Université Paul Valéry
Arts - Lettres - Langues
Sciences humaines & sociales

PSYCHOLOGIE - Licence 3
E56XPM
Statistique appliquée à la psychologie 4
Examen juin 2008

NOM	**:**
PRÉNOM	**:**
N° Carte d'étudiant	**:**
Signature	**:**

Durée : 2 heures.
• Matériel autorisé : une page manuscrite recto/verso, calculatrice et tables de lois.
• Téléphone portable interdit.
• Les calculatrices ne doivent en aucun cas circuler ou être échangées entre les étudiants.

Consignes
Tous les tests statistiques seront construits avec une erreur de 1re espèce de 1 %.

Les exercices 1 et 4 sont 2 questions de cours sans application numérique. Donner une réponse claire et succincte en précisant chacun des objets que vous introduisez.

Exercice 1 : Dans un texte ancien, on s'intéresse à l'apparition des 5 voyelles A, E, I, O, U et parmi les 3000 premières voyelles on compte le nombre d'occurrences de chacune d'elles. On cherche à éprouver l'hypothèse que chaque voyelle apparaît de façon équiprobable.
Construire le test approprié en spécifiant clairement les hypothèses, la statistique du test et la zone de rejet.

Le test approprié est le test du Khi2 d'adéquation à la loi uniforme à 5 catégories (le dé à 5 faces).
L'hypothèse H_0 est "la loi uniforme à 5 catégories (proba = 1/5)", l'alternative est non H_0
On pose N_A, N_E, N_I, N_O, N_U, le nombre d'occurrences OBSERVEES de chaque voyelle.
Ici l'effectif théorique est : $\tilde{n} = 3000/5 = 600$
La statistique de test est $T = \sum_{i=A,E,I,O,U} \frac{(N_i - \tilde{n})^2}{\tilde{n}}$,
le ddl = 4
on rejete H_0 si $T > 13.2767$

Exercice 2 : Un magasin d'habillement comporte 3 rayons distincts : le rayon "femme", le rayon "homme" et le rayon "enfant".

On a observé que sur 5 articles vendus, 3 provenaient du rayon "femme", 1 du rayon "homme" et 1 du rayon "enfant".

1. Exprimer par une variable aléatoire notée RF le fait qu'un article vendu provienne du rayon "femme" ; donner sa loi.
2. Quelle est la loi de la variable aléatoire notée NRF du nombre d'articles vendus provenant du rayon "femme" pour n articles vendus au total dans le magasin.
3. Donner l'expression de l'intervalle de dispersion de NRF pour n articles vendus au total dans le magasin (choisir un risque $\alpha = 2.8$ %).
4. Calculer les bornes de cet intervalle lorsque le nombre d'articles vendus dans la journée est de 750.

On note "RF" la variable aléatoire "l'article vendu provient du rayon femme" , "RF" prend les valeurs 1 si oui et 0 sinon et donc suit une loi de Bernouilli B(3/5)

Le nombre d'articles vendus provenant du rayon "femme" : NRF suit une loi binomialle de paramètre n et $3/5$.

On a $\frac{NRF-n*3/5}{\sqrt{n*3/5*2/5}}$ est N(0,1)donc l'Intervalle de dispersion du nombre d'articles vendus provenant du rayon "femme" sur n article est :

$$[n*3/5 - l_{\alpha/2}*\sqrt{n*3/5*2/5}, n*3/5 + l_{\alpha/2}*\sqrt{n*3/5*2/5}]$$

avec $l_{\alpha/2}$ = 2.1973

Le calcul des bornes pour n = 750
*750 * 0.6 - 2.1973 * sqrt(750 * 0.6 * 0.4) = 420.52*
*750 * 0.6 + 2.1973 * sqrt(750 * 0.6 * 0.4) = 479.48*

Exercice 3 : Dans ce magasin d'habillement, on a noté ces dernières années le prix moyen d'un article du rayon "enfant" qui est de 18 euros. La direction souhaite communiquer sur sa politique de promotion dans ce rayon par la baisse du prix moyen d'un article.

Pour cela, elle relève le prix de 121 articles du rayon (pris au hasard).

1. Construire le test statistique associé à cette démarche.
2. Calculer la zone de rejet de ce test, étant donné l'information suivante : sur les 121 articles on a calculé que la dispersion $\sum_{i=1}^{121}(p_i - \bar{p})^2 = 5929$ où p_i est le prix de l'article i et $\bar{p}$ le prix moyen observé de ces 121 articles.
3. Sur ces 121 articles le prix moyen observé au mois de mai 2008 est de 16.8 euros. Quel est la décision à prendre concernant ce test statistique ? Que proposeriez-vous alors à la direction ?
4. Sur ces 121 articles le prix moyen observé au début du mois de juin 2008 est de 16.2 euros. Quel est la décision à prendre concernant ce test statistique ?
5. Construire alors un intervalle du prix moyen d'un article du rayon "enfant" (choisir un risque $\alpha = 3.6$ %). Quel est le nom de cet intervalle ?

