

A decorative graphic on the right side of the page features three blue, 3D-style circles of varying sizes. The largest circle is at the bottom right, a medium one is at the top center, and a smaller one is in the middle. Thin blue lines connect the top-left corners of these circles, forming a triangular shape that points towards the top left.

Correction du contrôle N° :01

Module : ANALYSE DES DONNEES

Année Universitaire 2021-2022

Chala Adel
12/01/2022

Examen N° :01

Exercice 1

Une étude est réalisée en vue de comparer l'efficacité de deux fertilisants sur la croissance des plantes.

Partie N° :01

On mesure l'hauteur de deux lots de plantes différentes, chacun avec un fertilisant différent. Bien sûr, nous avons cultivé la même espèce dans des conditions environnementales identiques (ensoleillement, apports d'eau, température...). Les données relevées sont les suivantes:

Fertilisant I			Fertilisant II		
48	52,0	47	52,3	58,0	53,2
51	55,0	52,4	57,4	57,8	53,3
58	57,1	55,6	55,6	54,8	55
51	49,0	57,5	55,0	54,0	56
50	53	49	52	58	53
57	48,0	52	55	59	54
55	50	47	49,8	53,7	55,4
50,50	50,20	52,0	52,5	56,8	54,9
52,10	55,50	55,0	57,2	59,6	51,7

Table N° :01

Après introduire ces données dans la base du Logiciel SPSS 20, nous obtenons les résultats indiqués dans les tableaux suivants :

Récapitulatif du traitement des observations							
	fertilisant	Observations					
		Valide		Manquante		Total	
		N	Pourcent	N	Pourcent	N	Pourcent
hauteur	Fert1	27	100,0%	0	0,0%	27	100,0%
	Fert2	27	100,0%	0	0,0%	27	100,0%

Table N° :02

Descriptives ^a				
	Fertilisant	Statistique	Erreur standard	
hauteur	Fert1	Moyenne	52,2185	0,63699
		Médiane	52,0000	
		Variance	10,955	
		Ecart-type	3,30990	

		Intervalle	11,00	
		Intervalle interquartile	5,00	
		Aplatissement	-1,029	0,872
	Fert2	Moyenne	55,0000	0,46639
		Médiane	55,0000	
		Variance	5,873	
		Ecart-type	2,42344	
		Intervalle	9,80	
		Intervalle interquartile	4,00	
		Aplatissement	-0,488	0,872

Table N° :03

Tests de normalité ^a							
	fertilisant	Kolmogorov-Smirnov ^b			Shapiro-Wilk		
		Statistique	Ddl	Signification	Statistique	ddl	Signification
hanteur	Fert1	0,133	27	0,200*	0,951	27	0,221
	Fert2	0,093	27	0,200*	0,983	27	0,918

Table N° :04

Test d'homogénéité de la variance ^a					
		Statistique de Levene	ddl1	ddl2	Signification
hanteur	Basé sur la moyenne	3,280	1	52	0,076
	Basé sur la médiane	2,890	1	52	0,095
	Basé sur la médiane et avec ddl ajusté	2,890	1	48,784	0,095
	Basé sur la moyenne tronquée	3,213	1	52	0,079

Table N° :05

- 1) Déterminer l'objectif pour cette expérience et cela à partir des résultats obtenues dans les tableaux précédentes (table N° :1, table N° :2, table N° :3, table N° :4, table N° :5).
- 2) Déterminer la variable qualitative qui exprime les deux échantillons, et la variable quantitative a mesuré.
- 3) Déterminer l'hypothèse nulle et alternative pour la normalité et pour l'homogénéité de la distribution des données.
- 4) Avec un risque de signification de 3%, que peut-on dire pour la normalité? Et sur l'homogénéité de la distribution des données?

Partie N° :02

D'après les questions du partie N° :01, si on accepte que les données sont normalement distribuées, ainsi que les données sont homogènes pour la variance.

