Introduction

L'analyse de variance simple ou ANOVA simple (**An**alysis **o**f **Va**riance) est un test statistique qui sert à comparer les moyennes quand vous en avez trois ou plus. Pour faire une analyse de variance simple, il est important que vos conditions soient effectuées dans un plan simple (c'est-à-dire un participant prend part à une seule condition). L'ANOVA simple se substitue donc au test t pour échantillons indépendants quand vous avez plus de deux groupes.

Quand on fait une ANOVA, on cherche à calculer la **statistique *F***. Pour cela, il faut diviser les **carrés moyens intergroupe** par les **carrés moyens intragroupe**. Le calcul de l'ANOVA se fait donc en trois grands paliers: 1) Calculer les carrés moyens intergroupe ; 2) Calculer les carrés moyens intragroupe ; 3) Calculer la statistique *F*. Les carrés moyens inter et intragroupe sont eux-même calculés en trois étapes : 1) Somme des carrés inter/intagroupe 2) degré de liberté inter/intragroupe 3) Diviser la somme des carrés inter/intragroupe par le degré de liberté inter/intragroupe. L'ANOVA nécessite donc plus de calculs, mais on refait globalement toujours la même chose.

On peut se demander pourquoi s'embêter à faire une ANOVA, alors qu'on pourrait simplement faire plusieurs tests t. Par exemple, nous avons trois groupes et on fait un un test t entre groupe 1 et groupe 2, entre groupe 1 et groupe 3 et entre groupe 2 et 3. En fait, cette méthode pose deux principaux problèmes. Premièrement, cela ce fait certes relativement bien avec trois groupes, mais plus il y aura de groupes, plus il y aura de tests à faire et la quantité de tests augmentera de façon exponentielle. Par exemple, avec cinq groupes, il faudra faire dix tests t et avec dix groupes il faudrait en faire 45. Le second problème que cela pose est qu'en statistique en essaie généralement de faire le moins de tests possible car chaque nouveau test augmente les chances de commettre une erreur de type 1 (c'est-à-dire trouver un effet, alors qu'il n'y en a pas).

Formules

1. Somme des carrés intergroupe

Moyenne du groupe j

Moyenne de tous les groupes

Somme des carrés intergroupe

$$SC_{inter} = \sum_{j=1}^{K} n_j (\bar{X}_j - \bar{X})^2$$

Taille du groupe j (permet de pondérer pour la taille du groupe dans le cas de groupes de tailles inégales)

4. Somme des carrés intragroupe

Valeur de la VD pour un participant

Somme des carrés intragroupe

$$SC_{intra} = \sum_{j=1}^{K} \sum_{i=1}^{n_i} (X_{ij} - \bar{X}_j)^2$$

Moyenne du groupe j

Somme à travers les groupes

Somme à travers les participants d'un groupe

2. Degré de liberté intergroupe

$$dl_{inter} = K-1$$

Nombre de groupes

5. Degré de liberté intragroupe

Nombre de groupes

$$dl_{intra} = N-K$$

Nombre total d'observations

3. Carrées moyens intergroupe

Carré moyen intergroupe

$$CM_{inter} = \frac{SC_{inter}}{dl_{inter}}$$

Somme des carrés intergroupe

Degrés de liberté

6. Carrées moyens intergroupe

Carré moyen intragroupe

$$CM_{intra} = \frac{SC_{intra}}{dl_{intra}}$$

Somme des carrés intragroupe

Degrés de liberté

7. Statistique F

Statistique F

$$F = \frac{CM_{inter}}{CM_{intra}}$$

Carrés moyen intergroupe

Carrés moyen intragroupe

Mise en situation

Mario a une compagnie de plombiers et possède un centre dans chaque arrondissement de la ville avec chacun 5 employés. La ville dans laquelle Mario a son entreprise est constituée de 3 arrondissements : 1) Toad Town ; 3) Mushroom City ; 3) Koopa Village. Il demande à chaque employé combien d'heures par semaine est-ce qu'ils font afin de voir si un groupe d'employés fait, en moyenne, plus d'heures que les autres. Voici ses résultats:

Employé	**Arr. 1**	**Arr. 2**	**Arr. 3**
1	35	40	50
2	30	30	40
3	20	20	25
4	15	30	20
5	25	30	40
Moyennes	**25**	**30**	**35**

Les employés de l'arrondissement de Koopa Village semblent faire plus d'heure en moyenne, suivi de Mushroom City et de Toad Town, mais les résultats sont-ils significatifs ?

