République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Mohamed Khider, Biskra

Faculté des Sciences Exactes

Département de Mathématiques

Mémoire présenté en vue de l'obtention du Diplôme :

MASTER en Mathématiques

Option : Statistique

Par

BENBRAHIM Hakima

Titre :

# Anova multivariée et application

Membres du Comité d'Examen :

| | | | |
|---|---|---|---|
| Dr. | Zouaoui Nour Elhouda | UMKB | Président |
| MAA. | Roubi Affef | UMKB | Encadreur |
| Dr. | Khemissi Zahia | UMKB | Examinatrice |

Juin 2025

## Dédicace

Avec l'aide de dieu, j'ai pu réaliser ce modeste travail que je dédie :

A mes parents, que je ne peux remarcierai assez pour leur soutien, amour et affection tout au long de mon parcours mes études.

A mes trés chéres soeurs.

A mon seul frére.

A mes chéres amies.

A tout mes proffeseurs que j 'ai connus durant mes études.

A ceux qui partage ma réussite de prés ou loin

**Ben Brahim hakima**

## REMERCIEMENTS

Tout d'abord, nous tenons à remercier notre Bon Dieu Tout Puissant (ALLAH) source de volonté, de patience et de courage.

Nos vifs remerciements à l'egard de **Madame Roubi Affef** pour sa profonde patience, ses séance de lecteure, de correction, pour ses orientations et ses consiels pleins de courage et de renforcements.

Je tiens également à remercier chaleureusement les membres du jury **Dr.Zouaoui Nour Elhouda** et **Dr.Khemissi Zahia** davoir accepté d'évaluer et de juger mon travail.

Mes vifs remerciements vont également à tout ceux qui ont contribués de prés ou de loin à la réalisation de ce travail.

A nos camarades et nos amies sans exception.

# Table des matières

# Liste des tableaux

# Introduction

L'évolution de la science statistique au fil des siècles a conduit à l'émergence de nombreux outils statistiques qui contribuent à l'interprétation et à l'analyse des données de manière précise et objective. Parmi ces outils, l'analyse de la variance (ANOVA) occupe une place essentielle.

Cette dernière est considérée comme une méthode qui consiste à comparer plusieurs moyennes. Bien qu'elle soit efficace dans nombreuses applications. Mais dans certains cas les données peuvent êtres complexes et multidimensionnels, ce qui limite les capacités de l'ANOVA à les traiter. C'est dans ce contexte que l'analyse de la variance multivariée (MANOVA) s'impose comme un outil avancé, permettant ainsi aux chercheurs à comparer les moyennes à condition d'être des moyennes vectorielles de plusieurs échantillons, et offrant une vision plus complète et précise des données multidimensionnelles. ([1], [3], [11])

Ce mémoire à pour objectif d'étudier à la fois l'analyse de la variance (ANOVA) et l'analyse de la variance multivariée (MANOVA) dans les cas de présence d'un ou deux facteurs, en expliquant tout d'abord les concepts théoriques puis en passant à l'aspect pratique.

Pour atteindre cet objectif, le mémoire sera structuré en trois chapitres :

**Le premier chapitre** sera consacré à l'étude des bases théoriques de l'ANOVA à un et deux facteurs, en mettant en évidence ses modèles et ses étapes.

**Le deuxième chapitre** portera sur la méthode de l'analyse de la variance multiple

(MANOVA) à un et à deux facteurs, avec une présentations de ses modèles, ses hypothèses et ses tests statistiques les plus courant (Lambda Wilk's, Trace de Pillai, etc.).

**Le troisième chapitre** est consacré à la présentation des applications pratiques à l'aide du logiciel R afin d'analyser des données réelles pour comprendre plus les deux chapitres précédents.

# Chapitre 1

# Analyse de la variance univariée

Dans ce chapitre, nous aborderons la méthode d'analyse de la variance univariée (ANOVA) à un et à deux facteurs. Cette méthode est une technique qui sert à tester l'influence d'un (ou de plusieurs) facteur (s) qualitative (s) sur une variable quantitative.

## 1.1 ANOVA à un facteur (ANOVA 1)

**Définition 1.1.1** *L'analyse de la variance à un facteur teste l'effet d'un facteur contrôlé $A$ ayant $I$ modalités (groupes) sur les moyennes d'une variable quantitative $Y$.[3]*

### 1.1.1 Les données et le modèle

#### Les données

Nous voulons étudier l'effet d'un facteur $A$, que nous supposerons à $I$ niveaux, sur une variable quantitative $Y$. L'hypothèse est que le facteur $A$ influe uniquement sur les moyennes des distributions de chacun des $I$ groupes et non sur leur variance. Pour chaque niveau $i$ du facteur $A$ (avec $1 \leq i \leq I$), nous disposons de $n_i$ mesures de $Y$, notées $Y_{ij}$ avec $j = 1, ...., n_i$. Dans la suite, nous noterons par $n$ le nombre total d'observations, ie. $n = \sum_{i=1}^{I} n_i$. Les données sont généralement présentées sous la forme du tableau 1.1

| Niveau du Facteur$A$ | $A_1$ | $A_2$ | $\ldots$ | $A_i$ | $\ldots$ | $A_I$ |
|---|---|---|---|---|---|---|
| | $Y_{11}$<br>$Y_{12}$<br>$\vdots$<br>$Y_{1n_1}$ | $Y_{21}$<br>$Y_{22}$<br>$\vdots$<br>$Y_{2n_2}$ | $\ldots$ | $Y_{i1}$<br>$Y_{i2}$<br>$\vdots$<br>$Y_{in_i}$ | $\ldots$ | $Y_{I1}$<br>$Y_{I2}$<br>$\vdots$<br>$Y_{In_I}$ |
| Effectifs | $n_1$ | $n_2$ | $\ldots$ | $n_i$ | $\ldots$ | $n_I$ |

TAB. 1.1 – Les données de l'ANOVA à 1 facteur.

### Le modèle de l'ANOVA 1

Le modèle est donné par

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \text{ avec } i = \overline{1, I} \text{ , } j = \overline{1, n_i} \text{ et } \varepsilon_{ij} \rightsquigarrow \mathcal{N}(0, \sigma^2), \tag{1.1}$$

où

$Y_{ij}$ : est la $j^{ème}$ réalisation de la variable quantitative $Y$ dans le $i^{ème}$ échantillon

$\varepsilon_{ij}$ : sont les erreurs de mesure.

**Remarque 1.1.1** *Nous décomposons parfois $\mu_i$ en :*

$$\mu_i = \mu + \alpha_i \quad \textit{avec} \quad \sum_{i=1}^{I} n_i \alpha_i = 0,$$

tels que

$\mu$ : représente un effet global inconnu du facteur ;

$\alpha_i$ : représente l'effet principal (spécifique) inconnu du niveau $i$ du facteur $A$.

Donc, le modèle dans ce cas s'écrit :

$$Y_{ij} = \underbrace{\mu + \alpha_i}_{\mu_i} + \varepsilon_{ij} \text{ avec } \varepsilon_{ij} \rightsquigarrow \mathcal{N}(0, \sigma^2).$$

Si on retient le modèle (1.1) alors le test à réaliser est défini par

$$H_0 : \mu_1 = \mu_2 = .... = \mu_I = \mu \quad \text{contre } H_1 : \exists i, j \in \{1, 2..., I\} \text{ telle que } \mu_i \neq \mu_j. \tag{1.2}$$

Pour réaliser le test défini par (1.2), il faut suivre les étapes suivantes.

### 1.1.2 Les étapes de l'ANOVA 1

**Etape 1 : (Vérification des conditions)**

– Pour effectuer le test décrit dans (1.2), trois conditions doivent être préalablement remplies

1. Les $I$ échantillons comparés sont indépendants.
2. La variable quantitative étudiée suit une loi normale dans chacune des $I$ populations comparées.
3. Les $I$ populations comparées ont la même variance, ce qui signifie qu'il y a homogénéité des variances ou homoscédasticité.

**Etape 2 : (Calcul des moyennes et variances)**

Si ces conditions sont respectées, nous peuvons appliquer la méthode d'ANOVA 1 pour réaliser le test défini par (1.2). Pour cela, nous aurons besoin des quantités statistiques suivantes

– La moyenne de toutes les observations

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i} Y_{ij} \text{ avec } n = \sum_{i=1}^{I} n_i \,;$$

– La moyenne de chaque échantillon

$$\overline{Y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij} \text{ pour } i = \overline{1, I} \,;$$

– La variance de chaque échantillon

$$\widehat{\sigma}_i^2 = \frac{1}{n_i}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_i)^2 \text{ pour } i = \overline{1, I}\,;$$

– La variance de toutes les observations

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y})^2.$$

**Etape 3 : (Décomposition de la variance)**

En remarquant que

$$Y_{ij} - \overline{Y} = Y_{ij} - \overline{Y}_i + \overline{Y}_i - \overline{Y}$$

alors

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y})^2 &= \frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i}((Y_{ij} - \overline{Y}_i) + (\overline{Y}_i - \overline{Y}))^2 \\ &= \frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_i)^2 + \frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i}(\overline{Y}_i - \overline{Y})^2 + 2\frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_i)(\overline{Y}_i - \overline{Y})\end{aligned}$$

or

$$\sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_i)(\overline{Y}_i - \overline{Y}) = \sum_{i=1}^{I}(\overline{Y}_i - \overline{Y})\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_i) = 0$$

donc

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y})^2 = \frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_i)^2 + \frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i}(\overline{Y}_i - \overline{Y})^2, \tag{1.3}$$

en multipliant (1.3), par $n$ on obtient :

$$\underbrace{\sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y})^2}_{SC_{Tot}} = \underbrace{\sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_i)^2}_{SC_{Rés}} + \underbrace{\sum_{i=1}^{I}\sum_{j=1}^{n_i}(\overline{Y}_i - \overline{Y})^2}_{SC_{Fac}},$$

où

- $SC_{Tot}$ : est la variation totale qui représente la dispersion des données autour de la moyenne générale.
- $SC_{Fac}$ : est la variation due au facteur (variation inter-groupes) qui représente la dispersion des moyennes autour de la moyenne générale.
- $SC_{Rés}$ : est la variation résiduelle (variation intra-groupes) qui représente la dispersion des données à l'intérieur de chaque échantillon autour de sa moyenne.

**Etape 4 : (Décision)**

Si le facteur $A$ n'a pas d'influence sur la variable $Y$, l'hypothèse nulle à tester est

$$H_0 : \mu_1 = \mu_2 = ... = \mu_I \iff \alpha_1 = \alpha_2 = ... = \alpha_I = 0.$$

Sous l'hypothèse de normalité des erreurs $\varepsilon_i$, nous avons la statistique $F_{obs}$ définie par le rapport des carrés moyens associés au facteur ($CM_{Fac}$) et les carrés moyens résiduels ($CM_{Rés}$)

$$F_{obs} = \frac{CM_{Fac}}{CM_{Rés}}$$

suit une loi de Fisher de degrés de liberté $(I-1)$ et $(n-I)$ notée par $F_{(I-1,n-I)}$, tels que

$$CM_{Fac} = \frac{SC_{Fac}}{I-1}, \, CM_{Rés} = \frac{SC_{Rés}}{n-I}.$$

**Remarque 1.1.2** *Les statistiques $CM_{Fac}$ et $CM_{Rés}$ suivent une loi de Khi-deux de degrés de liberté $(I-1)$ et $(n-I)$ respectivement ($CM_{Fac} \sim \mathcal{X}^2_{I-1}$ et $CM_{Rés} \sim \mathcal{X}^2_{n-I}$).*

Pour un seuil de risque donné $\alpha$, les tables de Fisher nous fournissent une valeur critique $f_\alpha$ telle que

$$P(F_{obs} < f_\alpha) = 1 - \alpha,$$

alors :

si $f_{obs} < f_\alpha \Rightarrow$ nous ne pouvons pas rejeter $H_0$ (le facteur n'a aucune influence sur le caractère étudié),

si $f_{obs} \geq f_\alpha \Rightarrow$ nous rejetons $H_0$ (le facteur influe sur le caractère étudié), avec $f_{obs}$ est la réalisation de la variable (statistique) $F_{obs}$.

Les résultats d'une ANOVA 1 sont souvent présentés dans un tableau sous la forme suivante

| Variation | $ddl$ | $SC$ | $CM$ | $F_{obs}$ |
|---|---|---|---|---|
| Facteur | $I-1$ | $SC_{Fac}$ | $CM_{Fac}$ | $F_{obs} = \frac{CM_{Fac}}{CM_{Rés}}$ |
| Résiduelle | $n-I$ | $SC_{Rés}$ | $CM_{Rés}$ | |
| Total | $n-1$ | $SC_{Tot}$ | | |

TAB. 1.2 – Tableau d'analyse de la variance à un facteur.