Alors on veut faire la comparaison entre les deux fertilisantes et d'établir qu'elle est la meilleur fertilisante, pour cela on doit introduire les données dans la table N° :01 sous logiciel SPSS 20, et on obtient les résultats suivantes :

		Test-t pour égalité des moyennes						
		T	ddl	Sig. (bilatérale)	Différence moyenne	Différence écart-type	Intervalle de confiance 97% de la différence	
							Inférieure	Supérieure
hanteur	Hypothèse de variances égales	-3,523	52	0,001	-2,78148	0,78948	-4,54302	-1,01994
	Hypothèse de variances inégales	-3,523	47,654	0,001	-2,78148	0,78948	-4,54756	-1,01540

Table N° :06

- 1) Déterminer les conditions des validations de test de Student.
- 2) Déterminer l'objectif pour cette expérience et cela à partir des résultats obtenues dans les tableaux précédentes (table N° :06).
- 3) Déterminer la variable qualitative qui exprime les deux échantillons, et la variable quantitative a mesuré.
- 4) Déterminer l'hypothèse nulle et alternative la comparaison entre les deux moyennes des deux fertilisantes.
- 5) Dans la table N° :06, Donner la formule explicite qui calcul la valeur du statistique (T), de plus faire une application numérique pour obtenir la valeur $T = -3,523$.
- 6) Avec un risque de signification de 3%, que peut-on dire sur l'homogénéité des deux type des fertilisantes?
- 7) Avec un risque de signification de 3%, que peut-on dire sur l'efficacité des deux type des fertilisantes?

Exercice N° :02

Une étude est réalisée en vue de comparer l'efficacité de deux fertilisants sur la croissance des plantes. On mesure l'hanteur de deux lots de plantes différentes, chacun avec un fertilisant différent. Bien sûr, nous avons cultivé la même espèce dans des conditions environnementales identiques (ensoleillement, apports d'eau, température...). Les données relevées sont les suivantes:

Fertilisant I			Fertilisant II		
48	52,0	47	52,3	58,0	53,2
51	55,0	52,4	57,4	57,8	53,3
58	57,1	55,6	55,6	54,8	55
51	49,0	57,5	55,0	54,0	56
50	53	49	52		

Avec un seuil de signification 05%, Nous désirons savoir s'il existe une différence significative entre les deux types de fertilisants.

Exercice N° :03

Dans un groupe de 200 malades atteints du cancer du col de l'utérus, un traitement par application locale du radium a donné 50 guérisons, et les autres non guéris.

Un autre groupe de 150 sujets atteints de la même maladie a été traité par chirurgie, on a trouvé 54 guérisons.

Que peut-on dire sur la liaison entre le type de traitement et les guérisons (Oui et Non)? On prendra un risque $\alpha=5\%$.

Indication :

$T_{26}^{0,05} = 2,056$	$\chi_2^{0,05} = 5,99$
$T_{20}^{0,05} = 2,086$	$\chi_{10}^{0,05} = 18,31$
$T_{15}^{0,05} = 2,131$	$\chi_1^{0,05} = 3,84$
$T_{13}^{0,05} = 2,160$	$\chi_4^{0,05} = 9,49$

Correction

Exercice 1

Une étude est réalisée en vue de comparer l'efficacité de deux fertilisants sur la croissance des plantes.

Partie N° :01

- 1) Notre objectif est de valider la condition du test de Student (test d'homogénéité des distributions, et test de normalité).
- 2) La variable qualitative c'est : Fertilisante. Variable quantitative c'est Hauteur.
- 3) **Pour l'homogénéité :**

H0 : (La distribution est homogène),

H1 : (La distribution n'est pas homogène).

Pour la normalité :

H0 : (La distribution est gaussienne),

H1 : (La distribution n'est pas gaussienne).

- 4) Lorsque $n=27 < 30$, le cas des petits échantillons, alors on utilise test de Shapiro-Wilk :

Pour la Fert 1 : $\text{Sig}=0,221 > 0,03$; alors on accepte de H0, d'où la normalité de la distribution pour l'échantillon « Fert 1 ».

Pour la Fert 2 : $\text{Sig}=0,918 > 0,03$; alors on accepte de H0, d'où la normalité de la distribution pour l'échantillon « Fert 2 ».

Pour le test d'homogénéité de la variance, on utilise seulement qui est basé sur la moyenne :

$\text{Sig}=0,076 > 0,03$; alors on accepte de H0, d'où l'homogénéité de la distribution pour l'échantillon total.

