Université de Bretagne Occidentale
U.F.R. de Lettres et Sciences Humaines
CS 93837 - 29238 BREST CEDEX 3

Année Universitaire 2011-2012

Section : Psychologie - Licence 3è année

Enseignant responsable : F.-G. Carpentier

Correction de l'épreuve de Statistiques et Informatique appliquées à la Psychologie

N.B. Calculatrices, tables des lois statistiques et résumé de cours autorisés.

Exercice 1

Dans un article intitulé "Les facteurs psychophysiologiques de la précocité intellectuelle : résultats d'une enquête comparative chez l'enfant entre 8 et 11 ans"[1], des chercheurs ont étudié une population de 195 enfants précoces.

1) Les auteurs indiquent notamment :

> Le niveau moyen du quotient intellectuel dans notre population d'enfants précoces était de 136.46 avec un écart type de 7.87. La tranche 130-140 de QI global représentait 40.3% de la population.

Si le QI global est distribué selon une loi normale de moyenne 136.46 et d'écart type 7.87, quelle proportion de la population se situe dans la tranche 130-140 ? Comparer, au niveau descriptif, la proportion indiquée par les auteurs et la proportion théorique.

On introduit la variable Z centrée réduite associée au QI : $Z = \dfrac{QI - 136.46}{7.87}$. Pour $QI = 130$, on a : $Z = -0.82$ et pour $QI = 140$, $Z = 0.45$. Si QI, et donc Z suivaient une loi normale, on aurait : $P(130 \leq QI \leq 140) = P(-0.82 \leq Z \leq 0.45) = P(0 \leq Z \leq 0.82) + P(0 \leq Z \leq 0.45) = 0.1736 + 0.2939 = 0.4675$. Or, on constate que la classe centrale $[130, 140]$ ne représente en réalité que 40.3% de la population. La distribution réelle est donc plus dispersée (aplatie) que la loi normale considérée ici.

2) Plus loin, les auteurs indiquent :

> Les différents niveaux de QI Verbal et de QI Performance étaient les suivants : $QIV = 135.82$ (écart type : 8.66) et $QIP = 125.45$ (écart type : 9.99). La différence entre le niveau de QIV et QIP était en moyenne de 14.22 points en valeur absolue.

a) En considérant les 195 enfants comme un échantillon tiré au hasard dans la population des enfants précoces, donner un intervalle de confiance avec un degré de confiance de 95% pour le QI Verbal dans cette population.

[1] J. Louis, O. Revol, C. Nemoz, R.M. Dulac, P. Fourneret, Archives de Pédiatrie, 2004

Soit $\overline{QIV}$ la variable "moyenne du QI verbal sur un échantillon de taille 195" et μ la moyenne (inconnue) du QI Verbal dans la population considérée. La distribution de la loi suivie par la variable $\overline{QIV}$ est la distribution d'échantillonnage. Il s'agit d'une loi normale de moyenne μ et d'écart type $E = \dfrac{8.66}{\sqrt{195}} = 0.6202$. On introduit la variable $Z = \dfrac{\overline{QIV} - \mu}{0.6202}$, qui suit une loi normale centrée réduite. $Z_{obs} = \dfrac{135.82 - \mu}{0.6202}$ est une valeur observée de Z et on peut affirmer, avec un degré de confiance de 95%, que : $-1.96 \leq Z_{obs} \leq 1.96$ (lecture des tables).
De $-1.96 \leq \dfrac{135.82 - \mu}{0.6202}$, on tire : $\mu \leq 137.04$. De $\dfrac{135.82 - \mu}{0.6202} \leq 1.96$, on tire : $\mu \geq 134.60$.
Finalement, l'intervalle de confiance cherché est : $[134.60, 137.04]$.

b) Comment peut-on expliquer que la dernière valeur indiquée (14.22) ne corresponde pas à la différence entre les moyennes de QIV et QIP ?

