République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITÉ MOHAMED KHIDER, BISKRA

FACULTÉ des SCIENCES EXACTES et des SCIENCES de la NATURE et de la VIE

DÉPARTEMENT DE MATHÉMATIQUES

Mémoire présenté en vue de l'obtention du Diplôme **:**

MASTER en Mathématiques

Option : **Statistique**

Par

BESSIOUD Ziad

Titre :

Analyse de la variance multivariée

Membres du Comité d'Examen :

Pr.	**Meraghni Djamel**	UMKB	Président
Dr.	**ROUBI Affef**	UMKB	Encadreur
Dr.	**Hassouna Houda**	UMKB	Examinateur

Juin 2018

Dédicace

Je dédie ce travail modeste :

A mes chèrs parants ,qui ont bien élevés, aidés, soutenus et encouragés durant toutes ces années d'étude, qu'AllAh la protège .

A mes très chers fréres .

A mes très chéres soeurs

A mes amis durant mes années d'études qui ma beacoup encouragé

-A toutes mes proffeseurs que j'ai connus durant mes études.

-A mes connaissances de proche ou de loin .

-A tous ceux que J'aime et me souhaitent la réussite pour toute ma vie .

Bessioud ziad

REMERCIEMENTS

Louange à Allah , seigneur de l'univers, Avant tout puissant de m'avior donné la volonté et la force pour réaliser ce travail et sans qui ce travail n'aurait , sans doute , pas pu voir le jour Mes vifs remerciments sont adressés à mon encadreur Madame Roubi Affef pour ces conseils et ces orientations qui m'ont été d'une grande utilité au cours de l'élaboration de mon mémoire .

Je tiens à remercie spécialement Monsieur Meraghni Djamel (Proffesur à l'université de Biskra) qui m'a fait l'honneur de présider mon jury de mémoire .

Je tiens à remercie également Mademoisselle Hassouna Houda pour avoir pris le temps de lire et juger ce travail.

Je remercie tous les enseignants qui aider pour finaliser ce travail surtout le proffeseur Meraghni Djamel et Mon encadreur Roubi Affef

Ainsi que tous les employé du Département de Mathématique .

Je remercie tout particulièrement mes parents pour leurs encouragements et soutien sur tous les aspects et aussi tout ma famille .

Je remercie tous ceux qui ont finalisé ce mémoire soit de prés du loin.

Table des matières

Table des figures

Introduction

La Statistique est la science qui à partir de l'observation des phénomènes on peut obtenir des informations ; elle utilise principalement les méthodes de la statistique mathématique pour obtenir des résultats qui forment l'information statistique, ils peuvent porter sur tous les domaines : économie, médecine, technique...etc. Dont le rôle de statisticien est de prendre des décisions sur la base de résultats expérimentaux, en étant conscient qu.il y' a un risque d'erreur lié à l'incertitude des observations ou des résultats expérimentaux, avant de prendre une telle décision, il testera une hypothèse statistique correspondant à son problème.

Dans le cadre des tests d'hypothèses, nous avons émis des hypothèses concernant la moyenne d'une population puis comparé les moyennes de deux populations. L'analyse de la variance (ANOVA) est la méthode employée pour comparer plusieurs moyennes. Au cœur de cette méthode est la décomposition de la variabilité totale selon les différentes sources présentes dans les données. La variabilité totale est décomposée en deux sources : la variabilité intra dûe à l'erreur expérimentale et la variabilité inter dûe aux écarts de moyenne entre les différentes modalités d'un facteur.

Une autre méthode appelée analyse de la variance multiple (MANOVA) sert à comparer des moyennes mais des moyennes vectorielles de plusieurs échantillons, ou d'une autre façon elle permet d'étudier l'effet d'un ou des plusieurs variables qualitatives (facteurs) sur une variable multidimensionnelle.

L'objectif de ce mémoire est d'étudier le cas où deux facteurs se présentent. Pour arriver à la réalisation de nos objectifs, nous proposons le plan de ce mémoire suivant

Chapitre 1 : Nous traitons dans ce chapitre, la technique d'analyse de la variance à un et à deux facteurs (ANOVA 1 et ANOVA 2), leurs différentes types, principes, ainsi leurs différentes étapes.

Chapitre 2 : Ce chapitre est consacré à l'étude en détails de la méthode de l'analyse de la variance multiple à deux facteurs (MANOVA 2).

Chapitre 3 : Ce dernier chapitre est consacré à l'application de tous ce que nous avons parlé dans les chapitres précédents sur des données réelles sous le logiciel R.

Chapitre 1

Analyse de la variance univariée

Ce chapitre est consacré à une méthode statistique appelée l'analyse de la variance (ANOVA) qui a pour objectif d'étudier l'effet d'un ou de plusieurs variables qualitatives sur une varaible quantitative.

1.1 Généralité sur l'ANOVA

Dans cette section ,on va parler sur l'analyse de la variance (ANOVA), leurs types et ses principes.

1.1.1 *Différents types D'ANOVA*

Il existe trois types d'ANOVA

Type 1 (effets fixes)

Les traitements sont déterminés par le chercheur.

Type 2 (effets aléatoires)

Les traitements ne sont pas sous le controle de l'expérimentateur .

Type 03 (modèle mixte)

Au moins un facteur du type 1 et au moins un du type 2.

1.1.2 Les principes d'ANOVA

L'analyse de la variance a pour objectif de tester l'effet d'un ou de plusieurs variables qualitatives sur une variable aléatoire continue.

Chaque variable qualitative est appellé un facteur et chaque facteur peut avoir deux ou plusieurs niveaux ou traitements.

Une ANOVA teste si toutes les moyennes sont égales donc

$$\begin{cases} H_0 & : \text{égalité.} \\ H_1 & : \text{au moins une différence .} \end{cases}.$$

A utiliser quand le nombre de niveaux est supérieur à deux.

Conditions d'applications de l'ANOVA

IL y a trois conditions pour l'application d'ANOVA

1 / Indépandance les échantillons comparés sont indépendants .

2 / Homoscédasticité les échantillons comparés ont même variance (Homogénéité des variances). Il y' a plusieurs tests permettent de tester l'égalité de plusieurs variances (le test le plus utilisé est le test de Bartlett).

1.2 Analyse de la variance à un facteur (ANOVA 1)

Définition 1.2.1 *L'analyse de la variance à un facteur nous permet de tester l'effet d'un facteur contrôlé A ayant p modalités sur les moyennes d'une variable quantitative Y.*

Facteur	A_1	A_2	$\dots$	A_i	$\dots$	A_p
	Y_{11}	Y_{21}	$\dots$	Y_{i1}	$\dots$	Y_{p1}
	Y_{12}	Y_{22}	$\dots$	Y_{i2}	$\dots$	Y_{p2}
	$\vdots$	$\vdots$	$\dots$	$\vdots$	$\dots$	$\vdots$
	Y_{1n_1}	Y_{2n_2}	$\dots$	Y_{in_i}	$\dots$	Y_{pn_p}
Moyennes	$\overline{Y}_1$	$\overline{Y}_2$		$\overline{Y}_i$		$\overline{Y}_p$

TAB. 1.1 – Les données d'ANOVA 1.

1.2.1 Structure des données

Un facteur contrôlé A se présente sous p modalités, chacune d'entre elles étant notée A_i. Pour chacune des modalités nous effectuons $n_i \geq 2$ mesures d'une réponse Y qui est une varible continue. Nous notons $n = \sum_{i=1}^{p} n_i$ le nombre total de mesures ayants été effectuées. Les données relatives à une analyse de variance à un facteur contrôlé sont structurées dans un tableau du type suivant

Conséderons p échantillons Y_i d'effectifs n_i, issu des p populations qui suivent p lois normales $\mathcal{N}(\mu_i, \sigma^2)$ de même variance. Chaque observation s'écrit Y_{ij}, avec

$$i = \overline{1, p} \quad \text{et} \quad j = \overline{1, n_i}.$$

L'effectif total est

$$n = \sum_{i=1}^{p} n_i.$$

• Moyenne de chaque échantillon

$$\overline{Y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij} \backsim \mathcal{N}\left(\mu_i, \frac{\sigma^2}{n_i}\right), i = 1, \dots, p.$$

• Globalement, la moyenne de toutes les observations

$$\overline{Y} = \frac{1}{n} \sum_{i=1}^{p} \sum_{j=1}^{n_i} Y_{ij} \backsim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \text{ avec } \mu = \frac{1}{n} \sum_{i=1}^{p} (n_i \mu_i).$$

• Variance de chaque échantillon

$$s_i^2\left(Y\right) = \frac{1}{n_i}\sum_{j=1}^{n_i}\left(Y_{ij} - \overline{Y}_i\right)^2, i = 1, \ldots, p.$$

• Variance de toutes les observations

$$s^2\left(Y\right) = \frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{n_i}\left(Y_{ij} - \overline{Y}\right)^2.$$

