



HAL
open science

Estimation de la répétabilité et de la reproductibilité des mesures expérimentales. Utilisation des procédures VARCOMP et NESTED du logiciel SAS/STAT (version 6.12)

Jean-Pierre Gauchi

► **To cite this version:**

Jean-Pierre Gauchi. Estimation de la répétabilité et de la reproductibilité des mesures expérimentales. Utilisation des procédures VARCOMP et NESTED du logiciel SAS/STAT (version 6.12). [Rapport Technique] RT 2000-4, auto-saisine. 2000, 44 p. <hal-01604565>

HAL Id: hal-01604565

<https://hal.science/hal-01604565v1>

Submitted on 7 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-SA 4.0 - Attribution - ShareAlike - International License



Institut National de la Recherche Agronomique
Centre de Recherches de Versailles
Unité de Biométrie
Equipe Analyse Quantitative des Risques Alimentaires

Estimation
de la Répétabilité et de la Reproductibilité
des mesures expérimentales

Utilisation
des procédures VARCOMP et NESTED
du logiciel SAS/STAT (version 6.12)

Jean-Pierre Gauchi

Rapport technique 2000-4

13 Avril 2000

Sommaire

Introduction.....	p. 2
<u>Partie théorique</u>	
1. Définitions.....	p. 4
2. Les questions posées.....	p. 6
3. Les modèles.....	p. 6
4. Intervalles de répétabilité et de reproductibilité.....	p. 8
5. Résultats de base.....	p. 9
6. Espérances des carrés moyens et estimation des variances de répétabilité et reproductibilité.....	p. 11
7. Intervalles de confiance des différentes variances.....	p. 12
<u>Partie pratique</u>	
8. Présentation de la procédure VARCOMP.....	p. 15
9. Présentation de la procédure NESTED.....	p. 16
10. Exemples.....	p. 17
- Exemple 1 : exemple artificiel.....	p. 18
- Exemple 2 : Analyse labo 1.....	p. 22
- Exemple 3 : Analyse labo 2.....	p. 30
- Exemple 4 : Vitesse de réaction.....	p. 32
- Exemple 5 : Mycotoxine.....	p. 34
- Exemple 6 : Calcium.....	p. 43
Conclusion.....	p. 44

Introduction

Ce document a pour but de montrer comment estimer des variances de répétabilité et de reproductibilité des mesures expérimentales en se plaçant dans le contexte du modèle de l'analyse de la variance. Il a été rédigé à l'attention spéciale des membres de la cellule opérationnelle du projet PREVIUS en vue de leur apporter une aide lors de la mise en œuvre de tels calculs. Les mesures expérimentales concernées sont typiquement celles des laboratoires d'analyse (chimie analytique, biologie, microbiologie,...), des procédés industriels (par exemple agro-alimentaires), ou des comparaisons inter-laboratoires. Les calculs sont réalisés avec les procédures VARCOMP et NESTED du logiciel SAS/STAT (version 6.12). On supposera ici que les erreurs sont gaussiennes et les mesures indépendantes. On sait en effet que s'il n'y a pas indépendance, ces procédures ne sont pas adaptées et il faut alors s'orienter vers la procédure MIXED qui fera l'objet d'un rapport technique ultérieur.

L'originalité scientifique de ce document est faible mais son aspect synthétique nous semble très utile pour les expérimentateurs, non statisticiens par définition, ayant à faire face à de nombreuses notices et normes où ces notions de répétabilité et reproductibilité sont certes bien définies mais où les procédures de calcul fournies sont rarement rigoureuses et justes. Ainsi, même dans les normes ISO récentes de 1994 à 1998 n'apparaissent pas clairement les modèles postulés, la distinction entre les effets fixes et les effets aléatoires, les tableaux usuels de l'analyse de la variance et d'autre part, aucun logiciel n'est indiqué pour mener à bien ce type de calcul, ce qui, il est vrai, n'est pas l'objectif premier de ces normes. Nous montrons ici comment le contexte du modèle de l'analyse de variance peut conduire d'une part à une présentation plus élégante de ces notions de répétabilité et reproductibilité et d'autre part à un calcul facilité de celles-ci. En outre, la procédure VARCOMP du logiciel SAS/STAT, très simple d'utilisation, fournit notamment la matrice de variance des estimateurs des variances de répétabilité et reproductibilité, information indispensable mais pourtant assez rare dans les logiciels du marché et totalement absente des normes officielles AFNOR et ISO.

Cet exposé emprunte des éléments des documents suivants :

- Hunter, J.S., 1985. Measurement Error *in* Encyclopaedia of Statistical Sciences, Kotz & Johnson (Eds), Wiley.

- Tenenhaus, M., 1993. Répétabilité et Reproductibilité de la mesure. Document du Groupe de Réflexion en Analyse des Données (interne au Groupe Rhône-Poulenc).
- Procédure VARCOMP, documentation SAS réf :56045, année 1994.
- Procédure NESTED, documentation SAS réf :56045, année 1994.

Partie théorique

1. Définitions

Au plan international, les différents comités de normalisation AFNOR et ISO proposent les définitions suivantes pour la répétabilité et la reproductibilité.

Répétabilité

Le concept de répétabilité correspond à l'application de la même méthode de mesure par le même opérateur, avec le même instrument de mesure, au même lieu, dans les mêmes conditions d'emploi et **en répétant le mesurage** sur une courte période de temps. La répétabilité se caractérise par la dispersion plus ou moins grande des résultats de mesure et elle peut s'exprimer, par exemple, par une **variance de répétabilité ou un écart-type de répétabilité**, ou tout autre paramètre statistique caractéristique de la dispersion des résultats. On notera ici σ_r^2 la variance de répétabilité.

Remarque :

En termes plus statistiques, la variance de répétabilité serait donc une variance d'erreur « pure ».

Reproductibilité

Le concept de reproductibilité correspond à la variation d'au moins une des conditions qui ont été énumérées pour la répétabilité, les autres restant alors fixées. La reproductibilité s'exprime par une **variance de reproductibilité ou un écart-type de reproductibilité**. On notera ici σ_R^2 la variance de reproductibilité.

Remarques :

Donc, il suffit qu'une des conditions **au moins** du dispositif expérimental, entièrement figé lors de la mesure de répétabilité, ait varié entre deux mesures pour que cette condition contribue à la variance de reproductibilité. Par exemple, la méthode de mesure est fixée mais

elle est mise en œuvre dans deux laboratoires différents. Il se peut aussi que l'ensemble des conditions varie.

Par ailleurs, il nous faut préciser que dans les normes officielles la reproductibilité est calculée différemment de notre σ_R^2 , elle est en général formée de la somme de σ_r^2 et d'une variance que l'on notera ici $\sigma_{R'}^2$ qui ne prend pas en compte les interactions entre les effets comme on verra plus loin sur l'exemple 5. On notera $\sigma_{Repro-normes}^2$ la variance de reproductibilité calculée dans les normes. On notera $\sigma_{r\&R}^2$ la variance totale de la mesure formée de la somme de σ_r^2 et de $\sigma_{R'}^2$. En résumé on a :

- σ_r^2 : variance de répétabilité, c'est la variance d'erreur « pure »,
- σ_R^2 : variance de reproductibilité, c'est la somme de toutes les sources de la variabilité de la variance de la réponse, les interactions entre facteurs sont prises en compte, hors la variance de répétabilité σ_r^2 ,
- $\sigma_{R'}^2$: variance de reproductibilité basée sur la somme des sources de la variabilité de la variance de la réponse, hors la variance de répétabilité σ_r^2 et hors celles relatives aux interactions éventuelles entre facteurs,
- $\sigma_{r\&R}^2 = \sigma_r^2 + \sigma_{R'}^2$,
- $\sigma_{Repro-normes}^2 = \sigma_r^2 + \sigma_{R'}^2$.

Nous renvoyons le lecteur à la bibliographie suivante pour d'autres aspects concernant les mesures expérimentales :

- Guide pour l'expression de l'incertitude de mesure. Première Edition 1995. ISBN 92-67-20188-3. Organisation internationale de normalisation.
- Exactitude (justesse et fidélité) des résultats et méthodes de mesure. NF ISO 5725 en 6 parties, Décembre 1994.

2. Les questions posées

La question centrale qui nous intéresse ici est l'estimation des variances de répétabilité et de reproductibilité. Néanmoins, d'autres questions corollaires sont importantes comme :

- savoir si la variance de reproductibilité est significativement supérieure à la variance de répétabilité, ce qui impliquerait, en cas de confirmation, un **mesurage robuste** à la variation des conditions expérimentales,
- comparer entre eux, et à la variance de répétabilité, les termes constitutifs de la variance de reproductibilité à l'aide d'un test statistique ; c'est le cadre statistique usuel des composantes de la variance.

