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

PSYCHOLOGIE - Licence 3
E64XP1
Statistique appliquée à la psychologie 4
Examen du 16 mai 2011

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

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

Consignes pour le choix des risques dans tout l'examen :
Pour la construction d'un intervalle, on choisira un risque $\alpha = 2$ %.
Pour la construction des tests, on choisira un risque $\alpha = 1$ %.

Le taux d'absentéisme (maladie ou accident du travail) des salariés en France est aujourd'hui de 4%. On souhaite analyser les disparités qui pourraient exister sur ce taux selon différents critères.

Dans tout le sujet, la notion "absence/absentéisme" sera comprise comme "absence/absentéisme pour maladie ou accident du travail".

Exercice 1 : L'Ile-de-France est supposée être la région française avec un taux d'absentéisme plus faible que la moyenne nationale. Pour vérifier cette hypothèse, on propose d'élaborer le sondage suivant :

Chaque enquêteur tire au hasard une lettre de l'alphabet noté "ϕ", visite des entreprises et recueille les informations suivantes :
i) Quels sont vos salariés dont le nom commence par la lettre "ϕ" ?
ii) Parmi eux, quels sont ceux qui sont absents aujourd'hui pour raison de maladie ou accident du travail ?

Cette stratégie de recueil des données est supposée être de bonne qualité : en regroupant les données de tous les enquêteurs on obtient une liste de salariés et pour chaque salarié le fait qu'il soit absent ou non.

Décrire la modélisation en précisant :

1. Quelle est la population visée par l'étude afin de vérifier l'hypothèse formulée ?

2. Quelle réponse associe-t-on à chaque individu ? Quelle est la loi de la variable induite ?

On construit un test statistique afin de vérifier l'hypothèse formulée :

3. Quel test statistique va-t-on construire ? spécifier les 2 hypothèses du test par une écriture mathématique et une phrase.

4. Si on suppose qu'un échantillon de taille 9216 a été obtenu au total par l'ensemble des enquêteurs, quelle est la variable (statistique de test) qui résume l'information recueillie et qui permet de construire ce test ?
5. Spécifier la loi de cette variable pour chaque hypothèse du test.

6. Donner l'expression de la zone de rejet (ne pas faire de calcul).

7. Calculer maintenant la borne de rejet.

8. On lit dans les médias la phrase suivante : l'Ile-de-France arbore le plus faible taux d'absentéisme avec 3,57%, sous entendu "il est plus petit que le taux d'absentéisme national". Que pensez-vous de cette affirmation ?

9. Même question pour une taille d'échantillon de 12000 salariés et le même taux d'absentéisme observé.

Exercice 2 : On se propose maintenant d'évaluer l'hypothèse selon laquelle les plus de 50 ans sont plus souvent absents que leurs collègues.
On découpe alors la population des salariés en 3 tranches d'âge : les moins de 35 ans ; les 35 à 50 ; les plus de 50 ans. On sait que la répartition des salariés selon ces 3 tranches d'âge est de :
34 % pour les - de 35 ; 39 % pour les 35 à 50 ans ; 27 % pour les plus de 50 ans.

On effectue alors 2 types de sondage qui amène les résultats suivants :

- Le premier consiste à compter les salariés absents un jour donné et on obtient le tableau suivant sur un échantillon de taille 500 :

Classe d'âge	moins de 35 ans	les 35 à 50	plus de 50 ans
Effectifs	155	188	157

- Le deuxième sondage consiste à compter sur un autre échantillon le nombre de salarié par tranche d'âge et par nombre de jours d'absence dans l'année 2010, répartis en 4 classes. On obtient le tableau suivant :

Classe d'âge	moins de 35 ans	les 35 à 50	plus de 50 ans
moins de 7 jours	155	188	127
de 7 à 14 jours	125	138	127
de 14 à 20 jours	105	118	127
+ de 20 jours	95	98	127

1. Pour chacun de ces 2 tableaux, proposer un test statistique adapté afin d'évaluer l'hypothèse de travail. Spécifier clairement dans chaque cas les 2 hypothèses du test que vous souhaitez construire.

2. Pour chaque cas, donner l'expression de la statistique de test (sans faire de calcul) et sa borne de rejet.

3. Pour chaque cas, donner la valeur de l'effectif théorique sous H_0 de la 1re cellule.

Deux autres stratégies (en terme de recueil de données) sont possibles :

- On note pour chaque salarié de l'échantillon son âge exact et son nombre exact de jours d'absence en 2010 (pas de classe ni pour l'âge ni pour les jours d'absence). Quel test pourriez-vous alors construire en utilisant cette information. Donner la statistique de test et sa borne de rejet pour un échantillon de taille 500.

- On note pour chaque salarié de l'échantillon son nombre exact de jours d'absence en 2010 et sa tranche d'âge. Quel test pourriez-vous alors construire en utilisant cette information ? Donner la statistique de test et sa borne de rejet pour un échantillon de taille 500.

Exercice 3 :

Le nombre de jours d'absence par salarié dans une année est supposé suivre une loi de Poisson de paramètre 14,5 jours.

1. Dans une entreprise de 300 salariés, donner un intervalle de dispersion du nombre moyen de jours d'absence que l'on devrait observer en 2011.

2. La direction de l'entreprise pense que ses salariés sont plus souvent absents que la moyenne nationale. Proposez un test statistique basé sur le nombre de jours d'absence qui permette ou non de valider cette impression.

3. Donner la zone de rejet et calculer la borne de rejet (toujours pour une entreprise de 300 salariés).