1.2.2 Modèle d'ANOVA 1

Soit $\mu = \frac{1}{n}\sum_{i=1}^{p}\left(n_i\mu_i\right)$ et $\mu_i = E\left[Y_i\right]$ En terme d'observation on a

$$\underbrace{\left(Y_{ij} - \overline{Y}\right)}_{\text{écart total}} = \underbrace{\left(Y_{ij} - \overline{Y}_i\right)}_{\text{écart résiduel}} + \underbrace{\left(\overline{Y}_i - \overline{Y}\right)}_{\text{écart factoriel}},$$

Le modèle théorique et

$$Y_{ij} - \mu = Y_{ij} - \mu + \mu_i - \mu$$

$$\begin{cases} Y_{ij} - \mu_i = \varepsilon_{ij} \text{ est l'effet résiduel} \\ \mu_i - \mu = \alpha_i \text{ est l'effet principal} \end{cases}.$$

Nous introduisons le modèle

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, i = 1, ..., p,\ j = 1, ..., n_i;$$

avec $\varepsilon_{ij} \backsim \mathcal{N}\left(\mu, \sigma^2\right)$ et la contrainte supplémentaire $\sum_{i=1}^{p}\alpha_i = 0.$

1.2.3 L'équation fondamentale d'ANOVA 1

A partir du modèle fondamentale

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

et avec les estimateurs des paramètres des modèles

• Sous H_0

$$Y_{ij} = \mu + \varepsilon_{ij},\ \widehat{\mu} \ = \overline{Y}\,;$$

• Sous H_1

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij},\ \widehat{\mu} + \widehat{\alpha}_i = \overline{Y}_i\,;$$

d'où

$$\begin{cases} \widehat{\alpha}_i = \overline{Y}_i - \overline{Y} \\ \widehat{\varepsilon}_{ij} = Y_{ij} - \widehat{\mu} - \widehat{\alpha}_i = Y_{ij} - \overline{Y}_i \end{cases}$$

en remarquant que

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

il vient facilement

$$\frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{n_i}\left(Y_{ij} - \overline{Y}\right)^2 = \frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{n_i}\left(Y_{ij} - \overline{Y}_i\right)^2 + \frac{1}{n}\sum_{i=1}^{p} n_i\left(\overline{Y}_i - \overline{Y}\right)^2.$$

Remarque 1.2.1 *L'equation ne contient pas de double produit car la somme des doubles produits est nulle en raison de la nullité de la somme de écarts par rapport à la moyenne*

En effet

$$2\sum_{i=1}^{p}\sum_{j=1}^{n_i}\left(Y_{ij} - \overline{Y}_i\right)\left(\overline{Y}_i - \overline{Y}\right) = 2\sum_{j=1}^{n_i}\left[\left(\overline{Y}_i - \overline{Y}\right)\sum_{i=1}^{p}\left(Y_{ij} - \overline{Y}_i\right)\right] = 0.$$

formule qui n'est autre que celle de la variance totale décomposée en variance des moyennes et moyenne des variances

$$s^2\left(Y\right) = \frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{n_i}\left(Y_{ij} - \overline{Y}_i\right)^2 = \frac{1}{p}\sum_{i=1}^{p}\left(\overline{Y}_i - \overline{Y}\right)^2 + \frac{1}{p}\sum_{i=1}^{p} s_i^2\left(Y\right) \tag{2.1}$$

on multiple l'équation (2.1) par n, on obtient l'équation fondamentale de l'analyse de la variance

$$\underbrace{\sum_{i=1}^{p}\sum_{j=1}^{n_i}\left(Y_{ij}-\overline{Y}\right)^2}_{SC_{TOT}} = \underbrace{\sum_{i=1}^{p} n_i\left(\overline{Y}_i-\overline{Y}\right)^2}_{SC_{Fac}} + \sum_{i=1}^{p}\underbrace{\left(\sum_{j=1}^{n_i}\left(Y_{ij}-\overline{Y}_i\right)^2\right)}_{SC_{\mathrm{Re}\,s}},$$

où

$$SC_{Fac} : \text{est la variation due au facteur,}$$

$$SC_{\mathrm{Re}\,s} : \text{est la variation résiduelle,}$$

$$SC_{Tot} : \text{est la variation totale.}$$

On a alors la relation fondamentale de l'ANOVA

$$SC_{Tot} = SC_{Fac} + SC_{\mathrm{Re}\,s}.$$

On y associe des degrés de libertés

$$n-1 = (p-1)+(n-p)$$

Les sommes des carrés des écarts peuvent être divisées par leur nombres de degré de liberté respectifs, on obtient alors, les carrés moyens

$$\begin{cases} CM_{Tot} = SC_{Tot} \diagup (n-1) \\ CM_{Fac} = SC_{Fac} \diagup (p-1) \\ CM_{\mathrm{Re}\,s} = SC_{\mathrm{Re}\,s} \diagup (n-p) \end{cases}.$$

1.2.4 Tests d'hypothèses

Nous construisons le tableau d'analyse de la variance à partir des informations précédentes

• F comparée à f_α, obtenue à l'aide d'une table de Fisher pour un seuil donné α, avec la loi de Fisher étant définie comme le rapport de deux lois du χ^2_α.

Nous souhaitons faire le test d'hypothèse suivant

Sources de variation	ddl	Sommes des carrés	Carrés moyens	F	Décision
Inter-group (Fac)	$p-1$	SC_{Fac}	CM_{Fac}	$F=\frac{CM_{Fac}}{CM_{\mathrm{Re}\,s}}$	H_0 ou H_1
Intra-group (Rés)	$n-p$	$SC_{\mathrm{Re}\,s}$	$CM_{\mathrm{Re}\,s}$		
Total	$n-1$	SC_{Tot}			

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

$$\begin{cases} H_0 : \mu_1 = \mu_2 = \cdots = \mu_p = \mu, \\ H_1 : \exists\,(k,l) \in \{1,2,...,p\} \text{ tel que } \mu_K \neq \mu_l, \end{cases} \Longleftrightarrow \begin{cases} H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_p = 0, \\ H_1 : \exists i \in \{1,2,...,p\} \text{ tel que } \alpha_i \neq 0. \end{cases}.$$

On a

$$CM_{Fac} = \frac{SC_{Fac}}{p-1} \backsim \chi^2_{p-1}, \qquad CM_{\mathrm{Re}\,s} = \frac{SC_{\mathrm{Re}\,s}}{n-p} \backsim \chi^2_{n-p} \text{ et } \quad F = \frac{CM_{Fac}}{CM_{\mathrm{Re}\,s}} \backsim F\,(p-1, n-p)\,.$$

Nous concluons alors à l'aide des tables, en utilisant le quantille f_α de loi de Fisher à $p-1$ et $n-p$ degrés de liberté que

si $F < f_\alpha$ on dit que H_0 est accepté et on conclut qu'il n'existe pas une influence significative du facteur A.

si $F \geq f_\alpha$, on a H_0 est rejetée et conclut qu'il existe une influence significative du facteur A.

• Nous concluons alors à l'aide de la p-valeur, rejet de H_0 si elle est inférieure ou égale au seuil α du test.

Lorsque l'hypothèse nulle H_0 est rejetée, nous pouvons procéder à des comparaisons multiples des différents effets des niveaux du facteur.

1.3 Analyse de la variance à 2 facteurs (ANOVA 2)

Cette section est consacrée à l'étude des situations expérimentales dans lesquelles l'effet de deux facteurs (variables qualitatives) est étudié simultanément, c'est-à-dire dans le même protocole expérimental. En cela, elle constitue une extension à la situation précédente dans laquelle on n'étudiait qu 'un seul facteur à la fois (ANOVA d'ordre 1).

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

TAB. 1.3 – Les données d'ANOVA2 avec répétitions.

L'identification de l'ANOVA d'ordre 2 (ANOVA 2) au sens littéraire peut être résumé dans la définition suivante

Définition 1.3.1 *L'analyse de la variance à deux facteurs teste l'effet de deux facteurs controlés A et B (variables qualitatives) ayant respectivement I et J modalités sur les moyennes d'une variable quantitative Y.*

Les problèmes concernés par la technique ANOVA 2 se présentent en générale de la manière suivante

et son modèle mathématique est donné par

$$Y_{ijk} = \mu_{ij} + \epsilon_{ijk}, \text{avec } \ i = \overline{1, J} \ \text{ et } \ k = \overline{1, K}, \tag{1.1}$$

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 ϵ_{ijk} sont les erreurs de mesure (inconnues) de plus $\varepsilon_{ijk} \backsim \mathcal{N}(0, \sigma_\varepsilon^2)$.

Ce modèle peut-être réécrit sous sa forme détaillée comme suit

$$Y_{ijk} = \mu + a_i + b_j + c_{ij} + \epsilon_{ijk}, \qquad \text{avec } i = \overline{1, I}, j = \overline{1, J} \ \text{ et } k = \overline{1, K},$$

ce qui s'explique que la réalisation de la variable Y est un cumule d'une constante μ (indépendante des deux facteurs), de l'effet du premier facteur A, de l'effet du deuxième facteur B et de l'effet d'interaction des deux facteurs c et de l'erreur de mesure ϵ .

