Quelques modèles d'analyse de la variance.[1]

1. Notations

1.1. La notation synthétique d'un total : application aux effectifs

On considère un tableau de données provenant de résultats expérimentaux effectués dans certaines conditions symbolisées par un ou plusieurs indices. Ces expériences peuvent être répétées ou non en fonction des indices que l'on considère, chacune des cases du tableau comportant ainsi une ou plusieurs données. On note l'effectif de chaque case du tableau en remplaçant y par n et en conservant les indices associés à y. La somme des effectifs suivant l'un des indices est notée en remplaçant cet indice par le symbole $+$. On adopte cette notation, classique en statistique, pour alléger et uniformiser les notations. Ainsi par exemple, dans les cas où nos données sont indexées par un, deux ou trois indices, on obtient les notations ci-dessous.

1. On a observé n valeurs de la variable Y. Chacune d'entre elle correspond à un résultat expérimental effectué dans la condition i, $1 \leqslant i \leqslant I$. On construit donc un tableau comportant I lignes. L'effectif de chaque case du tableau est alors noté n_i, l'effectif total n_+ est évidemment donné par la formule :
$$n_+ = \sum_{i=1}^{I} n_i.$$
Bien entendu on a $n_+ = n$.
2. On a observé n valeurs de la variable Y que l'on regroupe en fonction des valeurs des deux indices i, $1 \leqslant i \leqslant I$, et j, $1 \leqslant j \leqslant J$, associés aux conditions expérimentales. On construit ainsi un tableau à I lignes et J colonnes.
   - L'effectif de la case (i, j) du tableau, située à l'intersection de la $i-$ème ligne et de la $j-$ème colonne est alors noté $n_{i,j}$.
   - La somme des effectifs par rapport à l'indice i, c'est-à-dire le nombre de données dans la colonne j du tableau, est notée $n_{+,j}$.
$$n_{+,j} = \sum_{i=1}^{I} n_{i,j}.$$

[1] Les références [3], [4], [7], [5], [8], [11] ayant servi à l'élaboration de ce document sont mentionnées dans la bibliographie.

– La somme des effectifs par rapport à l'indice j, c'est-à-dire le nombre de données dans la ligne i du tableau, est notée $n_{i,+}$.

$$n_{i,+} = \sum_{j=1}^{J} n_{i,j}.$$

– La somme des effectifs par rapport aux indices i et j, c'est-à-dire le nombre de données dans le tableau, est notée $n_{+,+}$.

$$n_{+,+} = \sum_{i=1}^{I} \sum_{j=1}^{J} n_{i,j}.$$

Bien entendu on a $n_{+,+} = n$.

2. On a observé n valeurs de la variable Y que l'on regroupe en fonction des valeurs des trois indices i, $1 \leqslant i \leqslant I$, j, $1 \leqslant j \leqslant J$, et k, $1 \leqslant k \leqslant K$, associés aux conditions expérimentales. On construit ainsi un multitableau à I lignes, J colonnes et une profondeur de K.
   – L'effectif de la case (i, j, k) du multitableau, située à l'intersection de la $i-$ème ligne et de la $j-$ème colonne à la profondeur k est alors noté $n_{i,j,k}$.
   – La somme des effectifs par rapport à l'indice i, c'est-à-dire le nombre de données dans la colonne j à la profondeur k du multitableau, est notée $n_{+,j,k}$.

$$n_{+,j,k} = \sum_{i=1}^{I} n_{i,j,k}.$$

   – La somme des effectifs par rapport à l'indice j, c'est-à-dire le nombre de données dans la ligne i à la profondeur k du multitableau, est notée $n_{i,+,k}$.

$$n_{i,+,k} = \sum_{j=1}^{J} n_{i,j,k}.$$

   – La somme des effectifs par rapport à l'indice i, c'est-à-dire le nombre de données dans les cases transversales, c'est-à-dire sur toute la profondeur du tableau, situées aux intersections de la ligne i et de la colonne j du multitableau, est notée $n_{i,j,+}$.

$$n_{i,j,+} = \sum_{k=1}^{K} n_{i,j,k}.$$

   – La somme des effectifs par rapport aux indices i et j, c'est-à-dire le nombre de données dans la tranche de profondeur k du multitableau, est notée $n_{+,+,k}$.

$$n_{+,+,k} = \sum_{i=1}^{I} \sum_{j=1}^{J} n_{i,j,k}.$$

– La somme des effectifs par rapport aux indices i et k, c'est-à-dire le nombre de données dans la tranche verticale de position j du multitableau, est notée $n_{+,j,+}$.

$$n_{+,j,+} = \sum_{i=1}^{I} \sum_{k=1}^{K} n_{i,j,k}.$$

– La somme des effectifs par rapport aux indices j et k, c'est-à-dire le nombre de données dans la tranche horizontale de position i du multitableau, est notée $n_{i,+,+}$.

$$n_{i,+,+} = \sum_{j=1}^{J} \sum_{k=1}^{K} n_{i,j,k}.$$

– La somme des effectifs par rapport aux indices i, j et k, c'est-à-dire le nombre de données dans le tableau, est notée $n_{+,+,+}$.

$$n_{+,+,+} = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} n_{i,j}.$$

Bien entendu on a $n_{+,+,+} = n$.

On peut bien sûr généraliser cette notation lorsque l'on considère plus de trois indices.

1.2. La notation synthétique d'une moyenne

Lorsque l'on dispose d'un tableau de données indexées par un ou plusieurs indices on note la moyenne par rapport à l'un de ces indices en le remplaçant par le symbole $\bullet$. On adopte cette notation, classique en statistique, pour alléger et uniformiser les notations. Ainsi par exemple, dans les cas où nos données sont indexées par un, deux ou trois indices, on obtient les notations ci-dessous.

1. On a observé $n = I$ valeurs de la variable Y indexée par un indice i, notées $y_1, \ldots, y_n$. La moyenne de ces valeurs $y_1, \ldots, y_n$ par rapport à l'indice i est donc notée $y_\bullet$. Dans ce cas il s'agit simplement de la moyenne $\overline{y}$:

$$y_\bullet = \frac{1}{I} \sum_{i=1}^{I} y_i = \frac{1}{I} y_+.$$

2. On a observé $n = I \times J$ valeurs de la variable Y indexée par deux indices i et j, notées $y_{1,1}, \ldots, y_{1,J}$, $y_{2,1}, \ldots, y_{2,J}, \ldots, y_{I,J}, \ldots, y_{I,J}$.
   – La moyenne de ces valeurs $y_{1,1}, \ldots, y_{I,J}$ par rapport à l'indice i est notée $y_{\bullet,j}$. Il s'agit simplement de la moyenne $\overline{y_j}$ des valeurs de la $j-$ème colonne du tableau :

$$y_{\bullet,j} = \frac{1}{I} \sum_{i=1}^{I} y_{i,j} = \frac{1}{I} y_{+,j}.$$

– La moyenne de ces valeurs $y_{1,1}, \ldots, y_{I,J}$ par rapport à l'indice j est notée $y_{i,\bullet}$. Il s'agit simplement de la moyenne $\overline{y_i}$ des valeurs de la $i-$ème ligne :

$$y_{i,\bullet} = \frac{1}{J}\sum_{j=1}^{J} y_{i,j} = \frac{1}{J} y_{i,+}.$$

– La moyenne de ces valeurs $y_{1,1}, \ldots, y_{I,J}$ par rapport aux indices i et j est notée $y_{\bullet,\bullet}$. Il s'agit simplement de la moyenne $\overline{y}$, parfois aussi appelée grande moyenne du tableau :

$$y_{\bullet,\bullet} = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} y_{i,j} = \frac{1}{IJ} y_{+,+}.$$

3. On a observé $n = I \times J \times K$ valeurs de la variable Y indexée par trois indices i, j et k, notées $y_{1,1,1}, \ldots, y_{1,1,K}$, $y_{2,1,1}, \ldots, y_{2,1,K}$, $\ldots$,$y_{I,1,1}, \ldots, y_{I,1,K}, \ldots, y_{1,2,1}, \ldots, y_{1,2,K}, \ldots$, $y_{I,J,1}, \ldots, y_{I,J,K}$.
   – La moyenne de ces valeurs $y_{1,1,1}, \ldots, y_{I,J,K}$ par rapport à l'indice i est notée $y_{\bullet,j,k}$:

$$y_{\bullet,j,k} = \frac{1}{I}\sum_{i=1}^{I} y_{i,j,k} = \frac{1}{I} y_{+,j,k}.$$

   – La moyenne de ces valeurs $y_{1,1,1}, \ldots, y_{I,J,K}$ par rapport à l'indice j est notée $y_{i,\bullet,k}$:

$$y_{i,\bullet,k} = \frac{1}{J}\sum_{j=1}^{J} y_{i,j,k} = \frac{1}{J} y_{i,+,k}.$$

   – La moyenne de ces valeurs $y_{1,1,1}, \ldots, y_{I,J,K}$ par rapport à l'indice k est notée $y_{i,j,\bullet}$:

$$y_{i,j,\bullet} = \frac{1}{K}\sum_{k=1}^{K} y_{i,j,k} = \frac{1}{K} y_{i,j,+}.$$

   – La moyenne de ces valeurs $y_{1,1,1}, \ldots, y_{I,J,K}$ par rapport aux indices i et j est notée $y_{\bullet,\bullet,k}$:

$$y_{\bullet,\bullet,k} = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} y_{i,j,k} = \frac{1}{IJ} y_{+,+,k}.$$

   – La moyenne de ces valeurs $y_{1,1,1}, \ldots, y_{I,J,K}$ par rapport aux indices i et k est notée $y_{\bullet,j,\bullet}$:

$$y_{\bullet,j,\bullet} = \frac{1}{IK}\sum_{i=1}^{I}\sum_{k=1}^{K} y_{i,j,k} = \frac{1}{IK} y_{+,j,+}.$$

– La moyenne de ces valeurs $y_{1,1,1}, \ldots, y_{I,J,K}$ par rapport aux indices j et k est notée $y_{i,\bullet,\bullet}$:

$$y_{i,\bullet,\bullet} = \frac{1}{JK} \sum_{j=1}^{J} \sum_{k=1}^{K} y_{i,j,k} = \frac{1}{JK} y_{i,+,+}.$$

– La moyenne de ces valeurs $y_{1,1,1}, \ldots, y_{I,J,K}$ par rapport aux indices i, j et k, est notée $y_{\bullet,\bullet,\bullet}$. Il s'agit simplement de la moyenne $\overline{y}$, parfois aussi appelée grande moyenne du tableau :

$$y_{\bullet,\bullet,\bullet} = \frac{1}{IJK} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} y_{i,j,k} = \frac{1}{IJK} y_{+,+,+}.$$

On étend cette notation à la situation où les valeurs que la variable Y va prendre ne sont pas encore connues, c'est-à-dire lorsque l'on remplace les nombres y_i, $y_{i,j}$ ou $y_{i,j,k}$ par les variables aléatoires Y_i, $Y_{i,j}$ ou $Y_{i,j,k}$. Ceci permettra d'obtenir des formules simples pour les estimateurs des paramètres des modèles étudiés par la suite.

On peut bien sûr généraliser cette notation lorsque les valeurs prises par la variable Y dépendent de plus de trois indices.

2. Introduction

2.1. Un peu de vocabulaire

i. On appelle **réponse** une variable dont on cherche à comprendre le comportement. On considèrera dans le cadre de ce cours des réponses quantitatives continues. Une réponse est généralement notée Y et c'est le cas systématiquement dans la suite.

ii. On appelle **facteur** toute variable dont on veut se servir pour analyser les variations de la réponse. On note généralement les facteurs par des lettres romaines majuscules A, B, ... Dans le contexte de l'analyse de la variance tous les facteurs sont considérés comme des variables qualitatives. Les niveaux ou modalités d'un facteur sont notés en indiçant la majuscule romaine associée au facteur : A_1, ..., A_I où I désigne le nombre total de modalités du facteur A. Pour l'analyse d'une réponse à l'aide de variables quantitatives on utilisera des techniques de régression, linéaire ou non, et pour l'analyse d'une réponse à l'aide d'un mélange de variables quantitatives et qualitatives on pourra se servir de l'analyse de la covariance.

iii. Un **modèle statistique** est une équation reliant les différentes mesures $Y_{i,j,\ldots}$ de la réponse Y aux effets des niveaux $A_1, \ldots, A_I$, $B_1, \ldots, B_J$, ... des facteurs A, B, ... pour lesquels ont été réalisés ces différentes mesures au travers d'une relation

fonctionnelle f et tout en modélisant les fluctuations expérimentales à l'aide de variables aléatoires d'erreur généralement notée $\epsilon_{i,j,\ldots}$ sur lesquelles on fait porter certaines hypothèses :

$$Y_{i,j,\ldots} = f(A_i, B_j, \ldots) + \epsilon_{i,j,\ldots}.$$

iv. Les modèles associés à l'analyse de la variance ont une forme particulière : ce sont des modèles linéaires. Dans le cas la relation fonctionnelle f s'écrit simplement comme une somme de différents **termes**. Ainsi on souhaite par exemple étudier une réponse continue Y à l'aide d'un facteur qualitatif A à effets fixes, un nombre identique J de répétitions ayant été effectuées pour chacun des niveaux du facteur.

On introduit le modèle :

$$Y_{i,j} = \mu + \alpha_i + \epsilon_{i,j}, \quad i = 1 \ldots I,\ j = 1 \ldots J,$$
$$\text{avec la contrainte supplémentaire } \sum_{i=1}^{I} \alpha_i = 0,$$

où $Y_{i,j}$ est la valeur prise par la réponse Y dans la condition A_i lors de la $j-$ème répétition. On postule les hypothèses classiques suivantes pour les erreurs :

$$\forall\ (i,j), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ \mathcal{L}(\epsilon_{i,j}) = \mathcal{N}(0, \sigma^2),$$
$$\mathrm{Cov}(\epsilon_{i,j}, \epsilon_{k,l}) = 0 \text{ si } (i,j) \neq (k,l) \text{ avec } 1 \leqslant i,k \leqslant I, \text{ et } 1 \leqslant j,l \leqslant J.$$

Les μ, α_1, $\ldots$, α_I sont donc des **termes du modèle**. On dit parfois que $\epsilon_{i,j}$ est le terme d'erreur dans le modèle.

On remarque que les hypothèses faites sur les termes d'erreur $\epsilon_{i,j}$ font partie intégrante de la définiton du modèle. Ceci explique le soin particulier que l'on doit avoir pour **vérifier que les conditions d'utilisation des modèles sont bien remplies**, la **validité des conclusions obtenues** à l'aide du modèle dépendant fortement de son utilisation opportune ou non.

v. En fonction du type d'effet auquel les termes du modèle sont associés, ces termes ont un nom différent.

- Un terme associé à l'effet direct d'un des facteurs du modèle est appelé **effet principal**, on le note généralement par la lettre grecque minuscule associée à la majuscule romaine désignant le facteur, par exemple pour le facteur A on notera α_i l'effet du niveau A_i, pour $1 \leqslant i \leqslant I$.
- L'ensemble des termes associés aux effets de l'interaction de deux des facteurs du modèle est appelé **interaction d'ordre 1**. On les note généralement par les lettres grecques minuscules mises entre parenthèses associées aux majuscules romaines désignant les deux facteurs dont on étudie l'interaction et indicées par les niveaux des deux facteurs. Par exemple si l'interaction d'ordre 2 met en jeu les facteurs A et B, on note les termes associés à ces effets par $(\alpha\beta)_{1,1}$, $(\alpha\beta)_{1,2}$, $\ldots$, $(\alpha\beta)_{1,I}$, $\ldots(\alpha\beta)_{2,1}$, $\ldots(\alpha\beta)_{i,j}$, $\ldots(\alpha\beta)_{I,J}$ pour $1 \leqslant i \leqslant I$ et $1 \leqslant j \leqslant J$.

- Plus généralement l'ensemble des termes associés aux effets de l'interaction de k des facteurs du modèle, $k \geqslant 2$, est appelé **interaction d'ordre $k-1$**. On les note généralement par les lettres grecques minuscules mises entre parenthèses associées aux majuscules romaines désignant les k facteurs dont on étudie l'interaction et indicées par les niveaux de ces k facteurs. Par exemple $(\alpha\beta\gamma)_{i,j,k}$ pour l'interaction d'ordre $k-1$ entre les facteurs A, B et C.
  Ainsi on souhaite par exemple étudier une réponse continue Y à l'aide de trois facteurs qualitatifs à effets fixes A, B et C pour lesquels on a fait un nombre de répétitions identiques, noté K, pour chaque combinaison des niveaux des trois facteurs. Le facteur A a I modalités, le facteur B a J modalités et le facteur C a K modalités.
  On introduit le modèle :

$$Y_{i,j,k,l} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{i,j} + (\alpha\gamma)_{i,k} + (\beta\gamma)_{j,k} + (\alpha\beta\gamma)_{i,j,k} + \epsilon_{i,j,k,l},$$
$$i = 1 \dots I,\ j = 1 \dots J,\ k = 1 \dots K,\ l = 1 \dots L,$$

avec les contraintes supplémentaires $\displaystyle\sum_{i=1}^{I} \alpha_i = 0,\ \sum_{j=1}^{J} \beta_j = 0$ et $\displaystyle\sum_{k=1}^{K} \gamma_k = 0,$

$$\sum_{i=1}^{I} (\alpha\beta)_{i,j} = 0,\ \forall j \in \{1, \dots, J\} \text{ et } \sum_{j=1}^{J} (\alpha\beta)_{i,j} = 0,\ \forall i \in \{1, \dots, I\},$$
$$\sum_{i=1}^{I} (\alpha\gamma)_{i,k} = 0,\ \forall k \in \{1, \dots, K\} \text{ et } \sum_{k=1}^{K} (\alpha\gamma)_{i,k} = 0,\ \forall i \in \{1, \dots, I\},$$
$$\sum_{j=1}^{J} (\beta\gamma)_{j,k} = 0,\ \forall k \in \{1, \dots, K\} \text{ et } \sum_{k=1}^{K} (\beta\gamma)_{j,k} = 0,\ \forall j \in \{1, \dots, J\},$$
$$\sum_{i=1}^{I} (\alpha\beta\gamma)_{i,j,k} = 0,\ \forall (j,k) \in \{1, \dots, J\} \times \{1, \dots, K\},$$
$$\sum_{j=1}^{J} (\alpha\beta\gamma)_{i,j,k} = 0,\ \forall (i,k) \in \{1, \dots, I\} \times \{1, \dots, K\},$$
$$\sum_{k=1}^{K} (\alpha\beta\gamma)_{i,j,k} = 0,\ \forall (i,j) \in \{1, \dots, I\} \times \{1, \dots, J\},$$

où $Y_{i,j,k,l}$ est la valeur prise par la réponse Y dans la condition (A_i, B_j, C_k) lors de la $l-$ème répétition. On postule les hypothèses classiques suivantes pour les erreurs :

$$\forall\ (i,j,k,l), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K,\ 1 \leqslant l \leqslant L,\ \mathcal{L}(\epsilon_{i,j,k,l}) = \mathcal{N}(0, \sigma^2),$$
$$\mathrm{Cov}(\epsilon_{i,j,k,l}, \epsilon_{m,n,o,p}) = 0 \text{ si } (i,j,k,l) \neq (m,n,o,p)$$
$$\text{avec } 1 \leqslant i, m \leqslant I,\ 1 \leqslant j, n \leqslant J,\ 1 \leqslant k, o \leqslant K \text{ et } 1 \leqslant l, p \leqslant L.$$

Les termes α_i, β_j et γ_k sont donc les **effets principaux** associés aux trois facteurs A, B et C. Les termes $(\alpha\beta)_{i,j}$ sont associés aux **interactions d'ordre 2** entre les facteurs A et B, les termes $(\alpha\gamma)_{i,k}$ sont associés aux **interactions d'ordre 2** entre les facteurs A et C, les termes $(\beta\gamma)_{j,k}$ sont associés aux **interactions d'ordre 2** entre les facteurs B et C, les termes $(\alpha\beta\gamma)_{i,j,k}$ sont associés à **l'interaction d'ordre 3** entre les facteurs A, B et C.
On remarque qu'à nouveau les hypothèses faites sur les termes d'erreur $\epsilon_{i,j,k,l}$ font partie intégrante de la définiton du modèle. Ceci explique le soin particulier que l'on doit avoir pour **vérifier que les conditions d'utilisation des modèles sont bien remplies**, **la validité des conclusions obtenues** à l'aide du modèle dépendant fortement de son utilisation opportune ou non.

- Dans certaines situations particulières, celles où **l'un ou plusieurs des facteurs sont dits emboîtés dans un autre facteur**, on introduira d'autres notations qui seront décrites à la section 5 et à la section 7.

2.2. Plans équilibrés, plans déséquilibrés

De manière générale, tout dispositif expérimental comportant au moins un facteur et comportant un nombre identique de répétitions dans chacune des modalités des facteurs est un plan équilibré. Dans la suite on appelera plan équilibré tout plan vérifiant cette condition.[2]

Pour de multiples raisons on conseille, lors de la phase de planification expérimentale, de prévoir d'utiliser des plans équilibrés pour l'analyse des effets ou des interactions d'un ou plusieurs facteurs qualitatifs sur une réponse expérimentale modélisée par une variable continue. En particulier dans le cas où le plan est équilibré :

1. Les estimations des paramètres d'un modèle où les facteurs sont à effets fixes seront faciles à calculer puisque l'on est dans une situation que l'on dénomme en statistique par plan d'expérience orthogonal. Concrètement ceci signifie que pour estimer l'effet d'un des termes du modèle, il n'est pas nécessaire d'estimer les effets des autres termes du modèle. Ceci se traduit dans les formules par le fait que l'estimation des effets des termes du modèle se fait comme si le modèle ne comportait pas d'autres termes que celui que l'on cherche à estimer.
2. Lorsque l'on ne dispose d'aucune information a priori sur les possibles résultats de l'expérience, la situation la plus propice à mettre en évidence une différence, si elle

[2]Cette condition est plus restrictive que celle que l'on peut définir plus généralement en statistique pour étudier des modèles d'analyse de la variance. On peut aussi facilement traiter le cas où le plan est déséquilibré mais à effectifs proportionnels. Par exemple, dans le cas où l'on considère deux facteurs, un dispositif expérimental déséquilibré à effectifs proportionnels est tel que :

$$n_{i,j} = \frac{n_{i,+} \times n_{+,j}}{n_{+,+}},$$

où les notattions $n_{i,j}$, $n_{i,+}$, $n_{+,j}$ et $n_{+,+}$ ont été définies au paragraphe 1.1. Cette notion se généralise directement au cas où le plan comporte k facteurs avec $k \geqslant 2$.

existe, entre les différents effets pour lesquels des tests sont réalisés est associée à une répartition des essais où le plan équilibré. On renvoie à la section 11 pour un exposé plus détaillé du calcul de la puissance associée aux différents tests réalisés lors d'une analyse de la variance.

Néanmoins il est possible d'utiliser tous les modèles exposés dans les sections suivantes dans le contexte particulier des plans équilibrés lorsqu'en fait le plan est déséquilibré, c'est-à-dire lorsque le nombre de répétitions varie d'une case à l'autre du tableau de données. **En aucun cas on ne supprimera des observations pour se ramener à une situation où le plan est équilibré.** Il y a deux faits qu'il faut absolument avoir à l'esprit lorsque l'on se sert de ce type de modèles, donc pour lesquels le plan est déséquilibré.

1. D'une part, l'estimation des effets des différents niveaux des facteurs ou de leurs interactions n'est plus aussi simple que précédemment,sauf dans le cas d'un modèle d'analyse de la variance à un facteur ou lorsque le plan est dit à effectifs proportionnels. En effet, ces estimations ne sont plus indépendantes les unes des autres. Voir par exemple la situation étudiée dans le paragraphe 12.2.5 où l'on insiste sur les différences qui découlent de cette situation. On renvoie à [4] pour plus de détails.
2. D'autre part, les lois des statistiques de test que l'on utilise pour tester les différentes hypothèses présentes dans le tableau de l'analyse de la variance ne sont, dans la majorité des cas, connues qu'approximativement. On renvoie à nouveau à [4] pour plus de détails.

Dans la suite, on se bornera à exposer la théorie de l'analyse de la variance dans le cadre des plans équilibrés, c'est-à-dire dans le cas où le nombre de répétitions est constant d'une case à l'autre du tableau.

2.3. Risque global associé au tableau de l'analyse de la variance

Lorsque l'on réalise une étude de la significativité des effets de deux ou de plus de facteurs, en s'intéressant ou non à la significativité de leurs interactions, avec un modèle d'analyse de la variance, le tableau de l'analyse de la variance que l'on étudie permet de prendre plusieurs décisions, chacune associée au test d'une hypothèse nulle. On prend ainsi en fait plusieurs décisions chacune d'entre elles étant associées à un risque de première ou de seconde espèce. Pour plus de détails sur les méthodes de gestion du risque de première espèce associée à la lecture de tableaux récapitulatifs de tests statistiques voir le chapitre de cours consacré à ce problème.

Exemple 2.1. Analyse de la variance à deux facteurs à effets fixes, aléatoires ou mixtes et avec répétitions

Pour un exposé complet du modèle d'analyse de la variance à deux facteurs contrôlés avec répétitions voir le paragraphe 4.1.2.
Avec un tel modèle il est possible de tester trois hypothèses nulles, $\mathcal{H}_0^{'}$ d'absence d'effet du premier facteur, $\mathcal{H}_0^{''}$ d'absence d'effet du second facteur, $\mathcal{H}_0^{'''}$ d'absence d'effet de l'interaction des deux facteurs. Chacune d'elles peut être testée avec un risque différent, $\alpha^{'}$, $\alpha^{''}$ et $\alpha^{'''}$. Le risque global α vérifiera alors la relation :

$$\alpha \leqslant 1 - (1-\alpha^{'})(1-\alpha^{''})(1-\alpha^{'''}) \leqslant \alpha^{'} + \alpha^{''} + \alpha^{'''}.$$

La dernière inégalité est une application de **l'inégalité de Bonferonni**. Ainsi si l'on veut prendre une décision associée à un risque global de $\alpha = 5$ %, on pourra fixer un seuil individuel de $\alpha^{'} = \alpha^{''} = \alpha^{'''} = 5/3$ % $\approx 2,67$ % pour chacun des tests élémentaires de chaque hypothèse $\mathcal{H}_0^{'}$, $\mathcal{H}_0^{''}$ et $\mathcal{H}_0^{'''}$.

L'exemple précédent met en évidence le fait que, plus le tableau de l'analyse de la variance étudié comporte de tests, plus la correction à apporter au niveau individuel de significativité de chacun des tests réalisés sera importante. Cette correction vise à réduire le nombre de faux positifs, effets jugés à tort comme significatifs, que l'on a détectés. Elle est d'autant plus intéressante à faire que le nombre de tests réalisés est important.

Remarque 2.1. La nécessité de l'utilisation de ce type de correction dépend de la formulation des résultats qui est faite par l'utilisateur. Si celui-ci s'intéresse à un risque global portant sur l'ensemble des décisions prises à l'aide d'un modèle d'analyse de la variance il privilégiera une maîrise du risque global. Si l'utilisateur s'intéresse au risque associé à chacune des décisions, il maîtrisera le risque individuel associé à chacune d'entre elles. Quoi qu'il advienne il faut savoir que l'on entre dans le domaine de l'analyse de tableaux de résultats. On pourra consulter [8] ou [10] pour plus de précisions.

3. Analyse de la variance à un facteur

Dans toute cette section, on utilise les notations définies à la section 1.

3.1. Modèles à effets fixes

Un facteur contrôlé A se présente sous I modalités, chacune d'entre elles étant notée A_i. Pour chacune des modalités on effectue $J \geqslant 2$ mesures d'une réponse Y qui est une variable continue. On note $n = I \times J$ le nombre total de mesures ayant été effectuées.

On introduit le modèle :

$$Y_{i,j} = \mu + \alpha_i + \epsilon_{i,j}, \quad i = 1 \dots I,\ j = 1 \dots J,$$
$$\text{avec la contrainte supplémentaire } \sum_{i=1}^{I} \alpha_i = 0,$$

où $Y_{i,j}$ est la valeur prise par la réponse Y dans la condition A_i lors de la $j-$ème répétition. On postule les hypothèses classiques suivantes pour les erreurs :

$$\forall\ (i,j), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ \mathcal{L}(\epsilon_{i,j}) = \mathcal{N}(0, \sigma^2),$$
$$\mathrm{Cov}(\epsilon_{i,j}, \epsilon_{k,l}) = 0 \text{ si } (i,j) \neq (k,l) \text{ avec } 1 \leqslant i,k \leqslant I, \text{ et } 1 \leqslant j,l \leqslant J.$$

On suppose que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fera l'objet d'un autre paragraphe.

On regroupe les valeurs que peut prendre la réponse Y dans la condition A_i lors de la $j-$ème répétition dans le tableau suivant :

Facteur A	Y
A_1	$Y_{1,1}, \dots, Y_{1,J}$
$\vdots$	$\vdots$
A_i	$Y_{i,1}, \dots, Y_{i,J}$
$\vdots$	$\vdots$
A_I	$Y_{I,1}, \dots, Y_{I,J}$

On rappelle que la variation théorique due au facteur A est définie par :

$$SC_F = J \sum_{i=1}^{I} (Y_{i,\bullet} - Y_{\bullet,\bullet})^2.$$

La variation résiduelle théorique est quant à elle définie par :

$$SC_R = \sum_{i=1}^{I} \left(\sum_{j=1}^{J} (Y_{i,j} - Y_{i,\bullet})^2 \right).$$

Enfin la variation totale théorique est égale à :

$$SC_{TOT} = \sum_{i=1}^{I} \sum_{j=1}^{J} (Y_{i,j} - Y_{\bullet,\bullet})^2.$$

On rappelle la relation fondamentale de l'ANOVA :

$$SC_{TOT} = SC_F + SC_R.$$

La liste $\boldsymbol{y}$ des données expérimentales $y_{1,1}, \ldots, y_{1,J}, \ldots, y_{2,1}, \ldots, y_{2,J}, \ldots, y_{I,J}$ permet de construire une réalisation du tableau précédent :

Facteur A	$\boldsymbol{y}$
A_1	$y_{1,1}, \ldots, y_{1,J}$
$\vdots$	$\vdots$
A_i	$y_{i,1}, \ldots, y_{i,J}$
$\vdots$	$\vdots$
A_I	$y_{I,1}, \ldots, y_{I,J}$

La variation due au facteur A observée sur la liste de données $\boldsymbol{y}$ est définie par :

$$sc_F = J \sum_{i=1}^{I} (y_{i,\bullet} - y_{\bullet,\bullet})^2.$$

La variation résiduelle observée sur la liste de données $\boldsymbol{y}$ est quant à elle définie par :

$$sc_R = \sum_{i=1}^{I} \left(\sum_{j=1}^{J} (y_{i,j} - y_{i,\bullet})^2 \right).$$

Enfin la variation totale observée sur la liste de données $\boldsymbol{y}$ est égale à :

$$sc_{TOT} = \sum_{i=1}^{I} \sum_{j=1}^{J} (y_{i,j} - y_{\bullet,\bullet})^2.$$

La relation fondamentale de l'ANOVA reste valable lorsqu'elle est évaluée sur la liste de données $\boldsymbol{y}$:

$$sc_{TOT} = sc_F + sc_R.$$

On résume ces informations dans le tableau de l'ANOVA ci-dessous :

Source	Variation	Degrés de liberté	Carré Moyen	F	Décision
Facteur	sc_F	$I-1$	$s_F^2 = \dfrac{sc_F}{I-1}$	$f = \dfrac{s_F^2}{s_R^2}$	$\mathcal{H}_0$ ou $\mathcal{H}_1$
Résiduelle	sc_R	$n-I = I(J-1)$	$s_R^2 = \dfrac{sc_R}{n-I}$		
Totale	sc_{TOT}	$n-1 = IJ-1$			

On souhaite faire le test d'hypothèse suivant :

$$\boxed{\mathcal{H}_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \text{Il existe } i_0 \in \{1, 2, \ldots, I\} \text{ tel que } \alpha_{i_0} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0$ précédente d'absence d'effet du facteur A et lorsque les conditions de validité du modèle sont respectées, f est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $I-1$ et $n-I$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f est supérieure ou égale à la valeur critique issue de la table. Lorsque l'hypothèse nulle $\mathcal{H}_0$ est rejetée on peut procéder à des comparaisons multiples des différents effets des niveaux du facteur voir la section 10.

Les estimateurs $\widehat{\mu}$, $\widehat{\alpha_1}$, $\ldots$, $\widehat{\alpha_I}$, $\widehat{\sigma^2}$ des paramètres μ, α_1, $\ldots$, α_I, σ^2 du modèle sont donnés par les formules suivantes :

$$\widehat{\mu} = Y_{\bullet,\bullet} = \overline{Y},\ \widehat{\alpha_i} = Y_{i,\bullet} - \widehat{\mu},\ 1 \leqslant i \leqslant I,$$
$$\widehat{\sigma^2} = \frac{SC_R}{n-I} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{i,j} - Y_{i,\bullet})^2}{n-I} = S_R^2.$$

Ce sont des estimateurs sans biais.

Les estimations, obtenues pour la liste de données $\boldsymbol{y}$ et notées $\widehat{\mu}(\boldsymbol{y})$, $\widehat{\alpha_1}(\boldsymbol{y})$, $\ldots$, $\widehat{\alpha_I}(\boldsymbol{y})$, $\widehat{\sigma^2}(\boldsymbol{y})$, des paramètres μ, α_1, $\ldots$, α_I, σ^2 du modèle se déduisent des formules ci-dessus :

$$\widehat{\mu}(\boldsymbol{y}) = y_{\bullet,\bullet} = \overline{y},\ \widehat{\alpha_i}(\boldsymbol{y}) = y_{i,\bullet} - \widehat{\mu}(\boldsymbol{y}),\ 1 \leqslant i \leqslant I,$$
$$\widehat{\sigma^2}(\boldsymbol{y}) = \frac{sc_R}{n-I} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J}(y_{i,j} - y_{i,\bullet})^2}{n-I} = s_R^2.$$

3.2. Modèles à effets aléatoires

Dans ce cas les A_i représentent un échantillon de taille I prélevé dans une population importante. Nous admettrons que les effets des A_i, les α_i, sont distribués suivant une loi normale centrée de variance σ_A^2. Le modèle ne dépend plus que de trois paramètres μ, σ^2 et σ_A^2. Pour chacune des modalités on effectue $J \geqslant 2$ mesures d'une réponse Y qui est une variable continue. On note $n = I \times J$ le nombre total de mesures ayant été effectuées.

On introduit le modèle :

$$Y_{i,j} = \mu + \alpha_i + \epsilon_{i,j}, \quad i = 1 \dots I,\ j = 1 \dots J,$$

où $Y_{i,j}$ est la valeur prise par la réponse Y dans la condition A_i lors de la $j-$ème répétition. On suppose que :

$$\mathcal{L}\left(\alpha_i\right) = \mathcal{N}(0, \sigma_A^2),\ \forall\ i, 1 \leqslant i \leqslant I,$$

ainsi que l'indépendance des effets aléatoires :

$$\mathrm{Cov}(\alpha_i, \alpha_j) = 0 \text{ si } i \neq j \text{ et } 1 \leqslant i, j \leqslant I.$$

On postule les hypothèses supplémentaires pour les erreurs :

$$\begin{gathered}\forall\ (i,j), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ \mathcal{L}(\epsilon_{i,j}) = \mathcal{N}(0, \sigma^2), \\ \mathrm{Cov}(\epsilon_{i,j}, \epsilon_{k,l}) = 0 \text{ si } (i,j) \neq (k,l) \text{ avec } 1 \leqslant i, k \leqslant I,\ \text{ et } 1 \leqslant j, l \leqslant J,\end{gathered}$$

ainsi que l'indépendance des effets aléatoires et des erreurs :

$$\mathrm{Cov}(\alpha_i, \epsilon_{j,k}) = 0 \text{ si } 1 \leqslant i, j \leqslant I,\ \text{ et } 1 \leqslant k \leqslant J.$$

On suppose que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fera l'objet d'un autre paragraphe.

On utilise les quantités SC_F, SC_R, SC_{TOT}, sc_F, sc_R, sc_{TOT} introduites à la section 3.1.

On rappelle la relation fondamentale de l'ANOVA :

$$SC_{TOT} = SC_F + SC_R.$$

On résume les informations dans le tableau de l'ANOVA ci-dessous :

Source	Variation	Degrés de liberté	Carré Moyen	F	Décision
Facteur	sc_F	$I-1$	$s_F^2 = \dfrac{sc_F}{I-1}$	$f = \dfrac{s_F^2}{s_R^2}$	$\mathcal{H}_0$ ou $\mathcal{H}_1$
Résiduelle	sc_R	$n-I = I(J-1)$	$s_R^2 = \dfrac{sc_R}{n-I}$		
Totale	sc_{TOT}	$n-1 = IJ-1$			

On souhaite faire le test d'hypothèse suivant :

$$\boxed{\mathcal{H}_0 : \sigma_A^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \sigma_A^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0$ précédente d'absence d'effet du facteur A et lorsque les conditions de validité du modèle sont respectées, f est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $I-1$ et $n-I$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f est supérieure ou égale à la valeur critique issue de la table.

Les estimateurs $\widehat{\mu}$, $\widehat{\sigma_A^2}$, $\widehat{\sigma^2}$ des paramètres μ, σ_A^2, σ^2 du modèle sont donnés par les formules suivantes :

$$\widehat{\mu} = Y_{\bullet,\bullet} = \overline{Y},\ \widehat{\sigma_A^2} = \frac{1}{J}\left(S_F^2 - S_R^2\right),\ \widehat{\sigma^2} = \frac{SC_R}{n-I} = S_R^2,$$

où $S_F^2 = \dfrac{SC_F}{I-1}$ et $S_R^2 = \dfrac{SC_R}{n-I}$.
Ce sont des estimateurs sans biais.

Les estimations, obtenues pour la liste de données $\boldsymbol{y}$ et notées $\widehat{\mu}(\boldsymbol{y})$, $\widehat{\sigma_A^2}(\boldsymbol{y})$, $\widehat{\sigma^2}(\boldsymbol{y})$, des paramètres μ, σ_A^2, σ^2 du modèle se déduisent des formules ci-dessus :

$$\widehat{\mu}(\boldsymbol{y}) = y_{\bullet,\bullet} = \overline{y},\ \widehat{\sigma_A^2}(\boldsymbol{y}) = \frac{1}{J}\left(s_F^2 - s_R^2\right),\ \widehat{\sigma^2}(\boldsymbol{y}) = \frac{sc_R}{n-I} = s_R^2.$$

4. Analyse de la variance à deux facteurs

Dans toute cette section, on utilise les notations définies à la section 1.

4.1. Modèles à effets fixes

4.1.1. Sans répétition

Un facteur contrôlé A se présente sous I modalités, chacune d'entre elles étant notée A_i. Un facteur contrôlé B se présente sous J modalités, chacune d'entre elles étant notée B_j. Pour chacun des couples de modalités (A_i, B_j) on effectue une mesure d'une réponse Y qui est une variable continue. On note $n = I \times J$ le nombre total de mesures ayant été effectuées.

On introduit le modèle :

$$Y_{i,j} = \mu + \alpha_i + \beta_j + \epsilon_{i,j}, \quad i = 1 \dots I,\ j = 1 \dots J,$$
$$\text{avec les contraintes supplémentaires } \sum_{i=1}^{I} \alpha_i = 0 \text{ et } \sum_{j=1}^{J} \beta_j = 0,$$

où $Y_{i,j}$ est la valeur prise par la réponse Y dans les conditions (A_i, B_j). On postule les hypothèses classiques suivantes pour les erreurs :

$$\forall\ (i,j), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ \mathcal{L}(\epsilon_{i,j}) = \mathcal{N}(0, \sigma^2),$$
$$\mathrm{Cov}(\epsilon_{i,j}, \epsilon_{k,l}) = 0 \text{ si } (i,j) \neq (k,l) \text{ avec } 1 \leqslant i,k \leqslant I \text{ et } 1 \leqslant j,l \leqslant J.$$

On suppose que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fera l'objet d'un autre paragraphe.

On regroupe les valeurs que peut prendre la réponse Y dans les conditions (A_i, B_j) dans le tableau suivant :

Facteur A	Facteur B				
	B_1	$\dots$	B_j	$\dots$	B_J
A_1	$Y_{1,1}$	$\dots$	$Y_{1,j}$	$\dots$	$Y_{1,J}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_i	$Y_{i,1}$	$\dots$	$Y_{i,j}$	$\dots$	$Y_{i,J}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_I	$Y_{I,1}$	$\dots$	$Y_{1,j}$	$\dots$	$Y_{I,J}$

On rappelle que la variation théorique due au facteur A est définie par :

$$SC_A = J \sum_{i=1}^{I} (Y_{i,\bullet} - Y_{\bullet,\bullet})^2.$$

On rappelle que la variation théorique due au facteur B est définie par :

$$SC_B = I \sum_{j=1}^{J} (Y_{\bullet,j} - Y_{\bullet,\bullet})^2.$$

La variation résiduelle théorique est quant à elle définie par :

$$SC_R = \sum_{i=1}^{I} \sum_{j=1}^{J} (Y_{i,j} - Y_{i,\bullet} - Y_{\bullet,j} + Y_{\bullet,\bullet})^2.$$

Enfin la variation totale théorique est égale à :

$$SC_{TOT} = \sum_{i=1}^{I} \sum_{j=1}^{J} (Y_{i,j} - Y_{\bullet,\bullet})^2.$$

On rappelle la relation fondamentale de l'ANOVA :

$$SC_{TOT} = SC_A + SC_B + SC_R.$$

La liste $\boldsymbol{y}$ des données expérimentales $y_{1,1}, \ldots, y_{1,J}, \ldots, y_{2,1}, \ldots, y_{2,J}, \ldots, y_{I,J}$ permet de construire une réalisation du tableau précédent :

Facteur A	Facteur B				
	B_1	$\ldots$	B_j	$\ldots$	B_J
A_1	$y_{1,1}$	$\ldots$	$y_{1,j}$	$\ldots$	$y_{1,J}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_i	$y_{i,1}$	$\ldots$	$y_{i,j}$	$\ldots$	$y_{i,J}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_I	$y_{I,1}$	$\ldots$	$y_{1,j}$	$\ldots$	$y_{I,J}$

La variation due au facteur A observée sur la liste de données $\boldsymbol{y}$ est définie par :

$$sc_A = J \sum_{i=1}^{I} (y_{i,\bullet} - y_{\bullet,\bullet})^2.$$

La variation due au facteur B observée sur la liste de données $\boldsymbol{y}$ est définie par :

$$sc_B = I \sum_{j=1}^{J} (y_{\bullet,j} - y_{\bullet,\bullet})^2.$$

La variation résiduelle observée sur la liste de données $\boldsymbol{y}$ est quant à elle définie par :

$$sc_R = \sum_{i=1}^{I} \sum_{j=1}^{J} (y_{i,j} - y_{i,\bullet} - y_{\bullet,j} + y_{\bullet,\bullet})^2.$$

Enfin la variation totale observée sur la liste de données $\boldsymbol{y}$ est égale à :

$$sc_{TOT} = \sum_{i=1}^{I} \sum_{j=1}^{J} (y_{i,j} - y_{\bullet,\bullet})^2.$$

La relation fondamentale de l'ANOVA reste valable lorsqu'elle est évaluée sur la liste de données $\boldsymbol{y}$:

$$sc_{TOT} = sc_A + sc_B + sc_R.$$

On introduit les dégres de liberté (Ddl) associés à chaque ligne du tableau de l'ANOVA :

Source	Degrés de liberté
Facteur A	$n_A = I - 1$
Facteur B	$n_B = J - 1$
Résiduelle	$n_R = (I-1)(J-1)$
Totale	$n_{TOT} = IJ - 1$

On résume ces informations dans le tableau de l'ANOVA ci-dessous :

Source	Variation	Ddl	Carré Moyen	F	Décision
Facteur A	sc_A	n_A	$s_A^2 = \dfrac{sc_A}{n_A}$	$f_A = \dfrac{s_A^2}{s_R^2}$	$\mathcal{H}_0^{'}$ ou $\mathcal{H}_1^{'}$
Facteur B	sc_B	n_B	$s_B^2 = \dfrac{sc_B}{n_B}$	$f_B = \dfrac{s_B^2}{s_R^2}$	$\mathcal{H}_0^{''}$ ou $\mathcal{H}_1^{''}$
Résiduelle	sc_R	n_R	$s_R^2 = \dfrac{sc_R}{n_R}$		
Totale	sc_{TOT}	n_{TOT}			

On souhaite faire les tests d'hypothèse suivants :

$$\boxed{\mathcal{H}_0^{'} : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0}$$

contre

$$\boxed{\mathcal{H}_1^{'} : \text{Il existe } i_0 \in \{1, 2, \ldots, I\} \text{ tel que } \alpha_{i_0} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{'}$ précédente d'absence d'effet du facteur A et lorsque les conditions de validité du modèle sont respectées, f_A est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $I-1$ et $(I-1)(J-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_A est supérieure ou égale à la valeur critique issue de la table. Lorsque l'hypothèse nulle $\mathcal{H}_0^{'}$ est rejetée on peut procéder à des comparaisons multiples des différents effets des niveaux du facteur voir la section 10.

$$\boxed{\mathcal{H}_0^{''} : \beta_1 = \beta_2 = \cdots = \beta_J = 0}$$

contre

$$\boxed{\mathcal{H}_1^{''} : \text{Il existe } j_0 \in \{1, 2, \ldots, J\} \text{ tel que } \beta_{j_0} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{''}$ précédente d'absence d'effet du facteur B et lorsque les conditions de validité du modèle sont respectées, f_B est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $J-1$ et $(I-1)(J-1)$ degrés de liberté. On conclut alors à

l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_B est supérieure ou égale à la valeur critique issue de la table. Lorsque l'hypothèse nulle $\mathcal{H}_0''$ est rejetée on peut procéder à des comparaisons multiples des différents effets des niveaux du facteur voir la section 10.

Les estimateurs $\widehat{\mu}$, $\widehat{\alpha_1}$, ..., $\widehat{\alpha_I}$, $\widehat{\beta_1}$, ..., $\widehat{\beta_J}$, $\widehat{\sigma^2}$ des paramètres μ, α_1, ..., α_I, β_1, ..., β_J, σ^2 du modèle sont donnés par les formules suivantes :

$$\widehat{\mu} = Y_{\bullet,\bullet} = \overline{Y},\ \widehat{\alpha_i} = Y_{i,\bullet} - \widehat{\mu},\ 1 \leqslant i \leqslant I,\ \widehat{\beta_j} = Y_{\bullet,j} - \widehat{\mu},\ 1 \leqslant j \leqslant J,$$
$$\widehat{\sigma^2} = \frac{SC_R}{(I-1)(J-1)} = s_R^2.$$

Ce sont des estimateurs sans biais.

Les estimations, obtenues pour la liste de données $\boldsymbol{y}$ et notées $\widehat{\mu}(\boldsymbol{y})$, $\widehat{\alpha_1}(\boldsymbol{y})$, ..., $\widehat{\alpha_I}(\boldsymbol{y})$, $\widehat{\beta_1}(\boldsymbol{y})$, ..., $\widehat{\beta_J}(\boldsymbol{y})$, $\widehat{\sigma^2}(\boldsymbol{y})$, des paramètres μ, α_1, ..., α_I, β_1, ..., β_J, σ^2 du modèle se déduisent des formules ci-dessus :

$$\widehat{\mu}(\boldsymbol{y}) = y_{\bullet,\bullet} = \overline{y},\ \widehat{\alpha_i}(\boldsymbol{y}) = y_{i,\bullet} - \widehat{\mu}(\boldsymbol{y}),\ 1 \leqslant i \leqslant I,\ \widehat{\beta_j}(\boldsymbol{y}) = y_{\bullet,j} - \widehat{\mu}(\boldsymbol{y}),\ 1 \leqslant j \leqslant J,$$
$$\widehat{\sigma^2}(\boldsymbol{y}) = \frac{sc_R}{(I-1)(J-1)} = s_R^2.$$

4.1.2. Avec répétitions

Un facteur contrôlé A se présente sous I modalités, chacune d'entre elles étant notée A_i. Un facteur contrôlé B se présente sous J modalités, chacune d'entre elles étant notée B_j. Pour chacun des couples de modalités (A_i, B_j) on effectue $K \geqslant 2$ mesures d'une réponse Y qui est une variable continue. On note $n = I \times J \times K$ le nombre total de mesures ayant été effectuées.

On introduit le modèle :

$$Y_{i,j,k} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{i,j} + \epsilon_{i,j,k}, \quad i = 1 \ldots I,\ j = 1 \ldots J,\ k = 1 \ldots K$$
$$\text{avec les contraintes supplémentaires } \sum_{i=1}^{I} \alpha_i = 0,\ \sum_{j=1}^{J} \beta_j = 0,$$
$$\sum_{i=1}^{I} (\alpha\beta)_{i,j} = 0,\ \forall j \in \{1, \ldots, J\} \text{ et } \sum_{j=1}^{J} (\alpha\beta)_{i,j} = 0,\ \forall i \in \{1, \ldots, I\},$$

où $Y_{i,j,k}$ est la valeur prise par la réponse Y dans les conditions (A_i, B_j) lors du $k-$ème essai. On postule les hypothèses classiques suivantes pour les erreurs :

$$\forall\ (i,j,k), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K,\ \mathcal{L}(\epsilon_{i,j,k}) = \mathcal{N}(0, \sigma^2),$$
$$\text{et } \mathrm{Cov}(\epsilon_{i,j,k}, \epsilon_{l,m,n}) = 0 \text{ si } (i,j,k) \neq (l,m,n) \text{ avec}$$
$$1 \leqslant i, l \leqslant I,\ 1 \leqslant j, m \leqslant J \text{ et } 1 \leqslant k, n \leqslant K.$$

On suppose que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fera l'objet d'un autre paragraphe.

On regroupe les valeurs que peut prendre la réponse Y dans les conditions (A_i, B_j) dans le tableau suivant :

Facteur A	Facteur B				
	B_1	$\dots$	B_j	$\dots$	B_J
A_1	$Y_{1,1,1} \dots Y_{1,1,K}$	$\dots$	$Y_{1,j,1} \dots Y_{1,j,K}$	$\dots$	$Y_{1,J,1} \dots Y_{1,J,K}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_i	$Y_{i,1,1} \dots Y_{i,1,K}$	$\dots$	$Y_{i,j,1} \dots Y_{i,j,K}$	$\dots$	$Y_{i,J,1} \dots Y_{i,J,K}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_I	$Y_{I,1,1} \dots Y_{I,1,K}$	$\dots$	$Y_{1,j,1} \dots Y_{1,j,K}$	$\dots$	$Y_{I,J,1} \dots Y_{I,J,K}$

On rappelle que la variation théorique due au facteur A est définie par :

$$SC_A = J \sum_{i=1}^{I} (Y_{i,\bullet,\bullet} - Y_{\bullet,\bullet,\bullet})^2.$$

On rappelle que la variation théorique due au facteur B est définie par :

$$SC_B = I \sum_{j=1}^{J} (Y_{\bullet,j,\bullet} - Y_{\bullet,\bullet,\bullet})^2.$$

On rappelle que la variation théorique due à l'interaction des facteurs A et B est définie par :

$$SC_{AB} = I \sum_{j=1}^{J} (Y_{i,j,\bullet} - Y_{i,\bullet,\bullet} - Y_{\bullet,j,\bullet} + Y_{\bullet,\bullet,\bullet})^2.$$

La variation résiduelle théorique est quant à elle définie par :

$$SC_R = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (Y_{i,j,k} - Y_{i,j,\bullet})^2.$$

Enfin la variation totale théorique est égale à :

$$SC_{TOT} = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (Y_{i,j,k} - Y_{\bullet,\bullet,\bullet})^2.$$

On rappelle la relation fondamentale de l'ANOVA :

$$SC_{TOT} = SC_A + SC_B + SC_{AB} + SC_R.$$

La liste $\boldsymbol{y}$ des données expérimentales $y_{1,1,1}, \ldots, y_{1,1,K}, \ldots, y_{1,2,1}, \ldots, y_{1,2,K}, \ldots, y_{I,J,K}$ permet de construire une réalisation du tableau précédent :

Facteur A	Facteur B				
	B_1	$\ldots$	B_j	$\ldots$	B_J
A_1	$y_{1,1,1} \ldots y_{1,1,K}$	$\ldots$	$y_{1,j,1} \ldots y_{1,j,K}$	$\ldots$	$y_{1,J,1} \ldots y_{1,J,K}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_i	$y_{i,1,1} \ldots y_{i,1,K}$	$\ldots$	$y_{i,j,1} \ldots y_{i,j,K}$	$\ldots$	$y_{i,J,1} \ldots y_{i,J,K}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_I	$y_{I,1,1} \ldots y_{I,1,K}$	$\ldots$	$y_{1,j,1} \ldots y_{1,j,K}$	$\ldots$	$y_{I,J,1} \ldots y_{I,J,K}$

La variation due au facteur A observée sur la liste de données $\boldsymbol{y}$ est définie par :

$$sc_A = J \sum_{i=1}^{I} (y_{i,\bullet,\bullet} - y_{\bullet,\bullet,\bullet})^2.$$

La variation due au facteur B observée sur la liste de données $\boldsymbol{y}$ est définie par :

$$sc_B = I \sum_{j=1}^{J} (y_{\bullet,j,\bullet} - y_{\bullet,\bullet,\bullet})^2.$$

La variation due à l'interaction des facteurs A et B observée sur la liste de données $\boldsymbol{y}$ est définie par :

$$sc_{AB} = I \sum_{j=1}^{J} (y_{i,j,\bullet} - y_{i,\bullet,\bullet} - y_{\bullet,j,\bullet} + y_{\bullet,\bullet,\bullet})^2.$$

La variation résiduelle observée sur la liste de données $\boldsymbol{y}$ est quant à elle définie par :

$$sc_R = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (y_{i,j,k} - y_{i,j,\bullet})^2.$$

Enfin la variation totale observée sur la liste de données $\boldsymbol{y}$ est égale à :

$$sc_{TOT} = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (y_{i,j,k} - y_{\bullet,\bullet,\bullet})^2.$$

La relation fondamentale de l'ANOVA reste valable lorsqu'elle est évaluée sur la liste de données $\boldsymbol{y}$:

$$sc_{TOT} = sc_A + sc_B + sc_{AB} + sc_R.$$

On introduit les dégres de liberté (Ddl) associés à chaque ligne du tableau de l'ANOVA :

Source	Degrés de liberté
Facteur A	$n_A = I - 1$
Facteur B	$n_B = J - 1$
Interaction AB	$n_{AB} = (I-1)(J-1)$
Résiduelle	$n_R = IJ(K-1)$
Totale	$n_{TOT} = IJK - 1$

On résume ces informations dans le tableau de l'ANOVA ci-dessous :

Source	Variation	Ddl	Carré Moyen	F	Décision
Facteur A	sc_A	n_A	$s_A^2 = \dfrac{sc_A}{n_A}$	$f_A = \dfrac{s_A^2}{s_R^2}$	$\mathcal{H}_0^{'}$ ou $\mathcal{H}_1^{'}$
Facteur B	sc_B	n_B	$s_B^2 = \dfrac{sc_B}{n_B}$	$f_B = \dfrac{s_B^2}{s_R^2}$	$\mathcal{H}_0^{''}$ ou $\mathcal{H}_1^{''}$
Interaction	sc_{AB}	n_{AB}	$s_{AB}^2 = \dfrac{sc_{AB}}{n_{AB}}$	$f_{AB} = \dfrac{s_{AB}^2}{s_R^2}$	$\mathcal{H}_0^{'''}$ ou $\mathcal{H}_1^{'''}$
Résiduelle	sc_R	n_R	$s_R^2 = \dfrac{sc_R}{n_R}$		
Totale	sc_{TOT}	n_{TOT}			

On souhaite faire les tests d'hypothèse suivants :

$$\boxed{\mathcal{H}_0^{'} : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0}$$

contre

$$\boxed{\mathcal{H}_1^{'} : \text{Il existe } i_0 \in \{1, 2, \ldots, I\} \text{ tel que } \alpha_{i_0} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{'}$ précédente d'absence d'effet du facteur A et lorsque les conditions de validité du modèle sont respectées, f_A est la réalisation d'une variable aléatoire

qui suit une loi de Fisher à $I-1$ et $IJ(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_A est supérieure ou égale à la valeur critique issue de la table. Lorsque l'hypothèse nulle $\mathcal{H}_0^{'}$ est rejetée on peut procéder à des comparaisons multiples des différents effets des niveaux du facteur voir la section 10.

$$\boxed{\mathcal{H}_0^{''} : \beta_1 = \beta_2 = \cdots = \beta_J = 0}$$

contre

$$\boxed{\mathcal{H}_1^{''} : \text{Il existe } j_0 \in \{1, 2, \ldots, J\} \text{ tel que } \beta_{j_0} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{''}$ précédente d'absence d'effet du facteur B et lorsque les conditions de validité du modèle sont respectées, f_B est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $J-1$ et $IJ(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_B est supérieure ou égale à la valeur critique issue de la table. Lorsque l'hypothèse nulle $\mathcal{H}_0^{''}$ est rejetée on peut procéder à des comparaisons multiples des différents effets des niveaux du facteur voir la section 10.

$$\boxed{\mathcal{H}_0^{'''} : (\alpha\beta)_{1,1} = (\alpha\beta)_{1,2} = \cdots = (\alpha\beta)_{1,J} = (\alpha\beta)_{2,1} = \cdots = (\alpha\beta)_{I,J} = 0}$$

contre

$$\boxed{\mathcal{H}_1^{'''} : \text{Il existe } (i_0, j_0) \in \{1, 2, \ldots, I\} \times \{1, 2, \ldots, J\} \text{ tel que } (\alpha\beta)_{i_0,j_0} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{'''}$ précédente d'absence d'effet de l'interaction des facteurs A et B et lorsque les conditions de validité du modèle sont respectées, f_{AB} est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $(I-1)(J-1)$ et $IJ(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_{AB} est supérieure ou égale à la valeur critique issue de la table.

Les estimateurs $\widehat{\mu}$, $\widehat{\alpha_1}, \ldots, \widehat{\alpha_I}$, $\widehat{\beta_1}, \ldots, \widehat{\beta_J}$, $\widehat{(\alpha\beta)_{1,1}}$, $\widehat{(\alpha\beta)_{1,2}}, \ldots, \widehat{(\alpha\beta)_{1,J}}$, $\widehat{(\alpha\beta)_{2,1}}, \ldots, \widehat{(\alpha\beta)_{I,J}}$, $\widehat{\sigma^2}$ des paramètres μ, $\alpha_1, \ldots, \alpha_I$, $\beta_1, \ldots, \beta_J$, $(\alpha\beta)_{1,1}$, $(\alpha\beta)_{1,2}, \ldots, (\alpha\beta)_{1,J}$, $(\alpha\beta)_{2,1}, \ldots, (\alpha\beta)_{I,J}$, σ^2 du modèle sont donnés par les formules suivantes :

$$\begin{aligned}
&\widehat{\mu} = Y_{\bullet,\bullet,\bullet} = \overline{Y},\ \widehat{\alpha_i} = Y_{i,\bullet,\bullet} - \widehat{\mu},\ 1 \leqslant i \leqslant I,\ \widehat{\beta_j} = Y_{\bullet,j,\bullet} - \widehat{\mu},\ 1 \leqslant j \leqslant J,\\
&\widehat{(\alpha\beta)_{i,j}} = Y_{i,j,\bullet} - Y_{i,\bullet,\bullet} - Y_{\bullet,j,\bullet} + Y_{\bullet,\bullet,\bullet},\ 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\\
&\widehat{\sigma^2} = \frac{SC_R}{IJ(K-1)} = S_R^2.
\end{aligned}$$

Ce sont des estimateurs sans biais.

Les estimations, obtenues pour la liste de données $\boldsymbol{y}$ et notées $\widehat{\mu}(\boldsymbol{y})$, $\widehat{\alpha_1}(\boldsymbol{y})$, ..., $\widehat{\alpha_I}(\boldsymbol{y})$, $\widehat{\beta_1}(\boldsymbol{y})$, ..., $\widehat{\beta_J}(\boldsymbol{y})$, $\widehat{(\alpha\beta)_{1,1}}(\boldsymbol{y})$, $\widehat{(\alpha\beta)_{1,2}}(\boldsymbol{y})$, ..., $\widehat{(\alpha\beta)_{1,J}}(\boldsymbol{y})$, $\widehat{(\alpha\beta)_{2,1}}(\boldsymbol{y})$, ..., $\widehat{(\alpha\beta)_{I,J}}(\boldsymbol{y})$, $\widehat{\sigma^2}(\boldsymbol{y})$, des paramètres μ, α_1, ..., α_I, β_1, ..., β_J, $(\alpha\beta)_{1,1}$, $(\alpha\beta)_{1,2}$, ..., $(\alpha\beta)_{1,J}$, $(\alpha\beta)_{2,1}$, ..., $(\alpha\beta)_{I,J}$, σ^2 du modèle se déduisent des formules ci-dessus :

$$\begin{aligned}
&\widehat{\mu}(\boldsymbol{y}) = y_{\bullet,\bullet,\bullet} = \overline{y},\ \widehat{\alpha_i}(\boldsymbol{y}) = y_{i,\bullet,\bullet} - \widehat{\mu}(\boldsymbol{y}),\ 1 \leqslant i \leqslant I,\ \widehat{\beta_j}(\boldsymbol{y}) = y_{\bullet,j,\bullet} - \widehat{\mu}(\boldsymbol{y}),\ 1 \leqslant j \leqslant J,\\
&\widehat{(\alpha\beta)_{i,j}}(\boldsymbol{y}) = y_{i,j,\bullet} - y_{i,\bullet,\bullet} - y_{\bullet,j,\bullet} + y_{\bullet,\bullet,\bullet},\ 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\\
&\widehat{\sigma^2}(\boldsymbol{y}) = \frac{sc_R}{IJ(K-1)} = s_R^2.
\end{aligned}$$

4.2. Modèles à effets aléatoires

4.2.1. Sans répétition

Dans ce cas les A_i représentent un échantillon de taille I prélevé dans une population importante. Nous admettrons que les effets des A_i, les α_i, sont distribués suivant une loi normale centrée de variance σ_A^2. Les β_j représentent un échantillon de taille J prélevé dans une population importante. Nous admettrons que les effets des B_j, les β_j, sont distribués suivant une loi normale centrée de variance σ_B^2. Pour chacun des couples de modalités (A_i, B_j) on effectue une mesure d'une réponse Y qui est une variable continue. On note $n = I \times J$ le nombre total de mesures ayant été effectuées.

On introduit le modèle :

$$Y_{i,j} = \mu + \alpha_i + \beta_j + \epsilon_{i,j}, \quad i = 1 \ldots I,\ j = 1 \ldots J,$$

où $Y_{i,j}$ est la valeur prise par la réponse Y dans les conditions (A_i, B_j). On suppose que :

$$\begin{aligned}
\mathcal{L}\left(\alpha_i\right) &= \mathcal{N}(0, \sigma_A^2),\ \forall\ i, 1 \leqslant i \leqslant I,\\
\mathcal{L}\left(\beta_j\right) &= \mathcal{N}(0, \sigma_B^2),\ \forall\ j, 1 \leqslant j \leqslant J,
\end{aligned}$$

ainsi que l'indépendance des effets aléatoires :

$$\begin{gathered}
\mathrm{Cov}(\alpha_i, \alpha_j) = 0 \text{ si } i \neq j \text{ et } 1 \leqslant i, j \leqslant I, \quad \mathrm{Cov}(\beta_i, \beta_j) = 0 \text{ si } i \neq j \text{ et } 1 \leqslant i, j \leqslant J,\\
\mathrm{Cov}(\alpha_i, \beta_j) = 0 \text{ si } 1 \leqslant i \leqslant I \text{ et } 1 \leqslant j \leqslant J.
\end{gathered}$$

On postule les hypothèses classiques suivantes pour les erreurs :

$$\begin{gathered}
\forall\ (i, j), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ \mathcal{L}(\epsilon_{i,j}) = \mathcal{N}(0, \sigma^2),\\
\mathrm{Cov}(\epsilon_{i,j}, \epsilon_{k,l}) = 0 \text{ si } (i, j) \neq (k, l) \text{ avec } 1 \leqslant i, k \leqslant I \text{ et } 1 \leqslant j, l \leqslant J,
\end{gathered}$$

ainsi que l'indépendance des effets aléatoires et des erreurs :

$$\begin{gathered}
\mathrm{Cov}(\alpha_i, \epsilon_{j,k}) = 0 \text{ si } 1 \leqslant i, j \leqslant I \text{ et } 1 \leqslant k \leqslant J,\\
\mathrm{Cov}(\beta_i, \epsilon_{j,k}) = 0 \text{ si } 1 \leqslant j \leqslant I \text{ et } 1 \leqslant i, k \leqslant J.
\end{gathered}$$

On suppose que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fera l'objet d'un autre paragraphe.

On utilise les quantités SC_A, SC_B, SC_R, SC_{TOT}, sc_A, sc_B, sc_R et sc_{TOT} introduites à la section 4.1.1.

On rappelle la relation fondamentale de l'ANOVA :

$$SC_{TOT} = SC_A + SC_B + SC_R.$$

On introduit les dégres de liberté (Ddl) associés à chaque ligne du tableau de l'ANOVA :

Source	Degrés de liberté
Facteur A	$n_A = I - 1$
Facteur B	$n_B = J - 1$
Résiduelle	$n_R = (I-1)(J-1)$
Totale	$n_{TOT} = IJ - 1$

On résume ces informations dans le tableau de l'ANOVA ci-dessous :

Source	Variation	Ddl	Carré Moyen	F	Décision
Facteur A	sc_A	n_A	$s_A^2 = \dfrac{sc_A}{n_A}$	$f_A = \dfrac{s_A^2}{s_R^2}$	$\mathcal{H}_0'$ ou $\mathcal{H}_1'$
Facteur B	sc_B	n_B	$s_B^2 = \dfrac{sc_B}{n_B}$	$f_B = \dfrac{s_B^2}{s_R^2}$	$\mathcal{H}_0''$ ou $\mathcal{H}_1''$
Résiduelle	sc_R	n_R	$s_R^2 = \dfrac{sc_R}{n_R}$		
Totale	sc_{TOT}	n_{TOT}			

On souhaite faire les tests d'hypothèse suivants :

$$\boxed{\mathcal{H}_0' : \sigma_A^2 = 0}$$
contre
$$\boxed{\mathcal{H}_1' : \sigma_A^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0'$ précédente d'absence d'effet du facteur A et lorsque les conditions de validité du modèle sont respectées, f_A est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $I-1$ et $(I-1)(J-1)$ degrés de liberté.

$$\boxed{\mathcal{H}_0'' : \sigma_B^2 = 0}$$
contre
$$\boxed{\mathcal{H}_1'' : \sigma_B^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0''$ précédente d'absence d'effet du facteur B et lorsque les conditions de validité du modèle sont respectées, f_B est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $J-1$ et $(I-1)(J-1)$ degrés de liberté.

Les estimateurs $\widehat{\mu}$, $\widehat{\sigma_A^2}$, $\widehat{\sigma_B^2}$, $\widehat{\sigma^2}$ des paramètres μ, σ_A^2, σ_B^2, σ^2 du modèle sont donnés par les formules suivantes :

$$\begin{aligned}
&\widehat{\mu} = Y_{\bullet,\bullet} = \overline{Y},\\
&\widehat{\sigma_A^2} = \frac{1}{J}\left(S_A^2 - S_R^2\right),\ \widehat{\sigma_B^2} = \frac{1}{I}\left(S_B^2 - S_R^2\right)\\
&\widehat{\sigma^2} = \frac{SC_R}{(I-1)(J-1)} = S_R^2,
\end{aligned}$$

où $S_A^2 = \dfrac{SC_A}{n_A}$, $S_B^2 = \dfrac{SC_B}{n_B}$ et $S_R^2 = \dfrac{SC_R}{n_R}$.
Ce sont des estimateurs sans biais.

Les estimations, obtenues pour la liste de données $\boldsymbol{y}$ et notées $\widehat{\mu}(\boldsymbol{y})$, $\widehat{\sigma_A^2}(\boldsymbol{y})$, $\widehat{\sigma_B^2}(\boldsymbol{y})$, $\widehat{\sigma^2}(\boldsymbol{y})$, des paramètres μ, σ_A^2, σ_B^2, σ^2 du modèle se déduisent des formules ci-desses :

$$\begin{aligned}
&\widehat{\mu}(\boldsymbol{y}) = y_{\bullet,\bullet} = \overline{y},\\
&\widehat{\sigma_A^2}(\boldsymbol{y}) = \frac{1}{J}\left(s_A^2 - s_R^2\right),\ \widehat{\sigma_B^2}(\boldsymbol{y}) = \frac{1}{I}\left(s_B^2 - s_R^2\right)\\
&\widehat{\sigma^2}(\boldsymbol{y}) = \frac{sc_R}{(I-1)(J-1)} = s_R^2.
\end{aligned}$$

4.2.2. Avec répétitions

Dans ce cas les A_i représentent un échantillon de taille I prélevé dans une population importante. Nous admettrons que les effets des A_i, les α_i, sont distribués suivant une loi

normale centrée de variance σ_A^2. Les β_j représentent un échantillon de taille J prélevé dans une population importante. Nous admettrons que les effets des B_j, les β_j, sont distribués suivant une loi normale centrée de variance σ_B^2. Pour chacun des couples de modalités (A_i, B_j) on effectue $K \geqslant 2$ mesures d'une réponse Y qui est une variable continue. On note $n = I \times J \times K$ le nombre total de mesures ayant été effectuées.

On introduit le modèle :

$$Y_{i,j,k} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{i,j} + \epsilon_{i,j,k}, \quad i = 1 \dots I,\ j = 1 \dots J,\ k = 1 \dots K,$$

où $Y_{i,j,k}$ est la valeur prise par la réponse Y dans les conditions (A_i, B_j) lors du $k-$ème essai. On suppose que

$$\begin{gathered}
\mathcal{L}\left(\alpha_i\right) = \mathcal{N}(0, \sigma_A^2),\ \forall\ i, 1 \leqslant i \leqslant I, \\
\mathcal{L}\left(\beta_j\right) = \mathcal{N}(0, \sigma_B^2),\ \forall\ j, 1 \leqslant j \leqslant J, \\
\mathcal{L}\left((\alpha\beta)_{i,j}\right) = \mathcal{N}(0, \sigma_{AB}^2),\ \forall\ (i,j), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,
\end{gathered}$$

ainsi que l'indépendance des effets aléatoires :

$$\begin{gathered}
\mathrm{Cov}(\alpha_i, \alpha_j) = 0 \text{ si } i \neq j \text{ et } 1 \leqslant i, j \leqslant I, \quad \mathrm{Cov}(\beta_i, \beta_j) = 0 \text{ si } i \neq j \text{ et } 1 \leqslant i, j \leqslant J, \\
\mathrm{Cov}((\alpha\beta)_{i,j}, (\alpha\beta)_{k,l}) = 0 \text{ si } (i,j) \neq (k,l) \text{ avec } 1 \leqslant i, k \leqslant I \text{ et } 1 \leqslant j, l \leqslant J, \\
\mathrm{Cov}(\alpha_i, \beta_j) = 0 \text{ si } 1 \leqslant i \leqslant I \text{ et } 1 \leqslant j \leqslant J, \\
\mathrm{Cov}(\alpha_i, (\alpha\beta)_{j,k}) = 0 \text{ si } 1 \leqslant i, j \leqslant I \text{ et } 1 \leqslant k \leqslant J, \\
\mathrm{Cov}(\beta_i, (\alpha\beta)_{j,k}) = 0 \text{ si } 1 \leqslant j \leqslant I \text{ et } 1 \leqslant i, k \leqslant J.
\end{gathered}$$

On postule les hypothèses classiques suivantes pour les erreurs :

$$\begin{gathered}
\forall\ (i,j,k), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K,\ \mathcal{L}(\epsilon_{i,j,k}) = \mathcal{N}(0, \sigma^2), \\
\text{et } \mathrm{Cov}(\epsilon_{i,j,k}, \epsilon_{l,m,n}) = 0 \text{ si } (i,j,k) \neq (l,m,n) \text{ avec} \\
1 \leqslant i, l \leqslant I,\ 1 \leqslant j, m \leqslant J \text{ et } 1 \leqslant k, n \leqslant K,
\end{gathered}$$

ainsi que l'indépendance des effets aléatoires et des erreurs :

$$\begin{gathered}
\mathrm{Cov}(\alpha_i, \epsilon_{j,k,l}) = 0 \text{ si } 1 \leqslant i, j \leqslant I, 1 \leqslant k \leqslant J, \text{ et } 1 \leqslant l \leqslant K, \\
\mathrm{Cov}(\beta_i, \epsilon_{j,k,l}) = 0 \text{ si } 1 \leqslant j \leqslant I, 1 \leqslant i, k \leqslant J, \text{ et } 1 \leqslant l \leqslant K, \\
\mathrm{Cov}((\alpha\beta)_{i,j}, \epsilon_{k,l,m}) = 0 \text{ si } 1 \leqslant i, k \leqslant I, 1 \leqslant j, l \leqslant J, \text{ et } 1 \leqslant m \leqslant K.
\end{gathered}$$

On suppose que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fera l'objet d'un autre paragraphe.

On utilise les quantités SC_A, SC_B, SC_{AB}, SC_R, SC_{TOT}, sc_A, sc_B, sc_{AB}, sc_R et sc_{TOT} introduites à la section 4.1.2.

On rappelle la relation fondamentale de l'ANOVA :

$$SC_{TOT} = SC_A + SC_B + SC_{AB} + SC_R.$$

On introduit les dégres de liberté (Ddl) associés à chaque ligne du tableau de l'ANOVA :

Source	Degrés de liberté
Facteur A	$n_A = I - 1$
Facteur B	$n_B = J - 1$
Interaction AB	$n_{AB} = (I-1)(J-1)$
Résiduelle	$n_R = IJ(K-1)$
Totale	$n_{TOT} = IJK - 1$

On résume ces informations dans le tableau de l'ANOVA ci-dessous :

Source	Variation	Ddl	Carré Moyen	F	Décision
Facteur A	sc_A	n_A	$s_A^2 = \frac{sc_A}{n_A}$	$f_A = \frac{s_A^2}{s_{AB}^2}$	$\mathcal{H}_0^{'}$ ou $\mathcal{H}_1^{'}$
Facteur B	sc_B	n_B	$s_B^2 = \frac{sc_B}{n_B}$	$f_B = \frac{s_B^2}{s_{AB}^2}$	$\mathcal{H}_0^{''}$ ou $\mathcal{H}_1^{''}$
Interaction	sc_{AB}	n_{AB}	$s_{AB}^2 = \frac{sc_{AB}}{n_{AB}}$	$f_{AB} = \frac{s_{AB}^2}{s_R^2}$	$\mathcal{H}_0^{'''}$ ou $\mathcal{H}_1^{'''}$
Résiduelle	sc_R	n_R	$s_R^2 = \frac{sc_R}{n_R}$		
Totale	sc_{TOT}	n_{TOT}			

On souhaite faire les tests d'hypothèse suivants :

$$\boxed{\mathcal{H}_0^{'} : \sigma_A^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1^{'} : \sigma_A^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{'}$ précédente d'absence d'effet du facteur A et lorsque les conditions de validité du modèle sont respectées, f_A est la réalisation d'une variable aléatoire

qui suit une loi de Fisher à $I-1$ et $IJ(K-1)$ degrés de liberté.

$$\boxed{\mathcal{H}_0'' : \sigma_B^2 = 0}$$
contre
$$\boxed{\mathcal{H}_1'' : \sigma_B^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0''$ précédente d'absence d'effet du facteur B et lorsque les conditions de validité du modèle sont respectées, f_B est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $J-1$ et $IJ(K-1)$ degrés de liberté.

$$\boxed{\mathcal{H}_0''' : \sigma_{AB}^2 = 0}$$
contre
$$\boxed{\mathcal{H}_1''' : \sigma_{AB}^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0'''$ précédente d'absence d'effet de l'interaction entre les facteurs A et B et lorsque les conditions de validité du modèle sont respectées, f_{AB} est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $(I-1)(J-1)$ et $IJ(K-1)$ degrés de liberté.

Les estimateurs $\widehat{\mu}$, $\widehat{\sigma_A^2}$, $\widehat{\sigma_B^2}$, $\widehat{\sigma_{AB}^2}$, $\widehat{\sigma^2}$ des paramètres μ, σ_A^2, σ_B^2, σ_{AB}^2, σ^2 du modèle sont donnés par les formules suivantes :

$$\begin{aligned}
&\widehat{\mu} = Y_{\bullet,\bullet,\bullet} = \overline{Y},\\
&\widehat{\sigma_A^2} = \frac{1}{JK}\left(S_A^2 - S_{AB}^2\right),\ \widehat{\sigma_B^2} = \frac{1}{IK}\left(S_B^2 - S_{AB}^2\right)\\
&\widehat{\sigma_{AB}^2} = \frac{1}{K}\left(S_{AB}^2 - S_R^2\right)\\
&\widehat{\sigma^2} = \frac{SC_R}{(I-1)(J-1)} = S_R^2,
\end{aligned}$$

où $S_A^2 = \dfrac{SC_A}{n_A}$, $S_B^2 = \dfrac{SC_B}{n_B}$, $S_{AB}^2 = \dfrac{SC_{AB}}{n_{AB}}$ et $S_R^2 = \dfrac{SC_R}{n_R}$.
Ce sont des estimateurs sans biais.

