FICHE MÉTHODOLOGIQUE

Série de Fiches Méthodologiques en Recherche et en Rédaction Scientifique
Fiche Méthodologique n°9:
Comment comparer plusieurs moyennes par le test d'Analyse de Variance (ANOVA) ?
How to compare several means using test of Analysis of Variance (ANOVA)?

كيف نقارن بين معدلات عدة مؤشرات

Rym Mallekh[1], Hajer Hannachi[1], Hyem Khiari[1], Faouzi Mehdi[2], Mohamed Hsairi[1], Réseau Maghrébin PRP2S*

** Réseau- Maghrébin: Pédagogie- Recherche- Publication en Sciences de Santé (PRP2S)*
1 : Service d'Epidémiologie – Institut de Cancérologie Salah Azaiz Tunis
2 : Direction Générale de la Santé Militaire - Tunis

Cette série…

Le Réseau Maghrébin PRP2S et la Rédaction de la revue « La Tunisie Médicale » ont l'honneur de vous présenter, régulièrement à partir du numéro de janvier 2020, une série des fiches techniques en épidémiologie et en bio statistique. Ces fiches méthodologiques décrivent, d'une manière standardisée, les modes d'usage des concepts, des outils et des méthodes utilisés lors des différentes phases de la rédaction médicale scientifique depuis la phase de la recherche documentaire jusqu'à la phase de la communication médicale scientifique.
Cette série est rédigée par des experts de méthodologie de recherche dans les universités du Grand Maghreb et les facultés sœurs au Nord de la Méditerranée. Chaque fiche répond à trois questions essentielles (Quoi ? Pourquoi ? Comment) du concept étudié, en se basant sur un article publié dans la revue Tunis Med.
Le coordinateur de la série « Fiches Méthodologiques ».

Professeur Ahmed Ben Abdelaziz (Président du Réseau Maghrébin PRP2S)
ahmedbenabdelaziz.prp2s@gmail.com

Série des Fiches méthodologiques
Sommaire

Correspondance
Ahmed Ben Abdelaziz
Laboratoire de Recherche LR19SP01 « Mesure et Appui de la Performance des Etablissements de Santé ». Université de Sousse (Tunisie)
ahmedbenabdelaziz.prp2s@gmail.com

ETUDE DE CAS

Dans un article de la revue «La Tunisie Médicale», intitulé «Pneumopathies communautaires chez l'enfant» [1], les auteurs ont analysé les aspects cliniques, étiologiques et évolutifs des pneumopathies communautaires chez l'enfant, ont recherché les arguments cliniques et biologiques et ont précisé la place des examens microbiologiques dans l'orientation étiologique. Il s'agissait d'une enquête prospective, ayant porté sur 72 dossiers hospitaliers du service de médecine infantile B de l'Hôpital d'Enfants de Tunis, sur une période de 7 mois allant du 1er décembre 2004 au 30 juin 2005. Les enfants âgés de 3 mois à 10 ans, hospitalisés pour pneumonie aigue communautaire, définie par un ou plusieurs foyers radiologiques de condensation parenchymateuse dans un contexte de fièvre associée à des signes respiratoires et/ou digestifs, ont été retenus. Au niveau de l'analyse statistique, on lisait: «Les comparaisons de plusieurs moyennes sur séries indépendantes ont été effectuées au moyen du test F de Snedecor d'analyse de la variance paramétrique (ANOVA à un facteur) et en cas de faible effectif par le test H de Krusckall-Wallis d'analyse de la variance non paramétrique ». Le tableau 3 de l'article a résumé les comparaisons des moyennes des paramètres biologiques selon le type de pneumopathie (bactérienne, atypique ou non déterminée).

