiple se fait en trois étapes. La première étape consiste à créer M ($M > 1$) ensembles de données complètes en substituant les données manquantes M fois en utilisant M tirages indépendants d'un modèle d'imputation. Le modèle d'imputation est construit pour permettre une approximation raisonnable de la vraie relation entre la distribution des valeurs manquantes et la distribution des valeurs observées. Dans la deuxième étape, les M ensembles de données complètes imputés sont analysés en traitant chaque ensemble de données complètes imputé comme un réel ensemble de données complètes. Les logiciels et les procédures usuelles pour les données complètes peuvent être utilisés directement. Dans la troisième étape, les résultats de l'analyse basée sur les M ensembles de données complètes sont combinés pour obtenir les inférences de l'imputation répétée (Rubin, 1987). Les variances des estimations combinées se composent des variances intra et inter imputation. Ainsi l'incertitude des données imputées est incorporée dans l'inférence finale. Cette méthode surmonte les inconvénients de l'imputation simple qui sous estime l'erreur standard de l'estimation. La première étape, le tirage aléatoire des échantillons à partir d'un modèle d'imputation est le plus fondamental pour l'imputation multiple. Une imputation multiple est dite appropriée (Rubin, 1987) si (1) le modèle d'imputation préserve les corrélations entre les valeurs observées et les valeurs manquantes, (2) les valeurs manquantes sont imputées par des tirages aléatoires indépendants issus du modèle d'imputation, et (3) l'estimation de θ , le paramètre inconnu du modèle d'imputation, est approximativement sans biais pour la distribution des données imputées.

Si le modèle d'imputation ne préserve pas les corrélations entre les valeurs observées et les valeurs manquantes, les inférences au sujet de ces corrélations, basées sur des données complètes imputées, sont biaisées. Si l'imputation multiple n'est pas basée sur les tirages aléatoires issus du modèle d'imputation, la variance inter imputation qui en résulte est sous estimée. Enfin, l'estimation approximativement sans biais de θ pour le modèle d'imputation assure que les inférences basées sur des imputations répétées sont valides du point de vue fréquentiste.

La théorie de l'inférence de l'imputation répétée ou combinée provient d'un modèle bayésien (Rubin , 1987). Soit Q , une quantité scalaire générique à estimer telle que l'effet du traitement, le rapport de cotes (odds ratio) ou les coefficients de régression. Alors la distribution *a posteriori* des données observées de Q est donnée par :

$$P(Q|Y^{(o)}) = \int P(Q|Y^{(o)}, Y^{(m)})P(Y^{(m)}|Y^{(o)})dY^{(m)}$$

où $Y^{(o)}$ et $Y^{(m)}$ sont respectivement le vecteur des données observées et le vecteur des données manquantes. La distribution *a posteriori* des données observées est le produit de la distribution moyenne *a posteriori* des données complètes avec la distribution prédictive. Ainsi les moments peuvent être obtenus à partir de la distribution *a posteriori* des données observées :

$$E(Q|Y^{(o)}) = E[E(Q|Y^{(o)}, Y^{(m)})|Y^{(o)}]$$

et

$$V(Q|Y^{(o)}) = E[V(Q|Y^{(o)}, Y^{(m)})|Y^{(o)}] + V[E(Q|Y^{(o)}, Y^{(m)})|Y^{(o)}]$$

où \hat{Q} et \hat{U} sont respectivement des estimateurs de Q et $V(Q)$ basées sur les données complètes imputées (voir Schafer, 1997). Le calcul des données imputées estimées est basé sur les M imputations indépendantes $\hat{Q}_{(k)} = \hat{Q}(y^{(o)}, y_{(k)}^{(m)})$ avec leurs variances estimées $\hat{U}_{(k)} = \hat{U}(y^{(o)}, y_{(k)}^{(m)})$, $k = 1, \dots, M$. L'estimation combinée

de Q est simplement la moyenne de ces estimations imputées

$$\bar{Q} = \frac{1}{M} \sum_{k=1}^M \hat{Q}_{(k)}$$

et la variance totale de \bar{Q} est

$$T = [(1 + M^{-1})B + \bar{U}]$$

où $B = \frac{1}{M-1} \sum_{k=1}^M (\hat{Q}_{(k)} - \bar{Q})^2$ est la variance inter imputation et $\bar{U} = \frac{1}{M} \sum_{k=1}^M \hat{U}_{(k)}$ est la variance intra imputation. Le test d'hypothèse et l'intervalle de confiance sont alors basés sur une approximation de la distribution t de Student

$$(Q - \bar{Q})/T^{1/2} \sim t_{\nu}$$

avec degrés de liberté

$$\nu = (M - 1)[1 + r^{-1}]^2$$

où r est la croissance relative de la variance due aux données manquantes :

$$r = (1 + M^{-1})B/\bar{U}.$$

Dans le cas où l'imputation n'est pas multiple, la variance inter imputation B est zéro et la variance intra imputation \bar{U} est la variance des données complètes. Ainsi r est la mesure de la croissance de la variance causée par l'imputation multiple. La valeur de l'indice r diminue tandis que la valeur de ν augmente au fur et à mesure que le nombre d'imputation M augmente. Si $B = 0$, alors $r = 0$, ν tend vers l'infini et par conséquent $(Q - \bar{Q})/T^{1/2}$ suit une loi normale. Un autre indice de diagnostic important pour l'imputation multiple est la fraction de l'information manquante pour Q :

$$\hat{\lambda} = \frac{r + 2/(\nu + 3)}{r + 1}$$

(voir Schafer, 1997). Cet indice mesure la façon dont les données manquantes influencent l'incertitude de l'estimation de Q . Sa valeur est déterminée par la proportion des données manquantes et la performance de l'imputation multiple.

Lorsque le nombre d'imputation M augmente, la valeur de $\hat{\lambda}$ diminue. L'inférence de l'imputation répétée ci-dessus est valide aussi longtemps que les imputations $y_{(1)}^{(m)}, y_{(2)}^{(m)}, \dots, y_{(M)}^{(m)}$ sont correctement générées. Les imputations doivent en moyenne donner des prédictions raisonnables pour les valeurs manquantes et la variation entre les imputations doit refléter un degré d'incertitude approprié. Pour plus de détails au sujet de l'imputation multiple, voir Rubin(1987) et Schafer(1997).

1.6. LA VRAISEMBLANCE MAXIMALE

La vraisemblance maximale est une approche générale de l'estimation statistique qui peut être utilisée pour traiter le problème des données manquantes. Le principe de base de l'estimation à vraisemblance maximale est de choisir comme estimateurs, les valeurs qui maximiseraient la probabilité d'observer ce qui a en réalité été observé. Pour accomplir cela, il faut une expression qui exprime la probabilité des données comme fonction de données et des paramètres inconnus. Lorsque les observations d'un échantillon sont indépendantes, la vraisemblance globale pour l'échantillon est simplement le produit des vraisemblances de chacune des observations de l'échantillon. Supposons que nous essayons d'estimer un paramètre θ . Si Y est un vecteur aléatoire, y un échantillon de n observations et $f(y|\theta)$ la fonction de densité conjointe ou la fonction de probabilité conjointe de y par rapport à θ , la vraisemblance pour un échantillon y de n observations est

$$L(\theta) = \prod_{i=1}^n f(y_i|\theta).$$

L'estimateur à vraisemblance maximale (EVM) de θ est la valeur $\hat{\theta}$ qui maximise la fonction $L(\theta)$. Les estimateurs à vraisemblance maximale ont plusieurs propriétés. Sous certaines conditions, ils sont convergents, asymptotiquement efficaces et asymptotiquement normaux. La convergence implique que les estimations sont

approximativement non biaisées pour des grands échantillons. L'efficacité implique que les vraies erreurs standards sont au moins aussi petites que les erreurs standards pour tout autre estimateur convergent.

Lorsque les données sont incomplètes et le mécanisme des données manquantes est non informatif, la vraisemblance des données observées de θ est donnée par

$$L(\theta|Y^{(o)}) = \int P(Y^{(o)}, Y^{(m)}|\theta) dY^{(m)} \quad (1.6.1)$$

où $Y^{(o)}$ et $Y^{(m)}$ sont respectivement le vecteur des données observées et le vecteur des données manquantes. L'estimateur à vraisemblance maximale de θ est obtenu dans ce cas en maximisant la fonction de vraisemblance des données observées par rapport à θ . Formellement, il n'y a pas de différence entre la méthode d'estimation à vraisemblance maximale pour les données complètes et pour les données incomplètes (Little et Rubin, 1987, p88).

Les procédures de l'estimation à vraisemblance maximale peuvent être récapitulées de la manière suivante (Little et Rubin, 1987, p128) : si la fonction de vraisemblance est différentiable et unimodale, l'EVM de θ peut être obtenue en résolvant l'équation de vraisemblance

$$S(\theta|Y^{(o)}) \equiv \frac{\partial \ell(\theta|Y^{(o)})}{\partial \theta} = 0 \quad (1.6.2)$$

où $\ell(\theta|Y^{(o)}) = \ln L(\theta|Y^{(o)})$ est la fonction log-vraisemblance des données observées et $S(\theta|Y^{(o)})$ la fonction de score des données observées. Lorsqu'une solution de l'équation (1.6.2) ne peut être trouvée, des méthodes itératives seront appliquées.

L'algorithme de Newton-Raphson est l'une de ces méthodes : étant donné l'estimation initiale de θ , disons $\theta^{(0)}$, soit $\theta^{(t)}$ l'estimation à la t^e itération, $t = 1, 2, \dots$, alors la $(t + 1)^e$ itération est définie par l'équation

$$\theta^{(t+1)} = \theta^{(t)} + I^{-1}(\theta^{(t)}|Y^{(o)})S(\theta^{(t)}|Y^{(o)}) \quad (1.6.3)$$

où $I(\theta|Y^{(o)})$ est la matrice de l'information observée, c'est-à-dire la dérivée de second ordre du log-vraisemblance des données observées $\ell(\theta|Y^{(o)})$ par rapport à θ :

$$I(\theta|Y^{(o)}) = -\frac{\partial^2 \ell(\theta|Y^{(o)})}{\partial \theta \partial \theta'}.$$

Sous des hypothèses générales sur la fonction de vraisemblance des données observées (1.6.1), la séquence des itérations $\theta^{(t)}$ converge vers une estimation ponctuelle $\hat{\theta}$, qui est l'EVM de θ .

Une variante de cette procédure est la méthode de score, où l'information observée en (1.6.3) est remplacée par l'information attendue

$$J(\theta) = E[I(\theta|Y^{(o)})|\theta] = -\int \frac{\partial^2 \ell(\theta|Y^{(o)})}{\partial \theta \partial \theta'} P(Y^{(o)}|\theta) dY^{(o)}.$$

Pour ces deux méthodes, il faut déterminer la matrice des dérivées de second ordre de la fonction log-vraisemblance.

Une troisième alternative d'algorithme (Berndt *et al.*, 1974) est de remplacer l'information observée en (1.6.3) par la matrice de covariances échantillonnale de la fonction de score $S(\theta|Y^{(o)})$, puisque cette matrice est une estimation convergente de la matrice d'information dans le voisinage de $\hat{\theta}$. L'équation itérative qui en résulte est

$$\theta^{(t+1)} = \theta^{(t)} + \lambda_t Q^{-1}(\theta^{(t)}) S(\theta^{(t)}|Y^{(o)})$$

où $Q(\theta) = \sum_{i=1}^n (\partial \ell_i / \partial \theta)(\partial \ell_i / \partial \theta)'$, ℓ_i est la vraisemblance des données observées pour la i^e observation et λ_t est une échelle positive conçue pour assurer la convergence vers un maximum local (voir Berndt *et al.*, 1974).

Ces méthodes itératives vont générer des estimateurs à vraisemblance maximale de θ non biaisés aussi longtemps que le mécanisme des données manquantes peut être ignoré sans risque. Ces méthodes exigent le calcul ou l'approximation des dérivées de second ordre de la fonction de vraisemblance des données observées. Lorsque les données sont incomplètes, la vraisemblance des données

observées (1.6.1) peut être compliquée et la matrice des dérivées de second ordre a tendance à être une fonction compliquée de θ . Par conséquent, le calcul ou l'approximation de cette matrice peut devenir très compliqué (Little and Rubin, 1987, p128). Une approche alternative, laquelle n'exige pas le calcul ou l'approximation des dérivées de second ordre est l'algorithme EM (Dempster *et al.*, 1977).

1.6.1. L'algorithme EM

L'algorithme EM complète les données manquantes $Y^{(m)}$ sur la base de l'estimation de θ et ré-estime ensuite θ sur la base de $Y^{(o)}$ et les $Y^{(m)}$ complétées en procédant de façon itérative jusqu'à ce que les estimateurs convergent. Le terme EM (Expectation-Maximization) fut introduit par Dempster *et al.*, (1977).

Le fondement théorique de l'algorithme EM est le suivant. Le modèle de probabilité des données complètes peut être factorisé comme

$$P(Y|\theta) = P(Y^{(o)}|\theta)P(Y^{(m)}|Y^{(o)}, \theta) \quad (1.6.4)$$

et le log-vraisemblance peut être exprimé comme

$$\ell(\theta|Y) = \ell(\theta|Y^{(o)}) + \ln P(Y^{(m)}|Y^{(o)}, \theta) + c \quad (1.6.5)$$

où $\ell(\theta|Y) = \ln P(Y|\theta)$ est le log-vraisemblance des données complètes, le terme $P(Y^{(m)}|Y^{(o)}, \theta)$ est la distribution conditionnelle de $Y^{(m)}$ étant donné $Y^{(o)}$ et θ , c est une constante arbitraire. Lorsque vu comme la distribution des probabilités de $Y^{(m)}$, il résume l'information au sujet de $Y^{(m)}$ pour toute valeur de θ donnée ; lorsque vu comme une fonction de θ , il apporte l'information au sujet de θ qui est contenue dans $Y^{(m)}$ au delà de celle déjà fournie par $Y^{(o)}$. Le log-vraisemblance des données complètes $\ell(\theta|Y)$ ne peut pas être calculé directement parce que $P(Y^{(m)}|Y^{(o)}, \theta)$ ne peut pas être calculé car les $Y^{(m)}$ sont inconnues. Au lieu de cela, le log-vraisemblance attendu pour la distribution des données manquantes est calculé. Étant donné les données observées $Y^{(o)}$ et l'estimateur $\theta^{(t)}$ de θ à

l'étape t , $t = 1, 2, \dots$,

$$Q(\theta|\theta^{(t)}) = \ell(\theta|Y^{(o)}) + H(\theta|\theta^{(t)}) + c \quad (1.6.6)$$

où

$$Q(\theta|\theta^{(t)}) = \int \ell(\theta|Y)P(Y^{(m)}|Y^{(o)}, \theta^{(t)})dY^{(m)}$$

et

$$H(\theta|\theta^{(t)}) = \int \ln P(Y^{(m)}|Y^{(o)}, \theta)P(Y^{(m)}|Y^{(o)}, \theta^{(t)})dY^{(m)}.$$

Supposant que $\theta^{(t+1)}$ soit la valeur qui maximise $Q(\theta|\theta^{(t)})$, c'est-à-dire

$$Q(\theta|\theta^{(t)}) \leq Q(\theta^{(t+1)}|\theta^{(t)}) \quad \forall \theta,$$

Dempster *et al.*, (1977) ont montré que $\theta^{(t+1)}$ est un meilleur estimateur que $\theta^{(t)}$ dans le sens que le log-vraisemblance de ses données observées est au moins aussi grand que celui de $\theta^{(t)}$:

$$\ell(\theta^{(t)}|Y^{(o)}) \leq \ell(\theta^{(t+1)}|Y^{(o)}) \quad \forall \theta. \quad (1.6.7)$$

Donc, la séquence $\theta^{(t)}$ augmente les valeurs du log-vraisemblance des données observées à chaque itération et si le log-vraisemblance des données observées $\ell(\theta|Y^{(o)})$ converge vers un maximum, la séquence $\ell(\theta^{(t)}|Y^{(o)})$ converge vers une valeur stationnaire de $\ell(\theta|Y^{(o)})$.

L'itération de EM consiste en deux étapes distinctes. L'étape E ("Expectation") consiste à trouver l'espérance conditionnelle des données manquantes étant donné les données observées et l'estimateur courant de θ , et ensuite à remplacer les données manquantes par ces espérances. Remarquons que ce n'est pas la donnée manquante qui est substituée par son espérance, mais plutôt la fonction des données manquantes apparaissant dans le log-vraisemblance des données complètes qui est substituée. Spécifiquement, l'étape E consiste à trouver la valeur attendue du log-vraisemblance des données complètes pour la distribution des

données manquantes $Y^{(m)}$ étant donné les données observées $Y^{(o)}$ et l'estimation courante de θ , soit $\theta^{(t)}$, c'est-à-dire

$$Q(\theta|\theta^{(t)}) = \int \ell(\theta|Y)P(Y^{(m)}|Y^{(o)}, \theta^{(t)})dY^{(m)}.$$

L'étape M ("Maximisation") consiste à déterminer l'estimateur à vraisemblance maximale de θ comme s'il n'y a pas de données manquantes, c'est-à-dire comme si la maximisation est basée sur le log-vraisemblance des données complètes. Ainsi la maximisation de $Q(\theta|\theta^{(t)})$ utilise la même méthode de calcul que pour la maximisation du log-vraisemblance des données complètes. Exécuter alternativement les étapes E et M en commençant avec une valeur $\theta^{(0)}$ génère une séquence des valeurs d'estimateurs $\theta^{(t)}$ et la séquence des valeurs du log-vraisemblance des données observées $\ell(\theta^{(t)}|Y^{(o)})$. Dempster *et al.*, (1977) ont fourni les conditions sous lesquelles la suite de log-vraisemblance converge presque sûrement vers une valeur stationnaire du log-vraisemblance des données observées. La convergence est linéaire avec un taux proportionnel à la fraction de l'information observée au sujet de θ qui est contenue dans $\ell(\theta|Y)$. La vraisemblance maximale peut être une méthode pratique et efficace pour traiter les données qui sont manquantes aléatoirement (MAR). Dans ce cas, les estimateurs à vraisemblance maximale sont optimaux pour des grands échantillons.

Chapitre 2

PLAN D'EXPÉRIENCE À DEUX FACTEURS DONT UN COMPORTE DES MESURES RÉPÉTÉES.

2.1. INTRODUCTION

Les plans d'expériences à mesures répétées utilisent le même sujet pour chaque niveau d'un ou plusieurs facteurs. Le sujet sert comme un bloc et les unités expérimentales à l'intérieur du bloc peuvent être vues comme différentes périodes pendant lesquelles un traitement est appliqué au sujet. Un plan d'expérience à mesures répétées peut concerner un ou plusieurs traitements évalués sous diverses conditions expérimentales ou à des périodes différentes dans le temps.

L'avantage principal des plans à mesures répétées est qu'ils fournissent une bonne précision pour la comparaison des niveaux du facteur temps car toutes les sources de variabilité entre les sujets sont exclues de l'erreur expérimentale. Seulement la variation intra-sujet est présente dans l'erreur expérimentale puisque chaque deux niveaux du facteur temps peuvent être comparés directement pour un même sujet. Un autre avantage des plans à mesures répétées est qu'ils nécessitent moins de sujets. Cet aspect est particulièrement important lorsque seulement un petit nombre de sujets peut participer à l'expérience. De plus, lorsque

l'intérêt est dans les effets du traitement dans le temps, il est souvent souhaitable d'observer le même sujet à différentes périodes dans le temps.

Cependant, les plans à mesures répétées ont aussi un sérieux désavantage potentiel car ils peuvent comporter plusieurs types d'interférence. Comme exemples, on peut citer l'effet de la séquence qui est défini comme l'effet d'interférence due à l'ordre dans lequel les traitements sont administrés aux individus et l'effet d'interférence relié aux traitements précédents. Plusieurs dispositions peuvent être prises pour réduire les effets d'interférence. La randomisation des ordres des traitements indépendamment pour chaque sujet permet d'analyser les données comme si les termes d'erreur étaient indépendants. La réduction de l'effet d'interférence relié aux traitements précédents se fait en augmentant dans l'expérience l'intervalle de temps entre deux traitements.

Un plan à mesures répétées est dit équilibré si les périodes de prélèvement des mesures sont les mêmes pour toutes les unités expérimentales. Un plan à mesures répétées est dit complet si les mesures sont disponibles pour chaque période et pour chaque unité expérimentale. Dans ce chapitre nous considérons un plan d'analyse de variance à deux facteurs dont le premier facteur (facteur A avec a niveaux) est constitué de groupes d'individus et le deuxième facteur (facteur B avec b niveaux) renferme des mesures répétées sur tous les individus de chaque niveau du facteur A . Chaque individu doit être assigné à un seul niveau (groupe) du facteur A de façon aléatoire. Pour le facteur B , l'ordre des traitements (niveaux des mesures répétées) doit être distribué aléatoirement pour chaque individu. Notant cependant que la randomisation ne s'applique pas au facteur B à mesures répétées lorsque celui-ci est un facteur temps. De même, lorsque le facteur A est une caractéristique des individus telles que l'âge, le poids ou la taille, il n'est pas possible de faire une randomisation pour ce facteur (voir Neter *et al.*, 1996, p 1185).

Les unités expérimentales sont issues de a populations différentes ou d'une population comportant a sous populations. En choisissant au hasard n_j individus

dans chaque sous population, on obtient a groupes de sujets. À chacun des n_j individus de chacun des a groupes, on prend b mesures dans le temps.

Ainsi le plan d'expérience à deux facteurs comportant des mesures répétées sur un facteur peut être représenté schématiquement de la manière suivante :

	B_1	...	B_k	...	B_b
A_1	G_1	...	G_1	...	G_1
.					
.					
.					
A_j	G_j	...	G_j	...	G_j
.					
.					
.					
A_a	G_a	...	G_a	...	G_a

Les symboles $G_1, \dots, G_j, \dots, G_a$ représentent des groupes de $n_1, \dots, n_j, \dots, n_a$ sujets chacun respectivement. Les sujets du groupe G_1 sont observés sous les combinaisons des traitements $A_1B_1, \dots, A_1B_k, \dots, A_1B_b$, les sujets du groupe G_j sont observés sous les combinaisons des traitements $A_jB_1, \dots, A_jB_k, \dots, A_jB_b$ et les sujets du groupe G_a sont observés sous les combinaisons des traitements $A_aB_1, \dots, A_aB_k, \dots, A_aB_b$. Les observations des n_j sujets de chaque groupe G_j ($j = 1, \dots, a$) peuvent être représentées schématiquement de la manière suivante :

groupe	sujet	B_1	...	B_k	...	B_b
G_j	1	Y_{1j1}	...	Y_{1jk}	...	Y_{1jb}
	.					
	.					
	.					
	i	Y_{ij1}	...	Y_{ijk}	...	Y_{ijb}
	.					
	n_j	Y_{n_jj1}	...	Y_{n_jjk}	...	Y_{n_jjb}

Pour ce type de plan d'expérience, les comparaisons entre les moyennes pour les niveaux du facteur A impliquent aussi bien les différences entre les sujets que les différences associées aux différents niveaux du facteur A . Les comparaisons entre les moyennes des niveaux du facteur B (temps) pour le même niveau du facteur A sont basées sur les mêmes sujets, impliquant donc seulement les différences liées aux temps. Ainsi, pour ces dernières comparaisons, chaque sujet sert comme son propre contrôle. On dit que les effets du facteur A sont confondus avec les différences entre les groupes de sujets, tandis que les effets du facteur B ne comportent pas ce genre de confusion. Pour cette raison, les tests sur les effets du facteur B sont plus sensibles (la probabilité de rejeter une hypothèse nulle fautive est plus élevée) que les tests sur les effets du facteur A .

2.2. LE MODÈLE MATHÉMATIQUE.

Soit Y_{ijk} , la mesure prise au k^e temps pour le i^e individu appartenant au j^e groupe, $i = 1, \dots, n_j$; $j = 1, \dots, a$; $k = 1, \dots, b$.

$$\text{Posons : } N = \sum_j n_j, \quad \bar{Y}_{ij.} = \frac{1}{b} \sum_k Y_{ijk}, \quad \bar{Y}_{.jk} = \frac{1}{n_j} \sum_i Y_{ijk}, \quad \bar{Y}_{.j.} = \frac{1}{n_j b} \sum_i \sum_k Y_{ijk},$$

$$\bar{Y}_{..k} = \frac{1}{N} \sum_i \sum_j Y_{ijk} \quad \text{et} \quad \bar{Y}_{...} = \frac{1}{bN} \sum_i \sum_j \sum_k Y_{ijk}.$$

Le développement du modèle pour un plan d'expérience à deux facteurs comportant des mesures répétées sur un facteur se fait en introduisant dans le modèle les effets aléatoires des sujets, les effets fixes du facteur A , les effets fixes du facteur B et les effets de l'interaction entre les facteurs A et B .

