

Modèles pour données répétées

Résumé

*Les données répétées, ou données longitudinales, constituent un domaine à la fois important et assez particulier de la statistique. On entend par données répétées des données telles que, pour chaque individu considéré, on dispose d'observations à différents instants, autrement dit répétées dans le temps. Les principaux domaines d'application de ce type de données sont la médecine et la biologie, lors d'expérimentations humaines ou animales. La difficulté majeure dans le traitement statistique de ces données provient de ce qu'il n'est en général pas réaliste de supposer que les observations réalisées sur un même individu, au cours du temps, sont indépendantes. Il est donc nécessaire d'introduire une structure de covariance pour les variables aléatoires associées à chaque individu, afin de tenir compte de cette situation particulière. On va ainsi retrouver des modèles proches de ceux vus au chapitre 5 avec l'analyse de variance multidimensionnelle. Par ailleurs, il est fréquent, dans les modèles pour données répétées, de considérer, en plus des facteurs à effets fixes que l'on souhaite étudier dans le modèle, des effets aléatoires associés aux individus. On aura pour cela recours à des modèles mixtes, tels que nous les avons introduits au chapitre 6. Enfin, on notera que, dans le cadre de la modélisation des données répétées, le terme statistique d'individu est souvent remplacé par celui de **sujet**. Nous utiliserons indifféremment l'un ou l'autre par la suite.*

Retour au [plan du cours](#)

1 Introduction

Pour la bibliographie, nous conseillons les ouvrages de Brown & Prescott (1999), de Davis (2002) et de Verbeke & Molenberghs (2000).

Ce chapitre est donc consacré à la modélisation de ce que l'on appelle les *données répétées*, ou encore les *données longitudinales*. Sauf cas particulier, il s'agit de données répétées au cours du temps, sur les mêmes individus.

De manière générale, on peut penser que les observations faites à différents instants, sur un individu donné, sont corrélées. D'où la nécessité d'introduire une "structure de covariance" pour ces observations dans les modèles pour données répétées (autrement dit, de considérer une matrice de variances-covariances \mathbf{R} , de dimension $T \times T$, si T est le nombre d'instant d'observation). En ce sens, ces modèles ressemblent aux modèles de MANOVA vus au chapitre 5, puisque la réponse de chaque individu est multidimensionnelle. Mais, au lieu de correspondre à différentes composantes observées à un instant donné, cette réponse correspond à la même variable observée à différents instants. De plus, on va maintenant considérer le cadre général des modèles linéaires gaussiens mixtes, introduits au chapitre 6, pour pouvoir expliquer l'évolution au cours du temps de la variable réponse à la fois par des facteurs à effets fixes et par des facteurs à effets aléatoires, notamment des facteurs individuels. De façon naturelle, le temps lui-même interviendra comme facteur à effets fixes.

Les modèles pour données répétées sont donc assez complexes et d'un maniement plutôt délicat. Dans leur forme la plus générale (qui ne sera abordée qu'au paragraphe 7.6), ils nécessitent en effet l'estimation des effets fixes, celle des composantes de la variance et celle des éléments de la matrice \mathbf{R} . Néanmoins, certaines structures particulières pour la matrice \mathbf{R} pourront être envisagées, réduisant ainsi le nombre de paramètres à estimer. Parallèlement, il faudra tester la significativité des différents effets initialement considérés, dans le but de simplifier au maximum le modèle finalement retenu. Enfin, si plusieurs modèles sont envisageables, il faudra procéder à un choix de modèle en utilisant des indicateurs du type AIC ou BIC.

On notera, pour terminer cette présentation générale, que la procédure de SAS la plus usuelle pour traiter les modèles pour données répétées est la procédure MIXED, mais que la procédure GLM peut aussi être utilisée de manière complémentaire.

Nous noterons toujours Y la variable aléatoire réelle réponse que l'on souhaite modéliser. Elle est observée sur différents individus et à différents instants. L'objet de ce chapitre est de définir des modèles prenant en compte d'une part un ou plusieurs facteurs susceptibles d'avoir un effet sur les valeurs de Y , d'autre part le temps. Jusqu'au paragraphe 7.6, nous ne considérerons qu'un seul facteur à effets fixes autre que le temps (la généralisation à plusieurs fac-

teurs de même nature ne pose pas de problème particulier). Ce facteur sera noté F (il pourra s'agir, par exemple, d'un traitement médical), le nombre de ses niveaux sera noté J , ces derniers étant indicés par j .

Au chapitre 3, pour une analyse de variance (ANOVA) à un seul facteur F , nous avons écrit le modèle sous la forme suivante :

$$Y_{ij} = \mu + \alpha_j + U_{ij} ; \quad j = 1, \dots, J ; \quad i = 1, \dots, n_j .$$

Maintenant, chaque mesure Y_{ij} est répétée à différents instants notés t ($t = 1, \dots, T$) et les v.a.r. correspondantes sont notées Y_{ijt} , de sorte que le modèle est réécrit sous la forme suivante :

$$Y_{ijt} = \mu + \alpha_j^1 + \alpha_t^2 + \gamma_{jt} + U_{ijt} .$$

On pourrait envisager de traiter ce modèle comme un modèle d'ANOVA à deux facteurs croisés (F et le temps), mais cela reviendrait à supposer que les observations réalisées sur un même individu au cours du temps sont indépendantes, ce qui n'est pas réaliste (sauf cas très particulier). Il est donc nécessaire d'introduire ici des modèles spécifiques pour ce type de données, ces modèles devant prendre en compte une structure de covariance entre observations réalisées sur un même individu à différents instants.

Remarque. — Dans tout ce chapitre, nous supposons d'une part que les instants d'observation sont les mêmes pour tous les individus, d'autre part que tous les individus sont observés à chaque instant considéré (autrement dit, il n'y a pas de données manquantes). Malheureusement, cette situation, relativement commode en ce qui concerne la modélisation, n'est pas la plus courante dans la pratique, ce qui complique d'autant plus les choses.

2 Analyses préliminaires

Deux types d'analyses préliminaires sont possibles avec les données répétées. Elles n'abordent pas réellement le problème de la modélisation de ces dernières, mais peuvent apporter des compléments intéressants à cette modélisation. Notons que ces deux types d'analyses sont réalisées de façon systématique avec la procédure GLM de SAS.

2.1 ANOVA réalisée à chaque instant t

Dans une première approche, on peut réaliser l'ANOVA de Y en fonction de F à chaque instant d'observation t . Cela donne une première idée de l'influence de F sur Y , mais ne modélise pas l'évolution de Y au cours du temps. C'est donc largement insuffisant.

