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

**PSYCHOLOGIE - Licence 3**
**E56XPM**
**Statistique appliquée à la psychologie 4**
**Examen du 30 janvier 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 pour le choix des risques : dans tout l'examen**
Pour la construction d'un intervalle de dispersion on choisira un risque $\alpha = 1.2$ %.
Pour la construction d'un intervalle de confiance on choisira un risque $\alpha = 2.8$ %.
Tous les tests statistiques seront construits avec une erreur de 1re espèce de 1 %.

*quand $\alpha$ = 1.2 % la valeur $l_{\alpha/2}$ (associée à .994) = 2.5121*
*quand $\alpha$ = 2.8 % la valeur $l_{\alpha/2}$ (associée à .986) = 2.1973*

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

**Exercice 1 :** Dans une séquence ADN de longueur 3500, formée des lettres A, C, G, T, on compte le nombre d'occurrences de chacune des lettres. Et on cherche à éprouver l'hypothèse que chaque lettre 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 4 catégories (le dé à 4 faces).*
*L'hypothèse $H_0$ est "la loi uniforme à 4 catégories (proba = 1/4)", l'alternative est non $H_0$*
*On pose $N_A, N_C, N_G, N_T$, le nombre d'occurrences OBSERVEES de chacune des lettres.*
*Ici l'effectif théorique est : $\tilde{n} = 3500/4 = 875$*
*La statistique de test est $T = \sum_{i=A,C,G,T} \frac{(N_i - \tilde{n})^2}{\tilde{n}}$,*
*le ddl = 3*
*on rejete $H_0$ si $T > 11.3449$*

**Exercice 2 :** On dispose de 200 observations $x_1, x_2, ....., x_{200}$ de la variable aléatoire continue $X$.
Donner l'expression de l'intervalle de confiance pour le paramètre de l'espérance mathématique : $E(X)$.

*On pose $\bar{X}$ la moyenne des 200 observations , $S_X^2 = \frac{1}{n}\sum(x_i - \bar{X})^2$ la variance empirique des 200 observations.*
*La variable $\bar{X}$ est gaussienne $(\mu, \sigma^2)$ de part le TCL, l'intervalle de confiance de $\mu = E(X)$ est (comme $\sigma^2$ est inconnu)*

$$[\bar{X} - l_{\alpha/2}\frac{S}{\sqrt{n}}; \bar{X} + l_{\alpha/2}\frac{S}{\sqrt{n}}]$$

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

**Exercice 3 :** On suppose que la distribution de la variable $Y$ est $\mathcal{N}(2, \sigma^2 = .0625)$ mais on a de bonnes raisons de penser qu'en réalité le paramètre de l'espérance mathématique a diminué.

1. Construire le test statistique qui permette d'éprouver cette idée avec un échantillon de taille $n$ : donner les hypothèses, la zone de rejet.
2. Calculer la borne de rejet associée à un échantillon de 36 individus.

*On construit le test de $H_0$ : $\mu = 2$ (le paramètre de l'espérance mathématique est bien = 2) contre $H_1$ : $\mu < 2$ (le paramètre de l'espérance mathématique a bien diminué) (test unilatéral)*
*On pose $\bar{Y}$ la moyenne des observations de échantillon de taille n. $\bar{Y}$ est gaussienne $(2, \frac{.25}{\sqrt{n}})$ (sqrt(.0625) = 0.25)*
*On rejettera alors $H_0$ lorsque $\bar{Y} < 2 - l_\alpha \frac{.25}{\sqrt{n}}$*
*avec $l_\alpha$ = 2.326348*
*Pour un échantillon de 36 individus la borne de rejet est : 2 - 2.3263 * .25 / 6 = 1.903071*

**Exercice 4 :** On a mesuré sur 380 individus 2 variables $Z$ et $T$.
On obtient les résultats suivants : $\text{cov}(Z, T) = 136$ et $\sigma_Z = 14$ et $\sigma_T = 19$.
Peut on considérer qu'il existe une liaison linéaire entre $Z$ et $T$ ?

*Le coefficient de corrélation est : cor = 136 / (14*19) = 0.5112782*

*La statistique de test de $H_0$ "cor = 0" contre $H_1$ "cor $\neq$ 0" est "cor / $\sqrt{n-1}$"*
*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(379) = 0.1323100 donc cor = 0.5112782 est significativement NON nul (rejet de $H_0$)*

**Exercice 5 :** Dans la séquence ADN formée des lettres A, C, G et T, on suppose que ces lettres apparaissent de façon équiprobable et indépendante.

