

Section: Psychologie - Licence

Enseignant responsable : F.G. Carpentier

**CORRIGÉ DE L'ÉPREUVE PARTIELLE ET PONCTUELLE DE STATISTIQUES
 ET INFORMATIQUE APPLIQUÉES À LA PSYCHOLOGIE**

Exercice 1

Dans une expérience réalisée en 1992, M. Hupert, F. Nef et M. Maroy ont analysé et comparé le langage oral spontané chez des adultes jeunes et âgés.

Soixante adultes se sont prêtés à l'étude : 30 adultes jeunes (15 hommes et 15 femmes) dont l'âge moyen est de 24 ans et 30 adultes âgés (15 hommes et 15 femmes) dont l'âge moyen est de 70 ans.

1) Dans un premier temps, on se propose de comparer les fonctions cognitives des sujets des deux groupes. On soumet notamment les sujets à une épreuve de fluence orthographique : le sujet doit produire en une minute le plus grand nombre possible de mots différents commençant par la lettre "L".

Le tableau 1 ci-dessous est un extrait du tableau des scores observés sur cette épreuve. Dans ce tableau, les notations n , \bar{x} et s_c désignent respectivement l'effectif, la moyenne et l'écart type corrigé dans chacun des deux échantillons.

Les performances des sujets jeunes et âgés sont-elles significativement différentes sur cette épreuve? Répondre à cette question à l'aide d'un test de comparaison de moyennes, en utilisant un seuil de 5%.

Fluence orthographique		
	Jeunes	Agés
	9	11

	14	12
	11	13
n	30	30
\bar{x}	11.4	11.0
s_c	3.1	3.2

Tableau 1: Fluence orthographique

Il s'agit ici d'un test de comparaison de moyennes sur deux groupes indépendants. La taille de chacun des échantillons étant égale à 30, on pourra utiliser aussi bien le test pour "grands échantillons" que celui pour "petits échantillons". C'est ce dernier que nous utilisons ici. Notons μ_1 et μ_2 les moyennes sur les deux populations parentes.

Hypothèse nulle $H_0 : \mu_1 = \mu_2$

Hypothèse alternative $H_1 : \mu_1 \neq \mu_2$ (on fait ici un test bilatéral).

La statistique de test, $T = \frac{\bar{x}_1 - \bar{x}_2}{E}$, avec $E^2 = \frac{s_{c,1}^2}{n_1} + \frac{s_{c,2}^2}{n_2}$ suit une loi de Student à $30 + 30 - 2 = 58$ ddl.

Pour un seuil de 5%, la valeur critique déduite de la table est $t_c = 2.0017$. La règle de décision est donc :

- si $-2.0017 \leq t_{obs} \leq 2.0017$, on retient H_0 .
- si $t_{obs} < -2.0017$ ou $t_{obs} > 2.0017$, on rejette H_0 et on retient H_1 .

Avec les données observées, $E^2 = \frac{3.1^2}{30} + \frac{3.2^2}{30} = 0.6617$ d'où $E = 0.81$ et $t_{obs} = \frac{11.4 - 11.0}{0.81} = 0.4917$.

On retient donc l'hypothèse H_0 : on n'a pas mis en évidence de différence significative de la fluence verbale.

2) Dans un deuxième temps, on demande aux sujets de s'exprimer sur un thème donné et on analyse le discours correspondant aux 200 premiers mots énoncés par chaque sujet. On évalue notamment, pour chaque discours, le nombre moyen de propositions relatives.

Les résultats observés sont rassemblés dans la plage B2:C31 d'une feuille Excel.

On effectue ensuite à l'aide d'Excel les calculs suivants :

Résultats

	A	B	C
32	Moyenne	3.4	5.1
33	Ecart type corrigé	2.09	2.59
34	Test Student		0.70%

Formules correspondantes

	A	B	C
32	Moyenne	=MOYENNE(B2:B31)	=MOYENNE(C2:C31)
33	Ecart type corrigé	=ECARTYPE(B2:B31)	=ECARTYPE(C2:C31)
34	Test Student		=TEST.STUDENT(B2:B31;C2:C31;2;2)

N.B. : Signification des deux derniers paramètres de la fonction TEST.STUDENT :

- Troisième paramètre = 2 : test bilatéral ;

- Quatrième paramètre = 2 : deux échantillons indépendants de même variance.

a) Quelle interprétation peut-on donner du pourcentage (0.70%) obtenu en cellule C34 ?

