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

Hassouni ouafa

Titre :

Analyse de variance à un facteur et deux facteurs

Membres du Comité d'Examen :

Pr.	MERAGANI DJAMEL	UMKB	Président
Pr.	CHINE AMEL	UMKB	Encadreur
Dr.	SOLTANE LUIZA	UMKB	Examinateur

11Juin 2024

Dédicace

À mon cher père Mustafa, ma référence et mon inspiration dans la vie ;c'est lui qui m'a appris à vivre avec dignité et fierté.

À ma chère mère... Les mots ne suffisent pas à lui rendre justice, elle est une épopée d'amour et une joie de vivre, un exemple de dévouement et de don de soi.

À mon époux,

À la famille de mon époux : CHAHDI

À mes sœurs Sanaa, Sakina, Salima, Nabila, Chaker, Karima, Saïda, Mohamed et Majda, qui ont partagé mes joies et mes peines.

À mes précieux fils Raif et Rassim.

À la merveilleuse personne qui n'a jamais cessé de prier pour moi, celle qui a une réputation impeccable,

Ma chère grand-mère.

Mes amis à l'université, Hanane et Mohamed.

Mes professeurs qui m'ont beaucoup aidé pendant mes études.

À chacun de mes étudiants, individuellement.

À tous ceux qui ont prié pour moi.

REMERCIEMENTS

Je suis très reconnaissante envers Dieu Tout-Puissant qui m'a donné la force et la foi pour mener à bien mon projet de recherche de fin d'études de cette manière.

Un grand merci à ma superviseure, **Dr. AMEL CHINE**, pour son soutien considérable, ses conseils et son suivi dans la réalisation de cette recherche et pour en avoir fait un succès.

Je tiens également à remercier chaleureusement les membres du jury **Pr.MERAGANI DJAMEL** et **Dr. SOLTANE LUIZA** d'avoir accepté d'évaluer et de juger mon travail.

Je saisis cette opportunité pour remercier tous les professeurs de la première et de la deuxième année de master pour leur généreux et précieux soutien qui nous a permis d'atteindre cette étape.

J'exprime ma gratitude envers le **Dr. JIBRAN YAHYA** pour son excellent service et l'aide qu'il a apportée dans le passé. Votre contribution est inestimable pour nous. Nous vous souhaitons plein succès dans vos projets futurs et que Dieu vous récompense pour tout le bien que vous avez fait pour nous.

Enfin, je tiens à exprimer ma profonde gratitude à tous ceux qui m'ont apporté leur aide et leur soutien pour mener à bien cette étude, en particulier

MOHAMED EL AMINE DAAS.

Table des matières

Table des figures

Liste des tableaux

Introduction

L'analyse de variance (ANOVA) constitue une méthode statistique employée afin d'examiner le comportement d'une variable quantitative à expliquer, aussi appelée variable d'intérêt, en relation avec une ou plusieurs variables qualitatives. En d'autres termes, elle vise à évaluer l'impact d'un facteur, voire plusieurs, sur une variable quantitative, en utilisant une série de modèles statistiques pour comparer les moyennes des divers échantillons indépendants. Chaque échantillon correspond à une modalité différente de la variable qualitative, et les moyennes sont calculées en référence à la variable quantitative.

Pour mener à bien cette tâche, il est essentiel de vérifier l'indépendance des échantillons, la normalité des distributions et l'homogénéité des variances. L'ajout des covariables dans le modèle permet de réduire considérablement la composante de variabilité associée à l'erreur aléatoire, augmentant ainsi la puissance du modèle.

En résumé, la structure de ce travail est la suivante :

Le premier chapitre : Le premier chapitre est consacré aux généralités des statistiques appliquées, telles que les tests d'hypothèses, ainsi que le test de normalilté et le test d''homogénéité des variances, où ces deux derniers tests sont des conditions nécessaires pour appliquer et réaliser l'analyse de variance.

Le deuxième Chapitre : Ce chapitre présente l'analyse de variance à un facteur ANOVA 1, nous discutons dans ce chapitre sur la nature des données de l'analyse de variance à un facteur et les différentes étapes qui nous permet de faire le test de l'ANOVA 1 et prendre la décision.

Le troisième Chapitre : Ce dernier chapitre est consacré à l'analyse de variance à deux facteurs ANOVA 2, l'objectif de cette méthode et le même que l'ANOVA 1, sauf que les données sont différentes un peux ou nous avons dans L'ANOVA 2 deux variables qualitatitives. nous présentons dans ce chapitre les même étapes qu'on a vu dans le deuxième chapitre mais cette fois on a deux types des données, données avec répétition et sans répétition.

Chapitre 1

Généralités

1.1 Hypothèse statistique

Un test d'hypothèses est une règle de décision qui permet, sur la base des données observées et avec des risques d'erreur déterminés,[14] d'accepter ou de refuser une hypothèse statistique.

1.1.1 Hypothèses

L'hypothèse nulle notée H_0 est l'hypothèse que l'on désire contrôler : elle consiste à dire qu'il n'existe pas de différence entre les paramètres comparés ou que la différence observée n'est pas significative et est due aux fluctuations d'échantillonnage.

Cette hypothèse est formulée dans le but d'être rejetée.

L'hypothèse alternative notée H_1 est la négation de H_0, elle est équivalente à dire « H_0 est fausse » [19]La décision de rejeter H_0 signifie que H_1 est réalisée ou H_1 est vraie.

1.1.2 Erreurs

La décision d'un test consiste à choisir entre H_0 et H_1, il y a donc quatre cas possibles :

1. Accepter H_0 alors qu'elle est vraie (Rejeter H_1)
2. Rejeter H_0 alors qu'elle est fausse.(Accepter H_1)
3. Accepter H_0 alors qu'elle est fausse (Rejeter H_1)
4. Rejeter H_0 alors qu'elle est vraie(Accepter H_1)

La décision prise est bonne dans les deux premiers cas, mais elle est erronée dans les deux dernières. On peut résumer ceci dans le tableau suivant :

	Réalité	
Décision / Vérité	H_0 vraie	H_0 fausse (H_1 vraie)
Accepter H_0	Pas d'erreur	Erreur de deuxième espèce
Rejeter H_0 (Accepter H_1)	Erreur de première espèce	Pas d'erreur

Tab. 1.1 – Table de décision d'un test statistique

On ne pourra jamais conclure avec certitude dans un test statistique. Il y aura toujours des erreurs de décision.

Pour effectuer le test statistique, il faudra choisir un certain risque d'erreur qui est la probabilité de se tromper en prenant la décision retenue. Il existe deux types d'erreurs :

Erreur de première espèce : on rejeter H_0 alors qu'elle est vraie.

Erreur de deuxième espèce : on accepter H_0 alors qu'elle est fausse.

1.1.3 Risque

La probabilité de l'erreur de première espèce est appelé risque de première espèce, on la note généralement par $\alpha(\theta)$ où

$$\begin{aligned} \alpha(\theta) &= \mathcal{P}_{H_0}(rejeter\ H_0). \\ &= \mathcal{P}(rejeter\ H_0/H_0\ est\ vraie). \\ &= \mathcal{P}(H_0/H_0). \end{aligned}$$

La probabilité de commettre une erreur de deuxième espèce est appelé risque de deuxième espèce, on la note généralement par $\beta(\theta)$ où :

$$\begin{aligned} \beta(\theta) &= \mathcal{P}_{H_1}(accepter\ H_0) \\ &= \mathcal{P}(accepter\ H_0\ /H_0\ est\ fausse) \\ &= \mathcal{P}(H_0\ /\ H_1). \end{aligned}$$

Le tableau suivant résume simplement le rôle de ces probabilités :

Décision / Vérité	H_0	H_1
H_0	$1-\alpha$	β
H_1	α	$1-\beta$

TAB. 1.2 – Table de probabilité de bonne et mauvaise décision dans un test d'hypothèse

1.1.4 Niveau de signification

Le niveau de signification (ou seuil de signification) est égale un risque de première espèce maximum. On le note par α. Les valeurs usuelles sont $1\%, 5\%$ et 10%.

1.1.5 P-valeur

En statistique, la p-valeur représente la probabilité, sous l'hypothèse nulle, d'observer une valeur aussi extrême ou plus extrême que celle effectivement observée dans l'échantillon pour un modèle statistique donné. Elle est utilisée pour évaluer la crédibilité de l'hypothèse nulle.

Pour prendre une décision basée sur la p-valeur, on la compare généralement à un seuil de signification prédéfini, noté α. Voici comment interpréter cette comparaison :

▷ **Si** la p-valeur est inférieure ou égale à α, on rejette l'hypothèse nulle H_0.

▷ **Si** la p-valeur est supérieure à α, on accepte l'hypothèse nulle H_0.

En résumé, une p-valeur faible suggère que les données observées sont peu probables sous l'hypothèse nulle, ce qui conduit au rejet de cette hypothèse. En revanche, une p-valeur élevée indique que les données observées sont cohérentes avec l'hypothèse nulle, donc on ne la rejette pas.

