

# Éléments d'analyse de la variance

FRÉDÉRIC BERTRAND <sup>1</sup>

Université de Strasbourg  
Institut de Recherche Mathématique Avancée<sup>2</sup>



9 septembre 2019

1. Courriel : [fbertran@math.unistra.fr](mailto:fbertran@math.unistra.fr)
2. 7, rue René Descartes, 67084 STRASBOURG CEDEX



# Préface

Ce manuel de cours a pour but de vous guider dans vos réflexions pour les études statistiques que vous aurez à réaliser à l'aide de modèle d'analyse de la variance. Nous présentons un vaste paysage des modèles de l'analyse de la variance à un ou plusieurs facteurs. Ces facteurs peuvent être croisés ou hiérarchisés et à effets fixes ou aléatoires. La mise en évidence d'un effet de l'un des termes du modèle statistique est généralement suivie par l'analyse des effets comparés des modalités de ce terme et repose sur l'utilisation de procédures de comparaisons multiples. Nous avons rassemblé et détaillé la plupart des procédures qui sont utilisées dans la pratique ainsi que deux d'entre elles dont l'utilisation est recommandée par l'Encyclopédie de Statistique.

Pour chacun des modèles étudiés, nous commençons par détailler la modélisation statistique, ses conditions d'utilisation et ses limitations afin que vous puissiez être ensuite capable de l'utiliser avec pertinence. Nous proposons ensuite des exercices, le plus souvent intégralement corrigés avec le logiciel.

Les références [Dag98a], [Dag98b], [KRB96], [Fal05], [Par74], [SA00] ayant servi à l'élaboration de ce manuel sont mentionnées dans la bibliographie.

Vos remarques, vos commentaires, vos suggestions, et même vos encouragements, seront accueillis avec plaisir.

`fbertran@math.unistra.fr`



# Table des matières

Préface	iii
Table des matières	iv
Avant-propos	ix
<b>1 Introduction</b>	<b>1</b>
1.1 Un peu de vocabulaire . . . . .	1
1.2 Plans équilibrés, plans déséquilibrés . . . . .	4
1.3 Risque global associé au tableau de l'analyse de la variance . . . . .	5
<b>2 Analyse de la variance à un facteur</b>	<b>7</b>
2.1 Modèles à effets fixes . . . . .	7
2.2 Modèles à effets aléatoires . . . . .	10
<b>3 Analyse de la variance à deux facteurs</b>	<b>13</b>
3.1 Modèles à effets fixes . . . . .	13
3.1.1 Sans répétition . . . . .	13
3.1.2 Avec répétitions . . . . .	17
3.2 Modèles à effets aléatoires . . . . .	21
3.2.1 Sans répétition . . . . .	21
3.2.2 Avec répétitions . . . . .	24
3.3 Modèles à effets mixtes . . . . .	27
3.3.1 Sans répétition . . . . .	27
3.3.2 Avec répétitions . . . . .	30
<b>4 Analyse de la variance à deux facteurs emboîtés</b>	<b>35</b>
4.1 Modèles à effets fixes . . . . .	36
4.1.1 Avec répétitions . . . . .	36
4.2 Modèles à effets aléatoires . . . . .	39
4.2.1 Avec répétitions . . . . .	39
4.3 Modèles à effets mixtes . . . . .	42
4.3.1 Avec répétitions . . . . .	42
<b>5 Analyse de la variance à trois facteurs</b>	<b>47</b>
5.1 Modèles à effets fixes . . . . .	47
5.1.1 Sans répétition . . . . .	47
5.1.2 Avec répétitions . . . . .	54

5.2	Modèles à effets aléatoires . . . . .	62
5.2.1	Sans répétition . . . . .	62
5.2.2	Avec répétitions . . . . .	68
5.3	Modèles à effets mixtes . . . . .	74
5.3.1	Sans répétition . . . . .	74
5.3.2	Avec répétitions . . . . .	86
<b>6</b>	<b>Analyse de la variance à trois facteurs totalement emboîtés</b>	<b>101</b>
6.1	Modèles à effets fixes . . . . .	101
6.1.1	Avec répétitions . . . . .	101
6.2	Modèles à effets aléatoires . . . . .	107
6.2.1	Avec répétitions . . . . .	107
6.3	Modèles à effets mixtes . . . . .	110
6.3.1	Avec répétitions . . . . .	110
<b>7</b>	<b>Analyse de la variance à trois facteurs partiellement emboîtés</b>	<b>117</b>
7.1	Introduction . . . . .	117
7.2	Les différents modèles partiellement emboîtés . . . . .	117
7.2.1	Nécessitant des répétitions . . . . .	117
7.2.2	Ne nécessitant pas de répétitions . . . . .	118
<b>8</b>	<b>Analyse de la variance à <math>l</math> facteurs, <math>l &gt; 2</math></b>	<b>121</b>
<b>9</b>	<b>Comparaisons multiples</b>	<b>123</b>
9.1	Contrastes . . . . .	123
9.1.1	Définition . . . . .	123
9.1.2	Orthogonalité . . . . .	124
9.1.3	Estimation . . . . .	124
9.1.4	Test d'une hypothèse impliquant un contraste . . . . .	125
9.2	Comparaisons multiples sous l'hypothèse d'homoscédasticité . . . . .	126
9.2.1	La méthode de Tukey . . . . .	127
9.2.2	La méthode de Tukey-Kramer . . . . .	129
9.2.3	La méthode de Hochberg . . . . .	130
9.2.4	La méthode de Scheffé . . . . .	131
9.2.5	La méthode de la plus petite différence significative de Fisher . . . . .	133
9.2.6	La méthode de Bonferroni . . . . .	134
9.2.7	La méthode de rejet séquentiel de Bonferroni de Holm . . . . .	135
9.2.8	La méthode de Dunn-Šidák . . . . .	135
9.2.9	La méthode de Newman-Keuls . . . . .	136
9.2.10	La méthode de Duncan . . . . .	137
9.2.11	La méthode de Dunnett . . . . .	138
9.3	Comparaisons multiples sous l'hypothèse d'hétéroscédasticité . . . . .	139
9.3.1	La méthode de Games et Howell . . . . .	139
9.4	Deux méthodes recommandées . . . . .	139
9.4.1	Le test multiple des étendues de Ryan-Einot-Gabriel-Welsch . . . . .	140
9.4.2	Tests $F$ -multiples . . . . .	142
9.5	Comparaisons avec le <i>meilleur</i> . . . . .	143
9.5.1	La méthode de Hsu . . . . .	144

9.5.2	La méthode de Edwards-Hsu . . . . .	144
9.6	Deux facteurs sans répétition . . . . .	146
9.6.1	Contrastes . . . . .	146
9.6.2	Méthode de Tukey . . . . .	147
9.6.3	Méthode de Scheffé . . . . .	147
9.6.4	Méthode de Bonferroni et méthodes associées . . . . .	148
9.7	Deux facteurs avec répétitions . . . . .	148
9.7.1	Contrastes . . . . .	149
9.7.2	Méthode de Tukey . . . . .	150
9.7.3	Méthode de Scheffé . . . . .	151
9.7.4	Méthode de Bonferroni et méthodes associées . . . . .	152
9.7.5	Cas d'un plan non équilibré et des comparaisons deux à deux . . . . .	154
9.8	Deux facteurs emboîtés . . . . .	154
9.8.1	Contrastes . . . . .	154
9.8.2	Méthode de Tukey . . . . .	155
9.8.3	Méthode de Scheffé . . . . .	156
9.8.4	Méthode de Bonferroni et méthodes associées . . . . .	158
9.8.5	Cas d'un plan non équilibré et des comparaisons deux à deux . . . . .	159

**10 Puissance des tests de l'analyse de la variance** **161**

10.1	Analyse de la variance à un facteur . . . . .	162
10.1.1	Modèle à effets fixes . . . . .	162
10.1.2	Modèle à effets aléatoires . . . . .	164
10.2	Analyse de la variance à deux facteurs sans interaction . . . . .	165
10.2.1	Modèle à effets fixes . . . . .	165
10.2.2	Modèle à effets aléatoires . . . . .	166
10.2.3	Modèle à effets mixtes . . . . .	167
10.3	Analyse de la variance à deux facteurs avec interaction . . . . .	168
10.3.1	Modèle à effets fixes . . . . .	169
10.3.2	Modèle à effets aléatoires . . . . .	170
10.3.3	Modèle à effets mixtes . . . . .	171
10.4	Analyse de la variance à deux facteurs emboîtés . . . . .	173
10.4.1	Modèle à effets fixes . . . . .	173
10.4.2	Modèle à effets aléatoires . . . . .	174
10.4.3	Modèle à effets mixtes . . . . .	175
10.5	Analyse de la variance à trois facteurs sans interaction . . . . .	176
10.5.1	Modèle à effets fixes . . . . .	177
10.5.2	Modèle à effets aléatoires . . . . .	178
10.5.3	Modèle à effets mixtes . . . . .	179
10.6	Analyse de la variance à trois facteurs avec interaction . . . . .	183
10.6.1	Modèle à effets fixes . . . . .	183
10.6.2	Modèle à effets aléatoires . . . . .	184
10.6.3	Modèle à effets mixtes . . . . .	186
10.7	Analyse de la variance à trois facteurs totalement emboîtés . . . . .	191
10.7.1	Modèle à effets fixes . . . . .	192
10.7.2	Modèle à effets aléatoires . . . . .	193
10.7.3	Modèle à effets mixtes . . . . .	195





# Avant-propos

## Notations

Nous commençons par quelques notations, couramment utilisées en statistique, à la fois pour alléger les formules et pour les rendre plus compréhensibles.

### La notation synthétique d'un total : application aux effectifs

Nous considérons 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 nous considérons, chacune des cases du tableau comportant ainsi une ou plusieurs données. Nous notons 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  $+$ . Nous adoptons 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, nous obtenons les notations ci-dessous.

1. Nous avons observé  $n$  valeurs de la variable  $Y$ . Chacune d'entre elle correspond à un résultat expérimental effectué dans la condition  $i$ ,  $1 \leq i \leq I$ . Nous construisons 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 nous avons  $n_+ = n$ .

2. Nous avons observé  $n$  valeurs de la variable  $Y$  que nous regroupons en fonction des valeurs des deux indices  $i$ ,  $1 \leq i \leq I$ , et  $j$ ,  $1 \leq j \leq J$ , associés aux conditions expérimentales. Nous construisons 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  $n_{+,j}$ .

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

- 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 nous avons  $n_{+,+} = n$ .

2. Nous avons observé  $n$  valeurs de la variable  $Y$  que nous regroupons en fonction des valeurs des trois indices  $i$ ,  $1 \leq i \leq I$ ,  $j$ ,  $1 \leq j \leq J$ , et  $k$ ,  $1 \leq k \leq K$ , associés aux conditions expérimentales. Nous construisons 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  $n_{+,+,+}$ .

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

Bien entendu nous avons  $n_{+,+,+} = n$ .

Nous pouvons bien sûr généraliser cette notation lorsque nous considérons plus de trois indices.

### La notation synthétique d'une moyenne

Lorsque nous disposons d'un tableau de données indexées par un ou plusieurs indices nous notons la moyenne par rapport à l'un de ces indices en le remplaçant par le symbole  $\bullet$ . Nous adoptons 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, nous obtenons les notations ci-dessous.

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

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

2. Nous avons observé  $n = I \times J$  valeurs de la variable  $Y$  indexée par deux indices  $i$  et  $j$ , notées  $y_{1,1}, \dots, y_{1,J}, y_{2,1}, \dots, y_{2,J}, \dots, y_{I,1}, \dots, y_{I,J}$ .
  - La moyenne de ces valeurs  $y_{1,1}, \dots, y_{I,J}$  par rapport à l'indice  $i$  est notée  $y_{\bullet,j}$ . Il s'agit simplement de la moyenne  $\bar{y}_j$  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  $y_{1,1}, \dots, y_{I,J}$  par rapport à l'indice  $j$  est notée  $y_{i,\bullet}$ . Il s'agit simplement de la moyenne  $\bar{y}_i$  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  $y_{1,1}, \dots, y_{I,J}$  par rapport aux indices  $i$  et  $j$  est notée  $y_{\bullet,\bullet}$ . Il s'agit simplement de la moyenne  $\bar{y}$ , 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. Nous avons observé  $n = I \times J \times K$  valeurs de la variable  $Y$  indexée par trois indices  $i, j$  et  $k$ , notées  $y_{1,1,1}, \dots, y_{1,1,K}, y_{2,1,1}, \dots, y_{2,1,K}, \dots, y_{I,1,1}, \dots, y_{I,1,K}, \dots, y_{1,2,1}, \dots, y_{1,2,K}, \dots, y_{I,J,1}, \dots, y_{I,J,K}$ .

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

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

Nous étendons cette notation à la situation où les valeurs que la variable  $Y$  va prendre ne sont pas encore connues, c'est-à-dire lorsque nous remplaçons 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.

Nous pouvons bien sûr généraliser cette notation lorsque les valeurs prises par la variable  $Y$  dépendent de plus de trois indices.



# Chapitre 1

## Introduction

### 1.1. Un peu de vocabulaire

- i. Nous appelons **réponse** une variable dont nous cherchons à comprendre le comportement. Nous considérerons 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. Nous appelons **facteur** toute variable dont nous voulons nous servir pour analyser les variations de la réponse. Nous notons généralement les facteurs par des lettres romaines majuscules  $A, B, \dots$ . 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 indiquant la majuscule romaine associée au facteur :  $A_1, \dots, A_I$  où  $I$  désigne le nombre total de modalités du facteur  $A$ . Pour l'analyse d'une réponse à l'aide de variables quantitatives nous utiliserons 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 nous pourrions nous servir de l'analyse de la covariance.
- iii. Un **modèle statistique** est une équation reliant les différentes mesures  $Y_{i,j,\dots}$  de la réponse  $Y$  aux effets des niveaux  $A_1, \dots, A_I, B_1, \dots, B_J, \dots$  des facteurs  $A, B, \dots$  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  $\epsilon_{i,j,\dots}$  sur lesquelles nous faisons porter certaines hypothèses :

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

- 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 nous souhaitons 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.

Nous introduisons le modèle :

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

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. Nous postulons les hypothèses classiques suivantes pour les erreurs :

$$\text{Indép. } \epsilon_{i,j} \perp\!\!\!\perp \epsilon_{k,l} \text{ si } (i,j) \neq (k,l) \text{ avec } 1 \leq i,k \leq I, \text{ et } 1 \leq j,l \leq J,$$

$$\forall (i,j), 1 \leq i \leq I, 1 \leq j \leq J, \mathcal{L}(\epsilon_{i,j}) = \mathcal{N}(0, \sigma^2).$$

Les  $\mu, \alpha_1, \dots, \alpha_I$  sont donc des **termes du modèle**.  $\epsilon_{i,j}$  est parfois appelé le terme d'erreur dans le modèle.

Nous remarquons que les hypothèses faites sur les termes d'erreur  $\epsilon_{i,j}$  font partie intégrante de la définition du modèle. Ceci explique le soin particulier que nous devons 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**, nous le notons généralement par la lettre grecque minuscule associée à la majuscule romaine désignant le facteur, par exemple pour le facteur  $A$  nous noterons  $\alpha_i$  l'effet du niveau  $A_i$ , pour  $1 \leq i \leq I$ .
- L'ensemble des termes associés aux effets de l'interaction de deux des facteurs du modèle est appelé **interaction d'ordre 1**. Nous les notons généralement par les lettres grecques minuscules mises entre parenthèses associées aux majuscules romaines désignant les deux facteurs dont nous étudions 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  $B$ , nous notons les termes associés à ces effets par  $(\alpha\beta)_{1,1}, (\alpha\beta)_{1,2}, \dots, (\alpha\beta)_{1,I}, \dots, (\alpha\beta)_{2,1}, \dots, (\alpha\beta)_{i,j}, \dots, (\alpha\beta)_{I,J}$  pour  $1 \leq i \leq I$  et  $1 \leq j \leq J$ .
- Plus généralement l'ensemble des termes associés aux effets de l'interaction de  $k$  des facteurs du modèle,  $k \geq 2$ , est appelé **interaction d'ordre  $k - 1$** . Nous les notons généralement par les lettres grecques minuscules mises entre parenthèses associées aux majuscules romaines désignant les  $k$  facteurs dont nous étudions 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 nous souhaitons par exemple étudier une réponse continue  $Y$  à l'aide de trois facteurs qualitatifs à effets fixes  $A, B$  et  $C$  pour lesquels nous avons 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.



## 1.1 Un peu de vocabulaire

---

Nous introduisons 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  $\sum_{i=1}^I \alpha_i = 0$ ,  $\sum_{j=1}^J \beta_j = 0$  et  $\sum_{k=1}^K \gamma_k = 0$ ,

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

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

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

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

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

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

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. Nous postulons les hypothèses classiques suivantes pour les erreurs :

$$\forall (i, j, k, l), 1 \leq i \leq I, 1 \leq j \leq J, 1 \leq k \leq K, 1 \leq l \leq L, \mathcal{L}(\epsilon_{i,j,k,l}) = \mathcal{N}(0, \sigma^2),$$

$$\epsilon_{i,j,k,l} \perp\!\!\!\perp \epsilon_{m,n,o,p} \text{ si } (i, j, k, l) \neq (m, n, o, p)$$

avec  $1 \leq i, m \leq I, 1 \leq j, n \leq J, 1 \leq k, o \leq K$  et  $1 \leq l, p \leq L$ .

Les termes  $\alpha_i$ ,  $\beta_j$  et  $\gamma_k$  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  $C$ .

Nous remarquons qu'à nouveau les hypothèses faites sur les termes d'erreur  $\epsilon_{i,j,k,l}$  font partie intégrante de la définition du modèle. Ceci explique le soin particulier que nous devons 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**, nous introduirons d'autres notations qui seront décrites à la section 4 et à la section 6.

## 1.2. 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 nous appellerons plan équilibré tout plan vérifiant cette condition.<sup>1</sup>

Pour de multiples raisons nous conseillons, 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 nous sommes dans une situation dénommée 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 nous cherchons à estimer.
2. Lorsque nous ne disposons 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 existe, entre les différents effets pour lesquels des tests sont réalisés est associée à une répartition des essais où le plan est équilibré. Nous renvoyons au chapitre 10 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 chapitres suivants 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 il n'est légitime de supprimer 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 nous nous servons 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.

---

1. Cette condition est plus restrictive que celle que nous pouvons définir plus généralement en statistique pour étudier des modèles d'analyse de la variance. Nous pouvons aussi facilement traiter le cas où le plan est déséquilibré mais à effectifs proportionnels. Par exemple, dans le cas où nous considérons deux facteurs, un dispositif expérimental déséquilibré à effectifs proportionnels est tel que :

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

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

Voir par exemple la situation étudiée dans le paragraphe ?? où nous insistons sur les différences qui découlent de cette situation. Nous renvoyons à [Dag98b] pour plus de détails.

2. D'autre part, les lois des statistiques de test que nous utilisons 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. Nous renvoyons à nouveau à [Dag98b] pour plus de détails.

Dans la suite, nous nous bornerons 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.

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

Lorsque nous réalisons 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 nous étudions permet de prendre plusieurs décisions, chacune associée au test d'une hypothèse nulle. Nous prenons 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 1.3.1.** 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 3.1.2.

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  $\alpha$  vérifiera alors la relation :

$$\alpha \leq 1 - (1 - \alpha')(1 - \alpha'')(1 - \alpha''') \leq \alpha' + \alpha'' + \alpha'''.$$

La dernière inégalité est une application de **l'inégalité de Bonferroni**. Ainsi si nous voulons prendre une décision associée à un risque global de  $\alpha = 5\%$ , nous pourrions fixer un seuil individuel de  $\alpha' = \alpha'' = \alpha''' = 5/3\% \approx 1,67\%$  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 nous avons détectés. Elle est d'autant plus intéressante à faire que le nombre de tests réalisés est important.

**Remarque 1.3.1.** 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 maîtrise 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 adienne il faut savoir que nous entrons dans le domaine de l'analyse de tableaux de résultats. Nous pourrions consulter [Par74] ou [Ric89] pour plus de précisions.

# Chapitre 2

## Analyse de la variance à un facteur

Dans tout ce chapitre, nous utilisons les notations définies à la section Notations, page ix.

### 2.1. 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 nous effectuons  $J \geq 2$  mesures d'une réponse  $Y$  qui est une variable continue. Nous notons  $n = I \times J$  le nombre total de mesures ayant été effectuées.

Nous introduisons le modèle :

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

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. Nous postulons les hypothèses classiques suivantes pour les erreurs :

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

Nous supposons que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fait l'objet d'un paragraphe du chapitre ??.

Nous regroupons les valeurs que peut prendre la réponse  $Y$  dans la condition  $A_i$  lors des  $J$  répétitions dans le tableau suivant :

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

Nous rappelons 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 :

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

Nous rappelons la relation fondamentale de l'ANOVA :

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

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

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

La variation due au facteur  $A$  observée sur la liste de données  $\mathbf{y}$  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  $\mathbf{y}$  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  $\mathbf{y}$  est égale à :

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

## 2.1 Modèles à effets fixes

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

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

Nous résumons ces informations dans le tableau de l'ANOVA ci-dessous :

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

Nous souhaitons faire le test d'hypothèse suivant :

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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, nous pouvons procéder à des comparaisons multiples des différents effets des niveaux du facteur voir le chapitre 9.

Les estimateurs  $\hat{\mu}, \hat{\alpha}_1, \dots, \hat{\alpha}_I, \hat{\sigma}^2$  des paramètres  $\mu, \alpha_1, \dots, \alpha_I, \sigma^2$  du modèle sont donnés par les formules suivantes :

$$\hat{\mu} = Y_{\bullet, \bullet} = \bar{Y}, \quad \hat{\alpha}_i = Y_{i, \bullet} - \hat{\mu}, \quad 1 \leq i \leq I,$$

$$\hat{\sigma}^2 = \frac{sc_R}{n - I} = \frac{\sum_{i=1}^I \sum_{j=1}^J (Y_{i,j} - Y_{i, \bullet})^2}{n - I} = S_R^2.$$

Ce sont des estimateurs sans biais<sup>1</sup>.

1. La notion de biais d'un estimateur est définie dans le chapitre ??.

Les estimations, obtenues pour la liste de données  $\mathbf{y}$  et notées  $\widehat{\mu}(\mathbf{y})$ ,  $\widehat{\alpha}_1(\mathbf{y})$ ,  $\dots$ ,  $\widehat{\alpha}_I(\mathbf{y})$ ,  $\widehat{\sigma}^2(\mathbf{y})$ , des paramètres  $\mu$ ,  $\alpha_1$ ,  $\dots$ ,  $\alpha_I$ ,  $\sigma^2$  du modèle se déduisent des formules ci-dessus :

$$\begin{aligned}\widehat{\mu}(\mathbf{y}) &= y_{\bullet,\bullet} = \bar{y}, \quad \widehat{\alpha}_i(\mathbf{y}) = y_{i,\bullet} - \widehat{\mu}(\mathbf{y}), \quad 1 \leq i \leq I, \\ \widehat{\sigma}^2(\mathbf{y}) &= \frac{SC_R}{n - I} = \frac{\sum_{i=1}^I \sum_{j=1}^J (y_{i,j} - y_{i,\bullet})^2}{n - I} = s_R^2.\end{aligned}$$

## 2.2. 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  $\alpha_i$ , sont distribués suivant une loi normale centrée de variance  $\sigma_A^2$ . Le modèle ne dépend plus que de trois paramètres  $\mu$ ,  $\sigma^2$  et  $\sigma_A^2$ . Pour chacune des modalités nous effectuons  $J \geq 2$  mesures d'une réponse  $Y$  qui est une variable continue. Nous notons  $n = I \times J$  le nombre total de mesures ayant été effectuées.

Nous introduisons le modèle :

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

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

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

ainsi que l'indépendance des effets aléatoires :

$$\alpha_i \perp\!\!\!\perp \alpha_j \text{ si } i \neq j \text{ et } 1 \leq i, j \leq I.$$

Nous postulons les hypothèses supplémentaires pour les erreurs :

$$\begin{aligned}\forall (i, j), 1 \leq i \leq I, 1 \leq j \leq J, \quad \mathcal{L}(\epsilon_{i,j}) &= \mathcal{N}(0, \sigma^2), \\ \epsilon_{i,j} \perp\!\!\!\perp \epsilon_{k,l} \text{ si } (i, j) \neq (k, l) \text{ avec } 1 \leq i, k \leq I, \text{ et } 1 \leq j, l \leq J,\end{aligned}$$

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

$$\alpha_i \perp\!\!\!\perp \epsilon_{j,k} \text{ si } 1 \leq i, j \leq I, \text{ et } 1 \leq k \leq J.$$

Nous supposons que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fait l'objet d'un paragraphe du chapitre ??.

Nous utilisons les quantités  $SC_F$ ,  $SC_R$ ,  $SC_{TOT}$ ,  $sc_F$ ,  $sc_R$ ,  $sc_{TOT}$  introduites à la section 2.1.

Nous rappelons la relation fondamentale de l'ANOVA :

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

Nous résumons les informations dans le tableau de l'ANOVA ci-dessous :



## 2.2 Modèles à effets aléatoires

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

Nous souhaitons 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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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  $\hat{\mu}$ ,  $\widehat{\sigma}_A^2$ ,  $\widehat{\sigma}^2$  des paramètres  $\mu$ ,  $\sigma_A^2$ ,  $\sigma^2$  du modèle sont donnés par les formules suivantes :

$$\hat{\mu} = Y_{\bullet,\bullet} = \bar{Y}, \quad \widehat{\sigma}_A^2 = \frac{1}{J} (S_F^2 - S_R^2), \quad \widehat{\sigma}^2 = \frac{SC_R}{n - I} = S_R^2,$$

où  $S_F^2 = \frac{SC_F}{I - 1}$  et  $S_R^2 = \frac{SC_R}{n - I}$ .

Ce sont des estimateurs sans biais<sup>2</sup>.

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

$$\hat{\mu}(\mathbf{y}) = y_{\bullet,\bullet} = \bar{y}, \quad \widehat{\sigma}_A^2(\mathbf{y}) = \frac{1}{J} (s_F^2 - s_R^2), \quad \widehat{\sigma}^2(\mathbf{y}) = \frac{sc_R}{n - I} = s_R^2.$$

---

2. La notion de biais d'un estimateur est définie dans le chapitre ??.



# Chapitre 3

## Analyse de la variance à deux facteurs

Dans tout ce chapitre, nous utilisons les notations définies à la section Notations, page ix.

### 3.1. Modèles à effets fixes

#### 3.1.1. 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  $(A_i, B_j)$  nous effectuons une mesure d'une réponse  $Y$  qui est une variable continue. Nous notons  $n = I \times J$  le nombre total de mesures ayant été effectuées.

Nous introduisons le modèle :

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

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

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

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

Nous supposons que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fait l'objet d'un paragraphe du chapitre ??.

Nous regroupons les valeurs que peut prendre la réponse  $Y$  dans les conditions  $(A_i, B_j)$  dans le tableau suivant :

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

Nous rappelons que la variation théorique due au facteur  $A$  est définie par :

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

Nous rappelons 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 :

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

Nous rappelons la relation fondamentale de l'ANOVA :

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

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

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

### 3.1 Modèles à effets fixes

---

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

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

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

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

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

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

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

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

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

$$s_{TOT} = s_{CA} + s_{CB} + s_{CR}.$$

Nous introduisons les degrés 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	$n_{TOT} = IJ - 1$

Nous résumons 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 $\mathcal{H}''_1$
Résiduelle	$sc_R$	$n_R$	$s_R^2 = \frac{sc_R}{n_R}$		
Totale	$sc_{TOT}$	$n_{TOT}$			

Nous souhaitons faire les tests d'hypothèse suivants :

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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, nous pouvons procéder à des comparaisons multiples des différents effets des niveaux du facteur voir le chapitre 9.

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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, nous pouvons procéder à des comparaisons multiples des différents effets des niveaux du facteur voir le chapitre 9.

### 3.1 Modèles à effets fixes

Les estimateurs  $\widehat{\mu}$ ,  $\widehat{\alpha}_1, \dots, \widehat{\alpha}_I, \widehat{\beta}_1, \dots, \widehat{\beta}_J, \widehat{\sigma}^2$  des paramètres  $\mu, \alpha_1, \dots, \alpha_I, \beta_1, \dots, \beta_J, \sigma^2$  du modèle sont donnés par les formules suivantes :

$$\widehat{\mu} = Y_{\bullet,\bullet} = \bar{Y}, \quad \widehat{\alpha}_i = Y_{i,\bullet} - \widehat{\mu}, \quad 1 \leq i \leq I, \quad \widehat{\beta}_j = Y_{\bullet,j} - \widehat{\mu}, \quad 1 \leq j \leq J,$$

$$\widehat{\sigma}^2 = \frac{SC_R}{(I-1)(J-1)} = s_R^2.$$

Ce sont des estimateurs sans biais<sup>1</sup>.

Les estimations, obtenues pour la liste de données  $\mathbf{y}$  et notées  $\widehat{\mu}(\mathbf{y}), \widehat{\alpha}_1(\mathbf{y}), \dots, \widehat{\alpha}_I(\mathbf{y}), \widehat{\beta}_1(\mathbf{y}), \dots, \widehat{\beta}_J(\mathbf{y}), \widehat{\sigma}^2(\mathbf{y})$ , des paramètres  $\mu, \alpha_1, \dots, \alpha_I, \beta_1, \dots, \beta_J, \sigma^2$  du modèle se déduisent des formules ci-dessus :

$$\widehat{\mu}(\mathbf{y}) = y_{\bullet,\bullet} = \bar{y}, \quad \widehat{\alpha}_i(\mathbf{y}) = y_{i,\bullet} - \widehat{\mu}(\mathbf{y}), \quad 1 \leq i \leq I, \quad \widehat{\beta}_j(\mathbf{y}) = y_{\bullet,j} - \widehat{\mu}(\mathbf{y}), \quad 1 \leq j \leq J,$$

$$\widehat{\sigma}^2(\mathbf{y}) = \frac{sc_R}{(I-1)(J-1)} = s_R^2.$$

#### 3.1.2. 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  $(A_i, B_j)$  nous effectuons  $K \geq 2$  mesures d'une réponse  $Y$  qui est une variable continue. Nous notons  $n = I \times J \times K$  le nombre total de mesures ayant été effectuées.

Nous introduisons le modèle :

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

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

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

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

$$\forall (i, j, k), 1 \leq i \leq I, \quad 1 \leq j \leq J, \quad 1 \leq k \leq K, \quad \mathcal{L}(\epsilon_{i,j,k}) = \mathcal{N}(0, \sigma^2),$$

$$\text{et } \epsilon_{i,j,k} \perp\!\!\!\perp \epsilon_{l,m,n} \text{ si } (i, j, k) \neq (l, m, n) \text{ avec}$$

$$1 \leq i, l \leq I, \quad 1 \leq j, m \leq J \text{ et } 1 \leq k, n \leq K.$$

Nous supposons que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fait l'objet d'un paragraphe du chapitre ??.

Nous regroupons les valeurs que peut prendre la réponse  $Y$  dans les conditions  $(A_i, B_j)$  lors des  $K$  répétitions dans le tableau suivant :

1. La notion de biais d'un estimateur est définie dans le chapitre ??.

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

Nous rappelons que la variation théorique due au facteur  $A$  est définie par :

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

Nous rappelons que la variation théorique due au facteur  $B$  est définie par :

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

Nous rappelons que la variation théorique due à l'interaction des facteurs  $A$  et  $B$  est définie par :

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

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 :

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

Nous rappelons la relation fondamentale de l'ANOVA :

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

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

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



### 3.1 Modèles à effets fixes

---

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

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

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

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

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

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

La variation résiduelle observée sur la liste de données  $\mathbf{y}$  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  $\mathbf{y}$  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  $\mathbf{y}$  :

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

Nous introduisons les degrés 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	$n_{TOT} = IJK - 1$

Nous résumons 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 $\mathcal{H}''_1$
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	$sc_{TOT}$	$n_{TOT}$			

Nous souhaitons faire les tests d'hypothèse suivants :

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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, nous pouvons procéder à des comparaisons multiples des différents effets des niveaux du facteur voir le chapitre 9.

$$\mathcal{H}''_0 : \beta_1 = \beta_2 = \dots = \beta_J = 0$$

contre

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

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. Nous concluons alors

## 3.2 Modèles à effets aléatoires

à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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, nous pouvons procéder à des comparaisons multiples des différents effets des niveaux du facteur voir le chapitre 9.

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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  $\hat{\mu}, \hat{\alpha}_1, \dots, \hat{\alpha}_I, \hat{\beta}_1, \dots, \hat{\beta}_J, \widehat{(\alpha\beta)}_{1,1}, \widehat{(\alpha\beta)}_{1,2}, \dots, \widehat{(\alpha\beta)}_{1,J}, \widehat{(\alpha\beta)}_{2,1}, \dots, \widehat{(\alpha\beta)}_{I,J}, \widehat{\sigma}^2$  des paramètres  $\mu, \alpha_1, \dots, \alpha_I, \beta_1, \dots, \beta_J, (\alpha\beta)_{1,1}, (\alpha\beta)_{1,2}, \dots, (\alpha\beta)_{1,J}, (\alpha\beta)_{2,1}, \dots, (\alpha\beta)_{I,J}, \sigma^2$  du modèle sont donnés par les formules suivantes :

$$\begin{aligned} \hat{\mu} &= Y_{\bullet,\bullet,\bullet} = \bar{Y}, \quad \hat{\alpha}_i = Y_{i,\bullet,\bullet} - \hat{\mu}, \quad 1 \leq i \leq I, \quad \hat{\beta}_j = Y_{\bullet,j,\bullet} - \hat{\mu}, \quad 1 \leq j \leq J, \\ \widehat{(\alpha\beta)}_{i,j} &= Y_{i,j,\bullet} - Y_{i,\bullet,\bullet} - Y_{\bullet,j,\bullet} + Y_{\bullet,\bullet,\bullet}, \quad 1 \leq i \leq I, \quad 1 \leq j \leq J, \\ \widehat{\sigma}^2 &= \frac{SC_R}{IJ(K-1)} = S_R^2. \end{aligned}$$

Ce sont des estimateurs sans biais<sup>2</sup>.

Les estimations, obtenues pour la liste de données  $\mathbf{y}$  et notées  $\hat{\mu}(\mathbf{y}), \hat{\alpha}_1(\mathbf{y}), \dots, \hat{\alpha}_I(\mathbf{y}), \hat{\beta}_1(\mathbf{y}), \dots, \hat{\beta}_J(\mathbf{y}), \widehat{(\alpha\beta)}_{1,1}(\mathbf{y}), \widehat{(\alpha\beta)}_{1,2}(\mathbf{y}), \dots, \widehat{(\alpha\beta)}_{1,J}(\mathbf{y}), \widehat{(\alpha\beta)}_{2,1}(\mathbf{y}), \dots, \widehat{(\alpha\beta)}_{I,J}(\mathbf{y}), \widehat{\sigma}^2(\mathbf{y})$ , des paramètres  $\mu, \alpha_1, \dots, \alpha_I, \beta_1, \dots, \beta_J, (\alpha\beta)_{1,1}, (\alpha\beta)_{1,2}, \dots, (\alpha\beta)_{1,J}, (\alpha\beta)_{2,1}, \dots, (\alpha\beta)_{I,J}, \sigma^2$  du modèle se déduisent des formules ci-dessus :

$$\begin{aligned} \hat{\mu}(\mathbf{y}) &= y_{\bullet,\bullet,\bullet} = \bar{y}, \quad \hat{\alpha}_i(\mathbf{y}) = y_{i,\bullet,\bullet} - \hat{\mu}(\mathbf{y}), \quad 1 \leq i \leq I, \quad \hat{\beta}_j(\mathbf{y}) = y_{\bullet,j,\bullet} - \hat{\mu}(\mathbf{y}), \quad 1 \leq j \leq J, \\ \widehat{(\alpha\beta)}_{i,j}(\mathbf{y}) &= y_{i,j,\bullet} - y_{i,\bullet,\bullet} - y_{\bullet,j,\bullet} + y_{\bullet,\bullet,\bullet}, \quad 1 \leq i \leq I, \quad 1 \leq j \leq J, \\ \widehat{\sigma}^2(\mathbf{y}) &= \frac{sc_R}{IJ(K-1)} = s_R^2. \end{aligned}$$

## 3.2. Modèles à effets aléatoires

### 3.2.1. 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  $\alpha_i$ , sont distribués suivant une loi

2. La notion de biais d'un estimateur est définie dans le chapitre ??.

normale centrée de variance  $\sigma_A^2$ . Les  $\beta_j$  représentent un échantillon de taille  $J$  prélevé dans une population importante. Nous admettrons que les effets des  $B_j$ , les  $\beta_j$ , sont distribués suivant une loi normale centrée de variance  $\sigma_B^2$ . Pour chacun des couples de modalités  $(A_i, B_j)$  nous effectuons une mesure d'une réponse  $Y$  qui est une variable continue. Nous notons  $n = I \times J$  le nombre total de mesures ayant été effectuées.

Nous introduisons le modèle :

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

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

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

ainsi que l'indépendance des effets aléatoires :

$$\begin{aligned} \alpha_i \perp\!\!\!\perp \alpha_j \text{ si } i \neq j \text{ et } 1 \leq i, j \leq I, \quad \beta_i \perp\!\!\!\perp \beta_j \text{ si } i \neq j \text{ et } 1 \leq i, j \leq J, \\ \alpha_i \perp\!\!\!\perp \beta_j \text{ si } 1 \leq i \leq I \text{ et } 1 \leq j \leq J. \end{aligned}$$

Nous postulons les hypothèses classiques suivantes pour les erreurs :

$$\begin{aligned} \forall (i, j), 1 \leq i \leq I, 1 \leq j \leq J, \quad \mathcal{L}(\epsilon_{i,j}) = \mathcal{N}(0, \sigma^2), \\ \epsilon_{i,j} \perp\!\!\!\perp \epsilon_{k,l} \text{ si } (i, j) \neq (k, l) \text{ avec } 1 \leq i, k \leq I \text{ et } 1 \leq j, l \leq J, \end{aligned}$$

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

$$\begin{aligned} \alpha_i \perp\!\!\!\perp \epsilon_{j,k} \text{ si } 1 \leq i, j \leq I \text{ et } 1 \leq k \leq J, \\ \beta_i \perp\!\!\!\perp \epsilon_{j,k} \text{ si } 1 \leq j \leq I \text{ et } 1 \leq i, k \leq J. \end{aligned}$$

Nous supposons que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fait l'objet d'un paragraphe du chapitre ??.

Nous utilisons les quantités  $SC_A, SC_B, SC_R, SC_{TOT}, sc_A, sc_B, sc_R$  et  $sc_{TOT}$  introduites à la section 3.1.1.

Nous rappelons la relation fondamentale de l'ANOVA :

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

Nous introduisons les degrés 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	$n_{TOT} = IJ - 1$

### 3.2 Modèles à effets aléatoires

---

Nous résumons 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 $\mathcal{H}''_1$
Résiduelle	$sc_R$	$n_R$	$s_R^2 = \frac{sc_R}{n_R}$		
Totale	$sc_{TOT}$	$n_{TOT}$			

Nous souhaitons 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  $\mu$ ,  $\sigma_A^2$ ,  $\sigma_B^2$ ,  $\sigma^2$  du modèle sont donnés par les formules suivantes :

$$\begin{aligned}\widehat{\mu} &= Y_{\bullet,\bullet} = \bar{Y}, \\ \widehat{\sigma}_A^2 &= \frac{1}{J} (S_A^2 - S_R^2), \quad \widehat{\sigma}_B^2 = \frac{1}{I} (S_B^2 - S_R^2), \\ \widehat{\sigma}^2 &= \frac{SC_R}{(I-1)(J-1)} = S_R^2,\end{aligned}$$

où  $S_A^2 = \frac{SC_A}{n_A}$ ,  $S_B^2 = \frac{SC_B}{n_B}$  et  $S_R^2 = \frac{SC_R}{n_R}$ .

Ce sont des estimateurs sans biais<sup>3</sup>.

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

$$\begin{aligned}\widehat{\mu}(\mathbf{y}) &= y_{\bullet,\bullet} = \bar{y}, \\ \widehat{\sigma}_A^2(\mathbf{y}) &= \frac{1}{J} (s_A^2 - s_R^2), \quad \widehat{\sigma}_B^2(\mathbf{y}) = \frac{1}{I} (s_B^2 - s_R^2), \\ \widehat{\sigma}^2(\mathbf{y}) &= \frac{sc_R}{(I-1)(J-1)} = s_R^2.\end{aligned}$$

### 3.2.2. Avec répétitions

Dans ce cas les  $A_i$  représentent un échantillon de taille  $I$  prélevé dans une population importante. Nous admettons que les effets des  $A_i$ , les  $\alpha_i$ , sont distribués suivant une loi normale centrée de variance  $\sigma_A^2$ . Les  $B_j$  représentent un échantillon de taille  $J$  prélevé dans une population importante. Nous admettons que les effets des  $B_j$ , les  $\beta_j$ , sont distribués suivant une loi normale centrée de variance  $\sigma_B^2$ . Pour chacun des couples de modalités  $(A_i, B_j)$  nous effectuons  $K \geq 2$  mesures d'une réponse  $Y$  qui est une variable continue. Nous notons  $n = I \times J \times K$  le nombre total de mesures ayant été effectuées.

