MAÎTRISE DE MATHÉMATIQUES 2003-2004
STATISTIQUES

**8. Tests statistiques : modèle linéaire, tests non-paramétriques**

**8.1.** On observe les variables aléatoires $X_i = \beta + \gamma t_i + \varepsilon_i$, $i = 1, \dots, n$, où $(\varepsilon_1, \dots, \varepsilon_n)$ est un $n$-échantillon de loi $\mathcal{N}(0, \sigma^2)$ et $t_1, \dots, t_n$ sont des réels connus. On veut estimer le paramètre $\theta = (\beta, \gamma, \sigma^2)$.

**a)** Construire l'estimateur du maximum de vraisemblance pour $\hat{\theta}^* = (\widehat{\beta^*}, \widehat{\gamma^*}, \widehat{\sigma^{2,*}})$ de $\theta = (\beta, \gamma, \sigma^2)$.

**b)** Montrer que $\widehat{\beta^*}$ et $\widehat{\gamma^*}$ sont sans biais et calculer leur variances. Trouver leurs lois. Montrer que si $\bar{t} = \frac{1}{n}\sum_{i=1}^n = 0$ alors $\widehat{\beta^*}$ et $\widehat{\gamma^*}$ sont non-corrélées.

**c)** Construire le test de rapport de vraisemblance pour $\beta = 0$ contre $\beta \neq 0$ de niveau $\alpha \in ]0, 1[$. Construire le test de rapport de vraisemblance pour $\gamma = 0$ contre $\gamma \neq 0$ de niveau $\alpha \in ]0, 1[$. Application: $n = 27$ et $\alpha = 0,05$.

**d)** Écrire des intervalles de confiance pour $\beta$, pour $\gamma$ de coefficients de sécurité $1 - \alpha \in ]0, 1[$. Construire une région de confiance pour $(\beta, \gamma)$ de coefficient de sécurité $1-\alpha \in ]0, 1[$. Application: $n = 27$ et $\alpha = 0,05$.

**8.2.** Appliquer les résultats de l'exercice précédent pour les données suivantes :

| | | | | | | |
|---|---|---|---|---|---|---|
| $t$ | 5 | 10 | 15 | 20 | 25 | 30 |
| $x$ | 0,10 | 0,21 | 0,30 | 0,35 | 0,44 | 0,62 |

et $\alpha = 0,05$. Pouvez-vous prédire la valeur de $X$ en $t_0 = 17$.

**8.3.** On observe les variables aléatoires $X_i = \beta + \gamma(t_i - \bar{t}) + \varepsilon_i$, $i = 1, \dots, n$, où $(\varepsilon_1, \dots, \varepsilon_n)$ est un $n$-échantillon de loi $\mathcal{N}(0, \sigma^2)$, $t_1, \dots, t_n$ sont des réels connus et $\bar{t} = \frac{1}{n}\sum_{i=1}^n = 0$.

**a)** Construire le test de rapport de vraisemblance pour $\gamma = 0$ contre $\gamma \neq 0$ de niveau $\alpha \in ]0, 1[$.

**b)** Écrire des intervalles de confiance pour $\gamma$ de coefficients de sécurité $1 - \alpha \in ]0, 1[$.

**8.4.** Appliquer les résultats de l'exercice précédent pour les données suivantes :

| | | | | | |
|---|---|---|---|---|---|
| $t$ | 3,80 | 3,72 | 3,67 | 3,60 | 3,54 |
| $x$ | 1,36 | 1,23 | 1,09 | 0,82 | 0,61 |

et $\alpha = 0,05$. Pouvez-vous prédire la valeur de $X$ en $t_0 = 17$.

**8.5.** Un dé est lancé 600 fois. On observe que les faces apparaissent un certain nombre de fois indiqué dans le tableau suivant :

| | | | | | |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |
| 100 | 94 | 103 | 89 | 110 | 104 |

Au niveau $\alpha = 0,1$ tester si le dé est truqué.

**8.6.** Une espèce de tulipes peut avoir trois couleurs : rouge, rose et violet. L'expérience montre que ces couleurs apparaissent avec probabilités, respectivement $\frac{1}{12}$, $\frac{3}{12}$ et $\frac{8}{12}$. Un statisticien commande 60 tulipes au hasard et il reçoit 6 rouges, 18 roses et 36 violets. Tester la validité du modèle au niveau $\alpha = 0,05$.

