

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 Biologie



Module Biostatistique Master 1
Chapitre 01
Tests ANOVA à un Facteur :

Par
Prof : CHALA ADEL

2021-2022

Je dédie ce travail.
A mes parents ils m'ont tous,
avec leurs moyens, soutenu et donné
la force d'aller toujours
plus loin.
A ma chère femme Houda.
A l'esprit du professeur Bahlali Seid

Table des matières

Table des Matière	ii
Introduction	iii
1 Introduction aux théories des probabilités et statistique :	1
2 Test Analyse de la variance à un facteur	3
2.1 Structure des données	3
2.2 Conditions d'application de l'analyse	4
2.3 Modèle de l'analyse de variance	5
2.4 Equation fondamentale de l'analyse de variance	5
2.4.1 Pratique de l'analyse de variance	6
2.4.2 Application et tableau de variation	7
3 Exercice sur Analyse de la Variance à un Facteur	8

Introduction

En Master 1, les étudiants de Sciences de la Nature et de la Vie se voient proposer des formations spécialisées nécessitant des connaissances et un savoir-faire statistiques qui ne peuvent être acquis en tronc commun. D'autre part, certains étudiants, sans envisager a priori une spécialisation statistique, peuvent désirer acquérir une formation approfondie en méthodes statistiques.

Cette formation est particulièrement appréciée pour un débouché professionnel dans les domaines de l'expérimentation, et préparation de Master et post Doctoral.

Chaque méthode statistique est motivée par une présentation de problèmes concrets, par des utilisateurs dans différents domaines : agronomie, écologie, génétique, médecine, ...

Les connaissances acquises concernent l'estimation des paramètres, les tests statistiques (validité du modèle, effet des variables explicatives), la prévision et la sélection de variables dans le cadre du modèle linéaire (régression simple, analyse de la variance à plusieurs facteurs, analyse de la covariance).

Choix du modèle en fonction du type de données, structuration des données, traitement statistique et informatique (logiciel SPSS) pour les modèles de régression multiple et d'analyse de la variance ou de la covariance, ainsi que pour des extensions de ce modèle (ACP, AFC).

Ce module apporte une formation solide en statistique inférentielle directement exploitable dans de nombreux Masters Végétale ou Master Biologie Moléculaires, ou options de Poste Doctoral. Il donne des compétences indispensables pour la collecte et le traitement de données expérimentales. En cela, il constitue un pré requis important pour des formations en génétique, écologie et en sciences de l'environnement (Master Biologie)

Mots clés : Statistique inférentielle, modèle linéaire (régression, analyse de la variance et de la covariance), modèles mixtes (effet aléatoire), sélection de variables

Chapitre 1

Introduction aux théories des probabilités et statistique :

Lorsqu'on veut étudier les données relatives aux caractéristiques d'un ensemble d'individus ou d'objets il est difficile d'observer toutes les données lorsque leurs nombres sont élevés. Au lieu d'examiner l'ensemble qu'on appelle population on examine un nombre restreint qu'on appelle échantillon, pour être représentatif l'échantillon doit être pris au hasard (une population peut-être finie ou infinie).

Population : C'est l'ensemble sur lequel porte l'étude statistique.

Individus : Les éléments de cet ensemble.

Echantillon : Est un sous-ensemble de la population.

Caractère : C'est le trait (ou propriété) choisi pour l'étude statistique.

Modalités : Les différentes positions que peut prendre un caractère. Usage on numérote les modalités de 1 à k la modalité numéro i est notée C_i

Effectifs : Lorsque la population est répartie sur les différentes modalités nous obtenons pour chacune d'elles un nombre c'est le nombre des individus ayant cette modalité. On note habituellement n_i l'effectif correspondant à la modalité C_i : les fréquence absolu.

Fréquence relative : Par définition c'est le rapport entre n_i et N , où N est la somme totale des individus.

Nous allons ainsi adopter les définitions suivantes :

*Un caractère est dit quantitatif quand ses différentes modalités sont mesurables par des nombres qui en indiquent l'intensité.