Le modèle pour le plan d'expérience à deux facteurs dont un facteur comporte les mesures répétées est le suivant :

$$Y_{ijk} = \mu + \alpha_j + \beta_k + \rho_{i(j)} + (\alpha\beta)_{jk} + \varepsilon_{ijk} \quad (2.2.1)$$

où μ est la moyenne générale, μ représente un effet général, les α_j sont des constantes vérifiant la condition $\sum_j \alpha_j = 0$, α_j représente l'effet du facteur A pour le j^e groupe, les β_k sont des constantes vérifiant la condition $\sum_k \beta_k = 0$, β_k représente l'effet du facteur B pour le k^e temps, les $(\alpha\beta)_{jk}$ sont des constantes vérifiant la condition $\sum_j (\alpha\beta)_{jk} = 0$ pour tout k et $\sum_k (\alpha\beta)_{jk} = 0$ pour tout j , $(\alpha\beta)_{jk}$ est l'effet de l'interaction AB à l'intersection du j^e groupe et du k^e temps, le terme $\rho_{i(j)}$ est l'effet du i^e sujet appartenant au j^e groupe, ε_{ijk} représente l'erreur au k^e temps pour le i^e individu appartenant au j^e groupe, α_j est estimé par $(\bar{Y}_{.j} - \bar{Y}_{...})$, β_k par $(\bar{Y}_{..k} - \bar{Y}_{...})$, $(\alpha\beta)_{jk}$ par $(\bar{Y}_{.jk} - \bar{Y}_{.j} - \bar{Y}_{..k} + \bar{Y}_{...})$ et $\rho_{i(j)}$ par $(\bar{Y}_{ij.} - \bar{Y}_{.j.})$.

Les variables aléatoires $\rho_{i(j)}$ et ε_{ijk} sont supposées mutuellement indépendantes, normalement distribuées, de moyennes nulles et de variances respectives σ_ρ^2 et σ_ε^2 . Les observations Y_{ijk} du modèle (2.2.1) ont alors les propriétés suivantes :

$$E(Y_{ijk}) = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk}, \quad (2.2.2a)$$

$$Var(Y_{ijk}) = \sigma_Y^2 = Var(\rho_{i(j)} + \varepsilon_{ijk}) = \sigma_\rho^2 + \sigma_\varepsilon^2, \quad (2.2.2b)$$

$$Cov(Y_{ijk}, Y_{ijk'}) = \sigma_\rho^2 \quad k \neq k', \quad (2.2.2c)$$

$$Cov(Y_{ijk}, Y_{i'j'k'}) = 0 \quad i \neq i' \quad \text{et/ou} \quad j \neq j'. \quad (2.2.2d)$$

Notons que les observations Y_{ijk} ont une variance constante. Chaque deux observations de différents niveaux du facteur B pour le même sujet ont une covariance constante pour tous les sujets, tandis que les observations pour les sujets différents sont indépendantes.

2.3. ANALYSE DE VARIANCE.

La somme des carrés totale de l'analyse de variance pour le modèle à mesures répétées (2.2.1) est donnée par :

$$SC_{totale} = \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{...})^2$$

et peut être décomposée en deux parties. Une première partie est attribuable à la variation intra-sujet, c'est-à-dire la variation entre les observations du même sujet pour divers temps :

$$SC_{intra} = \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{ij.})^2.$$

La deuxième partie est attribuable à la variation entre les sujets, c'est la somme des carrés inter-sujets :

$$SC_{inter} = b \sum_i \sum_j (\bar{Y}_{ij.} - \bar{Y}_{...})^2.$$

Nous avons donc la relation $SC_{totale} = SC_{intra} + SC_{inter}$.

D'une part, la somme des carrés intra-sujet se décompose en trois parties. Une première partie nommée la somme des carrés des temps (Facteur B) :

$$SCB = N \sum_k (\bar{Y}_{..k} - \bar{Y}_{...})^2$$

est la variation attribuable aux différences entre les temps. La deuxième partie est attribuable à l'interaction groupe*temps (Interaction AB) :

$$SCAB = \sum_j n_j \sum_k (\bar{Y}_{.jk} - \bar{Y}_{.j.} - \bar{Y}_{..k} + \bar{Y}_{...})^2$$

et une troisième partie nommée la somme des carrés résiduelle :

$$SCB.S(A) = \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{.jk} - \bar{Y}_{ij.} + \bar{Y}_{.j.})^2$$

est attribuable à l'interaction temps*individu à l'intérieur des groupes. Nous pouvons écrire $SC_{intra} = SCB + SCAB + SCB.S(A)$.

D'autre part, la somme des carrés inter-sujets se décompose en deux parties : une première partie attribuable aux différences entre les groupes (Facteur A) :

$$SCA = b \sum_j n_j (\bar{Y}_{.j.} - \bar{Y}_{...})^2$$

et une seconde partie attribuable aux différences entre les individus à l'intérieur des groupes :

$$SCS(A) = b \sum_i \sum_j (\bar{Y}_{ij.} - \bar{Y}_{.j.})^2,$$

c'est la somme des carrés résiduelle due aux groupes. Ainsi,

$SC_{inter} = SCA + SCS(A)$. La somme des carrés utilisée pour estimer la variance de l'erreur s'avère être la somme des carrés de l'interaction $SCB.S(A)$ (voir Neter *et al.*, 1996).

Le tableau ANOVA suivant présente les sommes des carrés de l'analyse de variance et les degrés de liberté correspondants au plan d'expérience du modèle à mesures répétées (2.2.1).

La moyenne des carrés (MC) est le rapport de la somme des carrés sur le nombre de degrés de liberté correspondant. Les moyennes des carrés attendues (EMC) pour l'analyse de variance du tableau 2.3.1 sont présentées dans le tableau 2.3.2.

TAB. 2.3.1. Tableau des sommes des carrés des facteurs.

Source de variation	Somme des carrés	Degrés de liberté
Facteur A	SCA	$a - 1$
Facteur B	SCB	$b - 1$
Interaction AB	$SCAB$	$(a - 1)(b - 1)$
Sujets	$SCS(A)$	$N - a$
Erreur	$SCB.S(A)$	$(N - a)(b - 1)$
Totale	SC_{totale}	$bN - 1$

TAB. 2.3.2. Tableau des moyennes des carrés des facteurs.

Source de variation	Moyenne des carrés MC	EMC
Facteur A	MCA	$\sigma_{\epsilon}^2 + b\sigma_{\rho}^2 + b\frac{\sum_j n_j \alpha_j^2}{a-1}$
Facteur B	MCB	$\sigma_{\epsilon}^2 + N\frac{\sum_k \beta_k^2}{b-1}$
Interaction AB	$MCAB$	$\sigma_{\epsilon}^2 + \frac{\sum_j n_j \sum_k (\alpha\beta)_{jk}^2}{(a-1)(b-1)}$
Sujets	$MCS(A)$	$\sigma_{\epsilon}^2 + b\sigma_{\rho}^2$
Erreur	$MCB.S(A)$	σ_{ϵ}^2

2.4. TESTS DES EFFETS DES FACTEURS.

Pour tester les effets de l'interaction AB :

$$H_0 : (\alpha\beta)_{jk} = 0, \text{ pour tout } j = 1, \dots, a \text{ et } k = 1, \dots, b \quad (2.4.1)$$

$$\text{contre } H_{ab} : \text{il existe } j \text{ et } k \text{ tel que } (\alpha\beta)_{jk} \neq 0, \quad (2.4.2)$$

on utilise la statistique

$$F^* = \frac{MCAB}{MCB.S(A)}. \quad (2.4.3)$$

La règle de décision pour contrôler l'erreur de type 1 au niveau α est : rejeter H_0 si

$$F^* > F[1 - \alpha; (a - 1)(b - 1), (N - a)(b - 1)]$$

et ne pas rejeter H_0 sinon. La quantité $F[1 - \alpha; (a - 1)(b - 1), (N - a)(b - 1)]$ représente le percentile d'ordre $100(1 - \alpha)$ de la distribution F avec $(a - 1)(b - 1)$ et $(N - a)(b - 1)$ degrés de liberté respectivement au numérateur et au dénominateur. Le test pour les effets du facteur A est défini de façon analogue :

$$H_0 : \alpha_j = 0 \quad \text{pour tout } j = 1, \dots, a \quad (2.4.4)$$

$$\text{contre } H_a : \text{ il existe } j \text{ tel que } \alpha_j \neq 0. \quad (2.4.5)$$

La statistique du test est :

$$F^* = \frac{MCA}{MCS(A)}, \quad (2.4.6)$$

et la règle de décision pour contrôler l'erreur de type 1 au niveau α est : rejeter H_0 si

$$F^* > F[1 - \alpha; a - 1, N - a]$$

et ne pas rejeter H_0 sinon.

Enfin le test pour les effets du facteur B :

$$H_0 : \beta_k = 0 \quad \text{pour tout } k = 1, \dots, b \quad (2.4.7)$$

$$\text{contre } H_b : \text{ il existe } k \text{ tel que } \beta_k \neq 0. \quad (2.4.8)$$

La statistique du test est

$$F^* = \frac{MCB}{MCB.S(A)}, \quad (2.4.9)$$

et la règle de décision pour contrôler l'erreur de type 1 au niveau α est : rejeter H_0 si

$$F^* > F[1 - \alpha; b - 1, (N - a)(b - 1)]$$

et ne pas rejeter H_0 sinon.

2.5. VÉRIFICATION DES PRÉSUPPOSÉS DU MODÈLE À MESURES RÉPÉTÉES.

Une série de graphiques permet de vérifier si le modèle choisi est adéquat. Le graphique des résidus par individu permettent d'évaluer si la variance des erreurs est constante et s'il y a présence des effets d'interférence et la normalité peut être détectée à l'aide des graphiques de probabilité normale. Le modèle (2.2.1) exige que la variance inter-sujets soit constante pour tous les niveaux du facteur A . Cette hypothèse peut être examinée par les nuages des points des effets estimés $(\bar{Y}_{ij} - \bar{Y}_{.j})$ des sujets à l'intérieur de chaque niveau du facteur A .

Nous pouvons aussi tester l'égalité des variances inter-sujets en notant que la variabilité entre les sujets à l'intérieur des groupes du facteur A , $SCS(A)$ peut être décomposée en terme des composantes des niveaux du facteur A :

$$SCS(A) = SCS(A_1) + SCS(A_2) + \dots + SCS(A_a) \quad (2.5.1)$$

où

$$SCS(A_j) = b \sum_i (\bar{Y}_{ij} - \bar{Y}_{.j})^2, \quad j = 1, \dots, a. \quad (2.5.2)$$

Chaque composante des sommes des carrés possède $(n_j - 1)$ degrés de liberté. Nous pouvons donc faire un test d'égalité des variances inter-sujets au moyen du test de Hartley lorsque les tailles d'échantillons sont égales ou du test modifié de Levene pour tester l'égalité des variances de l'erreur σ_p^2 pour les différents niveaux du facteur A (voir Neter *et al.*, 1996).

De façon similaire, l'erreur de variation, $SCB.S(A)$, peut être décomposée en terme des composantes des niveaux du facteur A :

$$SCB.S(A) = SCB.S(A_1) + SCB.S(A_2) + \dots + SCB.S(A_a) \quad (2.5.3)$$

où

$$SCB.S(A_j) = \sum_i \sum_k (Y_{ijk} - \bar{Y}_{ij} - \bar{Y}_{.jk} + \bar{Y}_{.j})^2, \quad j = 1, \dots, a. \quad (2.5.4)$$

Chaque composante possède $(n_j - 1)(b - 1)$ degrés de liberté. Le test de Hartley et le test modifié de Levene peuvent être utilisés pour tester l'égalité des variances de l'erreur σ_ε^2 pour les différents niveaux du facteur A .

Le test de Hartley présuppose la normalité et est sensible à ce présupposé. Ainsi il est recommandé de vérifier l'hypothèse de normalité avant d'appliquer ce test. De façon générale, considérons σ_j^2 et s_j^2 respectivement la variance inter-sujets et la variance échantillonnale inter-sujets pour le j^e niveau du facteur A , $j = 1, \dots, a$. Pour vérifier l'hypothèse

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_a^2 \quad (2.5.5)$$

$$\text{contre } H_a : \text{ les } \sigma_j^2 \text{ ne sont pas toutes égales} \quad (2.5.6)$$

à l'aide du test de Hartley, on utilise la statistique du test, notée H , qui est basée uniquement sur la variance échantillonnale maximale $\max(s_j^2)$ et la variance échantillonnale minimale $\min(s_j^2)$:

$$H = \frac{\max(s_j^2)}{\min(s_j^2)}. \quad (2.5.7)$$

La règle de décision pour contrôler l'erreur de type 1 au niveau α est : rejeter H_0 si

$$H > H[1 - \alpha; a, n - 1]$$

où $n = n_j$ pour tout $j = 1, \dots, a$ et ne pas rejeter H_0 sinon (voir Neter *et al.*, 1996). La quantité $H[1 - \alpha; a, n - 1]$ représente le percentile d'ordre $100(1 - \alpha)$ de la distribution H avec a populations et $n - 1$ degrés de liberté pour chaque variance échantillonnale (voir Neter *et al.*, 1996 pour plus de détails au sujet de la distribution H).

Le test modifié de Levene est plus robuste et relativement insensible à la non-normalité et ne requiert pas l'égalité des tailles des échantillons. Pour conduire le test (2.5.5) en utilisant le test modifié de Levene, il faut tout d'abord déterminer $d_{ij.}$, la différence absolue entre la moyenne estimée $\bar{Y}_{ij.}$ et la médiane des moyennes

estimées $\bar{Y}_{1j}, \dots, \bar{Y}_{nj}$. notée $\tilde{Y}_{j.}$. Nous pouvons écrire $d_{ij} = |\bar{Y}_{ij} - \tilde{Y}_{j.}|$. Le test modifié de Levene détermine si les valeurs attendues des différences absolues d_{ij} pour les a niveaux du facteur A sont égales et si c'est le cas, alors les a variances σ_j^2 sont aussi égales. La statistique du test de Levene est une statistique F^* comme dans (2.4.6) non pas basée sur les différences entre les moyennes des niveaux du facteur A , mais sur les différences absolues d_{ij} :

$$F_L^* = \frac{MCA_L}{MCS(A)_L}, \quad (2.5.8)$$

où

$$MCA_L = \frac{\sum_j n_j (\bar{d}_{.j} - \bar{d}_{...})^2}{a - 1}, \quad (2.5.9)$$

$$MCS(A)_L = \frac{\sum_i \sum_j n_j (d_{ij} - \bar{d}_{.j})^2}{N - a}, \quad (2.5.10)$$

$$\bar{d}_{.j} = \frac{\sum_i d_{ij}}{n_j}, \quad (2.5.11)$$

$$\bar{d}_{...} = \frac{\sum \sum d_{ij}}{N}, \quad (2.5.12)$$

$N = \sum_j n_j$ et n_j est la taille échantillonnale du j^e niveau du facteur A (voir Neter *etal.*, 1996). La règle de décision pour contrôler l'erreur de type 1 au niveau α est : rejeter H_0 si

$$F_L^* > F[1 - \alpha; a - 1, N - a]$$

et ne pas rejeter H_0 sinon. La quantité $F[1 - \alpha; a - 1, N - a]$ représente le percentile d'ordre $100(1 - \alpha)$ de la distribution F avec $a - 1$ et $N - a$ degrés de liberté respectivement au numérateur et au dénominateur.

2.5.1. La symétrie composée.

Certains auteurs se sont interrogés sur la nécessité des présupposés concernant l'homogénéité des variances. En effet, certaines hypothèses doivent être satisfaites pour que le rapport $F^* = \frac{MCB}{MCB.S(A)}$ suive une loi F avec $(b-1)$ et $(N-a)(b-1)$ degrés de liberté. D'une part, on suppose que les variances à l'intérieur des niveaux du facteur temps sont homogènes pour chaque groupe, les observations entre les individus sont indépendantes et suivent une loi de distribution multinormale. D'autre part, on suppose l'égalité des variances des niveaux du facteur temps et l'égalité des covariances de chaque paire des niveaux du facteur temps. Ainsi la matrice Σ avec variances égales et covariances égales est dite avoir une structure symétrie composée ou matrice de type S , c'est-à-dire

$$\Sigma = \sigma_\epsilon^2 I_b + \sigma_\rho^2 J_b = \begin{pmatrix} \sigma_\epsilon^2 + \sigma_\rho^2 & \sigma_\rho^2 & \dots & \sigma_\rho^2 \\ \sigma_\rho^2 & \sigma_\epsilon^2 + \sigma_\rho^2 & \dots & \sigma_\rho^2 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \sigma_\rho^2 & \sigma_\rho^2 & \dots & \sigma_\epsilon^2 + \sigma_\rho^2 \end{pmatrix}$$

où I_b est la matrice identité ($b \times b$) et J_b est la matrice unité ($b \times b$) dont chaque élément est égal à 1.

La sphéricité est le cas particulier de la symétrie composée avec $\sigma_\rho^2 = 0$. Soit $Y : (b \times 1)$ un vecteur aléatoire de loi multinormale $N_b(\mu, \Sigma)$. Le test de sphéricité de Mauchly confronte

$$H_0 : \Sigma = \sigma_\epsilon^2 I_b \quad (2.5.13)$$

$$\text{contre } H_a : \Sigma \neq \sigma_\epsilon^2 I_b. \quad (2.5.14)$$

Étant donné les $Y^{(i)} : (b \times 1)$, $i = 1, \dots, N$, N observations indépendantes $N_b(\mu, \Sigma)$, nous pouvons déterminer la matrice $S : (b \times b)$ telle que

$$S = \sum_{i=1}^N (Y^{(i)} - \bar{Y})(Y^{(i)} - \bar{Y})'$$

où $\bar{Y} = \sum_{i=1}^N Y^{(i)}/N$. Mauchly(1940) a obtenu la statistique du rapport de vraisemblance pour tester H_0 sous la forme

$$\lambda = \frac{|S|^{\frac{1}{2}N}}{|\frac{1}{b}\sum_{i=1}^b S_{ii}|^{\frac{1}{2}bN}} \quad (2.5.15)$$

où S_{ii} est l'élément à la i^e ligne et la i^e colonne de la matrice S . Le terme $|S|$ est le déterminant de la matrice S et le terme $|\frac{1}{b}\sum_{i=1}^b S_{ii}|$ est la valeur absolue de $\frac{1}{b}\sum_{i=1}^b S_{ii}$. La règle de décision pour contrôler l'erreur de type 1 au niveau α est : rejeter H_0 si $\lambda \leq \lambda_0$ où λ_0 est déterminé sous l'hypothèse nulle et le niveau de signification α (voir Mauchly(1940) pour plus de détails concernant le test de sphéricité).

Si $Y^{(i)} : (b \times 1)$ est un vecteur aléatoire de loi elliptique avec paramètre d'aplatissement k , la distribution asymptotique de $-2 \log W / (1 + k)$ lorsque H_0 est vraie est $\chi_{(b-1)(b+2)/2}^2$ où $W = \lambda^{2/N}$ (voir Muirhead, 1982). Ce résultat est aussi vrai pour le cas particulier de la loi multinormale avec $k = 0$. Notant que le vecteur aléatoire $Y^{(i)} : (b \times 1)$ est dit de loi elliptique avec paramètre $\mu : (b \times 1)$ et $V : (b \times b)$ si sa fonction de densité est de la forme

$$c_b(|V|^{-1/2} h((y - \mu)' V^{-1} (y - \mu))),$$

pour une fonction h où V est définie positive (voir Muirhead, (1982) pour des plus amples détails au sujet de la loi elliptique).

2.5.2. Ajustement des degrés de liberté

Box(1954) et Geisser et Greenhouse(1958) ont examiné les effets du non-respect de la condition d'application de sphéricité pour la matrice de covariances. Ils ont prouvé que malgré la forme de la matrice de covariances, le rapport F issu

de la partie intra-sujet de l'analyse de variance sera plus ou moins distribué comme une loi F avec $(b-1)\varepsilon$ et $(N-a)(b-1)\varepsilon$ degrés de liberté pour les effets du facteur temps et $(a-1)(b-1)\varepsilon$ et $(N-a)(b-1)\varepsilon$ degrés de liberté pour les effets de l'interaction $A * B$. Le facteur de correction ε est estimé par :

$$\hat{\varepsilon} = \frac{b^2(\bar{D} - \overline{Cov}_T)^2}{(b-1)(\sum Cov_{ij}^2 - 2b \sum \overline{Cov}_i^2 + b^2 \overline{Cov}_T^2)} \quad (2.5.16)$$

où \bar{D} est la moyenne des variances (valeurs de la diagonale de la matrice de covariances), \overline{Cov}_T est la moyenne de tous les éléments de la matrice de covariances, Cov_{ij}^2 est le carré de l'élément de la i^e ligne et de la j^e colonne de la matrice de covariances, et \overline{Cov}_i est la moyenne des éléments de la ligne i de la matrice de covariances.

Plus $\hat{\varepsilon}$ est proche de 1, plus fort est le niveau de sphéricité. Par conséquent, lorsque $\hat{\varepsilon}$ n'est pas proche de 1, le test d'une hypothèse basée sur la statistique F^* est plus valide lorsque les degrés de liberté du numérateur et les degrés de liberté du dénominateur sont multipliés par le facteur de correction $\hat{\varepsilon}$. D'après Huynh et Feldt (1976), si le facteur de correction $\varepsilon > 0,75$ (la valeur de ε varie de $1/(b-1)$ à 1), alors l'ajustement des degrés de liberté par $\hat{\varepsilon}$ donne un test trop conservateur, c'est-à-dire que plusieurs mauvaises hypothèses H_0 ne seront pas rejetées. Ainsi Huynh et Feldt (1976) recommandent une estimation moins conservatrice

$$\tilde{\varepsilon} = \frac{N(b-1)\hat{\varepsilon} - 2}{(b-1)[N-a-(b-1)\hat{\varepsilon}]} \quad (2.5.17)$$

Dans certaines circonstances, $\tilde{\varepsilon}$ dépassera 1, auquel cas il est fixé à 1. En pratique, on utilise la correction par $\hat{\varepsilon}$ de Geisser et Greenhouse si $\varepsilon > 0,75$ et la correction par $\tilde{\varepsilon}$ de Huynh et Feldt si $\varepsilon < 0,75$.

2.5.3. L'analyse multivariée

L'analyse usuelle de variance à mesures répétées nécessite une condition d'application de sphéricité de la matrice de covariances pour le facteur intra-sujet.

Les facteurs de correction proposés par Greenhouse et Geisser (1958, 1959) et par Huynh et Feldt (1976) sont appliqués aux degrés de liberté pour remédier certains problèmes associés au non-respect de la condition d'application de sphéricité. Ces ajustements fonctionnent bien pour autant qu'on se limite à des effets globaux, simple ou de l'interaction. Boik (1981) a montré que l'analyse à mesures répétées est remarquablement sensible aux violations de la condition de sphéricité dans le cas où l'on effectue des analyses des contrastes, à moins qu'on adopte des termes d'erreur séparés pour chaque contraste.

Plusieurs auteurs suggèrent d'appliquer une analyse de variance multivariée (MANOVA) dans le cas de non-respect de la condition d'application de sphéricité. Fondamentalement une analyse de variance multivariée est une analyse portant sur plusieurs variables dépendantes à la fois. Pour exprimer le modèle multivarié d'analyse de la variance à mesures répétées, nous devons exclure du modèle univarié (2.2.1) l'effet du i^e individu appartenant au j^e groupe. Cela revient à ignorer la condition d'application de sphéricité. Ainsi le nouveau modèle devient :

$$Y_{ijk} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \varepsilon_{ijk} \quad (2.5.18)$$

où $i = 1, \dots, n_j$, $j = 1, \dots, a$ et $k = 1, \dots, b$. Les paramètres du modèle (2.5.18) sont définis de la même façon que dans le modèle (2.2.1). On suppose que les vecteurs aléatoires $\varepsilon_{ij} = (\varepsilon_{ij1}, \dots, \varepsilon_{ijb})'$ sont $N_b(0_b, \Sigma)$, où Σ est une matrice ($b \times b$) définie positive. Soient c_{ij} , $i = 1, \dots, N$; $j = 1, \dots, a$, les variables indicatrices de groupe telles que $c_{ij} = 1$ si l'individu i appartient au groupe j et $c_{ij} = 0$ sinon. En posant $c'_i = (c_{i1}, \dots, c_{ia})$, on peut récrire le modèle (2.5.18) en terme du modèle multivarié suivant :

$$Y'_i = c'_i \tau + \varepsilon'_i \quad (2.5.19)$$

où $Y'_i = (Y_{i1}, \dots, Y_{ib})$, $\varepsilon'_i = (\varepsilon_{i1}, \dots, \varepsilon_{ib})$ et

$$\tau = \begin{pmatrix} \tau_{11} & \tau_{12} & \dots & \tau_{1b} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \tau_{a1} & \tau_{a2} & \dots & \tau_{ab} \end{pmatrix}$$

est une matrice $(a \times b)$ des paramètres de position définis par

$\tau_{jk} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk}$ pour $j = 1, \dots, a$ et $k = 1, \dots, b$. Les τ_{jk} représentent les moyennes des cellules associées au j^e groupe et au k^e temps. En posant $Y = (Y_1, \dots, Y_N)'$, $C = (c_1, \dots, c_N)'$ et $\varepsilon = (\varepsilon_1, \dots, \varepsilon_N)'$, le modèle (2.5.19) devient :

$$Y = C\tau + \varepsilon \quad (2.5.20)$$

Anderson (1984) décrit quatre tests pour les effets des facteurs dans le cadre d'une analyse de variance multivariée. Ce sont le rapport de vraisemblance maximale de Wilks, la trace de Bartlett-Nanda-Pillai, la trace de Hotelling-Lawley et la valeur propre maximale (vpm) de Roy. Ces quatre tests permettent de contrôler l'hypothèse de la forme

$$H_0 : T\tau M = 0 \quad (2.5.21)$$

où T est une matrice $(s \times a, s \leq a)$ formée de s vecteurs-lignes linéairement indépendant et M est une matrice $(b \times p, b \leq p)$ de plein rang b .