## 1.2 ANOVA à deux facteurs (ANOVA 2) (Plan équilibré)

**Définition 1.2.1** *L'analyse de la variance à deux facteurs teste l'effet de deux facteurs contrôlés $A$ et $B$(variables qualitatives) ayant respectivement $I$ et $J$ modalités sur les moyennes d'une variable quantitative $Y$.[3]*

### 1.2.1 Les données et le modèle

#### Les données

Nous cherchons donc à étudier l'effet de deux facteurs qualitatifs $A$ et $B$ sur une variable quantitative $Y$. Nous supposons que le facteur $A$ a $I$ niveaux (modalités) et que le facteur $B$ a $J$ niveaux. Pour chaque couple de niveaux $(i, j)$ nous disposons de $K$ mesures de $Y$ notées $Y_{ijk}$ avec $i = \overline{1, I}$, $j = \overline{1, J}$ et $k = \overline{1, K}$.

Nous présentons généralement les données à l'aide du tableau suivant :

| $N^0$ | $B_1$ | $B_2$ | $\cdots$ | $B_J$ |
|---|---|---|---|---|
| $A_1$ | $Y_{111}$ $Y_{112}$ $\vdots$ $Y_{11K}$ | $Y_{121}$ $Y_{122}$ $\vdots$ $Y_{12K}$ | $\cdots$ | $Y_{1J1}$ $Y_{1J2}$ $\vdots$ $Y_{1JK}$ |
| $A_2$ | $Y_{211}$ $Y_{212}$ $\vdots$ $Y_{21K}$ | $Y_{221}$ $Y_{222}$ $\vdots$ $Y_{22K}$ | $\cdots$ | $Y_{2J1}$ $Y_{2J2}$ $\vdots$ $Y_{2JK}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\cdots$ | $\vdots$ |
| $A_I$ | $Y_{I11}$ $Y_{I12}$ $\vdots$ $Y_{I1K}$ | $Y_{I21}$ $Y_{I22}$ $\vdots$ $Y_{I2K}$ | $\cdots$ | $Y_{IJ1}$ $Y_{IJ2}$ $\vdots$ $Y_{IJK}$ |

TAB. 1.3 – Les données d'ANOVA 2 avec répétitions.

**Le modèle de l'ANOVA 2**

Le modèle s'écrit sous la forme

$$Y_{ijk} = \mu_{ij} + \varepsilon_{ijk} \quad \text{avec} \quad i = \overline{1, I}, j = \overline{1, J} \text{ et } k = \overline{1, K}, \tag{1.4}$$

où $Y_{ijk}$ est la $k^{ème}$ réalisation de la variable quantitative $Y$, lorsque on fixe le premier facteur à la $i^{ème}$ modalité et le deuxième facteur à la $j^{ème}$ modalité et $\varepsilon_{ijk}$ sont les erreurs de mesure (inconnues) de plus $E(\varepsilon_{ijk}) = 0$.

Le modèle (1.4) peut être réécrit sous sa forme détaillée comme suit

$$Y_{ijk} = \mu + \alpha_i + \beta_j + c_{ij} + \varepsilon_{ijk}, \text{ avec } i = \overline{1, I}, j = \overline{1, J} \text{ et } k = \overline{1, K}\,; \tag{1.5}$$

ce qui s'explique que la réalisation de la variable $Y$ est un cumule d'une constante $\mu$ (indépendante des deux facteurs), de l'effet du premier facteur $\alpha$, de l'effet du deuxième facteur $\beta$ et de l'effet d'interaction des deux facteurs $c$ et de l'erreur de mesure $\varepsilon$.

Les contraintes naturelles sont

$$\sum_{i=1}^{I}\alpha_i = 0 \ ,\sum_{j=1}^{J}\beta_j = 0 \ ,\sum_{i=1}^{I}c_{ij} = \sum_{j=1}^{J}c_{ij} = 0, \text{ pour } i = 1, ..., I \text{ et } j = 1, ..., J.$$

Les hypothèses à tester associées au modèle (1.4) sont

$$H_0 :'' \forall \left\{ \begin{array}{c} i \in \{1, 2, ..., I\} \\ j \in \{1, 2, ..., J\} \end{array} \right\}, \ \mu_{ij} = \mu'',$$

contre

$$H_1 :'' \exists \left\{ \begin{array}{c} i \in \{1, 2, ..., I\} \\ j \in \{1, 2, ..., J\} \end{array} \right\}, \text{ telle que } \mu_{ij} \neq \mu''.$$

En revanche, si le modèle de référence choisi est le modèle (1.5), alors l'analyse de variance à deux facteurs avec répétitions implique la réalisation simultanée de trois tests de Fisher, dont la formulation est :

**Effet du premier facteur**

$H_0$ : "les paramètres $\alpha_i$ sont tous nuls" contre $H_1$ : "les paramètres $\alpha_i$ ne sont pas tous nuls".

**Effet du second facteur**

$H_0$ : "les paramètres $\beta_j$ sont tous nuls " contre $H_1$ : "les paramètres $\beta_j$ ne sont pas tous nuls".

**Effet de l'interaction des deux facteurs**

$H_0$ : "les paramètres $c_{ij}$ sont tous nuls" contre $H_1$ : "les paramètres $c_{ij}$ ne sont pas tous nuls."

### 1.2.2 Les étapes de l'ANOVA 2

L'application d'une ANOVA à deux facteurs se déroule généralement en quatre étapes principales.

**Etape 1 :(Conditions)**

1. Les $IJ$ échantillons comparés sont multuellement indépendants.
2. La variable quantitative étudiée suit une loi normale dans chacune des $IJ$ populations comparées.
3. Les $IJ$ populations comparées ont la même variance, ce qui implique une homogénéité des variances (ou homoscédasticité).

**Etape 2 :(Calcul de moyennes et variances)**

– La moyenne globale de toutes les observations

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}Y_{ijk} \text{ avec } n = IJK\,;$$

– La moyenne de chaque échantillon

$$\overline{Y}_{ij\cdot} = \frac{1}{K}\sum_{k=1}^{K}Y_{ijk} \text{ pour } i = \overline{1, I} \text{ et } j = \overline{1, J}\,;$$

– La moyenne de chaque modalité du premier facteur

$$\overline{Y}_{i\cdot\cdot} = \frac{1}{JK}\sum_{j=1}^{J}\sum_{k=1}^{K}Y_{ijk} \text{ pour } i = \overline{1, I}\,;$$

– La moyenne de chaque modalité du deuxième facteur

$$\overline{Y}_{\cdot j\cdot} = \frac{1}{IK}\sum_{i=1}^{I}\sum_{k=1}^{K}Y_{ijk} \text{ pour } j = \overline{1, J}\,;$$

– La somme des carrés des erreurs totale

$$SC_{Tot} = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(Y_{ijk} - \overline{Y})^2\,;$$

– La somme des carrés des erreurs résiduelles

$$SC_{Rés} = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(Y_{ijk} - \overline{Y}_{ij\cdot})^2\,;$$

– La somme des carrés des erreurs du premier facteur

$$SC_{\alpha} = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(\overline{Y}_{i\cdot\cdot} - \overline{Y})^2\,;$$

– La somme des carrés des erreurs du deuxième facteur

$$SC_{\beta} = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(\overline{Y}_{\cdot j\cdot} - \overline{Y})^2\,;$$

– La somme des carrées des erreurs des deux facteurs

$$SC_{c} = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(\overline{Y}_{ij\cdot} - \overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdot j\cdot} + \overline{Y})^2\,;$$

L'équation d'analyse de la variance dans ce cas s'écrit

$$SC_{Tot} = SC_{Rés} + SC_{\alpha} + SC_{\beta} + SC_{c}. \tag{1.6}$$

**Etape 3 :(Calcul des Carrés moyens)**

À partir de la décomposition (1.6), l'idée la plus naturelle est que le facteur ou l'interaction des facteurs n' a pas d'impact sur le caractère étudié si la variation intergroupes (engendrée par les deux facteurs ou/et leur interaction) associée au caractère est négligeable par rapport aux fluctuations individuelles. Pour effectuer une comparaison entre ces quantités, nous considérons les carrés moyens suivants

Carré moyen due aux fluctuations individuelles : $CM_{Rés} = \frac{SC_{Rés}}{IJ(K-1)}$.

Carré moyen de mesure de l'effet du premier facteur : $CM_{\alpha} = \frac{SC_{\alpha}}{(I-1)}$.

Carré moyen de mesure de l'effet du second facteur : $CM_{\beta} = \frac{SC_{\beta}}{(J-1)}$.

Carré moyen de mesure de l'effet de l'interaction entre les deux facteurs : $CM_c = \frac{SC_c}{(I-1)(J-1)}$.

Il convient de noter que, si les trois conditions mentionnées précédemment (Indépendance, Normalité et Homogénéité) sont satisfaites, alors sous l'hypothèse nulle,

$$\frac{CM_{\alpha}}{CM_{Rés}} \sim F_{((I-1),IJ(K-1))},$$

$$\frac{CM_{\beta}}{CM_{Rés}} \sim F_{((J-1),IJ(K-1))},$$

$$\frac{CM_c}{CM_{Rés}} \sim F_{((I-1)(J-1),IJ(K-1))}.$$

**Etape 4 : (Décision)**

Pour un seuil de risque donné $\alpha$, nous déterminons les valeurs critiques $f_{\alpha}$, $f_{\beta}$, $f_c$ par la lecture sur la table de Fisher, telles que

$$P(\frac{CM_{\alpha}}{CM_{Rés}} < f_{\alpha}) = 1-\alpha,$$

$$P(\frac{CM_{\beta}}{CM_{Rés}} < f_{\beta}) = 1-\alpha,$$

$$P(\frac{CM_c}{CM_{Rés}} < f_c) = 1 - \alpha.$$

Les décisions se font comme suit

**Décision sur le premier facteur**

- si $\frac{CM_\alpha}{CM_{Rés}} < f_\alpha$, alors le premier facteur n'a pas une influence significative sur le caractère étudié.
- si $\frac{CM_\alpha}{CM_{Rés}} \geqslant f_\alpha$, alors le premier facteur a une influence significative sur le caractère étudié.

**Décision sur le deuxième facteur**

- si $\frac{CM_\beta}{CM_{Rés}} < f_\beta$, alors le deuxième facteur n'a pas une influence significative sur le caractère étudié.
- si $\frac{CM_\beta}{CM_{Rés}} \geqslant f_\beta$, alors le deuxième facteur a une influence significative sur le caractère étudié.

**Décision sur l'interaction des deux facteurs**

- si $\frac{CM_c}{CM_{Rés}} < f_c$, alors l'interaction des deux facteurs n'a pas une influence significative sur le caractère étudié.
- si $\frac{CM_c}{CM_{Rés}} \geqslant f_c$, alors l'interaction des deux facteurs a une influence significative sur le caractère étudié.

Les résultats d'ANOVA 2 avec répétition sont résumés dans le tableau suivant

| Variation | $SC$ | $ddl$ | $CM$ | $F_{obs}$ | $Fisher$ |
|---|---|---|---|---|---|
| Facteur $A$ | $SC_\alpha$ | $(I-1)$ | $CM_\alpha$ | $CM_\alpha/CM_{Rés}$ | $f_\alpha$ |
| Facteur $B$ | $SC_\beta$ | $(J-1)$ | $CM_\beta$ | $CM_\beta/CM_{Rés}$ | $f_\beta$ |
| Facteur $A*B$ | $SC_c$ | $(I-1)(J-1)$ | $CM_c$ | $CM_c/CM_{Rés}$ | $f_c$ |
| Résiduelle | $SC_{Rés}$ | $IJ(K-1)$ | $CM_{Rés}$ | | |
| Totale | $SC_{Tol}$ | $n-1$ | | | |

TAB. 1.4 – Tableau de l'ANOVA 2 avec répétition.

**Remarque 1.2.1** *Dans la littérature nous distinguons trois types d'ANOVA (I, II et III) et cela selon la nature du/des facteur(s) étudié :*

**Type 1** : Le type 1 est le modèle à effets fixés, ou les niveaux de chaque facteur sont déterminés délibérément (c'est-à-dire, fixés) par l'expérimentateur. Cela s'applique à la plupart des protocoles expérimentaux.

**Type 2** : Le type 2 correspond à un modèle à effets aléatoires, dans lequel les niveaux du facteur étudié sont choisis de manière aléatoire. L'intérêt principal porte alors sur la variabilité entre les échantillons par rapport à celle observée au sein d'un même échantillon.

**Type 3** : Le type 3 ou le modèle mixte intégre à la fois des effets fixes et aléatoires, apparait uniquement dans les ANOVA comportant plusieurs facteurs classification.