Nous introduisons le modèle :

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

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. Nous supposons que

$$\begin{aligned}\mathcal{L}(\alpha_i) &= \mathcal{N}(0, \sigma_A^2), \quad \forall i, 1 \leq i \leq I, \\ \mathcal{L}(\beta_j) &= \mathcal{N}(0, \sigma_B^2), \quad \forall j, 1 \leq j \leq J, \\ \mathcal{L}((\alpha\beta)_{i,j}) &= \mathcal{N}(0, \sigma_{AB}^2), \quad \forall (i, j), 1 \leq i \leq I, 1 \leq j \leq J,\end{aligned}$$

ainsi que l'indépendance des effets aléatoires :

$$\begin{aligned}\alpha_i \perp\!\!\!\perp \alpha_j &\text{ si } i \neq j \text{ et } 1 \leq i, j \leq I, \quad \beta_i \perp\!\!\!\perp \beta_j \text{ si } i \neq j \text{ et } 1 \leq i, j \leq J, \\ (\alpha\beta)_{i,j} \perp\!\!\!\perp (\alpha\beta)_{k,l} &\text{ si } (i, j) \neq (k, l) \text{ avec } 1 \leq i, k \leq I \text{ et } 1 \leq j, l \leq J, \\ \alpha_i \perp\!\!\!\perp \beta_j &\text{ si } 1 \leq i \leq I \text{ et } 1 \leq j \leq J, \\ \alpha_i \perp\!\!\!\perp (\alpha\beta)_{j,k} &\text{ si } 1 \leq i, j \leq I \text{ et } 1 \leq k \leq J, \\ \beta_i \perp\!\!\!\perp (\alpha\beta)_{j,k} &\text{ si } 1 \leq j \leq I \text{ et } 1 \leq i, k \leq J.\end{aligned}$$

---

3. La notion de biais d'un estimateur est définie dans le chapitre ??.

### 3.2 Modèles à effets aléatoires

---

Nous postulons les hypothèses classiques suivantes pour les erreurs :

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

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

$$\begin{aligned} \alpha_i \perp\!\!\!\perp \epsilon_{j,k,l} \text{ si } 1 \leq i, j \leq I, 1 \leq k \leq J, \text{ et } 1 \leq l \leq K, \\ \beta_i \perp\!\!\!\perp \epsilon_{j,k,l} \text{ si } 1 \leq j \leq I, 1 \leq i, k \leq J, \text{ et } 1 \leq l \leq K, \\ (\alpha\beta)_{i,j} \perp\!\!\!\perp \epsilon_{k,l,m} \text{ si } 1 \leq i, k \leq I, 1 \leq j, l \leq J, \text{ et } 1 \leq m \leq K. \end{aligned}$$

Nous supposons que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fait l'objet d'un paragraphe du chapitre ??.

Nous utilisons 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 3.1.2.

Nous rappelons la relation fondamentale de l'ANOVA :

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

Nous introduisons les degrés 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	$n_{TOT} = IJK - 1$

Nous résumons 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	$sc_{TOT}$	$n_{TOT}$			

Nous souhaitons 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.

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

contre

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



### 3.3 Modèles à effets mixtes

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  $\mu$ ,  $\sigma_A^2$ ,  $\sigma_B^2$ ,  $\sigma_{AB}^2$ ,  $\sigma^2$  du modèle sont donnés par les formules suivantes :

$$\begin{aligned}\widehat{\mu} &= Y_{\bullet,\bullet,\bullet} = \bar{Y}, \\ \widehat{\sigma}_A^2 &= \frac{1}{JK} (S_A^2 - S_{AB}^2), \quad \widehat{\sigma}_B^2 = \frac{1}{IK} (S_B^2 - S_{AB}^2), \\ \widehat{\sigma}_{AB}^2 &= \frac{1}{K} (S_{AB}^2 - S_R^2), \\ \widehat{\sigma}^2 &= \frac{SC_R}{(I-1)(J-1)} = S_R^2,\end{aligned}$$

où  $S_A^2 = \frac{SC_A}{n_A}$ ,  $S_B^2 = \frac{SC_B}{n_B}$ ,  $S_{AB}^2 = \frac{SC_{AB}}{n_{AB}}$  et  $S_R^2 = \frac{SC_R}{n_R}$ .

Ce sont des estimateurs sans biais<sup>4</sup>.

Les estimations, obtenues pour la liste de données  $\mathbf{y}$  et notées  $\widehat{\mu}(\mathbf{y})$ ,  $\widehat{\sigma}_A^2(\mathbf{y})$ ,  $\widehat{\sigma}_B^2(\mathbf{y})$ ,  $\widehat{\sigma}_{AB}^2(\mathbf{y})$ ,  $\widehat{\sigma}^2(\mathbf{y})$ , des paramètres  $\mu$ ,  $\sigma_A^2$ ,  $\sigma_B^2$ ,  $\sigma_{AB}^2$ ,  $\sigma^2$  du modèle se déduisent des formules ci-dessus :

$$\begin{aligned}\widehat{\mu}(\mathbf{y}) &= y_{\bullet,\bullet,\bullet} = \bar{y}, \\ \widehat{\sigma}_A^2(\mathbf{y}) &= \frac{1}{JK} (s_A^2 - s_{AB}^2), \quad \widehat{\sigma}_B^2(\mathbf{y}) = \frac{1}{IK} (s_B^2 - s_{AB}^2), \\ \widehat{\sigma}_{AB}^2(\mathbf{y}) &= \frac{1}{K} (s_{AB}^2 - s_R^2), \\ \widehat{\sigma}^2(\mathbf{y}) &= \frac{s_{CR}}{(I-1)(J-1)} = s_R^2.\end{aligned}$$

## 3.3. Modèles à effets mixtes

### 3.3.1. Sans répétition

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

Nous introduisons le modèle :

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

avec les contraintes supplémentaires  $\sum_{i=1}^I \alpha_i = 0$ ,

4. La notion de biais d'un estimateur est définie dans le chapitre ??.

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

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

ainsi que l'indépendance des effets aléatoires :

$$\beta_i \perp\!\!\!\perp \beta_j \text{ si } i \neq j \text{ et } 1 \leq i, j \leq J.$$

Nous postulons les hypothèses classiques suivantes pour les erreurs :

$$\begin{aligned} & \forall (i, j), 1 \leq i \leq I, 1 \leq j \leq J, \mathcal{L}(\epsilon_{i,j}) = \mathcal{N}(0, \sigma^2), \\ & \epsilon_{i,j} \perp\!\!\!\perp \epsilon_{k,l} \text{ si } (i, j) \neq (k, l) \text{ avec } 1 \leq i, k \leq I \text{ et } 1 \leq j, l \leq J, \end{aligned}$$

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

$$\beta_i \perp\!\!\!\perp \epsilon_{j,k} \text{ si } 1 \leq j \leq I \text{ et } 1 \leq i, k \leq J.$$

Nous supposons que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fait l'objet d'un paragraphe du chapitre ??.

Nous utilisons les quantités  $SC_A, SC_B, SC_R, SC_{TOT}, sc_A, sc_B, sc_R$  et  $sc_{TOT}$  introduites à la section 3.1.1.

Nous rappelons la relation fondamentale de l'ANOVA :

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

Nous introduisons les degrés 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	$n_{TOT} = IJ - 1$

Nous résumons ces informations dans le tableau de l'ANOVA ci-dessous :

### 3.3 Modèles à effets mixtes

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 $\mathcal{H}''_1$
Résiduelle	$sc_R$	$n_R$	$s_R^2 = \frac{sc_R}{n_R}$		
Totale	$sc_{TOT}$	$n_{TOT}$			

Nous souhaitons faire les tests d'hypothèse suivants :

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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, nous pouvons procéder à des comparaisons multiples des différents effets des niveaux du facteur voir le chapitre 9.

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

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  $\hat{\mu}, \hat{\alpha}_1, \dots, \hat{\alpha}_I, \hat{\sigma}_B^2, \hat{\sigma}^2$  des paramètres  $\mu, \alpha_1, \dots, \alpha_I, \sigma_B^2, \sigma^2$  du modèle sont donnés par les formules suivantes :

$$\begin{aligned}\hat{\mu} &= Y_{\bullet,\bullet,\bullet} = \bar{Y}, \quad \hat{\alpha}_i = Y_{i,\bullet,\bullet} - \hat{\mu}, \quad 1 \leq i \leq I, \\ \hat{\sigma}_B^2 &= \frac{1}{I} (S_B^2 - S_R^2), \\ \hat{\sigma}^2 &= \frac{SC_R}{(I-1)(J-1)} = S_R^2,\end{aligned}$$

où  $S_B^2 = \frac{SC_B}{n_B}$  et  $S_R^2 = \frac{SC_R}{n_R}$ .

Ce sont des estimateurs sans biais<sup>5</sup>.

Les estimations, obtenues pour la liste de données  $\mathbf{y}$  et notées  $\hat{\mu}(\mathbf{y}), \hat{\alpha}_1(\mathbf{y}), \dots, \hat{\alpha}_I(\mathbf{y}), \hat{\sigma}_B^2(\mathbf{y}), \hat{\sigma}^2(\mathbf{y})$ , des paramètres  $\mu, \alpha_1, \dots, \alpha_I, \sigma_B^2, \sigma^2$  du modèle se déduisent des formules ci-dessus :

$$\begin{aligned}\hat{\mu} &= y_{\bullet,\bullet,\bullet} = \bar{y}, \quad \hat{\alpha}_i = y_{i,\bullet,\bullet} - \hat{\mu}, \quad 1 \leq i \leq I, \\ \hat{\sigma}_B^2 &= \frac{1}{I} (s_B^2 - s_R^2), \\ \hat{\sigma}^2 &= \frac{s_{CR}}{(I-1)(J-1)} = s_R^2.\end{aligned}$$

### 3.3.2. Avec répétitions

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

Nous introduisons le modèle, dit restreint :

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

avec les contraintes supplémentaires  $\sum_{i=1}^I \alpha_i = 0,$

$$\sum_{i=1}^I (\alpha\beta)_{i,j} = 0, \quad \forall j \in \{1, \dots, J\},$$

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. Nous supposons que

$$\begin{aligned}\mathcal{L}(\beta_j) &= \mathcal{N}(0, \sigma_B^2), \quad \forall j, 1 \leq j \leq J, \\ \mathcal{L}((\alpha\beta)_{i,j}) &= \mathcal{N}(0, \sigma_{AB}^2), \quad \forall (i, j), 1 \leq i \leq I, 1 \leq j \leq J,\end{aligned}$$

---

5. La notion de biais d'un estimateur est définie dans le chapitre ??.

### 3.3 Modèles à effets mixtes

---

ainsi que l'indépendance des effets aléatoires :

$$\begin{aligned} \beta_i &\perp\!\!\!\perp \beta_j \text{ si } i \neq j \text{ et } 1 \leq i, j \leq J, \\ \beta_i &\perp\!\!\!\perp (\alpha\beta)_{j,k} \text{ si } 1 \leq j \leq I \text{ et } 1 \leq i, k \leq J. \end{aligned}$$

Nous postulons les hypothèses classiques suivantes pour les erreurs :

$$\begin{aligned} \forall (i, j, k), 1 \leq i \leq I, 1 \leq j \leq J, 1 \leq k \leq K, \mathcal{L}(\epsilon_{i,j,k}) &= \mathcal{N}(0, \sigma^2), \\ \text{et } \epsilon_{i,j,k} &\perp\!\!\!\perp \epsilon_{l,m,n} \text{ si } (i, j, k) \neq (l, m, n) \text{ avec} \\ 1 \leq i, l \leq I, 1 \leq j, m \leq J \text{ et } 1 \leq k, n \leq K, \end{aligned}$$

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

$$\begin{aligned} \beta_i &\perp\!\!\!\perp \epsilon_{j,k,l} \text{ si } 1 \leq j \leq I, 1 \leq i, k \leq J, \text{ et } 1 \leq l \leq K, \\ (\alpha\beta)_{i,j} &\perp\!\!\!\perp \epsilon_{k,l,m} \text{ si } 1 \leq i, k \leq I, 1 \leq j, l \leq J, \text{ et } 1 \leq m \leq K. \end{aligned}$$

Dans un modèle mixte restreint, les effets aléatoires croisant des facteurs à effets fixes et à effets aléatoires, ici les  $(\alpha\beta)_{i,j}$ , ne sont pas mutuellement indépendants à cause des contraintes portant sur leur somme,  $\sum_{i=1}^I (\alpha\beta)_{i,j} = 0, \forall j \in \{1, \dots, J\}$ . Ils le sont par contre dès que nous ne les considérons pas tous en même temps.

Certains logiciels travaillent par défaut avec un modèle non-restreint dont la définition est proche mais pas identique à celle du modèle précédent.

Nous introduisons le modèle, dit non restreint :

$$\begin{aligned} Y_{i,j,k} &= \mu + \alpha_i + \beta_j + (\alpha\beta)_{i,j} + \epsilon_{i,j,k}, \quad i = 1 \dots I, j = 1 \dots J, k = 1 \dots K, \\ \text{avec la contrainte supplémentaire } &\sum_{i=1}^I \alpha_i = 0, \end{aligned}$$

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

$$\begin{aligned} \mathcal{L}(\beta_j) &= \mathcal{N}(0, \sigma_B^2), \quad \forall j, 1 \leq j \leq J, \\ \mathcal{L}((\alpha\beta)_{i,j}) &= \mathcal{N}(0, \sigma_{AB}^2), \quad \forall (i, j), 1 \leq i \leq I, 1 \leq j \leq J, \end{aligned}$$

ainsi que l'indépendance des effets aléatoires :

$$\begin{aligned} \beta_i &\perp\!\!\!\perp \beta_j \text{ si } i \neq j \text{ et } 1 \leq i, j \leq J, \\ (\alpha\beta)_{i,j} &\perp\!\!\!\perp (\alpha\beta)_{k,l} \text{ si } (i, j) \neq (k, l) \text{ avec } 1 \leq i, k \leq I \text{ et } 1 \leq j, l \leq J, \\ \beta_i &\perp\!\!\!\perp (\alpha\beta)_{j,k} \text{ si } 1 \leq j \leq I \text{ et } 1 \leq i, k \leq J. \end{aligned}$$

Nous postulons les hypothèses classiques suivantes pour les erreurs :

$$\begin{aligned} \forall (i, j, k), 1 \leq i \leq I, 1 \leq j \leq J, 1 \leq k \leq K, \mathcal{L}(\epsilon_{i,j,k}) &= \mathcal{N}(0, \sigma^2), \\ \text{et } \epsilon_{i,j,k} &\perp\!\!\!\perp \epsilon_{l,m,n} \text{ si } (i, j, k) \neq (l, m, n) \text{ avec} \\ 1 \leq i, l \leq I, 1 \leq j, m \leq J \text{ et } 1 \leq k, n \leq K, \end{aligned}$$

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

$$\begin{aligned} \beta_i &\perp\!\!\!\perp \epsilon_{j,k,l} \text{ si } 1 \leq j \leq I, 1 \leq i, k \leq J, \text{ et } 1 \leq l \leq K, \\ (\alpha\beta)_{i,j} &\perp\!\!\!\perp \epsilon_{k,l,m} \text{ si } 1 \leq i, k \leq I, 1 \leq j, l \leq J, \text{ et } 1 \leq m \leq K. \end{aligned}$$

Dans un modèle mixte non restreint, les effets aléatoires croisant des facteurs à effets fixes et à effets aléatoires, ici les  $(\alpha\beta)_{i,j}$ , sont mutuellement indépendants.

Il n'existe pas consensus sur une raison statistique qui permettrait de privilégier plutôt l'une ou l'autre de ces deux approches. Nous utiliserons toujours des modèles restreints.

Nous supposons que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fait l'objet d'un paragraphe du chapitre ??.

Nous utilisons 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 3.1.2.

Nous rappelons la relation fondamentale de l'ANOVA :

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

Nous introduisons les degrés 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	$n_{TOT} = IJK - 1$

Nous résumons ces informations dans le tableau de l'ANOVA ci-dessous :

### 3.3 Modèles à effets mixtes

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 $B$	$sc_B$	$n_B$	$s_B^2 = \frac{sc_B}{n_B}$	$f_B = \frac{s_B^2}{s_R^2}$	$\mathcal{H}''_0$ 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	$sc_{TOT}$	$n_{TOT}$			

Nous souhaitons faire les tests d'hypothèse suivants :

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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, nous pouvons procéder à des comparaisons multiples des différents effets des niveaux du facteur voir le chapitre 9.

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

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  $\hat{\mu}, \hat{\alpha}_1, \dots, \hat{\alpha}_I, \hat{\sigma}_B^2, \hat{\sigma}_{AB}^2, \hat{\sigma}^2$  des paramètres  $\mu, \alpha_1, \dots, \alpha_I, \sigma_B^2, \sigma_{AB}^2, \sigma^2$  du modèle sont donnés par les formules suivantes :

$$\begin{aligned} \hat{\mu} &= Y_{\bullet\bullet} = \bar{Y}, \quad \hat{\alpha}_i = Y_{i\bullet} - \hat{\mu}, \quad 1 \leq i \leq I, \\ \hat{\sigma}_B^2 &= \frac{1}{IK} (S_B^2 - S_R^2) \\ \hat{\sigma}_{AB}^2 &= \frac{1}{K} (S_{AB}^2 - S_R^2) \\ \hat{\sigma}^2 &= \frac{SC_R}{(I - 1)(J - 1)} = S_R^2, \end{aligned}$$

où  $S_B^2 = \frac{SC_B}{n_B}$ ,  $S_{AB}^2 = \frac{SC_{AB}}{n_{AB}}$  et  $S_R^2 = \frac{SC_R}{n_R}$ .

Ce sont des estimateurs sans biais<sup>6</sup>.

Les estimations, obtenues pour la liste de données  $\mathbf{y}$  et notées  $\hat{\mu}(\mathbf{y}), \hat{\alpha}_1(\mathbf{y}), \dots, \hat{\alpha}_I(\mathbf{y}), \hat{\sigma}_B^2(\mathbf{y}), \hat{\sigma}_{AB}^2(\mathbf{y}), \hat{\sigma}^2(\mathbf{y})$ , des paramètres  $\mu, \alpha_1, \dots, \alpha_I, \sigma_B^2, \sigma_{AB}^2, \sigma^2$  du modèle se déduisent des formules ci-dessus :

$$\begin{aligned} \hat{\mu}(\mathbf{y}) &= y_{\bullet\bullet} = \bar{y}, \quad \hat{\alpha}_i(\mathbf{y}) = y_{i\bullet} - \hat{\mu}(\mathbf{y}), \quad 1 \leq i \leq I, \\ \hat{\sigma}_B^2(\mathbf{y}) &= \frac{1}{IK} (s_B^2 - s_R^2) \\ \hat{\sigma}_{AB}^2(\mathbf{y}) &= \frac{1}{K} (s_{AB}^2 - s_R^2) \\ \hat{\sigma}^2(\mathbf{y}) &= \frac{SC_R}{(I - 1)(J - 1)} = s_R^2. \end{aligned}$$

---

6. La notion de biais d'un estimateur est définie dans le chapitre ??.



# Chapitre 4

## Analyse de la variance à deux facteurs emboîtés

Dans tout ce chapitre, nous utilisons les notations définies à la section Notations, page ix.

Nous sommes 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.

Nous ne pouvons nous servir d'un modèle où les facteurs sont emboîtés<sup>1</sup>, que si nous disposons 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 nous devrons utiliser pour analyser les données sera l'un de ceux exposés au chapitre 2.

Ainsi par exemple un fabricant de détergents alimente plusieurs chaînes de distribution :  $A_1, A_2, \dots, A_I$ . Nous pensons que les boîtes de produit livrées à certaines chaînes de distribution contiennent une masse de détergent inférieure à celle des autres chaînes de distribution. Pour étudier cette situation, nous décidons 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  $A$  nous notons les niveaux du second facteur  $B : B_{j(i)}, 1 \leq i \leq I$  et  $1 \leq j \leq J$ .

---

1. Ces types de modèles sont également appelés des modèles hiérarchiques ou en anglais *hierarchical* ou *nested models*.

## 4.1. Modèles à effets fixes

### 4.1.1. 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  $(A_i, B_{j(i)})$  nous effectuons  $K \geq 2$  mesures d'une réponse  $Y$  qui est une variable continue. Nous notons  $n = I \times J \times K$  le nombre total de mesures ayant été effectuées.

Nous introduisons le modèle :

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

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. Nous postulons les hypothèses classiques suivantes pour les erreurs :

$$\forall (i, j, k), 1 \leq i \leq I, 1 \leq j \leq J, 1 \leq k \leq K, \mathcal{L}(\epsilon_{i,j,k}) = \mathcal{N}(0, \sigma^2),$$

et  $\epsilon_{i,j,k} \perp \epsilon_{l,m,n}$  si  $(i, j, k) \neq (l, m, n)$  avec  
 $1 \leq i, l \leq I, 1 \leq j, m \leq J$  et  $1 \leq k, n \leq K$ .

Nous supposons que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fait l'objet d'un paragraphe du chapitre ??.

Nous regroupons les valeurs que peut prendre la réponse  $Y$  dans les conditions  $(A_i, B_{j(i)})$  lors des  $K$  essais dans le tableau suivant :

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

Nous rappelons 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 :

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

## 4.1 Modèles à effets fixes

---

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 :

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

Nous rappelons la relation fondamentale de l'ANOVA :

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

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

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

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

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

La variation résiduelle observée sur la liste de données  $\mathbf{y}$  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  $\mathbf{y}$  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  $\mathbf{y}$  :

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

Nous reconnaissons parmi les quantités définies ci-dessus, des quantités similaires à celles,  $SC_A, SC_B, SC_{AB}, SC_R, SC_{TOT}, sc_A, sc_B, sc_{AB}, sc_R$  et  $sc_{TOT}$  introduites à la section 3.1.2. Nous remarquons également que les nouvelles quantités,  $SC_{B|A}$  et  $sc_{B|A}$ , sont liées aux précédentes par les relations :

$$SC_{B|A} = SC_B + SC_{AB},$$

$$sc_{B|A} = sc_B + sc_{AB}.$$

Nous introduisons les degrés 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	$n_{TOT} = IJK - 1$

Nous résumons 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 $B$ dans $A$	$sc_{B A}$	$n_{B A}$	$s_{B A}^2 = \frac{sc_{B A}}{n_{B A}}$	$f_{B A} = \frac{s_{B A}^2}{s_R^2}$	$\mathcal{H}''_0$ ou $\mathcal{H}''_1$
Résiduelle	$sc_R$	$n_R$	$s_R^2 = \frac{sc_R}{n_R}$		
Totale	$sc_{TOT}$	$n_{TOT}$			

Nous souhaitons faire les tests d'hypothèse suivants :

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

contre

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

## 4.2 Modèles à effets aléatoires

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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, nous pouvons procéder à des comparaisons multiples des différents effets des niveaux du facteur voir le chapitre 9.

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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  $\hat{\mu}, \hat{\alpha}_1, \dots, \hat{\alpha}_I, \hat{\beta}_{1(1)}, \hat{\beta}_{2(1)}, \dots, \hat{\beta}_{J(1)}, \hat{\beta}_{1(2)}, \dots, \hat{\beta}_{J(I)}, \hat{\sigma}^2$  des paramètres  $\mu, \alpha_1, \dots, \alpha_I, \beta_{1(1)}, \beta_{2(1)}, \dots, \beta_{J(1)}, \beta_{1(2)}, \dots, \beta_{J(I)}, \sigma^2$  du modèle sont donnés par les formules suivantes :

$$\begin{aligned} \hat{\mu} &= Y_{\bullet,\bullet,\bullet} = \bar{Y}, \quad \hat{\alpha}_i = Y_{i,\bullet,\bullet} - \hat{\mu}, \quad 1 \leq i \leq I, \\ \hat{\beta}_{j(i)} &= Y_{i,j,\bullet} - Y_{i,\bullet,\bullet}, \quad 1 \leq i \leq I, \quad 1 \leq j \leq J, \\ \hat{\sigma}^2 &= \frac{SC_R}{IJ(K - 1)} = S_R^2. \end{aligned}$$

Ce sont des estimateurs sans biais<sup>2</sup>.

Les estimations, obtenues pour la liste de données  $\mathbf{y}$  et notées  $\hat{\mu}(\mathbf{y}), \hat{\alpha}_1(\mathbf{y}), \dots, \hat{\alpha}_I(\mathbf{y}), \hat{\beta}_{1(1)}(\mathbf{y}), \hat{\beta}_{2(1)}(\mathbf{y}), \dots, \hat{\beta}_{J(1)}(\mathbf{y}), \hat{\beta}_{1(2)}(\mathbf{y}), \dots, \hat{\beta}_{J(I)}(\mathbf{y}), \hat{\sigma}^2(\mathbf{y})$ , des paramètres  $\mu, \alpha_1, \dots, \alpha_I, \beta_{1(1)}, \beta_{2(1)}, \dots, \beta_{J(1)}, \beta_{1(2)}, \dots, \beta_{J(I)}, \sigma^2$  du modèle se déduisent des formules suivantes :

$$\begin{aligned} \hat{\mu}(\mathbf{y}) &= y_{\bullet,\bullet,\bullet} = \bar{y}, \quad \hat{\alpha}_i(\mathbf{y}) = y_{i,\bullet,\bullet} - \hat{\mu}(\mathbf{y}), \quad 1 \leq i \leq I, \\ \hat{\beta}_{j(i)}(\mathbf{y}) &= y_{i,j,\bullet} - y_{i,\bullet,\bullet}, \quad 1 \leq i \leq I, \quad 1 \leq j \leq J, \\ \hat{\sigma}^2(\mathbf{y}) &= \frac{SC_R}{IJ(K - 1)} = s_R^2. \end{aligned}$$

## 4.2. Modèles à effets aléatoires

### 4.2.1. 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

2. La notion de biais d'un estimateur est définie dans le chapitre ??.

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  $\sigma_{B|A}^2$ . Pour chacun des couples de modalités  $(A_i, B_{j(i)})$  nous effectuons  $K \geq 2$  mesures d'une réponse  $Y$  qui est une variable continue. Nous notons  $n = I \times J \times K$  le nombre total de mesures ayant été effectuées.

Nous introduisons le modèle :

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

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. Nous supposons que

$$\begin{aligned} \mathcal{L}(\alpha_i) &= \mathcal{N}(0, \sigma_A^2), \quad \forall i, 1 \leq i \leq I, \\ \mathcal{L}(\beta_{j(i)}) &= \mathcal{N}(0, \sigma_{B|A}^2), \quad \forall (i, j), 1 \leq i \leq I, 1 \leq j \leq J, \end{aligned}$$

ainsi que l'indépendance des effets aléatoires :

$$\begin{aligned} \alpha_i &\perp\!\!\!\perp \alpha_j \text{ si } i \neq j \text{ et } 1 \leq i, j \leq I, \\ \beta_{j(i)} &\perp\!\!\!\perp \beta_{l(k)} \text{ si } (i, j) \neq (k, l) \text{ avec } 1 \leq i, k \leq I \text{ et } 1 \leq j, l \leq J, \\ \alpha_i &\perp\!\!\!\perp \beta_{k(j)} \text{ si } 1 \leq i, j \leq I \text{ et } 1 \leq k \leq J. \end{aligned}$$

Nous postulons les hypothèses classiques suivantes pour les erreurs :

$$\begin{aligned} \forall (i, j, k), 1 \leq i \leq I, 1 \leq j \leq J, 1 \leq k \leq K, \quad \mathcal{L}(\epsilon_{i,j,k}) &= \mathcal{N}(0, \sigma^2), \\ \text{et } \epsilon_{i,j,k} &\perp\!\!\!\perp \epsilon_{l,m,n} \text{ si } (i, j, k) \neq (l, m, n) \text{ avec} \\ 1 \leq i, l \leq I, 1 \leq j, m \leq J \text{ et } 1 \leq k, n \leq K, \end{aligned}$$

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

$$\begin{aligned} \alpha_i &\perp\!\!\!\perp \epsilon_{j,k,l} \text{ si } 1 \leq i, j \leq I, 1 \leq k \leq J, \text{ et } 1 \leq l \leq K, \\ \beta_{j(i)} &\perp\!\!\!\perp \epsilon_{k,l,m} \text{ si } 1 \leq i, k \leq I, 1 \leq j, l \leq J, \text{ et } 1 \leq m \leq K. \end{aligned}$$

Nous supposons que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fait l'objet d'un paragraphe du chapitre ??.

Nous utilisons 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 4.1.1.

Nous rappelons la relation fondamentale de l'ANOVA :

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

Nous introduisons les degrés 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	$n_{TOT} = IJK - 1$

## 4.2 Modèles à effets aléatoires

---

Nous résumons 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 $A$	$sc_{B A}$	$n_{B A}$	$s_{B A}^2 = \frac{sc_{B A}}{n_{B A}}$	$f_{B A} = \frac{s_{B A}^2}{s_R^2}$	$\mathcal{H}''_0$ ou $\mathcal{H}''_1$
Résiduelle	$sc_R$	$n_R$	$s_R^2 = \frac{sc_R}{n_R}$		
Totale	$sc_{TOT}$	$n_{TOT}$			

Nous souhaitons 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

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

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

$$\begin{aligned}\widehat{\mu} &= Y_{\bullet,\bullet,\bullet} = \bar{Y}, \quad \widehat{\sigma}_A^2 = \frac{1}{JK} (S_A^2 - S_{B|A}^2), \\ \widehat{\sigma}_{B|A}^2 &= \frac{1}{K} (S_{B|A}^2 - S_R^2), \\ \widehat{\sigma}^2 &= \frac{SC_R}{(I-1)(J-1)} = S_R^2,\end{aligned}$$

où  $S_A^2 = \frac{SC_A}{n_A}$ ,  $S_{B|A}^2 = \frac{SC_{B|A}}{n_{B|A}}$  et  $S_R^2 = \frac{SC_R}{n_R}$ .

Ce sont des estimateurs sans biais<sup>3</sup>.

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

$$\begin{aligned}\widehat{\mu}(\mathbf{y}) &= y_{\bullet,\bullet,\bullet} = \bar{y}, \quad \widehat{\sigma}_A^2(\mathbf{y}) = \frac{1}{JK} (s_A^2 - s_{B|A}^2), \\ \widehat{\sigma}_{B|A}^2(\mathbf{y}) &= \frac{1}{K} (s_{B|A}^2 - s_R^2), \\ \widehat{\sigma}^2(\mathbf{y}) &= \frac{sc_R}{(I-1)(J-1)} = s_R^2.\end{aligned}$$

### 4.3. Modèles à effets mixtes

Pour le plupart des auteurs d'ouvrages sur l'analyse de la variance, voir [Dag98b] à ce sujet par exemple, un facteur emboîté dans un facteur aléatoire doit être considéré comme aléatoire<sup>4</sup>. Ainsi le seul modèle mixte possible est le cas où le facteur  $A$  est fixe et le facteur que nous emboîtons dans  $A$ , le facteur  $B$ , est aléatoire.

#### 4.3.1. Avec répétitions

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

Nous introduisons le modèle :

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

avec les contraintes supplémentaires  $\sum_{i=1}^I \alpha_i = 0$ ,

---

3. La notion de biais d'un estimateur est définie dans le chapitre ??.

4. Il existe néanmoins certains cas où nous pouvons 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.



### 4.3 Modèles à effets mixtes

---

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. Nous supposons que

$$\mathcal{L}(\beta_{j(i)}) = \mathcal{N}(0, \sigma_{B|A}^2), \quad \forall (i, j), 1 \leq i \leq I, 1 \leq j \leq J,$$

ainsi que l'indépendance des effets aléatoires :

$$\beta_{j(i)} \perp\!\!\!\perp \beta_{l(k)} \text{ si } (i, j) \neq (k, l) \text{ avec } 1 \leq i, k \leq I \text{ et } 1 \leq j, l \leq J.$$

Nous postulons les hypothèses classiques suivantes pour les erreurs :

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

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

$$\beta_{j(i)} \perp\!\!\!\perp \epsilon_{k,l,m} \text{ si } 1 \leq i, k \leq I, 1 \leq j, l \leq J, \text{ et } 1 \leq m \leq K.$$

Nous supposons que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fait l'objet d'un paragraphe du chapitre ??.

Nous utilisons 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 4.1.1.

Nous rappelons la relation fondamentale de l'ANOVA :

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

Nous introduisons les degrés 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	$n_{TOT} = IJK - 1$

Nous résumons 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 $A$	$sc_{B A}$	$n_{B A}$	$s_{B A}^2 = \frac{sc_{B A}}{n_{B A}}$	$f_{B A} = \frac{s_{B A}^2}{s_R^2}$	$\mathcal{H}''_0$ ou $\mathcal{H}''_1$
Résiduelle	$sc_R$	$n_R$	$s_R^2 = \frac{sc_R}{n_R}$		
Totale	$sc_{TOT}$	$n_{TOT}$			

Nous souhaitons faire les tests d'hypothèse suivants :

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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, nous pouvons procéder à des comparaisons multiples des différents effets des niveaux du facteur voir le chapitre 9.

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

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.

### 4.3 Modèles à effets mixtes

---

Les estimateurs  $\widehat{\mu}, \widehat{\alpha}_1, \dots, \widehat{\alpha}_I, \widehat{\sigma}_{B|A}^2, \widehat{\sigma}^2$  des paramètres  $\mu, \alpha_1, \dots, \alpha_I, \sigma_{B|A}^2, \sigma^2$  du modèle sont donnés par les formules suivantes :

$$\begin{aligned}\widehat{\mu} &= Y_{\bullet,\bullet,\bullet} = \bar{Y}, \quad \widehat{\alpha}_i = Y_{i,\bullet,\bullet} - \widehat{\mu}, \quad 1 \leq i \leq I, \\ \widehat{\sigma}_{B|A}^2 &= \frac{1}{K} (S_{B|A}^2 - S_R^2), \\ \widehat{\sigma}^2 &= \frac{SC_R}{(I-1)(J-1)} = S_R^2,\end{aligned}$$

où  $S_{B|A}^2 = \frac{SC_{B|A}}{n_{B|A}}$  et  $S_R^2 = \frac{SC_R}{n_R}$ .

Ce sont des estimateurs sans biais<sup>5</sup>.

Les estimations, obtenues pour la liste de données  $\mathbf{y}$  et notées  $\widehat{\mu}(\mathbf{y}), \widehat{\alpha}_1(\mathbf{y}), \dots, \widehat{\alpha}_I(\mathbf{y}), \widehat{\sigma}_{B|A}^2(\mathbf{y}), \widehat{\sigma}^2(\mathbf{y})$ , des paramètres  $\mu, \alpha_1, \dots, \alpha_I, \sigma_{B|A}^2, \sigma^2$  du modèle se déduisent des formules ci-dessus :

$$\begin{aligned}\widehat{\mu}(\mathbf{y}) &= y_{\bullet,\bullet,\bullet} = \bar{y}, \quad \widehat{\alpha}_i(\mathbf{y}) = y_{i,\bullet,\bullet} - \widehat{\mu}(\mathbf{y}), \quad 1 \leq i \leq I, \\ \widehat{\sigma}_{B|A}^2(\mathbf{y}) &= \frac{1}{K} (s_{B|A}^2 - s_R^2), \\ \widehat{\sigma}^2(\mathbf{y}) &= \frac{sc_R}{(I-1)(J-1)} = s_R^2.\end{aligned}$$

---

5. La notion de biais d'un estimateur est définie dans le chapitre ??.



# Chapitre 5

## Analyse de la variance à trois facteurs

Dans tout ce chapitre, nous utilisons les notations définies à la section Notations, page ix.

### 5.1. Modèles à effets fixes

#### 5.1.1. 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  $(A_i, B_j, C_k)$  nous effectuons une mesure d'une réponse  $Y$  qui est une variable continue. Nous notons  $n = I \times J \times K$  le nombre total de mesures ayant été effectuées.

Nous introduisons le modèle :

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

avec les contraintes supplémentaires  $\sum_{i=1}^I \alpha_i = 0$ ,  $\sum_{j=1}^J \beta_j = 0$  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  $(A_i, B_j, C_k)$ . Nous postulons les hypothèses classiques suivantes pour les erreurs :

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

Nous supposons que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fait l'objet d'un paragraphe du chapitre ??.

Nous regroupons les valeurs que peut prendre la réponse  $Y$  dans les conditions  $(A_i, B_j, C_k)$  dans le tableau suivant :

Facteur $A$	Facteur $B$	Facteur $C$	Réponse $Y$
$A_1$	$B_1$	$C_1$	$Y_{1,1,1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$A_I$	$B_1$	$C_1$	$Y_{I,1,1}$
$A_1$	$B_2$	$C_1$	$Y_{1,2,1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$A_I$	$B_J$	$C_1$	$Y_{I,J,1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$A_1$	$B_1$	$C_2$	$Y_{1,1,2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$A_I$	$B_J$	$C_K$	$Y_{I,J,K}$

Nous rappelons que la variation théorique due au facteur  $A$  est définie par :

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

Nous rappelons que la variation théorique due au facteur  $B$  est définie par :

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

Nous rappelons que la variation théorique due au facteur  $C$  est définie par :

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

Nous rappelons que la variation théorique due à l'interaction des facteurs  $A$  et  $B$  est définie par :

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

Nous rappelons que la variation théorique due à l'interaction des facteurs  $B$  et  $C$  est définie par :

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

## 5.1 Modèles à effets fixes

---

Nous rappelons que la variation théorique 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 théorique est quant à elle définie par :

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

Enfin la variation totale théorique est égale à :

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

Nous rappelons la relation fondamentale de l'ANOVA :

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

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

Facteur $A$	Facteur $B$	Facteur $C$	Réponse $Y$
$A_1$	$B_1$	$C_1$	$y_{1,1,1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$A_I$	$B_1$	$C_1$	$y_{I,1,1}$
$A_1$	$B_2$	$C_1$	$y_{1,2,1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$A_I$	$B_J$	$C_1$	$y_{I,J,1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$A_1$	$B_1$	$C_2$	$y_{1,1,2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$A_I$	$B_J$	$C_K$	$y_{I,J,K}$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$s_{TOT} = s_{CA} + s_{CB} + s_{CC} + s_{AB} + s_{AC} + s_{BC} + s_{R}.$$

Nous introduisons les degrés 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	$n_{TOT} = IJK - 1$



## 5.1 Modèles à effets fixes

Nous résumons 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	$sc_{TOT}$	$n_{TOT}$			

Nous souhaitons faire les tests d'hypothèse suivants :

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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, nous pouvons procéder à des comparaisons multiples des différents effets des niveaux du facteur voir le chapitre 9.

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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, nous pouvons procéder à des comparaisons multiples des différents effets des niveaux du facteur voir le chapitre 9.

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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, nous pouvons procéder à des comparaisons multiples des différents effets des niveaux du facteur voir le chapitre 9.

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

contre

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

Sous l'hypothèse nulle  $\mathcal{H}^{(4)}_0$  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  $(I - 1)(J - 1)(K - 1)$  degrés de liberté. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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.

## 5.1 Modèles à effets fixes

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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.

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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 estimateurs  $\widehat{\mu}$ ,  $\widehat{\alpha}_1, \dots, \widehat{\alpha}_I$ ,  $\widehat{\beta}_1, \dots, \widehat{\beta}_J$ ,  $\widehat{\gamma}_1, \dots, \widehat{\gamma}_K$ ,  $\widehat{(\alpha\beta)}_{1,1}, \dots, \widehat{(\alpha\beta)}_{I,J}$ ,  $\widehat{(\alpha\gamma)}_{1,1}, \dots, \widehat{(\alpha\gamma)}_{I,K}$ ,  $\widehat{(\beta\gamma)}_{1,1}, \dots, \widehat{(\beta\gamma)}_{J,K}$ ,  $\widehat{\sigma}^2$  des paramètres  $\mu, \alpha_1, \dots, \alpha_I, \beta_1, \dots, \beta_J, \gamma_1, \dots, \gamma_K, (\alpha\beta)_{1,1}, \dots, (\alpha\beta)_{I,J}, (\alpha\gamma)_{1,1}, \dots, (\alpha\gamma)_{I,K}, (\beta\gamma)_{1,1}, \dots, (\beta\gamma)_{J,K}, \sigma^2$  du modèle sont donnés par les formules suivantes :

$$\begin{aligned} \widehat{\mu} &= Y_{\bullet,\bullet,\bullet} = \bar{Y}, \\ \widehat{\alpha}_i &= Y_{i,\bullet,\bullet} - \widehat{\mu}, \quad 1 \leq i \leq I, \\ \widehat{\beta}_j &= Y_{\bullet,j,\bullet} - \widehat{\mu}, \quad 1 \leq j \leq J, \\ \widehat{\gamma}_k &= Y_{\bullet,\bullet,k} - \widehat{\mu}, \quad 1 \leq k \leq K, \\ \widehat{(\alpha\beta)}_{i,j} &= Y_{i,j,\bullet} - Y_{i,\bullet,\bullet} - Y_{\bullet,j,\bullet} + \widehat{\mu}, \quad 1 \leq i \leq I, \quad 1 \leq j \leq J, \\ \widehat{(\alpha\gamma)}_{i,k} &= Y_{i,\bullet,k} - Y_{i,\bullet,\bullet} - Y_{\bullet,\bullet,k} + \widehat{\mu}, \quad 1 \leq i \leq I, \quad 1 \leq k \leq K, \\ \widehat{(\beta\gamma)}_{j,k} &= Y_{\bullet,j,k} - Y_{\bullet,j,\bullet} - Y_{\bullet,\bullet,k} + \widehat{\mu}, \quad 1 \leq j \leq J, \quad 1 \leq k \leq K, \\ \widehat{\sigma}^2 &= \frac{SC_R}{(I-1)(J-1)(K-1)} = S_R^2. \end{aligned}$$

Ce sont des estimateurs sans biais<sup>1</sup>.