**8.7.** Le devoir de statistique d'un groupe sont notés de 0 à 100 mais on rend seulement des qualificatifs : TB pour points entre 90 et 100, B entre 75 et 89, M entre 60 et 74, F entre 50 et 59 et enfin I pour points entre 0 et 49. Le correcteur aimerait avoir une répartition du nombre de points obtenus de loi gaussienne de paramètres 75 et 81. Après avoir corrigé il trouve :

| | | | | |
|---|---|---|---|---|
| $TB$ | $B$ | $M$ | $F$ | $I$ |
| 3 | 12 | 10 | 4 | 1 |

Tester l'hypothèse du correcteur au niveau $\alpha = 0,05$.

**8.8.** Même question pour le tableau suivant :

| | | | | | |
|---|---|---|---|---|---|
| $n \leqslant 90$ | $90 < n \leqslant 100$ | $100 < n \leqslant 110$ | $110 < n \leqslant 120$ | $120 < n \leqslant 130$ | $n > 130$ |
| 10 | 18 | 23 | 22 | 18 | 9 |

Pour estimer $m$ et $\sigma^2$ on pourra prendre les points au milieu des intervalles ainsi que 65 pour le premier intervalle et 160 pour le dernier.

**8.9.** On observe les temps de panne d'une machine. On utilise un appareil qui donne en fait les logarithmes de ces temps de panne. Les observations sont: 2,88; 3,36; 3,50; 3,73; 3,74; 3,82; 3,88; 3,95; 3,95; 3,99; 4,02; 4,22; 4,23; 4,23; 4,23; 4,43; 4,53; 4,59; 4,66; 4,66; 4,85; 4,85, 5,16.

**a)** On veut tester à l'aide d'un test de $\chi^2$ l'hypothèse que la loi des logarithmes des temps de panne est une loi normale de paramètres $\log(50) = 3,912$ et variance 0,25. On pourra décomposer $\mathbb{R}$ en intervalles ayant chacun la probabilité 1/4. Montrer que pour la loi $\mathcal{N}(3,912; 0,25)$ il s'agit des intervalles $]-\infty; 3,575]$, $]3,575; 3,912]$, $]3,912; 4,249]$ et $]4,249; \infty[$. Appliquer le test de $\chi^2$ à un niveau plus petit que 0,3, ainsi qu'à un niveau plus grand que 0,4.

**b)** On reprend les mêmes données mais cette fois-ci on va tester que la loi est une loi normale, mais sans connaître les paramètres. Montrer que 4,150 et 0,2722 sont des estimations sans biais pour l'espérance et la variance. On décompose $\mathbb{R}$ avec les mêmes intervalles qu'au point précédent.Calculer les probabilités de ces intervalles pour la nouvelle loi et appliquer ensuite le test de $\chi^2$ à un niveau plus petit que 0,27. Préciser le nombre de degrés de liberté.

**c)** Utiliser le test de Kolmogorov-Smirnov pour tester l'hypothèse du point **a)**, c'est-à-dire que les observations proviennent d'une loi $\mathcal{N}(3,912;0,25)$.

**8.10.** Un magasin vend deux types de céréales pour le petit déjeuner A et B. Le propriétaire observe que les hommes et les femmes n'achètent pas au hasard et il veut savoir s'il y a indépendance entre le sexe de l'acheteur et la marque de céréales achetée. Il bâtit le tableau suivant:

| | *A* | *B* |
|---|---|---|
| Homme | 9 | 6 |
| Femme | 13 | 16 |

Construire un test de $\chi^2$ d'indépendance et conclure.

**8.11** Deux traitements sont appliqués et on note trois types de résultats:

| | guérison | amélioration | état stationnaire |
|---|---|---|---|
| traitement A | 280 | 210 | 110 |
| traitement B | 220 | 90 | 90 |

Construire un test de $\chi^2$ d'homogénéité pour étudier si les traitements sont différents?

**8.12.** On observe un caractère 21 fois. On veut tester si ce caractère est de loi normale de paramètres 2 et 1 par le test de Kolmogorov-Smirnov : 0,3; 0,7; 0,9; 1,2; 1,4; 1,4; 1,5; 1,5; 1,6; 1,9; 2,0; 2,1; 2,1; 2,3; 2,5; 2,6; 2,7; 3,0; 3,8; 3,9; 4,0.