*Un caractère est dit qualitatif quand ses différentes modalités ne peuvent

être désignées que par leurs qualités.

*Une variable statistique est dite discrète lorsque ses modalités ne peuvent être que des nombres isolés.

*Une variable statistique est dite continue quand elle peut prendre n'importe quelle valeur dans un intervalle donné.

* **Le mode** : c'est la valeur la plus fréquente.

* **La médiane** : C'est la valeur de la variable statistique qui partage la population en deux populations d'effectifs égaux.

* **Les quartiles** : Comme on a définie la médiane on peut définir des paramètres qui la répartissent en quarts.

* **La moyenne arithmétique** : est égale par définition $\bar{X}_n = \frac{1}{n} \sum_1^n X_i$.

b/ Caractéristiques de dispersion :

* **L'étendue** : C'est la longueur de l'intervalle sur lequel se disperse la variable.

* **L'écart-interquartiles** : C'est la différence entre les deux quartiles Q_1 et Q_3 .

* **La variance** : C'est la caractéristique qui est réellement utilisée pour mesurer la dispersion :

Chapitre 2

Test Analyse de la variance à un facteur

Dans le cadre des tests d'hypothèses, nous avons émis des hypothèses concernant la moyenne d'une population (test de conformité) puis comparé les moyennes de deux populations (test d'homogénéité). Ce chapitre a trait à la comparaison des moyennes de plusieurs populations (> 2). L'analyse de variance peut être vue comme une comparaison multiple de moyennes. Dans tous les cas, la variable étudiée est un caractère quantitatif de type continu qui suit une loi normale. Il existe différents types d'analyse de variance qui se distinguent par le nombre de facteurs étudiés (un facteur, deux facteurs, deux facteurs avec répétitions, etc), la nature du facteur (caractère qualitatif ou quantitatif) et la nature des modalités associées au facteur (modèle fixe, modèle aléatoire, modèle mixte).

Nous ne traiterons que le cas de l'analyse de variance à un facteur contrôlé (modèle fixe). Les différentes modalités du facteur correspondent aux différentes modalités d'un caractère qualitatif. Ces modalités sont déterminées par l'expérimentateur.

2.1 Structure des données

Les données relatives à une analyse de variance à un facteur contrôlé sont :

Modalité 1	Modalité 2	...	Modalité i.....	Modalité p
x_{11}	x_{12}	...	x_{1i}	x_{1p}
x_{21}	x_{22}	...	x_{2i}	x_{2p}
...
x_{n_11}	x_{n_22}	...	$x_{n_i i}$	$x_{n_p p}$
$\bar{x}_{.1}$	$\bar{x}_{.2}$...	$\bar{x}_{.i}$	$\bar{x}_{.p}$

Notation :

Le facteur contrôlé A présente p modalités ($1 \leq i \leq p$). On parle aussi de niveaux ou traitements. Le nombre de répétitions j pour une modalité i est notée n_i . Le nombre de répétitions pour chaque modalité du facteur n'est pas forcément le même. La valeur prise par la variable aléatoire X pour la modalité i du facteur et la répétition j est notée x_{ij} et la valeur moyenne pour chaque modalité notée \bar{x}_{ij} .

2.2 Conditions d'application de l'analyse

Indépendance

L'indépendance entre les différentes valeurs de la variable mesurée x_{ij} est une condition essentielle à la réalisation de l'analyse de variance.

Les p échantillons comparés sont indépendants.

Normalité

La variable quantitative étudiée suit une loi normale dans les p populations comparées. La variable aléatoire étudiée X dont x_{ij} est une représentation, suit une loi normale $\mathcal{N}(\mu_i, \sigma)$ sous \mathbf{H}_0 .

En d'autres termes les échantillons suivant une loi normale peuvent faire l'objet d'une analyse de variance paramétrique. Pour vérifier que la distribution d'un échantillon suit une loi normale, il est possible d'utiliser, dans XL Stat 2009 ou bien SPSS, le test descriptif d'aplatissement et d'asymétrie (de kurtosis and skewness, en anglais).

