Université d'Orléans - Licence Economie et Gestion Statistique

C. Hurlin. Examen Décembre 2010

Exercice 1 *Tests d'hypothèses. Barème : 17 points*

On considère une variable aléatoire X distribuée selon une loi normale $\mathcal{N}\left(a, b^2\right)$ où a et b **sont des paramètres inconnus**. On dispose d'un N-échantillon $\{X_1, .., X_N\}$ *i.i.d.* de même que loi que X. On considère le test d'hypothèses simples :

$$H_0 : a = a_0 = 2 \tag{1}$$

$$H_1 : a = a_1 = 3 \tag{2}$$

Partie I : Test d'hypothèses simples (10 points)

Question 1 (1 point) Montrez que le rapport des vraisemblances de l'échantillon sous les hypothèses nulle et alternative peut s'écrire sous la forme suivante :

$$\frac{L\left(X_1, .., X_{10}, a_0\right)}{L\left(X_1, .., X_{10}, a_1\right)} = \exp\left[\frac{1}{2b^2}\left(\sum_{i=1}^{N}(X_i - a_1)^2 - \sum_{i=1}^{N}(X_i - a_0)^2\right)\right] \tag{3}$$

Question 2 (1 point) A partir du résultat de la question précédente, démontrez que la région critique W du test UPP est de la forme :

$$W = \left\{ X_1, .., X_{10} \mid \overline{X}_N > A \right\} \tag{4}$$

où $\overline{X}_N = (1/N) \sum_{i=1}^{N} X_i$ désigne la moyenne empirique des poids et où A est une constante déterminée par le niveau de risque de première espèce.

Question 3 (2 points) Montrez que le seuil critique A, obtenu pour un niveau de risque de première espèce α, dépend du paramètre inconnu (variance) b^2. Qu'en concluez vous ?

Question 4 (2 points) On considère à présent la statistique Z_N définie par la relation :

$$Z_N = \frac{\overline{X}_N - a}{S_N / \sqrt{N}} \sim t_{(N-1)} \tag{5}$$

où $S_N^2 = (N-1)^{-1} \sum_{i=1}^{N} \left(X_i - \overline{X}_N\right)$ désigne la variance empirique corrigée et $t_{(N-1)}$ désigne la loi de Student à $N-1$ degrés de liberté. En déduire que le seuil critique A peut se réécrire sous la forme :

$$A = a_0 + G^{-1}(1-\alpha) \frac{S_N}{\sqrt{N}} \tag{6}$$

où $G(.)$ désigne la fonction de répartition de la loi $t_{(N-1)}$.

Question 5 (1 point) On suppose que pour un échantillon de taille $N = 19$, on observe une moyenne empirique $\overline{X}_N$ égale à 2.4 et une variance empirique corrigée S_N^2 égale à 4. Que pouvez conclure pour un risque de première espèce de 5% ?

Question 6 (2 points) Montrez que la puissance de ce test est égale à :

$$Puissance = 1 - G\left(\frac{a_0 - a_1}{S_N/\sqrt{N}} + G^{-1}(1-\alpha)\right) \tag{7}$$

Question 7 (1 point) Commentez la forme générale de la fonction puissance en fonction de l'écart $a_0 - a_1$? Que vaut la puissance dans le cas de ce test pour $N = 19$, $S_N^2 = 4$ et $\alpha = 5\%$?

Partie II : Test d'hypothèse simple contre hypothèse multiple (6 points)

On considère à présent le test suivant :

$$H_0 : a = 2$$

$$H_1 : a > 2$$

Question 1 (1 point) En reprenant vos résultats concernant la région critique du test $H_0 : a = 2$ contre $H_1 : a = 3$, justifiez l'existence d'un test UPP pour $H_0 : a = 2$ contre $H_1 : a > 2$. On observe $N = 19$ et $S_N^2 = 4$. Montrez que la région critique du test UPP pour un niveau de risque de première espèce $\alpha = 5\%$ est :

$$W = \left\{X_1, .., X_{10} \mid \overline{X}_N > 2.795\right\} \tag{8}$$

Question 2 (2 points) On souhaite enfin tester l'hypothèse :

$$H_0 : a = 2 \tag{9}$$

$$H_1 : m \neq 2 \tag{10}$$

Construisez la région critique associée à ce test pour un niveau de risque de première espèce de 5%. On suppose que pour un échantillon de taille $N = 19$, on observe une moyenne empirique $\overline{X}_N$ égale à 2.4 et une variance empirique corrigée S_N^2 égale à 4. Que pouvez conclure pour un risque de première espèce de 5% ?

Question 3 (2 points) Donnez la formule de la puissance du test bilatéral en fonction de la valeur de a. Que vaut la puissance de ce test pour $a = 3$ sous les hypothèses de la question 2 ?

Question 4 (2 point) Donnez un intervalle de confiance sur la moyenne a au seuil de risque de 5% sous les hypothèses de la question 2.

Exercice 2 *Analyse de retour de campagne marketing. Barème : 4 points.*

Une entreprise souhaite analyser l'impact d'une campagne marketing suivant les canaux de diffusion utilisés (emailing, courriers et appels téléphoniques). Pour cela, elle dispose du tableau de contingence des achats moyens en euros pour 1500 clients ayant été contactés par l'un de ces trois médias. *Exemple : 220 clients contactés par mail ont acheté en moyenne entre 50 et 100 euros de produits.* Testez l'indépendance entre le montant moyen des achats (variable X) et le média (variable Y) au seuil de risque de 10%. *Vous détaillerez précisément votre démarche.*

$X \setminus Y$	Emailing	Courriers	Appels	Total
50-100 euros	220	200	50	470
100 - 200 euros	140	250	100	490
plus de 200 euros	140	50	350	540
Total	500	500	500	1500

Bon courage...