Plus globalement, il faut prendre en compte la nature des facteurs de l'ensemble du dispositif expérimental pour faire une estimation correcte de ces variances: les facteurs sont-ils fixes ou aléatoires ? Si tous les facteurs sont fixes, ce n'est pas un problème de répétabilité-reproductibilité qu'il faut traiter mais plutôt un problème d'estimation des effets qui jouent significativement sur le niveau moyen de la réponse. Le paragraphe suivant distingue deux types de modèle fréquents où au moins un facteur est aléatoire.

3. Les modèles

La première difficulté pour postuler un modèle d'analyse de la variance de la grandeur mesurée (la réponse) en fonction des facteurs susceptibles d'influencer son niveau moyen et/ou sa variance est de déterminer soigneusement quels sont les facteurs qu'on qualifera de fixes et ceux que l'on qualifera d'aléatoires. En outre, il faut décider si le modèle est croisé (avec ou sans interaction) ou emboîté (on dit aussi hiérarchisé). On expose ci-dessous les trois formes qui nous intéressent dans le cadre de la répétabilité et de la reproductibilité, pour deux facteurs A et B , elles sont généralisables à un nombre quelconque de facteurs :

- le modèle mixte croisé (1) à un facteur fixe, un facteur aléatoire, et l'interaction (aléatoire) de ces deux facteurs,

- le modèle mixte hiérarchisé (2) à un facteur primaire fixe et un facteur secondaire aléatoire, ou bien où tous les facteurs sont aléatoires,
- le modèle aléatoire hiérarchisé (3) où tous les facteurs sont aléatoires.

Les données peuvent être déséquilibrées, c'est-à-dire qu'il y a le même nombre de répétitions pour chaque combinaison de niveaux de tous les facteurs.

Le modèle mixte croisé (1)

On pose :

$$Y_{ijk} = \eta + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \quad (1)$$

avec :

- Y_{ijk} : variable aléatoire dont on observe la réalisation y_{ijk} c'est-à-dire la k -ième mesure réalisée pour la modalité i du facteur A et la modalité j du facteur B ,
- η : une valeur de référence,
- α_i : un effet, **fixe**, de la modalité i d'un facteur A à p modalités, $i = 1, \dots, p$,
- β_j : un effet, **aléatoire**, de la modalité j d'un facteur B à q modalités, $j = 1, \dots, q$, choisies parmi un ensemble très grand ou infini de modalités,
- γ_{ij} : un effet d'interaction, **aléatoire**, du croisement de la modalité i du facteur A et de la modalité j du facteur B ,
- ε_{ijk} : l'erreur aléatoire associée à la k -ième mesure, $k = 1, \dots, n_{ij}$, soit n_{ij} répétitions, avec $n_{ij} = n$, $\forall i, j$ si les données sont équilibrées.

On suppose $\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon)$, et $\text{cov}(\varepsilon_{ijk}, \varepsilon_{ijk'}) = 0$, $k \neq k'$, et de plus que $\beta_j \sim N(0, \sigma_\beta)$, $\gamma_{ij} \sim N(0, \sigma_\gamma)$, ainsi que l'indépendance entre les aléas issus des trois lois gaussiennes. Les exemples 2 et 3 illustrent ce modèle.

Le modèle mixte hiérarchique (2)

On pose :

$$Y_{ijk} = \eta + \alpha_i + \beta_{ij} + \varepsilon_{ijk} \quad (2)$$

avec β_{ij} l'effet, **aléatoire**, de la modalité j du facteur B **emboîtée** dans la modalité i du facteur A . Les autres notations sont les mêmes que celles du modèle (1). On suppose aussi que $\beta_{ij} \sim N(0, \sigma_\beta)$ et l'indépendance entre tous les aléas. L'exemple 4 donne un exemple de modèle mixte emboîté.

Le modèle aléatoire hiérarchique (3)

C'est la même forme que le modèle (2) mais tous les effets sont aléatoires (voir l'exemple 6).

4. Intervalles de répétabilité et de reproductibilité

Définissons maintenant les intervalles de répétabilité et de reproductibilité.

Intervalle de Répétabilité

L'intervalle de répétabilité I_r est l'intervalle de probabilité (en général à 95%) de la **différence entre deux mesures** réalisées par le même opérateur sur le même échantillon, dans les mêmes conditions expérimentales. On a :

$$Y_{ijk} - Y_{ijk'} = \varepsilon_{ijk} - \varepsilon_{ijk'}$$

et comme :

$$\varepsilon_{ijk} - \varepsilon_{ijk'} \sim N(0, \sqrt{2\sigma_r^2})$$

on déduit :

$$I_r = [-2\sqrt{2}\sigma_r ; +2\sqrt{2}\sigma_r].$$

Intervalle de Reproductibilité

L'intervalle de reproductibilité I_R est l'intervalle de probabilité (en général à 95%) de la différence entre deux mesures réalisées dans des conditions expérimentales différentes. Par exemple, pour le modèle (1) ce pourrait être un opérateur (une modalité d'un facteur fixe « opérateur » présentant deux modalités, c'est-à-dire deux opérateurs) dosant un produit donné dans deux échantillons différents, tirés au hasard d'un même lot (facteur aléatoire « échantillon » à deux modalités). On aurait alors :

$$Y_{ijk} - Y_{ij'k'} = \beta_j + \gamma_{ij} + \varepsilon_{ijk} - \beta_{j'} - \gamma_{ij'} - \varepsilon_{ij'k'}.$$

Cette différence suivant une loi : $N(0, \sqrt{2(\sigma_R^2 + \sigma_r^2)})$, on déduit :

$$I_R = [-2\sqrt{2(\sigma_r^2 + \sigma_R^2)} ; +2\sqrt{2(\sigma_r^2 + \sigma_R^2)}].$$

5. Résultats de base

Le théorème fondamental qui nous intéresse ici est le suivant :

Théorème

Soit un vecteur aléatoire Y de moyenne μ et de variance V .

Alors :

$$E(Y'QY) = \text{Trace}(QV) + \mu'Q\mu.$$

Appliquons ce théorème aux trois types de modèle suivants :

Application au modèle à effets fixes

Ce modèle ne nous concerne pas dans le cadre de l'étude répétabilité-reproductibilité mais il permet de mieux appréhender les formules des deux autres types.

Si on pose $Y = X\beta + \varepsilon$ où X est la matrice du plan d'expériences, avec $E(Y) = X\beta$ et $V = \sigma_\varepsilon^2 I$ on obtient :

$$E(Y'QY) = \text{Trace}(Q\sigma_\varepsilon^2 I) + \beta' X' QX \beta = \sigma_\varepsilon^2 \text{Trace}(Q) + \beta' X' QX \beta \quad (3)$$

Cette formule générale permet de calculer l'espérance des sommes de carrés usuelles.

Application au modèle mixte

Le vecteur β des paramètres se décompose en paramètres fixes et aléatoires. On pose :

$\beta' = [\beta_1', \beta_A', \dots, \beta_K']$ où le vecteur β_1 est fixe et le vecteur $\beta_\theta' = [\beta_A', \dots, \beta_K']$ regroupant les effets aléatoires. On suppose $E(\beta_\theta) = 0$, $V(\beta_\theta) = \sigma_\theta^2 I_{N_\theta}$ avec N_θ le nombre total de modalités de tous les facteurs aléatoires, et $\text{cov}(\beta_\theta, \beta_{\theta'}) = 0, \theta \neq \theta'$. Si on présente le modèle (1) sous la forme matricielle :

$$Y = X_1 \beta_1 + \sum_{\theta=A}^K X_\theta \beta_\theta + \varepsilon$$

on a :

$$E(Y) = X_1 \beta_1$$

$$V(Y) = \sum_{\theta=A}^K \sigma_\theta^2 X_\theta X_\theta' + \sigma_\varepsilon^2 I_N$$

où N est le nombre de lignes du vecteur Y . Alors l'application du théorème donne :

$$E(Y'QY) = \beta_1' X_1' Q X_1 \beta_1 + \sum_{\theta=A}^K \sigma_\theta^2 \text{Trace}(Q X_\theta X_\theta') + \sigma_\varepsilon^2 \text{Trace}(Q) \quad (4)$$

Application au modèle à effets aléatoires

Tous les effets étant aléatoires, le vecteur β_1 se réduit à la constante μ et X_1 se réduit au vecteur $\mathbf{1}$. On a donc :

$$E(Y'QY) = \mu^2 \mathbf{1}' Q \mathbf{1} + \sum_{\theta=A}^K \sigma_\theta^2 \text{Trace}(Q X_\theta X_\theta') + \sigma_\varepsilon^2 \text{trace}(Q) \quad (5)$$

6. Espérances des carrés moyens et estimation des variances σ_r^2 et σ_R^2

Pour le modèle (1)

L'application de la formule (4) au modèle (1) permet d'obtenir l'espérance des carrés moyens (on considère ici des données équilibrées) :

- $E(CM(A)) = \sigma_\varepsilon^2 + n\sigma_\gamma^2 + \frac{qn}{p-1} \sum_{i=1}^p (\alpha_i - \bar{\alpha})^2$,
- $E(CM(B)) = \sigma_\varepsilon^2 + n\sigma_\gamma^2 + pn\sigma_\beta^2$,
- $E(CM(A \times B)) = \sigma_\varepsilon^2 + n\sigma_\gamma^2$,
- $E(CM(\text{résidu})) = \sigma_\varepsilon^2$.