On considère que l'échantillon suit une loi normale à 95 % lorsque la valeur de son aplatissement est compris entre -2 et +2.

Homoscédasticité

Les p populations comparées ont même variance. Le facteur A agit seulement sur la moyenne de la variable X et ne change pas sa variance.

Différents tests permettent de vérifier l'égalité des variances relatives aux p populations. On calcule les différentes variances pour chacun des échantillons à comparer, et on fait le rapport de la plus grande sur la plus petite, ce rapport est F .

Cette valeur est comparée, dans une table de Hartley (ou du Fmax), à une valeur théorique et doit lui être inférieure pour un seuil de risque choisi (par exemple, 95 %) pour conserver l'hypothèse d'homogénéité des variances.

Les d.d.l. sont, pour la colonne de la table du Fmax, le nombre de traitements ($=k$) (colonnes de données), et pour la ligne de la table, le nombre de données du plus grand échantillon - 1 ($=n-1$).

2.3 Modèle de l'analyse de variance

Modèle sous H_0 : homogénéité des données

L'analyse de variance à un facteur teste l'effet d'un facteur contrôlé A ayant p modalités sur les moyennes d'une variable quantitative X . L'hypothèse nulle testée est la suivante :

"Il n'y a pas d'effet du facteur A sur les différentes modalités, et les p moyennes sont égales à une même moyenne ".

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \dots = \mu_i = \dots \mu_p.$$

Modèle sous H_1 : Hétérogénéité des données. L'hypothèse alternative est la suivante : "Il y a un effet du facteur A et il existe au moins deux moyennes significativement différentes."

$$H_1 : \exists i \neq j \quad \text{tel que } \mu_i \neq \mu_j.$$

Ainsi il existe une différence entre les moyennes de la variable selon les modalités du facteur contrôlé.

2.4 Equation fondamentale de l'analyse de variance

Notation :

SCEtotale = Somme des écarts totaux ou variation totale.

SCE_{inter} = Somme des écarts liés aux effets du facteur A ou variation inter (entre modalités).

SCE_{intra} = Somme des écarts résiduels ou variation intra (interne à chaque modalité).

2.4.1 Pratique de l'analyse de variance

Principe du test

Soit l'équation de décomposition de la variation totale

$$\sum_{j=1}^{j=p} \sum_{i=1}^{i=n_j} (x_{ij} - \bar{x}_{..})^2 = \sum_{j=1}^{j=p} n_j (\bar{x}_{.j} - \bar{x}_{..})^2 + \sum_{j=1}^{j=p} \sum_{i=1}^{i=n_j} (x_{ij} - \bar{x}_{.j})^2.$$

Alors

$$SCE_{totale} = \sum_{j=1}^{j=p} \sum_{i=1}^{i=n_j} (x_{ij} - \bar{x}_{..})^2 = \sum_{j=1}^{j=p} \sum_{i=1}^{i=n_j} x_{ij}^2 - N (\bar{x}_{..})^2.$$

$$SCE_{inter} = \sum_{j=1}^{j=p} n_j (\bar{x}_{.j} - \bar{x}_{..})^2 = \sum_{j=1}^{j=p} n_j \bar{x}_{.j}^2 - N \bar{x}_{..}^2.$$

$$SCE_{intra} = \sum_{j=1}^{j=p} \sum_{i=1}^{i=n_j} (x_{ij} - \bar{x}_{.j})^2 = \sum_{j=1}^{j=p} \sum_{i=1}^{i=n_j} x_{ij}^2 - \sum_{j=1}^{j=p} n_j \bar{x}_{.j}^2.$$

Alors il est facile de voir que

$$SCE_{totale} = SCE_{inter} + SCE_{intra}$$

L'estimation des variances associées ou carré moyen sont :

$$\text{Variance Totale } (CM_T) = \frac{SCE_T}{N-1}, \quad \text{avec } N = \sum_1^n n_j.$$

$$\text{Variance due au facteur } A \text{ } (CM_{inter}) = \frac{SCE_{inter}}{p-1}.$$

$$\text{Variance résiduelle } (CM_R) = (CM_{intra}) = \frac{SCE_{intra}}{N-p}.$$

Remarque : L'équation fondamentale de l'analyse de variance ne s'applique pas aux variances : Variance totale \neq Variance inter + Variance intra.