1. Exprimer par une variable aléatoire (dont on donnera la loi) le fait qu'une lettre est un "A".
2. Donner l'expression de l'intervalle de dispersion du nombre de "A" dans un séquence de longueur $n$.
3. Calculer les bornes de cet intervalle pour une séquence de longueur 350.

*On note "$L_A$" la variable aléatoire "la lettre est A" , "$L_A$" suit une loi de Bernouilli B(1/4)*

*Le nombre de "A" : $NL_A$ est binomial et on a $\frac{NL_A - n*1/4}{\sqrt{n*1/4*3/4}}$ est N(0,1)*
*donc l'Intervalle de dispersion du nombre de "A" dans un séquence de longueur n est :*

$$[n*1/4 - l_{\alpha/2}*\sqrt{n*1/4*3/4}, n*1/4 + l_{\alpha/2}*\sqrt{n*1/4*3/4}]$$

*avec $l_{\alpha/2}$ = 2.5121*
*Le calcul des bornes pour n = 350*
*350 * .25 - 2.5121 * sqrt( 350 * .25 * .75) = 67.14966*
*350 * .25 + 2.5121 * sqrt( 350 * .25 * .75) = 107.8503*

**Exercice 6 :** Dans un service hospitalier, on a l'habitude de déclarer aux patients atteints d'une certaine pathologie que l'examen qu'ils doivent subir comporte un risque de complication de 15 %. Cette valeur fait référence dans le milieu médical depuis plusieurs années.

1. Il y a en France de l'ordre de 1500 personnes qui doivent subir cet examen chaque année. Donner sous forme d'intervalle, le taux de complications auquel on peut s'attendre au cours d'une année.

   *On a ici une loi de Bernouilli B(.15)*
   *Idem exo precedent avec le "taux de complications" (ou moyenne des Oui)*
   *Soit T le taux de complications on a $\frac{T-.15}{\sqrt{\frac{.15*.85}{n}}}$ est N(0,1)*
   *donc l'Intervalle de dispersion du "taux de complications" pour 1500 patients :*

$$[.15 - l_{\alpha/2}*\sqrt{\frac{.15*.85}{1500}}, n*1/4 + l_{\alpha/2}*\sqrt{\frac{.15*.85}{1500}}]$$

   *avec $l_{\alpha/2}$ = 2.5121*

   *.15 - 2.5121 * sqrt( .15 * .85 /1500) = 0.1268396*
   *.15 + 2.5121 * sqrt( .15 * .85 /1500) = 0.1731604*

2. Une nouvelle procédure vient récemment d'être mise au point de sorte à diminuer ce risque de complication. On envisage donc de procéder à une expérimentation sur 200 patients afin de confirmer cette hypothèse. Proposer une modélisation et une règle de décision qui permettent de répondre à la question.

   *On construit le test de*
   $H_0$ *:* $p = .15$ *(la nouvelle procédure ne fait pas diminuer le risque de complication) contre*
   $H_1$ *:* $p < .15$ *(la nouvelle procédure permet de diminuer le risque de complication)*

   *Sur 200 patients, on note* $\bar{Y}$ *le taux de complication observé et on adopte la regle de decision*

   *Rejet de* $H_0$ *si* $\bar{Y} < .15 - l_\alpha * \sqrt{(}.15 * .85/200)$
   *avec* $l_\alpha = -2.3263$

   *la borne de rejet est donc*
   *.15 - 2.3263 * sqrt( .15 * .85 /200) = 0.0912638*

   *(si vous travailler sur la variable "Nbr de Complications"*
   *la borne de rejet sera : 200 * .15 - 2.3263 * sqrt( 200 * .15 * .85) = 18.25276*)

3. Quelle devrait être la conclusion à apporter à cette expérimentation lorsque on observe 22 patients (sur 200) atteints de complication.

   *NON rejet de* $H_0$

4. Quelle devrait être la conclusion à apporter à cette expérimentation lorsque on observe 15 patients (sur 200) atteints de complication.

   *REJET de* $H_0$

5. À partir des conclusions de la question précédente, discuter de la pertinence de construire un intervalle pour la valeur du nouveau risque de complication que l'on pourrait annoncer aux patients. Calculer alors les bornes de cet intervalle.

   *Puisque on REJETE* $H_0$ *le paramètre p le la loi de Bernouilli assocée à la variable "avoir une complication" est maintenant inconnu, il est donc pertinent de construire un intervalle de confiance de ce nouveau paramètre (le nouveau risque de complication)*
   *15/200 - 2.1973 * sqrt( 15/200 * 185/200 / 200 ) = 0.03407621*
   *15/200 + 2.1973 * sqrt( 15/200 * 185/200 / 200 ) = 0.1159238*