1. La notion de biais d'un estimateur est définie dans le chapitre ??.

Les estimations, obtenues pour la liste de données  $\mathbf{y}$  et notées  $\widehat{\mu}(\mathbf{y})$ ,  $\widehat{\alpha}_1(\mathbf{y})$ ,  $\dots$ ,  $\widehat{\alpha}_I(\mathbf{y})$ ,  $\widehat{\beta}_1(\mathbf{y})$ ,  $\dots$ ,  $\widehat{\beta}_J(\mathbf{y})$ ,  $\widehat{\gamma}_1(\mathbf{y})$ ,  $\dots$ ,  $\widehat{\gamma}_K(\mathbf{y})$ ,  $\widehat{(\alpha\beta)}_{1,1}(\mathbf{y})$ ,  $\dots$ ,  $\widehat{(\alpha\beta)}_{I,J}(\mathbf{y})$ ,  $\widehat{(\alpha\gamma)}_{1,1}(\mathbf{y})$ ,  $\dots$ ,  $\widehat{(\alpha\gamma)}_{I,K}(\mathbf{y})$ ,  $\widehat{(\beta\gamma)}_{1,1}(\mathbf{y})$ ,  $\dots$ ,  $\widehat{(\beta\gamma)}_{J,K}(\mathbf{y})$ ,  $\widehat{\sigma^2}(\mathbf{y})$  des paramètres  $\mu$ ,  $\alpha_1, \dots, \alpha_I$ ,  $\beta_1, \dots, \beta_J$ ,  $\gamma_1, \dots, \gamma_K$ ,  $(\alpha\beta)_{1,1}, \dots, (\alpha\beta)_{I,J}$ ,  $(\alpha\gamma)_{1,1}, \dots, (\alpha\gamma)_{I,K}$ ,  $(\beta\gamma)_{1,1}, \dots, (\beta\gamma)_{J,K}$ ,  $\sigma^2$  du modèle se déduisent des formules ci-dessus :

$$\begin{aligned}\widehat{\mu}(\mathbf{y}) &= y_{\bullet,\bullet,\bullet} = \bar{y}, \\ \widehat{\alpha}_i(\mathbf{y}) &= y_{i,\bullet,\bullet} - \widehat{\mu}(\mathbf{y}), \quad 1 \leq i \leq I, \\ \widehat{\beta}_j(\mathbf{y}) &= y_{\bullet,j,\bullet} - \widehat{\mu}(\mathbf{y}), \quad 1 \leq j \leq J, \\ \widehat{\gamma}_k(\mathbf{y}) &= y_{\bullet,\bullet,k} - \widehat{\mu}(\mathbf{y}), \quad 1 \leq k \leq K, \\ \widehat{(\alpha\beta)}_{i,j}(\mathbf{y}) &= y_{i,j,\bullet} - y_{i,\bullet,\bullet} - y_{\bullet,j,\bullet} + \widehat{\mu}(\mathbf{y}), \quad 1 \leq i \leq I, \quad 1 \leq j \leq J, \\ \widehat{(\alpha\gamma)}_{i,k}(\mathbf{y}) &= y_{i,\bullet,k} - y_{i,\bullet,\bullet} - y_{\bullet,\bullet,k} + \widehat{\mu}(\mathbf{y}), \quad 1 \leq i \leq I, \quad 1 \leq k \leq K, \\ \widehat{(\beta\gamma)}_{j,k}(\mathbf{y}) &= y_{\bullet,j,k} - y_{\bullet,j,\bullet} - y_{\bullet,\bullet,k} + \widehat{\mu}(\mathbf{y}), \quad 1 \leq j \leq J, \quad 1 \leq k \leq K, \\ \widehat{\sigma^2}(\mathbf{y}) &= \frac{SC_R}{(I-1)(J-1)(K-1)} = S_R^2.\end{aligned}$$

### 5.1.2. 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  $(A_i, B_j, C_k)$  nous effectuons  $L \geq 2$  mesures d'une réponse  $Y$  qui est une variable continue. Nous notons  $n = I \times J \times K \times L$  le nombre total de mesures ayant été effectuées.

Nous introduisons le modèle :

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

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

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

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

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

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

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

## 5.1 Modèles à effets fixes

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

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. Nous postulons les hypothèses classiques suivantes pour les erreurs :

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

Nous supposons que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fait l'objet d'un paragraphe du chapitre ??.

Nous regroupons les valeurs que peut prendre la réponse  $Y$  dans les conditions  $(A_i, B_j, C_k)$  lors des  $L$  essais dans le tableau suivant :

Facteur $A$	Facteur $B$	Facteur $C$	Essai n°	Réponse $Y$
$A_1$	$B_1$	$C_1$	1	$Y_{1,1,1,1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$A_1$	$B_1$	$C_1$	$L$	$Y_{1,1,1,L}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$A_I$	$B_1$	$C_1$	1	$Y_{I,1,1,1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$A_I$	$B_1$	$C_1$	$L$	$Y_{I,1,1,L}$
$A_1$	$B_2$	$C_1$	1	$Y_{1,2,1,1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$A_I$	$B_J$	$C_1$	$L$	$Y_{I,J,1,L}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$A_1$	$B_1$	$C_2$	1	$Y_{1,1,2,1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$A_I$	$B_J$	$C_K$	$L$	$Y_{I,J,K,L}$

Nous rappelons que la variation théorique due au facteur  $A$  est définie par :

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

Nous rappelons que la variation théorique due au facteur  $B$  est définie par :

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

Nous rappelons que la variation théorique due au facteur  $C$  est définie par :

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

Nous rappelons que la variation théorique due à l'interaction des facteurs  $A$  et  $B$  est définie par :

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

Nous rappelons que la variation théorique due à l'interaction des facteurs  $B$  et  $C$  est définie par :

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

Nous rappelons que la variation théorique 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 théorique due à l'interaction d'ordre 2 entre les facteurs  $A$ ,  $B$  et  $C$  est définie par :

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

La variation résiduelle théorique est quant à elle définie par :

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

Enfin la variation totale théorique est égale à :

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

Nous rappelons la relation fondamentale de l'ANOVA :

$$SC_{TOT} = SC_A + SC_B + SC_C + SC_{AB} + SC_{AC} + SC_{BC} + SC_{ABC} + SC_R.$$

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

## 5.1 Modèles à effets fixes

Facteur $A$	Facteur $B$	Facteur $C$	Essai n°	Donnée expérimentale $y$
$A_1$	$B_1$	$C_1$	1	$y_{1,1,1,1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$A_1$	$B_1$	$C_1$	$L$	$y_{1,1,1,L}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$A_I$	$B_1$	$C_1$	1	$y_{I,1,1,1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$A_I$	$B_1$	$C_1$	$L$	$y_{I,1,1,L}$
$A_1$	$B_2$	$C_1$	1	$y_{1,2,1,1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$A_I$	$B_J$	$C_1$	$L$	$y_{I,J,1,L}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$A_1$	$B_1$	$C_2$	1	$y_{1,1,2,1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$A_I$	$B_J$	$C_K$	$L$	$y_{I,J,K,L}$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$s_{CTOT} = s_{CA} + s_{CB} + s_{CC} + s_{CAB} + s_{CAC} + s_{CBC} + s_{CABC} + s_{CR}.$$

Nous introduisons les degrés 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	$n_{TOT} = IJKL - 1$

Nous résumons ces informations dans le tableau de l'ANOVA ci-dessous :



## 5.1 Modèles à effets fixes

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	$sc_{TOT}$	$n_{TOT}$			

Nous souhaitons faire les tests d'hypothèse suivants :

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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, nous pouvons procéder à des comparaisons multiples des différents effets des niveaux du facteur voir le chapitre 9.

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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, nous pouvons procéder à des comparaisons multiples des différents effets des niveaux du facteur voir le chapitre 9.

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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, nous pouvons procéder à des comparaisons multiples des différents effets des niveaux du facteur voir le chapitre 9.

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

contre

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

Sous l'hypothèse nulle  $\mathcal{H}^{(4)}_0$  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  $IJK(L - 1)$  degrés de liberté. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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.

## 5.1 Modèles à effets fixes

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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.

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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.

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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 estimateurs  $\widehat{\mu}$ ,  $\widehat{\alpha}_1, \dots, \widehat{\alpha}_I$ ,  $\widehat{\beta}_1, \dots, \widehat{\beta}_J$ ,  $\widehat{\gamma}_1, \dots, \widehat{\gamma}_K$ ,  $\widehat{(\alpha\beta)}_{1,1}, \dots, \widehat{(\alpha\beta)}_{I,J}$ ,  $\widehat{(\alpha\gamma)}_{1,1}, \dots, \widehat{(\alpha\gamma)}_{I,K}$ ,  $\widehat{(\beta\gamma)}_{1,1}, \dots, \widehat{(\beta\gamma)}_{J,K}$ ,  $\widehat{(\alpha\beta\gamma)}_{1,1,1}, \dots, \widehat{(\alpha\beta\gamma)}_{I,J,K}$ ,  $\widehat{\sigma}^2$  des paramètres  $\mu$ ,  $\alpha_1, \dots, \alpha_I$ ,  $\beta_1, \dots, \beta_J$ ,  $\gamma_1, \dots, \gamma_K$ ,  $(\alpha\beta)_{1,1}, \dots, (\alpha\beta)_{I,J}$ ,  $(\alpha\gamma)_{1,1}, \dots, (\alpha\gamma)_{I,K}$ ,  $(\beta\gamma)_{1,1}, \dots, (\beta\gamma)_{J,K}$ ,

$(\alpha\beta\gamma)_{1,1,1}, \dots, (\alpha\beta\gamma)_{I,J,K}, \sigma^2$  du modèle sont donnés par les formules suivantes :

$$\begin{aligned}\widehat{\mu} &= Y_{\bullet,\bullet,\bullet,\bullet} = \bar{Y}, \\ \widehat{\alpha}_i &= Y_{i,\bullet,\bullet,\bullet} - \widehat{\mu}, \quad 1 \leq i \leq I, \\ \widehat{\beta}_j &= Y_{\bullet,j,\bullet,\bullet} - \widehat{\mu}, \quad 1 \leq j \leq J, \\ \widehat{\gamma}_k &= Y_{\bullet,\bullet,k,\bullet} - \widehat{\mu}, \quad 1 \leq k \leq K, \\ \widehat{(\alpha\beta)}_{i,j} &= Y_{i,j,\bullet,\bullet} - Y_{i,\bullet,\bullet,\bullet} - Y_{\bullet,j,\bullet,\bullet} + \widehat{\mu}, \quad 1 \leq i \leq I, \quad 1 \leq j \leq J, \\ \widehat{(\alpha\gamma)}_{i,k} &= Y_{i,\bullet,k,\bullet} - Y_{i,\bullet,\bullet,\bullet} - Y_{\bullet,\bullet,k,\bullet} + \widehat{\mu}, \quad 1 \leq i \leq I, \quad 1 \leq k \leq K, \\ \widehat{(\beta\gamma)}_{j,k} &= Y_{\bullet,j,k,\bullet} - Y_{\bullet,j,\bullet,\bullet} - Y_{\bullet,\bullet,k,\bullet} + \widehat{\mu}, \quad 1 \leq j \leq J, \quad 1 \leq k \leq K, \\ \widehat{(\alpha\beta\gamma)}_{i,j,k} &= Y_{i,j,k,\bullet} - Y_{i,j,\bullet,\bullet} - Y_{i,\bullet,k,\bullet} - Y_{\bullet,j,k,\bullet} + Y_{i,\bullet,\bullet,\bullet} + Y_{\bullet,j,\bullet,\bullet} + Y_{\bullet,\bullet,k,\bullet} - \widehat{\mu}, \\ &\quad 1 \leq i \leq I, \quad 1 \leq j \leq J, \quad 1 \leq k \leq K, \\ \widehat{\sigma}^2 &= \frac{SC_R}{IJK(L-1)} = S_R^2.\end{aligned}$$

Ce sont des estimateurs sans biais<sup>2</sup>.

Les estimations, obtenues pour la liste de données  $\mathbf{y}$  et notées  $\widehat{\mu}(\mathbf{y}), \widehat{\alpha}_1(\mathbf{y}), \dots, \widehat{\alpha}_I(\mathbf{y}), \widehat{\beta}_1(\mathbf{y}), \dots, \widehat{\beta}_J(\mathbf{y}), \widehat{\gamma}_1(\mathbf{y}), \dots, \widehat{\gamma}_K(\mathbf{y}), \widehat{(\alpha\beta)}_{1,1}(\mathbf{y}), \dots, \widehat{(\alpha\beta)}_{I,J}(\mathbf{y}), \widehat{(\alpha\gamma)}_{1,1}(\mathbf{y}), \dots, \widehat{(\alpha\gamma)}_{I,K}(\mathbf{y}), \widehat{(\beta\gamma)}_{1,1}(\mathbf{y}), \dots, \widehat{(\beta\gamma)}_{J,K}(\mathbf{y}), \widehat{(\alpha\beta\gamma)}_{1,1,1}(\mathbf{y}), \dots, \widehat{(\alpha\beta\gamma)}_{I,J,K}(\mathbf{y}), \widehat{\sigma}^2(\mathbf{y})$  des paramètres  $\mu, \alpha_1, \dots, \alpha_I, \beta_1, \dots, \beta_J, \gamma_1, \dots, \gamma_K, (\alpha\beta)_{1,1}, \dots, (\alpha\beta)_{I,J}, (\alpha\gamma)_{1,1}, \dots, (\alpha\gamma)_{I,K}, (\beta\gamma)_{1,1}, \dots, (\beta\gamma)_{J,K}, (\alpha\beta\gamma)_{1,1,1}, \dots, (\alpha\beta\gamma)_{I,J,K}, \sigma^2$  du modèle se déduisent des formules ci-dessus :

$$\begin{aligned}\widehat{\mu}(\mathbf{y}) &= y_{\bullet,\bullet,\bullet,\bullet} = \bar{y}, \\ \widehat{\alpha}_i(\mathbf{y}) &= y_{i,\bullet,\bullet,\bullet} - \widehat{\mu}(\mathbf{y}), \quad 1 \leq i \leq I, \\ \widehat{\beta}_j(\mathbf{y}) &= y_{\bullet,j,\bullet,\bullet} - \widehat{\mu}(\mathbf{y}), \quad 1 \leq j \leq J, \\ \widehat{\gamma}_k(\mathbf{y}) &= y_{\bullet,\bullet,k,\bullet} - \widehat{\mu}(\mathbf{y}), \quad 1 \leq k \leq K, \\ \widehat{(\alpha\beta)}_{i,j}(\mathbf{y}) &= y_{i,j,\bullet,\bullet} - y_{i,\bullet,\bullet,\bullet} - y_{\bullet,j,\bullet,\bullet} + \widehat{\mu}(\mathbf{y}), \quad 1 \leq i \leq I, \quad 1 \leq j \leq J, \\ \widehat{(\alpha\gamma)}_{i,k}(\mathbf{y}) &= y_{i,\bullet,k,\bullet} - y_{i,\bullet,\bullet,\bullet} - y_{\bullet,\bullet,k,\bullet} + \widehat{\mu}(\mathbf{y}), \quad 1 \leq i \leq I, \quad 1 \leq k \leq K, \\ \widehat{(\beta\gamma)}_{j,k}(\mathbf{y}) &= y_{\bullet,j,k,\bullet} - y_{\bullet,j,\bullet,\bullet} - y_{\bullet,\bullet,k,\bullet} + \widehat{\mu}(\mathbf{y}), \quad 1 \leq j \leq J, \quad 1 \leq k \leq K, \\ \widehat{(\alpha\beta\gamma)}_{i,j,k}(\mathbf{y}) &= y_{i,j,k,\bullet} - y_{i,j,\bullet,\bullet} - y_{i,\bullet,k,\bullet} - y_{\bullet,j,k,\bullet} + y_{i,\bullet,\bullet,\bullet} + y_{\bullet,j,\bullet,\bullet} + y_{\bullet,\bullet,k,\bullet} - \widehat{\mu}(\mathbf{y}), \\ &\quad 1 \leq i \leq I, \quad 1 \leq j \leq J, \quad 1 \leq k \leq K, \\ \widehat{\sigma}^2(\mathbf{y}) &= \frac{SC_R}{IJK(L-1)} = s_R^2.\end{aligned}$$

## 5.2. Modèles à effets aléatoires

### 5.2.1. Sans répétition

Les  $\alpha_i$  représentent un échantillon de taille  $I$  prélevé dans une population importante. Nous admettrons que les effets des  $A_i$ , les  $\alpha_i$ , sont distribués suivant une loi normale

2. La notion de biais d'un estimateur est définie dans le chapitre ??.

## 5.2 Modèles à effets aléatoires

---

centrée de variance  $\sigma_A^2$ . Les  $\beta_j$  représentent un échantillon de taille  $J$  prélevé dans une population importante. Nous admettrons que les effets des  $B_j$ , les  $\beta_j$ , sont distribués suivant une loi normale centrée de variance  $\sigma_B^2$ . Les  $\gamma_k$  représentent un échantillon de taille  $K$  prélevé dans une population importante. Nous admettrons que les effets des  $C_k$ , les  $\gamma_k$ , sont distribués suivant une loi normale centrée de variance  $\sigma_C^2$ . Pour chacun des couples de modalités  $(A_i, B_j, C_k)$  nous effectuons une mesure d'une réponse  $Y$  qui est une variable continue. Nous notons  $n = I \times J \times K$  le nombre total de mesures ayant été effectuées.

Nous introduisons le modèle :

$$Y_{i,j,k} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{i,j} + (\alpha\gamma)_{i,k} + (\beta\gamma)_{j,k} + \epsilon_{i,j,k},$$

$$i = 1 \dots I, j = 1 \dots J, k = 1 \dots K,$$

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

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

ainsi que l'indépendance des effets aléatoires  $\alpha_i, \beta_j, \gamma_k, (\alpha\beta)_{i,j}, (\alpha\gamma)_{i,k}$  et  $(\beta\gamma)_{j,k}$ . Nous postulons les hypothèses classiques suivantes pour les erreurs :

$$\begin{aligned} \forall (i, j, k), 1 \leq i \leq I, 1 \leq j \leq J, 1 \leq k \leq K, \quad \mathcal{L}(\epsilon_{i,j,k}) &= \mathcal{N}(0, \sigma^2), \\ \epsilon_{i,j,k} \perp\!\!\!\perp \epsilon_{l,m,n} \text{ si } (i, j, k) \neq (l, m, n) \text{ avec } 1 \leq i, l \leq I, 1 \leq j, m \leq J \text{ et } 1 \leq k, n \leq K, \end{aligned}$$

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

Nous supposons que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fait l'objet d'un paragraphe du chapitre ??.

Nous utilisons les quantités  $sc_A, sc_B, sc_C, sc_{AB}, sc_{AC}, sc_{BC}, sc_R$  et  $sc_{TOT}$  introduites à la section 5.1.1.

Nous rappelons la relation fondamentale de l'ANOVA :

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

Nous introduisons les degrés 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	$n_{TOT} = IJK - 1$

Nous résumons ces informations dans le tableau de l'ANOVA ci-dessous :

Source	Variation	Ddl	Carré Moyen	$F$	Décision
Facteur $A$	$sc_A$	$n_A$	$s_A^2 = \frac{sc_A}{n_A}$	$f_A = \frac{s_A^2}{s_{AB}^2 + s_{AC}^2 - s_R^2}$	$\mathcal{H}'_0$ ou $\mathcal{H}'_1$
Facteur $B$	$sc_B$	$n_B$	$s_B^2 = \frac{sc_B}{n_B}$	$f_B = \frac{s_B^2}{s_{AB}^2 + s_{BC}^2 - s_R^2}$	$\mathcal{H}''_0$ ou $\mathcal{H}''_1$
Facteur $C$	$sc_C$	$n_C$	$s_C^2 = \frac{sc_C}{n_C}$	$f_C = \frac{s_C^2}{s_{AC}^2 + s_{BC}^2 - s_R^2}$	$\mathcal{H}'''_0$ 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	$sc_{TOT}$	$n_{TOT}$			

Nous souhaitons 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**<sup>3</sup> une loi de Fisher à  $I - 1$  et  $n'$  degrés de liberté avec :

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

Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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**<sup>5</sup> une loi de Fisher à  $J - 1$  et  $n''$  degrés de liberté avec :

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

Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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

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

---

3. Nous utilisons 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** <sup>5</sup> une loi de Fisher à  $K - 1$  et  $n'''$  degrés de liberté avec :

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

Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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  $(I-1)(J-1)(K-1)$



## 5.2 Modèles à effets aléatoires

degrés de liberté. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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 estimateurs  $\widehat{\mu}$ ,  $\widehat{\sigma}_A^2$ ,  $\widehat{\sigma}_B^2$ ,  $\widehat{\sigma}_C^2$ ,  $\widehat{\sigma}_{AB}^2$ ,  $\widehat{\sigma}_{AC}^2$ ,  $\widehat{\sigma}_{BC}^2$ ,  $\widehat{\sigma}^2$  des paramètres  $\mu$ ,  $\sigma_A^2$ ,  $\sigma_B^2$ ,  $\sigma_C^2$ ,  $\sigma_{AB}^2$ ,  $\sigma_{AC}^2$ ,  $\sigma_{BC}^2$ ,  $\sigma^2$  du modèle sont donnés par les formules suivantes :

$$\begin{aligned}\widehat{\mu} &= Y_{\bullet,\bullet,\bullet} = \bar{Y}, \\ \widehat{\sigma}_A^2 &= \frac{1}{JK} (S_A^2 - S_{AB}^2 - S_{AC}^2 + S_R^2), \\ \widehat{\sigma}_B^2 &= \frac{1}{IK} (S_B^2 - S_{AB}^2 - S_{BC}^2 + S_R^2), \\ \widehat{\sigma}_C^2 &= \frac{1}{IJ} (S_C^2 - S_{AC}^2 - S_{BC}^2 + S_R^2), \\ \widehat{\sigma}_{AB}^2 &= \frac{1}{K} (S_{AB}^2 - S_R^2), \\ \widehat{\sigma}_{AC}^2 &= \frac{1}{J} (S_{AC}^2 - S_R^2), \\ \widehat{\sigma}_{BC}^2 &= \frac{1}{I} (S_{BC}^2 - S_R^2), \\ \widehat{\sigma}^2 &= \frac{SC_R}{(I-1)(J-1)(K-1)} = S_R^2,\end{aligned}$$

où  $S_A^2 = \frac{SC_A}{n_A}$ ,  $S_B^2 = \frac{SC_B}{n_B}$ ,  $S_C^2 = \frac{SC_C}{n_C}$ ,  $S_{AB}^2 = \frac{SC_{AB}}{n_{AB}}$ ,  $S_{AC}^2 = \frac{SC_{AC}}{n_{AC}}$ ,  $S_{BC}^2 = \frac{SC_{BC}}{n_{BC}}$  et  $S_R^2 = \frac{SC_R}{n_R}$ .

Ce sont des estimateurs sans biais<sup>6</sup>.

Les estimations, obtenues pour la liste de données  $\mathbf{y}$  et notées  $\widehat{\mu}(\mathbf{y})$ ,  $\widehat{\sigma}_A^2(\mathbf{y})$ ,  $\widehat{\sigma}_B^2(\mathbf{y})$ ,  $\widehat{\sigma}_C^2(\mathbf{y})$ ,  $\widehat{\sigma}_{AB}^2(\mathbf{y})$ ,  $\widehat{\sigma}_{AC}^2(\mathbf{y})$ ,  $\widehat{\sigma}_{BC}^2(\mathbf{y})$ ,  $\widehat{\sigma}^2(\mathbf{y})$  des paramètres  $\mu$ ,  $\sigma_A^2$ ,  $\sigma_B^2$ ,  $\sigma_C^2$ ,  $\sigma_{AB}^2$ ,  $\sigma_{AC}^2$ ,  $\sigma_{BC}^2$ ,  $\sigma^2$  du modèle se déduisent des formules ci-dessus :

$$\begin{aligned}\widehat{\mu}(\mathbf{y}) &= y_{\bullet,\bullet,\bullet} = \bar{y}, \\ \widehat{\sigma}_A^2(\mathbf{y}) &= \frac{1}{JK} (s_A^2 - s_{AB}^2 - s_{AC}^2 + s_R^2), \\ \widehat{\sigma}_B^2(\mathbf{y}) &= \frac{1}{IK} (s_B^2 - s_{AB}^2 - s_{BC}^2 + s_R^2), \\ \widehat{\sigma}_C^2(\mathbf{y}) &= \frac{1}{IJ} (s_C^2 - s_{AC}^2 - s_{BC}^2 + s_R^2), \\ \widehat{\sigma}_{AB}^2(\mathbf{y}) &= \frac{1}{K} (s_{AB}^2 - s_R^2), \\ \widehat{\sigma}_{AC}^2(\mathbf{y}) &= \frac{1}{J} (s_{AC}^2 - s_R^2), \\ \widehat{\sigma}_{BC}^2(\mathbf{y}) &= \frac{1}{I} (s_{BC}^2 - s_R^2), \\ \widehat{\sigma}^2(\mathbf{y}) &= \frac{sc_R}{(I-1)(J-1)(K-1)} = s_R^2.\end{aligned}$$

6. La notion de biais d'un estimateur est définie dans le chapitre ??.

**Remarque 5.2.1.** Ainsi lorsque les trois facteurs sont aléatoires, et que nous cherchons à tester l'existence d'un effet de l'un d'entre eux nous devons 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 5.1.1 et 5.3.1.

### 5.2.2. Avec répétitions

Les  $\alpha_i$  représentent un échantillon de taille  $I$  prélevé dans une population importante. Nous admettrons que les effets des  $A_i$ , les  $\alpha_i$ , sont distribués suivant une loi normale centrée de variance  $\sigma_A^2$ . Les  $\beta_j$  représentent un échantillon de taille  $J$  prélevé dans une population importante. Nous admettrons que les effets des  $B_j$ , les  $\beta_j$ , sont distribués suivant une loi normale centrée de variance  $\sigma_B^2$ . Les  $\gamma_k$  représentent un échantillon de taille  $K$  prélevé dans une population importante. Nous admettrons que les effets des  $C_k$ , les  $\gamma_k$ , sont distribués suivant une loi normale centrée de variance  $\sigma_C^2$ . Pour chacun des couples de modalités  $(A_i, B_j, C_k)$  nous effectuons  $L \geq 2$  mesures d'une réponse  $Y$  qui est une variable continue. Nous notons  $n = I \times J \times K \times L$  le nombre total de mesures ayant été effectuées.

Nous introduisons le modèle :

$$Y_{i,j,k,l} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{i,j} + (\alpha\gamma)_{i,k} + (\beta\gamma)_{j,k} + (\alpha\beta\gamma)_{i,j,k} + \epsilon_{i,j,k,l},$$

$$i = 1 \dots I, j = 1 \dots J, k = 1 \dots K, l = 1 \dots L,$$

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. Nous supposons que :

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

ainsi que l'indépendance des effets aléatoires  $\alpha_i, \beta_j, \gamma_k, (\alpha\beta)_{i,j}, (\alpha\gamma)_{i,k}, (\beta\gamma)_{j,k}$  et  $(\alpha\beta\gamma)_{i,j,k}$ . Nous postulons les hypothèses classiques suivantes pour les erreurs :

$$\begin{aligned} \forall (i, j, k, l), 1 \leq i \leq I, 1 \leq j \leq J, 1 \leq k \leq K, 1 \leq l \leq L, \quad \mathcal{L}(\epsilon_{i,j,k,l}) &= \mathcal{N}(0, \sigma^2), \\ \epsilon_{i,j,k,l} &\perp\!\!\!\perp \epsilon_{m,n,o,p} \text{ si } (i, j, k, l) \neq (m, n, o, p) \\ \text{avec } 1 \leq i, m \leq I, 1 \leq j, n \leq J, 1 \leq k, o \leq K \text{ et } 1 \leq l, p \leq L, \end{aligned}$$

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

Nous supposons que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fait l'objet d'un paragraphe du chapitre ??.

## 5.2 Modèles à effets aléatoires

---

Nous utilisons 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 5.1.2.

Nous rappelons la relation fondamentale de l'ANOVA :

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

Nous introduisons les degrés 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	$n_{TOT} = IJK - 1$

Nous résumons ces informations dans le tableau de l'ANOVA ci-dessous :

## Chapitre 5. Analyse de la variance à trois facteurs

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 $AB$	$sc_{AB}$	$n_{AB}$	$s_{AB}^2 = \frac{sc_{AB}}{n_{AB}}$	$f_{AB} = \frac{s_{AB}^2}{s_{ABC}^2}$	$\mathcal{H}^{(4)}_0$ ou $\mathcal{H}^{(4)}_1$
Interaction $AC$	$sc_{AC}$	$n_{AC}$	$s_{AC}^2 = \frac{sc_{AC}}{n_{AC}}$	$f_{AC} = \frac{s_{AC}^2}{s_{ABC}^2}$	$\mathcal{H}^{(5)}_0$ ou $\mathcal{H}^{(5)}_1$
Interaction $BC$	$sc_{BC}$	$n_{BC}$	$s_{BC}^2 = \frac{sc_{BC}}{n_{BC}}$	$f_{BC} = \frac{s_{BC}^2}{s_{ABC}^2}$	$\mathcal{H}^{(6)}_0$ ou $\mathcal{H}^{(6)}_1$
Interaction $ABC$	$sc_{ABC}$	$n_{ABC}$	$s_{ABC}^2 = \frac{sc_{ABC}}{n_{ABC}}$	$f_{ABC} = \frac{s_{ABC}^2}{s_R^2}$	$\mathcal{H}^{(7)}_0$ ou $\mathcal{H}^{(7)}_1$
Résiduelle	$sc_R$	$n_R$	$s_R^2 = \frac{sc_R}{n_R}$		
Totale	$sc_{TOT}$	$n_{TOT}$			

Nous souhaitons faire les tests d'hypothèse suivants :

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

contre

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

## 5.2 Modèles à effets aléatoires

---

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**<sup>7</sup> une loi de Fisher à  $I - 1$  et  $n'$  degrés de liberté avec :

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

Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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**<sup>6</sup> une loi de Fisher à  $J - 1$  et  $n''$  degrés de liberté avec :

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

Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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

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

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**<sup>6</sup> une loi de Fisher à  $K - 1$  et  $n'''$  degrés de liberté avec :

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

Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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.

---

7. Nous utilisons ici l'approximation dite de Satterthwaite.

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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.

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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.

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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.

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

contre

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

## 5.2 Modèles à effets aléatoires

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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 estimateurs  $\widehat{\mu}$ ,  $\widehat{\sigma}_A^2$ ,  $\widehat{\sigma}_B^2$ ,  $\widehat{\sigma}_C^2$ ,  $\widehat{\sigma}_{AB}^2$ ,  $\widehat{\sigma}_{AC}^2$ ,  $\widehat{\sigma}_{BC}^2$ ,  $\widehat{\sigma}_{ABC}^2$ ,  $\widehat{\sigma}^2$  des paramètres  $\mu$ ,  $\sigma_A^2$ ,  $\sigma_B^2$ ,  $\sigma_C^2$ ,  $\sigma_{AB}^2$ ,  $\sigma_{AC}^2$ ,  $\sigma_{BC}^2$ ,  $\sigma_{ABC}^2$ ,  $\sigma^2$  du modèle sont donnés par les formules suivantes :

$$\begin{aligned}\widehat{\mu} &= Y_{\bullet,\bullet,\bullet,\bullet} = \bar{Y}, \\ \widehat{\sigma}_A^2 &= \frac{1}{JKL} (S_A^2 - S_{AB}^2 - S_{AC}^2 + S_{ABC}^2), \\ \widehat{\sigma}_B^2 &= \frac{1}{IKL} (S_B^2 - S_{AB}^2 - S_{BC}^2 + S_{ABC}^2), \\ \widehat{\sigma}_C^2 &= \frac{1}{IJJ} (S_C^2 - S_{AC}^2 - S_{BC}^2 + S_{ABC}^2), \\ \widehat{\sigma}_{AB}^2 &= \frac{1}{KL} (S_{AB}^2 - S_{ABC}^2), \\ \widehat{\sigma}_{AC}^2 &= \frac{1}{JL} (S_{AC}^2 - S_{ABC}^2), \\ \widehat{\sigma}_{BC}^2 &= \frac{1}{IL} (S_{BC}^2 - S_{ABC}^2), \\ \widehat{\sigma}_{ABC}^2 &= \frac{1}{L} (S_{ABC}^2 - S_R^2), \\ \widehat{\sigma}^2 &= \frac{SC_R}{IJK(L-1)} = S_R^2.\end{aligned}$$

$$\text{où } S_A^2 = \frac{SC_A}{n_A}, S_B^2 = \frac{SC_B}{n_B}, S_C^2 = \frac{SC_C}{n_C}, S_{AB}^2 = \frac{SC_{AB}}{n_{AB}}, S_{AC}^2 = \frac{SC_{AC}}{n_{AC}}, S_{BC}^2 = \frac{SC_{BC}}{n_{BC}}, \\ S_{ABC}^2 = \frac{SC_{ABC}}{n_{ABC}} \text{ et } S_R^2 = \frac{SC_R}{n_R}.$$

Ce sont des estimateurs sans biais<sup>7</sup>.

Les estimations, obtenues pour la liste de données  $\mathbf{y}$  et notées  $\widehat{\mu}(\mathbf{y})$ ,  $\widehat{\sigma}_A^2(\mathbf{y})$ ,  $\widehat{\sigma}_B^2(\mathbf{y})$ ,  $\widehat{\sigma}_C^2(\mathbf{y})$ ,  $\widehat{\sigma}_{AB}^2(\mathbf{y})$ ,  $\widehat{\sigma}_{AC}^2(\mathbf{y})$ ,  $\widehat{\sigma}_{BC}^2(\mathbf{y})$ ,  $\widehat{\sigma}_{ABC}^2(\mathbf{y})$ ,  $\widehat{\sigma}^2(\mathbf{y})$  des paramètres  $\mu$ ,  $\sigma_A^2$ ,  $\sigma_B^2$ ,  $\sigma_C^2$ ,  $\sigma_{AB}^2$ ,  $\sigma_{AC}^2$ ,  $\sigma_{BC}^2$ ,  $\sigma_{ABC}^2$ ,  $\sigma^2$  du modèle se déduisent des formules ci-dessus :

$$\begin{aligned}\widehat{\mu}(\mathbf{y}) &= y_{\bullet,\bullet,\bullet,\bullet} = \bar{y}, \\ \widehat{\sigma}_A^2(\mathbf{y}) &= \frac{1}{JKL} (s_A^2 - s_{AB}^2 - s_{AC}^2 + s_{ABC}^2), \\ \widehat{\sigma}_B^2(\mathbf{y}) &= \frac{1}{IKL} (s_B^2 - s_{AB}^2 - s_{BC}^2 + s_{ABC}^2), \\ \widehat{\sigma}_C^2(\mathbf{y}) &= \frac{1}{IJJ} (s_C^2 - s_{AC}^2 - s_{BC}^2 + s_{ABC}^2), \\ \widehat{\sigma}_{AB}^2(\mathbf{y}) &= \frac{1}{KL} (s_{AB}^2 - s_{ABC}^2),\end{aligned}$$

7. La notion de biais d'un estimateur est définie dans le chapitre ??.

$$\begin{aligned}\widehat{\sigma}_{AC}^2(\mathbf{y}) &= \frac{1}{JL} (s_{AC}^2 - s_{ABC}^2), \\ \widehat{\sigma}_{BC}^2(\mathbf{y}) &= \frac{1}{IL} (s_{BC}^2 - s_{ABC}^2), \\ \widehat{\sigma}_{ABC}^2(\mathbf{y}) &= \frac{1}{L} (s_{ABC}^2 - s_R^2), \\ \widehat{\sigma}^2(\mathbf{y}) &= \frac{s_{CR}}{IJK(L-1)} = s_R^2.\end{aligned}$$

### 5.3. Modèles à effets mixtes

#### 5.3.1. 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  $\gamma_k$  représentent un échantillon de taille  $K$  prélevé dans une population importante. Nous admettrons que les effets des  $C_k$ , les  $\gamma_k$ , sont distribués suivant une loi normale centrée de variance  $\sigma_C^2$ . Pour chacun des couples de modalités  $(A_i, B_j, C_k)$  nous effectuons une mesure d'une réponse  $Y$  qui est une variable continue. Nous notons  $n = I \times J \times K$  le nombre total de mesures ayant été effectuées.

Nous introduisons le modèle :

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

$$\text{avec les contraintes supplémentaires } \sum_{i=1}^I \alpha_i = 0, \sum_{j=1}^J \beta_j = 0,$$

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

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

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



### 5.3 Modèles à effets mixtes

ainsi que l'indépendance des effets aléatoires  $\gamma_k$ .

Nous postulons les hypothèses classiques suivantes pour les erreurs :

$$\forall (i, j, k), 1 \leq i \leq I, 1 \leq j \leq J, 1 \leq k \leq K, \mathcal{L}(\epsilon_{i,j,k}) = \mathcal{N}(0, \sigma^2),$$

$$\epsilon_{i,j,k} \perp\!\!\!\perp \epsilon_{l,m,n} \text{ si } (i, j, k) \neq (l, m, n) \text{ avec } 1 \leq i, l \leq I, 1 \leq j, m \leq J \text{ et } 1 \leq k, n \leq K,$$

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

Dans un modèle mixte restreint, les effets aléatoires croisant des facteurs à effets fixes et à effets aléatoires, ici aussi bien les  $(\alpha\gamma)_{i,k}$  que les  $(\beta\gamma)_{j,k}$ , ne sont pas mutuellement

indépendants à cause des contraintes portant sur leur somme,  $\sum_{i=1}^I (\alpha\gamma)_{i,k} = 0, \forall k \in$

$\{1, \dots, K\}$  et  $\sum_{j=1}^J (\beta\gamma)_{j,k} = 0, \forall k \in \{1, \dots, K\}$ . Ils le sont par contre dès que nous ne

les considérons pas tous en même temps.

Certains logiciels travaillent par défaut avec un modèle non-restreint dont la définition est proche mais pas identique à celle du modèle précédent.

Nous introduisons le modèle :

$$Y_{i,j,k} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{i,j} + (\alpha\gamma)_{i,k} + (\beta\gamma)_{j,k} + \epsilon_{i,j,k},$$

$$i = 1 \dots I, j = 1 \dots J, k = 1 \dots K,$$

$$\text{avec les contraintes supplémentaires } \sum_{i=1}^I \alpha_i = 0, \sum_{j=1}^J \beta_j = 0,$$

$$\sum_{i=1}^I (\alpha\beta)_{i,j} = 0, \forall j \in \{1, \dots, J\}, \sum_{j=1}^J (\alpha\beta)_{i,j} = 0, \forall i \in \{1, \dots, I\},$$

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

$$\mathcal{L}(\gamma_k) = \mathcal{N}(0, \sigma_C^2), \forall k, 1 \leq k \leq K,$$

$$\mathcal{L}((\alpha\gamma)_{i,k}) = \mathcal{N}(0, \sigma_{AC}^2), \forall (i, k), 1 \leq i \leq I, 1 \leq k \leq K,$$

$$\mathcal{L}((\beta\gamma)_{j,k}) = \mathcal{N}(0, \sigma_{BC}^2), \forall (j, k), 1 \leq j \leq J, 1 \leq k \leq K,$$

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

Nous postulons les hypothèses classiques suivantes pour les erreurs :

$$\forall (i, j, k), 1 \leq i \leq I, 1 \leq j \leq J, 1 \leq k \leq K, \mathcal{L}(\epsilon_{i,j,k}) = \mathcal{N}(0, \sigma^2),$$

$$\epsilon_{i,j,k} \perp\!\!\!\perp \epsilon_{l,m,n} \text{ si } (i, j, k) \neq (l, m, n) \text{ avec } 1 \leq i, l \leq I, 1 \leq j, m \leq J \text{ et } 1 \leq k, n \leq K,$$

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

Dans un modèle mixte non restreint, les effets aléatoires croisant des facteurs à effets fixes et à effets aléatoires, ici les  $(\alpha\gamma)_{i,k}$  et les  $(\beta\gamma)_{j,k}$ , sont mutuellement indépendants.

