



1<sup>ère</sup> Année Médecine, 2022/2023

**TD sur l'ANOVA**

**Remarque :**

Dans les solutions des exercices et pour la comparaison de plusieurs variances on utilise la méthode EXCEL pour le calcul numérique et comme règle dans le rapport F des deux variances à considérer, on met la plus grande variance au numérateur et la plus petite au dénominateur.

**Nota :**

**Les différentes formules de l'analyse de la variance vues au cours :**

$$SCE_{fa} = \sum n_i (\bar{X}_i - \bar{x})^2 = \sum \frac{x_{i.}^2}{n_i} - \frac{x_{..}^2}{N} = \text{Somme des Carrés des Ecarts dûe au facteur.}$$

$$SCE_r = \sum_{i,j} x_{ij}^2 - \sum \frac{x_{i.}^2}{n_i} = \text{Somme des Carrés des Ecarts résiduelle.}$$

$$SCE_t = \sum_{i,j} (x_{ij} - \bar{x})^2 = \sum_{i,j} x_{ij}^2 - \frac{x_{..}^2}{N} = \text{Somme des Carrés des Ecarts totale.}$$

La fameuse équation de l'analyse de la variance : **SCE<sub>t</sub> = SCE<sub>fa</sub> + SCE<sub>r</sub>**

**Exercice 1**

Nous souhaitons comparer trois traitements, notés **A**, **B** et **C** contre l'asthme. Nous répartissons par tirage au sort les patients venant consulter dans un centre de soin en leur affectant l'un des trois traitements. Nous mesurons sur chaque patient la durée, en jours, le séparant de la prochaine crise d'asthme.

Les mesures sont reportées dans le tableau ci-dessous :

Traitement A	Traitement B	Traitement C
26. 27. 35. 36. 38.	29. 42. 42. 44. 45.	26. 26. 30. 30. 33.
38. 41. 42. 45. 50.	48. 48. 52. 56. 56.	36. 38. 38. 39. 46.
65	58. 58. 60. 61. 63.	47. 51. 51. 56. 75
	63. 69	

On suppose que les conditions de réaliser un test de comparaison de moyennes sont vérifiées.

Pouvons-nous conclure que les traitements ont une efficacité différente pour le critère : « temps séparant une crise à la prochaine (la crise suivante) ».

**Solution :**

- a) Après avoir calculé les variances de 3 échantillons, on teste l'égalité des deux variances (la plus grande et la plus petite en utilisant le test F) :

	<i>Traitement C</i>	<i>Traitement B</i>
Moyenne	$\bar{x}_C = 41,46666667$	$\bar{x}_B = 52,58823529$
Variance	$s_C^2 = 174,4095238$	$s_B^2 = 103,0073529$
Observations	$n_C = 15$	$n_B = 17$
Degré de liberté	$v_C = n_C - 1 = 14$	$v_B = n_B - 1 = 16$
$F = \frac{s_C^2}{s_B^2} = 1.693175476$	1,693175476	
P ( $F \leq f$ ) unilatéral	0,155668941	
Valeur critique pour F (unilatéral)	2,373318231	

**Interprétation du test :**

$H_0 = \{ \text{Les variances sont identiques} \}$  ; contre

$H_1 = \{ \text{Les variances ne sont pas identiques} \}$ .

Etant donné que la p-valeur calculée ( $0.155668941 \approx 15.57 \% > \alpha = 5 \%$ ) est supérieure au niveau de signification seuil  $\alpha = 0.05$ , on ne peut pas rejeter l'hypothèse nulle  $H_0$ , donc il y a égalité des variances.

- b) Analyse de la variance : un facteur

- Traitement : facteur à 3 niveaux (variable qualitative indépendante).
- Temps en jours (entre le début de l'étude et la prochaine crise d'asthme) : variable quantitative expliquée dépendante.

Statistiques descriptives :

Groupes	Nb échant	Somme	Moyenne	Variance	Ecart-type
Trait_A	$n_1. = 11$	$x_1. = 443$	$\bar{x}_1. = 40,27272727$	116,8181818	10.808246
Trait_B	$n_2. = 17$	$x_2. = 894$	$\bar{x}_2. = 52,58823529$	103,0073529	10.149254
Trait_C	$n_3. = 15$	$x_3. = 622$	$\bar{x}_3. = 41,46666667$	174,4095238	13.206420
$\Sigma$	$N = 43$	$x_{..} = 1959$	$\bar{\bar{x}} = 45.55813953$		

ANOVA : un facteur

Source	SS = SCE	DDL	MS = CM	F	P-Valeur	F critique
Inter groupes	1398,571853	2	699,285926	5,31975325	0,0089415	3,23172699
Intra groupes	5258,032799	40	131,45082			
Total	6656,604651	42				

**Interprétation du test :**

$H_0$  : moyennes égales ; contre  $H_1$  : Il existe au moins une des trois qui est différente.

