

Statistiques
Master Statistique et économétrie
TD - Feuille n° 4

V. Monbet

Master 1 - 2012-2013

Exercice 1 (CT 2011-2012)

On suppose que le nombre de clients N attendant à la caisse d'un grand magasin à 11h00 du matin peut être modélisé par une loi de Poisson de paramètre $\theta > 0$. On sait que pour les clients la limite du supportable en attente est de voir au maximum 5 clients devant une caisse. La question se pose de savoir si l'on doit augmenter ou non le nombre de personnes travaillant aux caisses : ceci conduit à un problème de test sur l'opportunité d'embaucher.

Pour la direction, peu désireuse d'embaucher, le problème se pose de la manière suivante : $H_0 : \theta \leq 5$ contre $H_1 : \theta > 5$ à tester avec le niveau 5%; ainsi, la direction veut que la probabilité d'embaucher alors qu'il n'y en a pas besoin est contrôlée (moins de 5%) sans s'intéresser à l'autre erreur possible (qui est de ne pas embaucher alors qu'il le fallait).

1. Déterminer les résultats des tests du rapport de vraisemblance sachant que l'on dispose
 - (a) d'une seule observation de valeur 6;
 - (b) de 100 observations de moyenne 5.4.
2. Quelles sont les propriétés du test de rapport de vraisemblance (a)? Justifier la réponse.

Exercice 2

Voici les chiffres (fictifs) du suivi d'une population de 100 personnes (50 fumeurs, 50 non fumeurs) pendant 20 ans.

	fumeur	non fumeur
cancer diagnostiqué	11	5
pas de cancer	39	45

On s'intéresse : la différence du nombre de cancers entre fumeurs et non fumeurs est-elle statistiquement significative? On note X_i la variable qui vaut 1 si le fumeur i a été atteint

d'un cancer et 0 sinon. De même, on note Y_i la variable qui vaut 1 si le non fumeur i a été atteint d'un cancer et 0 sinon. On suppose les X_i de loi Bernouilli de paramètre θ_f et les Y_i de loi Bernouilli de paramètre θ_{nf} .

1. Si $\theta_f \neq \theta_{nf}$, quelle est la limite de $\sqrt{n}|\bar{X}_n - \bar{Y}_n|$?
2. On suppose que $\theta_f = \theta_{nf} = \theta$ et on note $\hat{\theta} = (\bar{X}_n + \bar{Y}_n)/2$. Montrer que

$$\sqrt{\frac{n}{2\hat{\theta}(1-\hat{\theta})}}(\bar{X}_n - \bar{Y}_n)$$

converge en loi vers une variable de loi de Gauss centrée et réduite.

3. Proposez un test de niveau asymptotique 5% de H_0 : "le taux de cancer n'est pas différent" contre H_1 : "le taux de cancer est différent".
4. Supposons maintenant qu'une étude supplémentaire permet d'avoir le suivi de 300 personnes et que les proportions sont les mêmes :

	fumeur	non fumeur
cancer diagnostiqué	33	15
pas de cancer	117	135

Quelle est la conclusion avec ces données?

5. Revenons aux chiffres de la première étude : proposez un test de niveau asymptotique 5% de H_0 : "fumer n'a pas d'impact sur le taux de cancer" contre "fumer augmente le taux de cancer". Quelle est sa conclusion? Quelle est la p-valeur associée aux données?
6. Que retenir?

Exercice 3

Considérons n variables aléatoires indépendantes identiquement distribuées de densité

$$f(x; \theta) = \theta x^{\theta-1} \text{ pour } 0 \leq x \leq 1$$

avec θ un paramètre réel positif.

1. Montrer que le test uniformément plus puissant pour tester

$$H_0 : \theta \leq 1 \text{ contre } H_1 : \theta > 1$$

au niveau α rejette H_0 si

$$T(x_1, \dots, x_n) = \sum_{i=1}^n \ln(x_i) \geq k$$

2. Quelle est la distribution de $T(X_1, \dots, X_n)$ sous H_0 ?
3. Supposons que $\alpha = 0.05$ et que $n = 50$. Trouver une valeur exacte ou approchée pour k .

Exercice 4

Le taux normal de glycémie est de 1,0 g/L. On dose la glycémie chez 17 sujets diabétiques à jeun depuis 4 heures. La moyenne estimée est de 1,2 g/L avec un écart-type de 0,10 g/L.

Peut-on dire, au risque de 5%, que ces sujets sont hyperglycémiques en supposant que le taux de glycémie est distribué selon une loi normale?

Exercice 5

Le pourcentage de grossesses multiples en France en 1950, donc avant l'introduction des contraceptifs était de 1.25%. En 1991, dans un échantillon de 1000 grossesses de femmes ayant utilisé par le passé des contraceptifs oraux, 21 ont été multiples.

1. La proportion observée dans cet échantillon est-elle significativement supérieure à celle de 1950? On fera un test au niveau de signification nominal 5% et on donnera le niveau de signification réel du test.
2. Peut-on en déduire que l'emploi de contraceptifs oraux augmente le risque de grossesses multiples?