

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 02
Tests ANOVA à deux Facteurs :

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 Analyse de la variance à deux facteurs	1
1.1 Echantillons de plusieurs observations (de taille n)	1
1.2 Echantillons d'une seule observation	4
2 Exercice sur Analyse de la Variance à deux Facteurs	6
3 La correction	12

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

Analyse de la variance à deux facteurs

1.1 Echantillons de plusieurs observations (de taille n)

Ce plan est appliqué lorsqu'on dispose de plusieurs mesures pour chaque combinaison des niveaux des deux facteurs. Le bénéfice est qu'on peut tester l'interaction entre les deux facteurs. L'interaction dans le sens de l'ANOVA, elle, mesure l'influence que l'état d'un des facteurs a sur l'effet exercé par l'autre facteur sur la variable dépendante. En d'autres mots, il y a interaction lorsque l'effet des différents niveaux d'un facteur change selon le niveau de l'autre facteur.

Structure des données

Les données relatives à une analyse de variance à deux facteurs contrôlé avec

répétitions de même nombre, sont structurées dans un tableau du type suivant :

	Modalité 1	Modalité 2	Modalité 3	Moy Mar
Modalité 1	x_{111}, \dots, x_{11n}	x_{121}, \dots, x_{12n}	x_{131}, \dots, x_{13n}	
	\bar{x}_{11}	\bar{x}_{12}	\bar{x}_{13}	$\bar{x}_{1.}$
Modalité 2	x_{211}, \dots, x_{21n}	x_{221}, \dots, x_{22n}	x_{231}, \dots, x_{23n}	
	\bar{x}_{21}	\bar{x}_{21}	\bar{x}_{21}	$\bar{x}_{2.}$
Modalité 3	x_{311}, \dots, x_{31n}	x_{311}, \dots, x_{32n}	x_{311}, \dots, x_{33n}	
	\bar{x}_{31}	\bar{x}_{31}	\bar{x}_{31}	$\bar{x}_{3.}$
Modalité p	x_{p11}, \dots, x_{p1n}	x_{p11}, \dots, x_{p2n}	x_{p11}, \dots, x_{p3n}	
	\bar{x}_{p1}	\bar{x}_{p1}	\bar{x}_{p1}	$\bar{x}_{p.}$
Moyenne Marginal	$\bar{x}_{.1}$	$\bar{x}_{.2}$	$\bar{x}_{.3}$	$\bar{x}_{...}$

La décomposition des SCE

Soit l'équation de décomposition de la variation totale :

$$SCE_T = SCE_A + SCE_B + SCE_{AB} + SCE_R.$$

Calcul des SCE

$$\begin{aligned}
 SCE_T &= \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n (x_{ijk} - \bar{x}_{...})^2 = \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n x_{ijk}^2 - npq\bar{x}_{...}^2 \\
 SCE_R &= \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n (x_{ijk} - \bar{x}_{ij})^2 = \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n x_{ijk}^2 - n \sum_{i=1}^p \sum_{j=1}^q \bar{x}_{ij}^2 \\
 SCE_A &= qn \sum_{i=1}^p (\bar{x}_{i.} - \bar{x}_{...})^2 = qn \sum_{i=1}^p \bar{x}_{i.}^2 - npq\bar{x}_{...}^2 \\
 SCE_B &= pn \sum_{j=1}^q (\bar{x}_{.j} - \bar{x}_{...})^2 = pn \sum_{j=1}^q \bar{x}_{.j}^2 - npq\bar{x}_{...}^2 \\
 SCE_{AB} &= n \sum_{i=1}^p \sum_{j=1}^q (\bar{x}_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{...})^2 = SCE_T - SCE_R - SCE_A - SCE_B
 \end{aligned}$$

Hypothèses

Il y a maintenant trois hypothèses principales à tester :

H_0 : le facteur A n'a aucun effet sur les résultats, c'est-à-dire

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \dots = \mu_i \quad \text{avec} \quad i = \overline{1; p}$$

H'_0 : le facteur B n'a aucun effet sur les résultats, c'est-à-dire

$$H'_0 : \mu_1 = \mu_2 = \mu_3 = \dots = \mu_j \quad \text{avec} \quad j = \overline{1; q}$$

H''_0 : il n'y a pas d'interaction entre les facteurs A et B.

Hypothèses contraires :

H_1 : le facteur A a un effet sur les résultats, c'est-à-dire qu'au moins une des moyennes $\mu_{i..}$ n'est pas égale aux autres ;

H'_1 : le facteur B a un effet sur les résultats, c'est-à-dire qu'au moins une des moyennes $\mu_{.j}$ n'est pas égale aux autres.

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, et réciproquement.