Etant donné que la p-valeur calculée (0,0089415) est inférieure au niveau de signification seuil  $\alpha = 0.05$ , on doit rejeter l'hypothèse nulle  $H_0$ .

**Conclusion :** Il y a une différence significative entre les moyennes donc les traitements ont une efficacité différente.

### Utilisation manuelle des formules de l'analyse de la variance vues au cours :

$$\text{SCE}_{\text{fa}} = \sum n_i (\bar{X}_i - \bar{x})^2 = 11(40.27273 - 45.55814)^2 + 17(52.58824 - 45.55814)^2 + 15(41.46667 - 45.55814)^2 = \mathbf{1398.57185}$$

$$\text{SCE}_{\text{fa}} = \sum \frac{x_i^2}{n_i} - \frac{x_{..}^2}{N} = \frac{443^2}{11} + \frac{894^2}{17} + \frac{622^2}{15} - \frac{1959^2}{43} = 90\,646.9672 - 89\,248.39535 = \mathbf{1398.57185}$$

$$\text{SCE}_{\text{r}} = \sum_{i,j} x_{ij}^2 - \sum \frac{x_{i.}^2}{n_i} = 95\,905 - \left( \frac{443^2}{11} + \frac{770^2}{15} + \frac{491^2}{13} \right) = 95\,905 - 90\,646.9672 = \mathbf{5258.0328}$$

$$\text{SCE}_{\text{t}} = \sum_{i,j} (x_{ij} - \bar{x})^2 = \sum_{i,j} x_{ij}^2 - \frac{x_{..}^2}{N} = 95\,905 - 89\,248.39535 = \mathbf{6656.60465}$$

La fameuse équation de l'analyse de la variance :

$$\text{SCE}_{\text{t}} = \text{SCE}_{\text{fa}} + \text{SCE}_{\text{r}} = \mathbf{1398.57185} + \mathbf{5258.03280} = \mathbf{6656.60465}$$

### Exercice 2

On étudie l'activité d'un enzyme sérique, noté PDE, en fonction de différents facteurs dans l'espèce humaine. Les résultats sont exprimés en unités internationales par litres de sérum. On admettra que les populations considérées sont gaussiennes.

a) Chez deux groupes de femmes, enceintes et non enceintes, on obtient les résultats suivants :

Enceintes                    4.2 ; 5.5 ; 4.6 ; 5.4 ; 3.9 ; 5.4 ; 2.7 ; 3.9 ; 4.1 ; 4.1 ; 4.6 ; 3.9 ; 3.5  
Non Enceintes            1.5 ; 1.6 ; 1.4 ; 2.9 ; 2.2 ; 1.8 ; 2.7 ; 1.9 ; 2.2 ; 2.8 ; 2.1 ; 1.8 ; 3.7 ; 1.8 ; 3.1

La grossesse a-t-elle une influence significative ou hautement significative sur l'activité de la PDE ?

b) Afin d'évaluer la précocité de l'augmentation d'activité enzymatique lors de la grossesse, on pratique les dosages chez des femmes enceintes à différentes semaines d'aménorrhée. On suppose que les conditions de validité du test sont satisfaites et de ce fait, on obtient sur des échantillons indépendants les résultats suivants :

4 semaines	5 semaines	6 semaines	7 semaines	8 semaines
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
4.5	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 ? (On suppose que les conditions de validité du test sont satisfaites).

### Solution :

La grossesse a-t-elle une influence significative (où  $\alpha = 5\%$ ) ou hautement significative (où  $\alpha = 1\%$ ) sur l'activité de la PDE ?

a) L'égalité des variances est satisfaite par hypothèse pour pouvoir comparer les moyennes.

**Test de comparaison de deux moyennes**

	<i>Enceintes</i>	<i>Non Enceintes</i>
Moyenne	$\bar{x}_1 = 4,292307692$	$\bar{x}_2 = 2,233333333$
Variance $s^2$	$s_1^2 = 0,650769231$	$s_2^2 = 0,443809524$
Observations	$n_1 = 13$	$n_2 = 15$
Variance pondérée	0,539329389	
Différence hypothétique des moyennes	0	
Degré de liberté	$n_1 + n_2 - 2 = 26$	
Statistique de test observé	7,398815288	
P ( $T \leq t$ ) unilatéral	3,70567E-08	
Valeur critique de t (unilatéral)	1,70561792	
P ( $T \leq t$ ) bilatéral	7,41133E-08	
Valeur critique de t (bilatéral)	2,055529439	

**Interprétation du test :**

$H_0$  : les moyennes égales ; contre  $H_1$  : les moyennes ne sont pas égales.