Les estimations, obtenues pour la liste de données $\boldsymbol{y}$ et notées $\widehat{\mu}(\boldsymbol{y})$, $\widehat{\sigma_A^2}(\boldsymbol{y})$, $\widehat{\sigma_B^2}(\boldsymbol{y})$, $\widehat{\sigma_{AB}^2}(\boldsymbol{y})$, $\widehat{\sigma^2}(\boldsymbol{y})$, des paramètres μ, σ_A^2, σ_B^2, σ_{AB}^2, σ^2 du modèle se déduisent des formules ci-dessus :

$$\begin{aligned}
&\widehat{\mu}(\boldsymbol{y}) = y_{\bullet,\bullet,\bullet} = \overline{y},\\
&\widehat{\sigma_A^2}(\boldsymbol{y}) = \frac{1}{JK}\left(s_A^2 - s_{AB}^2\right),\ \widehat{\sigma_B^2}(\boldsymbol{y}) = \frac{1}{IK}\left(s_B^2 - s_{AB}^2\right)\\
&\widehat{\sigma_{AB}^2}(\boldsymbol{y}) = \frac{1}{K}\left(s_{AB}^2 - s_R^2\right)\\
&\widehat{\sigma^2}(\boldsymbol{y}) = \frac{sc_R}{(I-1)(J-1)} = s_R^2.
\end{aligned}$$

4.3. Modèles à effets mixtes

4.3.1. Sans répétition

Un facteur contrôlé A se présente sous I modalités, chacune d'entre elles étant notée A_i. Les β_j représentent un échantillon de taille J prélevé dans une population importante. Nous admettrons que les effets des B_j, les β_j, sont distribués suivant une loi normale centrée de variance σ_B^2. Pour chacun des couples de modalités (A_i, B_j) on effectue une mesure d'une réponse Y qui est une variable continue. On note $n = I \times J$ le nombre total de mesures ayant été effectuées.

On introduit le modèle :

$$Y_{i,j} = \mu + \alpha_i + \beta_j + \epsilon_{i,j}, \quad i = 1 \dots I,\ j = 1 \dots J,$$
$$\text{avec les contraintes supplémentaires } \sum_{i=1}^{I} \alpha_i = 0,$$

où $Y_{i,j}$ est la valeur prise par la réponse Y dans les conditions (A_i, B_j). On suppose que :

$$\mathcal{L}\left(\beta_j\right) = \mathcal{N}(0, \sigma_B^2),\ \forall\ j, 1 \leqslant j \leqslant J,$$

ainsi que l'indépendance des effets aléatoires :

$$\mathrm{Cov}(\beta_i, \beta_j) = 0 \text{ si } i \neq j \text{ et } 1 \leqslant i, j \leqslant J.$$

On postule les hypothèses classiques suivantes pour les erreurs :

$$\forall\ (i,j), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ \mathcal{L}(\epsilon_{i,j}) = \mathcal{N}(0, \sigma^2),$$
$$\mathrm{Cov}(\epsilon_{i,j}, \epsilon_{k,l}) = 0 \text{ si } (i,j) \neq (k,l) \text{ avec } 1 \leqslant i, k \leqslant I \text{ et } 1 \leqslant j, l \leqslant J,$$

ainsi que l'indépendance des effets aléatoires et des erreurs :

$$\mathrm{Cov}(\beta_i, \epsilon_{j,k}) = 0 \text{ si } 1 \leqslant j \leqslant I \text{ et } 1 \leqslant i, k \leqslant J.$$

On suppose que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fera l'objet d'un autre paragraphe.

On utilise les quantités SC_A, SC_B, SC_R, SC_{TOT}, sc_A, sc_B, sc_R et sc_{TOT} introduites à la section 4.1.1.

On rappelle la relation fondamentale de l'ANOVA :

$$SC_{TOT} = SC_A + SC_B + SC_R.$$

On introduit les dégres de liberté (Ddl) associés à chaque ligne du tableau de l'ANOVA :

Source	Degrés de liberté
Facteur A	$n_A = I - 1$
Facteur B	$n_B = J - 1$
Résiduelle	$n_R = (I-1)(J-1)$
Totale	$n_{TOT} = IJ - 1$

On résume ces informations dans le tableau de l'ANOVA ci-dessous :

Source	Variation	Ddl	Carré Moyen	F	Décision
Facteur A	sc_A	n_A	$s_A^2 = \dfrac{sc_A}{n_A}$	$f_A = \dfrac{s_A^2}{s_R^2}$	$\mathcal{H}_0^{'}$ ou $\mathcal{H}_1^{'}$
Facteur B	sc_B	n_B	$s_B^2 = \dfrac{sc_B}{n_B}$	$f_B = \dfrac{s_B^2}{s_R^2}$	$\mathcal{H}_0^{''}$ ou $\mathcal{H}_1^{''}$
Résiduelle	sc_R	n_R	$s_R^2 = \dfrac{sc_R}{n_R}$		
Totale	sc_{TOT}	n_{TOT}			

On souhaite faire les tests d'hypothèse suivants :

$$\boxed{\mathcal{H}_0^{'} : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0}$$

contre

$$\boxed{\mathcal{H}_1^{'} : \text{Il existe } i_0 \in \{1, 2, \ldots, I\} \text{ tel que } \alpha_{i_0} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{'}$ précédente d'absence d'effet du facteur A et lorsque les conditions de validité du modèle sont respectées, f_A est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $I-1$ et $(I-1)(J-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_A est supérieure ou égale à la valeur critique issue de la table. Lorsque l'hypothèse nulle $\mathcal{H}_0^{'}$ est rejetée on peut procéder à des comparaisons multiples des différents effets des niveaux du facteur voir la section 10.

$$\boxed{\mathcal{H}_0'' : \sigma_B^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1'' : \sigma_B^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0''$ précédente d'absence d'effet du facteur B et lorsque les conditions de validité du modèle sont respectées, f_B est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $J-1$ et $(I-1)(J-1)$ degrés de liberté.

Les estimateurs $\widehat{\mu}$, $\widehat{\alpha_1}$, ..., $\widehat{\alpha_I}$, $\widehat{\sigma_B^2}$, $\widehat{\sigma^2}$ des paramètres μ, α_1, ..., α_I, σ_B^2, σ^2 du modèle sont donnés par les formules suivantes :

$$\begin{aligned}
&\widehat{\mu} = Y_{\bullet,\bullet,\bullet} = \overline{Y},\ \widehat{\alpha}_i = Y_{i,\bullet,\bullet} - \widehat{\mu},\ 1 \leqslant i \leqslant I,\\
&\widehat{\sigma_B^2} = \frac{1}{I}\left(S_B^2 - S_R^2\right)\\
&\widehat{\sigma^2} = \frac{SC_R}{(I-1)(J-1)} = S_R^2,
\end{aligned}$$

où $S_B^2 = \dfrac{SC_B}{n_B}$ et $S_R^2 = \dfrac{SC_R}{n_R}$.
Ce sont des estimateurs sans biais.

Les estimations, obtenues pour la liste de données $\boldsymbol{y}$ et notées $\widehat{\mu}(\boldsymbol{y})$, $\widehat{\alpha_1}(\boldsymbol{y})$, ..., $\widehat{\alpha_I}(\boldsymbol{y})$, $\widehat{\sigma_B^2}(\boldsymbol{y})$, $\widehat{\sigma^2}(\boldsymbol{y})$, des paramètres μ, α_1, ..., α_I, σ_B^2, σ^2 du modèle se déduisent des formules ci-dessus :

$$\begin{aligned}
&\widehat{\mu} = y_{\bullet,\bullet,\bullet} = \overline{y},\ \widehat{\alpha}_i = y_{i,\bullet,\bullet} - \widehat{\mu},\ 1 \leqslant i \leqslant I,\\
&\widehat{\sigma_B^2} = \frac{1}{I}\left(s_B^2 - s_{AB}^2\right)\\
&\widehat{\sigma^2} = \frac{sc_R}{(I-1)(J-1)} = s_R^2.
\end{aligned}$$

4.3.2. Avec répétitions

Un facteur contrôlé A se présente sous I modalités, chacune d'entre elles étant notée A_i. Les β_j représentent un échantillon de taille J prélevé dans une population importante. Nous admettrons que les effets des B_j, les β_j, sont distribués suivant une loi normale centrée de variance σ_B^2. Pour chacun des couples de modalités (A_i, B_j) on effectue $K \geqslant 2$ mesures d'une réponse Y qui est une variable continue. On note $n = I \times J \times K$ le nombre total de mesures ayant été effectuées.

On introduit le modèle :

$$Y_{i,j,k} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{i,j} + \epsilon_{i,j,k}, \quad i = 1 \dots I,\ j = 1 \dots J,\ k = 1 \dots K$$
$$\text{avec les contraintes supplémentaires } \sum_{i=1}^{I} \alpha_i = 0,$$
$$\sum_{i=1}^{I} (\alpha\beta)_{i,j} = 0,\ \forall j \in \{1, \dots, J\},$$

où $Y_{i,j,k}$ est la valeur prise par la réponse Y dans les conditions (A_i, B_j) lors du $k-$ème essai. On suppose que

$$\mathcal{L}(\beta_j) = \mathcal{N}(0, \sigma_B^2),\ \forall\ j, 1 \leqslant j \leqslant J,$$
$$\mathcal{L}((\alpha\beta)_{i,j}) = \mathcal{N}(0, \sigma_{AB}^2),\ \forall\ (i,j), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,$$

ainsi que l'indépendance des effets aléatoires :

$$\mathrm{Cov}(\beta_i, \beta_j) = 0 \text{ si } i \neq j \text{ et } 1 \leqslant i, j \leqslant J,$$
$$\mathrm{Cov}((\alpha\beta)_{i,j}, (\alpha\beta)_{k,l}) = 0 \text{ si } (i,j) \neq (k,l) \text{ avec } 1 \leqslant i, k \leqslant I \text{ et } 1 \leqslant j, l \leqslant J,$$
$$\mathrm{Cov}(\beta_i, (\alpha\beta)_{j,k}) = 0 \text{ si } 1 \leqslant j \leqslant I \text{ et } 1 \leqslant i, k \leqslant J.$$

On postule les hypothèses classiques suivantes pour les erreurs :

$$\forall\ (i,j,k), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K,\ \mathcal{L}(\epsilon_{i,j,k}) = \mathcal{N}(0, \sigma^2),$$
$$\text{et } \mathrm{Cov}(\epsilon_{i,j,k}, \epsilon_{l,m,n}) = 0 \text{ si } (i,j,k) \neq (l,m,n) \text{ avec}$$
$$1 \leqslant i, l \leqslant I,\ 1 \leqslant j, m \leqslant J \text{ et } 1 \leqslant k, n \leqslant K,$$

ainsi que l'indépendance des effets aléatoires et des erreurs :

$$\mathrm{Cov}(\beta_i, \epsilon_{j,k,l}) = 0 \text{ si } 1 \leqslant j \leqslant I, 1 \leqslant i, k \leqslant J, \text{ et } 1 \leqslant l \leqslant K,$$
$$\mathrm{Cov}((\alpha\beta)_{i,j}, \epsilon_{k,l,m}) = 0 \text{ si } 1 \leqslant i, k \leqslant I, 1 \leqslant j, l \leqslant J, \text{ et } 1 \leqslant m \leqslant K.$$

On suppose que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fera l'objet d'un autre paragraphe.

On utilise les quantités SC_A, SC_B, SC_{AB}, SC_R, SC_{TOT}, sc_A, sc_B, sc_{AB}, sc_R et sc_{TOT} introduites à la section 4.1.2.

On rappelle la relation fondamentale de l'ANOVA :

$$SC_{TOT} = SC_A + SC_B + SC_{AB} + SC_R.$$

On introduit les dégres de liberté (Ddl) associés à chaque ligne du tableau de l'ANOVA :

Source	Degrés de liberté
Facteur A	$n_A = I - 1$
Facteur B	$n_B = J - 1$
Interaction AB	$n_{AB} = (I-1)(J-1)$
Résiduelle	$n_R = IJ(K-1)$
Totale	$n_{TOT} = IJK - 1$

On résume ces informations dans le tableau de l'ANOVA ci-dessous :

Source	Variation	Ddl	Carré Moyen	F	Décision
Facteur A	sc_A	n_A	$s_A^2 = \dfrac{sc_A}{n_A}$	$f_A = \dfrac{s_A^2}{s_{AB}^2}$	$\mathcal{H}_0^{'}$ ou $\mathcal{H}_1^{'}$
Facteur B	sc_B	n_B	$s_B^2 = \dfrac{sc_B}{n_B}$	$f_B = \dfrac{s_B^2}{s_R^2}$	$\mathcal{H}_0^{''}$ ou $\mathcal{H}_1^{''}$
Interaction	sc_{AB}	n_{AB}	$s_{AB}^2 = \dfrac{sc_{AB}}{n_{AB}}$	$f_{AB} = \dfrac{s_{AB}^2}{s_R^2}$	$\mathcal{H}_0^{'''}$ ou $\mathcal{H}_1^{'''}$
Résiduelle	sc_R	n_R	$s_R^2 = \dfrac{sc_R}{n_R}$		
Totale	sc_{TOT}	n_{TOT}			

On souhaite faire les tests d'hypothèse suivants :

$$\boxed{\mathcal{H}_0^{'} : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0}$$

contre

$$\boxed{\mathcal{H}_1^{'} : \text{Il existe } i_0 \in \{1, 2, \ldots, I\} \text{ tel que } \alpha_{i_0} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{'}$ précédente d'absence d'effet du facteur A et lorsque les conditions de validité du modèle sont respectées, f_A est la réalisation d'une variable aléatoire

qui suit une loi de Fisher à $I-1$ et $IJ(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_A est supérieure ou égale à la valeur critique issue de la table. Lorsque l'hypothèse nulle $\mathcal{H}_0^{'}$ est rejetée on peut procéder à des comparaisons multiples des différents effets des niveaux du facteur voir la section 10.

$$\boxed{\mathcal{H}_0^{''} : \sigma_B^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1^{''} : \sigma_B^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{''}$ précédente d'absence d'effet du facteur B et lorsque les conditions de validité du modèle sont respectées, f_B est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $J-1$ et $IJ(K-1)$ degrés de liberté.

$$\boxed{\mathcal{H}_0^{'''} : \sigma_{AB}^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1^{'''} : \sigma_{AB}^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{'''}$ précédente d'absence d'effet de l'interaction entre les facteurs A et B et lorsque les conditions de validité du modèle sont respectées, f_{AB} est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $(I-1)(J-1)$ et $IJ(K-1)$ degrés de liberté.

Les estimateurs $\widehat{\mu}$, $\widehat{\alpha_1}$, $\ldots$, $\widehat{\alpha_I}$, $\widehat{\sigma_B^2}$, $\widehat{\sigma_{AB}^2}$, $\widehat{\sigma^2}$ des paramètres μ, α_1, $\ldots$, α_I, σ_B^2, σ_{AB}^2, σ^2 du modèle sont donnés par les formules suivantes :

$$\begin{aligned}
&\widehat{\mu} = Y_{\bullet,\bullet} = \overline{Y},\ \widehat{\alpha_i} = Y_{i,\bullet} - \widehat{\mu},\ 1 \leqslant i \leqslant I,\\
&\widehat{\sigma_B^2} = \frac{1}{IK}\left(S_B^2 - S_R^2\right)\\
&\widehat{\sigma_{AB}^2} = \frac{1}{K}\left(S_{AB}^2 - S_R^2\right)\\
&\widehat{\sigma^2} = \frac{SC_R}{(I-1)(J-1)} = S_R^2,
\end{aligned}$$

où $S_B^2 = \dfrac{SC_B}{n_B}$, $S_{AB}^2 = \dfrac{SC_{AB}}{n_{AB}}$ et $S_R^2 = \dfrac{SC_R}{n_R}$.
Ce sont des estimateurs sans biais.

Les estimations, obtenues pour la liste de données $\boldsymbol{y}$ et notées $\widehat{\mu}(\boldsymbol{y})$, $\widehat{\alpha_1}(\boldsymbol{y})$, ..., $\widehat{\alpha_I}(\boldsymbol{y})$, $\widehat{\sigma_B^2}(\boldsymbol{y})$, $\widehat{\sigma_{AB}^2}(\boldsymbol{y})$, $\widehat{\sigma^2}(\boldsymbol{y})$, des paramètres μ, α_1, ..., α_I, σ_B^2, σ_{AB}^2, σ^2 du modèle se déduisent des formules ci-dessus :

$$\begin{aligned}
&\widehat{\mu}(\boldsymbol{y}) = y_{\bullet,\bullet} = \overline{y},\ \widehat{\alpha_i}(\boldsymbol{y}) = y_{i,\bullet} - \widehat{\mu}(\boldsymbol{y}),\ 1 \leqslant i \leqslant I,\\
&\widehat{\sigma_B^2}(\boldsymbol{y}) = \frac{1}{IK}\left(s_B^2 - s_R^2\right)\\
&\widehat{\sigma_{AB}^2}(\boldsymbol{y}) = \frac{1}{K}\left(s_{AB}^2 - s_R^2\right)\\
&\widehat{\sigma^2}(\boldsymbol{y}) = \frac{sc_R}{(I-1)(J-1)} = s_R^2.
\end{aligned}$$

5. Analyse de la variance à deux facteurs emboîtés

Dans toute cette section, on utilise les notations définies à la section 1.

On est dans la situation particulière où les effets des niveaux du facteur B n'ont pas de signification concrète, par exemple ces niveaux dépendent du niveau du facteur A considéré et une étude des effets principaux du facteur B n'a pas de pertinence.

On ne peut se servir d'un modèle où les facteurs sont emboîtés[3], que si l'on dispose de répétitions. Dans le cas contraire où les essais ne seraient pas répétés, l'effet dû au facteur B ne pourra être étudié et le modèle que l'on devra utiliser pour étudiés les données sera l'un de ceux exposés à la section 3.

Ainsi par exemple un fabriquant de détergents alimente plusieurs chaînes de distribution : A_1, A_2, ..., A_I. On pense que les boîtes de produits livrées à certaines chaînes de distribution contiennent une masse de détergent inférieure à celle des autres chaînes de distribution. Pour étudier cette situation, on décide de prélever K boîtes dans J magasins de chaque chaîne. Ainsi le second facteur B_j, associé au $j-$ème magasin dans la chaîne, est un repère qui n'a aucune signification réelle : il n'y a, par exemple aucune relation entre le magasin n°3 de la chaîne 1 et le magasin n°3 de la chaîne 4. Il n'y a donc aucun intérêt à introduire un terme dans le modèle caractérisant l'effet principal du facteur B. Pour indiquer la dépendance des niveaux du second facteur B aux niveaux du premier facteur A on note les niveaux du second facteur B : $B_{j(i)}$, $1 \leqslant i \leqslant I$ et $1 \leqslant j \leqslant J$.

[3]On appelle aussi ce type de modèles, des modèles hiérarchiques ou en anglais *hierarchical* ou *nested models*.

5.1. Modèles à effets fixes

5.1.1. Avec répétitions

Un facteur contrôlé A se présente sous I modalités, chacune d'entre elles étant notée A_i. Un facteur contrôlé B se présente sous J modalités, chacune d'entre elles dépendant du niveau A_i du facteur A et étant alors notée $B_{j(i)}$. Pour chacun des couples de modalités $(A_i, B_{j(i)})$ on effectue $K \geqslant 2$ mesures d'une réponse Y qui est une variable continue. On note $n = I \times J \times K$ le nombre total de mesures ayant été effectuées.

On introduit le modèle :

$$Y_{i,j,k} = \mu + \alpha_i + \beta_{j(i)} + \epsilon_{i,j,k}, \quad i = 1 \dots I,\ j = 1 \dots J,\ k = 1 \dots K$$

$$\text{avec les contraintes supplémentaires } \sum_{i=1}^{I} \alpha_i = 0,\ \sum_{j=1}^{J} \beta_{j(i)} = 0,\ \forall i \in \{1, \dots, I\}$$

où $Y_{i,j,k}$ est la valeur prise par la réponse Y dans les conditions $(A_i, B_{j(i)})$ lors du $k-$ème essai. On postule les hypothèses classiques suivantes pour les erreurs :

$$\forall\ (i,j,k), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K,\ \mathcal{L}(\epsilon_{i,j,k}) = \mathcal{N}(0, \sigma^2),$$
$$\text{et } \mathrm{Cov}(\epsilon_{i,j,k}, \epsilon_{l,m,n}) = 0 \text{ si } (i,j,k) \neq (l,m,n) \text{ avec}$$
$$1 \leqslant i, l \leqslant I,\ 1 \leqslant j, m \leqslant J \text{ et } 1 \leqslant k, n \leqslant K.$$

On suppose que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fera l'objet d'un autre paragraphe.

On regroupe les valeurs que peut prendre la réponse Y dans les conditions $(A_i, B_{j(i)})$ dans le tableau suivant :

A_1			...			A_I		
$B_{1(1)}$	...	$B_{J(1)}$	...	...	...	$B_{1(I)}$	...	$B_{J(I)}$
$Y_{1,1,1}$	...	$Y_{1,J,1}$	...	...	...	$Y_{I,1,1}$	...	$Y_{I,J,1}$
$\vdots$	$\vdots$	$\vdots$	...	...	...	$\vdots$	$\vdots$	$\vdots$
$Y_{1,1,K}$	...	$Y_{1,J,K}$	...	...	...	$Y_{I,1,K}$	...	$Y_{I,J,K}$

On rappelle que la variation théorique due au facteur A est définie par :

$$SC_A = JK \sum_{i=1}^{I} (Y_{i,\bullet,\bullet} - Y_{\bullet,\bullet,\bullet})^2.$$

La variation théorique du facteur B dans le facteur A est définie par :

$$SC_{B|A} = K \sum_{j=1}^{J} (Y_{i,j,\bullet} - Y_{i,\bullet,\bullet})^2.$$

La variation résiduelle théorique est quant à elle définie par :

$$SC_R = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(Y_{i,j,k} - Y_{i,j,\bullet}\right)^2.$$

Enfin la variation totale théorique est égale à :

$$SC_{TOT} = \sum_{i=1}^{I}\sum_{j=1}^{J}(\sum_{k=1}^{K}(Y_{i,j,k} - Y_{\bullet,\bullet,\bullet})^2.$$

On rappelle la relation fondamentale de l'ANOVA :

$$SC_{TOT} = SC_A + SC_{B|A} + SC_R.$$

La liste $\boldsymbol{y}$ des données expérimentales $y_{1,1,1}, \ldots, y_{1,1,K}, \ldots, y_{1,2,1}, \ldots, y_{1,2,K}, \ldots, y_{I,J,K}$ permet de construire une réalisation du tableau précédent :

A_1			$\ldots$			A_I		
$B_{1(1)}$	$\ldots$	$B_{J(1)}$	$\ldots$	$\ldots$	$\ldots$	$B_{1(I)}$	$\ldots$	$B_{J(I)}$
$y_{1,1,1}$	$\ldots$	$y_{1,J,1}$	$\ldots$	$\ldots$	$\ldots$	$y_{I,1,1}$	$\ldots$	$y_{I,J,1}$
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\ldots$	$\ldots$	$\vdots$	$\vdots$	$\vdots$
$y_{1,1,K}$	$\ldots$	$y_{1,J,K}$	$\ldots$	$\ldots$	$\ldots$	$y_{I,1,K}$	$\ldots$	$y_{I,J,K}$

La variation due au facteur A observée sur la liste de données $\boldsymbol{y}$ est définie par :

$$sc_A = JK\sum_{i=1}^{I}(y_{i,\bullet,\bullet} - y_{\bullet,\bullet,\bullet})^2.$$

La variation du facteur B dans le facteur A observée sur la liste de données $\boldsymbol{y}$ est définie par :

$$sc_{B|A} = K\sum_{j=1}^{J}(y_{i,j,\bullet} - y_{i,\bullet,\bullet})^2.$$

La variation résiduelle observée sur la liste de données $\boldsymbol{y}$ est quant à elle définie par :

$$sc_R = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(y_{i,j,k} - y_{i,j,\bullet}\right)^2.$$

Enfin la variation totale observée sur la liste de données $\boldsymbol{y}$ est égale à :

$$sc_{TOT} = \sum_{i=1}^{I}\sum_{j=1}^{J}(\sum_{k=1}^{K}(y_{i,j,k} - y_{\bullet,\bullet,\bullet})^2.$$

La relation fondamentale de l'ANOVA reste valable lorsqu'elle est évaluée sur la liste de données $\boldsymbol{y}$:

$$sc_{TOT} = sc_A + sc_{B|A} + sc_R.$$

On introduit les dégres de liberté (Ddl) associés à chaque ligne du tableau de l'ANOVA :

Source	Degrés de liberté
Facteur A	$n_A = I - 1$
Facteur B dans A	$n_{B\|A} = I(J-1)$
Résiduelle	$n_R = IJ(K-1)$
Totale	$n_{TOT} = IJK - 1$

On résume ces informations dans le tableau de l'ANOVA ci-dessous :

Source	Variation	Ddl	Carré Moyen	F	Décision
Facteur A	sc_A	n_A	$s_A^2 = \dfrac{sc_A}{n_A}$	$f_A = \dfrac{s_A^2}{s_R^2}$	$\mathcal{H}_0'$ ou $\mathcal{H}_1'$
Facteur B dans A	$sc_{B\|A}$	$n_{B\|A}$	$s_{B\|A}^2 = \dfrac{sc_{B\|A}}{n_{B\|A}}$	$f_{B\|A} = \dfrac{s_{B\|A}^2}{s_R^2}$	$\mathcal{H}_0''$ ou $\mathcal{H}_1''$
Résiduelle	sc_R	n_R	$s_R^2 = \dfrac{sc_R}{n_R}$		
Totale	sc_{TOT}	n_{TOT}			

On souhaite faire les tests d'hypothèse suivants :

$$\boxed{\mathcal{H}_0' : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0}$$

contre

$$\boxed{\mathcal{H}_1' : \text{Il existe } i_0 \in \{1, 2, \ldots, I\} \text{ tel que } \alpha_{i_0} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{'}$ précédente d'absence d'effet du facteur A et lorsque les conditions de validité du modèle sont respectées, f_A est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $I-1$ et $IJ(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_A est supérieure ou égale à la valeur critique issue de la table. Lorsque l'hypothèse nulle $\mathcal{H}_0^{'}$ est rejetée on peut procéder à des comparaisons multiples des différents effets des niveaux du facteur voir la section 10.

$$\boxed{\mathcal{H}_0^{''} : \beta_{1(1)} = \beta_{2(1)} = \cdots = \beta_{J(1)} = \beta_{1(2)} = \cdots = \beta_{J(I)} = 0}$$

contre

$$\boxed{\mathcal{H}_1^{''} : \text{Il existe } (i_0, j_0) \in \{1, 2, \ldots, I\} \times \{1, 2, \ldots, J\} \text{ tel que } \beta_{j_0(i_0)} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{''}$ précédente d'absence d'effet des facteurs B dans le facteur A et lorsque les conditions de validité du modèle sont respectées, $f_{B|A}$ est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $I(J-1)$ et $IJ(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur $f_{B|A}$ est supérieure ou égale à la valeur critique issue de la table.

Les estimateurs $\widehat{\mu}$, $\widehat{\alpha_1}$, $\ldots$, $\widehat{\alpha_I}$, $\widehat{\beta_{1(1)}}$, $\widehat{\beta_{2(1)}}$, $\ldots$, $\widehat{\beta_{J(1)}}$, $\widehat{\beta_{1(2)}}$, $\ldots$, $\widehat{\beta_{J(I)}}$, $\widehat{\sigma^2}$ des paramètres μ, α_1, $\ldots$, α_I, $\beta_{1(1)}$, $\beta_{2(1)}$, $\ldots$, $\beta_{J(1)}$, $\beta_{1(2)}$, $\ldots$, $\beta_{J(I)}$, σ^2 du modèle sont donnés par les formules suivantes :

$$\begin{aligned}
&\widehat{\mu} = Y_{\bullet,\bullet,\bullet} = \overline{Y},\ \widehat{\alpha_i} = Y_{i,\bullet,\bullet} - \widehat{\mu},\ 1 \leqslant i \leqslant I,\\
&\widehat{\beta_{j(i)}} = Y_{i,j,\bullet} - Y_{i,\bullet,\bullet},\ 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\\
&\widehat{\sigma^2} = \frac{SC_R}{IJ(K-1)} = S_R^2.
\end{aligned}$$

Ce sont des estimateurs sans biais.

Les estimations, obtenues pour la liste de données $\boldsymbol{y}$ et notées $\widehat{\mu}(\boldsymbol{y})$, $\widehat{\alpha_1}(\boldsymbol{y})$, $\ldots$, $\widehat{\alpha_I}(\boldsymbol{y})$, $\widehat{\beta_{1(1)}}(\boldsymbol{y})$, $\widehat{\beta_{2(1)}}(\boldsymbol{y})$, $\ldots$, $\widehat{\beta_{J(1)}}(\boldsymbol{y})$, $\widehat{\beta_{1(2)}}(\boldsymbol{y})$, $\ldots$, $\widehat{\beta_{J(I)}}(\boldsymbol{y})$, $\widehat{\sigma^2}(\boldsymbol{y})$, des paramètres μ, α_1, $\ldots$, α_I, $\beta_{1(1)}$, $\beta_{2(1)}$, $\ldots$, $\beta_{J(1)}$, $\beta_{1(2)}$, $\ldots$, $\beta_{J(I)}$, σ^2 du modèle se déduisent des formules suivantes :

$$\begin{aligned}
&\widehat{\mu}(\boldsymbol{y}) = y_{\bullet,\bullet,\bullet} = \overline{y},\ \widehat{\alpha_i}(\boldsymbol{y}) = y_{i,\bullet,\bullet} - \widehat{\mu}(\boldsymbol{y}),\ 1 \leqslant i \leqslant I,\\
&\widehat{\beta_{j(i)}}(\boldsymbol{y}) = y_{i,j,\bullet} - y_{i,\bullet,\bullet},\ 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\\
&\widehat{\sigma^2}(\boldsymbol{y}) = \frac{sc_R}{IJ(K-1)} = s_R^2.
\end{aligned}$$

5.2. Modèles à effets aléatoires

5.2.1. Avec répétitions

Un facteur contrôlé A se présente sous I modalités, chacune d'entre elles étant notée A_i. Les $\beta_{j(i)}$ représentent un échantillon de taille J prélevé dans une population importante dépendant du niveau A_i du facteur A. Nous admettrons que les effets des $B_{j(i)}$, les $\beta_{j(i)}$, sont distribués suivant une loi normale centrée de variance $\sigma^2_{B|A}$. Pour chacun des couples de modalités $(A_i, B_{j(i)})$ on effectue $K \geqslant 2$ mesures d'une réponse Y qui est une variable continue. On note $n = I \times J \times K$ le nombre total de mesures ayant été effectuées.

On introduit le modèle :

$$Y_{i,j,k} = \mu + \alpha_i + \beta_{j(i)} + \epsilon_{i,j,k}, \quad i = 1 \dots I,\ j = 1 \dots J,\ k = 1 \dots K$$

où $Y_{i,j,k}$ est la valeur prise par la réponse Y dans les conditions $(A_i, B_{j(i)})$ lors du $k-$ème essai. On suppose que

$$\mathcal{L}\left(\alpha_i\right) = \mathcal{N}(0, \sigma^2_A),\ \forall\ i, 1 \leqslant i \leqslant I,$$
$$\mathcal{L}\left(\beta_{j(i)}\right) = \mathcal{N}(0, \sigma^2_{B|A}),\ \forall\ (i,j), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,$$

ainsi que l'indépendance des effets aléatoires :

$$\mathrm{Cov}(\alpha_i, \alpha_j) = 0 \text{ si } i \neq j \text{ et } 1 \leqslant i, j \leqslant I,$$
$$\mathrm{Cov}(\beta_{j(i)}, \beta_{l(k)}) = 0 \text{ si } (i,j) \neq (k,l) \text{ avec } 1 \leqslant i, k \leqslant I \text{ et } 1 \leqslant j, l \leqslant J,$$
$$\mathrm{Cov}(\alpha_i, \beta_{k(j)}) = 0 \text{ si } 1 \leqslant i, j \leqslant I \text{ et } 1 \leqslant k \leqslant J.$$

On postule les hypothèses classiques suivantes pour les erreurs :

$$\forall\ (i,j,k), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K,\ \mathcal{L}(\epsilon_{i,j,k}) = \mathcal{N}(0, \sigma^2),$$
$$\text{et } \mathrm{Cov}(\epsilon_{i,j,k}, \epsilon_{l,m,n}) = 0 \text{ si } (i,j,k) \neq (l,m,n) \text{ avec}$$
$$1 \leqslant i, l \leqslant I,\ 1 \leqslant j, m \leqslant J \text{ et } 1 \leqslant k, n \leqslant K,$$

ainsi que l'indépendance des effets aléatoires et des erreurs :

$$\mathrm{Cov}(\alpha_i, \epsilon_{j,k,l}) = 0 \text{ si } 1 \leqslant i, j \leqslant I, 1 \leqslant k \leqslant J, \text{ et } 1 \leqslant l \leqslant K,$$
$$\mathrm{Cov}(\beta_{j(i)}, \epsilon_{k,l,m}) = 0 \text{ si } 1 \leqslant i, k \leqslant I, 1 \leqslant j, l \leqslant J, \text{ et } 1 \leqslant m \leqslant K.$$

On suppose que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fera l'objet d'un autre paragraphe.

On utilise les quantités SC_A, $SC_{B|A}$, SC_R, SC_{TOT}, sc_A, $sc_{B|A}$, sc_R et sc_{TOT} introduites à la section 5.1.1.

On rappelle la relation fondamentale de l'ANOVA :

$$SC_{TOT} = SC_A + SC_{B|A} + SC_R.$$

On introduit les dégres de liberté (Ddl) associés à chaque ligne du tableau de l'ANOVA :

Source	Degrés de liberté
Facteur A	$n_A = I - 1$
Facteur B dans A	$n_{B\|A} = I(J-1)$
Résiduelle	$n_R = IJ(K-1)$
Totale	$n_{TOT} = IJK - 1$

On résume ces informations dans le tableau de l'ANOVA ci-dessous :

Source	Variation	Ddl	Carré Moyen	F	Décision
Facteur A	sc_A	n_A	$s_A^2 = \dfrac{sc_A}{n_A}$	$f_A = \dfrac{s_A^2}{s_{B\|A}^2}$	$\mathcal{H}_0^{'}$ ou $\mathcal{H}_1^{'}$
Facteur B dans A	$sc_{B\|A}$	$n_{B\|A}$	$s_{B\|A}^2 = \dfrac{sc_{B\|A}}{n_{B\|A}}$	$f_{B\|A} = \dfrac{s_{B\|A}^2}{s_R^2}$	$\mathcal{H}_0^{''}$ ou $\mathcal{H}_1^{''}$
Résiduelle	sc_R	n_R	$s_R^2 = \dfrac{sc_R}{n_R}$		
Totale	sc_{TOT}	n_{TOT}			

On souhaite faire les tests d'hypothèse suivants :

$$\boxed{\mathcal{H}_0^{'} : \sigma_A^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1^{'} : \sigma_A^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{'}$ précédente d'absence d'effet du facteur A et lorsque les conditions de validité du modèle sont respectées, f_A est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $I-1$ et $I(J-1)$ degrés de liberté.

$$\boxed{\mathcal{H}_0^{''} : \sigma_{B|A}^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1'' : \sigma^2_{B|A} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0''$ précédente d'absence d'effet du facteur B dans le facteur A et lorsque les conditions de validité du modèle sont respectées, $f_{B|A}$ est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $I(J-1)$ et $IJ(K-1)$ degrés de liberté.

Les estimateurs $\widehat{\mu}$, $\widehat{\sigma^2_A}$, $\widehat{\sigma^2_{B|A}}$, $\widehat{\sigma^2}$ des paramètres μ, σ^2_A, $\sigma^2_{B|A}$, σ^2 du modèle se déduisent des formules suivantes :

$$\begin{aligned}
&\widehat{\mu} = Y_{\bullet,\bullet,\bullet} = \overline{Y},\ \widehat{\sigma^2_A} = \frac{1}{JK}\left(S^2_A - S^2_{B|A}\right),\\
&\widehat{\sigma^2_{B|A}} = \frac{1}{K}\left(S^2_{B|A} - S^2_R\right),\\
&\widehat{\sigma^2} = \frac{SC_R}{(I-1)(J-1)} = S^2_R,
\end{aligned}$$

où $S^2_A = \dfrac{SC_A}{n_A}$, $S^2_{B|A} = \dfrac{SC_{B|A}}{n_{B|A}}$ et $S^2_R = \dfrac{SC_R}{n_R}$.
Ce sont des estimateurs sans biais.

Les estimations, obtenues pour la liste de données $\boldsymbol{y}$ et notées $\widehat{\mu}(\boldsymbol{y})$, $\widehat{\sigma^2_A}(\boldsymbol{y})$, $\widehat{\sigma^2_{B|A}}(\boldsymbol{y})$, $\widehat{\sigma^2}(\boldsymbol{y})$, des paramètres μ, σ^2_A, $\sigma^2_{B|A}$, σ^2 du modèle se déduisent des formules ci-dessus :

$$\begin{aligned}
&\widehat{\mu}(\boldsymbol{y}) = y_{\bullet,\bullet,\bullet} = \overline{y},\ \widehat{\sigma^2_A}(\boldsymbol{y}) = \frac{1}{JK}\left(s^2_A - s^2_{B|A}\right),\\
&\widehat{\sigma^2_{B|A}}(\boldsymbol{y}) = \frac{1}{K}\left(s^2_{B|A} - s^2_R\right),\\
&\widehat{\sigma^2}(\boldsymbol{y}) = \frac{sc_R}{(I-1)(J-1)} = s^2_R.
\end{aligned}$$

5.3. Modèles à effets mixtes

Pour le plupart des auteurs d'ouvrages sur l'analyse de la variance, voir [4] à ce sujet par exemple, un facteur emboîté dans un facteur aléatoire doit être considéré comme aléatoire[4]. Ainsi le seul modèle mixte possible est le cas où le facteur A est fixe et le facteur que l'on emboîte dans A, le facteur B, est aléatoire.

5.3.1. Avec répétitions

Un facteur contrôlé A se présente sous I modalités, chacune d'entre elles étant notée A_i. Les β_j représentent un échantillon de taille J prélevé dans une population importante.

[4]Il existe néanmoins certains cas où l'on peut utiliser un modèle mixte où le facteur emboîté est à effets fixes tandis que le facteur dans lequel il est emboîté est à effets aléatoires.

Nous admettrons que les effets des $B_{j(i)}$, les $\beta_{j(i)}$, sont distribués suivant une loi normale centrée de variance $\sigma^2_{B|A}$. Pour chacun des couples de modalités $(A_i, B_{j(i)})$ on effectue $K \geqslant 2$ mesures d'une réponse Y qui est une variable continue. On note $n = I \times J \times K$ le nombre total de mesures ayant été effectuées.

On introduit le modèle :

$$Y_{i,j,k} = \mu + \alpha_i + \beta_{j(i)} + \epsilon_{i,j,k}, \quad i = 1 \dots I,\ j = 1 \dots J,\ k = 1 \dots K$$
$$\text{avec les contraintes supplémentaires } \sum_{i=1}^{I} \alpha_i = 0,$$

où $Y_{i,j,k}$ est la valeur prise par la réponse Y dans les conditions $(A_i, B_{j(i)})$ lors du $k-$ème essai. On suppose que

$$\mathcal{L}\left(\beta_{j(i)}\right) = \mathcal{N}(0, \sigma^2_{B|A}),\ \forall\ (i,j), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,$$

ainsi que l'indépendance des effets aléatoires :

$$\mathrm{Cov}(\beta_{j(i)}, \beta_{l(k)}) = 0 \text{ si } (i,j) \neq (k,l) \text{ avec } 1 \leqslant i,k \leqslant I \text{ et } 1 \leqslant j,l \leqslant J.$$

On postule les hypothèses classiques suivantes pour les erreurs :

$$\forall\ (i,j,k), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K,\ \mathcal{L}(\epsilon_{i,j,k}) = \mathcal{N}(0, \sigma^2),$$
$$\text{et } \mathrm{Cov}(\epsilon_{i,j,k}, \epsilon_{l,m,n}) = 0 \text{ si } (i,j,k) \neq (l,m,n) \text{ avec}$$
$$1 \leqslant i,l \leqslant I,\ 1 \leqslant j,m \leqslant J \text{ et } 1 \leqslant k,n \leqslant K,$$

ainsi que l'indépendance des effets aléatoires et des erreurs :

$$\mathrm{Cov}(\beta_{j(i)}, \epsilon_{k,l,m}) = 0 \text{ si } 1 \leqslant i,k \leqslant I, 1 \leqslant j,l \leqslant J, \text{ et } 1 \leqslant m \leqslant K.$$

On suppose que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fera l'objet d'un autre paragraphe.

On utilise les quantités SC_A, $SC_{B|A}$, SC_R, SC_{TOT}, sc_A, $sc_{B|A}$, sc_R et sc_{TOT} introduites à la section 5.1.1.

On rappelle la relation fondamentale de l'ANOVA :

$$SC_{TOT} = SC_A + SC_{B|A} + SC_R.$$

On introduit les dégres de liberté (Ddl) associés à chaque ligne du tableau de l'ANOVA :

Source	Degrés de liberté
Facteur A	$n_A = I - 1$
Facteur B dans A	$n_{B\|A} = I(J-1)$
Résiduelle	$n_R = IJ(K-1)$
Totale	$n_{TOT} = IJK - 1$

On résume ces informations dans le tableau de l'ANOVA ci-dessous :

Source	Variation	Ddl	Carré Moyen	F	Décision
Facteur A	sc_A	n_A	$s_A^2 = \dfrac{sc_A}{n_A}$	$f_A = \dfrac{s_A^2}{s_{B\|A}^2}$	$\mathcal{H}_0^{'}$ ou $\mathcal{H}_1^{'}$
Facteur B dans A	$sc_{B\|A}$	$n_{B\|A}$	$s_{B\|A}^2 = \dfrac{sc_{B\|A}}{n_{B\|A}}$	$f_{B\|A} = \dfrac{s_{B\|A}^2}{s_R^2}$	$\mathcal{H}_0^{''}$ ou $\mathcal{H}_1^{''}$
Résiduelle	sc_R	n_R	$s_R^2 = \dfrac{sc_R}{n_R}$		
Totale	sc_{TOT}	n_{TOT}			

On souhaite faire les tests d'hypothèse suivants :

$$\boxed{\mathcal{H}_0^{'} : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0}$$

contre

$$\boxed{\mathcal{H}_1^{'} : \text{Il existe } i_0 \in \{1, 2, \ldots, I\} \text{ tel que } \alpha_{i_0} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{'}$ précédente d'absence d'effet du facteur A et lorsque les conditions de validité du modèle sont respectées, f_A est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $I-1$ et $I(J-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_A est supérieure ou égale à la valeur critique issue de la table. Lorsque l'hypothèse nulle $\mathcal{H}_0^{'}$ est rejetée on peut procéder à des comparaisons multiples des différents effets des niveaux du facteur voir la section 10.

$$\boxed{\mathcal{H}_0^{''} : \sigma_{B|A}^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1^{''} : \sigma_{B|A}^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0''$ précédente d'absence d'effet du facteur B dans A et lorsque les conditions de validité du modèle sont respectées, $f_{B|A}$ est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $I(J-1)$ et $IJ(K-1)$ degrés de liberté.

Les estimateurs $\widehat{\mu}$, $\widehat{\alpha_1}$, ..., $\widehat{\alpha_I}$, $\widehat{\sigma^2_{B|A}}$, $\widehat{\sigma^2}$ des paramètres μ, α_1, ..., α_I, $\sigma^2_{B|A}$, σ^2 du modèle sont donnés par les formules suivantes :

$$\begin{aligned}
&\widehat{\mu} = Y_{\bullet,\bullet,\bullet} = \overline{Y},\ \widehat{\alpha}_i = Y_{i,\bullet,\bullet} - \widehat{\mu},\ 1 \leqslant i \leqslant I,\\
&\widehat{\sigma^2_{B|A}} = \frac{1}{K}\left(S^2_{B|A} - S^2_R\right),\\
&\widehat{\sigma^2} = \frac{SC_R}{(I-1)(J-1)} = S^2_R,
\end{aligned}$$

où $S^2_{B|A} = \dfrac{SC_{B|A}}{n_{B|A}}$ et $S^2_R = \dfrac{SC_R}{n_R}$.
Ce sont des estimateurs sans biais.

Les estimations, obtenues pour la liste de données $\boldsymbol{y}$ et notées $\widehat{\mu}(\boldsymbol{y})$, $\widehat{\alpha_1}(\boldsymbol{y})$, ..., $\widehat{\alpha_I}(\boldsymbol{y})$, $\widehat{\sigma^2_{B|A}}(\boldsymbol{y})$, $\widehat{\sigma^2}(\boldsymbol{y})$, des paramètres μ, α_1, ..., α_I, $\sigma^2_{B|A}$, σ^2 du modèle se déduisent des formules ci-dessus :

$$\begin{aligned}
&\widehat{\mu}(\boldsymbol{y}) = y_{\bullet,\bullet,\bullet} = \overline{y},\ \widehat{\alpha}_i(\boldsymbol{y}) = y_{i,\bullet,\bullet} - \widehat{\mu}(\boldsymbol{y}),\ 1 \leqslant i \leqslant I,\\
&\widehat{\sigma^2_{B|A}}(\boldsymbol{y}) = \frac{1}{K}\left(s^2_{B|A} - s^2_R\right),\\
&\widehat{\sigma^2}(\boldsymbol{y}) = \frac{sc_R}{(I-1)(J-1)} = s^2_R.
\end{aligned}$$

6. Analyse de la variance à trois facteurs

Dans toute cette section, on utilise les notations définies à la section 1.

6.1. Modèles à effets fixes

6.1.1. Sans répétition

Un facteur contrôlé A se présente sous I modalités, chacune d'entre elles étant notée A_i. Un facteur contrôlé B se présente sous J modalités, chacune d'entre elles étant notée B_j. Un facteur contrôlé C se présente sous K modalités, chacune d'entre elles étant notée C_k. Pour chacun des couples de modalités (A_i, B_j, C_k) on effectue une mesure d'une réponse Y qui est une variable continue. On note $n = I \times J \times K$ le nombre total de mesures ayant été effectuées.

On introduit le modèle :

$$Y_{i,j,k} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{i,j} + (\alpha\gamma)_{i,k} + (\beta\gamma)_{j,k} + \epsilon_{i,j,k},$$
$$i = 1 \dots I,\ j = 1 \dots J,\ k = 1 \dots K,$$
$$\text{avec les contraintes supplémentaires } \sum_{i=1}^{I} \alpha_i = 0,\ \sum_{j=1}^{J} \beta_j = 0 \text{ et } \sum_{k=1}^{K} \gamma_k = 0,$$
$$\sum_{i=1}^{I} (\alpha\beta)_{i,j} = 0,\ \forall j \in \{1, \dots, J\} \text{ et } \sum_{j=1}^{J} (\alpha\beta)_{i,j} = 0,\ \forall i \in \{1, \dots, I\},$$
$$\sum_{i=1}^{I} (\alpha\gamma)_{i,k} = 0,\ \forall k \in \{1, \dots, K\} \text{ et } \sum_{k=1}^{K} (\alpha\gamma)_{i,k} = 0,\ \forall i \in \{1, \dots, I\},$$
$$\sum_{j=1}^{J} (\beta\gamma)_{j,k} = 0,\ \forall k \in \{1, \dots, K\} \text{ et } \sum_{k=1}^{K} (\beta\gamma)_{j,k} = 0,\ \forall j \in \{1, \dots, J\},$$

où $Y_{i,j,k}$ est la valeur prise par la réponse Y dans les conditions (A_i, B_j, C_k). On postule les hypothèses classiques suivantes pour les erreurs :

$$\forall\ (i,j,k), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K,\ \mathcal{L}(\epsilon_{i,j,k}) = \mathcal{N}(0, \sigma^2),$$
$$\mathrm{Cov}(\epsilon_{i,j,k}, \epsilon_{l,m,n}) = 0 \text{ si } (i,j,k) \neq (l,m,n) \text{ avec } 1 \leqslant i, l \leqslant I,\ 1 \leqslant j, m \leqslant J \text{ et } 1 \leqslant k, n \leqslant K.$$

On suppose que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fera l'objet d'un autre paragraphe.

On rappelle que la variation due au facteur A est définie par :

$$sc_A = JK \sum_{i=1}^{I} (y_{i,\bullet,\bullet} - y_{\bullet,\bullet,\bullet})^2.$$

On rappelle que la variation due au facteur B est définie par :

$$sc_B = IK \sum_{j=1}^{J} (y_{\bullet,j,\bullet} - y_{\bullet,\bullet,\bullet})^2.$$

On rappelle que la variation due au facteur C est définie par :

$$sc_C = IJ \sum_{k=1}^{K} (y_{\bullet,\bullet,k} - y_{\bullet,\bullet,\bullet})^2.$$

On rappelle que la variation due à l'interaction des facteurs A et B est définie par :

$$sc_{AB} = K \sum_{i=1}^{I} \sum_{j=1}^{J} (y_{i,j,\bullet} - y_{i,\bullet,\bullet} - y_{\bullet,j,\bullet} + y_{\bullet,\bullet,\bullet})^2.$$

On rappelle que la variation due à l'interaction des facteurs B et C est définie par :

$$sc_{BC} = I \sum_{j=1}^{J} \sum_{k=1}^{K} (y_{\bullet,j,k} - y_{\bullet,j,\bullet} - y_{\bullet,\bullet,k} + y_{\bullet,\bullet,\bullet})^2.$$

On rappelle que la variation due à l'interaction des facteurs A et C est définie par :

$$sc_{AC} = J \sum_{i=1}^{I} \sum_{k=1}^{K} (y_{i,\bullet,k} - y_{i,\bullet,\bullet} - y_{\bullet,\bullet,k} + y_{\bullet,\bullet,\bullet})^2.$$

La variation résiduelle est quant à elle définie par :

$$sc_R = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (y_{i,j,k} - y_{i,j,\bullet} - y_{i,\bullet,k} - y_{\bullet,j,k} + y_{i,\bullet,\bullet} + y_{\bullet,j,\bullet} + y_{\bullet,\bullet,k} - y_{\bullet,\bullet,\bullet})^2.$$

Enfin la variation totale est égale à :

$$sc_{TOT} = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (y_{i,j,k} - y_{\bullet,\bullet,\bullet})^2.$$

On rappelle la relation fondamentale de l'ANOVA :

$$sc_{TOT} = sc_A + sc_B + sc_C + sc_{AB} + sc_{AC} + sc_{BC} + sc_R.$$

On introduit les dégres de liberté (Ddl) associés à chaque ligne du tableau de l'ANOVA :

Source	Degrés de liberté
Facteur A	$n_A = I - 1$
Facteur B	$n_B = J - 1$
Facteur C	$n_C = K - 1$
Interaction AB	$n_{AB} = (I-1)(J-1)$
Interaction AC	$n_{AC} = (I-1)(K-1)$
Interaction BC	$n_{BC} = (J-1)(K-1)$
Résiduelle	$n_R = (I-1)(J-1)(K-1)$
Totale	$n_{TOT} = IJK - 1$

On résume ces informations dans le tableau de l'ANOVA ci-dessous :

Source	Variation	Ddl	Carré Moyen	F	Décision
Facteur A	sc_A	n_A	$s_A^2 = \frac{sc_A}{n_A}$	$f_A = \frac{s_A^2}{s_R^2}$	$\mathcal{H}_0^{'}$ ou $\mathcal{H}_1^{'}$
Facteur B	sc_B	n_B	$s_B^2 = \frac{sc_B}{n_B}$	$f_B = \frac{s_B^2}{s_R^2}$	$\mathcal{H}_0^{''}$ ou $\mathcal{H}_1^{''}$
Facteur C	sc_C	n_C	$s_C^2 = \frac{sc_C}{n_C}$	$f_C = \frac{s_C^2}{s_R^2}$	$\mathcal{H}_0^{'''}$ ou $\mathcal{H}_1^{'''}$
Interaction AB	sc_{AB}	n_{AB}	$s_{AB}^2 = \frac{sc_{AB}}{n_{AB}}$	$f_{AB} = \frac{s_{AB}^2}{s_R^2}$	$\mathcal{H}_0^{(4)}$ ou $\mathcal{H}_1^{(4)}$
Interaction AC	sc_{AC}	n_{AC}	$s_{AC}^2 = \frac{sc_{AC}}{n_{AC}}$	$f_{AC} = \frac{s_{AC}^2}{s_R^2}$	$\mathcal{H}_0^{(5)}$ ou $\mathcal{H}_1^{(5)}$
Interaction BC	sc_{BC}	n_{BC}	$s_{BC}^2 = \frac{sc_{BC}}{n_{BC}}$	$f_{BC} = \frac{s_{BC}^2}{s_R^2}$	$\mathcal{H}_0^{(6)}$ ou $\mathcal{H}_1^{(6)}$
Résiduelle	sc_R	n_R	$s_R^2 = \frac{sc_R}{n_R}$		
Totale	sc_{TOT}	n_{TOT}			

On souhaite faire les tests d'hypothèse suivants :

$$\boxed{\mathcal{H}_0^{'} : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0}$$

contre

$$\boxed{\mathcal{H}_1^{'} : \text{Il existe } i_0 \in \{1, 2, \ldots, I\} \text{ tel que } \alpha_{i_0} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{'}$ précédente d'absence d'effet du facteur A et lorsque les conditions de validité du modèle sont respectées, f_A est la réalisation d'une variable aléatoire

qui suit une loi de Fisher à $I-1$ et $(I-1)(J-1)(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_A est supérieure ou égale à la valeur critique issue de la table. Lorsque l'hypothèse nulle $\mathcal{H}_0^{'}$ est rejetée on peut procéder à des comparaisons multiples des différents effets des niveaux du facteur voir la section 10.

$$\boxed{\mathcal{H}_0^{''} : \beta_1 = \beta_2 = \cdots = \beta_J = 0}$$

contre

$$\boxed{\mathcal{H}_1^{''} : \text{Il existe } j_0 \in \{1, 2, \ldots, J\} \text{ tel que } \beta_{j_0} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{''}$ précédente d'absence d'effet du facteur B et lorsque les conditions de validité du modèle sont respectées, f_B est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $J-1$ et $(I-1)(J-1)(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_B est supérieure ou égale à la valeur critique issue de la table. Lorsque l'hypothèse nulle $\mathcal{H}_0^{''}$ est rejetée on peut procéder à des comparaisons multiples des différents effets des niveaux du facteur voir la section 10.

$$\boxed{\mathcal{H}_0^{'''} : \gamma_1 = \gamma_2 = \cdots = \gamma_K = 0}$$

contre

$$\boxed{\mathcal{H}_1^{'''} : \text{Il existe } k_0 \in \{1, 2, \ldots, K\} \text{ tel que } \gamma_{k_0} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{'''}$ précédente d'absence d'effet du facteur C et lorsque les conditions de validité du modèle sont respectées, f_C est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $K-1$ et $(I-1)(J-1)(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_C est supérieure ou égale à la valeur critique issue de la table. Lorsque l'hypothèse nulle $\mathcal{H}_0^{'''}$ est rejetée on peut procéder à des comparaisons multiples des différents effets des niveaux du facteur voir la section 10.

$$\boxed{\mathcal{H}_0^{(4)} : (\alpha\beta)_{1,1} = (\alpha\beta)_{1,2} = \cdots = (\alpha\beta)_{1,J} = (\alpha\beta)_{2,1} = \cdots = (\alpha\beta)_{I,J} = 0}$$

contre

$$\boxed{\mathcal{H}_1^{(4)} : \text{Il existe } (i_0, j_0) \in \{1, 2, \ldots, I\} \times \{1, 2, \ldots, J\} \text{ tel que } (\alpha\beta)_{i_0,j_0} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{(4)}$ précédente d'absence d'effet de l'interaction des facteurs A et B et lorsque les conditions de validité du modèle sont respectées, f_{AB} est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $(I-1)(J-1)$ et $(I-1)(J-1)(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou

égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_{AB} est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0^{(5)} : (\alpha\gamma)_{1,1} = (\alpha\gamma)_{1,2} = \cdots = (\alpha\gamma)_{1,K} = (\alpha\gamma)_{2,1} = \cdots = (\alpha\gamma)_{I,K} = 0}$$

contre

$$\boxed{\mathcal{H}_1^{(5)} : \text{Il existe } (i_0, k_0) \in \{1, 2, \ldots, I\} \times \{1, 2, \ldots, K\} \text{ tel que } (\alpha\gamma)_{i_0,k_0} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{(5)}$ précédente d'absence d'effet de l'interaction des facteurs A et C et lorsque les conditions de validité du modèle sont respectées, f_{AC} est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $(I-1)(K-1)$ et $(I-1)(J-1)(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_{AC} est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0^{(6)} : (\beta\gamma)_{1,1} = (\beta\gamma)_{1,2} = \cdots = (\beta\gamma)_{1,K} = (\beta\gamma)_{2,1} = \cdots = (\beta\gamma)_{J,K} = 0}$$

contre

$$\boxed{\mathcal{H}_1^{(6)} : \text{Il existe } (j_0, k_0) \in \{1, 2, \ldots, J\} \times \{1, 2, \ldots, K\} \text{ tel que } (\beta\gamma)_{j_0,k_0} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{(6)}$ précédente d'absence d'effet de l'interaction des facteurs B et C et lorsque les conditions de validité du modèle sont respectées, f_{BC} est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $(J-1)(K-1)$ et $(I-1)(J-1)(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_{BC} est supérieure ou égale à la valeur critique issue de la table.

Les estimations $\widehat{\mu}$, $\widehat{\alpha_1}$, $\ldots$, $\widehat{\alpha_I}$, $\widehat{\beta_1}$, $\ldots$, $\widehat{\beta_J}$, $\widehat{\gamma_1}$, $\ldots$, $\widehat{\gamma_K}$, $\widehat{(\alpha\beta)_{1,1}}$, $\ldots$, $\widehat{(\alpha\beta)_{I,J}}$, $\widehat{(\alpha\gamma)_{1,1}}$, $\ldots$, $\widehat{(\alpha\gamma)_{I,K}}$, $\widehat{(\beta\gamma)_{1,1}}$, $\ldots$, $\widehat{(\beta\gamma)_{J,K}}$, $\widehat{\sigma^2}$ des paramètres μ, α_1, $\ldots$, α_I, β_1, $\ldots$, β_J, γ_1, $\ldots$, γ_K, $(\alpha\beta)_{1,1}$, $\ldots$, $(\alpha\beta)_{I,J}$, $(\alpha\gamma)_{1,1}$, $\ldots$, $(\alpha\gamma)_{I,K}$, $(\beta\gamma)_{1,1}$, $\ldots$, $(\beta\gamma)_{J,K}$, σ^2 du modèle se déduisent des formules suivantes :

$$\begin{aligned}
&\widehat{\mu} = y_{\bullet,\bullet,\bullet} = \overline{y},\\
&\widehat{\alpha_i} = y_{i,\bullet,\bullet} - \widehat{\mu},\ 1 \leqslant i \leqslant I,\ \widehat{\beta_j} = y_{\bullet,j,\bullet} - \widehat{\mu},\ 1 \leqslant j \leqslant J,\ \widehat{\gamma_k} = y_{\bullet,\bullet,k} - \widehat{\mu},\ 1 \leqslant k \leqslant K,\\
&\widehat{(\alpha\beta)_{i,j}} = y_{i,j,\bullet} - y_{i,\bullet,\bullet} - y_{\bullet,j,\bullet} + \widehat{\mu},\ 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\\
&\widehat{(\alpha\gamma)_{i,k}} = y_{i,\bullet,k} - y_{i,\bullet,\bullet} - y_{\bullet,\bullet,k} + \widehat{\mu},\ 1 \leqslant i \leqslant I,\ 1 \leqslant k \leqslant K,\\
&\widehat{(\beta\gamma)_{j,k}} = y_{\bullet,j,k} - y_{\bullet,j,\bullet} - y_{\bullet,\bullet,k} + \widehat{\mu},\ 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K,\\
&\widehat{\sigma^2} = \frac{sc_R}{(I-1)(J-1)(K-1)} = s_R^2.
\end{aligned}$$

6.1.2. Avec répétitions

Un facteur contrôlé A se présente sous I modalités, chacune d'entre elles étant notée A_i. Un facteur contrôlé B se présente sous J modalités, chacune d'entre elles étant notée B_j. Un facteur contrôlé C se présente sous K modalités, chacune d'entre elles étant notée C_k. Pour chacun des couples de modalités (A_i, B_j, C_k) on effectue $L \geqslant 2$ mesures d'une réponse Y qui est une variable continue. On note $n = I \times J \times K \times L$ le nombre total de mesures ayant été effectuées.

On introduit le modèle :

$$Y_{i,j,k,l} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{i,j} + (\alpha\gamma)_{i,k} + (\beta\gamma)_{j,k} + (\alpha\beta\gamma)_{i,j,k} + \epsilon_{i,j,k,l},$$
$$i = 1 \dots I,\ j = 1 \dots J,\ k = 1 \dots K,\ l = 1 \dots L,$$

avec les contraintes supplémentaires $\displaystyle\sum_{i=1}^{I} \alpha_i = 0,\ \sum_{j=1}^{J} \beta_j = 0$ et $\displaystyle\sum_{k=1}^{K} \gamma_k = 0,$

$$\sum_{i=1}^{I} (\alpha\beta)_{i,j} = 0,\ \forall j \in \{1, \dots, J\} \text{ et } \sum_{j=1}^{J} (\alpha\beta)_{i,j} = 0,\ \forall i \in \{1, \dots, I\},$$
$$\sum_{i=1}^{I} (\alpha\gamma)_{i,k} = 0,\ \forall k \in \{1, \dots, K\} \text{ et } \sum_{k=1}^{K} (\alpha\gamma)_{i,k} = 0,\ \forall i \in \{1, \dots, I\},$$
$$\sum_{j=1}^{J} (\beta\gamma)_{j,k} = 0,\ \forall k \in \{1, \dots, K\} \text{ et } \sum_{k=1}^{K} (\beta\gamma)_{j,k} = 0,\ \forall j \in \{1, \dots, J\},$$
$$\sum_{i=1}^{I} (\alpha\beta\gamma)_{i,j,k} = 0,\ \forall (j,k) \in \{1, \dots, J\} \times \{1, \dots, K\},$$
$$\sum_{j=1}^{J} (\alpha\beta\gamma)_{i,j,k} = 0,\ \forall (i,k) \in \{1, \dots, I\} \times \{1, \dots, K\},$$
$$\sum_{k=1}^{K} (\alpha\beta\gamma)_{i,j,k} = 0,\ \forall (i,j) \in \{1, \dots, I\} \times \{1, \dots, J\},$$

où $Y_{i,j,k,l}$ est la valeur prise par la réponse Y dans les conditions (A_i, B_j, C_k) lors du $l-$ème essai. On postule les hypothèses classiques suivantes pour les erreurs :

$$\forall\ (i,j,k,l), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K,\ 1 \leqslant l \leqslant L,\ \mathcal{L}(\epsilon_{i,j,k,l}) = \mathcal{N}(0, \sigma^2),$$
$$\mathrm{Cov}(\epsilon_{i,j,k,l}, \epsilon_{m,n,o,p}) = 0 \text{ si } (i,j,k,l) \neq (m,n,o,p)$$
$$\text{avec } 1 \leqslant i, m \leqslant I,\ 1 \leqslant j, n \leqslant J,\ 1 \leqslant k, o \leqslant K \text{ et } 1 \leqslant l, p \leqslant L.$$

On suppose que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fera l'objet d'un autre paragraphe.

On rappelle que la variation due au facteur A est définie par :

$$sc_A = JKL \sum_{i=1}^{I} (y_{i,\bullet,\bullet,\bullet} - y_{\bullet,\bullet,\bullet,\bullet})^2.$$

On rappelle que la variation due au facteur B est définie par :

$$sc_B = IKL \sum_{j=1}^{J} (y_{\bullet,j,\bullet,\bullet} - y_{\bullet,\bullet,\bullet,\bullet})^2.$$

On rappelle que la variation due au facteur C est définie par :

$$sc_C = IJL \sum_{k=1}^{K} (y_{\bullet,\bullet,k,\bullet} - y_{\bullet,\bullet,\bullet,\bullet})^2.$$

On rappelle que la variation due à l'interaction des facteurs A et B est définie par :

$$sc_{AB} = KL \sum_{i=1}^{I} \sum_{j=1}^{J} (y_{i,j,\bullet,\bullet} - y_{i,\bullet,\bullet,\bullet} - y_{\bullet,j,\bullet,\bullet} + y_{\bullet,\bullet,\bullet,\bullet})^2.$$

On rappelle que la variation due à l'interaction des facteurs B et C est définie par :

$$sc_{BC} = IL \sum_{j=1}^{J} \sum_{k=1}^{K} (y_{\bullet,j,k,\bullet} - y_{\bullet,j,\bullet,\bullet} - y_{\bullet,\bullet,k,\bullet} + y_{\bullet,\bullet,\bullet,\bullet})^2.$$

On rappelle que la variation due à l'interaction des facteurs A et C est définie par :

$$sc_{AC} = JL \sum_{i=1}^{I} \sum_{k=1}^{K} (y_{i,\bullet,k,\bullet} - y_{i,\bullet,\bullet,\bullet} - y_{\bullet,\bullet,k,\bullet} + y_{\bullet,\bullet,\bullet,\bullet})^2.$$

La variation due à l'interaction d'ordre 2 entre les facteurs A, B et C est définie par :

$$sc_R = L \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (y_{i,j,k,\bullet} - y_{i,j,\bullet,\bullet} - y_{i,\bullet,k,\bullet} - y_{\bullet,j,k,\bullet} + y_{i,\bullet,\bullet,\bullet} + y_{\bullet,j,\bullet,\bullet} + y_{\bullet,\bullet,k,\bullet} - y_{\bullet,\bullet,\bullet,\bullet})^2.$$

La variation résiduelle est quant à elle définie par :

$$sc_R = L \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (y_{i,j,k,l} - y_{i,j,k,\bullet})^2.$$

Enfin la variation totale est égale à :

$$sc_{TOT} = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{l=1}^{L} (y_{i,j,k,l} - y_{\bullet,\bullet,\bullet,\bullet})^2.$$

On rappelle la relation fondamentale de l'ANOVA :

$$sc_{TOT} = sc_A + sc_B + sc_C + sc_{AB} + sc_{AC} + sc_{BC} + sc_{ABC} + sc_R.$$

On introduit les dégres de liberté (Ddl) associés à chaque ligne du tableau de l'ANOVA :

Source	Degrés de liberté
Facteur A	$n_A = I - 1$
Facteur B	$n_B = J - 1$
Facteur C	$n_C = K - 1$
Interaction AB	$n_{AB} = (I-1)(J-1)$
Interaction AC	$n_{AC} = (I-1)(K-1)$
Interaction BC	$n_{BC} = (J-1)(K-1)$
Interaction ABC	$n_{ABC} = (I-1)(J-1)(K-1)$
Résiduelle	$n_R = IJK(L-1)$
Totale	$n_{TOT} = IJKL - 1$

On résume ces informations dans le tableau de l'ANOVA ci-dessous :

Source	Variation	Ddl	Carré Moyen	F	Décision
Facteur A	sc_A	n_A	$s_A^2 = \frac{sc_A}{n_A}$	$f_A = \frac{s_A^2}{s_R^2}$	$\mathcal{H}_0^{'}$ ou $\mathcal{H}_1^{'}$
Facteur B	sc_B	n_B	$s_B^2 = \frac{sc_B}{n_B}$	$f_B = \frac{s_B^2}{s_R^2}$	$\mathcal{H}_0^{''}$ ou $\mathcal{H}_1^{''}$
Facteur C	sc_C	n_C	$s_C^2 = \frac{sc_C}{n_C}$	$f_C = \frac{s_C^2}{s_R^2}$	$\mathcal{H}_0^{'''}$ ou $\mathcal{H}_1^{'''}$
Interaction AB	sc_{AB}	n_{AB}	$s_{AB}^2 = \frac{sc_{AB}}{n_{AB}}$	$f_{AB} = \frac{s_{AB}^2}{s_R^2}$	$\mathcal{H}_0^{(4)}$ ou $\mathcal{H}_1^{(4)}$
Interaction AC	sc_{AC}	n_{AC}	$s_{AC}^2 = \frac{sc_{AC}}{n_{AC}}$	$f_{AC} = \frac{s_{AC}^2}{s_R^2}$	$\mathcal{H}_0^{(5)}$ ou $\mathcal{H}_1^{(5)}$
Interaction BC	sc_{BC}	n_{BC}	$s_{BC}^2 = \frac{sc_{BC}}{n_{BC}}$	$f_{BC} = \frac{s_{BC}^2}{s_R^2}$	$\mathcal{H}_0^{(6)}$ ou $\mathcal{H}_1^{(6)}$
Interaction ABC	sc_{ABC}	n_{ABC}	$s_{ABC}^2 = \frac{sc_{ABC}}{n_{ABC}}$	$f_{ABC} = \frac{s_{ABC}^2}{s_R^2}$	$\mathcal{H}_0^{(7)}$ ou $\mathcal{H}_1^{(7)}$
Résiduelle	sc_R	n_R	$s_R^2 = \frac{sc_R}{n_R}$		
Totale	sc_{TOT}	n_{TOT}			

On souhaite faire les tests d'hypothèse suivants :

$$\boxed{\mathcal{H}_0^{'} : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0}$$

contre

$$\boxed{\mathcal{H}_1^{'} : \text{Il existe } i_0 \in \{1, 2, \ldots, I\} \text{ tel que } \alpha_{i_0} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{'}$ précédente d'absence d'effet du facteur A et lorsque les conditions de validité du modèle sont respectées, f_A est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $I-1$ et $IJK(L-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_A est supérieure ou égale à la valeur critique issue de la table. Lorsque l'hypothèse nulle $\mathcal{H}_0^{'}$ est rejetée on peut procéder à des comparaisons multiples des différents effets des niveaux du facteur voir la section 10.

$$\boxed{\mathcal{H}_0^{''} : \beta_1 = \beta_2 = \cdots = \beta_J = 0}$$

contre

$$\boxed{\mathcal{H}_1^{''} : \text{Il existe } j_0 \in \{1, 2, \ldots, J\} \text{ tel que } \beta_{j_0} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{''}$ précédente d'absence d'effet du facteur B et lorsque les conditions de validité du modèle sont respectées, f_B est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $J-1$ et $IJK(L-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_B est supérieure ou égale à la valeur critique issue de la table. Lorsque l'hypothèse nulle $\mathcal{H}_0^{''}$ est rejetée on peut procéder à des comparaisons multiples des différents effets des niveaux du facteur voir la section 10.

$$\boxed{\mathcal{H}_0^{'''} : \gamma_1 = \gamma_2 = \cdots = \gamma_K = 0}$$

contre

$$\boxed{\mathcal{H}_1^{'''} : \text{Il existe } k_0 \in \{1, 2, \ldots, K\} \text{ tel que } \gamma_{k_0} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{'''}$ précédente d'absence d'effet du facteur C et lorsque les conditions de validité du modèle sont respectées, f_C est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $K-1$ et $IJK(L-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_C est supérieure ou égale à la valeur critique issue de la table. Lorsque l'hypothèse nulle $\mathcal{H}_0^{'''}$ est rejetée on peut procéder à des comparaisons multiples des différents effets des niveaux du facteur voir la section 10.

$$\boxed{\mathcal{H}_0^{(4)} : (\alpha\beta)_{1,1} = (\alpha\beta)_{1,2} = \cdots = (\alpha\beta)_{1,J} = (\alpha\beta)_{2,1} = \cdots = (\alpha\beta)_{I,J} = 0}$$

contre

$$\boxed{\mathcal{H}_1^{(4)} : \text{Il existe } (i_0, j_0) \in \{1, 2, \ldots, I\} \times \{1, 2, \ldots, J\} \text{ tel que } (\alpha\beta)_{i_0,j_0} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{(4)}$ précédente d'absence d'effet de l'interaction des facteurs A et B et lorsque les conditions de validité du modèle sont respectées, f_{AB} est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $(I-1)(J-1)$ et $IJK(L-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_{AB} est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0^{(5)} : (\alpha\gamma)_{1,1} = (\alpha\gamma)_{1,2} = \cdots = (\alpha\gamma)_{1,K} = (\alpha\gamma)_{2,1} = \cdots = (\alpha\gamma)_{I,K} = 0}$$

contre

$$\boxed{\mathcal{H}_1^{(5)} : \text{Il existe } (i_0, k_0) \in \{1, 2, \ldots, I\} \times \{1, 2, \ldots, K\} \text{ tel que } (\alpha\gamma)_{i_0,k_0} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{(5)}$ précédente d'absence d'effet de l'interaction des facteurs A et C et lorsque les conditions de validité du modèle sont respectées, f_{AC} est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $(I-1)(K-1)$ et $IJK(L-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_{AC} est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0^{(6)} : (\beta\gamma)_{1,1} = (\beta\gamma)_{1,2} = \cdots = (\beta\gamma)_{1,K} = (\beta\gamma)_{2,1} = \cdots = (\beta\gamma)_{J,K} = 0}$$

contre

$$\boxed{\mathcal{H}_1^{(6)} : \text{Il existe } (j_0, k_0) \in \{1, 2, \ldots, J\} \times \{1, 2, \ldots, K\} \text{ tel que } (\beta\gamma)_{j_0,k_0} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{(6)}$ précédente d'absence d'effet de l'interaction des facteurs B et C et lorsque les conditions de validité du modèle sont respectées, f_{BC} est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $(J-1)(K-1)$ et $IJK(L-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_{BC} est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0^{(7)} : (\alpha\beta\gamma)_{1,1,1} = (\alpha\beta\gamma)_{1,1,2} = \cdots = (\alpha\beta\gamma)_{1,1,K} = (\alpha\beta\gamma)_{2,1,1} = \cdots = (\alpha\beta\gamma)_{I,J,K} = 0}$$

contre

$$\boxed{\mathcal{H}_1^{(7)} : \exists\ (i_0, j_0, k_0) \in \{1, 2, \ldots, I\} \times \{1, 2, \ldots, J\} \times \{1, 2, \ldots, K\} \mid (\alpha\beta\gamma)_{i_0,j_0,k_0} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{(7)}$ précédente d'absence d'effet de l'interaction, d'ordre 3, des facteurs A, B et C et lorsque les conditions de validité du modèle sont respectées, f_{ABC} est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $(I-1)(J-1)(K-1)$ et $IJK(L-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est

inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_{ABC} est supérieure ou égale à la valeur critique issue de la table.

Les estimations $\widehat{\mu}$, $\widehat{\alpha_1}$, ..., $\widehat{\alpha_I}$, $\widehat{\beta_1}$, ..., $\widehat{\beta_J}$, $\widehat{\gamma_1}$, ..., $\widehat{\gamma_K}$, $\widehat{(\alpha\beta)_{1,1}}$, ..., $\widehat{(\alpha\beta)_{I,J}}$, $\widehat{(\alpha\gamma)_{1,1}}$, ..., $\widehat{(\alpha\gamma)_{I,K}}$, $\widehat{(\beta\gamma)_{1,1}}$, ..., $\widehat{(\beta\gamma)_{J,K}}$, $\widehat{(\alpha\beta\gamma)_{1,1,1}}$, ..., $\widehat{(\alpha\beta\gamma)_{I,J,K}}$, $\widehat{\sigma^2}$ des paramètres μ, α_1, ..., α_I, β_1, ..., β_J, γ_1, ..., γ_K, $(\alpha\beta)_{1,1}$, ..., $(\alpha\beta)_{I,J}$, $(\alpha\gamma)_{1,1}$, ..., $(\alpha\gamma)_{I,K}$, $(\beta\gamma)_{1,1}$, ..., $(\beta\gamma)_{J,K}$, $(\alpha\beta\gamma)_{1,1,1}$, ..., $(\alpha\beta\gamma)_{I,J,K}$, σ^2 du modèle se déduisent des formules suivantes :

$$
\begin{aligned}
&\widehat{\mu} = y_{\bullet,\bullet,\bullet,\bullet} = \overline{y},\\
&\widehat{\alpha_i} = y_{i,\bullet,\bullet,\bullet} - \widehat{\mu},\ 1 \leqslant i \leqslant I,\ \widehat{\beta_j} = y_{\bullet,j,\bullet,\bullet} - \widehat{\mu},\ 1 \leqslant j \leqslant J,\ \widehat{\gamma_k} = y_{\bullet,\bullet,k,\bullet} - \widehat{\mu},\ 1 \leqslant k \leqslant K,\\
&\widehat{(\alpha\beta)_{i,j}} = y_{i,j,\bullet,\bullet} - y_{i,\bullet,\bullet,\bullet} - y_{\bullet,j,\bullet,\bullet} + \widehat{\mu},\ 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\\
&\widehat{(\alpha\gamma)_{i,k}} = y_{i,\bullet,k,\bullet} - y_{i,\bullet,\bullet,\bullet} - y_{\bullet,\bullet,k,\bullet} + \widehat{\mu},\ 1 \leqslant i \leqslant I,\ 1 \leqslant k \leqslant K,\\
&\widehat{(\beta\gamma)_{j,k}} = y_{\bullet,j,k,\bullet} - y_{\bullet,j,\bullet,\bullet} - y_{\bullet,\bullet,k,\bullet} + \widehat{\mu},\ 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K,\\
&\widehat{(\alpha\beta\gamma)_{i,j,k}} = y_{i,j,k,\bullet} - y_{i,j,\bullet,\bullet} - y_{i,\bullet,k,\bullet} - y_{\bullet,j,k,\bullet} + y_{i,\bullet,\bullet,\bullet} + y_{\bullet,j,\bullet,\bullet} + y_{\bullet,\bullet,k,\bullet} - \widehat{\mu},\\
&1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K,\\
&\widehat{\sigma^2} = \frac{sc_R}{(I-1)(J-1)(K-1)} = s_R^2.
\end{aligned}
$$

6.2. Modèles à effets aléatoires

6.2.1. Sans répétition

Les α_i représentent un échantillon de taille I prélevé dans une population importante. Nous admettrons que les effets des A_i, les α_i, sont distribués suivant une loi normale centrée de variance σ_A^2. Les β_j représentent un échantillon de taille J prélevé dans une population importante. Nous admettrons que les effets des B_j, les β_j, sont distribués suivant une loi normale centrée de variance σ_B^2. Les γ_k représentent un échantillon de taille K prélevé dans une population importante. Nous admettrons que les effets des C_k, les γ_k, sont distribués suivant une loi normale centrée de variance σ_C^2. Pour chacun des couples de modalités (A_i, B_j, C_k) on effectue une mesure d'une réponse Y qui est une variable continue. On note $n = I \times J \times K$ le nombre total de mesures ayant été effectuées.