1.1.6 Variable de décision

La variable de décision ou la statistique de test est la variable aléatoire que l'on utilise pour tester l'hypothèse nulle, c'est la valeur calculée à partir de nos échantillons que nous allons comparer à la valeur critique. Il y a de nombreuses manières de calculer, ceci en fonction de la situation et de ce que nous souhaitons faire. Cette valeur se calcule en fonction du test que nous avons choisi initialement[14]. On choisit cette statistique de test pour pouvoir calculer sa loi sous H_0.

1.1.7 Région de rejet et Région d'acceptation

La région critique d'un test, est un ensemble des valeurs $(x_1, x_2, ..., x_n)$ de la variable de décision, qui conduisent à écarter H_0 au profit de H_1, et la note :

$$W = \{(x_1, x_2, ..., x_n) \in \mathbb{R}^n / \ H_0 \ est \ rejetée\}.$$

Elle est défini par :

$$\alpha = \mathcal{P}(W \setminus H_0).$$

La région complémentaire est appelé région d'acceptation, est on la note $\overline{W}$:Elle est défini par :

$$1 - \alpha = \mathcal{P}(\overline{W} \setminus H_0).$$

La forme de la région de rejet définit la latéralité du test :

1. **test multilatéral** :On veut rejetter H_0 si S_{obs} est trop grand ou trop petit,sans à priori. La région de rejet est alors de la forme $]-\infty \ , a] \cup [b \ , +\infty[$.
2. **test unilatéral à droite** :On veut rejetter H_0 seulement si S_{obs} est trop grand.La région de rejet est alors de la forme $[a \ , +\infty[$.
3. **test unilatéral à gauche** : On veut rejetter H_0 seulement si S_{obs} est trop petit. La région de rejet est alors de la forme $]-\infty,b]$

Définition 1.1.1 *Pour un test d'hypothèse nulle multiple ,on appelle niveau du test (ou seuil du test) si :[10]*

$$\alpha = \sup_{\theta \in \theta_0} \alpha(\theta).$$

1.2 Conditions d'application de l'analyse de la variance

L'application et la validité de l'analyse de variance reposent sur le test de Fischer, qui repose sur trois conditions, à savoir :

1.2.1 Indépendance des échantillons

Cette condition peut être vérifiée facilement. Chaque élément statistique doit être associé à une seule modalité de la variable qualitative, et seulement à celle-ci[5]. On peut également appliquer le test du chi-deux d'indépendance en comparant les échantillons deux par deux de manière réciproque.

1.2.2 Normalité des distributions

Il existe plusieurs tests pour affirmer la normalité d'une distribution mentionnez l'un d'eux test de Shapiro-Wilk, test de Kolmogorov-Smirnov, test de normalité de Lilliefors, test de Anderson-Darling. on s'interesse sur le test de Shapiro- Wilk, ce dernier permet de vérifier la normalité d'une distribution en évaluant si un échantillon $x_1, x_2, ..., x_n$ de taille $n < 30$ est issu d'une population distribuée selon une loi normale.

Test Shapiro-Wilk

On résume ce test dans les étapes suivantes :

- **Etape 01** : Les hypothèse à tester sont :

$$\begin{cases} H_0 : \text{La population est normalement distribuée} \\ H_1 : \text{La population n'est pas normalement distribuée} \end{cases}$$

• **Etape 02** : La statistique du test est :

$$W = \frac{\left(\sum_{j=1}^{k} a_j d_j\right)^2}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}.$$

avec : $d_j = X_{n-j+1} - X_j$

$$\begin{cases} k = \frac{n}{2} & \text{si } n \text{ est pair} \\ k = \frac{n-1}{2} & \text{si } n \text{ est impair} \end{cases}$$

et Les coefficients a_j étant donnés en fonction de n par une table, on dispose d'autre part d'une autre table donnant, en fonction de n, quelques fractiles usuels de la distribution de W.

• **Etape 03** : Décision pour un risque α

▷ Si la p-valeur $< \alpha$ alors on rejette l'hypothèse nulle. Cela signifie que la condition de la normalité n'est pas vérifiée.

▷ Si la p-valeur $> \alpha$ alors on ne rejette pas l'hypothèse nulle [16].Cela signifie que la condition de la normalité est vérifiée.

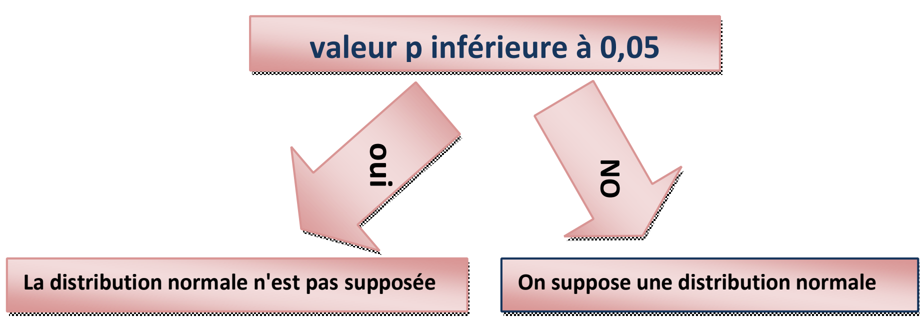

FIG. 1.1 – Un schéma illustrant la sélection α

1.2.3 Homogénéité des variances

les variances dans chaque groupe doivent être à peu près les mêmes. C'est un test d'homgénéité des variances, on résume ce test dans les étapes suivantes :

– **Etape 01** : Les hypothèses à tester sont :

$$\begin{cases} H_0 : \forall i \ \sigma_i^2 = \ \sigma^2 \\ H_1 : \exists i \ \ \sigma_i^2 \neq \sigma^2 \end{cases}.$$

– **Etape 02** : La statistique du test est en train d'être recalculée

$$B_{obs} = \frac{(n-p)\ln S^2 - \sum_{i=1}^{p}(n_i - 1)\ln S_i^2}{1 + \frac{1}{3(k-1)}\left(\sum_{i=1}^{p}\frac{1}{n_i - 1} - \frac{1}{n-p}\right)}.$$

Où S^2 est l'estimateur non biaisé de σ^2

$$S^2 = \frac{\sum_{i=1}^{p}(n_i - 1)S_i^2}{n-p}.$$

où : $n = \sum_{i=1}^{p} n_i$. p échantillons à comparer et S_i^2 la variance empirique du $i^{éme}$ échantillon et :

$$S_i^2 = \frac{1}{n_i}\sum_{j=1}^{n_i}(x_{ij} - \overline{x}_i)^2 .$$

– **Etape 03** : Sous l'hypothèse H_0 : $B_{obs} \sim \chi^2_{p-1}$. Alors la région critique est : $B_{obs} > \chi^2_{p-1,\alpha}$.

• L'ANOVA suppose l'homogénéité des variances entre les groupes que nous souhaitons comparer. Si cette hypothèse n'est pas respectée (c'est-à-dire si les variances des différents groupes sont significativement différentes), alors les résultats de l'ANOVA peuvent être faussés. Si l'homogénéité des variances n'est pas vérifiée, il existe plusieurs approches pour remédier à ce problème, notamment : utilisation de transformations,utilisation de tests alternatifs, et la regression robuste.

• Il est important de prendre en compte ces considérations lors de l'interprétation des résultats de l'ANOVA et de choisir la méthode appropriée en fonction de la structure des données et des objectifs de l'étude.

• L'analyse de la variance est une technique qui sert à tester l'influence d'un (ou de plusieurs) facteur (s) qualitative (s) sur une variable quantitative. [9]

• L'analyse de variance permet désormais de comparer plusieurs groupes simultanément. Pour deux groupes($k = 2$), l'analyse de variance est équivalente au test t. La variable indépendante est une variable nominale avec au moins deux valeurs distinctes, tandis que la variable dépendante est mesurée sur une échelle métrique. Dans le contexte de l'analyse de variance, la variable indépendante est appelée facteur. L'analyse de variance permet de déterminer s'il existe des différences entre au moins deux groupes.Cependant, elle ne permet pas d'identifier spécifiquement quels groupes diffèrent les uns des autres.

Chapitre 2

Analyse de variance à un facteur

L'analyse de variance à un facteur, ANOVA 1 est une technique qui permet d'étudier l'homogénéité entre plus de deux groupes, ou en d'autres termes, d'étudier l'effet d'une variable qualitative (le facteur) sur une variable quantitative. Dans ce chapitre, nous concentrons sur l'ANOVA 1, où nous commençons par définir les données et après avoir présenté toutes les étapes de calcul, nous terminons ce chapitre par un exemple.

2.1 Données, modèle et hypothèse

2.1.1 Tableau des données

Les données se présentent sous forme de tableaux de nombres comme suit. Notons A le facteur et $A_1, ..., A_p$ ses p modalités ou niveaux. Soient X la variable étudier et $x_1, x_2, ..., x_n$ ses observations. Pour chaque niveau A_j du facteur est associer n_j mesures de X : $x_{1j}, x_{2j}, ..., x_{n_j j}$..