Scherrer montre comment décomposer la variance totale des résultats en :

- variance "cellulaire" (c'est-à-dire la variance due à la variabilité au sein des "cellules" du tableau contenant chacune les t répétitions d'une combinaison donnée de niveaux des facteurs A et B) $(\frac{SCEC}{(pq-1)})$;
- variance due au facteur A $(\frac{SCEA}{(p-1)})$;
- variance due au facteur B $(\frac{SCEB}{(q-1)})$;
- variance due à l'interaction AB $(\frac{SCEAB}{[(p-1)(q-1)])}$;
- variance résiduelle $(\frac{SCER}{[pq(n-1)])}$.

Lorsque les populations sont normales, les échantillons indépendants et les hypothèses principales vraies, les variances dues au facteur A, au facteur B, à l'interaction et résiduelle sont quatre estimations de la variance s^2 de la même population.

Par conséquent, chaque hypothèse H_0 peut être testée à l'aide d'un test de Fisher comparant les variances dues à chacun des facteurs à la

variance des résidus :

$$F_A^{obs} = \frac{CM_A}{CM_R}$$

$$F_B^{obs} = \frac{CM_B}{CM_R}$$

$$F_{AB}^{obs} = \frac{CM_{AB}}{CM_R}$$

Lorsque H_0 est vraie et que les conditions d'application du test sont remplies, les trois variables auxiliaires F_A^α , F_B^α et F_{AB}^α sont distribuées respectivement comme les F de Fisher suivants :

$$F_{(p-1),[pq(n-1)]}^{\alpha}, \quad F_{(q-1),[pq(n-1)]}^{\alpha} \quad \text{et} \quad F_{(p-1)(q-1),[pq(n-1)]}^{\alpha}$$

Remarque :

1) Il faut remarquer ici que, contrairement au cas de l'ANOVA à un critère de classification, cette procédure n'est valide que dans le cas d'une ANOVA de modèle I, c'est-à-dire lorsque les deux facteurs sont contrôlés. Dans le cas d'ANOVA de modèle II (deux facteurs aléatoires) ou III (mixte : un facteur contrôlé et un aléatoire), les statistiques F et leurs degrés de liberté peuvent être différents.

2) Il est impératif de tester en premier l'effet de l'interaction, les tests des effets principaux de A et B ne pouvant être interprétés que si cette interaction est non significative (H_0'').

1.2 Echantillons d'une seule observation

Tableau des données

Les données relatives à une analyse de variance à deux facteurs contrôlés sans répétitions, sont structurées dans un tableau du type suivant :

	Modalité 1	Modalité 2	Modalité 3	Moy Mar
Modalité 1	x_{11}	x_{12}	x_{13}	$\bar{x}_{1.}$
Modalité 2	x_{21}	x_{21}	x_{21}	$\bar{x}_{2.}$
Modalité 3	x_{31}	x_{31}	x_{31}	$\bar{x}_{3.}$
Modalité p	x_{p1}	x_{p1}	x_{p1}	$\bar{x}_{p.}$
Moyenne Marginal	$\bar{x}_{.1}$	$\bar{x}_{.2}$	$\bar{x}_{.3}$	$\bar{x}_{...}$

La décomposition des SCE

Soit l'équation de décomposition de la variation totale :

$$SCET = SCEA + SCEB + SCEAB.$$

Calcul des SCE

Mêmes formules en remplaçant n par 1.

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

Sources des Variations	Degrés de Liberté	Sommes des carrées des écarts	Carrée Moyen	Test de Fisher-Snédecor
Totale	$N - 1$	SCET	$(CM_T) = \frac{SCE_T}{N-1}$	
Facteur A	$p - 1$	SCE A	$(CM_A) = \frac{SCE_{Ar}}{p-1}$	$F_{obs}^A = \frac{CM_A}{CM_{AB}}$
Facteur B	$q - 1$	SCE B	$(CM_B) = \frac{SCE_{Br}}{q-1}$	$F_{obs}^B = \frac{CM_B}{CM_{AB}}$
Facteur AB	$(p - 1)(q - 1)$	SCE Intra	$(CM_{AB}) = \frac{SCE_{AB}}{(p-1)(q-1)}$	

Chapitre 2

Exercice sur Analyse de la Variance à deux Facteurs

Exercice N°: 01

La qualité d'oxygène consommée par deux espèces de patelle, *A cmaca scabra* et *A digitalis* a été analysé pour différentes conditions halines. Les résultats sont les suivants :

	A cmaca scabra	A digitalis
100 % Eau de mer	7,16	6,14
	6,78	3,86
	13,6	10,4
	8,93	5,49
	8,26	6,14
75 % Eau de mer	5,2	4,47
	5,2	9,9
	7,17	5,75
	6,37	11,8
	13,20	4,95
50 % Eau de mer	11,11	9,63
	9,74	6,38
	18,8	13,4
	9,74	14,5
	10,5	14,5

1) Quelle sont les facteurs étudiés ?.