Le prix d'un article du rayon "enfant" est une variable aléatoire continue ;
Le test statistique associé à cette démarche est le test sur l'espérance de cette variable.
H_0 *"*$\mu = 18$*" contre* H_1 *"* $\mu < 18$*"*
La variable $\bar{p}$ *suit une loi normale et on a sous* H_0 $\frac{\bar{p}-18}{\sigma/\sqrt{n}}$ *est N(0,1)*

et on rejettera H_0 *si*

$$\bar{p} < 18 - l_\alpha \sigma/\sqrt{n}$$

avec $l_\alpha = 2.3263$

Comme on ne connaît pas σ^2 *, on peut l'estimer par "la dispersion / n" soit 5929/121 = 49 donc* σ *est estimé par 7.*

La zone de rejet est donc $18 - l_\alpha \sigma/\sqrt{n} = 18 - 2.3263 * 7/11 =$ *16.51963*

Au mois de mai, le prix moyen observé est de 16.8 euros ; on ne rejette donc H_0 *(le prix moyen ne baisse pas de façon significative) et on proposera alors à la direction d'accentuer sa sa politique de promotion*

Au mois de juin, le prix moyen observé est de 16.2 euros ; on rejette donc H_0 *(le prix moyen baisse de façon significative)*

On calcule alors un intervalle de CONFIANCE du prix moyen d'un article du rayon "enfant" :

$$[\bar{p} - l_{\alpha/2}\sigma/\sqrt{n}; \bar{p} + l_{\alpha/2}\sigma/\sqrt{n}]$$

avec $\bar{p} = 16.2$ *et* $l_{\alpha/2} = 2.0969$

et on trouve $[16.2 - 2.0969 * 7/11; 16.2 + 2.0969 * 7/11] =$ *[14.8656 , 17.5344]*

Exercice 4 : On dispose sur 800 individus de l'observation de 2 variables numériques notées X et Y que l'on note $(x_1, y_1, x_2, y_2, ..., x_{800}, y_{800})$. On cherche à mettre en évidence l'existence d'une liaison linéaire.
Construire le test approprié en spécifiant clairement les hypothèses, la statistique du test et la zone de rejet.

Le test approprié est le test sur le coefficient de corrélation ρ entre X et Y.

La statistique de test de H_0 "$\rho = 0$" contre H_1 "$\rho \neq 0$" est "$\sqrt{n-1}$. cor"
et on rejettera H_0 si $|cor| > \frac{l_{\alpha/2}}{\sqrt{n-1}}$ avec $l_{\alpha/2} = 2.5758$

Ici la borne de rejet est 2.5758 / sqrt(800-1) = 0.09112525

Exercice 5 : On a mesuré une variable continue X sur 4 groupes disjoints d'individus, et on a obtenu les résultats suivants :

	Taille du groupe	Moyenne	Variance
Groupe 1	380	46.4	300.9
Groupe 2	280	45.6	298.4
Groupe 3	455	52.2	294.0
Groupe 4	165	47.1	320.8

On se pose naturellement la question de savoir si ces 4 groupes ont des résultats similaires au vu de cette variable X. Construire le test approprié en spécifiant clairement les hypothèses, la statistique du test et la zone de rejet. Proposer votre conclusion.

Le test approprié est le test d'analyse de la variance.

Si μ_i est l'espérance de la variable X pour le groupe i
on construit le test de H_0 "l'egalité des μ_i" contre H_1 "les μ_i ne sont pas tous égaux"

*dispersion INTRA = 380 * 300.9 + 280 * 298.4 + 455 * 294.0 + 165 * 320.8 = 384596*

*moyenne générale = (380 * 46.4 + 280 * 45.6 + 455 * 52.2 + 165 * 47.1) / 1280 = 48.37695 = 48.38*

dispersion INTER $= 380 * (46.4 - 48.38)^2 + 280 * (45.6 - 48.38)^2 + 455 * (52.2 - 48.38)^2 + 165 * (47.1 - 48.38)^2$ *= 10563.58*

*la statistique du test est Inter / Intra * (1280 - 4) = 217.6774*

ddl = 3 ; la zone de rejet = 11.3449,

On rejette donc l'hypothèse H_0 "egalité des μ_i"