On notera que ces analyses sont les premières sorties fournies par la procédure GLM de SAS avec des données répétées, ce qui peut être trompeur (en laissant supposer que cela relève de la modélisation du temps, ce qui n'est pas le cas).

Toutefois, de même qu'il est conseillé de faire une étude univariée de chaque variable quantitative considérée dans une Analyse en Composantes Principales avant de réaliser cette dernière, de même les T ANOVA dont il est ici question sont utiles pour bien maîtriser les données considérées.

2.2 ANOVA réalisée sur la moyenne temporelle des observations

Il s'agit de l'ANOVA de Y en fonction de F réalisée sur les moyennes temporelles

$$\bar{y}_{ij\bullet} = \frac{1}{T} \sum_{t=1}^T y_{ijt} .$$

Bien entendu, cette analyse ne prend nullement en compte le temps, puisqu'on fait au préalable la moyenne des différentes valeurs obtenues aux différents instants d'observation. En fait, on teste ainsi l'influence marginale (indépendamment du temps) du facteur F . Comme précédemment, cela apporte un complément intéressant à la modélisation des données répétées, même si cela n'en fait pas partie.

On notera que, bizarrement, dans la procédure GLM de SAS, cette analyse marginale est faite en utilisant \sqrt{T} au lieu de T au dénominateur de l'expression ci-dessus. Toutefois, cela ne change strictement rien au résultat du test de significativité marginale du facteur F (test de Fisher usuel).

3 Modèle à un facteur à effets fixes pour données répétées

Contrairement aux analyses décrites en 7.2, le modèle présenté ici est spécifique aux données répétées. Il convient lorsqu'on ne souhaite pas prendre en compte des facteurs à effets aléatoires pour modéliser les données. La généralisation à plus d'un facteur à effets fixes autre que le temps ne pose pas de problème particulier.

3.1 Principe

Le principe général de ce modèle est le suivant : pour un niveau j du facteur F ($j = 1, \dots, J$), et pour un individu i pris en compte à ce niveau là ($i = 1, \dots, n_j$), on considère le vecteur aléatoire $Y_{ij} = (Y_{ij1}, \dots, Y_{ijT})'$ de \mathbb{R}^T et l'on pose :

$$Y_{ij} \sim \mathcal{N}_T(\mu_j, \mathbf{R}).$$

On notera que, contrairement à ce qui a été fait au chapitre 5, le vecteur aléatoire Y_{ij} est ici considéré comme un vecteur colonne. Dans l'écriture ci-dessus :

- le vecteur μ_j désigne le vecteur moyenne de la loi gaussienne au niveau j de F ; les composantes de ce vecteur sont liées au temps :

$$\mu_j = \begin{pmatrix} \mu_{j1} \\ \vdots \\ \mu_{jT} \end{pmatrix};$$

par ailleurs, on pose : $\mu_{jt} = \mu + \alpha_j^1 + \alpha_t^2 + \gamma_{jt}$ (μ est l'effet général, α_j^1 est l'effet principal du niveau j du facteur F , α_t^2 est l'effet principal de l'instant t et γ_{jt} est l'effet d'interaction entre le niveau j et l'instant t ; les paramètres α_j^1 et α_t^2 sont centrés et les paramètres γ_{jt} sont doublement centrés) ;

- la matrice \mathbf{R} est $T \times T$, symétrique et strictement définie positive ; elle représente la structure de covariance des observations répétées sur un même individu et doit être estimée ; nous verrons, au prochain paragraphe, les principales structures de covariance usuelles, la plupart permettant de diminuer sensiblement le nombre de paramètres à estimer. On notera ici

que la matrice \mathbf{R} ne dépend pas de j , autrement dit que le modèle considéré est encore homoscédastique ; il est certes possible de considérer des modèles hétéroscédastiques en présence de données répétées, mais cela entraîne encore une augmentation du nombre de paramètres à estimer et peut devenir très problématique.

Le modèle considéré s'écrit donc :

$$Y_{ijt} = \mu_{jt} + U_{ijt},$$

le paramètre de moyenne μ_{jt} s'écrivant comme indiqué plus haut et la v.a.r. erreur U_{ijt} étant la t -ième composante du vecteur aléatoire gaussien centré U_{ij} de \mathbb{R}^T , de matrice de variances-covariances \mathbf{R} : les composantes de ce vecteur aléatoire ne sont donc pas indépendantes (sauf si \mathbf{R} est diagonale, ce qui n'est en général pas le cas).

Dans la pratique, on commence par tester les effets d'interactions temps \times facteur puis, s'ils ne sont pas significatifs, on teste successivement l'effet du temps et l'effet du facteur (ce dernier point a été précisé en 7.2.2). Des indications sur les tests liés au temps sont données dans le point suivant. Dans le modèle retenu, on estime les différents paramètres, dont les éléments de la matrice \mathbf{R} .

3.2 Terminologie

Nous précisons ici quelques points de terminologie utiles dans l'utilisation des logiciels statistiques tels que SAS. On notera que les individus (ou unités statistiques) sont souvent appelés "sujets" dans le contexte des données répétées.

- Les effets principaux du facteur F (les α_j^1), non liés au temps, sont appelés les effets entre les sujets, ou inter-sujets (*between subjects effects*). On a vu en 7.2.2 comment les tester.
- Les effets liés au temps, à savoir les effets principaux du temps (les α_t^2) et les effets d'interactions temps \times facteur (les γ_{jt}), sont appelés les effets intra-sujets (*within subjects effects*). On les teste avec les tests multidimensionnels introduits en MANOVA (voir le chapitre 5), en privilégiant toujours le test de Wilks. Ces tests sont définis en dimension $D = T - 1$ et portent sur des différences du type : $Y_{ij2} - Y_{ij1}, Y_{ij3} - Y_{ij1}, \dots, Y_{ijT} - Y_{ij1}$.

- Finalement, dans le modèle multidimensionnel complet, on distingue :
 - 1 effet général, μ ;
 - J effets principaux du facteur F , les α_j^1 (avec une contrainte de centrage) ;
 - T effets principaux du temps, les α_t^2 (également avec une contrainte de centrage) ;
 - $J \times T$ effets d'interactions, les γ_{jt} (avec $J + T - 1$ contraintes pour le double centrage).

Les deux premiers types d'effets sont des effets inter-sujets (indépendants du temps), tandis que les deux derniers types sont des effets intra-sujets (ils dépendent du temps).