On en déduit les estimations des variances σ_β^2 , σ_γ^2 et σ_ε^2 :

- $\hat{\sigma}_\varepsilon^2 = \frac{1}{pq(n-1)} \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n (y_{ijk} - \bar{y}_{ij\bullet})^2$, estimation usuelle de σ_ε^2 ,
- $\hat{\sigma}_\gamma^2 = [CM(A \times B) - CM(\text{résidu})]/n$,
- $\hat{\sigma}_\beta^2 = [CM(B) - CM(A \times B)]/np$,

Enfin, on aboutit à l'estimation des variances de répétabilité et de reproductibilité :

- $\hat{\sigma}_r^2 = \hat{\sigma}_\varepsilon^2 = CM(\text{résidu})$,
- $\hat{\sigma}_R^2 = \hat{\sigma}_\beta^2 + \hat{\sigma}_\gamma^2 = [CM(B)]/pn + [(p-1) \times CM(A \times B)]/pn - CM(\text{résidu})/n$,
- $\hat{\sigma}_{r\&R}^2 = \hat{\sigma}_r^2 + \hat{\sigma}_R^2 = (1/pn) \times [CM(B) + (p-1) \times CM(A \times B) + p(n-1)CM(\text{résidu})]$

Pour le modèle (2)

L'application de la formule (5) au modèle (2), où le facteur primaire est fixe (facteur opérateur) et le facteur secondaire est aléatoire (échantillon), permet d'obtenir l'espérance des carrés moyens :

- $E(CM(A)) = \sigma_\varepsilon^2 + n\sigma_\beta^2 + \frac{qn}{p-1} \sum_{i=1}^p (\alpha_i - \bar{\alpha})^2$,
- $E(CM(B(A))) = \sigma_\varepsilon^2 + n\sigma_\beta^2$, se lit « B dans A »,
- $E(CM(\text{résidu})) = \sigma_\varepsilon^2$.

On en déduit les estimations des variances σ_β^2 et σ_ε^2 :

- $\hat{\sigma}_\varepsilon^2 = \frac{1}{pq(n-1)} \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n (y_{ijk} - \bar{y}_{ij\bullet})^2$, estimation usuelle de σ_ε^2 ,
- $\hat{\sigma}_\beta^2 = [CM(B(A)) - CM(\text{résidu})]/n$,

De même, on aboutit à l'estimation des variances de répétabilité et de reproductibilité :

- $\hat{\sigma}_r^2 = \hat{\sigma}_\varepsilon^2$
- $\hat{\sigma}_R^2 = \hat{\sigma}_\beta^2$
- $\hat{\sigma}_{r\&R}^2 = \hat{\sigma}_r^2 + \hat{\sigma}_R^2$

7. Intervalles de confiance des variances σ_r^2 , σ_R^2 et $\sigma_{r\&R}^2$

On peut les calculer en utilisant l'approximation de Satterthwaite.

Approximation de Satterthwaite

On dispose de k carrés moyens CM_1, \dots, CM_k basés sur les degrés de liberté respectifs f_1, \dots, f_k . On considère la combinaison linéaire : $Q = q_1 CM_1 + \dots + q_k CM_k$. On a alors :

$$\frac{rQ}{E(Q)} \sim \chi^2(r)$$

avec :

$$r = \frac{Q^2}{\sum_{i=1}^k \frac{(q_i CM_i)^2}{f_i}}$$

d'où l'on déduit un intervalle de confiance de $E(Q)$ à 95% :

$$\Pr \left[\frac{rQ}{\chi_{0.975}^2(r)} \leq E(Q) \leq \frac{rQ}{\chi_{0.025}^2(r)} \right] = 0.95$$

Par exemple, l'application de l'approximation de Satterthwaite au modèle (1) conduit à :

L'intervalle de confiance de σ_r^2 :

$$\left[\frac{pq(n-1)\hat{\sigma}_\varepsilon^2}{\chi_{0.975}^2(pq(n-1))}; \frac{pq(n-1)\hat{\sigma}_\varepsilon^2}{\chi_{0.025}^2(pq(n-1))} \right]$$

L'intervalle de confiance de σ_R^2 :

Calculons d'abord r :

$$r = \frac{(\hat{\sigma}_R^2)^2}{\left[\frac{\left(\frac{1}{np} CM(B) \right)^2}{q-1} + \frac{\left(\frac{p-1}{np} CM(A \times B) \right)^2}{(p-1)(q-1)} + \frac{\left(\frac{1}{n} CM(residu) \right)^2}{pq(n-1)} \right]}$$

d'où l'intervalle cherché :

$$\left[\frac{r\hat{\sigma}_R^2}{\chi_{0.975}^2(r)}; \frac{r\hat{\sigma}_R^2}{\chi_{0.025}^2(r)} \right]$$

L'intervalle de confiance de $\sigma_{r\&R}^2$:

Calculons d'abord r :

$$r = \frac{(\hat{\sigma}_{r\&R}^2)^2}{\left[\frac{\left(\frac{1}{np} CM(B) \right)^2}{q-1} + \frac{\left(\frac{p-1}{np} CM(A \times B) \right)^2}{(p-1)(q-1)} + \frac{\left(\frac{n-1}{n} CM(residu) \right)^2}{pq(n-1)} \right]}$$

d'où l'intervalle cherché :

$$\left[\frac{r\hat{\sigma}_{r\&R}^2}{\chi_{0.975}^2(r)}, \frac{r\hat{\sigma}_{r\&R}^2}{\chi_{0.025}^2(r)} \right].$$

Partie pratique

8. Présentation de la procédure VARCOMP

La procédure VARCOMP estime les composantes de la variance dans un modèle linéaire général où au moins un effet aléatoire est présent. Plus précisément, elle estime la contribution de chacun des effets aléatoires à la variance de la variable dépendante (la réponse). Les effets étudiés doivent être à modalités discrètes : facteurs qualitatifs, de classe, de groupe ou facteurs continus discrétisés. On peut étudier aussi leurs interactions et des formes emboîtées. Enfin, les données peuvent ne pas être équilibrées. Pour le codage la matrice du plan d'expérience X ne contient que des zéros et des 1, aucune reparamétrisation n'est effectuée.

Quatre méthodes d'estimation existent dans cette procédure :

- Méthode TYPE1 : elle est basée sur le calcul des sommes de carrés de type I (type séquentiel), pour chaque effet. Les variances sont estimées en égalant chaque somme de carrés de type I, notée $R(\beta_i | \beta_1, \dots, \beta_{i-1})$, pour β_i **aléatoire**, à son espérance, puis en résolvant le système linéaire obtenu.
- Méthode MIVQUE0 : c'est une amélioration de la méthode précédente. Les effets aléatoires sont ajustés seulement par rapport aux effets fixes, c'est la méthode par défaut de la procédure. Pour cette méthode et la précédente les variances estimées peuvent parfois être négatives compte tenu de certains aspects pathologiques dans les données (outliers, très grande variabilité, ...).
- Méthode du Maximum de Vraisemblance (ML) : elle calcule les estimations des composantes de la variance qui maximisent la vraisemblance en utilisant la transformation W de Hemmerle & Hartley [1973]. Dans un deuxième temps la méthode est itérative.