Il n'existe pas consensus sur une raison statistique qui permettrait de privilégier plutôt l'une ou l'autre de ces deux approches. Nous utiliserons toujours des modèles restreints.

Nous supposons que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fait l'objet d'un paragraphe du chapitre ??.

Nous utilisons les quantités  $sc_A$ ,  $sc_B$ ,  $sc_C$ ,  $sc_{AB}$ ,  $sc_{AC}$ ,  $sc_{BC}$ ,  $sc_R$  et  $sc_{TOT}$  introduites à la section 5.1.1.

Nous rappelons la relation fondamentale de l'ANOVA :

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

Nous introduisons les degrés 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	$n_{TOT} = IJK - 1$

Nous résumons ces informations dans le tableau de l'ANOVA ci-dessous :

### 5.3 Modèles à effets mixtes

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 $AB$	$sc_{AB}$	$n_{AB}$	$s_{AB}^2 = \frac{sc_{AB}}{n_{AB}}$	$f_{AB} = \frac{s_{AB}^2}{s_R^2}$	$\mathcal{H}^{(4)}_0$ ou $\mathcal{H}^{(4)}_1$
Interaction $AC$	$sc_{AC}$	$n_{AC}$	$s_{AC}^2 = \frac{sc_{AC}}{n_{AC}}$	$f_{AC} = \frac{s_{AC}^2}{s_R^2}$	$\mathcal{H}^{(5)}_0$ ou $\mathcal{H}^{(5)}_1$
Interaction $BC$	$sc_{BC}$	$n_{BC}$	$s_{BC}^2 = \frac{sc_{BC}}{n_{BC}}$	$f_{BC} = \frac{s_{BC}^2}{s_R^2}$	$\mathcal{H}^{(6)}_0$ ou $\mathcal{H}^{(6)}_1$
Résiduelle	$sc_R$	$n_R$	$s_R^2 = \frac{sc_R}{n_R}$		
Totale	$sc_{TOT}$	$n_{TOT}$			

Nous souhaitons faire les tests d'hypothèse suivants :

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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, nous pouvons procéder à des comparaisons multiples des différents effets des niveaux du facteur voir le chapitre 9.

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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, nous pouvons procéder à des comparaisons multiples des différents effets des niveaux du facteur voir le chapitre 9.

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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.

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

contre

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

Sous l'hypothèse nulle  $\mathcal{H}^{(4)}_0$  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  $(I - 1)(J - 1)(K - 1)$  degrés de liberté. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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.

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

contre

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

### 5.3 Modèles à effets mixtes

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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 estimateurs  $\widehat{\mu}$ ,  $\widehat{\alpha}_1, \dots, \widehat{\alpha}_I$ ,  $\widehat{\beta}_1, \dots, \widehat{\beta}_J$ ,  $\widehat{\sigma}_C^2$ ,  $\widehat{(\alpha\beta)}_{1,1}, \dots, \widehat{(\alpha\beta)}_{I,J}$ ,  $\widehat{\sigma}_{AC}^2$ ,  $\widehat{\sigma}_{BC}^2$ ,  $\widehat{\sigma}^2$  des paramètres  $\mu$ ,  $\alpha_1, \dots, \alpha_I$ ,  $\beta_1, \dots, \beta_J$ ,  $\sigma_C^2$ ,  $(\alpha\beta)_{1,1}, \dots, (\alpha\beta)_{I,J}$ ,  $\sigma_{AC}^2$ ,  $\sigma_{BC}^2$ ,  $\sigma^2$  du modèle sont donnés par les formules suivantes :

$$\begin{aligned}\widehat{\mu} &= Y_{\bullet,\bullet,\bullet} = \bar{Y}, \\ \widehat{\alpha}_i &= Y_{i,\bullet,\bullet} - \widehat{\mu}, \quad 1 \leq i \leq I, \quad \widehat{\beta}_j = Y_{\bullet,j,\bullet} - \widehat{\mu}, \quad 1 \leq j \leq J, \\ \widehat{\sigma}_C^2 &= \frac{1}{IJ} (S_C^2 - S_R^2), \\ \widehat{(\alpha\beta)}_{i,j} &= Y_{i,j,\bullet} - Y_{i,\bullet,\bullet} - Y_{\bullet,j,\bullet} + \widehat{\mu}, \quad 1 \leq i \leq I, \quad 1 \leq j \leq J, \\ \widehat{\sigma}_{AC}^2 &= \frac{1}{J} (S_{AC}^2 - S_R^2), \\ \widehat{\sigma}_{BC}^2 &= \frac{1}{I} (S_{BC}^2 - S_R^2), \\ \widehat{\sigma}^2 &= \frac{SC_R}{(I-1)(J-1)(K-1)} = S_R^2.\end{aligned}$$

$$\text{où } S_C^2 = \frac{SC_C}{n_C}, S_{AC}^2 = \frac{SC_{AC}}{n_{AC}}, S_{BC}^2 = \frac{SC_{BC}}{n_{BC}} \text{ et } S_R^2 = \frac{SC_R}{n_R}.$$

Ce sont des estimateurs sans biais<sup>8</sup>.

Les estimations, obtenues pour la liste de données  $\mathbf{y}$  et notées  $\widehat{\mu}(\mathbf{y})$ ,  $\widehat{\alpha}_1(\mathbf{y})$ ,  $\dots$ ,  $\widehat{\alpha}_I(\mathbf{y})$ ,  $\widehat{\beta}_1(\mathbf{y})$ ,  $\dots$ ,  $\widehat{\beta}_J(\mathbf{y})$ ,  $\widehat{\sigma}_C^2(\mathbf{y})$ ,  $\widehat{(\alpha\beta)}_{1,1}(\mathbf{y})$ ,  $\dots$ ,  $\widehat{(\alpha\beta)}_{I,J}(\mathbf{y})$ ,  $\widehat{\sigma}_{AC}^2(\mathbf{y})$ ,  $\widehat{\sigma}_{BC}^2(\mathbf{y})$ ,  $\widehat{\sigma}^2(\mathbf{y})$  des paramètres  $\mu$ ,  $\alpha_1, \dots, \alpha_I$ ,  $\beta_1, \dots, \beta_J$ ,  $\sigma_C^2$ ,  $(\alpha\beta)_{1,1}, \dots, (\alpha\beta)_{I,J}$ ,  $\sigma_{AC}^2$ ,  $\sigma_{BC}^2$ ,  $\sigma^2$  du modèle se

8. La notion de biais d'un estimateur est définie dans le chapitre ??.

déduisent des formules ci-dessus :

$$\begin{aligned}\widehat{\mu}(\mathbf{y}) &= y_{\bullet,\bullet,\bullet} = \bar{y}, \\ \widehat{\alpha}_i(\mathbf{y}) &= y_{i,\bullet,\bullet} - \widehat{\mu}(\mathbf{y}), \quad 1 \leq i \leq I, \quad \widehat{\beta}_j(\mathbf{y}) = y_{\bullet,j,\bullet} - \widehat{\mu}(\mathbf{y}), \quad 1 \leq j \leq J, \\ \widehat{\sigma}_C^2(\mathbf{y}) &= \frac{1}{IJ} (s_C^2 - s_R^2), \\ \widehat{(\alpha\beta)}_{i,j}(\mathbf{y}) &= y_{i,j,\bullet} - y_{i,\bullet,\bullet} - y_{\bullet,j,\bullet} + \widehat{\mu}(\mathbf{y}), \quad 1 \leq i \leq I, \quad 1 \leq j \leq J, \\ \widehat{\sigma}_{AC}^2(\mathbf{y}) &= \frac{1}{J} (s_{AC}^2 - s_R^2), \\ \widehat{\sigma}_{BC}^2(\mathbf{y}) &= \frac{1}{I} (s_{BC}^2 - s_R^2), \\ \widehat{\sigma}^2(\mathbf{y}) &= \frac{s_{CR}}{(I-1)(J-1)(K-1)} = s_R^2.\end{aligned}$$

**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  $\beta_j$  représentent un échantillon de taille  $J$  prélevé dans une population importante. Nous admettrons que les effets des  $B_j$ , les  $\beta_j$ , sont distribués suivant une loi normale centrée de variance  $\sigma_B^2$ . Les  $\gamma_k$  représentent un échantillon de taille  $K$  prélevé dans une population importante. Nous admettrons que les effets des  $C_k$ , les  $\gamma_k$ , sont distribués suivant une loi normale centrée de variance  $\sigma_C^2$ . Pour chacun des couples de modalités  $(A_i, B_j, C_k)$  nous effectuons une mesure d'une réponse  $Y$  qui est une variable continue. Nous notons  $n = I \times J \times K$  le nombre total de mesures ayant été effectuées.

Nous introduisons le modèle :

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

$$\text{avec les contraintes supplémentaires } \sum_{i=1}^I \alpha_i = 0,$$

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

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

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

ainsi que l'indépendance des effets aléatoires  $\beta_j$ ,  $\gamma_k$ , et  $(\beta\gamma)_{j,k}$ .

Nous postulons les hypothèses classiques suivantes pour les erreurs :

$$\begin{aligned}\forall (i, j, k), 1 \leq i \leq I, \quad 1 \leq j \leq J, \quad 1 \leq k \leq K, \quad \mathcal{L}(\epsilon_{i,j,k}) &= \mathcal{N}(0, \sigma^2), \\ \epsilon_{i,j,k} \perp\!\!\!\perp \epsilon_{l,m,n} \text{ si } (i, j, k) \neq (l, m, n) \text{ avec } 1 \leq i, l \leq I, \quad 1 \leq j, m \leq J \text{ et } 1 \leq k, n \leq K,\end{aligned}$$

### 5.3 Modèles à effets mixtes

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

Dans un modèle mixte restreint, les effets aléatoires croisant des facteurs à effets fixes et à effets aléatoires, ici aussi bien les  $(\alpha\beta)_{i,j}$  que les  $(\alpha\gamma)_{i,k}$ , ne sont pas mutuellement

indépendants à cause des contraintes portant sur leur somme,  $\sum_{i=1}^I (\alpha\beta)_{i,j} = 0$ ,  $\forall j \in$

$\{1, \dots, J\}$  et  $\sum_{i=1}^I (\alpha\gamma)_{i,k} = 0$ ,  $\forall k \in \{1, \dots, K\}$ . Ils le sont par contre dès que nous ne

les considérons pas tous en même temps.

Certains logiciels travaillent par défaut avec un modèle non-restreint dont la définition est proche mais pas identique à celle du modèle précédent.

Nous introduisons le modèle :

$$Y_{i,j,k} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{i,j} + (\alpha\gamma)_{i,k} + (\beta\gamma)_{j,k} + \epsilon_{i,j,k},$$

$$i = 1 \dots I, j = 1 \dots J, k = 1 \dots K,$$

$$\text{avec les contraintes supplémentaires } \sum_{i=1}^I \alpha_i = 0,$$

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

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

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

Nous postulons les hypothèses classiques suivantes pour les erreurs :

$$\begin{aligned} \forall (i, j, k), 1 \leq i \leq I, 1 \leq j \leq J, 1 \leq k \leq K, \quad \mathcal{L}(\epsilon_{i,j,k}) &= \mathcal{N}(0, \sigma^2), \\ \epsilon_{i,j,k} \perp\!\!\!\perp \epsilon_{l,m,n} \text{ si } (i, j, k) &\neq (l, m, n) \text{ avec } 1 \leq i, l \leq I, 1 \leq j, m \leq J \text{ et } 1 \leq k, n \leq K, \end{aligned}$$

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

Dans un modèle mixte non restreint, les effets aléatoires croisant des facteurs à effets fixes et à effets aléatoires, ici les  $(\alpha\beta)_{i,j}$  et les  $(\alpha\gamma)_{i,k}$ , sont mutuellement indépendants.

Il n'existe pas consensus sur une raison statistique qui permettrait de privilégier plutôt l'une ou l'autre de ces deux approches. Nous utiliserons toujours des modèles restreints.

Nous supposons que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fait l'objet d'un paragraphe du chapitre ??.

Nous utilisons les quantités  $sc_A$ ,  $sc_B$ ,  $sc_C$ ,  $sc_{AB}$ ,  $sc_{AC}$ ,  $sc_{BC}$ ,  $sc_R$  et  $sc_{TOT}$  introduites à la section 5.1.1.

Nous rappelons la relation fondamentale de l'ANOVA :

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

Nous introduisons les degrés 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	$n_{TOT} = IJK - 1$

Nous résumons ces informations dans le tableau de l'ANOVA ci-dessous :



### 5.3 Modèles à effets mixtes

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

Nous souhaitons faire les tests d'hypothèse suivants :

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

contre

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

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**<sup>9</sup> une loi de Fisher à  $I - 1$  et  $n'$  degrés de liberté avec :

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

Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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, nous pouvons procéder à des comparaisons multiples des différents effets des niveaux du facteur voir le chapitre 9.

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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.

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

contre

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

Sous l'hypothèse nulle  $\mathcal{H}^{(4)}_0$  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  $(I - 1)(J - 1)(K - 1)$  degrés de liberté. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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.

---

9. Nous utilisons toujours l'approximation de Satterthwaite.

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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 estimateurs  $\hat{\mu}, \hat{\alpha}_1, \dots, \hat{\alpha}_I, \hat{\sigma}_B^2, \hat{\sigma}_C^2, \hat{\sigma}_{AB}^2, \hat{\sigma}_{AC}^2, \hat{\sigma}_{BC}^2, \hat{\sigma}^2$  des paramètres  $\mu, \alpha_1, \dots, \alpha_I, \sigma_B^2, \sigma_C^2, \sigma_{AB}^2, \sigma_{AC}^2, \sigma_{BC}^2, \sigma^2$  du modèle sont donnés par les formules suivantes :

$$\begin{aligned} \hat{\mu} &= Y_{\bullet,\bullet,\bullet} = \bar{y}, \\ \hat{\alpha}_i &= Y_{i,\bullet,\bullet} - \hat{\mu}, \quad 1 \leq i \leq I, \\ \hat{\sigma}_B^2 &= \frac{1}{IK} (S_B^2 - S_{BC}^2), \quad \hat{\sigma}_C^2 = \frac{1}{IJ} (S_C^2 - S_{BC}^2), \\ \hat{\sigma}_{AB}^2 &= \frac{1}{K} (S_{AB}^2 - S_R^2), \\ \hat{\sigma}_{AC}^2 &= \frac{1}{J} (S_{AC}^2 - S_R^2), \\ \hat{\sigma}_{BC}^2 &= \frac{1}{I} (S_{BC}^2 - S_R^2), \\ \hat{\sigma}^2 &= \frac{SC_R}{(I-1)(J-1)(K-1)} = S_R^2. \end{aligned}$$

$$\text{où } S_B^2 = \frac{SC_B}{n_B}, S_C^2 = \frac{SC_C}{n_C}, S_{AB}^2 = \frac{SC_{AB}}{n_{AB}}, S_{AC}^2 = \frac{SC_{AC}}{n_{AC}}, S_{BC}^2 = \frac{SC_{BC}}{n_{BC}} \text{ et } S_R^2 = \frac{SC_R}{n_R}.$$

Ce sont des estimateurs sans biais<sup>10</sup>.

---

10. La notion de biais d'un estimateur est définie dans le chapitre ??.

Les estimations, obtenues pour la liste de données  $\mathbf{y}$  et notées  $\widehat{\mu}(\mathbf{y})$ ,  $\widehat{\alpha}_1(\mathbf{y})$ ,  $\dots$ ,  $\widehat{\alpha}_I(\mathbf{y})$ ,  $\widehat{\sigma}_B^2(\mathbf{y})$ ,  $\widehat{\sigma}_C^2(\mathbf{y})$ ,  $\widehat{\sigma}_{AB}^2(\mathbf{y})$ ,  $\widehat{\sigma}_{AC}^2(\mathbf{y})$ ,  $\widehat{\sigma}_{BC}^2(\mathbf{y})$ ,  $\widehat{\sigma}^2(\mathbf{y})$  des paramètres  $\mu$ ,  $\alpha_1, \dots, \alpha_I$ ,  $\sigma_B^2$ ,  $\sigma_C^2$ ,  $\sigma_{AB}^2$ ,  $\sigma_{AC}^2$ ,  $\sigma_{BC}^2$ ,  $\sigma^2$  du modèle se déduisent des formules ci-dessus :

$$\begin{aligned}\widehat{\mu}(\mathbf{y}) &= y_{\bullet,\bullet,\bullet} = \bar{y}, \\ \widehat{\alpha}_i(\mathbf{y}) &= y_{i,\bullet,\bullet} - \widehat{\mu}(\mathbf{y}), \quad 1 \leq i \leq I, \\ \widehat{\sigma}_B^2(\mathbf{y}) &= \frac{1}{IK} (s_B^2 - s_{BC}^2), \quad \widehat{\sigma}_C^2(\mathbf{y}) = \frac{1}{IJ} (s_C^2 - s_{BC}^2), \\ \widehat{\sigma}_{AB}^2(\mathbf{y}) &= \frac{1}{K} (s_{AB}^2 - s_R^2), \\ \widehat{\sigma}_{AC}^2(\mathbf{y}) &= \frac{1}{J} (s_{AC}^2 - s_R^2), \\ \widehat{\sigma}_{BC}^2(\mathbf{y}) &= \frac{1}{I} (s_{BC}^2 - s_R^2), \\ \widehat{\sigma}^2(\mathbf{y}) &= \frac{s_{CR}}{(I-1)(J-1)(K-1)} = s_R^2.\end{aligned}$$

**Remarque 5.3.1.** Ainsi lorsque deux facteurs sont aléatoires, et que nous cherchons à tester l'existence d'un effet du troisième facteur, ici à effets fixes, nous devons 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.

### 5.3.2. 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  $\gamma_k$  représentent un échantillon de taille  $K$  prélevé dans une population importante. Nous admettrons que les effets des  $C_k$ , les  $\gamma_k$ , sont distribués suivant une loi normale centrée de variance  $\sigma_C^2$ . Pour chacun des couples de modalités  $(A_i, B_j, C_k)$  nous effectuons  $L \geq 2$  mesures d'une réponse  $Y$  qui est une variable continue. Nous notons  $n = I \times J \times K \times L$  le nombre total de mesures ayant été effectuées.

Nous introduisons le modèle :

$$Y_{i,j,k,l} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{i,j} + (\alpha\gamma)_{i,k} + (\beta\gamma)_{j,k} + (\alpha\beta\gamma)_{i,j,k} + \epsilon_{i,j,k,l},$$

$$i = 1 \dots I, \quad j = 1 \dots J, \quad k = 1 \dots K, \quad l = 1 \dots L,$$

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

$$\sum_{i=1}^I (\alpha\beta)_{i,j} = 0, \quad \forall j \in \{1, \dots, J\}, \quad \sum_{j=1}^J (\alpha\beta)_{i,j} = 0, \quad \forall i \in \{1, \dots, I\},$$

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

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. Nous supposons que :

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

ainsi que l'indépendance des effets aléatoires  $\gamma_k$ .

Nous postulons les hypothèses classiques suivantes pour les erreurs :

$$\begin{aligned} \forall (i, j, k, l), 1 \leq i \leq I, 1 \leq j \leq J, 1 \leq k \leq K, 1 \leq l \leq L, \quad \mathcal{L}(\epsilon_{i,j,k,l}) &= \mathcal{N}(0, \sigma^2), \\ \epsilon_{i,j,k,l} &\perp\!\!\!\perp \epsilon_{m,n,o,p} \text{ si } (i, j, k, l) \neq (m, n, o, p) \\ \text{avec } 1 \leq i, m \leq I, 1 \leq j, n \leq J, 1 \leq k, o \leq K \text{ et } 1 \leq l, p \leq L, \end{aligned}$$

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

Dans un modèle mixte restreint, les effets aléatoires croisant des facteurs à effets fixes et à effets aléatoires, ici aussi bien les  $(\alpha\gamma)_{i,k}$  que les  $(\beta\gamma)_{j,k}$  et que les  $(\alpha\beta\gamma)_{i,j,k}$ , ne sont pas mutuellement indépendants à cause des contraintes portant sur leur somme,

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

Ils le sont par contre dès que nous ne les considérons pas tous en même temps.

Certains logiciels travaillent par défaut avec un modèle non-restreint dont la définition est proche mais pas identique à celle du modèle précédent.

Nous introduisons le modèle :

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

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

sons que :

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

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

$$\begin{aligned} \forall (i, j, k, l), 1 \leq i \leq I, 1 \leq j \leq J, 1 \leq k \leq K, 1 \leq l \leq L, \quad \mathcal{L}(\epsilon_{i,j,k,l}) &= \mathcal{N}(0, \sigma^2), \\ \epsilon_{i,j,k,l} &\perp\!\!\!\perp \epsilon_{m,n,o,p} \text{ si } (i, j, k, l) \neq (m, n, o, p) \\ \text{avec } 1 \leq i, m \leq I, 1 \leq j, n \leq J, 1 \leq k, o \leq K \text{ et } 1 \leq l, p \leq L, \end{aligned}$$

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

Dans un modèle mixte non restreint, les effets aléatoires croisant des facteurs à effets fixes et à effets aléatoires, ici les  $(\alpha\gamma)_{i,k}$ , les  $(\beta\gamma)_{j,k}$  et les  $(\alpha\beta\gamma)_{i,j,k}$ , sont mutuellement indépendants.

Il n'existe pas consensus sur une raison statistique qui permettrait de privilégier plutôt l'une ou l'autre de ces deux approches. Nous utiliserons toujours des modèles restreints. Nous supposons que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fait l'objet d'un paragraphe du chapitre ??.

Nous utilisons 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 5.1.2.

Nous rappelons la relation fondamentale de l'ANOVA :

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

Nous introduisons les degrés de liberté (Ddl) associés à chaque ligne du tableau de l'ANOVA :

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

Nous résumons ces informations dans le tableau de l'ANOVA ci-dessous :

### 5.3 Modèles à effets mixtes

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	$sc_{TOT}$	$n_{TOT}$			

Nous souhaitons faire les tests d'hypothèse suivants :

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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, nous pouvons procéder à des comparaisons multiples des différents effets des niveaux du facteur voir le chapitre 9.

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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, nous pouvons procéder à des comparaisons multiples des différents effets des niveaux du facteur voir le chapitre 9.

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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.

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

contre

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

Sous l'hypothèse nulle  $\mathcal{H}^{(4)}_0$  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  $(I - 1)(J - 1)(K - 1)$  degrés de liberté. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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.



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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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.

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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.

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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 estimateurs  $\hat{\mu}$ ,  $\hat{\alpha}_1, \dots, \hat{\alpha}_I$ ,  $\hat{\beta}_1, \dots, \hat{\beta}_J$ ,  $\hat{\sigma}_C^2$ ,  $\widehat{(\alpha\beta)}_{1,1}, \dots, \widehat{(\alpha\beta)}_{I,J}$ ,  $\hat{\sigma}_{AC}^2$ ,  $\hat{\sigma}_{BC}^2$ ,  $\hat{\sigma}_{ABC}^2$ ,  $\hat{\sigma}^2$  des paramètres  $\mu$ ,  $\alpha_1, \dots, \alpha_I$ ,  $\beta_1, \dots, \beta_J$ ,  $\sigma_C^2$ ,  $(\alpha\beta)_{1,1}, \dots, (\alpha\beta)_{I,J}$ ,  $\sigma_{AC}^2$ ,  $\sigma_{BC}^2$ ,  $\sigma_{ABC}^2$ ,  $\sigma^2$

du modèle sont donnés par les formules suivantes :

$$\begin{aligned}
 \widehat{\mu} &= Y_{\bullet,\bullet,\bullet,\bullet} = \bar{Y}, \\
 \widehat{\alpha}_i &= Y_{i,\bullet,\bullet,\bullet} - \widehat{\mu}, \quad 1 \leq i \leq I, \quad \widehat{\beta}_j = Y_{\bullet,j,\bullet,\bullet} - \widehat{\mu}, \quad 1 \leq j \leq J, \\
 \widehat{\sigma}_C^2 &= \frac{1}{IJL} (S_C^2 - S_R^2), \\
 \widehat{(\alpha\beta)}_{i,j} &= Y_{i,j,\bullet,\bullet} - Y_{i,\bullet,\bullet,\bullet} - Y_{\bullet,j,\bullet,\bullet} + \widehat{\mu}, \quad 1 \leq i \leq I, \quad 1 \leq j \leq J, \\
 \widehat{\sigma}_{AC}^2 &= \frac{1}{JL} (S_{AC}^2 - S_R^2), \\
 \widehat{\sigma}_{BC}^2 &= \frac{1}{IL} (S_{BC}^2 - S_R^2), \\
 \widehat{\sigma}_{ABC}^2 &= \frac{1}{L} (S_{ABC}^2 - S_R^2), \\
 \widehat{\sigma}^2 &= \frac{SC_R}{IJK(L-1)} = S_R^2.
 \end{aligned}$$

où  $S_C^2 = \frac{SC_C}{n_C}$ ,  $S_{AC}^2 = \frac{SC_{AC}}{n_{AC}}$ ,  $S_{BC}^2 = \frac{SC_{BC}}{n_{BC}}$ ,  $S_{ABC}^2 = \frac{SC_{ABC}}{n_{ABC}}$  et  $S_R^2 = \frac{SC_R}{n_R}$ .

Ce sont des estimateurs sans biais<sup>11</sup>.

Les estimations, obtenues pour la liste de données  $\mathbf{y}$  et notées  $\widehat{\mu}(\mathbf{y})$ ,  $\widehat{\alpha}_1(\mathbf{y})$ ,  $\dots$ ,  $\widehat{\alpha}_I(\mathbf{y})$ ,  $\widehat{\beta}_1(\mathbf{y})$ ,  $\dots$ ,  $\widehat{\beta}_J(\mathbf{y})$ ,  $\widehat{\sigma}_C^2(\mathbf{y})$ ,  $\widehat{(\alpha\beta)}_{1,1}(\mathbf{y})$ ,  $\dots$ ,  $\widehat{(\alpha\beta)}_{I,J}(\mathbf{y})$ ,  $\widehat{\sigma}_{AC}^2(\mathbf{y})$ ,  $\widehat{\sigma}_{BC}^2(\mathbf{y})$ ,  $\widehat{\sigma}_{ABC}^2(\mathbf{y})$ ,  $\widehat{\sigma}^2(\mathbf{y})$  des paramètres  $\mu$ ,  $\alpha_1$ ,  $\dots$ ,  $\alpha_I$ ,  $\beta_1$ ,  $\dots$ ,  $\beta_J$ ,  $\sigma_C^2$ ,  $(\alpha\beta)_{1,1}$ ,  $\dots$ ,  $(\alpha\beta)_{I,J}$ ,  $\sigma_{AC}^2$ ,  $\sigma_{BC}^2$ ,  $\sigma_{ABC}^2$ ,  $\sigma^2$  du modèle se déduisent des formules ci-dessus :

$$\begin{aligned}
 \widehat{\mu}(\mathbf{y}) &= y_{\bullet,\bullet,\bullet,\bullet} = \bar{y}, \\
 \widehat{\alpha}_i(\mathbf{y}) &= y_{i,\bullet,\bullet,\bullet} - \widehat{\mu}(\mathbf{y}), \quad 1 \leq i \leq I, \quad \widehat{\beta}_j(\mathbf{y}) = y_{\bullet,j,\bullet,\bullet} - \widehat{\mu}(\mathbf{y}), \quad 1 \leq j \leq J, \\
 \widehat{\sigma}_C^2(\mathbf{y}) &= \frac{1}{IJL} (s_C^2 - s_R^2), \\
 \widehat{(\alpha\beta)}_{i,j}(\mathbf{y}) &= y_{i,j,\bullet,\bullet} - y_{i,\bullet,\bullet,\bullet} - y_{\bullet,j,\bullet,\bullet} + \widehat{\mu}, \quad 1 \leq i \leq I, \quad 1 \leq j \leq J, \\
 \widehat{\sigma}_{AC}^2(\mathbf{y}) &= \frac{1}{JL} (s_{AC}^2 - s_R^2), \\
 \widehat{\sigma}_{BC}^2(\mathbf{y}) &= \frac{1}{IL} (s_{BC}^2 - s_R^2), \\
 \widehat{\sigma}_{ABC}^2(\mathbf{y}) &= \frac{1}{L} (s_{ABC}^2 - s_R^2), \\
 \widehat{\sigma}^2(\mathbf{y}) &= \frac{SC_R}{IJK(L-1)} = s_R^2.
 \end{aligned}$$

**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  $\beta_j$  représentent un échantillon de taille  $J$  prélevé dans une population importante. Nous admettrons que les effets des  $B_j$ , les  $\beta_j$ , sont distribués suivant une loi normale centrée de variance  $\sigma_B^2$ . Les  $\gamma_k$  représentent un échantillon de taille  $K$  prélevé dans une

---

11. La notion de biais d'un estimateur est définie dans le chapitre ??.

### 5.3 Modèles à effets mixtes

population importante. Nous admettrons que les effets des  $C_k$ , les  $\gamma_k$ , sont distribués suivant une loi normale centrée de variance  $\sigma_C^2$ . Pour chacun des couples de modalités  $(A_i, B_j, C_k)$  nous effectuons  $L \geq 2$  mesures d'une réponse  $Y$  qui est une variable continue. Nous notons  $n = I \times J \times K \times L$  le nombre total de mesures ayant été effectuées.

Nous introduisons le modèle :

$$Y_{i,j,k,l} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{i,j} + (\alpha\gamma)_{i,k} + (\beta\gamma)_{j,k} + (\alpha\beta\gamma)_{i,j,k} + \epsilon_{i,j,k,l},$$

$$i = 1 \dots I, j = 1 \dots J, k = 1 \dots K, l = 1 \dots L,$$

$$\text{avec les contraintes supplémentaires } \sum_{i=1}^I \alpha_i = 0,$$

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

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

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

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

ainsi que l'indépendance des effets aléatoires  $\beta_j$ ,  $\gamma_k$  et  $(\beta\gamma)_{j,k}$ .

Nous postulons les hypothèses classiques suivantes pour les erreurs :

$$\forall (i, j, k, l), 1 \leq i \leq I, 1 \leq j \leq J, 1 \leq k \leq K, 1 \leq l \leq L, \mathcal{L}(\epsilon_{i,j,k,l}) = \mathcal{N}(0, \sigma^2),$$

$$\epsilon_{i,j,k,l} \perp\!\!\!\perp \epsilon_{m,n,o,p} \text{ si } (i, j, k, l) \neq (m, n, o, p)$$

$$\text{avec } 1 \leq i, m \leq I, 1 \leq j, n \leq J, 1 \leq k, o \leq K \text{ et } 1 \leq l, p \leq L,$$

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

Dans un modèle mixte restreint, les effets aléatoires croisant des facteurs à effets fixes et à effets aléatoires, ici aussi bien les  $(\alpha\beta)_{i,j}$  que les  $(\alpha\gamma)_{i,k}$  et que les  $(\alpha\beta\gamma)_{i,j,k}$ , ne sont pas mutuellement indépendants à cause des contraintes portant sur leur somme,

$\sum_{i=1}^I (\alpha\beta)_{i,j} = 0, \forall j \in \{1, \dots, J\}, \sum_{i=1}^I (\alpha\gamma)_{i,k} = 0, \forall k \in \{1, \dots, K\}$  et  $\sum_{i=1}^I (\alpha\beta\gamma)_{i,j,k} = 0, \forall (j, k) \in \{1, \dots, J\} \times \{1, \dots, K\}$ . Ils le sont par contre dès que nous ne les considérons pas tous en même temps.

Certains logiciels travaillent par défaut avec un modèle non-restreint dont la définition est proche mais pas identique à celle du modèle précédent.

Nous introduisons 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  $\sum_{i=1}^I \alpha_i = 0,$

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

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

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

$$\begin{aligned} \forall (i, j, k, l), 1 \leq i \leq I, 1 \leq j \leq J, 1 \leq k \leq K, 1 \leq l \leq L, \quad \mathcal{L}(\epsilon_{i,j,k,l}) &= \mathcal{N}(0, \sigma^2), \\ \epsilon_{i,j,k,l} &\perp\!\!\!\perp \epsilon_{m,n,o,p} \text{ si } (i, j, k, l) \neq (m, n, o, p) \\ \text{avec } 1 \leq i, m \leq I, 1 \leq j, n \leq J, 1 \leq k, o \leq K \text{ et } 1 \leq l, p \leq L, \end{aligned}$$

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

Dans un modèle mixte non restreint, les effets aléatoires croisant des facteurs à effets fixes et à effets aléatoires, ici les  $(\alpha\beta)_{i,j}, (\alpha\gamma)_{i,k}$  et les  $(\alpha\beta\gamma)_{i,j,k}$ , sont mutuellement indépendants.

Il n'existe pas consensus sur une raison statistique qui permettrait de privilégier plutôt l'une ou l'autre de ces deux approches. Nous utiliserons toujours des modèles restreints. Nous supposons que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fait l'objet d'un paragraphe du chapitre ??.

Nous utilisons 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 5.1.2.

Nous rappelons la relation fondamentale de l'ANOVA :

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

Nous introduisons les degrés de liberté (Ddl) associés à chaque ligne du tableau de l'ANOVA :

### 5.3 Modèles à effets mixtes

---

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	$n_{TOT} = IJK - 1$

Nous résumons ces informations dans le tableau de l'ANOVA ci-dessous :

**Chapitre 5. Analyse de la variance à trois facteurs**

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 $AB$	$sc_{AB}$	$n_{AB}$	$s_{AB}^2 = \frac{sc_{AB}}{n_{AB}}$	$f_{AB} = \frac{s_{AB}^2}{s_{ABC}^2}$	$\mathcal{H}^{(4)}_0$ ou $\mathcal{H}^{(4)}_1$
Interaction $AC$	$sc_{AC}$	$n_{AC}$	$s_{AC}^2 = \frac{sc_{AC}}{n_{AC}}$	$f_{AC} = \frac{s_{AC}^2}{s_{ABC}^2}$	$\mathcal{H}^{(5)}_0$ ou $\mathcal{H}^{(5)}_1$
Interaction $BC$	$sc_{BC}$	$n_{BC}$	$s_{BC}^2 = \frac{sc_{BC}}{n_{BC}}$	$f_{BC} = \frac{s_{BC}^2}{s_R^2}$	$\mathcal{H}^{(6)}_0$ ou $\mathcal{H}^{(6)}_1$
Interaction $ABC$	$sc_{ABC}$	$n_{ABC}$	$s_{ABC}^2 = \frac{sc_{ABC}}{n_{ABC}}$	$f_{ABC} = \frac{s_{ABC}^2}{s_R^2}$	$\mathcal{H}^{(7)}_0$ ou $\mathcal{H}^{(7)}_1$
Résiduelle	$sc_R$	$n_R$	$s_R^2 = \frac{sc_R}{n_R}$		
Totale	$sc_{TOT}$	$n_{TOT}$			

Nous souhaitons faire les tests d'hypothèse suivants :

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

contre

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

### 5.3 Modèles à effets mixtes

---

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**<sup>12</sup> une loi de Fisher à  $I - 1$  et  $n'$  degrés de liberté avec :

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

Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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, nous pouvons procéder à des comparaisons multiples des différents effets des niveaux du facteur voir le chapitre 9.

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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.

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

contre

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

Sous l'hypothèse nulle  $\mathcal{H}^{(4)}_0$  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  $(I - 1)(J - 1)(K - 1)$

---

12. Nous utilisons toujours l'approximation de Satterthwaite.

degrés de liberté. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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.



### 5.3 Modèles à effets mixtes

Les estimateurs  $\widehat{\mu}, \widehat{\alpha}_1, \dots, \widehat{\alpha}_I, \widehat{\sigma}_B^2, \widehat{\sigma}_C^2, \widehat{\sigma}_{AB}^2, \widehat{\sigma}_{AC}^2, \widehat{\sigma}_{BC}^2, \widehat{\sigma}_{ABC}^2, \widehat{\sigma}^2$  des paramètres  $\mu, \alpha_1, \dots, \alpha_I, \sigma_B^2, \sigma_C^2, \sigma_{AB}^2, \sigma_{AC}^2, \sigma_{BC}^2, \sigma_{ABC}^2, \sigma^2$  du modèle sont donnés par les formules suivantes :

$$\begin{aligned}\widehat{\mu} &= Y_{\bullet, \bullet, \bullet, \bullet} = \bar{Y}, \\ \widehat{\alpha}_i &= Y_{i, \bullet, \bullet, \bullet} - \widehat{\mu}, \quad 1 \leq i \leq I, \\ \widehat{\sigma}_B^2 &= \frac{1}{IKL} (S_B^2 - S_{BC}^2), \quad \widehat{\sigma}_C^2 = \frac{1}{IJL} (S_C^2 - S_{BC}^2), \\ \widehat{\sigma}_{AB}^2 &= \frac{1}{KL} (S_{AB}^2 - S_{ABC}^2), \\ \widehat{\sigma}_{AC}^2 &= \frac{1}{JL} (S_{AC}^2 - S_{ABC}^2), \\ \widehat{\sigma}_{BC}^2 &= \frac{1}{IL} (S_{BC}^2 - S_R^2), \\ \widehat{\sigma}_{ABC}^2 &= \frac{1}{L} (S_{ABC}^2 - S_R^2), \\ \widehat{\sigma}^2 &= \frac{SC_R}{IJK(L-1)} = S_R^2.\end{aligned}$$

où  $S_B^2 = \frac{SC_B}{n_B}$ ,  $S_C^2 = \frac{SC_C}{n_C}$ ,  $S_{AB}^2 = \frac{SC_{AB}}{n_{AB}}$ ,  $S_{AC}^2 = \frac{SC_{AC}}{n_{AC}}$ ,  $S_{BC}^2 = \frac{SC_{BC}}{n_{BC}}$ ,  $S_{ABC}^2 = \frac{SC_{ABC}}{n_{ABC}}$   
et  $S_R^2 = \frac{SC_R}{n_R}$ .

Ce sont des estimateurs sans biais<sup>13</sup>.

Les estimations, obtenues pour la liste de données  $\mathbf{y}$  et notées  $\widehat{\mu}(\mathbf{y}), \widehat{\alpha}_1(\mathbf{y}), \dots, \widehat{\alpha}_I(\mathbf{y}), \widehat{\sigma}_B^2(\mathbf{y}), \widehat{\sigma}_C^2(\mathbf{y}), \widehat{\sigma}_{AB}^2(\mathbf{y}), \widehat{\sigma}_{AC}^2(\mathbf{y}), \widehat{\sigma}_{BC}^2(\mathbf{y}), \widehat{\sigma}_{ABC}^2(\mathbf{y}), \widehat{\sigma}^2(\mathbf{y})$  des paramètres  $\mu, \alpha_1, \dots, \alpha_I, \sigma_B^2, \sigma_C^2, \sigma_{AB}^2, \sigma_{AC}^2, \sigma_{BC}^2, \sigma_{ABC}^2, \sigma^2$  du modèle se déduisent des formules ci-dessus :

$$\begin{aligned}\widehat{\mu}(\mathbf{y}) &= y_{\bullet, \bullet, \bullet, \bullet} = \bar{y}, \\ \widehat{\alpha}_i(\mathbf{y}) &= y_{i, \bullet, \bullet, \bullet} - \widehat{\mu}(\mathbf{y}), \quad 1 \leq i \leq I, \\ \widehat{\sigma}_B^2(\mathbf{y}) &= \frac{1}{IKL} (s_B^2 - s_{BC}^2), \quad \widehat{\sigma}_C^2(\mathbf{y}) = \frac{1}{IJL} (s_C^2 - s_{BC}^2), \\ \widehat{\sigma}_{AB}^2(\mathbf{y}) &= \frac{1}{KL} (s_{AB}^2 - s_{ABC}^2), \\ \widehat{\sigma}_{AC}^2(\mathbf{y}) &= \frac{1}{JL} (s_{AC}^2 - s_{ABC}^2), \\ \widehat{\sigma}_{BC}^2(\mathbf{y}) &= \frac{1}{IL} (s_{BC}^2 - s_R^2), \\ \widehat{\sigma}_{ABC}^2(\mathbf{y}) &= \frac{1}{L} (s_{ABC}^2 - s_R^2), \\ \widehat{\sigma}^2(\mathbf{y}) &= \frac{s_{CR}}{IJK(L-1)} = s_R^2.\end{aligned}$$

13. La notion de biais d'un estimateur est définie dans le chapitre ??.



# Chapitre 6

## Analyse de la variance à trois facteurs totalement emboîtés

Dans tout ce chapitre, nous utilisons les notations définies à la section Notations, page ix. Nous sommes dans la situation particulière, semblable à celle du chapitre 4, 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. Nous compliquons la situation en rajoutant un niveau d'imbrication par rapport aux modèles exposés au chapitre 4. 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<sup>1</sup>. Il n'est utilisable que si nous disposons 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 nous devons utiliser pour étudier les données sera l'un de ceux exposés au chapitre 4.