Niveau du Facteur A	A_1	A_2	...	A_i	...	A_p
	x_{11}	x_{21}		x_{i1}		x_{p1}
	x_{12}	x_{12}	$\cdots$	x_{i2}	$\cdots$	x_{p2}
	$\vdots$	$\vdots$		$\vdots$		$\vdots$
	x_{1n_1}	x_{2n_2}		x_{in_i}		x_{pn_p}
Effectifs	n_1	n_2	$\cdots$	n_i	$\cdots$	n_p
Moyennes empiriques	$\overline{x}_{1\bullet}$	$\overline{x}_{1\bullet}$	$\cdots$	$\overline{x}_{i\bullet}$	$\cdots$	$\overline{x}_{p\bullet}$

TAB. 2.1 – Tableau des données de l'ANOVA 1.

où :

- x_{ij} : le $i^{éme}$ observation du $j^{éme}$ niveau ;
- n_j : la taille de l'échantillon $x_{\bullet j}$;
- $\overline{x}_j$: la moyenne de la classe j;
- n : la taille de l'échantillon X
- $\overline{x}$: la moyenne totale.

Les différentes moyennes que l'on peut calculer de ces données sont notées comme suit :

$$x_{i\bullet} = \frac{1}{q}\sum_{j=1}^{q} x_{ij} \qquad \text{moyenne de ligne}$$

$$x_{\bullet j} = \frac{1}{p}\sum_{i=1}^{p} x_{ij} \qquad \text{moyenne de colonnes}$$

$$x_{\bullet\bullet} = \frac{1}{pq}\sum_{i=1}^{p}\sum_{j=1}^{q} x_{ij} = \frac{1}{p}\sum_{i=1}^{p} x_{i\bullet} = \frac{1}{q}\sum_{j=1}^{q} x_{\bullet j} \qquad \text{moyenne générale}$$

2.1.2 Modèle de l'ANOVA 1

Le modèle de l'ANOVA 1 peut s'écrire :

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

où :

x_{ij} : valeur de la variable de réponse j-ième essai modalité i du facteur

μ : moyenne globale de la variable dépendante

α_i : effet de la modalité i sur la variable dépendante

ε_{ij} : termes d'erreur aléatoire, tels que l'on ait indépendance des ε_{ij} et que $\varepsilon_{ij} \sim N\left(0, \sigma^2\right)$;

$i = 1, 2, ..., p$ et $j = 1, 2, ..., n_i$

Les hypothèses classiques pour les erreurs sont :

$$E\left[\varepsilon_{ij}\right] = 0 \quad \text{pour tout} \left(i; j\right).$$

$$Var\left[\varepsilon_{ij}\right] = \sigma^2 \quad \text{pour tout} \left(i; j\right).$$

Les (ε_{ij}) sont indépendantes, ce qui est assuré par la manière dont a été fait l'échantillonnage.

Les (x_{ij}) (et donc les ε_{ij}) sont des variables gaussiennes.[12].

2.1.3 Hypothèses de l'ANOVA 1

Les hypothèses de l'ANOVA 1 H_0 et H_1 sont données comme suit :

Hypothèse nulle H_0 : Il n'y a pas de différences entre les moyennes des p groupes, en d'autre terme :

$$H_0 : \mu_1 = \mu_2 = ... = \mu_p. \tag{2.1}$$

Hypothèse alternative H_1 : Il existe des différences aux moins entre deux moyennes des p groupes, en d'autre terme :

$$H_1 : \mu_i \neq \mu_j \quad \text{pour} \quad i, j \in \{1, ..., p\} \quad \text{et} \quad i \neq j. \tag{2.2}$$

2.2 Etapes de l'analyse de la variance

2.2.1 Conditions d'application de l'analyse de la variance

Afin de réaliser le test définie dans 2.1 et 2.2 , trois conditions doit être vérifiées préalablement, à savoir :

1. Les p échantillons comparés sont indépendants.
2. La variable quantitative étudiée suit une loi normale dans les p populations comparées.
3. Les p populations comparées ont même variance : Homogénéité des variances ou homoscédasticité.

Si ces dernières conditions sont vérifiées alors, on peut utiliser la technique ANOVA 1 [5]. Pour réaliser le test , et pour ce faire nous avons besoin les quantités (statistiques) suivantes :

2.2.2 Moyennes et variances

– **Moyenne de toutes les observations :** Elle est définie par :

$$\overline{X} = \frac{1}{n} \sum_{j=1}^{q} \sum_{i=1}^{p} x_{ij} \ , \qquad \text{avec} \qquad n = \sum_{j=1}^{q} n_j.$$

– **Moyenne de chaque échantillon :** Elle est définie par :

$$\overline{X}_j = \frac{1}{q}\sum_{i=1}^{p} x_{ij}\ , \qquad \text{pour} \qquad j = \overline{1,q}.$$

– **Variance de chaque échantillon :** Elle est définie par :

$$\hat{\sigma}_i^2 = \frac{1}{q}\sum_{i=1}^{p}\left(x_{ij} - \overline{X}_j\right)^2, \qquad \text{pour} \qquad j = \overline{1,q}.$$

– **Variance de toutes les observations** : Elle est définie par :

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{q}\sum_{i=1}^{p}\left(x_{ij} - \overline{X}\right)^2, \qquad \text{avec} \qquad n = \sum_{j=1}^{q} n_j.$$

2.2.3 Décomposition des carrés

– **Somme des carées totale :** la variation totale théorique, ou somme des carrés totale des écarts, noté SCE_T est égale à :

$$SCE_T = \sum_{i=1}^{p}\sum_{j=1}^{q}\left(x_{ij} - \overline{x}_{\bullet\bullet}\right)^2.$$

– **Somme des carées factorielle :** On appelle variation théorique due au facteur A, noté SCE_F, la quantité :

$$SCE_F = q\sum_{i=1}^{p}\left(\overline{x}_{i\bullet} - \overline{x}_{\bullet\bullet}\right)^2.$$

– **Somme des carées résiduelle :** On appelle variation résiduelle, noté SCE_R, la quantité :

$$SCE_R = \sum_{i=1}^{p}\left(\sum_{j=1}^{q}\left(x_{ij} - \overline{x}_{i\bullet}\right)^2\right).$$

Proposition 2.2.1 *La décomposition suivante est toujours vraie*

$$SCE_T = SCE_R + SCE_F.$$

En effet, on a :

$$\begin{aligned}SCE_T &= \sum_{i=1}^{p}\sum_{j=1}^{q}(x_{ij}-\overline{x}_{..})^2 = \sum_{i=1}^{p}\sum_{j=1}^{q}((x_{ij}-\overline{x}_{i.})+(\overline{x}_{i.}-\overline{x}_{..}))^2 \\ &= \sum_{i=1}^{p}\sum_{j=1}^{q}(x_{ij}-\overline{x}_{i.})^2 + \sum_{i=1}^{p}\sum_{j=1}^{q}(\overline{x}_{i.}-\overline{x}_{..})^2 + 2\sum_{i=1}^{p}\sum_{j=1}^{q}(x_{ij}-\overline{x}_{i.})(\overline{x}_{i.}-\overline{x}_{..}).\end{aligned}$$

Or

$$\sum_{i=1}^{p}\sum_{j=1}^{q}(x_{ij}-\overline{x}_{i.})(\overline{x}_{i.}-\overline{x}_{..}) = \sum_{i=1}^{p}(\overline{x}_{i.}-\overline{x}_{..})\sum_{j=1}^{q}(x_{ij}-\overline{x}_{i.}) = 0.$$

car $\sum_{j=1}^{q}(x_{ij}-x_{i.}) = qx_{i.}-qx_{i.} = 0$ pour $i = 1,2,...,p$. Donc

$$\begin{aligned}SCE_T &= \sum_{i=1}^{p}\sum_{j=1}^{q}(x_{ij}-\overline{x}_{i.})^2 + \sum_{i=1}^{p}\sum_{j=1}^{q}(\overline{x}_{i.}-\overline{x}_{..})^2 = \sum_{i=1}^{p}\sum_{j=1}^{q}(x_{ij}-\overline{x}_{i.})^2 + \sum_{i=1}^{p}j(x_{ij}-\overline{x}_{i.})^2 \\ &= SCE_R + SCE_F.\end{aligned}$$

Cette relation montre que la somme des carrés des écarts par rapport à la moyenne générale également appelé SCE_T peut être divisée en 2 composantes additives :[12]

– Une SCE factorielle (SCE entre échantillon ou entre traitement) SCE_F.

– Et une SCE résiduelle (ou variation á l'intérieur des échantillons) SCE_R[1]

Proposition 2.2.2 *Sous les hypothèses de normalité et d'égalité des variances, on a :*

$$MCF_A = \frac{SCE_F}{\sigma^2} \sim \chi^2_{p-1}\left(q\sum_{i=1}^{p}\alpha_i^2\right).$$

Sous les hypothèses usuelles de l'analyse de la variance, on a :

$$MCR = \frac{SCE_R}{\sigma^2} \sim \chi^2_{n-p}.$$

En effet, posons :

$$S_i^2 = \frac{1}{q-1} \sum_{j=1}^{q} (y_{ij} - y_{i.})^2 .$$

Pour $i = 1, 2, ..., p$ on a :

$$\frac{(q-1) S_i^2}{\sigma^2} \sim \chi^2_{q-1}.$$

Par indépendance des échantillons,

$$\frac{SCE_R}{\sigma^2} = \frac{\sum_{i=1}^{p} (q-1) S_i^2}{\sigma^2} \sim \chi^2_{n-p}.$$

puisque $\sum_{i=1}^{p} (q-1) = n - p$.