Tableau 3 : Valeurs moyennes des tests biologiques selon l'étiologie de la pneumopathie

Examen biologique	n	Pneumopathie bactérienne	Pneumopathie atypique	Non déterminé	p
GB (mm3)	39	18983	18550	20052	0,9
PNN (mm3)	36	16220	13807	13807	0,6
CRP (mg/l)	37	38,4	126	55,6	0,03
Procalcitonie (ng/ml)	34	6,67	4,34	3,87	0,4

Quiz :

1. L'analyse de la variance ou ANOVA :
    a. *Permet de comparer plusieurs moyennes.*
    b. *Permet de comparer plusieurs distributions.*
    c. *Permet de comparer plusieurs variances.*
2. Le tableau 3 de l'article vous semble-t-il avoir inclus toutes les informations nécessaires
    a. *Il manque le nombre de sujets pour chaque groupe,*
    b. *Il manque les valeurs des écarts types.*
    c. *a+b*
3. Les conditions d'application de l'ANOVA ont-elles été vérifiées ?
    a. *La normalité des distributions n'a pas été vérifiée.*
    b. *L'égalité des variances n'a pas été vérifiée.*
    c. *a+b*

INTRODUCTION

Pour comparer deux moyennes à partir de deux échantillons indépendants, on utilise le test T de Student ou le test de l'écart réduit si la taille de l'échantillon est assez grande. L'analyse de la variance ou ANOVA est une méthode de comparaison globale de plus de deux moyennes, qui généralise le test T de Student [2]. Cette méthode peut inclure un seul facteur ou plus d'un facteur. Nous commencerons par expliquer le principe de l'analyse de la variance: pourquoi est-on amené à comparer des variances pour résoudre un problème de comparaison de moyennes ? Nous définirons ensuite la méthode et en exposerons les étapes.

ANOVA : POURQUOI ?

Pour comparer deux moyennes sur deux échantillons indépendants, on peut utiliser les tests paramétriques: test de l'écart réduit et test T de Student. Lorsqu'on compare plus de deux moyennes, Il est incorrect d'effectuer des comparaisons deux à deux; car la multiplication des tests multiplierait le risque d'erreur. L'analyse de variance ANOVA permet de remédier à cette situation en utilisant un seul test, qui suit la distribution de Fisher [2,3].

ANOVA: QUOI ?

Il s'agit de comparer les moyennes d'une variable quantitative X dans k populations à partir de k échantillons indépendants de tailles nj. Soient μj les moyennes respectives de X dans les populations d'origine. Soit μ la moyenne générale dans la population regroupant les k populations (valable si la répartition de l'échantillon dans les k populations est celle de l'ensemble de la population). L'hypothèse nulle Ho serait que les échantillons pour chaque groupe proviennent de populations ayant les mêmes valeurs des moyennes. Ho: $\mu 1=\mu 2=\ldots=\mu k=\mu$. L'hypothèse alternative H1: l'une au moins des moyennes μj diffère des autres. L'analyse de variance est basée sur la variabilité («variance») entre les groupes; cela signifie que si les moyennes des groupes sont suffisamment éloignées, cela suggère que les groupes appartiennent à des populations différentes. Le processus consiste à décomposer la variabilité globale en deux composantes de variabilité: i) La variabilité entre les groupes ou «variance entre les groupes» (ii) la variabilité restante non due à des différences entre les groupes: «variance résiduelle».

Si les groupes sont vraiment différents, la variance entre les groupes sera beaucoup plus grande que la variance résiduelle. Ceci est testé en utilisant le rapport des deux variances: le rapport. $F_{Observé} = \frac{s^2_F}{s^2_R}$ où s²F est l'estimation de la variance intergroupe ou liée au facteur F et s²R est l'estimation de la variance intragroupe ou résiduelle [2,3]. Si la variabilité entre les groupes n'est pas supérieure à ce à quoi nous nous attendions au hasard seul, alors les deux estimations seront similaires et le rapport F sera proche de 1,0. Si le rapport F est bien supérieur à 1,0; les deux estimations doivent être très différentes, fournissant la preuve que les moyennes des groupes sont différentes.