- 2) Quelle est l'objectif de notre travail ?.
- 3) Etablir les hypothèses nulles pour les deux facteurs.
- 4) Calculer les moyennes marginales pour les deux facteurs, ainsi que la moyenne totale ?.
- 5) Décrire la table de décomposition de l'ANOVA, mais avec un seuil de signification de 5 %.
- 6) Etudier l'effet de l'interaction des facteurs étudiés ?.
- 7) Est-ce que les deux espèces de patelle sont coïncident ou bien non ? Et les trois types d'eaux ont-ils le même effet.
- 8) conclusion.

Exercice N°: 02

Dans le cadre d'une expérience sur la durée de vie des piles, le but est de déterminer la durée de vie en fonction du type de pile. Comme on sait que les piles ont une durée de vie qui dépend de la température d'utilisation, un plan à deux facteurs (type de pile et température d'utilisation). Les résultats sont dans le tableau suivant

	15°C	70°C	125°C
Type I	130// 155// 74// 180,	34// 40// 80// 75,	20// 70// 82// 58.
Type II	150// 188// 159// 126,	136// 122// 106// 115,	25// 70// 58// 45.
Type III	138// 110// 168// 160,	174// 120// 150// 139,	96// 104// 82// 60.

On doit conclure qu'il y a interaction entre le type de pile et la température avec un niveau de 5% et de plus les facteurs ont une influence sur la durée de vie

Exemple 5 :

Dans une étude sur la plombémie, 240 personnes habitant 3 villes différentes ont subi un prélèvement sanguin. Les prélèvements ont été soumis à 4 laboratoires qui ont mesuré la plombémie X de chaque personne. Le tableau 1 donne les moyennes (en μ / dL) de la plombémie mesurée par les laboratoires selon les villes. L'échantillon de prélèvement a été fait de sorte que chaque «case» du tableau contienne le même nombre de sujets (c'est-à-dire 20).

	Ville 1	Ville 2	Ville 3
Labo 1	17.97	13.135	10.60
Labo 2	15.30	11.49	12.09
Labo 3	14.69	11.94	9.41
Labo 4	16.10	12.11	11.69

Observe-t-on une différence entre les villes ? entre les laboratoires ?

Exercice N°: 03

On évalue l'efficacité d'un nouveau traitement ayant pour objet d'améliorer le développement global des enfants atteints de trisomie 21. Pour cela, une étude a été menée auprès de 12 enfants. Six d'entre eux ont reçu un produit actif alors que 6 autres ont reçu un placebo, et ce pendant 6 mois. Un indice de développement global de chaque enfant est calculé avant et après le début de l'étude par un même psychologue. Cet indice de développement global résume l'ensemble des capacités en terme de coordination, posture, langage et sociabilité. La nature de traitement donné n'est connue ni de la famille de patient, ni du psychologue. Deux psychologues ont participé à l'étude. Les données ont été fournies par L'institut J. Lejeune

Enfant	Traitement	Psychologue	Augmentation de l'indice de développement global
1	Produit actif	1	93,7
2	Placebo	2	36,8
3	Placebo	1	56,6
4	Placebo	1	37,7
5	Produit actif	2	70,3
6	Placebo	1	24,2
7	Produit actif	2	42,5
8	Placebo	2	50,9
9	Produit actif	1	71,8
10	Produit actif	2	72,5
11	Produit actif	1	49,3
12	Placebo	2	37,3

Propose un modèle permettant de mettre en évidence un éventuel effet traitement et Psychologue.

Exercice N°: 04

I) Dans une réplique d'une expérience de Smith (1979), l'auteur désire mettre en évidence l'effet facilitateur du contexte d'acquisition sur le rappel.

Dans un premier temps, les sujets apprennent une liste de quatre-vingts mots dans une pièce orange décorée avec des draperies, posters, tableaux, ... etc... Les sujets subissent quelques minutes après leur apprentissage un test de reconnaissance dont le but est simplement de leur laisser croire que l'expérience s'arrête

là. Le lendemain, on demande aux sujets de noter les mots dont ils peuvent se souvenir. Le rappel se déroule dans différentes conditions qui définissent cinq groupes expérimentaux (les sujets étant affectés aléatoirement à chaque groupe).

(1) Même contexte : *les sujets se trouvent dans la même pièce orange où ils ont appris la liste la veille.*

(2) Contexte différent : *les sujets se trouvent dans une pièce nettement différente de celle où ils ont appris la veille.*

(3) Contexte imaginé : *les sujets sont dans la pièce du groupe (2). Mais on leur conseille de se rappeler d'abord la pièce orange (on les aide en leur posant des questions).*

(4) Contexte photographié : *les sujets se trouvent dans la condition (3), mais on leur montre plusieurs photographies de la pièce orange.*

(5) Contexte placebo : *les sujets sont dans la même pièce que le groupe (2). Mais on leur demande de se rappeler d'abord leur chambre à coucher à titre « d'échauffement intellectuel ».*

Les observations figurent dans le tableau ci-dessous

Groupe I Même contexte	Groupe II Contexte différent	Groupe III Contexte imaginé	Groupe IV Contexte photographié	Groupe V Placebo
16	12	14	17	12
20	22	15	21	19
19	10	20	27	10
22	7	10	10	7
25	8	12	15	5
13	15	21	11	12
14	12	16	19	15

1) Quelle est la variable étudiée ?, 2) Quelles sont les modalités ? et quelle est le nombre des modalités ?