Sous les hypothèses habituelles de l'analyse de la variance, les statistiques SCE_F et SCR sont indépendantes et on a :

$$\frac{SCE_T}{\sigma^2} \sim \chi^2_{n-1} \left(\sum_{i=1}^{p} q (\mu_i - \mu)^2 / \sigma^2 \right) = \chi^2_{n-1} \left(q \sum_{i=1}^{p} \alpha_i^2 / \sigma^2 \right).$$

Pour $i = 1, 2, ..., p$ on a les statistiques $x_{i.}$ et S_i^2 sont indépendantes.[15]

Les statistiques SCF_A et SCE_R sont donc indépendantes puisque la première est une fonction de $\{x_{1.}, x_{2.}, ..., x_{p.}\}$. et la deuxième est une fonction de $\left\{S_1^2, S_2^2, ..., S_p^2\right\}$.

$$SCE_T = SCF_A + SCE_R \sim \chi^2_{n-1} \left(\sum_{i=1}^{p} q (\mu_i - \mu)^2 / \sigma^2 \right).$$

Sous H_0, on a $\mu_1 = \mu_2 = ... = \mu_p = \mu$, d'ou $\sum_{i=1}^{p} q (\mu_i - \mu)^2 / \sigma^2 = 0$.

Sous cette hypothèse, les trois statistiques SCE_T, SCF_A et SCE_R sont distribuées selon des lois du Khi-deux centrées[20], à $n-1$, $p-1$ et $n-1$degrés de liberté respectivement .

En divisant les SCE sur leurs ddl respectifs, on obtient ce qu'on appelle les carrés moyens qui serviront de base pour rejeter ou accepter H_0.

- **Carré moyen total** : on l'obtient en divisant la somme des carrée totale sur $n-1$:
$$CMT = \frac{SCE_T}{n-1}.$$
- **Carré moyen factorielle** : on l'obtient en divisant la somme des carrée factorielle sur $p-1$:
$$CMF_A = \frac{SCF_A}{p-1}.$$
- **Carré moyen résiduel** : on l'obtient par en divisant la somme des carrée résiduelle sur $n-p$:
$$CMR = \frac{SCE_R}{n-p}.$$

2.2.4 Test statistique

Statistique de test : Ce test statistique est employé pour déterminer si la variance factorielle est significativement supérieure à la variance résiduelle. Il s'agit du test du Fisher du rapport de ces deux variances. Sous les hypothèses habituelles de l'analyse de la variance, posons :
$$F_{obs} = \frac{SCF_A/\left(p-1\right)}{SCR/\left(n-p\right)}.$$
ou
$$F_{obs} = \frac{CMF_A}{CMR}.$$
Cette statistique suit la loi du Fisher[20] à $p-1$ et $n-p$ degrés de liberté. Alors la

valeur théorique $F_{théorique}$ est lu à partir de la table de Fischer – snedecor à $V_1 = p-1$ et $V_2 = n - p$ de degré de liberté pour un seuil α donnée.

Décision : Pour un seuil de risque donné α les tables de Fisher nous fournissent une valeur critique f_α telle que :

$$\mathcal{P}\left(\frac{CMF_A}{CMR} < f_\alpha\right) = 1 - \alpha.$$

Si $F_{obs} < f_\alpha\,(p-1, n-p)$ (ou $p > \alpha$) on accepte $H_0 \Rightarrow$ Il n'y a aucun effet des traitement sur les variables dans ce cas il faut voir la puissance de l'essai.

Si $F_{obs} > f_{1-\alpha}\,(p-1, n-p)$ (ou $p < \alpha$) on rejette $H_0 \Rightarrow$ l'effet des traitements est significatif [8], dans ce cas[7] on passe à la CMM (comparaison multiple des moyennes).

2.2.5 Tableau de L'ANOVA 1

Tous les calculs de L'ANOVA 1 sont résumé dans la table de l'analyse de variance, comme indique la table 2.2 :

Source de variation	Somme des carrés	Degrés de liberté	Carré moyen	F de Fisher
Facteur A	SCF_A	$p-1$	$CMF_A = \frac{SCF_A}{p-1}$	$F_A = \frac{CMF_A}{CMR}$
Totale	SCE_T	$n-1$	$CMT = \frac{SCT}{n-1}$	
Résiduelle	SCE_R	$n-p$	$CMR = \frac{SCR}{n-p}$	

TAB. 2.2 – Tableau d'analyse de variance à un facteur.

.

2.2.6 Comparaisons multiples : Méthode de Bonferroni

Si l'hypothèse nulle est rejetée, la question suivante consiste à rechercher quelles sont les groupes ou cellules qui possèdent des moyennes significativement différentes ($\exists \mu_i, \mu_{i'} / \mu_i \neq \mu_{i\prime}$). On peut donc chercher à identifier les couples $(i, i\prime)$ pour lesquels($\mu_i = \mu_{i\prime}$). Dans ce cas nous utilisons la méthode de Bonferrouni basée sur la comparaison deux à deux des couple($\mu_i ; \mu_{i\prime}$)

$$\begin{cases} H_0 : \mu_i - \mu_{i\prime} = 0 \\ \qquad i \neq i^{'} \\ H_1 : \mu_i - \mu_{i\prime} \neq 0 \end{cases}$$

Il y a donc $m = C_p^2$ comparaisons possibles tests de Student à faire. on accepte $H_0\,(\mu_i = \mu_{i\prime})$ si $|T| < t_{1-\frac{\alpha}{2m}(n-p)}$ où

$$T = \frac{\overline{y}_i - \overline{y}_{i'}}{\sqrt{\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right) MCR}}.$$

Exemple 2.2.1 *On désire comparer 3 types d'aliments en ce qui concerne leur effet sur la production laitière, à cet effet on prend* 15 *vaches, on sert le premier aliment* A_1 *aux* 5 *premières vaches choisi au hasard, l'aliment* A_2 *sera administré aux* 5 *autres vaches, l'aliment* A_3 *sera administré aux* 5 *dernières vaches et ceci au hasard [1]. Les données sont présentés dans la table 2.3*

A_1	A_2	A_3
38	42	30
40	45	32
41	43	41
35	44	34
36	39	33

TAB. 2.3 – production laitière de 15 vaches selon trois aliments.

Question : Tester l'hypothèse que les aliments n'ont aucun effet sur la production laitière des vaches pour un seuil de sgnification $\alpha = 5\%$.

Réponse : on pose H_0 :"les 3 aliments n'ont aucun effets sur la production laitière". Aprés les calculs des moyennes et des carré moyenne et la statistisque du test, on résume tous ça dans la table 2.4

Source de variation	ddl	SC	CM	F	P	ET	CV
Variation factorielle	2	185,2	92,60	9,48	0,0035	3,13	8,2%
Variation résiduelle	12	117,2	9,77				
Variation totale	14	302,4					

TAB. 2.4 – Tableau de l'analyse de variance.

La valeur critique $f_\alpha = 3.89$ (lu de la table de Fischer-Snédecore[20] á $\alpha = 0,05$ et ddl : 2 et 12). On observe que $F_{obs} > f_\alpha$ donc H_0 est rejetée , c'est à dire l'alimentation a bien un effet sur la production laitière sur la production laitière.

Chapitre 3

Analyse de variance à deux facteurs

L'analyse de la variance à deux facteurs ANOVA 2 explore l'impact de deux variables qualitatives indépendantes (facteurs) sur une variable quantitative dépendante. Elle enrichit l'analyse de variance à un facteur en incorporant une variable indépendante supplémentaire. Encore une fois, l'objectif est de déterminer si les moyennes des groupes diffèrent de manière significative. En incorporant deux facteurs dans l'analyse, on peut mieux comprendre comment ces facteurs interagissent et influent sur les résultats de l'expérience. Cette extension permet une analyse plus approfondie des relations entre les variables étudiées. L'objectif de ce chapitre est de présenter et de détailler cette méthode.

Définition 3.0.1 *L'analyse de la variance à deux facteurs teste l'effet de deux facteurs contrôlés* A *et* B *(variables qualitatives) ayant respectivement* p *et* q *modalités sur les moyennes d'une variable quantitative* X*.*

3.1 Modèle de l'ANOVA 2 avec répétition

3.1.1 Données

On cherche donc à étudier l'effet de deux facteurs qualitatifs A et B sur une variable quantitative X. On suppose que le facteur A a p niveaux (modalités) notons $A_1, A_2, ..., A_p$ et $B_1, B_2, ..., B_q$ les q niveaux du facteur B. Pour chaque couple $(i; j)$ de niveau, on dispose de n_{ij} mesures de X , notées x_{ijk} avec $i = 1, ..., p$, $j = 1, ..., q$ et $k = 1, ..., n_{ij}$. On note $n_{i.} = \sum_{j=1}^{q} n_{ij}$, le nombre de mesures de X pour la modalité i du facteur A, et $n_{.j} = \sum_{i=1}^{p} n_{ij}$, le nombre de mesures de X pour la modalité j du facteur B.