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

### 6.1. Modèles à effets fixes

#### 6.1.1. 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  $(A_i, B_{j(i)}, C_{k(j(i))})$  nous effectuons  $L \geq 2$  mesures d'une réponse  $Y$  qui est une variable continue. Nous notons  $n = I \times J \times K \times L$  le nombre total de mesures ayant été effectuées.

---

1. en anglais *completely nested model* ou *completely hierarchical model*.

Nous introduisons le modèle :

$$Y_{i,j,k,l} = \mu + \alpha_i + \beta_{j(i)} + \gamma_{k(j(i))} + \epsilon_{i,j,k,l},$$

$$i = 1 \dots I, j = 1 \dots J, k = 1 \dots K, l = 1 \dots L,$$

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. Nous postulons les hypothèses classiques suivantes pour les erreurs :

$$\forall (i, j, k, l), 1 \leq i \leq I, 1 \leq j \leq J, 1 \leq k \leq K, 1 \leq l \leq L, \mathcal{L}(\epsilon_{i,j,k,l}) = \mathcal{N}(0, \sigma^2),$$

$$\epsilon_{i,j,k,l} \perp\!\!\!\perp \epsilon_{m,n,o,p} \text{ si } (i, j, k, l) \neq (m, n, o, p)$$

avec  $1 \leq i, m \leq I, 1 \leq j, n \leq J, 1 \leq k, o \leq K$  et  $1 \leq l, p \leq L.$

Nous supposons que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fait l'objet d'un paragraphe du chapitre ??.

Nous regroupons les valeurs que peut prendre la réponse  $Y$  dans les conditions  $(A_i, B_{j(i)}, C_{k(j(i))})$  lors des  $L$  essais dans le tableau 6.1.

Nous rappelons que la variation théorique due au facteur  $A$  est définie par :

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

Nous rappelons que la variation théorique due au facteur  $B$  dans  $A$  est définie par :

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

La variation théorique due au facteur  $C$  dans  $B$  qui lui-même est emboîté 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 théorique est quant à elle définie par :

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

$A_I$			$A_I$		
$B_{1(1)}$	$\dots$	$B_{J(1)}$	$\dots$	$B_{1(I)}$	$B_{J(I)}$
$C_{1(1(1))}$	$\dots$	$C_{K(1(1))}$	$\dots$	$C_{1(I)}$	$C_{K(J(I))}$
$Y_{1,1,1,1}$	$\dots$	$Y_{1,J,K,1}$	$\dots$	$Y_{I,1,1,1}$	$Y_{I,J,K,1}$
$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$	$\vdots$
$Y_{1,1,1,L}$	$\dots$	$Y_{1,J,K,L}$	$\dots$	$Y_{I,1,1,L}$	$Y_{I,J,K,L}$

TABLE 6.1 – Représentation des variables aléatoires de réponse en fonction des niveaux des trois facteurs d'un plan complètement emboîté

$A_I$			$A_I$		
$B_{1(1)}$	$\dots$	$B_{J(1)}$	$\dots$	$B_{1(I)}$	$B_{J(I)}$
$C_{1(1(1))}$	$\dots$	$C_{K(1(1))}$	$\dots$	$C_{1(I)}$	$C_{K(J(I))}$
$y_{1,1,1,1}$	$\dots$	$y_{1,J,K,1}$	$\dots$	$y_{I,1,1,1}$	$y_{I,J,K,1}$
$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$	$\vdots$
$y_{1,1,1,L}$	$\dots$	$y_{1,J,K,L}$	$\dots$	$y_{I,1,1,L}$	$y_{I,J,K,L}$

TABLE 6.2 – Représentation de la liste de données expérimentales  $\mathbf{y}$  en fonction des niveaux des trois facteurs d'un plan complètement emboîté

Enfin la variation totale théorique est égale à :

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

Nous rappelons la relation fondamentale de l'ANOVA :

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

La liste  $\mathbf{y}$  des données expérimentales  $y_{1,1,1,1}, \dots, y_{1,1,1,L}, \dots, y_{I,1,1,1}, \dots, y_{I,1,1,L}, y_{1,2,1,1}, \dots, y_{I,J,1,L}, y_{1,1,2,1}, \dots, y_{I,J,K,L}$  permet de construire le tableau 6.2 qui est une réalisation du tableau 6.1.

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

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

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

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

La variation due au facteur  $C$  dans  $B$  qui lui même est dans  $A$  observée sur la liste de données  $\mathbf{y}$  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 observée sur la liste de données  $\mathbf{y}$  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 observée sur la liste de données  $\mathbf{y}$  est égale à :

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

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

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

Nous introduisons les degrés 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	$n_{TOT} = IJKL - 1$

## 6.1 Modèles à effets fixes

Nous résumons 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 $B$ dans $A$	$sc_{B A}$	$n_{B A}$	$s_{B A}^2 = \frac{sc_{B A}}{n_{B A}}$	$f_{B A} = \frac{s_{B A}^2}{s_R^2}$	$\mathcal{H}''_0$ ou $\mathcal{H}''_1$
Facteur $C$ dans $B$	$sc_{C B A}$	$n_{C B A}$	$s_{C B A}^2 = \frac{sc_{C B A}}{n_{C B A}}$	$f_{C B A} = \frac{s_{C B A}^2}{s_R^2}$	$\mathcal{H}'''_0$ ou $\mathcal{H}'''_1$
Résiduelle	$sc_R$	$n_R$	$s_R^2 = \frac{sc_R}{n_R}$		
Totale	$sc_{TOT}$	$n_{TOT}$			

Nous souhaitons faire les tests d'hypothèse suivants :

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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, nous pouvons procéder à des comparaisons multiples des différents effets des niveaux du facteur voir le chapitre 9.

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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.

$$\mathcal{H}_0''' : \gamma_{1(1(1))} = \gamma_{2(1(1))} = \dots = \gamma_{K(1(1))} = \gamma_{1(1(2))} = \dots = \gamma_{K(J(I))} = 0$$

contre

$$\mathcal{H}_1''' : \exists (i_0, j_0, k_0) \in \{1, 2, \dots, I\} \times \{1, 2, \dots, J\} \times \{1, 2, \dots, K\} \mid \gamma_{k_0(j_0(i_0))} \neq 0.$$

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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 estimateurs  $\widehat{\mu}$ ,  $\widehat{\alpha}_1, \dots, \widehat{\alpha}_I$ ,  $\widehat{\beta}_{1(1)}, \dots, \widehat{\beta}_{J(I)}$ ,  $\widehat{\gamma}_{1(1(1))}, \dots, \widehat{\gamma}_{K(J(I))}$ ,  $\widehat{\sigma}^2$  des paramètres  $\mu, \alpha_1, \dots, \alpha_I, \beta_{1(1)}, \dots, \beta_{J(I)}, \gamma_{1(1(1))}, \dots, \gamma_{K(J(I))}, \sigma^2$  du modèle sont donnés par les formules suivantes :

$$\begin{aligned} \widehat{\mu} &= Y_{\bullet, \bullet, \bullet, \bullet} = \bar{Y}, \\ \widehat{\alpha}_i &= Y_{i, \bullet, \bullet, \bullet} - \widehat{\mu}, \quad 1 \leq i \leq I, \\ \widehat{\beta}_{j(i)} &= Y_{i, j, \bullet, \bullet} - Y_{i, \bullet, \bullet, \bullet}, \quad 1 \leq i \leq I, \quad 1 \leq j \leq J, \\ \widehat{\gamma}_{k(j(i))} &= Y_{i, j, k, \bullet} - Y_{i, j, \bullet, \bullet}, \quad 1 \leq i \leq I, \quad 1 \leq j \leq J, \quad 1 \leq k \leq K, \\ \widehat{\sigma}^2 &= \frac{SC_R}{IJK(L - 1)} = S_R^2. \end{aligned}$$

Ce sont des estimateurs sans biais<sup>2</sup>.

Les estimations, obtenues pour la liste de données  $\mathbf{y}$  et notées  $\widehat{\mu}(\mathbf{y})$ ,  $\widehat{\alpha}_1(\mathbf{y}), \dots, \widehat{\alpha}_I(\mathbf{y})$ ,  $\widehat{\beta}_{1(1)}(\mathbf{y}), \dots, \widehat{\beta}_{J(I)}(\mathbf{y})$ ,  $\widehat{\gamma}_{1(1(1))}(\mathbf{y}), \dots, \widehat{\gamma}_{K(J(I))}(\mathbf{y})$ ,  $\widehat{\sigma}^2(\mathbf{y})$  des paramètres  $\mu, \alpha_1, \dots, \alpha_I, \beta_{1(1)}, \dots, \beta_{J(I)}, \gamma_{1(1(1))}, \dots, \gamma_{K(J(I))}, \sigma^2$  du modèle se déduisent des formules ci-dessus :

$$\begin{aligned} \widehat{\mu}(\mathbf{y}) &= y_{\bullet, \bullet, \bullet, \bullet} = \bar{y}, \\ \widehat{\alpha}_i(\mathbf{y}) &= y_{i, \bullet, \bullet, \bullet} - \widehat{\mu}(\mathbf{y}), \quad 1 \leq i \leq I, \\ \widehat{\beta}_{j(i)}(\mathbf{y}) &= y_{i, j, \bullet, \bullet} - y_{i, \bullet, \bullet, \bullet}, \quad 1 \leq i \leq I, \quad 1 \leq j \leq J, \\ \widehat{\gamma}_{k(j(i))}(\mathbf{y}) &= y_{i, j, k, \bullet} - y_{i, j, \bullet, \bullet}, \quad 1 \leq i \leq I, \quad 1 \leq j \leq J, \quad 1 \leq k \leq K, \\ \widehat{\sigma}^2(\mathbf{y}) &= \frac{SC_R}{IJK(L - 1)} = s_R^2. \end{aligned}$$

2. La notion de biais d'un estimateur est définie dans le chapitre ??.



## 6.2. Modèles à effets aléatoires

### 6.2.1. Avec répétitions

Les  $\alpha_i$  représentent un échantillon de taille  $I$  prélevé dans une population importante. Nous admettrons que les effets des  $A_i$ , les  $\alpha_i$ , sont distribués suivant une loi normale centrée de variance  $\sigma_A^2$ . 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  $\sigma_{C|B|A}^2$ . Pour chacun des couples de modalités  $(A_i, B_{j(i)}, C_{k(j(i))})$  nous effectuons  $L \geq 2$  mesures d'une réponse  $Y$  qui est une variable continue. Nous notons  $n = I \times J \times K \times L$  le nombre total de mesures ayant été effectuées.

Nous introduisons 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. Nous supposons que :

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

$$\mathcal{L}(\beta_{j(i)}) = \mathcal{N}(0, \sigma_{B|A}^2), \quad \forall (i, j), 1 \leq i \leq I, 1 \leq j \leq J,$$

$$\mathcal{L}(\gamma_{k(j(i))}) = \mathcal{N}(0, \sigma_{C|B|A}^2), \quad \forall (i, j, k), 1 \leq i \leq I, 1 \leq j \leq J, 1 \leq k \leq K,$$

ainsi que l'indépendance des effets aléatoires  $\alpha_i$ ,  $\beta_{j(i)}$ , et  $\gamma_{k(j(i))}$ .

Nous postulons les hypothèses classiques suivantes pour les erreurs :

$$\forall (i, j, k, l), 1 \leq i \leq I, 1 \leq j \leq J, 1 \leq k \leq K, 1 \leq l \leq L, \mathcal{L}(\epsilon_{i,j,k,l}) = \mathcal{N}(0, \sigma^2),$$

$$\epsilon_{i,j,k,l} \perp\!\!\!\perp \epsilon_{m,n,o,p} \text{ si } (i, j, k, l) \neq (m, n, o, p)$$

avec  $1 \leq i, m \leq I, 1 \leq j, n \leq J, 1 \leq k, o \leq K$  et  $1 \leq l, p \leq L$ ,

ainsi que l'indépendance des effets aléatoires  $\alpha_i$ ,  $\beta_{j(i)}$ , et  $\gamma_{k(j(i))}$  et des erreurs  $\epsilon_{i,j,k,l}$ .

Nous supposons que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fait l'objet d'un paragraphe du chapitre ??.

Nous utilisons les quantités  $sc_A$ ,  $sc_{B|A}$ ,  $sc_{C|B|A}$ ,  $sc_R$  et  $sc_{TOT}$  introduites à la section 6.1.1.

Nous rappelons la relation fondamentale de l'ANOVA :

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

Nous introduisons les degrés 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	$n_{TOT} = IJKL - 1$

Nous résumons 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	$sc_{TOT}$	$n_{TOT}$			

Nous souhaitons 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. Nous concluons alors à

## 6.2 Modèles à effets aléatoires

---

l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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.

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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.

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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 estimateurs  $\widehat{\mu}$ ,  $\widehat{\sigma}_A^2$ ,  $\widehat{\sigma}_{B|A}^2$ ,  $\widehat{\sigma}_{C|B|A}^2$ ,  $\widehat{\sigma}^2$  du modèle sont donnés par les formules suivantes :

$$\begin{aligned}\widehat{\mu} &= Y_{\bullet, \dots, \bullet} = \bar{Y}, \\ \widehat{\sigma}_A^2 &= \frac{1}{JKL} (S_A^2 - S_{B|A}^2), \\ \widehat{\sigma}_{B|A}^2 &= \frac{1}{KL} (S_{B|A}^2 - S_{C|B|A}^2), \\ \widehat{\sigma}_{C|B|A}^2 &= \frac{1}{L} (S_{C|B|A}^2 - S_R^2), \\ \widehat{\sigma}^2 &= \frac{SC_R}{IJK(L-1)} = S_R^2,\end{aligned}$$

$$\text{où } S_A^2 = \frac{SC_A}{n_A}, S_{B|A}^2 = \frac{SC_{B|A}}{n_{B|A}}, S_{C|B|A}^2 = \frac{SC_{C|B|A}}{n_{C|B|A}} \text{ et } S_R^2 = \frac{SC_R}{n_R}.$$

Ce sont des estimateurs sans biais<sup>3</sup>.

---

3. La notion de biais d'un estimateur est définie dans le chapitre ??.

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

$$\begin{aligned}\widehat{\mu}(\mathbf{y}) &= y_{\bullet,\bullet,\bullet,\bullet} = \bar{y}, \\ \widehat{\sigma}_A^2(\mathbf{y}) &= \frac{1}{JKL} (s_A^2 - s_{B|A}^2), \\ \widehat{\sigma}_{B|A}^2(\mathbf{y}) &= \frac{1}{KL} (s_{B|A}^2 - s_{C|B|A}^2), \\ \widehat{\sigma}_{C|B|A}^2(\mathbf{y}) &= \frac{1}{L} (s_{C|B|A}^2 - s_R^2), \\ \widehat{\sigma}^2(\mathbf{y}) &= \frac{s_{CR}}{IJK(L-1)} = s_R^2.\end{aligned}$$

## 6.3. Modèles à effets mixtes

### 6.3.1. 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  $\sigma_{C|B|A}^2$ . Pour chacun des couples de modalités  $(A_i, B_{j(i)}, C_{k(j(i))})$  nous effectuons  $L \geq 2$  mesures d'une réponse  $Y$  qui est une variable continue. Nous notons  $n = I \times J \times K \times L$  le nombre total de mesures ayant été effectuées.

Nous introduisons 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. Nous supposons que :

$$\mathcal{L}(\gamma_{k(j(i))}) = \mathcal{N}(0, \sigma_{C|B|A}^2), \forall (i, j, k), 1 \leq i \leq I, 1 \leq j \leq J, 1 \leq k \leq K,$$

ainsi que l'indépendance des effets aléatoires  $\gamma_{k(j(i))}$ .

Nous postulons les hypothèses classiques suivantes pour les erreurs :

$$\forall (i, j, k, l), 1 \leq i \leq I, 1 \leq j \leq J, 1 \leq k \leq K, 1 \leq l \leq L, \mathcal{L}(\epsilon_{i,j,k,l}) = \mathcal{N}(0, \sigma^2),$$

$$\epsilon_{i,j,k,l} \perp\!\!\!\perp \epsilon_{m,n,o,p} \text{ si } (i, j, k, l) \neq (m, n, o, p)$$

$$\text{avec } 1 \leq i, m \leq I, 1 \leq j, n \leq J, 1 \leq k, o \leq K \text{ et } 1 \leq l, p \leq L,$$

### 6.3 Modèles à effets mixtes

ainsi que l'indépendance des effets aléatoires  $\gamma_{k(j(i))}$  et des erreurs  $\epsilon_{i,j,k,l}$ .

Nous supposons que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fait l'objet d'un paragraphe du chapitre ??.

Nous utilisons les quantités  $sc_A$ ,  $sc_{B|A}$ ,  $sc_{C|B|A}$ ,  $sc_R$  et  $sc_{TOT}$  introduites à la section 6.1.1.

Nous rappelons la relation fondamentale de l'ANOVA :

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

Nous introduisons les degrés 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	$n_{TOT} = IJKL - 1$

Nous résumons ces informations dans le tableau de l'ANOVA ci-dessous :

Source	Variation	Ddl	Carré Moyen	$F$	Décision
Facteur $A$	$sc_A$	$n_A$	$s_A^2 = \frac{sc_A}{n_A}$	$f_A = \frac{s_A^2}{s_{C B A}^2}$	$\mathcal{H}'_0$ ou $\mathcal{H}'_1$
Facteur $B$ dans $A$	$sc_{B A}$	$n_{B A}$	$s_{B A}^2 = \frac{sc_{B A}}{n_{B A}}$	$f_{B A} = \frac{s_{B A}^2}{s_{C B A}^2}$	$\mathcal{H}''_0$ 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	$sc_{TOT}$	$n_{TOT}$			

Nous souhaitons faire les tests d'hypothèse suivants :

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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, nous pouvons procéder à des comparaisons multiples des différents effets des niveaux du facteur voir le chapitre 9.

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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.

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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.

### 6.3 Modèles à effets mixtes

Les estimateurs  $\widehat{\mu}$ ,  $\widehat{\alpha}_1, \dots, \widehat{\alpha}_I, \widehat{\beta}_{1(1)}, \dots, \widehat{\beta}_{J(I)}, \widehat{\sigma}_{C|B|A}^2, \widehat{\sigma}^2$  des paramètres  $\mu, \alpha_1, \dots, \alpha_I, \beta_{1(1)}, \dots, \beta_{J(I)}, \sigma_{C|B|A}^2, \sigma^2$  du modèle sont donnés par les formules suivantes :

$$\begin{aligned}\widehat{\mu} &= Y_{\bullet, \bullet, \bullet, \bullet} = \bar{Y}, \\ \widehat{\alpha}_i &= Y_{i, \bullet, \bullet, \bullet} - \widehat{\mu}, \quad 1 \leq i \leq I, \\ \widehat{\beta}_{j(i)} &= Y_{i, j, \bullet, \bullet} - Y_{i, \bullet, \bullet, \bullet}, \quad 1 \leq i \leq I, \quad 1 \leq j \leq J, \\ \widehat{\sigma}_{C|B|A}^2 &= \frac{1}{L} (S_{C|B|A}^2 - S_R^2), \\ \widehat{\sigma}^2 &= \frac{SC_R}{IJK(L-1)} = S_R^2,\end{aligned}$$

où  $S_{C|B|A}^2 = \frac{SC_{C|B|A}}{n_{C|B|A}}$  et  $S_R^2 = \frac{SC_R}{n_R}$ .

Ce sont des estimateurs sans biais<sup>4</sup>.

Les estimations, obtenues pour la liste de données  $\mathbf{y}$  et notées  $\widehat{\mu}(\mathbf{y}), \widehat{\alpha}_1(\mathbf{y}), \dots, \widehat{\alpha}_I(\mathbf{y}), \widehat{\beta}_{1(1)}(\mathbf{y}), \dots, \widehat{\beta}_{J(I)}(\mathbf{y}), \widehat{\sigma}_{C|B|A}^2(\mathbf{y}), \widehat{\sigma}^2(\mathbf{y})$  des paramètres  $\mu, \alpha_1, \dots, \alpha_I, \beta_{1(1)}, \dots, \beta_{J(I)}, \sigma_{C|B|A}^2, \sigma^2$  du modèle se déduisent des formules ci-dessus :

$$\begin{aligned}\widehat{\mu}(\mathbf{y}) &= y_{\bullet, \bullet, \bullet, \bullet} = \bar{y}, \\ \widehat{\alpha}_i(\mathbf{y}) &= y_{i, \bullet, \bullet, \bullet} - \widehat{\mu}, \quad 1 \leq i \leq I, \\ \widehat{\beta}_{j(i)}(\mathbf{y}) &= y_{i, j, \bullet, \bullet} - y_{i, \bullet, \bullet, \bullet}, \quad 1 \leq i \leq I, \quad 1 \leq j \leq J, \\ \widehat{\sigma}_{C|B|A}^2(\mathbf{y}) &= \frac{1}{L} (s_{C|B|A}^2 - s_R^2), \\ \widehat{\sigma}^2(\mathbf{y}) &= \frac{s_{C_R}}{IJK(L-1)} = s_R^2.\end{aligned}$$

**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  $\beta_{j(i)}$  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  $\sigma_{C|B|A}^2$ . Pour chacun des couples de modalités  $(A_i, B_{j(i)}, C_{k(j(i))})$  nous effectuons  $L \geq 2$  mesures d'une réponse  $Y$  qui est une variable continue. Nous notons  $n = I \times J \times K \times L$  le nombre total de mesures ayant été effectuées.

Nous introduisons le modèle :

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

4. La notion de biais d'un estimateur est définie dans le chapitre ??.

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. Nous supposons que :

$$\begin{aligned} \mathcal{L}(\beta_{j(i)}) &= \mathcal{N}(0, \sigma_{B|A}^2), \quad \forall (i, j), 1 \leq i \leq I, 1 \leq j \leq J, \\ \mathcal{L}(\gamma_{k(j(i))}) &= \mathcal{N}(0, \sigma_{C|B|A}^2), \quad \forall (i, j, k), 1 \leq i \leq I, 1 \leq j \leq J, 1 \leq k \leq K, \end{aligned}$$

ainsi que l'indépendance des effets aléatoires  $\beta_{j(i)}$ , et  $\gamma_{k(j(i))}$ .

Nous postulons les hypothèses classiques suivantes pour les erreurs :

$$\begin{aligned} \forall (i, j, k, l), 1 \leq i \leq I, 1 \leq j \leq J, 1 \leq k \leq K, 1 \leq l \leq L, \quad \mathcal{L}(\epsilon_{i,j,k,l}) &= \mathcal{N}(0, \sigma^2), \\ \epsilon_{i,j,k,l} &\perp\!\!\!\perp \epsilon_{m,n,o,p} \text{ si } (i, j, k, l) \neq (m, n, o, p) \\ \text{avec } 1 \leq i, m \leq I, 1 \leq j, n \leq J, 1 \leq k, o \leq K \text{ et } 1 \leq l, p \leq L, \end{aligned}$$

ainsi que l'indépendance des effets aléatoires  $\beta_{j(i)}$ , et  $\gamma_{k(j(i))}$  et des erreurs  $\epsilon_{i,j,k,l}$ .

Nous supposons que les conditions d'utilisation de ce modèle sont bien remplies, l'étude de leur vérification fait l'objet d'un paragraphe du chapitre ??.

Nous utilisons les quantités  $sc_A$ ,  $sc_{B|A}$ ,  $sc_{C|B|A}$ ,  $sc_R$  et  $sc_{TOT}$  introduites à la section 6.1.1.

Nous rappelons la relation fondamentale de l'ANOVA :

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

Nous introduisons les degrés 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	$n_{TOT} = IJKL - 1$

Nous résumons ces informations dans le tableau de l'ANOVA ci-dessous :



### 6.3 Modèles à effets mixtes

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	$sc_{TOT}$	$n_{TOT}$			

Nous souhaitons faire les tests d'hypothèse suivants :

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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, nous pouvons procéder à des comparaisons multiples des différents effets des niveaux du facteur voir le chapitre 9.

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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.

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

contre

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

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. Nous concluons alors à l'aide de la  $p$ -valeur, rejet si elle est inférieure ou égale au seuil  $\alpha$  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 estimateurs  $\widehat{\mu}, \widehat{\alpha}_1, \dots, \widehat{\alpha}_I, \widehat{\sigma}_{B|A}^2, \widehat{\sigma}_{C|B|A}^2, \widehat{\sigma}^2$  des paramètres  $\mu, \alpha_1, \dots, \alpha_I, \sigma_{B|A}^2, \sigma_{C|B|A}^2, \sigma^2$  du modèle sont donnés par les formules suivantes :

$$\begin{aligned} \widehat{\mu} &= Y_{\bullet, \bullet, \bullet, \bullet} = \bar{Y}, \\ \widehat{\alpha}_i &= Y_{i, \bullet, \bullet, \bullet} - \widehat{\mu}, \quad 1 \leq i \leq I, \\ \widehat{\sigma}_{B|A}^2 &= \frac{1}{KL} (S_{B|A}^2 - S_{C|B|A}^2), \\ \widehat{\sigma}_{C|B|A}^2 &= \frac{1}{L} (S_{C|B|A}^2 - S_R^2), \\ \widehat{\sigma}^2 &= \frac{SC_R}{IJK(L-1)} = S_R^2, \end{aligned}$$

où  $S_{B|A}^2 = \frac{SC_{B|A}}{n_{B|A}}, S_{C|B|A}^2 = \frac{SC_{C|B|A}}{n_{C|B|A}}$  et  $S_R^2 = \frac{SC_R}{n_R}$ .

Ce sont des estimateurs sans biais<sup>5</sup>.

Les estimations, obtenues pour la liste de données  $\mathbf{y}$  et notées  $\widehat{\mu}(\mathbf{y}), \widehat{\alpha}_1(\mathbf{y}), \dots, \widehat{\alpha}_I(\mathbf{y}), \widehat{\sigma}_{B|A}^2(\mathbf{y}), \widehat{\sigma}_{C|B|A}^2(\mathbf{y}), \widehat{\sigma}^2(\mathbf{y})$  des paramètres  $\mu, \alpha_1, \dots, \alpha_I, \sigma_{B|A}^2, \sigma_{C|B|A}^2, \sigma^2$  du modèle se déduisent des formules ci-dessus :

$$\begin{aligned} \widehat{\mu}(\mathbf{y}) &= y_{\bullet, \bullet, \bullet, \bullet} = \bar{y}, \\ \widehat{\alpha}_i(\mathbf{y}) &= y_{i, \bullet, \bullet, \bullet} - \widehat{\mu}(\mathbf{y}), \quad 1 \leq i \leq I, \\ \widehat{\sigma}_{B|A}^2(\mathbf{y}) &= \frac{1}{KL} (s_{B|A}^2 - s_{C|B|A}^2), \\ \widehat{\sigma}_{C|B|A}^2(\mathbf{y}) &= \frac{1}{L} (s_{C|B|A}^2 - s_R^2), \\ \widehat{\sigma}^2(\mathbf{y}) &= \frac{SC_R}{IJK(L-1)} = s_R^2. \end{aligned}$$

5. La notion de biais d'un estimateur est définie dans le chapitre ??.

# Chapitre 7

## Analyse de la variance à trois facteurs partiellement emboîtés

### 7.1. Introduction

Un modèle d'analyse de la variance est dit partiellement emboîté ou partiellement hiérarchisé si, contrairement aux modèles complètement emboîtés présentés aux chapitres 4 et 6, ce modèle fait intervenir certaines interactions. Nous remarquons 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 nous pouvons obtenir ainsi, nous nous bornerons à exposer ci-dessous les équations des différents modèles comportant trois facteurs et partiellement emboîtés qu'il est possible de 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, nous renvoyons le lecteur à [Dag98b].

### 7.2. Les différents modèles partiellement emboîtés

#### 7.2.1. 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 nous considérons 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  $1 \leq i \leq I$ ,  $1 \leq j \leq J$ ,  $1 \leq k \leq K$ ,  $1 \leq l \leq L$ .

Pour illustrer les deux règles exposées à la fin de la section 7.1 concernant les effets fixes ou aléatoires, nous détaillons les différentes déclinaisons de modèles que nous pouvons 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, nous supposons que le facteur  $A$  est à effets fixes et que c'est le facteur  $B$  qui est à effets aléatoires. Alors **nécessairement** nous devons 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  $\alpha_1, \dots, \alpha_I$  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.

### 7.2.2. 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 nous considérons 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  $1 \leq i \leq I$ ,  $1 \leq j \leq J$ ,  $1 \leq k \leq K$ ,  $1 \leq l \leq L$ . Dans le cas, nous pouvons avoir le nombre de répétitions,  $L$ , égal à un, c'est-à-dire que nous pouvons utiliser un plan d'expérience ne comportant pas de répétitions.

Pour illustrer les deux règles exposées à la fin de la section 7.1 concernant les effets fixes ou aléatoires, nous détaillons les différentes déclinaisons de modèles que nous pouvons 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 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 nous considérons que les facteurs sont aléatoires ou fixes, les conditions habituelles sur les lois, les variances et l'indépendance des erreurs et  $1 \leq i \leq I$ ,  $1 \leq j \leq J$ ,  $1 \leq k \leq K$ ,  $1 \leq l \leq L$ . Dans ce cas, nous pouvons avoir un nombre de répétitions,  $L$ , égal à un, c'est-à-dire que nous pouvons utiliser un plan d'expérience ne comportant pas de répétitions puis l'analyser à l'aide de ce modèle.

Pour illustrer les deux règles exposées à la fin de la section 7.1 concernant les facteurs à effets fixes ou aléatoires, nous détaillons les différentes déclinaisons de modèles que nous pouvons 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.



# Chapitre 8

## Analyse de la variance à $l$ facteurs croisés, partiellement ou totalement emboîtés, $l \geq 3$

Nous donnons dans ce chapitre quelques règles générales qui étendent les indications qui ont été exposées, aux chapitres 2, 3, 4, 5, 6, 7, dans les cas où nous avons 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  $l - 1$ <sup>1</sup> 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 nous faisons justement l'hypothèse que l'interaction d'ordre  $l - 1$  est négligeable.
2. Si le plan comporte des répétitions, nous pouvons 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. Nous pourrions néanmoins nous servir de certains types de modèles partiellement emboîtés.
4. Si le plan comporte des répétitions, nous pouvons utiliser un modèle totalement emboîté.
5. Nous pouvons comme dans les chapitres 2, 3, 5 utiliser des modèles dont certains des facteurs sont à effets aléatoires et d'autres à effets fixes.
6. Lorsque les facteurs dont nous cherchons 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.

---

1. 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, nous devons recourir à une transformation afin d'essayer de limiter son effet. En effet, les statistiques des tests dont nous nous servons sont exactes sous l'hypothèse que l'interaction d'ordre  $l - 1$  est nulle. Pour un exposé de ce test, nous renvoyons au livre [Dag98b].

8. Dans le cas où nous souhaitons 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. Nous rappelons 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, nous conseillons au lecteur de consulter [Dag98b]. Par exemple il existe 48 modèles différents partiellement emboîtés lorsque nous considérons un modèle d'analyse de la variance à quatre facteurs et en présence de répétitions. Nous renvoyons à nouveau le lecteur vers le livre [Dag98b] s'il souhaite obtenir des informations complémentaires.



# Chapitre 9

## Comparaisons multiples

### 9.1. Contrastes

#### 9.1.1. Définition

Pour introduire la notion de contraste, nous considérons le cas d'un modèle d'analyse de la variance pour un facteur  $A$  à effets fixes. Nous notons  $A_i$ , pour  $1 \leq i \leq I$  les modalités contrôlées du facteur  $A$  et  $\alpha_i$  les effets de ces différentes modalités, toujours pour  $1 \leq i \leq I$ .

Nous introduisons le modèle :

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

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. Nous postulons les hypothèses classiques suivantes pour les erreurs :

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

Nous appelons **contraste**  $L$  des  $I$  moyennes  $\mu_1, \dots, \mu_I$  la somme :

$$L = l_1\mu_1 + l_2\mu_2 + \dots + l_I\mu_I,$$

où  $l_1, \dots, l_I$  sont  $I$  nombres réels tels que  $\sum_i l_i = 0$  et  $\mu_1 = \mu + \alpha_1, \dots, \mu_I = \mu + \alpha_I$  sont tels que  $\alpha_1, \dots, \alpha_I$  sont les  $I$  différents effets des  $I$  niveaux du facteur  $A$ .

Les expressions suivantes sont des exemples de contrastes :

$$\mu_1 - \mu_3, \text{ un tel contraste permettra par exemple de comparer } \mu_1 \text{ et } \mu_3.$$
$$\mu_1 - 2\mu_2 + \mu_3, \text{ un tel contraste permettra par exemple de comparer } \mu_1 + \mu_3 \text{ et } 2\mu_2.$$

### 9.1.2. Orthogonalité

Nous considérons deux contrastes  $L_1$  et  $L_2$  :

$$L_1 = l_1\mu_1 + l_2\mu_2 + \dots + l_I\mu_I$$

et

$$L_2 = l'_1\mu_1 + l'_2\mu_2 + \dots + l'_I\mu_I.$$

Nous avons donc les relations  $l_1, \dots, l_I$  sont  $I$  nombres réels tels que  $\sum_i^I l_i = 0$  et  $l'_1, \dots, l'_I$  sont  $I$  nombres réels tels que  $\sum_i^I l'_i = 0$ ,  $\mu_1 = \mu + \alpha_1, \dots, \mu_I = \mu + \alpha_I$  sont tels que  $\alpha_1, \dots, \alpha_I$  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 :

$$l_1l'_1 + l_2l'_2 + \dots + l_I l'_I = 0.$$

Par exemple, les deux contrastes suivant sont des contrastes orthogonaux :

$$\begin{aligned} L_1 &= \mu_1 - \mu_2 \\ L_2 &= \mu_1 + \mu_2 - \mu_3 - \mu_4. \end{aligned}$$

### 9.1.3. Estimation

Soit  $L$  un contraste. Un estimateur sans biais<sup>1</sup>  $\widehat{L}$  de  $L$  est obtenu de la manière suivante :

$$\widehat{L} = l_1\widehat{\mu}_1 + l_2\widehat{\mu}_2 + \dots + l_I\widehat{\mu}_I,$$

où  $\widehat{\mu}_i = \widehat{\mu} + \widehat{\alpha}_i = \widehat{\mu} + \widehat{\alpha}_i$ , avec  $1 \leq i \leq I$ .

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  $A_i$ , nous obtenons la formule suivante :

$$SC_{\widehat{L}} = \frac{\left( \sum_{i=1}^I l_i \widehat{\mu}_i \right)^2}{\sum_{i=1}^I \frac{l_i^2}{n_i}}.$$

1. La notion de biais d'un estimateur est définie dans le chapitre ??.

### 9.1.4. Test d'une hypothèse impliquant un contraste

Pour le modèle considéré de l'analyse de la variance à un facteur, les estimateurs  $Y_{i,\bullet}$  des  $\mu_i$  sont indépendants, même si le plan n'est pas équilibré. Ainsi la variance de l'estimateur  $\widehat{L}$  du contraste  $L$  est égale à :

$$\begin{aligned}\text{Var} [\widehat{L}] &= \sum_{i=1}^I (l_i^2 \text{Var} [Y_{i,\bullet}]) \\ &= \sigma^2 \sum_{i=1}^I \frac{l_i^2}{n_i}.\end{aligned}$$

Un estimateur sans biais de cette variance est alors :

$$\widehat{\text{Var}} [\widehat{L}] = s_R^2 \sum_{i=1}^I \frac{l_i^2}{n_i}.$$

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 :

$$\frac{\widehat{L} - L}{\sqrt{\widehat{\text{Var}} [\widehat{L}]}} \sim t_{n-I},$$

où  $t_{n-I}$  est la loi de Student à  $n - I$  degrés de liberté.

**Remarque :**

$\sqrt{\widehat{\text{Var}} [\widehat{L}]}$  n'est pas un estimateur sans biais de  $\sqrt{\text{Var} [\widehat{L}]}$ . Il est possible de corriger le biais de cet estimateur en posant  $\sqrt{\widehat{\text{Var}} [\widehat{L}]_{\text{corrigé}}}$  défini par :

$$\sqrt{\widehat{\text{Var}} [\widehat{L}]_{\text{corrigé}}} = \sqrt{\frac{n-1}{2} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}} \sqrt{\widehat{\text{Var}} [\widehat{L}]},$$

où  $\Gamma$  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 [Cha73], propose la formule approximative :

$$\sqrt{\widehat{\text{Var}} [\widehat{L}]_{\text{corrigé}}} \approx \sqrt{\widehat{\text{Var}} [\widehat{L}]} \sqrt{\frac{2n-2}{2n-3}} \approx \sqrt{\widehat{\text{Var}} [\widehat{L}]} \left(1 + \frac{1}{4n-6}\right).$$

Nous perdons alors la propriété ci-dessous qui permet de connaître la loi de la statistique

$$\frac{\widehat{L} - L}{\sqrt{\widehat{\text{Var}} [\widehat{L}]}}$$

c'est pourquoi nous préférons utiliser l'estimateur biaisé  $\sqrt{\widehat{\text{Var}}[\widehat{L}]}$  à la place de l'estimateur sans biais  $\sqrt{\widehat{\text{Var}}[\widehat{L}]_{\text{corrigé}}}$  lorsque nous cherchons à obtenir une estimation par intervalle de  $\sqrt{\text{Var}}[\widehat{L}]$  et à la rigueur nous nous servirions de  $\sqrt{\widehat{\text{Var}}[\widehat{L}]_{\text{corrigé}}}$  pour obtenir une estimation ponctuelle de  $\sqrt{\text{Var}}[\widehat{L}]$  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  $\frac{\widehat{L} - L}{\sqrt{\widehat{\text{Var}}[\widehat{L}]}}$  permet de déterminer un intervalle de confiance de niveau  $100(1 - \alpha) \%$  pour la valeur du contraste  $L$ .

Nous nous donnons désormais un nombre réel  $L_0$  et nous souhaitons 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,

$$l = \frac{\widehat{L}(\mathbf{y}) - L_0}{\sqrt{s_R^2 \sum_{i=1}^I \frac{l_i^2}{n_i}}}$$

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  $\alpha$  pour une loi de Student à  $n - I$  degrés de liberté nous pouvons 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.

## 9.2. Comparaisons multiples sous l'hypothèse d'homoscédasticité

Dans le paragraphe 9.1 ci-dessus, nous nous sommes intéressés au test **d'une seule hypothèse** concernant les effets des niveaux du facteur  $A$ . Le contexte des tests de comparaisons multiples<sup>2</sup> est radicalement différent puisque nous cherchons à comparer tous les effets des niveaux  $A_i$  du facteur  $A$  entre eux ou avec un niveau de référence dit de contrôle. Nous devons ainsi réaliser  $I(I - 1)/2$  comparaisons dans la première situation ou  $I - 1$  comparaisons dans la seconde situation où nous comparons les effets à un niveau

2. Ces tests sont souvent appelés *post-hoc tests* ou *multiple comparisons tests* en anglais.

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  $L = \mu_i - \mu_j$ . Nous exposons dans la suite les procédures de test simultané de plusieurs contrastes en gardant à l'esprit que nous appliquerons principalement les résultats au cas où ces contrastes sont des différences de moyennes.

Nous rappelons que nous n'utiliserons l'un des tests de comparaisons multiples que si le facteur étudié est à effets fixes et que nous avons rejeté l'hypothèse nulle d'absence d'effet de ce facteur sur la réponse.

Nous exposons 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  $\alpha$ .

### 9.2.1. 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 9.2.2 sur la méthode de Tukey-Kramer.

Si  $L$  est un contraste dont un estimateur est  $\hat{L}$ , alors un intervalle de confiance de niveau simultané  $100(1 - \alpha) \%$  pour tous les contrastes considérés est donné par la formule suivante :

$$\hat{L}(\mathbf{y}) - T\sqrt{\frac{s_R^2}{J}} \left( \frac{1}{2} \sum_{i=1}^I |l_i| \right) < L < \hat{L}(\mathbf{y}) + T\sqrt{\frac{s_R^2}{J}} \left( \frac{1}{2} \sum_{i=1}^I |l_i| \right),$$

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, nous décidons que le contraste n'est pas significativement différent de 0 au seuil  $\alpha$ . Au contraire si l'intervalle de confiance ne contient pas 0, alors nous décidons que le contraste est significativement différent de 0 au seuil  $\alpha$ .

L'intérêt de cette procédure est que, si nous fixons  $\alpha$ , les intervalles définis ci-dessus sont valables **simultanément** pour tous les contrastes qu'il est possible de construire !