En comptant les éléments de la matrice \mathbf{R} , cela fait au total $(J \times T) + \frac{T(T+1)}{2}$ paramètres indépendants à estimer dans le modèle complet, avec $n \times T$ observations ($n = \sum_{j=1}^J n_j$). On veillera donc à disposer d'un nombre total de sujets n vérifiant : $n > J + \frac{T+1}{2}$.

3.3 Mise en œuvre

Le logiciel statistique SAS dispose de deux procédures permettant de mettre en œuvre les modèles pour données répétées : la procédure GLM et la procédure MIXED. Nous renvoyons aux paragraphes 7.6 et 7.7 pour des précisions et des illustrations.

On notera que la procédure MIXED, plus récente, est mieux adaptée à ces modèles. Toutefois, la procédure GLM présente plusieurs avantages qui rendent son utilisation intéressante, en complément de MIXED : tout d'abord, elle réalise les différents tests multidimensionnels relatifs aux effets liés au temps¹, ce que ne fait pas MIXED ; ensuite, elle met en œuvre le test de sphéricité de Mauchly (voir le point 7.5.3), indiquant si la structure *compound symmetry* est acceptable ou non.

1. Pour des précisions sur ces tests, on se reportera à l'Annexe E.

4 Les structures usuelles de covariance pour R

La procédure MIXED de SAS propose différentes structures classiques pour la matrice \mathbf{R} des covariances entre v.a.r. associées aux observations répétées sur un même individu. La structure choisie se déclare dans la commande `repeated` de cette procédure, au sein de l'option `type=...`. Par défaut, MIXED utilise l'option *simple*, correspondant à la structure d'indépendance (voir plus loin).

Nous précisons ci-dessous les principales options possibles pour la structure de \mathbf{R} . Dans les cas nécessitant une illustration, nous la donnons en utilisant le cas particulier $T = 4$ (on notera qu'on ne parle de données répétées que pour $T \geq 2$ et que le choix de structure n'a vraiment d'intérêt que pour $T \geq 3$). La liste donnée ci-dessous n'est pas exhaustive et l'on pourra trouver d'autres possibilités pour \mathbf{R} en consultant l'aide en ligne de SAS.

Pour un modèle donné, le choix de la structure de covariance peut se faire en utilisant un critère usuel de choix de modèle (AIC ou BIC, voir l'Annexe D).

Absence de structure

C'est le cas le plus général : la matrice \mathbf{R} n'a aucune structure particulière dans ce cas (elle est simplement symétrique et définie positive) et tous ses éléments doivent être estimés, soit $\frac{T(T+1)}{2}$. C'est, bien entendu, le cas le plus coûteux en nombre de paramètres à estimer. L'option correspondant à cette structure est appelée *unstructured* dans MIXED et se déclare avec l'option `type=un`.

Structure d'indépendance

À l'opposé de la précédente, cette structure est la plus simple qui soit. Elle consiste à poser $\mathbf{R} = \sigma^2 \mathbf{I}_T$, où σ^2 est le seul paramètre (supposé strictement positif) à estimer et où \mathbf{I}_T est la matrice identité d'ordre T . Cette structure suppose donc que les observations réalisées sur tout individu aux différents instants sont indépendantes, ce qui n'est, en général, pas réaliste et n'a donc pas de réel intérêt dans la pratique (on notera que cela revient à faire une ANOVA

dans laquelle le temps est un facteur ordinaire). Sauf cas très particulier, cette structure n'est donc pas utilisée dans les modèles pour données répétées. L'option correspondante s'appelle *simple* dans la procédure MIXED de SAS et se déclare avec l'option `type=simple`.

Attention : c'est l'option par défaut.

Structure symétrie composée, ou "compound symmetry"

Même si elle n'est souvent pas très réaliste, on rencontre fréquemment cette structure car elle présente différentes propriétés intéressantes. Elle consiste à poser $\mathbf{R} = \sigma_1^2 \mathbf{I}_T + \sigma_2^2 \mathbb{K}_{T \times T}$, où σ_1^2 et σ_2^2 sont deux paramètres strictement positifs à estimer, $\mathbb{K}_{T \times T}$ désignant la matrice carrée d'ordre T dont tous les éléments valent 1. Ainsi, la variance de la v.a.r. associée à toute observation vaut $\sigma_1^2 + \sigma_2^2$, la covariance des v.a.r. associées à deux observations réalisées sur le même individu à deux instants différents valant constamment σ_2^2 . L'option correspondante s'appelle *compound symmetry* (symétrie composée) et se déclare avec l'option `type=cs`.

Outre la simplicité d'une telle structure de covariance, nous en verrons, au paragraphe 7.5, d'autres propriétés.

Structure auto-régressive d'ordre 1

L'intérêt de cette structure est de ne nécessiter que l'estimation de deux paramètres (comme la précédente), tout en étant telle que la corrélation des v.a.r. associées à deux observations réalisées sur le même individu à deux instants différents soit inversement proportionnelle à l'écart entre ces deux instants. Pour $T = 4$, la matrice \mathbf{R} s'écrit

$$\mathbf{R} = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{pmatrix},$$

avec : $\sigma^2 > 0$; $0 < \rho < 1$. Ainsi, la variance de la v.a.r. associée à toute observation vaut σ^2 , la corrélation des v.a.r. associées à deux observations réalisées sur le même individu à deux instants différents i et i' valant $\sigma^2 \rho^{|i-i'|}$. L'option correspondante s'appelle auto-régressive d'ordre 1 et se déclare avec

l'option `type=ar(1)`. Il s'agit d'une structure très courante en pratique.

Structure de Toeplitz à deux bandes

Il n'y a encore que deux paramètres à estimer dans cette structure, un peu moins générale que la précédente. Pour $T = 4$, la matrice \mathbf{R} s'écrit

$$\mathbf{R} = \begin{pmatrix} \sigma_1^2 & \sigma_2 & 0 & 0 \\ \sigma_2 & \sigma_1^2 & \sigma_2 & 0 \\ 0 & \sigma_2 & \sigma_1^2 & \sigma_2 \\ 0 & 0 & \sigma_2 & \sigma_1^2 \end{pmatrix},$$

avec $\sigma_1^2 > 0$. Le paramètre σ_2 peut être négatif, ce qui permet de traiter des cas un peu particuliers. La structure correspondante se déclare avec l'option `type=toep(2)`.

