Université 08Mai 1945 Guelma
1 Licence Sciences Alimentaires (Procédés agroalimentaires) (1 COFFEE)
STATISTIQUES APPLIQUEES

Corrigé-type de Série N°5 de Statistiques Appliques

(Etude de corrélation et Régression 2)

Exercice 1

On veut savoir si l'addition de substances adjuvantes à un vaccin modifie la production d'anticorps. Pour cela, on mesure les quantités d'anticorps produites par des sujets après administration de quantités égales du vaccin, additionnées ou non d'une substance adjuvante. On obtient les taux dans le tableau suivant :

Sans substance adjuvante	Avec de l'alumine	Avec des phosphates
1, 3, 3, 0, 1	2, 4, 5, 4, 3, 6	1, 4, 2, 3, 3

a) Quelles hypothèses faut-il faire pour appliquer la technique de l'analyse de la variance à la résolution du problème posé ?
b) Sous les hypothèses adéquates, tester l'hypothèse selon laquelle les populations dont sont extraits les 3 échantillons ont la même variance.
c) En précisant toujours les hypothèses adéquates, l'efficacité du vaccin dépend-elle
   i) De la présence de substances adjuvantes ?
   ii) De leur nature ?

Solution :

a) Les échantillons doivent être indépendants, gaussiens et avoir des variances égales.

Statistiques descriptives des données :

IC(μ) à 95 % = $[\mathbf{x}_{i.} \pm t_{(1-\alpha/2\,;\,N-a)} \sqrt{\frac{CM_E}{n_i}}] = [1.6 \pm 2.1604 \sqrt{\frac{1,723076923}{5}}] = [0.33176 ; 2.86824]$

Variable	Obs.	Moy $\bar{x}_{i.}$	IC(μ) à 95 %	Som	$(\mathbf{x}_{i.})^2$	$(\mathbf{x}_{ij})^2$	Ecart-type s	Variance
Sans adjuvant	$n_1 = 5$	$\bar{x}_{1.} = 1,6$	[0.332 ; 2.868]	$\mathbf{x}_{1.} = 8$	64	20	$s_1 = 1,341640786$	$s_1^2 = 1.8$
Alumine	$n_2 = 6$	$\bar{x}_{2.} = 4$	[2.842 ; 5.158]	$\mathbf{x}_{2.} = 24$	576	106	$s_2 = 1,414213562$	$s_2^2 = 2$
Phosphate	$n_3 = 5$	$\bar{x}_{3.} = 2,6$	[1.332 ; 3.868]	$\mathbf{x}_{3.} = 13$	169	39	$s_3 = 1,140175425$	$s_3^2 = 1.3$
x_{ij} globale	n = 16	$\bar{\bar{x}} = 2.8125$		$\mathbf{x}_{..} = 45$		165		
Ecart-type	Regroupé	*= 1,31266*	$\Rightarrow CM_{erreur} =$	$1,31266^2 =$	*1.723*			

b) Le test d'hypothèse selon lequel les populations dont sont extraits les 3 échantillons ont la même variance.

Test unilatéral *d'égalité des variances* avec le logiciel EXCEL

On considère la plus grande variance ($S_A^2 = 2$) et la plus petite variance ($S_P^2 = 1.3$) :

H_0 = {Les variances sont identiques} = $\{\sigma_1^2 = \sigma_2^2\}$; contre **H_1** = $\{\sigma_2^2 < \sigma_1^2\}$.

	Alumine	Phosphate	
Moyenne	4	2,6	Seuil critique : $F_{5\,;\,4\,;\,0.95} = 6.2561$
Variance	$S^2_A = 2$	$S^2_P = 1{,}3$	Si $F_{obs} > F_{5\,;\,4\,;\,0.95}$ on rejette H_0.
Observations	6	5	On obtient : $F \approx 1.54 < F_{5\,;\,4\,;\,0.95} \approx 6.26$
Degré de liberté : ddl	5	4	Décision : On ne peut pas rejeter H_0.
$F_{obs} = 2/1.3$	1,5385		C'est-à-dire les variances sont égales.
P (F ≤ f) unilatéral	0,3487		
F critique (unilatéral)	6,2561		

Conclusion pratique : les variances sont égales ; et donc on peut procéder à la comparaison des moyennes de ces 3 échantillons à l'aide d'un test d'analyse de variance (ANOVA).

c) *Test d'égalité des moyennes* avec le tableau ANOVA à un facteur ; ***méthode EXCEL***

- Substance adjuvante : c'est le facteur à 3 niveaux (variable indépendante qualitative).
- Quantité d'anticorps : c'est la variable expliquée quantitative dépendante.