Un traitement (ij) est une combinaison des niveaux des 2 facteurs, k (≥ 1) est un indice de répétition du traitement (ij). Le tableau 3.1 présente les données de L'ANOVA 2.

N^0	B_1	...	B_j	...	B_q
A_1	$x_{111}, x_{112}, ...x_{11r}$	...	$x_{1j1}, x_{1j2}, ..., x_{1jr}$	...	$x_{1q1}, x_{1q2}, ..., x_{1qr}$
A_2	$x_{211}, x_{122}, ..., x_{12r}$	...	$x_{2j1}, x_{2j2}, ..., x_{2jr}$	...	$x_{2q1}, x_{2q2}, ..., x_{2qr}$
$\vdots$	$\vdots$	...	$\vdots$	...	$\vdots$
A_i	$x_{i11}, x_{i12}, ..., x_{i1r}$	...	$x_{ij1}, x_{ij2}, ..., x_{ijr}$	...	$x_{iq1}, x_{iq2}, ..., x_{iqr}$
$\vdots$	$\vdots$	...	$\vdots$	...	$\vdots$
A_p	$x_{p11}, x_{p12}, ..., x_{p1r}$	...	$x_{pj1}, x_{pj2}, ..., x_{pjr}$	...	$x_{pq1}; x_{pq2}, ..., x_{pqr}$

TAB. 3.1 – Les données d'ANOVA2 avec répétitions

3.1.2 Plan de l'expérience

L'ensemble des (n_{ij}) définit le plan d'expérience qui est déterminant pour la suite de l'étude. Pour pouvoir étudier simplement l'effet de chacun des deux facteurs A et B, il faut un plan d'expérience dit orthogonal.

On dira que le plan d'expérience est :

1. complet si $n_{ij} > 0$ pour tout traitement$(i;j)$;
2. à répétition si $n_{ij} > 1$ pour tout traitement $(i;j)$;
3. équilibré si $n_{ij} = r > 0$ pour tout traitement $(i;j)$;
4. orthogonal lorsque $n_{ij} = \frac{n_{i.} \times n_{.j}}{n_{..}}$pour tout traitement $(i;j)$ [13]

3.2 Modèle de l'ANOVA 2

Le modèle à deux facteurs croisés s'écrit sous la forme :

$$x_{ijk} = \mu_{ij} + \varepsilon_{ijk} \qquad \text{avec} \quad i = \overline{1,p} \quad \text{et} \quad k = \overline{1,r}.$$

où x_{ijk} est la $k^{ème}$ réalisation de la variable quantitative X , 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 $\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2)$.

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

$$x_{ijk} = \underbrace{\mu}_{\text{moyenne générale}} + \underbrace{\alpha_i}_{\text{effet de A}} + \underbrace{\beta_j}_{\text{effet de B}} + \underbrace{\gamma_{ij}}_{\text{effet d'intéraction}} + \underbrace{\varepsilon_{ijk}}_{\text{effet résiduel}} .$$

où

$$\begin{cases} \alpha_i = \mu_{i.} - \mu, & \text{effet de } A \text{ au niveau } i = 1, ..., p \\ \beta_j = \mu_{.j} - \mu, & \text{effet de } B \text{ au niveau } j = 1, ..., q \\ \gamma_{ij} = \mu_{ij} - \mu_{i.} - \mu_{.j} + \mu, & \text{effet d'interaction entre } A \text{ et } B \\ \varepsilon_{ijk} = x_{ijk} - \mu_{ij} \ , & \text{erreur } \ (k = 1, ..., r) \end{cases}$$

et sous les contraintes suivantes :

$$\sum_{i=1}^{p} \alpha_i = 0, \ \sum_{j=1}^{q} \beta_j = 0, \ \forall k, \sum_{j=1}^{q} \gamma_{jk} = 0 \text{ et } \forall j, \sum_{k=1}^{r} \gamma_{jk}.$$

qui découlent de la définition des effets et assurent l'unicité de la solution.

3.2.1 Hypothèses de l'ANOVA 2

Il y a maintenant trois hypothèses principales à tester :

– **<u>Pour le facteur</u>** <u>A</u>

$$\begin{cases} H_0 : \mu_1 = \mu_2 = ... = \mu_i \qquad \text{avec } \ i = \overline{1; p} \\ H_1 : \text{le facteur} A \text{ a un effet sur les résultats} \end{cases}$$

– **Pour le facteur** B

$$\begin{cases} H_0' : \mu_1 = \mu_2 = ... = \mu_j \qquad \text{avec } \ j = \overline{1; q} \\ H_1' : \text{le facteur B a un effet sur les résultats} \end{cases}$$

– **L'interaction entre les facteurs** A **et** B :

$$\begin{cases} H_0'' : \text{il n'y a pas d'interaction entre les facteurs } A \text{ et } B. \\ H_1'' : \text{les facteurs A et B interagissent sur les résultats} \end{cases}$$

– L'hypothèse H_1 : le facteur A a un effet sur les résultats, c'est-à-dire qu'au moins une des moyennes μ_i n'est pas égale aux autres.

– L'hypothèse H_1' : le facteur B a un effet sur les résultats, c'est-à-dire qu'au moins une des moyennes μ_j. n'est pas égale aux autres.

– L'hypothèse H_1'' : les facteurs A et B interagissent sur les résultats. Autrement dit, l'état du facteur A influence la réponse face au facteur B, [2] et réciproquement.

3.2.2 Etapes de L'ANOVA 2

Conditions d'application du test : Afin de réaliser une analyse de la variance à deux facteurs, les conditions suivantes doivent être vérifiées préalablement :

– Les pq échantillons comparés sont mutuellement indépendants.

– La variable quantitative étudiée suit une loi normale dans les pq populations comparées.

– Les pq populations comparées ont même variance : Homogénéité des variances.

Moyennes et somme des carrées : La première étape pour réaliser L'ANOVA 2 aprés la vérification des conditions, C'est le calcul des moyennes et les sommes des carrées, comme suit :

1. **moyenne de la ligne i :** Elle est définie par :

$$\overline{x}_{i\bullet\bullet} = \frac{1}{qr}\sum_{j=1}^{q}\sum_{k=1}^{r} x_{ijk}.$$

2. **moyenne de la colonne j :** Elle est définie par :

$$\overline{x}_{\bullet j\bullet} = \frac{1}{pr}\sum_{i=1}^{p}\sum_{k=1}^{r} x_{ijk}.$$

3. **moyenne de la case(i, j) :** Elle est définie par :

$$\overline{x}_{ij\bullet} = \frac{1}{r}\sum_{k=1}^{r} x_{ijk}.$$

4. **moyenne totale :** Elle est définie par :

$$\overline{x} = \overline{x}_{\bullet\bullet\bullet} = \frac{1}{pqr}\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r} x_{ijk} = \frac{1}{p}\sum_{i=1}^{p}\overline{x}_{i\bullet\bullet} = \frac{1}{q}\sum_{j=1}^{q}\overline{x}_{\bullet j\bullet} = \frac{1}{pq}\sum_{i=1}^{p}\sum_{j=1}^{q}\overline{x}_{ij\bullet}$$

où $n = pqr$.

<u>**Somme des carrées :**</u> Nous calculons les sommes suivantes :

– <u>**Somme des carrés totale**</u> : Elle égale à :

$$SC_T = \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r} \left(x_{ijk} - \overline{x}\right)^2.$$

– <u>**Somme des carrés des erreurs résiduelles**</u> : Elle égale à :

$$SC_R = \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r} \left(x_{ijk} - \overline{x}_{ij\bullet}\right)^2.$$

– <u>**Somme des carrés des erreurs du premier facteur**</u> : Elle égale à :

$$SC_A = \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r} \left(\overline{x}_{i\bullet\bullet} - \overline{x}\right)^2.$$

– <u>**Somme des carrés des erreurs des deux facteurs**</u> : Elle égale à :

$$SC_B = \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r} \left(\overline{x}_{\bullet j\bullet} - \overline{x}\right)^2.$$

– <u>**Somme des carrés des erreurs des deux facteurs**</u> : Elle égale à :

$$SC_{AB} = \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r} \left(\overline{x}_{ij\bullet} - \overline{x}_{i\bullet\bullet} - x_{\bullet j\bullet} + \overline{x}\right)^2.$$