Nous souhaitons tester l'hypothèse

$$\boxed{\mathcal{H}_0 : L = 0}$$

contre

$$\boxed{\mathcal{H}_1 : L \neq 0.}$$

Le test est significatif au seuil  $\alpha$  et nous décidons de rejeter l'hypothèse nulle  $\mathcal{H}_0$  «  $L = 0$  » en faveur de l'hypothèse alternative  $\mathcal{H}_1$  «  $L \neq 0$  » si :

$$\frac{|\widehat{L}(\mathbf{y})|}{T\sqrt{\frac{s_R^2}{J} \left( \frac{1}{2} \sum_{i=1}^I |l_i| \right)}} \geq q(I, I(J-1); 1-\alpha).$$

Le test n'est pas significatif au seuil  $\alpha$  et nous décidons de conserver par défaut l'hypothèse nulle  $\mathcal{H}_0$  «  $L = 0$  » si :

$$\frac{|\widehat{L}(\mathbf{y})|}{T\sqrt{\frac{s_R^2}{J} \left( \frac{1}{2} \sum_{i=1}^I |l_i| \right)}} < q(I, I(J-1); 1-\alpha).$$

Appliqué au contexte des comparaisons multiples<sup>3</sup>, 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 :

$$\widehat{\mu}_i(\mathbf{y}) - \widehat{\mu}_{i'}(\mathbf{y}) - T\sqrt{\frac{s_R^2}{J}} < \mu_i - \mu_{i'} < \widehat{\mu}_i(\mathbf{y}) - \widehat{\mu}_{i'}(\mathbf{y}) + T\sqrt{\frac{s_R^2}{J}}.$$

Les hypothèses et les statistiques des tests se transforment, quant à elles, en :

$$\boxed{\mathcal{H}_0 : \mu_i = \mu_{i'}}$$

contre

$$\boxed{\mathcal{H}_1 : \mu_i \neq \mu_{i'}}.$$

Nous rappelons 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  $\alpha$  et nous décidons 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 :

$$\frac{|\widehat{\mu}_i(\mathbf{y}) - \widehat{\mu}_{i'}(\mathbf{y})|}{\sqrt{\frac{s_R^2}{J}}} \geq q(I, I(J-1); 1-\alpha).$$

3. Cette procédure est souvent appelée Tukey's HSD, pour *Tukey's Honestly Significance Difference*.

## 9.2 Comparaisons multiples sous l'hypothèse d'homoscédasticité

---

Le test n'est pas significatif au seuil  $\alpha$  et nous décidons de conserver par défaut l'hypothèse nulle  $\mathcal{H}_0$  «  $\mu_i = \mu_{i'}$  » si :

$$\frac{|\widehat{\mu}_i(\mathbf{y}) - \widehat{\mu}_{i'}(\mathbf{y})|}{\sqrt{\frac{s_R^2}{J}}} < q(I, I(J-1); 1-\alpha).$$

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  $A$ , **nous sommes assurés** que la probabilité qu'aucune des différences n'est significative est exactement de  $1-\alpha$ .

### 9.2.2. 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 nous voulons comparer les deux moyennes  $\mu_i$  et  $\mu_{i'}$ , nous remplaçons 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<sup>4</sup> 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 :

$$\widehat{\mu}_i(\mathbf{y}) - \widehat{\mu}_{i'}(\mathbf{y}) - T \sqrt{\frac{s_R^2}{2} \left( \frac{1}{n_i} + \frac{1}{n_{i'}} \right)} < \mu_i - \mu_{i'} < \widehat{\mu}_i(\mathbf{y}) - \widehat{\mu}_{i'}(\mathbf{y}) + T \sqrt{\frac{s_R^2}{2} \left( \frac{1}{n_i} + \frac{1}{n_{i'}} \right)},$$

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

$$\boxed{\mathcal{H}_1 : \mu_i \neq \mu_{i'}}$$

Nous rappelons que ce jeu d'hypothèses est équivalent à celui-ci :

$$\boxed{\mathcal{H}_0 : \alpha_i = \alpha_{i'}}$$

contre

$$\boxed{\mathcal{H}_1 : \alpha_i \neq \alpha_{i'}}$$

---

4. Nous rappelons que la moyenne harmonique  $\text{Harm}(a, b)$  de deux nombres réels strictement positifs  $a$  et  $b$  est définie par :

$$\text{Harm}(a, b) = \frac{1}{\frac{1}{2} \left( \frac{1}{a} + \frac{1}{b} \right)}$$

Le test est significatif au seuil  $\alpha$  et nous décidons 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 :

$$\frac{|\widehat{\mu}_i(\mathbf{y}) - \widehat{\mu}_{i'}(\mathbf{y})|}{\sqrt{\frac{s_R^2}{2} \left( \frac{1}{n_i} + \frac{1}{n_{i'}} \right)}} \geq q(I, n - I; 1 - \alpha).$$

Le test n'est pas significatif au seuil  $\alpha$  et nous décidons de conserver par défaut l'hypothèse nulle  $\mathcal{H}_0$  «  $\mu_i = \mu_{i'}$  » si :

$$\frac{|\widehat{\mu}_i(\mathbf{y}) - \widehat{\mu}_{i'}(\mathbf{y})|}{\sqrt{\frac{s_R^2}{2} \left( \frac{1}{n_i} + \frac{1}{n_{i'}} \right)}} < q(I, n - I; 1 - \alpha).$$

### 9.2.3. 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 9.2.2 :

Les intervalles de confiance ci-dessus sont modifiés en :

$$\widehat{\mu}_i(\mathbf{y}) - \widehat{\mu}_{i'}(\mathbf{y}) - m \sqrt{s_R^2 \left( \frac{1}{n_i} + \frac{1}{n_{i'}} \right)} < \mu_i - \mu_{i'} < \widehat{\mu}_i(\mathbf{y}) - \widehat{\mu}_{i'}(\mathbf{y}) + m \sqrt{s_R^2 \left( \frac{1}{n_i} + \frac{1}{n_{i'}} \right)}.$$

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

$$\boxed{\mathcal{H}_1 : \mu_i \neq \mu_{i'}.$$

Nous rappelons 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  $\alpha$  et nous décidons 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 :

$$\frac{|\widehat{\mu}_i(\mathbf{y}) - \widehat{\mu}_{i'}(\mathbf{y})|}{\sqrt{\frac{s_R^2}{2} \left( \frac{1}{n_i} + \frac{1}{n_{i'}} \right)}} \geq m(I, n - I; 1 - \alpha).$$

Le test n'est pas significatif au seuil  $\alpha$  et nous décidons de conserver par défaut l'hypothèse nulle  $\mathcal{H}_0$  «  $\mu_i = \mu_{i'}$  » si :

$$\frac{|\widehat{\mu}_i(\mathbf{y}) - \widehat{\mu}_{i'}(\mathbf{y})|}{\sqrt{\frac{s_R^2}{2} \left( \frac{1}{n_i} + \frac{1}{n_{i'}} \right)}} < m(I, n - I; 1 - \alpha).$$



### 9.2.4. La méthode de Scheffé

Contrairement à la méthode de Tukey exposée au paragraphe 9.2.1, 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  $\hat{L}$  alors un intervalle de confiance de niveau  $100(1 - \alpha) \%$  est donné par la formule suivante :

$$\hat{L}(\mathbf{y}) - S \sqrt{(I - 1) s_R^2 \left( \sum_{i=1}^I \frac{l_i^2}{n_i} \right)} < L$$

et

$$L < \hat{L}(\mathbf{y}) + S \sqrt{(I - 1) s_R^2 \left( \sum_{i=1}^I \frac{l_i^2}{n_i} \right)},$$

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, nous décidons que le contraste n'est pas significativement différent de 0 au seuil  $\alpha$ . Au contraire si l'intervalle de confiance ne contient pas 0, alors nous décidons que le contraste est significativement différent de 0 au seuil  $\alpha$ .

L'intérêt de cette procédure est que, si nous fixons  $\alpha$ , les intervalles définis ci-dessus sont valables **simultanément** pour tous les contrastes qu'il est possible de construire !

Nous souhaitons tester l'hypothèse

$$\mathcal{H}_0 : L = 0$$

contre

$$\mathcal{H}_1 : L \neq 0.$$

Le test est significatif au seuil  $\alpha$  et nous décidons de rejeter l'hypothèse nulle  $\mathcal{H}_0$  «  $L = 0$  » en faveur de l'hypothèse alternative  $\mathcal{H}_1$  «  $L \neq 0$  » si :

$$\frac{|\hat{L}(\mathbf{y})|}{\sqrt{(I - 1) s_R^2 \left( \sum_{i=1}^I \frac{l_i^2}{n_i} \right)}} \geq \sqrt{F(I - 1, n - I; 1 - \alpha)}.$$

Le test n'est pas significatif au seuil  $\alpha$  et nous décidons de conserver par défaut l'hypothèse nulle  $\mathcal{H}_0$  «  $L = 0$  » si :

$$\frac{|\hat{L}(\mathbf{y})|}{\sqrt{(I - 1) s_R^2 \left( \sum_{i=1}^I \frac{l_i^2}{n_i} \right)}} < \sqrt{F(I - 1, n - I; 1 - \alpha)}.$$

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 multiples, les intervalles de confiance ci-dessus se simplifient en :

$$\begin{aligned} \widehat{\mu}_i(\mathbf{y}) - \widehat{\mu}_{i'}(\mathbf{y}) - S\sqrt{(I-1)s_R^2\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)} &< \mu_i - \mu_{i'} \\ \text{et} \\ \mu_i - \mu_{i'} &< \widehat{\mu}_i(\mathbf{y}) - \widehat{\mu}_{i'}(\mathbf{y}) + S\sqrt{(I-1)s_R^2\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)}. \end{aligned}$$

Les hypothèses et les statistiques des tests se transforment, quant à elles, en :

$$\mathcal{H}_0 : \mu_i = \mu_{i'}$$

contre

$$\mathcal{H}_1 : \mu_i \neq \mu_{i'}.$$

Nous rappelons que ce jeu d'hypothèses est équivalent à celui-ci :

$$\mathcal{H}_0 : \alpha_i = \alpha_{i'}$$

contre

$$\mathcal{H}_1 : \alpha_i \neq \alpha_{i'}.$$

Le test est significatif au seuil  $\alpha$  et nous décidons 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 :

$$\frac{|\widehat{\mu}_i(\mathbf{y}) - \widehat{\mu}_{i'}(\mathbf{y})|}{\sqrt{(I-1)s_R^2\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)}} \geq \sqrt{F(I-1, n-I; 1-\alpha)}.$$

Le test n'est pas significatif au seuil  $\alpha$  et nous décidons de conserver par défaut l'hypothèse nulle  $\mathcal{H}_0$  «  $\mu_i = \mu_{i'}$  » si :

$$\frac{|\widehat{\mu}_i(\mathbf{y}) - \widehat{\mu}_{i'}(\mathbf{y})|}{\sqrt{(I-1)s_R^2\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)}} < \sqrt{F(I-1, n-I; 1-\alpha)}.$$

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  $A$ , **nous sommes assurés** que la probabilité qu'aucune des différences n'est significative est exactement de  $1 - \alpha$ .

Dans le cas où le plan est équilibré nous obtenons l'intervalle de confiance de niveau  $100(1 - \alpha)$  suivant :

$$\widehat{\mu}_i(\mathbf{y}) - \widehat{\mu}_{i'}(\mathbf{y}) - S\sqrt{\frac{2(I-1)s_R^2}{J}} < \mu_i - \mu_{i'} < \widehat{\mu}_i(\mathbf{y}) - \widehat{\mu}_{i'}(\mathbf{y}) + S\sqrt{\frac{2(I-1)s_R^2}{J}}.$$

### 9.2.5. La méthode de la plus petite différence significative de Fisher

La méthode de la plus petite différence significative de Fisher<sup>5</sup> est une méthode jugée par beaucoup de statisticiens comme inadéquate, voir par exemple [Phi82], puisqu'elle ne tient pas compte du contexte même des comparaisons multiples c'est-à-dire du fait que nous réalisons  $I(I - 1)/2$  comparaisons qui ne sont pas indépendantes. Nous l'exposons 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 :

$$\widehat{\mu}_i(\mathbf{y}) - \widehat{\mu}_{i'}(\mathbf{y}) - t\sqrt{s_R^2\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)} < \mu_i - \mu_{i'} < \widehat{\mu}_i(\mathbf{y}) - \widehat{\mu}_{i'}(\mathbf{y}) + t\sqrt{s_R^2\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)},$$

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 nous ne pouvons pas rejeter l'hypothèse que le contraste n'est pas significativement différent de 0 au seuil  $\alpha$ . Au contraire si l'intervalle de confiance ne contient pas 0, alors nous décidons que le contraste est significativement différent de 0 au seuil  $\alpha$ .

Nous souhaitons tester l'hypothèse :

$$\mathcal{H}_0 : \mu_i = \mu_{i'}$$

contre

$$\mathcal{H}_1 : \mu_i \neq \mu_{i'}$$

Nous rappelons que ce jeu d'hypothèses est équivalent à celui-ci :

$$\mathcal{H}_0 : \alpha_i = \alpha_{i'}$$

contre

$$\mathcal{H}_1 : \alpha_i \neq \alpha_{i'}$$

Le test est significatif au seuil  $\alpha$  et nous décidons 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 :

$$\frac{|\widehat{\mu}_i(\mathbf{y}) - \widehat{\mu}_{i'}(\mathbf{y})|}{\sqrt{s_R^2\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)}} \geq t_{n-I;1-\alpha/2}$$

Le test n'est pas significatif au seuil  $\alpha$  et nous décidons de conserver par défaut l'hypothèse nulle  $\mathcal{H}_0$  «  $\mu_i = \mu_{i'}$  » si :

$$\frac{|\widehat{\mu}_i(\mathbf{y}) - \widehat{\mu}_{i'}(\mathbf{y})|}{\sqrt{s_R^2\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)}} < t_{n-I;1-\alpha/2}$$

5. En anglais ce test de comparaisons multiples est souvent noté *Fisher's LSD* pour *Fisher's Least Significant Difference*.

### 9.2.6. La méthode de Bonferroni

Supposons que nous souhaitons réaliser  $k$  comparaisons de moyennes ou  $k$  tests de contrastes et prendre une décision conjointe au risque de première espèce  $\alpha$ . Nous notons  $E_i$  l'évènement associé au rejet de l'absence de différence significative au seuil  $\alpha_{\text{ind}}$  lors de la  $i$ -ème comparaison ou le test du  $i$ -ème contraste. La méthode de Bonferroni est basée sur l'inégalité suivante :

$$\begin{aligned} \alpha &= \mathbb{P}[E_1 \cup E_2 \cup \dots \cup E_k] \\ &\leq \mathbb{P}[E_1] + \mathbb{P}[E_2] + \dots + \mathbb{P}[E_k] \\ &\leq k * \alpha_{\text{ind}}. \end{aligned}$$

Ainsi si nous voulons être sûr que le risque de première espèce  $\alpha$  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  $\alpha_0$  fixée à l'avance, il suffit de choisir :

$$\alpha_{\text{ind}} \leq \frac{\alpha_0}{k}.$$

Nous procédons alors à des comparaisons des moyennes deux à deux avec un test  $t$  de Student de seuil  $\alpha_0/k$  ou à un test de chacun des contrastes avec la statistique exposée au paragraphe 9.1.4 au seuil  $\alpha_0/k$ . Cette procédure s'applique donc si le plan est équilibré ou non.

Les intervalles de confiance pour  $k$  comparaisons de deux moyennes  $\mu_i$  et  $\mu_{i'}$  de deux groupes d'effectifs  $n_i$  et  $n_{i'}$  sont :

$$\widehat{\mu}_i(\mathbf{y}) - \widehat{\mu}_{i'}(\mathbf{y}) - t_B \sqrt{s_R^2 \left( \frac{1}{n_i} + \frac{1}{n_{i'}} \right)} < \mu_i - \mu_{i'} < \widehat{\mu}_i(\mathbf{y}) - \widehat{\mu}_{i'}(\mathbf{y}) + t_B \sqrt{s_R^2 \left( \frac{1}{n_i} + \frac{1}{n_{i'}} \right)},$$

où  $t_B = t(n - I; 1 - \frac{\alpha}{2k})$  est le  $100(1 - \frac{\alpha}{2k})$  quantile de la loi de Student à  $n - I$  degrés de liberté.

**Remarque :**

Il est conseillé, par Fleiss voir [Fle86], d'utiliser la procédure de Bonferroni dans les cas suivants :

1. Le nombre de comparaisons  $k$  n'est pas très élevé, cette procédure étant trop conservative<sup>6</sup> si  $k$  est élevé.
2. Nous préférons la procédure de Bonferroni à celle de Scheffé si le nombre de comparaisons est strictement inférieur à  $I^2$ .
3. Nous préférons la procédure de Bonferroni à celle de Tukey si le nombre de comparaisons est strictement inférieur à  $I(I - 1)/2$  ou si nous souhaitons tester en plus un petit nombre de comparaisons autres que celles des effets principaux des modalités  $A_i$  du facteur  $A$ .

---

6. Une procédure est dite **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 nous avons fixé en avance pour le test.

4. Nous préférons la procédure de Bonferroni à celle de Dunnett si nous souhaitons 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 nous souhaitons 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 recommandations s'appliquent si nous considérons des tests portant sur des contrastes.

### 9.2.7. La méthode de rejet séquentiel de Bonferroni de Holm

La méthode de rejet séquentiel de Bonferroni de Holm<sup>7</sup> est dérivée de la méthode de Bonferroni exposée au paragraphe 9.2.6. 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 nous effectuons 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 9.2.6 et nous pouvons aussi bien nous en servir pour des plans équilibrés que des plans déséquilibrés.

### 9.2.8. 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,  $\alpha_{\text{ind}}$ , est modifiée.

Supposons que nous souhaitons réaliser  $k$  comparaisons de moyennes ou  $k$  tests de contrastes et prendre une décision conjointe au risque de première espèce  $\alpha$ . Nous notons  $E_i$  l'évènement associé au rejet de l'absence de différence significative au seuil  $\alpha_{\text{ind}}$  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 conservatrice que celle de Dunn-Šidák. Ainsi si nous voulons être sûr que le risque de première espèce  $\alpha$  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  $\alpha_0$  fixée à l'avance, il suffit de choisir :

$$\alpha_{\text{ind}} \leq 1 - \sqrt[k]{1 - \alpha_0}.$$

Nous procédons 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 9.1.4 au seuil  $1 - \sqrt[k]{1 - \alpha_0}$ . Cette procédure s'applique donc si le

---

7. En anglais *Holm's SRB criterion* pour *Holm's Sequentially Rejective Bonferroni criterion*.

plan est équilibré ou non.

L'intervalle de confiance de niveau  $100(1 - \alpha_0)$  % pour  $\mu_i - \mu_{i'}$  est alors le suivant :

$$\widehat{\mu}_i(\mathbf{y}) - \widehat{\mu}_{i'}(\mathbf{y}) - t_{DS} \sqrt{s_R^2 \left( \frac{1}{n_i} + \frac{1}{n_{i'}} \right)} < \mu_i - \mu_{i'} < \widehat{\mu}_i(\mathbf{y}) - \widehat{\mu}_{i'}(\mathbf{y}) + t_{DS} \sqrt{s_R^2 \left( \frac{1}{n_i} + \frac{1}{n_{i'}} \right)},$$

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.

### 9.2.9. 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. Nous commençons par ordonner les estimations des moyennes  $\widehat{\mu}_1(\mathbf{y}), \dots, \widehat{\mu}_I(\mathbf{y})$  par ordre croissant. Nous obtenons donc  $\widehat{\mu}_{(1)}(\mathbf{y}) \leq \dots \leq \widehat{\mu}_{(I)}(\mathbf{y})$ . Nous ordonnons de la même manière les moyennes  $\mu_1, \dots, \mu_I$ .
2. Nous calculons la quantité  $\widehat{\mu}_{(I)}(\mathbf{y}) - \widehat{\mu}_{(1)}(\mathbf{y})$ , qui est toujours positive, et nous la comparons 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  $\widehat{\mu}_{(I)}(\mathbf{y}) - \widehat{\mu}_{(1)}(\mathbf{y}) \geq W_{(I)}$  alors nous décidons que le groupe de moyennes  $\mu_{(1)}, \dots, \mu_{(I)}$  n'est pas homogène et nous créons les deux groupes  $\mu_{(1)}, \dots, \mu_{(I-1)}$  et  $\mu_{(2)}, \dots, \mu_{(I)}$  sinon nous considérons que le groupe de moyennes  $\mu_{(1)}, \dots, \mu_{(I)}$  est homogène et la procédure s'arrête là.

3. Nous calculons les quantités  $\widehat{\mu}_{(I-1)}(\mathbf{y}) - \widehat{\mu}_{(1)}(\mathbf{y})$  et  $\widehat{\mu}_{(I)}(\mathbf{y}) - \widehat{\mu}_{(2)}(\mathbf{y})$ , qui sont toujours positives, et nous les comparons à  $W_{(I-1)}$  où :

$$W_{(I-1)} = q(I - 1, I(J - 1); 1 - \alpha) \sqrt{\frac{s_R^2}{J}},$$

avec  $q(I - 1, I(J - 1); 1 - \alpha)$  le  $100(1 - \alpha)$  quantile de la loi de l'étendue Studentisée à  $I - 1$  et  $I(J - 1)$  degrés de liberté.

Si  $\widehat{\mu}_{(I-1)}(\mathbf{y}) - \widehat{\mu}_{(1)}(\mathbf{y}) \geq W_{(I-1)}$  alors nous décidons que le groupe de moyennes  $\mu_{(1)}, \dots, \mu_{(I-1)}$  n'est pas homogène et nous créons les deux groupes  $\mu_{(1)}, \dots, \mu_{(I-2)}$  et  $\mu_{(2)}, \dots, \mu_{(I-1)}$  sinon nous considérons que le groupe de moyennes  $\mu_{(1)}, \dots, \mu_{(I-1)}$  est homogène et la procédure s'arrête là pour ce groupe. Nous pratiquons la même analyse pour le groupe  $\mu_{(2)}, \dots, \mu_{(I)}$ .

4. Nous réitérons 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 nous testons 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 nous obtiendrons  $I$  groupes distincts ce qui prouve que cette procédure 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 nous remplaçons, 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, nous comparerons la quantité  $\widehat{\mu}_{(I)}(\mathbf{y}) - \widehat{\mu}_{(1)}(\mathbf{y})$ , 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.

### 9.2.10. La méthode de Duncan

Le test de l'étendue multiple de Duncan<sup>8</sup> 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 9.2.9. 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 nous remplaçons, 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 nous cherchons à tester l'homogénéité de  $k$  moyennes dont l'estimation la plus petite est pour  $\mu_{(i)}$  et l'estimation la plus grande pour  $\mu_{(i')}$ , nous comparerons la quantité  $\widehat{\mu}_{(i)}(\mathbf{y}) - \widehat{\mu}_{(i')}(\mathbf{y})$  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.

---

8. En anglais Duncan's Multiple Range Test.

### 9.2.11. 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  $\mu_i$  obtenue pour la modalité  $A_i$  du facteur  $A$  et de la moyenne  $\mu_c$  de la réponse obtenue pour le niveau de contrôle  $A_c$  du facteur  $A$ .

Nous souhaitons tester l'hypothèse :

$$\boxed{\mathcal{H}_0 : \mu_i = \mu_c}$$

contre

$$\boxed{\mathcal{H}_1 : \mu_i \neq \mu_c.}$$

Nous rappelons 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  $\alpha$  et nous décidons 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 :

$$|\widehat{\mu}_i(\mathbf{y}) - \widehat{\mu}_c(\mathbf{y})| \geq D(I - 1, n - I; 1 - \alpha) \sqrt{s_R^2 \left( \frac{1}{n_i} + \frac{1}{n_c} \right)},$$

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  $\alpha$  et nous décidons de conserver par défaut l'hypothèse nulle  $\mathcal{H}_0$  «  $\mu_i = \mu_c$  » si :

$$|\widehat{\mu}_i(\mathbf{y}) - \widehat{\mu}_c(\mathbf{y})| < D(I - 1, n - I; 1 - \alpha) \sqrt{s_R^2 \left( \frac{1}{n_i} + \frac{1}{n_c} \right)},$$

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 nous souhaitons 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.



## 9.3. Comparaisons multiples sous l'hypothèse d'hétéroscédasticité

### 9.3.1. 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  $\sigma_i^2$  des niveaux  $A_i$  du facteur  $A$  ainsi que d'un éventuel déséquilibre du plan. Nous remplaçons ainsi  $S_R^2$  par :

$$S_R^2 = \sqrt{\frac{S_i^2}{n_i} + \frac{S_{i'}^2}{n_{i'}}},$$

où

$$S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (Y_{i,j} - Y_{i,\bullet})^2,$$

$$S_{i'}^2 = \frac{1}{n_{i'} - 1} \sum_{j=1}^{n_{i'}} (Y_{i',j} - Y_{i',\bullet})^2,$$

sont les estimateurs sans biais des variances au sein des niveaux  $A_i$  et  $A_{i'}$  du facteur. Nous désignons par  $s_i^2$  et  $s_{i'}^2$  les estimations de  $\sigma_i^2$  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  $\sigma_i^2$  et  $\sigma_{i'}^2$  ou entre le nombre de répétitions  $n_i$  et  $n_{i'}$  d'un niveau du facteur  $A$  à l'autre.

Nous modifions également un des degrés de liberté de la loi de l'étendue Studentisée dont nous nous servons pour trouver les quantiles nécessaires au test. Plus précisément, nous utilisons  $q(I, \nu_{i,i'}; 1 - \alpha)$  avec :

$$\nu_{i,i'} = \frac{\left(\frac{s_i^2}{n_i} + \frac{s_{i'}^2}{n_{i'}}\right)^2}{\frac{\left(\frac{S_i^2}{n_i}\right)^2}{n_i - 1} + \frac{\left(\frac{S_{i'}^2}{n_{i'}}\right)^2}{n_{i'} - 1}}.$$

## 9.4. 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. Il a été 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 [KRB96].

### 9.4.1. Le test multiple des étendues de Ryan-Einot-Gabriel-Welsch

Nous souhaitons utiliser des tests multiples de l'étendue, nous utilisons les valeurs suivantes pour les seuils :

$$\alpha_m = \begin{cases} 1 - (1 - \alpha)^{\frac{m}{I}} & \text{si } m = 2, \dots, I - 2, \\ \alpha & \text{si } m = I - 1, I. \end{cases} \quad (9.4.1)$$

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  $\mathbb{I} = \{1, \dots, I\}$ , 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  $\mathbb{K}$ . Nous ne considérerons 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 (9.4.1),  $\mathcal{H}_{\mathbb{I}} : \mu_1 = \dots = \mu_I$ . Si  $\mathcal{H}_{\mathbb{I}}$  n'est pas rejetée au seuil  $\alpha = \alpha_I$ , alors nous décidons que nous ne pouvons rejeter aucune des hypothèses  $\mathcal{H}_{\mathbb{K}}$  pour  $\mathbb{K} \subset \mathbb{I}$  et nous arrêtons le procédé à cette étape. Sinon nous décidons qu'il existe  $i_0$  et  $j_0$  dans  $\{1, \dots, I\}$  tels que  $\mu_{i_0} \neq \mu_{j_0}$  et nous nous rendons à l'étape 2..
2. Tester au seuil  $\alpha = \alpha_{|\mathbb{I}-1|} = \alpha_{I-1}$ , donné par l'équation (9.4.1), 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 nous décidons que nous ne pouvons rejeter aucune des hypothèses  $\mathcal{H}_{\mathbb{K}}$  pour  $\mathbb{K} \subset \mathbb{L}$ . Sinon nous décidons 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, nous avonssrête la procédure; dans le cas contraire, nous nous rendons à l'étape 3..
3. Tester au seuil  $\alpha = \alpha_{|\mathbb{L}|}$ , donné par l'équation (9.4.1), 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 nous décidons que nous ne pouvons rejeter aucune des hypothèses  $\mathcal{H}_{\mathbb{K}}$  pour  $\mathbb{K} \subset \mathbb{L}$ . Sinon nous décidons 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  $\alpha_{|\mathbb{L}|}$ .
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 nous testons les hypothèses  $\mathcal{H}_{\mathbb{L}}$  au seuil  $\alpha_{|\mathbb{L}|}$  de la manière suivante :

- Nous ne pouvons pas rejeter l'hypothèse  $\mathcal{H}_{\mathbb{L}}$  au seuil  $\alpha_{|\mathbb{L}|}$  si

$$\max_{i,j \in \mathbb{L}} \widehat{\mu}_i(\mathbf{y}) - \widehat{\mu}_j(\mathbf{y}) < c_{\alpha_{|\mathbb{L}|}} \sqrt{\frac{2s_R^2}{n}},$$

## 9.4 Deux méthodes recommandées

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 (9.4.1) pour les valeurs des  $\alpha_m$ ,  $1 \leq m \leq I$  nous obtenons, si  $2 \leq |\mathbb{L}| \leq I - 2$  :

$$1 - (1 - \alpha)^{\frac{|\mathbb{L}|}{I}} = \mathbb{P} \left[ \max_{i,j \in \mathbb{L}} \frac{\widehat{\mu}_i - \widehat{\mu}_j}{\sqrt{\frac{2S_R^2}{n}}} > c_{\alpha_{|\mathbb{L}|}} \right],$$

et si  $I - 1 \leq |\mathbb{L}| \leq I$  :

$$\alpha = \mathbb{P} \left[ \max_{i,j \in \mathbb{L}} \frac{\widehat{\mu}_i - \widehat{\mu}_j}{\sqrt{\frac{2S_R^2}{n}}} > c_{\alpha_{|\mathbb{L}|}} \right].$$

Ainsi  $c_{\alpha_{|\mathbb{L}|}}$  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.

- Nous décidons de rejeter l'hypothèse  $\mathcal{H}_{\mathbb{L}}$  au seuil  $\alpha_{|\mathbb{L}|}$  si

$$\max_{i,j \in \mathbb{L}} \widehat{\mu}_i(\mathbf{y}) - \widehat{\mu}_j(\mathbf{y}) \geq c_{\alpha_{|\mathbb{L}|}} \sqrt{\frac{2s_R^2}{n}},$$

où  $c_{\alpha_{|\mathbb{L}|}}$  est défini comme ci-dessus.

Si les valeurs critiques  $c_m$ ,  $1 \leq m \leq I$  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  $\alpha_m$  donnée par l'équation (9.4.1), il n'y a aucune garantie que la valeur critique  $c_m$  soit croissante avec  $m$ . Pour comprendre pourquoi cette condition est absolument nécessaire, nous pourrions consulter l'exemple page 5067 dans [KRB96]

Ainsi, nous préférons utiliser, si les valeurs critiques calculées à l'aide de l'équation (9.4.1) ne sont pas monotones, c'est-à-dire si elles ne sont pas croissantes, les valeurs critiques modifiées définies par :

$$c'_m = \max_{2 \leq j \leq m} c_j, \quad m = 2, \dots, I.$$

Attention à bien vérifier ce point si nous ne voulons pas obtenir de résultats incohérents.

**Remarque :**

Nous aurions 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, nous utiliserions alors l'inégalité de Bonferroni et les valeurs suivantes pour les seuils :

$$\alpha_m = \begin{cases} \frac{m}{I}\alpha & \text{si } m = 2, \dots, I-2, \\ \alpha & \text{si } m = I-1, I. \end{cases}$$

Toutefois cette procédure possède le même défaut que celle de Bonferroni par rapport à celle de la méthode de Dunn-Šidák : elle est plus conservatrice, voir le paragraphe 9.2.8.

### 9.4.2. 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  $\mathbb{L} \subseteq \mathbb{I} = \{1, \dots, I\}$ , 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 nous ne pouvons nous 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 nous ne voulons pas obtenir de résultats incohérents au seuil  $\alpha$ .
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  $\alpha$  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 nous testons les hypothèses  $\mathcal{H}_{\mathbb{L}}$  au seuil  $\alpha_{|\mathbb{L}|}$  de la manière suivante :

- Nous ne pouvons pas rejeter l'hypothèse  $\mathcal{H}_{\mathbb{L}}$  au seuil  $\alpha_{|\mathbb{L}|}$  si

$$\frac{\sum_{i \in \mathbb{L}} n \left( \widehat{\mu}_i(\mathbf{y}) - \widehat{\mu}_{\mathbb{L}}(\mathbf{y}) \right)^2}{(|\mathbb{L}| - 1) s_R^2} < c_{\alpha_{|\mathbb{L}|}},$$

où  $\widehat{\mu}_{\mathbb{L}}$  est défini par

$$\widehat{\mu}_{\mathbb{L}} = \frac{\sum_{i \in \mathbb{L}} \sum_{a=1}^n Y_{i,a}}{\sum_{i \in \mathbb{L}} n},$$

ainsi que  $c_{\alpha_{|\mathbb{L}|}}$  par

$$\alpha_{|\mathbb{L}|} = \mathbb{P} \left[ \frac{\sum_{i \in \mathbb{L}} n \left( \widehat{\mu}_i - \widehat{\mu}_{\mathbb{L}} \right)^2}{(|\mathbb{L}| - 1) S_R^2} > c_{\alpha_{|\mathbb{L}|}} \right].$$

## 9.5 Comparaisons avec le *meilleur*

---

Ainsi avec le choix de l'équation (9.4.1) pour les valeurs des  $\alpha_i$ ,  $1 \leq i \leq I$  nous obtenons, si  $2 \leq |\mathbb{L}| \leq I - 2$  :

$$1 - (1 - \alpha)^{\frac{|\mathbb{L}|}{I}} = \mathbb{P} \left[ \frac{\sum_{i \in \mathbb{L}} n (\widehat{\mu}_i - \widehat{\mu}_{\mathbb{L}})^2}{(|\mathbb{L}| - 1) S_R^2} > c_{\alpha_{|\mathbb{L}|}} \right],$$

et si  $I - 1 \leq |\mathbb{L}| \leq I$  :

$$\alpha = \mathbb{P} \left[ \frac{\sum_{i \in \mathbb{L}} n (\widehat{\mu}_i - \widehat{\mu}_{\mathbb{L}})^2}{(|\mathbb{L}| - 1) S_R^2} > c_{\alpha_{|\mathbb{L}|}} \right].$$

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.

- Nous décidons de rejeter l'hypothèse  $\mathcal{H}_{\mathbb{L}}$  au seuil  $\alpha_{|\mathbb{L}|}$  si

$$\frac{\sum_{i \in \mathbb{L}} n (\widehat{\mu}_i(\mathbf{y}) - \widehat{\mu}_{\mathbb{L}}(\mathbf{y}))^2}{(|\mathbb{L}| - 1) s_R^2} \geq c_{\alpha_{|\mathbb{L}|}},$$

où  $c_{\alpha_{|\mathbb{L}|}}$  est défini comme ci-dessus.

## 9.5. 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 nous avons observé et que nous considérons comme le « meilleur ». Ce type de problématique ne peut être traité à l'aide d'une approche basée sur la méthode de Dunnett puisque nous ne savons pas *a priori* pour quelle modalité nous allons observer le « meilleur » effet.

Nous utilisons alors des techniques de comparaisons multiples avec le « meilleur<sup>9</sup> », le plus souvent celle de Hsu ou celle de Edwards-Hsu.

Nous nous intéressons ici donc non pas à toutes les comparaisons possibles mais seulement à celles-ci :

- Si le « meilleur » est la valeur la plus élevée nous utilisons :

$$\mu_i - \max_{j \neq i} \mu_j, \quad i = 1, \dots, I,$$

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 nous avons néanmoins  $\mu_{i_2} - \max_{j \neq i_2} \mu_j > -\delta$  avec  $\delta$  un petit nombre réel strictement positif alors la modalité  $i_2$  est proche de la « meilleure ».

---

9. Ceci est souvent abrégé en anglais en *MCB* pour « Multiple comparisons with the Best ».

- Si le « meilleur » est la valeur la moins élevée nous utilisons :

$$\mu_i - \min_{j \neq i} \mu_j, \quad i = 1, \dots, I,$$

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 nous avons néanmoins  $\mu_{i_2} - \min_{j \neq i_2} \mu_j > \delta$  avec  $\delta$  un petit nombre réel strictement positif alors la modalité  $i_2$  est proche de la « meilleure ».

### 9.5.1. La méthode de Hsu

Hsu a développé des intervalles de confiance simultanés de niveau de confiance  $100(1 - \alpha)$  pour  $\mu_i - \max_{j \neq i} \mu_j$ ,  $i = 1, \dots, I$ . Ces intervalles contiennent tous la valeur 0 et sont construits ainsi : pour chaque  $i$ , soit  $d_i$  la valeur critique telle que

$$\mathbb{P} \left[ \widehat{\mu}_i - \mu_i > \widehat{\mu}_j - \mu_j - d_i \sqrt{S_R^2 \left( \frac{1}{n_i} + \frac{1}{n_j} \right)}, \quad \forall j | j \neq i \right] = 1 - \alpha.$$

Nous remarquons 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  $[D_i^-, D_i^+]$ ,  $i = 1, \dots, I$  où

$$\begin{aligned} D_i^+ &= \max \left\{ 0, \min_{j \neq i} \left( \widehat{\mu}_i - \widehat{\mu}_j - d_i \sqrt{S_R^2 \left( \frac{1}{n_i} + \frac{1}{n_j} \right)} \right) \right\} \\ G &= \{ i \mid D_i^+ > 0 \} \\ D_i^- &= \begin{cases} 0 & \text{si } i \notin G \\ \min_{j \in G, j \neq i} \left( \widehat{\mu}_i - \widehat{\mu}_j - d_j \sqrt{S_R^2 \left( \frac{1}{n_i} + \frac{1}{n_j} \right)} \right) & \text{si } i \in G \end{cases} \end{aligned}$$

forment un ensemble d'intervalles de confiance simultanés de niveau de confiance  $100(1 - \alpha)$  pour  $\mu_i - \max_{j \neq i} \mu_j$ ,  $i = 1, \dots, I$ .

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  $1 \leq i \leq I$  et  $d = d_1 = \dots = d_I$ .

### 9.5.2. La méthode de Edwards-Hsu

Si nous cherchons à avoir également des bornes sur l'écart minimum entre la valeur du « meilleur » traitement et celle des autres traitements que celui dont l'estimatinous avons

## 9.5 Comparaisons avec le *meilleur*

---

été la meilleure, alors nous pouvons utiliser la méthode de Edwards-Hsu. Celle-ci permet d'obtenir des intervalles de confiance simultanés pour  $\mu_i - \max_{i \leq j \leq I} \mu_j$ ,  $i = 1, \dots, I$  que nous ne forçons plus à contenir la valeur 0.

Pour chaque  $i$ , soit  $|d|_i$  la valeur critique telle que

$$\mathbb{P} \left[ \widehat{\mu}_i - \mu_i - (\widehat{\mu}_j - \mu_j) < |d|_i \sqrt{S_R^2 \left( \frac{1}{n_i} + \frac{1}{n_j} \right)}, \forall j | j \neq i \right] = 1 - \alpha.$$

Nous remarquons 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

$$S = \left\{ i \text{ tels que } \min_{j \neq i} \left\{ \widehat{\mu}_i - \widehat{\mu}_j + |d|_i \sqrt{S_R^2 \left( \frac{1}{n_i} + \frac{1}{n_j} \right)} \right\} > 0 \right\},$$

$$L_{ij} = \begin{cases} 0 & \text{si } i = j \\ \widehat{\mu}_i - \widehat{\mu}_j - |d|_j \sqrt{S_R^2 \left( \frac{1}{n_i} + \frac{1}{n_j} \right)} & \text{si } i \neq j \end{cases},$$

$$U_{ij} = \begin{cases} 0 & \text{si } i = j \\ -\min \left\{ 0, \left( \widehat{\mu}_i - \widehat{\mu}_j + |d|_j \sqrt{S_R^2 \left( \frac{1}{n_i} + \frac{1}{n_j} \right)} \right) \right\} & \text{si } i \neq j \end{cases},$$

$$L_i = \min_{j \in S} L_{ij} \quad U_i = \max_{j \in S} U_{ij}$$

pour  $i = 1, \dots, I$ , forment un ensemble d'intervalles de confiances simultanés de niveau  $100(1 - \alpha)$  pour  $\mu_i - \max_{i \leq j \leq I} \mu_j$ ,  $i = 1, \dots, I$ .

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 :

$$\mathbb{P} \left[ \widehat{\mu}_i - \mu_i - (\widehat{\mu}_j - \mu_j) < |d| \sqrt{\frac{2S_R^2}{n}}, \forall j | j \neq i \right] = 1 - \alpha.$$

Nous remarquons 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

$$S = \left\{ i \text{ tels que } \min_{j \neq i} \left\{ \widehat{\mu}_i - \widehat{\mu}_j + |d| \sqrt{\frac{2S_R^2}{n}} \right\} > 0 \right\},$$

$$L_{ij} = \begin{cases} 0 & \text{si } i = j \\ \widehat{\mu}_i - \widehat{\mu}_j - |d| \sqrt{\frac{2S_R^2}{n}} & \text{si } i \neq j \end{cases},$$

$$U_{ij} = \begin{cases} 0 & \text{si } i = j \\ -\min \left\{ 0, \left( \widehat{\mu}_i - \widehat{\mu}_j + |d| \sqrt{\frac{2S_R^2}{n}} \right) \right\} & \text{si } i \neq j \end{cases},$$

$$L_i = \min_{j \in S} L_{ij} \quad U_i = \max_{j \in S} U_{ij}$$

pour  $i = 1, \dots, I$ , forment un ensemble d'intervalles de confiances simultanés de niveau  $100(1 - \alpha)$  pour  $\mu_i - \max_{i \leq j \leq I} \mu_j$ ,  $i = 1, \dots, I$ .