# Chapitre 2

# Analyse de la variance multivariée

Dans ce chapitre, nous étudions la méthode de l'analyse de la variance multiple (MANOVA), ou analyse de variance multidimensionnelle, qui est une extension de l'ANOVA. Elle permet d'étudier l'effet de facteurs et/ou leurs interactions sur plusieurs variables quantitatives. La MANOVA permet alors de tester l'égalité de plusieurs vecteurs de moyennes.

## 2.1 MANOVA à un seul facteur (MANOVA 1)

### 2.1.1 Présentation des données et de modèle

#### Les données

Comme dans le cas d'ANOVA 1, nous considérons ici un unique facteur $A$ possédant $I$ niveaux ($i = 1, ..., I$). Pour chaque niveau $i$ de $A$, nous réalisons $n_i$ observations du vecteur $Y$ de $\mathcal{R}^p$, nous posons $n = \sum_{i=1}^{I} n_i$. Nous notons $Y_{ij}$ le vecteur associé à la $j^{ème}$ observation réalisée au niveau $i$ de $A$, telle que

$$Y_{ij} = (Y_{ij1}, Y_{ij2}, ..., Y_{ijp})^t, \forall j = \overline{1, n_i}\,; \forall i = \overline{1, I},$$

les données sont présentées dans le tableau 2.1.

| échantillon 1 | échantillon 2 | $\cdots$ | échantillon $I$ |
|---|---|---|---|
| $Y_{11}$ | $Y_{21}$ | $\cdots$ | $Y_{I1}$ |
| $Y_{12}$ | $Y_{22}$ | $\cdots$ | $Y_{I2}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $Y_{1n_1}$ | $Y_{2n_2}$ | $\cdots$ | $Y_{In_I}$ |

TAB. 2.1 – Tableau des données d'une MANOVA 1.

.

**Notation 2.1.1** *Dans ce qui suit, nous allons utiliser les notations suivantes*

$$\begin{aligned} Y_{i\cdot} &= \sum_{j=1}^{n_i} Y_{ij}, \\ Y_{\cdot\cdot} &= \sum_{i=1}^{I}\sum_{j=1}^{n_i} Y_{ij}, \\ \overline{Y}_{i\cdot} &= \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}, \\ \overline{Y}_{\cdot\cdot} &= \frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i} Y_{ij}. \end{aligned}$$

### Le modèle

Le modèle mathématique associé à chaque observation est donné par :

$$\begin{aligned} Y_{ij} &= \mu + \alpha_i + \varepsilon_{ij} \\ &= \mu_i + \varepsilon_{ij} \quad \text{avec } i = 1, 2, ...I\,; \quad j = 1, 2, ..., n_i,\ \varepsilon_{ij} \sim \mathcal{N}(0_{\mathbb{R}^p}, \Sigma), \end{aligned}$$

qui peut être reformulé en sa forme matricielle comme suit :

$$\begin{pmatrix} Y_{ij1} \\ Y_{ij2} \\ \vdots \\ Y_{ijp} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix} + \begin{pmatrix} \alpha_{i1} \\ \alpha_{i2} \\ \vdots \\ \alpha_{ip} \end{pmatrix} + \begin{pmatrix} \varepsilon_{ij1} \\ \varepsilon_{ij2} \\ \vdots \\ \varepsilon_{ijp} \end{pmatrix} = \begin{pmatrix} \mu_{i1} \\ \mu_{i2} \\ \vdots \\ \mu_{ip} \end{pmatrix} + \begin{pmatrix} \varepsilon_{ij1} \\ \varepsilon_{ij1} \\ \vdots \\ \varepsilon_{ijp} \end{pmatrix}.$$

Ainsi, le modèle pour la $r^{ème}$ variable $(r = 1, 2, ..., p)$ dans chaque vecteur $Y_{ij}$ est :

$$Y_{ijr} = \mu_r + \alpha_{ir} + \varepsilon_{ijr} = \mu_{ir} + \varepsilon_{ijr}.$$

Notre objectif est de réaliser un test de comparaison de $I$ vecteurs moyens, qui peut être formuler comme suit

$$H_0 : "\mu_1 = \mu_2 = ... = \mu_I" \text{ contre } H_1 : "\exists\ i_1, i_2 \in \{1, 2, ...I\} \text{ tel que } \mu_{i_1} \neq \mu_{i_2}"$$

ou sous sa forme matricielle suivante

$$H_0 : \begin{pmatrix} \mu_{11} \\ \mu_{12} \\ \vdots \\ \mu_{1p} \end{pmatrix} = \begin{pmatrix} \mu_{21} \\ \mu_{22} \\ \vdots \\ \mu_{2p} \end{pmatrix} = ... = \begin{pmatrix} \mu_{I1} \\ \mu_{I2} \\ \vdots \\ \mu_{Ip} \end{pmatrix}$$

contre

$$\exists\ i_1, i_2 \in \{1, 2, ...I\} \text{ tel que}$$

$$H_1 : \begin{pmatrix} \mu_{i_1 1} \\ \mu_{i_1 1} \\ \vdots \\ \mu_{i_1 p} \end{pmatrix} \neq \begin{pmatrix} \mu_{i_2 1} \\ \mu_{i_2 1} \\ \vdots \\ \mu_{i_2 p} \end{pmatrix}.$$

### 2.1.2 Les étapes de MANOVA 1

Pour effectuer MANOVA à un facteur, nous suivons les mêmes étapes citées dans le chapitre précédent.

**Etape 1 : (Vérification des conditions)**

1. Les échantillons aléatoires de différentes populations sont indépendants.
2. Toutes les populations ont une matrice de variance-covariance commune $\Sigma$.

3. Chaque population suit une loi normale multivariée.

**Etape 2 : (Décomposition de la variation)**

Le principe de l'analyse de la variance à un facteur est de diviser la somme des carrés des écarts totale en deux parties qui sont la somme des carrés des écarts factorielle (variation inter-groupe) et la somme des carrés des écarts résiduelle (variation intra-groupe). Ce principe peut être étendu au cas de $p$ variables comme suit

À partir de la décomposition du vecteur d'observation nous avons

$$Y_{ij} = \overline{Y} + (\overline{Y}_{i\cdot} - \overline{Y}) + (Y_{ij} - \overline{Y}_{i\cdot})$$

nous observons que le produit

$$(Y_{ij} - \overline{Y})(Y_{ij} - \overline{Y})^t$$

peut être écrit comme suit

$$\begin{aligned}(Y_{ij} - \overline{Y})(Y_{ij} - \overline{Y})^t &= [(Y_{ij} - \overline{Y}_{i\cdot}) + (\overline{Y}_{i\cdot} - \overline{Y})][(Y_{ij} - \overline{Y}_{i\cdot}) + (\overline{Y}_{i\cdot} - \overline{Y})]^t \\ &= (Y_{ij} - \overline{Y}_{i\cdot})(Y_{ij} - \overline{Y}_{i\cdot})^t + (Y_{ij} - \overline{Y}_{i\cdot})(\overline{Y}_{i\cdot} - \overline{Y})^t \\ &+ (\overline{Y}_{i\cdot} - \overline{Y})(Y_{ij} - \overline{Y}_{i\cdot})^t + (\overline{Y}_{i\cdot} - \overline{Y})(\overline{Y}_{i\cdot} - \overline{Y})^t,\end{aligned}$$

par l'addition sur $j$, les termes au milieu ($2^{ème}$ et $3^{ème}$) donnent la matrice nulle, car

$$\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i\cdot}) = 0.$$

Par conséquent, nous obtenons l'équation fondamentale de l'analyse de la variance multiple à un facteur

$$\underbrace{\sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y})(Y_{ij}-\overline{Y})^t}_{T} = \underbrace{\sum_{i=1}^{I} n_i(\overline{Y}_{i\cdot}-\overline{Y})(\overline{Y}_{i\cdot}-\overline{Y})^t}_{H} + \underbrace{\sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_{i\cdot})(Y_{ij}-\overline{Y}_{i\cdot})^t}_{E}$$

$$= \underbrace{\sum_{i=1}^{I}\frac{1}{n_i}Y_{i\cdot}Y_{i\cdot}^t - \frac{1}{n}Y_{\cdot\cdot}Y_{\cdot\cdot}^t}_{H} + \underbrace{\sum_{i=1}^{I}\sum_{j=1}^{n_i}Y_{ij}Y_{ij}^t - \sum_{i=1}^{I}\frac{1}{n_i}Y_{i\cdot}Y_{i\cdot}^t}_{E} \tag{2.1}$$

$T$, $H$ et $E$ étant les matrices de somme de carrés d'écarts et de somme de produit d'écarts.

$T$ est la matrice totale, $H$ est la matrice factorielle et $E$ est la matrice résiduelle. Ces matrices sont des matrices carrées symétriques, de dimensions $(p * p)$. Les éléments diagonaux de ces matrices correspondent bien aux sommes des carrés des écarts factorielles et résiduelles de l'ANOVA 1.

**Remarque 2.1.1** *Les résultats de l'analyse de variance multiple à un facteur sont aussi résumés dans un tableau de la forme*

| source de variation | Matrice des sommes des carrés et des produits | *ddl* |
|---|---|---|
| Factorielle | $H$ | $I-1$ |
| Résiduelle | $E$ | $n-I$ |
| Totale | $T=H+E$ | $n-1$ |

TAB. 2.2 – Table de MANOVA à un facteur.

Après le calcul des matrices $H$ et $E$, nous passons directement aux tests statistiques.

**Etape 3 : (Tests statistiques)**

Lorsque les dernières conditions d'applicabilités sont vérifiées, différents tests ont été proposés pour tester l'hypothèse $(H_0)$. Ces tests sont définis à partir des deux matrices $H$ et $E$.

**Test de Wilks**

Il s'agit du test le plus courant dans le contexte de la MANOVA.

La statistique de ce test est

$$\Lambda = \frac{|E|}{|E+H|}.$$

Sous l'hypothèse $H_0$, la statistique $\Lambda$ suit une loi de Wilks de paramètres $p$, $\vartheta_H$ et $\vartheta_E$ tels que

$$\begin{aligned} p &= \text{nombre des variables,} \\ \vartheta_H &= \text{degré de liberté associé au } H, \\ \vartheta_E &= \text{degré de liberté associé au erreur } E. \end{aligned}$$

Pour un seuil de risque donné $\alpha$, nous rejetons $H_0$ si $\Lambda_{obs} \leq \Lambda_{\alpha,p,\vartheta_H,\vartheta_E}$ avec $\Lambda_{obs}$ est la réalisation de la statistique $\Lambda$ et $\Lambda_{\alpha,p,\vartheta_H,\vartheta_E}$ est la valeur critique de la loi de Wilks de paramètres $p$, $\vartheta_H$ et $\vartheta_E$.

**Remarque 2.1.2** – *La statistique $\Lambda$ peut aussi s'écrire*

$$\Lambda = \Pi_{i=1}^{s} \frac{1}{1+\lambda_i}$$

où les $\lambda_i$ sont les valeurs propres de la matrice $HE^{-1}$ (ou $E^{-1}H$) et $s = min(p, \vartheta_H)$ est le nombre de valeurs propres non nulles de cette matrice.

– Dans certaines conditions particulières, lorsque $\vartheta_H$ (pour $\vartheta_H = 1$ ou 2) ou lorsque ($p = 1$ ou 2) la statistique de Wilks peut être transformée de manière exacte en une statistique $F$ de Fisher. Les formules de transformation vers $F$ pour ces cas spécifiques sont présentées dans le tableau suivant

| Les paramètres $p$ et $\vartheta_H$ | La statistique $F$ correspondante | Degrés de liberté |
|---|---|---|
| $\vartheta_H = 1,\ p \geq 1$ | $\frac{1-\Lambda}{\Lambda}\frac{\vartheta_E-p+1}{p}$ | $p$,$\vartheta_E - p + 1$ |
| $\vartheta_H = 2,\ p \geq 1$ | $\frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}}\frac{\vartheta_E-p+1}{p}$ | $2p$,$2(\vartheta_E - p + 1)$ |
| $p = 1,\ \vartheta_H \geq 1$ | $\frac{1-\Lambda}{\Lambda}\frac{\vartheta_E}{\vartheta_H}$ | $\vartheta_H$,$\vartheta_E$ |
| $p = 1,\ \vartheta_H \geq 1$ | $\frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}}\frac{\vartheta_E-1}{\vartheta_H}$ | $2\vartheta_H$, $2(\vartheta_E - 1)$ |

En dehors de ces quatre cas particuliers, une statistique $F$ approximative, qui suit une loi de Fisher, est donnée par la formule suivante

$$F = \frac{1 - \Lambda^{1/t}}{\Lambda^{1/t}} \frac{ddl_2}{ddl_1}$$

avec $ddl_1$ et $ddl_2$ degrés de liberté, où

$$ddl_1 = p\vartheta_H, \qquad ddl_2 = \omega t - \frac{1}{2}(p\vartheta_H - 2),$$
$$\omega = \vartheta_E + \vartheta_H - \frac{1}{2}(p + \vartheta_H + 1), \qquad t = \sqrt{\frac{p^2\vartheta_H^2 - 4}{p^2 + \vartheta_H^2 - 5}},$$

(pour plus détail voir [1]).