Proposition 3.2.1 *La décomposition [18] suivante est toujours vraie.*

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

En effet, on a :

$$\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r}\left(x_{ijk}-\overline{x}\right)^2=\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r}\left(x_{ijk}-\overline{x}_{ij.}\right)^2+qr\sum_{i=1}^{p}\left(\overline{x}_{i..}-\overline{x}\right)^2+pr\sum_{j=1}^{q}\left(\overline{x}_{.j.}-\overline{x}\right)^2$$
$$+\,r\sum_{i=1}^{p}\sum_{j=1}^{q}\left(\overline{x}_{ij.}-\overline{x}_{i..}-x_{.j.}+\overline{x}\right)^2.$$

En d'autre terme

$$\underbrace{\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r}x_{ijk}^2-n\overline{x}^2}_{SC_T}=\underbrace{\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r}\left(x_{ijk}-\overline{x}_{ij.}\right)^2}_{SC_R}+\underbrace{qr\sum_{i=1}^{p}\overline{x}_{i..}^2-n\overline{x}^2}_{SC_A}+\underbrace{pr\sum_{j=1}^{q}\overline{x}_{.j.}^2-n\overline{x}^2}_{SC_B}$$
$$+\underbrace{r\sum_{i=1}^{p}\sum_{j=1}^{q}\left(\overline{x}_{ij.}-\overline{x}_{i..}-\overline{x}_{.j.}+\overline{x}\right)^2}_{SC_{AB}}.$$

Carrés moyens : On définit en suite les moyennes des carrées par :

$$MC_A=\frac{SC_A}{p-1}.$$
$$MC_B=\frac{SC_B}{q-1}.$$
$$MC_{AB}=\frac{SC_{AB}}{(p-1)(q-1)}.$$
$$MC_R=\frac{SC_E}{pq(r-1)}.$$

Notons que, si les trois conditions citées précédemment sont vérifiées alors sous l'hypothèse H_0.

$$\frac{MC_A}{MC_R}\rightsquigarrow F_{((p-1),pq(r-1))}.$$
$$\frac{MC_B}{MC_R}\rightsquigarrow F_{((q-1),pq(r-1))}.$$
$$\frac{MC_{AB}}{MC_R}\rightsquigarrow F_{((p-1)(q-1),pq(r-1))}.$$

où F est la loi de Fisher

Proposition 3.2.2 *Les statistiques MC_A, MC_B, MC_{AB} et MC_R sont indépendantes et distribuées selon*

$$\frac{(p-1)\,MCA}{\sigma^2} \sim \chi^2_{p-1}\left(\frac{qn}{\sigma^2}\sum_{i=1}^{p}\alpha_i^2\right).$$
$$\frac{(q-1)\,MCB}{\sigma^2} \sim \chi^2_{q-1}\left(\frac{pn}{\sigma^2}\sum_{j=1}^{q}B_j^2\right).$$
$$\frac{(p-1)\,(q-1)\,MC\,(AB)}{\sigma^2} \sim \chi^2_{(p-1)(q-1)}\left(\frac{n}{\sigma^2}\sum_{i=1}^{p}\sum_{j=1}^{q}\gamma_{ij}^2\right).$$
$$\frac{pq\,(r-1)\,MCR}{\sigma^2} \sim \chi^2_{pq(r-1)}.$$

3.2.3 Tableau d'ANOVA 2

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

Variation	ddl	SC	MC	F
Fact.A	$p-1$	SCA	MCA	F_A
Fact.B	$q-1$	SCB	MCB	F_B
Fact.A,B	$(p-1)\,(q-1)$	SCAB	MCAB	F_{AB}
Résiduelle	$pq\,(r-1)$	SCR	MCR	
Totale	$n-1$	SCT		

TAB. 3.2 – Tableau de l'ANOVA 2.

3.2.4 Décision

Pour un seuil de risque donné α, nous quantifions les valeurs critiques f_A, f_B et f_{AB} (par la lecture sur la table de Fisher) [7], telle que :

$$P\left(\frac{MCA}{MCR} < f_{_A}\right) = 1 - \alpha.$$
$$P\left(\frac{MCB}{MCR} < f_{_B}\right) = 1 - \alpha.$$
$$P\left(\frac{MC\,(AB)}{MCR} < f_{_{AB}}\right) = 1 - \alpha.$$

Ainsi les décisions des test se font comme suit :

Décision sur le premier facteur :

▷ Si $\frac{MCA}{MCR} < f_{_A}$, alors le premier facteur n'a pas une influence significative sur le caractère étudié, c'est à dire on accepte H_0.

▷ Si$\frac{MCA}{MCR} \geq f_{_A}$, alors le premier facteur a une influence significative sur le caractère étudié, c'est à dire on rejette H_0.

Décision sur le deuxième facteur :

▷ Si$\frac{MCB}{MCR} < f_{_B}$, alors le deuxième facteur n'a pas une influence significative sur le caractère étudié, c'est à dire on accepte H_0'.

▷ Si$\frac{MCB}{MCR} \geq f_{_B}$, alors le deuxième facteur a une influence significative sur le caractère étudié, c'est à dire on rejette H_0'.

Décision sur l'interaction des deux facteurs :

▷ Si $\frac{MC(AB)}{MCR} < f_{AB}$, alors l'interaction des deux facteurs n'a pas une influence significative sur le caractère étudié, c'est à dire on accepte H_0''

▷ Si$\frac{MC(AB)}{MCR} \geq f_{AB}$, alors l'interaction des deux facteurs a une influence significative sur le caractère étudié,[7] c'est à dire on rejette H_0''.

3.3 Modèle d'ANOVA 2 sans répétitions

Supposons dans ce cas que chaque case contient une seule observation, c'est à dire $r = 1$. [5] On dit aussi qu'une seule mesure de la variable X pour chaque couple $(A_i;$ $B_j)$.

Le modèle sans répétition est défini par :

$$x_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \text{ avec } i = \overline{1,p} \quad j = \overline{1,q}.$$

où :

x_{ij} : la valeur prise par la variable à expliquer X pour le couple(i, j).

μ : la moyenne et c'est l'effet global

α_i : l'effet de la modalité i du premier facteur A.

β_j : l'effet de la modalité j du deuxième facteur B.

ε_{ij} : l'erreur aléatoire

Les valeurs que peut prendre la variable x sont présentées dans le tableau suivant :

	B_1	$\cdots$	B_J	$\cdots$	B_q
A_1	x_{11}	$\cdots$	x_{11}	$\cdots$	x_{11}
A_2	x_{21}	$\cdots$	x_{2j}	$\cdots$	x_{2q}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_i	x_{i1}	$\cdots$	x_{ij}	$\cdots$	x_{iq}
$\vdots$	$\vdots$	...	$\vdots$	...	$\vdots$
A_P	x_{p1}	...	x_{pj}	...	x_{pq}

TAB. 3.3 – Table des données de L'ANOVA 2 sans répétition.

3.3.1 Équation d'ANOVA 2 sans répétitions

Présentation de l'équation

La somme des carrés totale SC_T qui est une simple addition de la somme des carrés liée au facteur A (SC_A) et la somme des carrés liée au facteur B (SC_B) et la somme des carrées des écarts ou des résidus (SC_R), représente l'équation fondamentale de l'ANOVA2 sans répétitions. Car les trois sources de variations sont les deux facteurs fixes A et B et les facteurs inconnus (résidus) :

$$\sum_{i=1}^{p}\sum_{j=1}^{q}\left(x_{ij}-\overline{x}\right)^2=\sum_{i=1}^{p}\sum_{j=1}^{q}\left(x_{ij}-\overline{x}_{i\bullet}-\overline{x}_{\bullet j}+\overline{x}\right)^2+q\sum_{i=1}^{p}\left(\overline{x}_{i\bullet}-\overline{x}\right)^2+p\sum_{j=1}^{q}\left(\overline{x}_{\bullet j}-\overline{x}\right)^2.$$
$$\sum_{i=1}^{p}\sum_{j=1}^{q}\left(x_{ij}-\overline{x}\right)^2=\sum_{i=1}^{p}\sum_{j=1}^{q}\left(x_{ij}-\overline{x}_{i\bullet}-\overline{x}_{\bullet j}+\overline{x}\right)^2+q\sum_{i=1}^{p}\overline{x}_{i\bullet}^2-n\overline{x}^2+p\sum_{j=1}^{q}\overline{x}_{\bullet j}^2-n\overline{x}^2.$$

donc

$$SCT=SCA+SCB+SCR.$$

avec :

$$SCT=\sum_{i=1}^{p}\sum_{j=1}^{q}\left(x_{ij}-\overline{x}\right)^2=\sum_{i=1}^{p}\sum_{j=1}^{q}\left(x_{ij}-\overline{x}\right)^2.$$
$$SCR=\sum_{i=1}^{p}\sum_{j=1}^{q}\left(x_{ij}-\overline{x}_{i\bullet}-\overline{x}_{\bullet j}+\overline{x}\right)^2.$$
$$SCA=q\sum_{i=1}^{p}\left(\overline{x}_{i\bullet}-\overline{x}\right)^2=q\sum_{i=1}^{p}\overline{x}_{i\bullet}^2-n\overline{x}^2.$$
$$SCB=p\sum_{j=1}^{q}\left(\overline{x}_{\bullet j}-\overline{x}\right)^2=p\sum_{j=1}^{q}\overline{x}_{\bullet j}^2-n\overline{x}^2.$$

Différents tests de Fischer

Dans le cas de l'ANOVA2 sans répétition, nous avons deux tests de Fischer à réaliser :

– Si l'objectif est d'étudier l'effet du facteur A sur la variable X, les hypothèses sont :

Test 1 $\begin{cases} H_0 : \text{Le facteur } A \text{ n'a pas d'effet significatif sur } X \text{ ie : } \forall i \;\; \alpha_i = 0. \\ H_1 : \text{Le facteur } A \text{ a un effet significatif sur } X \text{ ie : } \exists\, i \; \alpha_i \neq 0. \end{cases}$

Sous l'hypothèse nulle, la statistique du test de Fischer est :

$$F = \frac{\frac{SCA}{p-1}}{\frac{SCR}{(p-1)(q-1)}} \sim F_{((p-1),(p-1)(q-1))}.$$

avec $(p-1),(p-1)(q-1)$ ddl.