On introduit le modèle :

$$
\begin{gathered}
Y_{i,j,k} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{i,j} + (\alpha\gamma)_{i,k} + (\beta\gamma)_{j,k} + \epsilon_{i,j,k},\\
i = 1 \ldots I,\ j = 1 \ldots J,\ k = 1 \ldots K,
\end{gathered}
$$

où $Y_{i,j,k}$ est la valeur prise par la réponse Y dans les conditions (A_i, B_j, C_k). On suppose

que :

$$\begin{aligned}
\mathcal{L}\left(\alpha_i\right) &= \mathcal{N}(0,\sigma_A^2),\ \forall\ i, 1 \leqslant i \leqslant I,\\
\mathcal{L}\left(\beta_j\right) &= \mathcal{N}(0,\sigma_B^2),\ \forall\ j, 1 \leqslant j \leqslant J,\\
\mathcal{L}\left(\gamma_k\right) &= \mathcal{N}(0,\sigma_C^2),\ \forall\ k, 1 \leqslant k \leqslant K,\\
\mathcal{L}\left((\alpha\beta)_{i,j}\right) &= \mathcal{N}(0,\sigma_{AB}^2),\ \forall\ (i,j), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\\
\mathcal{L}\left((\alpha\gamma)_{i,k}\right) &= \mathcal{N}(0,\sigma_{AC}^2),\ \forall\ (i,k), 1 \leqslant i \leqslant I,\ 1 \leqslant k \leqslant K,\\
\mathcal{L}\left((\beta\gamma)_{j,k}\right) &= \mathcal{N}(0,\sigma_{BC}^2),\ \forall\ (j,k), 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K,
\end{aligned}$$

ainsi que l'indépendance des effets aléatoires α_i, β_j, γ_k, $(\alpha\beta)_{i,j}$, $(\alpha\gamma)_{i,k}$ et $(\beta\gamma)_{j,k}$.
On postule les hypothèses classiques suivantes pour les erreurs :

$$\begin{gathered}
\forall\ (i,j,k), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K,\ \mathcal{L}(\epsilon_{i,j,k}) = \mathcal{N}(0,\sigma^2),\\
\mathrm{Cov}(\epsilon_{i,j,k},\epsilon_{l,m,n}) = 0 \text{ si } (i,j,k) \neq (l,m,n) \text{ avec } 1 \leqslant i,l \leqslant I,\ 1 \leqslant j,m \leqslant J \text{ et } 1 \leqslant k,n \leqslant K,
\end{gathered}$$

ainsi que l'indépendance des effets aléatoires α_i, β_j, γ_k, $(\alpha\beta)_{i,j}$, $(\alpha\gamma)_{i,k}$ et $(\beta\gamma)_{j,k}$ et des erreurs $\epsilon_{i,j,k}$.

On suppose que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fera l'objet d'un autre paragraphe.

On utilise les quantités sc_A, sc_B, sc_C, sc_{AB}, sc_{AC}, sc_{BC}, sc_R et sc_{TOT} introduites à la section 6.1.1.

On rappelle la relation fondamentale de l'ANOVA :

$$sc_{TOT} = sc_A + sc_B + sc_C + sc_{AB} + sc_{AC} + sc_{BC} + sc_R.$$

On introduit les dégres de liberté (Ddl) associés à chaque ligne du tableau de l'ANOVA :

Source	Degrés de liberté
Facteur A	$n_A = I - 1$
Facteur B	$n_B = J - 1$
Facteur C	$n_C = K - 1$
Interaction AB	$n_{AB} = (I-1)(J-1)$
Interaction AC	$n_{AC} = (I-1)(K-1)$
Interaction BC	$n_{BC} = (J-1)(K-1)$
Résiduelle	$n_R = (I-1)(J-1)(K-1)$
Totale	$n_{TOT} = IJK - 1$

On résume ces informations dans le tableau de l'ANOVA ci-dessous :

Source	Variation	Ddl	Carré Moyen	F	Décision
Facteur A	sc_A	n_A	$s_A^2 = \dfrac{sc_A}{n_A}$	$f_A = \dfrac{s_A^2}{s_{AB}^2 + s_{AC}^2 - s_R^2}$	$\mathcal{H}_0^{'}$ ou $\mathcal{H}_1^{'}$
Facteur B	sc_B	n_B	$s_B^2 = \dfrac{sc_B}{n_B}$	$f_B = \dfrac{s_B^2}{s_{AB}^2 + s_{BC}^2 - s_R^2}$	$\mathcal{H}_0^{''}$ ou $\mathcal{H}_1^{''}$
Facteur C	sc_C	n_C	$s_C^2 = \dfrac{sc_C}{n_C}$	$f_C = \dfrac{s_C^2}{s_{AC}^2 + s_{BC}^2 - s_R^2}$	$\mathcal{H}_0^{'''}$ ou $\mathcal{H}_1^{'''}$
Interaction AB	sc_{AB}	n_{AB}	$s_{AB}^2 = \dfrac{sc_{AB}}{n_{AB}}$	$f_{AB} = \dfrac{s_{AB}^2}{s_R^2}$	$\mathcal{H}_0^{(4)}$ ou $\mathcal{H}_1^{(4)}$
Interaction AC	sc_{AC}	n_{AC}	$s_{AC}^2 = \dfrac{sc_{AC}}{n_{AC}}$	$f_{AC} = \dfrac{s_{AC}^2}{s_R^2}$	$\mathcal{H}_0^{(5)}$ ou $\mathcal{H}_1^{(5)}$
Interaction BC	sc_{BC}	n_{BC}	$s_{BC}^2 = \dfrac{sc_{BC}}{n_{BC}}$	$f_{BC} = \dfrac{s_{BC}^2}{s_R^2}$	$\mathcal{H}_0^{(6)}$ ou $\mathcal{H}_1^{(6)}$
Résiduelle	sc_R	n_R	$s_R^2 = \dfrac{sc_R}{n_R}$		
Totale	sc_{TOT}	n_{TOT}			

On souhaite faire les tests d'hypothèse suivants :

$$\boxed{\mathcal{H}_0^{'} : \sigma_A^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1^{'} : \sigma_A^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{'}$ précédente d'absence d'effet du facteur A et lorsque les conditions de validité du modèle sont respectées, f_A est la réalisation d'une variable aléatoire qui suit **approximativement**[5] une loi de Fisher à $I-1$ et $n^{'}$ degrés de liberté avec :

$$n^{'} = \frac{\left(s_{AB}^2 + s_{AC}^2 - s_R^2\right)^2}{\dfrac{\left(s_{AB}^2\right)^2}{(I-1)(J-1)} + \dfrac{\left(s_{AC}^2\right)^2}{(I-1)(K-1)} + \dfrac{\left(s_R^2\right)^2}{(I-1)(J-1)(K-1)}}.$$

On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_A est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0^{''} : \sigma_B^2 = 0}$$
contre
$$\boxed{\mathcal{H}_1^{''} : \sigma_B^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{''}$ précédente d'absence d'effet du facteur B et lorsque les conditions de validité du modèle sont respectées, f_B est la réalisation d'une variable aléatoire qui suit **approximativement**[5] une loi de Fisher à $J-1$ et $n^{''}$ degrés de liberté avec :

$$n^{''} = \frac{\left(s_{AB}^2 + s_{BC}^2 - s_R^2\right)^2}{\dfrac{\left(s_{AB}^2\right)^2}{(I-1)(J-1)} + \dfrac{\left(s_{BC}^2\right)^2}{(J-1)(K-1)} + \dfrac{\left(s_R^2\right)^2}{(I-1)(J-1)(K-1)}}.$$

On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_B est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0^{'''} : \sigma_C^2 = 0}$$
contre
$$\boxed{\mathcal{H}_1^{'''} : \sigma_C^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{'''}$ précédente d'absence d'effet du facteur C et lorsque les conditions de validité du modèle sont respectées, f_C est la réalisation d'une variable aléatoire qui suit **approximativement**[5] une loi de Fisher à $K-1$ et $n^{'''}$ degrés de liberté avec :

$$n^{'''} = \frac{\left(s_{AC}^2 + s_{BC}^2 - s_R^2\right)^2}{\dfrac{\left(s_{AC}^2\right)^2}{(I-1)(K-1)} + \dfrac{\left(s_{BC}^2\right)^2}{(J-1)(K-1)} + \dfrac{\left(s_R^2\right)^2}{(I-1)(J-1)(K-1)}}.$$

[5]On utilise ici l'approximation dite de Satterthwaite.

On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_C est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0^{(4)} : \sigma_{AB}^2 = 0}$$
contre
$$\boxed{\mathcal{H}_1^{(4)} : \sigma_{AB}^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{(4)}$ précédente d'absence d'effet de l'interaction des facteurs A et B et lorsque les conditions de validité du modèle sont respectées, f_{AB} est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $(I-1)(J-1)$ et $(I-1)(J-1)(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_{AB} est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0^{(5)} : \sigma_{AC}^2 = 0}$$
contre
$$\boxed{\mathcal{H}_1^{(5)} : \sigma_{AC}^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{(5)}$ précédente d'absence d'effet de l'interaction des facteurs A et C et lorsque les conditions de validité du modèle sont respectées, f_{AC} est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $(I-1)(K-1)$ et $(I-1)(J-1)(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_{AC} est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0^{(6)} : \sigma_{BC}^2 = 0}$$
contre
$$\boxed{\mathcal{H}_1^{(6)} : \sigma_{BC}^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{(6)}$ précédente d'absence d'effet de l'interaction des facteurs B et C et lorsque les conditions de validité du modèle sont respectées, f_{BC} est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $(J-1)(K-1)$ et $(I-1)(J-1)(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_{BC} est supérieure ou égale à la valeur critique issue de la table.

Les estimations $\widehat{\mu}$, $\widehat{\sigma_A^2}$, $\widehat{\sigma_B^2}$, $\widehat{\sigma_C^2}$, $\widehat{\sigma_{AB}^2}$, $\widehat{\sigma_{AC}^2}$, $\widehat{\sigma_{BC}^2}$, $\widehat{\sigma^2}$ des paramètres μ, σ_A^2, σ_B^2, σ_C^2, σ_{AB}^2, σ_{AC}^2, σ_{BC}^2, σ^2 du modèle se déduisent des formules suivantes :

$$\begin{aligned}
\widehat{\mu} &= y_{\bullet,\bullet,\bullet} = \overline{y},\\
\widehat{\sigma_A^2} &= \frac{1}{JK}\left(s_A^2 - s_{AB}^2 - s_{AC}^2 + s_R^2\right),\\
\widehat{\sigma_B^2} &= \frac{1}{IK}\left(s_B^2 - s_{AB}^2 - s_{BC}^2 + s_R^2\right),\\
\widehat{\sigma_C^2} &= \frac{1}{IJ}\left(s_C^2 - s_{AC}^2 - s_{BC}^2 + s_R^2\right),\\
\widehat{\sigma_{AB}^2} &= \frac{1}{K}\left(s_{AB}^2 - s_R^2\right),\\
\widehat{\sigma_{AC}^2} &= \frac{1}{J}\left(s_{AC}^2 - s_R^2\right),\\
\widehat{\sigma_{BC}^2} &= \frac{1}{I}\left(s_{BC}^2 - s_R^2\right),\\
\widehat{\sigma^2} &= \frac{sc_R}{(I-1)(J-1)(K-1)} = s_R^2.
\end{aligned}$$

Remarque 6.1. Ainsi lorsque les trois facteurs sont aléatoires, et que l'on cherche à tester l'existence d'un effet de l'un d'entre eux on doit utiliser une approximation. Cette situation se démarque nettement de celle où tous les facteurs sauf au plus un sont à effets fixes voir les paragraphes 6.1.1 et 6.3.1.

6.2.2. Avec répétitions

Les α_i représentent un échantillon de taille I prélevé dans une population importante. Nous admettrons que les effets des A_i, les α_i, sont distribués suivant une loi normale centrée de variance σ_A^2. Les β_j représentent un échantillon de taille J prélevé dans une population importante. Nous admettrons que les effets des B_j, les β_j, sont distribués suivant une loi normale centrée de variance σ_B^2. Les γ_k représentent un échantillon de taille K prélevé dans une population importante. Nous admettrons que les effets des C_k, les γ_k, sont distribués suivant une loi normale centrée de variance σ_C^2. Pour chacun des couples de modalités (A_i, B_j, C_k) on effectue $L \geqslant 2$ mesures d'une réponse Y qui est une variable continue. On note $n = I \times J \times K \times L$ le nombre total de mesures ayant été effectuées.

On introduit le modèle :

$$\begin{gathered}
Y_{i,j,k,l} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{i,j} + (\alpha\gamma)_{i,k} + (\beta\gamma)_{j,k} + (\alpha\beta\gamma)_{i,j,k} + \epsilon_{i,j,k,l},\\
i = 1\ldots I,\ j = 1\ldots J,\ k = 1\ldots K,\ l = 1\ldots L,
\end{gathered}$$

où $Y_{i,j,k,l}$ est la valeur prise par la réponse Y dans les conditions (A_i, B_j, C_k) lors du $l-$ème

essai. On suppose que :

$$\begin{gathered}
\mathcal{L}\left(\alpha_i\right) = \mathcal{N}(0, \sigma_A^2),\ \forall\ i, 1 \leqslant i \leqslant I, \\
\mathcal{L}\left(\beta_j\right) = \mathcal{N}(0, \sigma_B^2),\ \forall\ j, 1 \leqslant j \leqslant J, \\
\mathcal{L}\left(\gamma_k\right) = \mathcal{N}(0, \sigma_C^2),\ \forall\ k, 1 \leqslant k \leqslant K, \\
\mathcal{L}\left((\alpha\beta)_{i,j}\right) = \mathcal{N}(0, \sigma_{AB}^2),\ \forall\ (i,j), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J, \\
\mathcal{L}\left((\alpha\gamma)_{i,k}\right) = \mathcal{N}(0, \sigma_{AC}^2),\ \forall\ (i,k), 1 \leqslant i \leqslant I,\ 1 \leqslant k \leqslant K, \\
\mathcal{L}\left((\beta\gamma)_{j,k}\right) = \mathcal{N}(0, \sigma_{BC}^2),\ \forall\ (j,k), 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K, \\
\mathcal{L}\left((\alpha\beta\gamma)_{i,j,k}\right) = \mathcal{N}(0, \sigma_{ABC}^2),\ \forall\ (i,j,k), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K,
\end{gathered}$$

ainsi que l'indépendance des effets aléatoires α_i, β_j, γ_k, $(\alpha\beta)_{i,j}$, $(\alpha\gamma)_{i,k}$, $(\beta\gamma)_{j,k}$ et $(\alpha\beta\gamma)_{i,j,k}$. On postule les hypothèses classiques suivantes pour les erreurs :

$$\begin{gathered}
\forall\ (i,j,k,l), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K,\ 1 \leqslant l \leqslant L,\ \mathcal{L}(\epsilon_{i,j,k,l}) = \mathcal{N}(0, \sigma^2), \\
\mathrm{Cov}(\epsilon_{i,j,k,l}, \epsilon_{m,n,o,p}) = 0 \text{ si } (i,j,k,l) \neq (m,n,o,p) \\
\text{avec } 1 \leqslant i, m \leqslant I,\ 1 \leqslant j, n \leqslant J,\ 1 \leqslant k, o \leqslant K \text{ et } 1 \leqslant l, p \leqslant L,
\end{gathered}$$

ainsi que l'indépendance des effets aléatoires α_i, β_j, γ_k, $(\alpha\beta)_{i,j}$, $(\alpha\gamma)_{i,k}$, $(\beta\gamma)_{j,k}$ et $(\alpha\beta\gamma)_{i,j,k}$ et des erreurs $\epsilon_{i,j,k}$.

On suppose que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fera l'objet d'un autre paragraphe.

On utilise les quantités sc_A, sc_B, sc_C, sc_{AB}, sc_{AC}, sc_{BC}, sc_{ABC}, sc_R et sc_{TOT} introduites à la section 6.1.2.

On rappelle la relation fondamentale de l'ANOVA :

$$sc_{TOT} = sc_A + sc_B + sc_C + sc_{AB} + sc_{AC} + sc_{BC} + s_{ABC} + sc_R.$$

On introduit les dégres de liberté (Ddl) associés à chaque ligne du tableau de l'ANOVA :

Source	Degrés de liberté
Facteur A	$n_A = I - 1$
Facteur B	$n_B = J - 1$
Facteur C	$n_C = K - 1$
Interaction AB	$n_{AB} = (I-1)(J-1)$
Interaction AC	$n_{AC} = (I-1)(K-1)$
Interaction BC	$n_{BC} = (J-1)(K-1)$
Interaction ABC	$n_{ABC} = (I-1)(J-1)(K-1)$
Résiduelle	$n_R = IJK(L-1)$
Totale	$n_{TOT} = IJK - 1$

On résume ces informations dans le tableau de l'ANOVA ci-dessous :

Source	Variation	Ddl	Carré Moyen	F	Décision
Facteur A	sc_A	n_A	$s_A^2 = \frac{sc_A}{n_A}$	$f_A = \frac{s_A^2}{s_{AB}^2 + s_{AC}^2 - s_{ABC}^2}$	$\mathcal{H}_0'$ ou $\mathcal{H}_1'$
Facteur B	sc_B	n_B	$s_B^2 = \frac{sc_B}{n_B}$	$f_B = \frac{s_B^2}{s_{AB}^2 + s_{BC}^2 - s_{ABC}^2}$	$\mathcal{H}_0''$ ou $\mathcal{H}_1''$
Facteur C	sc_C	n_C	$s_C^2 = \frac{sc_C}{n_C}$	$f_C = \frac{s_C^2}{s_{AC}^2 + s_{BC}^2 - s_{ABC}^2}$	$\mathcal{H}_0'''$ ou $\mathcal{H}_1'''$
Interaction AB	sc_{AB}	n_{AB}	$s_{AB}^2 = \frac{sc_{AB}}{n_{AB}}$	$f_{AB} = \frac{s_{AB}^2}{s_{ABC}^2}$	$\mathcal{H}_0^{(4)}$ ou $\mathcal{H}_1^{(4)}$
Interaction AC	sc_{AC}	n_{AC}	$s_{AC}^2 = \frac{sc_{AC}}{n_{AC}}$	$f_{AC} = \frac{s_{AC}^2}{s_{ABC}^2}$	$\mathcal{H}_0^{(5)}$ ou $\mathcal{H}_1^{(5)}$
Interaction BC	sc_{BC}	n_{BC}	$s_{BC}^2 = \frac{sc_{BC}}{n_{BC}}$	$f_{BC} = \frac{s_{BC}^2}{s_{ABC}^2}$	$\mathcal{H}_0^{(6)}$ ou $\mathcal{H}_1^{(6)}$
Interaction ABC	sc_{ABC}	n_{ABC}	$s_{ABC}^2 = \frac{sc_{ABC}}{n_{ABC}}$	$f_{ABC} = \frac{s_{ABC}^2}{s_R^2}$	$\mathcal{H}_0^{(7)}$ ou $\mathcal{H}_1^{(7)}$
Résiduelle	sc_R	n_R	$s_R^2 = \frac{sc_R}{n_R}$		
Totale	sc_{TOT}	n_{TOT}			

On souhaite faire les tests d'hypothèse suivants :

$$\boxed{\mathcal{H}_0^{'} : \sigma_A^2 = 0}$$
contre
$$\boxed{\mathcal{H}_1^{'} : \sigma_A^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{'}$ précédente d'absence d'effet du facteur A et lorsque les conditions de validité du modèle sont respectées, f_A est la réalisation d'une variable aléatoire qui suit **approximativement**[6] une loi de Fisher à $I-1$ et $n^{'}$ degrés de liberté avec :

$$n^{'} = \frac{\left(s_{AB}^2 + s_{AC}^2 - s_{ABC}^2\right)^2}{\dfrac{\left(s_{AB}^2\right)^2}{(I-1)(J-1)} + \dfrac{\left(s_{AC}^2\right)^2}{(I-1)(K-1)} + \dfrac{\left(s_{ABC}^2\right)^2}{(I-1)(J-1)(K-1)}}.$$

On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_A est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0^{''} : \sigma_B^2 = 0}$$
contre
$$\boxed{\mathcal{H}_1^{''} : \sigma_B^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{''}$ précédente d'absence d'effet du facteur B et lorsque les conditions de validité du modèle sont respectées, f_B est la réalisation d'une variable aléatoire qui suit **approximativement**[6] une loi de Fisher à $J-1$ et $n^{''}$ degrés de liberté avec :

$$n^{''} = \frac{\left(s_{AB}^2 + s_{BC}^2 - s_{ABC}^2\right)^2}{\dfrac{\left(s_{AB}^2\right)^2}{(I-1)(J-1)} + \dfrac{\left(s_{BC}^2\right)^2}{(J-1)(K-1)} + \dfrac{\left(s_{ABC}^2\right)^2}{(I-1)(J-1)(K-1)}}.$$

On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_C est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0^{'''} : \sigma_C^2 = 0}$$
contre
$$\boxed{\mathcal{H}_1^{'''} : \sigma_C^2 \neq 0.}$$

[6] On utilise ici l'approximation dite de Satterthwaite.

Sous l'hypothèse nulle $\mathcal{H}_0^{'''}$ précédente d'absence d'effet du facteur C et lorsque les conditions de validité du modèle sont respectées, f_C est la réalisation d'une variable aléatoire qui suit **approximativement**[6] une loi de Fisher à $K-1$ et $n^{'''}$ degrés de liberté avec :

$$n^{'''} = \frac{\left(s_{AC}^2 + s_{BC}^2 - s_{ABC}^2\right)^2}{\dfrac{\left(s_{AC}^2\right)^2}{(I-1)(K-1)} + \dfrac{\left(s_{BC}^2\right)^2}{(J-1)(K-1)} + \dfrac{\left(s_{ABC}^2\right)^2}{(I-1)(J-1)(K-1)}}.$$

On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_C est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0^{(4)} : \sigma_{AB}^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1^{(4)} : \sigma_{AB}^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{(4)}$ précédente d'absence d'effet de l'interaction des facteurs A et B et lorsque les conditions de validité du modèle sont respectées, f_{AB} est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $(I-1)(J-1)$ et $(I-1)(J-1)(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_{AB} est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0^{(5)} : \sigma_{AC}^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1^{(5)} : \sigma_{AC}^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{(5)}$ précédente d'absence d'effet de l'interaction des facteurs A et C et lorsque les conditions de validité du modèle sont respectées, f_{AC} est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $(I-1)(K-1)$ et $(I-1)(J-1)(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_{AC} est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0^{(6)} : \sigma_{BC}^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1^{(6)} : \sigma_{BC}^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{(6)}$ précédente d'absence d'effet de l'interaction des facteurs B et C et lorsque les conditions de validité du modèle sont respectées, f_{BC} est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $(J-1)(K-1)$ et $(I-1)(J-1)(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_{BC} est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0^{(7)} : \sigma^2_{ABC} = 0}$$

contre

$$\boxed{\mathcal{H}_1^{(7)} : \sigma^2_{ABC} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{(7)}$ précédente d'absence d'effet de l'interaction d'ordre trois des facteurs A, B et C et lorsque les conditions de validité du modèle sont respectées, f_{ABC} est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $(I-1)(J-1)(K-1)$ et $IJK(L-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_{ABC} est supérieure ou égale à la valeur critique issue de la table.

Les estimations $\widehat{\mu}$, $\widehat{\sigma^2_A}$, $\widehat{\sigma^2_B}$, $\widehat{\sigma^2_C}$, $\widehat{\sigma^2_{AB}}$, $\widehat{\sigma^2_{AC}}$, $\widehat{\sigma^2_{BC}}$, $\widehat{\sigma^2_{ABC}}$, $\widehat{\sigma^2}$ des paramètres μ, σ^2_A, σ^2_B, σ^2_C, σ^2_{AB}, σ^2_{AC}, σ^2_{BC}, σ^2_{ABC}, σ^2 du modèle se déduisent des formules suivantes :

$$\begin{aligned}
\widehat{\mu} &= y_{\bullet,\bullet,\bullet,\bullet} = \overline{y},\\
\widehat{\sigma^2_A} &= \frac{1}{JKL}\left(s^2_A - s^2_{AB} - s^2_{AC} + s^2_{ABC}\right),\\
\widehat{\sigma^2_B} &= \frac{1}{IKL}\left(s^2_B - s^2_{AB} - s^2_{BC} + s^2_{ABC}\right),\\
\widehat{\sigma^2_C} &= \frac{1}{IJL}\left(s^2_C - s^2_{AC} - s^2_{BC} + s^2_{ABC}\right),\\
\widehat{\sigma^2_{AB}} &= \frac{1}{KL}\left(s^2_{AB} - s^2_{ABC}\right),\\
\widehat{\sigma^2_{AC}} &= \frac{1}{JL}\left(s^2_{AC} - s^2_{ABC}\right),\\
\widehat{\sigma^2_{BC}} &= \frac{1}{IL}\left(s^2_{BC} - s^2_{ABC}\right),\\
\widehat{\sigma^2_{ABC}} &= \frac{1}{L}\left(s^2_{ABC} - s^2_R\right),\\
\widehat{\sigma^2} &= \frac{sc_R}{IJK(L-1)} = s^2_R.
\end{aligned}$$

6.3. Modèles à effets mixtes

6.3.1. Sans répétition

Il existe deux possibilités : soit deux facteurs sont fixes et un est aléatoire, soit un facteur est fixe et deux sont aléatoires.

Premier cas : Deux facteurs sont à effets fixes et un facteur est à effets aléatoires.

Un facteur contrôlé A se présente sous I modalités, chacune d'entre elles étant notée A_i. Un facteur contrôlé B se présente sous J modalités, chacune d'entre elles étant notée B_j. Les γ_k représentent un échantillon de taille K prélevé dans une population importante. Nous admettrons que les effets des C_k, les γ_k, sont distribués suivant une loi normale centrée de variance σ_C^2. Pour chacun des couples de modalités (A_i, B_j, C_k) on effectue une mesure d'une réponse Y qui est une variable continue. On note $n = I \times J \times K$ le nombre total de mesures ayant été effectuées.

On introduit le modèle :

$$Y_{i,j,k} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{i,j} + (\alpha\gamma)_{i,k} + (\beta\gamma)_{j,k} + \epsilon_{i,j,k},$$
$$i = 1 \dots I,\ j = 1 \dots J,\ k = 1 \dots K,$$
$$\text{avec les contraintes supplémentaires } \sum_{i=1}^{I} \alpha_i = 0,\ \sum_{j=1}^{J} \beta_j = 0,$$
$$\sum_{i=1}^{I} (\alpha\beta)_{i,j} = 0,\ \forall j \in \{1, \dots, J\},\ \sum_{j=1}^{J} (\alpha\beta)_{i,j} = 0,\ \forall i \in \{1, \dots, I\},$$
$$\sum_{i=1}^{I} (\alpha\gamma)_{i,k} = 0,\ \forall k \in \{1, \dots, K\} \text{ et } \sum_{j=1}^{K} (\beta\gamma)_{j,k} = 0,\ \forall k \in \{1, \dots, K\},$$

où $Y_{i,j,k}$ est la valeur prise par la réponse Y dans les conditions (A_i, B_j, C_k). On suppose que :

$$\mathcal{L}(\gamma_k) = \mathcal{N}(0, \sigma_C^2),\ \forall\ k, 1 \leqslant k \leqslant K,$$
$$\mathcal{L}((\alpha\gamma)_{i,k}) = \mathcal{N}(0, \sigma_{AC}^2),\ \forall\ (i,k), 1 \leqslant i \leqslant I,\ 1 \leqslant k \leqslant K,$$
$$\mathcal{L}((\beta\gamma)_{j,k}) = \mathcal{N}(0, \sigma_{BC}^2),\ \forall\ (j,k), 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K,$$

ainsi que l'indépendance des effets aléatoires γ_k, $(\alpha\gamma)_{i,k}$ et $(\beta\gamma)_{j,k}$.
On postule les hypothèses classiques suivantes pour les erreurs :

$$\forall\ (i,j,k), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K,\ \mathcal{L}(\epsilon_{i,j,k}) = \mathcal{N}(0, \sigma^2),$$
$$\mathrm{Cov}(\epsilon_{i,j,k}, \epsilon_{l,m,n}) = 0 \text{ si } (i,j,k) \neq (l,m,n) \text{ avec } 1 \leqslant i, l \leqslant I,\ 1 \leqslant j, m \leqslant J \text{ et } 1 \leqslant k, n \leqslant K,$$

ainsi que l'indépendance des effets aléatoires γ_k, $(\alpha\gamma)_{i,k}$ et $(\beta\gamma)_{j,k}$ et des erreurs $\epsilon_{i,j,k}$.

On suppose que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fera l'objet d'un autre paragraphe.

On utilise les quantités sc_A, sc_B, sc_C, sc_{AB}, sc_{AC}, sc_{BC}, sc_R et sc_{TOT} introduites à la section 6.1.1.

On rappelle la relation fondamentale de l'ANOVA :

$$sc_{TOT} = sc_A + sc_B + sc_C + sc_{AB} + sc_{AC} + sc_{BC} + sc_R.$$

On introduit les dégres de liberté (Ddl) associés à chaque ligne du tableau de l'ANOVA :

Source	Degrés de liberté
Facteur A	$n_A = I - 1$
Facteur B	$n_B = J - 1$
Facteur C	$n_C = K - 1$
Interaction AB	$n_{AB} = (I-1)(J-1)$
Interaction AC	$n_{AC} = (I-1)(K-1)$
Interaction BC	$n_{BC} = (J-1)(K-1)$
Résiduelle	$n_R = (I-1)(J-1)(K-1)$
Totale	$n_{TOT} = IJK - 1$

On résume ces informations dans le tableau de l'ANOVA ci-dessous :

Source	Variation	Ddl	Carré Moyen	F	Décision
Facteur A	sc_A	n_A	$s_A^2 = \frac{sc_A}{n_A}$	$f_A = \frac{s_A^2}{s_{AC}^2}$	$\mathcal{H}_0^{'}$ ou $\mathcal{H}_1^{'}$
Facteur B	sc_B	n_B	$s_B^2 = \frac{sc_B}{n_B}$	$f_B = \frac{s_B^2}{s_{BC}^2}$	$\mathcal{H}_0^{''}$ ou $\mathcal{H}_1^{''}$
Facteur C	sc_C	n_C	$s_C^2 = \frac{sc_C}{n_C}$	$f_C = \frac{s_C^2}{s_R^2}$	$\mathcal{H}_0^{'''}$ ou $\mathcal{H}_1^{'''}$
Interaction AB	sc_{AB}	n_{AB}	$s_{AB}^2 = \frac{sc_{AB}}{n_{AB}}$	$f_{AB} = \frac{s_{AB}^2}{s_R^2}$	$\mathcal{H}_0^{(4)}$ ou $\mathcal{H}_1^{(4)}$
Interaction AC	sc_{AC}	n_{AC}	$s_{AC}^2 = \frac{sc_{AC}}{n_{AC}}$	$f_{AC} = \frac{s_{AC}^2}{s_R^2}$	$\mathcal{H}_0^{(5)}$ ou $\mathcal{H}_1^{(5)}$
Interaction BC	sc_{BC}	n_{BC}	$s_{BC}^2 = \frac{sc_{BC}}{n_{BC}}$	$f_{BC} = \frac{s_{BC}^2}{s_R^2}$	$\mathcal{H}_0^{(6)}$ ou $\mathcal{H}_1^{(6)}$
Résiduelle	sc_R	n_R	$s_R^2 = \frac{sc_R}{n_R}$		
Totale	sc_{TOT}	n_{TOT}			

On souhaite faire les tests d'hypothèse suivants :

$$\boxed{\mathcal{H}_0^{'} : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0}$$

contre

$$\boxed{\mathcal{H}_1^{'} : \text{Il existe } i_0 \in \{1, 2, \ldots, I\} \text{ tel que } \alpha_{i_0} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{'}$ précédente d'absence d'effet du facteur A et lorsque les conditions de validité du modèle sont respectées, f_A est la réalisation d'une variable aléatoire

qui suit une loi de Fisher à $I-1$ et $(I-1)(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_A est supérieure ou égale à la valeur critique issue de la table. Lorsque l'hypothèse nulle $\mathcal{H}_0^{'}$ est rejetée on peut procéder à des comparaisons multiples des différents effets des niveaux du facteur voir la section 10.

$$\boxed{\mathcal{H}_0^{''} : \beta_1 = \beta_2 = \cdots = \beta_J = 0}$$

contre

$$\boxed{\mathcal{H}_1^{''} : \text{Il existe } j_0 \in \{1, 2, \ldots, J\} \text{ tel que } \beta_{j_0} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{''}$ précédente d'absence d'effet du facteur B et lorsque les conditions de validité du modèle sont respectées, f_B est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $J-1$ et $(J-1)(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_B est supérieure ou égale à la valeur critique issue de la table. Lorsque l'hypothèse nulle $\mathcal{H}_0^{''}$ est rejetée on peut procéder à des comparaisons multiples des différents effets des niveaux du facteur voir la section 10.

$$\boxed{\mathcal{H}_0^{'''} : \sigma_C^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1^{'''} : \sigma_C^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{'''}$ précédente d'absence d'effet du facteur C et lorsque les conditions de validité du modèle sont respectées, f_C est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $K-1$ et $(I-1)(J-1)(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_C est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0^{(4)} : (\alpha\beta)_{1,1} = (\alpha\beta)_{1,2} = \cdots = (\alpha\beta)_{1,J} = (\alpha\beta)_{2,1} = \cdots = (\alpha\beta)_{I,J} = 0}$$

contre

$$\boxed{\mathcal{H}_1^{(4)} : \text{Il existe } (i_0, j_0) \in \{1, 2, \ldots, I\} \times \{1, 2, \ldots, J\} \text{ tel que } (\alpha\beta)_{i_0,j_0} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{(4)}$ précédente d'absence d'effet de l'interaction des facteurs A et B et lorsque les conditions de validité du modèle sont respectées, f_{AB} est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $(I-1)(J-1)$ et $(I-1)(J-1)(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_{AB} est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0^{(5)} : \sigma_{AC}^2 = 0}$$
contre
$$\boxed{\mathcal{H}_1^{(5)} : \sigma_{AC}^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{(5)}$ précédente d'absence d'effet de l'interaction des facteurs A et C et lorsque les conditions de validité du modèle sont respectées, f_{AC} est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $(I-1)(K-1)$ et $(I-1)(J-1)(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_{AC} est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0^{(6)} : \sigma_{BC}^2 = 0}$$
contre
$$\boxed{\mathcal{H}_1^{(6)} : \sigma_{BC}^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{(6)}$ précédente d'absence d'effet de l'interaction des facteurs B et C et lorsque les conditions de validité du modèle sont respectées, f_{BC} est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $(J-1)(K-1)$ et $(I-1)(J-1)(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_{BC} est supérieure ou égale à la valeur critique issue de la table.

Les estimations $\widehat{\mu}$, $\widehat{\alpha_1}$, ..., $\widehat{\alpha_I}$, $\widehat{\beta}_1$, ..., $\widehat{\beta_J}$, $\widehat{\sigma_C^2}$, $\widehat{(\alpha\beta)_{1,1}}$, ..., $\widehat{(\alpha\beta)_{I,J}}$, $\widehat{\sigma_{AC}^2}$, $\widehat{\sigma_{BC}^2}$, $\widehat{\sigma^2}$ des paramètres μ, α_1, ..., α_I, β_1, ..., β_J, σ_C^2, $(\alpha\beta)_{1,1}$, ..., $(\alpha\beta)_{I,J}$, σ_{AC}^2, σ_{BC}^2, σ^2 du modèle se déduisent des formules suivantes :

$$\begin{aligned}
&\widehat{\mu} = y_{\bullet,\bullet,\bullet} = \overline{y},\\
&\widehat{\alpha_i} = y_{i,\bullet,\bullet} - \widehat{\mu},\ 1 \leqslant i \leqslant I,\ \widehat{\beta_j} = y_{\bullet,j,\bullet} - \widehat{\mu},\ 1 \leqslant j \leqslant J,\\
&\widehat{\sigma_C^2} = \frac{1}{IJ}\left(s_C^2 - s_R^2\right),\\
&\widehat{(\alpha\beta)_{i,j}} = y_{i,j,\bullet} - y_{i,\bullet,\bullet} - y_{\bullet,j,\bullet} + \widehat{\mu},\ 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\\
&\widehat{\sigma_{AC}^2} = \frac{1}{J}\left(s_{AC}^2 - s_R^2\right),\\
&\widehat{\sigma_{BC}^2} = \frac{1}{I}\left(s_{BC}^2 - s_R^2\right),\\
&\widehat{\sigma^2} = \frac{sc_R}{(I-1)(J-1)(K-1)} = s_R^2.
\end{aligned}$$

Deuxième cas : Un facteur est à effets fixes et deux facteurs sont à effets aléatoires.

Un facteur contrôlé A se présente sous I modalités, chacune d'entre elles étant notée A_i. Les β_j représentent un échantillon de taille J prélevé dans une population importante. Nous admettrons que les effets des B_j, les β_j, sont distribués suivant une loi normale centrée de variance σ_B^2. Les γ_k représentent un échantillon de taille K prélevé dans une population importante. Nous admettrons que les effets des C_k, les γ_k, sont distribués suivant une loi normale centrée de variance σ_C^2. Pour chacun des couples de modalités (A_i, B_j, C_k) on effectue une mesure d'une réponse Y qui est une variable continue. On note $n = I \times J \times K$ le nombre total de mesures ayant été effectuées.

On introduit le modèle :

$$Y_{i,j,k} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{i,j} + (\alpha\gamma)_{i,k} + (\beta\gamma)_{j,k} + \epsilon_{i,j,k},$$
$$i = 1 \dots I,\ j = 1 \dots J,\ k = 1 \dots K,$$
$$\text{avec les contraintes supplémentaires } \sum_{i=1}^{I} \alpha_i = 0,$$
$$\sum_{i=1}^{I} (\alpha\beta)_{i,j} = 0,\ \forall j \in \{1, \dots, J\} \text{ et } \sum_{i=1}^{I} (\alpha\gamma)_{i,k} = 0,\ \forall k \in \{1, \dots, K\},$$

où $Y_{i,j,k}$ est la valeur prise par la réponse Y dans les conditions (A_i, B_j, C_k). On suppose que :

$$\mathcal{L}(\beta_j) = \mathcal{N}(0, \sigma_B^2),\ \forall\ j, 1 \leqslant j \leqslant J,$$
$$\mathcal{L}(\gamma_k) = \mathcal{N}(0, \sigma_C^2),\ \forall\ k, 1 \leqslant k \leqslant K,$$
$$\mathcal{L}((\alpha\beta)_{i,j}) = \mathcal{N}(0, \sigma_{AB}^2),\ \forall\ (i,j), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,$$
$$\mathcal{L}((\alpha\gamma)_{i,k}) = \mathcal{N}(0, \sigma_{AC}^2),\ \forall\ (i,k), 1 \leqslant i \leqslant I,\ 1 \leqslant k \leqslant K,$$
$$\mathcal{L}((\beta\gamma)_{j,k}) = \mathcal{N}(0, \sigma_{BC}^2),\ \forall\ (j,k), 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K,$$

ainsi que l'indépendance des effets aléatoires β_j, γ_k, $(\alpha\beta)_{i,j}$, $(\alpha\gamma)_{i,k}$ et $(\beta\gamma)_{j,k}$.
On postule les hypothèses classiques suivantes pour les erreurs :

$$\forall\ (i,j,k), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K,\ \mathcal{L}(\epsilon_{i,j,k}) = \mathcal{N}(0, \sigma^2),$$
$$\mathrm{Cov}(\epsilon_{i,j,k}, \epsilon_{l,m,n}) = 0 \text{ si } (i,j,k) \neq (l,m,n) \text{ avec } 1 \leqslant i, l \leqslant I,\ 1 \leqslant j, m \leqslant J \text{ et } 1 \leqslant k, n \leqslant K,$$

ainsi que l'indépendance des effets aléatoires β_j, γ_k, $(\alpha\beta)_{i,j}$, $(\alpha\gamma)_{i,k}$ et $(\beta\gamma)_{j,k}$ et des erreurs $\epsilon_{i,j,k}$.

On suppose que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fera l'objet d'un autre paragraphe.

On utilise les quantités sc_A, sc_B, sc_C, sc_{AB}, sc_{AC}, sc_{BC}, sc_R et sc_{TOT} introduites à la section 6.1.1.

On rappelle la relation fondamentale de l'ANOVA :

$$sc_{TOT} = sc_A + sc_B + sc_C + sc_{AB} + sc_{AC} + sc_{BC} + sc_R.$$

On introduit les dégres de liberté (Ddl) associés à chaque ligne du tableau de l'ANOVA :

Source	Degrés de liberté
Facteur A	$n_A = I - 1$
Facteur B	$n_B = J - 1$
Facteur C	$n_C = K - 1$
Interaction AB	$n_{AB} = (I-1)(J-1)$
Interaction AC	$n_{AC} = (I-1)(K-1)$
Interaction BC	$n_{BC} = (J-1)(K-1)$
Résiduelle	$n_R = (I-1)(J-1)(K-1)$
Totale	$n_{TOT} = IJK - 1$

On résume ces informations dans le tableau de l'ANOVA ci-dessous :

Source	Variation	Ddl	Carré Moyen	F	Décision
Facteur A	sc_A	n_A	$s_A^2 = \dfrac{sc_A}{n_A}$	$f_A = \dfrac{s_A^2}{s_{AB}^2 + s_{AC}^2 - s_R^2}$	$\mathcal{H}_0^{'}$ ou $\mathcal{H}_1^{'}$
Facteur B	sc_B	n_B	$s_B^2 = \dfrac{sc_B}{n_B}$	$f_B = \dfrac{s_B^2}{s_{BC}^2}$	$\mathcal{H}_0^{''}$ ou $\mathcal{H}_1^{''}$
Facteur C	sc_C	n_C	$s_C^2 = \dfrac{sc_C}{n_C}$	$f_C = \dfrac{s_C^2}{s_{BC}^2}$	$\mathcal{H}_0^{'''}$ ou $\mathcal{H}_1^{'''}$
Interaction AB	sc_{AB}	n_{AB}	$s_{AB}^2 = \dfrac{sc_{AB}}{n_{AB}}$	$f_{AB} = \dfrac{s_{AB}^2}{s_R^2}$	$\mathcal{H}_0^{(4)}$ ou $\mathcal{H}_1^{(4)}$
Interaction AC	sc_{AC}	n_{AC}	$s_{AC}^2 = \dfrac{sc_{AC}}{n_{AC}}$	$f_{AC} = \dfrac{s_{AC}^2}{s_R^2}$	$\mathcal{H}_0^{(5)}$ ou $\mathcal{H}_1^{(5)}$
Interaction BC	sc_{BC}	n_{BC}	$s_{BC}^2 = \dfrac{sc_{BC}}{n_{BC}}$	$f_{BC} = \dfrac{s_{BC}^2}{s_R^2}$	$\mathcal{H}_0^{(6)}$ ou $\mathcal{H}_1^{(6)}$
Résiduelle	sc_R	n_R	$s_R^2 = \dfrac{sc_R}{n_R}$		
Totale	sc_{TOT}	n_{TOT}			

On souhaite faire les tests d'hypothèse suivants :

$$\boxed{\mathcal{H}_0^{'} : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0}$$

contre

$$\boxed{\mathcal{H}_1^{'} : \text{Il existe } i_0 \in \{1, 2, \ldots, I\} \text{ tel que } \alpha_{i_0} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{'}$ précédente d'absence d'effet du facteur A et lorsque les conditions de validité du modèle sont respectées, f_A est la réalisation d'une variable aléatoire

qui suit **approximativement**[7] une loi de Fisher à $I-1$ et $n^{'}$ degrés de liberté avec :

$$n^{'} = \frac{\left(s_{AB}^2 + s_{AC}^2 - s_R^2\right)^2}{\dfrac{\left(s_{AB}^2\right)^2}{(I-1)(J-1)} + \dfrac{\left(s_{AC}^2\right)^2}{(I-1)(K-1)} + \dfrac{\left(s_R^2\right)^2}{(I-1)(J-1)(K-1)}}.$$

On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_A est supérieure ou égale à la valeur critique issue de la table. Lorsque l'hypothèse nulle $\mathcal{H}_0^{'}$ est rejetée on peut procéder à des comparaisons multiples des différents effets des niveaux du facteur voir la section 10.

$$\boxed{\mathcal{H}_0^{''} : \sigma_B^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1^{''} : \sigma_B^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{''}$ précédente d'absence d'effet du facteur B et lorsque les conditions de validité du modèle sont respectées, f_B est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $J-1$ et $(J-1)(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_B est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0^{'''} : \sigma_C^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1^{'''} : \sigma_C^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{'''}$ précédente d'absence d'effet du facteur C et lorsque les conditions de validité du modèle sont respectées, f_C est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $K-1$ et $(J-1)(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_C est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0^{(4)} : \sigma_{AB}^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1^{(4)} : \sigma_{AB}^2 \neq 0.}$$

[7] On utilise toujours l'approximation de Satterthwaite.

Sous l'hypothèse nulle $\mathcal{H}_0^{(4)}$ précédente d'absence d'effet de l'interaction des facteurs A et B et lorsque les conditions de validité du modèle sont respectées, f_{AB} est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $(I-1)(J-1)$ et $(I-1)(J-1)(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_{AB} est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0^{(5)} : \sigma_{AC}^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1^{(5)} : \sigma_{AC}^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{(5)}$ précédente d'absence d'effet de l'interaction des facteurs A et C et lorsque les conditions de validité du modèle sont respectées, f_{AC} est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $(I-1)(K-1)$ et $(I-1)(J-1)(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_{AC} est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0^{(6)} : \sigma_{BC}^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1^{(6)} : \sigma_{BC}^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{(6)}$ précédente d'absence d'effet de l'interaction des facteurs B et C et lorsque les conditions de validité du modèle sont respectées, f_{BC} est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $(J-1)(K-1)$ et $(I-1)(J-1)(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_{BC} est supérieure ou égale à la valeur critique issue de la table.

Les estimations $\widehat{\mu}$, $\widehat{\alpha_1}$, ..., $\widehat{\alpha_I}$, $\widehat{\sigma_B^2}$, $\widehat{\sigma_C^2}$, $\widehat{\sigma_{AB}^2}$, $\widehat{\sigma_{AC}^2}$, $\widehat{\sigma_{BC}^2}$, $\widehat{\sigma^2}$ des paramètres μ, α_1, ..., α_I,

σ_B^2, σ_C^2, σ_{AB}^2, σ_{AC}^2, σ_{BC}^2, σ^2 du modèle se déduisent des formules suivantes :

$$\begin{aligned}
&\widehat{\mu} = y_{\bullet,\bullet,\bullet} = \overline{y},\\
&\widehat{\alpha_i} = y_{i,\bullet,\bullet} - \widehat{\mu},\ 1 \leqslant i \leqslant I,\\
&\widehat{\sigma_B^2} = \frac{1}{IK}\left(s_B^2 - s_R^2\right),\ \widehat{\sigma_C^2} = \frac{1}{IJ}\left(s_C^2 - s_R^2\right),\\
&\widehat{\sigma_{AB}^2} = \frac{1}{K}\left(s_{AB}^2 - s_R^2\right),\\
&\widehat{\sigma_{AC}^2} = \frac{1}{J}\left(s_{AC}^2 - s_R^2\right),\\
&\widehat{\sigma_{BC}^2} = \frac{1}{I}\left(s_{BC}^2 - s_R^2\right),\\
&\widehat{\sigma^2} = \frac{sc_R}{(I-1)(J-1)(K-1)} = s_R^2.
\end{aligned}$$

Remarque 6.2. Ainsi lorsque deux facteurs sont aléatoires, et que l'on cherche à tester l'existence d'un effet du troisième facteur, ici à effets fixes, on doit utiliser une approximation, ici celle de Satterthwaite. Cette situation se démarque nettement de celle où tous les facteurs sauf au plus un sont à effets fixes.

6.3.2. Avec répétitions

Il existe deux possibilités : soit deux facteurs sont fixes et un est aléatoire, soit un facteur est fixe et deux sont aléatoires.

Premier cas : Deux facteurs sont à effets fixes et un facteur est à effets aléatoires.

Un facteur contrôlé A se présente sous I modalités, chacune d'entre elles étant notée A_i. Un facteur contrôlé B se présente sous J modalités, chacune d'entre elles étant notée B_j. Les γ_k représentent un échantillon de taille K prélevé dans une population importante. Nous admettrons que les effets des C_k, les γ_k, sont distribués suivant une loi normale centrée de variance σ_C^2. Pour chacun des couples de modalités (A_i, B_j, C_k) on effectue $L \geqslant 2$ mesures d'une réponse Y qui est une variable continue. On note $n = I \times J \times K \times L$ le nombre total de mesures ayant été effectuées.

On introduit le modèle :

$$Y_{i,j,k,l} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{i,j} + (\alpha\gamma)_{i,k} + (\beta\gamma)_{j,k} + (\alpha\beta\gamma)_{i,j,k} + \epsilon_{i,j,k,l},$$
$$i = 1 \dots I,\ j = 1 \dots J,\ k = 1 \dots K,\ l = 1 \dots L,$$
$$\text{avec les contraintes supplémentaires } \sum_{i=1}^{I} \alpha_i = 0,\ \sum_{j=1}^{J} \beta_j = 0,$$
$$\sum_{i=1}^{I} (\alpha\beta)_{i,j} = 0,\ \forall j \in \{1, \dots, J\},\ \sum_{j=1}^{J} (\alpha\beta)_{i,j} = 0,\ \forall i \in \{1, \dots, I\},$$

$$\sum_{i=1}^{I}(\alpha\gamma)_{i,k} = 0,\ \forall k \in \{1,\ldots,K\} \text{ et } \sum_{j=1}^{J}(\beta\gamma)_{j,k} = 0,\ \forall k \in \{1,\ldots,K\},$$

$$\sum_{i=1}^{I}(\alpha\beta\gamma)_{i,j,k} = 0,\ \forall (j,k) \in \{1,\ldots,J\} \times \{1,\ldots,K\},$$

$$\sum_{j=1}^{J}(\alpha\beta\gamma)_{i,j,k} = 0,\ \forall (i,k) \in \{1,\ldots,I\} \times \{1,\ldots,K\},$$

où $Y_{i,j,k}$ est la valeur prise par la réponse Y dans les conditions (A_i, B_j, C_k) lors du $l-$ème essai. On suppose que :

$$\begin{gathered}
\mathcal{L}\left(\gamma_k\right) = \mathcal{N}(0,\sigma_C^2),\ \forall\ k, 1 \leqslant k \leqslant K,\\
\mathcal{L}\left((\alpha\gamma)_{i,k}\right) = \mathcal{N}(0,\sigma_{AC}^2),\ \forall\ (i,k), 1 \leqslant i \leqslant I,\ 1 \leqslant k \leqslant K,\\
\mathcal{L}\left((\beta\gamma)_{j,k}\right) = \mathcal{N}(0,\sigma_{BC}^2),\ \forall\ (j,k), 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K,\\
\mathcal{L}\left((\alpha\beta\gamma)_{i,j,k}\right) = \mathcal{N}(0,\sigma_{ABC}^2),\ \forall\ (i,j,k), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K,
\end{gathered}$$

ainsi que l'indépendance des effets aléatoires γ_k, $(\alpha\gamma)_{i,k}$, $(\beta\gamma)_{j,k}$, $(\alpha\beta\gamma)_{i,j,k}$.
On postule les hypothèses classiques suivantes pour les erreurs :

$$\begin{gathered}
\forall\ (i,j,k,l), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K,\ 1 \leqslant l \leqslant L,\ \mathcal{L}(\epsilon_{i,j,k,l}) = \mathcal{N}(0,\sigma^2),\\
\mathrm{Cov}(\epsilon_{i,j,k,l}, \epsilon_{m,n,o,p}) = 0 \text{ si } (i,j,k,l) \neq (m,n,o,p)\\
\text{avec } 1 \leqslant i,m \leqslant I,\ 1 \leqslant j,n \leqslant J,\ 1 \leqslant k,o \leqslant K \text{ et } 1 \leqslant l,p \leqslant L,
\end{gathered}$$

ainsi que l'indépendance des effets aléatoires γ_k, $(\alpha\gamma)_{i,k}$, $(\beta\gamma)_{j,k}$ et $(\alpha\beta\gamma)_{i,j,k}$ et des erreurs $\epsilon_{i,j,k,l}$.

On suppose que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fera l'objet d'un autre paragraphe.

On utilise les quantités sc_A, sc_B, sc_C, sc_{AB}, sc_{AC}, sc_{BC}, sc_{ABC}, sc_R et sc_{TOT} introduites à la section 6.1.2.

On rappelle la relation fondamentale de l'ANOVA :

$$sc_{TOT} = sc_A + sc_B + sc_C + sc_{AB} + sc_{AC} + sc_{BC} + sc_{ABC} + sc_R.$$

On introduit les dégres de liberté (Ddl) associés à chaque ligne du tableau de l'ANOVA :

Source	Degrés de liberté
Facteur A	$n_A = I - 1$
Facteur B	$n_B = J - 1$
Facteur C	$n_C = K - 1$
Interaction AB	$n_{AB} = (I-1)(J-1)$
Interaction AC	$n_{AC} = (I-1)(K-1)$
Interaction BC	$n_{BC} = (J-1)(K-1)$
Interaction ABC	$n_{ABC} = (I-1)(J-1)(K-1)$
Résiduelle	$n_R = IJK(L-1)$
Totale	$n_{TOT} = IJKL - 1$

On résume ces informations dans le tableau de l'ANOVA ci-dessous :

Source	Variation	Ddl	Carré Moyen	F	Décision
Facteur A	sc_A	n_A	$s_A^2 = \frac{sc_A}{n_A}$	$f_A = \frac{s_A^2}{s_{AC}^2}$	$\mathcal{H}_0^{'}$ ou $\mathcal{H}_1^{'}$
Facteur B	sc_B	n_B	$s_B^2 = \frac{sc_B}{n_B}$	$f_B = \frac{s_B^2}{s_{BC}^2}$	$\mathcal{H}_0^{''}$ ou $\mathcal{H}_1^{''}$
Facteur C	sc_C	n_C	$s_C^2 = \frac{sc_C}{n_C}$	$f_C = \frac{s_C^2}{s_R^2}$	$\mathcal{H}_0^{'''}$ ou $\mathcal{H}_1^{'''}$
Interaction AB	sc_{AB}	n_{AB}	$s_{AB}^2 = \frac{sc_{AB}}{n_{AB}}$	$f_{AB} = \frac{s_{AB}^2}{s_{ABC}^2}$	$\mathcal{H}_0^{(4)}$ ou $\mathcal{H}_1^{(4)}$
Interaction AC	sc_{AC}	n_{AC}	$s_{AC}^2 = \frac{sc_{AC}}{n_{AC}}$	$f_{AC} = \frac{s_{AC}^2}{s_R^2}$	$\mathcal{H}_0^{(5)}$ ou $\mathcal{H}_1^{(5)}$
Interaction BC	sc_{BC}	n_{BC}	$s_{BC}^2 = \frac{sc_{BC}}{n_{BC}}$	$f_{BC} = \frac{s_{BC}^2}{s_R^2}$	$\mathcal{H}_0^{(6)}$ ou $\mathcal{H}_1^{(6)}$
Interaction ABC	sc_{ABC}	n_{ABC}	$s_{ABC}^2 = \frac{sc_{ABC}}{n_{ABC}}$	$f_{ABC} = \frac{s_{ABC}^2}{s_R^2}$	$\mathcal{H}_0^{(7)}$ ou $\mathcal{H}_1^{(7)}$
Résiduelle	sc_R	n_R	$s_R^2 = \frac{sc_R}{n_R}$		
Totale	sc_{TOT}	n_{TOT}			

On souhaite faire les tests d'hypothèse suivants :

$$\boxed{\mathcal{H}_0^{'} : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0}$$

contre

$$\boxed{\mathcal{H}_1^{'} : \text{Il existe } i_0 \in \{1, 2, \ldots, I\} \text{ tel que } \alpha_{i_0} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{'}$ précédente d'absence d'effet du facteur A et lorsque les conditions de validité du modèle sont respectées, f_A est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $I-1$ et $(I-1)(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_A est supérieure ou égale à la valeur critique issue de la table. Lorsque l'hypothèse nulle $\mathcal{H}_0^{'}$ est rejetée on peut procéder à des comparaisons multiples des différents effets des niveaux du facteur voir la section 10.

$$\boxed{\mathcal{H}_0^{''} : \beta_1 = \beta_2 = \cdots = \beta_J = 0}$$

contre

$$\boxed{\mathcal{H}_1^{''} : \text{Il existe } j_0 \in \{1, 2, \ldots, J\} \text{ tel que } \beta_{j_0} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{''}$ précédente d'absence d'effet du facteur B et lorsque les conditions de validité du modèle sont respectées, f_B est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $J-1$ et $(J-1)(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_B est supérieure ou égale à la valeur critique issue de la table. Lorsque l'hypothèse nulle $\mathcal{H}_0^{''}$ est rejetée on peut procéder à des comparaisons multiples des différents effets des niveaux du facteur voir la section 10.

$$\boxed{\mathcal{H}_0^{'''} : \sigma_C^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1^{'''} : \sigma_C^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{'''}$ précédente d'absence d'effet du facteur C et lorsque les conditions de validité du modèle sont respectées, f_C est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $K-1$ et $IJK(L-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_C est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0^{(4)} : (\alpha\beta)_{1,1} = (\alpha\beta)_{1,2} = \cdots = (\alpha\beta)_{1,J} = (\alpha\beta)_{2,1} = \cdots = (\alpha\beta)_{I,J} = 0}$$

contre

$$\boxed{\mathcal{H}_1^{(4)} : \text{Il existe } (i_0, j_0) \in \{1, 2, \ldots, I\} \times \{1, 2, \ldots, J\} \text{ tel que } (\alpha\beta)_{i_0,j_0} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{(4)}$ précédente d'absence d'effet de l'interaction des facteurs A et B et lorsque les conditions de validité du modèle sont respectées, f_{AB} est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $(I-1)(J-1)$ et $(I-1)(J-1)(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_{AB} est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0^{(5)} : \sigma_{AC}^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1^{(5)} : \sigma_{AC}^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{(5)}$ précédente d'absence d'effet de l'interaction des facteurs A et C et lorsque les conditions de validité du modèle sont respectées, f_{AC} est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $(I-1)(K-1)$ et $IJK(L-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_{AC} est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0^{(6)} : \sigma_{BC}^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1^{(6)} : \sigma_{BC}^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{(6)}$ précédente d'absence d'effet de l'interaction des facteurs B et C et lorsque les conditions de validité du modèle sont respectées, f_{BC} est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $(J-1)(K-1)$ et $IJK(L-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_{BC} est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0^{(7)} : \sigma_{ABC}^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1^{(7)} : \sigma_{ABC}^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{(7)}$ précédente d'absence d'effet de l'interaction d'ordre trois des facteurs A, B et C et lorsque les conditions de validité du modèle sont respectées, f_{ABC} est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $(I-1)(J-1)(K-1)$ et $IJK(L-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est

inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_{ABC} est supérieure ou égale à la valeur critique issue de la table.

Les estimations $\widehat{\mu}$, $\widehat{\alpha_1}$, ..., $\widehat{\alpha_I}$, $\widehat{\beta_1}$, ..., $\widehat{\beta_J}$, $\widehat{\sigma^2_C}$, $\widehat{(\alpha\beta)_{1,1}}$, ..., $\widehat{(\alpha\beta)_{I,J}}$, $\widehat{\sigma^2_{AC}}$, $\widehat{\sigma^2_{BC}}$, $\widehat{\sigma^2_{ABC}}$, $\widehat{\sigma^2}$ des paramètres μ, α_1, ..., α_I, β_1, ..., β_J, σ^2_C, $(\alpha\beta)_{1,1}$, ..., $(\alpha\beta)_{I,J}$, σ^2_{AC}, σ^2_{BC}, σ^2_{ABC}, σ^2 du modèle se déduisent des formules suivantes :

$$
\begin{aligned}
&\widehat{\mu} = y_{\bullet,\bullet,\bullet,\bullet} = \overline{y},\\
&\widehat{\alpha_i} = y_{i,\bullet,\bullet,\bullet} - \widehat{\mu},\ 1 \leqslant i \leqslant I,\ \widehat{\beta_j} = y_{\bullet,j,\bullet,\bullet} - \widehat{\mu},\ 1 \leqslant j \leqslant J,\\
&\widehat{\sigma^2_C} = \frac{1}{IJL}\left(s^2_C - s^2_R\right),\\
&\widehat{(\alpha\beta)_{i,j}} = y_{i,j,\bullet,\bullet} - y_{i,\bullet,\bullet,\bullet} - y_{\bullet,j,\bullet,\bullet} + \widehat{\mu},\ 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\\
&\widehat{\sigma^2_{AC}} = \frac{1}{JL}\left(s^2_{AC} - s^2_R\right),\\
&\widehat{\sigma^2_{BC}} = \frac{1}{IL}\left(s^2_{BC} - s^2_R\right),\\
&\widehat{\sigma^2_{ABC}} = \frac{1}{L}\left(s^2_{ABC} - s^2_R\right),\\
&\widehat{\sigma^2} = \frac{sc_R}{IJK(L-1)} = s^2_R.
\end{aligned}
$$

Deuxième cas : Un facteur est à effets fixes et deux facteurs sont à effets aléatoires.

Un facteur contrôlé A se présente sous I modalités, chacune d'entre elles étant notée A_i. Les β_j représentent un échantillon de taille J prélevé dans une population importante. Nous admettrons que les effets des B_j, les β_j, sont distribués suivant une loi normale centrée de variance σ^2_B. Les γ_k représentent un échantillon de taille K prélevé dans une population importante. Nous admettrons que les effets des C_k, les γ_k, sont distribués suivant une loi normale centrée de variance σ^2_C. Pour chacun des couples de modalités (A_i, B_j, C_k) on effectue $L \geqslant 2$ mesures d'une réponse Y qui est une variable continue. On note $n = I \times J \times K \times L$ le nombre total de mesures ayant été effectuées.

On introduit le modèle :

$$
\begin{gathered}
Y_{i,j,k,l} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{i,j} + (\alpha\gamma)_{i,k} + (\beta\gamma)_{j,k} + (\alpha\beta\gamma)_{i,j,k} + \epsilon_{i,j,k,l},\\
i = 1 \ldots I,\ j = 1 \ldots J,\ k = 1 \ldots K,\ l = 1 \ldots L,\\
\text{avec les contraintes supplémentaires } \sum_{i=1}^{I} \alpha_i = 0,\\
\sum_{i=1}^{I} (\alpha\beta)_{i,j} = 0,\ \forall j \in \{1, \ldots, J\} \text{ et } \sum_{i=1}^{I} (\alpha\gamma)_{i,k} = 0,\ \forall k \in \{1, \ldots, K\},
\end{gathered}
$$

$$\sum_{i=1}^{I}(\alpha\beta\gamma)_{i,j,k} = 0,\ \forall (j,k) \in \{1,\ldots,J\} \times \{1,\ldots,K\},$$

où $Y_{i,j,k}$ est la valeur prise par la réponse Y dans les conditions (A_i, B_j, C_k) lors du $l-$ème essai. On suppose que :

$$\begin{gathered}
\mathcal{L}\left(\beta_j\right) = \mathcal{N}(0,\sigma_B^2),\ \forall\ j, 1 \leqslant j \leqslant J,\\
\mathcal{L}\left(\gamma_k\right) = \mathcal{N}(0,\sigma_C^2),\ \forall\ k, 1 \leqslant k \leqslant K,\\
\mathcal{L}\left((\alpha\beta)_{i,j}\right) = \mathcal{N}(0,\sigma_{AB}^2),\ \forall\ (i,j), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\\
\mathcal{L}\left((\alpha\gamma)_{i,k}\right) = \mathcal{N}(0,\sigma_{AC}^2),\ \forall\ (i,k), 1 \leqslant i \leqslant I,\ 1 \leqslant k \leqslant K,\\
\mathcal{L}\left((\beta\gamma)_{j,k}\right) = \mathcal{N}(0,\sigma_{BC}^2),\ \forall\ (j,k), 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K,\\
\mathcal{L}\left((\alpha\beta\gamma)_{i,j,k}\right) = \mathcal{N}(0,\sigma_{ABC}^2),\ \forall\ (i,j,k), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K,
\end{gathered}$$

ainsi que l'indépendance des effets aléatoires β_j, γ_k, $(\alpha\beta)_{i,j}$, $(\alpha\gamma)_{i,k}$, $(\beta\gamma)_{j,k}$ et $(\alpha\beta\gamma)_{i,j,k}$. On postule les hypothèses classiques suivantes pour les erreurs :

$$\begin{gathered}
\forall\ (i,j,k,l), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K,\ 1 \leqslant l \leqslant L,\ \mathcal{L}(\epsilon_{i,j,k,l}) = \mathcal{N}(0,\sigma^2),\\
\mathrm{Cov}(\epsilon_{i,j,k,l}, \epsilon_{m,n,o,p}) = 0 \text{ si } (i,j,k,l) \neq (m,n,o,p)\\
\text{avec } 1 \leqslant i,m \leqslant I,\ 1 \leqslant j,n \leqslant J,\ 1 \leqslant k,o \leqslant K \text{ et } 1 \leqslant l,p \leqslant L,
\end{gathered}$$

ainsi que l'indépendance des effets aléatoires β_j, γ_k, $(\alpha\beta)_{i,j}$, $(\alpha\gamma)_{i,k}$, $(\beta\gamma)_{j,k}$ et $(\alpha\beta\gamma)_{i,j,k}$ et des erreurs $\epsilon_{i,j,k,l}$.

On suppose que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fera l'objet d'un autre paragraphe.

On utilise les quantités sc_A, sc_B, sc_C, sc_{AB}, sc_{AC}, sc_{BC}, sc_{ABC}, sc_R et sc_{TOT} introduites à la section 6.1.2.

On rappelle la relation fondamentale de l'ANOVA :

$$sc_{TOT} = sc_A + sc_B + sc_C + sc_{AB} + sc_{AC} + sc_{BC} + s_{ABC} + sc_R.$$

On introduit les dégres de liberté (Ddl) associés à chaque ligne du tableau de l'ANOVA :

Source	Degrés de liberté
Facteur A	$n_A = I - 1$
Facteur B	$n_B = J - 1$
Facteur C	$n_C = K - 1$
Interaction AB	$n_{AB} = (I-1)(J-1)$
Interaction AC	$n_{AC} = (I-1)(K-1)$
Interaction BC	$n_{BC} = (J-1)(K-1)$
Interaction ABC	$n_{ABC} = (I-1)(J-1)(K-1)$
Résiduelle	$n_R = IJK(L-1)$
Totale	$n_{TOT} = IJK - 1$

On résume ces informations dans le tableau de l'ANOVA ci-dessous :

Source	Variation	Ddl	Carré Moyen	F	Décision
Facteur A	sc_A	n_A	$s_A^2 = \frac{sc_A}{n_A}$	$f_A = \frac{s_A^2}{s_{AB}^2 + s_{AC}^2 - s_{ABC}^2}$	$\mathcal{H}_0'$ ou $\mathcal{H}_1'$
Facteur B	sc_B	n_B	$s_B^2 = \frac{sc_B}{n_B}$	$f_B = \frac{s_B^2}{s_{BC}^2}$	$\mathcal{H}_0''$ ou $\mathcal{H}_1''$
Facteur C	sc_C	n_C	$s_C^2 = \frac{sc_C}{n_C}$	$f_C = \frac{s_C^2}{s_{BC}^2}$	$\mathcal{H}_0'''$ ou $\mathcal{H}_1'''$
Interaction AB	sc_{AB}	n_{AB}	$s_{AB}^2 = \frac{sc_{AB}}{n_{AB}}$	$f_{AB} = \frac{s_{AB}^2}{s_{ABC}^2}$	$\mathcal{H}_0^{(4)}$ ou $\mathcal{H}_1^{(4)}$
Interaction AC	sc_{AC}	n_{AC}	$s_{AC}^2 = \frac{sc_{AC}}{n_{AC}}$	$f_{AC} = \frac{s_{AC}^2}{s_{ABC}^2}$	$\mathcal{H}_0^{(5)}$ ou $\mathcal{H}_1^{(5)}$
Interaction BC	sc_{BC}	n_{BC}	$s_{BC}^2 = \frac{sc_{BC}}{n_{BC}}$	$f_{BC} = \frac{s_{BC}^2}{s_R^2}$	$\mathcal{H}_0^{(6)}$ ou $\mathcal{H}_1^{(6)}$
Interaction ABC	sc_{ABC}	n_{ABC}	$s_{ABC}^2 = \frac{sc_{ABC}}{n_{ABC}}$	$f_{ABC} = \frac{s_{ABC}^2}{s_R^2}$	$\mathcal{H}_0^{(7)}$ ou $\mathcal{H}_1^{(7)}$
Résiduelle	sc_R	n_R	$s_R^2 = \frac{sc_R}{n_R}$		
Totale	sc_{TOT}	n_{TOT}			

On souhaite faire les tests d'hypothèse suivants :

$$\boxed{\mathcal{H}_0^{'} : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0}$$

contre

$$\boxed{\mathcal{H}_1^{'} : \text{Il existe } i_0 \in \{1, 2, \ldots, I\} \text{ tel que } \alpha_{i_0} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{'}$ précédente d'absence d'effet du facteur A et lorsque les conditions de validité du modèle sont respectées, f_A est la réalisation d'une variable aléatoire qui suit **approximativement**[8] une loi de Fisher à $I-1$ et $n^{'}$ degrés de liberté avec :

$$n^{'} = \frac{\left(s_{AB}^2 + s_{AC}^2 - s_R^2\right)^2}{\dfrac{\left(s_{AB}^2\right)^2}{(I-1)(J-1)} + \dfrac{\left(s_{AC}^2\right)^2}{(I-1)(K-1)} + \dfrac{\left(s_R^2\right)^2}{(I-1)(J-1)(K-1)}}.$$

On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_A est supérieure ou égale à la valeur critique issue de la table. Lorsque l'hypothèse nulle $\mathcal{H}_0^{'}$ est rejetée on peut procéder à des comparaisons multiples des différents effets des niveaux du facteur voir la section 10.

$$\boxed{\mathcal{H}_0^{''} : \sigma_B^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1^{''} : \sigma_B^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{''}$ précédente d'absence d'effet du facteur B et lorsque les conditions de validité du modèle sont respectées, f_B est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $J-1$ et $(J-1)(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_B est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0^{'''} : \sigma_C^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1^{'''} : \sigma_C^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{'''}$ précédente d'absence d'effet du facteur C et lorsque les conditions de validité du modèle sont respectées, f_C est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $K-1$ et $(J-1)(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_C est supérieure ou égale à la valeur critique issue de la table.

[8] On utilise toujours l'approximation de Satterthwaite.

$$\boxed{\mathcal{H}_0^{(4)} : \sigma_{AB}^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1^{(4)} : \sigma_{AB}^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{(4)}$ précédente d'absence d'effet de l'interaction des facteurs A et B et lorsque les conditions de validité du modèle sont respectées, f_{AB} est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $(I-1)(J-1)$ et $(I-1)(J-1)(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_{AB} est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0^{(5)} : \sigma_{AC}^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1^{(5)} : \sigma_{AC}^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{(5)}$ précédente d'absence d'effet de l'interaction des facteurs A et C et lorsque les conditions de validité du modèle sont respectées, f_{AC} est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $(I-1)(K-1)$ et $(I-1)(J-1)(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_{AC} est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0^{(6)} : \sigma_{BC}^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1^{(6)} : \sigma_{BC}^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{(6)}$ précédente d'absence d'effet de l'interaction des facteurs B et C et lorsque les conditions de validité du modèle sont respectées, f_{BC} est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $(J-1)(K-1)$ et $IJK(L-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_{BC} est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0^{(7)} : \sigma_{ABC}^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1^{(7)} : \sigma_{ABC}^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{(7)}$ précédente d'absence d'effet de l'interaction d'ordre trois des facteurs A, B et C et lorsque les conditions de validité du modèle sont respectées, f_{ABC} est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $(I-1)(J-1)(K-1)$ et $IJK(L-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_{ABC} est supérieure ou égale à la valeur critique issue de la table.

Les estimations $\widehat{\mu}$, $\widehat{\alpha_1}$, ..., $\widehat{\alpha_I}$, $\widehat{\sigma_B^2}$, $\widehat{\sigma_C^2}$, $\widehat{\sigma_{AB}^2}$, $\widehat{\sigma_{AC}^2}$, $\widehat{\sigma_{BC}^2}$, $\widehat{\sigma_{ABC}^2}$, $\widehat{\sigma^2}$ des paramètres μ, α_1, ..., α_I, σ_B^2, σ_C^2, σ_{AB}^2, σ_{AC}^2, σ_{BC}^2, σ_{ABC}^2, σ^2 du modèle se déduisent des formules suivantes :

$$
\begin{aligned}
&\widehat{\mu} = y_{\bullet,\bullet,\bullet,\bullet} = \overline{y},\\
&\widehat{\alpha_i} = y_{i,\bullet,\bullet,\bullet} - \widehat{\mu},\ 1 \leqslant i \leqslant I,\\
&\widehat{\sigma_B^2} = \frac{1}{IKL}\left(s_B^2 - s_R^2\right),\ \widehat{\sigma_C^2} = \frac{1}{IJL}\left(s_C^2 - s_R^2\right),\\
&\widehat{\sigma_{AB}^2} = \frac{1}{KL}\left(s_{AB}^2 - s_R^2\right),\\
&\widehat{\sigma_{AC}^2} = \frac{1}{JL}\left(s_{AC}^2 - s_R^2\right),\\
&\widehat{\sigma_{BC}^2} = \frac{1}{IL}\left(s_{BC}^2 - s_R^2\right),\\
&\widehat{\sigma_{ABC}^2} = \frac{1}{L}\left(s_{ABC}^2 - s_R^2\right),\\
&\widehat{\sigma^2} = \frac{sc_R}{IJK(L-1)} = s_R^2.
\end{aligned}
$$

7. Analyse de la variance à trois facteurs totalement emboîtés

Dans toute cette section, on utilise les notations définies à la section 1.

On est dans la situation particulière, semblable à celle de la section 5, où les effets des niveaux du facteur B n'ont pas de signification concrète, par exemple ces niveaux dépendent du niveau du facteur A considéré et une étude des effets principaux du facteur B n'a pas de pertinence. On complique la situation en rajoutant un niveau d'imbrication par rapport aux modèles exposés à la section 5. Ainsi le facteur C est emboîté dans le facteur B qui lui même est emboîté dans le facteur A.
Un tel plan d'expérience est dit complètement hiérarchisé ou totalement emboîté[9]. Il n'est utilisable que si l'on dispose de répétitions. Dans le cas contraire où les essais ne seraient pas répétés, l'effet dû au facteur C ne pourra être étudié et le modèle que l'on devra utiliser pour étudier les données sera l'un de ceux exposés à la section 5.

[9]en anglais *completely nested model* ou *completely hierarchical model.*

Pour signaler cette situation particulière on note les effets du facteur B par $\beta_{j(i)}$ et ceux du facteur C par $\gamma_{k(j(i))}$.

7.1. Modèles à effets fixes

7.1.1. Avec répétitions

Un facteur contrôlé A se présente sous I modalités, chacune d'entre elles étant notée A_i. Un facteur contrôlé B se présente sous J modalités, chacune d'entre elles dépendant du niveau A_i du facteur A et étant alors notée $B_{j(i)}$. Un facteur contrôlé C se présente sous K modalités, chacune d'entre elles dépendant du niveau $B_{j(i)}$ du facteur B et donc du niveau A_i du facteur A et étant alors notée $C_{k(j(i))}$. Pour chacun des couples de modalités $(A_i, B_{j(i)}, C_{k(j(i))})$ on effectue $L \geqslant 2$ mesures d'une réponse Y qui est une variable continue. On note $n = I \times J \times K \times L$ le nombre total de mesures ayant été effectuées.

On introduit le modèle :

$$Y_{i,j,k,l} = \mu + \alpha_i + \beta_{j(i)} + \gamma_{k(j(i))} + \epsilon_{i,j,k,l},$$
$$i = 1 \dots I,\ j = 1 \dots J,\ k = 1 \dots K,\ l = 1 \dots L,$$
$$\text{avec les contraintes supplémentaires } \sum_{i=1}^{I} \alpha_i = 0,\ \sum_{j=1}^{J} \beta_{j(i)} = 0,\ \forall i \in \{1, \dots, I\},$$
$$\text{et } \sum_{k=1}^{K} \gamma_{k(j(i))} = 0,\ \forall (i,j) \in \{1, \dots, I\} \times \{1, \dots, J\},$$

où $Y_{i,j,k,l}$ est la valeur prise par la réponse Y dans les conditions $(A_i, B_{j(i)}, C_{k(j(i))})$ lors du $l-$ème essai. On postule les hypothèses classiques suivantes pour les erreurs :

$$\forall\ (i,j,k,l), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K,\ 1 \leqslant l \leqslant L,\ \mathcal{L}(\epsilon_{i,j,k,l}) = \mathcal{N}(0, \sigma^2),$$
$$\mathrm{Cov}(\epsilon_{i,j,k,l}, \epsilon_{m,n,o,p}) = 0 \text{ si } (i,j,k,l) \neq (m,n,o,p)$$
$$\text{avec } 1 \leqslant i, m \leqslant I,\ 1 \leqslant j, n \leqslant J,\ 1 \leqslant k, o \leqslant K \text{ et } 1 \leqslant l, p \leqslant L.$$

On suppose que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fera l'objet d'un autre paragraphe.