### 2.1.3 Autres tests statistiques

Il existe trois autres tests statistiques qui peuvent être utilisés pour répondre au test d'analyse de la variance multiple (MANOVA). Ce sont : test de Pillai, Trace de Hottelling et test de Roy.

**Test de ROY**

Appelé aussi test de la plus grande racine de Roy. Ce test est basé sur $\lambda_1$ la plus grande des valeurs propres de la matrice $E^{-1}H$. Il est défini par la statistique

$$\theta = \frac{\lambda_1}{1+\lambda_1}.$$

Les valeurs critiques pour $\theta$ sont tabulées. Nous rejetons $H_0$ si $\theta \geq \theta_{\alpha,s,m,N}$, avec

$$\begin{aligned} s &= \min(\vartheta_H, p), \\ m &= \frac{1}{2}(|\vartheta_H - p| - 1), \\ N &= \frac{1}{2}(\vartheta_E - p - 1). \end{aligned}$$

Il existe une statistique approximative de Fisher qui permette de mettre en oeuvre ce test, elle est définie par

$$F = \frac{(\vartheta_E - d - 1)\lambda_1}{d} \sim F_{(d, \vartheta_E - d - 1)}$$

où $d = \max(p, \vartheta_H)$.

**Test de Pillai**

Le test de Pillai ou test de Bartlett-Pillai est basé sur les valeurs propres de la matrice $E^{-1}H$. Il est défini par

$$V^{(s)} = tr[(E+H)^{-1}H] = \sum_{i=1}^{s} \frac{\lambda_i}{1+\lambda_i}.$$

Nous rejetons $H_0$ pour $V^{(s)} \geq V_\alpha^{(s)}$, où les valeurs critiques $V_\alpha^{(s)}$ sont tabulées et indexées par les paramètres $s$, $m$ and $N$ qui sont les mêmes paramètres que ceux du test de Roy.

**Remarque 2.1.3** *Pour les valeurs ne sont pas tabulées, nous peuvons utiliser une approximation par une statistique de Fisher*

| *Statistique approximative de Fisher* | *degrés de liberté* |
|---|---|
| $F_1 = \frac{(2N+s+1)V^{(s)}}{(2m+s+1)(s-V^{(s)})}$ | $s(2m+s+1)$ *et* $s(2N+s+1)$ |
| $F_2 = \frac{s(\vartheta_E-\vartheta_H+s)V^{(s)}}{p\vartheta_H(s-V^{(s)})}$ | $p\vartheta_H$ *et* $s(\vartheta_E-\vartheta_H+s)$ |
| $F_3 = \frac{(\vartheta_E-p+s)V^{(s)}}{d(s-V^{(s)})}$ | $sd,\ s(\vartheta_E-p+s)$ *et* $d = max(p, \vartheta_H)$ |

**Test de Lawley–Hotelling**

Connu aussi sous le nom de test $T^2$ généralisée de Hotelling. La statistique de ce dernier test est donnée par

$$U^{(s)} = tr(E^{-1}H) = \sum_{i=1}^{s} \lambda_i$$

où les $\lambda_i$ sont les valeurs propres de la matrice $HE^{-1}$.

Nous rejetons $H_0$ pour les grandes valeurs de la statistique.

**Remarque 2.1.4** *Pour des valeurs* ($p$, $\vartheta_H$ *et* $\vartheta_E$) *ne sont pas tabulées, nous peuvons utiliser une approximation par une statistique de Fisher*

| *Statistique approximative de Fisher* | *degrés de liberté* |
|---|---|
| $F_1 = \frac{U^{(s)}}{c}$ | $a = p\vartheta_H$<br>$b = 4 + \frac{a+2}{B-1}$<br>$c = \frac{a(b-2)}{b(\vartheta_E-p-1)}$<br>$B = \frac{(\vartheta_E+\vartheta_H-p-1)(\vartheta_E-1)}{(\vartheta_E-p-3)(\vartheta_E-p)}$ |
| $F_2 = \frac{2(sN+1)U^{(s)}}{s^2(2m+s+1)}$ | $s(2m+s+1)$ *et* $2(sN+1)$ |
| $F_3 = \frac{[s(\vartheta_E-\vartheta_H-1)+2]U^{(s)}}{sp\vartheta_H}$ | $sp\vartheta_H$ *et* $s(\vartheta_E-\vartheta_H-1)$ |

### 2.1.4 Exemple sur MANOVA 1

Les données résumées dans le tableau 2.3 sont issues d'une expérience dans laquelle la concentration de calcium ($mg/100\ ml$) dans le plasma et le taux d'eau perdu ($mg/min$) ont été mesurés chez 12 oiseaux des deux sexes ayant subi ou non l'administration d'un traitement hormonal ([12]). Dans cet exemple l'objectif est de tester l'existence d'un effet de traitement hormonal sur la concentration de calcium dans le plasma et sur le taux d'eau perdu.

| Pas de Traitement hormonal | | | | Traitement hormonal | | | |
|---|---|---|---|---|---|---|---|
| $Ca$ | $H_2O$ | $Ca$ | $H_2O$ | $Ca$ | $H_2O$ | $Ca$ | $H_2O$ |
| 16.5 | 76 | 14.5 | 80 | 39.1 | 71 | 32.0 | 65 |
| 18.4 | 71 | 11.0 | 72 | 26.2 | 70 | 23.8 | 69 |
| 12.7 | 64 | 10.80 | 77 | 21.3 | 63 | 28.8 | 67 |

TAB. 2.3 – Les valeurs de Ca et H2O.

Nous utilisons les résultats suivants pour calculer les matrices $H$, $E$ et $T$

$$Y_{1.} = \begin{pmatrix} 83.9 \\ 440 \end{pmatrix},$$

$$Y_{2.} = \begin{pmatrix} 171.2 \\ 405 \end{pmatrix},$$

$$Y_{..} = \begin{pmatrix} 255.1 \\ 845 \end{pmatrix}.$$

En utilisant la formule de calcul de $H$ donnée dans (2.1), nous obtenons

$$H = \left(\frac{1}{6} * \left(\begin{pmatrix} 83.9 \\ 440 \end{pmatrix} * \begin{pmatrix} 83.9 & 440 \end{pmatrix}\right)\right) + \left(\frac{1}{6} * \left(\begin{pmatrix} 171.2 \\ 405 \end{pmatrix} * \begin{pmatrix} 171.2 & 405 \end{pmatrix}\right)\right)$$

$$- \frac{1}{12} * \left(\begin{pmatrix} 255.1 \\ 845 \end{pmatrix} * \begin{pmatrix} 255.1 & 845 \end{pmatrix}\right)$$

$$= \begin{pmatrix} 6058.1083 & 17708.6667 \\ 17708.6667 & 59604.1667 \end{pmatrix} - \begin{pmatrix} 5423.0008 & 17963.2917 \\ 17963.2917 & 59502.0833 \end{pmatrix}$$

$$= \begin{pmatrix} 635.1075 & -254.625 \\ -254.625 & 102.0834 \end{pmatrix}.$$

Pour $E$ nous avons, par (2.1)

$$Y_{1j}Y_{1j}^t = \begin{pmatrix} 16.5 & 18.4 & 12.7 & 14.5 & 11 & 10.8 \\ 76 & 71 & 64 & 80 & 72 & 77 \end{pmatrix} * \begin{pmatrix} 16.5 & 76 \\ 18.4 & 71 \\ 12.7 & 64 \\ 14.5 & 80 \\ 11 & 72 \\ 10.8 & 77 \end{pmatrix}$$

$$= \begin{pmatrix} 1220.0 & 6156.8 \\ 6156.8 & 32426. \end{pmatrix},$$

$$Y_{2j}Y_{2j}^t = \begin{pmatrix} 39.1 & 26.2 & 21.3 & 32 & 23.8 & 28.8 \\ 71 & 70 & 63 & 65 & 69 & 67 \end{pmatrix} * \begin{pmatrix} 39.1 & 71 \\ 26.2 & 70 \\ 21.3 & 63 \\ 32.0 & 65 \\ 23.8 & 69 \\ 28.8 & 67 \end{pmatrix}$$

$$= \begin{pmatrix} 5088.8 & 11603.8 \\ 11603.8 & 27385 \end{pmatrix},$$

d'où

$$\sum\sum Y_{ij}Y_{ij}^t = \begin{pmatrix} 1220.0 & 6156.8 \\ 6156.8 & 32426 \end{pmatrix} + \begin{pmatrix} 5088.8 & 11603.8 \\ 11603.8 & 27385 \end{pmatrix}$$

$$= \begin{pmatrix} 6308.8 & 17760.6 \\ 17760.6 & 59811 \end{pmatrix}.$$

donc

$$E = \begin{pmatrix} 6308.8 & 17760.6 \\ 17760.6 & 59811 \end{pmatrix} - \begin{pmatrix} 6058.1083 & 17708.6667 \\ 17708.6667 & 59604.1667 \end{pmatrix}$$

$$= \begin{pmatrix} 250.6917 & 51.9333 \\ 51.9333 & 206.8333 \end{pmatrix}.$$

La matrice totale $T$ est obtenue en utilisant les deux matrices précédant $H$ et $E$

alors

$$\begin{aligned}
T &= H + E \\
&= \begin{pmatrix} 635.1075 & -254.625 \\ -254.625 & 102.0834 \end{pmatrix} + \begin{pmatrix} 250.6917 & 51.9333 \\ 51.9333 & 206.8333 \end{pmatrix} \\
&= \begin{pmatrix} 885.7992 & 202.6917 \\ 202.6917 & 308.9167 \end{pmatrix}.
\end{aligned}$$

Pour tester l'effet du traitement hormonal, nous utilisons la statistique de Wilks, alors

$$\Lambda = \frac{|E|}{|E+H|} = \frac{49154.3239}{232554.2405} = 0.211367 < \Lambda_{0.05,2,1,10} = 0.514.$$

Nous concluons que les moyennes des groupes sont significativement différentes.

## 2.2 MANOVA à deux facteurs (MANOVA 2)

### 2.2.1 Modèle de MANOVA 2 à effets fixe avec interaction

#### Les données

Les données d'une expérience concernée par la technique d'analyse de variance multiple à deux facteurs ($A$ : à $I$ niveaux) et ($B$ : à $J$ niveaux) se présentent sous la forme du tableau 2.4

| $N^0$ | $B_1$ | $B_2$ | $\cdots$ | $B_J$ |
|---|---|---|---|---|
| $A_1$ | $Y_{111}$<br>$Y_{112}$<br>$\vdots$<br>$Y_{11K}$ | $Y_{121}$<br>$Y_{122}$<br>$\vdots$<br>$Y_{12K}$ | $\cdots$ | $Y_{1J1}$<br>$Y_{1J2}$<br>$\vdots$<br>$Y_{1JK}$ |
| $A_2$ | $Y_{211}$<br>$Y_{212}$<br>$\vdots$<br>$Y_{21K}$ | $Y_{221}$<br>$Y_{222}$<br>$\vdots$<br>$Y_{22K}$ | $\cdots$ | $Y_{2J1}$<br>$Y_{2J2}$<br>$\vdots$<br>$Y_{2JK}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\cdots$ | $\vdots$ |
| $A_I$ | $Y_{I11}$<br>$Y_{I12}$<br>$\vdots$<br>$Y_{I1K}$ | $Y_{I21}$<br>$Y_{I22}$<br>$\vdots$<br>$Y_{I2K}$ | $\cdots$ | $Y_{IJ1}$<br>$Y_{IJ2}$<br>$\vdots$<br>$Y_{IJK}$ |

TAB. 2.4 – Tableau des données d'une MANOVA 2.

où

$$Y_{ijk} = (Y_{ijk1}, Y_{ijk2}, ..., Y_{ijkp})^t, \forall i = \overline{1, I}; \forall j = \overline{1, J}; \forall k = \overline{1, K}.$$

Le modèle associé au problème de MANOVA 2 sera donné par

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} = \mu_{ij} + \varepsilon_{ijk}, \tag{2.2}$$
$$i = 1, 2, ..., I, \quad j = 1, 2, ..., J, \quad k = 1, 2, ..., K,$$

où $\alpha_i$, $\beta_j$ et $\gamma_{ij}$ représentent les effets principaux des facteurs $A$ et $B$ et leur interaction sur les $p$ variables en $Y_{ijk}$. Nous utilisons les contraintes

$$\sum_i \alpha_i = \sum_j \beta_j = \sum_i \gamma_{ij} = \sum_j \gamma_{ij} = 0$$

et nous supposons que $\varepsilon_{ijk}$ sont distribués indépendamment selon $\mathcal{N}_p(0, \Sigma)$.