Si le modèle de référence retenu est le modèle (2.1), alors le test pour lequel nous intéressons à réaliser sera formulé comme suit

H_0 : "$\forall i \in \begin{cases} \{1,\ldots,I\} \\ j \in \{1,\ldots,J\} \end{cases} \mu_{ij} = \mu$" contre H_1 : "$\exists \begin{cases} i_1, i_2 \in \{1,\ldots,I\} \\ j_1, j_2 \in \{1,\ldots,J\} \end{cases}$ tel que $\mu_{i_1 j_1} \neq \mu_{i_2 j_2}$".

Par contre, si le modèle de référence retenu est le modèle (2.2), alors l'analyse de la variance à deux facteurs avec répétitions consiste en réalisation de trois tests de Fisher à la fois, dont la formulation est

1/ Effet du premier facteur

H_0 :"Les paramètres a_i sont tous nuls " contre H_1 : " les paramètres a_i ne sont pas tous nuls ".

2/ Effet du second facteur

H_0 :"Les paramètres b_j sont tous nuls " contre H_1 :"les paramètres b_j ne sont pas tous nuls ".

3/ 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 ".

1.3.1 Les étapes de L'ANOVA 2

La mise en oeuvre d'une ANOVA 2, se fait principalement en 4 étapes. Les détails de ces étapes sont comme suit

Étape 1 (Conditions)

Afin de réaliser une analyse de la variance à deux facteurs, les conditions suivantes doivent être vérifiées préalablement

• Les IJ échantillons comparés sont mutuellement indépendants.

• La variable quantitative étudiée suit une loi normale dans les IJ populations comparées.

• Les IJ populations comparées ont même variance : Homogénéité des variances.

Étape 2 (Moyennes et variances)

Quantifier les différentes statistiques intervenant dans L'ANOVA à deux facteurs et qui sont

• La moyenne globale de toutes les observations

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

• Moyenne de chaque échantillon

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

• Moyenne de chaque modalité du premier facteur

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

- Moyenne de chaque modalité du deuxième facteur

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

- La somme des carrés des erreurs totale

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

- La somme des carrés des erreurs résiduelles

$$SC_{\mathrm{Re}\,s} = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(Y_{ijk} - \overline{Y}_{ij\bullet}\right)^2;$$

- La somme des carrés des erreurs du premier facteur

$$SC_A = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(\overline{Y}_{i\bullet\bullet} - \overline{Y}_{...}\right)^2;$$

- la somme des carrés des erreurs des deux facteurs

$$SC_B = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(\overline{Y}_{\bullet j\bullet} - \overline{Y}_{...}\right)^2;$$

- La somme des carrés des erreurs des deux facteurs

$$SC_{AB} = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(\overline{Y}_{ij\bullet} - \overline{Y}_{i\bullet\bullet} - \overline{Y}_{\bullet j\bullet} + \overline{Y}_{...}\right)^2;$$

Avec le même raisonnement que dans l'ANOVA 1, on peut démontrer que la variation quadratique totale des observations autour de la moyenne $\overline{Y}_{...}$ peut être décomposée comme suit

$$SC_{Tot} = SC_{\mathrm{Re}\,s} + SC_A + SC_B + SC_{AB}.$$

Étape 2 (Les carrés moyens)

A partir de la décompositon précédente, 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 inter-groupes (engendrée par les deux facteurs ou/et leurs interaction) associée au caractère est négligeable par rapport aux fluctuations individuelles. Pour comparer ces quantités,on considère les carrés moyens suivants

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

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

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

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

Notons que, si les trois conditions citées précédemment (Indépandance,Normalité et Homogénéité) sont vérifiées alors sous l'hypothèse H_0

$$\frac{CM_A}{CM_{\operatorname{Re}s}} \backsim F_{((I-1),IJ(K-1))},$$

$$\frac{CM_B}{CM_{\operatorname{Re}s}} \backsim F_{((J-1),IJ(K-1))},$$

$$\frac{CM_{AB}}{CM_{\operatorname{Re}s}} \backsim F_{((I-1)(J-1),IJ(K-1))},$$

Étape. 4 : (Décision)

Pour un seuil de risque α ,nous quantifions les valeurs critiques f_a, f_b, f_c (par la lecture sur le table de Fisher), telle que

$P\left(\frac{CM_A}{CM_{\operatorname{Re}s}} < f_a\right) = 1-\alpha, P\left(\frac{CM_B}{CM_{\operatorname{Re}s}} < f_b\right) = 1-\alpha, P\left(\frac{CM_{AB}}{CM_{\operatorname{Re}s}} < f_c\right) = 1-\alpha.$

Ainsi les décisions des tests se font comme suit

Décision sur le premier facteur •Si $\frac{CM_A}{CM_{\operatorname{Re}s}} < f_a$, alors le premier facteur n'a pas une influence significative sur le caractère étudié.

Source de variation	SC	ddl	CM	F_{obs}	Fisher
due à F_A	SC_A	$(I-1)$	CM_A	$CM_A/CM_{\mathrm{Re}\,s}$	f_a
due à F_B	SC_B	$(J-1)$	CM_B	$CMB/CM_{\mathrm{Re}\,s}$	f_b
due à $F_A * F_B$	SC_{AB}	$(I-1)(J-1)$	CM_{AB}	$CM_{AB}/CM_{\mathrm{Re}\,s}$	f_c
Résiduelle	$SC_{\mathrm{Re}\,s}$	$IJ(K-1)$	$CM_{\mathrm{Re}\,s}$		
Total	SC_{Tot}	$n-1$			

Tab. 1.4 – Tableau d'ANOVA 2 avec plan équilibré.

• Si $\frac{CM_A}{CM_{\mathrm{Re}\,s}} \geq f_a$, alors le premier facteur a une influence significative sur le caractère étudié.

Décision sur le deuxième facteur • Si $\frac{CM_B}{CM_{\mathrm{Re}\,s}} < f_b$, alors le deuxième facteur n'a pas une influence significative sur le caractère étudié.

• Si $\frac{CMB}{CM_{\mathrm{Re}\,s}} \geq f_b$, alors le deuxième facteur a une influence significative sur le caractère étudié.

Décision sur l'interaction des deux facteurs • Si $\frac{CM_{AB}}{CM_{\mathrm{Re}\,s}} < f_c$, alors l'interaction des deux facteurs n'a pas une influence significative sur le caractère étudié .

• Si $\frac{CM_{AB}}{CM_{\mathrm{Re}\,s}} \geq f_c$, alors l'interaction des deux facteurs a une influence significative sur le caractère étudié.

Remarque 1.3.1 *Les résultats d'une ANOVA 2 sont souvent présentés dans un tableau de la forme suivante*

Chapitre 2

Analyse de la variance multivariée à deux facteurs (MANOVA 2)

Dans ce chapitre on s'intéresse à la comparaison des vecteurs moyens par la méthode MANOVA 2 (l'analyse de la variance multiple a deux facteurs) La MANOVA est souvent considérée comme extension de l'anova à deux facteurs pour des situations où il ya deux variables ou plus dépendantes. Le but principale du MANOVA 2 est de tester s'il y'a un effet de l' interaction entre les deux variables indépandantes sur deux variables ou plus dépendantes.

2.1 Généralité

Dans le cas de l'analyse de la variance univariée, l'hypothèse nulle a été

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_K,$$

où K représente le nombre total des niveaux du facteur.

Pour la MANOVA à un facteur, l'hypothèse nulle serait

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_K,$$

où μ_i $(i = \overline{1, K})$ est un vecteur des moyens.

Nous voulons tester si le vecteur des moyens est égale pour plusieurs groupes indépendants, et notre nouvelle hypothèse nulle serait

$$H_0 : \begin{pmatrix} \mu_{11} \\ \mu_{12} \\ \vdots \\ \mu_{1p} \end{pmatrix} = \begin{pmatrix} \mu_{21} \\ \mu_{22} \\ \vdots \\ \mu_{2p} \end{pmatrix} = \cdots = \begin{pmatrix} \mu_{K1} \\ \mu_{K2} \\ \vdots \\ \mu_{Kp} \end{pmatrix},$$

où p représente le nombre total de variables dépendantes pour les niveaux du facteur.

Rappelons que pour l'ANOVA, la statistique de test est le rapport entre les carrés moyens associés au facteur et les carrés moyen résiduels.

Pour MANOVA 1 (Analyse de la variance multiple à un facteur), notre statistique de test est calculée comme le rapport entre les déterminants de deux matrices comme suit

$$\Lambda = \frac{|E|}{|T|} \backsim \Lambda_{p,\vartheta_H,\vartheta_E}, \text{ (Λ est une variable de Wilks de paramètre p est de degrés de liberté ϑ_H et ϑ_E)},$$

où $|E|$ et $|T|$ sont les déterminants des matrices des sommes carées et des produits résiduelles et totales respectivement, que nous les étudions dans la section suivante.

On rejette l'hypothèse H_0 si $\Lambda \leq \Lambda_{p,\vartheta_H,\vartheta_E}$.

Les paramètres dans la distrubution de Wilks sont

p : nombre des variables,

ϑ_E ,ϑ_H : sont les degrés de liberté associés au facteur et au résiduelle.

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

TAB. 2.1 – Les données d'une MANOVA 2.