On rappelle que la variation due au facteur A est définie par :

$$sc_A = JKL \sum_{i=1}^{I} (y_{i,\bullet,\bullet,\bullet} - y_{\bullet,\bullet,\bullet,\bullet})^2.$$

On rappelle que la variation due au facteur B dans A est définie par :

$$sc_{B|A} = KL \sum_{i=1}^{I} \sum_{j=1}^{J} (y_{i,j,\bullet,\bullet} - y_{i,\bullet,\bullet,\bullet})^2.$$

On rappelle que la variation due au facteur C dans B qui lui même est dans A est définie par :

$$sc_{C|B|A} = L \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (y_{i,j,k,\bullet} - y_{i,j,\bullet,\bullet})^2.$$

La variation résiduelle est quant à elle définie par :

$$sc_R = L \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (y_{i,j,k,l} - y_{i,j,k,\bullet})^2.$$

Enfin la variation totale est égale à :

$$sc_{TOT} = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{l=1}^{L} (y_{i,j,k,l} - y_{\bullet,\bullet,\bullet,\bullet})^2.$$

On rappelle la relation fondamentale de l'ANOVA :

$$sc_{TOT} = sc_A + sc_{B|A} + sc_{C|B|A} + sc_R.$$

On introduit les dégres de liberté (Ddl) associés à chaque ligne du tableau de l'ANOVA :

Source	Degrés de liberté
Facteur A	$n_A = I - 1$
Facteur B dans A	$n_{B\|A} = I(J-1)$
Facteur C dans B	$n_{C\|B\|A} = IJ(K-1)$
Résiduelle	$n_R = IJK(L-1)$
Totale	$n_{TOT} = IJKL - 1$

On résume ces informations dans le tableau de l'ANOVA ci-dessous :

Source	Variation	Ddl	Carré Moyen	F	Décision
Facteur A	sc_A	n_A	$s_A^2 = \dfrac{sc_A}{n_A}$	$f_A = \dfrac{s_A^2}{s_R^2}$	$\mathcal{H}_0^{'}$ ou $\mathcal{H}_1^{'}$
Facteur B dans A	$sc_{B\|A}$	$n_{B\|A}$	$s_{B\|A}^2 = \dfrac{sc_{B\|A}}{n_{B\|A}}$	$f_{B\|A} = \dfrac{s_{B\|A}^2}{s_R^2}$	$\mathcal{H}_0^{''}$ ou $\mathcal{H}_1^{''}$
Facteur C dans B	$sc_{C\|B\|A}$	$n_{C\|B\|A}$	$s_{C\|B\|A}^2 = \dfrac{sc_{C\|B\|A}}{n_{C\|B\|A}}$	$f_{C\|B\|A} = \dfrac{s_{C\|B\|A}^2}{s_R^2}$	$\mathcal{H}_0^{'''}$ ou $\mathcal{H}_1^{'''}$
Résiduelle	sc_R	n_R	$s_R^2 = \dfrac{sc_R}{n_R}$		
Totale	sc_{TOT}	n_{TOT}			

On souhaite faire les tests d'hypothèse suivants :

$$\boxed{\mathcal{H}_0^{'} : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0}$$

contre

$$\boxed{\mathcal{H}_1^{'} : \text{Il existe } i_0 \in \{1, 2, \ldots, I\} \text{ tel que } \alpha_{i_0} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{'}$ précédente d'absence d'effet du facteur A et lorsque les conditions de validité du modèle sont respectées, f_A est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $I-1$ et $IJK(L-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_A est supérieure ou égale à la valeur critique issue de la table. Lorsque l'hypothèse nulle $\mathcal{H}_0^{'}$ est rejetée on peut procéder à des comparaisons multiples des différents effets des niveaux du facteur voir la section 10.

$$\boxed{\mathcal{H}_0^{''} : \beta_{1(1)} = \beta_{2(1)} = \cdots = \beta_{J(1)} = \beta_{1(2)} = \cdots = \beta_{J(I)} = 0}$$

contre

$$\boxed{\mathcal{H}_1^{''} : \text{Il existe } (i_0, j_0) \in \{1, 2, \ldots, I\} \times \{1, 2, \ldots, J\} \text{ tel que } \beta_{j_0(i_0)} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0''$ précédente d'absence d'effet du facteur B dans A et lorsque les conditions de validité du modèle sont respectées, $f_{B|A}$ est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $I(J-1)$ et $IJK(L-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur $f_{B|A}$ est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0''' : \gamma_{1(1(1))} = \gamma_{2(1(1))} = \cdots = \gamma_{K(1(1))} = \gamma_{1(1(2))} = \cdots = \gamma_{K(J(I))} = 0}$$

contre

$$\boxed{\mathcal{H}_1''' : \exists\ (i_0, j_0, k_0) \in \{1, 2, \ldots, I\} \times \{1, 2, \ldots, J\} \times \{1, 2, \ldots, K\} \mid \gamma_{k_0(j_0(i_0))} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0'''$ précédente d'absence d'effet du facteur C dans B et lorsque les conditions de validité du modèle sont respectées, $f_{C|B|A}$ est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $IJ(K-1)$ et $IJK(L-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur $f_{C|B|A}$ est supérieure ou égale à la valeur critique issue de la table.

Les estimations $\widehat{\mu}$, $\widehat{\alpha_1}$, ..., $\widehat{\alpha_I}$, $\widehat{\beta_{1(1)}}$, ..., $\widehat{\beta_{J(I)}}$, $\widehat{\gamma_{1(1(1))}}$, ..., $\widehat{\gamma_{K(J(I))}}$, $\widehat{\sigma^2}$ des paramètres μ, α_1, ..., α_I, $\beta_{1(1)}$, ..., $\beta_{J(I)}$, $\gamma_{1(1(1))}$, ..., $\gamma_{K(J(I))}$, σ^2 du modèle se déduisent des formules suivantes :

$$\begin{aligned}
&\widehat{\mu} = y_{\bullet,\bullet,\bullet,\bullet} = \overline{y},\\
&\widehat{\alpha_i} = y_{i,\bullet,\bullet,\bullet} - \widehat{\mu},\ 1 \leqslant i \leqslant I,\\
&\widehat{\beta_{j(i)}} = y_{i,j,\bullet,\bullet} - y_{i,\bullet,\bullet,\bullet},\ 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\\
&\widehat{\gamma_{k(j(i))}} = y_{i,j,k,\bullet} - y_{i,j,\bullet,\bullet},\ 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K,\\
&\widehat{\sigma^2} = \frac{sc_R}{IJK(L-1)} = s_R^2.
\end{aligned}$$

7.2. Modèles à effets aléatoires

7.2.1. Avec répétitions

Les α_i représentent un échantillon de taille I prélevé dans une population importante. Nous admettrons que les effets des A_i, les α_i, sont distribués suivant une loi normale centrée de variance σ_A^2. Les $\beta_{j(i)}$ représentent un échantillon de taille J prélevé dans une population importante dépendant du niveau A_i du facteur A. Nous admettrons que les effets des $B_{j(i)}$, les $\beta_{j(i)}$, sont distribués suivant une loi normale centrée de variance $\sigma_{B|A}^2$. Les $\gamma_{k(j(i))}$ représentent un échantillon de taille K prélevé dans une population importante dépendant du niveau $B_{j(i)}$ du facteur B et donc du niveau A_i du facteur A. Nous admettrons que les effets des $C_{k(j(i))}$, les $\gamma_{k(j(i))}$, sont distribués suivant une loi normale centrée

de variance $\sigma^2_{C|B|A}$. Pour chacun des couples de modalités $(A_i, B_{j(i)}, C_{k(j(i))})$ on effectue $L \geqslant 2$ mesures d'une réponse Y qui est une variable continue. On note $n = I \times J \times K \times L$ le nombre total de mesures ayant été effectuées.

On introduit le modèle :

$$Y_{i,j,k,l} = \mu + \alpha_i + \beta_{j(i)} + \gamma_{k(j(i))} + \epsilon_{i,j,k,l},$$
$$i = 1 \dots I,\ j = 1 \dots J,\ k = 1 \dots K,\ l = 1 \dots L,$$

où $Y_{i,j,k,l}$ est la valeur prise par la réponse Y dans les conditions $(A_i, B_{j(i)}, C_{k(j(i))})$ lors du $l-$ème essai. On suppose que :

$$\mathcal{L}\left(\alpha_i\right) = \mathcal{N}(0, \sigma^2_A),\ \forall\ i, 1 \leqslant i \leqslant I,$$
$$\mathcal{L}\left(\beta_{j(i)}\right) = \mathcal{N}(0, \sigma^2_{B|A}),\ \forall\ (i,j), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,$$
$$\mathcal{L}\left(\gamma_{k(j(i))}\right) = \mathcal{N}(0, \sigma^2_{C|B|A}),\ \forall\ (i,j,k), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K,$$

ainsi que l'indépendance des effets aléatoires α_i, $\beta_{j(i)}$, et $\gamma_{k(j(i))}$.
On postule les hypothèses classiques suivantes pour les erreurs :

$$\forall\ (i,j,k,l), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K,\ 1 \leqslant l \leqslant L,\ \mathcal{L}(\epsilon_{i,j,k,l}) = \mathcal{N}(0, \sigma^2),$$
$$\mathrm{Cov}(\epsilon_{i,j,k,l}, \epsilon_{m,n,o,p}) = 0 \text{ si } (i,j,k,l) \neq (m,n,o,p)$$
$$\text{avec } 1 \leqslant i, m \leqslant I,\ 1 \leqslant j, n \leqslant J,\ 1 \leqslant k, o \leqslant K \text{ et } 1 \leqslant l, p \leqslant L,$$

ainsi que l'indépendance des effets aléatoires α_i, $\beta_{j(i)}$, et $\gamma_{k(j(i))}$ et des erreurs $\epsilon_{i,j,k,l}$.

On suppose que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fera l'objet d'un autre paragraphe.

On utilise les quantités sc_A, $sc_{B|A}$, $sc_{C|B|A}$, sc_R et sc_{TOT} introduites à la section 7.1.1.

On rappelle la relation fondamentale de l'ANOVA :

$$sc_{TOT} = sc_A + sc_{B|A} + sc_{C|B|A} + sc_R.$$

On introduit les dégres de liberté (Ddl) associés à chaque ligne du tableau de l'ANOVA :

Source	Degrés de liberté
Facteur A	$n_A = I - 1$
Facteur B dans A	$n_{B\|A} = I(J-1)$
Facteur C dans B	$n_{C\|B\|A} = IJ(K-1)$
Résiduelle	$n_R = IJK(L-1)$
Totale	$n_{TOT} = IJKL - 1$

On résume ces informations dans le tableau de l'ANOVA ci-dessous :

Source	Variation	Ddl	Carré Moyen	F	Décision
Facteur A	sc_A	n_A	$s_A^2 = \frac{sc_A}{n_A}$	$f_A = \frac{s_A^2}{s_{B\|A}^2}$	$\mathcal{H}_0^{'}$ ou $\mathcal{H}_1^{'}$
Facteur B dans A	$sc_{B\|A}$	$n_{B\|A}$	$s_{B\|A}^2 = \frac{sc_{B\|A}}{n_{B\|A}}$	$f_{B\|A} = \frac{s_{B\|A}^2}{s_{C\|B\|A}^2}$	$\mathcal{H}_0^{''}$ ou $\mathcal{H}_1^{''}$
Facteur C dans B	$sc_{C\|B\|A}$	$n_{C\|B\|A}$	$s_{C\|B\|A}^2 = \frac{sc_{C\|B\|A}}{n_{C\|B\|A}}$	$f_{C\|B\|A} = \frac{s_{C\|B\|A}^2}{s_R^2}$	$\mathcal{H}_0^{'''}$ ou $\mathcal{H}_1^{'''}$
Résiduelle	sc_R	n_R	$s_R^2 = \frac{sc_R}{n_R}$		
Totale	sc_{TOT}	n_{TOT}			

On souhaite faire les tests d'hypothèse suivants :

$$\boxed{\mathcal{H}_0^{'} : \sigma_A^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1^{'} : \sigma_A^2 \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{'}$ précédente d'absence d'effet du facteur A et lorsque les conditions de validité du modèle sont respectées, f_A est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $I-1$ et $I(J-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_A est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0'' : \sigma^2_{B|A} = 0}$$

contre

$$\boxed{\mathcal{H}_1'' : \sigma^2_{B|A} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0''$ précédente d'absence d'effet du facteur B dans A et lorsque les conditions de validité du modèle sont respectées, $f_{B|A}$ est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $I(J-1)$ et $IJ(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur $f_{B|A}$ est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0''' : \sigma^2_{C|B|A} = 0}$$

contre

$$\boxed{\mathcal{H}_1''' : \sigma^2_{C|B|A} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0'''$ précédente d'absence d'effet du facteur C dans B et lorsque les conditions de validité du modèle sont respectées, $f_{C|B|A}$ est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $IJ(K-1)$ et $IJK(L-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur $f_{C|B|A}$ est supérieure ou égale à la valeur critique issue de la table.

Les estimations $\widehat{\mu}$, $\widehat{\sigma^2_A}$, $\widehat{\sigma^2_{B|A}}$, $\widehat{\sigma^2_{C|B|A}}$, $\widehat{\sigma^2}$ des paramètres μ, σ^2_A, $\sigma^2_{B|A}$, $\sigma^2_{C|B|A}$, σ^2 du modèle se déduisent des formules suivantes :

$$\begin{aligned}
&\widehat{\mu} = y_{\bullet,\bullet,\bullet,\bullet} = \overline{y},\\
&\widehat{\sigma^2_A} = \frac{1}{JKL}\left(s^2_A - s^2_{B|A}\right),\\
&\widehat{\sigma^2_{B|A}} = \frac{1}{KL}\left(s^2_{B|A} - s^2_{C|B|A}\right),\\
&\widehat{\sigma^2_{C|B|A}} = \frac{1}{L}\left(s^2_{C|B|A} - s^2_R\right),\\
&\widehat{\sigma^2} = \frac{sc_R}{IJK(L-1)} = s^2_R.
\end{aligned}$$

7.3. Modèles à effets mixtes

7.3.1. Avec répétitions

Premier cas : Deux facteurs sont à effets fixes et un facteur est à effets aléatoires.

Un facteur contrôlé A se présente sous I modalités, chacune d'entre elles étant notée A_i. Un facteur contrôlé B se présente sous J modalités, chacune d'entre elles dépendant du niveau A_i du facteur A et étant alors notée $B_{j(i)}$. Nous admettrons que les effets des $C_{k(j(i))}$, les $\gamma_{k(j(i))}$, sont distribués suivant une loi normale centrée de variance $\sigma^2_{C|B|A}$. Pour chacun des couples de modalités $(A_i, B_{j(i)}, C_{k(j(i))})$ on effectue $L \geqslant 2$ mesures d'une réponse Y qui est une variable continue. On note $n = I \times J \times K \times L$ le nombre total de mesures ayant été effectuées.

On introduit le modèle :

$$Y_{i,j,k,l} = \mu + \alpha_i + \beta_{j(i)} + \gamma_{k(j(i))} + \epsilon_{i,j,k,l},$$
$$i = 1 \dots I,\ j = 1 \dots J,\ k = 1 \dots K,\ l = 1 \dots L,$$
$$\text{avec les contraintes supplémentaires } \sum_{i=1}^{I} \alpha_i = 0,\ \sum_{j=1}^{J} \beta_{j(i)} = 0,\ \forall i \in \{1, \dots, I\},$$

où $Y_{i,j,k,l}$ est la valeur prise par la réponse Y dans les conditions $(A_i, B_{j(i)}, C_{k(j(i))})$ lors du $l-$ème essai. On suppose que :

$$\mathcal{L}\left(\gamma_{k(j(i))}\right) = \mathcal{N}(0, \sigma^2_{C|B|A}),\ \forall\ (i,j,k), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K,$$

ainsi que l'indépendance des effets aléatoires $\gamma_{k(j(i))}$.
On postule les hypothèses classiques suivantes pour les erreurs :

$$\forall\ (i,j,k,l), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K,\ 1 \leqslant l \leqslant L,\ \mathcal{L}(\epsilon_{i,j,k,l}) = \mathcal{N}(0, \sigma^2),$$
$$\mathrm{Cov}(\epsilon_{i,j,k,l}, \epsilon_{m,n,o,p}) = 0 \text{ si } (i,j,k,l) \neq (m,n,o,p)$$
$$\text{avec } 1 \leqslant i, m \leqslant I,\ 1 \leqslant j, n \leqslant J,\ 1 \leqslant k, o \leqslant K \text{ et } 1 \leqslant l, p \leqslant L,$$

ainsi que l'indépendance des effets aléatoires $\gamma_{k(j(i))}$ et des erreurs $\epsilon_{i,j,k,l}$.

On suppose que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fera l'objet d'un autre paragraphe.

On utilise les quantités sc_A, $sc_{B|A}$, $sc_{C|B|A}$, sc_R et sc_{TOT} introduites à la section 7.1.1.

On rappelle la relation fondamentale de l'ANOVA :

$$sc_{TOT} = sc_A + sc_{B|A} + sc_{C|B|A} + sc_R.$$

On introduit les dégres de liberté (Ddl) associés à chaque ligne du tableau de l'ANOVA :

Source	Degrés de liberté
Facteur A	$n_A = I - 1$
Facteur B dans A	$n_{B\|A} = I(J-1)$
Facteur C dans B	$n_{C\|B\|A} = IJ(K-1)$
Résiduelle	$n_R = IJK(L-1)$
Totale	$n_{TOT} = IJKL - 1$

On résume ces informations dans le tableau de l'ANOVA ci-dessous :

Source	Variation	Ddl	Carré Moyen	F	Décision
Facteur A	sc_A	n_A	$s_A^2 = \dfrac{sc_A}{n_A}$	$f_A = \dfrac{s_A^2}{s_{C\|B\|A}^2}$	$\mathcal{H}_0^{'}$ ou $\mathcal{H}_1^{'}$
Facteur B dans A	$sc_{B\|A}$	$n_{B\|A}$	$s_{B\|A}^2 = \dfrac{sc_{B\|A}}{n_{B\|A}}$	$f_{B\|A} = \dfrac{s_{B\|A}^2}{s_{C\|B\|A}^2}$	$\mathcal{H}_0^{''}$ ou $\mathcal{H}_1^{''}$
Facteur C dans B	$sc_{C\|B\|A}$	$n_{C\|B\|A}$	$s_{C\|B\|A}^2 = \dfrac{sc_{C\|B\|A}}{n_{C\|B\|A}}$	$f_{C\|B\|A} = \dfrac{s_{C\|B\|A}^2}{s_R^2}$	$\mathcal{H}_0^{'''}$ ou $\mathcal{H}_1^{'''}$
Résiduelle	sc_R	n_R	$s_R^2 = \dfrac{sc_R}{n_R}$		
Totale	sc_{TOT}	n_{TOT}			

On souhaite faire les tests d'hypothèse suivants :

$$\boxed{\mathcal{H}_0^{'} : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0}$$

contre

$$\boxed{\mathcal{H}_1^{'} : \text{Il existe } i_0 \in \{1, 2, \ldots, I\} \text{ tel que } \alpha_{i_0} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{'}$ précédente d'absence d'effet du facteur A et lorsque les conditions de validité du modèle sont respectées, f_A est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $I-1$ et $IJ(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_A est supérieure ou égale à la valeur critique issue de la table. Lorsque l'hypothèse nulle $\mathcal{H}_0^{'}$ est rejetée on peut procéder à des comparaisons multiples des différents effets des niveaux du facteur voir la section 10.

$$\boxed{\mathcal{H}_0'' : \beta_{1(1)} = \beta_{2(1)} = \cdots = \beta_{J(1)} = \beta_{1(2)} = \cdots = \beta_{J(I)} = 0}$$

contre

$$\boxed{\mathcal{H}_1'' : \text{Il existe } (i_0, j_0) \in \{1, 2, \ldots, I\} \times \{1, 2, \ldots, J\} \text{ tel que } \beta_{j_0(i_0)} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0''$ précédente d'absence d'effet du facteur B dans A et lorsque les conditions de validité du modèle sont respectées, $f_{B|A}$ est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $I(J-1)$ et $IJ(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur $f_{B|A}$ est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0''' : \sigma^2_{C|B|A} = 0}$$

contre

$$\boxed{\mathcal{H}_1''' : \sigma^2_{C|B|A} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0'''$ précédente d'absence d'effet du facteur C dans B et lorsque les conditions de validité du modèle sont respectées, $f_{C|B|A}$ est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $IJ(K-1)$ et $IJK(L-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur $f_{C|B|A}$ est supérieure ou égale à la valeur critique issue de la table.

Les estimations $\widehat{\mu}$, $\widehat{\alpha_1}$, $\ldots$, $\widehat{\alpha_I}$, $\widehat{\beta_{1(1)}}$, $\ldots$, $\widehat{\beta_{J(I)}}$, $\widehat{\sigma^2_{C|B|A}}$, $\widehat{\sigma^2}$ des paramètres μ, α_1, $\ldots$, α_I, $\beta_{1(1)}$, $\ldots$, $\beta_{J(I)}$, $\sigma^2_{C|B|A}$, σ^2 du modèle se déduisent des formules suivantes :

$$\begin{aligned}
&\widehat{\mu} = y_{\bullet,\bullet,\bullet,\bullet} = \overline{y},\\
&\widehat{\alpha_i} = y_{i,\bullet,\bullet,\bullet} - \widehat{\mu},\ 1 \leqslant i \leqslant I,\\
&\widehat{\beta_{j(i)}} = y_{i,j,\bullet,\bullet} - y_{i,\bullet,\bullet,\bullet},\ 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\\
&\widehat{\sigma^2_{C|B|A}} = \frac{1}{L}\left(s^2_{C|B|A} - s^2_R\right),\\
&\widehat{\sigma^2} = \frac{sc_R}{IJK(L-1)} = s^2_R.
\end{aligned}$$

Deuxième cas : Un facteur est à effets fixes et deux facteurs sont à effets aléatoires.

Un facteur contrôlé A se présente sous I modalités, chacune d'entre elles étant notée A_i. Les $\beta_{j(i)}$ représentent un échantillon de taille J prélevé dans une population importante dépendant du niveau A_i du facteur A. Nous admettrons que les effets des $B_{j(i)}$, les $\beta_{j(i)}$, sont distribués suivant une loi normale centrée de variance $\sigma^2_{B|A}$. Les $\gamma_{k(j(i))}$ représentent

un échantillon de taille K prélevé dans une population importante dépendant du niveau $B_{j(i)}$ du facteur B et donc du niveau A_i du facteur A. Nous admettrons que les effets des $C_{k(j(i))}$, les $\gamma_{k(j(i))}$, sont distribués suivant une loi normale centrée de variance $\sigma^2_{C|B|A}$. Pour chacun des couples de modalités $(A_i, B_{j(i)}, C_{k(j(i))})$ on effectue $L \geqslant 2$ mesures d'une réponse Y qui est une variable continue. On note $n = I \times J \times K \times L$ le nombre total de mesures ayant été effectuées.

On introduit le modèle :

$$Y_{i,j,k,l} = \mu + \alpha_i + \beta_{j(i)} + \gamma_{k(j(i))} + \epsilon_{i,j,k,l},$$
$$i = 1 \dots I,\ j = 1 \dots J,\ k = 1 \dots K,\ l = 1 \dots L,$$
$$\text{avec les contraintes supplémentaires } \sum_{i=1}^{I} \alpha_i = 0,$$

où $Y_{i,j,k,l}$ est la valeur prise par la réponse Y dans les conditions $(A_i, B_{j(i)}, C_{k(j(i))})$ lors du $l-$ème essai. On suppose que :

$$\mathcal{L}\left(\beta_{j(i)}\right) = \mathcal{N}(0, \sigma^2_{B|A}),\ \forall\ (i,j), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,$$
$$\mathcal{L}\left(\gamma_{k(j(i))}\right) = \mathcal{N}(0, \sigma^2_{C|B|A}),\ \forall\ (i,j,k), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K,$$

ainsi que l'indépendance des effets aléatoires $\beta_{j(i)}$, et $\gamma_{k(j(i))}$.
On postule les hypothèses classiques suivantes pour les erreurs :

$$\forall\ (i,j,k,l), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K,\ 1 \leqslant l \leqslant L,\ \mathcal{L}(\epsilon_{i,j,k,l}) = \mathcal{N}(0, \sigma^2),$$
$$\mathrm{Cov}(\epsilon_{i,j,k,l}, \epsilon_{m,n,o,p}) = 0 \text{ si } (i,j,k,l) \neq (m,n,o,p)$$
$$\text{avec } 1 \leqslant i, m \leqslant I,\ 1 \leqslant j, n \leqslant J,\ 1 \leqslant k, o \leqslant K \text{ et } 1 \leqslant l, p \leqslant L,$$

ainsi que l'indépendance des effets aléatoires $\beta_{j(i)}$, et $\gamma_{k(j(i))}$ et des erreurs $\epsilon_{i,j,k,l}$.

On suppose que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fera l'objet d'un autre paragraphe.

On utilise les quantités sc_A, $sc_{B|A}$, $sc_{C|B|A}$, sc_R et sc_{TOT} introduites à la section 7.1.1.

On rappelle la relation fondamentale de l'ANOVA :

$$sc_{TOT} = sc_A + sc_{B|A} + sc_{C|B|A} + sc_R.$$

On introduit les dégres de liberté (Ddl) associés à chaque ligne du tableau de l'ANOVA :

Source	Degrés de liberté
Facteur A	$n_A = I - 1$
Facteur B dans A	$n_{B\|A} = I(J-1)$
Facteur C dans B	$n_{C\|B\|A} = IJ(K-1)$
Résiduelle	$n_R = IJK(L-1)$
Totale	$n_{TOT} = IJKL - 1$

On résume ces informations dans le tableau de l'ANOVA ci-dessous :

Source	Variation	Ddl	Carré Moyen	F	Décision
Facteur A	sc_A	n_A	$s_A^2 = \frac{sc_A}{n_A}$	$f_A = \frac{s_A^2}{s_{B\|A}^2}$	$\mathcal{H}_0^{'}$ ou $\mathcal{H}_1^{'}$
Facteur B dans A	$sc_{B\|A}$	$n_{B\|A}$	$s_{B\|A}^2 = \frac{sc_{B\|A}}{n_{B\|A}}$	$f_{B\|A} = \frac{s_{B\|A}^2}{s_{C\|B\|A}^2}$	$\mathcal{H}_0^{''}$ ou $\mathcal{H}_1^{''}$
Facteur C dans B	$sc_{C\|B\|A}$	$n_{C\|B\|A}$	$s_{C\|B\|A}^2 = \frac{sc_{C\|B\|A}}{n_{C\|B\|A}}$	$f_{C\|B\|A} = \frac{s_{C\|B\|A}^2}{s_R^2}$	$\mathcal{H}_0^{'''}$ ou $\mathcal{H}_1^{'''}$
Résiduelle	sc_R	n_R	$s_R^2 = \frac{sc_R}{n_R}$		
Totale	sc_{TOT}	n_{TOT}			

On souhaite faire les tests d'hypothèse suivants :

$$\boxed{\mathcal{H}_0^{'} : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0}$$

contre

$$\boxed{\mathcal{H}_1^{'} : \text{Il existe } i_0 \in \{1, 2, \ldots, I\} \text{ tel que } \alpha_{i_0} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0^{'}$ précédente d'absence d'effet du facteur A et lorsque les conditions de validité du modèle sont respectées, f_A est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $I-1$ et $I(J-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur f_A est supérieure ou égale à la valeur critique issue de la table. Lorsque l'hypothèse nulle $\mathcal{H}_0^{'}$ est rejetée on peut procéder à des comparaisons multiples des différents effets des niveaux du facteur voir la section 10.

$$\boxed{\mathcal{H}_0'' : \sigma^2_{B|A} = 0}$$
contre
$$\boxed{\mathcal{H}_1'' : \sigma^2_{B|A} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0''$ précédente d'absence d'effet du facteur B dans A et lorsque les conditions de validité du modèle sont respectées, $f_{B|A}$ est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $I(J-1)$ et $IJ(K-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur $f_{B|A}$ est supérieure ou égale à la valeur critique issue de la table.

$$\boxed{\mathcal{H}_0''' : \sigma^2_{C|B|A} = 0}$$
contre
$$\boxed{\mathcal{H}_1''' : \sigma^2_{C|B|A} \neq 0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0'''$ précédente d'absence d'effet du facteur C dans B et lorsque les conditions de validité du modèle sont respectées, $f_{C|B|A}$ est la réalisation d'une variable aléatoire qui suit une loi de Fisher à $IJ(K-1)$ et $IJK(L-1)$ degrés de liberté. On conclut alors à l'aide de la $p-$valeur, rejet si elle est inférieure ou égale au seuil α du test, ou à l'aide d'une table, rejet si la valeur $f_{C|B|A}$ est supérieure ou égale à la valeur critique issue de la table.

Les estimations $\widehat{\mu}$, $\widehat{\alpha_1}$, ..., $\widehat{\alpha_I}$, $\widehat{\sigma^2_{B|A}}$, $\widehat{\sigma^2_{C|B|A}}$, $\widehat{\sigma^2}$ des paramètres μ, α_1, ..., α_I, $\sigma^2_{B|A}$, $\sigma^2_{C|B|A}$, σ^2 du modèle se déduisent des formules suivantes :

$$\begin{aligned}
&\widehat{\mu} = y_{\bullet,\bullet,\bullet,\bullet} = \overline{y},\\
&\widehat{\alpha_i} = y_{i,\bullet,\bullet,\bullet} - \widehat{\mu},\ 1 \leqslant i \leqslant I,\\
&\widehat{\sigma^2_{B|A}} = \frac{1}{KL}\left(s^2_{B|A} - s^2_{C|B|A}\right),\\
&\widehat{\sigma^2_{C|B|A}} = \frac{1}{L}\left(s^2_{C|B|A} - s^2_R\right),\\
&\widehat{\sigma^2} = \frac{sc_R}{IJK(L-1)} = s^2_R.
\end{aligned}$$

8. Analyse de la variance à trois facteurs partiellement emboîtés

8.1. Introduction

On dit qu'un modèle d'analyse de la variance est partiellement emboîté ou partiellement hiérarchisé si, contrairement aux modèles complètement emboîtés présentés aux sections 5

et 7, ce modèle fait intervenir certaines interactions. On remarque que pour qu'un modèle soit partiellement emboîté, il doit contenir au moins trois facteurs. Ce type de modèle nécessite généralement des répétitions, ceci sera précisé au cas par cas pour chacun des modèles exposés.

À nouveau chacun des facteurs peut être à effets fixes ou à effets aléatoires à condition de respecter les deux règles suivantes :

1. Toute interaction mettant en jeu un facteur à effets aléatoires est à effets aléatoires.
2. Tout terme emboîté dans un facteur à effets aléatoires est à effets aléatoires.

À cause de la grande multiplicité des modèles que l'on peut construire ainsi on se bornera à exposer ci-dessous les équations des différents modèles comportant trois facteurs et partiellement emboîtés que l'on peut construire.

Il est commode d'utiliser une représentation graphique pour améliorer la compréhension des modèles partiellement emboîtés. Pour plus de détails sur ce type de schéma, on consultera [4].

8.2. Les différents modèles partiellement emboîtés

8.2.1. Nécessitant des répétitions

Il s'agit du cas où le troisième facteur C est emboîté dans les différents termes d'un modèle de l'analyse de la variance à deux facteurs comportant les facteurs A et B.

$$Y_{i,j,k,l} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{i,j} + \gamma_{k|(i,j)} + \epsilon_{i,j,k,l}$$

avec les hypothèses de rigueur associées au fait que l'on considère que les facteurs sont à effets aléatoires ou à effets fixes, les conditions habituelles sur les lois, les variances et l'indépendance des erreurs $\epsilon_{i,j,k,l}$ et $1 \leqslant i \leqslant I$, $1 \leqslant j \leqslant J$, $1 \leqslant k \leqslant K$, $1 \leqslant l \leqslant L$.

Pour illustrer les deux règles exposées à la fin de la section 8.1 concernant les effets fixes ou aléatoires, on détaille les différentes déclinaisons de modèles que l'on peut considérer dans la situation de ce paragraphe :

- Tous les facteurs, A, B et C, sont à effets fixes.
- Les facteurs A et B sont à effets fixes, le facteur emboîté C est à effets aléatoires.
- L'un des facteurs, A ou B, est à effets aléatoires. Puisque ces deux facteurs jouent un rôle symétrique dans la définition du modèle, on suppose que le facteur A est à effets fixes et que c'est le facteur B qui est à effets aléatoires. Alors **nécessairement** on doit considérer que le facteur C, qui est emboîté dans le facteur B, est à effets aléatoires. De plus le terme d'interaction de A et de B doit aussi être **nécessairement** considéré

comme à effets aléatoires. Ainsi seuls les termes $\alpha_1, \ldots, \alpha_I$ associés aux effets du facteur A sont fixes.
- Les effets des deux facteurs A et B sont aléatoires et de ce fait C doit être nécessairement considéré comme un facteur à effets aléatoires : tous les facteurs du modèles sont à effets aléatoires.

8.2.2. Ne nécessitant pas de répétitions

Deux facteurs croisés B et C sont emboîtés dans un facteur A
Il s'agit du cas où les différents termes d'un modèle de l'analyse de la variance à deux facteurs comportant les facteurs B et C sont emboîtés dans un facteur A.

$$Y_{i,j,k,l} = \mu + \alpha_i + \beta_{j|i} + \gamma_{k|i} + (\beta\gamma)_{(j,k)|i} + \epsilon_{i,j,k,l}$$

avec les hypothèses de rigueur associées au fait que l'on considère que les facteurs sont à effets aléatoires ou à effets fixes, les conditions habituelles sur les lois, les variances et l'indépendance des erreurs et $1 \leqslant i \leqslant I$, $1 \leqslant j \leqslant J$, $1 \leqslant k \leqslant K$, $1 \leqslant l \leqslant L$. Dans le cas, on peut avoir le nombre de répétitions, L, égal à un, c'est-à-dire que l'on peut utiliser un plan d'expérience ne comportant pas de répétitions.

Pour illustrer les deux règles exposées à la fin de la section 8.1 concernant les effets fixes ou aléatoires, on détaille les différentes déclinaisons de modèles que l'on peut considérer dans la situation de ce paragraphe :
- Tous les facteurs, A, B et C, sont à effets fixes.
- Le facteur A est à effets aléatoires et de ce fait les facteurs B et C, qui sont emboîtés dans A, sont **nécessairement** à effets aléatoires.
- L'un des facteurs B ou C est à effets aléatoires et de ce fait l'interaction entre B et C est **nécessairement** à effets aléatoires.
- Tous les facteurs sont à effets aléatoires.

Une situation plus complexe
Il s'agit du cas où l'on le troisième facteur C est emboîté dans le facteur B, qui lui-même est croisé avec le facteur A.

$$Y_{i,j,k,l} = \mu + \alpha_i + \beta_j + \gamma_{k|j} + (\alpha\beta)_{i,j} + (\alpha\gamma)_{i,k|j} + \epsilon_{i,j,k,l}$$

avec les hypothèses de rigueurs associées au fait que l'on considère que les facteurs sont aléatoires ou fixes, les conditions habituelles sur les lois, les variances et l'indépendance des erreurs et $1 \leqslant i \leqslant I$, $1 \leqslant j \leqslant J$, $1 \leqslant k \leqslant K$, $1 \leqslant l \leqslant L$. Dans le cas, on peut avoir le

nombre de répétitions, L, égal à un, c'est-à-dire l'on peut utiliser un plan d'expérience ne comportant pas de répétitions.

Pour illustrer les deux règles exposées à la fin de la section 8.1 concernant les facteurs à effets fixes ou aléatoires, on détaille les différentes déclinaisons de modèles que l'on peut considérer dans la situation de ce paragraphe :
- Tous les facteurs, A, B et C, sont à effets fixes.
- Si le facteur B est à effets aléatoires, le facteur C l'est aussi **nécessairement** puisque ce facteur est emboîté dans le facteur B. Il en va de même de l'interaction entre le facteur A et le facteur B.
- Le facteur A est à effets aléatoires. **Nécessairement** ses interactions avec les facteurs B et C sont à effets aléatoires.
- Si le facteur C est à effets aléatoires, son interaction avec le facteur A est **nécessairement** à effets aléatoires.

9. Analyse de la variance à l facteurs croisés, partiellement ou totalement emboîtés, $l \geqslant 3$

On donne dans cette section quelques règles générales qui étendent les indications qui ont été exposées, aux sections 3, 4, 5, 6, 7, 8, dans les cas où l'on a deux ou trois facteurs et dont le domaine de validité est en fait celui de l'analyse de la variance à l facteurs.

1. Si le plan d'expérience ne comporte pas de répétitions, il sera impossible d'estimer l'interaction d'ordre $l-1$[10] entre les facteurs mais toutes les interactions d'ordre inférieur ou égal à $l-2$ pourront être estimées et donc la significativité de leur influence pourra être testée si l'on fait justement l'hypothèse que l'interaction d'ordre $l-1$ est négligeable.
2. Si le plan comporte des répétitions, on peut utiliser un modèle incluant jusqu'à l'interaction d'ordre l entre les facteurs.
3. Si le plan d'expérience ne comporte pas de répétitions, il sera impossible d'utiliser un modèle totalement emboîté. On pourra néanmoins se servir de certains types de modèles partiellement emboîtés.
4. Si le plan comporte des répétitions, on peut utiliser un modèle totalement emboîté.

[10] Il existe un test, le test de Tukey, permettant de s'intéresser à la question de la significativité de l'interaction d'ordre $l-1$ lorsque le plan comporte l facteurs sans répétition. Lorsque qu'une telle interaction n'est pas négligeable, on doir recourir à une transformation afin d'essayer de limiter son effet. En effet, les statistiques des tests dont l'on se sert sont exactes sous l'hypothèse que l'interaction d'ordre $l-1$ est nulle. Pour un exposé de ce test, on pourra se référer au livre [4].

5. On peut comme dans les sections 3, 4, 6 utiliser des modèles dont certains des facteurs sont à effets aléatoires et d'autres à effets fixes.
6. Lorsque les facteurs dont on cherche à tester l'influence sont en présence d'un facteur à effets aléatoires, il faudra peut-être utiliser l'approximation de Satterthwaite. Il en va de même pour certaines interactions. Les tests associés ne seront pas basés sur une statistique qui suit exactement une loi de Fisher.
7. Toute interaction mettant en jeu un facteur à effets aléatoires est **nécessairement** considérée comme un terme du modèle à effets aléatoires.
8. Dans le cas où l'on souhaite utiliser un modèle totalement emboîté ou partiellement emboîté, tout terme du modèle emboîté dans un facteur à effets aléatoires est **nécessairement** considéré comme un terme à effets aléatoires.
9. Il existe une très grande variété de modèles partiellement emboîtés, d'autant plus importante que le nombre de facteurs du modèle, l, est élevé. On rappelle qu'il est commode d'utiliser une représentation graphique pour améliorer la compréhension des modèles partiellement emboîtés. Pour plus de détails sur ce type de schéma, on consultera [4] pour plus de détails. Par exemple il existe 48 modèles differents partiellement emboîtés lorsque l'on considère un modèle d'analyse de la variance à quatre facteurs et en présence de répétitions. On pourra consulter [4] pour plus de détails.

10. Comparaisons multiples

10.1. Contrastes

10.1.1. Définition

Pour introduire la notion de contraste, on considère le cas d'un modèle d'analyse de la variance pour un facteur A à effets fixes. On note A_i, pour $1 \leqslant i \leqslant I$ les modalités contrôlées du facteur A et α_i les effets de ces différentes modalités, toujours pour $1 \leqslant i \leqslant I$.

On introduit le modèle :

$$Y_{i,j} = \mu + \alpha_i + \epsilon_{i,j}, \quad i = 1 \dots I,\ j = 1 \dots J,$$
$$\text{avec la contrainte supplémentaire } \sum_{i=1}^{I} \alpha_i = 0,$$

où $Y_{i,j}$ est la valeur prise par la réponse Y dans la condition A_i lors de la $j-$ème répétition. On postule les hypothèses classiques suivantes pour les erreurs :

$$\forall\ (i,j), 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ \mathcal{L}(\epsilon_{i,j}) = \mathcal{N}(0, \sigma^2),$$
$$\mathrm{Cov}(\epsilon_{i,j}, \epsilon_{k,l}) = 0 \text{ si } (i,j) \neq (k,l) \text{ avec } 1 \leqslant i,k \leqslant I, \text{ et } 1 \leqslant j,l \leqslant J.$$

On appelle **contraste** L des I moyennes $\mu_1, \ldots, \mu_I$ la somme :

$$L = l_1\mu_1 + l_2\mu_2 + \ldots + l_I\mu_I,$$

où $l_1, \ldots, l_I$ sont I nombres réels tels que $\sum_i^I l_i = 0$ et $\mu_1 = \mu + \alpha_1, \ldots, \mu_I = \mu + \alpha_I$ sont tels que $\alpha_1, \ldots, \alpha_I$ sont les I différents effets des I niveaux du facteur A.

Les expressions suivantes sont des exemples de contrastes :

$$\mu_1 - \mu_3, \text{un tel contraste permettra par exemple de comparer } \mu_1 \text{ et } \mu_3.$$
$$\mu_1 - 2\mu_2 + \mu_3, \text{un tel contraste permettra par exemple de comparer } \mu_1 + \mu_3 \text{ et } 2\mu_2.$$

10.1.2. Orthogonalité

On considère deux contrastes L_1 et L_2 :

$$L_1 = l_1\mu_1 + l_2\mu_2 + \ldots + l_I\mu_I$$

et

$$L_2 = l_1^{'}\mu_1 + l_2^{'}\mu_2 + \ldots + l_I^{'}\mu_I.$$

On a donc les relations $l_1, \ldots, l_I$ sont I nombres réels tels que $\sum_i^I l_i = 0$ et $l_1^{'}, \ldots, l_I^{'}$ sont I nombres réels tels que $\sum_i^I l_i^{'} = 0$, $\mu_1 = \mu + \alpha_1, \ldots, \mu_I = \mu + \alpha_I$ sont tels que $\alpha_1, \ldots, \alpha_I$ sont les I différents effets des I niveaux du facteur A.

L_1 et L_2 sont des contrastes dits **orthogonaux** si la relation suivante est vérifiée :

$$l_1l_1^{'} + l_2l_2^{'} + \ldots + l_Il_I^{'} = 0.$$

Par exemple, les deux contrastes suivant sont des contrastes orthogonaux :

$$L_1 = \mu_1 - \mu_2$$
$$L_2 = \mu_1 + \mu_2 - \mu_3 - \mu_4.$$

10.1.3. Estimation

Soit L un contraste. Un estimateur sans biais $\widehat{L}$ de L est obtenu de la manière suivante :

$$\widehat{L} = l_1\widehat{\mu_1} + l_2\widehat{\mu_2} + \ldots + l_I\widehat{\mu_I},$$

où $\widehat{\mu_i} = \widehat{\mu + \alpha_i} = \widehat{\mu} + \widehat{\alpha_i}$, avec $1 \leqslant i \leqslant I$.

La somme des carrés associée à l'estimateur $\widehat{L}$ du contraste L est définie par :

$$SC_{\widehat{L}} = J\frac{\left(\sum_{i=1}^{I} l_i\widehat{\mu}_i\right)^2}{\sum_{i=1}^{I} l_i^2}.$$

Si le nombre de répétitions diffère d'un niveau A_i du facteur A à l'autre, en notant n_i le nombre de répétitions pour la modalité A_i, on obtient la formule suivante :

$$SC_{\widehat{L}} = \frac{\left(\sum_{i=1}^{I} l_i\widehat{\mu}_i\right)^2}{\sum_{i=1}^{I} \frac{l_i^2}{n_i}}.$$

10.1.4. Test d'une hypothèse impliquant un constraste

Pour le modèle considéré de l'analyse de la variance à un facteur, les estimateurs $Y_{i,\bullet}$ des μ_i sont indépendants, même si le plan n'est pas équilibré. Ainsi la variance de l'estimateur $\widehat{L}$ du contraste L est égale à :

$$\begin{aligned}\mathrm{Var}\left[\widehat{L}\right] &= \sum_{i=1}^{I}\left(l_i^2\mathrm{Var}\left[Y_{i,\bullet}\right]\right)\\ &= \sigma^2\sum_{i=1}^{I}\frac{l_i^2}{n_i}.\end{aligned}$$

Un estimateur sans biais de cette variance est alors :

$$\widehat{\mathrm{Var}\left[\widehat{L}\right]} = s_R^2\sum_{i=1}^{I}\frac{l_i^2}{n_i}.$$

Les hypothèses du modèle utilisé impliquent alors que, puisque $\widehat{L}$ est une combinaison linéaire de variables aléatoires qui suivent une loi normale et qui sont indépendantes, l'estimateur $\widehat{L}$ suit aussi une loi normale. De ce fait :

$$\frac{\widehat{L} - L}{\sqrt{\widehat{\mathrm{Var}\left[\widehat{L}\right]}}} \sim t_{n-I},$$

où t_{n-I} est la loi de Student à $n - I$ degrés de liberté.

Remarque :

$\sqrt{\widehat{\mathrm{Var}\left[\widehat{L}\right]}}$ n'est pas un estimateur sans biais de $\sqrt{\mathrm{Var}\left[\widehat{L}\right]}$. Il est possible de corriger le biais de cet estimateur en posant $\widehat{\sqrt{\mathrm{Var}\left[\widehat{L}\right]}}_{\text{corrigé}}$ défini par :

$$\widehat{\sqrt{\mathrm{Var}\left[\widehat{L}\right]}}_{\text{corrigé}} = \sqrt{\frac{n-1}{2}}\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\sqrt{\widehat{\mathrm{Var}\left[\widehat{L}\right]}},$$

où Γ est une fonction classique en mathématiques définie à l'aide d'une intégrale.

Néanmoins l'obtention d'un intervalle de confiance reste délicate puisque la loi de cet estimateur n'est pas classique. P. Chapouille, dans son livre [1], propose la formule approximative :

$$\widehat{\sqrt{\mathrm{Var}\left[\widehat{L}\right]}}_{\text{corrigé}} \approx \sqrt{\widehat{\mathrm{Var}\left[\widehat{L}\right]}}\sqrt{\frac{2n-2}{2n-3}} \approx \sqrt{\widehat{\mathrm{Var}\left[\widehat{L}\right]}}\left(1+\frac{1}{4n-6}\right).$$

On perd alors la propriété ci-dessous qui permet de connaître la loi de la statistique

$$\frac{L-\widehat{L}}{\sqrt{\widehat{\mathrm{Var}\left[\widehat{L}\right]}}},$$

c'est pourquoi l'on préfère utiliser l'estimateur biaisé $\sqrt{\widehat{\mathrm{Var}\left[\widehat{L}\right]}}$ à la place de l'estimateur sans biais $\widehat{\sqrt{\mathrm{Var}\left[\widehat{L}\right]}}_{\text{corrigé}}$ lorsque l'on cherche à obtenir une estimation par intervalle de $\sqrt{\mathrm{Var}\left[\widehat{L}\right]}$ et à la rigueur on se servirait de $\widehat{\sqrt{\mathrm{Var}\left[\widehat{L}\right]}}_{\text{corrigé}}$ pour obtenir une estimation ponctuelle de $\sqrt{\mathrm{Var}\left[\widehat{L}\right]}$ avec la réserve qu'une estimation par intervalle est beaucoup plus intéressante qu'une simple estimation ponctuelle d'un paramètre et lui doit être systématiquement préférée lorsque c'est possible.

Le résultat ci-dessus sur la loi de $\dfrac{L-\widehat{L}}{\sqrt{\widehat{\mathrm{Var}\left[\widehat{L}\right]}}}$ permet de déterminer un intervalle de confiance de niveau $100(1-\alpha)$ % pour la valeur du contraste L.

On se donne désormais un nombre réel L_0 et l'on souhaite tester l'hypothèse

$$\boxed{\mathcal{H}_0 : L = L_0}$$

contre

$$\boxed{\mathcal{H}_1 : L \neq L_0.}$$

Sous l'hypothèse nulle $\mathcal{H}_0$ et les hypothèses du modèle,

$$l = \frac{\widehat{L}(\boldsymbol{y}) - L_0}{\sqrt{s_R^2 \sum_{i=1}^{I} \frac{l_i^2}{n_i}}}$$

est une réalisation d'une variable aléatoire suivant une loi de Student à $n - I$ degrés de liberté. En comparant la valeur l calculée à l'aide de l'échantillon à la valeur critique au seuil α pour une loi de Student à $n-I$ degrés de liberté on peut décider de la significativité du test. Certains logiciels peuvent vous fournir directement la $p-$valeur associée au test d'un contraste ce qui permet également de conclure quant à la significativité du test.

10.2. Comparaisons multiples sous l'hypothèse d'homoscédasticité

Dans le paragraphe 10.1 ci-dessus, on s'est intéressé au test **d'une seule hypothèse** concernant les effets des niveaux du facteur A. Le contexte des tests de comparaisons multiples[11] est radicalement différent puisque l'on cherche à comparer tous les effets des niveaux A_i du facteur A entre eux ou avec un niveau de référence dit de contrôle. On doit ainsi réaliser $I(I-1)/2$ comparaisons dans la première situation ou $I-1$ comparaisons dans la seconde situation où l'on compare les effets à un niveau de contrôle.

Tester l'égalité des effets de deux niveaux A_i et A_j, $i \neq j$, d'un facteur A revient à tester la nullité du contraste $L = \mu_i - \mu_j$. On expose dans la suite les procédures de test simultané de plusieurs contrastes en gardant à l'esprit que l'on appliquera principalement les résultats au cas où ces contrastes sont des différences de moyennes.

On rappelle que l'on n'utilisera l'un des tests de comparaisons multiples que si le facteur étudié est à effets fixes et que l'on a rejeté l'hypothèse nulle d'absence d'effet de ce facteur sur la réponse.
On expose ici la théorie des comparaisons multiples pour le cas d'un modèle à un facteur à effets fixes. De manière plus générale, il est **possible** de comparer les effets des différents niveaux d'un facteur si ceux-ci **sont à effets fixes**. Il n'est généralement **intéressant** de comparer les effets des différents niveaux d'un facteur que si aucun des termes d'interaction mettant en jeu ce facteur n'a un effet significatif au seuil α.

[11] Ces tests sont souvent appelés *post-hoc tests* ou *multiple comparisons tests* en anglais.

10.2.1. La méthode de Tukey

Cette méthode n'est valable que si le nombre de répétitions J_i d'une modalité à l'autre du facteur A est constant. Ce nombre commun de répétitions est alors noté J. Pour une version de la méthode de Tukey adaptée au cas où le plan n'est pas équilibré voir le paragraphe 10.2.2 sur la méthode de Tukey-Kramer.

Si L est un contraste dont un estimateur est $\widehat{L}$, alors un intervalle de confiance de niveau simultané $100(1-\alpha)$ % pour tous les contrastes considérés est donné par la formule suivante :

$$\widehat{L}(\boldsymbol{y}) - T\sqrt{\frac{s_R^2}{J}}\left(\frac{1}{2}\sum_{i=1}^{I}|l_i|\right) < L < \widehat{L}(\boldsymbol{y}) + T\sqrt{\frac{s_R^2}{J}}\left(\frac{1}{2}\sum_{i=1}^{I}|l_i|\right),$$

où $T = q(I, I(J-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de l'étendue Studentisée à I et $I(J-1)$ degrés de liberté. Si l'intervalle de confiance obtenu contient la valeur 0, on décide que le contraste n'est pas significativement différent de 0 au seuil α. Au contraire si l'intervalle de confiance ne contient pas 0, alors on décide que le contraste est significativement différent de 0 au seuil α.

L'intérêt de cette procédure est que, si l'on fixe α, les intervalles définis ci-dessus sont valables **simultanément** pour tous les contrastes qu'il est possible de construire !

On souhaite tester l'hypothèse

$$\boxed{\mathcal{H}_0 : L = 0}$$

contre

$$\boxed{\mathcal{H}_1 : L \neq 0.}$$

Le test est significatif au seuil α et l'on décide de rejeter l'hypothèse nulle $\mathcal{H}_0$ « $L = 0$ » en faveur de l'hypothèse alternative $\mathcal{H}_1$ « $L \neq 0$ » si :

$$\frac{\left|\widehat{L}(\boldsymbol{y})\right|}{T\sqrt{\frac{s_R^2}{J}}\left(\frac{1}{2}\sum_{i=1}^{I}|l_i|\right)} \geqslant q(I, I(J-1); 1-\alpha).$$

Le test n'est pas significatif au seuil α et l'on décide de conserver par défaut l'hypothèse nulle $\mathcal{H}_0$ « $L = 0$ » si :

$$\frac{\left|\widehat{L}(\boldsymbol{y})\right|}{T\sqrt{\frac{s_R^2}{J}}\left(\frac{1}{2}\sum_{i=1}^{I}|l_i|\right)} < q(I, I(J-1); 1-\alpha).$$

Appliqué au contexte des comparaisons multiples[12], l'intervalle de confiance ci-dessus se transforme de la manière suivante puisque les contrastes étudiés sont du type $L = \mu_i - \mu_j$, $i \neq j$:

$$\frac{1}{2}\sum_{i=1}^{I} |l_i| = \frac{1}{2}(|1| + |-1|) = \frac{1}{2}(1+1) = 1.$$

Ainsi, dans le cas des comparaisons multiples, les intervalles de confiance ci-dessus se simplifient en :

$$\widehat{\mu_i}(\boldsymbol{y}) - \widehat{\mu_{i'}}(\boldsymbol{y}) - T\sqrt{\frac{s_R^2}{J}} < \mu_i - \mu_{i'} < \widehat{\mu_i}(\boldsymbol{y}) - \widehat{\mu_{i'}}(\boldsymbol{y}) + T\sqrt{\frac{s_R^2}{J}}.$$

Les hypothèses et les statistiques des tests se transforment, quant à elles, en :

$$\boxed{\mathcal{H}_0 : \mu_i = \mu_{i'}}$$
contre
$$\boxed{\mathcal{H}_1 : \mu_i \neq \mu_{i'}.}$$

On rappelle que ce jeu d'hypothèses est équivalent à celui-ci :

$$\boxed{\mathcal{H}_0 : \alpha_i = \alpha_{i'}}$$
contre
$$\boxed{\mathcal{H}_1 : \alpha_i \neq \alpha_{i'}.}$$

Le test est significatif au seuil α et l'on décide de rejeter l'hypothèse nulle $\mathcal{H}_0$ « $\mu_i = \mu_{i'}$ » en faveur de l'hypothèse alternative $\mathcal{H}_1$ « $\mu_i \neq \mu_{i'}$ » si :

$$\frac{|\widehat{\mu_i}(\boldsymbol{y}) - \widehat{\mu_{i'}}(\boldsymbol{y})|}{\sqrt{\frac{s_R^2}{J}}} \geqslant q(I, I(J-1); 1-\alpha).$$

Le test n'est pas significatif au seuil α et l'on décide de conserver par défaut l'hypothèse nulle $\mathcal{H}_0$ « $\mu_i = \mu_{i'}$ » si :

$$\frac{|\widehat{\mu_i}(\boldsymbol{y}) - \widehat{\mu_{i'}}(\boldsymbol{y})|}{\sqrt{\frac{s_R^2}{J}}} < q(I, I(J-1); 1-\alpha).$$

En utilisant ces intervalles de confiance pour décider **simultanément** de la significativité des $I(I-1)/2$ différences entre les effets des modalités du facteur A, **on est garanti** que la probabilité qu'aucune des différences n'est significative est exactement de $1-\alpha$.

[12]Cette procédure est souvent appelée Tukey's HSD, pour *Tukey's Honestly Significance Difference.*

10.2.2. La méthode de Tukey-Kramer

Il s'agit d'une adaptation de la méthode de Tukey au cas où le plan expérimental n'est pas équilibré. Si l'on veut comparer les deux moyennes μ_i et $\mu_{i'}$, on remplace simplement la valeur J correspondant au nombre total constant d'essais réalisés dans des conditions les modalités A_i du facteur A par la moyenne harmonique[13] du nombre de répétitions effectuées dans la modalité A_i et dans la modalité $A_{i'}$.

Les intervalles de confiance ci-dessus se modifient en conséquence :

$$\widehat{\mu_i}(\boldsymbol{y}) - \widehat{\mu_{i'}}(\boldsymbol{y}) - T\sqrt{\frac{s_R^2}{2}\left(\frac{1}{n_i}+\frac{1}{n_i'}\right)} < \mu_i - \mu_{i'} < \widehat{\mu_i}(\boldsymbol{y}) - \widehat{\mu_{i'}}(\boldsymbol{y}) + T\sqrt{\frac{s_R^2}{2}\left(\frac{1}{n_i}+\frac{1}{n_i'}\right)},$$

où $T = q(I, n-I; 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de l'étendue Studentisée à I et $n-I$ degrés de liberté.
Les hypothèses sont toujours :

$$\boxed{\mathcal{H}_0 : \mu_i = \mu_{i'}}$$
contre
$$\boxed{\mathcal{H}_1 : \mu_i \neq \mu_{i'}.}$$

On rappelle que ce jeu d'hypothèses est équivalent à celui-ci :

$$\boxed{\mathcal{H}_0 : \alpha_i = \alpha_{i'}}$$
contre
$$\boxed{\mathcal{H}_1 : \alpha_i \neq \alpha_{i'}.}$$

Le test est significatif au seuil α et l'on décide de rejeter l'hypothèse nulle $\mathcal{H}_0$ « $\mu_i = \mu_{i'}$ » en faveur de l'hypothèse alternative $\mathcal{H}_1$ « $\mu_i \neq \mu_{i'}$ » si :

$$\frac{|\widehat{\mu_i}(\boldsymbol{y}) - \widehat{\mu_{i'}}(\boldsymbol{y})|}{\sqrt{\frac{s_R^2}{2}\left(\frac{1}{n_i}+\frac{1}{n_i'}\right)}} \geqslant q(I, n-I; 1-\alpha).$$

Le test n'est pas significatif au seuil α et l'on décide de conserver par défaut l'hypothèse nulle $\mathcal{H}_0$ « $\mu_i = \mu_{i'}$ » si :

$$\frac{|\widehat{\mu_i}(\boldsymbol{y}) - \widehat{\mu_{i'}}(\boldsymbol{y})|}{\sqrt{\frac{s_R^2}{2}\left(\frac{1}{n_i}+\frac{1}{n_i'}\right)}} < q(I, n-I; 1-\alpha).$$

[13] On rappelle que la moyenne harmonique $\mathrm{Harm}(a,b)$ de deux nombres réels strictement positifs a et b est définie par :

$$\mathrm{Harm}(a,b) = \frac{1}{\frac{1}{2}\left(\frac{1}{a}+\frac{1}{b}\right)}$$

.

10.2.3. La méthode de Hochberg

Dans le cadre des comparaisons deux à deux de moyennes, Hochberg a proposé la modification suivante de la méthode de Tukey-Kramer, exposée au paragraphe 10.2.2 :
Les intervalles de confiance ci-dessus sont modifiés en :

$$\widehat{\mu_i}(\boldsymbol{y}) - \widehat{\mu_{i'}}(\boldsymbol{y}) - m\sqrt{s_R^2\left(\frac{1}{n_i}+\frac{1}{n_i'}\right)} < \mu_i - \mu_{i'} < \widehat{\mu_i}(\boldsymbol{y}) - \widehat{\mu_{i'}}(\boldsymbol{y}) + m\sqrt{s_R^2\left(\frac{1}{n_i}+\frac{1}{n_i'}\right)}.$$

où $m = m(I, n-I; 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi du maximum du module Studentisé à I et $n-I$ degrés de liberté.
Les hypothèses sont toujours :

$$\boxed{\mathcal{H}_0 : \mu_i = \mu_{i'}}$$
contre
$$\boxed{\mathcal{H}_1 : \mu_i \neq \mu_{i'}.}$$

On rappelle que ce jeu d'hypothèses est équivalent à celui-ci :

$$\boxed{\mathcal{H}_0 : \alpha_i = \alpha_{i'}}$$
contre
$$\boxed{\mathcal{H}_1 : \alpha_i \neq \alpha_{i'}.}$$

Le test est significatif au seuil α et l'on décide de rejeter l'hypothèse nulle $\mathcal{H}_0$ « $\mu_i = \mu_{i'}$ » en faveur de l'hypothèse alternative $\mathcal{H}_1$ « $\mu_i \neq \mu_{i'}$ » si :

$$\frac{|\widehat{\mu_i}(\boldsymbol{y}) - \widehat{\mu_{i'}}(\boldsymbol{y})|}{\sqrt{\frac{s_R^2}{2}\left(\frac{1}{n_i}+\frac{1}{n_i'}\right)}} \geqslant m(I, n-I; 1-\alpha).$$

Le test n'est pas significatif au seuil α et l'on décide de conserver par défaut l'hypothèse nulle $\mathcal{H}_0$ « $\mu_i = \mu_{i'}$ » si :

$$\frac{|\widehat{\mu_i}(\boldsymbol{y}) - \widehat{\mu_{i'}}(\boldsymbol{y})|}{\sqrt{\frac{s_R^2}{2}\left(\frac{1}{n_i}+\frac{1}{n_i'}\right)}} < m(I, n-I; 1-\alpha).$$

10.2.4. La méthode de Scheffé

Contrairement à la méthode de Tukey exposée au paragraphe 10.2.1, la méthode de Scheffé est valide si le plan est déséquilibré. Soit n_i le nombre de répétitions effectuées pour la modalité A_i du facteur A.

Si L est un contraste quelconque estimé par $\widehat{L}$ alors un intervalle de confiance de niveau $100(1-\alpha)$ % est donné par la formule suivante :

$$\begin{aligned}\widehat{L}(\boldsymbol{y}) - S\sqrt{(I-1)s_R^2\left(\sum_{i=1}^{I}\frac{l_i^2}{n_i}\right)} &< L \\ &\text{et} \\ L &< \widehat{L}(\boldsymbol{y}) + S\sqrt{(I-1)s_R^2\left(\sum_{i=1}^{I}\frac{l_i^2}{n_i}\right)},\end{aligned}$$

où $S^2 = F(I-1, n-I; 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $I-1$ et $n-I$ degrés de liberté. Si l'intervalle de confiance obtenu contient la valeur 0, on décide que le contraste n'est pas significativement différent de 0 au seuil α. Au contraire si l'intervalle de confiance ne contient pas 0, alors on décide que le contraste est significativement différent de 0 au seuil α.

L'intérêt de cette procédure est que, si l'on fixe α, les intervalles définis ci-dessus sont valables **simultanément** pour tous les contrastes qu'il est possible de construire !

On souhaite tester l'hypothèse

$$\boxed{\mathcal{H}_0 : L = 0}$$

contre

$$\boxed{\mathcal{H}_1 : L \neq 0.}$$

Le test est significatif au seuil α et l'on décide de rejeter l'hypothèse nulle $\mathcal{H}_0$ « $L = 0$ » en faveur de l'hypothèse alternative $\mathcal{H}_1$ « $L \neq 0$ » si :

$$\frac{\widehat{L}(\boldsymbol{y})}{\sqrt{(I-1)s_R^2\left(\sum_{i=1}^{I}\frac{l_i^2}{n_i}\right)}} \geqslant \sqrt{F(I-1, n-I; 1-\alpha)}.$$

Le test n'est pas significatif au seuil α et l'on décide de conserver par défaut l'hypothèse nulle $\mathcal{H}_0$ « $L = 0$ » si :

$$\frac{\widehat{L}(\boldsymbol{y})}{\sqrt{(I-1)s_R^2\left(\sum_{i=1}^{I}\frac{l_i^2}{n_i}\right)}} < \sqrt{F(I-1, n-I; 1-\alpha)}.$$

Appliqué au contexte des comparaisons multiples, l'intervalle de confiance ci-dessus se transforme de la manière suivante, puisque les contrastes étudiés sont du type $L = \mu_i - \mu_j$,

$i \neq j$:

$$\sum_{i=1}^{I} \frac{l_i^2}{n_i} = \frac{1}{n_i} + \frac{(-1)^2}{n_{i'}} = \frac{1}{n_i} + \frac{1}{n_{i'}}.$$

Ainsi, dans le cas des comparaisons mulitples, les intervalles de confiance ci-dessus se simplifient en :

$$\begin{aligned} \widehat{\mu_i}(\boldsymbol{y}) - \widehat{\mu_{i'}}(\boldsymbol{y}) - S\sqrt{(I-1)s_R^2\left(\frac{1}{n_i}+\frac{1}{n_{i'}}\right)} &< \mu_i - \mu_{i'} \\ &\text{et} \\ \mu_i - \mu_{i'} &< \widehat{\mu_i}(\boldsymbol{y}) - \widehat{\mu_{i'}}(\boldsymbol{y}) + S\sqrt{(I-1)s_R^2\left(\frac{1}{n_i}+\frac{1}{n_{i'}}\right)}. \end{aligned}$$

Les hypothèses et les statistiques des tests se transforment, quant à elles, en :

$$\boxed{\mathcal{H}_0 : \mu_i = \mu_{i'}}$$

contre

$$\boxed{\mathcal{H}_1 : \mu_i \neq \mu_{i'}.}$$

On rappelle que ce jeu d'hypothèses est équivalent à celui-ci :

$$\boxed{\mathcal{H}_0 : \alpha_i = \alpha_{i'}}$$

contre

$$\boxed{\mathcal{H}_1 : \alpha_i \neq \alpha_{i'}.}$$

Le test est significatif au seuil α et l'on décide de rejeter l'hypothèse nulle $\mathcal{H}_0$ « $\mu_i = \mu_{i'}$ » en faveur de l'hypothèse alternative $\mathcal{H}_1$ « $\mu_i \neq \mu_{i'}$ » si :

$$\frac{|\widehat{\mu_i}(\boldsymbol{y}) - \widehat{\mu_{i'}}(\boldsymbol{y})|}{\sqrt{(I-1)s_R^2\left(\frac{1}{n_i}+\frac{1}{n_{i'}}\right)}} \geqslant \sqrt{F(I-1, n-I; 1-\alpha)}.$$

Le test n'est pas significatif au seuil α et l'on décide de conserver par défaut l'hypothèse nulle $\mathcal{H}_0$ « $\mu_i = \mu_{i'}$ » si :

$$\frac{|\widehat{\mu_i}(\boldsymbol{y}) - \widehat{\mu_{i'}}(\boldsymbol{y})|}{\sqrt{(I-1)s_R^2\left(\frac{1}{n_i}+\frac{1}{n_{i'}}\right)}} < \sqrt{F(I-1, n-I; 1-\alpha)}.$$

En utilisant ces intervalles de confiance pour décider **simultanément** de la significativité des $I(I-1)/2$ différences entre les effets des modalités A_i du facteur A, **on est garanti**

que la probabilité qu'aucune des différences n'est significative est exactment de $1-\alpha$.

Dans le cas où le plan est équilibré on obtient l'intervalle de confiance de niveau $100(1-\alpha)$ suivant :

$$\widehat{\mu_i}(\boldsymbol{y}) - \widehat{\mu_{i'}}(\boldsymbol{y}) - S\sqrt{\frac{2(I-1)s_R^2}{J}} < \mu_i - \mu_{i'} < \widehat{\mu_i}(\boldsymbol{y}) - \widehat{\mu_{i'}}(\boldsymbol{y}) + S\sqrt{\frac{2(I-1)s_R^2}{J}}.$$

10.2.5. La méthode de la plus petite différence significative de Fisher

La méthode de la plus petite différence significative de Fisher[14] est une méthode jugée par beaucoup de statisticiens comme inadéquate, voir par exemple [9], puisqu'elle ne tient pas compte du contexte même des comparaisons multiples c'est-à-dire du fait que l'on réalise $I(I-1)/2$ comparaisons qui ne sont pas indépendantes. On l'expose néanmoins brièvement ci-dessous. Cette méthode s'applique que le plan soit équilibré ou non.

Un intervalle de confiance au niveau $100(1-\alpha)$ pour la différence de deux moyennes $\mu_i - \mu_{i'}$ est :

$$\widehat{\mu_i}(\boldsymbol{y}) - \widehat{\mu_{i'}}(\boldsymbol{y}) - t\sqrt{s_R^2\left(\frac{1}{n_i}+\frac{1}{n_{i'}}\right)} < \mu_i - \mu_{i'} < \widehat{\mu_i}(\boldsymbol{y}) - \widehat{\mu_{i'}}(\boldsymbol{y}) + t\sqrt{s_R^2\left(\frac{1}{n_i}+\frac{1}{n_{i'}}\right)},$$

où $t = t_{n-I;1-\alpha/2}$ est le $100(1-\alpha/2)$ quantile de la loi de Student à $n-I$ degrés de liberté. Si l'intervalle de confiance obtenu contient la valeur 0, alors on ne peut pas rejeter l'hypothèse que le contraste n'est pas significativement différent de 0 au seuil α. Au contraire si l'intervalle de confiance ne contient pas 0, alors on décide que le contraste est significativement différent de 0 au seuil α.
On souhaite tester l'hypothèse :

$$\boxed{\mathcal{H}_0 : \mu_i = \mu_{i'}}$$

contre

$$\boxed{\mathcal{H}_1 : \mu_i \neq \mu_{i'}.}$$

On rappelle que ce jeu d'hypothèses est équivalent à celui-ci :

$$\boxed{\mathcal{H}_0 : \alpha_i = \alpha_{i'}}$$

contre

$$\boxed{\mathcal{H}_1 : \alpha_i \neq \alpha_{i'}.}$$

[14]En anglais ce test de comparaisons muliples est souvent noté *Fisher's LSD* pour *Fisher's Least Significant Difference.*

Le test est significatif au seuil α et l'on décide de rejeter l'hypothèse nulle $\mathcal{H}_0$ « $\mu_i = \mu_{i'}$ » en faveur de l'hypothèse alternative $\mathcal{H}_1$ « $\mu_i \neq \mu_{i'}$ » si :

$$\frac{|\widehat{\mu_i}(\boldsymbol{y}) - \widehat{\mu_{i'}}(\boldsymbol{y})|}{\sqrt{s_R^2\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)}} \geqslant t_{n-I;1-\alpha/2}.$$

Le test n'est pas significatif au seuil α et l'on décide de conserver par défaut l'hypothèse nulle $\mathcal{H}_0$ « $\mu_i = \mu_{i'}$ » si :

$$\frac{|\widehat{\mu_i}(\boldsymbol{y}) - \widehat{\mu_{i'}}(\boldsymbol{y})|}{\sqrt{s_R^2\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)}} < t_{n-I;1-\alpha/2}.$$

10.2.6. La méthode de Bonferroni

Supposons que l'on souhaite réaliser k comparaisons de moyennes ou k tests de contrastes et prendre une décision conjointe au risque de première espèce α. On note E_i l'évènement associé au rejet de l'absence de différence significative au seuil α_{ind} lors de la $i-$ème comparaison ou le test du $i-$ème contraste. La méthode de Bonferroni est basée sur l'inégalité suivante :

$$\begin{aligned} \alpha &= \mathbb{P}\left[E_1 \cup E_2 \cup \cdots \cup E_k\right] \\ &\leq \mathbb{P}\left[E_1\right] + \mathbb{P}\left[E_2\right] + \cdots + \mathbb{P}\left[E_k\right] \\ &\leq k * \alpha_{\text{ind}}. \end{aligned}$$

Ainsi si l'on veut être sûr que le risque de première espèce α associé globalement à la prise simultanée de toutes les décisions lors des k comparaisons ou des k tests de contrastes est plus petit qu'une valeur α_0 fixée à l'avance, il suffit de choisir :

$$\alpha_{\text{ind}} \leqslant \frac{\alpha_0}{k}.$$

On procède alors à des comparaisons des moyennes deux à deux avec un test t de Student de seuil α_0/k ou à un test de chacun des contrastes avec la statistique exposée au paragraphe 10.1.4 au seuil α_0/k. Cette procédure s'applique donc si le plan est équilibré ou non.

Les intervalles de confiance pour k comparaisons de deux moyennes μ_i et $\mu_{i'}$ de deux groupes d'effectifs n_i et $n_{i'}$ sont :

$$\widehat{\mu_i}(\boldsymbol{y}) - \widehat{\mu_{i'}}(\boldsymbol{y}) - t_B\sqrt{s_R^2\left(\frac{1}{n_i} + \frac{1}{n_i'}\right)} < \mu_i - \mu_{i'} < \widehat{\mu_i}(\boldsymbol{y}) - \widehat{\mu_{i'}}(\boldsymbol{y}) + t_B\sqrt{s_R^2\left(\frac{1}{n_i} + \frac{1}{n_i'}\right)},$$

où $t_B = t(n-I; 1-\dfrac{\alpha}{2k})$ est le $100(1-\dfrac{\alpha}{2k})$ quantile de la loi de Student à $n-I$ degrés de liberté.

Remarque :

Il est conseillé, par Fleiss voir [6], d'utiliser la procédure de Bonferroni dans les cas suivants :

1. Le nombre de comparaisons k n'est pas très élévé, cette procédure étant trop conservative[15] si k est élevé.

2. On préférera la procédure de Bonferroni à celle de Scheffé si le nombre de comparaisons est strictement inférieur à I^2.

3. On préférera la procédure de Bonferroni à celle de Tukey si le nombre de comparaisons est strictement inférieur à $I(I-1)/2$ ou si l'on souhaite tester en plus un petit nombre de comparaisons autres que celles des effets principaux des modalités A_i du facteur A.

4. On préférera la procédure de Bonferroni à celle de Dunnett si l'on souhaite tester en plus d'autres comparaisons que celles des effets principaux des modalités A_i du facteur A au niveau de contrôle A_{i_0} ou si l'on souhaite tester d'autres comparaisons que celles des effets principaux des modalités A_i du facteur A au niveau de contrôle A_{i_0}.

5. Les mêmes recommendations s'appliquent si l'on considère des tests portant sur des contrastes.

10.2.7. La méthode de rejet séquentiel de Bonferroni de Holm

La méthode de rejet séquentiel de Bonferroni de Holm[16] est dérivée de la méthode de Bonferroni exposée au paragraphe 10.2.6. Elle a été introduite pour augmenter la puissance de la méthode de Bonferroni. Elle consiste simplement à modifier le seuil des tests de comparaison de la manière suivante : si l'un des tests de comparaison est significatif au niveau $\alpha_{\text{ind}} = \alpha/k$, alors on effectue le test suivant au niveau $\alpha_{\text{ind}} = \alpha/(k-1)$ et ainsi de suite...

Les conseils d'utilisation sont les mêmes que ceux exposés au paragraphe 10.2.6 et l'on peut aussi bien s'en servir pour des plans équilibrés que des plans déséquilibrés.

10.2.8. La méthode de Dunn-Šidák

La méthode de Dunn-Šidák est basée sur la même idée que celle de Bonferroni, seule l'inégalité utilisée pour déterminer le seuil individuel de chacune des comparaisons, α_{ind}, est modifiée.

[15] On dit qu'une procédure est conservative si son niveau de confiance est supérieur à celui fixé, ici $1-\alpha_0$. La détection de différences significatives étant alors « trop difficile » par rapport au seuil que l'on avait fixé en avance pour le test.

[16] En anglais *Holm's SRB criterion* pour *Holm's Sequentially Rejective Bonferroni criterion.*

Supposons que l'on souhaite réaliser k comparaisons de moyennes ou k tests de contrastes et prendre une décision conjointe au risque de première espèce α. On note E_i l'évènement associé au rejet de l'absence de différence significative au seuil α_{ind} lors de la $i-$ème comparaison ou le test du $i-$ème contraste. La méthode de Dunn-Šidák est basée sur l'inégalité suivante :

$$\alpha \leq 1-(1-\alpha_{\text{ind}})^k.$$

L'inégalité de Bonferroni est plus grossière que celle de Dunn-Šidák et de ce fait la procédure de Bonferroni est plus conservative que celle de Dunn-Šidák. Ainsi si l'on veut être sûr que le risque de première espèce α associé globalement à la prise simultanée de toutes les décisions lors des k comparaisons ou des k tests de contrastes est plus petit qu'une valeur α_0 fixée à l'avance, il suffit de choisir :

$$\alpha_{\text{ind}} \leqslant 1-\sqrt[k]{1-\alpha_0}.$$

On procède alors à des comparaisons des moyennes deux à deux avec un test t de Student de seuil $1-\sqrt[k]{1-\alpha_0}$ ou à un test de chacun des contrastes avec la statistique exposée au paragraphe 10.1.4 au seuil $1-\sqrt[k]{1-\alpha_0}$. Cette procédure s'applique donc si le plan est équilibré ou non.

L'intervalle de confiance de niveau $100(1-\alpha_0)$ % pour $\mu_i-\mu_{i'}$ est alors le suivant :

$$\widehat{\mu_i}(\boldsymbol{y})-\widehat{\mu_{i'}}(\boldsymbol{y})-t_{DS}\sqrt{s_R^2\left(\frac{1}{n_i}+\frac{1}{n_i'}\right)}<\mu_i-\mu_{i'}<\widehat{\mu_i}(\boldsymbol{y})-\widehat{\mu_{i'}}(\boldsymbol{y})+t_{DS}\sqrt{s_R^2\left(\frac{1}{n_i}+\frac{1}{n_i'}\right)},$$

où $t_{DS}=t(n-I;\sqrt[k]{1-\alpha_0})$ est le $100\sqrt[k]{1-\alpha_0}$ quantile de la loi de Student à $n-I$ degrés de liberté.

10.2.9. La méthode de Newman-Keuls

Cette procédure est également connue sous le nom de test de Student-Newman-Keuls et est utilisable si le plan est équilibré ou déséquilibré.

