CORRIGE EXAMENS MODULAIRES

Semestre d'hiver 2008

Filière:	**Économie d'entreprise**	
Module:	**Branches instrumentales**	**Type: Écrit**
Unité de cours:	**Statistiques III**	
Date:	**13 Janvier 2009**	
Nombre de pages:	**6** (sans la présente page de garde)	

Étudiant-e

Nom: Prénom:

Enseignant-e

Nom: Prénom:

Veuillez laisser cet examen agrafé. Si vous choisissez d'enlever l'agrafe, tout manque de page sera sous votre entière responsabilité. La précision de vos calculs doit être de chiffres après la virgule.

Formulaire de 3 pages (recto-verso) autorisé et calculatrice.

Points: Note obtenue:

Problème 1 Compréhension

Le nombre de skieurs sur les pentes enneigées de Nax lors des dimanches ensoleillés suit une loi normale $\mathcal{N}(1000, 50)$.

a) Dimanche prochain sera une journée ensoleillée. Le manager Opti Mist suppose que le nombre moyen de skieurs sera supérieur à 1000. Son frère, le manager Pessi Mist suppose que le nombre moyen de skieurs sera inférieur à 1000. L'un d'eux a-t-il une plus forte probabilité d'avoir raison ? Justifiez.

   Réponse:

   Les deux probabilités sont équivalentes car la loi normale est symétrique. Ces probabilités valent 0.5 chacune.

b) Afin de vérifier si le nombre de skieurs suit bien une loi $\mathcal{N}(1000, 50)$, un échantillon de 10 dimanches a donné comme résultat l'intervalle de confiance à 90% suivant : $[910; 990]$. Interprétez ce résultat, en indiquant ce que signifie un tel intervalle de confiance.

   Réponse:

   Sur tous les échantillons possibles de taille 10, 90% comprennent la vraie moyenne (ici supposée 1000). Comme cet intervalle de confiance ne comprend pas la vraie moyenne, et en supposant qu'effectivement le nombre de skieurs par dimanche ensoleillé suive une loi $\mathcal{N}(1000, 50)$, alors cet intervalle de confiance fait partie des 10% ne comprenant pas la véritable moyenne $\mu = 1000$.

Problème 2 TCS : 7 points

Un responsable de centre d'interventions du TCS désire connaître la durée moyenne des interventions, dont la distribution suit une loi normale. Il relève la durée de 25 interventions de manière aléatoire, et calcule une durée moyenne de 30 minutes, avec un écart type de 10 minutes. Votre tâche est de l'aider à calculer un intervalle de confiance à 95% pour la durée moyenne d'une intervention.

Réponse:

a) (1 pt) Quel est le paramètre d'intérêt ?
   La population d'intérêt est l'ensemble de toutes les interventions et le paramètre d'intérêt est la durée moyenne de toutes les interventions.

b) (1 pt) Quelle loi (distribution) suit la statistique de test, et sous quelle condition ?
   la distribution de Student, ou t-distribution, qui demande une population suivant une loi normale.

c) (2 pt) Quelles sont les valeurs critiques ? Nombre de degrés de liberté vaut 25-1=24. Comme l'intervalle de confiance a un niveau de confiance de 0.95, cela signifie
   $t_{\alpha/2,dl} = t_{0.025,24} = 2.0639$ et $t_{1-\alpha/2,dl} = t_{0.975,24} = -2.0639$

d) (1 pt) Quelles sont les statistiques, et leur valeur, calculées sur l'échantillon ?
   La durée moyenne de l'échantillon est de $\bar{x} = 30$ minutes et l'écart type est de $s = 10$ minutes.

e) (1 pt) Quelle est l'erreur standard de la moyenne ?

$$\frac{s}{\sqrt{25}} = \frac{10}{\sqrt{25}} = 2$$

f) (1 pt) Quelle est l'intervalle de confiance ?

$$\bar{x} \pm t\frac{s}{\sqrt{n}} = 30 \pm 2.0639 \times 2 = 30 \pm 4.1278 = [25.8722; 34.1278]$$

Problème 3 Fertilisant (11 points)

Un producteur de céréales utilise un nouveau fertilisant, qui devrait permettre de réduire la variabilité de la production annuelle. Il fait un test sur 12 champs d'1 hectare et trouve les rendements suivant par hectare, qui sont issus d'une loi normale :

89	85	91	95
95	97	81	89
94	86	87	82

Á partir de ces données sont calculées les valeurs suivantes :
médiane = 89 ; moyenne = 89.25 ; variance = 27.84

a) (8 pts) Le fabricant du nouvel engrais assure que l'écart type des rendements est de 6 par hectare. Le producteur peut-il dire, avec un niveau de confiance de 90%, que le nouvel engrais génère une telle variabilité de la production ?
Réponse:

i) (1 pt) Quel est le paramètre d'intérêt ?
Le paramètre d'intérêt est l'écart type du rendement.

ii) (1 pt) Quelles sont les hypothèses nulles et alternatives ?