Soient $S_H = M'Y'A(A'A)^{-1}T'[T(A'A)^{-1}T']^{-1}T(A'A)^{-1}A'YM$ et

$S_E = M'Y'[I_N - A(A'A)^{-1}A']YM$ respectivement la matrice associée à l'hypothèse nulle considérée et la matrice résiduelle où A est une matrice de la forme

$$A_{N \times a} = \begin{pmatrix} \left\{ \begin{matrix} 1 & 0 & \dots & 0 \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \end{matrix} \right\}^{n_1} \\ \left\{ \begin{matrix} 1 & 0 & \dots & 0 \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \end{matrix} \right\} \dots \\ \left\{ \begin{matrix} \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \end{matrix} \right\} \dots \\ \left\{ \begin{matrix} 0 & 0 & \dots & 1 \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \end{matrix} \right\}^{n_a} \\ \left\{ \begin{matrix} 0 & 0 & \dots & 1 \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \end{matrix} \right\} \dots \end{pmatrix}$$

Pour le test du rapport de vraisemblance maximale de Wilks, l'hypothèse nulle est rejetée pour les grandes valeurs de la statistique

$$\Lambda = \frac{1}{|S_H S_E^{-1} + I_p|}.$$

Pour le test de la trace de Bartlett-Nanda-Pillai, l'hypothèse nulle est rejetée pour les grandes valeurs de la statistique

$$\text{tr} S_H (S_H + S_E)^{-1}.$$

Pour le test de la trace de Hotelling-Lawley, l'hypothèse nulle est rejetée pour les grandes valeurs de la statistique

$$\text{tr} (S_H S_E)^{-1}.$$

Pour le test de la valeur propre maximale de Roy, l'hypothèse nulle est rejetée pour les grandes valeurs de la statistique

$$\frac{\text{vpm}(S_H S_E^{-1})}{1 + \text{vpm}(S_H S_E^{-1})}.$$

Chacune de ces quatre statistiques permettent de calculer une statistique F qui suit approximativement une loi F (voir Anderson(1984)).

2.6. L'APPROCHE MIXTE

Une analyse de variance est dite à effets fixes si tous les effets des facteurs inclus dans le plan d'expérience sont fixes. Lorsque tous les effets des facteurs à l'étude sont aléatoires, on dit que le modèle d'analyse de variance est un modèle à effets aléatoires. Lorsque certains facteurs sont fixes et d'autres aléatoires, on dit que le modèle d'analyse de variance est un modèle mixte. Le modèle linéaire mixte présenté par Laird et Ware (1982) est une alternative très flexible pour l'analyse des plans à mesures répétées. Dans le modèle (2.2.1), les paramètres μ , α_j , β_k , $(\alpha\beta)_{jk}$ représentent des effets fixes tandis que les composantes $\rho_{i(j)}$ représentent les effets aléatoires. Ce modèle peut être convenablement représenté en utilisant les moyennes des cellules :

$$Y_{ij} = \tau_j + J_b b_{i(j)} + \varepsilon_{ij} \quad (2.6.1)$$

où $Y_{ij} = (Y_{ij1}, \dots, Y_{ijb})'$, τ_j et ε_{ij} sont tels que dans (2.5.19), J_b est la matrice unité ($b \times b$), $\tau_{jk} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk}$ et $b_{i(j)}$ est l'effet aléatoire $\rho_{i(j)}$. Soient c_{ij} et c'_i définis tels que dans (2.5.19), on peut récrire le modèle (2.6.1) en termes du modèle linéaire mixte comme

$$Y_i = X_i \tau + Z_i b_i + \varepsilon_i \quad (2.6.2)$$

où $Y'_i = (Y_{i1}, \dots, Y_{ib})$, $X_i = I_b \otimes c'_i$, $\tau' = (\tau_{11}, \dots, \tau_{ab})$ et $Z_i = J_b$. Le symbole \otimes représente le produit de Kronecker.

En posant $Y = (Y_1, \dots, Y_N)'$, $X = (X_1, \dots, X_N)'$, $Z = (Z_1, \dots, Z_N)'$, et $\varepsilon = (\varepsilon_1, \dots, \varepsilon_N)'$, le modèle (2.6.2) devient

$$Y = X\tau + Zb + \varepsilon \quad (2.6.3)$$

où b suit une loi multinormale de vecteur moyen 0 et de matrice de covariances G et ε suit une loi multinormale de vecteur moyen 0 et de matrice de covariances R . Alors, Y suit une loi multinormale de vecteur moyen $X\tau$ et de matrice de covariances $V = ZGZ' + R$.

L'approche mixte pour l'analyse des plans à mesures répétées permet de modéliser la structure de la matrice de covariances pour le facteur inter-sujets et le facteur intra-sujet. La symétrie composée n'est probablement pas la seule forme que peut prendre la matrice de covariances pour le facteur à mesures répétées car deux mesures observées à des périodes adjacentes sont vraisemblablement plus corrélées que deux mesures observées à des périodes de temps plus distantes. Ware (1985) présente une vue d'ensemble de l'application des modèles linéaires mixtes à des plans à mesures répétées. L'approche des modèles linéaires mixtes pour les mesures répétées est plus générale car en pratique, dans des études à mesures répétées, on observe souvent des données manquantes. De plus le nombre d'observations pour chaque unité expérimentale et le temps de prélèvement des mesures des observations varient souvent. De telles caractéristiques rendent difficile l'application des procédures multivariées. L'approche des modèles linéaires mixtes pour les mesures répétées peut être vue comme une analyse univariée de régression avec des erreurs corrélées. L'avantage majeur de cette méthodologie est qu'elle s'adapte bien à la complexité des ensembles de données à mesures répétées. Grâce à la disponibilité de certains logiciels statistiques, tels que la procédure MIXED de SAS, l'utilisation des modèles linéaires mixtes pour l'analyse de variance à mesures répétées devient de plus en plus courante. Une grande variété de structures de covariances intra-sujet et inter-sujets peut être considérée.

Voici une vue d'ensemble des structures de covariances qui peuvent être spécifiées dans la procédure MIXED de SAS. Pour $\sigma^2, \sigma_1^2, \dots, \sigma_n^2 > 0$ et pour $\rho, \dots, \rho^{n-1}, \rho_1, \dots, \rho_{n-1}$ chacune supérieure ou égale à 0 et inférieure ou égale à 1 en valeur absolue, on a :

$$\text{Non structurée : } \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \dots & \sigma_{2n} \\ \dots & \dots & \dots & \dots \\ \sigma_{1n} & \sigma_{2n} & \dots & \sigma_n^2 \end{pmatrix}$$

$$\text{Symétrie composée : } \begin{pmatrix} \sigma_1^2 + \sigma^2 & \sigma_1^2 & \dots & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 + \sigma^2 & \dots & \sigma_1^2 \\ \dots & \dots & \dots & \dots \\ \sigma_1^2 & \sigma_1^2 & \dots & \sigma_1^2 + \sigma^2 \end{pmatrix}$$

$$\text{Autorégressive d'ordre 1 : } \begin{pmatrix} \sigma^2 & \rho\sigma^2 & \dots & \rho^{n-1}\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \dots & \rho^{n-2}\sigma^2 \\ \dots & \dots & \dots & \dots \\ \rho^{n-1}\sigma^2 & \rho^{n-2}\sigma^2 & \dots & \sigma^2 \end{pmatrix}$$

$$\text{Symétrie composée hétérogène : } \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \dots & \rho\sigma_1\sigma_n \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \dots & \rho\sigma_2\sigma_n \\ \dots & \dots & \dots & \dots \\ \rho\sigma_1\sigma_n & \rho\sigma_2\sigma_n & \dots & \sigma_n^2 \end{pmatrix}$$

$$\text{Simple : } \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix}$$

$$\text{Composantes de variance : } \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_n^2 \end{pmatrix}$$

$$\text{En bande : } \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & 0 \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2n} \\ \dots & \dots & \dots & \dots \\ 0 & \sigma_{2n} & \dots & \sigma_n^2 \end{pmatrix}$$

$$\text{Toeplitz : } \begin{pmatrix} \sigma^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{12} & \sigma^2 & \dots & \sigma_{2n} \\ \dots & \dots & \dots & \dots \\ \sigma_{1n} & \sigma_{2n} & \dots & \sigma^2 \end{pmatrix}$$

$$\text{Autorégressive hétérogène d'ordre 1 : } \begin{pmatrix} \sigma^2 & \rho\sigma_1\sigma_2 & \dots & \rho^{n-1}\sigma_1\sigma_n \\ \rho\sigma_1\sigma_2 & \sigma^2 & \dots & \rho^{n-2}\sigma_2\sigma_n \\ \dots & \dots & \dots & \dots \\ \rho^{n-1}\sigma_1\sigma_n & \rho^{n-2}\sigma_2\sigma_n & \dots & \sigma^2 \end{pmatrix}$$

$$\text{Toeplitz hétérogène : } \begin{pmatrix} \sigma_1^2 & \rho_1\sigma_1\sigma_2 & \dots & \rho_{n-1}\sigma_1\sigma_n \\ \rho_1\sigma_1\sigma_2 & \sigma_2^2 & \dots & \rho_{n-2}\sigma_2\sigma_n \\ \dots & \dots & \dots & \dots \\ \rho_{n-1}\sigma_1\sigma_n & \rho_{n-2}\sigma_2\sigma_n & \dots & \sigma_n^2 \end{pmatrix}$$

Pour une structure de covariances donnée, les paramètres des matrices G et R sont estimés en utilisant la méthode du maximum de vraisemblance (ML) ou la méthode du maximum de vraisemblance restreint (REML). La méthode du maximum de vraisemblance consiste à minimiser -2 fois le log de la fonction de vraisemblance $L(G, R)$ par rapport à G et R :

$$\ln[L(G, R)] = -\frac{1}{2} \ln |V| - \frac{1}{2} r' V^{-1} r - \frac{N}{2} \ln(2\pi) \quad (2.6.4)$$

et la méthode du maximum de vraisemblance restreint consiste à minimiser -2 fois le log de la fonction de vraisemblance $L_R(G, R)$ par rapport à G et R :

$$\ln[L_R(G, R)] = -\frac{1}{2} \ln |V| - \frac{1}{2} \ln |X'V^{-1}X| - \frac{1}{2} r' V^{-1} r - \frac{N-p}{2} \ln(2\pi) \quad (2.6.5)$$

où $r = Y - X(X'V^{-1}X)^{-1}X'V^{-1}Y$, N est le nombre d'observations et p est le rang de la matrice X . L'algorithme d'optimisation de Newton-Raphson est privilégié, selon Lindstrom et Bates (1988) pour minimiser ces fonctions. Wolfinger *et al.*, (1994) présentent les détails concernant l'algorithme de Newton-Raphson dans l'optimisation du modèle général mixte.

Après avoir estimé G et R , les vecteurs τ et b sont estimés en résolvant les équations du modèle mixte (voir Henderson(1984)) :

$$\begin{pmatrix} X'\hat{R}^{-1}X & X'\hat{R}^{-1}Z \\ Z'\hat{R}^{-1}X & Z'\hat{R}^{-1}Z + \hat{G}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\tau} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} X'\hat{R}^{-1}Y \\ Z'\hat{R}^{-1}Y \end{pmatrix},$$

dont la solution est donnée par

$$\begin{pmatrix} \hat{\tau} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}Y \\ \hat{G}Z'\hat{V}^{-1}(Y - X\hat{\tau}) \end{pmatrix} = \hat{C} \begin{pmatrix} X'\hat{R}^{-1}Y \\ Z'\hat{R}^{-1}Y \end{pmatrix}$$

où

$$\hat{C} = \begin{pmatrix} X'\hat{R}^{-1}X & X'\hat{R}^{-1}Z \\ Z'\hat{R}^{-1}X & Z'\hat{R}^{-1}Z + \hat{G}^{-1} \end{pmatrix}^{-}$$

avec le signe exposant $-$ désignant l'inverse généralisé. Henderson(1984) et McLean et Sanders(1988) démontrent que \hat{C} peut également s'écrire comme suit :

$$\hat{C} = \begin{pmatrix} \hat{C}_{11} & \hat{C}'_{21} \\ \hat{C}_{21} & \hat{C}_{22} \end{pmatrix}$$

où $\hat{C}_{11} = (X'\hat{V}^{-1}X)^{-}$, $\hat{C}_{21} = -\hat{G}Z'\hat{V}^{-1}X\hat{C}_{11}$ et $\hat{C}_{22} = (Z'\hat{V}^{-1}Z + \hat{G}^{-1})^{-1} - \hat{C}_{21}X'\hat{V}^{-1}Z\hat{G}$.

Les tests des effets des facteurs pour le modèle linéaire mixte sont décrits dans Littell *et al.*,(1996) et résumés dans Cossette(2002). Pour tester l'effet du facteur groupe, l'effet du facteur temps et l'effet de l'interaction entre le facteur groupe et le facteur temps, l'hypothèse linéaire de la forme $L'\tau = 0$ est considérée (voir Searle, 1971). Le terme L représente une matrice des coefficients des combinaisons linéaires des paramètres pour l'effet du facteur considéré.

Pour tester l'effet du facteur groupe, l'hypothèse $L'_{gr}\tau = 0$ est considérée et la statistique F est donnée par

$$F = \frac{\hat{\tau}'L_{gr}(L'_{gr}\hat{C}_{11}L_{gr})^{-1}L'_{gr}\hat{\tau}}{\text{rang}(L_{gr})}. \quad (2.6.6)$$

Cette statistique F suit approximativement une loi F de Fisher à f_1 et f_2 degrés de liberté où $f_1 = \text{rang}(L_{gr})$ et f_2 est estimé en utilisant la procédure de Satterthwaite discutée par Jeske et Harville(1988).

De façon similaire, l'hypothèse $L'_{temps}\tau = 0$ est considérée pour tester l'effet du facteur temps et la statistique F est donnée par

$$F = \frac{\hat{\tau}' L_{temps} (L'_{temps} \hat{C}_{11} L_{temps})^{-1} L'_{temps} \hat{\tau}}{\text{rang}(L_{temps})}. \quad (2.6.7)$$

Cette statistique F suit approximativement une loi F de Fisher à f_1 et f_2 degrés de liberté où $f_1 = \text{rang}(L_{temps})$ et f_2 est estimé en utilisant la procédure de Satterthwaite.

Enfin, l'hypothèse $L'_{gr*temps}\tau = 0$ est considérée pour tester l'effet de l'interaction entre le facteur groupe et le facteur temps et la statistique F est donnée par

$$F = \frac{\hat{\tau}' L_{gr*temps} (L'_{gr*temps} \hat{C}_{11} L_{gr*temps})^{-1} L'_{gr*temps} \hat{\tau}}{\text{rang}(L_{gr*temps})}. \quad (2.6.8)$$

Cette statistique F suit approximativement une loi F de Fisher à f_1 et f_2 degrés de liberté où $f_1 = \text{rang}(L_{gr*temps})$ et f_2 est aussi estimé en utilisant la procédure de Satterthwaite.

2.6.1. Stratégies d'analyse

Les données à mesures répétées sont particulièrement intéressantes parce que les différentes mesures observées pour un même individu sont souvent corrélées et présentent généralement une variabilité hétérogène. Une analyse par la méthode des moindres carrés ordinaires (disponible par exemple dans la procédure GLM de SAS) est généralement appropriée dans le cas où ces mesures ne présentent pas une variabilité hétérogène tandis que dans le cas d'une variabilité hétérogène, d'autres procédures peuvent être plus appropriées, comme par exemple la procédure MIXED de SAS. Nous faisons ici une comparaison entre les procédures GLM et MIXED de SAS.

La première chose à remarquer au sujet de l'analyse de la variance à mesures répétées avec Proc GLM est que cette procédure exclut de l'analyse tout individu ayant des observations manquantes. Cette procédure exige que les données intra-sujet soient prélevées durant les mêmes périodes pour tous les sujets et ce, un même nombre de fois. Proc GLM utilise la méthode de moments pour estimer les paramètres du modèle et exécute un test pour la signification des effets inter-sujets et deux différents tests pour les effets intra-sujet dont un test univarié et un test multivarié. Proc GLM exécute aussi un test de sphéricité. Lorsque la valeur-p du test de sphéricité n'est pas significative, on considère les tests univariés pour les effets intra-sujet puisque sous l'hypothèse de sphéricité, les tests univariés sont plus robustes que les tests multivariés. Lorsque la valeur-p du test de sphéricité est significative, Proc GLM offre deux façons de tester la signification des effets intra-sujet. La première démarche consiste à considérer les tests univariés mais en faisant l'ajustement des degrés de liberté par la formule de Greenhouse et Geisser (1959) et l'ajustement moins conservateur de Huynh et Feldt(1976). La seconde démarche implique les quatre tests multivariés de la section 2.5.3 (voir Hand et Taylor, 1987).

La procédure MIXED utilise toutes les données disponibles et par conséquent n'exclut pas de l'analyse les sujets avec données manquantes. La procédure MIXED se sert d'une méthode d'estimation basée sur la vraisemblance. La différence fondamentale entre la procédure GLM et la procédure MIXED est que la procédure MIXED offre une grande flexibilité en ce sens que l'on peut spécifier différentes structures pour la matrice de covariances. Les tests statistiques pour les effets fixes des facteurs dépendent de la structure de la matrice de covariances spécifiée. Proc MIXED peut produire des résultats erronés si la structure de covariances est mal spécifiée, c'est pourquoi il est nécessaire de spécifier différentes structures de covariances et de choisir le modèle avec la structure de covariances pour laquelle les critères d'information AIC (Akaike Information Criterion) et BIC (Bayesian Information Criterion) ont les valeurs les plus petites.

Pour chaque i^e modèle alternatif proposé pour décrire les données, le AIC_i est donné par :

$$AIC_i = 2\log L(\hat{\theta}_i, Y) - 2(k + r_i) \quad (2.6.9)$$

où $k + r_i$ est le nombre de paramètres du modèle et $L(\hat{\theta}_i, Y)$ est la fonction de vraisemblance (voir Akaike, 1974).

De façon analogue, le BIC_i est donné par :

$$BIC_i = 2\log L(\hat{\theta}_i, Y) - (k + r_i)\log(N) \quad (2.6.10)$$

où N représente la taille d'échantillon (voir Schwarz, 1978).

La stratégie pour sélectionner la structure de covariance appropriée pour un modèle donné est présentée dans Wolfinger(1993).

Chapitre 3

ANALYSE DES RÉSULTATS

3.1. INTRODUCTION

Dans le but d'examiner l'influence des données manquantes dans un plan d'analyse de variance à deux facteurs dont un comporte des mesures répétées, nous avons considéré quatre scénarios distincts en utilisant quatre ensembles de données de loi multinormale ayant la même structure des moyennes et des structures différentes pour la matrice de covariances.

Dans ce chapitre, nous expliquons la méthodologie utilisée pour générer quatre ensembles de données et la manière dont ces données ont été transformées en créant des valeurs manquantes d'une part et d'autre part en remplaçant ces valeurs manquantes par des valeurs imputées.

Ensuite, pour chaque scénario élaboré, nous avons fait des analyses de variance à deux facteurs dont un comporte des mesures répétées dans le cas des données complètes, le cas des données avec valeurs manquantes, le cas des données avec valeurs manquantes imputées par la moyenne des valeurs observées pour le même sujet et le cas des données avec valeurs manquantes imputées par la moyenne locale des valeurs observées pour le même sujet. Pour ce faire, nous avons utilisé les procédures GLM et MIXED de SAS.

Enfin, l'interprétation des résultats obtenus de ces analyses nous a permis de tirer certaines conclusions.

3.2. MÉTHODOLOGIE DE GÉNÉRATION DES ENSEMBLES DE DONNÉES

Nous avons généré quatre ensembles de données de loi multinormale pour lesquels les matrices de covariances ont été fixées en vue de contrôler leurs structures. Chaque ensemble de données est constitué de 54 individus regroupés dans 3 groupes de même taille ($n_1 = n_2 = n_3 = 18$) sur lesquels les mesures sont disponibles à 5 périodes. Les données ont été générées séparément pour chacun des 3 groupes d'individus.

Tout d'abord, en se servant du générateur des nombres pseudo-aléatoires de S-Plus, nous avons généré indépendamment 5 variables aléatoires normales X_1, X_2, X_3, X_4 et X_5 de taille 18, de moyenne 0 et de variance 1 chacune. Ensuite dans le but de fixer chacune des quatre matrices de covariances, nous avons transformé l'ensemble de données obtenu de ces 5 variables aléatoires en utilisant la décomposition racine de Choleski pour que la matrice de covariances qui en résulte soit numériquement une matrice identité. Enfin, l'ensemble de données obtenu a été transformé pour créer quatre ensembles de données ayant chacun une matrice de covariance différente.

Mathématiquement, les opérations effectuées peuvent se résumer comme suit : soit $Y_1 : 5 \times 1$ un vecteur aléatoire défini tel que $Y_1 = (X_1, X_2, X_3, X_4, X_5)'$. $\hat{\Sigma}_{Y_1}$ est la matrice de covariances estimée à partir Y_1 . On calcule $Y = A^{-1}Y_1$ où A est une matrice triangulaire supérieure telle que $A'\hat{\Sigma}_{Y_1}^{-1}A = I$ et I est la matrice identité de dimension 5. Enfin on calcule chacun des quatre ensembles de données par $Z = CY + M$ où C est une matrice triangulaire supérieure telle que $C'\Sigma^{-1}C = I$ et Σ représente la matrice de covariances à affecter aux données pour le groupe correspondant. Pour chaque groupe correspondant, M est un vecteur fixe (5×1) composé de la moyenne de chacune des 5 périodes.

Les quatre matrices de covariances considérées sont de structures symétrie composée pour la première, non-structurée pour la deuxième, autorégressive d'ordre 1 pour la troisième et symétrie composée hétérogène pour la quatrième.

Les matrices de covariances utilisées sont les suivantes :

$$\text{Symétrie composée(CS)} : \begin{pmatrix} 49 & 10 & 10 & 10 & 10 \\ 10 & 49 & 10 & 10 & 10 \\ 10 & 10 & 49 & 10 & 10 \\ 10 & 10 & 10 & 49 & 10 \\ 10 & 10 & 10 & 10 & 49 \end{pmatrix}$$

$$\text{Non - structurée(UN)} : \begin{pmatrix} 50 & 9 & 20 & 18 & 22 \\ 9 & 40 & 12 & 18 & 32 \\ 20 & 12 & 35 & 24 & 24 \\ 18 & 18 & 24 & 35 & 20 \\ 22 & 32 & 24 & 20 & 45 \end{pmatrix}$$

$$\text{Autorégressive d'ordre 1(AR1)} : \begin{pmatrix} 49 & 44,1 & 39,69 & 35,72 & 32,15 \\ 44,1 & 49 & 44,1 & 39,69 & 35,72 \\ 39,69 & 44,1 & 49 & 44,1 & 39,69 \\ 35,72 & 39,69 & 44,1 & 49 & 44,1 \\ 32,15 & 35,72 & 39,69 & 44,1 & 49 \end{pmatrix}$$

$$\text{Symétrie composée hétérogène(CSH)} : \begin{pmatrix} 49 & 50,4 & 56,7 & 37,8 & 31,5 \\ 50,4 & 64 & 64,8 & 43,2 & 36 \\ 56,7 & 64,8 & 81 & 48,6 & 40,5 \\ 37,8 & 43,2 & 48,6 & 36 & 27 \\ 31,5 & 36 & 40,5 & 27 & 25 \end{pmatrix} .$$

Pour chaque scénario, la matrice de covariances correspondante se rapporte à un seul groupe d'individus à la fois et est la même pour chacun des trois groupe d'individus.

Pour construire chacune des matrices de covariances, les paramètres utilisés sont les suivants : pour la matrice de covariances de structure symétrie composée, nous avons spécifié une égalité des variances ($\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2 = 49$) et une égalité des covariances ($\sigma_{kk'} = 10$, $k \neq k'$, $k = 1, \dots, 5$ et $k' = 1, \dots, 5$). Pour la matrice de covariances de structure autorégressive d'ordre 1, l'autocorrélation d'ordre 1 équivaut à $\rho = 0,9$ et la variance $\sigma^2 = 49$. Pour la matrice de covariances de structure symétrie composée hétérogène, les paramètres spécifiés sont $\sigma_1 = 7$, $\sigma_2 = 8$, $\sigma_3 = 9$, $\sigma_4 = 6$, $\sigma_5 = 5$ et $\rho = 0,9$.

Nous avons utilisé le même vecteur des moyennes des groupes dans le temps pour les quatre scénarios. Le tableau suivant présente les moyennes des groupes dans le temps utilisées pour générer les données.