2.4.2 Application et tableau de variation

Le tableau de variation donne un résumé des calculs effectués pour l'analyse de variance.

Sources de variation	Degrés de liberté	Somme des Carrés des Ecart	Carré Moyen	Test de Fisher-Snédecor
Totale	$N - 1$	SCET		
Facteur	$p - 1$	SCEinter	$CMinter = \frac{SCEinter}{p - 1}$	$F_{obs} = \frac{CMinter}{CMintra}$
Résiduelle	$N - p$	SCEintra	$CMintra = \frac{SCEintra}{N - p}$	

Pour effectuer les calculs, des formules développées peuvent être utilisées.

$$CMinter = \frac{SCEinter}{p - 1} = \text{Variance des moyennes}$$

$$CMintra = \frac{SCEintra}{N - p} = \text{La moyenne des variances.}$$

Comparaison à la table

Une table pour chaque taux de risque donné en cours : 5% et 1%

Lire la valeur $F^{1-\alpha}(p - 1, n - p)$ de la ligne $(n - p)$ et la colonne $(p - 1)$ de la table choisie

*Si $F_{obs} > F^{1-\alpha}(p - 1, n - k)$ alors l'hypothèse nulle est rejetée, sinon elle est retenue.

Chapitre 3

Exercice sur Analyse de la Variance à un Facteur

Exercice N°: 01

Dans une clinique de réhabilitation on veut vérifier si la condition physique avant une intervention chirurgicale au genou a un effet sur le nombre de jours de physiothérapie pour conduire à une réhabilitation complète. La condition physique est évaluée selon un barème qui donne moyenne, sous la moyenne ou au dessous de la moyenne. Voici les données en jours de traitement pour obtenir une réhabilitation complète

Condition										
Inférieure	29	42	38	40	43	40	30	42		
Moyenne	30	35	39	28	31	31	29	35	29	33
Supérieure	26	32	21	20	23	22				

Donner l'évaluation du nombre moyen de jours de réhabilitation par groupe. Peut-on dire au niveau 5 % que la condition physique influence la temps de réhabilitation ? La conclusion resterait-elle la même pour un niveau de 1 %.

Exercice N°: 02

Un test psychologique a été passé par 30 sportifs évoluant à des niveaux de compétition différents : international, national, régional et « récréational ». Une des mesures réalisées porte sur l'anxiété des sportifs au moment de la compéti-

tion.

International	National	Régional	"Récréational"
24,8		33,4	31,1
26,7		34,6	35,7
27,5		36,4	37,3
30,6	45,6	39,1	39,4
32,4	41,1	43,8	40,4
38,2	34,3	47,9	44,5
40,5	37,6	49,9	45,4
42,9	39,5	51,2	49,8
			50,1

Celle-ci diffère-t-elle en fonction du niveau de compétition ?

Exercice N°: 03

Dans une école, un pédagogue veut évaluer 3 méthodes d'enseignement. Pour ce faire il divise un groupe de 24 sujets en 3 groupes de tailles égales. Chacun de ces groupes adopte une méthode différente d'apprentissage et un même test (sur 10 points) permet d'évaluer les connaissances de chaque sujet à la fin de la période. Les résultats sont les suivants :

	Test							
Méthode 1	3	5	2	4	8	4	3	9
Méthode 2	4	4	3	8	7	4	2	5
Méthode 3	6	7	8	6	7	9	10	9

On cherche à déterminer si les méthodes d'apprentissage donnent des résultats identiques au niveau de la note moyenne au test.