Conditions d'application:

- Données continues, normalement réparties dans chaque groupe.
- Variance égale (écart-type) dans chaque groupe. [2]

La distribution de Fisher:

Le rapport F suit une distribution F si l'hypothèse nulle est vraie, c'est-à-dire s'il n'y a pas de différence entre les moyennes. La distribution F est déterminée par ses deux paramètres, les degrés de liberté des deux estimations de variance: (i) nombre de groupes – 1, (ii) total des observations - nombre de groupes. Le rapport F a une valeur p correspondante, $p < 0,05$ est interprété comme indiquant que les moyennes des groupes sont globalement différentes les unes des autres.

ANOVA: COMMENT ?

Les calculs sont généralement effectués par un logiciel; mais peuvent être effectués manuellement. Dans le but d'éviter les formules qui peuvent être perçues comme compliquées, nous allons partir d'un exemple. Les données du tableau suivant proviennent d'une expérience en double aveugle de l'effet de la caféine sur la vitesse de taper du doigt comme mesure de la performance. Trente sujets ont reçu une des trois doses de caféine, 0 mg, 100 mg, ou 200 mg. Le nombre de fois de taper du doigt par minute a été enregistré pour chaque sujet.

Tableau 1. Nombre de fois de taper au doigt par minute pour chaque sujet en fonction de la dose de caféine

Dose de caféine (mg)	Nombre de fois de taper du doigt par minute pour chaque sujet									
0	242	245	244	248	247	248	242	244	246	242
100	248	246	245	247	248	250	247	246	243	244
200	246	248	250	252	248	250	246	248	245	250

Tableau 2. Moyennes et écarts-types du nombre de fois de taper du doigt par minute dans chaque groupe :

Dose de caféine (mg)	n	Moyenne	(Ecart type)	IC95% de la moyenne
0	10	244,8	2.39	243.1 -246.5
100	10	246,4	2.07	244.9 - 247.9
200	10	248,3	2.21	246.7 - 249.9

Conditions d'application: L'échantillon étant petit, il fallait vérifier la normalité de la distribution de la variable (nombre de fois de taper du doigt); plusieurs tests peuvent être utilisés, le plus courant étant celui de **Kolmogorov et Smirnov**. L'application de ce test est en faveur d'une distribution normale (figure 1).

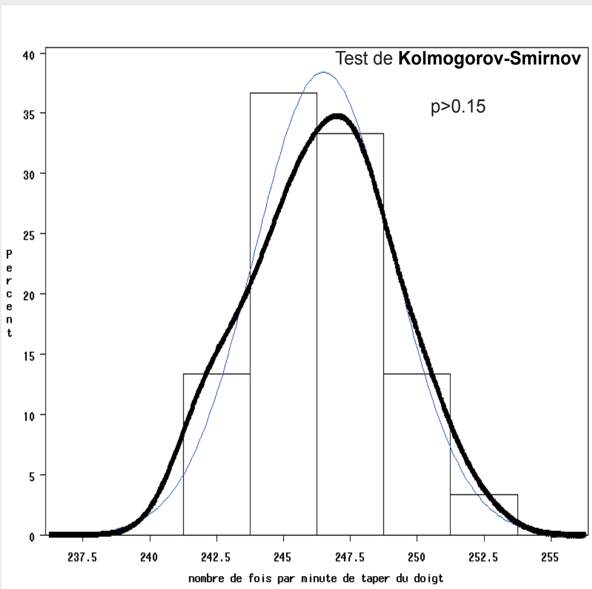

Figure 1. Test de la normalité de la distribution de la variable «nombre de fois de taper du doigt par minute»

En plus, les valeurs des écarts types sont relativement proches les unes et des autres; ce qui est en faveur de l'égalité des variances. Cette égalité des variances pourrait être vérifiée par le **test de Bartlett**. Il est à noter que le test d'ANOVA est robuste et il reste valide si cette condition d'égalité des variances n'est pas respectée [2].