Les auteurs indiquent qu'il s'agit d'une différence *en valeur absolue*. De façon plus précise, en calculant les différences algébriques entre les valeurs de QIV et de QIP sur chaque sujet, puis en calculant la moyenne de ces différences et enfin la valeur absolue de cette moyenne, on obtiendrait la valeur $|135.82 - 125.45| = 10.37$, différente de la valeur indiquée par les auteurs. Cette dernière a donc été calculée en calculant, pour chaque sujet, la valeur absolue de la différence entre QIV et QIP, puis en faisant la moyenne de la série de ces différences absolues. L'écart entre ces deux valeurs montre que pour un certain nombre de sujets, on a observé $QIV < QIP$, bien qu'en moyenne on ait : $\overline{QIV} > \overline{QIP}$.

Exercice 2

Une psychologue étudie l'efficacité de trois stratégies de contrôle de la douleur. 27 participants prennent part à l'expérience. Lors d'une séance expérimentale les participants doivent plonger une main dans l'eau glacée pendant une durée de 180 s. Trois groupes de 9 participants sont formés aléatoirement et un type de stratégie pour contrôler la douleur ressentie est utilisé par chacun des groupes. Les stratégies sont : (a) dire « ha ! » sans arrêt ; (b) se concentrer sur le rythme de sa respiration ; (c) imaginer un événement plaisant. Par ailleurs, à trois moments au cours de la séance (après 60 s, 120 s et 180 s), la psychologue leur demande d'évaluer l'intensité de la douleur ressentie sur une échelle de 1 (non douloureux) à 50 (douleur intolérable). Les résultats observés sont les suivants :

Dire "ha !"			Respiration			Imagerie		
60s	120s	180s	60s	120s	180s	60s	120s	180s
27	33	42	14	18	35	16	18	35
28	34	45	16	22	34	15	22	33
26	35	43	18	23	39	14	21	34
28	33	48	14	20	32	13	20	31
24	28	42	16	19	29	15	22	32
24	27	44	15	21	30	17	16	28
25	30	38	14	20	31	15	20	32
23	33	49	19	26	37	16	22	27
25	34	42	20	22	40	14	21	36

1) Dans la condition "Imagerie", l'intensité de la douleur ressentie est-elle la même après 60 s et après 120 s ? Répondre à cette question à l'aide d'un test paramétrique unilatéral de comparaison de moyennes, au seuil de 5%.

Il s'agit ici d'une situation de groupes appariés. Calculons les paramètres descriptifs de la série des différences individuelles, prises dans le sens “valeur observée à 120s” - “valeur observée à 60s”.

60s	120s	d_i	d_i^2
16	18	2	4
15	22	7	49
14	21	7	49
13	20	7	49
15	22	7	49
17	16	-1	1
15	20	5	25
16	22	6	36
14	21	7	49
	Σ	47	311

Moyenne : $\overline{d} = \frac{47}{9} = 5.22$. Variance : $s^2 = \frac{311}{9} - \frac{47^2}{9^2} = 7.28$. Variance corrigée : $s_c^2 = \frac{9}{8} s^2 = 8.19$.
Soient μ_{60} et μ_{120} les moyennes de la variable dépendante sur la population parente, dans les deux conditions. Pour le test unilatéral demandé, on prend l'hypothèse H_1 consistante avec les observations faites, c'est-à-dire correspondant à des valeurs à 120s supérieures à celles à 60s. Les hypothèses du test sont donc les suivantes :

- $H_0 : \mu_{60} = \mu_{120}$
- $H_1 : \mu_{60} < \mu_{120}$

La statistique de test est $t = \frac{\overline{d}}{E}$ avec $E^2 = \frac{s_c^2}{n}$. Elle suit une loi de Student à 8 *ddl*. Pour un seuil $\alpha = 5\%$ unilatéral, la valeur critique lue dans la table est $t_{lue} = 1.86$. Comme la moyenne des différences observées est positive et que l'hypothèse alternative a été prise en cohérence avec les observations, la valeur critique est $t_{crit} = 1.86$ et la règle de décision est donc :

– Si $t_{obs} \leq 1.86$, on retient H_0 ;
– Si $t_{obs} > 1.86$, on retient H_1.

Les calculs donnent ici : $E^2 = \frac{8.19}{9} = 0.91$, $E = \sqrt{0.91} = 0.95$ et $t_{obs} = \frac{5.22}{0.95} = 5.47$.
Cette valeur est supérieure à 1.86 et on retient donc H_1 : l'évaluation de l'intensité de la douleur est significativement plus élevée à 120s qu'à 60s.