Etant donné que la p-valeur calculée (presque nulle) est inférieure au niveau de signification seuil  $\alpha = 0.05$ , on ne peut pas accepter l'hypothèse nulle  $H_0$ .

**Conclusion** : Il y a une différence significative entre les moyennes ; c'est-à-dire que la grossesse a une influence *hautement significative* sur l'activité de la PDE.

Analyse de variance : un facteur						
RAPPORT DÉTAILLÉ						
Groupes	Nombre d'éch	Somme $\sum x$	Moyenne $\bar{x}$	Variance $s^2$		
Enceintes	13	55,8	4,29230769	0,65076923		
Non-Enceintes	15	33,5	2,23333333	0,44380952		
ANALYSE DE VARIANCE						
Source	SCE	DDL	Carrés Moyens	F	Probabilité	F-critique
Entre Groupes	29,5242216	1	29,5242216	54,7424677	7,4113E-08	4,22520127
Intra Groupes	14,0225641	26	0,53932939			
Total	43,5467857	27				

b) Evaluation de la précocité de l'augmentation d'activité enzymatique lors de la grossesse : test sur les moyennes en utilisant ANOVA.

**Rapport détaillé :**

Groupes	Obs	Somme	Moyenne	Variance
4_semaine	12	$x_1 = 61,1$	5,091666667	1,420833333
5_semaine	12	$x_2 = 62,5$	5,208333333	0,849924242
6_semaine	12	$x_3 = 83,2$	6,933333333	2,495151515
7_semaine	12	$x_4 = 69,2$	5,766666667	1,742424242
8_semaine	12	$x_5 = 89,7$	7,475	6,522045455
Total	60	$x_{..} = 365.7$	$\bar{\bar{x}} = 6.095$	

Analyse de la variance : un facteur

- Âge de grossesse : facteur à 5 niveaux (variable qualitative indépendante).

- Dosage de l'enzyme : variable quantitative expliquée dépendante.

Source	SS=SCE	DDL	MS=CM	F	P-Valeur	F critique
Inter groupes : SCE <sub>fa</sub>	54,0943333	4	13,52358333	5,189251807	0,001279507	2,53968863
Intra groupes : SCE <sub>r</sub>	143,334167	55	2,606075758			
Total : SCE <sub>t</sub>	197,4285	59				

Interprétation du test :

H<sub>0</sub> : les moyennes égales ; contre H<sub>1</sub> : les moyennes ne sont pas égales.

Etant donné que la p-value calculée (presque nulle = 0.13 %) est inférieure au niveau de signification seuil  $\alpha = 0.05 = 5\%$ , on ne peut pas accepter l'hypothèse nulle H<sub>0</sub>.

Conclusion : Il y a une différence significative entre les moyennes c'est-à-dire qu'il y a une influence de la grossesse sur l'activité de l'enzyme.

Utilisation manuelle des formules de l'analyse de la variance vues au cours :

$$SCE_{fa} = \sum n_i (\bar{X}_i - \bar{x})^2 = 12\{(5.0917 - 6.095)^2 + (5.2083 - 6.095)^2 + (6.9333 - 6.095)^2 + (5.7667 - 6.095)^2 + (7.475 - 6.095)^2\} = \mathbf{54.094333}$$

$$SCE_{fa} = \sum \frac{x_i^2}{n_i} - \frac{x^2}{N} = \frac{61.1^2 + 62.5^2 + 83.2^2 + 69.2^2 + 89.7^2}{12} - \frac{365.7^2}{60} = 2283.035833 - 2228.9415 = \mathbf{54.094333}$$

$$SCE_r = \sum_{i,j} x_{ij}^2 - \sum \frac{x_i^2}{n_i} = 2426.37 - \left( \frac{61.1^2 + 62.5^2 + 83.2^2 + 69.2^2 + 89.7^2}{12} \right) = 2426.37 - 2283.035833 = \mathbf{143.334167}$$

$$SCE_t = \sum_{i,j} (x_{ij} - \bar{x})^2 = \sum_{i,j} x_{ij}^2 - \frac{x^2}{N} = 2426.37 - 2228.9415 = \mathbf{197.4285}$$

L'équation de l'analyse de la variance :

$$SCE_t = SCE_{fa} + SCE_r = \mathbf{54.094333 + 143.334167 = 197.4285}$$