TAB. 3.2.1. Tableau des moyennes des groupes dans le temps.

	temps1	temps2	temps3	temps4	temps5
groupe1	175	180	185	190	185
groupe2	178	184	189	196	191
groupe3	182	186	191	194	188

Pour chaque ensemble de données obtenu, nous avons créé des données manquantes complètement aléatoirement (MCAR) en procédant de deux manières différentes. Dans un premier temps, les valeurs manquantes ont été générées aléatoirement pour chaque période de façon uniforme sans distinction des groupes avec les probabilités 0,05, 0,10 et 0,15 et dans un deuxième temps, les valeurs manquantes ont été générées aléatoirement pour chaque période avec les probabilités 0,05, 0,10 et 0,15 à l'intérieur de chacun des groupes 1 et 2 séparément tout en laissant le groupe 3 inchangé, sans aucune valeur manquante.

Ensuite nous avons utilisé deux méthodes d'imputation simple pour remplacer les valeurs manquantes par les valeurs estimées. Tenant compte de la dépendance des observations pour le même sujet, les méthodes d'imputation utilisées sont

l'imputation par la moyenne des valeurs observées pour le même sujet et l'imputation par la moyenne locale des valeurs observées pour le même sujet c'est-à-dire la moyenne obtenue en utilisant la valeur qui précède et celle qui suit la donnée manquante dans le cas où ces deux valeurs sont observées, ou en remplaçant la valeur manquante par la valeur voisine la plus proche dans le cas contraire.

Pour des raisons pratiques, nous avons choisi de ne pas appliquer la méthode d'imputation multiple. En effet les procédures MI et MIANALYSE de SAS qui permettent de faire de l'imputation multiple sont expérimentales dans la plus récente version de SAS (SAS 8.2).

Enfin nous avons fait des analyses de variance à mesures répétées avec chaque ensemble de données obtenu à l'aide des procédures GLM et MIXED de SAS.

Afin de rendre claire l'interprétation des résultats des analyses effectuées, nous avons jugé nécessaire de définir quelques termes.

Le scénario 1 représente les ensembles de données pour lesquels la matrice de covariances est de structure symétrie composée, le scénario 2 représente les ensembles de données pour lesquels la matrice de covariances est non structurée, le scénario 3 représente les ensembles de données pour lesquels la matrice de covariances est de structure autorégressive d'ordre 1 et le scénario 4 représente les ensembles de données pour lesquels la matrice de covariances est de structure symétrie composée hétérogène.

Le cas 1 représente les ensembles de données complètes d'origine, le cas 2 représente les ensembles de données avec valeurs manquantes c'est-à-dire les ensembles des données obtenus après avoir généré les valeurs manquantes dans les ensembles de données complètes d'origine, le cas 3 représente les ensembles des données obtenus suite à l'imputation des valeurs manquantes dans les ensembles de données avec valeurs manquantes par la moyenne des valeurs observées pour le même sujet et le cas 4 représente les ensembles des données obtenus suite à l'imputation des valeurs manquantes dans les ensembles de données avec valeurs

manquantes par la moyenne locale des valeurs observées pour le même sujet. p_{vm} symbolise la probabilité des données manquantes.

Le contexte 1 des valeurs manquantes est le contexte dans lequel les valeurs manquantes ont été générées aléatoirement sans distinction des groupes et le contexte 2 des valeurs manquantes est le contexte dans lequel les valeurs manquantes ont été générées aléatoirement à l'intérieur des groupes 1 et 2 séparément tout en laissant le groupe 3 sans modification.

Dans l'annexe A, on retrouve les données obtenues pour chacun des quatre scénarios, des quatre cas et des deux contextes des données manquantes. Signalons que pour le cas 1, les observations au temps 1 sont les mêmes pour les ensembles de données avec matrices de covariances symétrie composée, autorégressive d'ordre 1 et symétrie composée hétérogène à cause du fait que nous avons fixé la variance au temps 1 à 49 et le germe de départ est le même pour ces trois scénarios.

Les estimateurs des moyennes des groupes dans le temps pour le cas 1 et pour les quatre scénarios sont disponibles en annexe B.

3.3. RÉSULTATS ISSUS DE LA PROCÉDURE GLM

Les résultats des tests des effets des facteurs obtenus en utilisant la procédure GLM de SAS sont présentés dans les tableaux C.0.1 à C.0.8 de l'annexe C. Une valeur-p inférieure à 5% indique que l'effet du facteur est significatif.

Pour chaque scénario et dans chaque contexte des valeurs manquantes, nous faisons référence au cas 1 pour déterminer s'il y a changement ou non dans les résultats des tests des effets des facteurs pour les cas 2, 3 et 4 en se basant sur le fait que l'effet du facteur soit significatif ou non et cela simultanément pour l'effet de l'interaction entre le facteur groupe et le facteur temps, l'effet principal du facteur groupe et l'effet principal du facteur temps.

3.3.1. Résultats pour le scénario 1

Pour le scénario 1 relatif aux ensembles de données en rapport avec la matrice de covariances de structure symétrie composée, la valeur-p du test de sphéricité de Mauchly pour le cas 1 est égale à 1 indiquant une symétrie composée parfaite. Pour les cas 2, 3 et 4, les valeurs-p du test de sphéricité de Mauchly demeurent très élevées et supérieures à 0,05 dans les deux contextes des valeurs manquantes indiquant ainsi que la condition d'application de sphéricité pour la matrice de covariances des données à mesures répétées est satisfaite pour chaque cas et dans chaque contexte des valeurs manquantes (voir les tableaux C.0.1 et C.0.2 de l'annexe C pour la lecture des résultats concernant le scénario 1). Le signe (=) indique que les conclusions obtenues pour le cas 1 sont les mêmes que celles obtenues pour les cas correspondants en ce qui concerne simultanément l'effet principal du facteur groupe, l'effet principal du facteur temps et l'effet de l'interaction entre le facteur groupe et le facteur temps. Le signe (\neq) indique qu'il y a au moins une conclusion différente de celle obtenue pour le cas 1.

Pour le cas 1, l'effet de l'interaction groupe*periode est non significatif avec une valeur-p du test de 0,1129. Il y a un parallélisme entre les trois courbes des moyennes dans le temps pour les trois groupes. L'effet principal du groupe et l'effet principal de la periode sont donc indépendants. L'effet principal du groupe est significatif avec une valeur-p inférieure à 0,0001, ce qui veut dire que les moyennes des niveaux du facteur groupe ne sont pas toutes égales et les différences observées entre les moyennes des niveaux du facteur groupe sont sensiblement les mêmes pour chaque niveau du facteur temps. De même, il y a des différences entre les moyennes des niveaux du facteur temps et ces différences sont sensiblement les mêmes pour chaque niveau du facteur groupe.

Dans le contexte 1 des valeurs manquantes, nous observons que pour $p_{vm}=0,05$, les conclusions des tests des effets des facteurs pour le cas 2 sont les mêmes que celles pour le cas 1, tandis que pour les cas 3 et 4, l'effet de l'interaction

groupe*periode devient significatif avec une valeur-p de 0,0313 dans les deux cas. Le parallélisme observé dans le cas 1 entre les trois courbes des moyennes dans le temps pour les trois groupes est détruit et donc les effets principaux des facteurs ne sont plus indépendants. Il serait en principe nécessaire de faire une analyse des effets simples des facteurs, mais nous ne le ferons pas car dans le cadre de ce mémoire nous recherchons des conclusions générales. Pour $pvm=0,10$ et $pvm=0,15$, les conclusions des tests des effets des facteurs pour les cas 2, 3 et 4 demeurent les mêmes que celles obtenues pour le cas 1.

Dans le contexte 2 des valeurs manquantes, nous observons que pour $pvm=0,05$ et $pvm=0,10$, l'effet de l'interaction groupe*periode devient significatif pour le cas 2 avec des valeurs-p du test de 0,0360 et 0,0288 respectivement, tandis que pour les cas 3 et 4, les conclusions des tests des effets des facteurs demeurent les mêmes que celles pour le cas 1. Pour $pvm=0,15$, les conclusions des tests des effets des facteurs pour les cas 2, 3 et 4 demeurent les mêmes que celles obtenues pour le cas 1.

Comparativement au cas 1, les conclusions des tests des effets des facteurs s'accordent 4 fois sur 6 pour le cas 2 et 5 fois sur 6 pour les cas 3 et 4.

Le fait de la présence des valeurs manquantes dans le cas 2 influence les résultats de l'analyse de variance à mesures répétées. En effet, les conclusions des tests des effets des facteurs pour le cas 2 sont différentes 2 fois sur 6 de celles obtenues pour le cas 1 et lorsque ces conclusions s'accordent, nous constatons tout de même que la valeur-p du test de l'interaction est toujours assez différente de la valeur-p de 0,1129 obtenue dans le cas des données complètes (cas 1).

Un pourcentage des valeurs manquantes plus élevé ne semble pas exercer une plus grande influence sur les résultats de l'analyse de variance à mesures répétées car il n'y a pas une tendance unique qui se dégage dans les résultats en fonction de la croissance du pourcentage des valeurs manquantes.

Les deux méthodes d'imputations utilisées ne semblent pas être clairement supérieures à la méthode des cas complets et aucune des deux méthodes d'imputation n'est supérieure à l'autre de façon évidente.

3.3.2. Résultats pour le scénario 2

Pour le scénario 2 relatif aux ensembles de données en rapport avec la matrice de covariances non-structurée, la valeur-p du test de sphéricité de Mauchly est inférieure à 0,05 pour les quatre cas et cela dans les deux contextes des valeurs manquantes (voir les tableaux C.0.3 et C.0.4 de l'annexe C pour la lecture des résultats concernant le scénario 2). La condition d'application de sphéricité pour la matrice de covariances des données à mesures répétées n'est pas satisfaite pour chacun des cas. Ainsi pour les effets intra-sujet, nous considérons les résultats des tests des effets des facteurs ajustés par le facteur de correction $\bar{\epsilon}$ de Huynh et Feldt (1976). Notons que ces résultats sont presque identiques en utilisant le facteur de correction $\hat{\epsilon}$ de Greenhouse et Geisser (1959).

Pour le cas 1, l'effet de l'interaction groupe*periode est significatif avec une valeur-p du test de 0,0059. Il n'y a pas de parallélisme entre les trois courbes des moyennes dans le temps pour les trois groupes et ainsi, l'effet principal du groupe et l'effet principal de la période ne sont pas indépendants. Comme expliqué précédemment, notre intérêt se limite à des conclusions générales et donc nous ne ferons pas une analyse des effets simples des facteurs. L'effet principal du groupe est significatif avec une valeur-p du test de 0,0009 et l'effet principal de la période est significatif avec une valeur-p du test inférieure à 0,0001.

Dans le contexte 1 des valeurs manquantes, nous observons que pour $pvm=0,05$, les conclusions des tests des effets des facteurs pour les cas 2, 3 et 4 demeurent les mêmes que celles obtenues pour le cas 1. Pour $pvm=0,10$, l'effet de l'interaction groupe*periode devient non significatif pour les cas 2, 3 et 4 avec des valeurs-p du test de 0,2052, 0,1100 et 0,0915 respectivement, rendant ainsi les conclusions des tests des effets des facteurs pour les cas 2, 3 et 4 différentes de celles obtenues

pour le cas 1. Pour $pvm=0,15$, les conclusions des tests des effets des facteurs pour les cas 2 et 4 demeurent les mêmes que celles pour le cas 1, tandis que pour le cas 3, ces conclusions deviennent différentes car l'effet de l'interaction groupe*periode devient non significatif avec une valeur-p de 0,0732.

Dans le contexte 2 des valeurs manquantes, nous observons que pour $pvm=0,05$, les conclusions des tests des effets des facteurs pour les cas 2, 3 et 4 demeurent les mêmes que celles obtenues pour le cas 1. Pour $pvm=0,10$, les conclusions des tests des effets des facteurs pour le cas 4 demeurent les mêmes que celles obtenues pour le cas 1, par contre les conclusions des tests des effets des facteurs pour les cas 2 et 3 sont différentes de celles obtenues pour le cas 1 car l'effet principal du groupe devient non significatif pour le cas 2 avec une valeur-p du test de 0,0994 et l'effet de l'interaction groupe*periode devient non significatif pour le cas 3 avec une valeur-p du test de 0,1138. Pour $pvm=0,15$, l'effet principal du groupe devient non significatif pour le cas 2 avec une valeur-p du test de 0,1283 et l'effet de l'interaction groupe*periode devient non significatif pour le cas 3 avec une valeur-p du test de 0,0727, rendant ainsi les conclusions des tests des effets des facteurs pour les cas 2 et 3 différentes de celles obtenues pour le cas 1, tandis que pour le cas 4, les conclusions des tests des effets des facteurs demeurent les mêmes que celles obtenues pour le cas 1.

Comparativement au cas 1, les conclusions des tests des effets des facteurs s'accordent 3 fois sur 6 pour le cas 2, 2 fois sur 6 pour le cas 3 et 5 fois sur 6 pour le cas 4.

Les valeurs manquantes n'exercent pas une influence significative dans les résultats de l'analyse de variance à mesures répétées pour $pvm=5\%$. Dès lors que le pourcentage des valeurs manquantes devient grand ($pvm=10\%$ et $pvm=15\%$), les différences significatives surgissent entre le cas 1 et le cas 2 dans les résultats de l'analyse de variance à mesures répétées.

En se servant de l'imputation par la moyenne locale des valeurs observées pour le même sujet (cas 4), nous constatons que seulement dans une situation

sur 6 les résultats de l'analyse de variance à mesures répétées sont différents de ceux obtenus pour le cas des données complètes d'origine (cas 1).

La méthode d'imputation par la moyenne des valeurs observées pour le même sujet (cas 3) ne permet pas de réduire l'impact négatif causé par la présence des données manquantes dans les résultats de l'analyse de variance à mesures répétées suite à l'utilisation de la méthode des cas complets (cas 2).

La méthode d'imputation par la moyenne locale des valeurs observées pour le même sujet (cas 4) donne des meilleurs résultats comparativement à la méthode d'imputation par la moyenne des valeurs observées pour le même sujet (cas 3).

3.3.3. Résultats pour le scénario 3

Pour le scénario 3 relatif aux ensembles de données en rapport avec la matrice de covariances de structure autorégressive d'ordre 1, la valeur-p du test de sphéricité de Mauchly est inférieure à 0,001 pour les quatre cas et cela dans les deux contextes des valeurs manquantes (voir les tableaux C.0.5 et C.0.6 de l'annexe C pour la lecture des résultats concernant le scénario 3). Pour chaque cas, la condition d'application de sphéricité pour la matrice de covariances des données à mesures répétées n'est pas satisfaite. Nous considérons donc pour les effets intra-sujet, les résultats des tests des effets des facteurs ajustés par le facteur de correction $\bar{\epsilon}$ de Huynh et Feldt (1976).

Pour le cas 1, l'effet de l'interaction groupe* période est significatif avec une valeur-p du test de 0,001. Il n'y a pas de parallélisme entre les trois courbes des moyennes dans le temps pour les trois groupes et ainsi, l'effet principal du groupe et l'effet principal de la période ne sont pas indépendants. L'effet principal du groupe est non significatif avec une valeur-p du test de 0,0582 et l'effet principal de la période est significatif avec une valeur-p du test inférieure à 0,0001.

Dans le contexte 1 des valeurs manquantes, nous observons que pour $pvm=0,05$, les conclusions des tests des effets des facteurs pour le cas 2 demeurent les mêmes que celles obtenues pour le cas 1, tandis que pour les cas 3 et 4, l'effet principal

du groupe devient faiblement significatif avec des valeurs-p du test de 0,0468 et 0,0472 respectivement, conduisant ainsi à des conclusions des tests des effets des facteurs différentes de celles obtenues pour le cas 1. Pour $pvm=0,10$, les conclusions des tests des effets des facteurs pour le cas 3 demeurent les mêmes que celles obtenues pour le cas 1 tandis que ces conclusions sont différentes pour les cas 2 et 4 car l'effet principal du groupe devient fortement significatif pour le cas 2 et à peine significatif pour le cas 4 avec des valeurs-p du test de 0,0038 et 0,0457 respectivement. Pour $pvm=0,15$, l'effet principal du groupe devient significatif pour les cas 2, 3 et 4 avec des valeurs-p du test de 0,0081, 0,0402 et 0,0365 respectivement, conduisant ainsi à des conclusions des tests des effets des facteurs différentes de celles obtenues pour le cas 1.

Dans le contexte 2 des valeurs manquantes, nous observons que pour $pvm=0,05$ et $pvm=0,10$, les conclusions des tests des effets des facteurs pour les cas 2 et 4 demeurent les mêmes que celles obtenues pour le cas 1, tandis que pour le cas 3, l'effet de l'interaction groupe**période* devient non significatif avec des valeurs-p du test de 0,0545 et 0,2430 respectivement pour $pvm=0,05$ et $pvm=0,10$, conduisant ainsi à des conclusions des tests des effets des facteurs différentes de celles obtenues pour le cas 1. Pour $pvm=0,15$, les conclusions des tests des effets des facteurs pour le cas 2 demeurent les mêmes que celles obtenues pour le cas 1. Par contre pour les cas 3 et 4, l'effet de l'interaction groupe**période* devient non significatif avec des valeurs-p du test de 0,2806 et 0,1444 respectivement.

Comparativement au cas 1, les conclusions des tests des effets des facteurs s'accordent 4 fois sur 6 pour le cas 2, 1 fois sur 6 pour le cas 3 et 2 fois sur 6 pour le cas 4.

Le fait de la présence des valeurs manquantes pour le cas 2 influence les résultats de l'analyse de variance à mesures répétées et les résultats sont encore plus mauvais en se servant des deux méthodes d'imputation utilisées.

3.3.4. Résultats pour le scénario 4

Pour le scénario 4 relatif aux ensembles de données en rapport avec la matrice de covariances de structure symétrie composée hétérogène, la valeur-p du test de sphéricité de Mauchly est supérieure à 0,05 pour le cas 3 avec $pvm=0,05$ et les cas 2 avec les probabilités des données manquantes $pvm=0,10$ et $pvm=0,15$ dans le contexte 1 des valeurs manquantes, satisfaisant ainsi la condition d'application de sphéricité pour la matrice de covariances des données à mesures répétées pour ces cas (voir les tableaux C.0.7 et C.0.8 de l'annexe C pour la lecture des résultats concernant le scénario 4). Pour toutes les autres situations, la valeur-p du test de sphéricité de Mauchly est inférieure à 0,05, indiquant que la condition d'application de sphéricité pour la matrice de covariances des données à mesures répétées n'est pas satisfaite. Pour ces situations, nous considérons pour les effets intra-sujet, les résultats des tests des effets des facteurs ajustés par le facteur de correction $\bar{\epsilon}$ de Huynh et Feldt (1976).

Pour le cas 1, l'effet de l'interaction groupe*periode est significatif avec une valeur-p du test inférieure à 0,0001. Il n'y a pas de parallélisme entre les trois courbes des moyennes dans le temps pour les trois groupes. L'effet principal du groupe et l'effet principal de la période ne sont donc pas indépendants. L'effet principal du groupe est non significatif avec une valeur-p du test de 0,1779 et l'effet principal de la période est significatif avec une valeur-p du test inférieure à 0,0001.

Dans le contexte 1 des valeurs manquantes, nous observons que pour $pvm=0,05$ les conclusions des tests des effets des facteurs pour les cas 2, 3 et 4 demeurent les mêmes que celles obtenues pour le cas 1. Pour $pvm=0,10$ et $pvm=0,15$, l'effet principal du groupe pour le cas 2 devient significatif avec des valeurs-p du test de 0,0153 et 0,0298 respectivement, conduisant ainsi à des conclusions des tests des effets des facteurs différentes de celles obtenues pour le cas 1, tandis que pour les

cas 3 et 4, ces conclusions demeurent les mêmes que celles obtenues pour le cas 1.

Dans le contexte 2 des valeurs manquantes, nous observons que pour $pvm=0,05$, les conclusions des tests des effets des facteurs pour les cas 2, 3 et 4 demeurent les mêmes que celles obtenues pour le cas 1. Pour $pvm=0,10$ et $pvm=0,15$, l'effet de l'interaction groupe*periode pour le cas 3 devient non significatif avec des valeurs-p du test de 0,2065 et 0,2608 respectivement, conduisant ainsi à des conclusions des tests des effets des facteurs différentes de celles obtenues pour le cas 1, tandis que pour les cas 2 et 4, ces conclusions demeurent les mêmes que celles obtenues pour le cas 1.

Comparativement au cas 1, les conclusions des tests des effets des facteurs s'accordent 4 fois sur 6 pour le cas 2, 4 fois sur 6 pour le cas 3 et 6 fois sur 6 pour le cas 4.

En utilisant la méthode des cas complets (cas 2) pour faire face aux données manquantes, nous constatons que l'influence des valeurs manquantes est significative dans les résultats de l'analyse de variance à mesures répétées pour dès lors que le pourcentage des valeurs manquantes est grand ($pvm=10\%$ et $pvm=15\%$). Pour $pvm=5\%$, les conclusions des tests des effets des facteurs sont exactement les mêmes pour toutes les situations.

L'impact négatif dans les résultats de l'analyse de variance à mesures répétées dû à la présence des données manquantes pour le cas 2 a été complètement anéanti grâce à l'utilisation de la méthode d'imputation par la moyenne locale des valeurs observées pour le même sujet (cas 4) puisque les conclusions des tests des effets des facteurs trouvées pour le cas 4 sont exactement les mêmes pour le cas 1.

Pour la méthode d'imputation par la moyenne des valeurs observées pour le même sujet (cas 3), nous constatons que les résultats de l'analyse de variance à mesures répétées ne sont pas meilleurs comparativement à ceux obtenus suite à l'utilisation de la méthode des cas complets (cas 2).

Encore une fois, nous constatons que la méthode d'imputation par la moyenne locale des valeurs observées pour le même sujet (cas 4) donne des meilleurs résultats comparativement à la méthode d'imputation par la moyenne des valeurs observées pour le même sujet (cas 3).

3.4. RÉSULTATS ISSUS DE LA PROCÉDURE MIXED

Les valeurs des critères d'information AIC et BIC les plus petites sont celles sur lesquelles nous nous basons pour choisir la meilleure structure de la matrice de covariances pour chaque ensemble de données à mesures répétées en comparant ces valeurs entre elles pour les quatre structures de la matrice de covariances considérées. Ainsi pour le cas 1 et cela pour chacun des quatre scénarios considérés, les valeurs des critères d'information AIC et BIC les plus petites sont obtenues en spécifiant la structure de la matrice de covariances utilisée au départ pour générer chaque ensemble de données complètes d'origine correspondant.

Pour les quatre scénarios, les valeurs des critères d'information AIC et BIC sont présentées dans les tableaux D.0.1 à E.0.6 des annexes D et E. Les valeurs soulignées représentent les valeurs les plus petites pour chaque cas. Les résultats des tests des effets des facteurs sont présentés dans les tableaux F.0.1 à G.0.12 des annexes F et G. Les résultats considérés pour l'interprétation des tests des effets des facteurs sont ceux pour lesquels la structure de la matrice de covariances est la meilleure. Ces résultats sont soulignés dans les tableaux F.0.1 à G.0.12. Les autres résultats servent simplement pour des fins de comparaison pour indiquer comment les résultats seraient déformés advenant un mauvais choix de la structure de la matrice de covariances.

Une valeur-p inférieure à 5% indique que l'effet du facteur est significatif.

3.4.1. Résultats pour le scénario 1

Pour le cas 1, les valeurs des critères d'information AIC et BIC les moins élevées correspondent à celles pour la matrice de covariances de structure symétrie composée (voir les tableaux D.0.1 à D.0.6 pour les valeurs des critères d'information AIC et BIC pour le scénario 1). Donc pour le cas des données complètes d'origines, la structure symétrie composée pour la matrice de covariances est celle qui décrit mieux les données à mesures répétées. L'effet de l'interaction groupe*periode est non significatif avec une valeur-p du test de 0,1129 (voir les tableaux F.0.1 à F.0.6 pour la lecture des résultats des tests des effets des facteurs pour le scénario 1). Les trois courbes des moyennes dans le temps pour les trois groupes sont donc parallèles et les effets principaux du groupe et de la période sont indépendants. L'effet principal du groupe et l'effet principal de la période sont significatifs avec des valeurs-p inférieures à 0,0001 pour les deux effets principaux. Donc les moyennes des niveaux du facteur groupe ne sont pas toutes égales et les différences observées entre les moyennes des niveaux du facteur groupe sont sensiblement les mêmes pour chaque niveau du facteur temps. De même, il y a des différences entre les moyennes des niveaux du facteur temps et ces différences sont sensiblement les mêmes pour chaque niveau du facteur groupe.

Pour les cas 2, 3 et 4, nous constatons que la structure symétrie composée pour la matrice de covariances demeure celle qui présente les valeurs des critères d'information AIC et BIC les moins élevées sauf pour les cas 4 avec $pvm=0,10$ et $pvm=0,15$ dans le contexte 1 des valeurs manquantes où les valeurs des critères d'information AIC et BIC les moins élevées sont celles pour la matrice de covariances de structure autorégressive d'ordre 1.