Exercice N°: 04

Afin d'évaluer la précocité de l'augmentation d'activité enzymatique lors de la grossesse, on pratique des dosages chez des femmes enceintes à différentes semaines d'aménorrhée. On obtient les résultats suivants (sur des échantillons

indépendants) :

4 Semaine	5 Semaine	6 Semaine	7 Semaine	8 Semaine
7,2	4,9	10,4	4,6	6,1
4,3	4,8	4,6	5,6	11,4
5,5	4,7	8,4	8,3	8,2
4,6	5,4	6,1	6,9	5,7
4,7	4,7	8,1	4,5	6,6
5,5	4,7	5,4	4,7	6,6
6,6	6,2	6,7	6,7	6,3
5,3	5,6	7,5	4,8	5,9
5,4	3,2	6,4	5,0	5,8
3,9	6,1	5,6	5,0	4,8
5,5	6,7	6,3	5,3	9,1
2,7	5,5	7,7	7,8	13,2

L'âge de la grossesse a-t-il une influence sur l'activité de l'enzyme ?

Exercice N°: 05

Pour définir l'impact de la nature du sol sur la croissance d'une plante X , un botaniste a mesuré la hauteur des plantes pour 4 types de sol. Pour chaque type de sol, il disposait de 3 répliques.

Type I	Type II	Type III	Type IV
15	25	17	10
9	21	23	13
4	19	20	19

Que peut-on conclure sur cette expérience ?

Exercice N°: 06

Données Bransfor

On reprend une expérience de Bransford et *al.* (1972), dans laquelle on demande à des sujets d'écouter le texte suivant :

"Si les ballons éclatent, le son ne portera pas puisque tout sera bien trop loin du bon étage. Une fenêtre fermée empêchera également le son de porter, surtout depuis que les immeubles récents sont correctement isolés. Comme l'essentiel de l'opération dépend d'une arrivée correcte d'électricité, un l cassé causerait bien

des problèmes. Evidemment, le type peut hurler. Mais la voix humaine n'est pas assez puissante pour porter bien loin. Un problème supplémentaire serait qu'une corde casse sur l'instrument. Alors il serait impossible d'accompagner le message. C'est clair que la meilleure situation impliquerait la plus petite distance. Alors, il y aurait bien moins de problèmes potentiels. Avec un contact en face à face, un bien petit nombre de choses pourrait gêner"

Le but visé par Bransford et al. est de montrer l'importance du contexte dans la compréhension et la mémorisation d'un texte. Pour ce faire, ils utilisent quatre groupes expérimentaux :

1. Un groupe "sans contexte" entend simplement le texte.
2. Le groupe "avec contexte avant" regarde une guère suggérant un contexte approprié pendant qu'il entend le texte.
3. Le groupe "avec contexte après" entend le texte puis regarde la guère précédente.
4. Le groupe "avec contexte partiel" regarde une guère suggérant un contexte inapproprié pendant qu'il entend le texte.

Groupe 1	Groupe 2	Groupe 3	Groupe 4
3	5	2	5
3	9	4	4
2	8	5	3
4	4	4	5
3	9	1	4

Faire l'étude d'ANOVA convenable pour ce type problème.

Exercice N°: 07

Une entreprise alimentaire veut mettre sur la marché une nouvelle gamme de pâtes alimentaires. Les consultants en marketing ont proposé 4 emballages et pour déterminer le choix final, 10 marchés d'alimentation de même taille ont été sélectionnés pour faire une prévente du produit. Les emballages ont été assignés aléatoirement entre les marchés et voici les résultats sur les ventes :

Emballage		Ventes	
1	12	18	
2	14	12	13
3	19	17	21
4	24	30	

Répondre au question posée.

La correction :

Exercice N°: 01**Réponse :**

Le facteur étudié c'est condition physique avec $p = 3$, et $n_1 = 8$, $n_2 = 10$, $n_3 = 6$. . $N = \sum n_i = 24$.