2.2 Présentation des données d'une MANOVA 2

Soient A et B deux facteurs à I et J niveaux respectivement, pour chaque combinaison de ces deux facteurs, nous effectuons K mesures d'un vecteur Y de dimension p. Les résultats de l'expérience sont présentées sous la forme suivante

où

$$\mathbf{Y}_{ijk} = (Y_{ijk1}, Y_{ijk2}, ..., Y_{ijkp})', \forall i = \overline{1, I}, \forall j = \overline{1, J}, \forall k = \overline{1, K}.$$

2.3 Modèle de MANOVA 2 à effets fixe avec interaction

Le modèle mathématique associé à chaque observation est donné par

$$\mathbf{Y}_{ijk} = \mu + \mathbf{a}_i + \mathbf{b}_j + \mathbf{c}_{ij} + \varepsilon_{ijk} = \mu_{ij} + \varepsilon_{ijk},$$
$$i = 1, 2, \ldots, I,\ j = 1, 2, \ldots, J,\ k = 1, 2, \ldots, K,$$

où $\mathbf{a}_i$ est l'effet du $i^{ème}$ niveau d'un facteur A sur chacune des p variables de $\mathbf{Y}_{ijk}$, $\mathbf{b}_j$ est

Source de variation	Somme des carrés et des produits	df
A	$H_A = KJ\sum_i(\overline{\mathbf{Y}}_{i..} - \overline{\mathbf{Y}}_{...})\left(\overline{\mathbf{Y}}_{i..} - \overline{\mathbf{Y}}_{...}\right)'$	$I-1$
B	$H_B = KI\sum_j\left(\overline{\mathbf{Y}}_{.j.} - \overline{Y}_{...}\right)\left(\overline{\mathbf{Y}}_{.j.} - \overline{\mathbf{Y}}_{...}\right)'$	$J-1$
AB	$H_{AB} = K\sum_{ij}\left(\overline{\mathbf{Y}}_{ij.} - \overline{\mathbf{Y}}_{i..} - \overline{\mathbf{Y}}_{.j.} - \overline{\mathbf{Y}}_{...}\right)\left(\overline{\mathbf{Y}}_{ij.} - \overline{\mathbf{Y}}_{i..} - \overline{\mathbf{Y}}_{.j.} - \overline{\mathbf{Y}}_{...}\right)'$	$(I-1)(J-1)$
Erreur	$E = \sum_{ijk}\left(\mathbf{Y}_{ijk} - \overline{\mathbf{Y}}_{ij.}\right)\left(\mathbf{Y}_{ijk} - \overline{\mathbf{Y}}_{ij.}\right)'$	$IJ(K-1)$
Totale	$T = \sum_{ijk}\left(\mathbf{Y}_{ijk} - \overline{\mathbf{Y}}_{...}\right)\left(\mathbf{Y}_{ijk} - \overline{\mathbf{Y}}_{...}\right)'$	$IJK-1$

TAB. 2.2 – Table d'analyse de la variance multiple à 2 facteurs.

l'effet du $j^{ème}$ niveau de B et $\mathbf{c}_{ij}$ est l'effet d'interaction. Nous utilisons les conditions

$$\sum_i \mathbf{a}_i = \sum_j \mathbf{b}_j = \sum_i \mathbf{c}_{ij} = \sum_j \mathbf{c}_{ij} = \mathbf{0},$$

et les ε_{ijk} sont indépendamment et identiquementdistrubués selon la loi $\mathcal{N}_p\left(\mathbf{0},\ \sum\ \right)$. Sous la condition $\sum_i \mathbf{a}_i = 0$, l'effet de A est la moyenne sur les niveaux de B, donc $\mathbf{a}_i = \overline{\mu}_{i.} - \overline{\mu}_{..}$, où $\overline{\mu}_{i.} = \sum_j \mu_{ij}/J$, et $\overline{\mu}_{..} = \sum_{ij}\mu_{ij}/IJ$, il y'a des définitions similaires pour $\mathbf{b}_j$ et $\mathbf{c}_{ij}$. Comme dans le cas univarié, le vecteur moyen $\overline{\mathbf{Y}}_{i..}$ est défini par $\overline{\mathbf{Y}}_{i..} = \sum_{jk}\mathbf{Y}_{ijk}/KJ$. Les moyennes $\overline{\mathbf{Y}}_{.j.}$, $\overline{\mathbf{Y}}_{ij.}$, et $\overline{\mathbf{Y}}_{...}$, ont des définitions analogues : $\overline{\mathbf{Y}}_{.j.} = \sum_{ik}\mathbf{Y}_{ijk}/KI$, $\overline{\mathbf{Y}}_{ij.} = \sum_k \mathbf{Y}_{ijk}/K$, $\overline{\mathbf{Y}}_{...} = \sum_{ijk}\mathbf{Y}_{ijk}/KIJ$.

Le tableau de l'analyse de la variance multiple à deux facteurs contient les matrices des sommes carrés et des produits et les degrés de liberté associés et il est de la forme suivant. Noton que les degrés de liberté sont les mêmes dans le cas univarié.

Pour le modèle à deux facteurs plan équilibré, la matrice des sommes carrés et des produits totales et partitionnée comme

$$T = H_A + H_B + H_{AB} + E. \tag{2.1}$$

La structure des matrices H_A et H_B est similaire.

Par exemple H_A a sur la diagonale, la somme des carrés pour le facteur A pour chacune des p variables, les éléments hors diagonale de H_A sont les sommes correspondantes de produits pour toutes les paires de variables. Ainsi, le $r^{\grave{e}me}$ élément de H_A correspondant à la r $^{\grave{e}me}$ variable ($r = 1, 2, \ldots, p$) est donnée par

$$h_{Arr} = KJ\sum_{i=1}^{I}\left(\overline{Y}_{i\cdot\cdot r} - \overline{Y}_{\cdots r}\right)^2 = \sum_{i=1}^{I}\frac{Y_{i\cdot\cdot r}^2}{KJ} - \frac{Y_{\cdots r}^2}{KIJ}, \tag{2.2}$$

où $\overline{Y}_{i\cdot\cdot r}$ et $\overline{Y}_{\cdots r}$ représentent les $r^{\acute{e}mes}$ composantes de $\overline{\mathbf{Y}}_{i\cdot\cdot}$et $\overline{\mathbf{Y}}_{\cdots}$ respectivement, et $Y_{i\cdot\cdot r}$ et $Y_{\cdots r}$ sont des totaux correspondants aux $\overline{\mathbf{Y}}_{i\cdot\cdot r}$ et $\overline{\mathbf{Y}}_{\cdots r}$. Le $(rs)^{\grave{e}me}$ élément de H_A est

$$h_{Ars} = KJ\sum_{i=1}^{I}\left(\overline{Y}_{i\cdot\cdot r} - \overline{Y}_{\cdots r}\right)\left(\overline{Y}_{i\cdot\cdot s} - \overline{Y}_{\cdots s}\right) = \sum_{i=1}^{I}\frac{Y_{i\cdot\cdot r}Y_{i\cdot\cdot s}}{KJ} - \frac{Y_{\cdots r}Y_{\cdots s}}{KIJ}. \tag{2.3}$$

De la formule (2.1) et le tableau (2.1), nous obtenons

$$h_{ABrr} = \sum_{ij}\frac{Y_{ij\cdot r}^2}{K} - \frac{Y_{\cdots r}^2}{KIJ} - h_{Arr} - h_{Brr}, \tag{2.4}$$

$$h_{ABrs} = \sum_{ij}\frac{Y_{ij\cdot r}\ Y_{ij\cdot s}}{K} - \frac{Y_{\cdots r}Y_{\cdots s}}{KIJ} - h_{Ars} - h_{Brs}.$$

Pour la matrice E, les formules de clalcul de E sont basés sur (2.1)

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

Ainsi les éléments de E ont la forme

$$e_{rr} = \sum_{ijk}Y_{ijkr}^2 - \frac{Y_{\cdots r}^2}{KIJ} - h_{Arr} - h_{Brr} - h_{ABrr}, \tag{2.5}$$

$$e_{rs} = \sum_{ijk}Y_{ijkr}Y_{ijks} - \frac{Y_{\cdots r}Y_{\cdots s}}{KIJ} - h_{Ars} - h_{Brs} - h_{ABrs}.$$

2.4 Tests d'hypothèse

Comme dans le cas univarié, il existe trois hypothèses à tester, qui sont

1/ Effet du premier facteur

H_0 : "Les paramètres a_i sont tous nuls " contre H_1 : "les paramètres $\mathbf{a}_i$ ne sont pas tous nuls".

2/ Effet du second facteur

H_0 : "Les paramètres $\mathbf{b}_j$ sont tous nuls " contre H_1 :"les paramètres $\mathbf{b}_j$ ne sont pas tous nuls ".

3/ Effet de l'interaction des deux facteurs

H_0 : "Les paramètres $\mathbf{c}_{ij}$ sont tous nuls" contre H_1 :"les paramètres $\mathbf{c}_{ij}$ ne sont pas tous nuls ".

Le test de validité de ces hypothèses peut se faire par différentes méthodes, que nous examinerons dans ce qui suit.