Avec les contraintes ci-dessus, les estimation des paramètres du modèle sont

$$\hat{\mu} = \overline{Y}_{...}$$
$$\hat{\alpha}_i = \overline{Y}_{i..} - \overline{Y}_{...}$$
$$\hat{\beta}_j = \overline{Y}_{.j.} - \overline{Y}_{...}$$
$$\hat{\gamma}_{ij} = \overline{Y}_{ij.} - \overline{Y}_{i..} - \overline{Y}_{.j.} + \overline{Y}_{...},$$

comme dans le cas univarié, le vecteur moyen $\overline{Y}_{i..}$ est défini par $\overline{Y}_{i..} = \sum\limits_{jk} Y_{ijk}/KJ$.
Les moyens $\overline{Y}_{.j.}$, $\overline{Y}_{ij.}$ et $\overline{Y}_{...}$ ont des définitions analogues : $\overline{Y}_{.j.} = \sum\limits_{ik} Y_{ijk}/KI$, $\overline{Y}_{ij.} = \sum\limits_{k} Y_{ijk}/K$, $\overline{Y}_{...} = \sum\limits_{ijk} Y_{ijk}/KIJ$.
Pour le modèle à deux facteurs (plan équilibré), le tableau d'analyse de la variance multiple prend la forme suivante

| Source de variance | Matrice de somme des carrés et des produits | ddl |
|---|---|---|
| A | $H_A = KJ\sum\limits_i (\overline{Y}_{i..} - \overline{Y}_{...})(\overline{Y}_{i..} - \overline{Y}_{...})^t$ | $I-1$ |
| B | $H_B = KI\sum\limits_j (\overline{Y}_{.j.} - \overline{Y}_{...})(\overline{Y}_{.j.} - \overline{Y}_{...})^t$ | $J-1$ |
| Interaction AB | $H_{AB} = K\sum\limits_{ij} (\overline{Y}_{ij.} - \overline{Y}_{i..} - \overline{Y}_{.j.} + \overline{Y}_{...}) *$ $(\overline{Y}_{ij.} - \overline{Y}_{i..} - \overline{Y}_{.j.} + \overline{Y}_{...})^t$ | $(I-1)(J-1)$ |
| Erreur | $E = \sum\limits_{ijk} (Y_{ijk} - \overline{Y}_{ij.})(\overline{Y}_{ijk} - \overline{Y}_{ij.})^t$ | $IJ(K-1)$ |
| Totale | $T = \sum\limits_{ijk} (Y_{ijk} - \overline{Y}_{...})(\overline{Y}_{ijk} - \overline{Y}_{...})^t$ | $IJK-1$ |

TAB. 2.5 – Tableau d'analyse de la variance multiple à deux facteurs.

Donc, la matrice totale des sommes des carrés et des produits sera décomposée comme suit

$$T = H_A + H_B + H_{AB} + E \tag{2.3}$$

où $H_A$, $H_B$ et $H_{AB}$ sont respectivement les matrices des sommes des carrées et des produits associées aux hypothèses de nullité des effet principaux des facteurs $A$ et $B$ et de l'effet d'interaction de ces deux facteurs.

$E$ est la matrice résiduelle.

La structure de n'importe quelle matrice d'hypothèse $H$ ($H = H_A$, $H_B$ ou $H_{AB}$) est simillaire. Par exemple, les éléments diagonaux de $H_A$ sont les sommes des carrés pour le facteur $A$ pour chaque variable. Les éléments en dehors de la diagonale sont les sommes des produits croisés correspondants pour chaque paires de variables. Par conséquent, le $r^{ème}$ élément diagonal de $H_A$ correspondant à la $r^{ème}$ variable $(r = 1, 2, ..., p)$ est donné par :

$$h_{Arr} = KJ\sum_{i=1}^{I}(\overline{Y}_{i\cdot\cdot r} - \overline{Y}_{\dots r})^2 = \sum_{i=1}^{I}\frac{Y_{i\cdot\cdot r}^2}{KJ} - \frac{\overline{Y}_{\dots r}}{KIJ}, \tag{2.4}$$

où $\overline{Y}_{i\cdot\cdot r}$ et $\overline{Y}_{\dots r}$ sont les composantes de $\overline{Y}_{i\cdot\cdot}$ et $\overline{Y}_{\dots}$, tandis que $Y_{i\cdot\cdot r}$ et $Y_{\dots r}$ sont les totaux correspondants aux $\overline{Y}_{i\cdot\cdot r}$ et $\overline{Y}_{\dots r}$. Le $(r, s)^{ème}$ élément de $H_A$ est

$$h_{Ars} = KJ\sum_{i=1}^{I}(\overline{Y}_{i\cdot\cdot r} - \overline{Y}_{\dots r})(\overline{Y}_{i\cdot\cdot s} - \overline{Y}_{\dots s})^t = \sum_{i=1}^{I}\frac{Y_{i\cdot\cdot r}Y_{i\cdot\cdot s}}{KJ} - \frac{Y_{\dots r}Y_{\dots s}}{KIJ}. \tag{2.5}$$

À partir de (2.3) et du tableau 2.5, nous obtenons

$$\begin{aligned} h_{ABrr} &= \sum_{ij}\frac{Y_{ij\cdot r}^2}{K} - \frac{Y_{\dots r}^2}{KIJ} - h_{Arr} - h_{Brr}, \\ h_{ABrs} &= \sum_{ij}\frac{Y_{ij\cdot r}Y_{ij\cdot s}}{K} - \frac{Y_{\dots r}Y_{\dots s}}{KIJ} - h_{Ars} - h_{Brs}. \end{aligned} \tag{2.6}$$

Pour la matrice $E$, les formules de calcul sont basées sur (2.3)

$$E = T - H_A - H_B - H_{AB}.$$

Ainsi, les éléments de $E$ ont la forme

$$e_{rr} = \sum_{ijk} Y_{ijkr}^2 - \frac{Y_{\dots r}^2}{KIJ} - h_{Arr} - h_{Brr} - h_{ABrr}, \tag{2.7}$$
$$e_{rs} = \sum_{ijk} Y_{ijkr} Y_{ijks} - \frac{Y_{\dots r} Y_{\dots s}}{KIJ} - h_{Ars} - h_{Brs} - h_{ABrs}.$$

### 2.2.2 Test d'hypothèses

De la même manière comme dans le cas univarié, trois hypothèses doivent être testées

**Effet du premier facteur**

$H_0$ : "les paramètres $\alpha_i$ sont tous nuls" contre $H_1$ : "les paramètres $\alpha_i$ ne sont pas tous nuls".

**Effet du second facteur**

$H_0$ : "les paramètres $\beta_j$ sont tous nuls " contre $H_1$ : "les paramètres $\beta_j$ ne sont pas tous nuls".

**Effet de l'interaction des deux facteurs**

$H_0$ : "les paramètres $c_{ij}$ sont tous nuls" contre $H_1$ : "les paramètres $c_{ij}$ ne sont pas tous nuls."

Si les conditions d'application de MANOVA 2 (conditions de MANOVA 1) sont vérifiées, nous pouvons contrôler la validité de l'hypothèse nulle pour chacun des effets par le test de Wilks en comparant les matrices ($H_A$, $H_B$ et $H_{AB}$) avec la matrice $E$.

Alors, pour chaque cas, les statistiques Lambda de Wilks ($\Lambda$) utilisées pour tester l'effet de l'interaction, l'effet de $A$ et l'effet de $B$ sont définies comme suit

$$\begin{aligned}
\Lambda_A &= \frac{|E|}{|E+H_A|} \sim \Lambda_{p,I-1,IJ(K-1)}, \\
\Lambda_B &= \frac{|E|}{|E+H_B|} \sim \Lambda_{p,J-1,IJ(K-1)}, \\
\Lambda_{AB} &= \frac{|E|}{|E+H_{AB}|} \sim \Lambda_{p,(I-1)(J-1),IJ(K-1)}.
\end{aligned}$$

Dans chaque cas, la distribution indiquée est valide lorsque $H_0$ est vraie.

Notons que $H_0$ sera rejetée pour des petites valeurs de $\Lambda$.

**Remarque 2.2.1** *Pour tester l'effet d'interaction et l'effet des deux facteurs $A$ et $B$ nous pouvons utiliser les autres tests statistiques cités dans MANOVA1 (Pillai, Roy et Hotelling-Lowley) basés sur les valeurs propres des matrices $E^{-1}H_{AB}$, $E^{-1}H_A$ et $E^{-1}H_B$ respectivement.*

### 2.2.3 Exemple sur MANOVA 2

Nous prenons le même exemple précédent en supposant que les données sont aussi réparties selon le sexe (facteur $B$) de telle sorte que les valeurs dans les deux premières colonnes de chaque niveau du facteur traitement hormonal ( facteur $A$) du tableau 2.3 sont celles du sexe femalle ($F$) et les valeurs dans les deux autres colonnes sont celles du sexe male ($M$).

Les tableaux suivants résument les calculs des totaux pour chaque variable ($Ca$ et $H_2O$), pour chaque combinaison des deux facteurs (dans les cellules) et les toutaux marginaux pour chaque niveau de $A$ et $B$.

| $Ca$ | Pas de Traitement hormonal | Traitement hormonal | |
|---|---|---|---|
| $F$ | 47.6 | 86.6 | 134.2 |
| $M$ | 36.3 | 84.6 | 120.9 |
| | 83.9 | 171.2 | 255.1 |

TAB. 2.6 – Les totaux pour Ca.

| $H_2O$ | Pas de Traitement hormonal | Traitement hormonal | |
|---|---|---|---|
| $F$ | 211 | 204 | 415 |
| $M$ | 229 | 201 | 430 |
| | 440 | 405 | 845 |

TAB. 2.7 – Les totaux pour H2O.

En utilisant les formules de calcul de $h_{A_{rr}}$ dans (2.4), l'élément (1.1) de $H_A$ est donné par

$$h_{A11} = \frac{(83.9)^2 + (171.2)^2}{(3) * (2)} - \frac{(255.1)^2}{(2) * (2) * (3)} = 635.1075.$$

Pour l'élément $(2, 2)$ du $H_A$ nous avons

$$h_{A22} = \frac{(440)^2 + (405)^2}{(3) * (2)} - \frac{(845)^2}{(2) * (2) * (3)} = 102.0833.$$

Pour l'élément $(1, 2)$ du $H_A$ nous utilisons (2.5) pour obtenir $h_{A_{rs}}$

$$\begin{aligned} h_{A12} &= \frac{(83.9) * (440) + (171.2) * (405)}{(3) * (2)} - \frac{(255.1) * (845)}{(2) * (2) * (3)} \\ &= -254.625, \end{aligned}$$

donc

$$H_A = \begin{pmatrix} 635.1075 & -254.625 \\ -254.625 & 102.0833 \end{pmatrix}.$$

Nous obtenons $H_B$ de manière similaire

$$h_{B11} = \frac{(134.2)^2 + (120.9)^2}{(3) * (2)} - \frac{(255.1)^2}{(2) * (2) * (3)} = 14.7408,$$

$$h_{B22} = \frac{(415)^2 + (430)^2}{(3) * (2)} - \frac{(845)^2}{(2) * (2) * (3)} = 18.75,$$

$$\begin{aligned} h_{B12} &= \frac{(134.2) * (415) + (120.9) * (430)}{(3) * (2)} - \frac{(255.1) * (845)}{(2) * (2) * (3)} \\ &= -16.625, \end{aligned}$$

d'où

$$H_B = \begin{pmatrix} 14.7408 & -16.625 \\ -16.625 & 18.75 \end{pmatrix}.$$

Pour $H_{AB}$ nous avons, par (2.6)

$$\begin{aligned} h_{AB11} &= \frac{(47.6)^2 + (36.3)^2 + (86.6)^2 + (84.6)^2}{3} - \frac{(255.1)^2}{(2) * (2) * (3)} \\ &- 635.1075 - 14.7408 \\ &= 7.2075, \end{aligned}$$

$$\begin{aligned} h_{AB22} &= \frac{(211)^2 + (229)^2 + (204)^2 + (201)^2}{3} - \frac{(845)^2}{(2) * (2) * (3)} - 102.0833 - 18.75 \\ &= 36.75, \end{aligned}$$

$$\begin{aligned}
h_{AB12} &= \frac{(47.6)*(211)+(36.3)*(229)+(86.6)*(204)+(84.6)*(201)}{3} \\
&\quad - \frac{(255.1)^2*(845)}{(2)*(2)*(3)} - (-254.625) - (-16.625) \\
&= -16.275,
\end{aligned}$$

alors

$$H_{AB} = \begin{pmatrix} 7.207\,5 & -16.275 \\ -16.275 & 36.75 \end{pmatrix}.$$