3) Quelle est l'objectif pour cette expérience ? 4) Définir l'hypothèses nulles

5) Calculer les moyennes pour chaque modalité.

6) Etablir le rapport de Fisher, et calculer la valeur correspondant dans la table de Fisher.

7) conclure.

II) Les données de l'expérience précédente ne concernant que des hommes, on ajoute en plus 7 femmes dans chacun des contextes. Les observations sur celles-ci

sont les suivantes :

Groupe I Même contexte	Groupe II Contexte différent	Groupe III Contexte imaginé	Groupe IV Contexte photographié	Groupe V Placebo
11	5	21	16	14
11	17	17	28	10
22	18	16	22	5
25	8	16	19	4
23	5	25	31	5
20	11	14	25	9
16	13	25	30	5

1) Quelles sont les facteurs étudiés ? Et quelles sont leurs modalités ? 2) Ecrire l'hypothèse nulle. 3) Compléter le tableau. 4) Faire tableau de décompositions des ANOVA. 5) Conclure.

Exercice N°: 05

Au cours de l'expertise d'une trousse pour le dosage des folates, on étudie la reproductibilité pour 3 lots de réactifs (A, B et C) à 4 temps différents (T_0 = début d'utilisation de la trousse, T_1 et T_2 = temps intermédiaires et T_p = date de péremption). Chaque dosage du sérum de contrôle est réalisé en double.

	Lot A	Lot B	Lot C
T_0	4,2	3,7	4,3
	4,1	4,1	4,0
T_1	3,7	3,6	3,8
	4,2	4,1	3,9
T_2	4,3	4,6	4,1
	4,4	4,5	4,5
T_p	4,2	4,5	4,4
	4,1	4,3	4,6

- 1) Quelles sont les facteurs étudiés ?
- 2) Calculer les moyennes marginales pour les deux facteurs, ainsi que la moyenne totale ?
- 3) Compléter le tableau des données ?
- 4) Décrire la table de décomposition de l'ANOVA.
- 5) Étudier l'effet de l'interaction des facteurs étudiés ?

6) conclusion.

Exercice N°: 06

Pour tester la fiabilité de 4 laboratoires d'analyse, on utilise 4 solutions ayant le même titre de glucose dans du sérum physiologique additionné de quantités variables de galactose. Chaque laboratoire reçoit un échantillon de chaque solution et fournit le résultat de ses mesures. L'ensemble des résultats, exprimés en grammes de glucose par litre de solution, est regroupé dans le tableau suivant :

	Labo 1	Labo 2	Labo 3	Labo 4
Solution 1	1,05	1,15	1,08	1,13
Solution 2	1,12	1,15	1,11	1,09
Solution 3	1,02	1,10	1,04	1,05
Solution 4	1,09	1,11	1,07	1,10

Faire une étude ANOVA.

Chapitre 3

La correction

Exercice N°: 01

La réponse :

Facteur A : Pourcentage d'eau de mère $p = 3$

Facteur B : Type d'espèce . $q = 2$

$n = 5$, et $N = npq = 30$, le type d'ANOVA c'est ANOVA a deux facteurs avec répétitions des mêmes expériences.

	Es I	Es II	Moyenne marginale
100 %	8,946	6,406	7,676
75 %	7,428	7,374	7,401
25 %	11,978	11,682	11,830
Moyenne marginale	9,451	8,487	8,969

Modèle de l'analyse de variance à deux facteurs

$$SCE_T = SCE_A + SCE_B + SCE_{AB} + SCE_R$$

Calcul des termes

$$\sum \sum \sum x_{ijk}^2 = 2804, 225.$$

$$\sum \bar{x}_i^2 = 253, 644.$$

$$\sum \bar{x}_j^2 = 161, 331.$$

$$\sum \sum \bar{x}_{ij}^2 = 510, 560.$$

$$npq (\bar{x}^2) = 2413, 288$$

L'hypothèses nulles :

H_0 : Il n'y'a pas une influence des facteur A sur la quantité d'O₂.

H_0 : Il n'y'a pas une influence des facteur B sur la quantité d'O₂

H_0 : n'a aucun effet d'interactions.