1. On commence par ordonner les estimations des moyennes $\widehat{\mu_1}(\boldsymbol{y}),\ldots,\widehat{\mu_I}(\boldsymbol{y})$ par ordre croissant. On obtient donc $\widehat{\mu_{(1)}}(\boldsymbol{y})\leqslant\ldots\leqslant\widehat{\mu_{(I)}}(\boldsymbol{y})$. On ordonne de la même manière les moyennes $\mu_1,\ldots,\mu_I$.
2. On calcule la quantité $\widehat{\mu_{(I)}}(\boldsymbol{y})-\widehat{\mu_{(1)}}(\boldsymbol{y})$, qui est toujours positive, et on la compare à $W_{(I)}$ où :

$$W_{(I)}=q(I,I(J-1);1-\alpha)\sqrt{\frac{s_R^2}{J}},$$

avec $q(I, I(J-1); 1-\alpha)$ le $100(1-\alpha)$ quantile de la loi de l'étendue Studentisée à I et $I(J-1)$ degrés de liberté.
Si $\widehat{\mu_{(I)}}(\boldsymbol{y}) - \widehat{\mu_{(1)}}(\boldsymbol{y}) \geqslant W_{(I)}$ alors on décide que le groupe de moyennes $\mu_{(1)}, \ldots \mu_{(I)}$ n'est pas homogène et on crée les deux groupes $\mu_{(1)}, \ldots, \mu_{(I-1)}$ et $\mu_{(2)}, \ldots, \mu_{(I)}$ sinon on considère que le groupe de moyennes $\mu_{(1)}, \ldots \mu_{(I)}$ est homogène et la procèdure s'arrête là.

3. On calcule les quantités $\widehat{\mu_{(I-1)}}(\boldsymbol{y}) - \widehat{\mu_{(1)}}(\boldsymbol{y})$ et $\widehat{\mu_{(I)}}(\boldsymbol{y}) - \widehat{\mu_{(2)}}(\boldsymbol{y})$, qui sont toujours positives, et on les compare à $W_{(I-1)}$ où :

$$W_{(I-1)} = q(I-1, I(J-1); 1-\alpha)\sqrt{\frac{s_R^2}{J}},$$

avec $q(I-1, I(n-1); 1-\alpha)$ le $100(1-\alpha)$ quantile de la loi de l'étendue Studentisée à $I-1$ et $I(J-1)$ degrés de liberté.
Si $\widehat{\mu_{(I-1)}}(\boldsymbol{y}) - \widehat{\mu_{(1)}}(\boldsymbol{y}) \geqslant W_{(I)}$ alors on décide que le groupe de moyennes $\mu_{(1)}, \ldots \mu_{(I-1)}$ n'est pas homogène et on crée les deux groupes $\mu_{(1)}, \ldots, \mu_{(I-2)}$ et $\mu_{(2)}, \ldots, \mu_{(I-1)}$ sinon on considère que le groupe de moyennes $\mu_{(1)}, \ldots \mu_{(I-1)}$ est homogène et la procèdure s'arrête là pour ce groupe. On pratique la même analyse pour le groupe $\mu_{(2)}, \ldots, \mu_{(I)}$.

4. On réitère le processus, en utilisant la valeur critique $W_{(k)}$ avec :

$$W_{(k)} = q(k, I(J-1); 1-\alpha)\sqrt{\frac{s_R^2}{J}}$$

avec k le nombre de moyennes que comporte le groupe dont on teste l'homogénéité, jusqu'à ce que tous les groupes obtenus soient homogènes ou que toutes les moyennes aient été comparées entre elles. Dans la pire des configurations on obtiendra I groupes disctincts ce qui prouve que cette procèdure se termine bien.

L'avantage de cette procédure est d'aboutir à la constitution, au risque global de $\alpha = 5$ %, de groupes de moyennes homogènes.
Pour tenir compte d'un éventuel déséquilibre du plan expérimental on remplace, dans les expressions ci-dessus, J par la moyenne harmonique du nombre de répétitions dans les modalités pour lesquelles les moyennes sont la plus petite et la plus grande de l'ensemble considéré. Ainsi, par exemple, on comparera la quantité $\widehat{\mu_{(I)}}(\boldsymbol{y}) - \widehat{\mu_{(1)}}(\boldsymbol{y})$, qui est toujours positive, à $W_{(I)}$ défini par :

$$W_{(I)} = q(I, n-I; 1-\alpha)\sqrt{\frac{s_R^2}{2}\left(\frac{1}{n_{(1)}} + \frac{1}{n_{(I)}}\right)},$$

avec $q(I, n-I; 1-\alpha)$ le $100(1-\alpha)$ quantile de la loi de l'étendue Studentisée à I et $n-I$ degrés de liberté.

10.2.10. La méthode de Duncan

Le test de l'étendue multiple de Duncan[17] est basée sur le même algorithme, c'est-à-dire la même suite de décisions, que celle du test de Newman-Keuls exposée au paragraphe 10.2.9. La seule différence entre ces deux procédures réside dans le fait que les estimations des étendues des groupes de moyennes comportant k moyennes sont comparées à la valeur $D_{(k)}$ où $D_{(k)}$ est définie par :

$$D_{(k)} = R(k, I(J-1); 1-\alpha)\sqrt{\frac{s_R^2}{J}},$$

avec $R(k, I(J-1); 1-\alpha)$ le $100(1-\alpha)$ quantile de la loi de l'étendue multiple de Duncan à k et $I(J-1)$ degrés de liberté, I le nombre de niveaux du facteur A et J le nombre de répétitions commun à toutes les modalités A_i du facteur A.

Pour tenir compte d'un éventuel déséquilibre du plan expérimental on remplace, dans les expressions ci-dessus, J par la moyenne harmonique du nombre de répétitions dans les modalités pour lesquelles les moyennes sont la plus petite et la plus grande de l'ensemble considéré. Ainsi, par exemple si l'on cherche à tester l'homogénéité de k moyennes dont l'estimation la plus petite est pour $\mu_{(i)}$ et l'estimation la plus grande pour $\mu_{(i')}$, on comparera la quantité $\widehat{\mu_{(i)}}(\boldsymbol{y}) - \widehat{\mu_{(i')}}(\boldsymbol{y})$ à $D_{(k)}$ définie par :

$$D_{(k)} = q(k, n-I; 1-\alpha)\sqrt{\frac{s_R^2}{2}\left(\frac{1}{n_{(i)}} + \frac{1}{n_{(i')}}\right)},$$

avec $q(k, n-I; 1-\alpha)$ le $100(1-\alpha)$ quantile de la loi de l'étendue multiple de Duncan à k et $n-I$ degrés de liberté.

10.2.11. La méthode de Dunnett

Le test de Dunnett a été proposé pour le contexte particulier de la comparaison des effets des modalités A_i d'un facteur A avec un niveau de référence de ce facteur, souvent appelé niveau de contrôle et noté A_c. Cette procédure peut être utilisée si le plan est équilibré ou non.

Il s'agit simplement d'une modification de la valeur critique utilisée avec la statistique du test t de Student pour deux moyennes. Ainsi la valeur de la différence des estimations de la moyenne de la réponse μ_i obtenue pour la modalité A_i du facteur A et de la moyenne μ_c de la réponse obtenue pour le niveau de contrôle A_c du facteur A.
On souhaite tester l'hypothèse :

[17] En anglais Duncan's Multiple Range Test.

$$\boxed{\mathcal{H}_0 : \mu_i = \mu_c}$$

contre

$$\boxed{\mathcal{H}_1 : \mu_i \neq \mu_c.}$$

On rappelle que ce jeu d'hypothèses est équivalent à celui-ci :

$$\boxed{\mathcal{H}_0 : \alpha_i = \alpha_c}$$

contre

$$\boxed{\mathcal{H}_1 : \alpha_i \neq \alpha_c.}$$

Le test est significatif au seuil α et l'on décide de rejeter l'hypothèse nulle $\mathcal{H}_0$ « $\mu_i = \mu_c$ » en faveur de l'hypothèse alternative $\mathcal{H}_1$ « $\mu_i \neq \mu_c$ » si :

$$|\widehat{\mu}_i(\boldsymbol{y}) - \widehat{\mu}_c(\boldsymbol{y})| \geqslant D(I-1, n-I; 1-\alpha)\sqrt{s_R^2\left(\frac{1}{n_i} + \frac{1}{n_c}\right)},$$

où $D(I-1, n-I; 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Dunnett à $I-1$ et $n-I$ degrés de liberté. Le test n'est pas significatif au seuil α et l'on décide de conserver par défaut l'hypothèse nulle $\mathcal{H}_0$ « $\mu_i = \mu_c$ » si :

$$|\widehat{\mu}_i(\boldsymbol{y}) - \widehat{\mu}_c(\boldsymbol{y})| < D(I-1, n-I; 1-\alpha)\sqrt{s_R^2\left(\frac{1}{n_i} + \frac{1}{n_c}\right)},$$

où $D(I-1, n-I; 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Dunnett à $I-1$ et $n-I$ degrés de liberté.
Le test doit être préféré aux autres procédures lorsque l'on souhaite seulement comparer les effets des modalités A_i avec l'effet de la modalité de contrôle A_c d'un facteur A. En effet ce test est dans ce cas généralement plus puissant.

10.3. Comparaisons multiples sous l'hypothèse d'hétéroscédasticité

10.3.1. La méthode de Games et Howell

Il s'agit d'une modification des statistiques du test de Tukey-Kramer pour tenir compte de l'hétérogénéité des variances σ_i^2 des niveaux A_i du facteur A ainsi que d'un éventuel déséquilibre du plan. On remplace ainsi S_R^2 par :

$$S_R^2 = \sqrt{\frac{S_i^2}{n_i} + \frac{S_{i'}^2}{n_{i'}}},$$

où

$$S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} \left(Y_{i,j} - Y_{i,\bullet}\right)^2 ,$$

$$S_{i'}^2 = \frac{1}{n_i' - 1} \sum_{j=1}^{n_i'} \left(Y_{i',j} - Y_{i',\bullet}\right)^2 ,$$

sont les estimateurs sans biais des variances au sein des niveaux A_i et $A_{i'}$ du facteur. On désigne par s_i^2 et $s_{i'}^2$ les estimations de σ_i^2 et $\sigma_{i'}^2$ associées aux estimateurs S_i^2 et $S_{i'}^2$. Cette modification permet donc de tenir compte des éventuelles différences entre les variances σ_i^2 et $\sigma_{i'}^2$ ou entre le nombre de répétitions n_i et $n_{i'}$ d'un niveau du facteur A à l'autre.

On modifie également un des degrés de liberté de la loi de l'étendue Studentisée dont on se sert pour trouver les quantiles nécessaires au test. Plus précisement, on utilise $q(I, \nu_{i,i'}; 1-\alpha)$ avec :

$$\nu_{i,i'} = \frac{\left(\dfrac{s_i^2}{n_i} + \dfrac{s_{i'}^2}{n_{i'}}\right)^2}{\dfrac{\left(\dfrac{s_i^2}{n_i}\right)^2}{n_i - 1} + \dfrac{\left(\dfrac{s_{i'}^2}{n_{i'}}\right)^2}{n_{i'} - 1}}.$$

10.4. Deux méthodes recommandées

Les procédures de Newman-Keuls et de Duncan sont toutes deux basées sur des méthodes de tests multiples de l'étendue séquentiels. On a montré que ces procédures sont statistiquement invalides. Les deux méthodes suivantes sont valides et recommandées par l'encyclopédie de statistique, pour plus de détails voir [7].

10.4.1. Le test multiple des étendues de Ryan-Einot-Gabriel-Welsch

On souhaite utiliser des tests multiples de l'étendue, on utilise les valeurs suivantes pour les seuils :

$$\alpha_m = \begin{cases} 1 - (1-\alpha)^{\frac{m}{I}} & \text{si} \quad m = 2, \ldots, I-2, \\ \alpha & \text{si} \quad m = I-1, I. \end{cases} \tag{10.1}$$

La validité du test multiple des étendues de Ryan-Einot-Gabriel-Welsch est basée sur la validité du procédé plus général de comparaisons descendant décrit ci-dessous.

Pour tout sous-ensemble $\mathbb{K}$ inclus dans $\mathbb{I} = \{1, \ldots, I\}$, soit $\mathcal{H}_{\mathbb{K}}$ l'hypothèse :

$$\boxed{\mathcal{H}_{\mathbb{K}} : \mu_i = \mu_j \text{ pour tout } i, j \text{ de } \mathbb{K},}$$

et soit $|\mathbb{K}|$ le nombre d'éléments de $\mathbb{K}$. On ne considèrera dans ce paragraphe que des sous-ensembles de $\mathbb{I}$ formés d'entiers consécutifs.

1. Tester au seuil $\alpha = \alpha_{|\mathbb{I}|} = \alpha_I$, donné par l'équation (10.1), $\mathcal{H}_{\mathbb{I}} : \mu_1 = \ldots = \mu_I$. Si $\mathcal{H}_{\mathbb{I}}$ n'est pas rejetée au seuil $\alpha = \alpha_I$, alors on décide que l'on ne peut rejeter aucune des hypothèses $\mathcal{H}_{\mathbb{K}}$ pour $\mathbb{K} \subset \mathbb{I}$ et l'on arrête le procédé à cette étape. Sinon on décide qu'il existe i_0 et j_0 dans $\{1, \ldots, I\}$ tels que $\mu_{i_0} \neq \mu_{j_0}$ et l'on se rend à l'étape **2.**.

2. Tester au seuil $\alpha = \alpha_{|\mathbb{I}|-1} = \alpha_{I-1}$, donné par l'équation (10.1), chaque $\mathcal{H}_{\mathbb{L}}$ avec $|\mathbb{L}| = I - 1$ et en imposant la condition que les entiers qui appartiennent à $\mathbb{L}$ sont consécutifs qui n'a pas encore été acceptée. Si l'hypothèse $\mathcal{H}_{\mathbb{L}}$ est acceptée, alors on décide que l'on ne peut rejeter aucune des hypothèses $\mathcal{H}_{\mathbb{K}}$ pour $\mathbb{K} \subset \mathbb{L}$. Sinon on décide qu'il existe i_0 et j_0 dans $\mathbb{L}$ tels que $\mu_{i_0} \neq \mu_{j_0}$. Si aucune des $\mathcal{H}_{\mathbb{L}}$ n'est rejetée, on arrête la procédure ; dans le cas contraire, on se rend à l'étape **3.**.

3. Tester au seuil $\alpha = \alpha_{|\mathbb{L}|}$, donné par l'équation (10.1), chaque $\mathcal{H}_{\mathbb{L}}$ et en imposant la condition que les entiers qui appartiennent à $\mathbb{L}$ sont consécutifs qui n'a pas encore été acceptée lors d'une étape précédente. Si $\mathcal{H}_{\mathbb{L}}$ est acceptée, alors on décide que l'on ne peut rejeter aucune des hypothèses $\mathcal{H}_{\mathbb{K}}$ pour $\mathbb{K} \subset \mathbb{L}$. Sinon on décide qu'il existe i_0 et j_0 dans $\mathbb{L}$ tels que $\mu_{i_0} \neq \mu_{j_0}$.

4. Continuer jusqu'à ce qu'il ne reste plus aucune hypothèse $\mathcal{H}_{\mathbb{L}}$, pour laquelle les entiers qui appartiennent à $\mathbb{L}$ sont consécutifs, à tester au seuil $\alpha_{|\mathbb{J}|}$.

5. Enfin décider que $\mu_{i_0} \neq \mu_{j_0}$ si tout $\mathcal{H}_{\mathbb{L}}$ telle que $i_0, j_0 \in \mathbb{L}$ est rejetée.

Pour réaliser le test multiple des étendues de Ryan-Einot-Gabriel-Welsch on teste les hypothèses $\mathcal{H}_{\mathbb{L}}$ au seuil $\alpha_{|\mathbb{L}|}$ de la manière suivante :

- On ne peut pas rejeter l'hypothèse $\mathcal{H}_{\mathbb{L}}$ au seuil $\alpha_{|\mathbb{L}|}$ si

$$\max_{i,j \in \mathbb{L}} \widehat{\mu}_i(\boldsymbol{y}) - \widehat{\mu}_j(\boldsymbol{y}) < c_{\alpha_{|\mathbb{L}|}} \sqrt{\frac{2s_R^2}{n}},$$

où $c_{\alpha_{|\mathbb{L}|}}$ est défini par :

$$\alpha_{|\mathbb{L}|} = \mathbb{P}\left[\max_{i,j \in \mathbb{L}} \frac{\widehat{\mu}_i - \widehat{\mu}_j}{\sqrt{\frac{2S_R^2}{n}}} > c_{\alpha_{|\mathbb{L}|}}\right].$$

Ainsi avec le choix de l'équation (10.1) pour les valeurs des α_m, $1 \leqslant m \leqslant I$ on obtient, si $2 \leqslant |\mathbb{L}| \leqslant I - 2$:

$$1 - (1-\alpha)^{\frac{|\mathbb{L}|}{I}} = \mathbb{P}\left[\max_{i,j \in \mathbb{L}} \frac{\widehat{\mu}_i - \widehat{\mu}_j}{\sqrt{\frac{2S_R^2}{n}}} > c_{\alpha_{|\mathbb{L}|}}\right],$$

et si $I-1 \leqslant |\mathbb{L}| \leqslant I$:

$$\alpha = \mathbb{P}\left[\max_{i,j\in\mathbb{L}} \frac{\widehat{\mu}_i - \widehat{\mu}_j}{\sqrt{\dfrac{2s_R^2}{n}}} > c_{\alpha_{|\mathbb{L}|}}\right].$$

Ainsi $c_{\alpha_{|\mathbb{L}|}}$ est donc le $100(1-\alpha_{|\mathbb{L}|})$ quantile de la loi de l'étendue Studentisée à $|\mathbb{L}|$ et $n-I$ degrés de liberté.

- On décide de rejeter l'hypothèse $\mathcal{H}_{\mathbb{L}}$ au seuil $\alpha_{|\mathbb{L}|}$ si

$$\max_{i,j\in\mathbb{L}} \widehat{\mu}_i(\boldsymbol{y}) - \widehat{\mu}_j(\boldsymbol{y}) \geqslant c_{\alpha_{|\mathbb{L}|}} \sqrt{\frac{2s_R^2}{n}},$$

où $c_{\alpha_{|\mathbb{L}|}}$ est défini comme ci-dessus.

Si les valeurs critiques c_m, $1 \leqslant m \leqslant I$ utilisées sont croissantes avec m alors le test multiple des étendues de Ryan-Einot-Gabriel-Welsch est un cas particulier de la procédure descendante ci-dessous et de ce fait valide. Malheureusement, avec la définition de α_m donnée par l'équation (10.1), il n'y a aucune garantie que la valeur critique c_m soit croissante avec m. Pour comprendre pourquoi cette condition est absolument nécessaire, on pourra consulter l'exemple page 5067 dans [7]

Ainsi, on préférera utiliser, si les valeurs critiques calculées à l'aide de l'équation (10.1) ne sont pas monotones, c'est-à-dire si elles ne sont pas croissantes, les valeurs critiques modifiées définies par :

$$c_m^{'} = \max_{2\leqslant j\leqslant m} c_j, \quad m = 2, \ldots, I.$$

Attention à bien vérifier ce point si l'on ne veut pas obtenir de résultats incohérents.

Remarque :

On aurait pu utiliser la procédure ci-dessous avec une autre manière de tester les hypothèses $\mathcal{H}_{\mathbb{L}}$, par exemple des tests t à deux échantillons, on utilisera alors l'inégalité de Bonferroni et les valeurs suivantes pour les seuils :

$$\alpha_m = \begin{cases} \dfrac{m}{I}\alpha & \text{si} \quad m = 2, \ldots, I-2, \\ \alpha & \text{si} \quad m = I-1, I. \end{cases}$$

Toutefois cette procédure possède le même défaut que celle de Bonferroni par rapport à celle de la méthode de Dunn-Šidák : elle est plus conservative, voir le paragraphe 10.2.8.

10.4.2. Tests $F-$multiples

La validité de la méthode des tests $F-$multiples est basée sur la validité du procédé plus général de comparaisons descendant décrit ci-dessous.

Pour tout $\mathbb{L} \subseteq \mathbb{I} = \{1, \ldots, I\}$, soit $H_{\mathbb{L}}$ l'hypothèse :

$$\boxed{\mathcal{H}_{\mathbb{L}} : \mu_i = \mu_j \text{ pour tout } i, j \text{ de } \mathbb{L},}$$

et soit $|\mathbb{L}|$ le nombre d'éléments de $\mathbb{L}$.

1. Contrairement aux tests basés sur l'étendue on ne peut se contenter de tester des hypothèses $\mathcal{H}_{\mathbb{L}}$ pour lesquelles les indices appartenant à $\mathbb{L}$ sont consécutifs. Attention à bien faire tous les tests au seuil $\alpha_{\mathbb{L}}$ pour toutes les hypothèses $\mathcal{H}_{\mathbb{L}}$ possibles, avec $\mathbb{L} \subset \mathbb{I}$ si l'on ne veut pas obtenir de résultats incohérents au seuil α.

2. Continuer jusqu'à ce qu'il ne reste plus aucune hypothèse $\mathcal{H}_{\mathbb{L}}$ à tester.
3. Décider alors que $\mu_{i_0} \neq \mu_{j_0}$ au seuil α si toutes les hypothèses $\mathcal{H}_{\mathbb{L}}$ pour lesquelles $i_0, j_0 \in \mathbb{L}$ sont rejetées au seuil $\alpha_{\mathbb{L}}$.

Pour réaliser les tests $F-$multiples on teste les hypothèses $\mathcal{H}_{\mathbb{L}}$ au seuil $\alpha_{|\mathbb{L}|}$ de la manière suivante :

- On ne peut pas rejeter l'hypothèse $\mathcal{H}_{\mathbb{L}}$ au seuil $\alpha_{|\mathbb{L}|}$ si

$$\frac{\sum_{i \in \mathbb{L}} n \left(\widehat{\mu}_i(\boldsymbol{y}) - \widehat{\overline{\mu_{\mathbb{L}}}}(\boldsymbol{y})\right)^2}{(|\mathbb{L}| - 1)\, s_R^2} < c_{\alpha_{|\mathbb{L}|}},$$

où $\widehat{\overline{\mu_{\mathbb{L}}}}$ est défini par

$$\widehat{\overline{\mu_{\mathbb{L}}}} = \frac{\sum_{i \in \mathbb{L}} \sum_{a=1}^{n} Y_{i,a}}{\sum_{i \in \mathbb{L}} n},$$

ainsi que $c_{\alpha_{|\mathbb{L}|}}$ par

$$\alpha_{|\mathbb{L}|} = \mathbb{P}\left[\frac{\sum_{i \in \mathbb{L}} n \left(\widehat{\mu}_i - \widehat{\overline{\mu_{\mathbb{L}}}}\right)^2}{(|\mathbb{L}| - 1)\, S_R^2} > c_{\alpha_{|\mathbb{L}|}}\right].$$

Ainsi avec le choix de l'équation (10.1) pour les valeurs des α_i, $1 \leqslant i \leqslant I$ on obtient, si $2 \leqslant |\mathbb{L}| \leqslant I - 2$:

$$1 - (1 - \alpha)^{\frac{|\mathbb{L}|}{I}} = \mathbb{P}\left[\frac{\sum_{i \in \mathbb{L}} n \left(\widehat{\mu}_i - \widehat{\overline{\mu_{\mathbb{L}}}}\right)^2}{(|\mathbb{L}| - 1)\, S_R^2} > c_{\alpha_{|\mathbb{L}|}}\right],$$

et si $I-1 \leqslant |\mathbb{L}| \leqslant I$:

$$\alpha = \mathbb{P}\left[\frac{\sum_{i\in\mathbb{L}} n\left(\widehat{\mu}_i - \widehat{\overline{\mu_{\mathbb{L}}}}\right)^2}{(|\mathbb{L}|-1)\, S_R^2} > c_{\alpha_{|\mathbb{L}|}}\right].$$

Ainsi $c_{\alpha_{|\mathbb{L}|}}$ est donc le $100(1-\alpha_{|\mathbb{L}|})$ quantile de la loi de Fisher à $|\mathbb{L}|-1$ et $n-I$ degrés de liberté.

- On décide de rejeter l'hypothèse $\mathcal{H}_{\mathbb{L}}$ au seuil $\alpha_{|\mathbb{L}|}$ si

$$\frac{\sum_{i\in\mathbb{L}} n\left(\widehat{\mu}_i(\boldsymbol{y}) - \widehat{\overline{\mu_{\mathbb{L}}}}(\boldsymbol{y})\right)^2}{(|\mathbb{L}|-1)\, s_R^2} \geqslant c_{\alpha_{|\mathbb{L}|}},$$

où $c_{\alpha_{|\mathbb{L}|}}$ est défini comme ci-dessus.

10.5. Comparaisons avec le « meilleur »

Dans certains cas il peut être intéressant de comparer les effets de tous les niveaux d'un facteur avec l'effet extrême, maximal ou minimal, que l'on a observé et que l'on considère comme le « meilleur ». Ce type de problématique ne peut être traité à l'aide d'une approche basée sur la méthode de Dunnett puisque l'on ne sait pas ***a priori*** pour quelle modalité l'on va observer le « meilleur » effet.

On utilise alors des techniques de comparaisons multiples avec le « meilleur[18] », le plus souvent celle de Hsu ou celle de Edwards-Hsu.

On s'intéresse ici donc non pas à toutes les comparaisons possibles mais seulement à celles-ci :

- Si le « meilleur » est la valeur la plus élévée on utilise :

$$\mu_i - \max_{j\neq i} \mu_j, \quad i = 1, \ldots, I,$$

ainsi si $\mu_{i_0} - \max_{j\neq i_0} \mu_j > 0$ alors la modalité i_0 est celle pour laquelle l'effet est maximal et de ce fait la « meilleure ». Par contre si $\mu_{i_1} - \max_{j\neq i_1} \mu_j < 0$ alors la modalité i_1 n'est pas celle pour laquelle l'effet est maximal.
Si l'effet de la modalité i_2 n'est pas le plus important mais que l'on a néanmoins $\mu_{i_2} - \max_{j\neq i_2} \mu_j > -\delta$ avec δ un petit nombre réel strictement positif alors la modalité i_2 est proche de la « meilleure ».

[18] Ceci est souvent abrégé en anglais en *MCB* pour « Multiple comparisons with the Best ».

- Si le « meilleur » est la valeur la moins élévée on utilise :

$$\mu_i - \min_{j \neq i} \mu_j, \quad i = 1, \ldots, I,$$

ainsi si $\mu_{i_0} - \min_{j \neq i_0} \mu_j < 0$ alors la modalité i_0 est celle pour laquelle l'effet est minimal et de ce fait la « meilleure ». Par contre si $\mu_{i_1} - \min_{j \neq i_1} \mu_j > 0$ alors la modalité i_1 n'est pas celle pour laquelle l'effet est minimal.
Si l'effet de la modalité i_2 n'est pas le plus important mais que l'on a néanmoins $\mu_{i_2} - \min_{j \neq i_2} \mu_j > \delta$ avec δ un petit nombre réel strictement positif alors la modalité i_2 est proche de la « meilleure ».

10.5.1. La méthode de Hsu

Hsu a développé des intervalles de confiance simultanés de niveau de confiance $100(1-\alpha)$ pour $\mu_i - \max_{j \neq i} \mu_j,\ i = 1, \ldots, I$. Ces intervalles contiennent tous la valeur 0 et sont construits ainsi : pour chaque i, soit d_i la valeur critique telle que

$$\mathbb{P}\left[\widehat{\mu}_i - \mu_i > \widehat{\mu}_j - \mu_j - d_i \sqrt{S_R^2 \left(\frac{1}{n_i} + \frac{1}{n_j}\right)},\ \forall\, j | j \neq i\right] = 1 - \alpha.$$

On remarque que d_i est la valeur critique pour les comparaisons multiples unilatérales en considérant le $i-$ème niveau comme le niveau de contrôle.

Les intervalles fermés $\left[D_i^-, D_i^+\right]$, $i = 1, \ldots, I$ où

$$\begin{aligned}
D_i^+ &= \max\left\{0, \min_{j \neq i}\left(\widehat{\mu}_i - \widehat{\mu}_j - d_i \sqrt{S_R^2 \left(\frac{1}{n_i} + \frac{1}{n_j}\right)}\right)\right\} \\
G &= \left\{i \mid D_i^+ > 0\right\} \\
D_i^- &= \begin{cases} 0 & \text{si } i \notin G \\ \min_{j \in G, j \neq i}\left(\widehat{\mu}_i - \widehat{\mu}_j - d_j \sqrt{S_R^2 \left(\frac{1}{n_i} + \frac{1}{n_j}\right)}\right) & \text{si } i \in G \end{cases}
\end{aligned}$$

forment un ensemble d'intervalles de confiance simultanés de niveau de confiance $100(1-\alpha)$ pour $\mu_i - \max_{j \neq i} \mu_j,\ i = 1, \ldots, I$.

Lorsque le plan est équilibré les intervalles se transforment en :

$$\left[-\min\left\{0, \left(\widehat{\mu}_i - \max_{j \neq i} \widehat{\mu}_j - d\sqrt{\frac{2S_R^2}{n}}\right)\right\}, \max\left\{0, \left(\widehat{\mu}_i - \max_{j \neq i} \widehat{\mu}_j - d\sqrt{\frac{2S_R^2}{n}}\right)\right\}\right],$$

où $1 \leqslant i \leqslant I$ et $d = d_1 = \ldots = d_I$.

10.5.2. La méthode de Edwards-Hsu

Si l'on cherche à avoir également des bornes sur l'écart minimum entre la valeur du « meilleur » traitement et celle des autres traitements que celui dont l'estimation a été la meilleure, alors on peut utiliser la méthode de Edwards-Hsu. Celle-ci permet d'obtenir des intervalles de confiance simultanés pour $\mu_i - \max_{i \leqslant j \leqslant I} \mu_j$, $i = 1, \ldots, I$ que l'on ne force plus à contenir la valeur 0.

Pour chaque i, soit $|d|_i$ la valeur critique telle que

$$\mathbb{P}\left[\widehat{\mu}_i - \mu_i - (\widehat{\mu}_j - \mu_j) < |d|_i \sqrt{S_R^2 \left(\frac{1}{n_i} + \frac{1}{n_j}\right)},\ \forall\ j | j \neq i\right] = 1 - \alpha.$$

On remarque que d_i est la valeur critique pour les comparaisons multiples avec un contrôle en considérant le $i-$ème niveau comme le niveau de contrôle et de ce fait le nombre $|d|_i$ est supérieur à d_i qui est utilisé dans la méthode de Hsu.

Les intervalles $[L_i, U_i]$ définis par

$$S = \left\{ i \text{ tels que } \min_{j \neq i} \left\{ \widehat{\mu}_i - \widehat{\mu}_j + |d|_i \sqrt{S_R^2 \left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \right\} > 0 \right\},$$

$$L_{ij} = \begin{cases} 0 & \text{si } i = j \\ \widehat{\mu}_i - \widehat{\mu}_j - |d|_j \sqrt{S_R^2 \left(\frac{1}{n_i} + \frac{1}{n_j}\right)} & \text{si } i \neq j \end{cases},$$

$$U_{ij} = \begin{cases} 0 & \text{si } i = j \\ -\min\left\{0, \left(\widehat{\mu}_i - \widehat{\mu}_j + |d|_j \sqrt{S_R^2 \left(\frac{1}{n_i} + \frac{1}{n_j}\right)}\right)\right\} & \text{si } i \neq j \end{cases},$$

$$L_i = \min_{j \in S} L_{ij} \qquad U_i = \max_{j \in S} U_{ij}$$

pour $i = 1, \ldots, I$, forment un ensemble d'intervalles de confiances simultanés de niveau $100(1 - \alpha)$ pour $\mu_i - \max_{i \leqslant j \leqslant I} \mu_j$, $i = 1, \ldots, I$.

Lorsque le plan est équilibré les quantités utilisées pour construire les intervalles se transforment de la manière suivante.

Il n'y a qu'une seule valeur critique $|d| = |d|_1 = |d|_I$ définie par :

$$\mathbb{P}\left[\widehat{\mu}_i - \mu_i - (\widehat{\mu}_j - \mu_j) < |d| \sqrt{\frac{2S_R^2}{n}},\ \forall\ j | j \neq i\right] = 1 - \alpha.$$

On remarque que d est la valeur critique pour les comparaisons multiples avec un contrôle en considérant l'un quelconque des niveaux comme le niveau de contrôle et de ce fait le nombre $|d|$ est supérieur à d qui est utilisé dans la méthode de Hsu.

Les intervalles $[L_i, U_i]$ définis par

$$S = \left\{ i \text{ tels que } \min_{j \neq i} \left\{ \widehat{\mu}_i - \widehat{\mu}_j + |d| \sqrt{\frac{2S_R^2}{n}} \right\} > 0 \right\},$$

$$\begin{aligned} L_{ij} &= \begin{cases} 0 & \text{si } i = j \\ \widehat{\mu}_i - \widehat{\mu}_j - |d| \sqrt{\dfrac{2S_R^2}{n}} & \text{si } i \neq j \end{cases}, \\ U_{ij} &= \begin{cases} 0 & \text{si } i = j \\ -\min \left\{ 0, \left(\widehat{\mu}_i - \widehat{\mu}_j + |d| \sqrt{\dfrac{2S_R^2}{n}} \right) \right\} & \text{si } i \neq j \end{cases}, \end{aligned}$$

$$L_i = \min_{j \in S} L_{ij} \qquad U_i = \max_{j \in S} U_{ij}$$

pour $i = 1, \ldots, I$, forment un ensemble d'intervalles de confiances simultanés de niveau $100(1-\alpha)$ pour $\mu_i - \max_{i \leqslant j \leqslant I} \mu_j$, $i = 1, \ldots, I$.

10.6. Deux facteurs sans répétition

On peut adapter les méthodes de comparaison des effets des différentes modalités, exposées aux paragraphes 10.1, 10.2 et 10.3 dans le cas de l'analyse de la variance à un facteur à effets fixes, au cas où l'on considère un modèle à deux facteurs sans interaction à effets fixes ou à effets mixtes. On commence par détailler les modifications à faire pour utiliser les méthodes de Tukey et de Scheffé.

On reprend ici les notations du paragraphe 4.1.1 ou du paragraphe 4.3.1 en fonction de la situation étudiée.

10.6.1. Contrastes

Les contrastes mettent en jeu des moyennes associées soit au premier facteur A, s'il est à effets fixes, soit au second facteur B, s'il est à effets fixes. Il y a ainsi deux types de contrastes :

$$\begin{aligned} L &= l_1\alpha_1 + l_2\alpha_2 + \cdots + l_I\alpha_I, \\ L' &= l'_1\beta_1 + l'_2\beta_2 + \cdots + l'_J\beta_J. \end{aligned}$$

avec $\sum_{i=1}^{I} l_i = 0$ et $\sum_{j=1}^{J} l_j^{'} = 0$.
Des estimateurs sans biais de L et $L^{'}$ sont alors :

$$\widehat{L} = l_1\widehat{\alpha_1} + l_2\widehat{\alpha_2} + \cdots + l_I\widehat{\alpha_I},$$
$$\widehat{L'} = l_1^{'}\widehat{\beta}_1 + l_2^{'}\widehat{\beta}_2 + \cdots + l_J^{'}\widehat{\beta_J}.$$

Ainsi des estimations sans biais de L et $L^{'}$ s'écrivent :

$$\widehat{L}(\boldsymbol{y}) = l_1 y_{1,\bullet} + l_2 y_{2,\bullet} + \cdots + l_I y_{I,\bullet},$$
$$\widehat{L'}(\boldsymbol{y}) = l_1^{'} y_{\bullet,1} + l_2^{'} y_{\bullet,2} + \cdots + l_J^{'} y_{\bullet,J}.$$

10.6.2. Méthode de Tukey

Si un contraste L met en jeu des moyennes associées au facteur A à effets fixes, on utilise l'estimation s_R^2 calculée pour le modèle incluant les deux facteurs et on change le nombre de degrés de liberté de la loi de l'étendue Studentisées en I et $(I-1)(J-1)$.
On décide que le contraste L est significativement différent de 0 au seuil α si :

$$\frac{\widehat{L}(\boldsymbol{y})}{\sqrt{\frac{s_R^2}{J}}\left(\frac{1}{2}\sum_{i=1}^{I}|l_i|\right)} \geqslant q(I,(I-1)(J-1);1-\alpha).$$

où $q(I,(I-1)(J-1);1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de l'étendue Studentisée à I et $(I-1)(J-1)$ degrés de liberté. Sinon on ne peut rejeter l'hypothèse d'absence de différence entre L et 0 au seuil α.

De même si le contraste $L^{'}$ met en jeu des moyennes associées au facteur B à effets fixes, on utilise l'estimation s_R^2 calculée pour le modèle incluant les deux facteurs et on change le nombre de degrés de liberté de la loi de l'étendue Studentisées en J et $(I-1)(J-1)$.
On décide que le contraste $L^{'}$ est significativement différent de 0 au seuil α si :

$$\frac{\widehat{L'}(\boldsymbol{y})}{\sqrt{\frac{s_R^2}{I}}\left(\frac{1}{2}\sum_{j=1}^{J}\left|l_j^{'}\right|\right)} \geqslant q(J,(I-1)(J-1);1-\alpha).$$

où $q(J,(I-1)(J-1);1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de l'étendue Studentisée à J et $(I-1)(J-1)$ degrés de liberté. Sinon on ne peut rejeter l'hypothèse d'absence de différence entre $L^{'}$ et 0 au seuil α.

10.6.3. Méthode de Scheffé

Si un contraste L met en jeu des moyennes associées au facteur A à effets fixes, on utilise l'estimation s_R^2 calculée pour le modèle incluant les deux facteurs et on change le nombre

de degrés de liberté de la loi de Fisher en $I-1$ et $(I-1)(J-1)$.
On décide que le contraste L est significativement différent de 0 au seuil α si :

$$\frac{\widehat{L}(\boldsymbol{y})}{\sqrt{\dfrac{(I-1)s_R^2}{J}\displaystyle\sum_{i=1}^{I} l_i^2}} \geqslant \sqrt{F(I-1,(I-1)(J-1);1-\alpha)},$$

où $F(I-1,(I-1)(J-1);1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $I-1$ et $(I-1)(J-1)$ degrés de liberté. Sinon on ne peut rejeter l'hypothèse d'absence de différence entre L et 0 au seuil α.

De même si le contraste $L^{'}$ met en jeu des moyennes associées au facteur B à effets fixes, on utilise l'estimation s_R^2 calculée pour le modèle incluant les deux facteurs et on change le nombre de degrés de liberté de la loi de Fisher en $J-1$ et $(I-1)(J-1)$.
On décide que le contraste $L^{'}$ est significativement différent de 0 au seuil α si :

$$\frac{\widehat{L^{'}}(\boldsymbol{y})}{\sqrt{\dfrac{(J-1)s_R^2}{I}\displaystyle\sum_{j=1}^{J} l_j'^2}} \geqslant \sqrt{F(J-1,(I-1)(J-1);1-\alpha)},$$

où $F(J-1,(I-1)(J-1);1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $J-1$ et $(I-1)(J-1)$ degrés de liberté. Sinon on ne peut rejeter l'hypothèse d'absence de différence entre $L^{'}$ et 0 au seuil α.

10.6.4. Méthode de Bonferroni et méthodes associées

On adapte ici, de manière similaire, la procédure de Bonferroni et toutes celles qui en découle.

Ainsi si l'on ne souhaite **comparer que deux niveaux** d'un même des deux facteurs, ici par exemple deux niveaux du facteur A, on pourra utiliser la valeur critique

$$\sqrt{2}\; t((I-1)(J-1);1-\frac{\alpha}{2})$$

à la place de

$$q(I,(I-1)(J-1);1-\alpha)$$

où $t((I-1)(J-1);1-\alpha/2)$ est le $100(1-\alpha/2)$ quantile de la loi de Student à $(I-1)(J-1)$ degrés de liberté.

Si l'on souhaite réaliser **un nombre fini k de comparaisons** on pourra utiliser la valeur critique

$$\sqrt{2}\, t((I-1)(J-1); 1-\frac{\alpha}{2k})$$

à la place de

$$q(I, (I-1)(J-1); 1-\alpha)$$

où $t((I-1)(J-1); 1-\frac{\alpha}{2k})$ est le $100(1-\frac{\alpha}{2k})$ quantile de la loi de Student à $(I-1)(J-1)$ degrés de liberté.

10.7. Deux facteurs avec répétitions

On peut adapter les méthodes de comparaison des effets des différentes modalités, exposées aux paragraphes 10.1, 10.2 et 10.3 dans le cas de l'analyse de la variance à un facteur à effets fixes, au cas où l'on considère un modèle à deux facteurs avec interaction à effets fixes ou à effets mixtes. On commence par détailler les modifications à faire pour utiliser les méthodes de Tukey et de Scheffé.

On reprend ici les notations du paragraphe 4.1.2 ou du paragraphe 4.3.2 en fonction de la situation considérée. La différence fondamentale entre ces deux situations résidant dans le fait que pour la première, **deux facteurs à effets fixes**, on utilisera le **carré moyen résiduel** s_R^2 dans les statistiques de test tandis que dans la seconde, **un facteur à effets fixes et un facteur à effets aléatoires**, on utilisera le **carré moyen associé au terme de l'interaction** s_{AB}^2 dans les statistiques de test.

10.7.1. Contrastes

Modèle à effets fixes

Dans le cas d'un modèle où les deux facteurs sont à effets fixes, c'est-à-dire le cas envisagé au paragraphe 4.1.2, il peut être intéressant, en fonction de la significativité des différents tests auxquels on a procédé, d'étudier les constrastes exposés ci-dessous.

$$\boxed{\mathcal{H}_0''' : (\alpha\beta)_{1,1} = (\alpha\beta)_{1,2} = \cdots = (\alpha\beta)_{1,J} = (\alpha\beta)_{2,1} = \cdots = (\alpha\beta)_{I,J} = 0}$$

contre

$$\boxed{\mathcal{H}_1''' : \text{Il existe } (i_0, j_0) \in \{1, 2, \ldots, I\} \times \{1, 2, \ldots, J\} \text{ tel que } (\alpha\beta)_{i_0,j_0} \neq 0.}$$

Si l'on rejete l'hypothèse nulle $\mathcal{H}_0'''$ il peut être intéressant de comparer entre elles les réponses moyennes dans chacun des couples de modalités A_i et B_j. Les contrastes L_{AB} qui permettent de comparer les moyennes de la réponse $\mu_{i,j} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{i,j}$ dans

deux différentes configurations expérimentales modalité i du facteur A et modalité j du facteur B contre modalité $i^{'}$ du facteur A et modalité $j^{'}$ du facteur B sont :

$$L_{AB} = \mu_{i,j} - \mu_{i^{'},j^{'}}.$$

Une estimation sans biais de L_{AB} est donnée par :

$$\widehat{L_{AB}}(\boldsymbol{y}) = y_{i,j,\bullet} - y_{i^{'},j^{'},\bullet}.$$

Si l'on ne peut rejeter l'hypothèse nulle $\mathcal{H}_0^{'''}$ il peut être intéressant de comparer entre elles les réponses moyennes dans chacune des modalités A_i ou dans chacune des modalités B_j.
Les contrastes mettent en jeu des moyennes associées soit au premier facteur A soit au second facteur B. Il y a ainsi deux types de contrastes :

$$\begin{aligned} L &= l_1\alpha_1 + l_2\alpha_2 + \cdots + l_I\alpha_I, \\ L^{'} &= l_1^{'}\beta_1 + l_2^{'}\beta_2 + \cdots + l_J^{'}\beta_J, \end{aligned}$$

avec $\sum_{i=1}^{I} l_i = 0$ et $\sum_{j=1}^{J} l_j^{'} = 0$.
Des estimateurs sans biais de L et $L^{'}$ sont alors :

$$\begin{aligned} \widehat{L} &= l_1\widehat{\alpha_1} + l_2\widehat{\alpha_2} + \cdots + l_I\widehat{\alpha_I}, \\ \widehat{L^{'}} &= l_1^{'}\widehat{\beta_1} + l_2^{'}\widehat{\beta_2} + \cdots + l_J^{'}\widehat{\beta_J}. \end{aligned}$$

Ainsi des estimations sans biais de L et $L^{'}$ s'écrivent :

$$\begin{aligned} \widehat{L}(\boldsymbol{y}) &= l_1 y_{1,\bullet,\bullet} + l_2 y_{2,\bullet,\bullet} + \cdots + l_I y_{I,\bullet,\bullet}, \\ \widehat{L^{'}}(\boldsymbol{y}) &= l_1^{'} y_{\bullet,1,\bullet} + l_2^{'} y_{\bullet,2,\bullet} + \cdots + l_J^{'} y_{\bullet,J,\bullet}. \end{aligned}$$

Modèle à effets mixtes

Dans le cas d'un modèle mixte, c'est-à-dire le cas envisagé au paragraphe 4.3.2, il peut être intéressant, en fonction de la significativité du test de l'effet du facteur à effets fixes, d'étudier les constrastes du type :

$$L = l_1\alpha_1 + l_2\alpha_2 + \cdots + l_I\alpha_I,$$

avec $\sum_{i=1}^{I} l_i = 0$.
Un estimateur sans biais de L est alors :

$$\widehat{L} = l_1\widehat{\alpha_1} + l_2\widehat{\alpha_2} + \cdots + l_I\widehat{\alpha_I}.$$

Ainsi une estimation sans biais de L s'écrit :

$$\widehat{L}(\boldsymbol{y}) = l_1 y_{1,\bullet,\bullet} + l_2 y_{2,\bullet,\bullet} + \cdots + l_I y_{I,\bullet,\bullet}.$$

10.7.2. Méthode de Tukey

Modèle à effets fixes

L_{AB} est significativement différent de 0 au seuil α si, en supposant que $y_{i,j,\bullet} \geqslant y_{i',j',\bullet}$, on a :

$$\frac{\widehat{L}(\boldsymbol{y})}{\sqrt{\frac{s_R^2}{K}}} \geqslant q(IJ, IJ(K-1); 1-\alpha),$$

où $q(IJ, IJ(K-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de l'étendue Studentisée à IJ et $IJ(K-1)$ degrés de liberté. Sinon on ne peut rejeter l'hypothèse d'absence de différence entre L_{AB} et 0 au seuil α.

Si un contraste L met en jeu des moyennes associées au facteur A à effets fixes, on utilise l'estimation s_R^2 calculée pour le modèle incluant les deux facteurs et on change le nombre de degrés de liberté de la loi de l'étendue Studentisées en I et $IJ(K-1)$.
On décide que le contraste L est significativement différent de 0 au seuil α si :

$$\frac{\widehat{L}(\boldsymbol{y})}{\sqrt{\frac{s_R^2}{JK}}\left(\frac{1}{2}\sum_{i=1}^{I}|l_i|\right)} \geqslant q(I, IJ(K-1); 1-\alpha).$$

où $q(I, IJ(K-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de l'étendue Studentisée à I et $IJ(K-1)$ degrés de liberté. Sinon on ne peut rejeter l'hypothèse d'absence de différence entre L et 0 au seuil α.

De même si le contraste $L^{'}$ met en jeu des moyennes associées au facteur B à effets fixes, on utilise l'estimation s_R^2 calculée pour le modèle incluant les deux facteurs et on change le nombre de degrés de liberté de la loi de l'étendue Studentisées en J et $IJ(K-1)$.
On décide que le contraste $L^{'}$ est significativement différent de 0 au seuil α si :

$$\frac{\widehat{L'}(\boldsymbol{y})}{\sqrt{\frac{s_R^2}{IK}}\left(\frac{1}{2}\sum_{j=1}^{J}\left|l_j'\right|\right)} \geqslant q(J, IJ(K-1); 1-\alpha).$$

où $q(J, IJ(K-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de l'étendue Studentisée à J et $IJ(K-1)$ degrés de liberté. Sinon on ne peut rejeter l'hypothèse d'absence de différence entre $L^{'}$ et 0 au seuil α.

Modèle à effets mixtes

Si un contraste L met en jeu des moyennes associées au facteur A à effets fixes, on utilise l'estimation s^2_{AB} calculée pour le modèle incluant les deux facteurs et on change le nombre de degrés de liberté de la loi de l'étendue Studentisées en I et $(I-1)(J-1)$.
On décide que le contraste L est significativement différent de 0 au seuil α si :

$$\frac{\widehat{L}(\boldsymbol{y})}{\sqrt{\frac{s^2_{AB}}{JK}}\left(\frac{1}{2}\sum_{i=1}^{I}|l_i|\right)} \geqslant q(I,(I-1)(J-1);1-\alpha).$$

où $q(I,(I-1)(J-1);1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de l'étendue Studentisée à I et $(I-1)(J-1)$ degrés de liberté. Sinon on ne peut rejeter l'hypothèse d'absence de différence entre L et 0 au seuil α.

10.7.3. Méthode de Scheffé

Modèle à effets fixes

L_{AB} est significativement différent de 0 au seuil α si, en supposant que $y_{i,j,\bullet} \geqslant y_{i',j',\bullet}$, on a :

$$\frac{\widehat{L}(\boldsymbol{y})}{\sqrt{(IJ-1)\frac{2s^2_R}{K}}} \geqslant \sqrt{F(IJ-1,IJ(K-1);1-\alpha)},$$

où $F(IJ-1,IJ(K-1);1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $IJ-1$ et $IJ(K-1)$ degrés de liberté. Sinon on ne peut rejeter l'hypothèse d'absence de différence entre L_{AB} et 0 au seuil α.

Si un contraste L met en jeu des moyennes associées au facteur A à effets fixes, on utilise l'estimation s^2_R calculée pour le modèle incluant les deux facteurs et on change le nombre de degrés de liberté de la loi de Fisher en $I-1$ et $IJ(K-1)$.
On décide que le contraste L est significativement différent de 0 au seuil α si :

$$\frac{\widehat{L}(\boldsymbol{y})}{\sqrt{\frac{(I-1)s^2_R}{JK}\sum_{i=1}^{I} l_i^2}} \geqslant \sqrt{F(I-1,IJ(K-1);1-\alpha)},$$

où $F(I-1,IJ(K-1);1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $I-1$ et $IJ(K-1)$ degrés de liberté. Sinon on ne peut rejeter l'hypothèse d'absence de différence entre L et 0 au seuil α.

De même si le contraste $L^{'}$ met en jeu des moyennes associées au facteur B à effets fixes, on utilise l'estimation s_R^2 calculée pour le modèle incluant les deux facteurs et on change le nombre de degrés de liberté de la loi de Fisher en $J-1$ et $IJ(K-1)$.
On décide que le contraste $L^{'}$ est significativement différent de 0 au seuil α si :

$$\frac{\widehat{L'}(\boldsymbol{y})}{\sqrt{\dfrac{(J-1)s_R^2}{IK}\displaystyle\sum_{j=1}^{J} l_j'^2}} \geqslant \sqrt{F(J-1, IJ(K-1); 1-\alpha)},$$

où $F(J-1, IJ(K-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $J-1$ et $IJ(K-1)$ degrés de liberté. Sinon on ne peut rejeter l'hypothèse d'absence de différence entre $L^{'}$ et 0 au seuil α.

Modèle à effets mixtes

Si un contraste L met en jeu des moyennes associées au facteur A à effets fixes, on utilise l'estimation s_{AB}^2 calculée pour le modèle incluant les deux facteurs et on change le nombre de degrés de liberté de la loi de Fisher en $I-1$ et $(I-1)(J-1)$.
On décide que le contraste L est significativement différent de 0 au seuil α si :

$$\frac{\widehat{L}(\boldsymbol{y})}{\sqrt{\dfrac{(I-1)s_{AB}^2}{JK}\displaystyle\sum_{i=1}^{I} l_i^2}} \geqslant F(I-1, (I-1)(J-1); 1-\alpha).$$

où $F(I-1, (I-1)(J-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à I et $(I-1)(J-1)$ degrés de liberté. Sinon on ne peut rejeter l'hypothèse d'absence de différence entre L et 0 au seuil α.

10.7.4. Méthode de Bonferroni et méthodes associées

On adapte ici, de manière similaire, la procédure de Bonferroni et toutes celles qui en découle.

Modèle à effets fixes

Si l'on souhaite réaliser **un nombre fini k de comparaisons** de contrastes associés au facteur A on pourra utiliser la valeur critique

$$t(IJ(K-1); 1-\frac{\alpha}{2k})\sqrt{\frac{s_R^2}{JK}\left(\sum_{i=1}^{I} l_i^2\right)}$$

à la place de

$$q(I, IJ(K-1); 1-\alpha)\sqrt{\frac{s_R^2}{JK}}\left(\frac{1}{2}\sum_{i=1}^{I}|l_i|\right)$$

où $t(IJ(K-1); 1-\frac{\alpha}{2k})$ est le $100(1-\frac{\alpha}{2k})$ quantile de la loi de Student à $IJ(K-1)$ degrés de liberté.

Si l'on souhaite réaliser **un nombre fini k' de comparaisons** de contrastes associés au facteur B on pourra utiliser la valeur critique

$$t(IJ(K-1); 1-\frac{\alpha}{2k'})\sqrt{\frac{s_R^2}{IK}\left(\sum_{j=1}^{J} l_j'^2\right)}$$

à la place de

$$q(J, IJ(K-1); 1-\alpha)\sqrt{\frac{s_R^2}{IK}}\left(\frac{1}{2}\sum_{j=1}^{J}|l_j|\right)$$

où $t(IJ(K-1); 1-\frac{\alpha}{2k'})$ est le $100(1-\frac{\alpha}{2k'})$ quantile de la loi de Student à $IJ(K-1)$ degrés de liberté.

Modèle à effets mixtes

Si l'on souhaite réaliser **un nombre fini k de comparaisons** de contrastes associés au facteur A on pourra utiliser la valeur critique

$$t((I-1)(J-1); 1-\frac{\alpha}{2k})\sqrt{\frac{s_{AB}^2}{JK}\left(\sum_{i=1}^{I} l_i^2\right)}$$

à la place de

$$q(I, (I-1)(J-1); 1-\alpha)\sqrt{\frac{s_{AB}^2}{JK}}\left(\frac{1}{2}\sum_{i=1}^{I}|l_i|\right)$$

où $t((I-1)(J-1); 1-\frac{\alpha}{2k})$ est le $100(1-\frac{\alpha}{2k})$ quantile de la loi de Student à $(I-1)(J-1)$ degrés de liberté.

Ainsi dans le cas de comparaisons deux à deux des effets des modalités du facteur A à effets fixes la valeur critique que l'on utilisera avec la procédure de Bonferroni est :

$$\sqrt{2}\; t((I-1)(J-1); 1-\frac{\alpha}{2k})\sqrt{\frac{s_{AB}^2}{JK}}$$

où $t((I-1)(J-1); 1-\frac{\alpha}{2k})$ est le $100(1-\frac{\alpha}{2k})$ quantile de la loi de Student à $(I-1)(J-1)$ degrés de liberté.

10.7.5. Cas d'un plan non équilibré et des comparaisons deux à deux

Dans les deux situations, effets fixes ou mixtes, on peut utiliser la procédure de Tukey en remplaçant le nombre de répétitions K par la moyenne harmonique K_h du nombre de répétitions K_i et $K_{i'}$ effectuées dans chacune des deux conditions dont on souhaite comparer les effets. On rappelle l'expression de K_h :

$$K_h = \frac{1}{\frac{1}{2}\left(\frac{1}{K_i} + \frac{1}{K_{i'}}\right)}$$

10.8. Deux facteurs emboîtés

On peut adapter les méthodes de comparaison des effets des différentes modalités, exposées aux paragraphes 10.1, 10.2 et 10.3 dans le cas de l'analyse de la variance à un facteur à effets fixes, au cas où l'on considère un modèle à deux facteurs sans interaction à effets fixes ou à effets mixtes. On commence par détailler les modifications à faire pour utiliser les méthodes de Tukey et de Scheffé.

On reprend ici les notations du paragraphe 5.1.1 ou du paragraphe 5.3.1 en fonction de la situation considérée. La différence fondamentale entre ces deux situations résidant dans le fait que pour la première, **deux facteurs à effets fixes**, on utilisera le **carré moyen résiduel** s_R^2 dans les statistiques de test tandis que dans la seconde, **un facteur à effets fixes et un facteur à effets aléatoires**, on utilisera le **carré moyen associé au terme emboîté** $s_{B|A}^2$ dans les statistiques de test.

10.8.1. Contrastes

Modèle à effets fixes

Dans le cas d'un modèle où les deux facteurs sont à effets fixes, c'est-à-dire le cas envisagé au paragraphe 4.1.2, il peut être intéressant, en fonction de la significativité des différents tests auxquels on a procédé, d'étudier les constrastes exposés ci-dessous.

Le contraste met en jeu des moyennes associées au premier facteur A.

$$L = l_1\alpha_1 + l_2\alpha_2 + \cdots + l_I\alpha_I,$$

avec $\sum_{i=1}^{I} l_i = 0$.
Un estimateur sans biais de L est alors :

$$\widehat{L} = l_1\widehat{\alpha_1} + l_2\widehat{\alpha_2} + \cdots + l_I\widehat{\alpha_I}.$$

Ainsi une estimation sans biais de L s'écrit :

$$\widehat{L}(\boldsymbol{y}) = l_1 y_{1,\bullet,\bullet} + l_2 y_{2,\bullet,\bullet} + \cdots + l_I y_{I,\bullet,\bullet}.$$

Le contraste met en jeu des moyennes associées au second facteur B emboîté dans le premier facteur A.

$$L^{'} = l^{'}_{1(1)}\beta_{1(1)} + l^{'}_{2(1)}\beta_{1(1)} + \cdots + l^{'}_{J(I)}\beta_{J(I)},$$

avec $\sum_{i=1}^{I}\sum_{j=1}^{J} l^{'}_{j(i)} = 0$.
Un estimateur sans biais de $L^{'}$ est alors :

$$\widehat{L^{'}} = l^{'}_{1(1)}\widehat{\beta_{1(1)}} + l^{'}_{2(1)}\widehat{\beta_{2(1)}} + \cdots + l^{'}_{J(I)}\widehat{\beta_{J(I)}}.$$

Ainsi une estimation sans biais de $L^{'}$ s'écrit :

$$\widehat{L^{'}}(\boldsymbol{y}) = l^{'}_{1(1)} y_{1,1,\bullet} + l^{'}_{2(1)} y_{2,1,\bullet} + \cdots + l^{'}_{J(I)} y_{J,I,\bullet}.$$

Modèle à effets mixtes

Dans le cas d'un modèle mixte, c'est-à-dire le cas envisagé au paragraphe 4.3.2, il peut être intéressant, en fonction de la significativité du test de l'effet du facteur à effets fixes, d'étudier les constrastes du type :

$$L = l_1\alpha_1 + l_2\alpha_2 + \cdots + l_I\alpha_I,$$

avec $\sum_{i=1}^{I} l_i = 0$.
Un estimateur sans biais de L est alors :

$$\widehat{L} = l_1\widehat{\alpha_1} + l_2\widehat{\alpha_2} + \cdots + l_I\widehat{\alpha_I}.$$

Ainsi une estimation sans biais de $L^{'}$ s'écrit :

$$\widehat{L}(\boldsymbol{y}) = l_1 y_{1,\bullet,\bullet} + l_2 y_{2,\bullet,\bullet} + \cdots + l_I y_{I,\bullet,\bullet}.$$

10.8.2. Méthode de Tukey

Modèle à effets fixes

Si un contraste L met en jeu des moyennes associées au facteur A à effets fixes, on utilise l'estimation s_R^2 calculée pour le modèle incluant les deux facteurs et on change le nombre de degrés de liberté de la loi de l'étendue Studentisées en I et $IJ(K-1)$.
On décide que le contraste L est significativement différent de 0 au seuil α si :

$$\frac{\widehat{L}(\boldsymbol{y})}{\sqrt{\frac{s_R^2}{JK}}\left(\frac{1}{2}\sum_{i=1}^{I}|l_i|\right)} \geqslant q(I, IJ(K-1); 1-\alpha).$$

où $q(I, IJ(K-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de l'étendue Studentisée à I et $IJ(K-1)$ degrés de liberté. Sinon on ne peut rejeter l'hypothèse d'absence de différence entre L et 0 au seuil α.

De même si le contraste $L^{'}$ met en jeu des moyennes associées aux niveaux du facteur B pour un niveau i du facteur A fixé, on utilise l'estimation s_R^2 calculée pour le modèle incluant les deux facteurs et on change le nombre de degrés de liberté de la loi de l'étendue Studentisées en J et $IJ(K-1)$.
On décide que le contraste $L^{'}$ est significativement différent de 0 au seuil α si :

$$\frac{\widehat{L}^{'}(\boldsymbol{y})}{\sqrt{\frac{s_R^2}{K}}\left(\frac{1}{2}\sum_{j=1}^{J}\left|l_j^{'}\right|\right)} \geqslant q(J, IJ(K-1); 1-\alpha).$$

où $q(J, IJ(K-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de l'étendue Studentisée à J et $IJ(K-1)$ degrés de liberté. Sinon on ne peut rejeter l'hypothèse d'absence de différence entre $L^{'}$ et 0 au seuil α.

Modèle à effets mixtes

Si un contraste L met en jeu des moyennes associées au facteur A à effets fixes, on utilise l'estimation $s_{B|A}^2$ calculée pour le modèle incluant les deux facteurs et on change le nombre de degrés de liberté de la loi de l'étendue Studentisées en I et $I(J-1)$.
On décide que le contraste L est significativement différent de 0 au seuil α si :

$$\frac{\widehat{L}(\boldsymbol{y})}{\sqrt{\frac{s_{B|A}^2}{JK}}\left(\frac{1}{2}\sum_{i=1}^{I}|l_i|\right)} \geqslant q(I, I(J-1); 1-\alpha).$$

où $q(I, I(J-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de l'étendue Studentisée à I et $I(J-1)$ degrés de liberté. Sinon on ne peut rejeter l'hypothèse d'absence de différence entre L et 0 au seuil α.

10.8.3. Méthode de Scheffé

Modèle à effets fixes

L_{AB} est significativement différent de 0 au seuil α si, en supposant que $y_{i,j,\bullet} \geqslant y_{i',j',\bullet}$, on a :

$$\frac{\widehat{L}(\boldsymbol{y})}{\sqrt{(IJ-1)\frac{2s_R^2}{K}}} \geqslant \sqrt{F(IJ-1, IJ(K-1); 1-\alpha)},$$

où $F(IJ-1, IJ(K-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $IJ-1$ et $IJ(K-1)$ degrés de liberté. Sinon on ne peut rejeter l'hypothèse d'absence de différence entre L_{AB} et 0 au seuil α.

Si un contraste L met en jeu des moyennes associées au facteur A à effets fixes, on utilise l'estimation s_R^2 calculée pour le modèle incluant les deux facteurs et on change le nombre de degrés de liberté de la loi de Fisher en $I-1$ et $IJ(K-1)$.
On décide que le contraste L est significativement différent de 0 au seuil α si :

$$\frac{\widehat{L}(\boldsymbol{y})}{\sqrt{\dfrac{(I-1)s_R^2}{JK}\displaystyle\sum_{i=1}^{I} l_i^2}} \geqslant \sqrt{F(I-1, IJ(K-1); 1-\alpha)},$$

où $F(I-1, IJ(K-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $I-1$ et $IJ(K-1)$ degrés de liberté. Sinon on ne peut rejeter l'hypothèse d'absence de différence entre L et 0 au seuil α.

De même si le contraste $L^{'}$ met en jeu des moyennes associées aux niveaux du facteur B pour un niveau i du facteur A fixé, on utilise l'estimation s_R^2 calculée pour le modèle incluant les deux facteurs et on change le nombre de degrés de liberté de la loi de Fisher en $J-1$ et $IJ(K-1)$.
On décide que le contraste $L^{'}$ est significativement différent de 0 au seuil α si :

$$\frac{\widehat{L'}(\boldsymbol{y})}{\sqrt{\dfrac{(J-1)s_R^2}{K}\displaystyle\sum_{j=1}^{J} l_j'^2}} \geqslant \sqrt{F(J-1, IJ(K-1); 1-\alpha)},$$

où $F(J-1, IJ(K-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $J-1$ et $IJ(K-1)$ degrés de liberté. Sinon on ne peut rejeter l'hypothèse d'absence de différence entre $L^{'}$ et 0 au seuil α.

Modèle à effets mixtes

Si un contraste L met en jeu des moyennes associées au facteur A à effets fixes, on utilise l'estimation $s_{B|A}^2$ calculée pour le modèle incluant les deux facteurs et on change le nombre de degrés de liberté de la loi de Fisher en $I-1$ et $I(J-1)$.
On décide que le contraste L est significativement différent de 0 au seuil α si :

$$\frac{\widehat{L}(\boldsymbol{y})}{\sqrt{\dfrac{(I-1)s_{B|A}^2}{JK}\displaystyle\sum_{i=1}^{I} l_i^2}} \geqslant F(I-1, I(J-1); 1-\alpha).$$

où $F(I-1, I(J-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à I et $I(J-1)$ degrés de liberté. Sinon on ne peut rejeter l'hypothèse d'absence de différence entre L et 0 au seuil α.

10.8.4. Méthode de Bonferroni et méthodes associées

On adapte ici, de manière similaire, la procédure de Bonferroni et toutes celles qui en découle.

Modèle à effets fixes

Si l'on souhaite réaliser **un nombre fini k de comparaisons** de contrastes associés au facteur A on pourra utiliser la valeur critique

$$t(IJ(K-1); 1-\frac{\alpha}{2k})\sqrt{\frac{s_R^2}{JK}\left(\sum_{i=1}^{I} l_i^2\right)}$$

à la place de

$$q(I, IJ(K-1); 1-\alpha)\sqrt{\frac{s_R^2}{JK}}\left(\frac{1}{2}\sum_{i=1}^{I}|l_i|\right)$$

où $t(IJ(K-1); 1-\frac{\alpha}{2k})$ est le $100(1-\frac{\alpha}{2k})$ quantile de la loi de Student à $IJ(K-1)$ degrés de liberté.

Si l'on souhaite réaliser **un nombre fini k' de comparaisons** de contrastes associés aux niveaux du facteur B pour un niveau i du facteur A fixé on pourra utiliser la valeur critique

$$t(IJ(K-1); 1-\frac{\alpha}{2k'})\sqrt{\frac{s_R^2}{K}\left(\sum_{j=1}^{J} l_j'^2\right)}$$

à la place de

$$q(J, IJ(K-1); 1-\alpha)\sqrt{\frac{s_R^2}{K}}\left(\frac{1}{2}\sum_{j=1}^{J}|l_j|\right)$$

où $t(IJ(K-1); 1-\frac{\alpha}{2k'})$ est le $100(1-\frac{\alpha}{2k'})$ quantile de la loi de Student à $IJ(K-1)$ degrés de liberté.

Modèle à effets mixtes

Si l'on souhaite réaliser **un nombre fini k de comparaisons** de contrastes associés au facteur A on pourra utiliser la valeur critique

$$t(I(J-1); 1-\frac{\alpha}{2k})\sqrt{\frac{s^2_{B|A}}{JK}\left(\sum_{i=1}^{I} l_i^2\right)}$$

à la place de

$$q(I, I(J-1); 1-\alpha)\sqrt{\frac{s^2_{B|A}}{JK}}\left(\frac{1}{2}\sum_{i=1}^{I}|l_i|\right)$$

où $t(I(J-1); 1-\frac{\alpha}{2k})$ est le $100(1-\frac{\alpha}{2k})$ quantile de la loi de Student à $I(J-1)$ degrés de liberté.

Ainsi dans le cas de comparaisons deux à deux des effets des modalités du facteur A à effets fixes la valeur critique que l'on utilisera avec la procédure de Bonferroni est :

$$\sqrt{2}\, t(I(J-1); 1-\frac{\alpha}{2k})\sqrt{\frac{s^2_{B|A}}{JK}}$$

où $t(I(J-1); 1-\frac{\alpha}{2k})$ est le $100(1-\frac{\alpha}{2k})$ quantile de la loi de Student à $I(J-1)$ degrés de liberté.

10.8.5. Cas d'un plan non équilibré et des comparaisons deux à deux

Dans les deux situations, effets fixes ou mixtes, on peut utiliser la procédure de Tukey en remplaçant le nombre de répétitions K par la moyenne harmonique K_h du nombre de répétitions K_i et $K_{i'}$ effectuées dans chacune des deux conditions dont on souhaite comparer les effets. On rappelle l'expression de K_h :

$$K_h = \frac{1}{\frac{1}{2}\left(\frac{1}{K_i}+\frac{1}{K_{i'}}\right)}.$$

11. Puissance des tests de l'analyse de la variance

Les moyens de calculs désormais à la disposition de l'expérimentateur permettent de calculer directement, c'est-à-dire sans avoir à recourir à une table ou à un abaque, la puissance

des tests, dits F, de l'analyse de la variance. Si nécessaire, on pourra se réferrer à [4] pour consulter la plupart des tables ou abaques servant au calcul de la puissance d'un test dans un tableau d'analyse de la variance.

La puissance d'un test est une fonction complexe des différents paramètres de ce test. On souligne certaines relations qu'il faut toujours avoir à l'esprit :

- Les erreurs de première espèce et de deuxième espèce sont antagonistes, c'est-à-dire qu'elles évoluent en sens contraire. Si les autres paramètres du test sont fixés, plus on diminue le risque de commettre une erreur de première espèce, plus on augmente celui de commettre une erreur de deuxième espèce et réciproquement plus on diminue le risque de commettre une erreur de deuxième espèce, plus on augmente celui de commettre une erreur de première espèce.
- Plus les effets du facteur sont marqués, soit en terme d'amplitude des écarts entre les modalités d'un facteur à effets fixes ou soit en terme de dispersion des effets des modalités d'un facteur à effets aléatoires, plus la puissance est grande.
- Plus la dispersion de l'erreur est faible, plus la puissance est grande.
- Plus le nombre de répétitions effectuées pour une modalité du facteur est important plus la puissance est grande.

Parmi tous les paramètres ci-dessous, le seuil du test α étant fixé, il n'est possible pour l'expérimentateur que d'agir sur le nombre de répétitions effectuées pour une modalité du facteur. C'est pourquoi l'expérimentateur a un intérêt particulier à bien choisir ce paramètre lorsqu'il élabore le protocole expérimental dont il va servir s'il souhaite garantir un niveau de puissance correct pour les tests qu'il va réaliser.

11.1. Analyse de la variance à un facteur

11.1.1. Modèle à effets fixes

On reprend ici les notations du paragraphe 3.1. On s'intéresse à la puissance $1-\beta$, où β est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \text{Il existe } i_0 \in \{1, 2, \ldots, I\} \text{ tel que } \alpha_{i_0} \neq 0.}$$

Cette puissance $1-\beta$ est donnée par la formule suivante :

$$1-\beta = \mathbb{P}\left[F^{'}(I-1, I(J-1); \lambda) > F(I-1, I(J-1); 1-\alpha)\right],$$

où $F(I-1, I(J-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $I-1$ et $I(J-1)$ degrés de liberté et $F^{'}(I-1, I(J-1); \lambda)$ est une variable aléatoire qui suit une loi de

Fisher non-centrale à $I-1$ et $I(J-1)$ degrés de liberté et de paramètre de non-centralité λ. Ce paramètre de non-centralité λ vaut :

$$\lambda = \frac{J}{2\sigma^2}\sum_{i=1}^{I}\alpha_i^2.$$

Lorsque l'on utilise une loi de Fisher non centrale à ν_1 et ν_2 degrés de liberté et de paramètre de non-centralité λ, on introduit souvent le paramètre de non-centralité normalisé ϕ défini par :

$$\phi = \sqrt{\frac{2\lambda}{v_1+1}}.$$

On obtient ainsi dans notre situation :

$$\begin{aligned}\phi &= \sqrt{\frac{2\lambda}{v_1+1}}\\ &= \sqrt{\frac{2\frac{J}{2\sigma^2}\sum_{i=1}^{I}\alpha_i^2}{I-1+1}}\\ &= \frac{1}{\sigma}\sqrt{\frac{J}{I}\sum_{i=1}^{I}\alpha_i^2}.\end{aligned}$$

Si le nombre de répétitions n_i effectué pour chaque modalité A_i du facteur A n'est pas constant, c'est-à-dire si le plan expérimental n'est pas équilibré, le paramètre de non-centralité λ devient :

$$\lambda = \frac{1}{2\sigma^2}\sum_{i=1}^{I}n_i\alpha_i^2.$$

Le paramètre de non-centralité normalisé ϕ est alors :

$$\phi = \frac{1}{\sigma}\sqrt{\frac{1}{I}\sum_{i=1}^{I}n_i\alpha_i^2}.$$

On constate que la puissance évolue bien comme indiqué au début de cette lorsque α_i, $1 \leqslant i \leqslant I$, I, n_i, $1 \leqslant i \leqslant I$, σ et α varient.

Puissance *a posteriori*

On obtient la puissance ***a posteriori*** du test de l'absence d'effet du facteur A en remplaçant dans la formule appropriée ci-dessus. Le choix se fait en fonction du fait que le

plan expérimental est équilibré ou non, les valeurs des paramètres par les estimations que l'on a obtenues en réalisant l'analyse de la variance. Généralement on considère qu'une puissance de 0,8 est satisfaisante et qu'alors la décision de ne pas rejeter l'hypothèse nulle $\mathcal{H}_0$ est « vraiment » associée à l'absence d'effet du facteur considéré.

Détermination du nombre de répétitions

Une autre approche, celle développée lorsque l'on fait de la planification expérimentale, serait de déterminer *a priori* le nombre de répétitions J nécessaires pour obtenir une valeur de puissance du test supérieure à un niveau fixé à l'avance. L'intérêt de cette démarche réside dans le fait que l'on ne connaît pas a priori si le test que l'on va réaliser une fois que les expériences ont été réalisées sera significatif ou non à un seuil α % fixé à l'avance. Le fait de ne pas rejeter l'hypothèse nulle en ayant un risque élevé de commettre une erreur de deuxième rendrait cette décision très peu fiable et ne permettrait pas de conclure avec une confiance suffisante à l'absence d'un effet du facteur étudié sur la réponse. C'est pourquoi dans de nombreux domaines comme les études cliniques, où les expériences peuvent durer plusieurs années, il est primordial de s'assurer que si une différence existe il y aura un faible risque de ne pas la mettre en évidence. Généralement on considère qu'une puissance de 0,8 est satisfaisante ; dans certains cas on vise même une puissance de 0,9.

On peut utiliser directement la formule ci-dessus pour déterminer le nombre de répétitions nécessaires à l'obtention d'un valeur minimale de puissance. Il faut néanmoins avoir une idée de la valeur minimale que peut prendre la somme $\sum_{i=1}^{I} n_i\alpha_i^2$ et la valeur maximale que peut avoir σ^2. Ces valeurs doivent être déterminées par un expert du domaine considéré.

Détermination du nombre de répétitions à l'aide de la plus petite différence détectable

Dans ce type d'étude prospective, la situation est compliquée par le fait qu'il est difficile d'évaluer le terme $\sum_{i=1}^{I} \alpha_i^2$. On introduit alors le concept de plus petite différence détectable Δ, ce qui revient à évaluer le sensibilité du test en terme d'amplitude entre les effets des différents niveaux du facteur étudié. Ainsi on cherchera à ce que la probabilité de détecter une amplitude $|\alpha_i - \alpha_j|$ entre les effets α_i et α_j de deux modalités A_i et A_j différentes du facteur étudié strictement supérieure à Δ soit élevée.

Ainsi pour faire le calcul de la puissance on se place dans le pire des cas, c'est-à-dire celui pour lequel tous les effets sont nuls sauf deux α_{i_0} et α_{j_0} pour lesquels il existe un écart en valeur absolue égal à Δ. Alors $|\alpha_{i_0}| = |\alpha_{j_0}| = \Delta/2$. On obtient alors :

$$\begin{aligned}\lambda &= \frac{J}{2\sigma^2}\sum_{i=1}^{I}\alpha_i^2 \\ &= \frac{J}{2\sigma^2}\left(\alpha_{i_0}^2 + \alpha_{j_0}^2\right)\end{aligned}$$

$$
\begin{aligned}
&= \frac{J}{2\sigma^2}\left(\left(\frac{\Delta}{2}\right)^2 + \left(\frac{\Delta}{2}\right)^2\right) \\
&= \frac{J}{4\sigma^2}\Delta^2.
\end{aligned}
$$

On utilise la formule ci-dessus pour déterminer les valeurs de J pour lesquelles la puissance $1-\beta$ est supérieure à une valeur $1-\beta_0$ fixée à l'avance, généralement 0,8 soit 80 %. Remarquons que là encore il est nécessaire de connaître σ^2 ou au moins d'avoir une idée précise de la valeur de ce paramètre ce qui n'est malheureusement généralement pas le cas. Dans cette situation on considère plutôt le paramètre de sensibilité Δ/σ à la place de Δ.

11.1.2. Modèle à effets aléatoires

On reprend ici les notations du paragraphe 3.2. On s'intéresse à la puissance $1-\beta$, où β est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \sigma_A^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \sigma_A^2 \neq 0.}$$

Cette puissance $1-\beta$ est donnée par la formule suivante :

$$1-\beta = \mathbb{P}\left[F(I-1, I(J-1)) > \frac{F(I-1, I(J-1); 1-\alpha)}{1 + J\dfrac{\sigma_A^2}{\sigma^2}}\right],$$

où $F(I-1, I(J-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $I-1$ et $I(J-1)$ degrés de liberté et $F(I-1, I(J-1))$ est une variable aléatoire qui suit une loi de Fisher à $I-1$ et $I(J-1)$ degrés de liberté.

La différence fondamentale entre ce cas et le cas où le facteur est à effets fixes, exposés au paragraphe 11.1.1, est que le calcul de la puissance repose une loi de Fisher et non sur une loi de Fisher non-centrale.