- Méthode du Maximum de Vraisemblance Restreint (REML) : c'est une méthode similaire à la précédente mais en premier lieu elle sépare la vraisemblance en deux parties, l'une contenant les effets fixes, l'autre les effets aléatoires. Pour cette méthode et la précédente les variances estimées ne peuvent pas être négatives car l'optimisation de la fonction objective se déroule sous la contrainte que les variances soient bornées inférieurement par zéro. De plus, les méthodes ML et REML fournissent des estimations des variances et covariances des estimations des variances.

Les instructions de la procédure VARCOMP

```
PROC VARCOMP DATA= SAS-data-set
```

```
    EPSILON= n → par défaut 10-8
```

```
    MAXITER= n → par défaut 50
```

```
    METHOD= TYPE1 | MIVQUE0 | ML | REML;
```

```
    CLASS variable-list;                                /* required */
```

```
    MODEL dependents=effects / option;                  /* required */
```

→ seule option : FIXED=n avec n le nombre d'effets

fixes apparaissant dans l'instruction MODEL (introduits avant les effets aléatoires).

```
    BY variable-list;
```

```
RUN ;
```

9. Présentation de la procédure NESTED

La procédure NESTED réalise l'analyse de variance et de covariance quand tous les effets du modèle postulé sont aléatoires et hiérarchisés. Si les données sont équilibrées le niveau de significativité des composantes est également fourni. Cette procédure peut sembler faire double emploi avec la précédente. Toutefois, elle est particulièrement optimisée pour des problèmes à nombreux facteurs emboîtés et nombreux niveaux.

Les instructions de la procédure NESTED

`PROC NESTED DATA=SAS-data-set options;` → *l'option possible est AOV (sortie uniquement de l'analyse de variance quand plusieurs réponses dépendantes présentes)*

```
    CLASS variable-list;                                /* required */
    VAR variable-list; → une ou plusieurs (analyse de covariance) variables dépendantes
    BY variable-list;
run ;
```

10.Exemples

Suivent quelques exemples variés et issus de problèmes réels (sauf le premier) qui illustrent l'utilisation des procédures VARCOMP et NESTED.

Exemple 1 :exemple artificiel (tiré de la documentation de SAS)

Ce premier exemple est un exemple artificiel ayant pour but de montrer les différences entre les 4 options de calcul : TYPE1, MIVQUE0, ML et REML.

Le code sas

```
data d0; input a b y @@; cards;
1 1 237 1 1 254 1 1 246
1 2 178 1 2 179
2 1 208 2 1 178 2 1 187
2 2 146 2 2 145 2 2 141
3 1 186 3 1 183
3 2 142 3 2 125 3 2 136
;

proc print;

proc varcomp data=d0 method=typel;
  class a b ;
  model y=a|b / fixed=1;
run;

proc varcomp data=d0 method=mivque0;
  class a b ;
  model y=a|b / fixed=1;
run;

proc varcomp data=d0 method=ml;
  class a b ;
  model y=a|b / fixed=1;
run;

proc varcomp data=d0 method=reml;
  class a b ;
  model y=a|b / fixed=1;
run;
```

→ l'effet a est fixe, b et l'interaction a×b sont aléatoires, pas de hiérarchie.

La sortie

OBS	A	B	Y
1	1	1	237
2	1	1	254
3	1	1	246
4	1	2	178
5	1	2	179
6	2	1	208
7	2	1	178
8	2	1	187
9	2	2	146
10	2	2	145
11	2	2	141
12	3	1	186
13	3	1	183
14	3	2	142
15	3	2	125
16	3	2	136

→ *Les données sont déséquilibrées.*

1/ Méthode TYPE1

Variance Components Estimation Procedure
Class Level Information

Class	Levels	Values
A	3	1 2 3
B	2	1 2

Number of observations in data set = 16

Variance Components Estimation Procedure

Dependent Variable: Y

Source	DF	Type I SS	Type I MS
A	2	11736.43750000	5868.21875000
B	1	11448.12564103	11448.12564103
A*B	2	299.04102564	149.52051282
Error	10	786.33333333	78.63333333
Corrected Total	15	24269.93750000	

Source	Expected Mean Square
A	Var(Error) + 2.725 Var(A*B) + 0.1 Var(B) + Q(A)
B	Var(Error) + 2.6308 Var(A*B) + 7.8 Var(B)
A*B	Var(Error) + 2.5846 Var(A*B)
Error	Var(Error)

→ *Les espérances ci-dessus sont obtenues par la formule (4).*

Variance Component	Estimate
Var(B)	1448.37683150
Var(A*B)	27.42658730
Var(Error)	78.63333333

2/ Méthode MIVQUE0

MIVQUE(0) Variance Component Estimation Procedure

SSQ Matrix

Source	B	A*B	Error	Y
B	60.84000000	20.52000000	7.80000000	89295.38000000
A*B	20.52000000	20.52000000	7.80000000	30181.30000000
Error	7.80000000	7.80000000	13.00000000	12533.50000000

Variance Component	Estimate
	Y
Var(B)	1466.12301587
Var(A*B)	-35.49170274
Var(Error)	105.73659674

3/ Méthode ML

Maximum Likelihood Variance Components Estimation Procedure

Dependent Variable: Y

Iteration	Objective	Var(B)	Var(A*B)	Var(Error)
0	78.38503712	1031.49069751	0	74.39097179
1	78.26370438	732.36064536	0	77.40116882
2	78.26354712	723.68674709	0	77.53017748
3	78.26354712	723.66583653	0	77.53049269

Convergence criteria met.

Asymptotic Covariance Matrix of Estimates

	Var(B)	Var(A*B)	Var(Error)
Var(B)	537826.14593	0	-107.3390452
Var(A*B)	0	0	0
Var(Error)	-107.3390452	0	858.71104234

4/ Méthode REML

Restricted Maximum Likelihood Variance Components Estimation Procedure

Dependent Variable: Y

Iteration	Objective	Var(B)	Var(A*B)	Var(Error)
0	63.41341449	1269.52701231	0	91.55811913
1	63.04468698	1601.84198824	32.76324172	76.93555625
2	63.03115305	1468.82931677	27.22581866	78.75482763
3	63.03112651	1464.33645861	26.95640530	78.84314765
4	63.03112651	1464.36727374	26.95885252	78.84238988

Convergence criteria met.

Asymptotic Covariance Matrix of Estimates

	Var(B)	Var(A*B)	Var(Error)
Var(B)	4401703.838	1.294	-273.397
Var(A*B)	1.294	3559.113	-502.852
Var(Error)	-273.397	-502.852	1249.699

→ Ces sorties seront commentées à l'occasion des exemples suivants.

Exemple 2 : Analyse labo 1

Ce deuxième exemple est classique des laboratoires d'analyse. C'est une application du modèle (1).

Le code sas

```
data d0; input num oper ech repet y ; cards;
```

→ donc : 3 opérateurs ($p=3$), 10 échantillons ($q=10$) et 2 répétitions.

1	1	10	1	0.60
2	1	4	1	0.85
3	1	9	1	1.00
4	1	3	1	0.85
5	1	8	1	0.85
6	1	2	1	1.00
7	1	6	1	1.00
8	1	7	1	0.95
9	1	1	1	0.65
10	1	5	1	0.55
11	1	4	2	0.95
12	1	9	2	1.00
13	1	6	2	1.00
14	1	10	2	0.70
15	1	2	2	1.00
16	1	5	2	0.45
17	1	3	2	0.80
18	1	8	2	0.80
19	1	7	2	0.95
20	1	1	2	0.60
21	2	7	1	0.95
22	2	3	1	0.80
23	2	4	1	0.80
24	2	8	1	0.75
25	2	1	1	0.55
26	2	10	1	0.55
27	2	5	1	0.40
28	2	9	1	1.00
29	2	6	1	1.00
30	2	2	1	1.05
31	2	3	2	0.75
32	2	5	2	0.40
33	2	7	2	0.90
34	2	4	2	0.75
35	2	6	2	1.05
36	2	2	2	0.95
37	2	1	2	0.55
38	2	8	2	0.70
39	2	10	2	0.50
40	2	9	2	0.95
41	3	3	1	0.80
42	3	10	1	0.85
43	3	2	1	1.05
44	3	6	1	1.00
45	3	9	1	1.05
46	3	8	1	0.80
47	3	4	1	0.80
48	3	7	1	0.95
49	3	1	1	0.50
50	3	5	1	0.45
51	3	5	2	0.50