L'hypothèse nulle \mathbf{H}_0 : " Il n'y'a pas une influence du facteur A sur les nombres des jours".

$$\bar{X}_1 = 38.$$

$$\bar{X}_2 = 32$$

$$\bar{X}_3 = 24$$

$$\bar{X}_{..} = 32 = \frac{1}{24} (8\bar{X}_1 + 10\bar{X}_2 + 6\bar{X}_3)$$

$$\sum \sum x_{ij}^2 = 25664$$

$$\sum n_j (\bar{x}_j)^2 = 25248$$

Alors on peut calculer les SCE :

$$SCE_{inter} = \sum n_j (\bar{x}_j)^2 - N (\bar{X}_{..})^2 = 25248 - 24 (32)^2 = 672 > 0.$$

$$SCE_{intra} = \sum \sum x_{ij}^2 - \sum n_j (\bar{x}_j)^2 = 25664 - 25248 = 416 > 0.$$

Tableaux des variations

Sources de variation	Degrés de liberté	Somme des Carrés des Ecarts	Carré Moyen	Test de Fisher-Snédecor
Totale	$N - 1 = 23$	///		
Facteur	$p - 1 = 2$	672	$CM_{inter} = 336$	$F_{obs} = 16,961$
Résiduelle	$N - p = 21$	416	$CM_{intra} = 19,810$	

$$F^{0,95}(p - 1, n - p) = F^{0,95}(2, 21) = 3,44$$

Il est clair que $F_{obs} > F^{0,95}(p - 1, n - p)$. Donc on rejette \mathbf{H}_0 . il y'a un effet du facteur A sur les nombres des jours.

Exercice N°: 02**Réponse :**

Le facteur étudié c'est le niveaux des compétitions avec $p = 4$, et $n_1 = 8$, $n_2 = 5$, $n_3 = 8$, et $n_4 = 9$. $N = \sum n_i = 30$.

L'hypothèse nulle \mathbf{H}_0 : " Il n'y'a pas une influence du facteur A sur les résultats".

$$\bar{X}_1 = 32,950.$$

$$\bar{X}_2 = 39,620$$

$$\bar{X}_3 = 42,038$$

$$\bar{X}_4 = 41,522$$

$$\bar{X}_{..} = 39,057 = \frac{1}{30} (8\bar{X}_1 + 5\bar{X}_2 + 8\bar{X}_3 + 9\bar{X}_4)$$

$$\sum \sum x_{ij}^2 = 47269,030$$

$$\sum n_j (\bar{x}_j)^2 = 46188,578$$

Alors on peut calculer les SCE :

$$SCE_{inter} = \sum n_j (\bar{x}_j)^2 - N (\bar{X}_{..})^2 = 46188,578 - 30 (39,057)^2 = 425,108 > 0.$$

$$SCE_{intra} = \sum \sum x_{ij}^2 - \sum n_j (\bar{x}_j)^2 = 47269,030 - 46188,578 = 1080,425 > 0.$$

Tableaux des variations

Sources de variation	Degrés de liberté	Somme des Carrés des Ecartés	Carré Moyen	Test de Fisher-Snédecor
Totale	$N - 1 = 30 - 1$	///		
Facteur	$p - 1 = 4 - 1$	425,108	$CM_{inter} = 141,700$	$F_{obs} = 3,410$
Résiduelle	$N - p = 30 - 4$	1080,425	$CM_{intra} = 41,556$	

$$F^{0,95}(p-1, n-p) = F^{0,95}(3, 26) = 2,98$$

Il est clair que $F_{obs} > F^{0,95}(p-1, n-p)$. Donc on rejette H_0 . il y'a un effet du facteur A sur l'anxiété des sportifs au moment des compétitions.

Exercice N°: 03

Réponse :

On veut donc confronter les hypothèses

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

$$H_1 : \mu_i \neq \mu_j$$

Au niveau $\alpha = 5\%$. Le facteur étudié c'est le Méthodes d'enseignement avec $p = 3$, et $n_1 = 8$. $n_2 = 8$. $n_3 = 8$. $N = \sum n_i = 24$.