Structure de Toeplitz générale

Moins simple que les précédentes, cette structure nécessite l'estimation de T paramètres. La matrice \mathbf{R} est dans ce cas une matrice dite de Toeplitz (quelconque). Elle s'écrit, toujours pour $T = 4$, sous la forme

$$\mathbf{R} = \begin{pmatrix} \sigma_1^2 & \sigma_2 & \sigma_3 & \sigma_4 \\ \sigma_2 & \sigma_1^2 & \sigma_2 & \sigma_3 \\ \sigma_3 & \sigma_2 & \sigma_1^2 & \sigma_2 \\ \sigma_4 & \sigma_3 & \sigma_2 & \sigma_1^2 \end{pmatrix},$$

avec $\sigma_1^2 > 0$. Les paramètres σ_2 , σ_3 et σ_4 peuvent encore être négatifs. La structure correspondante se déclare avec l'option `type=toep`.

Structure spatiale

La procédure MIXED de SAS propose différentes formes de structures spatiales (on pourra se reporter à la documentation de SAS pour plus de détails). La plus "naturelle" est une généralisation de la structure auto-régressive d'ordre 1. Elle consiste à poser (toujours pour $T = 4$) :

$$\mathbf{R} = \sigma^2 \begin{pmatrix} 1 & \rho^{t_2-t_1} & \rho^{t_3-t_1} & \rho^{t_4-t_1} \\ \rho^{t_2-t_1} & 1 & \rho^{t_3-t_2} & \rho^{t_4-t_2} \\ \rho^{t_3-t_1} & \rho^{t_3-t_2} & 1 & \rho^{t_4-t_3} \\ \rho^{t_4-t_1} & \rho^{t_4-t_2} & \rho^{t_4-t_3} & 1 \end{pmatrix},$$

avec : $\sigma^2 > 0$; $0 < \rho < 1$; t_1, t_2, t_3, t_4 désignent les instants d'observation. Pour des instants régulièrement espacés ($t = 1, \dots, T$), cette structure est identique à la structure auto-régressive d'ordre 1. Mais, dans le cas d'observations irrégulières (cas assez fréquent dans la pratique), ce n'est plus le cas et cette structure spatiale est alors commode. Elle se déclare avec l'option `type=sp (pow) (temps)`, si `temps` est le nom de la variable contenant les instants d'observation. On notera que `pow` signifie `power` et correspond à la fonction puissance utilisée ici. Les autres structures spatiales proposées par `MIXED` pour la matrice \mathbf{R} correspondent à d'autres fonctions que l'on doit déclarer à cet endroit là.

Laquelle choisir ?

Supposons un modèle choisi (modèle complet, modèle additif...) au moyen des tests indiqués au paragraphe 7.3 (par exemple, en ne spécifiant pas au départ de structure sur la matrice \mathbf{R}). Se pose alors le problème du choix de la structure de covariance "optimale" pour ce modèle. La façon usuelle de procéder consiste à regarder chacune des structures présentées ci-dessus et à choisir ensuite celle qui minimise un critère de choix de modèle : AIC ou BIC (lorsqu'il y a contradiction entre ces critères, les choses deviennent délicates...). On notera que les expressions des matrices \mathbf{H} et \mathbf{E} intervenant dans les tests multivariés introduits au chapitre 5 et utilisés ici ne font pas intervenir la matrice \mathbf{R} ; le choix de cette dernière n'a donc pas d'influence sur le résultat de ces tests, d'où l'intérêt de procéder comme indiqué (à condition, toutefois, que le nombre d'observations $n \times T$ soit suffisant).

5 Cas particulier : la structure "compound symmetry"

5.1 Propriété préliminaire

On a vu, au paragraphe précédent, que la structure de covariance dite *compound symmetry* consiste à poser : $\mathbf{R} = \sigma_1^2 \mathbf{I}_T + \sigma_2^2 \mathbf{J}_T$. D'un point de vue algébrique, on peut montrer que, dans \mathbb{R}^T , la matrice \mathbf{R} définie ci-dessus n'admet que deux valeurs propres : $\lambda_1 = \sigma_1^2$; $\lambda_2 = \sigma_1^2 + T\sigma_2^2$. La première de ces valeurs propres est d'ordre $T - 1$; la seconde est d'ordre 1. Par conséquent, il existe une matrice \mathbf{C} , de dimension $(T - 1) \times T$ et de rang $T - 1$, telle que,

en posant $Y_{ij}^* = \mathbf{C}Y_{ij}$, il vient

$$Y_{ij}^* \sim \mathcal{N}_{T-1}(\mathbf{C}\mu_j, \mathbf{C}\mathbf{R}\mathbf{C}'),$$

avec :

$$\mathbf{C}\mathbf{R}\mathbf{C}' = \sigma_1^2 \mathbf{I}_{T-1}.$$

On se reportera à l'Annexe F pour les justifications de ce qui précède.

5.2 Conséquences

En supposant que la structure *compound symmetry* soit appropriée aux données considérées, si l'on projette chaque vecteur aléatoire Y_{ij} sur le sous-espace propre associé à la valeur propre λ_1 définie ci-dessus, on obtient, pour le vecteur projeté, une structure de corrélation proportionnelle à la matrice identité, autrement dit on se ramène au cas standard où l'on peut considérer comme indépendantes les différentes observations d'un même individu réalisées aux différents instants. Les tests relatifs aux différents effets (facteur, temps et interactions) sont alors des tests de Fisher usuels, puisqu'on s'est ramené au cas standard unidimensionnel. La taille de l'échantillon correspondant est $N = n(T - 1)$, ce qu'on peut obtenir en travaillant sur des différences, par exemple entre chaque instant et l'instant initial.

Enfin, en posant $U_{ijt} = A_{ij} + U_{ijt}^*$ dans le modèle défini au paragraphe 7.3, avec $A_{ij} \sim \mathcal{N}(0, \sigma_2^2)$, on montre, de façon immédiate, que le modèle se ramène à un modèle mixte dont les effets fixes sont ceux du temps, du facteur et des interactions, et les effets aléatoires (les A_{ij}) sont ceux attachés aux individus observés. Cette remarque permet d'estimer les composantes de la variance de ce modèle mixte (σ_1^2 et σ_2^2) en utilisant, par exemple, la procédure `VARCOMP` de SAS. On a ainsi contourné toutes les difficultés provenant de la nature répétée des données...

5.3 Le test de sphéricité de Mauchly

Les commodités évoquées ci-dessus montrent que la structure de covariance *compound symmetry* présente des avantages certains lorsqu'on doit modéliser des données répétées (et cela était carrément vital lorsqu'on ne disposait pas de logiciel performant comme aujourd'hui). Encore faut-il s'assurer que cette structure soit valide. Un test permet de contrôler l'adéquation de cette structure aux données analysées : c'est le test dit de sphéricité de Mauchly.