Puissance *a posteriori*

On obtient alors la puissance ***a posteriori*** du test de l'absence d'effet du facteur A en remplaçant dans la formule appropriée ci-dessus, le choix se fait en fonction du fait que le plan expérimental est équilibré ou non, les valeurs paramètres par les estimations que l'on a obtenues en réalisant l'analyse de la variance. Généralement on considère qu'une puissance de 0,8 est satisfaisante et qu'alors la décision de ne pas rejeter l'hypothèse nulle $\mathcal{H}_0$ est « vraiment » associée à l'absence d'effet du facteur considéré.

11.2. Analyse de la variance à deux facteurs sans interaction

Les idées développées dans la section 11.1 précédente sur le calcul de la puissance *a posteriori* ou la détermination du nombre de répétitions minimal pour obtenir un niveau de puissance supérieur ou égal à une valeur cible $1 - \beta_0$ sont directement transférables aux tests étudiés dans cette section à condition de remplacer les formules par celles qui sont exposées ci-dessous.

La puissance calculée pour chacun des tests ne dépend pas du risque de première espèce fixé pour les autres tests. Ceci permettrait de calculer la puissance des tests du tableau de l'analyse de la variance même si le seuil de chaque test varie en fonction du test considéré. Voir le paragraphe 2.3 pour plus de détails sur la gestion des risques de première espèce des tests du tableau de l'analyse de la variance.

11.2.1. Modèle à effets fixes

On reprend ici les notations du paragraphe 4.1.1.

On s'intéresse à la puissance $1 - \beta_A$, où β_A est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \text{Il existe } i_0 \in \{1, 2, \ldots, I\} \text{ tel que } \alpha_{i_0} \neq 0.}$$

Cette puissance $1 - \beta_A$ est donnée par la formule suivante :

$$1 - \beta_A = \mathbb{P}\left[F'(I-1, (I-1)(J-1); \phi_A) > F(I-1, (I-1)(J-1); 1-\alpha)\right],$$

où $F(I-1, (I-1)(J-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $I-1$ et $(I-1)(J-1)$ degrés de liberté et $F'(I-1, (I-1)(J-1); \phi_A)$ est une variable aléatoire qui suit une loi de Fisher non-centrale à $I-1$ et $(I-1)(J-1)$ degrés de liberté et de paramètre de non-centralité normalisé ϕ_A. Ce paramètre de non-centralité normalisé ϕ_A vaut :

$$\phi_A = \frac{1}{\sigma}\sqrt{\frac{J}{I}\sum_{i=1}^{I}\alpha_i^2}.$$

On s'intéresse à la puissance $1 - \beta_B$, où β_B est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \beta_1 = \beta_2 = \cdots = \beta_J = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \text{Il existe } j_0 \in \{1, 2, \ldots, J\} \text{ tel que } \beta_{j_0} \neq 0.}$$

Cette puissance $1-\beta_B$ est donnée par la formule suivante :

$$1-\beta_B = \mathbb{P}\left[F^{'}(J-1,(I-1)(J-1);\phi_B) > F(J-1,(I-1)(J-1);1-\alpha)\right],$$

où $F(J-1,(I-1)(J-1);1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $J-1$ et $(I-1)(J-1)$ degrés de liberté et $F^{'}(J-1,(I-1)(J-1);\phi_B)$ est une variable aléatoire qui suit une loi de Fisher non-centrale à $J-1$ et $(I-1)(J-1)$ degrés de liberté et de paramètre de non-centralité normalisé ϕ_B. Ce paramètre de non-centralité normalisé ϕ_B vaut :

$$\phi_B = \frac{1}{\sigma}\sqrt{\frac{I}{J}\sum_{j=1}^{J}\beta_j^2}.$$

11.2.2. Modèle à effets aléatoires

On reprend ici les notations du paragraphe 4.2.1. On s'intéresse à la puissance $1-\beta_A$, où β_A est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \sigma_A^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \sigma_A^2 \neq 0.}$$

Cette puissance $1-\beta_A$ est donnée par la formule suivante :

$$1-\beta_A = \mathbb{P}\left[F(I-1,(I-1)(J-1)) > \frac{F(I-1,(I-1)(J-1);1-\alpha)}{1+J\dfrac{\sigma_A^2}{\sigma^2}}\right],$$

où $F(I-1,(I-1)(J-1);1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $I-1$ et $(I-1)(J-1)$ degrés de liberté et $F(I-1,(I-1)(J-1))$ est une variable aléatoire qui suit une loi de Fisher à $I-1$ et $(I-1)(J-1)$ degrés de liberté.

On s'intéresse à la puissance $1-\beta_B$, où β_B est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \sigma_B^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \sigma_B^2 \neq 0.}$$

Cette puissance $1-\beta_B$ est donnée par la formule suivante :

$$1-\beta_B = \mathbb{P}\left[F(J-1,(I-1)(J-1)) > \frac{F(J-1,(I-1)(J-1);1-\alpha)}{1+I\dfrac{\sigma_B^2}{\sigma^2}}\right],$$

où $F(J-1,(I-1)(J-1);1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $(J-1)$ et $(I-1)(J-1)$ degrés de liberté et $F(J-1,(I-1)(J-1))$ est une variable aléatoire qui suit une loi de Fisher à $J-1$ et $(I-1)(J-1)$ degrés de liberté.

La différence fondamentale entre ce cas et le cas où le facteur est à effets fixes, exposé au paragraphe 4.2.1, est que le calcul de la puissance repose une loi de Fisher et non sur une loi de Fisher non-centrale.

11.2.3. Modèle à effets mixtes

On reprend ici les notations du paragraphe 4.3.1. On s'intéresse à la puissance $1-\beta_A$, où β_A est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \text{Il existe } i_0 \in \{1,2,\ldots,I\} \text{ tel que } \alpha_{i_0} \neq 0.}$$

Cette puissance $1-\beta_A$ est donnée par la formule suivante :

$$1-\beta_A = \mathbb{P}\left[F'(I-1,(I-1)(J-1);\phi_A) > F(I-1,(I-1)(J-1);1-\alpha)\right],$$

où $F(I-1,(I-1)(J-1);1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $I-1$ et $(I-1)(J-1)$ degrés de liberté et $F'(I-1,I(J-1);\phi_A)$ est une variable aléatoire qui suit une loi de Fisher non-centrale à $I-1$ et $(I-1)(J-1)$ degrés de liberté et de paramètre de non-centralité normalisé ϕ_A. Ce paramètre de non-centralité normalisé ϕ_A vaut :

$$\phi_A = \frac{1}{\sigma}\sqrt{\frac{J}{I}\sum_{i=1}^{I}\alpha_i^2}.$$

On s'intéresse à la puissance $1-\beta_B$, où β_B est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \sigma_B^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \sigma_B^2 \neq 0.}$$

Cette puissance $1-\beta_B$ est donnée par la formule suivante :

$$1-\beta_B = \mathbb{P}\left[F(J-1,(I-1)(J-1)) > \frac{F(J-1,(I-1)(J-1);1-\alpha)}{1+I\dfrac{\sigma_B^2}{\sigma^2}}\right],$$

où $F(J-1,(I-1)(J-1);1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $(J-1)$ et $(I-1)(J-1)$ degrés de liberté et $F(J-1,(I-1)(J-1))$ est une variable aléatoire qui suit une loi de Fisher à $J-1$ et $(I-1)(J-1)$ degrés de liberté.

Les puissances des tests sont donc identiques à celles où le facteur à effets fixes seraient avec un autre facteur à effets fixes et le facteur à effets aléatoires serait avec un autre facteur à effets aléatoires.

11.3. Analyse de la variance à deux facteurs avec interaction

Les idées développées dans la section 11.1 précédente sur le calcul de la puissance *a posteriori* ou la détermination du nombre de répétitions minimal pour obtenir un niveau de puissance supérieur ou égal à une valeur cible $1-\beta_0$ sont directement transférables aux tests étudiés dans cette section à condition de remplacer les formules par celles qui sont exposées ci-dessous.

La puissance calculée pour chacun des tests ne dépend pas du risque de première espèce fixé pour les autres tests. Ceci permettrait de calculer la puissance des tests du tableau de l'analyse de la variance même si le seuil de chaque test varie en fonction du test considéré. Voir le paragraphe 2.3 pour plus de détails sur la gestion des risques de première espèce des tests du tableau de l'analyse de la variance.

11.3.1. Modèle à effets fixes

On reprend ici les notations du paragraphe 4.1.2.

On s'intéresse à la puissance $1-\beta_A$, où β_A est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \text{Il existe } i_0 \in \{1,2,\ldots,I\} \text{ tel que } \alpha_{i_0} \neq 0.}$$

Cette puissance $1-\beta_A$ est donnée par la formule suivante :

$$1-\beta_A = \mathbb{P}\left[F'(I-1,IJ(K-1);\phi_A) > F(I-1,IJ(K-1);1-\alpha)\right],$$

où $F(I-1, IJ(K-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $I-1$ et $IJ(K-1)$ degrés de liberté et $F^{'}(I-1, IJ(K-1); \phi_A)$ est une variable aléatoire qui suit une loi de Fisher non-centrale à $I-1$ et $IJ(K-1)$ degrés de liberté et de paramètre de non-centralité normalisé ϕ_A. Ce paramètre de non-centralité normalisé ϕ_A vaut :

$$\phi_A = \frac{1}{\sigma}\sqrt{\frac{JK}{I}\sum_{i=1}^{I}\alpha_i^2}.$$

On s'intéresse à la puissance $1-\beta_B$, où β_B est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \beta_1 = \beta_2 = \cdots = \beta_J = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \text{Il existe } j_0 \in \{1, 2, \ldots, J\} \text{ tel que } \beta_{j_0} \neq 0.}$$

Cette puissance $1-\beta_B$ est donnée par la formule suivante :

$$1-\beta_B = \mathbb{P}\left[F^{'}(J-1, IJ(K-1); \phi_B) > F(J-1, IJ(K-1); 1-\alpha)\right],$$

où $F(J-1, IJ(K-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $J-1$ et $IJ(K-1)$ degrés de liberté et $F^{'}(J-1, IJ(K-1); \phi_B)$ est une variable aléatoire qui suit une loi de Fisher non-centrale à $J-1$ et $IJ(K-1)$ degrés de liberté et de paramètre de non-centralité normalisé ϕ_B. Ce paramètre de non-centralité normalisé ϕ_B vaut :

$$\phi_B = \frac{1}{\sigma}\sqrt{\frac{IK}{J}\sum_{j=1}^{J}\beta_j^2}.$$

On s'intéresse à la puissance $1-\beta_{AB}$, où β_{AB} est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : (\alpha\beta)_{1,1} = (\alpha\beta)_{1,2} = \cdots = (\alpha\beta)_{1,J} = (\alpha\beta)_{2,1} = \cdots = (\alpha\beta)_{I,J} = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \text{Il existe } (i_0, j_0) \in \{1, 2, \ldots, I\} \times \{1, 2, \ldots, J\} \text{ tel que } (\alpha\beta)_{i_0,j_0} \neq 0.}$$

Cette puissance $1-\beta_{AB}$ est donnée par la formule suivante :

$$1-\beta_{AB} = \mathbb{P}\left[F^{'}((I-1)(J-1), IJ(K-1); \phi_{AB}) > F((I-1)(J-1), IJ(K-1); 1-\alpha)\right],$$

où $F((I-1)(J-1), IJ(K-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $(I-1)(J-1)$ et $IJ(K-1)$ degrés de liberté et $F^{'}((I-1)(J-1), IJ(K-1); \phi_{AB})$ est une variable aléatoire qui suit une loi de Fisher non-centrale à $(I-1)(J-1)$ et $IJ(K-1)$

degrés de liberté et de paramètre de non-centralité normalisé ϕ_{AB}. Ce paramètre de non-centralité normalisé ϕ_{AB} vaut :

$$\phi_{AB} = \frac{1}{\sigma}\sqrt{\frac{K\sum_{i=1}^{I}\sum_{j=1}^{J}(\alpha\beta)_{i,j}^2}{(I-1)(J-1)+1}}.$$

11.3.2. Modèle à effets aléatoires

On reprend ici les notations du paragraphe 4.2.2.

On s'intéresse à la puissance $1-\beta_A$, où β_A est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \sigma_A^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \sigma_A^2 \neq 0.}$$

Cette puissance $1-\beta_A$ est donnée par la formule suivante :

$$1-\beta_A = \mathbb{P}\left[F(I-1,(I-1)(J-1)) > \frac{F(I-1,(I-1)(J-1);1-\alpha)}{1+JK\dfrac{\sigma_A^2}{\sigma^2+K\sigma_{AB}^2}}\right],$$

où $F(I-1,(I-1)(J-1);1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $(I-1)$ et $(I-1)(J-1)$ degrés de liberté et $F(I-1,(I-1)(J-1))$ est une variable aléatoire qui suit une loi de Fisher à $I-1$ et $(I-1)(J-1)$ degrés de liberté.

On s'intéresse à la puissance $1-\beta_B$, où β_B est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \sigma_B^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \sigma_B^2 \neq 0.}$$

Cette puissance $1-\beta_B$ est donnée par la formule suivante :

$$1-\beta_B = \mathbb{P}\left[F(J-1,(I-1)(J-1)) > \frac{F(J-1,(I-1)(J-1);1-\alpha)}{1+IK\dfrac{\sigma_B^2}{\sigma^2+K\sigma_A^2}}\right],$$

où $F(J-1,(I-1)(J-1);1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $(J-1)$ et $(I-1)(J-1)$ degrés de liberté et $F(J-1,(I-1)(J-1))$ est une variable aléatoire qui suit une loi de Fisher à $J-1$ et $(I-1)(J-1)$ degrés de liberté.

On s'intéresse à la puissance $1-\beta_{AB}$, où β_{AB} est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \sigma_{AB}^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \sigma_{AB}^2 \neq 0.}$$

Cette puissance $1-\beta_{AB}$ est donnée par la formule suivante :

$$1-\beta_{AB} = \mathbb{P}\left[F((I-1)(J-1), IJ(K-1)) > \frac{F((I-1)(J-1), IJ(K-1); 1-\alpha)}{1+K\dfrac{\sigma_{AB}^2}{\sigma^2}}\right],$$

où $F((I-1)(J-1), IJ(K-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $(I-1)(J-1)$ et $IJ(K-1)$ degrés de liberté et $F((I-1)(J-1), IJ(K-1))$ est une variable aléatoire qui suit une loi de Fisher à $(I-1)(J-1)$ et $IJ(K-1)$ degrés de liberté.

La différence fondamentale entre ce cas et le cas où le facteur est à effets fixes, exposé au paragraphe 4.2.1, est que le calcul de la puissance repose une loi de Fisher et non sur une loi de Fisher non-centrale.

11.3.3. Modèle à effets mixtes

On reprend ici les notations du paragraphe 4.3.2.

On s'intéresse à la puissance $1-\beta_A$, où β_A est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \text{Il existe } i_0 \in \{1, 2, \ldots, I\} \text{ tel que } \alpha_{i_0} \neq 0.}$$

Cette puissance $1-\beta_A$ est donnée par la formule suivante :

$$1-\beta_A = \mathbb{P}\left[F'(I-1,(I-1)(J-1);\phi_A) > F(I-1,(I-1)(J-1);1-\alpha)\right],$$

où $F(I-1,(I-1)(J-1);1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $I-1$ et $(I-1)(J-1)$ degrés de liberté et $F'(I-1, I(J-1);\phi_A)$ est une variable aléatoire qui suit

une loi de Fisher non-centrale à $I-1$ et $(I-1)(J-1)$ degrés de liberté et de paramètre de non-centralité normalisé ϕ_A. Ce paramètre de non-centralité normalisé ϕ_A vaut :

$$\phi_A = \sqrt{\frac{JK\sum_{i=1}^{I}\alpha_i^2}{I\left(\sigma^2 + K\sigma_{AB}^2\right)}}.$$

On s'intéresse à la puissance $1-\beta_B$, où β_B est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \sigma_B^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \sigma_B^2 \neq 0.}$$

Cette puissance $1-\beta_B$ est donnée par la formule suivante :

$$1-\beta_B = \mathbb{P}\left[F(J-1, IJ(K-1)) > \frac{F(J-1, IJ(K-1); 1-\alpha)}{1 + IK\dfrac{\sigma_B^2}{\sigma^2 + K\sigma_A^2}}\right],$$

où $F(J-1, IJ(K-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $(J-1)$ et $IJ(K-1)$ degrés de liberté et $F(J-1, IJ(K-1))$ est une variable aléatoire qui suit une loi de Fisher à $J-1$ et $IJ(K-1)$ degrés de liberté.

On s'intéresse à la puissance $1-\beta_{AB}$, où β_{AB} est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \sigma_{AB}^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \sigma_{AB}^2 \neq 0.}$$

Cette puissance $1-\beta_{AB}$ est donnée par la formule suivante :

$$1-\beta_{AB} = \mathbb{P}\left[F((I-1)(J-1), IJ(K-1)) > \frac{F((I-1)(J-1), IJ(K-1); 1-\alpha)}{1 + K\dfrac{\sigma_{AB}^2}{\sigma^2}}\right],$$

où $F((I-1)(J-1), IJ(K-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $(I-1)(J-1)$ et $IJ(K-1)$ degrés de liberté et $F((I-1)(J-1), IJ(K-1))$ est une variable aléatoire qui suit une loi de Fisher à $(I-1)(J-1)$ et $IJ(K-1)$ degrés de liberté.

Les puissances des tests sont donc identiques à celles où le facteur à effets fixes serait avec un autre facteur à effets fixes, le facteur à effets aléatoires serait avec un autre facteur à effets aléatoires et l'interaction dans un modèle où les deux facteurs seraient aléatoires.

11.4. Analyse de la variance à deux facteurs emboîtés

Les idées développées dans la section 11.1 précédente sur le calcul de la puissance *a posteriori* ou la détermination du nombre de répétitions minimal pour obtenir un niveau de puissance supérieur ou égal à une valeur cible $1-\beta_0$ sont directement transférables aux tests étudiés dans cette section à condition de remplacer les formules par celles qui sont exposées ci-dessous.

La puissance calculée pour chacun des tests ne dépend pas du risque de première espèce fixé pour les autres tests. Ceci permettrait de calculer la puissance des tests du tableau de l'analyse de la variance même si le seuil de chaque test varie en fonction du test considéré. Voir le paragraphe 2.3 pour plus de détails sur la gestion des risques de première espèce des tests du tableau de l'analyse de la variance.

11.4.1. Modèle à effets fixes

On reprend ici les notations du paragraphe 5.1.1. On s'intéresse à la puissance $1-\beta_A$, où β_A est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \text{Il existe } i_0 \in \{1, 2, \ldots, I\} \text{ tel que } \alpha_{i_0} \neq 0.}$$

Cette puissance $1-\beta_A$ est donnée par la formule suivante :

$$1-\beta_A = \mathbb{P}\left[F^{'}(I-1, IJ(K-1); \phi_A) > F(I-1, IJ(K-1); 1-\alpha)\right],$$

où $F(I-1, IJ(K-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $I-1$ et $IJ(K-1)$ degrés de liberté et $F^{'}(I-1, IJ(K-1); \phi_A)$ est une variable aléatoire qui suit une loi de Fisher non-centrale à $I-1$ et $IJ(K-1)$ degrés de liberté et de paramètre de non-centralité normalisé ϕ_A. Ce paramètre de non-centralité normalisé ϕ_A vaut :

$$\phi_A = \frac{1}{\sigma}\sqrt{\frac{JK}{I}\sum_{i=1}^{I}\alpha_i^2}.$$

On s'intéresse à la puissance $1-\beta_{B(A)}$, où $\beta_{B(A)}$ est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \beta_{1(1)} = \beta_{2(1)} = \cdots = \beta_{J(1)} = \beta_{1(2)} = \cdots = \beta_{J(I)} = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \text{Il existe } (i_0, j_0) \in \{1, 2, \ldots, I\} \times \{1, 2, \ldots, J\} \text{ tel que } \beta_{j_0(i_0)} \neq 0.}$$

Cette puissance $1-\beta_{B(A)}$ est donnée par la formule suivante :

$$1-\beta_{B(A)} = \mathbb{P}\left[F^{'}(I(J-1), IJ(K-1); \phi_{B(A)}) > F(I(J-1), IJ(K-1); 1-\alpha)\right],$$

où $F(I(J-1), IJ(K-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $I(J-1)$ et $IJ(K-1)$ degrés de liberté et $F^{'}(I(J-1), IJ(K-1); \phi_{B(A)})$ est une variable aléatoire qui suit une loi de Fisher non-centrale à $I(J-1)$ et $IJ(K-1)$ degrés de liberté et de paramètre de non-centralité normalisé $\phi_{B(A)}$. Ce paramètre de non-centralité normalisé $\phi_{B(A)}$ vaut :

$$\phi_{B(A)} = \frac{1}{\sigma}\sqrt{\frac{K\sum_{i=1}^{I}\sum_{j=1}^{J}\beta_{j(i)}^2}{I(J-1)+1}}.$$

11.4.2. Modèle à effets aléatoires

On reprend ici les notations du paragraphe 5.2.1.

On s'intéresse à la puissance $1-\beta_A$, où β_A est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \sigma_A^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \sigma_A^2 \neq 0.}$$

Cette puissance $1-\beta_A$ est donnée par la formule suivante :

$$1-\beta_A = \mathbb{P}\left[F(I-1, I(J-1)) > \frac{F(I-1, I(J-1); 1-\alpha)}{1+JK\dfrac{\sigma_A^2}{\sigma^2+K\sigma_{B|A}^2}}\right],$$

où $F(I-1, I(J-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $(I-1)$ et $I(J-1)$ degrés de liberté et $F(I-1, I(J-1))$ est une variable aléatoire qui suit une loi de Fisher à $I-1$ et $I(J-1)$ degrés de liberté.

On s'intéresse à la puissance $1-\beta_{B(A)}$, où $\beta_{B(A)}$ est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \sigma_{B|A}^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \sigma_{B|A}^2 \neq 0.}$$

Cette puissance $1-\beta_{B(A)}$ est donnée par la formule suivante :

$$1-\beta_{B(A)} = \mathbb{P}\left[F(I(J-1), IJ(K-1)) > \frac{F(I(J-1), IJ(K-1); 1-\alpha)}{1+K\dfrac{\sigma^2_{B|A}}{\sigma^2}}\right],$$

où $F(I(J-1), IJ(K-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $I(J-1)$ et $IJ(K-1)$ degrés de liberté et $F(I(J-1), IJ(K-1))$ est une variable aléatoire qui suit une loi de Fisher à $I(J-1)$ et $IJ(K-1)$ degrés de liberté.

11.4.3. Modèle à effets mixtes

On reprend ici les notations du paragraphe 5.3.1.

On s'intéresse à la puissance $1-\beta_A$, où β_A est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \text{Il existe } i_0 \in \{1, 2, \ldots, I\} \text{ tel que } \alpha_{i_0} \neq 0.}$$

Cette puissance $1-\beta_A$ est donnée par la formule suivante :

$$1-\beta_A = \mathbb{P}\left[F'(I-1, I(J-1); \phi_A) > F(I-1, I(J-1); 1-\alpha)\right],$$

où $F(I-1, I(J-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $I-1$ et $I(J-1)$ degrés de liberté et $F'(I-1, I(J-1); \phi_A)$ est une variable aléatoire qui suit une loi de Fisher non-centrale à $I-1$ et $I(J-1)$ degrés de liberté et de paramètre de non-centralité normalisé ϕ_A. Ce paramètre de non-centralité normalisé ϕ_A vaut :

$$\phi_A = \sqrt{\frac{JK\sum_{i=1}^{I}\alpha_i^2}{I\left(\sigma^2 + K\sigma^2_{B|A}\right)}}.$$

On s'intéresse à la puissance $1-\beta_{B(A)}$, où $\beta_{B(A)}$ est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \sigma^2_{B|A} = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \sigma^2_{B|A} \neq 0.}$$

Cette puissance $1-\beta_{B(A)}$ est donnée par la formule suivante :

$$1-\beta_{B(A)} = \mathbb{P}\left[F(I(J-1), IJ(K-1)) > \frac{F(I(J-1), IJ(K-1); 1-\alpha)}{1+K\dfrac{\sigma^2_{B|A}}{\sigma^2}}\right],$$

où $F(I(J-1), IJ(K-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $I(J-1)$ et $IJ(K-1)$ degrés de liberté et $F(I(J-1), IJ(K-1))$ est une variable aléatoire qui suit une loi de Fisher à $I(J-1)$ et $IJ(K-1)$ degrés de liberté.

11.5. Analyse de la variance à trois facteurs sans interaction

Les idées développées dans la section 11.1 précédente sur le calcul de la puissance *a posteriori* ou la détermination du nombre de répétitions minimal pour obtenir un niveau de puissance supérieur ou égal à une valeur cible $1-\beta_0$ sont directement transférables aux tests étudiés dans cette section à condition de remplacer les formules par celles qui sont exposées ci-dessous.

La puissance calculée pour chacun des tests ne dépend pas du risque de première espèce fixé pour les autres tests. Ceci permettrait de calculer la puissance des tests du tableau de l'analyse de la variance même si le seuil de chaque test varie en fonction du test considéré. Voir le paragraphe 2.3 pour plus de détails sur la gestion des risques de première espèce des tests du tableau de l'analyse de la variance.

11.5.1. Modèle à effets fixes

On reprend ici les notations du paragraphe 6.1.1.

Pour des problèmes de concision on ne présentera que la puissance d'un type de test, effet principal, interaction d'ordre 1, interaction d'ordre 2, les valeurs de puissance associées aux tests restants s'en déduisant en remplaçant les quantités associées aux facteurs mis en jeu par celles associées aux facteurs pour lesquels on veut calculer la puissance du test.

On s'intéresse à la puissance $1-\beta_A$, où β_A est le risque de commettre une erreur de deuxième espèce, du pseudo test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \text{Il existe } i_0 \in \{1, 2, \ldots, I\} \text{ tel que } \alpha_{i_0} \neq 0.}$$

Cette puissance $1-\beta_A$ est donnée par la formule suivante :

$$1-\beta_A = \mathbb{P}\left[F^{'}(\nu_1,\nu_2;\phi_A) > F(\nu_1,\nu_2;1-\alpha)\right],$$

où $\nu_1 = I-1$ et $\nu_2 = (I-1)(J-1)(K-1)$, $F(I-1,(I-1)(J-1)(K-1);1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $I-1$ et $(I-1)(J-1)(K-1)$ degrés de liberté et $F^{'}(I-1,(I-1)(J-1)(K-1);\phi_A)$ est une variable aléatoire qui suit une loi de Fisher non-centrale à $I-1$ et $(I-1)(J-1)(K-1)$ degrés de liberté et de paramètre de non-centralité normalisé ϕ_A. Ce paramètre de non-centralité normalisé ϕ_A vaut :

$$\phi_A = \frac{1}{\sigma}\sqrt{\frac{JK}{I}\sum_{i=1}^{I}\alpha_i^2}.$$

On s'intéresse à la puissance $1-\beta_{AB}$, où β_{AB} est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : (\alpha\beta)_{1,1} = (\alpha\beta)_{1,2} = \cdots = (\alpha\beta)_{1,J} = (\alpha\beta)_{2,1} = \cdots = (\alpha\beta)_{I,J} = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \text{Il existe } (i_0,j_0) \in \{1,2,\ldots,I\}\times\{1,2,\ldots,J\} \text{ tel que } (\alpha\beta)_{i_0,j_0} \neq 0.}$$

Cette puissance $1-\beta_{AB}$ est donnée par la formule suivante :

$$1-\beta_{AB} = \mathbb{P}\left[F^{'}(\nu_1,\nu_2;\phi_{AB}) > F(\nu_1,\nu_2;1-\alpha)\right],$$

où $\nu_1 = (I-1)(J-1)$ et $\nu_2 = (I-1)(J-1)(K-1)$, $F((I-1)(J-1),(I-1)(J-1)(K-1);1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $(I-1)(J-1)$ et $(I-1)(J-1)(K-1)$ degrés de liberté et $F^{'}((I-1)(J-1),(I-1)(J-1)(K-1);\phi_{AB})$ est une variable aléatoire qui suit une loi de Fisher non-centrale à $(I-1)(J-1)$ et $(I-1)(J-1)(K-1)$ degrés de liberté et de paramètre de non-centralité normalisé ϕ_{AB}. Ce paramètre de non-centralité normalisé ϕ_{AB} vaut :

$$\phi_{AB} = \frac{1}{\sigma}\sqrt{\frac{K\sum_{i=1}^{I}\sum_{j=1}^{J}(\alpha\beta)_{ij}^2}{(I-1)(J-1)+1}}.$$

11.5.2. Modèle à effets aléatoires

On reprend ici les notations du paragraphe 6.2.1.

On s'intéresse à la puissance $1-\beta_A$, où β_A est le risque de commettre une erreur de deuxième espèce, du pseudo test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \sigma_A^2 = 0}$$
contre
$$\boxed{\mathcal{H}_1 : \sigma_A^2 \neq 0.}$$

La loi de la statistique de ce test n'est connue que de manière approximative. On peut néanmoins calculer la puissance de ce pseudo test F.
Cette puissance $1 - \beta_A$ est donnée par la formule suivante :

$$1 - \beta_A = \mathbb{P}\left[F(\nu_1, \nu_2) > \frac{F(\nu_1, \nu_2; 1-\alpha)}{1 + JK\dfrac{\sigma_A^2}{\sigma^2 + K\sigma_{AB}^2}}\right],$$

où $\nu_1 = I - 1$ et

$$\nu_2 = \frac{\left(s_{AB}^2 + s_{AC}^2 - s_R^2\right)^2}{\dfrac{\left(s_{AB}^2\right)^2}{(I-1)(J-1)} + \dfrac{\left(s_{AC}^2\right)^2}{(I-1)(K-1)} + \dfrac{\left(s_R^2\right)^2}{(I-1)(J-1)(K-1)}},$$

$F(I-1, \nu_2; 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $(I-1)$ et ν_2 degrés de liberté et $F(I-1, \nu_2)$ est une variable aléatoire qui suit une loi de Fisher à $I-1$ et ν_2 degrés de liberté.

On s'intéresse à la puissance $1 - \beta_{AB}$, où β_{AB} est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \sigma_{AB}^2 = 0}$$
contre
$$\boxed{\mathcal{H}_1 : \sigma_{AB}^2 \neq 0.}$$

Cette puissance $1 - \beta_{AB}$ est donnée par la formule suivante :

$$1 - \beta_{AB} = \mathbb{P}\left[F(\nu_1, \nu_2) > \frac{F(\nu_1, \nu_2; 1-\alpha)}{1 + K\dfrac{\sigma_{AB}^2}{\sigma^2}}\right],$$

où $\nu_1 = (I-1)(J-1)$ et $\nu_2 = (I-1)(J-1)(K-1)$, $F((I-1)(J-1), (I-1)(J-1)(K-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $(I-1)(J-1)$ et $(I-1)(J-1)(K-1)$ degrés de liberté et $F((I-1)(J-1), (I-1)(J-1)(K-1))$ est une variable aléatoire qui suit une loi de Fisher à $(I-1)(J-1)$ et $(I-1)(J-1)(K-1)$ degrés de liberté.

La différence fondamentale entre ce cas et le cas où le facteur est à effets fixes, exposés au paragraphe 11.5.1, est que le calcul de la puissance repose une loi de Fisher et non sur une loi de Fisher non-centrale.

11.5.3. Modèle à effets mixtes

On reprend ici les notations du paragraphe 6.3.1.

Premier cas : Deux facteurs sont à effets fixes, A et B, et un facteur C est à effets aléatoires.

On s'intéresse à la puissance $1-\beta_A$, où β_A est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \text{Il existe } i_0 \in \{1, 2, \ldots, I\} \text{ tel que } \alpha_{i_0} \neq 0.}$$

Cette puissance $1-\beta_A$ est donnée par la formule suivante :

$$1-\beta_A = \mathbb{P}\left[F^{'}(I-1,(I-1)(K-1);\phi_A) > F(I-1,(I-1)(K-1);1-\alpha)\right],$$

où $F(I-1,(I-1)(K-1);1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $I-1$ et $(I-1)(K-1)$ degrés de liberté et $F^{'}(I-1,(I-1)(K-1);\phi_A)$ est une variable aléatoire qui suit une loi de Fisher non-centrale à $I-1$ et $(I-1)(K-1)$ degrés de liberté et de paramètre de non-centralité normalisé ϕ_A. Ce paramètre de non-centralité normalisé ϕ_A vaut :

$$\phi_A = \sqrt{\frac{JK\sum_{i=1}^{I}\alpha_i^2}{I\left(\sigma^2 + J\sigma_{AC}^2\right)}}.$$

On s'intéresse à la puissance $1-\beta_C$, où β_C est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \sigma_C^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \sigma_C^2 \neq 0.}$$

Cette puissance $1-\beta_C$ est donnée par la formule suivante :

$$1-\beta_C = \mathbb{P}\left[F(K-1,(I-1)(J-1)(K-1)) > \frac{F(K-1,(I-1)(J-1)(K-1);1-\alpha)}{1+\dfrac{IJ\sigma_C^2}{\sigma^2}}\right],$$

où $F(K-1,(I-1)(J-1)(K-1);1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $K-1$ et $(I-1)(J-1)(K-1)$ degrés de liberté et $F(K-1,(I-1)(J-1)(K-1))$ est une

variable aléatoire qui suit une loi de Fisher à $K-1$ et $(I-1)(J-1)(K-1)$ degrés de liberté.

On s'intéresse à la puissance $1-\beta_{AB}$, où β_{AB} est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : (\alpha\beta)_{1,1} = (\alpha\beta)_{1,2} = \cdots = (\alpha\beta)_{1,J} = (\alpha\beta)_{2,1} = \cdots = (\alpha\beta)_{I,J} = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \text{Il existe } (i_0, j_0) \in \{1, 2, \ldots, I\} \times \{1, 2, \ldots, J\} \text{ tel que } (\alpha\beta)_{i_0,j_0} \neq 0.}$$

Cette puissance $1-\beta_{AB}$ est donnée par la formule suivante :

$$1-\beta_{AB} = \mathbb{P}\left[F'(\nu_1, \nu_2; \phi_{AB}) > F(\nu_1, \nu_2; 1-\alpha)\right],$$

où $\nu_1 = (I-1)(J-1)$ et $\nu_2 = (I-1)(J-1)(K-1)$, $F((I-1)(J-1), (I-1)(J-1)(K-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $(I-1)(J-1)$ et $(I-1)(J-1)(K-1)$ degrés de liberté et $F'((I-1)(J-1), (I-1)(J-1)(K-1); \phi_{AB})$ est une variable aléatoire qui suit une loi de Fisher non-centrale à $(I-1)(J-1)$ et $(I-1)(J-1)(K-1)$ degrés de liberté et de paramètre de non-centralité normalisé ϕ_{AB}. Ce paramètre de non-centralité normalisé ϕ_{AB} vaut :

$$\phi_{AB} = \frac{1}{\sigma}\sqrt{\frac{K\sum_{i=1}^{I}\sum_{j=1}^{J}(\alpha\beta)_{ij}^2}{(I-1)(J-1)+1}}.$$

On s'intéresse à la puissance $1-\beta_{AC}$, où β_{AC} est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \sigma_{AC}^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \sigma_{AC}^2 \neq 0.}$$

Cette puissance $1-\beta_{AC}$ est donnée par la formule suivante :

$$1-\beta_{AC} = \mathbb{P}\left[F(\nu_1, \nu_2) > \frac{F(\nu_1, \nu_2; 1-\alpha)}{1 + J\frac{\sigma_{AC}^2}{\sigma^2}}\right],$$

où $\nu_1 = (I-1)(K-1)$ et $\nu_2 = (I-1)(J-1)(K-1)$, $F((I-1)(K-1), (I-1)(J-1)(K-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $(I-1)(K-1)$ et $(I-1)(J-1)(K-1)$ degrés de liberté et $F((I-1)(K-1), (I-1)(J-1)(K-1))$ est une variable aléatoire qui suit une loi de Fisher à $(I-1)(K-1)$ et $(I-1)(J-1)(K-1)$ degrés de liberté.

Deuxième cas : Un facteur est à effets fixes et deux facteurs sont à effets aléatoires.

On s'intéresse à la puissance $1-\beta_A$, où β_A est le risque de commettre une erreur de deuxième espèce, du pseudo test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \text{Il existe } i_0 \in \{1, 2, \ldots, I\} \text{ tel que } \alpha_{i_0} \neq 0.}$$

La loi de la statistique de ce test n'est connue que de manière approximative. On peut néanmoins calculer la puissance de ce pseudo test F.
Cette puissance $1-\beta_A$ est donnée par la formule suivante :

$$1-\beta_A = \mathbb{P}\left[F^{'}(\nu_1, \nu_2, \phi_A) > F(\nu_1, \nu_2; 1-\alpha)\right],$$

où $\nu_1 = I-1$ et

$$\nu_2 = \frac{\left(s_{AB}^2 + s_{AC}^2 - s_R^2\right)^2}{\dfrac{\left(s_{AB}^2\right)^2}{(I-1)(J-1)} + \dfrac{\left(s_{AC}^2\right)^2}{(I-1)(K-1)} + \dfrac{\left(s_R^2\right)^2}{(I-1)(J-1)(K-1)}},$$

$F(I-1, \nu_2; 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $(I-1)$ et ν_2 degrés de liberté et $F^{'}(I-1, \nu_2)$ est une variable aléatoire qui suit une loi de Fisher non-centrale à $I-1$ et ν_2 degrés de liberté et de paramètre de non-centralité normalisé ϕ_{AB}. Ce paramètre de non-centralité normalisé ϕ_A vaut :

$$\phi_A = \sqrt{\frac{1}{I}\frac{JK\displaystyle\sum_{i=1}^{I}\alpha_i^2}{\sigma^2 + K\sigma_{AB}^2 + J\sigma_{AC}^2}}.$$

On s'intéresse à la puissance $1-\beta_B$, où β_B est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \sigma_B^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \sigma_B^2 \neq 0.}$$

Cette puissance $1-\beta_B$ est donnée par la formule suivante :

$$1-\beta_B = \mathbb{P}\left[F(J-1, (J-1)(K-1)) > \frac{F(J-1, (J-1)(K-1); 1-\alpha)}{1 + \dfrac{IK\sigma_B^2}{\sigma^2 + I\sigma_{BC}^2}}\right],$$

où $F(J-1,(J-1)(K-1);1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $J-1$ et $(J-1)(K-1)$ degrés de liberté et $F(J-1,(J-1)(K-1))$ est une variable aléatoire qui suit une loi de Fisher à $I-1$ et $(J-1)(K-1)$ degrés de liberté.

On s'intéresse à la puissance $1-\beta_{AB}$, où β_{AB} est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \sigma_{AB}^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \sigma_{AB}^2 \neq 0.}$$

Cette puissance $1-\beta_{AB}$ est donnée par la formule suivante :

$$1-\beta_{AB} = \mathbb{P}\left[F(\nu_1,\nu_2) > \frac{F(\nu_1,\nu_2;1-\alpha)}{1+K\dfrac{\sigma_{AB}^2}{\sigma^2}}\right],$$

où $\nu_1 = (I-1)(J-1)$ et $\nu_2 = (I-1)(J-1)(K-1)$, $F((I-1)(J-1),(I-1)(J-1)(K-1);1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $(I-1)(J-1)$ et $(I-1)(J-1)(K-1)$ degrés de liberté et $F((I-1)(J-1),(I-1)(J-1)(K-1))$ est une variable aléatoire qui suit une loi de Fisher à $(I-1)(J-1)$ et $(I-1)(J-1)(K-1)$ degrés de liberté.

11.6. Analyse de la variance à trois facteurs avec interaction

Les idées développées dans la section 11.1 précédente sur le calcul de la puissance *a posteriori* ou la détermination du nombre de répétitions minimal pour obtenir un niveau de puissance supérieur ou égal à une valeur cible $1-\beta_0$ sont directement transférables aux tests étudiés dans cette section à condition de remplacer les formules par celles qui sont exposées ci-dessous.

La puissance calculée pour chacun des tests ne dépend pas du risque de première espèce fixé pour les autres tests. Ceci permettrait de calculer la puissance des tests du tableau de l'analyse de la variance même si le seuil de chaque test varie en fonction du test considéré. Voir le paragraphe 2.3 pour plus de détails sur la gestion des risques de première espèce des tests du tableau de l'analyse de la variance.

11.6.1. Modèle à effets fixes

On reprend ici les notations du paragraphe 6.1.2.

Pour des problèmes de concision on ne présentera que la puissance d'un type de test, effet principal, interaction d'ordre 1, interaction d'ordre 2, les valeurs de puissance associées

aux tests restants s'en déduisant en remplaçant les quantités associées aux facteurs mis en jeu par celles associées aux facteurs pour lesquels on veut calculer la puissance du test.

On s'intéresse à la puissance $1-\beta_A$, où β_A est le risque de commettre une erreur de deuxième espèce, du pseudo test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \text{Il existe } i_0 \in \{1, 2, \ldots, I\} \text{ tel que } \alpha_{i_0} \neq 0.}$$

Cette puissance $1-\beta_A$ est donnée par la formule suivante :

$$1-\beta_A = \mathbb{P}\left[F^{'}(\nu_1, \nu_2; \phi_A) > F(\nu_1, \nu_2; 1-\alpha)\right],$$

où $\nu_1 = I-1$ et $\nu_2 = IJK(L-1)$, $F(I-1, IJK(L-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $I-1$ et $IJK(L-1)$ degrés de liberté et $F^{'}(I-1, IJK(L-1); \phi_A)$ est une variable aléatoire qui suit une loi de Fisher non-centrale à $I-1$ et $IJK(L-1)$ degrés de liberté et de paramètre de non-centralité normalisé ϕ_A. Ce paramètre de non-centralité normalisé ϕ_A vaut :

$$\phi_A = \frac{1}{\sigma}\sqrt{\frac{JKL}{I}\sum_{i=1}^{I}\alpha_i^2}.$$

On s'intéresse à la puissance $1-\beta_{AB}$, où β_{AB} est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : (\alpha\beta)_{1,1} = (\alpha\beta)_{1,2} = \cdots = (\alpha\beta)_{1,J} = (\alpha\beta)_{2,1} = \cdots = (\alpha\beta)_{I,J} = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \text{Il existe } (i_0, j_0) \in \{1, 2, \ldots, I\} \times \{1, 2, \ldots, J\} \text{ tel que } (\alpha\beta)_{i_0,j_0} \neq 0.}$$

Cette puissance $1-\beta_{AB}$ est donnée par la formule suivante :

$$1-\beta_{AB} = \mathbb{P}\left[F^{'}(\nu_1, \nu_2; \phi_{AB}) > F(\nu_1, \nu_2; 1-\alpha)\right],$$

où $\nu_1 = (I-1)(J-1)$ et $\nu_2 = IJK(L-1)$, $F((I-1)(J-1), IJK(L-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $(I-1)(J-1)$ et $IJK(L-1)$ degrés de liberté et $F^{'}((I-1)(J-1), IJK(L-1); \phi_{AB})$ est une variable aléatoire qui suit une loi de Fisher non-centrale à $(I-1)(J-1)$ et $IJK(L-1)$ degrés de liberté et de paramètre de non-centralité normalisé ϕ_{AB}. Ce paramètre de non-centralité normalisé ϕ_{AB} vaut :

$$\phi_{AB} = \frac{1}{\sigma}\sqrt{\frac{KL\sum_{i=1}^{I}\sum_{j=1}^{J}(\alpha\beta)_{ij}^2}{(I-1)(J-1)+1}}.$$

On s'intéresse à la puissance $1-\beta_{ABC}$, où β_{ABC} est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : (\alpha\beta\gamma)_{1,1,1} = (\alpha\beta\gamma)_{1,1,2} = \cdots = (\alpha\beta\gamma)_{1,1,K} = (\alpha\beta\gamma)_{2,1,1} = \cdots = (\alpha\beta\gamma)_{I,J,K} = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \exists\ (i_0, j_0, k_0) \in \{1, 2, \ldots, I\} \times \{1, 2, \ldots, J\} \times \{1, 2, \ldots, K\} \mid (\alpha\beta\gamma)_{i_0,j_0,k_0} \neq 0.}$$

Cette puissance $1 - \beta_{ABC}$ est donnée par la formule suivante :

$$1 - \beta_{ABC} = \mathbb{P}\left[F^{'}(\nu_1, \nu_2; \phi_{ABC}) > F(\nu_1, \nu_2; 1 - \alpha)\right],$$

où $\nu_1 = (I-1)(J-1)(K-1)$ et $\nu_2 = IJK(L-1)$, $F((I-1)(J-1)(K-1), IJK(L-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $(I-1)(J-1)(K-1)$ et $IJK(L-1)$ degrés de liberté et $F^{'}((I-1)(J-1)(K-1), IJK(L-1); \phi_{ABC})$ est une variable aléatoire qui suit une loi de Fisher non-centrale à $(I-1)(J-1)(K-1)$ et $IJK(L-1)$ degrés de liberté et de paramètre de non-centralité normalisé ϕ_{ABC}. Ce paramètre de non-centralité normalisé ϕ_{ABC} vaut :

$$\phi_{ABC} = \frac{1}{\sigma}\sqrt{\frac{L\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(\alpha\beta\gamma)_{ijk}^2}{(I-1)(J-1)(K-1)+1}}.$$

11.6.2. Modèle à effets aléatoires

On reprend ici les notations du paragraphe 6.2.2.

On s'intéresse à la puissance $1 - \beta_A$, où β_A est le risque de commettre une erreur de deuxième espèce, du pseudo test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \sigma_A^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \sigma_A^2 \neq 0.}$$

La loi de la statistique de ce test n'est connue que de manière approximative. On peut néanmoins calculer la puissance de ce pseudo test F.
Cette puissance $1 - \beta_A$ est donnée par la formule suivante :

$$1 - \beta_A = \mathbb{P}\left[F(\nu_1, \nu_2) > \frac{F(\nu_1, \nu_2; 1-\alpha)}{1 + \dfrac{JKL\sigma_A^2}{\sigma^2 + L\sigma_{ABC}^2 + KL\sigma_{AB}^2 + JL\sigma_{AC}^2}}\right],$$

où $\nu_1 = I - 1$ et

$$\nu_2 = \frac{\left(s_{AB}^2 + s_{AC}^2 - s_{ABC}^2\right)^2}{\dfrac{\left(s_{AB}^2\right)^2}{(I-1)(J-1)} + \dfrac{\left(s_{AC}^2\right)^2}{(I-1)(K-1)} + \dfrac{\left(s_{ABC}^2\right)^2}{(I-1)(J-1)(K-1)}},$$

$F(I-1,\nu_2;1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $(I-1)$ et ν_2 degrés de liberté et $F(I-1,\nu_2)$ est une variable aléatoire qui suit une loi de Fisher à $I-1$ et ν_2 degrés de liberté.

On s'intéresse à la puissance $1-\beta_{AB}$, où β_{AB} est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \sigma^2_{AB} = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \sigma^2_{AB} \neq 0.}$$

Cette puissance $1-\beta_{AB}$ est donnée par la formule suivante :

$$1-\beta_{AB} = \mathbb{P}\left[F(\nu_1,\nu_2) > \frac{F(\nu_1,\nu_2;1-\alpha)}{1+\dfrac{KL\sigma^2_{AB}}{\sigma^2+L\sigma^2_{ABC}}}\right],$$

où $\nu_1 = (I-1)(J-1)$ et $\nu_2 = (I-1)(J-1)(K-1)$, $F((I-1)(J-1),(I-1)(J-1)(K-1);1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $(I-1)(J-1)$ et $(I-1)(J-1)(K-1)$ degrés de liberté et $F((I-1)(J-1),(I-1)(J-1)(K-1))$ est une variable aléatoire qui suit une loi de Fisher à $(I-1)(J-1)$ et $(I-1)(J-1)(K-1)$ degrés de liberté.

On s'intéresse à la puissance $1-\beta_{ABC}$, où β_{ABC} est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \sigma^2_{ABC} = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \sigma^2_{ABC} \neq 0.}$$

Cette puissance $1-\beta_{ABC}$ est donnée par la formule suivante :

$$1-\beta_{ABC} = \mathbb{P}\left[F(\nu_1,\nu_2) > \frac{F(\nu_1,\nu_2;1-\alpha)}{1+L\dfrac{\sigma^2_{ABC}}{\sigma^2}}\right],$$

où $\nu_1 = (I-1)(J-1)(K-1)$ et $\nu_2 = IJK(L-1)$, $F((I-1)(J-1)(K-1),IJK(L-1);1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $(I-1)(J-1)(K-1)$ et $IJK(L-1)$ degrés de liberté et $F((I-1)(J-1)(K-1),IJK(L-1))$ est une variable aléatoire qui suit une loi de Fisher à $(I-1)(J-1)(K-1)$ et $IJK(L-1)$ degrés de liberté.

La différence fondamentale entre ce cas et le cas où le facteur est à effets fixes, exposé au paragraphe 11.6.1, est que le calcul de la puissance repose une loi de Fisher et non sur une loi de Fisher non-centrale.

11.6.3. Modèle à effets mixtes

On reprend ici les notations du paragraphe 6.3.2.

Premier cas : Deux facteurs sont à effets fixes, A et B, et un facteur C est à effets aléatoires.

On s'intéresse à la puissance $1-\beta_A$, où β_A est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \text{Il existe } i_0 \in \{1, 2, \ldots, I\} \text{ tel que } \alpha_{i_0} \neq 0.}$$

Cette puissance $1-\beta_A$ est donnée par la formule suivante :

$$1-\beta_A = \mathbb{P}\left[F^{'}(I-1,(I-1)(J-1);\phi_A) > F(I-1,(I-1)(J-1);1-\alpha)\right],$$

où $F(I-1,(I-1)(J-1);1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $I-1$ et $(I-1)(J-1)$ degrés de liberté et $F^{'}(I-1,(I-1)(J-1);\phi_A)$ est une variable aléatoire qui suit une loi de Fisher non-centrale à $I-1$ et $(I-1)(J-1)$ degrés de liberté et de paramètre de non-centralité normalisé ϕ_A. Ce paramètre de non-centralité normalisé ϕ_A vaut :

$$\phi_A = \sqrt{\frac{JKL\sum_{i=1}^{I}\alpha_i^2}{I\left(\sigma^2 + JL\sigma_{AC}^2\right)}}.$$

On s'intéresse à la puissance $1-\beta_C$, où β_C est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \sigma_C^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \sigma_C^2 \neq 0.}$$

Cette puissance $1-\beta_C$ est donnée par la formule suivante :

$$1-\beta_C = \mathbb{P}\left[F(K-1,IJK(L-1)) > \frac{F(K-1,IJK(L-1);1-\alpha)}{1+\dfrac{IJ\sigma_C^2}{\sigma^2}}\right],$$

où $F(K-1,IJK(L-1);1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $K-1$ et $IJK(L-1)$ degrés de liberté et $F(K-1,IJK(L-1))$ est une variable aléatoire qui suit

une loi de Fisher à $K-1$ et $IJK(L-1)$ degrés de liberté.

On s'intéresse à la puissance $1-\beta_{AB}$, où β_{AB} est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : (\alpha\beta)_{1,1} = (\alpha\beta)_{1,2} = \cdots = (\alpha\beta)_{1,J} = (\alpha\beta)_{2,1} = \cdots = (\alpha\beta)_{I,J} = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \text{Il existe } (i_0, j_0) \in \{1, 2, \ldots, I\} \times \{1, 2, \ldots, J\} \text{ tel que } (\alpha\beta)_{i_0,j_0} \neq 0.}$$

Cette puissance $1-\beta_{AB}$ est donnée par la formule suivante :

$$1-\beta_{AB} = \mathbb{P}\left[F^{'}(\nu_1, \nu_2; \phi_{AB}) > F(\nu_1, \nu_2; 1-\alpha)\right],$$

où $\nu_1 = (I-1)(J-1)$ et $\nu_2 = (I-1)(J-1)(K-1)$, $F((I-1)(J-1), (I-1)(J-1)(K-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $(I-1)(J-1)$ et $(I-1)(J-1)(K-1)$ degrés de liberté et $F^{'}((I-1)(J-1), (I-1)(J-1)(K-1); \phi_{AB})$ est une variable aléatoire qui suit une loi de Fisher non-centrale à $(I-1)(J-1)$ et $(I-1)(J-1)(K-1)$ degrés de liberté et de paramètre de non-centralité normalisé ϕ_{AB}. Ce paramètre de non-centralité normalisé ϕ_{AB} vaut :

$$\phi_{AB} = \sqrt{\frac{KL\displaystyle\sum_{i=1}^{I}\sum_{j=1}^{J}(\alpha\beta)_{ij}^2}{((I-1)(J-1)+1)(\sigma^2 + L\sigma_{ABC}^2)}}.$$

On s'intéresse à la puissance $1-\beta_{AC}$, où β_{AC} est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \sigma_{AC}^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \sigma_{AC}^2 \neq 0.}$$

Cette puissance $1-\beta_{AC}$ est donnée par la formule suivante :

$$1-\beta_{AC} = \mathbb{P}\left[F(\nu_1, \nu_2) > \frac{F(\nu_1, \nu_2; 1-\alpha)}{1 + JL\dfrac{\sigma_{AC}^2}{\sigma^2}}\right],$$

où $\nu_1 = (I-1)(K-1)$ et $\nu_2 = IJK(L-1)$, $F((I-1)(K-1), IJK(L-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $(I-1)(K-1)$ et $IJK(L-1)$ degrés de liberté et $F((I-1)(K-1), IJK(L-1))$ est une variable aléatoire qui suit une loi de Fisher à $(I-1)(K-1)$ et $IJK(L-1)$ degrés de liberté.

On s'intéresse à la puissance $1-\beta_{ABC}$, où β_{ABC} est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \sigma^2_{ABC} = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \sigma^2_{ABC} \neq 0.}$$

Cette puissance $1 - \beta_{ABC}$ est donnée par la formule suivante :

$$1 - \beta_{ABC} = \mathbb{P}\left[F(\nu_1, \nu_2) > \frac{F(\nu_1, \nu_2; 1-\alpha)}{1 + L\dfrac{\sigma^2_{ABC}}{\sigma^2}}\right],$$

où $\nu_1 = (I-1)(J-1)(K-1)$ et $\nu_2 = IJK(L-1)$, $F((I-1)(J-1)(K-1), IJK(L-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $(I-1)(J-1)(K-1)$ et $IJK(L-1)$ degrés de liberté et $F((I-1)(J-1)(K-1), IJK(L-1))$ est une variable aléatoire qui suit une loi de Fisher à $(I-1)(J-1)(K-1)$ et $IJK(L-1)$ degrés de liberté.

Deuxième cas : Un facteur est à effets fixes et deux facteurs sont à effets aléatoires.

On s'intéresse à la puissance $1 - \beta_A$, où β_A est le risque de commettre une erreur de deuxième espèce, du pseudo test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \text{Il existe } i_0 \in \{1, 2, \ldots, I\} \text{ tel que } \alpha_{i_0} \neq 0.}$$

La loi de la statistique de ce test n'est connue que de manière approximative. On peut néanmoins calculer la puissance de ce pseudo test F.
Cette puissance $1 - \beta_A$ est donnée par la formule suivante :

$$1 - \beta_A = \mathbb{P}\left[F'(\nu_1, \nu_2, \phi_A) > F(\nu_1, \nu_2; 1-\alpha)\right],$$

où $\nu_1 = I - 1$ et

$$\nu_2 = \frac{\left(s^2_{AB} + s^2_{AC} - s^2_{ABC}\right)^2}{\dfrac{\left(s^2_{AB}\right)^2}{(I-1)(J-1)} + \dfrac{\left(s^2_{AC}\right)^2}{(I-1)(K-1)} + \dfrac{\left(s^2_{ABC}\right)^2}{(I-1)(J-1)(K-1)}},$$

$F(I-1, \nu_2; 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $(I-1)$ et ν_2 degrés de liberté et $F'(I-1, \nu_2, \phi_A)$ est une variable aléatoire qui suit une loi de Fisher non-centrale à $I-1$ et ν_2 degrés de liberté et de paramètre de non-centralité normalisé ϕ_A. Ce paramètre de non-centralité normalisé ϕ_A vaut :

$$\phi_A = \sqrt{\frac{1}{I}\frac{JKL\displaystyle\sum_{i=1}^{I}\alpha_i^2}{\sigma^2 + L\sigma^2_{ABC} + KL\sigma^2_{AB} + JL\sigma^2_{AC}}}.$$

On s'intéresse à la puissance $1-\beta_B$, où β_B est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \sigma_B^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \sigma_B^2 \neq 0.}$$

Cette puissance $1-\beta_B$ est donnée par la formule suivante :

$$1-\beta_B = \mathbb{P}\left[F(J-1,(J-1)(K-1)) > \frac{F(J-1,(J-1)(K-1);1-\alpha)}{1+\dfrac{IKL\sigma_B^2}{\sigma^2+IL\sigma_{BC}^2}}\right],$$

où $F(J-1,(J-1)(K-1);1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $J-1$ et $(J-1)(K-1)$ degrés de liberté et $F(J-1,(J-1)(K-1))$ est une variable aléatoire qui suit une loi de Fisher à $I-1$ et $(J-1)(K-1)$ degrés de liberté.

On s'intéresse à la puissance $1-\beta_{AB}$, où β_{AB} est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \sigma_{AB}^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \sigma_{AB}^2 \neq 0.}$$

Cette puissance $1-\beta_{AB}$ est donnée par la formule suivante :

$$1-\beta_{AB} = \mathbb{P}\left[F(\nu_1,\nu_2) > \frac{F(\nu_1,\nu_2;1-\alpha)}{1+\dfrac{KL\sigma_{AB}^2}{\sigma^2+L\sigma_{ABC}^2}}\right],$$

où $\nu_1 = (I-1)(J-1)$ et $\nu_2 = (I-1)(J-1)(K-1)$, $F((I-1)(J-1),(I-1)(J-1)(K-1);1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $(I-1)(J-1)$ et $(I-1)(J-1)(K-1)$ degrés de liberté et $F((I-1)(J-1),(I-1)(J-1)(K-1))$ est une variable aléatoire qui suit une loi de Fisher à $(I-1)(J-1)$ et $(I-1)(J-1)(K-1)$ degrés de liberté.

On s'intéresse à la puissance $1-\beta_{BC}$, où β_{BC} est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \sigma_{BC}^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \sigma_{BC}^2 \neq 0.}$$

Cette puissance $1-\beta_{BC}$ est donnée par la formule suivante :

$$1-\beta_{BC} = \mathbb{P}\left[F(\nu_1,\nu_2) > \frac{F(\nu_1,\nu_2;1-\alpha)}{1+KL\dfrac{\sigma^2_{BC}}{\sigma^2}}\right],$$

où $\nu_1 = (J-1)(K-1)$ et $\nu_2 = IJK(L-1)$, $F((J-1)(K-1), IJK(L-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $(J-1)(K-1)$ et $IJK(L-1)$ degrés de liberté et $F((J-1)(K-1), IJK(L-1))$ est une variable aléatoire qui suit une loi de Fisher à $(J-1)(K-1)$ et $IJK(L-1)$ degrés de liberté.

On s'intéresse à la puissance $1-\beta_{ABC}$, où β_{ABC} est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \sigma^2_{ABC} = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \sigma^2_{ABC} \neq 0.}$$

Cette puissance $1-\beta_{ABC}$ est donnée par la formule suivante :

$$1-\beta_{ABC} = \mathbb{P}\left[F(\nu_1,\nu_2) > \frac{F(\nu_1,\nu_2;1-\alpha)}{1+L\dfrac{\sigma^2_{ABC}}{\sigma^2}}\right],$$

où $\nu_1 = (I-1)(J-1)(K-1)$ et $\nu_2 = IJK(L-1)$, $F((I-1)(J-1)(K-1), IJK(L-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $(I-1)(J-1)(K-1)$ et $IJK(L-1)$ degrés de liberté et $F((I-1)(J-1)(K-1), IJK(L-1))$ est une variable aléatoire qui suit une loi de Fisher à $(I-1)(J-1)(K-1)$ et $IJK(L-1)$ degrés de liberté.

11.7. Analyse de la variance à trois facteurs totalement emboîtés

Les idées développées dans la section 11.1 précédente sur le calcul de la puissance *a posteriori* ou la détermination du nombre de répétitions minimal pour obtenir un niveau de puissance supérieur ou égal à une valeur cible $1-\beta_0$ sont directement transférables aux tests étudiés dans cette section à condition de remplacer les formules par celles qui sont exposées ci-dessous.

La puissance calculée pour chacun des tests ne dépend pas du risque de première espèce fixé pour les autres tests. Ceci permettrait de calculer la puissance des tests du tableau de l'analyse de la variance même si le seuil de chaque test varie en fonction du test considéré. Voir le paragraphe 2.3 pour plus de détails sur la gestion des risques de première espèce des tests du tableau de l'analyse de la variance.

11.7.1. Modèle à effets fixes

On reprend ici les notations du paragraphe 7.1.1.

On s'intéresse à la puissance $1-\beta_A$, où β_A est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \text{Il existe } i_0 \in \{1, 2, \ldots, I\} \text{ tel que } \alpha_{i_0} \neq 0.}$$

Cette puissance $1-\beta_A$ est donnée par la formule suivante :

$$1-\beta_A = \mathbb{P}\left[F^{'}(I-1, IJK(L-1); \phi_A) > F(I-1, IJK(L-1); 1-\alpha)\right],$$

où $F(I-1, IJK(L-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $I-1$ et $IJK(L-1)$ degrés de liberté et $F^{'}(I-1, IJK(L-1); \phi_A)$ est une variable aléatoire qui suit une loi de Fisher non-centrale à $I-1$ et $IJK(L-1)$ degrés de liberté et de paramètre de non-centralité normalisé ϕ_A. Ce paramètre de non-centralité normalisé ϕ_A vaut :

$$\phi_A = \frac{1}{\sigma}\sqrt{\frac{JKL}{I}\sum_{i=1}^{I}\alpha_i^2}.$$

On s'intéresse à la puissance $1-\beta_{B(A)}$, où $\beta_{B(A)}$ est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \beta_{1(1)} = \beta_{2(1)} = \cdots = \beta_{J(1)} = \beta_{1(2)} = \cdots = \beta_{J(I)} = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \text{Il existe } (i_0, j_0) \in \{1, 2, \ldots, I\} \times \{1, 2, \ldots, J\} \text{ tel que } \beta_{j_0(i_0)} \neq 0.}$$

Cette puissance $1-\beta_{B(A)}$ est donnée par la formule suivante :

$$1-\beta_{B(A)} = \mathbb{P}\left[F^{'}(I(J-1), IJK(L-1); \phi_{B(A)}) > F(I(J-1), IJK(L-1); 1-\alpha)\right],$$

où $F(I(J-1), IJK(L-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $I(J-1)$ et $IJK(L-1)$ degrés de liberté et $F^{'}(I(J-1), IJK(L-1); \phi_B)$ est une variable aléatoire qui suit une loi de Fisher non-centrale à $I(J-1)$ et $IJK(L-1)$ degrés de liberté et de paramètre de non-centralité normalisé $\phi_{B(A)}$. Ce paramètre de non-centralité normalisé $\phi_{B(A)}$ vaut :

$$\phi_{B(A)} = \frac{1}{\sigma}\sqrt{\frac{KL\sum_{i=1}^{I}\sum_{j=1}^{J}\beta_{j(i)}^2}{I(J-1)+1}}.$$

On s'intéresse à la puissance $1-\beta_{C(B(A))}$, où $\beta_{C(B(A))}$ est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \gamma_{1(1(1))} = \gamma_{2(1(1))} = \cdots = \gamma_{K(1(1))} = \gamma_{1(1(2))} = \cdots = \gamma_{K(J(I))} = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \exists\ (i_0, j_0, k_0) \in \{1, 2, \ldots, I\} \times \{1, 2, \ldots, J\} \times \{1, 2, \ldots, K\} \mid \gamma_{k_0(j_0(i_0))} \neq 0.}$$

Cette puissance $1 - \beta_{C(B(A))}$ est donnée par la formule suivante :

$$1 - \beta_{C(B(A))} = \mathbb{P}\left[F'(\nu_1, \nu_2; \phi_{C(B(A))}) > F(\nu_1, \nu_2; 1 - \alpha)\right],$$

où $\nu_1 = IJ(L-1)$ et $\nu_2 = IJK(L-1)$, $F(IJ(L-1), IJK(L-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $IJ(L-1)$ et $IJK(L-1)$ degrés de liberté et $F'(IJ(L-1), IJK(L-1); \phi_{C(B(A))})$ est une variable aléatoire qui suit une loi de Fisher non-centrale à $IJ(L-1)$ et $IJK(L-1)$ degrés de liberté et de paramètre de non-centralité normalisé $\phi_{C(B(A))}$. Ce paramètre de non-centralité normalisé $\phi_{C(B(A))}$ vaut :

$$\phi_{C(B(A))} = \frac{1}{\sigma}\sqrt{\frac{L\sum_{i=1}^{I}\sum_{j=1}^{J}\gamma^2_{k(j(i))}}{IJ(L-1)+1}}.$$

11.7.2. Modèle à effets aléatoires

On reprend ici les notations du paragraphe 7.2.1.

On s'intéresse à la puissance $1 - \beta_A$, où β_A est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \sigma_A^2 = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \sigma_A^2 \neq 0.}$$

Cette puissance $1 - \beta_A$ est donnée par la formule suivante :

$$1 - \beta_A = \mathbb{P}\left[F(I-1, I(J-1)) > \frac{F(I-1, I(J-1); 1-\alpha)}{1 + JKL\dfrac{\sigma_A^2}{\sigma^2 + L\sigma^2_{C|B|A} + KL\sigma^2_{B|A}}}\right],$$

où $F(I-1, I(J-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $(I-1)$ et $I(J-1)$ degrés de liberté et $F(I-1, I(J-1))$ est une variable aléatoire qui suit une loi de Fisher à $I-1$ et $I(J-1)$ degrés de liberté.

On s'intéresse à la puissance $1 - \beta_{B(A)}$, où $\beta_{B(A)}$ est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \sigma^2_{B|A} = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \sigma^2_{B|A} \neq 0.}$$

Cette puissance $1 - \beta_{B(A)}$ est donnée par la formule suivante :

$$1 - \beta_{B(A)} = \mathbb{P}\left[F(I(J-1), IJ(K-1)) > \frac{F(I(J-1), IJ(K-1); 1-\alpha)}{1 + KL\dfrac{\sigma^2_{B|A}}{\sigma^2 + L\sigma C|B|A^2}}\right],$$

où $F(I(J-1), IJ(K-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $I(J-1)$ et $IJ(K-1)$ degrés de liberté et $F(I(J-1), IJ(K-1))$ est une variable aléatoire qui suit une loi de Fisher à $I(J-1)$ et $IJ(K-1)$ degrés de liberté.

On s'intéresse à la puissance $1-\beta_{C(B(A))}$, où $\beta_{C(B(A))}$ est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \sigma^2_{C|B|A} = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \sigma^2_{C|B|A} \neq 0.}$$

Cette puissance $1 - \beta_{C(B(A))}$ est donnée par la formule suivante :

$$1 - \beta_{C(B(A))} = \mathbb{P}\left[F(\nu_1, \nu_2) > \frac{F(\nu_1, \nu_2; 1-\alpha)}{1 + L\dfrac{\sigma^2_{C|B|A}}{\sigma^2}}\right],$$

où $\nu_1 = IJ(L-1)$ et $\nu_2 = IJK(L-1)$, $F(IJ(L-1), IJK(L-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $IJ(L-1)$ et $IJK(L-1)$ degrés de liberté et $F^{'}(IJ(L-1), IJK(L-1))$ est une variable aléatoire qui suit une loi de Fisher à $IJ(L-1)$ et $IJK(L-1)$ degrés de liberté.

11.7.3. Modèle à effets mixtes

On reprend ici les notations du paragraphe 7.3.

Premier cas : Deux facteurs sont à effets fixes et un facteur est à effets aléatoires.

On s'intéresse à la puissance $1 - \beta_A$, où β_A est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \text{Il existe } i_0 \in \{1, 2, \ldots, I\} \text{ tel que } \alpha_{i_0} \neq 0.}$$

Cette puissance $1 - \beta_A$ est donnée par la formule suivante :

$$1 - \beta_A = \mathbb{P}\left[F^{'}(I-1, IJ(K-1); \phi_A) > F(I-1, IJ(K-1); 1-\alpha)\right],$$

où $F(I-1, IJ(K-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $I-1$ et $IJ(K-1)$ degrés de liberté et $F^{'}(I-1, IJ(K-1); \phi_A)$ est une variable aléatoire qui suit une loi de Fisher non-centrale à $I-1$ et $IJ(K-1)$ degrés de liberté et de paramètre de non-centralité normalisé ϕ_A. Ce paramètre de non-centralité normalisé ϕ_A vaut :

$$\phi_A = \sqrt{\frac{JKL \sum_{i=1}^{I} \alpha_i^2}{I\left(\sigma^2 + K\sigma^2_{C|B|A}\right)}}.$$

On s'intéresse à la puissance $1 - \beta_{B(A)}$, où $\beta_{B(A)}$ est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \beta_{1(1)} = \beta_{2(1)} = \cdots = \beta_{J(1)} = \beta_{1(2)} = \cdots = \beta_{J(I)} = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \text{Il existe } (i_0, j_0) \in \{1, 2, \ldots, I\} \times \{1, 2, \ldots, J\} \text{ tel que } \beta_{j_0(i_0)} \neq 0.}$$

Cette puissance $1 - \beta_{B(A)}$ est donnée par la formule suivante :

$$1 - \beta_{B(A)} = \mathbb{P}\left[F^{'}(I(J-1), IJ(K-1); \phi_{B(A)}) > F(I(J-1), IJ(K-1); 1-\alpha)\right],$$

où $F(I(J-1), IJ(K-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $I(J-1)$ et $IJ(K-1)$ degrés de liberté et $F^{'}(I(J-1), IJ(K-1); \phi_{B(A)})$ est une variable aléatoire qui suit une loi de Fisher non-centrale à $I(J-1)$ et $IJ(K-1)$ degrés de liberté et de paramètre de non-centralité normalisé $\phi_{B(A)}$. Ce paramètre de non-centralité normalisé $\phi_{B(A)}$ vaut :

$$\phi_{B(A)} = \sqrt{\frac{KL}{I(J-1)+1} \frac{\sum_{i=1}^{I} \sum_{j=1}^{J} \beta^2_{j(i)}}{\sigma^2 + L\sigma^2_{C|B|A}}}.$$

On s'intéresse à la puissance $1 - \beta_{C(B(A))}$, où $\beta_{C(B(A))}$ est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \sigma^2_{C|B|A} = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \sigma^2_{C|B|A} \neq 0.}$$

Cette puissance $1 - \beta_{C(B(A))}$ est donnée par la formule suivante :

$$1 - \beta_{C(B(A))} = \mathbb{P}\left[F(\nu_1, \nu_2) > \frac{F(\nu_1, \nu_2; 1-\alpha)}{1 + L\dfrac{\sigma^2_{C|B|A}}{\sigma^2}}\right],$$

où $\nu_1 = IJ(L-1)$ et $\nu_2 = IJK(L-1)$, $F(IJ(L-1), IJK(L-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $IJ(L-1)$ et $IJK(L-1)$ degrés de liberté et $F^{'}(IJ(L-1), IJK(L-1))$ est une variable aléatoire qui suit une loi de Fisher à $IJ(L-1)$ et $IJK(L-1)$ degrés de liberté.

Deuxième cas : Un facteur est à effets fixes et deux facteurs sont à effets aléatoires.

On s'intéresse à la puissance $1 - \beta_A$, où β_A est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \text{Il existe } i_0 \in \{1, 2, \ldots, I\} \text{ tel que } \alpha_{i_0} \neq 0.}$$

Cette puissance $1 - \beta_A$ est donnée par la formule suivante :

$$1 - \beta_A = \mathbb{P}\left[F^{'}(I-1, I(J-1); \phi_A) > F(I-1, I(J-1); 1-\alpha)\right],$$

où $F(I-1, I(J-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $I-1$ et $I(J-1)$ degrés de liberté et $F^{'}(I-1, I(J-1); \phi_A)$ est une variable aléatoire qui suit une loi de Fisher non-centrale à $I-1$ et $I(J-1)$ degrés de liberté et de paramètre de non-centralité normalisé ϕ_A. Ce paramètre de non-centralité normalisé ϕ_A vaut :

$$\phi_A = \sqrt{\frac{JKL\displaystyle\sum_{i=1}^{I} \alpha_i^2}{I\left(\sigma^2 + KL\sigma^2_{B|A} + L\sigma^2_{C|B|A}\right)}}.$$

On s'intéresse à la puissance $1 - \beta_{B(A)}$, où $\beta_{B(A)}$ est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \sigma^2_{B|A} = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \sigma^2_{B|A} \neq 0.}$$

Cette puissance $1 - \beta_{B(A)}$ est donnée par la formule suivante :

$$1 - \beta_{B(A)} = \mathbb{P}\left[F(I(J-1), IJ(K-1)) > \frac{F(I(J-1), IJ(K-1); 1-\alpha)}{1 + KL\dfrac{\sigma^2_{B|A}}{\sigma^2 + L\sigma C|B|A^2}}\right],$$

où $F(I(J-1), IJ(K-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $I(J-1)$ et $IJ(K-1)$ degrés de liberté et $F(I(J-1), IJ(K-1))$ est une variable aléatoire qui suit une loi de Fisher à $I(J-1)$ et $IJ(K-1)$ degrés de liberté.

On s'intéresse à la puissance $1-\beta_{C(B(A))}$, où $\beta_{C(B(A))}$ est le risque de commettre une erreur de deuxième espèce, du test F d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \sigma^2_{C|B|A} = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \sigma^2_{C|B|A} \neq 0.}$$

Cette puissance $1 - \beta_{C(B(A))}$ est donnée par la formule suivante :

$$1 - \beta_{C(B(A))} = \mathbb{P}\left[F(\nu_1, \nu_2) > \frac{F(\nu_1, \nu_2; 1-\alpha)}{1 + L\dfrac{\sigma^2_{C|B|A}}{\sigma^2}}\right],$$

où $\nu_1 = IJ(L-1)$ et $\nu_2 = IJK(L-1)$, $F(IJ(L-1), IJK(L-1); 1-\alpha)$ est le $100(1-\alpha)$ quantile de la loi de Fisher à $IJ(L-1)$ et $IJK(L-1)$ degrés de liberté et $F'(IJ(L-1), IJK(L-1))$ est une variable aléatoire qui suit une loi de Fisher à $IJ(L-1)$ et $IJK(L-1)$ degrés de liberté.

12. Exemples

12.1. Analyse de la variance à un facteur

12.1.1. Carburateurs[19]

On veut tester quatre types de carburateurs : A_1, A_2, A_3 et A_4. Pour chaque type de carburateur on dispose de six pièces que l'on monte sucessivement en parallèle sur quatre

[19] Les données de cet exercice sont tirées du livre de Georges Parreins [8].

voitures que l'on suppose avoir des caractéristiques parfaitement identiques. Le tableau ci-dessous indique pour chacun des essais la valeur d'un paramètre lié à la consommation :

Essai	A_1	A_2	A_3	A_4
1	21	23	18	20
2	24	23	19	21
3	25	32	28	25
4	20	23	19	15
5	34	32	24	29
6	17	15	14	9

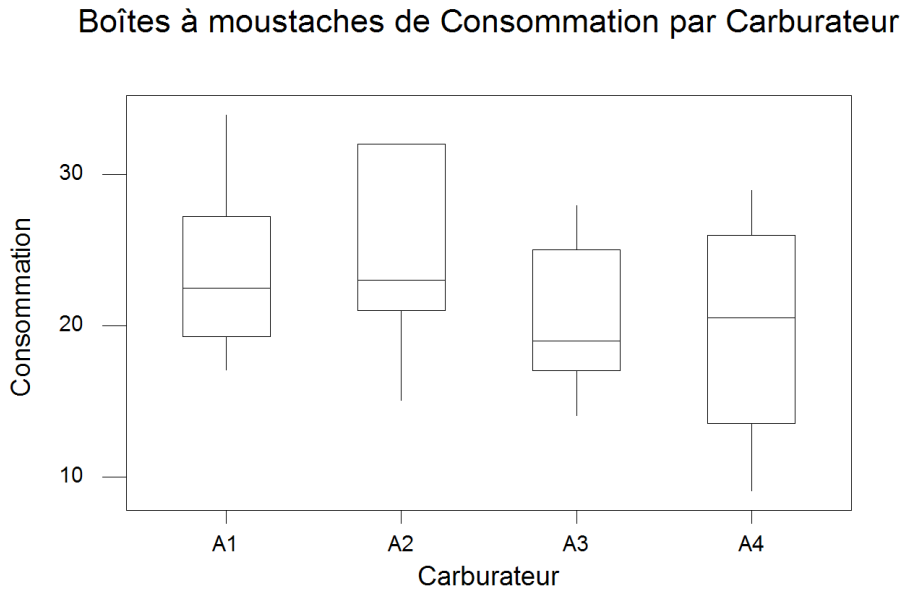
Partie I :

Dans cette partie **on ne tient pas compte** de la possible influence de l'ordre dans lequel les essais ont été effectués.

Une représentation graphique de la *Consommation* en fonction du *Carburateur* est donnée par la figure ci-après.

a) Proposer une méthode statistique permettant d'étudier l'influence des modalités du facteur *Carburateur* sur la *Consommation*. Énoncer le modèle et les hypothèses nécessaires au modèle que vous projetez d'utiliser. Ce modèle comporte-t-il des répétitions ?

b) Il y a-t-il des différences entre les *Carburateur* ? Quelles sont les estimations des coefficients du modèle ? Si nécessaire, comparer les différents niveaux du facteur *Carburateur*.



Éléments de corrigé : Carburateurs

Partie I :

Dans cette partie **on ne tient pas compte** de la possible influence de l'ordre dans lequel les essais ont été effectués.

a) La variable à expliquer, *Consommation*, est une variable continue. La variable explicative que l'on considère, le type de carburateur, *Carburateur*, est qualitative et contrôlée. Le plan qui a été utilisé pour réaliser l'expérience comporte des répétitions, on peut donc essayer de se servir d'un modèle d'analyse de la variance à un facteur contrôlé.