L'hypothèse nulle H_0 : " Il n'y'a pas une influence du facteur A sur les résultats".

$$\bar{X}_1 = 4,758.$$

$$\bar{X}_2 = 4,625$$

$$\bar{X}_3 = 7,750$$

$$\bar{X}_{..} = 5,708 = \frac{1}{24} (8\bar{X}_1 + 8\bar{X}_2 + 8\bar{X}_3) = \frac{1}{3} (\bar{X}_1 + \bar{X}_2 + \bar{X}_3)$$

$$\sum \sum x_{ij}^2 = 919$$

$$\sum n_j (\bar{x}_j)^2 = 8 \sum (\bar{x}_j)^2 = 832,125$$

Alors on peut calculer les SCE :

$$SCE_{inter} = \sum n_j (\bar{x}_j)^2 - N (\bar{X}_{..})^2 = 832,125 - 30 (5,708)^2 = 50,175 > 0.$$

$$SCE_{intra} = \sum \sum x_{ij}^2 - \sum n_j (\bar{x}_j)^2 = 919 - 832,125 = 86,875 > 0.$$

Tableaux des variations

Sources de variation	Degrés de liberté	Somme des Carrés des Ecartés	Carré Moyen	Test de Fisher-Snédecors
Totale	$N - 1 = 24 - 1$	///		
Facteur	$p - 1 = 3 - 1$	50,175	$CM_{inter} = 25,088$	$F_{obs} = 6,064$
Résiduelle	$N - p = 24 - 4$	86,875	$CM_{intra} = 4,137$	

$$F^{0,95}(p - 1, N - p) = F^{0,95}(2, 21) = 3,44$$

Il est clair que $F_{obs} > F^{0,95}(p - 1, n - p)$. Donc on rejette H_0 . il y'a un effet du facteur A sur l'évaluation.

Comparaison à la table

$$- \text{Table à } 5\% : F^{0,95}(2,21) = 3,44$$

L'hypothèse nulle est rejetée, donc au moins une des moyennes est différente des autres :

Exercice N°: 04

Réponse :

Le facteur étudié c'est L'âge des grossesses avec $p = 5$, et $n_1 = n_2 = n_3 = n_4 = n_5 = 12$, $N = \sum n_i = 60$.

L'hypothèse nulle H_0 : " Il n'y'a pas une influence du facteur A sur l'activités enzymatiques".

$$\bar{X}_1 = 5,100.$$

$$\bar{X}_2 = 5,208$$

$$\bar{X}_3 = 6,933$$

$$\bar{X}_4 = 5,767$$

$$\bar{X}_5 = 7,475$$

$$\bar{X}_{..} = 6,097 = \frac{1}{60} 12 (\bar{X}_1 + \bar{X}_2 + \bar{X}_3 + \bar{X}_4 + \bar{X}_5)$$

$$\sum \sum x_{ij}^2 = 2427,280$$

$$\sum n_j (\bar{x}_j)^2 = 2284,004$$

Alors on peut calculer les SCE :

$$SCE_{inter} = \sum n_j (\bar{x}_j)^2 - N (\bar{X}_{..})^2 = 2284,004 - 60 (6,097)^2 = 53,599 > 0.$$

$$SCE_{intra} = \sum \sum x_{ij}^2 - \sum n_j (\bar{x}_j)^2 = 2427,280 - 2284,004 = 143,276 > 0.$$

Tableaux des variations

Sources de variation	Degrés de liberté	Somme des Carrés des Ecarts	Carré Moyen	Test de Fisher-Snédecor
Totale	$N - 1 = 59$	///		
Facteur	$p - 1 = 4$	53,599	$CM_{inter} = 13,400$	$F_{obs} = 5,144$
Résiduelle	$N - p = 55$	143,276	$CM_{intra} = 2,605$	

$$F^{0,95}(p-1, n-p) = F^{0,95}(4, 55) = 2,53$$

Il est clair que $F_{obs} > F^{0,95}(p-1, n-p)$. Donc on rejette H_0 . il y'a un effet du facteur A sur l'activités enzymatique.

Exercice N°: 05

Réponse :

Le facteur étudié c'est La nature des sols avec $p = 4$, et $n_1 = n_2 = n_3 = n_4 = 3$, $N = \sum n_i = 12$.