## 9.6. Deux facteurs sans répétition

Nous pouvons adapter les méthodes de comparaison des effets des différentes modalités, exposées aux paragraphes 9.1, 9.2 et 9.3 dans le cas de l'analyse de la variance à un facteur à effets fixes, au cas où nous considérons un modèle à deux facteurs sans interaction à effets fixes ou à effets mixtes. Nous commençons par détailler les modifications à faire pour utiliser les méthodes de Tukey et de Scheffé.

Nous reprenons ici les notations du paragraphe 3.1.1 ou du paragraphe 3.3.1 en fonction de la situation étudiée.

### 9.6.1. 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 :

$$L = l_1 \alpha_1 + l_2 \alpha_2 + \dots + l_I \alpha_I,$$

$$L' = l'_1 \beta_1 + l'_2 \beta_2 + \dots + l'_J \beta_J.$$

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 :

$$\widehat{L} = l_1 \widehat{\alpha}_1 + l_2 \widehat{\alpha}_2 + \dots + l_I \widehat{\alpha}_I,$$

$$\widehat{L}' = l'_1 \widehat{\beta}_1 + l'_2 \widehat{\beta}_2 + \dots + l'_J \widehat{\beta}_J.$$

Ainsi des estimations sans biais de  $L$  et  $L'$  s'écrivent :

$$\widehat{L}(\mathbf{y}) = l_1 y_{1,\bullet} + l_2 y_{2,\bullet} + \dots + l_I y_{I,\bullet},$$

$$\widehat{L}'(\mathbf{y}) = l'_1 y_{\bullet,1} + l'_2 y_{\bullet,2} + \dots + l'_J y_{\bullet,J}.$$



### 9.6.2. Méthode de Tukey

Si un contraste  $L$  met en jeu des moyennes associées au facteur  $A$  à effets fixes, nous utilisons l'estimation  $s_R^2$  calculée pour le modèle incluant les deux facteurs et nous changeons le nombre de degrés de liberté de la loi de l'étendue Studentisées en  $I$  et  $(I - 1)(J - 1)$ . Nous décidons que le contraste  $L$  est significativement différent de 0 au seuil  $\alpha$  si :

$$\frac{|\widehat{L}(\mathbf{y})|}{\sqrt{\frac{s_R^2}{J} \left( \frac{1}{2} \sum_{i=1}^I |l_i| \right)}} \geq q(I, (I - 1)(J - 1); 1 - \alpha).$$

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 nous ne pouvons rejeter l'hypothèse d'absence de différence entre  $L$  et 0 au seuil  $\alpha$ .

De même si le contraste  $L'$  met en jeu des moyennes associées au facteur  $B$  à effets fixes, nous utilisons l'estimation  $s_R^2$  calculée pour le modèle incluant les deux facteurs et nous changeons le nombre de degrés de liberté de la loi de l'étendue Studentisées en  $J$  et  $(I - 1)(J - 1)$ .

Nous décidons que le contraste  $L'$  est significativement différent de 0 au seuil  $\alpha$  si :

$$\frac{|\widehat{L}'(\mathbf{y})|}{\sqrt{\frac{s_R^2}{I} \left( \frac{1}{2} \sum_{j=1}^J |l'_j| \right)}} \geq q(J, (I - 1)(J - 1); 1 - \alpha).$$

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 nous ne pouvons rejeter l'hypothèse d'absence de différence entre  $L'$  et 0 au seuil  $\alpha$ .

### 9.6.3. Méthode de Scheffé

Si un contraste  $L$  met en jeu des moyennes associées au facteur  $A$  à effets fixes, nous utilisons l'estimation  $s_R^2$  calculée pour le modèle incluant les deux facteurs et nous changeons le nombre de degrés de liberté de la loi de Fisher en  $I - 1$  et  $(I - 1)(J - 1)$ .

Nous décidons que le contraste  $L$  est significativement différent de 0 au seuil  $\alpha$  si :

$$\frac{|\widehat{L}(\mathbf{y})|}{\sqrt{\frac{(I - 1)s_R^2}{J} \sum_{i=1}^I l_i^2}} \geq \sqrt{F(I - 1, (I - 1)(J - 1); 1 - \alpha)},$$

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 nous ne pouvons rejeter l'hypothèse d'absence de différence entre  $L$  et 0 au seuil  $\alpha$ .

De même si le contraste  $L'$  met en jeu des moyennes associées au facteur  $B$  à effets fixes, nous utilisons l'estimation  $s_R^2$  calculée pour le modèle incluant les deux facteurs et nous

changeons le nombre de degrés de liberté de la loi de Fisher en  $J - 1$  et  $(I - 1)(J - 1)$ . Nous décidons que le contraste  $L'$  est significativement différent de 0 au seuil  $\alpha$  si :

$$\frac{|\widehat{L}'(\mathbf{y})|}{\sqrt{\frac{(J-1)s_R^2}{I} \sum_{j=1}^J l_j'^2}} \geq \sqrt{F(J-1, (I-1)(J-1); 1-\alpha)},$$

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 nous ne pouvons rejeter l'hypothèse d'absence de différence entre  $L'$  et 0 au seuil  $\alpha$ .

#### 9.6.4. Méthode de Bonferroni et méthodes associées

Nous adaptons ici, de manière similaire, la procédure de Bonferroni et toutes celles qui en découle.

Ainsi si nous ne souhaitons **comparer que deux niveaux** d'un même des deux facteurs, ici par exemple deux niveaux du facteur  $A$ , nous pourrons 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 nous souhaitons réaliser **un nombre fini  $k$  de comparaisons** nous pourrons 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.

### 9.7. Deux facteurs avec répétitions

Nous pouvons adapter les méthodes de comparaison des effets des différentes modalités, exposées aux paragraphes 9.1, 9.2 et 9.3 dans le cas de l'analyse de la variance à un facteur à effets fixes, au cas où nous considérons un modèle à deux facteurs avec interaction à

## 9.7 Deux facteurs avec répétitions

effets fixes ou à effets mixtes. Nous commençons par détailler les modifications à faire pour utiliser les méthodes de Tukey et de Scheffé.

Nous reprenons ici les notations du paragraphe 3.1.2 ou du paragraphe 3.3.2 en fonction de la situation considérée. La différence fondamentale entre ces deux situations réside dans le fait que pour la première, **deux facteurs à effets fixes**, nous utiliserons 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**, nous utiliserons le **carré moyen associé au terme de l'interaction**  $s_{AB}^2$  dans les statistiques de test.

### 9.7.1. 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 3.1.2, il peut être intéressant, en fonction de la significativité des différents tests auxquels nous avons procédé, d'étudier les contrastes exposés ci-dessous.

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

contre

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

Si nous rejetons 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 :

$$\widehat{L}_{AB}(\mathbf{y}) = y_{i,j,\bullet} - y_{i',j',\bullet}.$$

Si nous ne pouvons 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 :

$$\begin{aligned} L &= l_1\alpha_1 + l_2\alpha_2 + \dots + l_I\alpha_I, \\ L' &= l'_1\beta_1 + l'_2\beta_2 + \dots + l'_J\beta_J, \end{aligned}$$

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 :

$$\begin{aligned} \widehat{L} &= l_1\widehat{\alpha}_1 + l_2\widehat{\alpha}_2 + \dots + l_I\widehat{\alpha}_I, \\ \widehat{L}' &= l'_1\widehat{\beta}_1 + l'_2\widehat{\beta}_2 + \dots + l'_J\widehat{\beta}_J. \end{aligned}$$

Ainsi des estimations sans biais de  $L$  et  $L'$  s'écrivent :

$$\begin{aligned}\widehat{L}(\mathbf{y}) &= l_1 y_{1,\bullet,\bullet} + l_2 y_{2,\bullet,\bullet} + \cdots + l_I y_{I,\bullet,\bullet}, \\ \widehat{L}'(\mathbf{y}) &= l'_1 y_{\bullet,1,\bullet} + l'_2 y_{\bullet,2,\bullet} + \cdots + l'_J y_{\bullet,J,\bullet}.\end{aligned}$$

### Modèle à effets mixtes

Dans le cas d'un modèle mixte, c'est-à-dire le cas envisagé au paragraphe 3.3.2, il peut être intéressant, en fonction de la significativité du test de l'effet du facteur à effets fixes, d'étudier les contrastes 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 :

$$\widehat{L}(\mathbf{y}) = l_1 y_{1,\bullet,\bullet} + l_2 y_{2,\bullet,\bullet} + \cdots + l_I y_{I,\bullet,\bullet}.$$

### 9.7.2. Méthode de Tukey

#### Modèle à effets fixes

$L_{AB}$  est significativement différent de 0 au seuil  $\alpha$  si, en supposant que  $y_{i,j,\bullet} \geq y_{i',j',\bullet}$ , nous avons :

$$\frac{|\widehat{L}(\mathbf{y})|}{\sqrt{\frac{s_R^2}{K}}} \geq q(IJ, IJ(K-1); 1-\alpha),$$

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 nous ne pouvons rejeter l'hypothèse d'absence de différence entre  $L_{AB}$  et 0 au seuil  $\alpha$ .

Si un contraste  $L$  met en jeu des moyennes associées au facteur  $A$  à effets fixes, nous utilisons l'estimation  $s_R^2$  calculée pour le modèle incluant les deux facteurs et nous changeons le nombre de degrés de liberté de la loi de l'étendue Studentisées en  $I$  et  $IJ(K-1)$ .

Nous décidons que le contraste  $L$  est significativement différent de 0 au seuil  $\alpha$  si :

$$\frac{|\widehat{L}(\mathbf{y})|}{\sqrt{\frac{s_R^2}{JK} \left( \frac{1}{2} \sum_{i=1}^I |l_i| \right)}} \geq q(I, IJ(K-1); 1-\alpha).$$

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 nous ne pouvons rejeter l'hypothèse d'absence de

## 9.7 Deux facteurs avec répétitions

---

différence entre  $L$  et 0 au seuil  $\alpha$ .

De même si le contraste  $L'$  met en jeu des moyennes associées au facteur  $B$  à effets fixes, nous utilisons l'estimation  $s_R^2$  calculée pour le modèle incluant les deux facteurs et nous changeons le nombre de degrés de liberté de la loi de l'étendue Studentisée en  $J$  et  $IJ(K-1)$ .

Nous décidons que le contraste  $L'$  est significativement différent de 0 au seuil  $\alpha$  si :

$$\frac{|\widehat{L}'(\mathbf{y})|}{\sqrt{\frac{s_R^2}{IK} \left( \frac{1}{2} \sum_{j=1}^J |l'_j| \right)}} \geq q(J, IJ(K-1); 1-\alpha).$$

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 nous ne pouvons rejeter l'hypothèse d'absence de différence entre  $L'$  et 0 au seuil  $\alpha$ .

### Modèle à effets mixtes

Si un contraste  $L$  met en jeu des moyennes associées au facteur  $A$  à effets fixes, nous utilisons l'estimation  $s_{AB}^2$  calculée pour le modèle incluant les deux facteurs et nous changeons le nombre de degrés de liberté de la loi de l'étendue Studentisée en  $I$  et  $(I-1)(J-1)$ . Nous décidons que le contraste  $L$  est significativement différent de 0 au seuil  $\alpha$  si :

$$\frac{|\widehat{L}(\mathbf{y})|}{\sqrt{\frac{s_{AB}^2}{JK} \left( \frac{1}{2} \sum_{i=1}^I |l_i| \right)}} \geq q(I, (I-1)(J-1); 1-\alpha).$$

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 nous ne pouvons rejeter l'hypothèse d'absence de différence entre  $L$  et 0 au seuil  $\alpha$ .

### 9.7.3. Méthode de Scheffé

#### Modèle à effets fixes

$L_{AB}$  est significativement différent de 0 au seuil  $\alpha$  si, en supposant que  $y_{i,j,\bullet} \geq y_{i',j',\bullet}$ , nous avons :

$$\frac{|\widehat{L}(\mathbf{y})|}{\sqrt{(IJ-1) \frac{2s_R^2}{K}}} \geq \sqrt{F(IJ-1, IJ(K-1); 1-\alpha)},$$

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 nous ne pouvons rejeter l'hypothèse d'absence de différence entre  $L_{AB}$  et 0 au seuil  $\alpha$ .

Si un contraste  $L$  met en jeu des moyennes associées au facteur  $A$  à effets fixes, nous utilisons l'estimation  $s_R^2$  calculée pour le modèle incluant les deux facteurs et nous changeons le nombre de degrés de liberté de la loi de Fisher en  $I - 1$  et  $IJ(K - 1)$ .

Nous décidons que le contraste  $L$  est significativement différent de 0 au seuil  $\alpha$  si :

$$\frac{|\widehat{L}(\mathbf{y})|}{\sqrt{\frac{(I-1)s_R^2}{JK} \sum_{i=1}^I l_i^2}} \geq \sqrt{F(I-1, IJ(K-1); 1-\alpha)},$$

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 nous ne pouvons rejeter l'hypothèse d'absence de différence entre  $L$  et 0 au seuil  $\alpha$ .

De même si le contraste  $L'$  met en jeu des moyennes associées au facteur  $B$  à effets fixes, nous utilisons l'estimation  $s_R^2$  calculée pour le modèle incluant les deux facteurs et nous changeons le nombre de degrés de liberté de la loi de Fisher en  $J - 1$  et  $IJ(K - 1)$ .

Nous décidons que le contraste  $L'$  est significativement différent de 0 au seuil  $\alpha$  si :

$$\frac{|\widehat{L}'(\mathbf{y})|}{\sqrt{\frac{(J-1)s_R^2}{IK} \sum_{j=1}^J l_j'^2}} \geq \sqrt{F(J-1, IJ(K-1); 1-\alpha)},$$

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 nous ne pouvons rejeter l'hypothèse d'absence de différence entre  $L'$  et 0 au seuil  $\alpha$ .

### Modèle à effets mixtes

Si un contraste  $L$  met en jeu des moyennes associées au facteur  $A$  à effets fixes, nous utilisons l'estimation  $s_{AB}^2$  calculée pour le modèle incluant les deux facteurs et nous changeons le nombre de degrés de liberté de la loi de Fisher en  $I - 1$  et  $(I - 1)(J - 1)$ .

Nous décidons que le contraste  $L$  est significativement différent de 0 au seuil  $\alpha$  si :

$$\frac{|\widehat{L}(\mathbf{y})|}{\sqrt{\frac{(I-1)s_{AB}^2}{JK} \sum_{i=1}^I l_i^2}} \geq \sqrt{F(I-1, (I-1)(J-1); 1-\alpha)}.$$

où  $F(I-1, (I-1)(J-1); 1-\alpha)$  est le  $100(1-\alpha)$  quantile de la loi de Fisher à  $I$  et  $(I-1)(J-1)$  degrés de liberté. Sinon nous ne pouvons rejeter l'hypothèse d'absence de différence entre  $L$  et 0 au seuil  $\alpha$ .

#### 9.7.4. Méthode de Bonferroni et méthodes associées

Nous adaptons ici, de manière similaire, la procédure de Bonferroni et toutes celles qui en découlent.

## 9.7 Deux facteurs avec répétitions

---

### Modèle à effets fixes

Si nous souhaitons réaliser **un nombre fini  $k$  de comparaisons** de contrastes associés au facteur  $A$  nous pourrons 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 nous souhaitons réaliser **un nombre fini  $k'$  de comparaisons** de contrastes associés au facteur  $B$  nous pourrons 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 nous souhaitons réaliser **un nombre fini  $k$  de comparaisons** de contrastes associés au facteur  $A$  nous pourrons 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 nous utiliserons 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.

### 9.7.5. Cas d'un plan non équilibré et des comparaisons deux à deux

Dans les deux situations, effets fixes ou mixtes, nous pouvons 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 nous souhaitons comparer les effets. Nous rappelons l'expression de  $K_h$  :

$$K_h = \frac{1}{\frac{1}{2} \left( \frac{1}{K_i} + \frac{1}{K_{i'}} \right)}$$

## 9.8. Deux facteurs emboîtés

Nous pouvons adapter les méthodes de comparaison des effets des différentes modalités, exposées aux paragraphes 9.1, 9.2 et 9.3 dans le cas de l'analyse de la variance à un facteur à effets fixes, au cas où nous considérons un modèle à deux facteurs sans interaction à effets fixes ou à effets mixtes. Nous commençons par détailler les modifications à faire pour utiliser les méthodes de Tukey et de Scheffé.

Nous reprenons ici les notations du paragraphe 4.1.1 ou du paragraphe 4.3.1 en fonction de la situation considérée. La différence fondamentale entre ces deux situations réside dans le fait que pour la première, **deux facteurs à effets fixes**, nous utiliserons 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**, nous utiliserons le **carré moyen associé au terme emboîté**  $s_{B|A}^2$  dans les statistiques de test.

### 9.8.1. 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 3.1.2, il peut être intéressant, en fonction de la significativité des différents tests auxquels nous avons procédé, d'étudier les contrastes exposés ci-dessous.

Le contraste met en jeu des moyennes associées au premier facteur  $A$ .

$$L = l_1\alpha_1 + l_2\alpha_2 + \dots + l_I\alpha_I,$$



## 9.8 Deux facteurs emboîtés

---

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 :

$$\widehat{L}(\mathbf{y}) = l_1 y_{1,\bullet,\bullet} + l_2 y_{2,\bullet,\bullet} + \cdots + l_I y_{I,\bullet,\bullet}.$$

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 :

$$\widehat{L}'(\mathbf{y}) = l'_{1(1)} y_{1,1,\bullet} + l'_{2(1)} y_{2,1,\bullet} + \cdots + l'_{J(I)} y_{J,I,\bullet}.$$

### Modèle à effets mixtes

Dans le cas d'un modèle mixte, c'est-à-dire le cas envisagé au paragraphe 3.3.2, il peut être intéressant, en fonction de la significativité du test de l'effet du facteur à effets fixes, d'étudier les contrastes 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 :

$$\widehat{L}(\mathbf{y}) = l_1 y_{1,\bullet,\bullet} + l_2 y_{2,\bullet,\bullet} + \cdots + l_I y_{I,\bullet,\bullet}.$$

### 9.8.2. Méthode de Tukey

#### Modèle à effets fixes

Si un contraste  $L$  met en jeu des moyennes associées au facteur  $A$  à effets fixes, nous utilisons l'estimation  $s_R^2$  calculée pour le modèle incluant les deux facteurs et nous changeons le nombre de degrés de liberté de la loi de l'étendue Studentisées en  $I$  et  $IJ(K-1)$ .

Nous décidons que le contraste  $L$  est significativement différent de 0 au seuil  $\alpha$  si :

$$\frac{|\widehat{L}(\mathbf{y})|}{\sqrt{\frac{s_R^2}{JK} \left( \frac{1}{2} \sum_{i=1}^I |l_i| \right)}} \geq q(I, IJ(K-1); 1-\alpha).$$

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 nous ne pouvons rejeter l'hypothèse d'absence de différence entre  $L$  et 0 au seuil  $\alpha$ .

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, nous utilisons l'estimation  $s_R^2$  calculée pour le modèle incluant les deux facteurs et nous changeons le nombre de degrés de liberté de la loi de l'étendue Studentisées en  $J$  et  $IJ(K-1)$ .

Nous décidons que le contraste  $L'$  est significativement différent de 0 au seuil  $\alpha$  si :

$$\frac{|\widehat{L}'(\mathbf{y})|}{\sqrt{\frac{s_R^2}{K} \left( \frac{1}{2} \sum_{j=1}^J |l'_j| \right)}} \geq q(J, IJ(K-1); 1-\alpha).$$

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 nous ne pouvons rejeter l'hypothèse d'absence de différence entre  $L'$  et 0 au seuil  $\alpha$ .

### Modèle à effets mixtes

Si un contraste  $L$  met en jeu des moyennes associées au facteur  $A$  à effets fixes, nous utilisons l'estimation  $s_{B|A}^2$  calculée pour le modèle incluant les deux facteurs et nous changeons le nombre de degrés de liberté de la loi de l'étendue Studentisées en  $I$  et  $I(J-1)$ .

Nous décidons que le contraste  $L$  est significativement différent de 0 au seuil  $\alpha$  si :

$$\frac{|\widehat{L}(\mathbf{y})|}{\sqrt{\frac{s_{B|A}^2}{JK} \left( \frac{1}{2} \sum_{i=1}^I |l_i| \right)}} \geq q(I, I(J-1); 1-\alpha).$$

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 nous ne pouvons rejeter l'hypothèse d'absence de différence entre  $L$  et 0 au seuil  $\alpha$ .

### 9.8.3. Méthode de Scheffé

#### Modèle à effets fixes

$L_{AB}$  est significativement différent de 0 au seuil  $\alpha$  si, en supposant que  $y_{i,j,\bullet} \geq y_{i',j',\bullet}$ , nous avons :

$$\frac{|\widehat{L}(\mathbf{y})|}{\sqrt{(IJ-1) \frac{2s_R^2}{K}}} \geq \sqrt{F(IJ-1, IJ(K-1); 1-\alpha)},$$

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 nous ne pouvons rejeter l'hypothèse d'absence de

## 9.8 Deux facteurs emboîtés

---

différence entre  $L_{AB}$  et 0 au seuil  $\alpha$ .

Si un contraste  $L$  met en jeu des moyennes associées au facteur  $A$  à effets fixes, nous utilisons l'estimation  $s_R^2$  calculée pour le modèle incluant les deux facteurs et nous changeons le nombre de degrés de liberté de la loi de Fisher en  $I - 1$  et  $IJ(K - 1)$ .

Nous décidons que le contraste  $L$  est significativement différent de 0 au seuil  $\alpha$  si :

$$\frac{|\widehat{L}(\mathbf{y})|}{\sqrt{\frac{(I-1)s_R^2}{JK} \sum_{i=1}^I l_i^2}} \geq \sqrt{F(I-1, IJ(K-1); 1-\alpha)},$$

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 nous ne pouvons rejeter l'hypothèse d'absence de différence entre  $L$  et 0 au seuil  $\alpha$ .

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, nous utilisons l'estimation  $s_R^2$  calculée pour le modèle incluant les deux facteurs et nous changeons le nombre de degrés de liberté de la loi de Fisher en  $J-1$  et  $IJ(K-1)$ .

Nous décidons que le contraste  $L'$  est significativement différent de 0 au seuil  $\alpha$  si :

$$\frac{|\widehat{L}'(\mathbf{y})|}{\sqrt{\frac{(J-1)s_R^2}{K} \sum_{j=1}^J l_j'^2}} \geq \sqrt{F(J-1, IJ(K-1); 1-\alpha)},$$

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 nous ne pouvons rejeter l'hypothèse d'absence de différence entre  $L'$  et 0 au seuil  $\alpha$ .

### Modèle à effets mixtes

Si un contraste  $L$  met en jeu des moyennes associées au facteur  $A$  à effets fixes, nous utilisons l'estimation  $s_{B|A}^2$  calculée pour le modèle incluant les deux facteurs et nous changeons le nombre de degrés de liberté de la loi de Fisher en  $I-1$  et  $I(J-1)$ .

Nous décidons que le contraste  $L$  est significativement différent de 0 au seuil  $\alpha$  si :

$$\frac{|\widehat{L}(\mathbf{y})|}{\sqrt{\frac{(I-1)s_{B|A}^2}{JK} \sum_{i=1}^I l_i^2}} \geq \sqrt{F(I-1, I(J-1); 1-\alpha)}.$$

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 nous ne pouvons rejeter l'hypothèse d'absence de différence entre  $L$  et 0 au seuil  $\alpha$ .

### 9.8.4. Méthode de Bonferroni et méthodes associées

Nous adaptons ici, de manière similaire, la procédure de Bonferroni et toutes celles qui en découlent.

#### Modèle à effets fixes

Si nous souhaitons réaliser **un nombre fini  $k$  de comparaisons** de contrastes associés au facteur  $A$  nous pourrions 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 nous souhaitons réaliser **un nombre fini  $k'$  de comparaisons** de contrastes associés aux niveaux du facteur  $B$  pour un niveau  $i$  du facteur  $A$  fixé nous pourrions 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 nous souhaitons réaliser **un nombre fini  $k$  de comparaisons** de contrastes associés au facteur  $A$  nous pourrions utiliser la valeur critique

$$t(I(J-1); 1 - \frac{\alpha}{2k}) \sqrt{\frac{s_{B|A}^2}{JK} \left( \sum_{i=1}^I l_i^2 \right)}$$

à la place de

$$q(I, I(J-1); 1 - \alpha) \sqrt{\frac{s_{B|A}^2}{JK} \left( \frac{1}{2} \sum_{i=1}^I |l_i| \right)}$$

## 9.8 Deux facteurs emboîtés

---

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 nous utiliserons avec la procédure de Bonferroni est :

$$\sqrt{2} t(I(J-1); 1 - \frac{\alpha}{2k}) \sqrt{\frac{s_{B|A}^2}{JK}}$$

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.

### 9.8.5. Cas d'un plan non équilibré et des comparaisons deux à deux

Dans les deux situations, effets fixes ou mixtes, nous pouvons 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 nous souhaitons comparer les effets. Nous rappelons l'expression de  $K_h$  :

$$K_h = \frac{1}{\frac{1}{2} \left( \frac{1}{K_i} + \frac{1}{K_{i'}} \right)}$$



# Chapitre 10

## 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  $F$  ou de Fisher, de l'analyse de la variance. Si nécessaire, nous renvoyons le lecteur à l'ouvrage [Dag98b] 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. Nous soulignons 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 nous diminuons le risque de commettre une erreur de première espèce, plus nous augmentons celui de commettre une erreur de deuxième espèce et réciproquement plus nous diminuons le risque de commettre une erreur de deuxième espèce, plus nous augmentons 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  $\alpha$  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.

## 10.1. Analyse de la variance à un facteur

### 10.1.1. Modèle à effets fixes

Nous reprenons ici les notations du paragraphe 2.1. Nous nous intéressons à la puissance  $1 - \beta_A$ , où  $\beta_A$  est le risque de commettre une erreur de deuxième espèce, du test  $F$  d'analyse de la variance pour le test de l'hypothèse

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

contre

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

Cette puissance  $1 - \beta_A$  est donnée par la formule suivante :

$$1 - \beta_A = \mathbb{P} \left[ F'(I - 1, I(J - 1); \lambda) > F(I - 1, I(J - 1); 1 - \alpha) \right],$$

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  $F'(I - 1, I(J - 1); \lambda_A)$  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  $\lambda_A$ . Ce paramètre de non-centralité  $\lambda$  vaut :

$$\lambda_A = \frac{J}{2\sigma^2} \sum_{i=1}^I \alpha_i^2.$$

Lorsque nous utilisons une loi de Fisher non centrale à  $\nu_1$  et  $\nu_2$  degrés de liberté et de paramètre de non-centralité  $\lambda$ , nous introduisons souvent le paramètre de non-centralité normalisé  $\phi_A$  défini par :

$$\phi_A = \sqrt{\frac{2\lambda_A}{\nu_1 + 1}}.$$

Nous obtenons ainsi dans notre situation :

$$\begin{aligned} \phi_A &= \sqrt{\frac{2\lambda_A}{\nu_1 + 1}} \\ &= \sqrt{\frac{2 \frac{J}{2\sigma^2} \sum_{i=1}^I \alpha_i^2}{I - 1 + 1}} \\ &= \frac{1}{\sigma} \sqrt{\frac{J}{I} \sum_{i=1}^I \alpha_i^2}. \end{aligned}$$

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  $\lambda_A$  devient :

$$\lambda_A = \frac{1}{2\sigma^2} \sum_{i=1}^I n_i \alpha_i^2.$$



## 10.1 Analyse de la variance à un facteur

---

Le paramètre de non-centralité normalisé  $\phi_A$  est alors :

$$\phi_A = \frac{1}{\sigma} \sqrt{\frac{1}{I} \sum_{i=1}^I n_i \alpha_i^2}.$$

Nous constatons que la puissance évolue bien comme indiqué au début de cette lorsque  $\alpha_i$ ,  $1 \leq i \leq I$ ,  $I$ ,  $n_i$ ,  $1 \leq i \leq I$ ,  $\sigma$  et  $\alpha$  varient.

### Puissance *a posteriori*

Nous obtenons 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 nous avons obtenues en réalisant l'analyse de la variance. Généralement nous considérons 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 nous faisons 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 nous ne connaissons pas a priori si le test que nous allons réaliser une fois que les expériences ont été réalisées sera significatif ou non à un seuil  $\alpha$  % 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 nous considérons qu'une puissance de 0,8 est satisfaisante ; dans certains cas nous visons même une puissance de 0,9.

Nous pouvons utiliser directement la formule ci-dessus pour déterminer le nombre de répétitions nécessaires à l'obtention d'une 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  $\sigma^2$ . 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  $\sum_{i=1}^I \alpha_i^2$ . Nous introduisons alors le concept de plus petite différence détectable  $\Delta$ , ce qui revient à évaluer le sensibilité du test en terme d'amplitude entre les effets des différents niveaux du facteur étudié. Ainsi nous chercherons à ce que la probabilité de détecter une amplitude  $|\alpha_i - \alpha_j|$  entre les effets  $\alpha_i$  et  $\alpha_j$  de deux modalités  $A_i$  et

$A_j$  différentes du facteur étudié strictement supérieure à  $\Delta$  soit élevée.

Ainsi pour faire le calcul de la puissance nous nous plaçons dans le pire des cas, c'est-à-dire celui pour lequel tous les effets sont nuls sauf deux  $\alpha_{i_0}$  et  $\alpha_{j_0}$  pour lesquels il existe un écart en valeur absolue égal à  $\Delta$ . Alors  $|\alpha_{i_0}| = |\alpha_{j_0}| = \Delta/2$ . Nous obtenons alors :

$$\begin{aligned} \lambda_A &= \frac{J}{2\sigma^2} \sum_{i=1}^I \alpha_i^2 \\ &= \frac{J}{2\sigma^2} (\alpha_{i_0}^2 + \alpha_{j_0}^2) \\ &= \frac{J}{2\sigma^2} \left( \left(\frac{\Delta}{2}\right)^2 + \left(\frac{\Delta}{2}\right)^2 \right) \\ &= \frac{J}{4\sigma^2} \Delta^2. \end{aligned}$$

Nous utilisons la formule ci-dessus pour déterminer les valeurs de  $J$  pour lesquelles la puissance  $1 - \beta_A$  est supérieure à une valeur  $1 - \beta_{A,0}$  fixée à l'avance, généralement 0,8 soit 80 %. Remarquons que là encore il est nécessaire de connaître  $\sigma^2$  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 nous considérons plutôt le paramètre de sensibilité  $\Delta/\sigma$  à la place de  $\Delta$ .

### 10.1.2. Modèle à effets aléatoires

Nous reprenons ici les notations du paragraphe 2.2. Nous nous intéressons à la puissance  $1 - \beta_A$ , où  $\beta_A$  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 :

$$1 - \beta_A = \mathbb{P} \left[ F(I - 1, I(J - 1)) > \frac{F(I - 1, I(J - 1); 1 - \alpha)}{1 + \frac{J\sigma_A^2}{\sigma^2}} \right],$$

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.

Si nous cherchons à évaluer cette puissance à l'aide d'une table, nous serons le plus souvent amenés à évaluer la quantité  $\lambda_A$  suivante :

$$\lambda_A = \sqrt{1 + \frac{J\sigma_A^2}{\sigma^2}}.$$

La différence fondamentale entre ce cas et le cas où le facteur est à effets fixes, exposés au paragraphe 10.1.1, est que le calcul de la puissance repose une loi de Fisher et non sur une loi de Fisher non-centrale.

### Puissance *a posteriori*

Nous obtenons 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 nous avons obtenues en réalisant l'analyse de la variance. Généralement nous considérons 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.

## 10.2. Analyse de la variance à deux facteurs sans interaction

Les idées développées dans la section 10.1 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 1.3 pour plus de détails sur la gestion des risques de première espèce des tests du tableau de l'analyse de la variance.

### 10.2.1. Modèle à effets fixes

Nous reprenons ici les notations du paragraphe 3.1.1.

Nous nous intéressons à la puissance  $1 - \beta_A$ , où  $\beta_A$  est le risque de commettre une erreur de deuxième espèce, du test  $F$  d'analyse de la variance pour le test de l'hypothèse

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

contre

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

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  $\phi_A$ . Ce paramètre de non-centralité normalisé  $\phi_A$  vaut :

$$\phi_A = \frac{1}{\sigma} \sqrt{\frac{J}{I} \sum_{i=1}^I \alpha_i^2}.$$

Nous nous intéressons à la puissance  $1 - \beta_B$ , où  $\beta_B$  est le risque de commettre une erreur de deuxième espèce, du test  $F$  d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \beta_1 = \beta_2 = \dots = \beta_J = 0}$$

contre

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

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  $\phi_B$ . Ce paramètre de non-centralité normalisé  $\phi_B$  vaut :

$$\phi_B = \frac{1}{\sigma} \sqrt{\frac{I}{J} \sum_{j=1}^J \beta_j^2}.$$

### 10.2.2. Modèle à effets aléatoires

Nous reprenons ici les notations du paragraphe 3.2.1. Nous nous intéressons à la puissance  $1 - \beta_A$ , où  $\beta_A$  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 :

$$1 - \beta_A = \mathbb{P} \left[ F(I - 1, (I - 1)(J - 1)) > \frac{F(I - 1, (I - 1)(J - 1); 1 - \alpha)}{1 + \frac{J\sigma_A^2}{\sigma^2}} \right],$$

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. Si nous cherchons à

## 10.2 Analyse de la variance à deux facteurs sans interaction

évaluer cette puissance à l'aide d'une table, nous serons le plus souvent amenés à évaluer la quantité  $\lambda_A$  suivante :

$$\lambda_A = \sqrt{1 + \frac{J\sigma_A^2}{\sigma^2}}.$$

Nous nous intéressons à la puissance  $1 - \beta_B$ , où  $\beta_B$  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 :

$$1 - \beta_B = \mathbb{P} \left[ F(J - 1, (I - 1)(J - 1)) > \frac{F(J - 1, (I - 1)(J - 1); 1 - \alpha)}{1 + \frac{I\sigma_B^2}{\sigma^2}} \right],$$

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. Si nous cherchons à évaluer cette puissance à l'aide d'une table, nous serons le plus souvent amenés à évaluer la quantité  $\lambda_B$  suivante :

$$\lambda_B = \sqrt{1 + \frac{I\sigma_B^2}{\sigma^2}}.$$

La différence fondamentale entre ce cas et le cas où le facteur est à effets fixes, exposé au paragraphe 10.2.1, est que le calcul de la puissance repose une loi de Fisher et non sur une loi de Fisher non-centrale.

### 10.2.3. Modèle à effets mixtes

Nous reprenons ici les notations du paragraphe 3.3.1. Nous nous intéressons à la puissance  $1 - \beta_A$ , où  $\beta_A$  est le risque de commettre une erreur de deuxième espèce, du test  $F$  d'analyse de la variance pour le test de l'hypothèse

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

contre

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

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  $\phi_A$ . Ce paramètre de non-centralité normalisé  $\phi_A$  vaut :

$$\phi_A = \frac{1}{\sigma} \sqrt{\frac{J}{I} \sum_{i=1}^I \alpha_i^2}.$$

Nous nous intéressons à la puissance  $1 - \beta_B$ , où  $\beta_B$  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 :

$$1 - \beta_B = \mathbb{P} \left[ F(J - 1, (I - 1)(J - 1)) > \frac{F(J - 1, (I - 1)(J - 1); 1 - \alpha)}{1 + \frac{I\sigma_B^2}{\sigma^2}} \right],$$

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. Si nous cherchons à évaluer cette puissance à l'aide d'une table, nous serons le plus souvent amenés à évaluer la quantité  $\lambda_B$  suivante :

$$\lambda_B = \sqrt{1 + \frac{I\sigma_B^2}{\sigma^2}}.$$

Les puissances des tests sont donc identiques à celles où le facteur à effets fixes serait avec un autre facteur à effets fixes et le facteur à effets aléatoires serait avec un autre facteur à effets aléatoires.

### 10.3. Analyse de la variance à deux facteurs avec interaction

Les idées développées dans la section 10.1 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 1.3 pour plus de détails sur la gestion des risques de première espèce des tests du tableau de l'analyse de la variance.

### 10.3.1. Modèle à effets fixes

Nous reprenons ici les notations du paragraphe 3.1.2.

Nous nous intéressons à la puissance  $1 - \beta_A$ , où  $\beta_A$  est le risque de commettre une erreur de deuxième espèce, du test  $F$  d'analyse de la variance pour le test de l'hypothèse

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

contre

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

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  $\phi_A$ . Ce paramètre de non-centralité normalisé  $\phi_A$  vaut :

$$\phi_A = \frac{1}{\sigma} \sqrt{\frac{JK}{I} \sum_{i=1}^I \alpha_i^2}.$$

Nous nous intéressons à la puissance  $1 - \beta_B$ , où  $\beta_B$  est le risque de commettre une erreur de deuxième espèce, du test  $F$  d'analyse de la variance pour le test de l'hypothèse

$$\boxed{\mathcal{H}_0 : \beta_1 = \beta_2 = \dots = \beta_J = 0}$$

contre

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

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  $\phi_B$ . Ce paramètre de non-centralité normalisé  $\phi_B$  vaut :

$$\phi_B = \frac{1}{\sigma} \sqrt{\frac{IK}{J} \sum_{j=1}^J \beta_j^2}.$$

Nous nous intéressons à la puissance  $1 - \beta_{AB}$ , où  $\beta_{AB}$  est le risque de commettre une erreur de deuxième espèce, du test  $F$  d'analyse de la variance pour le test de l'hypothèse

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

contre

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

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  $\phi_{AB}$ . Ce paramètre de non-centralité normalisé  $\phi_{AB}$  vaut :

$$\phi_{AB} = \frac{1}{\sigma} \sqrt{\frac{K \sum_{i=1}^I \sum_{j=1}^J (\alpha\beta)_{i,j}^2}{(I-1)(J-1) + 1}}.$$

### 10.3.2. Modèle à effets aléatoires

Nous reprenons ici les notations du paragraphe 3.2.2.

Nous nous intéressons à la puissance  $1 - \beta_A$ , où  $\beta_A$  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 :

$$1 - \beta_A = \mathbb{P} \left[ F(I-1, (I-1)(J-1)) > \frac{F(I-1, (I-1)(J-1); 1 - \alpha)}{1 + \frac{JK\sigma_A^2}{\sigma^2 + K\sigma_{AB}^2}} \right],$$

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. Si nous cherchons à évaluer cette puissance à l'aide d'une table, nous serons le plus souvent amenés à évaluer la quantité  $\lambda_A$  suivante :

$$\lambda_A = \sqrt{1 + \frac{JK\sigma_A^2}{\sigma^2 + K\sigma_{AB}^2}}.$$

Nous nous intéressons à la puissance  $1 - \beta_B$ , où  $\beta_B$  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

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



### 10.3 Analyse de la variance à deux facteurs avec interaction

Cette puissance  $1 - \beta_B$  est donnée par la formule suivante :

$$1 - \beta_B = \mathbb{P} \left[ F(J - 1, (I - 1)(J - 1)) > \frac{F(J - 1, (I - 1)(J - 1); 1 - \alpha)}{1 + \frac{IK\sigma_B^2}{\sigma^2 + K\sigma_{AB}^2}} \right],$$

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. Si nous cherchons à évaluer cette puissance à l'aide d'une table, nous serons le plus souvent amenés à évaluer la quantité  $\lambda_B$  suivante :

$$\lambda_B = \sqrt{1 + \frac{IK\sigma_B^2}{\sigma^2 + K\sigma_{AB}^2}}.$$

Nous nous intéressons à la puissance  $1 - \beta_{AB}$ , où  $\beta_{AB}$  est le risque de commettre une erreur de deuxième espèce, du test  $F$  d'analyse de la variance pour le test de l'hypothèse

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

contre

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

Cette puissance  $1 - \beta_{AB}$  est donnée par la formule suivante :

$$1 - \beta_{AB} = \mathbb{P} \left[ F((I - 1)(J - 1), IJ(K - 1)) > \frac{F((I - 1)(J - 1), IJ(K - 1); 1 - \alpha)}{1 + \frac{K\sigma_{AB}^2}{\sigma^2}} \right],$$

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. Si nous cherchons à évaluer cette puissance à l'aide d'une table, nous serons le plus souvent amenés à évaluer la quantité  $\lambda_{AB}$  suivante :

$$\lambda_{AB} = \sqrt{1 + \frac{K\sigma_{AB}^2}{\sigma^2}}.$$

La différence fondamentale entre ce cas et le cas où le facteur est à effets fixes, exposé au paragraphe 3.1.2, est que le calcul de la puissance repose une loi de Fisher et non sur une loi de Fisher non-centrale.