Il consiste à estimer la matrice \mathbf{R} par $\hat{\mathbf{R}}$, sans supposer de structure (cas *unstructured*), puis à considérer une matrice \mathbf{C} , de dimension $(T - 1) \times T$, dont les lignes soient orthonormées (au sens de la métrique euclidienne classique de \mathbb{R}^T). Le test consiste alors à contrôler la sphéricité de la matrice $\mathbf{C}\hat{\mathbf{R}}\mathbf{C}'$, autrement dit sa proportionnalité avec la matrice identité d'ordre $T - 1$. Il s'agit d'un test asymptotique de type khi-deux dont le degré de liberté est $\frac{T(T - 1)}{2} - 1$.

La *p-value* de ce test est fournie par la procédure `GLM` de SAS, lorsqu'on a mis l'option `printe` dans la commande `repeated`, à la rubrique *Sphericity Tests* et à la ligne *Orthogonal Components* (cette ligne correspond à la transformation réalisée au moyen de la matrice \mathbf{C}).

Remarque. — On trouvera une présentation détaillée du test de sphéricité de Mauchly dans la section 7.2 de l'ouvrage de Rencher (1995).

Remarque. — On notera que la procédure `GLM` de SAS propose, lorsque le test de sphéricité a rejeté la structure de symétrie composée, deux modifications possibles du test de Fisher unidimensionnel, pour pouvoir continuer à l'appliquer (Greenhouse-Geisser et Huynh-Feldt). Nous déconseillons l'usage de ces tests. Il faut signaler qu'on était réduit à les utiliser lorsqu'on ne disposait ni de la procédure `MIXED` de SAS, ni d'autres logiciels permettant de modéliser des données répétées avec des structures de covariance variées. Ce n'est aujourd'hui plus le cas, et cette sorte de pis-aller n'a plus de raison d'être.

Remarque. — Lorsqu'on modélise des données répétées avec la procédure `GLM` de SAS, on dispose donc d'un ensemble de techniques permettant de faire un certain nombre de choses. Toutefois, il faut voir que les seules structures de covariance envisageables avec `GLM` sont l'absence de structure, la structure d'indépendance (en général inappropriée) et la symétrie composée. Dans le cas où cette dernière est rejetée par le test de sphéricité, on n'a donc pas d'alternative à l'absence de structure (dont les estimations ne sont d'ailleurs pas fournies directement), alors que la procédure `MIXED` offre toutes les possibilités signalées au paragraphe précédent.

Remarque. — On notera enfin que, lorsque la structure symétrie composée est acceptable, autrement dit lorsque le test de Mauchly n'est pas significatif, cela ne signifie pas que c'est la structure la mieux adaptée aux données considérées : les principales structures de covariance doivent être envisagée avec la procédure `MIXED`, avant de choisir celle qui minimise soit le critère AIC, soit

le critère BIC.

En conclusion, on peut dire que la modélisation de données répétées avec SAS nécessite en général les deux procédures `GLM` et `MIXED`.

6 Modèles mixtes pour données répétées

6.1 Principe

Dans la pratique des modèles pour données répétées, il est courant d'avoir affaire à des situations bien plus complexes que celle envisagée jusqu'à présent (un seul facteur à effets fixes).

On peut tout d'abord envisager plusieurs facteurs croisés, ou encore un mélange de facteurs et de covariables. Rappelons qu'on appelle covariable une variable quantitative explicative (toujours supposée contrôlée, c'est-à-dire non aléatoire), telle qu'on en utilise en analyse de covariance. Nous ne développons pas davantage ce point qui ne présente aucune difficulté particulière : on choisit d'abord les effets significatifs que l'on met dans le modèle (en utilisant, par exemple, les tests multidimensionnels), puis la structure de covariance (dans le catalogue présenté au paragraphe 7.4, en utilisant les critères indiqués). On estime enfin les paramètres du modèle retenu. À ce niveau, précisons que le temps est, par nature, un facteur à effets fixes : les instants d'observation ne sont pas choisis au hasard, ils correspondent aux moments où l'on souhaite étudier le phénomène considéré.

Mais il est surtout fréquent de considérer des modèles mixtes, mêlant des facteurs (et/ou des covariables) à effets fixes, comme indiqué ci-dessus, et des facteurs à effets aléatoires, en général liés aux individus observés. De façon très générale, en renumérotant les individus selon un indice i variant de 1 à n (si n est le nombre total de sujets observés), sans tenir compte du niveau du (ou des) facteur(s) correspondant(s), on pourra écrire un tel modèle sous la forme suivante :

$$Y_i = \mathbf{X}_i\beta + \mathbf{Z}_iA + U_i.$$

Dans cette écriture :

- Y_i est un vecteur aléatoire de \mathbb{R}^T , si T est le nombre d'instantants d'observation (de répétitions) ;
- \mathbf{X}_i est la matrice d'incidence associée au i -ième sujet et relative aux ef-

fets fixes ; elle est de dimension $T \times p$, si p est le nombre total d'effets fixes pris en compte dans le modèle, y compris les effets liés au temps ; elle ne comporte que des 0 et des 1 (sauf s'il y a une ou plusieurs covariables) ; pour un indice i fixé (donc pour un individu fixé), les lignes de \mathbf{X}_i correspondent aux différents instants d'observation et ne changent que pour les effets liés au temps ;

- β est le vecteur de \mathbb{R}^p contenant les paramètres correspondant aux effets fixes (y compris le temps) ;
- \mathbf{Z}_i est la matrice d'incidence associée au i -ième sujet et relative aux effets aléatoires ; elle est de dimension $T \times q$, si q est le nombre total d'effets aléatoires pris en compte dans le modèle ; elle aussi ne contient que des 0 et des 1 ; de plus, ses lignes sont constantes pour un indice i fixé ;
- A est le vecteur aléatoire de \mathbb{R}^q représentant les effets aléatoires ; on pose $A \sim \mathcal{N}_q(0, \mathbf{G})$, où \mathbf{G} définit la structure de covariance associée aux effets aléatoires ; le plus souvent, \mathbf{G} est choisie bloc-diagonale ; on notera que A a la même structure qu'en 6.3.1 et que \mathbf{G} correspond à la matrice introduite dans la remarque 68 ;
- U_i est le vecteur aléatoire erreur de \mathbb{R}^T ; sa loi est $\mathcal{N}_T(0, \mathbf{R})$; la structure de covariance associée aux données répétées est donc définie par la covariance de U_i (on notera que l'on n'a pas indicé \mathbf{R} par i , de sorte que l'on est encore dans le cadre d'un modèle homoscédastique) ; les vecteurs aléatoires U_i ($i = 1, \dots, n$) sont supposés i.i.d. et indépendants de A ;
- Ainsi, chaque vecteur aléatoire Y_i de \mathbb{R}^T est distribué selon une loi $\mathcal{N}_T(\mathbf{X}_i\beta, \mathbf{V})$, avec $\mathbf{V} = \mathbf{Z}_i\mathbf{G}\mathbf{Z}_i' + \mathbf{R}$.