L'hypothèse nulle H_0 : " Il n'y'a pas une influence du facteur A sur l'hauteur des plantes".

$$\bar{X}_1 = 9,333.$$

$$\bar{X}_2 = 20$$

$$\bar{X}_3 = 21,667$$

$$\bar{X}_4 = 14$$

$$\bar{X}_{..} = 16,258$$

$$\sum \sum x_{ij}^2 = 3597$$

$$\sum n_j (\bar{x}_j)^2 = 3457,691$$

Alors on peut calculer les SCE :

$$SCE_{inter} = \sum n_j (\bar{x}_j)^2 - N (\bar{X}_{..})^2 = 3457,691 - 12 (16,258)^2 = 288,941 > 0.$$

$$SCE_{intra} = \sum \sum x_{ij}^2 - \sum n_j (\bar{x}_j)^2 = 3597 - 3457,691 = 139,309 > 0.$$

Tableaux des variations

Sources de variation	Degrés de liberté	Somme des Carrés des Ecartés	Carré Moyen	Test de Fisher-Snédecor
Totale	$N - 1 = 11$	///		
Facteur	$p - 1 = 3$	288, 941	$CMinter = 96, 314$	$F_{obs} = 5, 531$
Résiduelle	$N - p = 8$	139, 309	$CMintra = 17, 414$	

$$F^{0,95}(p - 1, n - p) = F^{0,95}(3, 8) = 4, 07$$

Il est clair que $F_{obs} > F^{0,95}(p - 1, n - p)$. Donc on rejette H_0 . il y'a un effet du facteur A sur l'hauteur des plantes.

Exercice N°: 06**Réponse :**

Le facteur étudié c'est méthode d'apprentissage des contextes avec $p = 4$, et $n_1 = n_2 = n_3 = n_4 = 5$, $N = \sum n_i = 20$.

L'hypothèse nulle H_0 : " Il n'y'a pas une influence du facteur A sur la compréhension et la mémorisations des textes ".

$$\bar{X}_1 = 3.$$

$$\bar{X}_2 = 3, 2$$

$$\bar{X}_3 = 7$$

$$\bar{X}_4 = 4, 2$$

$$\bar{X}_{..} = 4, 350$$

$$\sum \sum x_{ij}^2 = 467$$

$$\sum n_j (\bar{x}_j)^2 = 429, 400$$

Alors on peut calculer les SCE :

$$SCE_{inter} = \sum n_j (\bar{x}_j)^2 - N (\bar{X}_{..})^2 = 429, 400 - 20 (4, 350)^2 = 50, 950 > 0.$$

$$SCE_{intra} = \sum \sum x_{ij}^2 - \sum n_j (\bar{x}_j)^2 = 467 - 429, 400 = 37, 600 > 0.$$

Tableaux des variations

Sources de variation	Degrés de liberté	Somme des Carrés des Ecartés	Carré Moyen	Test de Fisher-Snédecor
Totale	$N - 1 = 19$	///		
Facteur	$p - 1 = 3$	50, 950	$CMinter = 16, 983$	$F_{obs} = 7, 227$
Résiduelle	$N - p = 16$	37, 600	$CMintra = 2, 350$	

$$F^{0,95}(p - 1, n - p) = F^{0,95}(3, 16) = 3, 24$$

Il est clair que $F_{obs} > F^{0,95}(p - 1, n - p)$. Donc on rejette H_0 . il y'a un effet du facteur méthode d'apprentissage des contextes sur la compréhension et la mémorisations des textes .

Bibliographie

- [1] G. Saporta ; Probabilités, analyse des données et statistique. Technip, Paris, 1990.
- [2] B. Escofier, J. Pagés ; Analyse factorielles simple et multiple. Dunod, Paris, 1998.
- [3] Avner Bar-Hen : Cours de DEUG Probabilités et Statistiques, Université Aix-Marseille III, 2002–2003.
- [4] Admane, O., Hoang Ky., Ouakli N : Statistique (cours et exercices) Pour les étudiants du tronc commun bio-médical., OPU 1998.
- [5] M Vilain : Méthodes expérimentales en agronomie- Pratique et analyse. Edition Tec et Doc. 1999.