2.4.1 Le test de Wilks

Les matrices H_A, H_B et H_{AB} peut être comparées à E pour tester l'effet des deux facteurs A et B et l'effet d'interaction . Alors, pour chaque cas et sous l'hypothèse H_0, les statistiques Lambda de Wilks (Λ) utilisées pour tester l'effet de A, B et l'effet de l'interaction de cet ordre sont définies comme suit

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

Notons que H_0 sera rejetée pour petites valeurs de Λ.

2.4.2 Autres tests

Dans la littérature statistique, on trouve d'autres tests permettant d'aprouver la même hypothèse nulle. Nous donnons leur principe ci-dessous.

Notons que dans cette section , H représente (dénote) l'hypothèse testée dans le modèle du MANOVA. Par exemple, lorsque on veut tester l'effet d'interaction $H = H_{AB}$.

2.4.3 Le test de Roy

La statistique du test de Roy (appellé aussi le test de la plus grande racine de Roy).se base sur la plus grande valeur propre de la matrice $E^{-1}H$, qu'on la note λ_1 Pour tester H_0, on utilise la statistique

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

on rejette H_0 au risque α si $\theta \geq \theta_{\alpha,s,m,N}$,où

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

dont les valeurs critiques de sont données sous forme de tableaux.

Une autre approximation qui permette de mettre en oeuvre ce test est celle de Fisher, définie par

$$F = \frac{\left(\vartheta_E - d - 1\right)\lambda_1}{d},$$

avec d et $\vartheta_E - d - 1$ degrés de liberté, où $d = \max\left(p, \vartheta_H\right)$.

2.4.4 Le test de Pillai

Le test de Pillai est basé également sur les valeurs propres de la matrice $E^{-1}H$.

La statistique de ce test est

$$V^{(s)} = tr\left[(E+H)^{-1}H\right] = \sum_{i=1}^{s} \frac{\lambda_i}{1+\lambda_i}, s = \min\left(\vartheta_H, p\right).$$

On rejette H_0 , au seuil α si $V^{(s)} \geq V_\alpha^{(s)}$, où les valeurs critiques de $V_\alpha^{(s)}$ sont tabulées. La valeur $V_\alpha^{(s)}$ est indexée par les paramètres s,m et N qui sont les mêmes paramètres que ceux du test de Roy (pour plus de détail voir [8]).

2.4.5 Le test de lawley -Hotelling

La statistique de Lawley-Hotelling est définie par

$$U^{(s)} = tr\left(E^{-1}H\right) = \sum_{i=1}^{s} \lambda_i; s = \min\left(\vartheta_H, p\right).$$

2.5 Exemple d'application

Le tableau (2.2) contient des données concernant une expérience sur 32 pièces de barres en acier. L'objectif de cette expérience est d'étudier l'effet des deux facteurs vitesse de rotation (A) et lubrifiants (B) qui ont 2 et 4 niveaux respectivement sur deux variables mesurées sur chaque pièce de barre en acier, qui sont

Y_1 : le moment de la force,

Y_2 : la déformation.

Les tabeaux suivants résument les calculs des totaux pour chaque variable, pour chaque combinaison des deux facteurs (dans les cellules) et les toataux marginaux sont pour chaque niveau de A et B.

En utilisant les formules de calcul de h_{Arr} dans (2.2), l'élément (1,1) du H_A (correspondant

	A_1		A_2	
	Y_1	Y_2	Y_1	Y_2
B_1	7.80	90.40	7.12	85.1
	7.10	88.90	7.06	89.0
	7.89	85.90	7.45	75.9
	7.82	88.80	7.45	77.9
B_2	9.00	82.50	8.19	66.0
	8, 43	92, 40	8.25	74.5
	7.65	82.40	7.45	83.1
	7.70	87.40	7.45	86.4
B_3	7.28	79.6	7.15	81.2
	8.96	95.1	7.15	72.0
	7.75	90.2	7.70	79.9
	7.80	88.0	7.45	71.9
B_4	7.60	94.1	7.06	81.2
	7.70	86.6	7.04	79.9
	7.82	85.9	7.52	86.4
	7.80	88.8	7.70	76.4

TAB. 2.3 – Les valeurs de Y1 et Y2 selon les deux facteurs.

	A_1	A_2	
B_1	30.61	29.08	59.69
B_2	32.61	31.34	64.12
B_3	31.79	29.45	61.24
B_4	30.22	29.32	59.54
	125.40	119.19	244.59

TAB. 2.4 – Les totaux pour Y1.

	A_1	A_2	
B_1	354.0	327.9	681.90
B_2	344.70	310.0	654.70
B_3	352.90	305.0	657.90
B_4	355.40	323.90	679.3
	1407.0	1266.80	2673.8

TAB. 2.5 – Les totaux pour Y2.

à Y_1) est donné par

$$h_{A11} = \frac{(125.40)^2 + (119.19)^2}{(4)\,(4)} - \frac{(244.59)^2}{(4)\,(4)\,(2)} = 1.205.$$

Pour l'élément (2,2) du H_A (correspondant à Y_2) nous avons

$$h_{A22} = \frac{(1407.0)^2 + (1266.8)^2}{16} - \frac{(2673.8)^2}{32} = 614.25.$$

Pour l'élément (1,2) du H_A (correspondant à $Y_1\ Y_2$) nous utilisons (2.3) pour obtenir h_{Ars}

$$h_{A12} = \frac{(125.40)\,(1407.0) + (119.19)\,(1266.8)}{16} - \frac{(244.59)\,(2673.8)}{32} = 27.208.$$

Donc

$$H_A = \begin{pmatrix} 1.205 & 27.208 \\ 27.208 & 614.251 \end{pmatrix}.$$

Nous obtenons H_B de manière similaire

$$h_{B11} = \frac{(59.69)^2 + \cdots + (59.54)^2}{(4)\,(2)} - \frac{(244.59)^2}{32} = 1.694,$$

$$h_{B22} = \frac{(681.9)^2 + \cdots + (679.3)^2}{8} - \frac{(2673.8)^2}{32} = 74.874,$$

$$h_{B12} = \frac{(59.69)\,(681.9) + \cdots + (59.54)\,(679.3)}{8} - \frac{(244.59)\,(2673.8)}{32} = -9.862.$$

d'où

$$H_B = \begin{pmatrix} 1.694 & -9.862 \\ -9.862 & 74.874 \end{pmatrix}.$$

Pour H_{AB} nous avons, par (2.4)

$$h_{AB11} = \frac{(30.61)^2 + \cdots + (29.32)^2}{4} - \frac{(244.59)^2}{32} - 1.205 - 1.694 = 0.132,$$

$$h_{AB22} = \frac{(354.0)^2 + \cdots + (323.9)^2}{4} - \frac{(2673.8)^2}{32} - 614.25 - 74.874 = 32.244,$$

$$h_{AB12} = \frac{(30.61)(354.0) + \cdots + (29.32)(323.9)}{4} - \frac{(244.59)(2673.8)}{32} - 27.208 - (-9.862) = 1.585,$$

alors

$$H_{AB} = \begin{pmatrix} 0.132 & 1.585 \\ 1.585 & 32.244 \end{pmatrix}.$$

La matrice d'erreur E est obtenue en utilisant les formules de calcul données pour e_{rr} et e_{rs} dans (2.5). Par exemple, e_{11} et e_{12} sont calculés comme

$$e_{11} = (7.80)^2 + (7.10)^2 + \cdots + (7.70)^2 - \frac{(244.59)^2}{32} - 1.205 - 1.694 - 0.132 = 4.897,$$

$$e_{12} = (7.80)(90.40) + \cdots + (7.70)(76.4) - \frac{(244.59)(2673.8)}{32} - 27.208 - (-9.862) - 1.585 = -1.890.$$

En procédant de cette manière, nous obtenons

$$E = \begin{pmatrix} 4.897 & -1.890 \\ -1.890 & 736.390 \end{pmatrix}.$$

avec

$$\vartheta_E = IJ(K-1) = (2)(4)(4-1) = 24.$$

Pour tester l'effet principal de A avec le test de de Wilks, nous calculons

$$\Lambda_A = \frac{|E|}{|E + H_A|} = \frac{3602.2}{7600.2} = 0.474 < \Lambda_{0,05,2,1,24} = 0.771,$$

et nous concluons que la vitesse a un effet significatif sur Y_1 ou Y_2 ou sur les deux.

Pour l'effet principal de B, nous avons

$$\Lambda_B = \frac{|E|}{|E + H_B|} = \frac{3602.2}{5208.6} = 0.6916 > \Lambda_{0,05,2,3,24} = 0.591.$$

Nous concluons que l'effet des lubrifiants n'est pas significative.

Pour l'interaction entre A et B, on obtient

$$\Lambda_{AB} = \frac{|E|}{|E + H_{AB}|} = \frac{3602.2}{5208.6} = 0.6916 > \Lambda_{0,05,2,3,24} = 0.591,$$

par conséquent nous concluons que l'effet d'interaction n'est pas significative.

Chapitre 3

Application Numérique sous logiciel R

L'analyse de la variance multiple à deux facteurs (MANOVA 2) est une méthode visant à expliquer des variables quantitatives par deux variables explicatives qualitatives (appelées facteurs). Dans ce chapitre, nous allons présenter sous le logiciel R cette méthode en utilisant deux exemples numériques réalisées sur des données réelles.