– Si l'objectif est d'étudier l'effet du facteur B sur la variable X , les hypothèses sont :

Test 2 $\begin{cases} H_0 : \text{Le facteur } B \text{ n'a pas d'effet significatif sur } X \text{ ie : } \forall\, j \;\; \beta_j = 0. \\ H_1 : \text{Le facteur } B \text{ a un effet significatif sur } X \text{ ie : } \exists\;\, j \;\; \beta_j \neq 0. \end{cases}$

Sous l'hypothèse nulle, la statistique du test de Fischer est :

$$F = \frac{\frac{SCB}{q-1}}{\frac{SCR}{(p-1)(q-1)}} \sim F_{((q-1),(p-1)(q-1))}.$$

3.3.2 Tableau de variation de l'ANOVA2 sans répétition

Les résultats des tests sont généralement présentés dans la table suivantes :

Source de variation	**SC**	**ddl**	**MC**	**Statistique F**
Facteur A	SCA	$p-1$	$MCA = \frac{SCA}{p-1}$	$F_A = \frac{MCA}{MCR}$
FacteurB	SCB	$q-1$	$MCB = \frac{SCB}{q-1}$	$F_B = \frac{MCB}{MCR}$
Résiduelle	SCR	$(p-1)\,(q-1)$	$MCR = \frac{SCR}{(p-1)(q-1)}$	
Totale	SCT	$n-1$	$MCT = \frac{SCT}{n-1}$	

TAB. 3.4 – Tableau de l'ANOVA d'ordre 2 sans répétition

3.3.3 Règles de décision

Pour le test de l'effet de A sur X , on rejette H_0 au seuil α si la quantité de Fischer observée est supérieure à la valeur théorique (lue dans la table de la loi de Fischer-Snédécor) :

$$F_A > F_{\alpha,(p-1),(p-1)(q-1)}.$$

et Pour le test de l'effet de B sur X , on rejette H_0 au seuil α si la quantité de Fischer observée est supérieure à la valeur théorique[7] (lue dans la table de la loi de Fischer-Snédécor[20]) :

$$F_B > F_\alpha\left((q-1),(p-1)(q-1)\right).$$

Exemple 3.3.1 *La quantité d'oxygène consommé par deux espèces de patelle[3] : Acmaea Scabra et Acmaea Digitalis a été analysée pour différentes conditions halines :*

pourcentage d'eau ; les résultats suivants ont été obtenues :

% d'eau	**A.Scabra**					**A.Digitalis**				
100%	7,16	6,78	13,6	8,93	8,26	6,14	3,86	10,4	5,49	6,14
75%	5,2	5,2	7,18	6,37	13,2	4,47	9,9	5,75	11,8	4,95
50%	11,11	9,74	18,8	9,74	10,5	9,63	6,38	13,4	14,5	14,5

TAB. 3.5 – Table des moyennes

Question Tester l'hypothèse selon laquelle la quantité d'oxygène consommée par Acmaea Scabra et Acmaea Digitalis n'est pas affectée par les conditions salines au nivau sgnification $\alpha = 5\%$.

Réponse

Les résultats obtenus au cours de cette expériences au niveau : $\alpha = 5\%$ $p = 3$; $q = 2$; $r = 5$; $n = pqr = 30$ sont résumés dans les tables 3.6 et 3.7

	A.Scabra	A.Digitalis	$\overline{x}_{i..}$
100%	$\overline{x}_{11.} = 8,95$	$\overline{x}_{12.} = 6,41$	7,68
75%	$\overline{x}_{21.} = 7,43$	$\overline{x}_{22.} = 7,37$	7,4
50%	$\overline{x}_{31.} = 11,98$	$\overline{x}_{32.} = 11,68$	11,83
$\overline{x}_{.j.}$	9,45	8,49	$\overline{x} = 8,97$

TAB. 3.6 – Tabeau des moyennes de l'ANOVA2.

Le tabeau de variation de l'ANOVA 2 est donc

Variation	ddl	SC	MC	F
Fact. A	2	123,09	61,545	5,88
Fact. B	1	6,912	6,91	0,66
Fact. A,B	2	9,355	4,68	0,45
Résiduelle	24	251,185	10,47	
Totale	29	390,54		

TAB. 3.7 – Tabeau de variation de l'ANOVA 2.

$F_A = 5,88 > f_{0,95}(2;24) = 3,40 \Rightarrow H_1$ effet de A (% d'eau)

$F_B = 0,66 < f_{0,95}(1;24) = 4,26 \Rightarrow H_0$ absence d'effet de B (espèces de patelle)

$F_{AB} = 0,45 < f_{0,95}(2;24) = 3,40 \Rightarrow H_0$ absence d'effet d'interaction

Conclusion

Les points positifs de l'analyse de la variance (ANOVA) comprennent :

Efficacité dans la comparaison de plusieurs groupes : L'ANOVA permet de comparer les moyennes de trois groupes ou plus simultanément, ce qui est plus efficace que la comparaison paire à paire.

Gestion de la variabilité : L'ANOVA prend en compte à la fois la variabilité entre les groupes et la variabilité à l'intérieur des groupes, offrant ainsi une analyse plus complète des données.

Flexibilité :Il existe plusieurs variantes de l'ANOVA, telles que l'ANOVA à un facteur, à deux facteurs, et plus, qui peuvent être adaptées en fonction des besoins spécifiques de l'étude.

Interprétation intuitive : Les résultats de l'ANOVA sont généralement faciles à interpréter, fournissant des informations sur les différences significatives entre les groupes étudiés.

Robustesse : Dans de nombreux cas, l'ANOVA est robuste aux violations mineures des hypothèses sous-jacentes, telles que la normalité des données ou l'homoscédasticité.

En résumé, l'ANOVA est un outil statistique polyvalent et puissant pour comparer les moyennes de plusieurs groupes, offrant une analyse robuste et des conclusions interprétables.

Bibliographie

[1] AZOUZI, B.(2020). Polycopie de Biostatistique.

[2] Baccini, A., Besse, P., & Dejean, S. (2008). Analyse statistique de données d'expression.

[3] Benameur, S.Notes de Cours de Première Année Master.

[4] Besse, P. (2003). Pratique de la modélisation statistique.Publications du laboratoire de statistique et Probabilité.

[5] BOUDRAA, I., & ChEKRAOUI, L. (2021). Analyse de la variance et analyse de la covariance à deux facteurs (Doctoral dissertation, Université jijel).page 3-4.

[6] CANCEIL,M.(1971). L'ANALYSE DE VARIANCE(Programme "KR0AV").

[7] CHERFAOUI, M.(2020). Statistiques Appliquées à l'Expérimentation En Sciences Biologique,Université de Biskra,page 5 et 8.

[8] Chouquet, C.(2010).MODELES LINEAIRES,Laboratoire de Statistique et Probabilités - Université Paul Sabatier - Toulouse.

[9] DUSART,P(2012).Cours de Statistique ,Licence 2-S4 SI-MASS.

[10] Etienne Birmelé,Tests non paramétriques,M1IMSV.

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

[12] Godichon,A.Analyse de la Variance à 1 Facteur ,BaggioniINSA de Rouen – Génie Mathématique - 4ème année.

[13] Godichon,B.Analyse de la Variance à 2 Facteurs,BaggioniINSA de Rouen – Génie Mathématique - 4ème année.

[14] Houde,L.MODULE10 Tests d'hypothèses,Université du Québec à Trois-Rivières .

[15] MICHEL, C.(2015). STT- 2300,Cours d'Analyse de la Variance,Université de Laval, page 57-59.

[16] Morice, E. (1972). Tests de normalité d'une distribution observée. Revue de statistique appliquée, 20(2), 5-35.

[17] Ramousse, R., Le Berre, M., & Le Guelte, L. (1996). Introduction aux statistiques. Cons Dev.

[18] https ://www.techno-science.net/glossaire-definition/Analyse-de-la-variance-page-2.html.