52	3	10	2	0.80
53	3	6	2	1.05
54	3	9	2	1.05
55	3	3	2	0.80
56	3	1	2	0.55
57	3	4	2	0.80
58	3	8	2	0.80
59	3	2	2	1.00
60	3	7	2	0.95

```
proc varcomp data=d0 method=typel;
  class oper ech ;
  model y=oper|ech / fixed=1;
run;
```

→ 1/
→ l'opérateur est le facteur fixe

```
proc varcomp data=d0 method=reml;
  class oper ech ;
  model y=oper|ech / fixed=1;
run;
```

→ 2/

La sortie

1/

Variance Components Estimation Procedure Class Level Information

Class	Levels	Values
OPER	3	1 2 3
ECH	10	1 2 3 4 5 6 7 8 9 10

Number of observations in data set = 60

Variance Components Estimation Procedure

Dependent Variable: Y

Source	DF	Type I SS	Type I MS
OPER	2	0.04800000	0.02400000
ECH	9	2.05870833	0.22874537
OPER*ECH	18	0.10366667	0.00575926
Error	30	0.03875000	0.00129167
Corrected Total	59	2.24912500	

→ la méthode TYPE1 fournit le calcul explicite des espérances :

Source	Expected Mean Square
OPER	Var(Error) + 2 Var(OPER*ECH) + Q(OPER)
ECH	Var(Error) + 2 Var(OPER*ECH) + 6 Var(ECH)
OPER*ECH	Var(Error) + 2 Var(OPER*ECH)
Error	Var(Error)

Variance Component	Estimate	
Var(ECH)	0.03716435	} → variance de reproductibilité $\hat{\sigma}_R^2 = 0.0394$
Var(OPER*ECH)	0.00223380	
Var(Error)	0.00129167	→ variance de répétabilité $\hat{\sigma}_r^2$
		↓
		variance totale $\hat{\sigma}_{r\&R}^2 = 0.0407$

On déduit ainsi :

- l'intervalle de répétabilité $\hat{I}_r : \pm 2\sqrt{2}\hat{\sigma}_r = \pm 0.10165$
- l'intervalle de reproductibilité $\hat{I}_R : \pm 2\sqrt{2}\hat{\sigma}_{r\&R} = \pm 0.5706$

et en utilisant l'approximation de Satterthwaite on peut calculer les intervalles de confiance suivants (non fournis par la méthode TYPE1) :

- l'intervalle de confiance de $\hat{\sigma}_r^2$:

$$\left[\frac{30 \times 0.00129167}{\chi_{0.975}^2(30)} ; \frac{30 \times 0.00129167}{\chi_{0.025}^2(30)} \right] = [0.000824 ; 0.00231]$$

- *l'intervalle de confiance de $\hat{\sigma}_R^2$:*

on a :

$$r = \frac{(0.0394)^2}{\left[\frac{\left(\frac{1}{2 \times 3} \times 0.2287\right)^2}{10-1} + \frac{\left(\frac{3-1}{2 \times 3} \times 0.00575926\right)^2}{(3-1)(10-1)} + \frac{\left(\frac{1}{2} \times 0.00129167\right)^2}{3 \times 10 \times (2-1)} \right]} = 9.605$$

d'où l'intervalle :

$$\left[\frac{9.605 \times 0.0394}{\chi_{0.975}^2(9.605)} ; \frac{9.605 \times 0.0394}{\chi_{0.025}^2(9.605)} \right] = [0.0190 ; 0.1250]$$

- *l'intervalle de confiance de $\hat{\sigma}_{r\&R}^2$:*

on a :

$$r = \frac{(0.04069)^2}{\left[\frac{\left(\frac{1}{2 \times 3} \times 0.2287\right)^2}{10-1} + \frac{\left(\frac{3-1}{2 \times 3} \times 0.00575926\right)^2}{(3-1)(10-1)} + \frac{\left(\frac{2-1}{2} \times 0.00129167\right)^2}{3 \times 10 \times (2-1)} \right]} = 10.24$$

d'où l'intervalle :

$$\left[\frac{10.24 \times 0.04069}{\chi_{0.975}^2(10.24)} ; \frac{10.24 \times 0.04069}{\chi_{0.025}^2(10.24)} \right] = [0.020 ; 0.1232]$$

2/

Restricted Maximum Likelihood Variance Components Estimation Procedure

Dependent Variable: Y

Iteration	Objective	Var(ECH)	Var(OPER*ECH)	Var(Error)
0	-305.65600491	0.03716435	0.00223380	0.00129167
1	-305.65600491	0.03716435	0.00223380	0.00129167

Convergence criteria met.

Asymptotic Covariance Matrix of Estimates

	Var(ECH)	Var(OPER*ECH)	Var(Error)
Var(ECH)	0.0003230928	-.0000003071	0.0000000000
Var(OPER*ECH)	-.0000003071	0.0000009492	-.0000000556
Var(Error)	0.0000000000	-.0000000556	0.0000001112

→ Avec la méthode REML, l'approximation de Satterthwaite n'est pas utilisée.

D'où l'on déduit :

L'intervalle de confiance de $\hat{\sigma}_r^2$:

$$0.00129167 \pm 2 \times \sqrt{1.112 \times 10^{-7}} = [0.000625; 0.001959]$$

NB : l'intervalle obtenu est un peu différent de celui obtenu plus haut avec l'approximation de Satterthwaite.

L'intervalle de confiance de $\hat{\sigma}_{\text{échant}}^2$:

$$0.03716435 \pm 2 \times \sqrt{3.23 \times 10^{-4}} = [0.00122; 0.0731]$$

→ $\hat{\sigma}_{\text{échant}}^2$ est significativement différente de zéro.

L'intervalle de confiance de $\hat{\sigma}_{\text{oper} \times \text{échant}}^2$:

$$0.0022338 \pm 2 \times \sqrt{9.492 \times 10^{-7}} = [0.000285; 0.00418]$$

→ $\hat{\sigma}_{\text{oper} \times \text{échant}}^2$ est significativement différente de zéro.

Comparaison des variances

Par exemple comparons la variance de répétabilité à la variance de l'échantillon avec le test :

$$\begin{cases} H_0 : \sigma_r^2 = \sigma_{\text{échant}}^2 \\ H_1 : \sigma_r^2 \neq \sigma_{\text{échant}}^2 \end{cases}$$

On a :

$$\begin{aligned}\text{Var}(\hat{\sigma}_r^2 - \hat{\sigma}_{\text{échant.}}^2) &= \text{vâr}(\hat{\sigma}_r^2) + \text{vâr}(\hat{\sigma}_{\text{échant.}}^2) - 2\text{côv}(\hat{\sigma}_r^2, \hat{\sigma}_{\text{échant.}}^2) \\ &= 1.11 \times 10^{-7} + 3.23 \times 10^{-4} - 0 \approx 3.23 \times 10^{-4}\end{aligned}$$

On construit ensuite la variable réduite :

$$z = \frac{\hat{\sigma}_r^2 - \hat{\sigma}_{\text{échant.}}^2}{\text{écart-type}(\hat{\sigma}_r^2 - \hat{\sigma}_{\text{échant.}}^2)} \quad \text{qui suit sous } H_0 \text{ une } N(0,1).$$

Ici :

$$z = \frac{0.00129167 - 0.03716}{0.01797} = -1.996$$

→ on rejette donc H_0 si on choisit un risque de première espèce de 5% : la variance de l'échantillon est significativement supérieure à la variance de répétabilité.

Comparaison avec la procédure GLM de SAS

Le code serait :

```
proc glm ;  
  class oper ech repet ;  
  model y=oper ech oper*ech ;  
  random oper oper*ech/test
```

→ c'est cette option « test » qui permet d'avoir les tests corrects

```
run ;
```

Et la sortie correspondante (extrait) :

General Linear Models Procedure

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	29	2.21037500	0.07621983	59.01	0.0001
Error	30	0.03875000	0.00129167		
Corrected Total	59	2.24912500			

R-Square	C.V.	Root MSE	Y Mean
0.982771	4.450745	0.0359398	0.8075000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
OPER	2	0.04800000	0.02400000	18.58	0.0001
ECH	9	2.05870833	0.22874537	177.09	0.0001
OPER*ECH	18	0.10366667	0.00575926	4.46	0.0002



Faux car ce n'est pas la bonne variance d'erreur qui est choisie pour faire le test

General Linear Models Procedure

Source	Type III Expected Mean Square
OPER	Var(Error) + 2 Var(OPER*ECH) + Q(OPER)
ECH	Var(Error) + 2 Var(OPER*ECH) + 6 Var(ECH)
OPER*ECH	Var(Error) + 2 Var(OPER*ECH)

→ L'option « test » dans l'instruction « random » donne les bons tests :

General Linear Models Procedure
Tests of Hypotheses for Mixed Model Analysis of Variance

Dependent Variable: Y

Source: OPER
Error: MS(OPER*ECH)

DF	Type III MS	Denominator DF	Denominator MS	F Value	Pr > F
2	0.024	18	0.0057592593	4.1672	0.0326

Source: ECH
Error: MS(OPER*ECH)

DF	Type III MS	Denominator	Denominator	F Value	Pr > F
		DF	MS		
9	0.2287453704	18	0.0057592593	39.7178	0.0001

Source: OPER*ECH
Error: MS(Error)

DF	Type III MS	Denominator	Denominator	F Value	Pr > F
		DF	MS		
18	0.0057592593	30	0.0012916667	4.4588	0.0002

Exemple 3 : Analyse labo 2

C'est le même exemple que l'exemple 2 mais les données sont maintenant déséquilibrées.