La matrice d'erreur $E$ est obtenue en utilisant les formules de calcul données pour $e_{rr}$ et $e_{rs}$ dans (2.7). Par exemple, $e_{11}$ et $e_{12}$ sont calculés comme suit

$$\begin{aligned}
e_{11} &= (16.50)^2+(14.50)^2+(18.40)^2+(11)^2+(12.70)^2+(10.80)^2+(39.10)^2+(32.0)^2 \\
&+ (26.2)^2+(23.8)^2+(21.3)^2+(28.8)^2 - \frac{(255.1)^2}{12} - 635.1075 - 14.7408 - 7.2075 \\
&= 228.75,
\end{aligned}$$

$$\begin{aligned}
e_{22} &= (76)^2+(80)^2+(71)^2+(72)^2+(64)^2+(77)^2+(71)^2+(65)^2 \\
&+ (70)^2+(69)^2+(63)^2+(67)^2 - \frac{(845)^2}{12} - 102.0833 - 18.75 - 36.75 \\
&= 151.33,
\end{aligned}$$

$$\begin{aligned}
e_{12} &= (16.50*76)+(14.50*80)+(18.40*71)+(11*72)+(12.70*64)+(10.80*77) \\
&+ (39.10*71)+(32*65)+(26.2*70)+(23.8*69)+(21.3*63)+(28.8*67) \\
&- \frac{(255.1)*(845)}{12} + 254.625 + 16.625 + 16.275 \\
&= 84.833.
\end{aligned}$$

alors

$$E = \begin{pmatrix} 228.75 & 84.833 \\ 84.833 & 151.33 \end{pmatrix},$$

Pour tester l'effet principal du facteur $A$ ( traitement hormonal) par le test de Wilks, nous calculons

$$E{+}H_A = \begin{pmatrix} 228.75 & 84.833 \\ 84.833 & 151.33 \end{pmatrix} + \begin{pmatrix} 635.1075 & -254.625 \\ -254.625 & 102.0833 \end{pmatrix} = \begin{pmatrix} 863.86 & -169.79 \\ -169.79 & 253.41 \end{pmatrix},$$

donc

$$\Lambda_A = \frac{|E|}{|E + H_A|} = \frac{27420}{1.9008 * 10^5} = 0.1443 < \Lambda_{0.05,2,1,8} = 0.245.$$

et nous concluons que l'effet de $A$ est significatif.

Pour l'effet principal de $B$, nous avons

$$E + H_B = \begin{pmatrix} 228.75 & 84.833 \\ 84.833 & 151.33 \end{pmatrix} + \begin{pmatrix} 14.7408 & -16.625 \\ -16.625 & 18.75 \end{pmatrix} = \begin{pmatrix} 243.49 & 68.208 \\ 68.208 & 170.08 \end{pmatrix},$$

alors

$$\Lambda_B = \frac{|E|}{|E + H_B|} = \frac{27420}{36760} = 0.7459 > \Lambda_{0.05,2,1,8} = 0.245.$$

nous concluons que l'effet de $B$ n'est pas significatif.

Pour l'interaction entre $A$ et $B$, nous obtenons

$$E + H_{AB} = \begin{pmatrix} 228.75 & 84.833 \\ 84.833 & 151.33 \end{pmatrix} + \begin{pmatrix} 7.2075 & 68.558 \\ 68.558 & 36.75 \end{pmatrix} = \begin{pmatrix} 235.96 & 68.558 \\ 68.558 & 188.08 \end{pmatrix},$$

$$\Lambda_{AB} = \frac{|E|}{|E + H_{AB}|} = \frac{27420}{39679} = 0.6910 > \Lambda_{0.05,2,1,8} = 0.245.$$

Par conséquent nous concluons que l'effet d'interaction n'est pas significatif.

# Chapitre 3

# Application Numérique sous logiciel R

Dans ce chapitre, nous allons présenter sous le logiciel R la méthode d'analyse de la variance multiple en utilisant deux exemples numériques réalisés sur des données réelles.

## 3.1 Exemple 1

Dans cet exemple ([1]), des notes sont attribuées par des juges à des poissons préparés selon trois méthodes. Douze poissons ont été cuits selon chaque méthode et plusieurs juges ont goûté les échantillons de poissons et ils ont les notés selon quatre variables : $Y_1$ : arôme, $Y_2$ : saveur, $Y_3$ : texture et $Y_4$ : humidité. Chaque donnée correspond à la note moyenne attribuée par les juges à ce poisson.

Dans la présente application, nous voulons savoir s'il y'a un effet de la méthode de cuisson sur les variables étudiées, alors la technique adéquate pour réaliser notre but est MANOVA 1.

Alors, pour répondre à notre objectif nous réalisons cette technique sous le logiciel

R en suivant les étapes suivantes :

• Pour entrer les données sous R, nous utilisons les instructions suivantes

**Y1<-c(5.4,5.2,6.1,4.8,5.0,5.7,6.0,4.0,5.7,5.6,5.8,5.3,5.0,4.8,3.9,4.0,5.6,6.0, 5.2,5.3,5.9,6.1,6.2,5.1,4.8,5.4,4.9,5.7,4.2,6.0,5.1,4.8,5.3,4.6,4.5,4.4)**

**Y2<-c(6.0,6.5,5.9,5.0,5.7,6.1,6.0,5.0,5.4,5.2,6.1,5.9,5.3,4.9,4.0,5.1,5.4,5.5, 4.8,5.1,6.1,6.0,5.7,4.9,5.0,5.0,5.1,5.2,4.6,5.3,5.2,4.6,5.4,4.4,4.0,4.2)**

**Y3<-c(6.3,6.0,6.0,4.9,5.0,6.0,5.8,4.0,4.9,5.4,5.2,5.8,5.3,4.2,4.4,4.8,5.1,5.7, 5.4,5.8,5.7,6.1,5.9,5.3,6.5,6.0,5.9,6.4,5.3,5.8,6.2,5.7,6.8,5.7,5.0,5.6)**

**Y4<-c(6.7,5.8,7.0,5.0,6.5,6.6,6.0,5.0,5.0,5.8,6.4,6.0,6.5,5.6,5.0,5.8,6.2,6.0, 6.0,6.4,6.0,6.2,6.0,4.8,7.0,6.4,6.5,6.4,6.3,6.4,6.5,5.7,6.6,5.6,5.9,5.5)**

Dans l'instruction ci-dessous, la fonction gl() produit des facteurs en spécifiant leurs nombre de niveaux (le premier entier), leurs nombre de répétitions (le second le nombre) et leurs niveaux

**méthode<-factor(gl(3,12,labels=c("M1","M2","M3")))**

Nous mettons les quatre variables Y1, Y2, Y3 et Y4 dans un data.frame comme suit

**Y=cbind(Y1,Y2,Y3,Y4)**

• Après la saisie des données, nous effectuons une brève analyse statistique descriptives (calcul des moyennes) en tappant les instructions suivantes

**moyennesY1<-tapply(Y1,méthode,mean)**

**moyennesY2<-tapply(Y2,méthode,mean)**

**moyennesY3<-tapply(Y3,méthode,mean)**

**moyennesY4<-tapply(Y4,méthode,mean)**

Les moyennes de chacune des variables étudiées et pour chaque méthode sont présentées dans le tableau 3.1

| Méthode | | Moyenne |
|---|---|---|
| $Y1$ | $M1$ | 5.3833 |
| | $M2$ | 5.2583 |
| | $M3$ | 4.9750 |
| | Total | 5.2056 |
| $Y2$ | $M1$ | 5.7333 |
| | $M2$ | 5.2333 |
| | $M3$ | 4.8333 |
| | Total | 5.2667 |
| $Y3$ | $M1$ | 5.4417 |
| | $M2$ | 5.3083 |
| | $M3$ | 5.9083 |
| | Total | 5.5528 |
| $Y4$ | $M1$ | 5.9833 |
| | $M2$ | 5.8750 |
| | $M3$ | 6.2333 |
| | Total | 6.0305 |

TAB. 3.1 – Résultats du calcul des moyennes (Exemple1).

A partir de ces résultats préliminaires, il est possible de remarquer qu'il existe au moins une différence entre les méthodes de cuissons au niveau de chacune des variables étudiées.

• La fonction manova nous permet de tester l'effet global de la méthode de cuisson sur l'ensemble des variables sensorielles, alors par l'exécution de la commande suivante, nous obtenons le tableau ci dessous

**summary(manova(Y ~méthode))**

| | Df | Pillai | approx F | num Df | den Df | Pr(>F) | |
|---|---|---|---|---|---|---|---|
| méthode | 2 | 0.85987 | 5.845 | 8 | 62 | $1.465e-0.5$ | *** |
| Residuals | 33 | | | | | | |

Signif. codes : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

**Remarque 3.1.1** *Le logiciel R utilise la statistique de Pillai par défaut et les autres statistiques du test MANOVA sont disponibles, elles sont obtenues à l'aide des instructions suivantes* *summary(manova(Y ~méthode),test="W")*

*summary(manova(Y ~méthode),test="H")*

*summary(manova(Y ~méthode),test="R")*

L'exécution de ces instructions donne les résultats présentés dans le tableau suivant

| | Df | | Valeur | approx F | num Df | den Df | Pr(>F) | |
|---|---|---|---|---|---|---|---|---|
| méthode | 2 | Wilks | 0.22449 | 8.3294 | 8 | 60 | $1.609e-0.7$ | *** |
| | | Hotelling-Lawley | 3.0788 | 11.161 | 8 | 58 | $2.161e-0.9$ | *** |
| | | Roy | 2.9515 | 22.874 | 4 | 31 | $7.077e-0.9$ | *** |
| Résiduals | 33 | | | | | | | |

D'après les résultats obtenus et au seuil de signification $\alpha = 5\%$, nous constatons que les différents tests MANOVA (Pillai, Wilks, Hotelling-Lawley et Roy) indiquent un effet significatif de la méthode de cuisson sur l'ensemble des variables sensorielles (arôme, saveur, texture et humidité) car les $p$ valeurs associées sont toutes inférieures à 0.05. Cela signifie que les différentes méthodes n'ont pas les mêmes effets sur les caractéristiques évaluées du poisson. Autrement dit, le choix de la méthode de cuisson influe de manière significative sur la qualité sensorielle du poisson.

## 3.2 Exemple 2

L'expérimentation est menée sur des grains de deux variétés d'orge (Saida et Fouara). Les essais ont été conduits au laboratoire de biologie d'université Mohamed Khider de Biskra El-Hadjeb.

Selon le protocole expérimental suivi, 10 grains sont placés dans une boîte de Pétri et ils sont tapissés par deux couches de papier filtre humidifié par des différentes concentrations de $NaCl$ $(0g/l, 50g/l, 100g/l, 150g/l)$. Chaque traitement est répété trois fois.

La germination est observée quotidiennement pendant une semaine. Une graine est considérée comme germée lorsque la radicule atteint une longueur d'au moins $2mm$. A la fin de l'expérience, la longueur de la radicule et de tigelle est mesurée.

$$Y_{ij} = (l_1, l_2),$$

où

$l_1$ : la longueur de tigelle ;

$l_2$ : la longueur de radicule.

Dans cet exemple, nous voulons étudier l'effet de variété d'orge et de la concentration de $NaCl$ sur la longueur de radicule et la longueur de tigelle.

La technique la plus adéquate pour traiter ce problème est la MANOVA 2 (les deux facteurs simultanément). Notons que nous pouvons aussi faire une MANOVA1 (chaque facteur indépendamment).

Dans cette section, nous allons résoudre ce problème par ANOVA 2, MANOVA 2 et par MANOVA 1.

### 3.2.1 Application de L'ANOVA 2

Avant d'appliquer l'ANOVA 2, un calcul des moyennes pour chaque variable, pour chaque combinaison des facteurs et pour chaque niveau des deux facteurs nous fournis les résultats présentés dans le tableau 3.2.

L'analyse préliminaire des moyennes, montre que

- Les deux types de variétés ont des longueurs moyennes des radicules très proches et la même conclusion pour les longueurs moyennes des tigelles des deux variétés.