Dans le contexte 1 des valeurs manquantes, nous observons que pour $pvm=0,05$, les conclusions des tests des effets des facteurs pour le cas 2 demeurent les mêmes que celles obtenues pour le cas 1, tandis que pour les cas 3 et 4, l'effet de l'interaction entre le facteur groupe et le facteur temps devient significatif avec une

valeur-p de 0,0313 dans les deux cas, conduisant ainsi à des conclusions des tests des effets des facteurs différentes de celles obtenues pour le cas 1. Pour $pvm=0,10$ et $pvm=0,15$, les conclusions des tests des effets des facteurs pour les cas 2, 3 et 4 sont identiques à celles obtenues pour le cas 1.

Dans le contexte 2 des valeurs manquantes, nous observons que pour $pvm=0,05$, $pvm=0,10$ et $pvm=0,15$, les conclusions des tests des effets des facteurs pour les cas 2, 3 et 4 demeurent les mêmes que celles obtenues pour le cas 1.

Comparativement au cas 1, les conclusions des tests des effets des facteurs s'accordent 6 fois sur 6 pour le cas 2 et 5 fois sur 6 pour les cas 3 et 4.

En utilisant la méthode des cas disponibles pour faire face aux données manquantes, nous constatons que les données manquantes n'ont pas d'influence sur les résultats de l'analyse de variance à mesures répétées puisque les résultats obtenus pour le cas des données complètes d'origine (cas 1) sont exactement les mêmes que ceux pour la méthode des cas disponibles utilisée pour le cas 2 pour ce qui concerne les tests des effets principaux des facteurs et le test de l'effet de l'interaction entre le facteur groupe et le facteur temps.

3.4.2. Résultats pour le scénario 2

Pour le cas 1, les valeurs des critères d'information AIC et BIC les moins élevées sont celles pour la matrice de covariances non-structurée (voir les tableaux D.0.7 à D.0.12 pour les valeurs des critères d'information AIC et BIC pour le scénario 2). L'effet de l'interaction groupe*periode est significatif avec une valeur-p de 0,0003 (voir les tableaux F.0.7 à F.0.12 pour la lecture des résultats des tests des effets des facteurs pour le scénario 2). Les trois courbes des moyennes dans le temps pour les trois groupes ne sont pas parallèles et ainsi, l'effet principal du groupe et l'effet principal de la période ne sont pas indépendants. L'effet principal du groupe est significatif avec une valeur-p de 0,0009 et l'effet principal de la période est significatif avec une valeur-p inférieure à 0,0001. Donc les moyennes des niveaux du facteur groupe ne sont pas toutes égales et les

différences observées entre les moyennes des niveaux du facteur groupe ne sont les mêmes pour chaque niveau du facteur temps. De même, il y a des différences entre les moyennes des niveaux du facteur temps et ces différences ne sont pas les mêmes pour chaque niveau du facteur groupe. Il serait nécessaire de faire une analyse des effets simples des facteurs. Comme expliqué dans la section précédente sur les résultats des analyses des effets des facteurs avec GLM, notre intérêt dans le cadre de ce mémoire se limite à des conclusions générales.

Pour le cas 2, les valeurs des critères d'information AIC et BIC les moins élevées demeurent celles pour la matrice de covariances non structurée pour $pvm=0,05$. Pour les probabilités des données manquantes plus grandes $pvm=0,10$ et $pvm=0,15$ la valeur du critère d'information AIC la moins élevée est celle pour la matrice de covariances non structurée, par contre la valeur du critère d'information BIC la moins élevée est celle pour la matrice de covariances de structure symétrie composée. Pour le cas 3, la valeur du critère d'information AIC la moins élevée est celle pour la matrice de covariances non structurée pour les trois niveaux de la probabilité des données manquantes et cela dans les deux contextes des valeurs manquantes. La valeur du critère d'information BIC la moins élevée demeure celle pour la matrice de covariances non structurée pour $pvm=0,05$ et $pvm=0,15$ dans le contexte 1 des valeurs manquantes. Pour $pvm=0,10$ dans le contexte 1 des valeurs manquantes et pour les trois niveaux de la probabilité des données manquantes dans le contexte 2 des valeurs manquantes, la valeur du critère d'information BIC la moins élevée devient celle pour la matrice de covariances de structure symétrie composée. Pour le cas 4, nous observons les mêmes résultats comme pour le cas 3 en ce qui concerne les valeurs des critères d'information AIC et BIC.

Dans le contexte 1 des valeurs manquantes, nous observons que pour $pvm=0,05$ les conclusions des tests des effets des facteurs pour les cas 2, 3 et 4 demeurent les mêmes que celles obtenues pour le cas 1. Pour $pvm=0,10$, les conclusions des

tests des effets des facteurs pour le cas 2 demeurent les mêmes que celles obtenues pour le cas 1, tandis que les conclusions des tests des effets des facteurs pour les cas 3 et 4 sont différentes de celles obtenues pour le cas 1 car pour le cas 3, l'effet de l'interaction groupe*periode devient non significatif et pour le cas 4, en considérant les résultats pour la matrice de covariances de structure symétrie composée, on constate que l'effet de l'interaction groupe*periode devient non significatif. Pour $pvm=0,15$, les conclusions des tests des effets des facteurs pour les cas 2 et 4 demeurent les mêmes que celles obtenues pour le cas 1, par contre pour le cas 3, l'effet de l'interaction groupe*periode devient non significatif avec une valeur-p de 0,0628, conduisant ainsi à des conclusions pour les tests des effets des facteurs différentes de celles obtenues pour le cas 1.

Dans le contexte 2 des valeurs manquantes, nous observons que pour $pvm=0,05$, les conclusions des tests des effets des facteurs pour les cas 2, 3 et 4 demeurent les mêmes que celles obtenues pour le cas 1. Pour $pvm=0,10$ et $pvm=0,15$, les conclusions des tests des effets des facteurs pour les cas 2 et 4 demeurent les mêmes que celles obtenues pour le cas 1 tandis que pour le cas 3, l'effet de l'interaction groupe*periode devient non significatif, conduisant ainsi à des conclusions pour les tests des effets des facteurs différentes de celles obtenues pour le cas 1.

Comparativement au cas 1, les conclusions des tests des effets des facteurs s'accordent 6 fois sur 6 pour le cas 2, 2 fois sur 6 pour le cas 3 et 5 fois sur 6 pour le cas 4.

En utilisant la méthode des cas disponibles pour faire face aux données manquantes, nous constatons que les données manquantes n'ont pas d'influence sur les résultats de l'analyse de variance à mesures répétées puisque les résultats obtenus pour le cas des données complètes d'origine (cas 1) sont exactement les mêmes que ceux pour la méthode des cas disponibles utilisée pour le cas 2 pour ce qui concerne les tests des effets principaux des facteurs et le test de l'effet de l'interaction entre le facteur groupe et le facteur temps.

3.4.3. Résultats pour le scénario 3

Les valeurs des critères d'information AIC et BIC les moins élevées sont celles pour la matrice de covariances de structure autorégressive d'ordre 1 pour tous les cas et à tous les niveaux de la probabilité des données manquantes (voir les tableaux D.0.13 à D.0.18 pour les valeurs des critères d'information AIC et BIC pour le scénario 3). De ce fait les résultats considérés pour les tests des effets des facteurs sont uniquement ceux pour la matrice de covariances de structure autorégressive d'ordre 1.

Pour le cas 1, l'effet de l'interaction groupe* période est significatif avec une valeur-p du test de 0,0222 (voir les tableaux G.0.1 à G.0.6 pour la lecture des résultats des tests des effets des facteurs pour le scénario 3). Il n'y a pas de parallélisme entre les trois courbes des moyennes dans le temps pour les trois groupes et ainsi, l'effet principal du groupe et l'effet principal de la période ne sont pas indépendants. L'effet principal du groupe est non significatif avec une valeur-p du test de 0,0582 et l'effet de la période est significatif avec une valeur-p du test inférieure à 0,0001.

Dans le contexte 1 des valeurs manquantes, nous observons que pour $pvm=0,05$, les conclusions des tests des effets des facteurs pour le cas 2 demeurent les mêmes que celles obtenues pour le cas 1 tandis que pour les cas 3 et 4, les conclusions deviennent différentes de celles obtenues pour le cas 1 car l'effet principal du groupe devient significatif avec des valeurs-p du test de 0,0331 et 0,0410 respectivement. Pour $pvm=0,10$, l'effet principal du groupe devient significatif pour le cas 2 avec une valeur-p du test de 0,0341 et l'effet de l'interaction groupe* période devient non significatif pour les cas 3 et 4 avec des valeurs-p du test de 0,0772 et 0,0694 respectivement ; ainsi les conclusions des tests des effets des facteurs pour les cas 2, 3 et 4 deviennent différentes de celles obtenues pour le cas 1. Pour $pvm=0,15$, l'effet principal du groupe devient significatif pour les cas 2, 3 et 4 avec des valeurs-p du test de 0,0285, 0,0357 et 0,0399 respectivement tandis

que l'effet de l'interaction groupe*periode devient non significatif pour le cas 3 avec une valeur-p du test de 0,1043 ; ainsi les conclusions des tests des effets des facteurs pour les cas 2, 3 et 4 deviennent différentes de celles obtenues pour le cas 1.

Dans le contexte 2 des valeurs manquantes, nous observons que pour $pvm=0,05$, les conclusions des tests des effets des facteurs pour le cas 2 demeurent les mêmes que celles obtenues pour le cas 1 tandis que pour les cas 3 et 4, les conclusions deviennent différentes de celles obtenues pour le cas 1 car l'effet de l'interaction groupe*periode devient non significatif avec des valeurs-p du test de 0,3269 et 0,0994 respectivement. Pour $pvm=0,10$, les conclusions des tests des effets des facteurs pour le cas 2 demeurent les mêmes que celles obtenues pour le cas 1 tandis que pour les cas 3 et 4, les conclusions deviennent différentes de celles obtenues pour le cas 1 car l'effet de l'interaction groupe*periode devient non significatif avec des valeurs-p du test de 0,6800 et 0,1178 respectivement. Pour $pvm=0,15$, l'effet de l'interaction groupe*periode devient non significatif pour les cas 2, 3 et 4 avec des valeurs-p du test de 0,0545, 0,3512 et 0,2033 respectivement, conduisant ainsi à des conclusions pour les tests des effets des facteurs différentes de celles obtenues pour le cas 1.

Comparativement au cas 1, les conclusions des tests des effets des facteurs s'accordent 3 fois sur 6 pour le cas 2, 0 fois sur 6 pour les cas 3 et 4.

En utilisant la méthode des cas disponibles pour faire face aux données manquantes, nous constatons que les données manquantes ont une influence sur les résultats de l'analyse de variance à mesures répétées car les résultats obtenus pour le cas des données complètes d'origine (cas 1) ne s'accordent pas dans 3 situations sur 6 avec les résultats obtenus en utilisant la méthode des cas disponibles pour le cas 2.

Les deux méthodes d'imputation utilisées ne permettent pas de corriger l'impact négatif dans les résultats de l'analyse de variance à mesures répétées, du à

la présence des données manquantes dans le cas 2. Le fait d'imputer à l'aide des deux méthodes d'imputation utilisées donne des résultats encore plus mauvais.

3.4.4. Résultats pour le scénario 4

Les valeurs des critères d'information AIC et BIC les moins élevées sont celles pour la matrice de covariances de structure symétrie composée hétérogène pour tous les cas et à tous les niveaux de la probabilité des données manquantes excepté les cas 3 et 4 avec $pvm=0,10$ dans le contexte 1 des valeurs manquantes et le cas 3 avec $pvm=0,15$ dans le contexte 2 des valeurs manquantes où les valeurs du critère d'information BIC les plus faibles sont celles pour la matrice de covariances non-structurée (voir les tableaux E.0.1 à E.0.6 pour les valeurs des critères d'information AIC et BIC pour le scénario 4).

Pour le cas 1, l'effet de l'interaction groupe*periode est significatif avec une valeur-p du test inférieure à 0,0001 (voir les tableaux G.0.7 à G.0.12 pour la lecture des résultats des tests des effets des facteurs pour le scénario 4). Il n'y a pas de parallélisme entre les trois courbes des moyennes dans le temps pour les trois groupes et ainsi, l'effet principal du groupe et l'effet principal de la période ne sont pas indépendants. L'effet principal du groupe est non significatif avec une valeur-p du test de 0,1779 et l'effet de la période est significatif avec une valeur-p du test inférieure à 0,0001.

Dans le contexte 1 des valeurs manquantes, nous observons que pour $pvm=0,05$, $pvm=0,10$ et $pvm=0,15$, les conclusions des tests des effets des facteurs pour les cas 2, 3 et 4 demeurent les mêmes que celles obtenues pour le cas 1.

Dans le contexte 2 des valeurs manquantes, nous observons que pour $pvm=0,05$, les conclusions des tests des effets des facteurs pour les cas 2, 3 et 4 demeurent les mêmes que celles obtenues pour le cas 1. Pour $pvm=0,10$ et $pvm=0,15$, les conclusions des tests des effets des facteurs pour les cas 2 et 4 demeurent les

mêmes que celles obtenues pour le cas 1 tandis que pour le cas 3, ces conclusions sont différentes de celles obtenues pour le cas 1 car l'effet de l'interaction groupe*periode devient non significatif.

Comparativement au cas 1, les conclusions des tests des effets des facteurs s'accordent 6 fois sur 6 pour le cas 2 et 4 fois sur 6 pour le cas 3 et 6 fois sur 6 pour le cas 4.

Les données manquantes n'ont pas d'influence sur les résultats de l'analyse de variance à mesures répétées puisque les résultats obtenus pour le cas des données complètes d'origine (cas 1) sont exactement les mêmes que ceux pour la méthode des cas disponibles utilisée pour le cas 2 pour ce qui concerne les tests des effets principaux des facteurs et le test de l'effet de l'interaction entre le facteur groupe et le facteur temps.

La méthode d'imputation par la moyenne locale des observations pour le même sujet (cas 4) donne aussi les mêmes résultats obtenus pour le cas des données complètes d'origine (cas 1) dans toutes les situations, ce qui n'est pas le cas pour la méthode d'imputation par la moyenne des observations pour le même sujet (cas 3).

CONCLUSION

Les conclusions qui ressortent des analyses de variance à deux facteurs dont un est à mesures répétées exécutées dépendent des paramètres utilisées pour générer les quatre ensembles des données complètes d'origine, de la méthode de génération des données manquantes et des deux méthodes d'imputation utilisées. Ces conclusions pourraient être différentes en modifiant les paramètres utilisées pour générer les ensembles de données et en se servant des méthodes d'imputation plus robustes telle que la méthode d'imputation multiple.

En utilisant la procédure GLM de SAS, les résultats obtenus pour les tests de l'effet principal du facteur groupe, l'effet principal du facteur temps et de l'effet de l'interaction entre le facteur groupe et le facteur temps permettent de mettre en lumière ce qui suit :

pour la méthode des cas complets (cas 2), il y a accord des résultats avec ceux obtenus pour le cas des données complètes d'origine (cas 1) dans 15 situations sur 24 (soit un pourcentage de 62,5%) et désaccord des résultats dans 9 situations sur 24 (soit un pourcentage de 37,5%) parmi lesquelles 3 situations sont dues à des différences entre les effets de l'interaction groupe*periode et 6 situations sont dues à des différences entre les effets principaux du facteur groupe. La déduction qui en découle est que les données manquantes quoique manquantes complètement aléatoirement influencent les résultats de l'analyse de variance à deux facteurs dont un est à mesures répétées.

De façon plus générale, l'amplitude des différences dans les résultats de l'analyse

de variance à mesures répétées n'est pas fonction du pourcentage des données manquantes.

En utilisant la méthode d'imputation par la moyenne des valeurs observées pour le même sujet (cas 3), nous constatons que parmi les 9 situations sur 24 dans lesquelles les résultats du cas 2 sont en désaccord avec ceux du cas 1, il y a 5 situations sur 9 dans lesquelles les résultats deviennent en accord avec le cas 1 et parmi les 15 situations sur 24 dans lesquelles les résultats du cas 2 sont en accord avec ceux du cas 1, il y a 8 situations sur 15 dans lesquelles les résultats deviennent en désaccord avec le cas 1. Ainsi la méthode d'imputation par la moyenne des valeurs observées pour le même sujet a permis de corriger 55,56% des mauvais résultats du cas 2 et de déformer 53,33% des bons résultats du cas 2. Pour le total des 12 situations sur 24 dans lesquelles les résultats du cas 3 sont en désaccord avec ceux du cas 1, nous notons 10 situations dues à des différences entre les effets de l'interaction groupe*periode et 2 situations dues à des différences entre les effets principaux du facteur groupe. Nous pouvons conclure que cette méthode d'imputation ne semble pas performer mieux que la méthode des cas complets.

En utilisant la méthode d'imputation par la moyenne locale des valeurs observées pour le même sujet, nous constatons que parmi les 9 situations sur 24 dans lesquelles les résultats du cas 2 sont en désaccord avec ceux du cas 1, il y a 6 situations sur 9 dans lesquelles les résultats deviennent en accord avec le cas 1 et parmi les 15 situations sur 24 dans lesquelles les résultats du cas 2 sont en accord avec ceux du cas 1, il y a 3 situations sur 15 dans lesquelles les résultats deviennent en désaccord avec le cas 1. Ainsi la méthode d'imputation par la moyenne locale des valeurs observées pour le même sujet a permis de corriger 66,67% de mauvais résultats du cas 2 et de déformer 20% des bons résultats du cas 2. Pour le total des 6 situations sur 24 dans lesquelles les résultats du cas 4 sont en désaccord avec ceux du cas 1, nous notons 3 situations dues à des différences entre les effets de l'interaction groupe*periode et 3 situations dues à des différences entre les effets principaux du facteur groupe. Nous pouvons conclure

que la méthode d'imputation par la moyenne locale des valeurs observées pour le même sujet performe mieux que la méthode des cas complets. Par contre la méthode d'imputation par la moyenne des valeurs observées pour le même sujet ne performe pas mieux que la méthode des cas complets.

En utilisant la procédure MIXED de SAS, les résultats obtenus pour les tests de l'effet principal du facteur groupe, l'effet principal du facteur temps et de l'effet de l'interaction entre le facteur groupe et le facteur temps montrent ce qui suit :

pour la méthode des cas disponibles (cas 2), il y a accord des résultats avec ceux obtenus pour le cas des données complètes d'origine (cas 1) dans 21 situations sur 24 (soit un pourcentage de 87,5%) et désaccord des résultats dans 3 situations sur 24 (soit un pourcentage de 12,5%) parmi lesquelles 1 situation est due à des différences entre les effets de l'interaction groupe*periode et 2 situations sont dues à des différences entre les effets principaux du facteur groupe. La déduction qui en découle est que l'influence des données manquantes complètement aléatoirement sur les résultats de l'analyse de variance à deux facteurs dont un est à mesures répétées est très réduite lorsque l'on se sert de la procédure MIXED de SAS.

En utilisant la méthode d'imputation par la moyenne des valeurs observées pour le même sujet (cas 3), nous constatons que parmi les 3 situations sur 24 dans lesquelles les résultats du cas 2 sont en désaccord avec ceux du cas 1, il y a aucune situation dans laquelle les résultats deviennent en accord avec le cas 1 et parmi les 21 situations sur 24 dans lesquelles les résultats du cas 2 sont en accord avec ceux du cas 1, il y a 10 situations sur 21 dans lesquelles les résultats deviennent en désaccord avec le cas 1. Pour le total des 13 situations sur 24 dans lesquelles les résultats du cas 3 sont en désaccord avec ceux du cas 1, nous notons 12 situations dues à des différences entre les effets de l'interaction groupe*periode et 1 situation due à des différences entre les effets principaux du facteur groupe. Ainsi la méthode d'imputation par la moyenne des valeurs observées pour le même

sujet ne corrige pas de mauvais résultat du cas 2, mais déforme 47,62% des bons résultats du cas 2.

En utilisant la méthode d'imputation par la moyenne locale des valeurs observées pour le même sujet, nous constatons que parmi les 3 situations sur 24 dans lesquelles les résultats du cas 2 sont en désaccord avec ceux du cas 1, il y a aucune situation sur 3 dans laquelle les résultats deviennent en accord avec le cas 1 et parmi les 21 situations sur 24 dans lesquelles les résultats du cas 2 sont en accord avec ceux du cas 1, il y a 5 situations sur 21 dans lesquelles les résultats deviennent en désaccord avec le cas 1. Pour le total des 8 situations sur 24 dans lesquelles les résultats du cas 4 sont en désaccord avec ceux du cas 1, nous notons 6 situations dues à des différences entre les effets de l'interaction groupe*periode et 2 situations dues à des différences entre les effets principaux du facteur groupe. Ainsi la méthode d'imputation par la moyenne locale des valeurs observées pour le même sujet permet de corriger aucun mauvais résultats du cas 2, mais plutôt déforme 23,81% des bons résultats du cas 2. Nous pouvons donc conclure que la méthode des cas disponibles performe mieux que les deux méthodes d'imputation utilisées.

En présence des données manquantes complètement aléatoirement, l'utilisation de la procédure MIXED de SAS réduit considérablement l'influence des données manquantes dans les résultats de l'analyse de variance à deux facteurs dont un est à mesures répétées. Ce qui démontre que le modèle linéaire mixte s'adapte bien à l'analyse des données à mesure répétées lorsqu'il y a des données manquantes complètement aléatoirement. Par contre lorsque l'on utilise la procédure GLM de SAS, il s'avère nécessaire de faire de l'imputation. La méthode d'imputation par la moyenne locale des valeurs observées pour le même sujet peut être utilisée dans ce cas là.

Il est à souligner que les deux méthodes d'imputation simple utilisées dans la partie analytique de ce mémoire ne donnent pas de bons résultats. Des études plus avancées avec notamment l'utilisation de la méthode d'imputation multiple

devraient permettre d'élucider un plus grand nombre de questions concernant l'influence des valeurs manquantes dans un plan d'analyse de variance à deux facteurs dont un est à mesures répétées.