Ce type de modèle, très général, synthétise donc d'une part ce que nous avons vu au chapitre 6, avec les modèles mixtes, d'autre part les aspects spécifiques des modèles pour données répétées, étudiés précédemment dans ce chapitre. Nous ne le développons pas davantage sur le plan théorique, même s'il constitue le quotidien de la pratique des données répétées : c'est plutôt sa mise en œuvre qui nécessite des développements, ce qui relève de la pratique.

Remarque. — Pour retrouver la formule donnée en 7.3.1 pour un modèle sans effets aléatoires, il faut, dans la formule ci-dessus, annuler le terme A et, dans l'autre formule, d'une part considérer les vecteurs de \mathbb{R}^T des Y_{ijt} ($t = 1, \dots, T$), notés Y_{ij} , d'autre part renuméroter ces vecteurs selon un indice i variant de 1 à n ; le nouveau terme μ_i de cette formule (noté μ_j en 7.3.1) correspond alors à $\mathbf{X}_i\beta$.

6.2 Usage de la procédure MIXED

Lorsqu'on utilise la procédure MIXED de SAS pour traiter des données répétées, les trois commandes fondamentales de cette procédure sont explicitées ci-dessous (la seconde sera omise dans le cas d'un modèle sans effet aléatoire).

La commande "model"

Sa syntaxe est `model y = f1 f2 ...` ; elle sert à déclarer les effets fixes du modèle, y compris le temps ; elle permet à SAS de construire la "super" matrice d'incidence \mathbf{X} constituée des matrices \mathbf{X}_i mises les unes en dessous des autres (\mathbf{X} est donc de dimension $nT \times p$).

La commande "random"

Sa syntaxe est `random a1 a2 ...` ; elle sert à déclarer les effets aléatoires, donc à construire la "super" matrice d'incidence \mathbf{Z} constituée des matrices \mathbf{Z}_i mises les unes en dessous des autres (\mathbf{Z} est donc de dimension $nT \times q$) ; c'est dans cette commande que l'on doit déclarer la structure de covariance \mathbf{G} , avec l'option `type=...`

La commande "repeated"

Sa syntaxe est `repeated / <options>`, deux options étant ici indispensables : `sub=...`, pour déclarer la variable contenant les étiquettes des sujets (variable restant constante lorsque les observations sont répétées sur le même sujet) ; `type=...`, pour déclarer la structure de covariance intra-sujet (voir le paragraphe 7.4), ce qui permet de construire la matrice \mathbf{R} associée.

6.3 Inférence

On a déjà signalé, au chapitre 6, que les tests au sein des modèles mixtes sont assez délicats à mettre en œuvre. Pour un modèle mixte relatif à des données répétées, les choses sont, bien sûr, encore plus délicates puisque trois ensembles d'éléments sont maintenant à choisir pour déterminer le modèle retenu : les effets fixes, les effets aléatoires et la structure de covariance \mathbf{R} . Disons tout de suite qu'il n'y a pas de méthode réellement satisfaisante. On peut néanmoins envisager de procéder par étape : on commence, comme indiqué à la fin du point 7.4, par tester les effets fixes puis, avec les effets retenus, on choisit la structure de covariance. Enfin, on estime les composantes de la variance cor-

respondant aux effets aléatoires et on retient les effets les plus importants.

Concernant les estimations, c'est le plus souvent la méthode REML (maximum de vraisemblance restreint) qui est utilisée. On trouvera divers compléments sur l'inférence dans les modèles mixtes pour données répétées dans l'ouvrage de Verbeke & Molenberghs (2000).

7 Illustration

Les données

Il s'agit d'un exemple fictif de données répétées. Il comporte 20 individus (leurs codes, de 01 à 20, sont en première colonne), un facteur fixe à 4 niveaux (codés de 1 à 4, en seconde colonne), et une variable réponse, observée à 3 instants : les réponses aux instants 1, 2 et 3 sont dans les 3 dernières colonnes.

Les données sont reproduites ci-dessous.

01	1	60	70	80
02	1	65	70	75
03	1	60	75	75
04	1	70	75	80
05	1	75	75	90
06	2	60	65	85
07	2	80	90	100
08	2	65	80	95
09	2	85	90	95
10	2	90	90	90
11	3	90	95	100
12	3	85	90	95
13	3	85	85	100
14	3	80	90	100
15	3	70	80	90
16	4	80	80	104
17	4	85	95	114
18	4	90	95	109
19	4	95	100	114
20	4	90	90	129

Le programme SAS

Le programme SAS ci-dessous permet de réaliser différents traitements de ces données, en utilisant les procédures GLM (avec la commande REPEATED), puis MIXED (après transformation indispensable des données), enfin VARCOMP, pour retrouver certains résultats de MIXED et mieux comprendre le fonctionnement de ces traitements.

```

* ----- ;
*          IMPORTATION DES DONNEES          ;
* ----- ;
data fic1;
infile 'fic.don';
input indiv trait y1-y3;
run;
* ----- ;
*          GLM AVEC REPEATED                ;
* ----- ;
proc glm data=fic1;
class trait;
model y1-y3 = trait / ss3;
repeated temps / printh printe;
run;
* ----- ;
*          et avec CONTRAST                 ;
* ----- ;
proc glm data=fic1;
class trait;
model y1-y3 = trait / ss3 nouni;
repeated temps contrast(1)/ printm printh printe;
run;
quit;
* ----- ;
*          TRANSFORMATION DES DONNEES POUR MIXED ;
* ----- ;
data fic2;
set fic1;
y=y1; temps=1; output;
y=y2; temps=2; output;
y=y3; temps=3; output;

```

```

* ----- ;
*          PROC MIXED - REML          ;
*          COMPOUND SYMMETRY         ;
* ----- ;
proc mixed data=fic2 method=reml;
class trait temps;
model y = trait | temps;
repeated / sub=indiv type=cs;
run;
* ----- ;
*          sans interactions          ;
* ----- ;
proc mixed data=fic2 method=reml;
class trait temps;
model y = trait temps / s;
repeated / sub=indiv type=cs r;
run;
quit;
* ----- ;
*          PROC VARCOMP              ;
* ----- ;
proc varcomp data=fic2 method=reml;
class indiv trait temps;
model y = temps trait indiv / fixed=2;
run;
quit;