Le code sas

```
data d0; input num oper ech repet y ; cards;
```

```
1      1      10      1      0.60
2      1       4      1      0.85
3      1       9      1      .
4      1       3      1      0.85
5      1       8      1      .
6      1       2      1      .
7      1       6      1      1.00
8      1       7      1      0.95
9      1       1      1      0.65
10     1       5      1      0.55
11     1       4      2      .
12     1       9      2      1.00
13     1       6      2      1.00
14     1      10      2      0.70
15     1       2      2      1.00
16     1       5      2      0.45
17     1       3      2      0.80
18     1       8      2      0.80
19     1       7      2      0.95
20     1       1      2      0.60
21     2       7      1      0.95
22     2       3      1      0.80
23     2       4      1      .
24     2       8      1      0.75
25     2       1      1      .
26     2      10      1      .
27     2       5      1      0.40
28     2       9      1      1.00
29     2       6      1      1.00
30     2       2      1      1.05
31     2       3      2      0.75
32     2       5      2      0.40
33     2       7      2      0.90
34     2       4      2      .
35     2       6      2      1.05
36     2       2      2      0.95
37     2       1      2      0.55
38     2       8      2      .
39     2      10      2      .
40     2       9      2      0.95
41     3       3      1      0.80
42     3      10      1      0.85
43     3       2      1      .
44     3       6      1      1.00
45     3       9      1      1.05
46     3       8      1      0.80
47     3       4      1      0.80
48     3       7      1      .
49     3       1      1      0.50
50     3       5      1      0.45
51     3       5      2      0.50
52     3      10      2      .
53     3       6      2      1.05
54     3       9      2      .
55     3       3      2      0.80
56     3       1      2      0.55
```

57	3	4	2	0.80
58	3	8	2	.
59	3	2	2	1.00
60	3	7	2	0.95

```
proc varcomp data=d0 method=type1;
  class oper ech ;
  model y=oper|ech / fixed=1;
run;
```

La sortie

Variance Components Estimation Procedure

Dependent Variable: Y

Source	DF	Type I SS	Type I MS
OPER	2	0.00675149	0.00337574
ECH	9	1.68909498	0.18767722
OPER*ECH	16	0.04990353	0.00311897
Error	17	0.02625000	0.00154412
Corrected Total	44	1.77200000	

→ Le déséquilibre des données conduit à des calculs d'espérances impossibles à faire à la main :

Source	Expected Mean Square
OPER	Var(Error) + 1.7591 Var(OPER*ECH) + 0.2258 Var(ECH) + Q(OPER)
ECH	Var(Error) + 1.6676 Var(OPER*ECH) + 4.414 Var(ECH)
OPER*ECH	Var(Error) + 1.5449 Var(OPER*ECH)
Error	Var(Error)

Variance Component	Estimate
Var(ECH)	0.04178348
Var(OPER*ECH)	0.00101940
Var(Error)	0.00154412

Exemple 4 : Vitesse réaction

(aménagé à partir d'un exemple de la documentation de SAS)

On cherche à détecter les sources de variabilité d'une vitesse d'une réaction biologique (la réponse). L'expérimentateur cherche à savoir si c'est la souche étudiée ou bien le laboratoire qui joue le plus sur la variance de la réponse. Un plan d'expériences est mené : le facteur fixe est le facteur « TEMPERATURE » qui prend les trois niveaux 145-155-165 ; les facteurs aléatoires sont le « LABORATOIRE » et la « SOUCHE ». Trois laboratoires sont tirés au hasard et trois souches, tirées au hasard parmi de nombreuses souches possibles, sont testées à chaque combinaison température-laboratoire avec 4 répétitions. Le facteur « SOUCHE » est hiérarchisé par rapport cette dernière combinaison.

Le code sas

```
data vit ;
  input lab temp souche $ vit @@ ;
  cards ;
1 145 A 18.6 1 145 A 17.0 1 145 A 18.7 1 145 A 18.7
1 145 B 14.5 1 145 B 15.8 1 145 B 16.5 1 145 B 17.6
1 145 C 21.1 1 145 C 20.8 1 145 C 21.8 1 145 C 21.0
1 155 A 9.50 1 155 A 9.40 1 155 A 9.50 1 155 A 10.0
1 155 B 7.80 1 155 B 8.30 1 155 B 8.90 1 155 B 9.10
1 155 C 11.2 1 155 C 10.0 1 155 C 11.5 1 155 C 11.1
1 165 A 5.40 1 165 A 5.30 1 165 A 5.70 1 165 A 5.30
1 165 B 5.20 1 165 B 4.90 1 165 B 4.30 1 165 B 5.20
1 165 C 6.30 1 165 C 6.40 1 165 C 5.80 1 165 C 5.60
2 145 A 20.0 2 145 A 20.1 2 145 A 19.4 2 145 A 20.0
2 145 B 18.4 2 145 B 18.1 2 145 B 16.5 2 145 B 16.7
2 145 C 22.5 2 145 C 22.7 2 145 C 21.5 2 145 C 21.3
2 155 A 11.4 2 155 A 11.5 2 155 A 11.4 2 155 A 11.5
2 155 B 10.8 2 155 B 11.1 2 155 B 9.50 2 155 B 9.70
2 155 C 13.3 2 155 C 14.0 2 155 C 12.0 2 155 C 11.5
2 165 A 6.80 2 165 A 6.90 2 165 A 6.00 2 165 A 5.70
2 165 B 6.00 2 165 B 6.10 2 165 B 5.00 2 165 B 5.20
2 165 C 7.70 2 165 C 8.00 2 165 C 6.60 2 165 C 6.30
3 145 A 19.7 3 145 A 18.3 3 145 A 16.8 3 145 A 17.1
3 145 B 16.3 3 145 B 16.7 3 145 B 14.4 3 145 B 15.2
3 145 C 22.7 3 145 C 21.9 3 145 C 19.3 3 145 C 19.3
3 155 A 9.30 3 155 A 10.2 3 155 A 9.80 3 155 A 9.50
3 155 B 9.10 3 155 B 9.20 3 155 B 8.00 3 155 B 9.00
3 155 C 11.3 3 155 C 11.0 3 155 C 10.9 3 155 C 11.4
3 165 A 6.70 3 165 A 6.00 3 165 A 5.00 3 165 A 4.80
3 165 B 5.70 3 165 B 5.50 3 165 B 4.60 3 165 B 5.40
3 165 C 6.60 3 165 C 6.50 3 165 C 5.90 3 165 C 5.80
;

proc varcomp method=reml ;
  class temp lab souche ;
  model vit = temp|lab souche(lab temp) / fixed=1 ;
run ;
```

La sortie

Variance Components Estimation Procedure Class Level Information

Class	Levels	Values
TEMP	3	145 155 165
LAB	3	1 2 3
SOUCHE	3	A B C

Number of observations in data set = 108

Restricted Maximum Likelihood Variance Components Estimation Procedure

Dependent Variable: VIT

Iteration	Objective	Var(LAB)	Var(TEMP*LAB)	Var(SOUCHE(TEMP*LAB))	Var(Error)
0	13.45000603	0.50944643	0	2.40048886	0.57871852
1	13.08982622	0.31943483	0	2.08696369	0.60160053
2	13.08931256	0.31760480	0	2.07389061	0.60262172
3	13.08931256	0.31760171	0	2.07386855	0.60262346

Convergence criteria met.