H: La différence entre nos moyennes est supérieure (ou inférieure) à 0

H_0: La différence entre nos moyennes est de 0

Calculs

Étape 1 : Moyenne de tous les groupes

$$\underline{X} = \frac{25 + 30 + 35}{3} = \frac{90}{3} = 30$$

Étape 2 : Somme des carrés intergroupe

$$SC_{inter} = \sum_{j=1}^{K} \square\, n_j(\underline{X}_j - \underline{X})^2$$

$$5(25 - 30)^2 + 5(30 - 30)^2 + 5($$

$$5(-5)^2 + 5(0)^2 + 5(5)^2$$

$$5 * 25 + 5 * 0 + 5 * 25$$

$$125 + 0 + 125$$

250

Étape 3 : Degré de liberté intergroupe

$$dl_{inter} = K - 1$$

3-1=2

Étape 4 : Carrés moyens intergroupe

$$CM_{inter} = \frac{SC_{inter}}{dl_{inter}}$$

250/2=125

Étape 5 : Somme des carrés intragroupe

$$SC_{intra} = \sum_{j=1}^{K} \sum_{i=1}^{n_i} (X_{ij} - \underline{X_j})^2$$

Arr. 1 $(35 - 25)^2 + (30 - 25)^2 + (20 - 25)^2 + (15 - 25)^2 + (25 - 25)^2 = 250$
Arr. 2 $(40 - 30)^2 + (30 - 30)^2 + (20 - 30)^2 + (30 - 30)^2 + (30 - 30)^2 = 200$
Arr. 3 $(50 - 35)^2 + (40 - 35)^2 + (25 - 35)^2 + (20 - 35)^2 + (40 - 35)^2 = 600$

250+200+600=1050

Étape 6: Degré de liberté intragroupe

$$dl_{intra} = K - K$$

15-3=12

Étape 7 : Carrés moyens intergroupe

$$CM_{intra} = \frac{SC_{intra}}{dl_{intra}}$$

1050/12=87.5

Étape 8 : Calculer la statistique *F*

$$F = \frac{CM_{inter}}{CM_{intra}}$$

125/87.5=0.43

Étape 9 : Valeur Critique

Normalement, vous devriez avoir une table F fournie par votre professeur ou dans votre manuel. Pour cet exemple, nous allons travailler avec celle fournie par Tabachnick et Fidell dans leur manuel (Tabachnick et Fidell, 2021). Pour trouver la valeur critique dans la table F, il y a un degré de liberté de plus à prendre en compte. En effet, nous avons un degré de liberté intergroupe et un degré de liberté intragroupe. Habituellement les degrés de liberté intergroupe sont présentés à l'horizontal et les degrés de liberté intragroupe sont présentés à la verticale.

Dans notre cas, le degré de liberté intergroupe est de 2 et le degré de liberté intragroupe est de 12. Nous utiliserons un seuil de rejet de 0.05. Notre statistique *F* est de 0.43. Notre valeur critique est de 3.88.