```

PAGE 2

Dependent Variable: y1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1320.000000	440.000000	5.87	0.0067
Error	16	1200.000000	75.000000		
Corrected Total	19	2520.000000			

R-Square	Coeff Var	Root MSE	y1 Mean
0.523810	11.10289	8.660254	78.00000

Source	DF	Type III SS	Mean Square	F Value	Pr > F
trait	3	1320.000000	440.000000	5.87	0.0067

PAGE 3

Dependent Variable: y2

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1010.000000	336.666667	6.19	0.0054
Error	16	870.000000	54.375000		
Corrected Total	19	1880.000000			

R-Square	Coeff Var	Root MSE	y2 Mean
0.537234	8.778501	7.373941	84.00000

Les sorties de la procédure GLM

PAGE 1

The GLM Procedure

Class Level Information

Class	Levels	Values
trait	4	1 2 3 4

Number of Observations Read 20
Number of Observations Used 20

Source	DF	Type III SS	Mean Square	F Value	Pr > F
trait	3	1010.000000	336.666667	6.19	0.0054

PAGE 4

Dependent Variable: y3

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F	Partial Correlation Coefficients from the Error SSCP Matrix of the Variables Defined by the Specified Transformation / Prob > r		
Model	3	2950.000000	983.333333	22.16	<.0001			
Error	16	710.000000	44.375000			DF = 16	temps_1	temps_2
Corrected Total	19	3660.000000				temps_1	1.000000	0.793025 0.0001
	R-Square	Coeff Var	Root MSE	y3 Mean		temps_2	0.793025 0.0001	1.000000
	0.806011	6.939017	6.661456	96.00000				

Source	DF	Type III SS	Mean Square	F Value	Pr > F	Sphericity Tests				
trait	3	2950.000000	983.333333	22.16	<.0001	Variables	DF	Mauchly's Criterion	Chi-Square	Pr > ChiSq
						Transformed Variates	2	0.3688221	14.961612	0.0006
						Orthogonal Components	2	0.7564513	4.186756	0.1233

PAGE 5

Repeated Measures Analysis of Variance

PAGE 6

Repeated Measures Level Information

Dependent Variable	y1	y2	y3
Level of temps	1	2	3

H = Type III SSCP Matrix for temps

temps_N represents the contrast between the nth level of temps and the last

Partial Correlation Coefficients from the Error SSCP Matrix / Prob > |r|

DF = 16	y1	y2	y3
y1	1.000000	0.817215 <.0001	0.476687 0.0530
y2	0.817215 <.0001	1.000000	0.445327 0.0732
y3	0.476687 0.0530	0.445327 0.0732	1.000000

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of no temps Effect

H = Type III SSCP Matrix for temps
E = Error SSCP Matrix

S=1 M=0 N=6.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.13552715	47.84	2	15	<.0001
Pillai's Trace	0.86447285	47.84	2	15	<.0001
Hotelling-Lawley Trace	6.37859532	47.84	2	15	<.0001
Roy's Greatest Root	6.37859532	47.84	2	15	<.0001

temps_N represents the contrast between the nth level of temps and the last

	temps_1	temps_2
temps_1	1030	755
temps_2	755	880

H = Type III SSCP Matrix for temps*trait

temps_N represents the contrast between the nth level of temps and the last

	temps_1	temps_2
temps_1	450	555
temps_2	555	690

Greenhouse-Geisser Epsilon	0.8042
Huynh-Feldt Epsilon	1.0480

La procédure GLM avec CONTRAST

MANOVA Test Criteria and F Approximations for the Hypothesis of no temps*trait Effect
H = Type III SSCP Matrix for temps*trait
E = Error SSCP Matrix

PAGE 1

S=2 M=0 N=6.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.55370370	1.72	6	30	0.1507
Pillai's Trace	0.45037037	1.55	6	32	0.1940
Hotelling-Lawley Trace	0.79866221	1.94	6	18.326	0.1275
Roy's Greatest Root	0.78934068	4.21	3	16	0.0225

The GLM Procedure
Repeated Measures Analysis of Variance

Repeated Measures Level Information

Dependent Variable	y1	y2	y3
Level of temps	1	2	3

Partial Correlation Coefficients from the Error SSCP Matrix / Prob > |r|

DF = 16	y1	y2	y3
y1	1.000000	0.817215 <.0001	0.476687 0.0530
y2	0.817215 <.0001	1.000000	0.445327 0.0732
y3	0.476687 0.0530	0.445327 0.0732	1.000000

NOTE: F Statistic for Roy's Greatest Root is an upper bound.
NOTE: F Statistic for Wilks' Lambda is exact.

PAGE 7

Tests of Hypotheses for Between Subjects Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
trait	3	4890.000000	1630.000000	12.98	0.0001
Error	16	2010.000000	125.625000		

temps_N represents the contrast between the nth level of temps and the 1st

M Matrix Describing Transformed Variables

	y1	y2	y3
temps_2	-1.000000000	1.000000000	0.000000000
temps_3	-1.000000000	0.000000000	1.000000000

PAGE 8

Univariate Tests of Hypotheses for Within Subject Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
temps	2	3360.000000	1680.000000	69.82	<.0001
temps*trait	6	390.000000	65.000000	2.70	0.0309
Error(temps)	32	770.000000	24.062500		

E = Error SSCP Matrix

temps_N represents the contrast between the nth level of temps and the 1st

Source	Adj Pr > F	
	G - G	H - F
temps	<.0001	<.0001
temps*trait	0.0446	0.0309
Error(temps)		

	temps_2	temps_3
temps_2	400	275
temps_3	275	1030

Partial Correlation Coefficients from the Error SSCP Matrix of the

Variables Defined by the Specified Transformation / Prob > |r|

DF = 16	temps_2	temps_3
temps_2	1.000000	0.428434 0.0862
temps_3	0.428434 0.0862	1.000000

temps_2	30	-105
temps_3	-105	450

PAGE 3

PAGE 2

MANOVA Test Criteria and F Approximations for the Hypothesis of no temps*trait Effect
H = Type III SSCP Matrix for temps*trait
E = Error SSCP Matrix

S=2 M=0 N=6.5

Sphericity Tests

Variables	DF	Mauchly's Criterion	Chi-Square	Pr > ChiSq	Statistic	Value	F Value	Num DF	Den DF	Pr > F
Transformed Variates	2	0.6579784	6.278748	0.0433	Wilks' Lambda	0.55370370	1.72	6	30	0.1507
Orthogonal Components	2	0.7564513	4.186756	0.1233	Pillai's Trace	0.45037037	1.55	6	32	0.1940
					Hotelling-Lawley Trace	0.79866221	1.94	6	18.326	0.1275
					Roy's Greatest Root	0.78934068	4.21	3	16	0.0225

NOTE: F Statistic for Roy's Greatest Root is an upper bound.
NOTE: F Statistic for Wilks' Lambda is exact.