- Lorsque la concentration de NaCl augmente, la longueur des radicules et des tigelles

| | Concentration | Variété | Moyenne | | Concentration | Variété | Moyenne |
|---|---|---|---|---|---|---|---|
| Tigelle | $0g/l$ | Fouara | 126.4933 | Radicule | $0g/l$ | Fouara | 118.8633 |
| | | Saida | 122.8400 | | | Saida | 121.2267 |
| | | Total | 124.6667 | | | Total | 120.0450 |
| | $50g/l$ | Fouara | 104.8400 | | $50g/l$ | Fouara | 104.8400 |
| | | Saida | 91.3433 | | | Saida | 91.3433 |
| | | Total | 98.0917 | | | Total | 98.0917 |
| | $100g/l$ | Fouara | 73.5500 | | $100g/l$ | Fouara | 70.1433 |
| | | Saida | 85.0500 | | | Saida | 70.8833 |
| | | Total | 79.3000 | | | Total | 70.5133 |
| | $150g/l$ | Fouara | 58.1400 | | $150g/l$ | Fouara | 53.4667 |
| | | Saida | 44.8300 | | | Saida | 46.1467 |
| | | Total | 51.4850 | | | Total | 49.8067 |
| | Total | Fouara | 90.7558 | | Total | Fouara | 90.7558 |
| | | Saida | 86.0158 | | | Saida | 86.0158 |
| | | Total | 88.3858 | | | Total | 88.3858 |

TAB. 3.2 – Résultats du calcul des moyennes (Exemple 2).

diminue progressivement pour les deux variétés.

Les résultats de l'ANOVA 2, fournis par le logiciel R pour les deux variables longueur de radicule et longueur de tigelle, sont présentés dans le tableau 3.3.

D'après ces résultats et au seuil de risque de 0.05, nous concluons que :

- D'une part, il n'existe pas d'effet significatif du facteur variété sur la longueur des radicules ($p$ valeur $= 0.8106 > 0.05$), ni sur la longueur des tigelles ($p$ valeur$=$ $0.3296 > 0.05$).

- D'autre part, les effets de la concentration de $NaCl$ sont très hautement significatifs sur les deux variables ($p$ valeur$= 1.196e-07$ pour la radicule, $p$ valeur$= 7.173e-08$ pour la tigelle).

- L'interaction (variété : concentration) n'a pas d'effet significatif (car $p$ valeur $> 0.05$) dans les deux cas.

| Response Radicule : | Df | Sum Sq | Mean Sq | F value | Pr(>F) | |
|---|---|---|---|---|---|---|
| VARIETE | 1 | 7.9 | 7.9 | 0.0593 | 0.8106 | |
| CONCENTRATION | 3 | 15930.8 | 5310.3 | 39.8227 | $1.196e-07$ | *** |
| VARIETE : CONCENTRATION | 3 | 81.9 | 27.3 | 0.2070 | 0.8917 | |
| Residuals | 16 | 2133.6 | 133.6 | | | |

| Response Tigelle : | Df | Sum Sq | Mean Sq | F value | Pr(>F) | |
|---|---|---|---|---|---|---|
| VARIETE | 1 | 134.8 | 134.8 | 1.0111 | 0.3296 | |
| CONCENTRATION | 3 | 17128.4 | 5709.5 | 42.8240 | $7.173e-08$ | *** |
| VARIETE : CONCENTRATION | 3 | 622.6 | 207.5 | 1.5565 | 0.2387 | |
| Residuals | 16 | 2133.2 | 133.3 | | | |

Signif. codes : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

TAB. 3.3 – Résultats obtenus par l'ANOVA 2.

### 3.2.2 Application de MANOVA 2

L'application de la technique MANOVA 2 nous à fournis les résultats présentés dans le tableau 3.4.

De la comparaison des p valeurs de chaque terme indiquées dans le tableau du test MANOVA (tableau 3.4) au seuil de signification ($\alpha = 0.05$), nous concluons que :

**Effet du facteur variété**

D'une part, la $p$ valeur associée à ce facteur est strictement supérieure à 0.05 pour les quatre statistiques du test (Pillai, Wilks, Hotelling, Roy), donc il n'existe pas de différence significative entre les longueurs moyennes des radicules et des tigelles selon la variété.

Ce résultat est confirmé par le tableau 3.3 où nous avons observé que, pour un seuil de risque $\alpha = 0.05$, le facteur variété n'a pas d'effet significatif sur la longueur de la radicule ni sur celle de la tigelle.

**Effet du facteur concentration**

Les différents tests statistiques indiquent un effet hautement significatif du facteur concentration sur la longueur moyenne des radicules ainsi que sur celle des tigelles.

| | Df | Pillai | approx F | num Df | den F | Pr(>F) | |
|---|---|---|---|---|---|---|---|
| VARIETE | 1 | 0.07963 | 0.6489 | 2 | 15 | 0.5367 | |
| CONCENTRATION | 3 | 1.51215 | 16.5312 | 6 | 32 | $1.427e-08$ | *** |
| VARIETE : CONCENTRATION | 3 | 0.33930 | 1.0897 | 6 | 32 | 0.3896 | |
| Residuals | 16 | | | | | | |

| | Df | Wilks | approx F | num Df | den F | Pr(>F) | |
|---|---|---|---|---|---|---|---|
| VARIETE | 1 | 0.92037 | 0.6489 | 2 | 15 | 0.5367 | |
| CONCENTRATION | 3 | 0.03902 | 20.3108 | 6 | 30 | $2.459e-09$ | *** |
| VARIETE : CONCENTRATION | 3 | 0.67079 | 1.1049 | 6 | 30 | 0.3896 | |
| Residuals | 16 | | | | | | |

| | Df | Hotelling-Lawely | approx F | num Df | den F | Pr(>F) | |
|---|---|---|---|---|---|---|---|
| VARIETE | 1 | 0.0865 | 0.6489 | 2 | 15 | 0.5367 | |
| CONCENTRATION | 3 | 10.5014 | 24.5032 | 6 | 28 | $6.249e-10$ | *** |
| VARIETE : CONCENTRATION | 3 | 0.4757 | 1.1101 | 6 | 28 | 0.3815 | |
| Residuals | 16 | | | | | | |

| | Df | Roy | approx F | num Df | den F | Pr(>F) | |
|---|---|---|---|---|---|---|---|
| VARIETE | 1 | 0.0865 | 0.649 | 2 | 15 | 0.5367 | |
| CONCENTRATION | 3 | 8.9171 | 47.560 | 3 | 16 | $3.403e-08$ | *** |
| VARIETE : CONCENTRATION | 3 | 0.4417 | 2.356 | 3 | 16 | 0.1104 | |
| Residuals | 16 | | | | | | |

Signif. codes : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Tab. 3.4 – Résultats d'analyse de la variance multiple (MANOVA 2) avec les quatres statistiques du test.

Ces résultats sont conformes à ceux obtenus précédemment dans l'analyse univariée.

**Effet de l'interaction entre les deux facteurs**

Les résultats de l'analyse de la variance multiple indiquent que l'interaction entre les deux facteurs variété et concentration n'a pas d'effet significatif sur la longueur moyenne des radicules et des tigelles.

Ce résultat est cohérent avec les $p$-valeurs rapportées dans le tableau 3.4, toutes supérieures à 0.05, et confirme également les conclusions de l'ANOVA 2.

### 3.2.3 Application de MANOVA 1

La MANOVA 1 permet d'analyser séparément l'effet de chaque facteur (variété ou concentration) sur l'ensemble des variables dépendantes (longueur de radicule et de tigelle).

L'application de la technique MANOVA 1 nous à fournis les résultats suivants

| | Df | | Valeur | approx F | num Df | den Df | Pr(>F) | |
|---|---|---|---|---|---|---|---|---|
| Variété | 1 | Pillai | 0.020656 | 0.22146 | 2 | 21 | 0.8032 | |
| | | Wilks | 0.97934 | 0.22146 | 2 | 21 | 0.8032 | |
| | | Hotelling -Lawley | 0.021092 | 0.22146 | 2 | 21 | 0.8032 | |
| | | Roy | 0.021092 | 0.22146 | 2 | 21 | 0.8032 | |
| Résiduals | 22 | | | | | | | |
| | Df | | Valeur | approx F | num Df | den Df | Pr(>F) | |
| Concentration | 3 | Pillai | 1.4199 | 16.318 | 6 | 40 | $2.149e-09$ | *** |
| | | Wilks | 0.0529 | 21.199 | 6 | 38 | $9.596e-11$ | *** |
| | | Hotelling -Lawley | 8.9628 | 26.888 | 6 | 36 | $6.515e-12$ | *** |
| | | Roy | 7.8202 | 52.134 | 3 | 20 | $1.23e-09$ | *** |
| Résiduals | 20 | | | | | | | |

Signif. codes : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Tab. 3.5 – Résultats obtenus par MANOVA 1 (Exemple 2).

De la comparaison des $p$ valeurs de chaque terme indiquées dans les tableaux du test MANOVA 3.5 au seuil de signification $\alpha(\alpha = 0.05)$, nous concluons que

**Effet de la variété seule**

La p-valeur obtenue ($p$ valeur$= 0.8032$) est largement supérieure à 0.05, indiquant l'absence d'un effet significatif de la variété d'orge sur la croissance.

Ce résultat est cohérent avec ceux obtenus précédemment par MANOVA 2.

**Effet de la concentration seule**

La p-valeur très faible ($p$ valeur$= 2.15 \times 10^{-9}$) indique un effet hautement significatif de la concentration de $NaCl$ sur la croissance des plantules (radicule et tigelle).

Ce résultat confirme que le stress salin affecte fortement la germination, comme montré dans toutes les analyses précédentes.

**Remarque 3.2.1** *Les méthodes d'analyse appliquées ( ANOVA 2, MANOVA 2, MANOVA 1) confirment que la salinité influence fortement la croissance des jeunes plants d'orge, tandis que la variété n'a aucun effet significatif, et l'interaction entre les deux facteurs est également non significative.*

# Conclusion

En conclusion, nous pouvons confirmer que l'analyse de la variance multivariée (MANOVA) constitue un outil statistique efficace permettant d'étudier les effets de plusieurs variables indépendantes sur un ensemble de variables dépendantes simultanément. Elle se distingue particulièrement de l'analyse de variance univariée (ANOVA) par sa capacité à traiter plusieurs variables dépendantes en même temps, offrant ainsi aux chercheurs une compréhension plus globale et approfondie des données. De plus, la MANOVA ne se limite pas à mesurer les effets également d'analyser les effets d'interaction entre les variables, ce qui contribue à clarifier l'ensemble des relations statistiques au sein du modèle étudier. Grâce à cette approche, il devient possible de mettre en évidence les relations complexes entre les variables dépendantes, plutôt que de les traiter séparément. Par ailleurs, la MANOVA se distingue par sa capacité à réduire le nombre d'analyses nécessaires, en les intégrant dans un seul test global. Cela rend les résultats plus cohérents et minimise les risques d'obtenir des conclusions contradictoires dues à la multiplication des tests séparés. Ainsi, son utilisation permet non seulement un gain de temps et d'effort, mais renforce également la précision de l'interprétation statistique.

Ce travail ouvre la voie à plusieurs perspectives futures, notamment l'analyse des interactions entre facteurs, l'intégration de facteurs mixtes dans les modèles MANOVA, ainsi que l'amélioration des méthodes pour traiter des données réelles complexes.

# Bibliographie

[1] Alvin C Rencher. (2002). Method of Multivariate Analysis, second edition.

[2] Benameur, S. (2024). Notes de Cours de Premiére Année Master, Université Mohamed Khider Biskra.

[3] Cherfaoui, M. (2017). Polycopiés du cours Biostatistique, Statistique Appliquées à l'Expérimentation En Science Biologique, Université Mohamed Khider Biskra.

[4] Derkaoui, R. (2022). Cours Destiné aux étudiants deuxième année spécialité Mathématiques. L2, PEM, PES.

[5] Gilbert. Saporta. (2006). Probabilité, Analyse des données et Statistique, deuxième edition. Edition TECHNIP 27 rue Ginoux, 75737 PARIS, Cedex 15, FRANCE.

[6] Godichon, A. Analyse de la Variance 1 Facteur, Baggioni INSA de Rouen Génie Mathématique - $4^{ème}$ année.

[7] Godichon, B. Analyse de la Variance 2 Facteurs, Baggioni INSA de Rouen Génie Mathématique - $4^{ème}$ année.

[8] Lalanne, C ; Georges, S ; Pallier, C. (2005). Statistiques Appliquée sa l'Expérimentation en Sciences Humaines. En ligne

[9] Palm, R. (2000). L'analyse de la variance multivariée et l'analyse canonique discriminante : principes et applications. Notes de Statistique et d'Informatique, (1).

[10] Rakotomalala, R. (2013). Comparaison de populations : Tests paramétriques. Bartlett test, $7, 27 - 29$..

[11] Richard A Johnson, Dean W.Wechern. (2007). Applied Multivariate Statistical Analysis, sixth edition, Upper River, New Jersy.

[12] Wickerhauser, M. (2021). MA 322 : Biostatistics Homework Assignment 10.