Partie N° :01

- 1) Condition du validation de test de Student :
 - a) La variable mesuré c'est quantitative, et une variable qualitative avec deux exactement modalités.
 - b) La distribution gaussienne.
 - c) La distribution homogène
- 2) Notre objectif est de faire une comparaison entre l'effet de deux fertilisantes.
- 3) La variable qualitative c'est : Fertilisante. Variable quantitative c'est Hauteur.
- 4) Hypothèse nulle : H0 : (IL n' y a pas une différence significative entre les deux fertilisantes.
 $\bar{x}_1 = \bar{x}_2$.
Hypothèse alternative : H1 : (Il y a une différence significative entre les deux fertilisantes. $\bar{x}_1 \neq \bar{x}_2$.
- 5) La formule et application numérique

$$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{52,218 - 55,000}{\sqrt{\frac{10,955}{27} + \frac{5,873}{27}}} = -3,522.$$

- 6) $\text{Sig}=0,001 < 0,03$; Alors on accepte de H1, alors la fertilisante a des effet différentes significativement.

- 7) $\Delta\bar{x} \in [-4,543; -1,019]$, alors les deux bornes d'intervalle de la différence des moyennes sont négatives, ce qui signifie que $\Delta\bar{x} < 0$, alors $\bar{x}_1 - \bar{x}_2 < 0$, alors $\bar{x}_1 < \bar{x}_2$, alors la deuxième type de fertilisant est plus efficace.

Exercice 2

Proposition d'hypothèses :

Hypothèse nulle : $H_0 =$ (Il n y a pas une différence significative entre les deux types de fertilisantes).

Hypothèse alternative : $H_1 =$ (Il y a une différence significative entre les deux types de fertilisantes).

Les Calculs :

$n_1 = 15 < 30$	$n_2 = 13 < 30$
$\bar{x}_1 = 52,373$	$\bar{x}_2 = 54,953$
$\sigma_1 = 3,568$	$\sigma_2 = 1,987$
$\sigma_1^2 = 12,730$	$\sigma_2^2 = 3,948$

La variance pondérée

$$S^2 = \frac{n_1\sigma_1^2 + n_2\sigma_2^2}{n_1 + n_2 - 2} = \frac{15 \times 12,730 + 13 \times 3,948}{15 + 13 - 2} = 9,318.$$

Alors $S = \sqrt{S^2} = \sqrt{9,318} = 3,052$.

$$T_{obs} = \frac{|\bar{x}_1 - \bar{x}_2|}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{|52,373 - 54,953|}{3,052 \sqrt{\frac{1}{15} + \frac{1}{13}}} = 2,243.$$

Conclusion :

On lit dans la table du Student la valeur $T_\alpha = T_{n_1+n_2-2} = T_{26} = 2,056$

$T_{obs} = 2,243 > 2,056$; Alors on accepte de H_1 , alors la fertilisante a des effets différentes significativement.

Exercice 3

Proposition d'hypothèses :

Hypothèse nulle : $H_0 =$ (Les variables qualitatives sont indépendantes).

Hypothèse alternative : $H_1 =$ (Les variables qualitatives sont liées).

Les Calculs :

Tableau de contingence

		Guérie	Non Guérie	Total
T1	Effectifs observés	50	150	200
	Effectifs calculés	$0,297 \times 200 = 59,400$	$0,702 \times 150 = 140,400$	
T2	Effectifs observés	54	96	150
	Effectifs calculés	$0,297 \times 150 = 44,55$	$0,702 \times 150 = 105,300$	
Total	Total	104	246	350

	pourcentage	$\frac{104}{350} = 0,297.$	$\frac{246}{350} = 0,702.$	1
--	-------------	----------------------------	----------------------------	---

$$\chi_{obs}^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(O_{ij} - C_{ij})^2}{C_{ij}} = \frac{(50 - 59,4)^2}{59,4} + \frac{(150 - 140,4)^2}{140,4} + \frac{(54 - 44,55)^2}{44,55} + \frac{(96 - 105,3)^2}{105,3}$$

=4,968.

Conclusion et décision :

On lit dans la table du Khy-deux la valeur $\chi_{\alpha}^2 = \chi_{(2-1)(2-1)}^2 = \chi_1^2 = 3,84.$

Alors $\chi_{obs}^2 = 4,968 > \chi_{\alpha}^2 = 3,84.$

Alors on rejette H0 et on accepte H1, d'où la liaison entre les deux variables qualitatives (type de traitement et la guérison).