2) On utilise un logiciel de traitement statistique pour comparer, dans la condition “Imagerie”, les intensités de douleur ressentie au bout de 120 s et au bout de 180 s. Les résultats fournis par le logiciel sont les suivants :

	Test t pour des Echantillons Appariés (Feuille de données3) Différences significatives marquées à p <.05000							
	Moyenne	Ec-Type	N	Différ.	Ec-Type Différ.	t	dl	p
120s	20.22	2.0480						
180s	32.00	3.0000	9	-11.78	3.3458	-10.56	8	0.000006

Interpréter les résultats fournis par le logiciel.

On fait ici un test analogue au précédent, mais on compare cette fois la condition 120s à la condition 180s. Il s'agit d'un test bilatéral, et le logiciel indique une p-value égale à 6×10^{-6}. Au seuil de 5%, on conclut donc à une différence significative entre les deux conditions, la douleur ressentie étant plus élevée à 180s (moyenne : 32.0) qu'à 120s (moyenne 20.22).

Exercice 3

Un enseignant a mené une expérimentation visant à établir si de nouvelles activités de lecture dirigée aideraient les élèves à améliorer certains aspects de leurs savoir-faire en lecture. Il a fait suivre ces activités à une classe de 21 élèves pendant une période de 8 semaines. Une classe de contrôle de même niveau, comportant 23 élèves, a suivi le même programme, mais sans ces activités. A la fin des 8 semaines, tous les étudiants ont passé un test de lecture qui mesure les performances des sujets sur les savoir-faire que le traitement est censé améliorer. Les données sont les suivantes :

Groupe "Traitement"	Score au test
s1	24
s2	43
s3	58
s4	71
s5	43
s6	49
s7	61
s8	44
s9	67
s10	49
s11	53
s12	56
s13	59
s14	52
s15	62
s16	54
s17	57
s18	33
s19	46
s20	43
s21	57
Moyenne	51.48
Variance	115.39
Variance corrigée	121.16

Groupe "Contrôle"	Score au test
s'1	42
s'2	43
s'3	55
s'4	26
s'5	62
s'6	37
s'7	33
s'8	41
s'9	19
s'10	54
s'11	20
s'12	85
s'13	46
s'14	10
s'15	17
s'16	60
s'17	53
s'18	42
s'19	37
s'20	42
s'21	55
s'22	28
s'23	48
Moyenne	41.52
Variance	281.29
Variance corrigée	294.08

1) Faire un test bilatéral de comparaison de moyennes pour étudier les activités de lecture dirigée ont une influence sur la performance des élèves (seuil : 5%).

Soient respectivement μ_T et μ_C les moyennes pour les conditions "traitement" et "contrôle" dans les populations parentes. Les hypothèses du test bilatéral demandé s'écrivent :

- $H_0 : \mu_T = \mu_C$
- $H_1 : \mu_T \neq \mu_C$.

Soient respectivement $\overline{x}_T$, s_T^2, n_T et $\overline{x}_C$, s_C^2, n_C la moyenne et la variance de la variable "score au test" ainsi que l'effectif dans chacun des deux groupes.
La statistique de test est : $t = \dfrac{\overline{x}_T - \overline{x}_C}{E}$ avec $E^2 = \dfrac{n_T s_T^2 + n_C s_C^2}{n_T + n_C - 2}\left(\dfrac{1}{n_T} + \dfrac{1}{n_C}\right)$. Sous H_0, cette statistique suit une loi de Student à 42 ddl. Pour un seuil $\alpha = 5\%$ bilatéral, la valeur critique est $t_{crit} = 2.02$.
La règle de décision est donc :

– Si $|t_{obs}| \leq 2.02$, on retient H_0 ;
– Si $|t_{obs}| > 2.02$, on retient H_1.

Les calculs donnent : $E^2 = 19.29$; $E = 4.39$; $t_{obs} = \dfrac{51.48 - 41.52}{4.39} = 2.27$.
On conclut donc sur H_1 : il existe donc une différence significative des scores des deux groupes, au bénéfice du groupe “traitement”.