Annexe A

TABLEAUX DES DONNÉES

FIG. A.0.3. Les données pour le scénario 1, pvm=0,10.

sujet	Scénario 1 cas 2 pvm=10%		Scénario 1 cas 3 pvm=10%		Scénario 1 cas 4 pvm=10%		Scénario 1 cas 5 pvm=10%		Scénario 1 cas 6 pvm=10%		Scénario 1 cas 7 pvm=10%		Scénario 1 cas 8 pvm=10%		Scénario 1 cas 9 pvm=10%	
	cont	del	cont	del	cont	del	cont	del	cont	del	cont	del	cont	del	cont	del
1	172.82	170.66	180.36	179.78	176.94	176.18	181.30	180.36	177.00	176.94	179.39	178.00	183.20	181.43	176.34	179.39
1	172.82	170.66	180.36	179.78	176.94	176.18	181.30	180.36	177.00	176.94	179.39	178.00	183.20	181.43	176.34	179.39
3	174.02	171.99	178.54	183.22	178.63	174.02	171.99	178.54	183.22	178.63	174.02	171.99	178.54	183.22	178.63	174.02
4	179.94	174.29	192.47	192.59	191.97	179.94	191.97	192.47	192.59	191.97	179.94	191.97	192.47	192.59	191.97	179.94
5	175.14	184.38	180.20	195.66	187.72	175.14	184.38	180.20	195.66	187.72	175.14	184.38	180.20	195.66	187.72	175.14
6	193.64	174.14	183.00	185.58	177.09	193.64	174.14	183.00	185.58	177.09	193.64	174.14	183.00	185.58	177.09	193.64
7	169.82	167.86	182.94	196.39	182.99	169.82	167.86	182.94	196.39	182.99	169.82	167.86	182.94	196.39	182.99	169.82
8	169.82	167.86	182.94	196.39	182.99	169.82	167.86	182.94	196.39	182.99	169.82	167.86	182.94	196.39	182.99	169.82
9	179.18	182.60	182.60	182.60	182.60	179.18	182.60	182.60	182.60	182.60	179.18	182.60	182.60	182.60	182.60	179.18
10	164.60	177.14	185.00	181.84	178.10	164.60	177.14	185.00	181.84	178.10	164.60	177.14	185.00	181.84	178.10	164.60
11	182.45	179.30	189.30	196.16	200.44	182.45	179.30	189.30	196.16	200.44	182.45	179.30	189.30	196.16	200.44	182.45
12	182.45	179.30	189.30	196.16	200.44	182.45	179.30	189.30	196.16	200.44	182.45	179.30	189.30	196.16	200.44	182.45
13	195.76	179.30	195.76	195.76	195.76	195.76	179.30	195.76	195.76	195.76	195.76	179.30	195.76	195.76	195.76	195.76
14	170.13	182.39	181.81	177.68	169.04	170.13	182.39	181.81	177.68	169.04	170.13	182.39	181.81	177.68	169.04	170.13
15	170.13	182.39	181.81	177.68	169.04	170.13	182.39	181.81	177.68	169.04	170.13	182.39	181.81	177.68	169.04	170.13
16	179.41	182.39	181.81	177.68	169.04	179.41	182.39	181.81	177.68	169.04	179.41	182.39	181.81	177.68	169.04	179.41
17	165.42	193.60	187.47	198.69	178.48	165.42	193.60	187.47	198.69	178.48	165.42	193.60	187.47	198.69	178.48	165.42
18	180.01	184.24	187.31	192.70	185.05	180.01	184.24	187.31	192.70	185.05	180.01	184.24	187.31	192.70	185.05	180.01
19	177.38	177.42	205.37	192.52	177.28	177.38	177.42	205.37	192.52	177.28	177.38	177.42	205.37	192.52	177.28	177.38
20	177.38	177.42	205.37	192.52	177.28	177.38	177.42	205.37	192.52	177.28	177.38	177.42	205.37	192.52	177.28	177.38
21	177.38	177.42	205.37	192.52	177.28	177.38	177.42	205.37	192.52	177.28	177.38	177.42	205.37	192.52	177.28	177.38
22	177.45	177.08	186.55	177.45	177.08	186.55	177.45	177.08	186.55	177.45	177.08	186.55	177.45	177.08	186.55	177.45
23	188.90	180.32	194.15	185.63	188.90	180.32	194.15	185.63	188.90	180.32	194.15	185.63	188.90	180.32	194.15	185.63
24	188.90	180.32	194.15	185.63	188.90	180.32	194.15	185.63	188.90	180.32	194.15	185.63	188.90	180.32	194.15	185.63
25	170.21	181.27	180.15	170.21	181.27	180.15	170.21	181.27	180.15	170.21	181.27	180.15	170.21	181.27	180.15	170.21
26	174.27	187.28	203.36	208.52	193.19	174.27	187.28	203.36	208.52	193.19	174.27	187.28	203.36	208.52	193.19	174.27
27	188.30	181.55	188.30	181.55	188.30	181.55	188.30	181.55	188.30	181.55	188.30	181.55	188.30	181.55	188.30	181.55
28	188.30	181.55	188.30	181.55	188.30	181.55	188.30	181.55	188.30	181.55	188.30	181.55	188.30	181.55	188.30	181.55
29	185.14	188.26	185.14	188.26	185.14	188.26	185.14	188.26	185.14	188.26	185.14	188.26	185.14	188.26	185.14	188.26
30	177.65	197.00	180.71	177.65	197.00	180.71	177.65	197.00	180.71	177.65	197.00	180.71	177.65	197.00	180.71	177.65
31	177.65	197.00	180.71	177.65	197.00	180.71	177.65	197.00	180.71	177.65	197.00	180.71	177.65	197.00	180.71	177.65
32	174.49	185.93	195.73	193.71	174.49	185.93	195.73	193.71	174.49	185.93	195.73	193.71	174.49	185.93	195.73	193.71
33	174.49	185.93	195.73	193.71	174.49	185.93	195.73	193.71	174.49	185.93	195.73	193.71	174.49	185.93	195.73	193.71
34	183.69	189.38	199.23	195.49	190.05	183.69	189.38	199.23	195.49	190.05	183.69	189.38	199.23	195.49	190.05	183.69
35	183.69	189.38	199.23	195.49	190.05	183.69	189.38	199.23	195.49	190.05	183.69	189.38	199.23	195.49	190.05	183.69
36	185.06	187.79	188.29	194.08	183.50	185.06	187.79	188.29	194.08	183.50	185.06	187.79	188.29	194.08	183.50	185.06
37	175.02	180.81	182.38	208.54	184.47	175.02	180.81	182.38	208.54	184.47	175.02	180.81	182.38	208.54	184.47	175.02
38	183.53	184.28	198.77	199.96	196.38	183.53	184.28	198.77	199.96	196.38	183.53	184.28	198.77	199.96	196.38	183.53
39	188.78	194.79	202.45	191.82	188.78	194.79	202.45	191.82	188.78	194.79	202.45	191.82	188.78	194.79	202.45	191.82
40	177.55	174.01	184.75	190.25	181.91	177.55	174.01	184.75	190.25	181.91	177.55	174.01	184.75	190.25	181.91	177.55
41	186.59	189.10	189.23	192.71	186.59	189.10	189.23	192.71	186.59	189.10	189.23	192.71	186.59	189.10	189.23	192.71
42	184.65	191.19	201.00	187.82	181.89	184.65	191.19	201.00	187.82	181.89	184.65	191.19	201.00	187.82	181.89	184.65
43	184.65	191.19	201.00	187.82	181.89	184.65	191.19	201.00	187.82	181.89	184.65	191.19	201.00	187.82	181.89	184.65
44	150.66	175.74	190.89	192.54	187.46	150.66	175.74	190.89	192.54	187.46	150.66	175.74	190.89	192.54	187.46	150.66
45	171.50	196.69	188.99	192.89	193.38	171.50	196.69	188.99	192.89	193.38	171.50	196.69	188.99	192.89	193.38	171.50
46	171.50	196.69	188.99	192.89	193.38	171.50	196.69	188.99	192.89	193.38	171.50	196.69	188.99	192.89	193.38	171.50
47	172.93	187.78	195.73	208.51	196.99	172.93	187.78	195.73	208.51	196.99	172.93	187.78	195.73	208.51	196.99	172.93
48	191.43	187.28	189.07	195.73	190.05	191.43	187.28	189.07	195.73	190.05	191.43	187.28	189.07	195.73	190.05	191.43
49	181.25	197.10	190.79	199.73	200.29	181.25	197.10	190.79	199.73	200.29	181.25	197.10	190.79	199.73	200.29	181.25
50	181.25	197.10	190.79	199.73	200.29	181.25	197.10	190.79	199.73	200.29	181.25	197.10	190.79	199.73	200.29	181.25
51	186.03	183.81	187.38	183.34	182.13	186.03	183.81	187.38	183.34	182.13	186.03	183.81	187.38	183.34	182.13	186.03
52	186.79	190.12	187.55	188.87	199.40	186.79	190.12	187.55	188.87	199.40	186.79	190.12	187.55	188.87	199.40	186.79
53	194.77	201.20	202.84	195.47	197.48	194.77	201.20	202.84	195.47	197.48	194.77	201.20	202.84	195.47	197.48	194.77
54	195.33	195.81	182.61	175.48	189.57	195.33	195.81	182.61	175.48	189.57	195.33	195.81	182.61	175.48	189.57	195.33

FIG. A.0.4. Les données pour le scénario 1, pvm=0,15.

sujet	cote	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
		cas 1	cas 2	cas 3	cas 4	cas 1	cas 2	cas 3	cas 4	cas 1	cas 2	cas 3	cas 4	cas 1	cas 2	cas 3	cas 4
1	176.14	177.00	183.20	173.78	176.14	176.14	176.14	176.14	176.14	176.14	176.14	176.14	176.14	176.14	176.14	176.14	176.14
2	174.02	171.99	178.54	178.54	183.22	178.54	178.54	178.54	178.54	178.54	178.54	178.54	178.54	178.54	178.54	178.54	178.54
3	178.84	191.97	202.47	192.59	192.59	192.59	192.59	192.59	192.59	192.59	192.59	192.59	192.59	192.59	192.59	192.59	192.59
4	175.14	194.38	183.00	195.66	197.72	175.14	194.38	183.00	195.66	197.72	175.14	194.38	183.00	195.66	197.72	175.14	194.38
5	172.82	187.89	183.00	185.58	176.87	187.89	183.00	185.58	176.87	187.89	183.00	185.58	176.87	187.89	183.00	185.58	176.87
6	169.61	169.78	182.80	183.39	183.88	169.61	169.78	182.80	183.39	183.88	169.61	169.78	182.80	183.39	183.88	169.61	169.78
9	174.18	182.25	184.60	181.84	174.18	182.25	184.60	181.84	174.18	182.25	184.60	181.84	174.18	182.25	184.60	181.84	174.18
11	164.60	177.14	185.00	177.14	185.00	177.14	185.00	177.14	185.00	177.14	185.00	177.14	185.00	177.14	185.00	177.14	185.00
12	182.57	174.20	189.20	186.45	174.20	189.20	186.45	174.20	189.20	186.45	174.20	189.20	186.45	174.20	189.20	186.45	174.20
13	182.57	174.20	189.20	186.45	174.20	189.20	186.45	174.20	189.20	186.45	174.20	189.20	186.45	174.20	189.20	186.45	174.20
14	172.58	174.37	180.02	175.15	176.68	172.58	174.37	180.02	175.15	176.68	172.58	174.37	180.02	175.15	176.68	172.58	174.37
15	170.13	182.98	181.81	177.68	189.04	170.13	182.98	181.81	177.68	189.04	170.13	182.98	181.81	177.68	189.04	170.13	182.98
16	184.29	197.66	183.65	184.29	197.66	183.65	184.29	197.66	183.65	184.29	197.66	183.65	184.29	197.66	183.65	184.29	197.66
17	178.41	182.39	188.13	185.19	183.78	178.41	182.39	188.13	185.19	183.78	178.41	182.39	188.13	185.19	183.78	178.41	182.39
18	165.42	153.50	187.47	195.99	178.48	165.42	153.50	187.47	195.99	178.48	165.42	153.50	187.47	195.99	178.48	165.42	153.50
19	183.50	192.66	180.18	193.50	189.66	183.50	192.66	180.18	193.50	189.66	183.50	192.66	180.18	193.50	189.66	183.50	192.66
20	177.28	177.42	205.94	205.37	192.52	177.28	177.42	205.94	205.37	192.52	177.28	177.42	205.94	205.37	192.52	177.28	177.42
21	182.57	174.20	189.20	186.45	174.20	189.20	186.45	174.20	189.20	186.45	174.20	189.20	186.45	174.20	189.20	186.45	174.20
22	177.45	177.08	185.55	177.45	177.08	185.55	177.45	177.08	185.55	177.45	177.08	185.55	177.45	177.08	185.55	177.45	177.08
23	186.90	180.32	194.15	185.63	186.90	180.32	194.15	185.63	186.90	180.32	194.15	185.63	186.90	180.32	194.15	185.63	186.90
24	170.21	187.28	205.36	193.05	187.62	170.21	187.28	205.36	193.05	187.62	170.21	187.28	205.36	193.05	187.62	170.21	187.28
25	183.00	181.55	178.23	181.94	168.83	183.00	181.55	178.23	181.94	168.83	183.00	181.55	178.23	181.94	168.83	183.00	181.55
26	168.83	181.55	178.23	181.94	168.83	181.55	178.23	181.94	168.83	181.55	178.23	181.94	168.83	181.55	178.23	181.94	168.83
27	178.04	191.76	200.17	178.04	189.99	191.76	200.17	178.04	189.99	191.76	200.17	178.04	189.99	191.76	200.17	178.04	189.99
28	177.65	197.00	177.65	189.62	197.00	177.65	189.62	197.00	177.65	189.62	197.00	177.65	189.62	197.00	177.65	189.62	197.00
29	174.49	185.93	195.73	193.71	185.93	174.49	185.93	195.73	193.71	185.93	174.49	185.93	195.73	193.71	185.93	174.49	185.93
30	183.69	192.28	193.29	187.62	183.69	192.28	193.29	187.62	183.69	192.28	193.29	187.62	183.69	192.28	193.29	187.62	183.69
31	183.69	192.28	193.29	187.62	183.69	192.28	193.29	187.62	183.69	192.28	193.29	187.62	183.69	192.28	193.29	187.62	183.69
32	185.14	188.26	198.66	193.54	185.14	188.26	198.66	193.54	185.14	188.26	198.66	193.54	185.14	188.26	198.66	193.54	185.14
33	185.14	188.26	198.66	193.54	185.14	188.26	198.66	193.54	185.14	188.26	198.66	193.54	185.14	188.26	198.66	193.54	185.14
34	185.14	188.26	198.66	193.54	185.14	188.26	198.66	193.54	185.14	188.26	198.66	193.54	185.14	188.26	198.66	193.54	185.14
35	185.14	188.26	198.66	193.54	185.14	188.26	198.66	193.54	185.14	188.26	198.66	193.54	185.14	188.26	198.66	193.54	185.14
36	175.02	180.81	192.38	208.54	184.47	175.02	180.81	192.38	208.54	184.47	175.02	180.81	192.38	208.54	184.47	175.02	180.81
37	183.53	184.28	195.77	195.99	183.53	184.28	195.77	195.99	183.53	184.28	195.77	195.99	183.53	184.28	195.77	195.99	183.53
38	197.58	174.73	189.10	206.35	181.82	197.58	174.73	189.10	206.35	181.82	197.58	174.73	189.10	206.35	181.82	197.58	174.73
39	177.55	198.77	188.23	182.77	193.00	177.55	198.77	188.23	182.77	193.00	177.55	198.77	188.23	182.77	193.00	177.55	198.77
40	184.65	191.10	182.09	199.17	184.65	191.10	182.09	199.17	184.65	191.10	182.09	199.17	184.65	191.10	182.09	199.17	184.65
41	182.26	181.89	182.26	191.19	185.79	182.26	181.89	182.26	191.19	185.79	182.26	181.89	182.26	191.19	185.79	182.26	181.89
42	190.66	190.66	175.74	190.66	175.74	190.66	190.66	175.74	190.66	190.66	175.74	190.66	190.66	190.66	175.74	190.66	190.66
43	171.50	186.79	192.49	194.39	171.50	186.79	192.49	194.39	171.50	186.79	192.49	194.39	171.50	186.79	192.49	194.39	171.50
44	171.50	186.79	192.49	194.39	171.50	186.79	192.49	194.39	171.50	186.79	192.49	194.39	171.50	186.79	192.49	194.39	171.50
45	171.50	186.79	192.49	194.39	171.50	186.79	192.49	194.39	171.50	186.79	192.49	194.39	171.50	186.79	192.49	194.39	171.50
46	173.15	183.10	196.53	178.97	193.14	173.15	183.10	196.53	178.97	193.14	173.15	183.10	196.53	178.97	193.14	173.15	183.10
47	191.43	187.28	188.07	195.73	190.05	191.43	187.28	188.07	195.73	190.05	191.43	187.28	188.07	195.73	190.05	191.43	187.28
48	181.25	197.10	190.79	195.73	200.29	181.25	197.10	190.79	195.73	200.29	181.25	197.10	190.79	195.73	200.29	181.25	197.10
49	180.03	183.81	187.38	183.34	182.13	180.03	183.81	187.38	183.34	182.13	180.03	183.81	187.38	183.34	182.13	180.03	183.81
50	180.03	183.81	187.38	183.34	182.13	180.03	183.81	187.38	183.34	182.13	180.03	183.81	187.38	183.34	182.13	180.03	183.81
51	187.28	197.58	189.87	189.40	186.79	187.28	197.58	189.87	189.40	186.79	187.28	197.58	189.87	189.40	186.79	187.28	197.58
52	197.28	184.77	201.94	202.34	195.47	197.28	184.77	201.94	202.34	195.47	197.28	184.77	201.94	202.34	195.47	197.28	184.77
53	185.79	190.12	187.55	198.87	193.40	185.79	190.12	187.55	198.87	193.40	185.79	190.12	187.55	198.87	193.40	185.79	190.12
54	175.48	198.33	195.81	182.61	175.48	198.33	195.81	182.61	175.48	198.33	195.81	182.61	175.48	198.33	195.81	182.61	175.48

FIG. A.0.13. Les données pour le scénario 4, pvm=0,15.

sujet	cote	Scénario 4															
		contexte 1 cas 2 pvm=15%				contexte 1 cas 3 pvm=15%				contexte 1 cas 4 pvm=15%				contexte 2 cas 3 pvm=15%			
		per	per2	per3	per4	per1	per2	per3	per4	per1	per2	per3	per4	per1	per2	per3	per4
1	1	172,62	186,01	188,59	183,92	172,62	186,01	188,59	183,92	172,62	186,01	188,59	183,92	172,62	186,01	188,59	183,92
2	1	176,14	175,09	191,24	183,90	176,14	175,09	191,24	183,90	176,14	175,09	191,24	183,90	176,14	175,09	191,24	183,90
3	1	179,84	179,75	190,94	192,05	179,84	179,75	190,94	192,05	179,84	179,75	190,94	192,05	179,84	179,75	190,94	192,05
4	1	175,14	182,36	183,95	185,16	175,14	182,36	183,95	185,16	175,14	182,36	183,95	185,16	175,14	182,36	183,95	185,16
5	1	183,84	185,19	191,87	193,67	183,84	185,19	191,87	193,67	183,84	185,19	191,87	193,67	183,84	185,19	191,87	193,67
6	1	177,82	186,67	188,14	184,35	177,82	186,67	188,14	184,35	177,82	186,67	188,14	184,35	177,82	186,67	188,14	184,35
7	1	169,51	187,32	180,93	197,46	169,51	187,32	180,93	197,46	169,51	187,32	180,93	197,46	169,51	187,32	180,93	197,46
8	1	179,18	188,06	190,94	192,05	179,18	188,06	190,94	192,05	179,18	188,06	190,94	192,05	179,18	188,06	190,94	192,05
9	1	164,60	168,93	173,92	180,06	164,60	168,93	173,92	180,06	164,60	168,93	173,92	180,06	164,60	168,93	173,92	180,06
10	1	182,45	186,53	194,64	197,18	182,45	186,53	194,64	197,18	182,45	186,53	194,64	197,18	182,45	186,53	194,64	197,18
11	1	165,57	168,53	172,51	182,40	165,57	168,53	172,51	182,40	165,57	168,53	172,51	182,40	165,57	168,53	172,51	182,40
12	1	170,13	183,32	180,93	192,05	170,13	183,32	180,93	192,05	170,13	183,32	180,93	192,05	170,13	183,32	180,93	192,05
13	1	189,92	194,34	196,26	190,69	189,92	194,34	196,26	190,69	189,92	194,34	196,26	190,69	189,92	194,34	196,26	190,69
14	1	165,42	177,25	172,22	185,54	165,42	177,25	172,22	185,54	165,42	177,25	172,22	185,54	165,42	177,25	172,22	185,54
15	1	180,01	185,58	190,22	196,19	180,01	185,58	190,22	196,19	180,01	185,58	190,22	196,19	180,01	185,58	190,22	196,19
16	1	183,50	193,60	195,51	203,15	183,50	193,60	195,51	203,15	183,50	193,60	195,51	203,15	183,50	193,60	195,51	203,15
17	1	177,55	178,97	195,73	196,08	177,55	178,97	195,73	196,08	177,55	178,97	195,73	196,08	177,55	178,97	195,73	196,08
18	1	188,90	192,21	202,95	195,74	188,90	192,21	202,95	195,74	188,90	192,21	202,95	195,74	188,90	192,21	202,95	195,74
19	2	174,27	182,22	186,80	183,84	174,27	182,22	186,80	183,84	174,27	182,22	186,80	183,84	174,27	182,22	186,80	183,84
20	2	188,30	192,28	171,38	203,58	188,30	192,28	171,38	203,58	188,30	192,28	171,38	203,58	188,30	192,28	171,38	203,58
21	2	178,63	192,69	198,69	184,14	178,63	192,69	198,69	184,14	178,63	192,69	198,69	184,14	178,63	192,69	198,69	184,14
22	2	177,65	193,35	187,70	192,09	177,65	193,35	187,70	192,09	177,65	193,35	187,70	192,09	177,65	193,35	187,70	192,09
23	2	174,49	181,74	189,94	193,55	174,49	181,74	189,94	193,55	174,49	181,74	189,94	193,55	174,49	181,74	189,94	193,55
24	2	176,08	184,96	193,28	195,81	176,08	184,96	193,28	195,81	176,08	184,96	193,28	195,81	176,08	184,96	193,28	195,81
25	2	185,14	192,77	201,06	193,93	185,14	192,77	201,06	193,93	185,14	192,77	201,06	193,93	185,14	192,77	201,06	193,93
26	2	175,02	179,82	187,92	197,94	175,02	179,82	187,92	197,94	175,02	179,82	187,92	197,94	175,02	179,82	187,92	197,94
27	2	183,53	186,54	196,22	197,26	183,53	186,54	196,22	197,26	183,53	186,54	196,22	197,26	183,53	186,54	196,22	197,26
28	2	188,78	198,24	204,19	203,14	188,78	198,24	204,19	203,14	188,78	198,24	204,19	203,14	188,78	198,24	204,19	203,14
29	2	177,55	175,79	181,08	188,03	177,55	175,79	181,08	188,03	177,55	175,79	181,08	188,03	177,55	175,79	181,08	188,03
30	2	182,55	189,46	190,07	193,38	182,55	189,46	190,07	193,38	182,55	189,46	190,07	193,38	182,55	189,46	190,07	193,38
31	2	171,50	181,73	181,29	187,72	171,50	181,73	181,29	187,72	171,50	181,73	181,29	187,72	171,50	181,73	181,29	187,72
32	2	177,58	176,54	183,72	192,61	177,58	176,54	183,72	192,61	177,58	176,54	183,72	192,61	177,58	176,54	183,72	192,61
33	2	172,83	178,43	184,45	193,04	172,83	178,43	184,45	193,04	172,83	178,43	184,45	193,04	172,83	178,43	184,45	193,04
34	2	181,25	180,96	192,62	193,43	181,25	180,96	192,62	193,43	181,25	180,96	192,62	193,43	181,25	180,96	192,62	193,43
35	2	181,25	180,96	192,62	193,43	181,25	180,96	192,62	193,43	181,25	180,96	192,62	193,43	181,25	180,96	192,62	193,43
36	2	180,03	183,07	186,75	195,28	180,03	183,07	186,75	195,28	180,03	183,07	186,75	195,28	180,03	183,07	186,75	195,28
37	2	185,78	192,53	194,57	199,06	185,78	192,53	194,57	199,06	185,78	192,53	194,57	199,06	185,78	192,53	194,57	199,06
38	2	173,28	204,60	183,72	200,63	173,28	204,60	183,72	200,63	173,28	204,60	183,72	200,63	173,28	204,60	183,72	200,63
39	2	175,48	191,06	183,06	195,48	175,48	191,06	183,06	195,48	175,48	191,06	183,06	195,48	175,48	191,06	183,06	195,48

Annexe B

ESTIMATEURS DES MOYENNES DES GROUPES DANS LE TEMPS POUR LE CAS 1 ET POUR LES QUATRE SCÉNARIOS

TAB. B.0.1. Tableau des estimateurs des moyennes des groupes dans le temps pour le cas 1 et le scénario 1.

	temps1	temps2	temps3	temps4	temps5
groupe1	176.77	178.70	186.05	189.17	183.99
groupe2	177.20	184.40	193.08	196.44	189.61
groupe3	181.08	188.89	190.92	197.05	189.54

TAB. B.0.2. Tableau des estimateurs des moyennes des groupes dans le temps pour le cas 1 et le scénario 2.

	temps1	temps2	temps3	temps4	temps5
groupe1	176.79	178.82	186.10	189.83	184.55
groupe2	177.20	184.37	191.87	197.52	191.19
groupe3	181.07	188.62	190.93	196.21	189.59

TAB. B.0.3. Tableau des estimateurs des moyennes des groupes dans le temps pour le cas 1 et le scénario 3.

	temps1	temps2	temps3	temps4	temps5
groupe1	176.77	180.85	186.21	190.61	185.04
groupe2	177.20	183.54	190.46	197.27	191.26
groupe3	181.08	186.54	191.30	195.54	189.81

TAB. B.0.4. Tableau des estimateurs des moyennes des groupes dans le temps pour le cas 1 et le scénario 4.

	temps1	temps2	temps3	temps4	temps5
groupe1	176.77	180.98	187.09	190.83	185.61
groupe2	177.20	183.47	190.35	195.91	190.44
groupe3	181.08	186.62	190.56	194.71	188.26

Annexe C

TABLEAUX DES EFFETS DES FACTEURS RÉSULTANT DE LA PROCÉDURE GLM DE SAS

TAB. C.0.1. Résultats des tests avec GLM pour le cas de la matrice de covariances de structure symétrie composée (scénario 1) et des valeurs manquantes générées aléatoirement sans distinction des groupes (contexte 1).

Scénario 1, contexte 1 des valeurs manquantes	cas1	pvm=0,05			pvm=0,10			pvm=0,15		
		cas2	cas3	cas4	cas2	cas3	cas4	cas2	cas3	cas4
Valeur-p du test de sphéricité	1,0000	0,8899	0,7000	0,9759	0,9502	0,9545	0,5512	0,6728	0,5563	0,2954
groupe	<,0001	0,0024	0,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001
période	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001
Interaction groupe*période	0,1129	0,1946	0,0313	0,0313	0,4834	0,5508	0,6522	0,0707	0,2464	0,1997
Conclusions		=	≠	≠	=	=	=	=	=	=

TAB. C.0.2. Résultats des tests avec GLM pour le cas de la matrice de covariances de structure symétrie composée (scénario 1) et des valeurs manquantes générées aléatoirement à l'intérieur des groupes 1 et 2 séparément (contexte 2).