TABLE C.3 Continued

df_1 / df_2		1	2	3	4	5	6	8	12	24	∞
	1%	11.26	8.65	7.59	7.01	6.63	6.37	6.03	5.67	5.28	4.86
	2.5%	7.57	6.06	5.42	5.05	4.82	4.65	4.43	4.20	3.95	3.67
	5%	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.28	3.12	2.93
	10%	3.46	3.11	2.92	2.81	2.73	2.67	2.59	2.50	2.40	2.29
9	0.1	22.86	16.39	13.90	12.56	11.71	11.13	10.37	9.57	8.72	7.81
	0.5	13.61	10.11	8.72	7.96	7.47	7.13	6.69	6.23	5.73	5.19
	1	10.56	8.02	6.99	6.42	6.06	5.80	5.47	5.11	4.73	4.31
	2.5	7.21	5.71	5.08	4.72	4.48	4.32	4.10	3.87	3.61	3.33
	5	5.12	4.26	3.86	3.63	3.48	3.37	3.23	3.07	2.90	2.71
	10	3.36	3.01	2.81	2.69	2.61	2.55	2.47	2.38	2.28	2.16
10	0.1	21.04	14.91	12.55	11.28	10.48	9.92	9.20	8.45	7.64	6.76
	0.5	12.83	8.08	7.34	6.87	6.54	6.12	5.66	5.17	4.64	
	1	10.04	7.56	6.55	5.99	5.64	5.39	5.06	4.71	4.33	3.91
	2.5	6.94	5.46	4.83	4.47	4.24	4.07	3.85	3.62	3.37	3.08
	5	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.91	2.74	2.54
	10	3.28	2.92	2.73	2.61	2.52	2.46	2.38	2.28	2.18	2.06
11	0.1	19.69	13.81	11.56	10.35	9.58	9.05	8.35	7.63	6.85	6.00
	0.5	12.23	8.91	7.60	6.88	6.42	6.10	5.68	5.24	4.76	4.23
	1	9.65	7.20	6.22	5.67	5.32	5.07	4.74	4.40	4.02	3.60
	2.5	6.72	5.26	4.63	4.28	4.04	4.88	3.66	3.43	3.17	2.88
	5	4.84	3.98	3.59	3.36	3.20	3.09	2.95	2.79	2.61	2.40
	10	3.23	2.86	2.66	2.54	2.45	2.39	2.30	2.21	2.10	1.97
12	0.1	18.64	12.97	10.80	9.63	8.89	8.38	7.71	7.00	6.25	5.42
	0.5	11.75	8.51	7.23	6.52	6.07	5.76	5.35	4.91	4.43	3.90
	1	9.33	6.93	5.95	5.41	5.06	4.82	4.50	4.16	3.78	3.36
	2.5	6.55	5.10	4.47	4.12	3.89	3.73	3.51	3.28	3.02	2.72
	5	4.75	3.88	3.49	3.26	3.11	3.00	2.85	2.69	2.50	2.30
	10	3.18	2.81	2.61	2.48	2.39	2.33	2.24	2.15	2.04	1.90
13	0.1	17.81	12.31	10.21	9.07	8.35	7.86	7.21	6.52	5.78	4.97
	0.5	11.37	8.19	6.93	6.23	5.79	5.48	5.08	4.64	4.17	3.65
	1	9.07	6.70	5.74	5.20	4.86	4.62	4.30	3.96	3.59	3.16

Si votre statistique *t* est <u>**supérieure**</u> à votre valeur critique, vous pouvez rejeter l’hypothèse nulle.

$0.43<3.88$

Notre statistique *F* est plus petite que la valeur critique, nous ne pouvons pas rejeter l’hypothèse nulle.

Signification des symbole

X : Valeur

N : Nombre d’observations total

n : Nombre d’observations d’un groupe

- n_1 : nombre d'observations du groupe 1 ;
- n_2 : nombre d'observations du groupe 2 ;
- etc.

K : Nombre de groupes

Σ : Somme

$\bar{X}$: Moyenne d’un échantillon

μ : Moyenne d’une population

s : Écart-type d'un échantillon

σ : Écart-type d'une population

s^2 : Variance d'un échantillon

σ^2 : Variance d'une population

χ^2 : Khi-carré

A: Fréquence attendue

O: Fréquence observé

L: Ligne

C: Colonne

t : Statistique t (ou score t dans le cas d'une corrélation/régression linéaire)

r : Coefficient de Pearson

Z : Score Z

b : Coefficient de régression

a : Ordonnée à l'origine

$\hat{Y}$: Valeur de la VD qu'on veut prédire l'aide la VI

F : Statistique F

Ouvrages de référence

- Field, A. (2017). *Discovering Statistics Using IBM SPSS Statistics: North American Edition* (5th ed.). Sage Edge
- Howell. D.C. (2008). *Méthodes statistiques en sciences humaines* (M. Rogier, V. Yzerbyt, & Y. Bestgen, Trans.). (6th ed.). De boeck. (Original work published 2008)
- Tabachnick, B. G., & Fidell, L. S. (2021). *Using Multivariate Statistics* (7th ed.). Pearson