Asymptotic Covariance Matrix of Estimates

	Var(LAB)	Var(TEMP*LAB)
Var(LAB)	0.3245202664	0
Var(TEMP*LAB)	0	0
Var(SOUCHE(TEMP*LAB))	-0.049984938	0
Var(Error)	1.025912E-12	0

	Var(SOUCHE(TEMP*LAB))	Var(Error)
Var(LAB)	-0.049984938	1.025912E-12
Var(TEMP*LAB)	0	0
Var(SOUCHE(TEMP*LAB))	0.4504248653	-0.002241698
Var(Error)	-0.002241698	0.0089667909

→ *En conclusion la part de variabilité attribuable aux souches pour chaque combinaison température x laboratoire est beaucoup plus grande (2.074) que la part attribuable aux laboratoires eux-mêmes (0.318). En priorité, il est donc plus important de considérer la variabilité des souches plutôt qu'incriminer les différences de pratiques et d'appareillage entre laboratoires. En outre, la variance expérimentale n'étant que de 0.603, la variabilité due aux souches joue un rôle majeur sur la variabilité globale de la réponse.*

Exemple 5 : Mycotoxine

(aménagé à partir de l'exemple « Soufre » de la Norme ISO 5725-2)

C'est un exemple d'expérience de fidélité entre laboratoires. On cherche à déterminer la quantité d'une mycotoxine M chez un rat. Huit laboratoires prélèvent indépendamment 4 échantillons d'organe correspondant chacun à un organe précis. Sur chaque échantillon chaque laboratoire réalise au moins 3 mesures. On étudie l'influence du statut des facteurs sur les résultats.

Le code sas

```
data d0; input num labo organe myco; cards;
```

1	1	1	0.71
2	1	1	0.71
3	1	1	0.70
4	1	1	0.71
5	1	2	1.20
6	1	2	1.18
7	1	2	1.23
8	1	2	1.21
9	1	3	1.68
10	1	3	1.70
11	1	3	1.68
12	1	3	1.69
13	1	4	3.26
14	1	4	3.26
15	1	4	3.20
16	1	4	3.24
17	2	1	0.69
18	2	1	0.67
19	2	1	0.68
20	2	2	1.22
21	2	2	1.21
22	2	2	1.22
23	2	3	1.64
24	2	3	1.64
25	2	3	1.65
26	2	4	3.20
27	2	4	3.20
28	2	4	3.20
29	3	1	0.66
30	3	1	0.65
31	3	1	0.69
32	3	2	1.28
33	3	2	1.31
34	3	2	1.30
35	3	3	1.61
36	3	3	1.61
37	3	3	1.62
38	3	4	3.37
39	3	4	3.36
40	3	4	3.38
41	4	1	0.67

42	4	1	0.65
43	4	1	0.66
44	4	2	1.23
45	4	2	1.18
46	4	2	1.20
47	4	3	1.68
48	4	3	1.66
49	4	3	1.66
50	4	4	3.16
51	4	4	3.22
52	4	4	3.23
53	5	1	0.70
54	5	1	0.69
55	5	1	0.66
56	5	1	0.71
57	5	1	0.69
58	5	2	1.31
59	5	2	1.22
60	5	2	1.22
61	5	2	1.24
62	5	2	.
63	5	3	1.64
64	5	3	1.67
65	5	3	1.60
66	5	3	1.66
67	5	3	1.68
68	5	4	3.20
69	5	4	3.19
70	5	4	3.18
71	5	4	3.27
72	5	4	3.24
73	6	1	0.73
74	6	1	0.74
75	6	1	0.73
76	6	2	1.39
77	6	2	1.36
78	6	2	1.37
79	6	3	1.70
80	6	3	1.73
81	6	3	1.73
82	6	4	3.27
83	6	4	3.31
84	6	4	3.29
85	7	1	0.71
86	7	1	0.71
87	7	1	0.69
88	7	2	1.20
89	7	2	1.26
90	7	2	1.26
91	7	3	1.69
92	7	3	1.70
93	7	3	1.68
94	7	4	3.27
95	7	4	3.24
96	7	4	3.23
97	8	1	0.70
98	8	1	0.65
99	8	1	0.68
100	8	2	1.24
101	8	2	1.22
102	8	2	1.30
103	8	3	1.67
104	8	3	1.68
105	8	3	1.67
106	8	4	3.25
107	8	4	3.26
108	8	4	3.26

;

1/ Pas d'effet fixe

```
proc varcomp data=d0 method=typel;  
  class labo organe ;  
  model myco=labo|organe ;  
run;
```

```
proc varcomp data=d0 method=reml;  
  class labo organe ;  
  model myco=labo|organe ;  
run;
```

2/ Laboratoire = effet fixe

```
proc varcomp data=d0 method=typel;  
  class labo organe ;  
  model myco=labo|organe /fixed=1;  
run;
```

```
proc varcomp data=d0 method=reml;  
  class labo organe ;  
  model myco=labo|organe /fixed=1;  
run;
```

3/ Organe = effet fixe

```
proc varcomp data=d0 method=typel;  
  class labo organe ;  
  model myco=organe|labo /fixed=1;  
run;
```

```
proc varcomp data=d0 method=reml;  
  class labo organe ;  
  model myco=organe|labo /fixed=1;  
run;
```

4/ Laboratoire et Organe = effets fixes, pas d'effet aléatoire

```
proc glm data=d0 ;  
  class labo organe ;  
  model myco=labo organe labo*organe ;  
  
run;
```

Les sorties

1/ Pas d'effet fixe

Variance Components Estimation Procedure
Class Level Information

Class	Levels	Values
LABO	8	1 2 3 4 5 6 7 8
ORGANE	4	1 2 3 4

Number of observations in data set = 108

NOTE: Due to missing values, only 107 observations can be used in this analysis.

Variance Components Estimation Procedure

Dependent Variable: MYCO

Source	DF	Type I SS	Type I MS
LABO	7	0.07844150	0.01120593
ORGANE	3	97.56336304	32.52112101
LABO*ORGANE	21	0.09222398	0.00439162
Error	75	0.03771167	0.00050282
Corrected Total	106	97.77174019	

Source	Expected Mean Square
LABO	$\text{Var}(\text{Error}) + 3.3317 \text{Var}(\text{LABO} * \text{ORGANE}) + 0.0046 \text{Var}(\text{ORGANE}) + 13.308 \text{Var}(\text{LABO})$
ORGANE	$\text{Var}(\text{Error}) + 3.4614 \text{Var}(\text{LABO} * \text{ORGANE}) + 26.737 \text{Var}(\text{ORGANE})$
LABO*ORGANE	$\text{Var}(\text{Error}) + 3.3251 \text{Var}(\text{LABO} * \text{ORGANE})$
Error	$\text{Var}(\text{Error})$

Variance Component	Estimate
Var(LABO)	0.00008763
Var(ORGANE)	1.21617092
Var(LABO* ORGANE)	0.00116954
Var(Error)	0.00050282

NORME ISO 5725-2 : $\hat{\sigma}_r^2 = 4.84 \times 10^{-4}$ et $\hat{\sigma}_{r\&R}^2 = 2.023 \times 10^{-3}$

VARCOMP : $\hat{\sigma}_r^2 = 5.03 \times 10^{-4}$ et $\hat{\sigma}_{r\&R}^2 = 1.218$

Restricted Maximum Likelihood Variance Components Estimation Procedure

Dependent Variable: MYCO

Iteration	Objective	Var(LABO)	Var(ORGANE)	Var(LABO*ORGANE)	Var(Error)
0	-661.57341196	0	0.14857766	0.00081007	0.00111972
1	-670.08929918	0.00071704	0.09945264	0.00187756	0.00074854
.
15	-703.04493989	0.00061052	1.20972728	0.00122631	0.00050362

Convergence criteria met.