Scénario 1, contexte 2 des valeurs manquantes	cas1	pvm=0,05(g1,g2)			pvm=0,10(g1,g2)			pvm=0,15(g1,g2)		
		cas2	cas3	cas4	cas2	cas3	cas4	cas2	cas3	cas4
Valeur-p du test de sphéricité	1,0000	0,9996	0,9930	0,9978	1,0000	0,9293	0,8961	0,9988	0,9708	0,8450
groupe	<,0001	0,0039	0,0001	0,0002	0,0201	0,0003	0,0002	0,0317	0,0019	0,0006
période	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001
Interaction groupe*période	0,1129	0,0360	0,1249	0,0711	0,0288	0,2735	0,1045	0,0777	0,1226	0,1799
Conclusions		≠	=	=	≠	=	=	=	=	=

TAB. C.0.3. Résultats des tests avec GLM pour le cas de la matrice de covariances non structurée (scénario 2) et des valeurs manquantes générées aléatoirement sans distinction des groupes (contexte 1).

Scénario 2, contexte 1 des valeurs manquantes	cas1	pvm=0,05			pvm=0,10			pvm=0,15		
		cas2	cas3	cas4	cas2	cas3	cas4	cas2	cas3	cas4
Valeur-p du test de sphéricité	<,0001	<,0001	<,0001	<,0001	0,0004	<,0001	<,0001	0,0004	<,0001	<,0001
groupe	0,0009	0,0123	0,0010	0,0007	<,0001	0,0008	0,0003	0,0003	0,0004	0,0002
période	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001
Interaction groupe*période	0,0059	0,0227	0,0014	0,0011	0,2052	0,1100	0,0915	0,0355	0,0732	0,0292
Conclusions	=	=	=	=	≠	≠	≠	=	≠	=

TAB. C.0.4. Résultats des tests avec GLM pour le cas de la matrice de covariances non structurée (scénario 2) et des valeurs manquantes générées aléatoirement à l'intérieur des groupes 1 et 2 séparément (contexte 2).

Scénario 2, contexte 2 des valeurs manquantes	cas1	pvm=0,05(g1,g2)			pvm=0,10(g1,g2)			pvm=0,15(g1,g2)		
		cas2	cas3	cas4	cas2	cas3	cas4	cas2	cas3	cas4
Valeur-p du test de sphéricité	<,0001	<,0001	<,0001	<,0001	0,0012	<,0001	0,0039	0,0041	<,0001	0,0032
groupe	0,0009	0,0282	0,0017	0,0018	0,0994	0,0027	0,0022	0,1283	0,0096	0,0040
période	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001
Interaction groupe*période	0,0059	0,0019	0,0218	0,0045	0,0018	0,1138	0,0082	0,0091	0,0727	0,0331
Conclusions		=	=	=	≠	≠	=	≠	≠	=

TAB. C.0.5. Résultats des tests avec GLM pour le cas de la matrice de covariances de structure auto-régressive d'ordre 1 (scénario 3) et des valeurs manquantes générées aléatoirement sans distinction des groupes (contexte 1).

Scénario 3, contexte 1 des valeurs manquantes	cas1	pvm=0,05			pvm=0,10			pvm=0,15		
		cas2	cas3	cas4	cas2	cas3	cas4	cas2	cas3	cas4
Valeur-p du test de sphéricité	<,0001	<,0001	0,0002	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001
groupe	0,0582	0,2184	0,0468	0,0472	0,0038	0,0457	0,0081	0,0402	0,0365	
période	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001
Interaction groupe*période	0,0010	0,0088	0,0001	0,0001	0,0104	0,0091	0,0003	0,0284	0,0057	
Conclusions		=	≠	≠	≠	≠	≠	≠	≠	≠

TAB. C.0.6. Résultats des tests avec GLM pour le cas de la matrice de covariances de structure autorégressive d'ordre 1 (scénario 3) et des valeurs manquantes générées aléatoirement à l'intérieur des groupes 1 et 2 séparément (contexte 2).

Scénario 3, contexte 2 des valeurs manquantes	cas1	pvm=0,05(g1,g2)				pvm=0,10(g1,g2)				pvm=0,15(g1,g2)			
		cas2	cas3	cas4		cas2	cas3	cas4		cas2	cas3	cas4	
Valeur-p du test de sphéricité	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001
groupe	0,0582	0,1905	0,0717	0,0773	0,3619	0,0770	0,0794	0,2994	0,1191	0,1001	0,1001	0,1001	0,1001
période	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001
Interaction groupe*période	0,0010	0,0007	0,0545	0,0157	0,0004	0,2430	0,0380	0,0021	0,2806	0,1444	0,1444	0,1444	0,1444
Conclusions		=	≠	=	=	≠	=	=	≠	=	≠	=	≠

TAB. C.0.7. Résultats des tests avec GLM pour le cas de la matrice de covariances de structure symétrie composée hétérogène (scénario 4) et des valeurs manquantes générées aléatoirement sans distinction des groupes (contexte 1).

Scénario 4, contexte 1 des valeurs manquantes	cas1	pvm=0,05			pvm=0,10			pvm=0,15		
		cas2	cas3	cas4	cas2	cas3	cas4	cas2	cas3	cas4
Valeur-p du test de sphéricité	0,0003	0,0164	0,0866	0,0127	0,0690	0,0376	0,0111	0,1651	0,0356	0,0047
groupe	0,1779	0,3827	0,1537	0,1529	0,0153	0,2189	0,1707	0,0298	0,1917	0,1638
période	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001
Interaction groupe*période	<,0001	0,0009	<,0001	<,0001	0,0005	0,0118	0,0005	0,0001	0,0095	0,0003
Conclusions		=	=	=	≠	=	=	≠	=	=

TAB. C.0.8. Résultats des tests avec GLM pour le cas de la matrice de covariances de structure symétrie composée hétérogène (scénario 4) et des valeurs manquantes générées aléatoirement à l'intérieur des groupes 1 et 2 séparément (contexte 2).

Scénario 4, contexte 2 des valeurs manquantes	cas1	pvm=0,05(g1.g2)			pvm=0,10(g1.g2)			pvm=0,15(g1.g2)		
		cas2	cas3	cas4	cas2	cas3	cas4	cas2	cas3	cas4
Valeur-p du test de sphéricité	0,0003	0,0006	0,0074	0,0063	0,0063	0,0192	0,0012	0,0047	0,0049	0,0090
groupe	0,1779	0,2842	0,1835	0,2004	0,4613	0,2018	0,2058	0,3518	0,2866	0,2504
période	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001	<,0001
Interaction groupe*période	<,0001	<,0001	0,0114	0,0003	<,0001	0,2065	0,0008	0,0003	0,2608	0,0107
Conclusions		=	=	=	=	≠	=	=	≠	=

Annexe D

TABLEAUX DES VALEURS DES CRITÈRES D'INFORMATION AIC ET BIC.

TAB. D.0.1. Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 1 dans le contexte 1 des valeurs manquantes avec $pvm=0,05$.

structure	cas1		cas2		cas3		cas4	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
CS	1747,3	1751,3	1645,8	1649,8	1744,4	1748,4	1731,6	1735,6
UN	1773,3	1803,1	1669,7	1699,6	1761,7	1791,5	1754,3	1784,1
AR(1)	1754,8	1758,7	1655,6	1659,6	1763,3	1767,3	1744,0	1747,9
CSH	1755,3	1767,2	1653,7	1665,6	1751,9	1763,9	1739,6	1751,5

TAB. D.0.2. Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 1 dans le contexte 1 des valeurs manquantes avec $pvm=0,10$.

structure	cas1		cas2		cas3		cas4	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
CS	<u>1747,3</u>	<u>1751,3</u>	<u>1567,6</u>	<u>1571,6</u>	<u>1746,5</u>	<u>1750,5</u>	1751,3	1755,2
UN	1773,3	1803,1	1592,3	1622,1	1766,9	1796,7	1768,1	1798,0
AR(1)	1754,8	1758,7	1571,1	1575,1	1748,6	1752,6	<u>1746,3</u>	<u>1750,3</u>
CSH	1755,3	1767,2	1575,2	1587,1	1752,2	1764,1	1758,3	1770,2

TAB. D.0.3. Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 1 dans le contexte 1 des valeurs manquantes avec $pvm=0,15$.

structure	cas1		cas2		cas3		cas4	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
CS	<u>1747,3</u>	<u>1751,3</u>	<u>1474,5</u>	<u>1478,4</u>	<u>1736,0</u>	<u>1740,0</u>	1732,7	1736,7
UN	1773,3	1803,1	1497,0	1526,9	1752,5	1782,4	1747,2	1777,1
AR(1)	1754,8	1758,7	1476,4	1480,4	1736,7	1740,7	<u>1727,4</u>	<u>1731,3</u>
CSH	1755,3	1767,2	1482,2	1494,1	1742,2	1754,1	1740,1	1752,0

TAB. D.0.4. Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 1 dans le contexte 2 des valeurs manquantes avec $pvm=0,05$.

structure	cas1		cas2		cas3		cas4	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
CS	<u>1747,3</u>	<u>1751,3</u>	<u>1664,0</u>	<u>1668,0</u>	<u>1749,1</u>	<u>1753,1</u>	<u>1746,9</u>	<u>1750,8</u>
UN	1773,3	1803,1	1689,3	1719,2	1773,1	1802,9	1771,2	1801,0
AR(1)	1754,8	1758,7	1670,7	1674,7	1754,4	1758,3	1751,6	1755,6
CSH	1755,3	1767,2	1671,9	1683,9	1756,6	1768,5	1754,8	1766,7

TAB. D.0.5. Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 1 dans le contexte 2 des valeurs manquantes avec $pvm=0,10$.

structure	cas1		cas2		cas3		cas4	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
CS	<u>1747,3</u>	<u>1751,3</u>	<u>1614,8</u>	<u>1618,8</u>	<u>1752,2</u>	<u>1756,2</u>	<u>1750,1</u>	<u>1754,1</u>
UN	1773,3	1803,1	1639,6	1669,5	1774,0	1803,8	1771,3	1801,1
AR(1)	1754,8	1758,7	1621,4	1625,4	1759,1	1763,1	1751,1	1755,1
CSH	1755,3	1767,2	1622,7	1634,7	1759,7	1771,6	1758,0	1769,9

TAB. D.0.6. Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 1 dans le contexte 2 des valeurs manquantes avec $pvm=0,15$.

structure	cas1		cas2		cas3		cas4	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
CS	<u>1747,3</u>	<u>1751,3</u>	<u>1564,4</u>	<u>1568,4</u>	<u>1758,3</u>	<u>1762,3</u>	<u>1753,1</u>	<u>1757,0</u>
UN	1773,3	1803,1	1588,5	1618,3	1780,3	1810,1	1773,6	1803,4
AR(1)	1754,8	1758,7	1571,8	1575,8	1769,7	1773,7	1755,7	1759,7
CSH	1755,3	1767,2	1572,3	1584,3	1765,6	1777,5	1760,8	1772,7

TAB. D.0.7. Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 2 dans le contexte 1 des valeurs manquantes avec $pvm=0,05$.

structure	cas1		cas2		cas3		cas4	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
CS	1637,5	1641,5	1543,7	1547,7	1644,0	1648,0	1628,2	1632,2
UN	<u>1608,0</u>	<u>1637,8</u>	<u>1513,5</u>	<u>1543,4</u>	<u>1617,0</u>	<u>1646,8</u>	<u>1601,9</u>	<u>1631,7</u>
AR(1)	1679,8	1683,8	1586,3	1590,3	1689,7	1693,7	1671,7	1675,7
CSH	1639,9	1651,9	1545,2	1557,2	1646,7	1658,6	1631,7	1643,7

TAB. D.0.8. Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 2 dans le contexte 1 des valeurs manquantes avec $pvm=0,10$.

structure	cas1		cas2		cas3		cas4	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
CS	1637,5	1641,5	1469,1	<u>1473,1</u>	1655,1	<u>1659,1</u>	1647,6	<u>1651,6</u>
UN	<u>1608,0</u>	<u>1637,8</u>	<u>1454,7</u>	1484,6	<u>1632,1</u>	1661,9	<u>1630,2</u>	1660,1
AR(1)	1679,8	1683,8	1500,8	1504,7	1680,0	1684,0	1665,4	1669,3
CSH	1639,9	1651,9	1474,8	1486,7	1655,5	1667,4	1651,5	1663,4

TAB. D.0.9. Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 2 dans le contexte 1 des valeurs manquantes avec $pvm=0,15$.

structure	cas1		cas2		cas3		cas4	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
CS	1637,5	1641,5	1387,9	<u>1391,9</u>	1653,5	1657,4	1642,0	<u>1646,0</u>
UN	<u>1608,0</u>	<u>1637,8</u>	<u>1367,4</u>	1397,2	<u>1626,8</u>	<u>1656,6</u>	<u>1616,2</u>	<u>1646,0</u>
AR(1)	1679,8	1683,8	1415,5	1419,4	1674,5	1678,5	1657,2	1661,1
CSH	1639,9	1651,9	1392,3	1404,2	1651,8	1663,8	1645,3	1657,2

TAB. D.0.10. Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 2 dans le contexte 2 des valeurs manquantes avec $pvm=0,05$.

structure	cas1		cas2		cas3		cas4	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
CS	1637,5	1641,5	1563,3	1567,3	1654,6	<u>1658,6</u>	1642,0	<u>1646,0</u>
UN	<u>1608,0</u>	<u>1637,8</u>	<u>1537,0</u>	<u>1566,8</u>	<u>1629,7</u>	1659,5	<u>1622,1</u>	1651,9
AR(1)	1679,8	1683,8	1602,5	1606,4	1688,0	1692,0	1677,6	1681,6
CSH	1639,9	1651,9	1567,2	1579,1	1656,7	1668,6	1644,6	1656,5

TAB. D.0.11. Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 2 dans le contexte 2 des valeurs manquantes avec $pvm=0,10$.

structure	cas1		cas2		cas3		cas4	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
CS	1637,5	1641,5	1516,0	<u>1520,0</u>	1662,9	<u>1666,8</u>	1646,1	<u>1650,1</u>
UN	<u>1608,0</u>	<u>1637,8</u>	<u>1497,0</u>	1526,8	<u>1642,5</u>	1672,4	<u>1644,8</u>	1674,6
AR(1)	1679,8	1683,8	1548,7	1552,7	1691,1	1695,1	1670,0	1673,9
CSH	1639,9	1651,9	1519,8	1531,8	1662,9	1674,9	1647,9	1659,9

TAB. D.0.12. Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 2 dans le contexte 2 des valeurs manquantes avec $pvm=0,15$.

structure	cas1		cas2		cas3		cas4	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
CS	1637,5	1641,5	1468,9	<u>1472,9</u>	1673,0	<u>1677,0</u>	1650,4	<u>1654,4</u>
UN	<u>1608,0</u>	<u>1637,8</u>	<u>1449,8</u>	1479,6	<u>1651,6</u>	1681,4	<u>1646,8</u>	1676,7
AR(1)	1679,8	1683,8	1501,9	1505,9	1704,9	1708,8	1674,1	1678,0
CSH	1639,9	1651,9	1471,7	1483,6	1672,6	1684,5	1651,8	1663,8

TAB. D.0.13. Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 3 dans le contexte 1 des valeurs manquantes avec $pvm=0,05$.

structure	cas1		cas2		cas3		cas4	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
CS	1493,7	1497,7	1406,6	1410,6	1523,7	1527,6	1492,9	1496,9
UN	1450,6	1480,5	1376,7	1406,5	1509,3	1539,1	1454,1	1484,0
AR(1)	<u>1424,6</u>	<u>1428,6</u>	<u>1351,7</u>	<u>1355,6</u>	<u>1509,1</u>	<u>1513,1</u>	<u>1438,9</u>	<u>1442,9</u>
CSH	1500,9	1512,9	1413,7	1425,7	1526,6	1538,5	1498,8	1510,7

TAB. D.0.14. Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 3 dans le contexte 1 des valeurs manquantes avec $pvm=0,10$.

structure	cas1		cas2		cas3		cas4	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
CS	1493,7	1497,7	1351,9	1355,9	1561,4	1565,4	1531,1	1535,1
UN	1450,6	1480,5	1318,3	1348,1	1513,5	1543,3	1467,4	1497,2
AR(1)	<u>1424,6</u>	<u>1428,6</u>	<u>1294,0</u>	<u>1297,9</u>	<u>1507,7</u>	<u>1511,6</u>	<u>1452,0</u>	<u>1456,0</u>
CSH	1500,9	1512,9	1359,8	1371,7	1561,1	1573,0	1535,9	1547,8

TAB. D.0.15. Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 3 dans le contexte 1 des valeurs manquantes avec $pvm=0,15$.

structure	cas1		cas2		cas3		cas4	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
CS	1493,7	1497,7	1279,9	1283,8	1570,6	1574,6	1538,5	1542,5
UN	1450,6	1480,5	1245,0	1274,8	1526,8	1556,6	1464,6	1494,4
AR(1)	<u>1424,6</u>	<u>1428,6</u>	<u>1222,6</u>	<u>1226,6</u>	<u>1525,2</u>	<u>1529,2</u>	<u>1456,7</u>	<u>1460,7</u>
CSH	1500,9	1512,9	1287,4	1299,3	1567,9	1579,8	1542,1	1554,0

TAB. D.0.16. Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 3 dans le contexte 2 des valeurs manquantes avec $pvm=0,05$.

structure	cas1		cas2		cas3		cas4	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
CS	1493,7	1497,7	1432,7	1436,7	1544,0	1548,0	1517,7	1521,7
UN	1450,6	1480,5	1392,6	1422,4	1503,8	1533,6	1469,8	1499,7
AR(1)	<u>1424,6</u>	<u>1428,6</u>	<u>1367,0</u>	<u>1371,0</u>	<u>1486,9</u>	<u>1490,9</u>	<u>1448,1</u>	<u>1452,1</u>
CSH	1500,9	1512,9	1439,8	1451,8	1549,9	1561,8	1524,3	1536,2

TAB. D.0.17. Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 3 dans le contexte 2 des valeurs manquantes avec $pvm=0,10$.

structure	cas1		cas2		cas3		cas4	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
CS	1493,7	1497,7	1388,4	1392,4	1558,6	1562,6	1528,5	1532,5
UN	1450,6	1480,5	1352,5	1382,3	1516,5	1546,4	1472,2	1502,0
AR(1)	<u>1424,6</u>	<u>1428,6</u>	<u>1327,7</u>	<u>1331,7</u>	<u>1503,8</u>	<u>1507,8</u>	<u>1451,3</u>	<u>1455,3</u>
CSH	1500,9	1512,9	1394,2	1406,1	1563,0	1574,9	1534,7	1546,6

TAB. D.0.18. Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 3 dans le contexte 2 des valeurs manquantes avec $pvm=0,15$.

structure	cas1		cas2		cas3		cas4	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
CS	1493,7	1497,7	1346,4	1350,4	1581,4	1585,3	1536,0	1540,0
UN	1450,6	1480,5	1317,6	1347,4	1554,0	1583,8	1483,2	1513,0
AR(1)	<u>1424,6</u>	<u>1428,6</u>	<u>1292,9</u>	<u>1296,9</u>	<u>1547,7</u>	<u>1551,7</u>	<u>1462,1</u>	<u>1466,1</u>
CSH	1500,9	1512,9	1351,9	1363,8	1583,9	1595,9	1540,5	1552,4

Annexe E

TABLEAUX DES VALEURS DES CRITÈRES D'INFORMATION AIC ET BIC

TAB. E.0.1. Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 4 dans le contexte 1 des valeurs manquantes avec $pvm=0,05$.

structure	cas1		cas2		cas3		cas4	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
CS	1454.3	1458.3	1373.5	1377.5	1492.9	1496.9	1452.0	1456.0
UN	1386.7	1416.6	1306.9	1336.7	1445.2	1475.1	1394.9	1424.7
AR(1)	1476.0	1480.0	1392.6	1396.6	1521.9	1525.9	1471.8	1475.8
CSH	<u>1368.7</u>	<u>1380.7</u>	<u>1291.3</u>	<u>1303.2</u>	<u>1439.1</u>	<u>1451.0</u>	<u>1383.6</u>	<u>1395.5</u>

TAB. E.0.2. Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 4 dans le contexte 1 des valeurs manquantes avec $pvm=0,10$.

structure	cas1		cas2		cas3		cas4	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
CS	1454,3	1458,3	1320,8	1324,8	1531,3	1535,3	1490,4	1494,4
UN	1386,7	1416,6	1264,4	1294,2	<u>1488,2</u>	1518,0	<u>1450,4</u>	1480,3
AR(1)	1476,0	1480,0	1338,3	1342,3	1534,5	1538,5	1487,8	1491,8
CSH	<u>1368,7</u>	<u>1380,7</u>	<u>1247,4</u>	<u>1259,3</u>	1491,1	<u>1503,0</u>	1451,4	<u>1463,3</u>

TAB. E.0.3. Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 4 dans le contexte 1 des valeurs manquantes avec $pvm=0,15$.

structure	cas1		cas2		cas3		cas4	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
CS	1454,3	1458,3	1253,0	1257,0	1542,1	1546,1	1499,9	1503,9
UN	1386,7	1416,6	1192,0	1221,8	1504,6	1534,4	1468,8	1498,7
AR(1)	1476,0	1480,0	1267,2	1271,2	1545,9	1549,9	1488,6	1492,5
CSH	<u>1368,7</u>	<u>1380,7</u>	<u>1177,6</u>	<u>1189,6</u>	<u>1508,6</u>	<u>1520,5</u>	<u>1479,7</u>	<u>1491,6</u>

TAB. E.0.4. Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 4 dans le contexte 2 des valeurs manquantes avec $pvm=0,05$.

structure	cas1		cas2		cas3		cas4	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
CS	1454,3	1458,3	1394,9	1398,8	1509,7	1513,7	1476,4	1480,4
UN	1386,7	1416,6	1332,1	1361,9	1455,2	1485,1	1436,0	1465,8
AR(1)	1476,0	1480,0	1416,4	1420,3	1524,2	1528,2	1490,0	1494,0
CSH	<u>1368,7</u>	<u>1380,7</u>	<u>1314,5</u>	<u>1326,4</u>	<u>1451,4</u>	<u>1463,3</u>	<u>1423,0</u>	<u>1434,9</u>

TAB. E.0.5. Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 4 dans le contexte 2 des valeurs manquantes avec $pvm=0,10$.

structure	cas1		cas2		cas3		cas4	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
CS	1454,3	1458,3	1351,7	1355,7	1524,9	1528,9	1488,8	1492,8
UN	1386,7	1416,6	1299,4	1329,2	1486,4	1516,2	1444,9	1474,7
AR(1)	1476,0	1480,0	1375,7	1379,7	1539,7	1543,7	1490,3	1494,3
CSH	<u>1368,7</u>	<u>1380,7</u>	<u>1282,1</u>	<u>1294,1</u>	<u>1484,9</u>	<u>1496,8</u>	<u>1439,9</u>	<u>1451,8</u>

TAB. E.0.6. Sélection de la meilleure structure de covariances à l'aide des critères AIC et BIC pour le scénario 4 dans le contexte 2 des valeurs manquantes avec $pvm=0,15$.

structure	cas1		cas2		cas3		cas4	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
CS	1454,3	1458,3	1314,6	1318,6	1551,2	1555,2	1494,3	1498,3
UN	1386,7	1416,6	1265,5	1295,3	<u>1518,8</u>	1548,6	1470,7	1500,5
AR(1)	1476,0	1480,0	1336,4	1340,4	1566,3	1570,2	1496,5	1500,4
CSH	<u>1368,7</u>	<u>1380,7</u>	<u>1248,5</u>	<u>1260,4</u>	1524,9	<u>1536,8</u>	<u>1469,6</u>	<u>1481,5</u>

Annexe F

TABLEAUX DES EFFETS DES FACTEURS RÉSULTANT DE LA PROCÉDURE MIXED DE SAS

TAB. F.0.1. Effets de la structure de covariances sur les effets des facteurs pour le scénario 1 dans le contexte 1 des valeurs manquantes avec $pvm=0,05$.

structure	cas1			cas2			cas3			cas4		
	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per
CS	<,0001	<,0001	0,1129	<,0001	<,0001	0,0657	0,0001	<,0001	0,0313	<,0001	<,0001	0,0313
UN	<,0001	<,0001	0,1343	0,0001	<,0001	0,0804	0,0001	<,0001	0,0392	<,0001	<,0001	0,0454
AR(1)	<,0001	<,0001	0,1624	<,0001	<,0001	0,0929	<,0001	<,0001	0,0795	<,0001	<,0001	0,0654
CSH	<,0001	<,0001	0,1129	<,0001	<,0001	0,0699	0,0001	<,0001	0,0326	<,0001	<,0001	0,0330
Conclusions				=			≠			≠		

TAB. F.0.2. Effets de la structure de covariances sur les effets des facteurs pour le scénario 1 dans le contexte 1 des valeurs manquantes avec $pvm=0,10$.

structure	cas1			cas2			cas3			cas4		
	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per
CS	<,0001	<,0001	0,1129	<,0001	<,0001	0,5066	<,0001	<,0001	0,5508	<,0001	<,0001	0,6522
UN	<,0001	<,0001	0,1343	<,0001	<,0001	0,5281	<,0001	<,0001	0,5446	<,0001	<,0001	0,6618
AR(1)	<,0001	<,0001	0,1624	<,0001	<,0001	0,5602	<,0001	<,0001	0,5640	<,0001	<,0001	0,6572
CSH	<,0001	<,0001	0,1129	<,0001	<,0001	0,5060	<,0001	<,0001	0,5341	<,0001	<,0001	0,6456
Conclusions				=			=			=		