3.1 Exemple 1

Les données résumées dans le tableau (3.1) 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) a été mesurée chez 12 oiseaux des deux sexes ayant subi ou non l'administration d'un traitement hormonal ([26]). Dans cet exemple l'objectif est de tester l'existence d'un effet de sexe, de traitement et de leur interaction sur la concentration de calcium dans le plasma et le taux d'eau perdu.

Les deux facteurs sont codés de la manière suivante

Pas de Traitement hormonal : h1, Traitement hormonal : h2 ;

Femelle : s1, Male : s2.

Pas de Traitement hormonal				Traitement hormonal			
Femelle		Male		Femelle		Male	
Ca	H20	Ca	H20	Ca	H20	Ca	H20
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. 3.1 – Concentration en calcium (en mg/100 ml) et taux d'eau perdu (mg/min) des individus en fonction de leur sexe et de l'administration d'hormone.

Pour entrer les données sous R, on utilise ces instructions

```
CA<-c(16.50,14.50,18.40,11.0,12.70,10.80,39.10,32.0,26.2,23.8,21.3,28.8)

H20<-c(76,80,71,72,64,77,71,65,70,69,63,67)

hormone<-factor(gl(2,6,labels=c("h1","h2")))

sexe<-factor(gl(2,1,12,labels=c("s1","s2")))
```

On place les 2 variables CA et H2O dans un data.frame convenable comme suit

```
Y=cbind(CA,H2O)
```

Les sommes des carrées et des produits sont obtenues à l'aide de la fonction manova, comme on montre ci dessous

```
manova(Y~hormone*sexe)
```

Call :

manova(Y ~hormone * sexe)

Terms :

	hormone	sexe	hormone :sexe	Residuals
resp 1	635.1075	14.7408	7.2075	228.7533
resp 2	102.0833	18.7500	36.7500	151.3333
Deg. of Freedom	1	1	1	8

Residual standard error : 5.347351 4.349329

Estimated effects may be unbalanced

Les tests sur les effets principaux des 2 facteurs et l'effet d'interaction se fait en tapant la commande suivante

```
summary(manova(Y~hormone*sexe))
```

	Df	Pillai	approx F	num Df	den Df	Pr(>F)	
hormone	1	0.85574	20.7624	2	7	0.001140	**
sexe	1	0.25408	1.1922	2	7	0.358439	
hormone :sexe	1	0.30894	0.30894	2	7	0.274351	
Residuals	8						

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

Au risque $\alpha = 0.05$, les résultats des tests indiquent un effet significatif du facteur hormone sur les deux variables CA et H2O puisque la p valeur est inférieure à 0.05 ($0.001140 < 0.05$), tandis que l'effet du facteur (sexe) n'est pas significatif ($0.358439 > 0.05$). Il n'y a par ailleurs pas d'effet d'interaction entre les deux facteurs ($0.274351 > 0.05$) : l'effet du facteur hormone est le même quelque soit le sexe des individus.

Remarque 3.1.1 - *Les autres tests sont également disponibles sous R, et ils sont obtenus en exécutant ces instructions*

```
summary(manova(Y~hormone*sexe),test="W")
summary(manova(Y~hormone*sexe),test="H")
summary(manova(Y~hormone*sexe),test="R")
```

- *Ces trois tests donnent les mêmes résultats que ceux obtenus par le test de Pillai.*

3.2 Exemple 2

L'expérimentation est menée sur des grains de blé dur (Triticum durum Desf) fournies par C.C.L.S (coopératives de céréales et légumes secs) de Biskra, les génotypes qui sont retenus sont : **HEDBA**, **BELYOUNI**, **BIDI 17**. Les essais ont été conduits au laboratoire de biologie d'université Mohamed Khider de Biskra El-Hadjeb.

Selon le protocole expérimental suivi, 20 grains sont placés dans une boîte de Pétri et ils sont tapissés par deux couches de papier absorbant humidifiée par des différentes concentrations de NaCl (0 g/l, 5g/l, 10g/l, 15g/l). Chaque traitement est répété trois fois. Après 15 jours, la longueur de radicule (racine) et de l'épicotyle (feuille) est mesurée (en cm). Alors, chaque observation est présentée sous la forme suivante

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

où

l_1 : la longueur de radicule ;

l_2 : la longueur de l'épicotyle.

FIG. 3.1 – Radicules et épicotyles des graines germées.

Dans cet exemple, on veut étudier l'effet de variété de blé dur et de la concentration de NaCl sur la longueur de radicule et la longueur de l'épicotyle.

La technique la plus adéquate pour traiter ce problème est la MANOVA 2 (les deux facteurs simultanément). Notons qu'on peut aussi faire une MANOVA1 (chaque facteur indépendamment).

Dans cette section, on va résoudre ce problème par une ANOVA 2 (l'étude de l'effet des deux facteurs sur chaque variable séparément), puis on va appliquer la MANOVA 2.

	Variété	Concentration	Moyenne		Variété	Concentration	Moyenne
Radicule	v1	c1	5.4317	Epicotyle	v1	c1	3.8667
		c2	5.0800			c2	2.9217
		c3	3.2500			c3	0.7750
		c4	0.9567			c4	0.2167
		Total	3.6796			Total	1.9450
	v2	c1	6.8983		v2	c1	4.6433
		c2	4.6400			c2	$2,0900$
		c3	2.1767			c3	0.4033
		c4	0.5650			c4	0.0983
		Total	3.5700			Total	1.8087
	v3	c1	7.3967		v3	c1	$5,7500$
		c2	4.6050			c2	$2,6400$
		c3	2.1250			c3	0.3050
		c4	0.5633			c4	0.0600
		Total	3.6725			Total	2.1888
	Total	c1	6.5756		Total	c1	4.7533
		c2	$4.,7750$			c2	2.5506
		c3	2.5172			c3	0.4944
		c4	$0,695$			c4	0.1250
		Total	3.640694			Total	1.9808

TAB. 3.2 – Résultas du calcul des moyennes.

3.2.1 Application de L'ANOVA 2

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

Remarque 3.2.1 *On a utilisé les notations v1, v2, v3 pour le facteur variété et c1, c2, c3 et c4 pour les niveaux du facteur concentration.*

L'analyse préliminaire des moyennes, montre que

Les trois types de variétés ont une longueur moyenne des radicules presque similaire, tandis qu'elles ont des longueurs moyennes de l'épicotyle différentes.

Lorsque la concentration de NaCl augmente, la longueur des radicules et de l'épicotyle diminue pour toutes les variétés.

Response RADICULE :

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
VARIETE	2	1.8	0.90	0.7857	0.4562	
CONCENTRATION	3	3571.17	1190.36	1036.1911	<2e-16	***
VARIETE :CONCENTRATION	6	186.4	31.06	27.0411	<2e-16	***
Residuals	708	813.3	1.15			

Response EPICOTYLE :

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
VARIETE	2	17.79	8.90	18.042	2.279e-08	***
CONCENTRATION	3	2459.67	819.89	1662.946	<2.2e-16	***
VARIETE :CONCENTRATION	6	119.35	19.89	40.346	<2.2e-16	***
Residuals	708	349.07	0.49			

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

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

Les résultats de l'ANOVA 2 fournis par le logiciel R pour les deux variables radicule et épicotyle sont présentés respectivement dans le tableau (3.3).

D'après les résultats de l'analyse de la variance et au seuil de risque $\alpha = 0.05$, on constate que

D'une part, il n'existe pas un effet significatif du facteur variété sur la longueur des radicules puisque la p valeur est supérieure à α $(0.4562 > 0.05)$, mais il est hautement significatif sur la longueur de l'épicotyle (la p valeur est inférieure strictement à α).

D'une autre part, il y' a des effets très hautement significatifs de la concentration de NaCl et ainsi de l'interaction (variété : concentration) sur les deux variables (radicule, épicotyle).

3.2.2 Application de MANOVA 2

L'application de la technique MANOVA 2 nous à fournis les résultats suivants

De la comparaison des p valeurs de chaque terme indiquées dans les tableaux de test MANOVA (tableau (3.4)) au seuil de signification α $(\alpha = 0.05)$, on résulte que

Effet du facteur variété

D'une part, la p valeur associée à ce facteur est strictement inférieure à α (cela pour les

	Df	Pillai	approx F	num Df	den Df	Pr(>F)
VARIETE	2	0.05238	9.521	4	1416	1.350e-07
CONCENTRATION	3	1.11315	296.219	6	1416	<2.2e-16
VARIETE :CONCENTRATION	6	0.32792	23.142	12	1416	<2.2e-16
Residuals	708					
	Df	Wilks	approx F	num Df	den Df	Pr(>F)
VARIETE	2	0.94768	9.63	4	1414	1.108e-07
CONCENTRATION	3	0.08368	579.03	6	1414	<2.2e-16
VARIETE :CONCENTRATION	6	0.68408	24.63	12	1414	<2.2e-16
Residuals	708					
	Df	Hotelling-Lawley	approx F	num Df	den Df	Pr(>F)
VARIETE	2	0.0552	9.73	4	1412	9.101e-08
CONCENTRATION	3	8.5986	1011.77	6	1412	<2.2e-16
VARIETE :CONCENTRATION	6	0.4443	26.14	12	1412	<2.2e-16
Residuals	708					
	Df	Roy	approx F	num Df	den Df	Pr(>F)
VARIETE	2	0.0540	19.13	2	708	8.118e-09
CONCENTRATION	3	8.3158	1962.52	3	708	<2.2e-16
VARIETE :CONCENTRATION	6	0.4005	47.25	6	708	<2.2e-16
Residuals	708					

Tab. 3.4 – Analyse de la variance multiple (MANOVA 2) avec les quatres statistiques du test.

quatre statistiques du test), donc il existe une différence significative entre les longueurs moyennes des radicules des trois types de variétés ou les longueurs moyennes des épicotyles des trois types de variétés ou les deux.

