

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

Chapitre 06 :

Le 31/05/2020

Par
Dr : CHALA ADEL

BioStatistiques

2019-2020

Je dédie ce travail.....

A mes parents ils m'ont tous,
avec leurs moyens, soutenu et donné
la force d'aller toujours
plus loin.

Table des matières

Table des Matière	iii
1 Analyse de la variance	1
1.1 Analyse de la variance à un facteur	1
1.1.1 Structure des données	1
1.1.2 Conditions d'application de l'analyse	2
1.1.3 Modèle de l'analyse de variance	3
1.2 Equation fondamentale de l'analyse de variance	3
1.2.1 Pratique de l'analyse de variance	4
1.2.2 Application et tableau de variation	5
1.3 Analyse de la variance à deux facteurs	5
1.3.1 Echantillons de plusieurs observations (de taille n)	5
1.4 Echantillons d'une seule observation	8

Chapitre 1

Analyse de la variance

1.1 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.

1.1.1 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} .

1.1.2 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}_1 .

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, 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 ($=n-1$).

1.1.3 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.

1.2 Equation fondamentale de l'analyse de variance

Notation :

SCEtotale = Somme des écarts totaux ou variation totale.

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

SCEintra = Somme des écarts résiduels ou variation intra (interne à chaque modalité).

1.2.1 Pratique de l'analyse de variance

Principe du test

Soit l'équation de décomposition de la variation totale

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

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

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.

1.2.2 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.

1.3 Analyse de la variance à deux facteurs

1.3.1 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 M
Modalité 1	x_{111}, \dots, x_{11n}	x_{121}, \dots, x_{12n}	x_{131}, \dots, x_{13n}	
	Moyenne $C11 = \bar{x}_{11}$	Moyenne $C12 = \bar{x}_{12}$	Moyenne $C13 = \bar{x}_{13}$	$\bar{x}_{.1}$
Modalité 2	x_{211}, \dots, x_{21n}	x_{221}, \dots, x_{22n}	x_{231}, \dots, x_{23n}	
	Moyenne $C21 = \bar{x}_{21}$	Moyenne $C22 = \bar{x}_{22}$	Moyenne $C23 = \bar{x}_{23}$	$\bar{x}_{.2}$
Modalité 3	x_{311}, \dots, x_{31n}	x_{321}, \dots, x_{32n}	x_{331}, \dots, x_{33n}	
	Moyenne $C31 = \bar{x}_{31}$	Moyenne $C32 = \bar{x}_{32}$	Moyenne $C33 = \bar{x}_{33}$	$\bar{x}_{.3}$
Modalité p	x_{p11}, \dots, x_{p1n}	x_{p21}, \dots, x_{p2n}	x_{p31}, \dots, x_{p3n}	
	Moyenne $Cp1 = \bar{x}_{p1}$	Moyenne $Cp2 = \bar{x}_{p2}$	Moyenne $Cp3 = \bar{x}_{p3}$	$\bar{x}_{.p}$
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 :

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

Calcul des SCE

$$SCET = \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n (x_{ijk} - \bar{x}_{...})^2 = \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n x_{ijk}^2 - npq\bar{x}_{...}^2$$

$$SCER = \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n (x_{ijk} - \bar{x}_{ij})^2 = \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n x_{ijk}^2 - n \sum_{i=1}^p \sum_{j=1}^q \bar{x}_{ij}^2$$

$$SCEA = qn \sum_{i=1}^p (\bar{x}_{i.} - \bar{x}_{...})^2 = qn \sum_{i=1}^p \bar{x}_{i.}^2 - npq\bar{x}_{...}^2$$

$$SCEB = pn \sum_{j=1}^q (\bar{x}_{.j} - \bar{x}_{...})^2 = pn \sum_{j=1}^q \bar{x}_{.j}^2 - npq\bar{x}_{...}^2$$

$$SCEAB = n \sum_{i=1}^p \sum_{j=1}^q (\bar{x}_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{...})^2 = SCET - SCER - SCEA - SCEB$$

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 : les facteurs A et B n'interagissent pas sur les résultats.

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.

1.4 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 M
Modalité 1	Moyenne C11 = \bar{x}_{11}	Moyenne C12 = \bar{x}_{12}	Moyenne C13 = \bar{x}_{13}	$\bar{x}_{1.}$
Modalité 2	Moyenne C21 = \bar{x}_{21}	Moyenne C22 = \bar{x}_{22}	Moyenne C23 = \bar{x}_{23}	$\bar{x}_{2.}$
Modalité 3	Moyenne C31 = \bar{x}_{31}	Moyenne C32 = \bar{x}_{32}	Moyenne C33 = \bar{x}_{33}	$\bar{x}_{3.}$
Modalité p	Moyenne Cp1 = \bar{x}_{p1}	Moyenne Cp2 = \bar{x}_{p2}	Moyenne Cp3 = \bar{x}_{p3}	$\bar{x}_{p.}$
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	$(CM_{AB}) = \frac{SCE_{AB}}{(p-1)(q-1)}$	