Calculs des SCE

$$\begin{aligned}
 SCE_T &= \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n (x_{ijk} - \bar{x}_{...})^2 = \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n x_{ijk}^2 - npq\bar{x}_{...}^2 \\
 &= 2804,226 - 30(8,969)^2 = 390,937 > 0 \\
 SCE_R &= \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n (x_{ijk} - \bar{x}_{ij})^2 = \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n x_{ijk}^2 - n \sum_{i=1}^p \sum_{j=1}^q \bar{x}_{ij}^2 \\
 &= 2804,226 - 5(510,560) = 251,426 > 0 \\
 SCE_A &= qn \sum_{i=1}^p (\bar{x}_{i.} - \bar{x}_{...})^2 = qn \sum_{i=1}^p \bar{x}_{i.}^2 - npq\bar{x}_{...}^2 \\
 &= 10(253,645) - 30(8,969)^2 = 123,161 > 0 \\
 SCE_B &= pn \sum_{j=1}^q (\bar{x}_{.j} - \bar{x}_{...})^2 = pn \sum_{j=1}^q \bar{x}_{.j}^2 - npq\bar{x}_{...}^2 \\
 &= 15(161,331) - 30(8,969)^2 = 6,676 > 0 \\
 SCE_{AB} &= n \sum_{i=1}^p \sum_{j=1}^q (\bar{x}_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{...})^2 = SCE_T - SCE_R - SCE_A - SCE_B \\
 &= 9,673 > 0
 \end{aligned}$$

Tableau de ANOVA

Sources de variation	Degrés de liberté	Somme des Carrés des Ecart	Carré Moyen	Test de Fisher-Snédecor
Totale	$npq - 1 = 29$	390,937		
"type "A	$p - 1 = 2$	123,162.	61,581	$F_{obs}^A = 5,878$
"temp"B	$q - 1 = 2 - 1 = 1$	6,677	6,677	$F_{obs}^B = 0,647$
type*temp AB	$(p - 1)(q - 1) = 2$	6,673	4,836	$F_{obs}^{AB} = 0,461$
Résiduelle	$pq(n - 1) = 24$	251,425	10,476	

Comparaisons :

$$F_A^\alpha = F_{(p-1,pq(n-1))}^{0,05} = F_{(2,24)}^{0,05} = 3,40.$$

$$F_B^\alpha = F_{(q-1, pq(n-1))}^{0,05} = F_{(1,24)}^{0,05} = 4,26.$$

$$F_{AB}^\alpha = F_{((p-1)(q-1), pq(n-1))}^{0,05} = F_{(2,24)}^{0,05} = 3,40.$$

Alors $F_{obs}^{AB} = 0,461 < F_{AB}^\alpha = 3,40$. d'où acceptation de H_0'' , alors il y a effet d'interaction sur la quantité d'O₂.

Exercice N°: 02

La réponse :

Facteur A : Type des piles $p = 3$

Facteur B : Température $q = 3$

$n = 4$, et $N = npq = 36$, le type d'ANOVA c'est ANOVA a deux facteurs avec répétitions des mêmes expériences.

	15° C	70° C	125° C	Moyenne marginale
Type I	134,750	57,250	57,500	83,167
Type II	155,750	119,750	49,500	108,333
Type III	144	145,750	85,500	125,083
Moyenne marginale	144,833	107,583	64,167	105,528

Modèle de l'analyse de variance à deux facteurs

$$SCE_T = SCE_A + SCE_B + SCE_{AB} + SCE_R$$

Calcul des termes :

$$\sum \sum \sum x_{ijk}^2 = 478547.$$

$$\sum \bar{x}_i^2 = 34298,379.$$

$$\sum \bar{x}_j^2 = 36667,975.$$

$$\sum \sum \bar{x}_{ij}^2 = 115079,063.$$

$$npq\bar{x}^2 = 400894,118$$

Les hypothèses nulles :

H_0 : le facteur A n'a aucun effet sur les durées de vie.

H'_0 : le facteur B n'a aucun effet sur les durées de vie.

H''_0 : Il n'y à aucun effet d'interactions.

Calculs des SCE

$$\begin{aligned}
 SCE_T &= \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n (x_{ijk} - \bar{x}_{...})^2 = \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n x_{ijk}^2 - npq\bar{x}_{...}^2 \\
 &= 478547 - 36 (105, 528)^2 = 77642, 882 \\
 SCE_R &= \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n (x_{ijk} - \bar{x}_{ij})^2 = \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n x_{ijk}^2 - n \sum_{i=1}^p \sum_{j=1}^q \bar{x}_{ij}^2 \\
 &= 478547 - 4 (115079, 063) = 18230, 752 \\
 SCE_A &= qn \sum_{i=1}^p (\bar{x}_i - \bar{x}_{...})^2 = qn \sum_{i=1}^p \bar{x}_i^2 - npq\bar{x}_{...}^2 \\
 &= 12 (34298, 546) - 36 (105, 528)^2 = 10686, 430. \\
 SCE_B &= pn \sum_{j=1}^q (\bar{x}_j - \bar{x}_{...})^2 = pn \sum_{j=1}^q \bar{x}_j^2 - npq\bar{x}_{...}^2 \\
 &= 12 (36668, 104) - 36 (105, 528)^2 = 39121, 582 \\
 SCE_{AB} &= n \sum_{i=1}^p \sum_{j=1}^q (\bar{x}_{ij} - \bar{x}_i - \bar{x}_j + \bar{x}_{...})^2 = SCE_T - SCE_R - SCE_A - SCE_B \\
 &= 9614, 118
 \end{aligned}$$