Le pourcentage indiqué en cellule C34 est le niveau de significativité associé au test réalisé. Autrement dit, c'est la probabilité, sous H_0 , de tirer des échantillons au moins aussi "extrêmes" que ceux qui ont été observés. Plus formellement, on a, sous l'hypothèse H_0 : $P(|t| \geq |t_{obs}|) = 0.7\%$.

b) Quelles sont les hypothèses du test effectué et quelle conclusion peut-on en tirer ?

H_0 : Dans les populations parentes, égalité des moyennes des scores des sujets jeunes et âgés.

H_1 : Dans les populations, les moyennes des scores des sujets jeunes et âgés sont différentes (le test réalisé est un test bilatéral).

Si l'on choisit un seuil de 5% ou de 1%, on obtient un niveau de significativité inférieur au seuil.

On conclut donc sur H_1 : les scores des sujets varient selon l'âge.

Exercice 2

Dans une étude sur la recherche de différences liées au sexe chez le jeune enfant, B. Rogé et S. Ionescu ont observé 39 filles âgées de 26 mois en moyenne, en présence de chacun des deux parents. Ils ont noté le nombre de mimiques observées durant une séquence de 10 minutes, en présence de chacun des deux parents. Les résultats obtenus sont donnés dans le tableau 2.

	Mère	Père		Mère	Père		Mère	Père
s1	16	10	s14	4	0	s27	3	0
s2	12	2	s15	4	2	s28	3	1
s3	12	2	s16	3	2	s29	2	1
s4	8	9	s17	3	2	s30	1	0
s5	7	3	s18	3	2	s31	1	0
s6	7	0	s19	3	1	s32	1	2
s7	7	2	s20	3	2	s33	1	2
s8	6	2	s21	3	2	s34	0	2
s9	6	1	s22	3	1	s35	0	2
s10	5	5	s23	3	0	s36	0	1
s11	5	2	s24	3	8	s37	0	2
s12	4	2	s25	3	12	s38	0	1
s13	4	0	s26	3	8	s39	0	0

Tableau 2: Nombre de mimiques

1) A l'aide d'un test de comparaisons de moyennes, étudier si le nombre de mimiques est significativement différent selon le sexe du parent présent (seuil : 5%).

Il s'agit ici de comparer d'une comparaison de moyennes sur deux groupes appariés. Nous allons donc construire le protocole des différences individuelles (tableau 3).

	Mère	Père	d_i		Mère	Père	d_i		Mère	Père	d_i
s1	16	10	6	s14	4	0	4	s27	3	0	3
s2	12	2	10	s15	4	2	2	s28	3	1	2
s3	12	2	10	s16	3	2	1	s29	2	1	1
s4	8	9	-1	s17	3	2	1	s30	1	0	1
s5	7	3	4	s18	3	2	1	s31	1	0	1
s6	7	0	7	s19	3	1	2	s32	1	2	-1
s7	7	2	5	s20	3	2	1	s33	1	2	-1
s8	6	2	4	s21	3	2	1	s34	0	2	-2
s9	6	1	5	s22	3	1	2	s35	0	2	-2
s10	5	5	0	s23	3	0	3	s36	0	1	-1
s11	5	2	3	s24	3	8	-5	s37	0	2	-2
s12	4	2	2	s25	3	12	-9	s38	0	1	-1
s13	4	0	4	s26	3	8	-5	s39	0	0	0

Tableau 3: Nombre de mimiques

Pour calculer les paramètres de la série des différences individuelles, il sera commode de réaliser un tableau d'effectifs (tableau 4)

On obtient ainsi : $\bar{d} = \frac{56}{39} = 1.436$, $s^2 = \frac{602}{39} - 1.436^2 = 13.37$ et $s_c^2 = \frac{39}{38} s^2 = 13.73$.

Test proprement dit

Hypothèse nulle $H_0 : \mu_1 = \mu_2$ (dans la population dont est issu l'échantillon, le nombre moyen de mimiques est le même en présence de la mère et en présence du père)

Hypothèse alternative $H_1 : \mu_1 \neq \mu_2$ (on fait ici un test bilatéral).