Ce résultat est confirmé par les résultats présentés dans le tableau (3.3) où on a montré que pour un seuil de risque α, le facteur variété a un effet significatif sur l'épicotyle mais elle n'a pas un effet significative sur la longueur des radicules.

Effet du facteur concentration

Les différents tests statistiques indiquent un effet hautement significatif du facteur concentration sur la longueur des épicotyles ainsi que sur la longueur des radicules. Ces résultats sont confondus avec les résultats obtenus précédemment.

Effet de l'interaction entre les deux facteurs

Les résultats de l'analyse de la variance multiple indiquent aussi qu'il existe un effet d'interaction entre les deux facteurs variété et concentration très hautement significatif sur la longueur des radicules et la longueur des épicotyles. Ces résultats sont similaires à ceux présentés dans le tableau (3.3).

Conclusion

Enfin L'analyse de la variance multivariée (MANOVA) utilise dans le même cadre conceptuel que l'ANOVA. Il s'agit d'une extension de l'ANOVA permettant de prendre en compte une combinaison de variables dépendantes plutôt qu'une variable dépendante unique. Dans le cadre de la MANOVA, les variables explicatives sont souvent appelées facteurs.

Ainsi l'avantage de l'utilisation d'une MANOVA au lieu de plusieurs ANOVA simultanées réside dans le fait qu'elle prend en compte les corrélations entre les variables réponses et permet ainsi une meilleure utilisation des informations provenant des données. La combinaison des variables dépendantes peut représenter une variable non mesurable directement. Donc la MANOVA teste les effets de facteurs sur plusieurs variables réponses. Avec une MANOVA, il est donc possible de tester conjointement toutes les hypothèses testées par une série d'ANOVAs avec plus de chance d'observer un effet significatif.et utilisée pour effectuer des tests multivariés dont le but est de vérifier si les paramètres correspondant aux différentes modalités d'un facteur sont significativement différents ou non. Par exemple, on peut tester les effets de quatre traitements appliqués à des plantes sur une qualité de production représentée par une combinaison de variables, en incluant ou non des effets d'interaction.

L'objectif du présent mémoire est de discuter le cas de MANOVA 02 c'est -à-dire l'étude de l'effet d'un 2 facteur fixe sur plusieurs variables, la réalisations de l'analyse de variance multiple est équivalente à la réalisation d'un test ou plusieurs tests

Bibliographie

[1] Mouloud Cherfaoui, (2017) Polycopiés du cours Biostatistique, Statistique Appliquées à l'Expérimentation En Science Biologique, UNIVERSITE MOHAMED KHIDER , BISKRA

[2] Dr.J.Kyle Roberts(2002) one-Way Multivariate Analyse of Variance MANOVA Southern Methodist University Simmons School of Education and Human Devellopement Departement of Teaching and Learning .

[3] Pierre.Dagnelie , (1975). Analyse de la variance à plusieur variables. Les press agronomique de Gembeux ASBL.

[4] Fanny MEYER , Morgane CADRAN , Margaux GAILLARD ,Analyse de la variance M2 Statistique et Econometrie .

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

[6] P. Mayer (Mars 2011) Analyse de la variance , comparaison de Plusieurs Moyennes Laboratoire de Biostatistique et informatique Médicale , fac de Medecine de Strasbourg .

[7] Proffesseur P. Francour (22.05.2015) Analyse de variance à un ou plusieurs Facteurs , Régression , Analyse de covariance , Modèles Linéaires Généralisés .
francour@unice .fr

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

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

Annexe A : Rappel sur l'algèbre linéaire et statistique

Dans cette Annexe on va présenter quelques notions sur l'algèbre Linéaire (matrice, l'inverse d'une matrice,...) ainsi que les statistiques multidimensionnelles (notations, espérance ,...) ou nous focalisons principalement sur le cas d'un vecteur gaussien .

3.3 Rappel sur l'algèbre linéaire [3, 9]

Définition 3.3.1 (*Matrice*)

Une matrice X est un tableau rectangullaire de nombre.On dit que X est de taille $n * p$,si X a n lignes et p colonnes .Une telle matrice est repésentée de la manière suivante :

$$X = \begin{pmatrix} x_{11} & \dots & x_{1p} \\ \vdots & \vdots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix}$$

ou x_{ij} est l'élément de la $i^{ème}$ ligne de la $j^{ème}$ colonne , on le note aussi $(X)_{ij}$.

Définition 3.3.2 (*La matrice identité*)

La matrice identité est une matrice de dimension n, notée I_n, telle que :

$$x_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i = j \end{cases}$$

c'est-à dire :

$$I_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 \end{pmatrix}$$

Définition 3.3.3 (*la matrice nulle*) *Une matrice est dite nulle si ses éléments sont toutes nuls, c'est-à-dire :*

$$x_{ij} = 0 \; \forall i = \overline{1,n} \; \forall j = \overline{1,p}$$

Définition 3.3.4 (***Transposé***)

La transposée de la matrice$X$, notée X^t, est obtenue à partir de X en interchangeant les lignes et les colonnes, c'est-à-dire : $(X^t) = (X_{ji})$. Remarquer que si X est une matrice de taille $n * p$, alors X^t est de taille $p * n$.

Définition 3.3.5 (***Symétrie***)

On dit qu'une matrice est symétrique si $X^t = X$.

Définition 3.3.6 (***la trace***)

La trace d'une matrice carée X d'ordre p, notée $tr\,(X)$ est la somme de ses éléments diagonaux, c'est-à-dire :

$$tr\,(X) = \sum_{i=1}^{p} x_{ii}.$$

Définition 3.3.7 (***Le déterminant***)

A toute matrice carrée d'ordre p, correspond un nombre réel ; noté $|X|$ ou $\det(X)$; appelé le déterminant de X , qui peut être obtenu comme suit :

• Pour une matrice d'ordre 1 :

$$|x_{11}| = x_{11};$$

• Pour une matrice d'ordre 2 :

$$\begin{vmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{vmatrix} = x_{11}x_{22} - x_{21}x_{12};$$

• D'une façon générale , le déterminant d'une matrice carrée X peut être calculé grâce aux notions de mineurs ou de cofacteurs.

Le mineur de l'élément x_{ij} d'une matrice X est le déterminant que l'on obtient en éliminant la $i^{éme}$ ligne et la $j^{ème}$ colonne X .

Le cofacteur X_{ij} de x_{ij} est égal au mineur de x_{ij} multiplié par $(-1)^{i+j}$.

Le déterminant de X peut être calculer en effectuant la somme des produits des différents éléments d'une même colonne par leurs cofacteurs respectifs :

$$\begin{aligned} \det(X) &= \sum_{i=1}^{p} x_{jj} X_{jj} \quad \text{(pour tout } j) \\ &= \sum_{i=1}^{p} x_{ij} X_{ij} \quad \text{(pour tout } i) \end{aligned}$$

Définition 3.3.8 (***L'inverse d'une matrice***)

Notons que la transposée de la matrice des cofacteurs X_{ij} est appelée " matrice adjointe " est désignée par $adj(X)$.

La matrice inverse de X ; notée X^{-1} ; peut être obtenue notamment en divisant la matrice adjointe de X par le $det(X)$.

$$X^{-1} = \frac{adj\,(X)}{\det(X)}; \qquad \det(X) \neq 0.$$

Soit X une matrice de taille $n * n$, les trois conditions suivantes sont équivalentes :

1.$\det X = 0$.

2. il existe un vecteur colonne $V \in \mathbb{R}^n$ est un vecteur colonne de taille n, alors $XV = 0 \Longrightarrow 0_{\mathbb{R}^n}$.

3.X possède une matrice inverse , c'est-à-dire il existe une matrice X^{-1} telle que : $X^{-1}X = XX^{-1} = I_n$.