Asymptotic Covariance Matrix of Estimates

	Var(LABO)	Var(ORGANE)
Var(LABO)	2.7394759E-7	-1.828012E-9
Var(ORGANE)	-1.828012E-9	0.9759050454
Var(LABO*ORGANE)	-4.586087E-8	1.3055828E-8
Var(Error)	-2.4973E-11	-3.593629E-9

	Var(LABO*ORGANE)	Var(Error)
Var(LABO)	-4.586087E-8	-2.4973E-11
Var(ORGANE)	1.3055828E-8	-3.593629E-9
Var(LABO*ORGANE)	1.8413182E-7	-2.262114E-9
Var(Error)	-2.262114E-9	6.7852021E-9

2/ Laboratoire = effet fixe

Variance Components Estimation Procedure

Dependent Variable: MYCO

Source	DF	Type I SS	Type I MS
LABO	7	0.07844150	0.01120593
ORGANE	3	97.56336304	32.52112101
LABO*ORGANE	21	0.09222398	0.00439162
Error	75	0.03771167	0.00050282
Corrected Total	106	97.77174019	

Source	Expected Mean Square
LABO	Var(Error) + 3.3317 Var(LABO*ORGANE) + 0.0046 Var(ORGANE) + Q(LABO)
NIVEAU	Var(Error) + 3.4614 Var(LABO*ORGANE) + 26.737 Var(ORGANE)
LABO*ORGANE	Var(Error) + 3.3251 Var(LABO*ORGANE)
Error	Var(Error)

Variance Component	Estimate
Var(ORGANE)	1.21617092
Var(LABO*ORGANE)	0.00116954
Var(Error)	0.00050282

NORME ISO 5725-2 : $\hat{\sigma}_r^2 = 4.84 \times 10^{-4}$ et $\hat{\sigma}_{r\&R}^2 = 2.023 \times 10^{-3}$

VARCOMP : $\hat{\sigma}_r^2 = 5.03 \times 10^{-4}$ et $\hat{\sigma}_{r\&R}^2 = 1.218$

Restricted Maximum Likelihood Variance Components Estimation Procedure

Dependent Variable: MYCO

Iteration	Objective	Var(ORGANE)	Var(LABO*ORGANE)	Var(Error)
0	-638.22792427	0.11828436	0.00071221	0.00095300
1	-639.78256409	0.10664922	0.00111550	0.00085777
.
14	-672.42684743	1.20973139	0.00122626	0.00050350

Convergence criteria met.

Asymptotic Covariance Matrix of Estimates

	Var(ORGANE)	Var(LABO*ORGANE)	Var(Error)
Var(ORGANE)	0.9759116627	1.3014801E-8	-3.533385E-9
Var(LABO*ORGANE)	1.3014801E-8	1.8410999E-7	-2.261261E-9
Var(Error)	-3.533385E-9	-2.261261E-9	6.7786886E-9

3/ ORGANE = effet fixe , Laboratoire = effet aléatoire

Cette situation correspond à celle de la norme ISO 5725-2, mais elle n'est pas précisée clairement dans le document de cette norme.

Variance Components Estimation Procedure

Dependent Variable: MYCO

Source	DF	Type I SS	Type I MS
ORGANE	3	97.55916754	32.51972251
LABO	7	0.08263701	0.01180529
LABO*ORGANE	21	0.09222398	0.00439162
Error	75	0.03771167	0.00050282
Corrected Total	106	97.77174019	

Source	Expected Mean Square
ORGANE	Var(Error) + 3.4653 Var(LABO*ORGANE) + 0.0074 Var(LABO) + Q(ORGANE)
LABO	Var(Error) + 3.33 Var(LABO*ORGANE) + 13.305 Var(LABO)
LABO*ORGANE	Var(Error) + 3.3251 Var(LABO*ORGANE)
Error	Var(Error)

Variance Component	Estimate
Var(LABO)	0.00055676
Var(LABO*ORGANE)	0.00116954
Var(Error)	0.00050282

→ On retombe sur des résultats assez proches de ceux donnés dans la Norme ISO 5725-2 à la différence près que le terme d'interaction n'est pas pris en compte dans le calcul de la reproductibilité dans la Norme, et pourtant ce terme d'interaction est significatif (voir plus loin).

NORME ISO 5725-2 : $\hat{\sigma}_r^2 = 4.84 \times 10^{-4}$ et $\hat{\sigma}_{r\&R}^2 = 2.023 \times 10^{-3}$

VARCOMP : $\hat{\sigma}_r^2 = 5.03 \times 10^{-4}$ et $\hat{\sigma}_{r\&R}^2 = 2.229 \times 10^{-3}$

Restricted Maximum Likelihood Variance Components Estimation Procedure

Dependent Variable: MYCO

Iteration	Objective	Var(LABO)	Var(LABO*ORGANE)	Var(Error)
0	-711.57355533	0.00041650	0.00087740	0.00057514
1	-713.42848345	0.00064354	0.00130010	0.00049452
2	-713.47558677	0.00061116	0.00122849	0.00050335
3	-713.47562523	0.00061052	0.00122638	0.00050361
4	-713.47562524	0.00061051	0.00122635	0.00050362

Convergence criteria met.

Asymptotic Covariance Matrix of Estimates

	Var(LABO)	Var(LABO*ORGANE)	Var(Error)
Var(LABO)	2.7394907E-7	-4.586361E-8	-2.51067E-11
Var(LABO*ORGANE)	-4.586361E-8	1.8414249E-7	-2.261416E-9
Var(Error)	-2.51067E-11	-2.261416E-9	6.7849556E-9

→ *L'intervalle de confiance de $\sigma_{LABO \times ORGANE}^2$ vaut $[3.6810^{-4}; 2.08 \times 10^{-3}]$.*

4/ Laboratoire et ORGANE = effets fixes, pas d'effet aléatoire

General Linear Models Procedure

Dependent Variable: MYCO

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	31	97.73402852	3.15271060	6270.03	0.0001
Error	75	0.03771167	0.00050282		
Corrected Total	106	97.77174019			
	R-Square	C.V.	Root MSE	SOUFRE Mean	
	0.999614	1.304270	0.0224237	1.7192523	

Source	DF	Type III SS	Mean Square	F Value	Pr > F
LABO	7	0.08233043	0.01176149	23.39	0.0001
ORGANE	3	94.72499170	31.57499723	62795.55	0.0001
LABO*ORGANE	21	0.09222398	0.00439162	8.73	0.0001

→ Si on considère tous les effets fixes, ceux-ci apparaissent significatifs comme on le voit ci-dessus ; cela revient à mettre en évidence les différents facteurs de biais de la méthode et non pas les sources de la variance de reproductibilité. Dans certains cas favorables c'est-à-dire à nombres de répétitions suffisants pour chaque combinaison de facteurs, on pourrait envisager de tester l'homogénéité des variances de répétition, par exemple avec un test de Bartlett.

Exemple 6 : Calcium

(tiré de la documentation de la procédure NESTED)

On illustre ici l'utilisation de la procédure NESTED. Tous les effets sont aléatoires.

Le code sas

```
DATA TURNIP;
  DO PLANT=1 TO 4;
    DO LEAF=1 TO 3;
      DO SAMPLE=1 TO 2;
        INPUT CALCIUM @@;
        OUTPUT;
      END;
    END;
  END;
CARDS;
3.28 3.09 3.52 3.48 2.88 2.80
2.46 2.44 1.87 1.92 2.19 2.19
2.77 2.66 3.74 3.44 2.55 2.55
3.78 3.87 4.07 4.12 3.31 3.31
;
PROC NESTED;
  CLASSES PLANT LEAF;
  VAR CALCIUM;
RUN;
```

La sortie

Coefficients of Expected Mean Squares

Source	PLANT	LEAF	ERROR
PLANT	6	2	1
LEAF	0	2	1
ERROR	0	0	1

Nested Random Effects Analysis of Variance for Variable CALCIUM

Variance Source	Degrees of Freedom	Sum of Squares	F Value	Pr > F	Error Term
TOTAL	23	10.270396			
PLANT	3	7.560346	7.665	0.0097	<u>LEAF</u>
LEAF	8	2.630200	49.409	0.0000	<u>ERROR</u>
ERROR	12	0.079850			

Variance Source	Mean Square	Variance Component	Percent of Total
TOTAL	0.446539	0.532938	100.0000
PLANT	2.520115	0.365223	68.5302
LEAF	0.328775	0.161060	30.2212
ERROR	0.006654	0.006654	1.2486
	Mean		3.01208333
	Standard error of mean		0.32404445

→ Il apparaît une variabilité significative (différente de zéro) de plante à plante ($Pr > F$ inférieure à 0.0001). Remarquons aussi que la composante de la variance relative à la plante est testée en utilisant le carré moyen associé à « leaf » et non pas le terme d'erreur résiduelle, c'est-à-dire :

$$7.665 = \frac{7.560346 / 3}{0.07985 / 12} .$$

Conclusion

Ce document ne prétend évidemment pas être exhaustif ni sur les concepts de répétabilité et de reproductibilité, ni sur ceux, *a fortiori*, des composantes de la variance. Nous pensons qu'il constitue toutefois un guide pratique d'utilisation des procédures VARCOMP et NESTED, tout en redonnant des définitions simples mais rigoureuses de la répétabilité et la reproductibilité.