Etant donné la taille de l'échantillon ($n = 39$), la statistique de test, $Z = \frac{\bar{d}}{E}$, avec $E^2 = \frac{s_c^2}{n}$ suit une loi normale centrée réduite.

d_i	n_i	$n_i d_i$	$n_i d_i^2$
10	2	20	200
7	1	7	49
6	1	6	36
5	2	10	50
4	4	16	64
3	3	9	27
2	5	10	20
1	8	8	8
0	2	0	0
-1	5	-5	5
-2	3	-6	12
-5	2	-10	50
-9	1	-9	81
Somme	39	56	602

Tableau 4: Différences individuelles

Pour un seuil de 5%, la valeur critique déduite de la table est $z_c = 1.96$. La règle de décision est donc :

– si $-1.96 \leq z_{obs} \leq 1.96$, on retient H_0 .

– si $z_{obs} < -1.96$ ou $z_{obs} > 1.96$, on rejette H_0 et on retient H_1 .

Avec les données observées, $E^2 = \frac{13.73}{39} = 0.3520$ d'où $E = 0.5933$ et $z_{obs} = \frac{1.436}{0.5933} = 2.42$.

On retient donc l'hypothèse H_1 : le nombre de mimiques dépend significativement du sexe du parent présent.

2) Reprendre la question précédente à l'aide d'un test du signe.

H_0 : Dans la population parente, la fréquence des différences positives est 50%.

H_1 : Dans la population parente, cette fréquence n'est pas 50%.

L'échantillon observé comporte $N = 37$ différences non nulles, réparties en $D_+ = 26$ différences positives et $D_- = 11$ différences négatives.

Compte-tenu de la taille de l'échantillon, on utilise ici l'approximation de la statistique de test par la loi normale.

$$D = \max(D_+, D_-) = 26, \quad z_{obs} = \frac{2D - 1 - N}{\sqrt{N}} = \frac{52 - 1 - 37}{\sqrt{37}} = 2.30$$

La règle de décision est la même qu'à la question précédente, et on conclut encore sur l'hypothèse alternative: dans la population parente, la fréquence des différences positives n'est pas 50%.

Exercice 3

Le test d'empan de chiffres de la WAIS mesure la longueur de la plus longue série de chiffres que le sujet peut reproduire dans un ordre sériel correct. On considère généralement cette mesure comme une bonne évaluation de la capacité de la mémoire à court terme.

Sur un échantillon de 100 sujets, on a observé un score moyen de 8.8 avec un écart type corrigé de 2.1. Donner une estimation du score moyen observé dans la population, à l'aide d'un intervalle de confiance, avec un degré de confiance de 95%.

Il s'agit ici d'un grand échantillon. On sait que, dans ces conditions, un intervalle de confiance pour la moyenne μ sur la population, avec le degré de confiance $1 - \alpha$, est donné par :

$$\bar{x} - z_\alpha \frac{s_c}{\sqrt{n}} \leq \mu \leq \bar{x} + z_\alpha \frac{s_c}{\sqrt{n}}$$

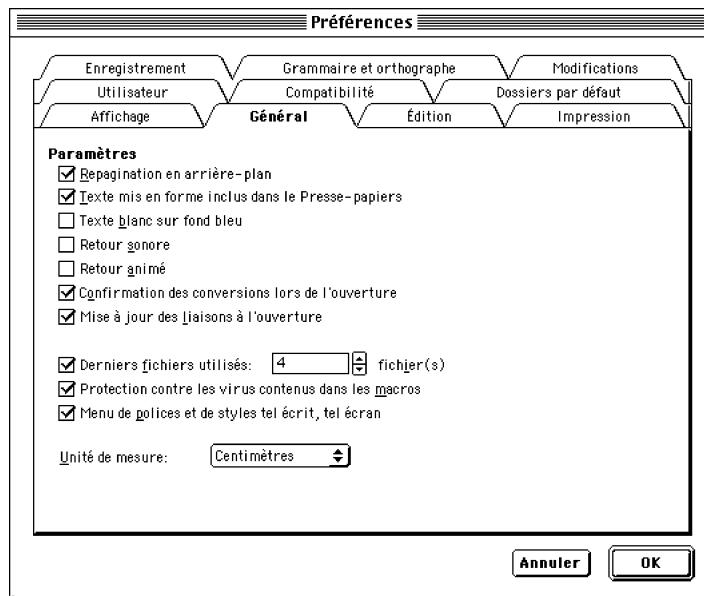
où z_α est la valeur lue dans la table de la loi normale centrée réduite telle que : $P(|Z| \geq z_\alpha) = \alpha$. Ici, $\bar{x} = 8.8$, $s_c = 2.1$, $n = 100$ et $z_\alpha = 1.96$. D'où l'intervalle de confiance :