# Annexe A : Rappel sur l'algèbre linéaire et quelques lois statistiques

Dans cette Annexe on va présenter quelques notions sur l'algèbre Linéaire ( matrice, l'inverse d'une matrice,...) ainsi que les lois statistiques.

## 3.3 Rappel sur l'algèbre linéaire

**Définition 3.3.1** *(Matrice)*

Soient $n$, $p$ deux entiers naturels non nuls.Nous appelons matrice à $n$ ligne et $p$ colonnes de dimensionn $n \times p$ (ou matrice de type $(n, p)$) à coefficients dans $\mathbb{k}$, un tableau de nombres $a_{ij}$ de $\mathbb{k}$ avec $i$ ($i \in \{1, ..., n\}$) désignant les lignes, et $j$ ($j \in \{1, ..., p\}$) désignant les colonnes. La matrice est notée $(a_{ij})_{\substack{1 \leq i \leq n \\ 1 \leq j \leq p}}$ c'est une application de $\{1, ..., n\} \times \{1, ..., p\}$ dans $\mathbb{k}^{n+p}$, alors

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ij} & \cdots & a_{ip} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nj} & \cdots & a_{np} \end{pmatrix},$$

Les nombres $(a_{ij})_{\substack{1 \leq i \leq n \\ 1 \leq j \leq p}}$ sont appelés coefficients ou termes de la matrice. $\mathcal{M}_{np}(\mathbb{k})$ désigne l'ensemble des matrices à $n$ lignes et $p$ colonnes à coefficients dans $\mathbb{k}$.

**Définition 3.3.2** ***(Matrice identitée)***

La matrice identitée est une matrice carrée dont tous les coefficients de la diagonale sont égaux à 1, et dont tous les autres sont nuls. Autrement dit $A$ est une matrice identité si

$$a_{ij} = 1 \text{ si } i = j \text{ et } a_{ij} = 0 \text{ si } i \neq j,$$

On note $I_n$ matrice identitée d'ordre $n$.

**Définition 3.3.3** ***(Matrice nulle)***

Nous appelons matrice nulle toute matrice carrée dont les coefficients sont tous nuls. Dans ce cas $a_{ij} = 0, \forall i, j \in \{1, ..., n\}$.

**Définition 3.3.4** ***(Matrice inversible)***

Une matrice $A$ de $\mathcal{M}_n(\mathbb{k})$ est dite inversible (ou régulière), s'il existe une matrice $B$ dans $\mathcal{M}_n(\mathbb{k})$ vérifiant $AB = I_n = BA$. Cette matrice $B$ est alors unique, est appelée l'inverse de $A$ et notée $A^{-1}$. L'ensemble des matrices de $\mathcal{M}_n(\mathbb{k})$ inversibles est noté $GL_n(\mathbb{k})$.

**Définition 3.3.5** ***(Matrice de transposition)***

Une matrice $A \in \mathcal{M}_n(\mathbb{k})$ avec $A = (a_{ij})$, la tranposée de $A$ est la matrice de $\mathcal{M}_n(\mathbb{k})$ défine par $A^t = (a_{ji})$ (les lignes de $A$ sont les colonnes de $A^t$ et inversement).

**Définition 3.3.6** ***(Matrice Symétrique)***

Nous disons qu'une matrice est symétrique si $A^t = A$.

**Définition 3.3.7** ***(La trace)***

La trace de la matrice $A \in \mathcal{M}_n(\mathbb{k})$ est définie par la somme de ses éléments diagonaux

$$tr(A) = a_{11} + a_{22} + ... + a_{nn} = \sum_{k=1}^{n} a_{kk}.$$

**Définition 3.3.8** ***(Le déterminant)***

**Déterminants d'ordre** 2

Soit $A \in \mathcal{M}_2(\mathbb{k})$. Nous appelons déterminant de $A$ le déterminant des vecteurs lignes (ou colonnes) de $A$.

Le déterminant de la matrice carrée d'ordre 2 : $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ est le nombre

$$\det A = |A| = a_{11}a_{22} - a_{12}a_{21}.$$

**Déterminant d'ordre 3**

Soit $A$ une matrice carrée d'ordre 3, son déterminant est défini comme suit

$$\begin{aligned}\det A = |A| = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ = a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}).\end{aligned}$$

**Déterminant d'ordre n**

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix},$$

Considérons un élément $a_{ij}$ de $A$. Si on raye dans $A$ la ligne et la colonne contenant $a_{ij}$, nous obtenons une matrice a $n-1$ lignes et $n-1$ colonnes notée $A_{ij}$. Son déterminant $A_{ij}$ s'appelle le mineur de $a_{ij}$ dans $A$.

Nous appelons cofacteur du terme $a_{ij}$ le produit $(-1)^{i+j}A_{ij}$

$$\det A = |A| = \sum_{i=1}^{n}(-1)^{i+j}a_{ij}\det A_{ij}.$$

## 3.4 Rappel sur quelques lois statistiques

**Définition 3.4.1** ***( La loi de Wishart )***

Une matrice $M(p,p)$ a une distribution de Wishart $W_p(n,\Sigma)$ si $M$ peut s'écrire $M = X^tX$ où $X$ est une matrice $(n,p)$ aléatoire définie de la façon suivante : les $n$ lignes de $X$ sont des vecteurs aléatoires gaussiens de même loi $N(0,\Sigma)$ indépendants.

La densité de probabilité de Wishart donnée par

$$f(M) = \frac{|M|^{(n-p-1)/2} \exp(-\frac{1}{2}tr\Sigma^{-1}M)}{2^{np/2}\pi^{p(p-1)/4}\,|\Sigma|^{n/2} \prod_{i=1}^{p} \Gamma(\frac{1}{2}(n+i-1))}.$$

**Définition 3.4.2** ***(La loi du $T^2$ de Hotelling)***

Soit $X$ un vecteur aléatoire multinormal $\mathcal{N}_p(0, I)$ et $M$ une matrice de Wishart $\mathcal{W}_p(n,I)$, indépendante de $X$, alors la variable aléatoire

$$T^2 = nX^tM^{-1}X\,;$$

suit une loi du $T^2$ de Hotelling de paramètres $p$ et $n$.

**Définition 3.4.3** ***(La loi du lambda ($\Lambda$) de Willk's)***

Soit $A$ et $B$ deux matrices de Wishart $\mathcal{W}_p(m,\Sigma)$ et $\mathcal{W}_p(m,\Sigma)$ indépendantes où $m \geq p$, alors le quotient

$$\Lambda = \frac{|A|}{|A+B|}\,;$$

a une distribution de Wilks de paramètres $p$, $m$ et $n$, notée $\Lambda_{(p,m,n)}$ (cette distribution ne dépend pas de $\Sigma$).

**Définition 3.4.4** ***(La loi du khi-deux)***

Soient $Z_1, ..., Z_p$ une suite de variables aléatoires gaussiennes centrées réduites ($E[Zi] = 0$ et $V(Zi) = 1, i = \overline{1,p}$), nous appelons loi du khi-deux à $p$ degrés de liberté ($\mathcal{X}_p^2$) la loi de variable aléatoire $\sum_{i=1}^{p} Z_i^2$.

Sa densité de probabilité notée $\mathcal{X}_p^2$ est donnée par

$$f(x) = \frac{1}{2^{p/2}\Gamma(\frac{p}{2})} e^{\frac{-x^2}{2}} (x^2)^{\frac{p}{2}-1},$$
$$\text{avec } \Gamma(x) = \int_0^x t^{x-1} e^{-t} dt.$$

**Définition 3.4.5** ***(La loi du Fisher Snedecor)***

$X_1$ et $X_2$ étant des variables suivant indépendamment des lois $\mathcal{X}_n^2$ et $\mathcal{X}_p^2$, alors

$$F = \frac{X_1/n}{X_2/p}$$

est une variable aléatoire qui suit une loi de Fisher a $n$ et $p$ degrés de liberté, avec la densité de probabilité correspondante est donnée par

$$f(x) = \frac{1}{\beta(\frac{n}{2}, \frac{p}{2})} \frac{(\frac{n}{p})^{n/2} x^{\frac{n}{2}-1}}{(1 + \frac{n}{p}x)^{(n+p)/2}} \quad \text{avec } \beta(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}.$$

# Annexe B : Abréviations et notations

Les différentes abréviations et notations utilisées tout au long de ce mémoire sont expliquées ci-dessous :

| | | |
|---|---|---|
| ANOVA | : | Analyse de la variance. |
| MANOVA | : | Analyse de la variance multivarée. |
| $H_0$ | : | Hypothèse nulle. |
| $H_1$ | : | Hypothèse alternative. |
| $SC_{Fac}$ | : | La variation due au facteur. |
| $SC_{Rés}$ | : | La variation résiduelle. |
| $SC_{Tol}$ | : | La variation totale. |
| $CM_{Fac}$ | : | Carrés moyens associés au facteur. |
| $CM_{Rés}$ | : | Carrés moyens résiduels. |
| $SC_{\alpha}$ | : | La somme des carrés des erreurs du premier facteur. |
| $SC_{\beta}$ | : | La somme des carrés des erreurs du deuxième facteur. |
| $SC_c$ | : | La somme des carrés des erreurs des deux facteurs. |
| $ddl$ | : | Degrée de liberté. |
| $\mathcal{N}(\mu, \sigma^2)$ | : | Loi normale despérence $\mu$ et de variance $\sigma^2$. |
| $\alpha_i$ | : | L'effet de première facteur. |
| $\beta_j$ | : | L'effet de deuxième facteur. |

| | | |
|---|---|---|
| $c_{ij}$ | : | L'effet d'interaction des deux facteurs. |
| $\varepsilon_{ij}$ | : | Sont les erreurs de mesure. |
| $\mu$ | : | Représente un effet global inconnu du facteur. |
| $E(\varepsilon_i)$ | : | Espérence mathématique de la variable $\varepsilon_i$. |
| $F_{(I-1,n-I)}$ | : | Une loi de Fisher de degrés de liberté $(I-1)$ et $(n-I)$. |
| $\mathcal{X}^2$ | : | Statistique de test de Khi-deux. |
| $H$ | : | La matrice factorielle. |
| $E$ | : | La matrice résiduelle. |
| $T$ | : | La matrice total $(H+E)$. |
| $\Lambda$ | : | Lambda de Wilks (statistique utilisée dans MANOVA). |
| $N^0$ | : | Niveaux du facteurs. |

# ملخص

قمنا في هذا العمل بعرض طريقتين إحصائيتين: تحليل التباين (ANOVA) وتحليل التباين المتعدد (MANOVA)، حيث يسمح تحليل (ANOVA) بدراسة تاثير متغير اوعدة متغيرات نوعية على متغير كمي ، من خلال اتباع منهجية دقيقة تبدأ من التحقق من الفرضيات وصولًا إلى تفسير النتائج باستخدام اختبارF. أما تحليل (MANOVA) فيوسّع هذا النهج ليشمل مجموعة من المتغيرات التابعة معتمدًا على مصفوفات مجموعات المربعات والجداءات ويستخدم اختبارات إحصائية مثل اختبار ويلكس او بيلاي و غيرها .مثالين من معطيات حقيقية تمت معالجتها باستخدام برنامج R.

**كلمات مفتاحية :**

تحليل التباين، تحليل التباين المتعدد، متوسطات، عينات، متغير كمي، متغير نوعي.

# Résumé

Dans ce travail, nous avons présenté deux méthodes statistiques: l'analyse de la variance univarée (ANOVA) et l'analyse de la variance multivariée (MANOVA). L'ANOVA permet d'étudier l'effet d'une ou de plusieurs variables qualitatives sur une variable quantitative, en suivant méthodologie précise allant de la vérification des hypothèses jusqu'à l'interprétation des résultats à l'aide du test F. La MANOVA étend cette approche à un ensemble de variables dépendantes, en s'appuyant sur des matrices de sommes de carrés et des produits, et en utilisant des tests statistiques tels que Wilks, Pillai et d'autres. Deux exemples de données réelles ont été traités à l'aide du logiciel R.

**Mots clés :** ANOVA, MANOVA , moyennes, échantillon, variable quantitative, variable qualitative.

# Abstract

In this work, we presented two statistical methods: univariate analysis of variance (ANOVA) and multivariate analysis of variance (MANOVA),where ANOVA allows studying the effect of one or several qualitative variables on a quantitative variable, by following a precise methodology starting from verifying the hypotheses to interpreting the results using the F-test. MANOVA extends this approach to include a set of dependent variables based on matrice of sums of squares and product, and uses statistical tests such as Wilks' Lambda, Pillai's Trace and others. Two examples of reel data are processed using the R program.

**Keywords :** ANOVA, MANOVA , mean , sample, quantitative Variable, qualitative Variable.