H = Type III SSCP Matrix for temps

temps_N represents the contrast between the nth level of temps and the 1st

	temps_2	temps_3
temps_2	720	2160
temps_3	2160	6480

PAGE 4

Tests of Hypotheses for Between Subjects Effects

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of no temps Effect
H = Type III SSCP Matrix for temps
E = Error SSCP Matrix

Source	DF	Type III SS	Mean Square	F Value	Pr > F
trait	3	4890.000000	1630.000000	12.98	0.0001
Error	16	2010.000000	125.625000		

PAGE 5

S=1 M=0 N=6.5

Univariate Tests of Hypotheses for Within Subject Effects

Statistic	Value	F Value	Num DF	Den DF	Pr > F	Source	DF	Type III SS	Mean Square	F Value	Pr > F
Wilks' Lambda	0.13552715	47.84	2	15	<.0001	temps	2	3360.000000	1680.000000	69.82	<.0001
Pillai's Trace	0.86447285	47.84	2	15	<.0001	temps*trait	6	390.000000	65.000000	2.70	0.0309
Hotelling-Lawley Trace	6.37859532	47.84	2	15	<.0001	Error (temps)	32	770.000000	24.062500		
Roy's Greatest Root	6.37859532	47.84	2	15	<.0001						

H = Type III SSCP Matrix for temps*trait

temps_N represents the contrast between the nth level of temps and the 1st

	temps_2	temps_3
--	---------	---------

Source	Adj G - G	Pr > F	H - F
temps	<.0001	<.0001	
temps*trait	0.0446	0.0309	

Error (temps)
 Greenhouse-Geisser Epsilon 0.8042
 Huynh-Feldt Epsilon 1.0480

Number of Observations Not Used 0

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	350.36360353	
1	1	334.64512218	0.00000000

Les sorties de la procédure MIXED

Il est tout d'abord nécessaire de procéder à une transformation des données, pour disposer les différents instants d'observation en lignes et non plus en colonnes. Ensuite, on a choisi la structure de covariance *compound symmetry* (puisque le test de Maucly n'est pas significatif) et utilisé d'abord un modèle avec interactions, puis un modèle additif.

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
CS	indiv	33.8542
Residual		24.0625

PAGE 1

The Mixed Procedure

Model Information

Data Set WORK.FIC2
 Dependent Variable y
 Covariance Structure Compound Symmetry
 Subject Effect indiv
 Estimation Method REML
 Residual Variance Method Profile
 Fixed Effects SE Method Model-Based
 Degrees of Freedom Method Between-Within

PAGE 2

Fit Statistics

-2 Res Log Likelihood	334.6
AIC (smaller is better)	338.6
AICC (smaller is better)	338.9
BIC (smaller is better)	340.6

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	15.72	<.0001

Class Level Information

Class	Levels	Values
trait	4	1 2 3 4
temps	3	1 2 3

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
trait	3	16	12.98	0.0001
temps	2	32	69.82	<.0001
trait*temps	6	32	2.70	0.0309

Dimensions

Covariance Parameters 2
 Columns in X 20
 Columns in Z 0
 Subjects 20
 Max Obs Per Subject 3

PAGE 3

Number of Observations

Number of Observations Read 60
 Number of Observations Used 60

Model Information

Data Set WORK.FIC2 PAGE 4
 Dependent Variable y -----
 Covariance Structure Compound Symmetry
 Subject Effect indiv
 Estimation Method REML
 Residual Variance Method Profile
 Fixed Effects SE Method Model-Based
 Degrees of Freedom Method Between-Within

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
CS	indiv	31.6996
Residual		30.5263

Class Level Information

Class	Levels	Values
trait	4	1 2 3 4
temps	3	1 2 3

Fit Statistics

-2 Res Log Likelihood	376.2
AIC (smaller is better)	380.2
AICC (smaller is better)	380.4
BIC (smaller is better)	382.2

Dimensions

Covariance Parameters	2
Columns in X	8
Columns in Z	0
Subjects	20
Max Obs Per Subject	3

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	12.68	0.0004

Number of Observations

Number of Observations Read	60
Number of Observations Used	60
Number of Observations Not Used	0

Solution for Fixed Effects

Effect	trait	temps	Estimate	Standard Error	DF	t Value	Pr > t
Intercept			108.00	3.0647	16	35.24	<.0001
trait	1		-25.0000	4.0927	16	-6.11	<.0001
trait	2		-14.0000	4.0927	16	-3.42	0.0035
trait	3		-9.0000	4.0927	16	-2.20	0.0429
trait	4		0
temps		1	-18.0000	1.7472	38	-10.30	<.0001
temps		2	-12.0000	1.7472	38	-6.87	<.0001
temps		3	0

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	388.88556695	
1	1	376.20962184	0.00000000

Convergence criteria met.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
trait	3	16	12.98	0.0001
temps	2	38	55.03	<.0001

Estimated R Matrix for Subject 1

Row	Col1	Col2	Col3
1	62.2259	31.6996	31.6996
2	31.6996	62.2259	31.6996
3	31.6996	31.6996	62.2259

Les sorties de la procédure VARCOMP

Elles n'apportent rien de nouveau. Elles permettent seulement de retrouver certains résultats de la procédure MIXED, moyennant une déclaration particulière des individus (voir le programme plus haut).

```

Variance Components Estimation Procedure

      Class Level Information

Class      Levels  Values
----      -
indiv      20     1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
trait      4       1 2 3 4
temps     3       1 2 3

      Number of Observations Read      60
      Number of Observations Used      60

      Dependent Variable:      y

      REML Iterations

Iteration      Objective      Var(indiv)      Var(Error)
-----
0      207.2392071872      31.6995614035      30.5263157895
1      207.2392071872      31.6995614035      30.5263157895

Convergence criteria met.

      REML Estimates

      Variance
      Component      Estimate
-----
      Var(indiv)      31.69956
      Var(Error)      30.52632

      Asymptotic Covariance Matrix of Estimates

      Var(indiv)      Var(Error)
-----
      Var(indiv)      224.63890      -16.34835
      Var(Error)      -16.34835      49.04505
  
```