Tableau de ANOVA

Sources de variation	Degrés de liberté	Somme des Carrés des Ecarts	Carré Moyen	Test de Fisher-Snédec
Totale	$npq - 1 = 3 \times 3 \times 4 = 35$	77645, 284		
"type "A	$p - 1 = 3 - 1 = 2$	10686, 430.	5343, 215	$F_{obs}^A = 7, 913$
"temp"B	$q - 1 = 3 - 1 = 2$	39121, 582	19560, 791	$F_{obs}^B = 28, 969$
type*temp AB	$(p - 1)(q - 1) = 4$	9614, 118	2403, 529	$F_{obs}^{AB} = 3, 559$
Résiduelle	$pq(n - 1) = 3 \times 3 \times 3 = 27$	18230, 752	675, 213	

Calculs valeurs fractiles de Fisher-Snédec :

$$F_{2,27}^A = 3, 35 \quad , F_{2,27}^B = 3, 35 \dots , F_{4,27}^{AB} = 2, 73$$

Comparaisons :

On peut remarquer que $F_{obs}^{AB} > F_{4,27}^{AB}$, donc on rejette H_0 (\exists effet d'interaction), alors les moyennes d'interactions sont différentes significativement.

Exercice N°: 03

La réponse :

Facteur A : Les laboratoires $p = 3$

Facteur B : Les villes. $q = 3$

$n = 1$, et $N = npq = 12$, le type d'ANOVA c'est ANOVA a deux facteurs sans répétitions.

Calcul des terme

$$\begin{aligned}\sum \sum x_{ij}^2 &= 2109,513. \\ \sum \bar{x}_{i.}^2 &= 682,429. \\ \sum \bar{x}_{.j}^2 &= 524,377. \\ pq\bar{x}^2 &= 2041,438\end{aligned}$$

Les hypothèses nulles :

H_0 : le facteur A n'a aucun effet sur les durées de vie.

H'_0 : le facteur B n'a aucun effet sur les durées de vie.

Tableau des données

	Ville 1	Ville 2	Ville 3	Moyennes
Labo 1	17.97	13.135	10.60	13,901
Labo2	15.30	11.49	12.09	12,960
Labo3	14.69	11.94	9.41	12,013
Labo 4	16.10	12.11	11.69	13,300
Moyennes	16,015	12,168	10,947	13,043

La décomposition des SCE

Soit l'équation de décomposition de la variation totale :

$$SCE_T = SCE_A + SCE_B + SCE_{AB}$$

Calcul des SCE

Mêmes formules en remplaçant n par 1

$$\begin{aligned}
 SCE_T &= \sum_{i=1}^p \sum_{j=1}^q (x_{ijk} - \bar{x}_{...})^2 = \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n x_{ijk}^2 - npq\bar{x}_{...}^2 \\
 &= 2109,513 - 2041,438 = 68,075 \\
 SCE_A &= q \sum_{i=1}^p (\bar{x}_{i.} - \bar{x}_{...})^2 = q \sum_{i=1}^p \bar{x}_{i.}^2 - pq\bar{x}_{...}^2 \\
 &= 3(682,429) - 2041,438 = 5,859. \\
 SCE_B &= p \sum_{j=1}^q (\bar{x}_{.j} - \bar{x}_{...})^2 = p \sum_{j=1}^q \bar{x}_{.j}^2 - pq\bar{x}_{...}^2 \\
 &= 4(524,377) - 2041,438 = 56,070 \\
 SCE_{AB} &= \sum_{i=1}^p \sum_{j=1}^q (\bar{x}_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{...})^2 = SCE_T - SCE_A - SCE_B \\
 &= 6,156
 \end{aligned}$$

Sources de variation	Degrés de liberté	Somme des Carrés des Ecart	Carré Moyen	Test de Fisher-Snécor
Totale	$npq = 3 \times 4 - 1 = 11$	67,762		
"type "A	$p - 1 = 4 - 1 = 3$	5,849.	1,949	$\frac{CM_A}{CM_{AB}} = 1,899$
"temp"B	$q - 1 = 3 - 1 = 2$	56,070	28,035	$\frac{CM_B}{CM_{AB}} = 27,324$
type*temp AB	$(p - 1)(q - 1) = 6$	6,156	1,026	

Calculs valeurs fractiles de Fisher-Snédecor :

$$F_{3,6}^A = 4,76 \quad , \quad F_{2,6}^B = 5,14.$$

Comparaisons :

On peut remarquer que $F_{obs}^A < F_{3,6}^A$, donc on accepte H_0 (il n'y a pas d'influence des laboratoir
 De plus on remarquer que $F_{obs}^B > F_{2,6}^B$, donc on rejette H'_0 (il y a une influence des villes sur les

Exercice N°: 04

La réponse :

Facteur A : Type de traitement $p = 2$

Facteur B : Psychologue . $q = 2$

$n = 3$, et $N = npq = 12$, le type d'ANOVA c'est ANOVA a deux facteurs

avec répétitions des mêmes expériences.