Définition 3.3.9 (***Les valeurs propres***)

Les valeurs propres ou les valeurs caractérestiques d'une matrice carrée X d'ordre p,sont les p solutions de l'equation caractérestique :

$$\det(X - \lambda I_p) = 0.$$

Les valeurs propres peuvent être réelles ou complexes, positives, nulles ou négatives, distinctes ou confondues , mais on s'interesse ici aux valeurs propres réelles, alors que leurs sommes est égale à la trace de la matrice X et leurs produits à son déterminant :

$$\sum_{i=1}^{p} \lambda_i = tr(X);$$
$$\prod_{i=1}^{p} \lambda_i = \det(X).$$

3.4 Rappel sur les statistiques multidimensionnelles [7]

3.4.1 Notations matricielles de vecteurs aléatoires

Pour établir les propriétés d'un p-uplets $(X_1,X_2,\ldots,X_p)$ de variables aléatoires, il est préférable d'adopter la notation matricielle. Ainsi, on note :

$$X = (X_1, X_2, \ldots, X_p)^t,$$

le vecteur de dimension p a valeur dans $\mathbb{R}^p$.

On définit alors l'espérance mathématique de X, noté $\mathbb{E}[X]$, par le vecteur des espérances mathématique (si elle existent) :

$$\mathbb{E}[X] = (\mathbb{E}[X_1], \mathbb{E}[X_2], \ldots, \mathbb{E}[X_p])^t.$$

Si les covariances des composantes prises 2 à 2 existent, la matrice d'élément (i,j) égal à $\operatorname{cov}(X_i, X_j)$ est appelle matrice des variances-covariances de X de taille $p * p$ et on la note $V(X)$ ou $\sum$,

$$V(X) = \sum = \begin{pmatrix} \sigma_{X_1}^2 & \operatorname{cov}(X_1, X_2) & \ldots & \operatorname{cov}(X_1, X_p) \\ \operatorname{cov}(X_2, X_2) & \sigma_{X_2}^2 & \cdots & \operatorname{cov}(X_2, X_p) \\ \vdots & \vdots & \ddots & \vdots \\ \operatorname{cov}(X_p, X_1) & \operatorname{cov}(X_p, X_2) & \cdots & \sigma_{X_p}^2 \end{pmatrix}$$

Notons que cette matrice est symétrique et que ses éléments diagonaux sont les variances des composantes du vecteur aléatoire X , en effet :

$$\begin{aligned}
\operatorname{cov}(X_i, X_j) &= \mathbb{E}\left[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])\right] \\
&= \mathbb{E}\left[(X_j - \mathbb{E}[X_j])(X_i - \mathbb{E}[X_i])\right] \text{(la commutativitée de produit dans } \mathbb{R}); \\
&= \operatorname{cov}(X_j, X_i),
\end{aligned}$$

et

$$\begin{aligned}
\operatorname{cov}(X_i, X_i) &= \mathbb{E}\left[(X_i - \mathbb{E}[X_i])(X_i - \mathbb{E}[X_i])\right] \\
&= \mathbb{E}\left[(X_i - \mathbb{E}[X_i])^2\right] \\
&= Var(X_i).
\end{aligned}$$

Soient maintenant A une matrice de taille $q * p$ et C un vecteur appartient à $\mathbb{R}^q$, Alors la relation $Y = AX+C$ définit un vecteur aléatoire a valeur dans $\mathbb{R}^q$, dont les caractérestiques sont résumées dans la proposition suivante :
Soit X un vecteur aléatoire d'espérance $\mathbb{E}[X]$ et de matrice de variance-covariance $V(X)$ et soit le vecteur Y tel que :

$$\mathbb{E}[Y] = A\mathbb{E}[X] + C \text{ et } V(Y) = AV(X)A^t.$$

3.4.2 Lois de Gauss multivariée[7, 9]

Dans ce passage nous considérons un vecteur aléatoire $X = (X_1, X_2, \ldots, X_p)$ tel que $X_i \rightsquigarrow \mathcal{N}(\mu_i, \sigma_i^2), \forall\ i = \overline{1,p}$.

Définition 3.4.1 *un vecteur aléatoire gaussien X de dimension p est parfaitement défini par son vecteur des espérances noté μ et sa matrice de variance-covariance notée $\sum$. Sa loi est noté $\mathcal{N}_p(\mu, \Sigma)$.*

la densité conjointe de ses composantes au point $x = (x_1, x_2, \ldots, x_p)$ est :

$$f_X(x_1, x_2, \ldots, x_p) = \frac{1}{(2\pi)(\det \Sigma)^{1/2}} \exp\left\{-\frac{1}{2}(X-\mu)^t \Sigma^{-1} (X-\mu)\right\}.$$

Proposition 3.4.1 *on dit qu'un vecteur aléatoire gussien a des composantes linéairement indépandantes si et seulement si sa matrice des variances-covariances est diagonale , c'est-à-dire si et seulement si les covariances des composantes prises deux à deux sont nulles* $(\operatorname{cov}(X_i, X_j) = 0\ , i \neq j)$.

3.5 Rappel sur quelque distributions [9]

Définition 3.5.1 (***La distrubution de WISHART***) *Les distrubutions d'échantillonage des matrices des sommes des carrés et des produits des écarts A sont des distributions de WISHART de paramètres k, p ,Σ , ou :*

p : le nombres des variables considérées.

n_i : l'effectif de l'échantillon i ;

k : le degrés de liberté,avec $k = n_i - 1$;

Σ : la matrice des variance − covariance de la population parent.

dont la densité de probabilité donnée par :

$$f(A) = c\frac{|A|^{(k-p-1)/2} \exp\left(-\frac{1}{2} trA^{-1}\Sigma\right)}{|\Sigma|^{k/2}};$$

avec c est une constante déterminée en fonction de p et n_i, de manière à donner la valeur 1 à l'integrale multiple de f .

Définition 3.5.2 (***La distrubution de WILKS***) *Les distrubutions de WILKS sont relatives à des expressions du type*

$$\Lambda = \frac{|A_1|}{|A_1 + A_2|};$$

ou les matrices A_1 et A_2 étant des variables de WISHART indépandantes à k_1 et k_2 degré de liberté et de même paramètres p et Σ .

Définition 3.5.3 (***la distribution de KHI-deux***)*Soit* $Z_1,\ldots,Z_v$ *une suite de variables aléatoires gaussiennes centrées réduites* $\left(\mathbb{E}\left[Z_i\right] = 0 \text{ et } V\left(Z_i\right) = 1 \text{ , } \forall\ i = \overline{1, v}\right)$, *alors* $X = \sum_{i=1}^{v} Z_i^2$ *est une variable aléatoire qui suit loi appelée* *"**loi de KHI-deux à** v **degrés de liberté** "* *, on la note* χ_v^2, *et sa densité de probabilité est donnée par :*

$$f\left(x\right) = \frac{1}{2^{v/2}\Gamma\left(\frac{v}{2}\right)} e^{\frac{-x^2}{2}} \left(x^2\right)^{\frac{p}{2}-1},$$

$$\text{avec } \Gamma\left(x\right) \ = \int_0^x t^{x-1} e^{-t} dt.$$

Définition 3.5.4 (***La distrubution de Fisher-Snedecor***) *étant deux variables aléatoires suivants indépandamment des lois* $\chi_{v_1}^2$ *et* $\chi_{v_2}^2$ *alors :*

$$F = \frac{X_1/v_1}{X_2/v_2},$$

est une variable aléatoire qui suit une loi de Fisher a v_1 et v_2 degrés de liberté, avec la densité de probabilité correspondante est donnée par :

$$f\left(x\right) = \frac{1}{\beta\left(\frac{v_1}{2}, \frac{v_2}{2}\right)} \frac{\left(\frac{v_1}{v_2}\right)^{v_{1/2}} x^{\frac{v_1}{2}-1}}{\left(1 + \frac{v_1}{v_2} x\right)^{(v_1+v_2)/2}}, \text{ avec } \beta\left(p, q\right) = \frac{\Gamma\left(p\right)\Gamma\left(q\right)}{\Gamma\left(p+q\right)} .$$

Annexe B : Quelques commandes de logiciel R

Le tableau ci-dessous présente les principales fonctions à utiliser afin d'effectuer d'analyse de varince et la régression

Commande	**Description**
plot($Y \backsim$ factor (X))	inspection graphique .
aov($Y \backsim$ factor (X))	construit un modéle d'analyse de la variance de la réponse y par le facteur.
summary(aov $(Y \backsim$ factor $(X)))$	tableau d'analyse de la variance $((ANOVA\ 1))$.
anova(Im $(Y \backsim$ factor $(X)))$	tableau d'analyse de la variance $(ANOVA\ 1)$.
pairwise.t.test( )	comparaisons deux à deux
shapiro.test	Permet de réaliser un test de normalité de **shapiro-wilk**.
bartlett.test	Permet de réaliser un test **Bartlett** d'égalité de variance.
interaction.plot(Y, factor (X), factor (Z)))	inspection graphique.
aov($Y \backsim$ factor $(X) *$ factor (Z))	analyse de la variance à 2 facteurs avec interaction.
summary(aov $(Y \backsim$ factor $(X) *$ factor $(Z)))$	tableau d'analyse de la variance $(ANOVA\ 2)$
names ( )	Liste les noms des variables dans un data.frame.
plot ( )	génère un nuage de points.
Im( )	Déterminer la ligne de régression des moindres carrés.
xyplot ( )	Commande Lattice pour produire un diagramme de dispersion.
confint( )	intervalle de confiance des paramètres de régression .
data.frame( )	Tableau à 2 dimensions dont les lignes sont des individus et les colonnes des variables. (numérique ou facteurs) .