Tableau de l'ANOVA d'un facteur :

Source de variabilité	*SCE*	*ddl*	*Carré moyen CM*	F_{obs}	*P-Value*	*F critique*
Inter groupes = SCE_{fa}	16,0375	2	8,01875	4,65374	0,0299	3,80556525
Intra groupes = SCE_r	22,4	13	1,723076923	----------	----------	---------------
Total = SCE_t	38,4375	15	--------------			

$CM = \frac{SCE}{ddl}$ $F_{obs} = \frac{CMfa}{CMr}$ F critique = $F_{2\,;\,13\,;\,0.95} = 3.8056$

Interprétation du test :

H_0 : les 3 moyennes égales ; contre **H_1** : Au moins l'une des moyennes est différente.

Comme la statistique de test observé $F_{obs} = 4.65 >$ F critique alors on rejette l'hypothèse nulle H_0.

Conclusion : les moyennes ne sont pas égales ; c'est-à-dire la présence ou non de la substance influe sur l'efficacité du vaccin.

MINITAB 18 : Comparaisons deux à deux de Tukey

Informations de groupement avec la méthode de Tukey et un niveau de confiance de 95 %

Facteur	N	Moyenne	Groupement	
Alumine	6	4,000	A	
Phosphate	5	2,600	A	B
Sans	5	1,600		B

Les moyennes ne partageant aucune lettre sont significativement différentes.

Il y a lieu de considérer la 2ème partie de la troisième question du problème pour tester s'il y a égalité des deux moyennes concernant l'ajout des adjuvants (alumine et phosphate).

2ème partie question c)

Test de comparaison de deux moyennes (alumine μ_1 et phosphate μ_2).

1) *Formulation des hypothèses* :

   H_0 = $\{\mu_1 - \mu_2 = 0\}$ contre **H_1** = $\{\mu_1 - \mu_2 \neq 0\}$ (Egalité contre inégalité)

2) *Le seuil critique* : $t_{(0.975\,;\,9)} = 2.2622$ si $T_o \in [t_{(0.025\,;\,9)}\,;\, + t_{(0.975\,;\,9)}] = [-2.262\,;\,+20262] \Rightarrow$ on ne peut pas rejeter **H_0**.

3) *Calcul de la statistique de test observé To.*

	Alumine	*Phosphate*
Moyenne	$\bar{x}_1 = 4$	$\bar{x}_2 = 2,6$
Variance	$s_1^2 = 2$	$s_2^2 = 1,3$
Observations	$n_1 = 6$	$n_2 = 5$
Variance pondérée	1,68888889	
Différence hypothétique des moyennes	0	
Degré de liberté : ddl	9	
T_o = Statistique de test observée t	1,77906486	
$P(T \leq t)$ unilatéral	0,05446847	
Valeur critique de t (unilatéral) = $t_{(0.95\,;\,9)}$	1,83311293	
$P(T \leq t)$ bilatéral	0,10893695	
Valeur critique de t (bilatéral) = $t_{(0.975\,;\,9)}$	2,26215716	

1) *Décision du test* :

Comme la statistique de test observé $T_o = 1.779$ appartient à la zone d'acceptation de H_0, alors on ne doit pas rejeter l'hypothèse nulle H_0.

Conclusion pratique : les moyennes des adjuvants (alumine et phosphate) sont égales ; c'est-à-dire la nature de la substance ajoutée n'influe pas sur l'efficacité du vaccin.