2) On utilise un logiciel de traitements statistiques pour faire un test non paramétrique sur les mêmes données. Le résultat obtenu est le suivant :

	Test U de Mann-Whitney (Feuille de données19) Tests significatifs marqués à p <.05000									
	SommeRgs Traitement	SommeRgs Contrôle	U	Z	niv. p	Z ajusté	niv.p	N Actifs Traitement	N Actifs Contrôle	2*(1-p) p exact
Score	579.00	411.00	135.0	2.5024	0.0123	2,5045	0,0123	21	23	0.0117

a) Quel est le test utilisé ? Enoncer l’hypothèse nulle et l’hypothèse alternative correspondantes.

Le test réalisé ici est un test de Wilcoxon Mann et Whitney. Les hypothèses nulle et alternative peuvent s’exprimer à l’aide des médianes θ_T et θ_C de la variable dépendante dans les populations parentes des deux groupes, sous la forme :

- $H_0 : \theta_T = \theta_C$
- $H_1 : \theta_T \neq \theta_C$.

On peut aussi exprimer les hypothèses du test sous la forme suivante :
H_0 : La probabilité qu’un score provenant de la première population soit supérieur à un score provenant de la seconde est de 50%.
H_1 : Cette probabilité est différente de 50%.

b) Interpréter les résultats fournis par le logiciel, en utilisant un seuil de 5%.

Etant donné la taille des deux échantillons (21 et 23), il est légitime d’utiliser ici l’approximation par une loi normale. En tenant compte de la correction pour les ex aequo, la statistique de test est $Z = 2.5045$ et le niveau de significativité correspondant est $p = 1.23\%$. Au seuil de 5%, on conclut donc à une différence significative entre les deux groupes, ce qui confirme le résultat précédent.

Exercice 4

On étudie la relation entre deux variables nominales X et Y. Sur un échantillon de 70 personnes on obtient $\chi^2 = 13.28$. Le test du Khi-2 indique $p = 1\%$. Que signifie cette valeur ? Parmi les propositions suivantes (A, B, C, D, E) deux seulement sont correctes. Indiquer lesquelles sont correctes.

A. Si on tirait tous les échantillons possibles de 70 personnes d’une population parente où il n’existe pas de liaison entre les deux variables X et Y, on trouverait 1% des échantillons avec un χ^2 supérieur ou égal à 13.28.

B. Si on tirait tous les échantillons possibles de 70 personnes d’une population parente où il n’existe pas de liaison entre les deux variables X et Y, on trouverait 1% des échantillons avec un χ^2 inférieur à 13.28.

C. Sachant que $\chi^2 = 13.28$ dans notre échantillon de 70 personnes, il y a seulement une chance sur 100 que l’hypothèse nulle soit vraie (qu’il n’y ait pas de liaison entre X et Y dans la population parente).

D. Si l'hypothèse nulle est vraie (s'il n'y a pas de liaison entre X et Y dans la population parente) et que l'on tire au hasard un échantillon de 70 personnes, on a une chance sur 100 d'obtenir un échantillon avec un χ^2 au moins égal à 13.28.

E. Sachant que $\chi^2 = 13.28$ dans notre échantillon de 70 personnes, il y a une chance sur 100 que l'hypothèse nulle soit fausse (qu'il y ait une liaison entre X et Y dans la population parente).

La loi du khi-2 envisagée est la loi suivie par la statistique du khi-2 lorsque H_0 est vraie, c'est-à-dire lorsqu'il n'existe pas de lien entre les deux variables. La p-value représente alors les chances que l'on a d'obtenir un échantillon "au moins aussi extrême que celui observé" lorsque l'on tire au hasard un échantillon, ce qui correspond à la formulation D. Le même résultat peut être exprimé sous une forme statistique, et non probabiliste : X et Y sont deux variables indépendantes ; on tire tous les échantillons possibles dans la population parente ; la p-value correspond alors à la fréquence des échantillons conduisant à une statistique du khi-2 supérieure ou égale à 13.28. Cela correspond à la formulation A.
Les propositions correctes sont donc A et D.