On introduit le modèle :

$$Y_{i,j} = \mu + \alpha_i + \epsilon_{i,j}, \quad i = 1 \dots 4,\ j = 1 \dots 6,$$
$$\text{avec la contrainte supplémentaire } \sum_{i=1}^{4} \alpha_i = 0,$$

où $Y_{i,j}$ la consommation de la voiture équipée du carburateur i lors du $j-$ème essai. On postule les hypothèses classiques suivantes pour les erreurs :

$$\forall\ (i,j), 1 \leqslant i \leqslant 4,\ 1 \leqslant j \leqslant 6,\ \mathcal{L}(\epsilon_{i,j}) = \mathcal{N}(0, \sigma^2),$$
$$\mathrm{Cov}(\epsilon_{i,j}, \epsilon_{k,l}) = 0 \text{ si } (i,j) \neq (k,l) \text{ avec } 1 \leqslant i, k \leqslant 4, \text{ et } 1 \leqslant j, l \leqslant 6.$$

Ce modèle comporte 6 répétitions pour chaque niveau du facteur. Il s'agit donc d'un plan expérimental équilibré.

b) On commence par vérifier que les conditions d'utilisation du modèle introduit à la question a) sont bien vérifiées. Pour cela on calcule les résidus du modèle.

1. Le protocole expérimental indique que l'on peut faire comme si chacun des 24 carburateurs utilisés au cours de l'expérience a été monté sur une même voiture. On suppose négligeable la dégradation des véhicules suite à leur utilisation répétée au cours de l'expérience. Ces considérations nous permettent de supposer que les erreurs $\epsilon_{i,j}$ sont indépendantes.
2. On effectue un test d'homogénéité des variances des erreurs :

$$\boxed{\mathcal{H}_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2}$$

contre

$$\boxed{\mathcal{H}_1 : \text{Il existe une variance différente des autres.}}$$

Puisque l'on ne connaît pas la loi des erreurs, on fait un test non-paramétrique, celui de Levene, sur les résidus du modèle :

```
Test de l'égalité des variances
Facteurs    Carburateur
Test de Levene (pour toute loi de probabilité continue)

Statistique du test : 0,194
P :                   0,899
```

Puisque la $p-$valeur du test est strictement supérieure au seuil $\alpha = 5$ %, le test est non significatif et l'on ne peut rejeter l'hypothèse nulle $\mathcal{H}_0$. Ainsi au niveau $\alpha = 5$ %, il n'y a pas de différences significatives entre les variances des erreurs.

3. On s'intéresse désormais à la normalité des erreurs. On cherche s'il l'on peut faire l'hypothèse que les résidus sont des réalisations d'une variable aléatoire ϵ dont la loi est une loi normale.

$$\boxed{\mathcal{H}_0 : \mathcal{L}(\epsilon) = \mathcal{N}}$$

contre

$$\boxed{\mathcal{H}_1 : \mathcal{L}(\epsilon) \neq \mathcal{N}.}$$

On utilise alors le test de Ryan-Joiner :

```
W-test pour la normalité
R:                              0,9868
Valeur de P (approximatif) : > 0,1000
```

Comme la $p-$valeur est strictement supérieure à $0,1$, elle est *a fortiori* strictement supérieure à $0,05$ qui est le seuil du test. On en déduit que le test n'est pas significatif au niveau $\alpha = 5$ % et que l'on ne peut donc pas rejeter l'hypothèse nulle $\mathcal{H}_0$. Il n'y a pas de différences significatives entre la loi de ϵ et une loi normale, au seuil $\alpha = 5$ %.

4. Comme l'hypothèse de normalité des erreurs n'a pas été rejetée à la question 3., on s'intéresse à nouveau à l'hypothèse d'homoscédasticité portant sur les erreurs.

$$\boxed{\mathcal{H}_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2}$$

contre

$$\boxed{\mathcal{H}_1 : \text{Il existe une variance différente des autres.}}$$

Il est maintenant possible d'utiliser le test paramétrique de Bartlett puisque ses conditions d'application, normalité des erreurs, sont vérifiées.

```
Test de l'égalité des variances
Facteurs    Carburateur
Test de Bartlett (loi normale)

Statistique du test : 0,650
P :                   0,885
```

Le test n'est pas significatif car $0,885$ est strictement supérieur à $\alpha = 5$ %. On ne peut rejeter l'hypothèse nulle $\mathcal{H}_0$ au seuil $\alpha = 5$ %.

Toutes les conditions d'application du modèle de l'analyse de la variance à un facteur sont vérifiées. Déterminons si le facteur *Carburateur* a un effet sur la *Consommation*. On teste donc les hypothèses :

$$\boxed{\mathcal{H}_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \text{Il existe } i_0 \in \{1,2,3,4\} \text{ tel que } \alpha_{i_0} \neq 0.}$$

On consulte alors le tableau de l'analyse de la variance

```
Facteur      Type Niveaux  Valeurs
Carburat     fixe       4  A1 A2 A3 A4

Analyse de la variance pour Consomma, en utilisant la SC
ajustée pour
les tests

Source     DL     SC séq   SC ajust   CM ajust       F      P
Carburat    3     100,83     100,83      33,61    0,89  0,464
Erreur     20     757,00     757,00      37,85
Total      23     857,83
```

La $p-$valeur associée, par la statistique de Fisher, à l'hypothèse nulle $\mathcal{H}_0$ est de $0,464$. Elle est strictement supérieure à $\alpha = 5$ % : le test n'est pas significatif à ce seuil. On ne peut rejeter l'hypothèse nulle $\mathcal{H}_0$ d'absence d'effet du facteur *Carburateur* sur la *Consommation*. Ainsi au niveau $\alpha = 5$ %, le facteur *Carburateur* n'a pas d'effet significatif sur la *Consommation*. En prenant cette décision, on risque de commettre une erreur de deuxième espèce, il convient donc de calculer la puissance du test. À l'aide de l'entrée **ANOVA à un facteur contrôlé** du sous menu **Puissance et effectif de l'échantillon** du menu **Stat** de Minitab, on obtient le résultat suivant :

```
Puissance et effectif de l'échantillon

ANOVA à un facteur contrôlé

Sigma = 6,152  Alpha = 0,05  Nombre de niveaux = 4

               Effectif de              Différence
Moyennes SC  l'échantillon  Puissance  maximale
    11,6838              6     0,1555  4,834
```

La *Différence maximale* correspond à l'écart le plus important en valeur absolue entre les estimations des différents niveaux du facteur. Ici c'est la valeur de $\widehat{A_4} - \widehat{A_2} = 4,834$ qui est la plus élevée. La puissance a posteriori est donc très faible : 0,1555 par rapport à la valeur de $0,80$ qui est considérée comme étant acceptable. On aurait pu améliorer la puissance du test en augmentant le nombre de répétitions.

```
Puissance et effectif de l'échantillon

ANOVA à un facteur contrôlé

Sigma = 6,152  Alpha = 0,05  Nombre de niveaux = 4

               Effectif de  Puissance  Puissance  Différence
Moyennes SC  l'échantillon      cible     réelle  maximale
    11,6838             37     0,8000     0,8084  4,834
```

On constate qu'il aurait fallu 37 répétitions, donc 148 carburateurs, pour obtenir une puissance d'au moins 0,80. On aurait multiplié par plus de six l'ampleur de l'étude qui a été réalisée...

Les estimations des coefficients du modèle sont :

```
Terme           Coef      Er-T coef        T      P
Constante     22,083          1,256    17,58  0,000
```

```
Carburat
A1            1,417          2,175     0,65  0,522
A2            2,583          2,175     1,19  0,249
A3           -1,750          2,175    -0,80  0,431
```

On en déduit que :

$$\widehat{\mu}(\boldsymbol{y}) = 22,083, \quad \widehat{\alpha_1}(\boldsymbol{y}) = 1,417, \quad \widehat{\alpha_2}(\boldsymbol{y}) = 2,583,$$
$$\widehat{\alpha_3}(\boldsymbol{y}) = -1,750, \quad \widehat{\alpha_4}(\boldsymbol{y}) = -2,250,$$

en utilisant la relation $\sum_{i=1}^{4} \widehat{\alpha_i}(\boldsymbol{y}) = 0$ qui se déduit de la relation vérifiée par les estimateurs $\sum_{i=1}^{4} \widehat{\alpha_i} = 0$.
On remarque que c'est le carburateur A_4 qui fait consommer le moins de carburant. Toutefois, puisque l'on n'a pas pu rejeter l'hypothèse nulle $\mathcal{H}_0$ d'absence d'effet du facteur *Carburateur* sur la *Consommation* au seuil $\alpha = 5$ %, ce comportement n'est pas significativement différent des autres au seuil $\alpha = 5$ %.

Au seuil $\alpha = 5$ %, l'hypothèse nulle $\mathcal{H}_0$ d'absence d'effet du facteur *Carburateur* sur la *Consommation* n'a pu être réfutée, **on de doit donc pas faire de comparaisons multiples** de l'effet des différents niveaux du facteur *Carburateur*.

12.2. Analyse de la variance à deux facteurs

12.2.1. Sans répétition : Carburateurs (suite et fin)[20]

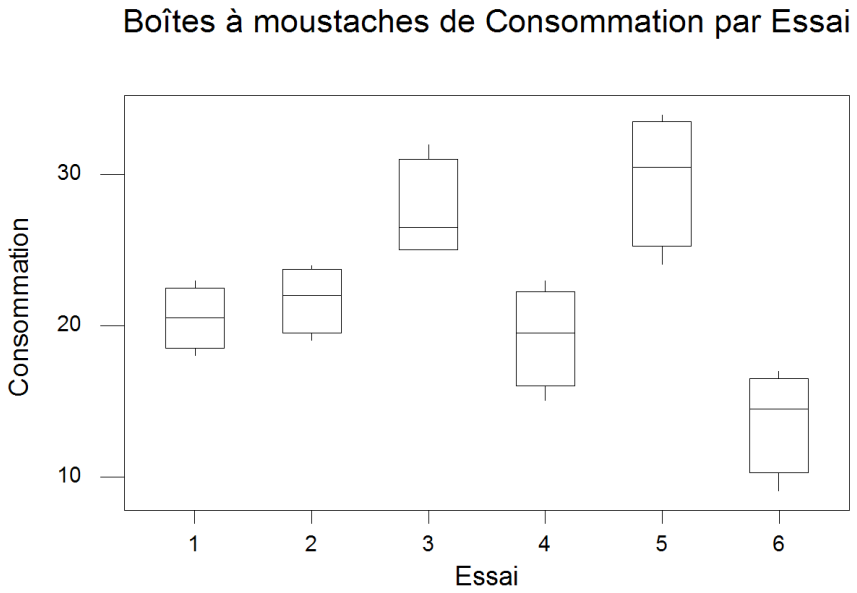
Partie II :

Après avoir effectué l'analyse de la première partie on apprend que tous les essais numéro 1 ont été réalisés un lundi, tous les essais numéro 2 un mardi, ..., tous les essais numéro 6 un samedi.

On décide donc dans cette partie **de tenir compte** de la possible influence de l'ordre de réalisation des essais, c'est-à-dire du facteur *Essai*.

Une représentation graphique de la *Consommation* en fonction de *Essai* est donnée par la figure au ci-dessous.

[20] Les données de cet exercice sont tirées du livre de Georges Parreins [8].



a) Proposer une méthode statistique permettant d'étudier conjointement l'influence du facteur *Carburateur* et du facteur *Essai* sur la consommation. Énoncer le modèle et les hypothèses nécessaires au modèle que vous projetez d'utiliser. Ce modèle comporte-t-il des répétitions ?

b) Il y a-t-il des différences entre les carburateurs ? Il y a-t-il des différences dues à l'ordre de réalisation des essais ? Quelles sont les estimations des paramètres du modèle ? Si nécessaire, comparer les différents niveaux du facteur *Carburateur* ainsi que les différents niveaux du facteur *Essai*.

. . . . . .

Éléments de corrigé : Carburateurs (suite et fin)

Partie II :

Après avoir effectué l'analyse de la première partie on apprend que tous les essais numéro 1 ont été réalisés un lundi, tous les essais numéro 2 un mardi, ..., tous les essais numéro 6 un samedi.

On décide donc dans cette partie **de tenir compte** de la possible influence de l'ordre de réalisation des essais, c'est-à-dire du facteur *Essai*.

a) La variable à expliquer, *Consommation*, est une variable continue. Les variables explicatives que l'on considère, le type de carburateur, *Carburateur*, et le jour de l'essai, *Essai*, sont qualitatives et contrôlées. Le plan qui a été utilisé pour réaliser l'expérience ne comporte pas de répétitions, on peut donc essayer de se servir d'un modèle d'analyse de la variance à deux facteurs contrôlés sans répétition.

On introduit le modèle :

$$Y_{i,j} = \mu + \alpha_i + \beta_j + \epsilon_{i,j}, \quad i = 1 \dots 4,\ j = 1 \dots 6,$$
$$\text{avec les contraintes supplémentaires } \sum_{i=1}^{4} \alpha_i = 0 \text{ et } \sum_{j=1}^{6} \beta_j = 0,$$

où $Y_{i,j}$ la consommation de la voiture équipée du carburateur i lors du $j-$ème essai. On postule les hypothèses classiques suivantes pour les erreurs :

$$\forall\ (i,j), 1 \leqslant i \leqslant 4,\ 1 \leqslant j \leqslant 6,\ \mathcal{L}(\epsilon_{i,j}) = \mathcal{N}(0, \sigma^2),$$
$$\mathrm{Cov}(\epsilon_{i,j}, \epsilon_{k,l}) = 0 \text{ si } (i,j) \neq (k,l) \text{ avec } 1 \leqslant i,k \leqslant 4 \text{ et } 1 \leqslant j,l \leqslant 6.$$

Ce modèle ne comporte aucune répétition pour aucun des niveaux du facteur. Il s'agit d'un plan expérimental équilibré.

b) On commence par vérifier que les conditions d'utilisation du modèle introduit à la question a) sont bien vérifiées. Pour cela on calcule les résidus du modèle.

1. Le protocole expérimental indique que l'on peut faire comme si chacun des 24 carburateurs utilisés au cours de l'expérience a été monté sur une même voiture. On suppose négligeable la dégradation des véhicules suite à leur utilisation répétée au cours de l'expérience. Ces considérations nous permettent de supposer que les erreurs $\epsilon_{i,j}$ sont indépendantes.
2. Il n'est pas possible de tester l'hypothèse d'égalité des variances des erreurs $\mathcal{H}_0$ ci-dessous à l'aide des données expérimentales dont on dispose puisque le plan utilisé ne comporte pas de répétitions.

$$\boxed{\mathcal{H}_0 : \sigma^2_{1,1} = \dots = \sigma^2_{4,1} = \dots = \sigma^2_{1,6} = \dots = \sigma^2_{4,6}}$$

contre

$$\boxed{\mathcal{H}_1 : \text{Il existe une variance différente des autres.}}$$

Par contre on sait que si l'hypothèse nulle globale $\mathcal{H}_0$ ci-dessus est vérifiée, alors chacune des deux hypothèses nulles $\mathcal{H}_0^{'}$ d'homogénéité des variances des erreurs par rapport au facteur *Carburateur* et $\mathcal{H}_0^{''}$ d'homogénéité des variances des erreurs par rapport au facteur *Essai* doivent l'être également. On effectue deux tests d'homogénéité des variances des erreurs :

$$\boxed{\mathcal{H}_0^{'} : \sigma_{c,1}^2 = \sigma_{c,2}^2 = \sigma_{c,3}^2 = \sigma_{c,4}^2}$$

contre

$$\boxed{\mathcal{H}_1^{'} : \text{Il existe une variance différente des autres pour } Carburateur.}$$

$$\boxed{\mathcal{H}_0^{''} : \sigma_{e,1}^2 = \sigma_{e,2}^2 = \sigma_{e,3}^2 = \sigma_{e,4}^2 = \sigma_{e,5}^2 = \sigma_{e,6}^2}$$

contre

$$\boxed{\mathcal{H}_1^{''} : \text{Il existe une variance différente des autres pour } Essai.}$$

Puisque l'on ne connaît pas la loi des erreurs, on utilise un test non-paramétrique, celui de Levene, sur les résidus du modèle :

```
Test de l'égalité des variances
Facteurs    Carburateur

Test de Levene (pour toute loi de probabilité continue)

Statistique du test : 0,842
P :                   0,487

Test de l'égalité des variances
Facteurs    Essai

Test de Levene (pour toute loi de probabilité continue)

Statistique du test : 0,925
P :                   0,488
```

Puisque les $p-$valeurs des tests sont strictement supérieures au seuil $\alpha = 5$ %, les tests sont non significatifs et l'on ne peut rejeter les hypothèses nulles $\mathcal{H}_0^{'}$ et $\mathcal{H}_0^{''}$.

3. On s'intéresse désormais à la normalité des erreurs. On cherche s'il l'on peut faire l'hypothèse que les résidus sont des réalisations d'une variable aléatoire ϵ dont la loi est une loi normale.

$$\boxed{\mathcal{H}_0 : \mathcal{L}(\epsilon) = \mathcal{N}}$$

contre

$$\boxed{\mathcal{H}_1 : \mathcal{L}(\epsilon) \neq \mathcal{N}.}$$

On utilise alors le test de Ryan-Joiner :

```
W-test pour la normalité
R:                              0,9725
Valeur de P (approximatif) : > 0,1000
```

Comme la $p-$valeur est strictement supérieure à $0,1$, elle est *a fortiori* strictement supérieure à $0,05$ qui est le seuil α du test. On en déduit que le test n'est pas significatif au niveau $\alpha = 5$ % et que l'on ne peut donc pas rejeter l'hypothèse nulle $\mathcal{H}_0$. Il n'y a pas de différences significatives entre la loi de ϵ et une loi normale, au seuil $\alpha = 5$ %.

4. Comme l'hypothèse de normalité des erreurs n'a pas été rejetée à la question 3., on s'intéresse à nouveau à l'hypothèse d'homoscédasticité portant sur les erreurs.
   Pour les mêmes raisons qu'au 2., il n'est pas possible de tester l'hypothèse d'égalité des variances des erreurs $\mathcal{H}_0$:

$$\boxed{\mathcal{H}_0 : \sigma^2_{1,1} = \cdots = \sigma^2_{4,1} = \cdots = \sigma^2_{1,6} = \cdots = \sigma^2_{4,6}}$$

contre

$$\boxed{\mathcal{H}_1 : \text{Il existe une variance différente des autres.}}$$

   On s'intéresse aux deux hypothèses nulles $\mathcal{H}_0^{'}$ d'homogénéité des variances des erreurs par rapport au facteur $Carburateur$ et $\mathcal{H}_0^{''}$ d'homogénéité des variances des erreurs par rapport au facteur $Essai$. On effectue deux tests d'homogénéité des variances des erreurs :

$$\boxed{\mathcal{H}_0^{'} : \sigma^2_{c,1} = \sigma^2_{c,2} = \sigma^2_{c,3} = \sigma^2_{c,4}}$$

contre

$$\boxed{\mathcal{H}_1^{'} : \text{Il existe une variance différente des autres pour } Carburateur.}$$

$$\boxed{\mathcal{H}_0^{''} : \sigma^2_{e,1} = \sigma^2_{e,2} = \sigma^2_{e,3} = \sigma^2_{e,4} = \sigma^2_{e,5} = \sigma^2_{e,6}}$$

contre

$$\boxed{\mathcal{H}_1^{''} : \text{Il existe une variance différente des autres pour } Essai.}$$

   Il est maintenant possible d'utiliser le test paramétrique de Bartlett puisque ses conditions d'application, normalité des erreurs, sont vérifiées.

```
Test de l'égalité des variances
Facteurs    Carburateur

Test de Bartlett (loi normale)

Statistique du test : 1,930
P :                   0,587

Test de l'égalité des variances
Facteurs    Essai

Test de Bartlett (loi normale)
```

```
Statistique du test : 3,525
P :                   0,620
```

Puisque les $p-$valeurs des tests sont strictement supérieures au seuil $\alpha = 5$ %, les tests sont non significatifs et l'on ne peut rejeter les hypothèses nulles $\mathcal{H}_0^{'}$ et $\mathcal{H}_0^{''}$.

Toutes les conditions d'application du modèle de l'analyse de la variance à deux facteurs contrôlés sans répétition sont vérifiées. Déterminons si le facteur *Carburateur* ou le facteur *Essai* a un effet sur la *Consommation*. On teste donc les hypothèses :

$$\boxed{\mathcal{H}_0^{'} : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0}$$

contre

$$\boxed{\mathcal{H}_1^{'} : \text{Il existe } i_0 \in \{1, 2, 3, 4\} \text{ tel que } \alpha_{i_0} \neq 0.}$$

$$\boxed{\mathcal{H}_0^{''} : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0}$$

contre

$$\boxed{\mathcal{H}_1^{''} : \text{Il existe } j_0 \in \{1, 2, 3, 4, 5, 6\} \text{ tel que } \beta_{j_0} \neq 0.}$$

On consulte alors le tableau de l'analyse de la variance

```
Facteur      Type Niveaux  Valeurs
Carburat     fixe       4  A1 A2 A3 A4
Essai        fixe       6  1 2 3 4 5 6

Analyse de la variance pour Consomma, en utilisant la SC
ajustée pour
les tests

Source     DL     SC séq   SC ajust   CM ajust       F      P
Carburat    3    100,833    100,833     33,611    5,99  0,007
Essai       5    672,833    672,833    134,567   23,98  0,000
Erreur     15     84,167     84,167      5,611
Total      23    857,833
```

La $p-$valeur associée, par la statistique de Fisher, à l'hypothèse nulle $\mathcal{H}_0^{'}$ est de $0,007$. Elle est inférieure ou égale à $\alpha = 5$ % : le test est significatif à ce seuil. On doit rejeter l'hypothèse nulle $\mathcal{H}_0^{'}$ d'absence d'effet du facteur *Carburateur* sur la *Consommation* et décider l'hypothèse alternative $\mathcal{H}_1^{'}$ d'existence d'un effet du facteur *Carburateur* sur la *Consommation*. Ainsi au niveau $\alpha = 5$ %, le facteur *Carburateur* a un effet significatif sur la *Consommation*. En prenant cette décision, on risque de commettre une erreur de première espèce, le risque associé à la décision est donc de 5 %.

La $p-$valeur associée, par la statistique de Fisher, à l'hypothèse nulle $\mathcal{H}_0^{''}$ est de $0,000$. Elle est inférieure ou égale à $\alpha = 5$ % : le test est significatif à ce seuil. On doit

rejeter l'hypothèse nulle $\mathcal{H}_0''$ d'absence d'effet du facteur *Essai* sur la *Consommation* et décider l'hypothèse alternative $\mathcal{H}_1''$ d'existence d'un effet du facteur *Essai* sur la *Consommation*. Ainsi au niveau $\alpha = 5$ %, le facteur *Essai* a un effet significatif sur la *Consommation*. En prenant cette décision, on risque de commettre une erreur de première espèce, le risque associé à la décision est donc de 5 %.
On constate que, contrairement à ce que l'on avait pu montrer dans la partie I, on arrive à mettre en évidence un effet du facteur *Carburateur* sur la consommation. Il semblait nécessaire de prendre en compte ce facteur supplémentaire, *Essai*, à la vue des boîtes à moustaches qui étaient reproduites dans l'énoncé. La variabilité importante associée au jour où l'essai a été réalisé, liée par exemple à la densité du traffic automobile, masquait les variations liées au facteur *Carburateur*.
Les estimations des coefficients du modèle sont :

Terme	Coef	Er-T coef	T	P
Constante	22,0833	0,4835	45,67	0,000
Carburat				
A1	1,4167	0,8375	1,69	0,111
A2	2,5833	0,8375	3,08	0,008
A3	-1,7500	0,8375	-2,09	0,054
Essai				
1	-1,583	1,081	-1,46	0,164
2	-0,333	1,081	-0,31	0,762
3	5,417	1,081	5,01	0,000
4	-2,833	1,081	-2,62	0,019
5	7,667	1,081	7,09	0,000

On en déduit que :

$$\begin{array}{lll} \widehat{\mu}(\boldsymbol{y}) = 22,083, & \widehat{\alpha_1}(\boldsymbol{y}) = 1,417, & \widehat{\alpha_2}(\boldsymbol{y}) = 2,583, \\ & \widehat{\alpha_3}(\boldsymbol{y}) = -1,750, & \widehat{\alpha_4}(\boldsymbol{y}) = -2,250, \\ \widehat{\beta}_1(\boldsymbol{y}) = -1,583, & \widehat{\beta}_2(\boldsymbol{y}) = -0,333, & \widehat{\beta}_3(\boldsymbol{y}) = 5,417, \\ \widehat{\beta}_4(\boldsymbol{y}) = -2,833, & \widehat{\beta}_5(\boldsymbol{y}) = 7,667, & \widehat{\beta}_6(\boldsymbol{y}) = -8,335, \end{array}$$

en utilisant les relations $\sum_{i=1}^{4} \widehat{\alpha}_i(\boldsymbol{y}) = 0$ et $\sum_{j=1}^{6} \widehat{\beta}_j(\boldsymbol{y}) = 0$ qui sont des conséquences de celles vérifiées par les estimateurs $\sum_{i=1}^{4} \widehat{\alpha}_i = 0$ et $\sum_{j=1}^{6} \widehat{\beta}_j = 0$.
On remarque que les estimations des coefficients associés aux carburateurs sont les mêmes que celles obtenues à la partie I. Il en va toujours de même lorsque le plan est équilibré.

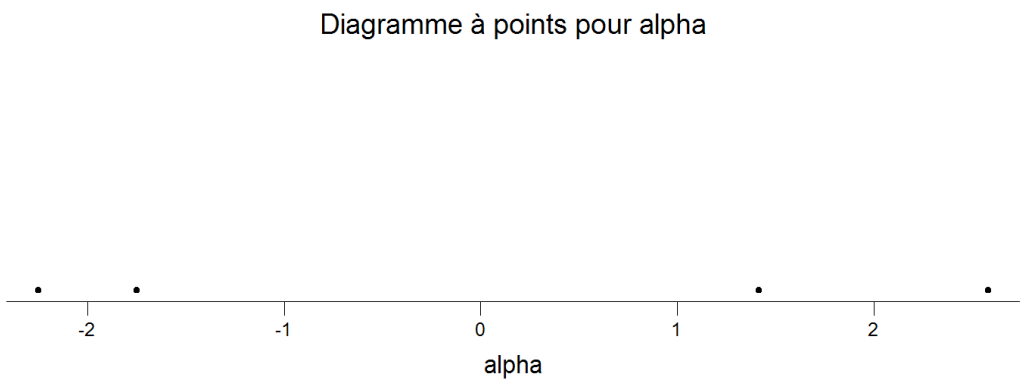
Au seuil $\alpha = 5$ %, l'hypothèse nulle $\mathcal{H}_0'$ d'absence d'effet du facteur *Carburateur* sur la *Consommation* a été rejetée, **on peut donc faire des comparaisons multiples** de l'effet des différents niveaux du facteur *Carburateur*.

```
Tests de simultanéité de Tukey
Variable de réponse Consomma
```

```
Toutes comparaisons deux à deux entre niveaux de Carburat

Carburat = A1 soustraites de :

Niveau          Différence        Er-T de     Valeur   Valeur ajustée
Carburat      des moyennes   la différence      de T            de P
A2                   1,167           1,368     0,853          0,8284
A3                  -3,167           1,368    -2,315          0,1385
A4                  -3,667           1,368    -2,681          0,0726

Carburat = A2 soustraites de :

Niveau          Différence        Er-T de     Valeur   Valeur ajustée
Carburat      des moyennes   la différence      de T            de P
A3                  -4,333           1,368    -3,169          0,0290
A4                  -4,833           1,368    -3,534          0,0142

Carburat = A3 soustraites de :

Niveau          Différence        Er-T de     Valeur   Valeur ajustée
Carburat      des moyennes   la différence      de T            de P
A4                 -0,5000           1,368   -0,3656          0,9827
```

Ainsi au seuil global de décision $\alpha_{global} = 5$ % les niveaux A_2 et A_3, A_2 et A_4 sont significativement différents, pour les autres l'hypothèse d'égalité ne peut être rejetée.

On a représenté, dans l'ordre de gauche à droite sur la figure, les estimations des valeurs des niveaux A_3, A_1, A_4, A_2 du facteur *Carburateur*.

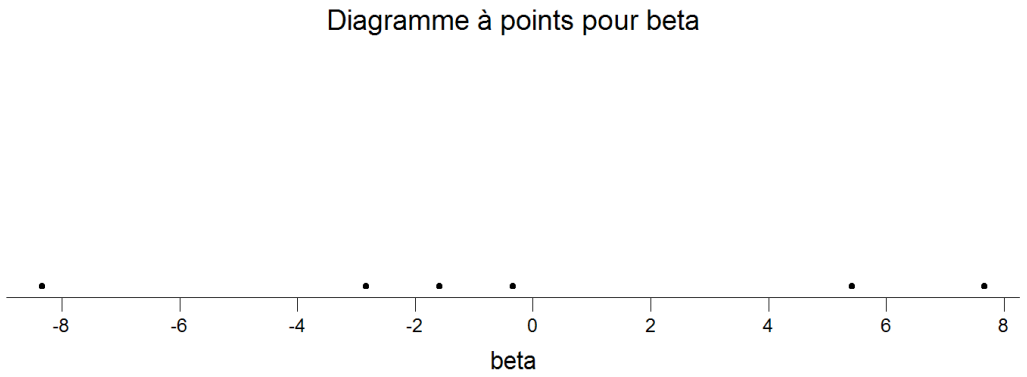


Au seuil $\alpha = 5$ %, l'hypothèse nulle $\mathcal{H}_0''$ d'absence d'effet du facteur *Essai* sur la *Consommation* a été rejetée, **on peut donc faire des comparaisons multiples** de l'effet des différents niveaux du facteur *Essai*.

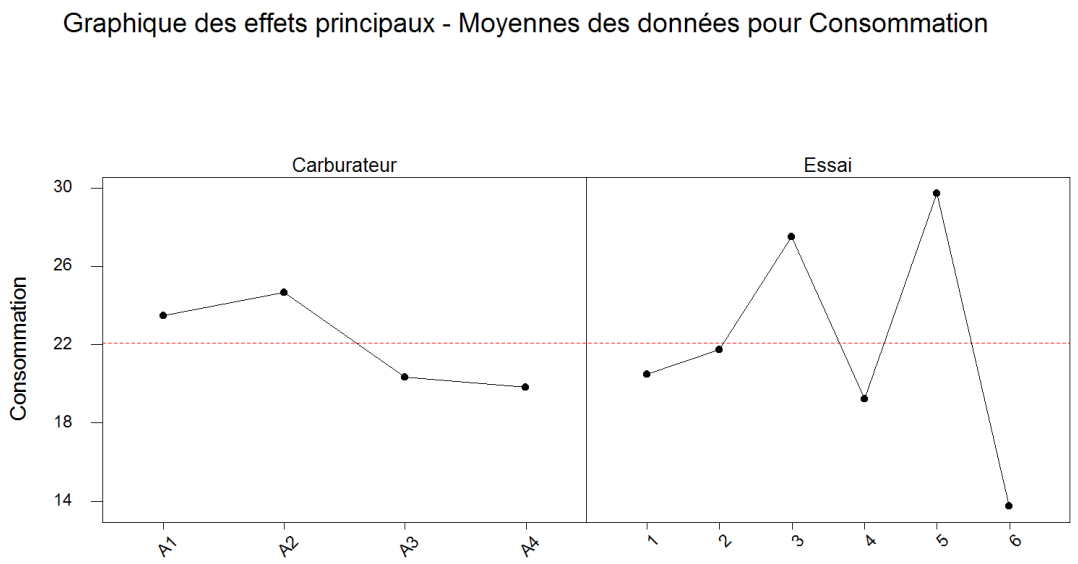
```
Tests de simultanéité de Tukey
Variable de réponse Consomma
Toutes comparaisons deux à deux entre niveaux de Essai

Essai = 1 soustraites de :

Niveau        Différence         Er-T de     Valeur   Valeur ajustée
Essai       des moyennes   la différence       de T             de P
2                  1,250           1,675      0,746           0,9725
3                  7,000           1,675      4,179           0,0086
4                 -1,250           1,675     -0,746           0,9725
5                  9,250           1,675      5,522           0,0007
6                 -6,750           1,675     -4,030           0,0114

Essai = 2 soustraites de :

Niveau        Différence         Er-T de     Valeur   Valeur ajustée
Essai       des moyennes   la différence       de T             de P
3                  5,750           1,675      3,433           0,0355
4                 -2,500           1,675     -1,493           0,6738
5                  8,000           1,675      4,776           0,0027
6                 -8,000           1,675     -4,776           0,0027

Essai = 3 soustraites de :

Niveau        Différence         Er-T de     Valeur   Valeur ajustée
Essai       des moyennes   la différence       de T             de P
4                  -8,25           1,675     -4,925           0,0021
5                   2,25           1,675      1,343           0,7578
6                 -13,75           1,675     -8,209           0,0000

Essai = 4 soustraites de :

Niveau        Différence         Er-T de     Valeur   Valeur ajustée
Essai       des moyennes   la différence       de T             de P
5                 10,500           1,675      6,269           0,0002
6                 -5,500           1,675     -3,284           0,0469

Essai = 5 soustraites de :

Niveau        Différence         Er-T de     Valeur   Valeur ajustée
Essai       des moyennes   la différence       de T             de P
6                 -16,00           1,675     -9,552           0,0000
```

Ainsi au seuil global de décision $\alpha_{global} = 5$ % les niveaux 1 et 3, 1 et 5, 2 et 3, 2 et 5, 2 et 6, 3 et 4, 3 et 6, 4 et 5, 5 et 6 sont significativement différents, pour les autres l'hypothèse d'égalité ne peut être rejetée.

On a représenté, dans l'ordre de gauche à droite sur la figure, les estimations des valeurs des niveaux 4, 1, 2, 3, 5, 6 du facteur *Essai*.



À l'aide de l'entrée **Graphique des effets principaux** du sous menu **ANOVA** du menu **Stat** de Minitab, on obtient la représentation graphique pertinente suivante :



La ligne en pointillé en rouge[21] est la valeur de l'estimation de la moyenne $\widehat{\mu}(\boldsymbol{y})$. Les $\widehat{\alpha}_i(\boldsymbol{y})$, $1 \leqslant i \leqslant 4$, et les $\widehat{\beta}_j(\boldsymbol{y})$, $1 \leqslant j \leqslant 6$, sont représentées par les valeurs des écarts à cette ligne.

[21] En gris si vous lisez une version imprimée du corrigé de cet exercice.

. . . . . .

12.2.2. Avec répétitions et plan équilibré I : Comparaison de l'appréciation de huit eaux gazeuses[22]

Un panel de 16 juges évalue 8 types différents d'eau gazeuze. Chaque évalue 4 fois chaque eau. On suppose que les hypothèses, relatives à l'utilisation du modèle d'analyse de la variance qui a permis d'obtenir les résultats reportés dans les tableaux ci-dessous, sont bien vérifiées.

1. Décrire succintement le jeu de données (nombre total d'observations, nature des variables, etc.)
2. Construire le test de significativité de l'interaction $eau \times juge$ (hypothèses, statistique de test, loi de cette statistique sous $\mathcal{H}_0$, décision à l'aide du tableau ci-dessous).
3. Commenter le tableau d'analyse de la variance ci-dessous.

	ddl	SCE	CM	f_{obs}	$Proba$
juge	15	731,3574	48,75716	25,03878	$5{,}91658\times10^{-48}$
eau	7	38,6543	5,52204	2,83579	0,006810201
juge : eau	105	508,5645	4,84347	2,48732	$1{,}19182\times10^{-10}$
Résiduelle	384	747,7500	1,94727		

4. L'estimation des coefficients pour la variable eau est fournie dans le listage ci-dessous. Quelle est la plus appréciée ? Lors d'un rassemblement avec beaucoup d'individus, utiliseriez-vous seulement cette eau ?

	Value	Std. Error	t value	$Pr(>\|t\|)$
(Intercept)	3,0722656	0,06167052	49,8174098	$0{,}000000\times10^{+000}$
St Yorre	-0,1660156	0,16316486	-1,017472	$3{,}095697\times10^{-001}$
Badoit	-0,1816406	0,16316486	-1,113234	$2{,}663044\times10^{-001}$
Vichy	-0,2441406	0,16316486	-1,496282	$1{,}354016\times10^{-001}$
Quézac	0,4277344	0,16316486	2,621486	$9{,}102684\times10^{-003}$
Arvie	-0,2128906	0,16316486	-1,304758	$1{,}927567\times10^{-001}$
Chateauneuf	0,2714844	0,16316486	1,663865	$9{,}695512\times10^{-002}$
Salvetat	0,3496094	0,16316486	2,142676	$3{,}276727\times10^{-001}$
Perrier	-0,2441406	0,16316486	-1,496282	$1{,}354016\times10^{-001}$

. . . . . .

[22] Les données de cet exercice sont tirées du livre d'exercices de François Husson et de Jérôme Pagès [2].

Éléments de corrigé : Comparaison de l'appréciation de huit eaux gazeuses

1. Il y a $8 \times 16 \times 4 = 512$ observations. Pour chaque observation, on dispose du nom du juge qui l'a effectuée (variable qualitative), de l'eau concernée (variable qualitative) et de la note mise (variable quantitative).
2. On introduit le modèle :

$$Y_{i,j,k} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{i,j} + \epsilon_{i,j,k}, \quad i = 1 \dots 8,\ j = 1 \dots 16,\ k = 1 \dots 4$$

$$\text{avec les contraintes supplémentaires } \sum_{i=1}^{8} \alpha_i = 0,\ \sum_{j=1}^{16} \beta_j = 0,$$

$$\sum_{i=1}^{8} (\alpha\beta)_{i,j} = 0,\ \forall j \in \{1, \dots, 16\} \text{ et } \sum_{j=1}^{16} (\alpha\beta)_{i,j} = 0,\ \forall i \in \{1, \dots, 8\},$$

où $Y_{i,j,k}$ est la note du juge i lors de l'évaluation de l'eau j au cours du $k-$ème essai. On postule les hypothèses classiques suivantes pour les erreurs :

$$\forall\ (i,j,k), 1 \leqslant i \leqslant 8,\ 1 \leqslant j \leqslant 16,\ 1 \leqslant k \leqslant 4,\ \mathcal{L}(\epsilon_{i,j,k}) = \mathcal{N}(0, \sigma^2),$$
$$\text{et } \mathrm{Cov}(\epsilon_{i,j,k}, \epsilon_{l,m,n}) = 0 \text{ si } (i,j,k) \neq (l,m,n) \text{ avec}$$
$$1 \leqslant i, l \leqslant 8,\ 1 \leqslant j, m \leqslant 16 \text{ et } 1 \leqslant k, n \leqslant 4.$$

Ce modèle comporte quatre répétitions pour chacun des niveaux du facteur. Il s'agit d'un plan expérimental équilibré.

Pour tester la significativité de l'interaction, on construit le test suivant :

$$\boxed{\mathcal{H}_0 : (\alpha\beta)_{1,1} = (\alpha\beta)_{1,2} = \dots = (\alpha\beta)_{1,J} = (\alpha\beta)_{2,1} = \dots = (\alpha\beta)_{I,J} = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \text{Il existe } (i_0, j_0) \in \{1, 2, \dots, I\} \times \{1, 2, \dots, J\} \text{ tel que } (\alpha\beta)_{i_0,j_0} \neq 0.}$$

La statistique de test est :

$$F = \frac{CM_{\text{interaction}}}{CM_{\text{résiduelle}}}.$$

Ici : $ddl_{\text{interaction}} = (I-1)(J-1) = (8-1)(16-1) = 105$ et $ddl_{\text{résiduelle}} = n - IJ = 512 - 128 = 384$. Sous l'hypothèse nulle $\mathcal{H}_0$ et les hypothèses du modèle, F suit une loi de Fisher à 105 et 384 degrés de liberté. Décision : la $p-$valeur vaut $1,192 \times 10^{10}$, le test est donc significatif au seuil $\alpha = 5$ %. On considère donc que l'interaction entre le facteur *juge* et le facteur *eau* est significative, c'est-à-dire que tous les juges n'ont pas les mêmes préférences.

3. Les tests sur les effets principaux *eau* et *juge* se construisent de la même façon que le test de l'interaction. Les effets *eau* et *juge* sont donc significatifs : certaines eaux sont globalement plus appréciées que d'autres ; certains juges mettent en moyenne des notes plus élevées que d'autres.

4. L'eau la plus appréciée est l'eau Quézac puisque cette eau a le coefficient $\widehat{\alpha}_i(\boldsymbol{y})$ le plus élevé (0,428). Mais comme l'interaction *juge* $\times$ *eau* est significative, tous les juges ne préfèrent pas forcément l'eau Quézac. A priori, lors d'un rassemblement, il sera bon de prévoir d'autres eaux minérales.

. . . . . .

12.2.3. Avec répétitions et plan équilibré II : Rats et régimes[23]

On teste l'influence de différents régimes alimentaires sur des rats de laboratoire.

Le gain de poids des rats est désigné par la variable $Poids$, exprimée en grammes, les deux facteurs sont les variables $Calorie$ et $Vitamine$. La variable $Calorie$ vaut 0 si les rats n'ont pas suivi un régime hypercalorique et 1 s'ils ont suivi un tel régime hypercalorique. La variable $Vitamine$ vaut 0 si les rats n'ont pas reçu de compléments vitaminés et 1 s'ils ont reçu de tels compléments.

[23] Les données, fictives, de cet exercice sont tirées du livre de Bruno Falissard [5].

Calorie	Vitamine	Poids	Calorie	Vitamine	Poids
1	1	84	1	1	66
1	2	62	1	2	59
2	1	87	2	1	89
2	2	103	2	2	90
1	1	66	1	1	56
1	2	84	1	2	74
2	1	92	2	1	101
2	2	107	2	2	116
1	1	82	1	1	79
1	2	73	1	2	74
2	1	77	2	1	95
2	2	95	2	2	112
1	1	62	1	1	89
1	2	75	1	2	74
2	1	88	2	1	91
2	2	96	2	2	92

1. Quels modèles d'analyse de la variance à deux facteurs pouvez-vous utiliser pour étudier ces données. On décide de retenir, pour répondre aux questions suivantes, le modèle le plus complet parmi ceux dont il est possible de se servir. Rappeler les hypothèses associées au modèle.
2. Procéder à l'étude à l'aide de Minitab.
3. Quelles sont les estimations des paramètres du modèle ?
4. Doit-on réaliser des comparaisons multiples ? Si oui pour quel facteur ? Le faire.

. . . . . .

Éléments de corrigé : Rats et régimes

1. La réponse observée, le gain de *Poids* exprimé en grammes, est considérée comme une variable quantitative. Le premier des deux facteurs, absence ou présence de compléments vitaminés, noté $Vitamine$, est qualitatif. Le second des deux facteurs, le type de régime suivi, hypocalorique ou non, noté $Calorie$, est également qualitatif. Le plan d'expérience comporte huit répétitions on peut de ce fait utiliser deux modèles d'analyse de la variance à deux facteurs, l'un ne comportant pas d'interaction entre les deux facteurs $Vitamine$ et $Calorie$, l'autre comportant cette interaction. Le modèle le plus complet parmi ces deux possibilités est celui de l'analyse de la variance avec répétitions et terme d'interaction. On décide de le retenir

pour la suite de l'étude.
On introduit le modèle :

$$Y_{i,j,k} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{i,j} + \epsilon_{i,j,k}, \quad i = 1,2,\ j = 1,2,\ k = 1\ldots 8$$
$$\text{avec les contraintes supplémentaires } \sum_{i=1}^{2} \alpha_i = 0,\ \sum_{j=1}^{2} \beta_j = 0,$$
$$\sum_{i=1}^{2} (\alpha\beta)_{i,j} = 0,\ \forall j \in \{1,2\} \text{ et } \sum_{j=1}^{2} (\alpha\beta)_{i,j} = 0,\ \forall i \in \{1,2\},$$

où $Y_{i,j,k}$ est le gain de poids dans la condition i du facteur $Vitamine$ et dans la condition j du facteur $Calorie$ lors de la $k-$ème répétition. On postule les hypothèses classiques suivantes pour les erreurs :

$$\forall\ (i,j,k), 1 \leqslant i \leqslant 2,\ 1 \leqslant j \leqslant 2,\ 1 \leqslant k \leqslant 8,\ \mathcal{L}(\epsilon_{i,j,k}) = \mathcal{N}(0,\sigma^2),$$
$$\text{et } \mathrm{Cov}(\epsilon_{i,j,k}, \epsilon_{l,m,n}) = 0 \text{ si } (i,j,k) \neq (l,m,n) \text{ avec}$$
$$1 \leqslant i,l \leqslant 2,\ 1 \leqslant j,m \leqslant 2 \text{ et } 1 \leqslant k,n \leqslant 8.$$

2. On commence par vérifier que les conditions d'utilisation du modèle introduit à la question 1. sont bien vérifiées. Pour cela, on calcule les résidus du modèle.

   1. Le protocole expérimental indique que l'expérience a bien porté sur 32 rats différents et que l'on n'est pas en présence de mesures qui seraient répétées sur un ou des mêmes individus. Ceci nous permet de supposer que les erreurs $\epsilon_{i,j,k}$ sont indépendantes.
   2. Le plan d'expérience utilisé comporte des répétitions ce qui permet de tester l'hypothèse $\mathcal{H}_0$ d'égalité des variances des erreurs $\epsilon_{i,j,k}$:

$$\boxed{\mathcal{H}_0 : \sigma^2_{1,1,1} = \sigma^2_{2,1,1} = \sigma^2_{1,2,1} = \sigma^2_{2,2,1} = \cdots = \sigma^2_{1,1,8} = \sigma^2_{2,1,8} = \sigma^2_{1,2,8} = \sigma^2_{2,2,8}}$$

contre

$$\boxed{\mathcal{H}_1 : \text{Il existe une variance différente des autres.}}$$

   Puisque l'on ne connaît pas la loi des erreurs, on utilise un test non-paramétrique, celui de Levene, sur les résidus du modèle :

```
Test de l'égalité des variances
Facteurs    Calorie  Vitamine

Test de Levene (pour toute loi de probabilité continue)

Statistique du test : 2,347
P :                   0,094
```

Puisque la $p-$valeur du test est strictement supérieure au seuil $\alpha = 5$ %, le test est non significatif et l'on ne peut rejeter l'hypothèses nulle $\mathcal{H}_0$ d'homoscédasticité.

3. On s'intéresse désormais à la normalité des erreurs. On cherche s'il l'on peut faire l'hypothèse que les résidus sont des réalisations d'une variable aléatoire ϵ dont la loi est une loi normale.

$$\boxed{\mathcal{H}_0 : \mathcal{L}(\epsilon) = \mathcal{N}}$$

contre

$$\boxed{\mathcal{H}_1 : \mathcal{L}(\epsilon) \neq \mathcal{N}.}$$

On utilise alors le test de Ryan-Joiner :

```
W-test pour la normalité
R:                              0,9906
Valeur de P (approximatif) : > 0,1000
```

Comme la $p-$valeur est strictement supérieure à $0,1$, elle est *a fortiori* strictement supérieure à $0,05$ qui est le seuil α du test. On en déduit que le test n'est pas significatif au niveau $\alpha = 5$ % et que l'on ne peut donc pas rejeter l'hypothèse nulle $\mathcal{H}_0$. Il n'y a pas de différences significatives entre la loi de ϵ et une loi normale, au seuil $\alpha = 5$ %.

4. Comme l'hypothèse de normalité des erreurs n'a pas été rejetée à la question 3., on s'intéresse à nouveau à l'hypothèse d'homoscédasticité portant sur les erreurs.
Le plan d'expérience utilisé comporte des répétitions ce qui permet de tester l'hypothèse $\mathcal{H}_0$ d'égalité des variances des erreurs $\epsilon_{i,j,k}$:

$$\boxed{\mathcal{H}_0 : \sigma^2_{1,1,1} = \sigma^2_{2,1,1} = \sigma^2_{1,2,1} = \sigma^2_{2,2,1} = \cdots = \sigma^2_{1,1,8} = \sigma^2_{2,1,8} = \sigma^2_{1,2,8} = \sigma^2_{2,2,8}}$$

contre

$$\boxed{\mathcal{H}_1 : \text{Il existe une variance différente des autres.}}$$

Il est maintenant possible d'utiliser le test paramétrique de Bartlett puisque ses conditions d'application, normalité des erreurs, sont vérifiées.

```
Test de l'égalité des variances
Facteurs    Calorie  Vitamine

Test de Bartlett (loi normale)

Statistique du test : 2,233
P :                   0,525
```

Puisque la $p-$valeur du test est strictement supérieure au seuil $\alpha = 5$ %, le test est non significatif et l'on ne peut rejeter l'hypothèse nulle $\mathcal{H}_0$ d'homoscédasticité.

Toutes les conditions d'application du modèle de l'analyse de la variance à deux facteurs contrôlés avec répétitions et terme d'interaction sont vérifiées. Déterminons si le facteur *Calorie*, le facteur *Vitamine* ou l'interaction des deux facteurs a un effet sur le gain de *Poids*. On teste donc les hypothèses :

$$\boxed{\mathcal{H}_0^{'} : \alpha_1 = \alpha_2 = 0}$$

contre

$$\boxed{\mathcal{H}_1^{'} : \text{Il existe } i_0 \in \{1,2\} \text{ tel que } \alpha_{i_0} \neq 0.}$$

$$\boxed{\mathcal{H}_0^{''} : \beta_1 = \beta_2 = 0}$$

contre

$$\boxed{\mathcal{H}_1^{''} : \text{Il existe } j_0 \in \{1,2\} \text{ tel que } \beta_{j_0} \neq 0.}$$

$$\boxed{\mathcal{H}_0^{'''} : (\alpha\beta)_{1,1} = (\alpha\beta)_{1,2} = (\alpha\beta)_{2,1} = (\alpha\beta)_{2,2} = 0}$$

contre

$$\boxed{\mathcal{H}_1^{'''} : \text{Il existe } (i_0, j_0) \in \{(1,1),(1,2),(2,1),(2,2)\} \text{ tel que } (\alpha\beta)_{i_0,j_0} \neq 0.}$$

On consulte alors le tableau de l'analyse de la variance

```
Facteur      Type Niveaux  Valeurs
Vitamine     fixe       2  1 2
Calorie      fixe       2  1 2

Analyse de la variance pour Poids, en utilisant la SC ajustée pour
les tests

Source       DL     SC séq   SC ajust   CM ajust       F      P
Vitamine      1      215,3      215,3      215,3    2,51  0,124
Calorie       1     4347,8     4347,8     4347,8   50,69  0,000
Vitamine
*Calorie      1      306,3      306,3      306,3    3,57  0,069
Erreur       28     2401,6     2401,6       85,8
Total        31     7271,0
```

La $p-$valeur associée, par la statistique de Fisher, à l'hypothèse nulle $\mathcal{H}_0^{'}$ est de $0,124$. Elle est strictement supérieure à $\alpha = 5$ % : le test n'est pas significatif à ce seuil. On ne peut pas rejeter l'hypothèse nulle $\mathcal{H}_0^{'}$ d'absence d'effet du facteur *Vitamine* sur le gain de *Poids*. Ainsi au niveau $\alpha = 5$ %, le facteur *Vitamine* n'a pas un effet significatif sur le gain de *Poids*. En prenant cette décision, on risque de commettre une erreur de seconde espèce, il serait très intéressant de calculer la puissance a posteriori.

La $p-$valeur associée, par la statistique de Fisher, à l'hypothèse nulle $\mathcal{H}_0^{''}$ est de

$0,000$. Elle est inférieure ou égale à $\alpha = 5$ % : le test est significatif à ce seuil. On doit rejeter l'hypothèse nulle $\mathcal{H}_0''$ d'absence d'effet du facteur *Calorie* sur le *Poids* et décider l'hypothèse alternative $\mathcal{H}_1''$ d'existence d'un effet du facteur *Calorie* sur le *Poids*. Ainsi au niveau $\alpha = 5$ %, le facteur *Calorie* a un effet significatif sur le *Poids*. En prenant cette décision, on risque de commettre une erreur de première espèce, le risque associé à la décision est donc de 5 %.
La $p-$valeur associée, par la statistique de Fisher, à l'hypothèse nulle $\mathcal{H}_0'''$ est de $0,069$. Elle est strictement supérieure à $\alpha = 5$ % : le test n'est pas significatif à ce seuil. On ne peut pas rejeter l'hypothèse nulle $\mathcal{H}_0'''$ d'absence d'effet de l'interaction des facteurs *Vitamine* et *Calorie* sur le gain de *Poids*. Ainsi au niveau $\alpha = 5$ %, l'interaction des facteurs *Vitamine* et *Calorie* n'a pas un effet significatif sur le gain de *Poids*. En prenant cette décision, on risque de commettre une erreur de seconde espèce, il serait très intéressant de calculer la puissance a posteriori.

Il est possible de calculer les puissances recherchées avec Minitab. On utilise le fait que le dispositif étudié est ce que l'on appelle aussi un plan factoriel complet à deux niveaux puisqu'il s'agit d'un modèle d'analyse de la variance à deux facteurs avec le terme d'interaction et surtout que le plan d'expérience est équilibré. À l'aide de l'entrée **Plan factoriel à 2 niveaux** du sous menu **Puissance et effectif de l'échantillon** du menu **Stat** de Minitab, on obtient le résultat suivant :

```
Puissance et effectif de l'échantillon

Plan factoriel à 2 niveaux

Sigma = 9,26121  Alpha = 0,05

Facteurs :        2   Plan de base :           2; 4
Blocs :       aucun

  Points
centraux
par bloc    Effet    Répétitions  Puissance
       0    5,187              8     0,3338
       0    6,188              8     0,4464
```

La valeur de l'*Effet* correspond à l'écart le plus important en valeur absolue entre les estimations des différents niveaux du facteur. Ainsi pour le facteur *Vitamine* l'*Effet* vaut $\widehat{\alpha_2}(\boldsymbol{y}) - \widehat{\alpha_1}(\boldsymbol{y}) = 2,594 - (-2,594) = 5,187$, il n'y avait en fait qu'une possibilité. Pour l'interaction entre les deux facteurs *Vitamine* et *Calorie* l'*Effet* vaut : $\widehat{(\alpha\beta)_{1,1}}(\boldsymbol{y}) - \widehat{(\alpha\beta)_{1,2}}(\boldsymbol{y}) = \widehat{(\alpha\beta)_{2,2}}(\boldsymbol{y}) - \widehat{(\alpha\beta)_{2,1}}(\boldsymbol{y}) = 3,094 - (-3,094) = 6,188$. Il y a deux facteurs, quatre sommets dans ce plan et aucun point central. La puissance a posteriori est donc relativement faible, 0,3338, dans le cas du facteur *Calorie* et à

peine meilleure, 0,4464, dans le cas de l'interaction entre les deux facteurs *Vitamine* et *Calorie*. On aurait pu améliorer la puissance du test en augmentant le nombre de répétitions. On détermine le nombre de répétitions nécessaires pour que l'on ait une puissance de 0,80, qui est généralement considéré comme un niveau satisfaisant pour la puissance d'un test.

```
Puissance et effectif de l'échantillon

Plan factoriel à 2 niveaux

Sigma = 9,26121  Alpha = 0,05

Facteurs :        2   Plan de base :          2; 4
Blocs :       aucun

  Points
centraux                            Puissance  Puissance
par bloc    Effet    Répétitions       cible     réelle
       0    5,187             26     0,8000     0,8074
       0    6,188             19     0,8000     0,8195
```

On constate qu'il aurait fallu 26 répétitions, donc 104 rats, pour obtenir une puissance d'au moins 0,80. On aurait multiplié par plus de trois l'ampleur de l'étude qui a été réalisée...

3. Les estimations des coefficients du modèle sont :

```
Terme                   Coef      Er-T coef        T      P
Constante             84,031          1,637    51,33  0,000
Vitamine
1                     -2,594          1,637    -1,58  0,124
Calorie
1                    -11,656          1,637    -7,12  0,000
Vitamine*Calorie
1        1             3,094          1,637     1,89  0,069
```

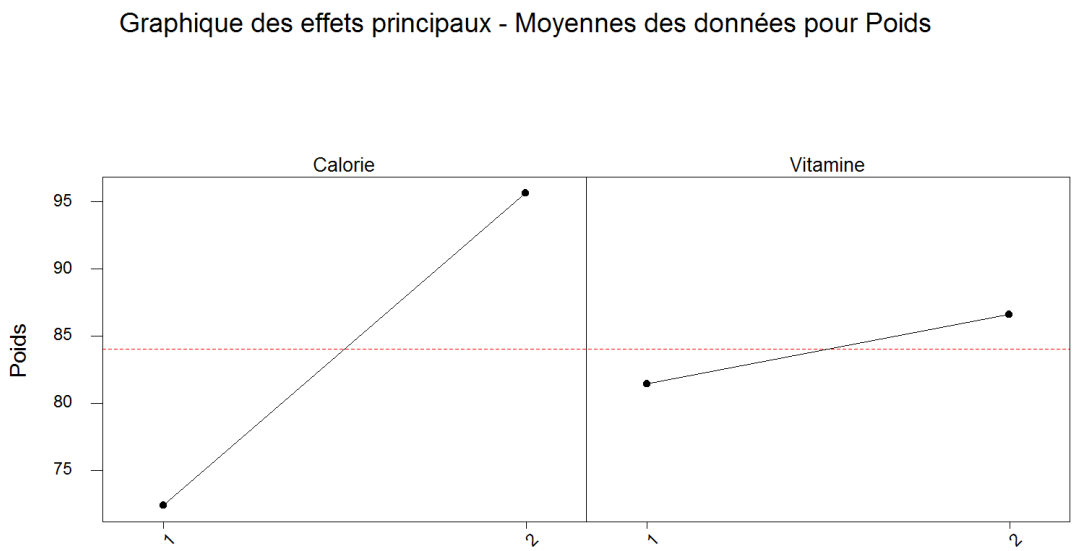
On en déduit que :

$$\widehat{\mu}(\boldsymbol{y}) = 84,031, \quad \widehat{\alpha_1}(\boldsymbol{y}) = -2,594, \quad \widehat{\alpha_2}(\boldsymbol{y}) = 2,594,$$
$$\widehat{\beta_1}(\boldsymbol{y}) = -11,656, \quad \widehat{\beta_2}(\boldsymbol{y}) = 11,656,$$
$$\widehat{(\alpha\beta)_{1,1}}(\boldsymbol{y}) = 3,094, \quad \widehat{(\alpha\beta)_{1,2}}(\boldsymbol{y}) = -3,094,$$
$$\widehat{(\alpha\beta)_{2,1}}(\boldsymbol{y}) = -3,094, \quad \widehat{(\alpha\beta)_{2,2}}(\boldsymbol{y}) = 3,094,$$

en utilisant les relations $\sum_{i=1}^{2} \widehat{\alpha_i} = 0$, $\sum_{j=1}^{2} \widehat{\beta_j} = 0$, $\widehat{(\alpha\beta)_{1,1}} + \widehat{(\alpha\beta)_{1,2}} = 0$, $\widehat{(\alpha\beta)_{1,1}} + \widehat{(\alpha\beta)_{2,1}} = 0$ et $\widehat{(\alpha\beta)_{2,1}} + \widehat{(\alpha\beta)_{2,2}} = 0$ qui sont en particulier valables pour les données expérimentales $\boldsymbol{y}$: $\sum_{i=1}^{2} \widehat{\alpha_i}(\boldsymbol{y}) = 0$, $\sum_{j=1}^{2} \widehat{\beta_j}(\boldsymbol{y}) = 0$, $\widehat{(\alpha\beta)_{1,1}}(\boldsymbol{y}) + \widehat{(\alpha\beta)_{1,2}}(\boldsymbol{y}) = 0$, $\widehat{(\alpha\beta)_{1,1}}(\boldsymbol{y}) + \widehat{(\alpha\beta)_{2,1}}(\boldsymbol{y}) = 0$ et $\widehat{(\alpha\beta)_{2,1}}(\boldsymbol{y}) + \widehat{(\alpha\beta)_{2,2}}(\boldsymbol{y}) = 0$

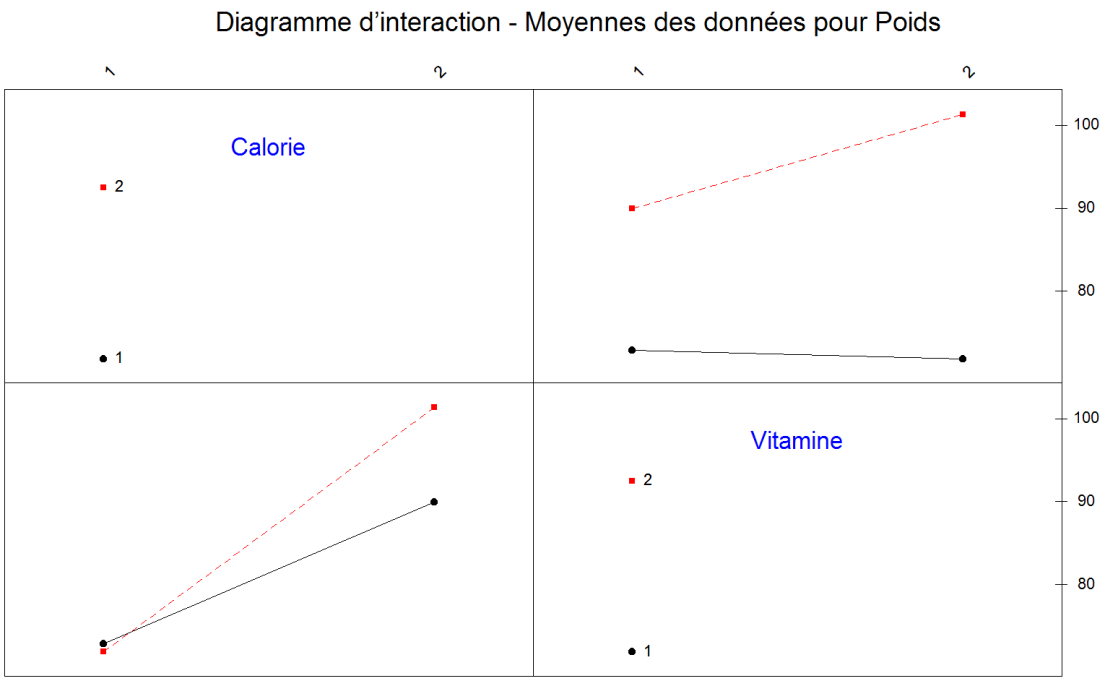
Remarquons que l'on n'a pas eu à utiliser la relation $\widehat{(\alpha\beta)_{1,2}} + \widehat{(\alpha\beta)_{2,2}} = 0$ ce qui était prévisible car, d'après le cours, l'on sait que l'on impose seulement $I + J - 1 = 2 + 2 - 1 = 3$ relations **indépendantes** aux coefficients $(\alpha\beta)_{i,j}$. On doit donc retrouver toutes les estimations $\widehat{(\alpha\beta)_{i,j}}(\boldsymbol{y})$ à l'aide des valeurs fournies par le logiciel et de seulement $I + J - 1 = 3$ des relations portant sur ces estimations.

À l'aide de l'entrée **Graphique des effets principaux** du sous menu **ANOVA** du menu **Stat** de Minitab, on obtient la représentation graphique des estimations des $\widehat{\mu}(\boldsymbol{y}) + \widehat{\alpha_1}(\boldsymbol{y})$, $\widehat{\mu}(\boldsymbol{y}) + \widehat{\alpha_2}(\boldsymbol{y})$, $\widehat{\mu}(\boldsymbol{y}) + \widehat{\beta_1}(\boldsymbol{y})$ et $\widehat{\mu}(\boldsymbol{y}) + \widehat{\beta_2}(\boldsymbol{y})$. On comprend ainsi visuellement pourquoi le facteur facteur *Calorie* a un effet significatif sur le gain de *Poids* au seuil $\alpha = 5$ %. L'influence du facteur *Vitamine* n'est, quant à elle, pas significative au seuil $\alpha = 5$ % ce qui est bien confirmé par le graphique ci-dessous.



À l'aide de l'entrée **Diagramme des interactions** du sous menu **ANOVA** du menu **Stat** de Minitab, on obtient la représentation graphique de l'effet des interactions. Sous l'hypothèse nulle $\mathcal{H}_0'''$ d'absence d'interaction entre les deux facteurs

Vitamine et *Calorie*, dans chacune des deux cases, inférieure gauche et supérieure droite, du graphique, les deux segments représentés en noir et rouge[24] sont parallèles. La force d'une interaction est associée au défaut de parallélisme de ces deux segments. On constate visuellement que dans les deux cases du graphique les deux segments ne sont pas parallèles mais ces interactions entre les deux facteurs ne sont pas assez importantes pour qu'elles soient significatives au seuil $\alpha = 5$ %. Par contre ces interactions auraient été significatives au seuil $\alpha = 10$ % puisque l'on a une $p-$valeur de $0,069$ associée au test de l'hypothèse nulle $\mathcal{H}_0'''$.



4. Puisque l'on a rejeté l'hypothèse nulle $\mathcal{H}_0''$ d'absence d'effet du facteur *Calorie* sur le gain de *Poids*. On peut réaliser des comparaisons multiples pour le facteur *Calorie*. Or ce facteur n'a que deux modalités ! Il n'est donc pas nécessaire de faire des investigations supplémentaires pour en déduire que ce sont les deux seuls niveaux du facteur *Calorie*, 1 et 2, qui sont significativement différents au seuil $\alpha = 5\%$.

. . . . . .

[24] En gris si vous lisez une version imprimée du corrigé de cet exercice.

12.2.4. Avec répétitions et plan équilibré III : Les préférences de biscuits sont-elles les mêmes d'un pays à l'autre ?[25]

Une analyse sensorielle a été organisée simultanément en France et au Pakistan. On a demandé à 150 français et à 163 pakistanais de donner une note d'appréciation à 8 biscuits (0 : je n'aime pas, 10 : j'aime beaucoup). Parmi ces biscuits, 4 sont fabriqués et vendus en France (biscuits notés F1, F2, F3 et F4) et 4 sont fabriqués et vendus au Pakistan (biscuits notés P1, P2, P3, et P4). L'objectif d'une telle analyse est de comparer les appréciations d'un pays à l'autre.
Un résumé des données est fourni dans le tableau ci-dessous.

	Origine française				Origine pakistanaise			
	F1	F2	F3	F4	P1	P2	P3	P4
Jury français	7,35	6,06	5,08	5,75	4,99	6,03	4,22	5,18
Jury pakistanais	6.29	5.10	5.23	5.38	7.49	6.74	5.56	6.32

1. Proposer un modèle d'analyse de la variance sur les données du tableau précédent permettant de répondre à la question : dans l'ensemble, les biscuits d'origine française et pakistanaise sont-ils évalués de la même manière par les deux jurys (le premier constitué de juges exclusivement de nationalité française et le second constitué exclusivement de juges de nationalité pakistanaise) ? Interpréter la variabilité résiduelle de ce modèle.
2. Le listage suivant donne les résultats du modèle d'analyse de la variance sur les seules données du tableau précédent. Que peut-on dire ?

	ddl	SCE	CM	f_{obs}	$Proba$
Nationalité jury	1	0,7439	0,7439	1,2398	0,28732
Origine biscuit	1	0,0053	0,0053	0,0088	0,92698
Nationalité jury : Origine biscuit	1	3,9303	3,9303	0,02503	0,02503
Résiduelle	12	7,2004	0,6000		

On analyse maintenant l'ensemble des données individuelles. Le tableau suivant donne les probabilités critiques de chacun des effets et de chacune des interactions calculées sur ces données brutes.

3. Pourquoi peut-il être plus intéressant de travailler sur les données brutes ? Que peut-on dire à partir du tableau d'analyse de la variance ci-dessous ? Interpréter.

[25] Les données de cet exercice sont tirées du livre d'exercices de François Husson et de Jérôme Pagès [2].

	ddl	*SCE*	*CM*	f_{obs}	*Proba*
Nationalité jury	1	1,5	1,5	0,245	0,621
Origine biscuit	1	0,06	0,06	0,0098	0,9212
Nationalité jury : Origine biscuit	1	410,17	410,17	66,91	$5,126 \times 10^{-16}$
Résiduelle	1907	11689,94	6,130		

4. Construire un graphique avec en abscisse les biscuits et en ordonnée les notes moyennes par jury. Relier les points d'un même jury. Que représente ce graphique ?
5. Quelle critique pouvez-vous formuler concernant ce modèle ?

. . . . . .

Éléments de corrigé : Les préférences de biscuits sont-elles les mêmes d'un pays à l'autre ?

1. Dire que les biscuits ne sont pas évalués de la même manière par les deux jurys revient à mettre en évidence une interaction entre les facteurs *biscuit* et *jury*. Si l'on considère l'ensemble des biscuits, le tableau ne fait pas apparaître de répétitions et il est impossible de mettre en évidence une interaction. En revanche, si l'on regroupe les biscuits par origine, alors on peut construire un modèle d'analyse de variance à deux facteurs avec interaction. La variable Y sera l'appréciation et les facteurs seront la nationalité du jury (française ou pakistanaise) et l'origine des biscuits (française ou pakistanaise) ainsi que l'interaction de ces deux facteurs.
On introduit le modèle :

$$Y_{i,j,k} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{i,j} + \epsilon_{i,j,k}, \quad i = 1,2,\ j = 1,2,\ k = 1\ldots 4$$
$$\text{avec les contraintes supplémentaires } \sum_{i=1}^{2} \alpha_i = 0,\ \sum_{j=1}^{2} \beta_j = 0,$$
$$\sum_{i=1}^{2} (\alpha\beta)_{i,j} = 0,\ \forall j \in \{1,2\} \text{ et } \sum_{j=1}^{2} (\alpha\beta)_{i,j} = 0,\ \forall i \in \{1,2\},$$

où $Y_{i,j,k}$ est la moyenne des appréciations du $k-$ème biscuit d'origine i évalué par l'ensemble des juges de nationalité j. On postule les hypothèses classiques suivantes pour les erreurs :

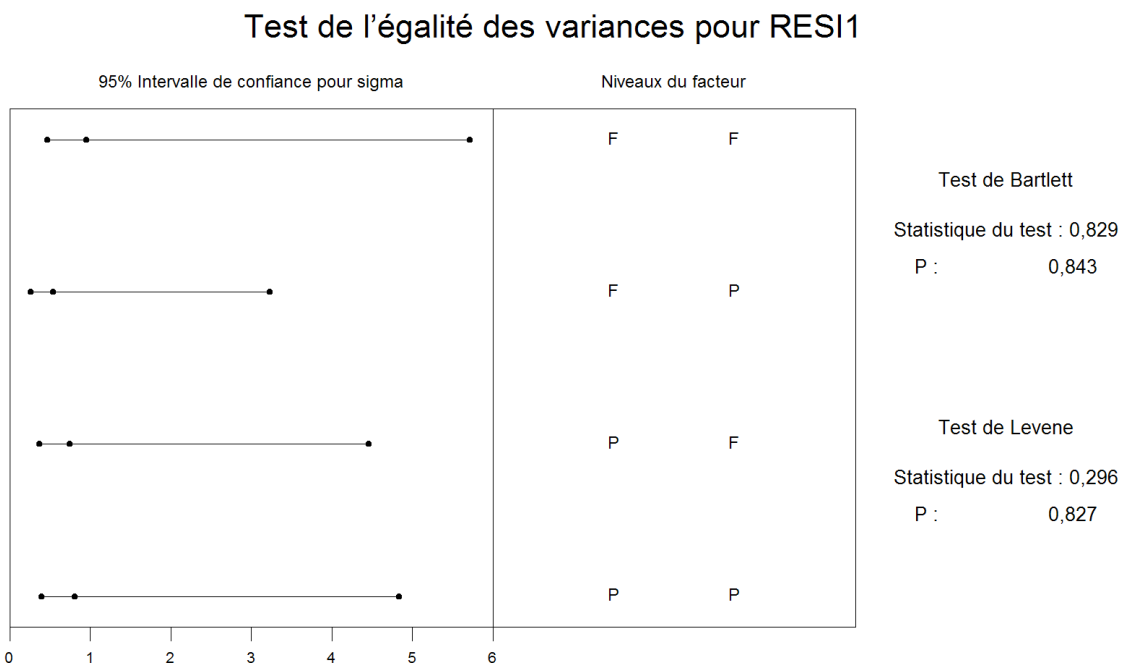
$$\forall\ (i,j,k), 1 \leqslant i \leqslant 2,\ 1 \leqslant j \leqslant 2,\ 1 \leqslant k \leqslant 4,\ \mathcal{L}(\epsilon_{i,j,k}) = \mathcal{N}(0,\sigma^2),$$
$$\text{et } \mathrm{Cov}(\epsilon_{i,j,k}, \epsilon_{l,m,n}) = 0 \text{ si } (i,j,k) \neq (l,m,n) \text{ avec}$$
$$1 \leqslant i,l \leqslant 2,\ 1 \leqslant j,m \leqslant 2 \text{ et } 1 \leqslant k,n \leqslant 4.$$

La variabilité résiduelle est alors celle des biscuits d'une même origine évalués par un même jury (ceci est dû à notre regroupement).

2. On s'intéresse à la validité des hypothèses du modèle.

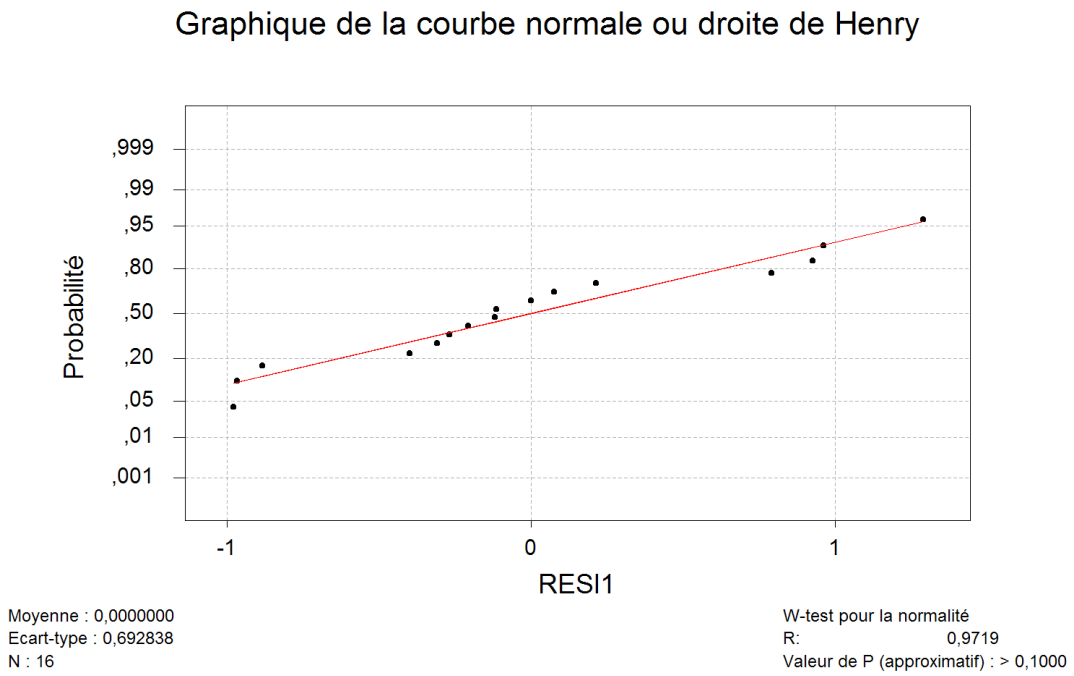
   Le test de l'homoscédasticité des résidus à l'aide d'un test de Levene n'est pas significatif :

```
Test de l'égalité des variances

Test de Levene (pour toute loi de probabilité continue)

Statistique du test : 0,296
P :                   0,827
```



   Le test de normalité à l'aide d'un test de Shapiro-Wilk (car $N = 45$) n'est pas significatif :

```
Valeur de P (approximatif) : > 0,1000
R:                             0,9719
W-test pour la normalité
```



On va donc faire également un test de Bartlett (plus puissant que Levene à cause de la normalité) :

```
Test de l'égalité des variances

Test de Bartlett (loi normale)

Statistique du test : 0,296
P :                   0,843
```

Ce test n'est pas significatif ce qui confirme bien l'homoscédasticité des résidus.

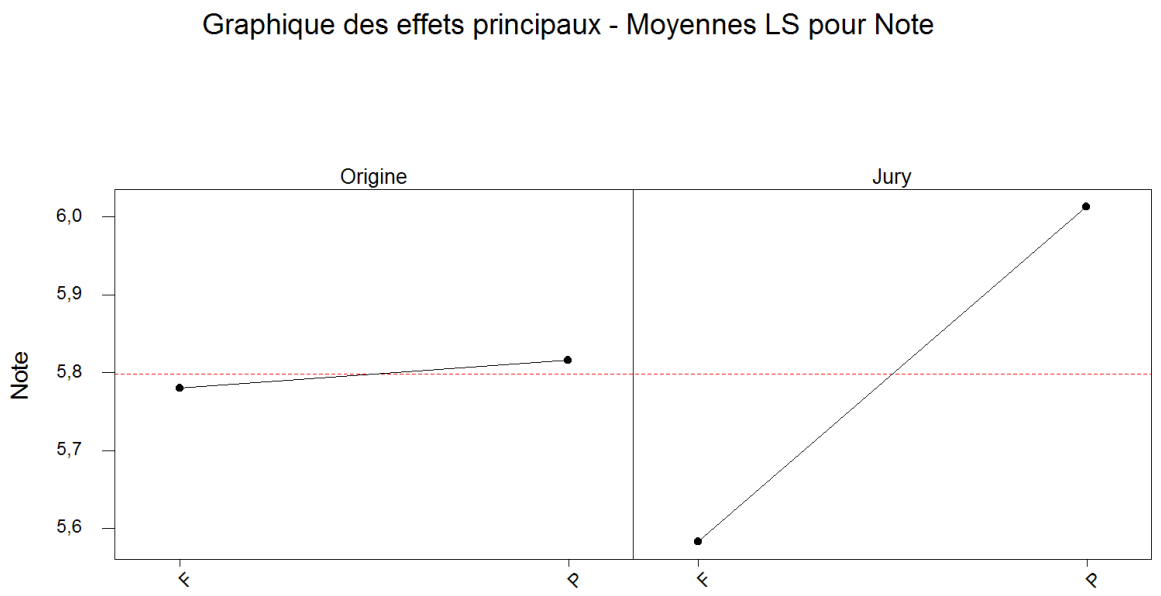
On obtient les résultats suivants, qui sont bien identiques à ceux qui figurent dans le tableau de l'énoncé.