Exercice 2

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

a) Tester l'égalité des variances.
b) Pouvons-nous conclure que les traitements ont une efficacité différente pour le critère « temps séparant une crise à la prochaine ».

Solution :

Statistiques descriptives des données : **n = 43** **$\bar{\bar{x}}$ = 45.55813953**

Groupes	Obs	$x_{i.}$	$(x_{i.})^2$	$(x_{ij})^2$	$\bar{x}_{i.}$	IC(μ) à 95 %	Variance
Trait_A	11	$x_{1.} = 443$	196249	19009	$\bar{x}_a = 40,27272727$	[33.286 ; 47.259]	$s_a^2 = 116,8181818$
Trait_B	17	$x_{2.} = 894$	799236	48662	$\bar{x}_b = 52,58823529$	[46.968 ; 58.208]	$s_b^2 = 103,0073529$
Trait_C	15	$x_{3.} = 622$	386884	28234	$\bar{x}_c = 41,46666667$	[35.484 ; 47.450]	$s_c^2 = 174,4095238$
Groupé	**43**	$x_{..} = 1959$		**95905**	**$\bar{\bar{x}} = 45.55813953$**		
Ecart type	*regroupé*	*= 11,4652*	*$\Rightarrow$ CMe=*	*11.4652^2*	*= 131.451*		

a) Test sur l'égalité des variances

Après avoir calculé les variances des 3 échantillons, on teste l'égalité des deux variances (la plus grande variance (174.41) et la plus petite variance (103.01) en utilisant le test F de EXCEL 2019.

Formulation des hypothèses du test :

H_0 = {Les variances sont identiques} ; contre

H_1 = {Les variances ne sont pas identiques} ; Excel fait un test unilatéral.

Seuil critique : $\alpha = 0,05 = P(\text{rejet de } H_0 \mid H_0 \text{ vraie}) \Rightarrow F_{14\,;\,16\,;\,0.95} = 2,373318231$.

Si le rapport des variances $F_{obs} = \frac{174.41}{103.01} \approx 1.693 < F_{14\,;\,16\,;\,0.95} \approx 2,3733$ on ne peut pas rejeter H_0.

Calcul de la statistique de test observé : voir tableau ci-dessous de Excel 2019.

	Trait_C	*Trait_B*
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 = ddl	14	16
$F_{obs} \approx 174.41/103.01$	1,693175476	
$P(F \leq f)$ unilatéral	0,155668941	
$F_{14\,;\,16\,;\,0.95}$ (unilatéral)	2,373318231	

Décision :

1) Comme $F_{obs} < F_{14\,;\,16\,;\,0.95}$ alors on ne peut pas rejeter H_0 et les variances sont égales.
2) Etant donné que la p-value = 0.155668941 calculée par Excel 2019 est supérieure au niveau de signification seuil $\alpha = 0.05$ alors on ne peut pas rejeter l'hypothèse nulle H_0, donc il y a égalité des variances.

b) Analyse de la variance : un facteur (utilisation d'EXCEL 2019)

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

ANOVA : un facteur

Source de variabilité	*SCE*	*DDL*	*CM*	*F_{obs}*	*P-Valeur*	*$F_{2\,;\,40\,;\,0.95} = F_{critique}$*
Inter groupes = SCE_{fa}	1398,571853	2	699,285926	5,31975325	0,0089415	3,23172699
Intra groupes = SCE_r	5258,032799	40	131,45082			
Total = SCEt	6656,604651	42				

Interprétation du test :

H_0 : moyennes égales ; contre **H_1** : Au moins l'une des trois est différente d'une autre.

Le risque étant $\alpha = 0.05 = P(\text{rejet de } H_0 \mid H_0 \text{ vraie}) \Rightarrow$ Si le rapport F_{obs} = 5.32 est supérieur au seuil critique $\mathbf{F_{2\,;\,40\,;\,0.95} = 3.23}$ alors on rejette H_0.

Décision :

Comme $\mathbf{F_{obs} > F_{2\,;\,40\,;\,0.95}}$ alors on rejette H_0.