[19] https ://mathsv.univ-lyon1.fr/app/cours/ ?theme=proba&chap=7

[20] http ://elearning.univ-biskra.dz/moodle2020/pluginfile.php/344988/mod_resource/content/

Annexe A : Logiciel R

• Dans ce chapitre, nous appliquerons les méthodes statistiques abordées dans les chapitres précédents en utilisant le programme R. Nous fournissons quelques exemples pour illustrer l'analyse unidirectionnelle de la variance et l'analyse unidirectionnelle de la variance.

Application d'analyse de la variance à un facteur

L'analyse de variance (ANOVA) est une méthode pour étudier le changement Moyenner le phénomène étudié X (variable quantitative) en fonction de l'influence d'un facteur, expérience qualitative Pour plusieurs facteurs, expérience spécifique (traitements).

Les données sont résumées dans la figure 3.1 .

```
Analysis of Variance Table

Response: d
          Df Sum Sq Mean Sq F value   Pr(>F)
a          2  185.2  92.600  9.4812 0.003389 **
Residuals 12  117.2   9.767
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

FIG. 3.1 – ANOVA 1 avec R.

Le tableau d'analyse de la variance renvoie le résultat du test de Fisher associé aux hypothèse :

$H_0 : \mu_1 = \mu_2 = ... = \mu_p$ et $H_1 : \exists i \neq i^{'} / \mu_i \neq \mu_{i'}$ (il existe au moins deux moyennes différentes). La valeur p = 0,003389

Vérification des conditions fondamentales d'ANOVA

Condition d'indépendance

Il n'existe pas, dans un contexte général, de test statistique simple permettant d'étudier l'indépendance. Ce sont les conditions de l'expérience qui vous permettront d'affirmer que cette condition est remplie.

Condition de normal

Nous souhaitons tester la normalité des variables d'erreur ε_{ij} à l'aide du test de **Shapiro-Wilk**. Nous utiliserons l'échantillon constitué des résidus pour effectuer ce test. Pour ce faire, en tapant les deux lignes de commande suivantes :

> residus<-residuals(model1)

> shapiro.test(residus)

```
        Shapiro-Wilk normality test

data:  residuals(ano)
W = 0.95055, p-value = 0.5332
```

FIG. 3.2 – Test de normalité.

Nous baserons notre conclusion sur la valeur p fournie par R. Des recherches ont démontré que lorsque la taille de l'échantillon est égale ou supérieure à 30, la puissance du test de Shapiro-Wilk est acceptable. Puisque la valeur p est strictement supérieure à 5%, nous concluons que l'hypothèse de normalité des variables d'erreur est confirmée.

Condition d'homogénéité des variances

Il existe plusieurs tests pour vérifier l'égalité de plusieurs variances. Parmi eux, le test de Bartlett est le plus couramment utilisé. Ce test requiert la normalité et l'indépendance des variables dont les variances sont comparées."

> bartlett.test(residus~forÍt,data=arbre)

```
        Bartlett test of homogeneity of variances

data:  d and a
Bartlett's K-squared = 1.5668, df = 2, p-value = 0.4569
```

FIG. 3.3 – Test d'homogénéité des variances.

La p-valeur est supérieur à 0,05, on peut conclure qu'il existe une homogénéité des variances.

application d'analyse de la variance à deux facteur

Les données résumées dans la figure suivante

```
> aov(d~factor(f1)*factor(f2))
Call:
   aov(formula = d ~ factor(f1) * factor(f2))

Terms:
                factor(f1) factor(f2) factor(f1):factor(f2) Residuals
Sum of Squares     6.96972  123.12659               9.38616 251.41852
Deg. of Freedom          1          2                     2        24

Residual standard error: 3.23663
Estimated effects may be unbalanced
> a=aov(d~factor(f1)*factor(f2))
> summary(a)
                      Df Sum Sq Mean Sq F value  Pr(>F)
factor(f1)             1   6.97    6.97   0.665 0.42271
factor(f2)             2 123.13   61.56   5.877 0.00837 **
factor(f1):factor(f2)  2   9.39    4.69   0.448 0.64414
Residuals             24 251.42   10.48
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ANOVA 2 avec R.

Annexe B : Abréviations et Notations

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

H_0	:	Hypothèse nulle
H_1	:	Hypothèse alternative
α	:	Risque de première espèce
β	:	Risque de deuxième espèce.
$\mathcal{P}$	:	La probabilité
X	:	Variable aléatoire
π	:	La fonction puissance
W	:	Région critique ou zone de rejet
$\overline{W}$	:	Région ou zone d'acceptation
$(x_1, x_2, ..., x_n)$	:	Echantillion de taille n de v.a X
ANOVA 1	:	Analyse de variance à un facteur
B_{obs}	:	Statistique de Shapiro-Wilk
$\overline{X}$	:	Moyenne empirique
σ^2	:	Variance mathématique d'une v-a
p	:	valeur représnte la probabilité

μ	:	Espérance mathématique d'une v-a
S_i^2	:	Variance empirique
χ_2	:	La statistique de khi-deux
A_i	:	le facteur
$N(0;1)$	:	La loi normale standard
$N(\mu;\sigma^2)$	:	Loi normale d'espérance μ et de variance σ^2
ε_{ij}	:	termes d'erreur aléatoire
x_{ij}	:	valeur de la variable de réponse j-ième essai modalité i du facteur
SCE_T	:	la variation totale théorique
SCE_F	:	la variation théorique due au facteur
SCE_R	:	la variation résiduelle
F	:	loi de Fisher non centrée
CMF_A	:	carré moyen factoriel
SCE_F	:	la variation théorique due au facteur
CM_T		carré moyen total
CM_R		carré moyen résiduel
SC_{AB}	:	somme des carrés des deux facteurs
ANOVA 2	:	Analyse de variance à deux facteur
ddl	:	degrée de liberté
α_i	:	effet du facteur A
β_j	:	effet du facteur B
γ_{ij}	:	effet d'intéraction
F	:	Loi de Fisher
T	:	Loi de Student

ملخص

تحليل التباين (ANOVA) هو عملية تهدف إلى دراسة تأثير عامل واحد أو أكثر على متغير كمي. يتمثل الهدف في مقارنة المتوسطات بين مجموعات متعددة (2 <). ويتمثل الهدف الفعلي في مقارنة تباين بين المجموعات (بين المجموعات المختلفة: الفارق بين متوسطات المجموعات والمتوسط الإجمالي) بتباين داخل المجموعة (مجموع التقلبات داخل كل مجموعة) . إذا لم يكن هناك فرق بين المجموعات، فإن هذين التباينين (تقريبًا) متساويين. وإلا فإن تباين بين المجموعات هو بالضرورة الأكبر. تعتمد شروط تطبيق الاختبار في الغالب على عدة فرضيات، بما في ذلك تجانس التباينات بين المجموعات والتوزيع الطبيعي للبيانات في كل مجموعة. يقدم تحليل التباين إحصائية الاختبار التي عادة ما تُسمىF ، التي تقارن التباين بين المجموعات بالتباين داخل المجموعات. إذا كانت هذه الإحصائية معنوية، فهذا يشير إلى أن واحدًا على الأقل من المجموعات يختلف بشكل معنوي عن الآخرين.

كلمات مفتاحية: تحليل التباين، اختبار التجانس، اختبار التوزيع الطبيعي، تحليل التباين ذو اتجاه واحد، تحليل التباين ذو اتجاهين.

Résumé

L'analyse de la variance (ANOVA) est un processus visant à étudier l'effet d'un ou de plusieurs facteurs sur une variable quantitative. Son objectif est de comparer les moyennes entre plusieurs groupes (> 2). L'objectif réel est de comparer la variance entre les groupes (différence entre les moyennes des groupes et la moyenne totale) à la variance à l'intérieur des groupes (somme des fluctuations au sein de chaque groupe). Si aucune différence n'est observée entre les groupes, ces deux variances sont (approximativement) égales. Sinon, la variance entre les groupes est nécessairement plus grande. Les conditions d'application du test sont généralement basées sur plusieurs hypothèses, y compris l'homogénéité des variances entre les groupes et la normalité de la distribution des données dans chaque groupe. L'ANOVA fournit une statistique de test, généralement appelée F, qui compare la variance entre les groupes à la variance à l'intérieur des groupes. Si cette statistique est significative, cela indique qu'au moins un des groupes diffère significativement des autres.

Mots clés : Analyse de variance, test d'homogénéité, test de normalité, analyse de variance à un facteur, analyse de variance à deux facteurs.

Abstract

ANOVA (Analysis of Variance) is a process aimed at studying the effect of one or more factors on a quantitative variable. Its objective is to compare means among multiple groups (> 2). The actual goal is to compare the variance between groups (difference between group means and overall mean) to the variance within groups (sum of fluctuations within each group). If no difference is observed between groups, these two variances are (approximately) equal. Otherwise, the between-group variance is necessarily larger. The conditions for applying the test are generally based on several assumptions, including homogeneity of variances between groups and normality of data distribution within each group. ANOVA provides a test statistic, typically called F, which compares the variance between groups to the variance within groups. If this statistic is significant, it indicates that at least one of the groups differs significantly from the others.

Keywords: Analysis of Variance, homogeneity test, normality test, one-way analysis of variance, two-way analysis of variance.