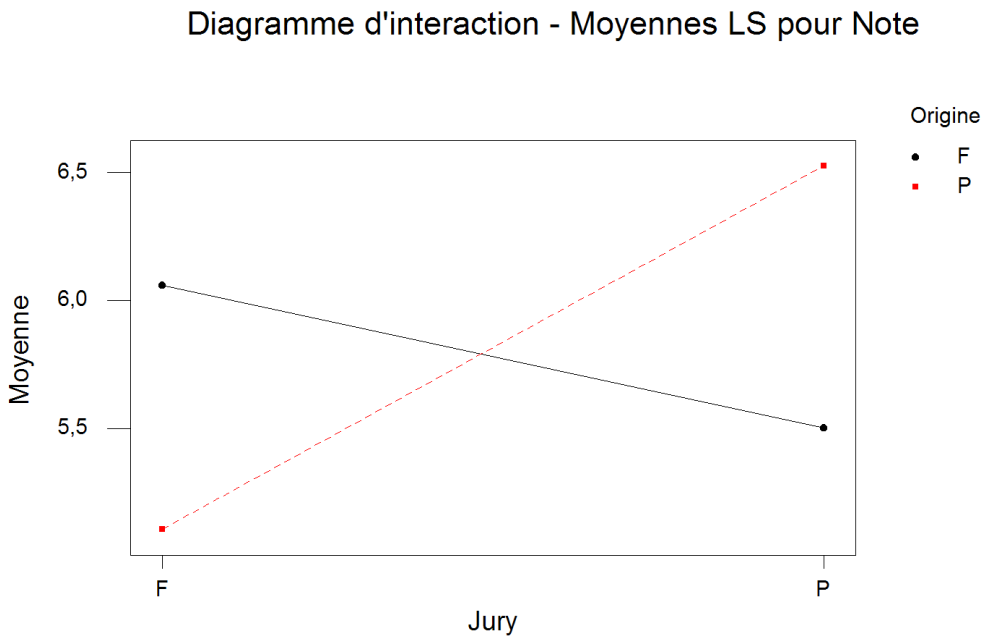
```
Modèle linéaire généralisé : Note en fonction de Origine; Jury


Facteur      Type Niveaux  Valeurs
Origine      fixe       2  F P
Jury         fixe       2  F P

Analyse de la variance pour Note, en utilisant la SC ajustée pour
les tests

Source          DL     SC séq   SC ajust   CM ajust       F      P
Origine          1     0,0053     0,0053     0,0053    0,01  0,927
Jury             1     0,7439     0,7439     0,7439    1,24  0,287
```

Origine*Jury	1	3,9303	3,9303	3,9303	6,55	0,025
Erreur	12	7,2004	7,2004	0,6000		
Total	15	11,8798				

Terme		Coef	Er-T coef	T	P
Constante		5,7981	0,1937	29,94	0,000
Origine					
F		-0,0181	0,1937	-0,09	0,927
Jury					
F		-0,2156	0,1937	-1,11	0,287
Origine*Jury					
F	F	0,4956	0,1937	2,56	0,025

Au seuil $\alpha = 5$ %, il y a un effet de l'interaction *nationalité du jury* $\times$ *origine des biscuits* mais il n'y a pas d'effet de la nationalité du jury ni d'effet de l'origine des biscuits. Cela signifie qu'aucun des jurys ne met en moyenne des notes plus élevées (pas d'effet jury), et que les biscuits français sont en moyenne autant appréciés que les biscuits pakistanais (pas d'effet origine) mais il existe une interaction. Donc les français n'évaluent pas les biscuits comme les pakistanais.





3. On introduit le modèle :

$$Y_{i,j,k,l} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{i,j} + \epsilon_{i,j,k,l},$$
$$i = 1,2,\ j = 1,2,\ k = 1\ldots 4,\ l = 1\ldots L(i,j,k),$$
$$\text{avec les contraintes supplémentaires } \sum_{i=1}^{2}\alpha_i = 0,\ \sum_{j=1}^{2}\beta_j = 0 \text{ et } \sum_{k=1}^{4}\gamma_k = 0,$$
$$\sum_{i=1}^{2}(\alpha\beta)_{i,j} = 0,\ \forall j \in \{1,2\} \text{ et } \sum_{j=1}^{2}(\alpha\beta)_{i,j} = 0,\ \forall i \in \{1,2\},$$

où $Y_{i,j,k,l}$ l'appréciation du $l-$ème juge de nationalité j évaluant le $k-$ème biscuit d'origine i. On postule les hypothèses classiques suivantes pour les erreurs :

$$\forall\ (i,j,k,l), 1 \leqslant i \leqslant 2,\ 1 \leqslant j \leqslant 2,\ 1 \leqslant k \leqslant 4,\ 1 \leqslant l \leqslant L(i,j,k),$$
$$\mathcal{L}(\epsilon_{i,j,k,l}) = \mathcal{N}(0,\sigma^2),$$
$$\mathrm{Cov}(\epsilon_{i,j,k,l}, \epsilon_{m,n,o,p}) = 0 \text{ si } (i,j,k,l) \neq (m,n,o,p)$$
$$\text{avec } 1 \leqslant i,m \leqslant 2,\ 1 \leqslant j,n \leqslant 2,\ 1 \leqslant k,o \leqslant 4,\ 1 \leqslant l \leqslant L(i,j,k)$$
$$\text{et } 1 \leqslant p \leqslant L(m,n,o).$$

Si l'on travaille sur les données brutes, on peut donc construire le modèle ci-dessus et la variablité résiduelle change donc de signification par rapport à la question 1. En effet deux indices n'interviennent qu'à travers la résiduelle : à la variabilité des biscuits d'une même origine s'ajoute la variabilité des juges d'un même jury. Mais parallèlement il y a beaucoup plus de données et donc les tests sont plus puissants.

Cependant malgré des tests plus puissants on ne met toujours pas en évidence l'effet de l'origine du biscuit ni l'effet de la nationalité du jury. L'interaction entre ces deux facteurs est par contre beaucoup mieux mise en évidence puisque la $p-$valeur associée est maintenant de $5,126 \times 10^{-16}$ (contre 0,025 précedemment).

4. Le graphique demandé représente l'interaction *origine du biscuit* $\times$ *jury*. S'il n'y avait pas d'interaction, les lignes brisées seraient parallèles. On peut voir ici que les pakistanais apprécient surtout les biscuits pakistanais (biscuits P1, P2, P3 et P4) tandis que les français préfèrent les biscuits français (biscuits F1 à F4). Peut-être y a-t-il une sorte de rejet devant un produit nouveau ou une sur-notation pour des produits nationaux qui sont reconnus et appréciés ?
5. On ne tient pas compte du facteur type de biscuit qui est emboîté dans le facteur *Origine* du biscuit.

. . . . . .

12.2.5. Avec répétitions et plan déséquilibré : Analyse sensorielle de trois chocolats[26]

Lors d'un test de dégustation hédonique, on s'intéresse à l'appréciation globale de trois chocolats. Pour cela, 45 juges ont participé à cette évaluation qui a eu lieu sur 2 jours (on dispose de 15 échantillons par chocolat). Les notes d'appréciation des juges, comprises en 0 et 7, sont données dans le tableau suivant. Chaque juge n'a évalué qu'un chocolat. Comme chacun choisit son jour de dégustation et le chocolat qu'il évalue, le nombre de données et la répartition des chocolats évalués ne sont pas les mêmes d'un jour à l'autre. On souhaite d'une part vérifier qu'il y a bien un effet *chocolat*, s'il y a un effet *jour* (les chocolats pouvant être plus ou moins appréciés lors du premier ou du deuxième jour) et un effet de l'interaction entre *chocolat* et *jour*.

	Chocolat 1			Chocolat 2			Chocolat 3		
Jour 1	5.2	6	5.4	4.2	5.2	5	4.6	5.6	5.2
	5.2	6.6		4.4	5.4	4.8	4.8	5.8	6.4
				4.4	5.6	6	5	6.2	5.2
							5.4	6.4	
Jour 2	3.2	4	3.6	3.2	3.6	4.2	3	3.8	3.4
	4.2	3.8	3.4	4	3.6	4	4.4		
	4	4.4	4.6						
		4							

[26] Les données de cet exercice sont tirées du livre d'exercices de François Husson et de Jérôme Pagès [2].

1. Écrire le modèle permettant de répondre à la problématique. Que signifie l'interaction entre les facteurs *chocolat* et *jour* ?
2. Comment (par quelle formule) sont estimés les paramètres de ce modèle ?
3. Il y a-t-il un effet *jour*, un effet *chocolat* et un effet de l'interaction ?
4. Par chocolat, calculer la moyenne des notes et la comparer avec la moyenne ajustée $(\widehat{\mu}(\boldsymbol{y}) + \widehat{\alpha}_i(\boldsymbol{y}))$. Qu'en pensez-vous ? Quel est le chocolat préféré ?
5. Tester l'hypothèse « $\mathcal{H}_0$: l'effet du chocolat 1 est nul » (préciser les hypothèses du test, la statistique de test, la loi de la statistique de test sous $\mathcal{H}_0$, et la décision)

. . . . . .

Éléments de corrigé : Analyse sensorielle de trois chocolats

1. On introduit le modèle :

$$Y_{i,j,k} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{i,j} + \epsilon_{i,j,k}, \quad i = 1\ldots 3,\ j = 1,2,\ k = 1\ldots K(i,j)$$
$$\text{avec les contraintes supplémentaires } \sum_{i=1}^{3} \alpha_i = 0,\ \sum_{j=1}^{2} \beta_j = 0,$$
$$\sum_{i=1}^{3} (\alpha\beta)_{i,j} = 0,\ \forall j \in \{1,2\} \text{ et } \sum_{j=1}^{2} (\alpha\beta)_{i,j} = 0,\ \forall i \in \{1,\ldots,3\},$$

   où $Y_{i,j,k}$ est la $k-$ème note obtenue le jour j par le chocolat i. On postule les hypothèses classiques suivantes pour les erreurs :

$$\forall\ (i,j,k), 1 \leqslant i \leqslant 3,\ 1 \leqslant j \leqslant 2,\ 1 \leqslant k \leqslant K(i,j),\ \mathcal{L}(\epsilon_{i,j,k}) = \mathcal{N}(0,\sigma^2),$$
$$\text{et } \mathrm{Cov}(\epsilon_{i,j,k}, \epsilon_{l,m,n}) = 0 \text{ si } (i,j,k) \neq (l,m,n) \text{ avec}$$
$$1 \leqslant i,l \leqslant 3,\ 1 \leqslant j,m \leqslant 2 \text{ et } 1 \leqslant k,n \leqslant K(i,j).$$

   Une interaction *Jour* $\times$ *Chocolat* signifie qu'un même chocolat n'est pas apprécié de la même façon les 2 jours.
2. Le vecteur des paramètres peut être estimé par moindres carrés.
3. On procède à l'ANOVA (en stockant les résidus afin de pouvoir vérifier s'ils vérifient les hypothèses) à l'aide de la fonction modèle linéaire généralisé et on obtient les résultats suivant :

```
Modèle linéaire généralisé : note en fonction de jour; chocolat

Facteur      Type Niveaux  Valeurs
jour         fixe       2  1 2
```

```
chocolat     fixe       3  1 2 3

Analyse de la variance pour note

Source          DL     SC séq   SC ajust   CM ajust       F      P
jour             1    26,3511    25,6301    25,6301   84,45  0,000
chocolat         2     1,3894     1,2021     0,6010    1,98  0,152
jour*chocolat    2     0,7662     0,7662     0,3831    1,26  0,294
Erreur          39    11,8364    11,8364     0,3035
Total           44    40,3431
```

Le test de l'homoscédasticité des résidus à l'aide d'un test de Levene n'est pas significatif :

```
Test de Levene (pour toute loi de probabilité continue)

Statistique du test : 0,613
P :                   0,691
```

Le test de normalité à l'aide d'un test de Shapiro-Wilk (car N=45) n'est pas significatif :

```
Valeur de P (approximatif) : > 0,1000
R:                             0,9885
W-test pour la normalité
```

On va donc faire également un test de Bartlett (plus puissant que Levene à cause de la normalité) :

```
Test de Bartlett (loi normale)

Statistique du test : 2,643
P :                   0,755
```

Ce test n'est pas significatif ce qui confirme bien l'homoscédasticité des résidus.

L'étude du protocole expérimental nous assure de l'indépendance des variables $\epsilon_{i,j,k}$.

Puisque les conditions d'application de l'ANOVA sont bien respectées, on peut déduire des résultats reportés dans le tableau ci-dessus qu'en considérant le modèle introduit à la première question il n'y a, au seuil $\alpha = 5$ %, ni effet *chocolat* $(0,152 > 0,05)$ ni effet de l'interaction $(0,294 > 0,05)$.

4. La moyenne des notes est de $4,676$. Les moyennes par chocolat sont :

```
Variable   chocolat          N    Moyenne    EcarType
note       1                15      4,507       0,982
           2                15      4,507       0,807
           3                15      5,013       1,041
```

Pour obtenir les moyennes ajustées $\widehat{\mu}(\boldsymbol{y}) + \widehat{\alpha_1}(\boldsymbol{y}), \widehat{\mu}(\boldsymbol{y}) + \widehat{\alpha_2}(\boldsymbol{y}), \widehat{\mu}(\boldsymbol{y}) + \widehat{\alpha_3}(\boldsymbol{y})$ on fait appel à la technique apprise dans le cours, basée sur la régression linéaire multiple, afin d'estimer les coefficients $\widehat{\mu}(\boldsymbol{y}), \widehat{\alpha_1}(\boldsymbol{y}), \widehat{\alpha_2}(\boldsymbol{y}), \widehat{\alpha_3}(\boldsymbol{y})$. Cette estimation des paramètres est encore obtenue à l'aide du modèle linéaire généralisé (pensez à afficher les résultats les plus détaillés). Les calculs renvoient les résultats ci-dessous :

Terme	Coef	Er-T coef	T	P
Constante	4,58763	0,08801	52,13	0,000
jour				
1	0,80874	0,08801	9,19	0,000
chocolat				
1	0,2124	0,1238	1,72	0,094
2	-0,2043	0,1215	-1,68	0,101
jour*chocolat				
1 1	0,0713	0,1238	0,58	0,568
1 2	-0,1921	0,1215	-1,58	0,122

On en déduit que :

$$\widehat{\mu}(\boldsymbol{y}) = 4,588, \quad \widehat{\alpha_1}(\boldsymbol{y}) = 0,212, \quad \widehat{\alpha_2}(\boldsymbol{y}) = -0,204, \quad \widehat{\alpha_3}(\boldsymbol{y}) = -0,008$$

et que les moyennes ajustées valent :

$$\widehat{\mu}(\boldsymbol{y}) + \widehat{\alpha_1}(\boldsymbol{y}) = 4,800, \quad \widehat{\mu}(\boldsymbol{y}) + \widehat{\alpha_2}(\boldsymbol{y}) = 4,384, \quad \widehat{\mu}(\boldsymbol{y}) + \widehat{\alpha_3}(\boldsymbol{y}) = 4,580.$$

Par conséquent c'est le chocolat 1 qui est le préféré. Le déséquilibre des données fait que la moyenne d'un chocolat n'est plus égale à sa moyenne ajustée. Ainsi contrairement à ce que la moyenne des notes semblaient indiquer ce n'est pas le chocolat 3 qui est le préféré.

5. $\mathcal{H}_0 \ : \ \alpha_1 = 0$ contre $\mathcal{H}_1 \ : \ \alpha_1 \neq 0$. La statistique du test est $\dfrac{\widehat{\alpha_1}}{\widehat{\sigma}_{\widehat{\alpha_1}}}$ et suit, sous l'hypothèse nulle $\mathcal{H}_0$, une loi de Student avec autant de degrés de liberté que ceux de la variance résiduelle (i.e. 39). La probabilité critique associée à cette valeur est de $0,094$, ce chocolat n'est donc pas significativement plus apprécié que les autres.

. . . . . .

12.3. Analyse de la variance à deux facteurs emboîtés

12.3.1. Plan déséquilibré : Les cyclamens[27]

Les résultats suivants sont relatifs à la vitesse moyenne de croissance, en *mm* par jour, des pédoncules floraux observés sur huit plantes de cyclamen, réparties au hasard entre deux milieux de culture.

[27] Les données de cet exercice sont tirées du livre de Pierre Dagnélie [4].

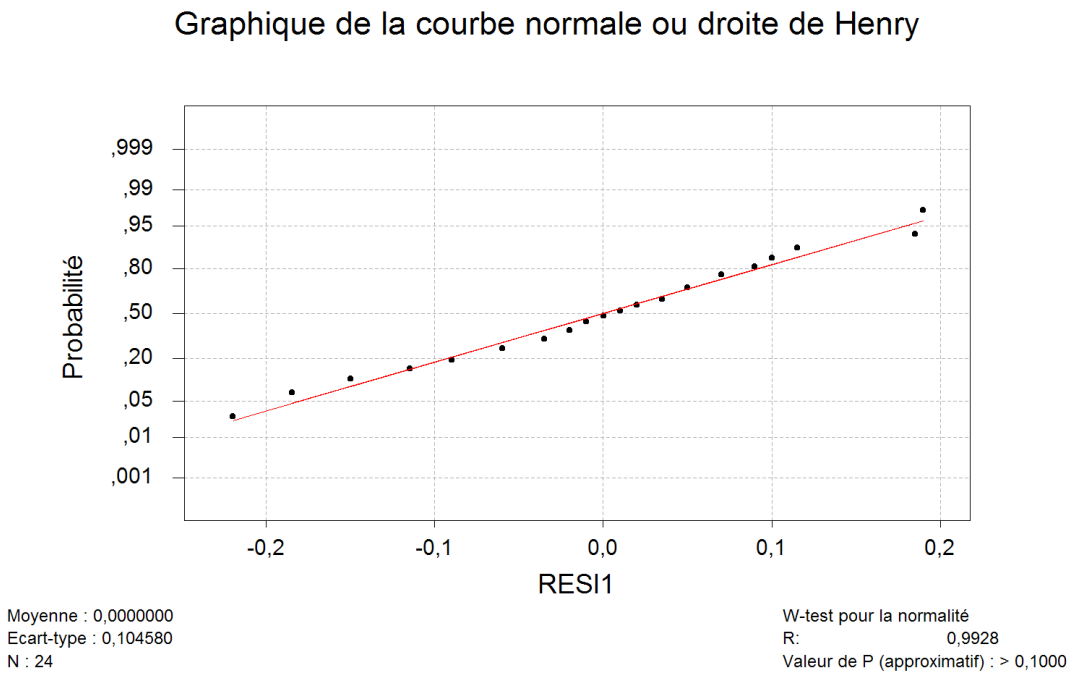
Milieu 1				Milieu 2			
Plante 1	Plante 2	Plante 3	Plante 4	Plante 5	Plante 6	Plante 7	Plante 8
0,76	0,97	0,58	0,86	0,81	1,07	0,62	0,93
0,66	0,74	0,57	0,93	0,77	0,82	0,99	1,07
0,61		0,51			1,02		0,95
0,65		0,62					0,79
							0,66

1. De quel modèle d'analyse de la variance peut-on se servir pour étudier ces données ?
2. Peut-on conclure sur cette base à l'existence d'une différence significative de vitesse de croissance entre les deux milieux ?

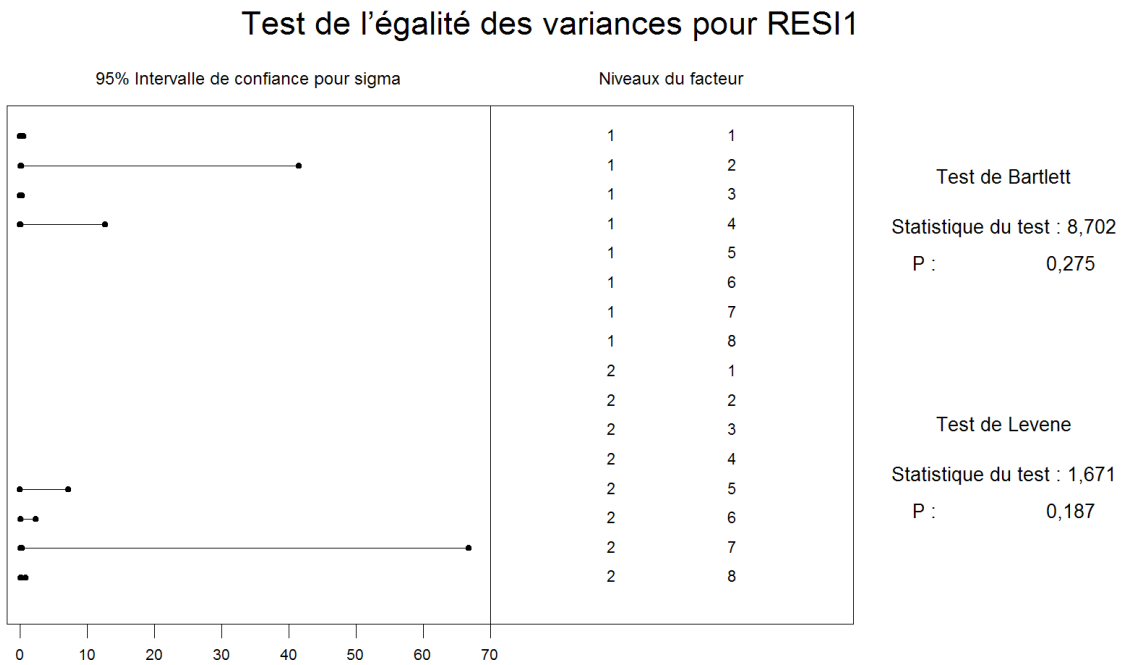
. . . . . .

Éléments de corrigé : Cyclamens

```
Test de l'égalité des variances

Réponse     RESI1
Facteurs    Milieu  Plante
NivConf     95,0000

Test de Levene (pour toute loi de probabilité continue)

Statistique du test : 1,671
P :                   0,187


W-test pour la normalité
R:                            0,9928)
Valeur de P (approximatif) : > 0,1000
```




```
Test de l'égalité des variances

Réponse      RESI1
Facteurs     Milieu  Plante
NivConf      95,0000

Test de Bartlett (loi normale)

Statistique du test : 8,702
P :                   0,275
```



Modèle linéaire généralisé : Croissance en fonction de Milieu; Plante

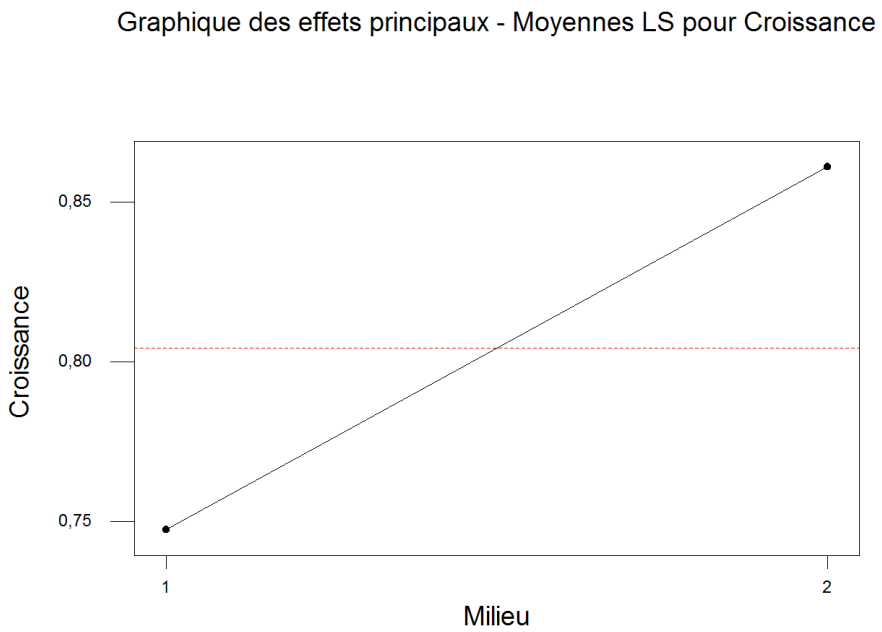
Facteur	Type	Niveaux	Valeurs
Milieu	fixe	2	1 2
Plante(Milieu)	fixe	8	1 2 3 4 1 2 3 4

Analyse de la variance pour Croissan, en utilisant la SC ajustée pour les tests

Source	DL	SC séq	SC ajust	CM ajust	F	P
Milieu	1	0,17340	0,06825	0,06825	4,34	0,054
Plante(Milieu)	6	0,24645	0,24645	0,04107	2,61	0,058
Erreur	16	0,25155	0,25155	0,01572		
Total	23	0,67140				

Terme		Coef	Er-T coef	T	P
Constante		0,80438	0,02730	29,47	0,000
Milieu					
1		-0,05687	0,02730	-2,08	0,054
(Milieu)Plante					
1	1	-0,07750	0,05864	-1,32	0,205
1	2	0,10750	0,07351	1,46	0,163
1	3	-0,17750	0,05864	-3,03	0,008

2	1	-0,07125	0,07374	-0,97	0,348
2	2	0,10875	0,06424	1,69	0,110
2	3	-0,05625	0,07374	-0,76	0,457



. . . . . .

12.4. Analyse de la variance à trois facteurs

12.4.1. Sans répétition : Détermination d'une fumure optimale pour le blé[28]

Une expérience en blocs aléatoires complets a été réalisée sur du blé au Rwanda. Trois doses d'acide phosphorique (100, 200 et 300 kg/ha) et trois doses de chaux (1000, 4500 et 8000 kg/ha), les neufs combinaisons des engrais ayant été affectées chacune au hasard et indépendamment à une parcelle au sein de chacun des trois blocs.

[28]Les données de cet exercice sont inspirées du livre de Pierre Dagnélie [4].

Fumures		Bloc		
P_2O_5	CaO	1	2	3
1	1	1,30	0,80	2,01
1	2	2,05	2,37	2,52
1	3	2,07	2,60	2,25
2	1	2,25	2,36	2,71
2	2	2,99	2,92	3,63
2	3	2,62	2,89	3,43
3	1	2,61	2,36	3,29
3	2	3,22	2,93	3,85
3	3	3,15	3,35	3,67

1. Quel modèle d'analyse de la variance peut-on utiliser avec ces données ? Le plan est-il équilibré ? Il y a-t-il des facteurs à effets aléatoires ?
2. On décide dans cette question de négliger l'influence du facteur *bloc*. Procéder à l'étude de ces données en utilisant le modèle d'analyse de la variance à deux facteurs le plus complet que vous pouvez utiliser.
2. On décide dans cette question de ne plus négliger l'influence du facteur *bloc*. Procéder à l'étude de ces données en utilisant le modèle d'analyse de la variance à trois facteurs le plus complet que vous pouvez utiliser.

. . . . . .

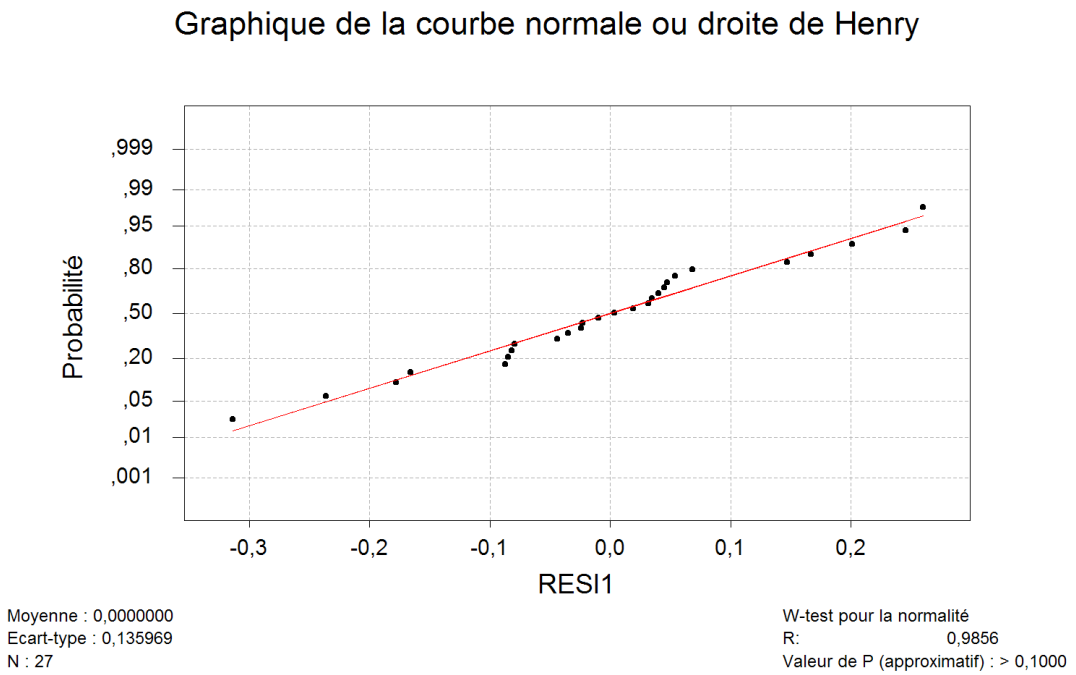
Éléments de corrigé : Rendement de blé

En absence de répétitions, on ne peut tester l'hypothèse d'homogénéité des variances des erreurs. On se contente de vérifier partiellement cette hypothèse par rapport à chacun des couples de trois variables $(P2O5, CaO)$, $(P2O5, Bloc)$ et $(CaO, Bloc)$.

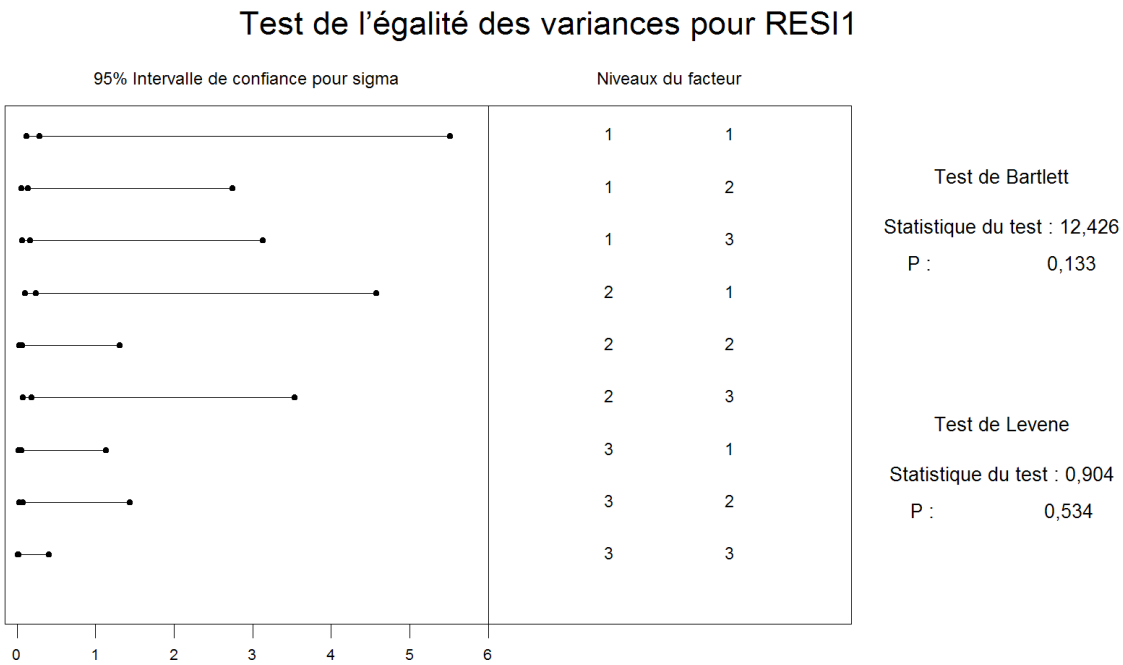
```
Test de l'égalité des variances

Réponse     RESI1
Facteurs    P2O5  CaO
NivConf     95,0000

Test de Levene (pour toute loi de probabilité continue)

Statistique du test : 0,904
P :                   0,534
```

```
Test de l'égalité des variances

Réponse     RESI1
Facteurs    P2O5  Bloc
NivConf     95,0000

Test de Levene (pour toute loi de probabilité continue)

Statistique du test : 0,626
P :                   0,746

Test de l'égalité des variances

Réponse     RESI1
Facteurs    CaO  Bloc
NivConf     95,0000

Test de Levene (pour toute loi de probabilité continue)

Statistique du test : 0,880
P :                   0,551
```

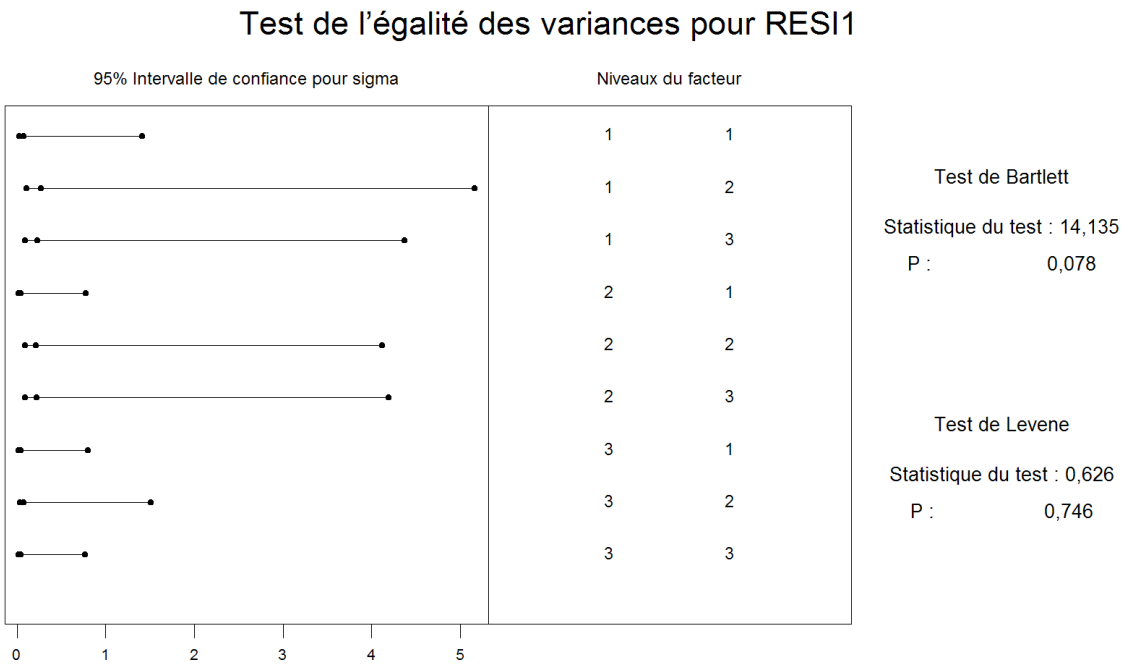
```
W-test pour la normalité
R:                            0,9856)
Valeur de P (approximatif) : > 0,1000
```



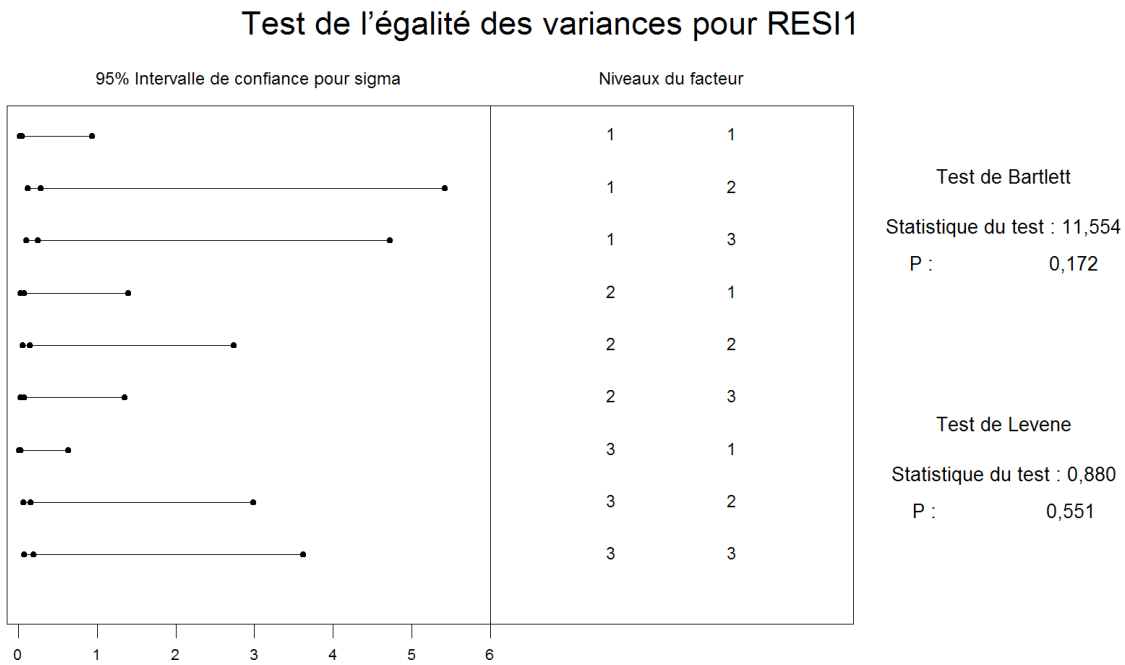

```
Test de l'égalité des variances

Réponse      RESI1
Facteurs     P2O5  CaO
NivConf      95,0000

Test de Bartlett (loi normale)

Statistique du test : 12,426
P :                    0,133
```




```
Test de l'égalité des variances

Réponse      RESI1
Facteurs     P2O5  Bloc
NivConf      95,0000

Test de Bartlett (loi normale)

Statistique du test : 14,135
P :                    0,078
```




```
Test de l'égalité des variances

Réponse      RESI1
Facteurs     CaO  Bloc
NivConf      95,0000

Test de Bartlett (loi normale)

Statistique du test : 11,554
P :                    0,172
```



Modèle linéaire généralisé : Rendement en fonction de P2O5; CaO; Bloc

Facteur	Type	Niveaux	Valeurs
P2O5	fixe	3	1 2 3
CaO	fixe	3	1 2 3
Bloc	aléatoire	3	1 2 3

Analyse de la variance pour Rendemen, en utilisant la SC ajustée pour les tests

Source	DL	SC séq	SC ajust	CM ajust	F	P	
P2O5	2	6,57916	6,57916	3,28958	103,02	0,000	
CaO	2	3,20379	3,20379	1,60189	16,72	0,011	
Bloc	2	1,81336	1,81336	0,90668	13,40	0,108	x
P2O5*CaO	4	0,18764	0,18764	0,04691	0,78	0,568	
P2O5*Bloc	4	0,12773	0,12773	0,03193	0,53	0,717	
CaO*Bloc	4	0,38330	0,38330	0,09583	1,59	0,266	
Erreur	8	0,48067	0,48067	0,06008			
Total	26	12,77565					

x Pas un test F exact.

Terme	Coef	Er-T coef	T	P

Constante		2,67407	0,04717	56,69	0,000
P2O5					
1		-0,67741	0,06671	-10,15	0,000
2		0,19259	0,06671	2,89	0,020
CaO					
1		-0,48630	0,06671	-7,29	0,000
2		0,26815	0,06671	4,02	0,004
Bloc					
1		-0,20074	0,06671	-3,01	0,017
2		-0,16519	0,06671	-2,48	0,038
P2O5*CaO					
1	1	-0,14037	0,09435	-1,49	0,175
1	2	0,04852	0,09435	0,51	0,621
2	1	0,05963	0,09435	0,63	0,545
2	2	0,04519	0,09435	0,48	0,645
P2O5*Bloc					
1	1	0,01074	0,09435	0,11	0,912
1	2	0,09185	0,09435	0,97	0,359
2	1	-0,04593	0,09435	-0,49	0,639
2	2	0,02185	0,09435	0,23	0,823
CaO*Bloc					
1	1	0,06630	0,09435	0,70	0,502
1	2	-0,18259	0,09435	-1,94	0,089
2	1	0,01185	0,09435	0,13	0,903
2	2	-0,03704	0,09435	-0,39	0,705

Remarque :

On remarque qu'ici Minitab 13.31 se comporte de manière étrange pour réaliser le test de l'effet du facteur *Bloc*. En effet, comme indiqué au paragraphe 6.3.1, la statistique de ce test suit une exactement une loi de Fisher à $K-1 = 2$ et $(I-1)(J-1)(K-1) = 2\times2\times2 = 8$ degrés de liberté. Minitab réalise ici l'approximation de Satterthwaite alors que cela n'est pas nécessaire. La version 14 de Minitab n'a pas corrigé ce problème.

Graphique des effets principaux - Moyennes LS pour Rendement

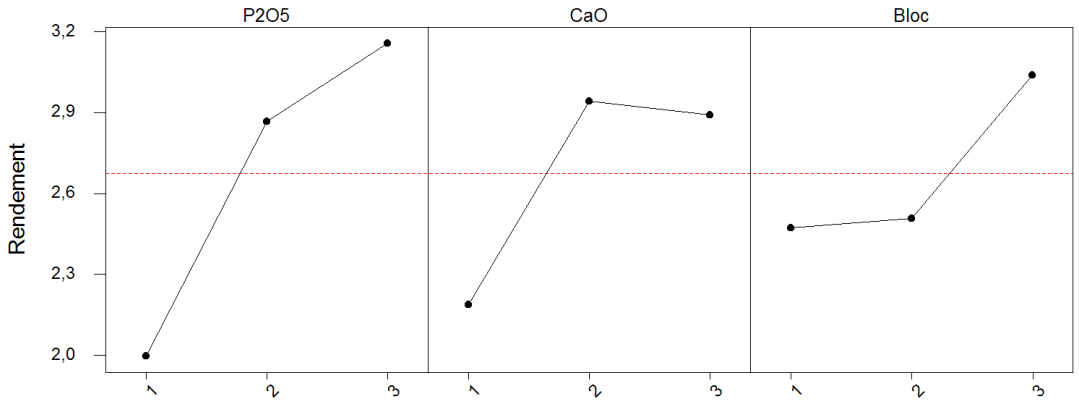
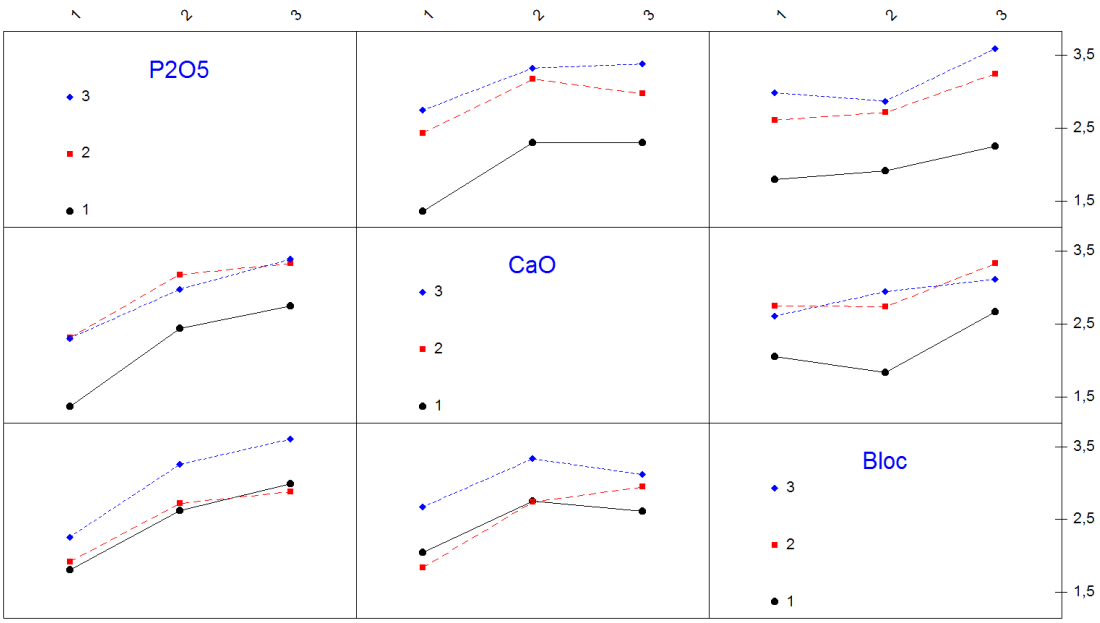


Diagramme d'interaction - Moyennes LS pour Rendement



. . . . . .

12.4.2. Avec répétitions I : Rendement de blé[29]

Vous étudiez le rendement du blé en fonction des trois variables suivantes : la présence ou l'absence d'irrigation, la quantité d'engrais apportée (faible, moyenne ou importante) et la présence ou l'absence de fongicide. Pour chaque combinaison des facteurs, deux essais ont été mis en place.

1. Définir le modèle (sous la forme indicée) que vous utilisez si vous considérez que toutes les interactions d'ordre 2 sont présentes.

Les interactions *fongicide* × *engrais* et *fongicide* × *irrigation* sont considérées a priori comme négligeables par l'expérimentateur.

2. Déterminer les valeurs de tous les coefficients du modèle.
3. Décrire le test de significativité de l'interaction *irrigation* × *engrais*. Conclure.
4. Y a-t-il un effet *engrais* (Décrire le test et la décision) ? Pour le facteur *engrais*, la modalité « importante » fait-elle varier le rendement significativement par rapport au rendement moyen (Décrire le test et la décision) ?
5. Quel rendement prédirez-vous dans les conditions suivantes : avec irrigation, avec fongicide et une dose d'engrais importante ?

. . . . . .

Éléments de corrigé : Rendement de blé

1. Le modèle s'écrit, en notant $Y_{i,j,k,l}$ le rendement de la $l-$ème parcelle cultivée avec l'irrigation i, le fongicide j et la dose d'engrais k :

$$Y_{i,j,k,l} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{i,j} + (\alpha\gamma)_{i,k} + (\beta\gamma)_{j,k} + \epsilon_{i,j,k,l}$$
$$\text{avec } i = 1,2,\ j = 1,2,\ k = 1\ldots 3,\ l = 1,2$$
$$\text{avec les contraintes supplémentaires}$$
$$\sum_{i=1}^{2} \alpha_i = 0,\ \sum_{j=1}^{2} \beta_j = 0 \text{ et } \sum_{k=1}^{3} \gamma_k = 0,$$
$$\sum_{i=1}^{2} (\alpha\beta)_{i,j} = 0,\ \forall j \in \{1,2\} \text{ et } \sum_{j=1}^{2} (\alpha\beta)_{i,j} = 0,\ \forall i \in \{1,2\},$$
$$\sum_{i=1}^{2} (\alpha\gamma)_{i,k} = 0,\ \forall k \in \{1,\ldots,3\} \text{ et } \sum_{k=1}^{3} (\alpha\gamma)_{i,k} = 0,\ \forall i \in \{1,2\},$$
$$\sum_{j=1}^{2} (\beta\gamma)_{j,k} = 0,\ \forall k \in \{1,\ldots,3\} \text{ et } \sum_{k=1}^{3} (\beta\gamma)_{j,k} = 0,\ \forall j \in \{1,2\},$$

[29] Les données de cet exercice sont tirées du livre d'exercices de François Husson et de Jérôme Pagès [2].

avec les hypothèses suivantes pour les résidus :

$\forall\ (i,j,k,l)\ \mathcal{L}(\epsilon_{i,j,k,l}) = \mathcal{N}(0,\sigma^2)$ et $\mathrm{Cov}(\epsilon_{i,j,k,l}, \epsilon_{m,n,o,p}) = 0$ si $(i,j,k,l) \neq (m,n,o,p)$.

2. Les contraintes imposées par le modèle donnent les relations suivantes :

$$\sum_{i=1}^{2} \widehat{\alpha_i}(\boldsymbol{y}) = 0 \text{ donc } \alpha_{\widehat{\text{avec irrigation}}}(\boldsymbol{y}) = -\alpha_{\widehat{\text{sans irrigation}}}(\boldsymbol{y}) = 4,8922$$

$$\sum_{j=1}^{2} \widehat{\beta_j}(\boldsymbol{y}) = 0 \text{ donc } \beta_{\widehat{\text{avec fongicide}}}(\boldsymbol{y}) = -\alpha_{\widehat{\text{sans fongicide}}}(\boldsymbol{y}) = 3,8917$$

$$\sum_{k=1}^{3} \widehat{\gamma_k}(\boldsymbol{y}) = 0 \text{ donc } \gamma_{\widehat{\text{engrais important}}}(\boldsymbol{y}) = -\gamma_{\widehat{\text{engrais moyen}}}(\boldsymbol{y}) - \gamma_{\widehat{\text{engrais faible}}}(\boldsymbol{y}) = 11,2751$$

$$\forall i \ \in \{1,2\}\,,\ \sum_{k=1}^{3} \widehat{\alpha\gamma}_{i,k}(\boldsymbol{y}) = 0 \text{ donc :}$$

$$\widehat{\alpha\gamma}_{\text{ss irr,eng imp}}(\boldsymbol{y}) = -\widehat{\alpha\gamma}_{\text{ss irr,eng faible}}(\boldsymbol{y}) - \widehat{\alpha\gamma}_{\text{ss irr,eng moy}}(\boldsymbol{y}) = 0,7960$$

$$\forall k \ \in \{1,2,3\}\,,\ \sum_{i=1}^{2} \widehat{\alpha\gamma}_{i,k}(\boldsymbol{y}) = 0 \text{ donc :}$$

$$\begin{aligned}
\widehat{\alpha\gamma}_{\text{ss irr,eng imp}}(\boldsymbol{y}) &= -\widehat{\alpha\gamma}_{\text{av irr,eng imp}}(\boldsymbol{y}) = -0,7960\\
\widehat{\alpha\gamma}_{\text{ss irr,eng moy}}(\boldsymbol{y}) &= -\widehat{\alpha\gamma}_{\text{av irr,eng moy}}(\boldsymbol{y}) = 0,9467\\
\widehat{\alpha\gamma}_{\text{ss irr,eng faib}}(\boldsymbol{y}) &= -\widehat{\alpha\gamma}_{\text{av irr,eng faib}}(\boldsymbol{y}) = -0,1507.
\end{aligned}$$

Les autres coefficients étaient renseignés dans le tableau fourni avec l'énoncé.

3. Test de significativité de l'interaction *irrigation* × *engrais* :

$$\boxed{\mathcal{H}_0 : \forall (i,k), (\alpha\gamma)_{i,k} = 0}$$

contre

$$\boxed{\mathcal{H}_1 : \exists (i_0,k_0), (\alpha\gamma)_{i_0,k_0} \neq 0.}$$

La statistique de test utilisée est :

$$F = \frac{CM_{\text{interaction}}}{CM_{\text{résiduelle}}}.$$

Ici $ddl_{\text{interaction}} = (I-1)(K-1) = 2$ et $ddl_{\text{résiduelle}} = n-1-(I-1)-(J-1)-(K-1)-(I-1)(K-1) = 24-1-1-1-2-2 = 17$. Sous l'hypothèse nulle $\mathcal{H}_0$ et une hypothèse de normalité des données, F suit une loi de Fisher à 2 et 17 degrés de liberté. Décision : soit on compare f_{obs} avec le quantile $1-\alpha$ de la loi de Fisher à 2 et 17 degrés de liberté, soit on compare la $p-$valeur avec le seuil α du test. Ici la $p-$valeur vaut $0,7535$, on accepte l'hypothèse nulle $\mathcal{H}_0$ au seuil de $\alpha = 5$ %. On considère donc qu'il n'y a pas d'interaction entre le facteur *irrigation* et le facteur *engrais*.

4. La probabilité critique du test pour l'effet *engrais* est inférieure à $0,0005$, on en déduit que le test est significatif : il y a un effet engrais. On se demande si le coefficient $\gamma_{\text{eng imp}}$ est significativement différent de 0 au seuil $\alpha = 5$ %. Pour cela on teste l'hypothèse $\mathcal{H}_0 \ : \gamma_{\text{eng imp}} = 0 \qquad \mathcal{H}_1 \ : \gamma_{\text{eng imp}} \neq 0$. La statistique du test est $\dfrac{\widehat{\gamma_{\text{eng imp}}}}{\widehat{\sigma}_{\widehat{\gamma_{\text{eng imp}}}}}$ et suit, sous l'hypothèse nulle, une loi de Student avec autant de degrés de liberté que ceux de la variance résiduelle (i.e. 17). La probabilité critique associée à cette valeur est $< 0,0005$ donc $< \alpha = 0,05$. Par conséquent les rendements observés pour un apport important d'engrais sont significativement supérieurs (en moyenne de 11,2751) au rendement moyen.
5. Le modèle retenu est le modèle sans interaction :

$$Y_{i,j,k,l} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{i,j,k,l}, \quad i = 1\ldots 2,\ j = 1,2,\ k = 1\ldots 3,\ l = 1,2$$
$$\text{avec les contraintes } \sum_{i=1}^{2} \alpha_i = 0, \sum_{j=1}^{2} \beta_j = 0 \text{ et } \sum_{k=1}^{3} \gamma_k = 0,$$

et les hypothèses suivantes pour les erreurs :

$$\forall\ (i,j,k,l)\ \mathcal{L}(\epsilon_{i,j,k,l}) = \mathcal{N}(0,\sigma^2) \text{ et } \mathrm{Cov}(\epsilon_{i,j,k,l}, \epsilon_{m,n,o,p}) = 0 \text{ si } (i,j,k,l) \neq (m,n,o,p).$$

Les estimations des paramètres des effets principaux sont identiques dans le modèle avec ou sans interaction car les données sont équilibrées (cf cours sur les plans d'expérience).

$$Y_{\widehat{\text{av irr,av fon,eng imp}}}(\boldsymbol{y}) = \widehat{\mu}(\boldsymbol{y}) + \widehat{\alpha_{\text{av irr}}}(\boldsymbol{y}) + \widehat{\beta_{\text{av fon}}}(\boldsymbol{y}) + \widehat{\gamma_{\text{eng imp}}}(\boldsymbol{y}) = 88,68 \text{ qx/ha}.$$

. . . . . .

12.4.3. Avec répétitions II : Résistance de panneaux de particules à l'arrachage des clous[30]

Une des qualités essentielles des panneaux agglomérés constitués de particules ou de fibres de bois est la résistance de ces panneaux à l'arrachage des clous. Au cours d'une étude

[30] Les données de cet exercice sont tirées du livre de Pierre Dagnélie [4].

on a étudié simultanément l'influence de trois facteurs : la grosseur des clous, le diamètre des anneaux sur lesquels sont déposées les éprouvettes soumises aux essais et la vitesse d'arrachage.
Les essais ont été effectués sur des éprouvettes carrées de 50 mm de côté, les modalités des trois facteurs sont :

-i- le diamètre de la tête des clou est soit de 6,5 mm soit de 8 mm,

-ii- le diamètre des anneaux servant de support est soit de 22 mm soit de 30 mm,

-iii- les vitesses d'arrachage ont été de 22, 45 et 90 par minute.

De plus 5 éprouvettes ont été utilisées pour chacune des 12 combinaisons des modalités du facteur.
On dispose ainsi au total de 60 résultats de la mesure de la résistance de panneaux de particules à l'arrachage des clous exprimée en kg.

grosseur	anneaux	vitesse	essai	resistance
1	1	1	1	54
1	1	1	2	56
1	1	1	3	58
1	1	1	4	51
1	1	1	5	57
1	1	2	1	57
1	1	2	2	58
1	1	2	3	61
1	1	2	4	59
1	1	2	5	55
1	1	3	1	60
1	1	3	2	68
1	1	3	3	61
1	1	3	4	61
1	1	3	5	67
1	2	1	1	54
1	2	1	2	51
1	2	1	3	47
1	2	1	4	51
1	2	1	5	59
1	2	2	1	52
1	2	2	2	56
1	2	2	3	52
1	2	2	4	52
1	2	2	5	53
1	2	3	1	63
1	2	3	2	54
1	2	3	3	65
1	2	3	4	62
1	2	3	5	60

grosseur	anneaux	vitesse	essai	resistance
2	1	1	1	67
2	1	1	2	71
2	1	1	3	72
2	1	1	4	70
2	1	1	5	81
2	1	2	1	79
2	1	2	2	80
2	1	2	3	81
2	1	2	4	80
2	1	2	5	85
2	1	3	1	78
2	1	3	2	78
2	1	3	3	77
2	1	3	4	83
2	1	3	5	79
2	2	1	1	67
2	2	1	2	66
2	2	1	3	62
2	2	1	4	71
2	2	1	5	69
2	2	2	1	72
2	2	2	2	67
2	2	2	3	75
2	2	2	4	70
2	2	2	5	71
2	2	3	1	78
2	2	3	2	81
2	2	3	3	67
2	2	3	4	76
2	2	3	5	75

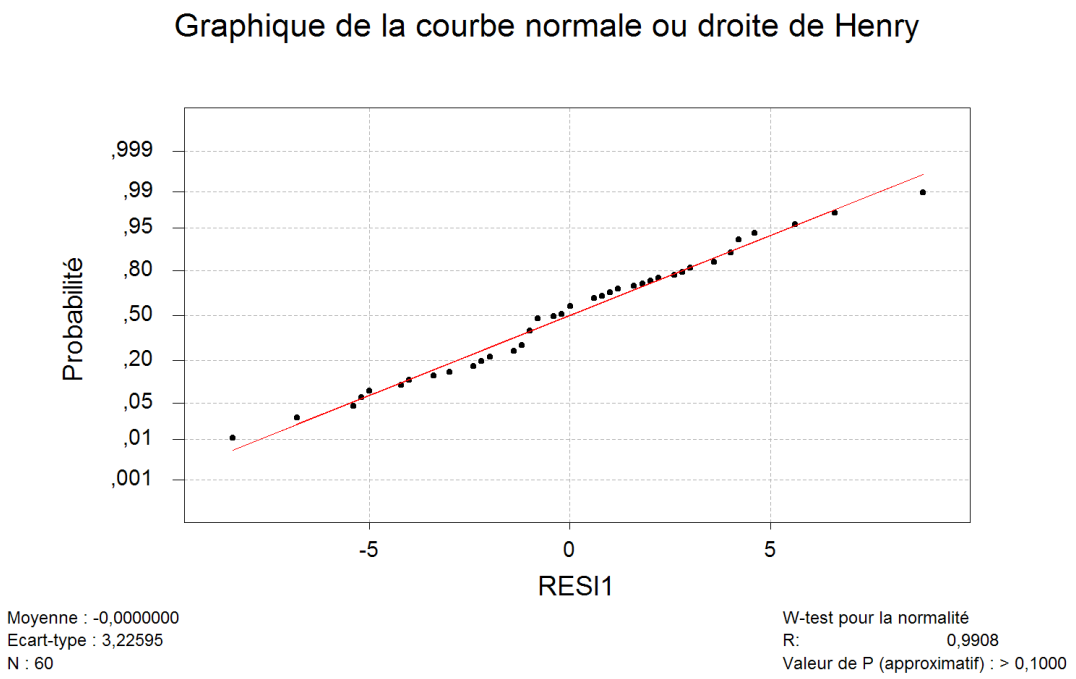
1. Quel modèle d'analyse de la variance peut-on utiliser avec ces données ? Le plan est-il équilibré ? Il y a-t-il des facteurs à effets aléatoires ?
2. On décide dans cette question de négliger l'influence du facteur *essai*. Procéder à l'étude de ces données en utilisant le modèle d'analyse de la variance à trois facteurs le plus complet que vous pouvez utiliser.

. . . . . .

Éléments de corrigé : Résistance de panneaux de particules à l'arrachage des clous

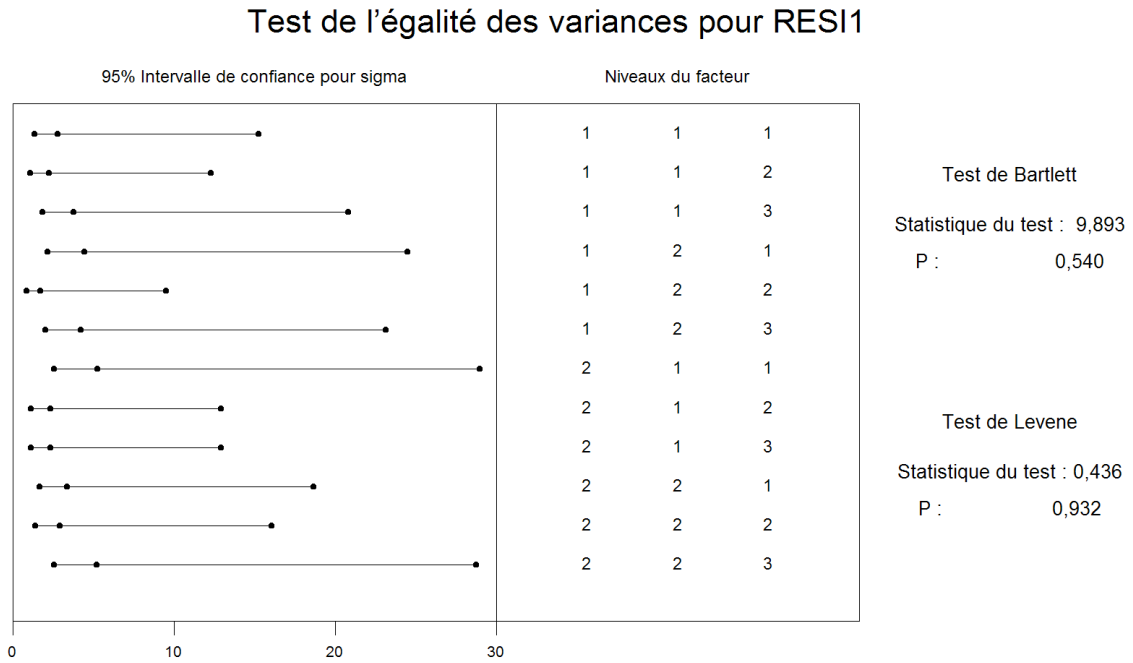
À compléter

```
Test de l'égalité des variances

Réponse     RESI1
Facteurs    grosseur  anneaux  vitesse
NivConf     95,0000

Test de Levene (pour toute loi de probabilité continue)

Statistique du test : 0,436
P :                   0,932
```




```
W-test pour la normalité
R:                            0,9908
Valeur de P (approximatif) : > 0,1000

Test de l'égalité des variances

Réponse     RESI1
Facteurs    grosseur  anneaux  vitesse
```

```
NivConf      95,0000

Test de Bartlett (loi normale)

Statistique du test :  9,893
P :                    0,540
```




```
Modèle linéaire généralisé : resistance en fonction de anneaux; grosseur; ...

Facteur       Type Niveaux  Valeurs
anneaux       fixe       2  1 2
grosseur      fixe       2  1 2
vitesse       fixe       3  1 2 3

Analyse de la variance pour resistan, en utilisant la SC ajustée pour
les tests

Source            DL     SC séq   SC ajust   CM ajust       F      P
anneaux            1     355,27     355,27     355,27   27,77  0,000
grosseur           1    4403,27    4403,27    4403,27  344,23  0,000
vitesse            2     632,10     632,10     316,05   24,71  0,000
anneaux*grosseur   1      29,40      29,40      29,40    2,30  0,136
grosseur*vitesse   2      86,23      86,23      43,12    3,37  0,043
anneaux*vitesse    2      54,03      54,03      27,02    2,11  0,132
```

anneaux*	2	10,30	10,30	5,15	0,40	0,671
grosseur*vitesse						
Erreur	48	614,00	614,00	12,79		
Total	59	6184,60				

Terme			Coef	Er-T coef	T	P
Constante			65,7000	0,4617	142,29	0,000
anneaux						
1			2,4333	0,4617	5,27	0,000
grosseur						
1			-8,5667	0,4617	-18,55	0,000
vitesse						
1			-4,0000	0,6530	-6,13	0,000
2			0,0500	0,6530	0,08	0,939
anneaux*grosseur						
1	1		-0,7000	0,4617	-1,52	0,136
grosseur*vitesse						
1	1		0,6667	0,6530	1,02	0,312
1	2		-1,6833	0,6530	-2,58	0,013
anneaux*vitesse						
1	1		-0,4333	0,6530	-0,66	0,510
1	2		1,3167	0,6530	2,02	0,049
anneaux*grosseur*vitesse						
1	1	1	0,1000	0,6530	0,15	0,879
1	1	2	-0,5500	0,6530	-0,84	0,404

Graphique des effets principaux - Moyennes LS pour resistance

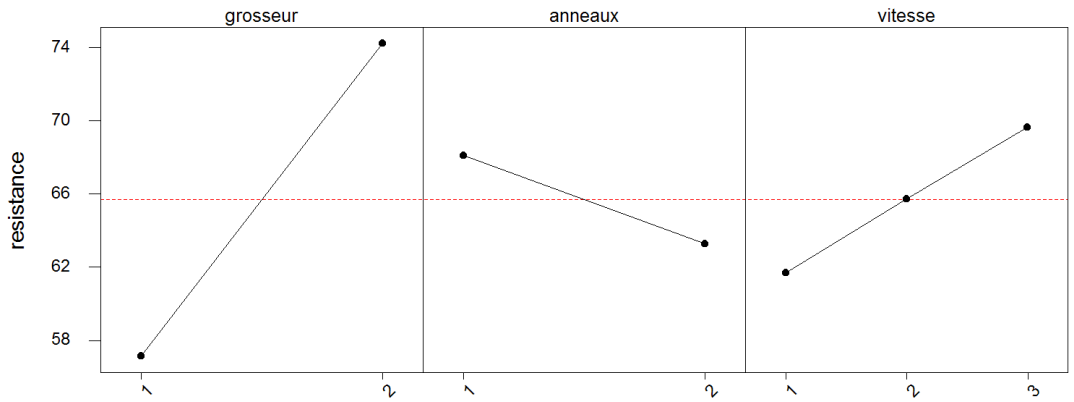
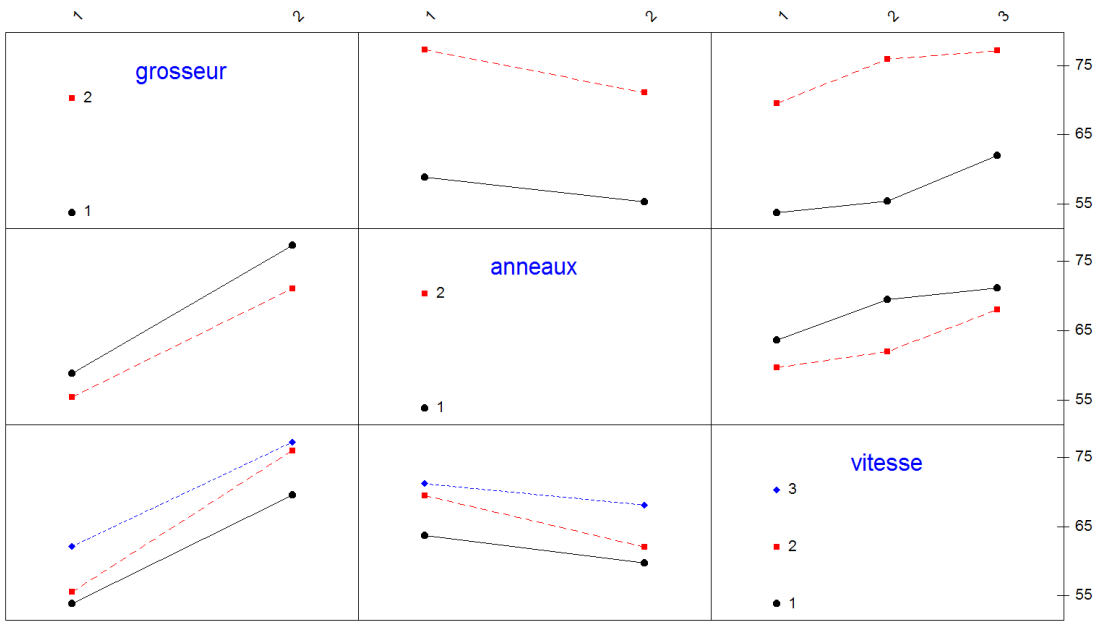


Diagramme d'interaction - Moyennes LS pour resistance



. . . . . .

12.5. Analyse de la variance à trois facteurs totalement emboîtés

Voir l'exemple page 337 du livre de Pierre Dagnélie [4].

. . . . . .

12.6. Analyse de la variance à trois facteurs partiellement emboîtés

Voir l'exemple page 340 du livre de Pierre Dagnélie [4].

. . . . . .

Table des matières

1 Notations **1**
1.1 La notation synthétique d'un total : application aux effectifs 1
1.2 La notation synthétique d'une moyenne 3

2 Introduction **5**
2.1 Un peu de vocabulaire 5
2.2 Plans équilibrés, plans déséquilibrés 8
2.3 Risque global associé au tableau de l'analyse de la variance 9

3 Analyse de la variance à un facteur **10**
3.1 Modèles à effets fixes 10
3.2 Modèles à effets aléatoires 14

4 Analyse de la variance à deux facteurs **16**
4.1 Modèles à effets fixes 16
4.1.1 Sans répétition 16
4.1.2 Avec répétitions 20
4.2 Modèles à effets aléatoires 25
4.2.1 Sans répétition 25
4.2.2 Avec répétitions 27
4.3 Modèles à effets mixtes 31
4.3.1 Sans répétition 31
4.3.2 Avec répétitions 33

5 Analyse de la variance à deux facteurs emboîtés **37**
5.1 Modèles à effets fixes 38
5.1.1 Avec répétitions 38
5.2 Modèles à effets aléatoires 42
5.2.1 Avec répétitions 42
5.3 Modèles à effets mixtes 44
5.3.1 Avec répétitions 44

6 Analyse de la variance à trois facteurs 47
6.1 Modèles à effets fixes . . . 47
6.1.1 Sans répétition . . . 47
6.1.2 Avec répétitions . . . 53
6.2 Modèles à effets aléatoires . . . 59
6.2.1 Sans répétition . . . 59
6.2.2 Avec répétitions . . . 64
6.3 Modèles à effets mixtes . . . 70
6.3.1 Sans répétition . . . 70
6.3.2 Avec répétitions . . . 80

7 Analyse de la variance à trois facteurs totalement emboîtés 91
7.1 Modèles à effets fixes . . . 92
7.1.1 Avec répétitions . . . 92
7.2 Modèles à effets aléatoires . . . 95
7.2.1 Avec répétitions . . . 95
7.3 Modèles à effets mixtes . . . 98
7.3.1 Avec répétitions . . . 98

8 Analyse de la variance à trois facteurs partiellement emboîtés 104
8.1 Introduction . . . 104
8.2 Les différents modèles partiellement emboîtés . . . 105
8.2.1 Nécessitant des répétitions . . . 105
8.2.2 Ne nécessitant pas de répétitions . . . 106

9 Analyse de la variance à l facteurs croisés, partiellement ou totalement emboîtés, $l \geqslant 3$ 107

10 Comparaisons multiples 108
10.1 Contrastes . . . 108
10.1.1 Définition . . . 108
10.1.2 Orthogonalité . . . 109
10.1.3 Estimation . . . 109
10.1.4 Test d'une hypothèse impliquant un constraste . . . 110
10.2 Comparaisons multiples sous l'hypothèse d'homoscédasticité . . . 112
10.2.1 La méthode de Tukey . . . 113
10.2.2 La méthode de Tukey-Kramer . . . 115
10.2.3 La méthode de Hochberg . . . 116
10.2.4 La méthode de Scheffé . . . 116
10.2.5 La méthode de la plus petite différence significative de Fisher . . . 119
10.2.6 La méthode de Bonferroni . . . 120
10.2.7 La méthode de rejet séquentiel de Bonferroni de Holm . . . 121
10.2.8 La méthode de Dunn-Šidák . . . 121
10.2.9 La méthode de Newman-Keuls . . . 122

10.2.10 La méthode de Duncan 124
10.2.11 La méthode de Dunnett 124
10.3 Comparaisons multiples sous l'hypothèse d'hétéroscédasticité 125
10.3.1 La méthode de Games et Howell 125
10.4 Deux méthodes recommandées 126
10.4.1 Le test multiple des étendues de Ryan-Einot-Gabriel-Welsch 126
10.4.2 Tests $F-$multiples 129
10.5 Comparaisons avec le « meilleur » 130
10.5.1 La méthode de Hsu 131
10.5.2 La méthode de Edwards-Hsu 132
10.6 Deux facteurs sans répétition 133
10.6.1 Contrastes 133
10.6.2 Méthode de Tukey 134
10.6.3 Méthode de Scheffé 134
10.6.4 Méthode de Bonferroni et méthodes associées 135
10.7 Deux facteurs avec répétitions 136
10.7.1 Contrastes 136
10.7.2 Méthode de Tukey 138
10.7.3 Méthode de Scheffé 139
10.7.4 Méthode de Bonferroni et méthodes associées 140
10.7.5 Cas d'un plan non équilibré et des comparaisons deux à deux 142
10.8 Deux facteurs emboîtés 142
10.8.1 Contrastes 142
10.8.2 Méthode de Tukey 143
10.8.3 Méthode de Scheffé 144
10.8.4 Méthode de Bonferroni et méthodes associées 146
10.8.5 Cas d'un plan non équilibré et des comparaisons deux à deux 147

11 Puissance des tests de l'analyse de la variance 147
11.1 Analyse de la variance à un facteur 148
11.1.1 Modèle à effets fixes 148
11.1.2 Modèle à effets aléatoires 151
11.2 Analyse de la variance à deux facteurs sans interaction 152
11.2.1 Modèle à effets fixes 152
11.2.2 Modèle à effets aléatoires 153
11.2.3 Modèle à effets mixtes 154
11.3 Analyse de la variance à deux facteurs avec interaction 155
11.3.1 Modèle à effets fixes 155
11.3.2 Modèle à effets aléatoires 157
11.3.3 Modèle à effets mixtes 158
11.4 Analyse de la variance à deux facteurs emboîtés 160
11.4.1 Modèle à effets fixes 160
11.4.2 Modèle à effets aléatoires 161

11.4.3 Modèle à effets mixtes 162
11.5 Analyse de la variance à trois facteurs sans interaction 163
11.5.1 Modèle à effets fixes 163
11.5.2 Modèle à effets aléatoires 164
11.5.3 Modèle à effets mixtes 166
11.6 Analyse de la variance à trois facteurs avec interaction 169
11.6.1 Modèle à effets fixes 169
11.6.2 Modèle à effets aléatoires 171
11.6.3 Modèle à effets mixtes 173
11.7 Analyse de la variance à trois facteurs totalement emboîtés 177
11.7.1 Modèle à effets fixes 178
11.7.2 Modèle à effets aléatoires 179
11.7.3 Modèle à effets mixtes 180

12 Exemples **183**
12.1 Analyse de la variance à un facteur 183
12.1.1 Carburateurs 183
12.2 Analyse de la variance à deux facteurs 189
12.2.1 Sans répétition : Carburateurs (suite et fin) 189
12.2.2 Avec répétitions et plan équilibré I : Eaux gazeuses 199
12.2.3 Avec répétitions et plan équilibré II : Les rats 201
12.2.4 Avec répétitions et plan équilibré III : Les biscuits 210
12.2.5 Avec répétitions et plan déséquilibré : Les chocolats 216
12.3 Analyse de la variance à deux facteurs emboîtés 219
12.3.1 Plan déséquilibré : Les cyclamens 219
12.4 Analyse de la variance à trois facteurs 223
12.4.1 Sans répétition : Détermination d'une fumure optimale pour le blé . 223
12.4.2 Avec répétitions I : Rendement de blé 232
12.4.3 Avec répétitions II : Résistance de panneaux de particules à l'arrachage des clous 234
12.5 Analyse de la variance à trois facteurs totalement emboîtés 241
12.6 Analyse de la variance à trois facteurs partiellement emboîtés 242

Références

[1] P. Chapouille. *Planification et analyse des expériences.* Masson, Paris, 1973.

[2] François Husson and Jérôme Pagès. *Statistiques générales pour utilisateurs.* PUR, Rennes, a trouver.

[3] P. Dagnélie. *Statistique Théorique et Appliquée*, volume 1. De Boeck & Larcier, Bruxelles, 1998.

[4] P. Dagnélie. *Statistique Théorique et Appliquée*, volume 2. De Boeck & Larcier, Bruxelles, 1998.

[5] B. Falissard. *Comprendre et utiliser les statistiques dans les sciences de la vie.* Abrégés. Masson, Paris, 3ème edition, 2005.

[6] J. L. Fleiss. *The Design and Analysis of Clinical Experiments.* John Wiley, New-York, 1986.

[7] S. Kotz, C. B. Read, and N. Balakrishnan, editors. *Encyclopædia Of Statistical Sciences.* Wiley-Interscience, 2nd edition, 1996.

[8] G. Parreins. *Techniques Statistiques : moyens rationnels de choix et de décision.* Dunod technique, Paris, 1974.

[9] G. Philippeau. *Théorie des plans d'expérience.* ICTF, Paris, 1982.

[10] W. R. Rice. Analysing tables of statistical tests. *Evolution*, 43(1) :223–225, 1989.

[11] H. Sahai and M. I. Ageel. *The Anamysis of Variance : Fixed, Random and Mixed Models.* Birkhäuser, Boston, 2000.