TAB. F.0.3. Effets de la structure de covariances sur les effets des facteurs pour le scénario 1 dans le contexte 1 des valeurs manquantes avec $pvm=0,15$.

structure	cas1			cas2			cas3			cas4		
	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per
CS	<,0001	<,0001	0,1129	<,0001	<,0001	0,1711	<,0001	<,0001	0,2464	<,0001	<,0001	0,1997
UN	<,0001	<,0001	0,1343	<,0001	<,0001	0,1715	<,0001	<,0001	0,2518	<,0001	<,0001	0,2191
AR(1)	<,0001	<,0001	0,1624	<,0001	<,0001	0,1551	<,0001	<,0001	0,2274	<,0001	<,0001	0,2012
CSH	<,0001	<,0001	0,1129	<,0001	<,0001	0,1722	<,0001	<,0001	0,2652	<,0001	<,0001	0,2096
Conclusions				=			=			=		

TAB. F.0.4. Effets de la structure de covariances sur les effets des facteurs pour le scénario 1 dans le contexte 2 des valeurs manquantes avec $pvm=0,05$.

structure	cas1			cas2			cas3			cas4		
	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per
CS	<,0001	<,0001	0,1129	0,0001	<,0001	0,0542	0,0001	<,0001	0,1249	0,0002	<,0001	0,0711
UN	<,0001	<,0001	0,1343	0,0001	<,0001	0,0716	0,0001	<,0001	0,1715	0,0002	<,0001	0,1042
AR(1)	<,0001	<,0001	0,1624	<,0001	<,0001	0,0771	<,0001	<,0001	0,1299	<,0001	<,0001	0,0821
CSH	<,0001	<,0001	0,1129	0,0001	<,0001	0,0532	0,0001	<,0001	0,1298	0,0002	<,0001	0,0709
Conclusions				=			=			=		

TAB. F.0.5. Effets de la structure de covariances sur les effets des facteurs pour le scénario 1 dans le contexte 2 des valeurs manquantes avec $pvm=0,10$.

structure	cas1			cas2			cas3			cas4		
	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per
CS	<,0001	<,0001	0,1129	0,0002	<,0001	0,0615	0,0003	<,0001	0,2735	0,0002	<,0001	0,1045
UN	<,0001	<,0001	0,1343	0,0002	<,0001	0,0799	0,0003	<,0001	0,3371	0,0002	<,0001	0,1588
AR(1)	<,0001	<,0001	0,1624	<,0001	<,0001	0,0900	<,0001	<,0001	0,2699	<,0001	<,0001	0,1001
CSH	<,0001	<,0001	0,1129	0,0002	<,0001	0,0588	0,0003	<,0001	0,2786	0,0002	<,0001	0,1044
Conclusions				=			=			=		

TAB. F.0.6. Effets de la structure de covariances sur les effets des facteurs pour le scénario 1 dans le contexte 2 des valeurs manquantes avec $pvm=0,15$.

structure	cas1			cas2			cas3			cas4		
	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per
CS	<,0001	<,0001	0,1129	0,0004	<,0001	0,0535	0,0019	<,0001	0,1226	0,0006	<,0001	0,1799
UN	<,0001	<,0001	0,1343	0,0005	<,0001	0,0434	0,0019	<,0001	0,2008	0,0006	<,0001	0,2256
AR(1)	<,0001	<,0001	0,1624	<,0001	<,0001	0,0862	0,0003	<,0001	0,1209	0,0002	<,0001	0,1426
CSH	<,0001	<,0001	0,1129	0,0004	<,0001	0,0542	0,0020	<,0001	0,1283	0,0006	<,0001	0,1740
Conclusions				=			=			=		

TAB. F.0.7. Effets de la structure de covariances sur les effets des facteurs pour le scénario 2 dans le contexte 1 des valeurs manquantes avec $pvm=0,05$.

structure	cas1			cas2			cas3			cas4		
	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per
CS	0,0009	<,0001	0,0028	0,0009	<,0001	0,0016	0,0010	<,0001	0,0006	0,0007	<,0001	0,0004
UN	<u>0,0009</u>	<u><,0001</u>	<u>0,0003</u>	<u>0,0012</u>	<u><,0001</u>	<u>0,0002</u>	<u>0,0010</u>	<u><,0001</u>	<u>0,0003</u>	<u>0,0007</u>	<u><,0001</u>	<u><,0001</u>
AR(1)	<,0001	<,0001	0,0519	<,0001	<,0001	0,0252	<,0001	<,0001	0,0189	<,0001	<,0001	0,0201
CSH	0,0010	<,0001	0,0058	0,0010	<,0001	0,0036	0,0010	<,0001	0,0012	0,0008	<,0001	0,0010
Conclusions					=			=			=	

TAB. F.0.8. Effets de la structure de covariances sur les effets des facteurs pour le scénario 2 dans le contexte 1 des valeurs manquantes avec $pvm=0,10$.

structure	cas1			cas2			cas3			cas4			
	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per	
CS	0,0009	<,0001	0,0028	0,0002	<,0001	0,0485	0,0008	<,0001	0,0945	0,0003	<,0001	0,0784	
UN	0,0009	<,0001	0,0003	0,0002	<,0001	0,0244	0,0008	<,0001	0,0790	0,0003	<,0001	0,0405	
AR(1)	<,0001	<,0001	0,0519	<,0001	<,0001	0,3249	0,0001	<,0001	0,2553	<,0001	<,0001	0,3552	
CSH	0,0010	<,0001	0,0058	0,0002	<,0001	0,0527	0,0010	<,0001	0,1194	0,0003	<,0001	0,0757	
Conclusions							=				≠		

TAB. F.0.9. Effets de la structure de covariances sur les effets des facteurs pour le scénario 2 dans le contexte 1 des valeurs manquantes avec $pvm=0,15$.

structure	cas1			cas2			cas3			cas4		
	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per
CS	0,0009	<,0001	0,0028	0,0001	<,0001	0,0187	0,0004	<,0001	0,0572	0,0002	<,0001	0,0195
UN	0,0009	<,0001	0,0003	0,0002	<,0001	0,0065	0,0004	<,0001	0,0628	0,0002	<,0001	0,0070
AR(1)	<,0001	<,0001	0,0519	<,0001	<,0001	0,0849	<,0001	<,0001	0,1340	<,0001	<,0001	0,1303
CSH	0,0010	<,0001	0,0058	0,0001	<,0001	0,0157	0,0005	<,0001	0,0736	0,0002	<,0001	0,0191
Conclusions				=			≠			=		

TAB. F.0.10. Effets de la structure de covariances sur les effets des facteurs pour le scénario 2 dans le contexte 2 des valeurs manquantes avec $pvm=0,05$.

structure	cas1			cas2			cas3			cas4		
	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per
CS	0,0009	<,0001	0,0028	0,0015	<,0001	0,0010	0,0017	<,0001	0,0148	0,0018	<,0001	0,0026
UN	0,0009	<,0001	0,0003	0,0014	<,0001	0,0002	0,0017	<,0001	0,0049	0,0018	<,0001	0,0011
AR(1)	<,0001	<,0001	0,0519	0,0001	<,0001	0,0211	0,0002	<,0001	0,0614	0,0002	<,0001	0,0276
CSH	0,0010	<,0001	0,0058	0,0016	<,0001	0,0026	0,0019	<,0001	0,0252	0,0019	<,0001	0,0052
Conclusions				=			=			=		

TAB. F.O.11. Effets de la structure de covariances sur les effets des facteurs pour le scénario 2 dans le contexte 2 des valeurs manquantes avec $pvm=0,10$.

structure	cas1			cas2			cas3			cas4		
	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per
CS	0,0009	<,0001	0,0028	0,0015	<,0001	0,0011	0,0027	<,0001	0,0992	0,0022	<,0001	0,0064
UN	0,0009	<,0001	0,0003	0,0016	<,0001	0,0005	0,0027	<,0001	0,0753	0,0022	<,0001	0,0140
AR(1)	<,0001	<,0001	0,0519	0,0002	<,0001	0,0196	0,0004	<,0001	0,1746	0,0004	<,0001	0,0308
CSH	0,0010	<,0001	0,0058	0,0016	<,0001	0,0028	0,0031	<,0001	0,1233	0,0025	<,0001	0,0098
Conclusions	=						≠			=		

TAB. F.O.12. Effets de la structure de covariances sur les effets des facteurs pour le scénario 2 dans le contexte 2 des valeurs manquantes avec $pvm=0,15$.

structure	cas1			cas2			cas3			cas4		
	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per
CS	0,0009	<,0001	0,0028	0,0022	<,0001	0,0012	0,0096	<,0001	0,0592	0,0040	<,0001	0,0282
UN	0,0009	<,0001	0,0003	0,0028	<,0001	0,0003	0,0096	<,0001	0,0655	0,0040	<,0001	0,0342
AR(1)	<,0001	<,0001	0,0519	0,0003	<,0001	0,0260	0,0019	<,0001	0,0842	0,0010	<,0001	0,0628
CSH	0,0010	<,0001	0,0058	0,0023	<,0001	0,0036	0,0109	<,0001	0,0843	0,0045	<,0001	0,0343
Conclusions	=						≠			=		

Annexe G

SUITE DES TABLEAUX DES EFFETS DES FACTEURS RÉSULTANT DE LA PROCÉDURE MIXED DE SAS

TAB. G.0.1. Effets de la structure de covariances sur les effets des facteurs pour le scénario 3 dans le contexte 1 des valeurs manquantes avec $pvm=0,05$.

structure	cas1			cas2			cas3			cas4		
	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per
CS	0,0582	<,0001	<,0001	0,0519	<,0001	<,0001	0,0468	<,0001	<,0001	0,0472	<,0001	<,0001
UN	0,0582	<,0001	0,0347	0,0527	<,0001	0,0200	0,0468	<,0001	0,0151	0,0472	<,0001	0,0144
AR(1)	0,0582	<,0001	0,0222	0,0526	<,0001	0,0164	0,0331	<,0001	0,0162	0,0410	<,0001	0,0094
CSH	0,0586	<,0001	<,0001	0,0522	<,0001	<,0001	0,0468	<,0001	<,0001	0,0474	<,0001	<,0001
Conclusions	=						≠					

TAB. G.0.2. Effets de la structure de covariances sur les effets des facteurs pour le scénario 3 dans le contexte 1 des valeurs manquantes avec $pvm=0,10$.

structure	cas1			cas2			cas3			cas4		
	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per
CS	0,0582	<,0001	<,0001	0,0358	<,0001	0,0007	0,0556	<,0001	0,0114	0,0457	<,0001	0,0014
UN	0,0582	<,0001	0,0347	0,0345	<,0001	0,0476	0,0556	<,0001	0,1496	0,0457	<,0001	0,1081
AR(1)	<u>0,0582</u>	<u><,0001</u>	<u>0,0222</u>	<u>0,0341</u>	<u><,0001</u>	<u>0,0382</u>	<u>0,0519</u>	<u><,0001</u>	<u>0,0772</u>	<u>0,0469</u>	<u><,0001</u>	<u>0,0694</u>
CSH	0,0586	<,0001	<,0001	0,0361	<,0001	0,0007	0,0567	<,0001	0,0218	0,0464	<,0001	0,0021
Conclusions				≠			≠			≠		

TAB. G.0.3. Effets de la structure de covariances sur les effets des facteurs pour le scénario 3 dans le contexte 1 des valeurs manquantes avec $pvm = 0,15$.

structure	cas1			cas2			cas3			cas4		
	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per
CS	0,0582	<,0001	<,0001	0,0287	<,0001	0,0003	0,0402	<,0001	0,0116	0,0365	<,0001	0,0005
UN	0,0582	<,0001	0,0347	0,0275	<,0001	0,0042	0,0402	<,0001	0,1274	0,0365	<,0001	0,0599
AR(1)	0,0582	<,0001	0,0222	0,0285	<,0001	0,0051	0,0357	<,0001	0,1043	0,0399	<,0001	0,0475
CSH	0,0586	<,0001	<,0001	0,0287	<,0001	0,0002	0,0414	<,0001	0,0210	0,0373	<,0001	0,0009
Conclusions				≠			≠			≠		

TAB. G.0.4. Effets de la structure de covariances sur les effets des facteurs pour le scénario 3 dans le contexte 2 des valeurs manquantes avec $pvm = 0,05$.

structure	cas1			cas2			cas3			cas4		
	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per
CS	0,0582	<,0001	<,0001	0,0692	<,0001	<,0001	0,0717	<,0001	0,0262	0,0773	<,0001	0,0038
UN	0,0582	<,0001	0,0347	0,0666	<,0001	0,0404	0,0717	<,0001	0,2339	0,0773	<,0001	0,1289
AR(1)	0,0582	<,0001	0,0222	0,0678	<,0001	0,0230	0,0721	<,0001	0,3269	0,0816	<,0001	0,0994
CSH	0,0586	<,0001	<,0001	0,0693	<,0001	<,0001	0,0726	<,0001	0,0322	0,0779	<,0001	0,0049
Conclusions	=						≠			≠		

TAB. G.0.5. Effets de la structure de covariances sur les effets des facteurs pour le scénario 3 dans le contexte 2 des valeurs manquantes avec $pvm = 0,10$.

structure	cas1			cas2			cas3			cas4		
	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per
CS	0,0582	<,0001	<,0001	0,0635	<,0001	<,0001	0,0770	<,0001	0,2136	0,0794	<,0001	0,0134
UN	0,0582	<,0001	0,0347	0,0669	<,0001	0,0243	0,0770	<,0001	0,5994	0,0794	<,0001	0,1888
AR(1)	<u>0,0582</u>	<u><,0001</u>	<u>0,0222</u>	<u>0,0703</u>	<u><,0001</u>	<u>0,0139</u>	<u>0,0829</u>	<u><,0001</u>	<u>0,6800</u>	<u>0,0861</u>	<u><,0001</u>	<u>0,1178</u>
CSH	0,0586	<,0001	<,0001	0,0638	<,0001	<,0001	0,0785	<,0001	0,2353	0,0802	<,0001	0,0166
Conclusions	=						≠			≠		

TAB. G.0.6. Effets de la structure de covariances sur les effets des facteurs pour le scénario 3 dans le contexte 2 des valeurs manquantes avec $pvm=0,15$.

structure	cas1			cas2			cas3			cas4		
	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per
CS	0,0582	<,0001	<,0001	0,0607	<,0001	<,0001	0,1191	<,0001	0,2645	0,1001	<,0001	0,1006
UN	0,0582	<,0001	0,0347	0,0639	<,0001	0,0883	0,1191	<,0001	0,3588	0,1001	<,0001	0,2482
AR(1)	0,0582	<,0001	0,0222	0,0705	<,0001	0,0545	0,1181	<,0001	0,3512	0,1097	<,0001	0,2033
CSH	0,0586	<,0001	<,0001	0,0607	<,0001	<,0001	0,1217	<,0001	0,3106	0,1014	<,0001	0,1194
Conclusions				≠			≠			≠		

TAB. G.0.7. Effets de la structure de covariances sur les effets des facteurs pour le scénario 4 dans le contexte 1 des valeurs manquantes avec $pvm=0,05$.

structure	cas1			cas2			cas3			cas4		
	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per
CS	0,1779	<,0001	<,0001	0,1741	<,0001	<,0001	0,1537	<,0001	<,0001	0,1529	<,0001	<,0001
UN	0,1779	<,0001	<,0001	0,1719	<,0001	<,0001	0,1537	<,0001	<,0001	0,1529	<,0001	<,0001
AR(1)	0,1063	<,0001	0,0832	0,1032	<,0001	0,0439	0,0795	<,0001	0,0283	0,0836	<,0001	0,0317
CSH	0,1779	<,0001	<,0001	0,1724	<,0001	<,0001	0,1531	<,0001	<,0001	0,1522	<,0001	<,0001
Conclusions				=			=			=		

TAB. G.0.8. Effets de la structure de covariances sur les effets des facteurs pour le scénario 4 dans le contexte 1 des valeurs manquantes avec $pvm=0,10$.

structure	cas1			cas2			cas3			cas4		
	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per
CS	0,1779	<,0001	<,0001	0,1583	<,0001	0,0006	0,2189	<,0001	0,0109	0,1707	<,0001	0,0004
UN	0,1779	<,0001	<,0001	0,1495	<,0001	<,0001	0,2189	<,0001	0,0401	0,1707	<,0001	0,0041
AR(1)	0,1063	<,0001	0,0832	0,0792	<,0001	0,2009	0,1540	<,0001	0,1169	0,1102	<,0001	0,1254
CSH	0,1779	<,0001	<,0001	0,1508	<,0001	<,0001	0,2196	<,0001	0,0116	0,1694	<,0001	<,0001
Conclusions				=			=			=		

TAB. G.0.9. Effets de la structure de covariances sur les effets des facteurs pour le scénario 4 dans le contexte 1 des valeurs manquantes avec $pvm=0,15$.

structure	cas1			cas2			cas3			cas4		
	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per
CS	0,1779	<,0001	<,0001	0,1626	<,0001	0,0003	0,1917	<,0001	0,0095	0,1638	<,0001	0,0002
UN	0,1779	<,0001	<,0001	0,1374	<,0001	<,0001	0,1917	<,0001	0,0239	0,1638	<,0001	0,0041
AR(1)	0,1063	<,0001	0,0832	0,0835	<,0001	0,0524	0,1319	<,0001	0,0702	0,1120	<,0001	0,0438
CSH	0,1779	<,0001	<,0001	0,1397	<,0001	<,0001	0,1930	<,0001	0,0096	0,1622	<,0001	<,0001
Conclusions				=			=			=		

TAB. G.0.10. Effets de la structure de covariances sur les effets des facteurs pour le scénario 4 dans le contexte 2 des valeurs manquantes avec $pvm=0,05$.

structure	cas1			cas2			cas3			cas4		
	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per
CS	0,1779	<,0001	<,0001	0,1828	<,0001	<,0001	0,1835	<,0001	0,0092	0,2004	<,0001	0,0002
UN	0,1779	<,0001	<,0001	0,1928	<,0001	<,0001	0,1835	<,0001	0,0030	0,2004	<,0001	0,0007
AR(1)	0,1063	<,0001	0,0832	0,1127	<,0001	0,0600	0,1163	<,0001	0,2297	0,1349	<,0001	0,0732
CSH	0,1779	<,0001	<,0001	0,1934	<,0001	<,0001	0,1833	<,0001	0,0044	0,1997	<,0001	<,0001
Conclusions				=			=			=		

TAB. G.0.1.1. Effets de la structure de covariances sur les effets des facteurs pour le scénario 4 dans le contexte 2 des valeurs manquantes avec $pvm=0,10$.

structure	cas1			cas2			cas3			cas4		
	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per
CS	0,1779	<,0001	<,0001	0,1801	<,0001	<,0001	0,2018	<,0001	0,2033	0,2058	<,0001	0,0004
UN	0,1779	<,0001	<,0001	0,1910	<,0001	<,0001	0,2018	<,0001	0,1367	0,2058	<,0001	0,0032
AR(1)	0,1063	<,0001	0,0832	0,1172	<,0001	0,0484	0,1387	<,0001	0,6458	0,1422	<,0001	0,0615
CSH	0,1779	<,0001	<,0001	0,1918	<,0001	<,0001	0,2012	<,0001	0,1803	0,2044	<,0001	0,0001
Conclusions				=			≠			=		

TAB. G.0.12. Effets de la structure de covariances sur les effets des facteurs pour le scénario 4 dans le contexte 2 des valeurs manquantes avec $pvm=0,15$.

structure	cas1			cas2			cas3			cas4		
	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per	groupe	période	gr*per
CS	0,1779	<,0001	<,0001	0,1805	<,0001	<,0001	0,2866	<,0001	0,2576	0,2504	<,0001	0,0090
UN	0,1779	<,0001	<,0001	0,1907	<,0001	<,0001	0,2866	<,0001	0,0629	0,2504	<,0001	0,0148
AR(1)	0,1063	<,0001	0,0832	0,1261	<,0001	0,1375	0,2125	<,0001	0,2423	0,1869	<,0001	0,0749
CSH	0,1779	<,0001	<,0001	0,1925	<,0001	<,0001	0,2882	<,0001	0,1692	0,2496	<,0001	0,0033
Conclusions				=			≠			=		

BIBLIOGRAPHIE

- [1] AKAIKE, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19 :716-723.
- [2] ANDERSON, T.W. (1984). *An Introduction to Multivariate Statistical Analysis*. John Wiley and Sons, New York.
- [3] BERNDT, E.B., HALL, B., HALL, R. AND HAUSMAN, J.A. (1974). Estimation and inference in nonlinear structural models. *Annals of Economic and Social Measurement*, 3 :653-665.
- [4] BOIK, R.J. (1981). A priori tests in repeated measures designs : Effects of nonsphericity. *Psychometrika*, 46(3) :241-255.
- [5] BOX, G.E.P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems. II, Effects of inequality of variance and of correlation between errors in the two-way classification. *Annals of Mathematical Statistics*, 25 :484-498.
- [6] BUCK, S.F. (1960). A method of estimation of missing values in multivariate data suitable for use with electronic computer. *Journal of the Royal Statistical Society*, 22 :302-306.
- [7] COSSETTE, M. (2002). Analyse de Variance à Mesures Répétées : Aspect Informatique. Rapport de stage. Département de mathématiques et statistique. Université de Montréal.
- [8] CROWDER, M.J. AND HAND, D.J. (1990). *Analysis of Repeated Measures*. Chapman and Hall, London.

- [9] DEMPSTER, A.P., LAIRD, N.M. AND RUBIN, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 39 :1-38.
- [10] DIGGLE, P.J., LIANG, K.Y. AND ZEGER, S.L. (1994). *Analysis of Longitudinal Data*. Oxford University Press, Oxford.
- [11] GEISSER, S. AND GREENHOUSE, S.W. (1958). An extension of Box's results on the use of the F distribution in multivariate analysis. *Annals of Mathematical Statistics*, 29 :885-891.
- [12] GREENHOUSE, S.W. AND GEISSER, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24 :95-112.
- [13] HAND, D.J. AND TAYLOR, C.C. (1987). *Multivariate Analysis of Variance and Repeated Measures*. Chapman and Hall, London.
- [14] HARVILLE, D.A. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika*, 61 :383-385.
- [15] HENDERSON, C.R. (1984). Applications of Linear Models in Animal Breeding. *University of Guelph Press*, Guelph.
- [16] HUYNH, H. ET FELDT L. (1976). Estimation of the Box Correction for Degrees of Freedom from Sample Data in the Randomized Block and Split-Plot Designs. *Journal of Educational Statistics*, 1 :69-82.
- [17] JESKE, D.R. ET HARVILLE, D.A. (1988). Prediction-Interval Procedures and (Fixed-Random) Confidence Interval Procedures for Mixed Linear Models. *Communications in Statistics*, 17 :1053-1087.
- [18] LAIRD, N.M. AND WARE, J.H. (1982). Random-effects models for longitudinal Data. *Biometrics*, 38 :963-974.
- [19] LINDSTROM, M.J. AND BATES, D.M. (1988). Newton-Raphson and EM algorithms for linear mixed-effects models for repeated measures data. *Journal of the American Statistical Association*, 83 :1014-1022.
- [20] LIANG, K.Y. AND ZEGER, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73 :13-22.

- [21] LITTELL, R.C., MILLIKEN, G.A., STROUP, W.W. AND WOLFINGER, R.D. (1996). *SAS System for Mixed Models*. SAS Institut Inc, Cary.
- [22] LITTLE, R.J.A. AND RUBIN, D.B. (1987). *Statistical Analysis with Missing Data*. John Wiley and Sons, New York.
- [23] MCLEAN, R.A. ET SANDERS, W.L. (1988). Approximation Degrees of Freedom for Standard Errors in Mixed Linear Models. *in Proceedings of the Statistical Computing Section. American Statistical Association*, 50-59.
- [24] MAUCHLY, J. W. (1940). Significance test for sphericity of a normal n -variate distribution. *Annals of Mathematical Statistics*, 29 :204-209.
- [25] MUIRHEAD, R.J. (1982) *Aspects of Multivariate Statistical Theory*. John Wiley and Sons, New York.
- [26] NETER, J., KUTNER, M.H. NACHTSHEIM, C.J. AND WASSERMAN, W. (1996). *Applied linear statistical models*. McGraw-Hill, New York.
- [27] RUBIN, D.B. (1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley and Sons, New York.
- [28] SCHAFER, J.L. (1997). *Analysis of Incomplete Multivariate Data*. Chapman and Hall, London.
- [29] SCHWARZ, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6 :461-464.
- [30] SEARLE, S.R. (1971). *Linear Models*. John Wiley and Sons, New York.
- [31] VERBEKE G. AND MOLENBERGHS G. (1997). *Linear Mixed Models in Practice*. Springer-Verlag, New York.
- [32] WARE, J.H. (1985). Linear models for the analysis of longitudinal studies. *The American Statistician*, 39 :95-101.
- [33] WOLFINGER, R.D. (1993). Covariance structure selection in general mixed models. *Communications in statistics*, 22 :1079-1106.
- [34] WOLFINGER, R.D., TOBIAS, R.D. AND SALL, J. (1994). Computing gaussian likelihoods and their derivatives for general mixed models. *SIAM Journal on Scientific Computing*, 15 :1294-1310.