#### 10.3.3. Modèle à effets mixtes

Nous reprenons ici les notations du paragraphe 3.3.2.

Nous nous intéressons à la puissance  $1 - \beta_A$ , où  $\beta_A$  est le risque de commettre une erreur de deuxième espèce, du test  $F$  d'analyse de la variance pour le test de l'hypothèse

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

contre

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

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  $\phi_A$ . Ce paramètre de non-centralité normalisé  $\phi_A$  vaut :

$$\phi_A = \sqrt{\frac{JK \sum_{i=1}^I \alpha_i^2}{I(\sigma^2 + K\sigma_{AB}^2)}}.$$

Nous nous intéressons à la puissance  $1 - \beta_B$ , où  $\beta_B$  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 :

$$1 - \beta_B = \mathbb{P} \left[ F(J - 1, IJ(K - 1)) > \frac{F(J - 1, IJ(K - 1); 1 - \alpha)}{1 + \frac{IK\sigma_B^2}{\sigma^2 + K\sigma_{AB}^2}} \right],$$

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. Si nous cherchons à évaluer cette puissance à l'aide d'une table, nous serons le plus souvent amenés à évaluer la quantité  $\lambda_B$  suivante :

$$\lambda_B = \sqrt{1 + \frac{IK\sigma_B^2}{\sigma^2 + K\sigma_{AB}^2}}.$$

Nous nous intéressons à la puissance  $1 - \beta_{AB}$ , où  $\beta_{AB}$  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 :

$$1 - \beta_{AB} = \mathbb{P} \left[ F((I-1)(J-1), IJ(K-1)) > \frac{F((I-1)(J-1), IJ(K-1); 1-\alpha)}{1 + \frac{K\sigma_{AB}^2}{\sigma^2}} \right],$$

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. Si nous cherchons à évaluer cette puissance à l'aide d'une table, nous serons le plus souvent amenés à évaluer la quantité  $\lambda_{AB}$  suivante :

$$\lambda_{AB} = \sqrt{1 + \frac{K\sigma_{AB}^2}{\sigma^2}}.$$

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.

## 10.4. Analyse de la variance à deux facteurs emboîtés

Les idées développées dans la section 10.1 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 1.3 pour plus de détails sur la gestion des risques de première espèce des tests du tableau de l'analyse de la variance.

### 10.4.1. Modèle à effets fixes

Nous reprenons ici les notations du paragraphe 4.1.1. Nous nous intéressons à la puissance  $1 - \beta_A$ , où  $\beta_A$  est le risque de commettre une erreur de deuxième espèce, du test  $F$  d'analyse de la variance pour le test de l'hypothèse

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

contre

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

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  $\phi_A$ . Ce paramètre de non-centralité normalisé  $\phi_A$  vaut :

$$\phi_A = \frac{1}{\sigma} \sqrt{\frac{JK}{I} \sum_{i=1}^I \alpha_i^2}.$$

Nous nous intéressons 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

$$\mathcal{H}_0 : \beta_{1(1)} = \beta_{2(1)} = \dots = \beta_{J(1)} = \beta_{1(2)} = \dots = \beta_{J(I)} = 0$$

contre

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

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 :

$$\phi_{B(A)} = \frac{1}{\sigma} \sqrt{\frac{K \sum_{i=1}^I \sum_{j=1}^J \beta_{j(i)}^2}{I(J-1) + 1}}.$$

### 10.4.2. Modèle à effets aléatoires

Nous reprenons ici les notations du paragraphe 4.2.1.

Nous nous intéressons à la puissance  $1 - \beta_A$ , où  $\beta_A$  est le risque de commettre une erreur de deuxième espèce, du test  $F$  d'analyse de la variance pour le test de l'hypothèse

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

contre

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

Cette puissance  $1 - \beta_A$  est donnée par la formule suivante :

$$1 - \beta_A = \mathbb{P} \left[ F(I-1, I(J-1)) > \frac{F(I-1, I(J-1); 1-\alpha)}{1 + \frac{JK\sigma_A^2}{\sigma^2 + K\sigma_{B|A}^2}} \right],$$

## 10.4 Analyse de la variance à deux facteurs emboîtés

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. Si nous cherchons à évaluer cette puissance à l'aide d'une table, nous serons le plus souvent amenés à évaluer la quantité  $\lambda_A$  suivante :

$$\lambda_A = \sqrt{1 + \frac{JK\sigma_A^2}{\sigma^2 + K\sigma_{B|A}^2}}.$$

Nous nous intéressons 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 :

$$1 - \beta_{B(A)} = \mathbb{P} \left[ F(I(J-1), IJ(K-1)) > \frac{F(I(J-1), IJ(K-1); 1-\alpha)}{1 + \frac{K\sigma_{B|A}^2}{\sigma^2}} \right],$$

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. Si nous cherchons à évaluer cette puissance à l'aide d'une table, nous serons le plus souvent amenés à évaluer la quantité  $\lambda_{B(A)}$  suivante :

$$\lambda_{B(A)} = \sqrt{1 + \frac{K\sigma_{B|A}^2}{\sigma^2}}.$$

### 10.4.3. Modèle à effets mixtes

Nous reprenons ici les notations du paragraphe 4.3.1.

Nous nous intéressons à la puissance  $1 - \beta_A$ , où  $\beta_A$  est le risque de commettre une erreur de deuxième espèce, du test  $F$  d'analyse de la variance pour le test de l'hypothèse

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

contre

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

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  $\phi_A$ . Ce paramètre de non-centralité normalisé  $\phi_A$  vaut :

$$\phi_A = \sqrt{\frac{JK \sum_{i=1}^I \alpha_i^2}{I(\sigma^2 + K\sigma_{B|A}^2)}}.$$

Nous nous intéressons 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 :

$$1 - \beta_{B(A)} = \mathbb{P} \left[ F(I(J - 1), IJ(K - 1)) > \frac{F(I(J - 1), IJ(K - 1); 1 - \alpha)}{1 + K \frac{\sigma_{B|A}^2}{\sigma^2}} \right],$$

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. Si nous cherchons à évaluer cette puissance à l'aide d'une table, nous serons le plus souvent amenés à évaluer la quantité  $\lambda_{B(A)}$  suivante :

$$\lambda_{B(A)} = \sqrt{1 + K \frac{\sigma_{B|A}^2}{\sigma^2}}.$$

## 10.5. Analyse de la variance à trois facteurs sans interaction

Les idées développées dans la section 10.1 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 1.3 pour plus de détails sur la gestion des risques de première espèce des tests du tableau de l'analyse de la variance.

### 10.5.1. Modèle à effets fixes

Nous reprenons ici les notations du paragraphe 5.1.1.

Pour des problèmes de concision nous ne présenterons 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 nous voulons calculer la puissance du test.

Nous nous intéressons à la puissance  $1 - \beta_A$ , où  $\beta_A$  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

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

contre

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

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  $\phi_A$ . Ce paramètre de non-centralité normalisé  $\phi_A$  vaut :

$$\phi_A = \frac{1}{\sigma} \sqrt{\frac{JK}{I} \sum_{i=1}^I \alpha_i^2}.$$

Nous nous intéressons à la puissance  $1 - \beta_{AB}$ , où  $\beta_{AB}$  est le risque de commettre une erreur de deuxième espèce, du test  $F$  d'analyse de la variance pour le test de l'hypothèse

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

contre

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

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  $\phi_{AB}$ . Ce paramètre de non-centralité normalisé  $\phi_{AB}$  vaut :

$$\phi_{AB} = \frac{1}{\sigma} \sqrt{\frac{K \sum_{i=1}^I \sum_{j=1}^J (\alpha\beta)_{ij}^2}{(I - 1)(J - 1) + 1}}.$$

### 10.5.2. Modèle à effets aléatoires

Nous reprenons ici les notations du paragraphe 5.2.1.

Nous nous intéressons à la puissance  $1 - \beta_A$ , où  $\beta_A$  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. Nous pouvons néanmoins calculer la puissance de ce pseudo test  $F$ .

Cette puissance  $1 - \beta_A$  est donnée par la formule suivante :

$$1 - \beta_A = \mathbb{P} \left[ F(\nu_1, \nu_2) > \frac{F(\nu_1, \nu_2; 1 - \alpha)}{1 + \frac{JK\sigma_A^2}{\sigma^2 + K\sigma_{AB}^2}} \right],$$

où  $\nu_1 = I - 1$  et

$$\nu_2 = \frac{(s_{AB}^2 + s_{AC}^2 - s_R^2)^2}{\frac{(s_{AB}^2)^2}{(I-1)(J-1)} + \frac{(s_{AC}^2)^2}{(I-1)(K-1)} + \frac{(s_R^2)^2}{(I-1)(J-1)(K-1)}},$$

$F(I - 1, \nu_2; 1 - \alpha)$  est le  $100(1 - \alpha)$  quantile de la loi de Fisher à  $(I - 1)$  et  $\nu_2$  degrés de liberté et  $F(I - 1, \nu_2)$  est une variable aléatoire qui suit une loi de Fisher à  $I - 1$  et  $\nu_2$  degrés de liberté. Si nous cherchons à évaluer cette puissance à l'aide d'une table, nous serons le plus souvent amenés à évaluer la quantité  $\lambda_A$  suivante :

$$\lambda_A = \sqrt{1 + \frac{JK\sigma_A^2}{\sigma^2 + K\sigma_{AB}^2}}.$$

Nous nous intéressons à la puissance  $1 - \beta_{AB}$ , où  $\beta_{AB}$  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 :

$$1 - \beta_{AB} = \mathbb{P} \left[ F(\nu_1, \nu_2) > \frac{F(\nu_1, \nu_2; 1 - \alpha)}{1 + \frac{K\sigma_{AB}^2}{\sigma^2}} \right],$$



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. Si nous cherchons à évaluer cette puissance à l'aide d'une table, nous serons le plus souvent amenés à évaluer la quantité  $\lambda_{AB}$  suivante :

$$\lambda_{AB} = \sqrt{1 + \frac{K\sigma_{AB}^2}{\sigma^2}}.$$

La différence fondamentale entre ce cas et le cas où le facteur est à effets fixes, exposés au paragraphe 10.5.1, est que le calcul de la puissance repose une loi de Fisher et non sur une loi de Fisher non-centrale.

### 10.5.3. Modèle à effets mixtes

Nous reprenons ici les notations du paragraphe 5.3.1.

**Premier cas : Deux facteurs sont à effets fixes,  $A$  et  $B$ , et un facteur  $C$  est à effets aléatoires.**

Nous nous intéressons à la puissance  $1 - \beta_A$ , où  $\beta_A$  est le risque de commettre une erreur de deuxième espèce, du test  $F$  d'analyse de la variance pour le test de l'hypothèse

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

contre

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

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  $\phi_A$ . Ce paramètre de non-centralité normalisé  $\phi_A$  vaut :

$$\phi_A = \sqrt{\frac{JK \sum_{i=1}^I \alpha_i^2}{I(\sigma^2 + J\sigma_{AC}^2)}}.$$

Nous nous intéressons à la puissance  $1 - \beta_C$ , où  $\beta_C$  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 :

$$1 - \beta_C = \mathbb{P} \left[ F(K - 1, (I - 1)(J - 1)(K - 1)) > \frac{F(K - 1, (I - 1)(J - 1)(K - 1); 1 - \alpha)}{1 + \frac{IJ\sigma_C^2}{\sigma^2}} \right],$$

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. Si nous cherchons à évaluer cette puissance à l'aide d'une table, nous serons le plus souvent amenés à évaluer la quantité  $\lambda_C$  suivante :

$$\lambda_C = \sqrt{1 + \frac{IJ\sigma_C^2}{\sigma^2}}.$$

Nous nous intéressons à la puissance  $1 - \beta_{AB}$ , où  $\beta_{AB}$  est le risque de commettre une erreur de deuxième espèce, du test  $F$  d'analyse de la variance pour le test de l'hypothèse

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

contre

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

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  $\phi_{AB}$ . Ce paramètre de non-centralité normalisé  $\phi_{AB}$  vaut :

$$\phi_{AB} = \frac{1}{\sigma} \sqrt{\frac{K \sum_{i=1}^I \sum_{j=1}^J (\alpha\beta)_{ij}^2}{(I - 1)(J - 1) + 1}}.$$

Nous nous intéressons à la puissance  $1 - \beta_{AC}$ , où  $\beta_{AC}$  est le risque de commettre une erreur de deuxième espèce, du test  $F$  d'analyse de la variance pour le test de l'hypothèse

$$\mathcal{H}_0 : \sigma_{AC}^2 = 0$$

contre

$$\mathcal{H}_1 : \sigma_{AC}^2 \neq 0.$$

Cette puissance  $1 - \beta_{AC}$  est donnée par la formule suivante :

$$1 - \beta_{AC} = \mathbb{P} \left[ F(\nu_1, \nu_2) > \frac{F(\nu_1, \nu_2; 1 - \alpha)}{1 + \frac{J\sigma_{AC}^2}{\sigma^2}} \right],$$

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. Si nous cherchons à évaluer cette puissance à l'aide d'une table, nous serons le plus souvent amenés à évaluer la quantité  $\lambda_{AC}$  suivante :

$$\lambda_{AC} = \sqrt{1 + \frac{J\sigma_{AC}^2}{\sigma^2}}.$$

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

Nous nous intéressons à la puissance  $1 - \beta_A$ , où  $\beta_A$  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

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

contre

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

La loi de la statistique de ce test n'est connue que de manière approximative. Nous pouvons 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

$$\nu_2 = \frac{(s_{AB}^2 + s_{AC}^2 - s_R^2)^2}{\frac{(s_{AB}^2)^2}{(I - 1)(J - 1)} + \frac{(s_{AC}^2)^2}{(I - 1)(K - 1)} + \frac{(s_R^2)^2}{(I - 1)(J - 1)(K - 1)}},$$

$F(I - 1, \nu_2; 1 - \alpha)$  est le  $100(1 - \alpha)$  quantile de la loi de Fisher à  $(I - 1)$  et  $\nu_2$  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  $\nu_2$  degrés de liberté et de paramètre de non-centralité normalisé  $\phi_{AB}$ . Ce paramètre de non-centralité normalisé  $\phi_A$  vaut :

$$\phi_A = \sqrt{\frac{1}{I} \frac{JK \sum_{i=1}^I \alpha_i^2}{\sigma^2 + K\sigma_{AB}^2 + J\sigma_{AC}^2}}.$$

Nous nous intéressons à la puissance  $1 - \beta_B$ , où  $\beta_B$  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 :

$$1 - \beta_B = \mathbb{P} \left[ F(J - 1, (J - 1)(K - 1)) > \frac{F(J - 1, (J - 1)(K - 1); 1 - \alpha)}{1 + \frac{IK\sigma_B^2}{\sigma^2 + I\sigma_{BC}^2}} \right],$$

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. Si nous cherchons à évaluer cette puissance à l'aide d'une table, nous serons le plus souvent amenés à évaluer la quantité  $\lambda_B$  suivante :

$$\lambda_B = \sqrt{1 + \frac{IK\sigma_B^2}{\sigma^2 + I\sigma_{BC}^2}}.$$

Nous nous intéressons à la puissance  $1 - \beta_{AB}$ , où  $\beta_{AB}$  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 :

$$1 - \beta_{AB} = \mathbb{P} \left[ F(\nu_1, \nu_2) > \frac{F(\nu_1, \nu_2; 1 - \alpha)}{1 + \frac{K\sigma_{AB}^2}{\sigma^2}} \right],$$

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. Si nous cherchons à évaluer cette puissance à l'aide d'une table, nous serons le plus souvent amenés à évaluer la quantité  $\lambda_{AB}$  suivante :

$$\lambda_{AB} = \sqrt{1 + \frac{K\sigma_{AB}^2}{\sigma^2}}.$$

## 10.6. Analyse de la variance à trois facteurs avec interaction

Les idées développées dans la section 10.1 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 1.3 pour plus de détails sur la gestion des risques de première espèce des tests du tableau de l'analyse de la variance.

### 10.6.1. Modèle à effets fixes

Nous reprenons ici les notations du paragraphe 5.1.2.

Pour des problèmes de concision nous ne présenterons 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 nous voulons calculer la puissance du test.

Nous nous intéressons à la puissance  $1 - \beta_A$ , où  $\beta_A$  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

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

contre

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

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  $\phi_A$ . Ce paramètre de non-centralité normalisé  $\phi_A$  vaut :

$$\phi_A = \frac{1}{\sigma} \sqrt{\frac{JKL}{I} \sum_{i=1}^I \alpha_i^2}.$$

Nous nous intéressons à la puissance  $1 - \beta_{AB}$ , où  $\beta_{AB}$  est le risque de commettre une erreur de deuxième espèce, du test  $F$  d'analyse de la variance pour le test de l'hypothèse

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

contre

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

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  $\phi_{AB}$ . Ce paramètre de non-centralité normalisé  $\phi_{AB}$  vaut :

$$\phi_{AB} = \frac{1}{\sigma} \sqrt{\frac{KL \sum_{i=1}^I \sum_{j=1}^J (\alpha\beta)_{ij}^2}{(I - 1)(J - 1) + 1}}.$$

Nous nous intéressons à la puissance  $1 - \beta_{ABC}$ , où  $\beta_{ABC}$  est le risque de commettre une erreur de deuxième espèce, du test  $F$  d'analyse de la variance pour le test de l'hypothèse

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

contre

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

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  $\phi_{ABC}$ . Ce paramètre de non-centralité normalisé  $\phi_{ABC}$  vaut :

$$\phi_{ABC} = \frac{1}{\sigma} \sqrt{\frac{L \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (\alpha\beta\gamma)_{ijk}^2}{(I - 1)(J - 1)(K - 1) + 1}}.$$

### 10.6.2. Modèle à effets aléatoires

Nous reprenons ici les notations du paragraphe 5.2.2.

Nous nous intéressons à la puissance  $1 - \beta_A$ , où  $\beta_A$  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. Nous pouvons néanmoins calculer la puissance de ce pseudo test  $F$ .

Cette puissance  $1 - \beta_A$  est donnée par la formule suivante :

$$1 - \beta_A = \mathbb{P} \left[ F(\nu_1, \nu_2) > \frac{F(\nu_1, \nu_2; 1 - \alpha)}{1 + \frac{JKL\sigma_A^2}{\sigma^2 + L\sigma_{ABC}^2 + KL\sigma_{AB}^2 + JL\sigma_{AC}^2}} \right],$$

où  $\nu_1 = I - 1$  et

$$\nu_2 = \frac{(s_{AB}^2 + s_{AC}^2 - s_{ABC}^2)^2}{\frac{(s_{AB}^2)^2}{(I-1)(J-1)} + \frac{(s_{AC}^2)^2}{(I-1)(K-1)} + \frac{(s_{ABC}^2)^2}{(I-1)(J-1)(K-1)}},$$

$F(I - 1, \nu_2; 1 - \alpha)$  est le  $100(1 - \alpha)$  quantile de la loi de Fisher à  $(I - 1)$  et  $\nu_2$  degrés de liberté et  $F(I - 1, \nu_2)$  est une variable aléatoire qui suit une loi de Fisher à  $I - 1$  et  $\nu_2$  degrés de liberté. Si nous cherchons à évaluer cette puissance à l'aide d'une table, nous serons le plus souvent amenés à évaluer la quantité  $\lambda_A$  suivante :

$$\lambda_A = \sqrt{1 + \frac{JKL\sigma_A^2}{\sigma^2 + L\sigma_{ABC}^2 + KL\sigma_{AB}^2 + JL\sigma_{AC}^2}}.$$

Nous nous intéressons à la puissance  $1 - \beta_{AB}$ , où  $\beta_{AB}$  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 :

$$1 - \beta_{AB} = \mathbb{P} \left[ F(\nu_1, \nu_2) > \frac{F(\nu_1, \nu_2; 1 - \alpha)}{1 + \frac{KL\sigma_{AB}^2}{\sigma^2 + L\sigma_{ABC}^2}} \right],$$

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. Si

nous cherchons à évaluer cette puissance à l'aide d'une table, nous serons le plus souvent amenés à évaluer la quantité  $\lambda_{AB}$  suivante :

$$\lambda_{AB} = \sqrt{1 + \frac{KL\sigma_{AB}^2}{\sigma^2 + L\sigma_{ABC}^2}}.$$

Nous nous intéressons à la puissance  $1 - \beta_{ABC}$ , où  $\beta_{ABC}$  est le risque de commettre une erreur de deuxième espèce, du test  $F$  d'analyse de la variance pour le test de l'hypothèse

$$\mathcal{H}_0 : \sigma_{ABC}^2 = 0$$

contre

$$\mathcal{H}_1 : \sigma_{ABC}^2 \neq 0.$$

Cette puissance  $1 - \beta_{ABC}$  est donnée par la formule suivante :

$$1 - \beta_{ABC} = \mathbb{P} \left[ F(\nu_1, \nu_2) > \frac{F(\nu_1, \nu_2; 1 - \alpha)}{1 + \frac{L\sigma_{ABC}^2}{\sigma^2}} \right],$$

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. Si nous cherchons à évaluer cette puissance à l'aide d'une table, nous serons le plus souvent amenés à évaluer la quantité  $\lambda_{ABC}$  suivante :

$$\lambda_{ABC} = \sqrt{1 + \frac{L\sigma_{ABC}^2}{\sigma^2}}.$$

La différence fondamentale entre ce cas et le cas où le facteur est à effets fixes, exposé au paragraphe 10.6.1, est que le calcul de la puissance repose une loi de Fisher et non sur une loi de Fisher non-centrale.

### 10.6.3. Modèle à effets mixtes

Nous reprenons ici les notations du paragraphe 5.3.2.

**Premier cas : Deux facteurs sont à effets fixes,  $A$  et  $B$ , et un facteur  $C$  est à effets aléatoires.**

Nous nous intéressons à la puissance  $1 - \beta_A$ , où  $\beta_A$  est le risque de commettre une erreur de deuxième espèce, du test  $F$  d'analyse de la variance pour le test de l'hypothèse

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

contre

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



## 10.6 Analyse de la variance à trois facteurs avec interaction

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  $\phi_A$ . Ce paramètre de non-centralité normalisé  $\phi_A$  vaut :

$$\phi_A = \sqrt{\frac{JKL \sum_{i=1}^I \alpha_i^2}{I(\sigma^2 + JL\sigma_{AC}^2)}}.$$

Nous nous intéressons à la puissance  $1 - \beta_C$ , où  $\beta_C$  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 :

$$1 - \beta_C = \mathbb{P} \left[ F(K - 1, IJK(L - 1)) > \frac{F(K - 1, IJK(L - 1); 1 - \alpha)}{1 + \frac{IJ\sigma_C^2}{\sigma^2}} \right],$$

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. Si nous cherchons à évaluer cette puissance à l'aide d'une table, nous serons le plus souvent amenés à évaluer la quantité  $\lambda_C$  suivante :

$$\lambda_C = \sqrt{1 + \frac{IJ\sigma_C^2}{\sigma^2}}.$$

Nous nous intéressons à la puissance  $1 - \beta_{AB}$ , où  $\beta_{AB}$  est le risque de commettre une erreur de deuxième espèce, du test  $F$  d'analyse de la variance pour le test de l'hypothèse

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

contre

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

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  $\phi_{AB}$ . Ce paramètre de non-centralité normalisé  $\phi_{AB}$  vaut :

$$\phi_{AB} = \sqrt{\frac{KL \sum_{i=1}^I \sum_{j=1}^J (\alpha\beta)_{ij}^2}{((I - 1)(J - 1) + 1)(\sigma^2 + L\sigma_{ABC}^2)}}.$$

Nous nous intéressons à la puissance  $1 - \beta_{AC}$ , où  $\beta_{AC}$  est le risque de commettre une erreur de deuxième espèce, du test  $F$  d'analyse de la variance pour le test de l'hypothèse

$$\mathcal{H}_0 : \sigma_{AC}^2 = 0$$

contre

$$\mathcal{H}_1 : \sigma_{AC}^2 \neq 0.$$

Cette puissance  $1 - \beta_{AC}$  est donnée par la formule suivante :

$$1 - \beta_{AC} = \mathbb{P} \left[ F(\nu_1, \nu_2) > \frac{F(\nu_1, \nu_2; 1 - \alpha)}{1 + \frac{JL\sigma_{AC}^2}{\sigma^2}} \right],$$

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. Si nous cherchons à évaluer cette puissance à l'aide d'une table, nous serons le plus souvent amenés à évaluer la quantité  $\lambda_{AC}$  suivante :

$$\lambda_{AC} = \sqrt{1 + \frac{JL\sigma_{AC}^2}{\sigma^2}}.$$

Nous nous intéressons à la puissance  $1 - \beta_{ABC}$ , où  $\beta_{ABC}$  est le risque de commettre une erreur de deuxième espèce, du test  $F$  d'analyse de la variance pour le test de l'hypothèse

$$\mathcal{H}_0 : \sigma_{ABC}^2 = 0$$

contre

$$\mathcal{H}_1 : \sigma_{ABC}^2 \neq 0.$$

Cette puissance  $1 - \beta_{ABC}$  est donnée par la formule suivante :

$$1 - \beta_{ABC} = \mathbb{P} \left[ F(\nu_1, \nu_2) > \frac{F(\nu_1, \nu_2; 1 - \alpha)}{1 + \frac{L\sigma_{ABC}^2}{\sigma^2}} \right],$$

## 10.6 Analyse de la variance à trois facteurs avec interaction

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. Si nous cherchons à évaluer cette puissance à l'aide d'une table, nous serons le plus souvent amenés à évaluer la quantité  $\lambda_{ABC}$  suivante :

$$\lambda_{ABC} = \sqrt{1 + \frac{L\sigma_{ABC}^2}{\sigma^2}}.$$

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

Nous nous intéressons à la puissance  $1 - \beta_A$ , où  $\beta_A$  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

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

contre

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

La loi de la statistique de ce test n'est connue que de manière approximative. Nous pouvons 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

$$\nu_2 = \frac{(s_{AB}^2 + s_{AC}^2 - s_{ABC}^2)^2}{\frac{(s_{AB}^2)^2}{(I-1)(J-1)} + \frac{(s_{AC}^2)^2}{(I-1)(K-1)} + \frac{(s_{ABC}^2)^2}{(I-1)(J-1)(K-1)}},$$

$F(I-1, \nu_2; 1-\alpha)$  est le  $100(1-\alpha)$  quantile de la loi de Fisher à  $(I-1)$  et  $\nu_2$  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  $\nu_2$  degrés de liberté et de paramètre de non-centralité normalisé  $\phi_A$ . Ce paramètre de non-centralité normalisé  $\phi_A$  vaut :

$$\phi_A = \sqrt{\frac{1}{I} \frac{JKL \sum_{i=1}^I \alpha_i^2}{\sigma^2 + L\sigma_{ABC}^2 + KL\sigma_{AB}^2 + JL\sigma_{AC}^2}}.$$

Nous nous intéressons à la puissance  $1 - \beta_B$ , où  $\beta_B$  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 :

$$1 - \beta_B = \mathbb{P} \left[ F(J-1, (J-1)(K-1)) > \frac{F(J-1, (J-1)(K-1); 1-\alpha)}{1 + \frac{IKL\sigma_B^2}{\sigma^2 + IL\sigma_{BC}^2}} \right],$$

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. Si nous cherchons à évaluer cette puissance à l'aide d'une table, nous serons le plus souvent amenés à évaluer la quantité  $\lambda_B$  suivante :

$$\lambda_B = \sqrt{1 + \frac{IKL\sigma_B^2}{\sigma^2 + IL\sigma_{BC}^2}}.$$

Nous nous intéressons à la puissance  $1 - \beta_{AB}$ , où  $\beta_{AB}$  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 :

$$1 - \beta_{AB} = \mathbb{P} \left[ F(\nu_1, \nu_2) > \frac{F(\nu_1, \nu_2; 1-\alpha)}{1 + \frac{KL\sigma_{AB}^2}{\sigma^2 + L\sigma_{ABC}^2}} \right],$$

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. Si nous cherchons à évaluer cette puissance à l'aide d'une table, nous serons le plus souvent amenés à évaluer la quantité  $\lambda_{AB}$  suivante :

$$\lambda_{AB} = \sqrt{1 + \frac{KL\sigma_{AB}^2}{\sigma^2 + L\sigma_{ABC}^2}}.$$

Nous nous intéressons à la puissance  $1 - \beta_{BC}$ , où  $\beta_{BC}$  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 :

$$1 - \beta_{BC} = \mathbb{P} \left[ F(\nu_1, \nu_2) > \frac{F(\nu_1, \nu_2; 1 - \alpha)}{1 + KL \frac{\sigma_{BC}^2}{\sigma^2}} \right],$$

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. Si nous cherchons à évaluer cette puissance à l'aide d'une table, nous serons le plus souvent amenés à évaluer la quantité  $\lambda_{BC}$  suivante :

$$\lambda_{BC} = \sqrt{1 + KL \frac{\sigma_{BC}^2}{\sigma^2}}.$$

Nous nous intéressons à la puissance  $1 - \beta_{ABC}$ , où  $\beta_{ABC}$  est le risque de commettre une erreur de deuxième espèce, du test  $F$  d'analyse de la variance pour le test de l'hypothèse

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

contre

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

Cette puissance  $1 - \beta_{ABC}$  est donnée par la formule suivante :

$$1 - \beta_{ABC} = \mathbb{P} \left[ F(\nu_1, \nu_2) > \frac{F(\nu_1, \nu_2; 1 - \alpha)}{1 + \frac{L\sigma_{ABC}^2}{\sigma^2}} \right],$$

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. Si nous cherchons à évaluer cette puissance à l'aide d'une table, nous serons le plus souvent amenés à évaluer la quantité  $\lambda_{ABC}$  suivante :

$$\lambda_{ABC} = \sqrt{1 + \frac{L\sigma_{ABC}^2}{\sigma^2}}.$$

## 10.7. Analyse de la variance à trois facteurs totalement emboîtés

Les idées développées dans la section 10.1 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 1.3 pour plus de détails sur la gestion des risques de première espèce des tests du tableau de l'analyse de la variance.

### 10.7.1. Modèle à effets fixes

Nous reprenons ici les notations du paragraphe 6.1.1.

Nous nous intéressons à la puissance  $1 - \beta_A$ , où  $\beta_A$  est le risque de commettre une erreur de deuxième espèce, du test  $F$  d'analyse de la variance pour le test de l'hypothèse

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

contre

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

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  $\phi_A$ . Ce paramètre de non-centralité normalisé  $\phi_A$  vaut :

$$\phi_A = \frac{1}{\sigma} \sqrt{\frac{JKL}{I} \sum_{i=1}^I \alpha_i^2}.$$

Nous nous intéressons 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

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

contre

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

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

## 10.7 Analyse de la variance à trois facteurs totalement emboîtés

paramètre de non-centralité normalisé  $\phi_{B(A)}$ . Ce paramètre de non-centralité normalisé  $\phi_{B(A)}$  vaut :

$$\phi_{B(A)} = \frac{1}{\sigma} \sqrt{\frac{KL \sum_{i=1}^I \sum_{j=1}^J \beta_{j(i)}^2}{I(J-1) + 1}}.$$

Nous nous intéressons 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

$$\mathcal{H}_0 : \gamma_{1(1(1))} = \gamma_{2(1(1))} = \dots = \gamma_{K(1(1))} = \gamma_{1(1(2))} = \dots = \gamma_{K(J(I))} = 0$$

contre

$$\mathcal{H}_1 : \exists (i_0, j_0, k_0) \in \{1, 2, \dots, I\} \times \{1, 2, \dots, J\} \times \{1, 2, \dots, K\} \mid \gamma_{k_0(j_0(i_0))} \neq 0.$$

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 :

$$\phi_{C(B(A))} = \frac{1}{\sigma} \sqrt{\frac{L \sum_{i=1}^I \sum_{j=1}^J \gamma_{k(j(i))}^2}{IJ(L-1) + 1}}.$$

### 10.7.2. Modèle à effets aléatoires

Nous reprenons ici les notations du paragraphe 6.2.1.

Nous nous intéressons à la puissance  $1 - \beta_A$ , où  $\beta_A$  est le risque de commettre une erreur de deuxième espèce, du test  $F$  d'analyse de la variance pour le test de l'hypothèse

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

contre

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

Cette puissance  $1 - \beta_A$  est donnée par la formule suivante :

$$1 - \beta_A = \mathbb{P} \left[ F(I-1, I(J-1)) > \frac{F(I-1, I(J-1); 1 - \alpha)}{1 + \frac{JKL\sigma_A^2}{\sigma^2 + L\sigma_{C|B|A}^2 + KL\sigma_{B|A}^2}} \right],$$

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. Si nous cherchons à évaluer cette puissance à l'aide d'une table, nous serons le plus souvent amenés à évaluer la quantité  $\lambda_A$  suivante :

$$\lambda_A = \sqrt{1 + \frac{JKL\sigma_A^2}{\sigma^2 + L\sigma_{C|B|A}^2 + KL\sigma_{B|A}^2}}.$$

Nous nous intéressons 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

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

contre

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

Cette puissance  $1 - \beta_{B(A)}$  est donnée par la formule suivante :

$$1 - \beta_{B(A)} = \mathbb{P} \left[ F(I(J-1), IJ(K-1)) > \frac{F(I(J-1), IJ(K-1); 1-\alpha)}{1 + \frac{KL\sigma_{B|A}^2}{\sigma^2 + L\sigma_{C|B|A}^2}} \right],$$

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. Si nous cherchons à évaluer cette puissance à l'aide d'une table, nous serons le plus souvent amenés à évaluer la quantité  $\lambda_{B(A)}$  suivante :

$$\lambda_{B(A)} = \sqrt{1 + \frac{KL\sigma_{B|A}^2}{\sigma^2 + L\sigma_{C|B|A}^2}}.$$

Nous nous intéressons 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

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

contre

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

Cette puissance  $1 - \beta_{C(B(A))}$  est donnée par la formule suivante :

$$1 - \beta_{C(B(A))} = \mathbb{P} \left[ F(\nu_1, \nu_2) > \frac{F(\nu_1, \nu_2; 1-\alpha)}{1 + \frac{L\sigma_{C|B|A}^2}{\sigma^2}} \right],$$

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.



$1), IJK(L - 1))$  est une variable aléatoire qui suit une loi de Fisher à  $IJ(L - 1)$  et  $IJK(L - 1)$  degrés de liberté. Si nous cherchons à évaluer cette puissance à l'aide d'une table, nous serons le plus souvent amenés à évaluer la quantité  $\lambda_{C(B(A))}$  suivante :

$$\lambda_{C(B(A))} = \sqrt{1 + \frac{L\sigma_{C|B|A}^2}{\sigma^2}}.$$

### 10.7.3. Modèle à effets mixtes

Nous reprenons ici les notations du paragraphe 6.3.

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

Nous nous intéressons à la puissance  $1 - \beta_A$ , où  $\beta_A$  est le risque de commettre une erreur de deuxième espèce, du test  $F$  d'analyse de la variance pour le test de l'hypothèse

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

contre

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

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  $\phi_A$ . Ce paramètre de non-centralité normalisé  $\phi_A$  vaut :

$$\phi_A = \sqrt{\frac{JKL \sum_{i=1}^I \alpha_i^2}{I(\sigma^2 + K\sigma_{C|B|A}^2)}}.$$

Nous nous intéressons 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

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

contre

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

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 :

$$\phi_{B(A)} = \sqrt{\frac{KL \sum_{i=1}^I \sum_{j=1}^J \beta_{j(i)}^2}{I(J-1) + 1 \sigma^2 + L\sigma_{C|B|A}^2}}.$$

Nous nous intéressons 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

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

contre

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

Cette puissance  $1 - \beta_{C(B(A))}$  est donnée par la formule suivante :

$$1 - \beta_{C(B(A))} = \mathbb{P} \left[ F(\nu_1, \nu_2) > \frac{F(\nu_1, \nu_2; 1 - \alpha)}{1 + \frac{L\sigma_{C|B|A}^2}{\sigma^2}} \right],$$

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. Si nous cherchons à évaluer cette puissance à l'aide d'une table, nous serons le plus souvent amenés à évaluer la quantité  $\lambda_{C(B(A))}$  suivante :

$$\lambda_{C(B(A))} = \sqrt{1 + \frac{L\sigma_{C|B|A}^2}{\sigma^2}}.$$

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

Nous nous intéressons à la puissance  $1 - \beta_A$ , où  $\beta_A$  est le risque de commettre une erreur de deuxième espèce, du test  $F$  d'analyse de la variance pour le test de l'hypothèse

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

contre

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

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

## 10.7 Analyse de la variance à trois facteurs totalement emboîtés

Fisher non-centrale à  $I - 1$  et  $I(J - 1)$  degrés de liberté et de paramètre de non-centralité normalisé  $\phi_A$ . Ce paramètre de non-centralité normalisé  $\phi_A$  vaut :

$$\phi_A = \sqrt{\frac{JKL \sum_{i=1}^I \alpha_i^2}{I(\sigma^2 + KL\sigma_{B|A}^2 + L\sigma_{C|B|A}^2)}}.$$

Nous nous intéressons 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

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

contre

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

Cette puissance  $1 - \beta_{B(A)}$  est donnée par la formule suivante :

$$1 - \beta_{B(A)} = \mathbb{P} \left[ F(I(J - 1), IJ(K - 1)) > \frac{F(I(J - 1), IJ(K - 1); 1 - \alpha)}{1 + \frac{KL\sigma_{B|A}^2}{\sigma^2 + L\sigma_{C|B|A}^2}} \right],$$

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. Si nous cherchons à évaluer cette puissance à l'aide d'une table, nous serons le plus souvent amenés à évaluer la quantité  $\lambda_{B(A)}$  suivante :

$$\lambda_{B(A)} = \sqrt{1 + \frac{KL\sigma_{B|A}^2}{\sigma^2 + L\sigma_{C|B|A}^2}}.$$

Nous nous intéressons 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

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

contre

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

Cette puissance  $1 - \beta_{C(B(A))}$  est donnée par la formule suivante :

$$1 - \beta_{C(B(A))} = \mathbb{P} \left[ F(\nu_1, \nu_2) > \frac{F(\nu_1, \nu_2; 1 - \alpha)}{1 + \frac{L\sigma_{C|B|A}^2}{\sigma^2}} \right],$$

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.

$1), IJK(L - 1)$  est une variable aléatoire qui suit une loi de Fisher à  $IJ(L - 1)$  et  $IJK(L - 1)$  degrés de liberté. Si nous cherchons à évaluer cette puissance à l'aide d'une table, nous serons le plus souvent amenés à évaluer la quantité  $\lambda_{C(B(A))}$  suivante :

$$\lambda_{C(B(A))} = \sqrt{1 + \frac{L\sigma_{C|B|A}^2}{\sigma^2}}.$$

# Bibliographie

- [Cha73] P. Chapouille. *Planification et analyse des expériences*. Masson, Paris, 1973.
- [Dag98a] P. Dagnélie. *Statistique Théorique et Appliquée*, volume 1. De Boeck & Larcier, Bruxelles, 1998.
- [Dag98b] P. Dagnélie. *Statistique Théorique et Appliquée*, volume 2. De Boeck & Larcier, Bruxelles, 1998.
- [DBBP97] G. Demengel, P. Bénichou, N. Boy, and J.-P. Pouget. *Probabilités, Statistique inférentielle, Fiabilité, Outils pour l'ingénieur*. Ellipses, 1997.
- [Dro01] J.J. Dreesbeke. *Éléments de Statistique*. Université de Bruxelles, Bruxelles, 4ème édition, 2001.
- [Fal05] B. Falissard. *Comprendre et utiliser les statistiques dans les sciences de la vie*. Abrégés. Masson, Paris, 3<sup>ème</sup> édition, 2005.
- [Fle86] J.L. Fleiss. *The Design and Analysis of Clinical Experiments*. John Wiley, New-York, 1986.
- [HPer] F. Husson and J. Pagès. *Statistiques générales pour utilisateurs*. PUR, Rennes, a trouver.
- [KRB96] S. Kotz, C.B. Read, and N. Balakrishnan, editors. *Encyclopædia Of Statistical Sciences*. Wiley-Interscience, 2nd edition, 1996.
- [Par74] G. Parreins. *Techniques Statistiques : moyens rationnels de choix et de décision*. Dunod technique, Paris, 1974.
- [Phi82] G. Philippeau. *Théorie des plans d'expérience*. ICTF, Paris, 1982.
- [Ric89] W.R. Rice. Analysing tables of statistical tests. *Evolution*, 43(1) :223–225, 1989.
- [SA00] H. Sahai and M.I. Ageel. *The Analysis of Variance : Fixed, Random and Mixed Models*. Birkhäuser, Boston, 2000.