Les conditions d'application de l'ANOVA étant réunies, nous pourrons dresser le tableau d'ANOVA (tableau 3). Ce tableau montre comment la variabilité totale est divisée en deux composantes: une composante en rapport avec la variabilité entre groupes et une deuxième composante restante qualifiée de variabilité aléatoire ou résiduelle.

Tableau 3. Tableau d'analyse de la variance

Source de variation	nombre de degrés de libertés	Somme des carrées des écarts	Variance	Valeur de F	P
Entre groupes	2	61,4	30,7	6,18	**0,006**
Résiduelle	27	134,1	4,97		
Totale	29	195,5			

Dans ce tableau, la ligne 2 donne les statistiques de la **variabilité entre les groupes,** la ligne 3 donne les statistiques de la **variabilité «résiduelle», l**a ligne 4 donne les totaux globaux. Le nombre de degré de liberté, pour la ligne 2, il est égal au nombre de groupes – 1= 2. Pour la ligne 4, il est égal au nombre total d'observations - 1 = 29, et pour la ligne 3, c'est la différence entre les deux précédents.

• La somme totale des carrés des écarts est calculée de manière similaire à une somme de carrés des écarts pour le calcul d'une variance.

$$(242–246,5)^2 + (245–246,5)^2 + (244–246,5)^2 +\dots + (250–246,5)^2 = 195,5$$

• La somme des carrés des écarts entre les groupes est basée sur la somme des carrés des écarts entre la moyenne de chaque groupe et la moyenne globale:

$$10 \times [(244,8–246,5)^2 + (246,4–246,5)^2 + (248,3–246,5)^2] = 61,4$$

• La somme résiduelle des carrés est obtenue par soustraction:

$$195,5 – 61,4 = 134,1$$

• La somme des carrés des écarts résiduelle est égale à la somme des carrés des écarts totale moins la somme des carrés des écarts entre les groupes:

• L'estimation de la variance est la somme des carrés des écarts divisée par le nombre de degrés de libertés:

Variance entre les groupes = 61,4 / 2 = 30,7
Variance résiduelle = 134,1 / 27 = 4,97
Le rapport F est le rapport de 2 variances: 30,7 / 4,97 = 6,18
La valeur p est la probabilité associée à une valeur F ≥ 6,18.
La distribution de Fisher est actuellement intégrée dans Excel.

Tableau 4. Distribution de Fisher à 2 et 29 degrés de libertés :

		Degré de signification (p)	Valeur de F
2	◄ **Nombre de degrés de liberté du numérateur**	0,1%	0,00
		0,5%	6,40
29	◄ **Nombre de degrés de liberté du dénominateur**	1,0%	5,42
		2,5%	4,20
		5,0%	3,33
		10,0%	2,50

Nous concluons que les moyennes des groupes sont globalement différentes. Il existe donc de bonnes preuves que la caféine est associée à la vitesse de prélèvement du sang du doigt.

Comparaisons des paires de moyennes:

Après avoir effectué une analyse de variance, il peut être souhaitable de comparer des paires de moyennes particulières. Il faut faire attention à la façon dont cela est fait pour empêcher les faux résultats significatifs qui se produiront lorsque de nombreuses comparaisons sont effectuées. Le test T de Student ne doit pas être utilisé pour tester toutes les combinaisons de comparaisons possibles. Des meilleures méthodes sont disponibles qui prennent en compte plusieurs tests en préservant le taux d'erreur de type 1 à 5%, comme les méthodes de Bonferroni, de Scheffé, de Newman, etc. Le choix dépend des données et des programmes statistiques disponibles. La *correction de Bonferroni* est une méthode simple pour corriger la coupure pour signification statistique pour les tests multiples. Elle est basée sur le fait que si l'hypothèse nulle est vraie et un test est réalisée avec p≤ 0,05, pris comme seuil de signification, la probabilité d'absence de différence significative qui est de 0,95. Il en résulte que si 10 tests indépendants sont effectués, alors la probabilité qu'aucun ne soit significatif est $0{,}95^{10} = 0{,}60$. Si α est le seuil de signification, alors pour conserver le niveau de signification à 0,05, nous avons besoin de $(1 – α)^{10} = 0{,}95$. Etant donné que α est petit, on peut montrer que $(1 – α)^{10}$ est approximativement égal à 1–10 α. Pour que cela soit égal à 0,95, nous devons avoir α = 0,05 / 10. Donc, en général, si n tests sont réalisée, le seuil de signification est de: 0,05 /n après correction de Bonferroni [4] **.** Il est à noter que la méthode d'analyse de variance à un facteur peut être étendue à plus d'un facteur. D'autre part, la comparaison de moyennes de mesures répétées dans le temps nécessite une méthodologie différente.

CONCLUSION

L'analyse de la variance est indiquée dans la comparaison de plusieurs moyennes d'une variable quantitative X dans les populations qui diffèrent par un facteur qualitatif F à plusieurs modalités. L'application de ce test requiert des conditions d'applications, qui sont la normalité des distributions les populations d'origine et l'égalité des variances dans les différents groupes. Toutefois, le test est robuste et il reste valide si les variances ne sont pas égales.

L'essentiel à retenir

*L'analyse de variance à un facteur contrôlé ou ANOVA a pour objectif de **tester l'effet d'un facteur F** sur une **variable aléatoire quantitative continue X**. Ceci revient à comparer les moyennes de plusieurs (k) populations normales et de même variance à partir d'échantillons aléatoires et indépendants les uns des autres. *Chaque échantillon est soumis ou correspond à une modalité du facteur F.

*Le terme ANOVA indique que la comparaison de plusieurs moyennes correspond à **la comparaison de deux variances** : la variance intergroupe ou liée au facteur : σ^2_F ; La variance intragroupe ou résiduelle : σ^2_R

*Par un test de Fisher-Snedecor où la statistique calculée est :

$F_{Observé} = \frac{s_F^2}{s_R^2}$ est comparée à : $F_{seuil} = F^{k-1}_{n-k,\alpha}$ à k-1 et n-k degrés de libertés, au risque d'erreur α.

* Un test significatif permet de conclure à une différence statistiquement significative entre les moyennes comparées.

Réponses aux questions de quizz

1. *a* - L'ANOVA permet de comparer plusieurs moyennes (en comparant les variances résiduelle et liée au facteur). Elle ne permet pas de comparer des distributions (tests non paramétriques), les distributions comparées étant à priori supposées normales et de variance égale. Bien que le nom du test soit analyse de variance, le test ne compare pas plusieurs variances.
2. *c* - Il fallait mentionner dans le tableau les effectifs de chaque groupe, ainsi que les valeurs des écarts types.
3. *c* - Etant donné que les effectifs par groupe sont réduits, il fallait vérifier la normalité des distributions des différents facteurs étudiés, sauf si on a la certitude (à travers des références) que les distributions sont normales dans la population. Il fallait vérifier aussi l'égalité des variances; toutefois, le test demeure robuste s'il y a un écart à cette condition.

Pour en savoir plus

1. Tinsa F, Boussetta K, Gharbi A, Bousnina D, Abdelaziz R, Brini I, Bousnina S. Pneumopathies communautaires chez l'enfant. Tunis Med 2009; 87(12): 851–6
2. Bouyer J. Méthodes statistiques: médecine-biologie. Estem Inserm; 1996.
3. Kim TK. Understanding one-way ANOVA using conceptual figures. Korean J Anesthesiol 2017; 70(1):22-6.
4. Sedgwick P. Multiple hypothesis testing and Bonferroni's correction. BMJ 2014;349: g6284-g6284.