$$8.8 - 1.96 \times \frac{2.1}{\sqrt{100}} \leq \mu \leq 8.8 + 1.96 \times \frac{2.1}{\sqrt{100}}$$

c'est-à-dire : $8.38 \leq \mu \leq 9.21$.

Exercice 4

1) La plupart des logiciels actuels comportent un item de menu nommé "Préférences" ou "Options", permettant de personnaliser dans une certaine mesure le fonctionnement du logiciel. Par exemple, on a représenté ci-dessous le dialogue affiché par l'item "Préférences" des menus de Microsoft Word.



a) Donner quelques raisons expliquant pourquoi les logiciels comportent généralement de telles fonctions de personnalisation.

Ces fonctions de personnalisation relèvent des propriétés d'adaptabilité de l'interface utilisateur. Grâce à ces fonctions, l'utilisateur pourra, dans une certaine mesure, adapter le logiciel à ses besoins propres (par exemple, ajouter dans un menu une commande peu accessible qu'il utilise fréquemment). Il pourra aussi adapter le logiciel à la configuration particulière de son poste de travail (dossiers par défaut pour l'enregistrement des documents, par exemple).

b) La fenêtre de dialogue affichée ci-dessus se présente sous forme "d'onglets" superposés. Pourquoi une telle disposition a-t-elle été choisie? Quelles qualités de l'interface utilisateur cette disposition permet-elle d'améliorer? Quelles sont au contraire les qualités qui sont ainsi perdues?

Cette disposition sous forme d'onglets a été choisie en raison du nombre élevé d'items à afficher. Cela permet de conserver une interface claire et concise. En revanche, on perd en visibilité: l'utilisateur ne peut pas visualiser sur un seul écran l'ensemble des paramètres susceptibles d'être modifiés. La présentation sous forme d'onglets est métaphorique et intuitive, mais encore faudrait-il que la répartition des paramètres entre les onglets soit "évidente", ce qui est loin d'être le cas: pour les trois logiciels (Word, Excel et Powerpoint) de la suite bureautique "Office" de Microsoft, le même paramètre se retrouve sous des onglets différents selon le logiciel.

2) Au cours des TD d'informatique, nous avons fréquemment utilisé "l'utilitaire d'analyse" d'Excel.

a) Ce composant permet-il de réaliser des traitements qui ne seraient pas réalisables par Excel autrement? Quel avantage y a-t-il cependant à utiliser l'utilitaire d'analyse?

L'utilitaire d'analyse est un ensemble de macro-commandes rédigées à l'aide du langage de programmation d'Excel, et utilisant les fonctions d'Excel. Les traitements qu'il propose peuvent tous être réalisés directement à l'aide des fonctions d'Excel. Cependant, l'utilitaire d'analyse permet de réaliser certains traitements sans connaître la syntaxe des fonctions utilisées, et il guide efficacement l'utilisateur à l'aide de fenêtres de dialogue bien documentées.

b) Parmi les fonctionnalités essentielles d'un tableur, quelle est celle qui ne s'applique pas aux résultats produits par l'utilitaire d'analyse?

L'utilitaire d'analyse effectue les calculs en s'appuyant sur les fonctions d'Excel et fournit les résultats en les affichant dans des cellules. Cependant, il ne fournit à l'utilisateur que les valeurs numériques ou textuelles des résultats, et non les formules de calcul qui ont servi à les obtenir. On perd ainsi une fonctionnalité fondamentale du tableur, à savoir le *recalcul automatique* de tous les résultats en cas de modification des données d'entrée.

c) Ce composant fait-il partie intégrante du "programme" Excel, ou en est-il un module complémentaire?

Comme nous l'avons constaté en travaux dirigés, ce composant n'est pas actif "par défaut" sur une installation standard d'Excel. De plus, comme nous l'avons dit au a), l'utilitaire d'analyse est un ensemble d'outils construits "au-dessus" d'Excel et non pas une partie du programme lui-même.