	Psy I	Psy II	Moyenne marginale
Produit Actif	93,7// 49,3//71,8	70,3//72,5//42,5	66,684
Placebo	56,6//24,2//37,7	37,3//50,9//36,8	40,584
Moyenne marginale	55,550	51,717	53,634

Modèle de l'analyse de variance à deux facteurs

$$SCE_T = SCE_A + SCE_B + SCE_{AB} + SCE_R$$

Calcule des termes

$$\sum \sum \sum x_{ijk}^2 = 38916,840.$$

$$\sum \bar{x}_i^2 = 6093,602.$$

$$\sum \bar{x}_j^2 = 5760,347.$$

$$\sum \sum \bar{x}_{ij}^2 = 12237,904.$$

$$npq\bar{x}_{...}^2 = 34517,984$$

L'hypothèses nulles :

H_0 : Il n'y'a pas une influence des facteur A sur l'indice de développement globale de chaque enfant.

H_0 : Il n'y'a pas une influence des facteur B sur l'indice de développement globale de chaque enfant.

H_0 : n'a aucun effet d'interactions.

Calculs des SCE

$$\begin{aligned}
 SCE_T &= \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n (x_{ijk} - \bar{x}_{...})^2 = \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n x_{ijk}^2 - npq\bar{x}_{...}^2 \\
 &= 4398,856 > 0 \\
 SCE_R &= \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n (x_{ijk} - \bar{x}_{ij})^2 = \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n x_{ijk}^2 - n \sum_{i=1}^p \sum_{j=1}^q \bar{x}_{ij}^2 \\
 &= 2203,128 > 0 \\
 SCE_A &= qn \sum_{i=1}^p (\bar{x}_i - \bar{x}_{...})^2 = qn \sum_{i=1}^p \bar{x}_i^2 - npq\bar{x}_{...}^2 \\
 &= 2043,628 > 0 \\
 SCE_B &= pn \sum_{j=1}^q (\bar{x}_j - \bar{x}_{...})^2 = pn \sum_{j=1}^q \bar{x}_j^2 - npq\bar{x}_{...}^2 \\
 &= 44,098 > 0 \\
 SCE_{AB} &= n \sum_{i=1}^p \sum_{j=1}^q (\bar{x}_{ij} - \bar{x}_i - \bar{x}_j + \bar{x}_{...})^2 = SCE_T - SCE_R - SCE_A - SCE_B \\
 &= 108,002 > 0
 \end{aligned}$$

Tableau de ANOVA

Sources de variation	Degrés de liberté	Somme des Carrés des Ecartés	Carré Moyen	Test de Fisher-Snédecor
Totale	$npq - 1 = 11$	4397,569		
"type "A	$p - 1 = 1$	2043,628	2043,628	$F_{obs}^A = 7,420$
"temp"B	$q - 1 = 1$	44,098	44,098	$F_{obs}^B = 0,160$
type*temp AB	$(p - 1)(q - 1) = 1$	108,002	108,002	$F_{obs}^{AB} = 0,392$
Résiduelle	$pq(n - 1) = 8$	2203,128	275,391	

Calculs valeurs fractiles de Student comparaisons :

$$\begin{aligned}
 F_A^\alpha &= F_{(p-1, pq(n-1))}^{0,05} = F_{(1,8)}^{0,05} = 5,32. \\
 F_B^\alpha &= F_{(q-1, pq(n-1))}^{0,05} = F_{(1,8)}^{0,05} = 5,32. \\
 F_{AB}^\alpha &= F_{((p-1)(q-1), pq(n-1))}^{0,05} = F_{(1,8)}^{0,05} = 5,32.
 \end{aligned}$$

Alors $F_{obs}^{AB} = 0,392 < F_{AB}^\alpha = 5,32$. d'où acceptation de H_0'' , alors il y a effet d'interaction sur l'indice de développement globale.

De plus $F_{obs}^A = 7,420 > F_A^\alpha = 5,32$. alors on accepte H_1 .

En fin $F_{obs}^B = 0,160 < F_B^\alpha = 5,32$. alors on accepte H_0' .

Exercice N°: 05

La réponse :

I)

1/ La variable étudiée c'est le nombre des mots mémoriser.

2/ $p=5$

3/ L'objectif : L'effet de type de contexte sur la mémorisation.

4/Le facteur étudié c'est Type de contexte, avec, et $n_i = 7.. . \quad N = \sum n_i = 35$.

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

$$\bar{X}_1 = 18,429.$$

$$\bar{X}_2 = 15,429$$

$$\bar{X}_3 = 12,286$$

$$\bar{X}_4 = 17,143$$

$$\bar{X}_5 = 11,429$$

$$\bar{X}_{..} = 14,943$$

$$\sum \sum x_{ij}^2 = 8777$$

$$\sum n_j (\bar{x}_j)^2 = 8071,927$$

Alors on peut calculer les SCE :

$$SCE_{inter} = \sum n_j (\bar{x}_j)^2 - N (\bar{X}_{..})^2 = 256,663 > 0.$$

$$SCE_{intra} = \sum \sum x_{ij}^2 - \sum n_j (\bar{x}_j)^2 = 705,073 > 0.$$

Tableaux des variations

Sources de variation	Degrés de liberté	Somme des Carrés des Ecartés	Carré Moyen	Test de Fisher-Snédecor
Totale	$N - 1 = 34$	///		
Facteur	$p - 1 = 4$	256,663	$CM_{inter} = 64,166$	$F_{obs} = 2,730$
Résiduelle	$N - p = 30$	705,073	$CM_{intra} = 23,502$	

$$F^{0,95}(p - 1, n - p) = F^{0,95}(4, 30) = 2,69$$

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 la variable.

II)

Facteur A : Sexe $p = 2$

Facteur B : Type des contextes . $q = 5$

$n = 7$, et $N = npq = 70$, le type d'ANOVA c'est ANOVA a deux facteurs avec répétitions des mêmes expériences.

	G I	G II	G III	G IV	G V	Moyenne marginale
Femme	18,286	11	19,143	24,420	27,429	16,057
Homme	18,429	12,286	15,429	17,143	11,429	14,943
Moyenne marginale	18,358	11,643	17,286	20,786	9,429	15,500

Modèle de l'analyse de variance à deux facteurs

$$SCE_T = SCE_A + SCE_B + SCE_{AB} + SCE_R$$

Calcul des moyennes

$$\begin{aligned} \sum \sum \sum x_{ijk}^2 &= 19857. \\ \sum \bar{x}_i^2 &= 481,120. \\ \sum \bar{x}_j^2 &= 1292,345. \\ \sum \sum \bar{x}_{ij}^2 &= 2626,931. \\ \bar{x}_{...}^2 &= 15,500 \end{aligned}$$

L'hypothèses nulles :

H_0 : Il n'y'a pas une influence des facteur A sur nombres des mots.

H_0 : Il n'y'a pas une influence des facteur B sur nombres des mots.

H_0 : n'a aucun effet d'interactions.

Calculs des SCE

$$\begin{aligned}
 SCE_T &= \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n (x_{ijk} - \bar{x}_{...})^2 = \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n x_{ijk}^2 - npq\bar{x}_{...}^2 \\
 &= 3039,500 > 0 \\
 SCE_R &= \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n (x_{ijk} - \bar{x}_{ij})^2 = \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n x_{ijk}^2 - n \sum_{i=1}^p \sum_{j=1}^q \bar{x}_{ij}^2 \\
 &= 1468,483 > 0 \\
 SCE_A &= qn \sum_{i=1}^p (\bar{x}_{i.} - \bar{x}_{...})^2 = qn \sum_{i=1}^p \bar{x}_{i.}^2 - npq\bar{x}_{...}^2 \\
 &= 21,700 > 0 \\
 SCE_B &= pn \sum_{j=1}^q (\bar{x}_{.j} - \bar{x}_{...})^2 = pn \sum_{j=1}^q \bar{x}_{.j}^2 - npq\bar{x}_{...}^2 \\
 &= 1275,300 > 0 \\
 SCE_{AB} &= n \sum_{i=1}^p \sum_{j=1}^q (\bar{x}_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{...})^2 = SCE_T - SCE_R - SCE_A - SCE_B \\
 &= 273,987 > 0
 \end{aligned}$$

Tableau de ANOVA

Sources de variation	Degrés de liberté	Somme des Carrés des Ecartés	Carré Moyen	Test de Fisher-Snédecors
Totale	$npq - 1 = 69$	3039,500		
"type "A	$p - 1 = 1$	21,700	21,700	$F_{obs}^A = 0,887$
"temp"B	$q - 1 = 4$	1275,300	318,825	$F_{obs}^B = 13,027$
type*temp AB	$(p - 1)(q - 1) = 4$	273,987	68,504	$F_{obs}^{AB} = 2,799$
Résiduelle	$pq(n - 1) = 60$	1468,483	24,475	

Calculs valeurs fractiles de Fisher-Snédecors comparaisons :

$$F_{AB}^\alpha = F_{((p-1)(q-1), pq(n-1))}^{0,05} = F_{(4,60)}^{0,05} = 2,53.$$

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.