$$\begin{aligned} H_0 &: \sigma^2 = 36 \\ H_1 &: \sigma^2 \neq 36 \end{aligned}$$

iii) (1 pt) Quelle loi (distribution) suit la statistique de test, et sous quelle condition ?
La statistique de test suit une loi du χ^2 à 11 degrés de liberté La population doit suivre une loi normale.

iv) (1 pt) Déterminez la ou les valeur(s) critique(s).

$$\chi^2_{1-\alpha/2,n-1} = \chi^2_{0.95,11} = 4.5748 \quad \chi^2_{\alpha/2,n-1} = \chi^2_{0.05,11} = 19.6751$$

v) (1 pt) Déterminez la région de rejet de l'hypothèse nulle. La région de rejet est donnée par toutes les valeurs inférieures à

$$\chi^2_{1-\alpha/2,n-1} = \chi^2_{0.95,11} = 4.5748$$

et toutes les valeurs supérieures à

$$\chi^2_{\alpha/2,n-1} = \chi^2_{0.05,11} = 19.6751$$

vi) (1 pt) Que vaut la statistique de test ? $s^2 = 27.84$ et

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = 8.5069$$

vii) (1 pt) Rejetez-vous l'hypothèse nulle (Justifiez) ?
Comme $\chi^2_{0.95,11} = 4.5748 < \chi^2 = 8.5069 < 19.6751 = \chi^2_{0.05,11}$, on ne peut pas rejeter H_0

viii) (1 pt) Concluez
Ainsi, le producteur ne peut pas conclure à une diminution de la variabilité due au nouvel engrais.

b) (3 pts) Calculez un intervalle de confiance à 90% sur la variabilité de la production annuelle avec le nouvel engrais. Indiquez les étapes (au minimum les valeurs critiques et l'intervalle de confiance).
Indication : certains résultats de la partie précédente peuvent être repris

Réponse:

Valeurs critiques :

$$\chi^2_{1-\alpha/2,n-1} = \chi^2_{0.95,11} = 4.5748 \quad \chi^2_{\alpha/2,n-1} = \chi^2_{0.05,11} = 19.6751$$

En reprenant les données du point précédent, nous trouvons que l'intervalle de confiance à 90% sur la variabilité de la production annuelle est le suivant :

$$\frac{(n-1)s^2}{\chi^2_{0.95}} = \frac{(12-1)27.84}{19.6752} = 15.56485 \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{0.05}} = \frac{(12-1)27.84}{4.5748} = 66.94063$$

Problème 4 Course alpine (5 points)

L'organisateur d'une course en montagne doit répartir les équipes médicales le long du parcours. Pour cela, il pense que la moitié des personnes nécessitent des soins avant 40 km de parcours, et l'autre moitié après 40 km de parcours. Il dispose des données de la précédente course, indiquant pour chacun des 7 coureurs ayant nécessité des soins le nombre de km parcourus. (Ces données ne proviennent pas d'une distribution normale).

31 48 23 56 28 29 44

Sur la base des ces données, l'organisateur a-t-il raison de penser que le 50% des soins doivent être dispensés avant 40km, et 50% après 40 km de course, avec un risque de première espèce de 0.1 ? (Justifiez et indiquez toutes les étapes)

Réponse:

$$H_0 : \tilde{\mu} = 40$$
$$H_1 : \tilde{\mu} \neq 40$$

Valeur	Différence	|Différence|	Rang	$R+$	$R-$
31	-9	9	3		3
48	8	8	2	2	
23	-17	17	7		7
56	16	16	6	6	
28	-12	12	5		5
29	-11	11	4		4
44	4	4	1	1	
				$W_{\text{inf}} = 9$	$W_{\text{sup}} = 19$

Il s'agit d'un test bilatéral, nous pouvons donc utiliser soit la somme des R+ soit la somme des R-. Nous avons $n = 7$ différences non nulles et $\alpha = 0.1$, ce qui donne des valeurs critiques de 3 et 25.
La région de rejet est R= $\mathbb{N}\backslash]3; 25[$
Comme $W_{\text{inf}} = 9 \notin R$ (ou $W_{\text{sup}} = 19 \notin R$), nous ne pouvons pas rejeter l'hypothèse nulle à un niveau de confiance de 0.9

```
x <- c(31, 48, 23, 56, 28, 29, 44)
summary(x)
wilcox.test(x, alternative = "two.sided", mu=40,
            exact = TRUE, conf.level=0.9, conf.int = TRUE)
```

Problème 5 Indépendance (3 points)

Un sondage portant sur la consultation des horoscopes publiés dans les journaux a fourni le tableau croisé suivant. La variable colonne "Fréquence de lecture" indique fréquence de consultation des horoscopes, et la variable ligne "Niveau d'étude" indique le niveau d'étude atteint.

	Jamais	Parfois	Souvent	Toujours
Obligatoire	84	50	44	16
Secondaire	82	64	34	10
Universitaire	44	21	8	5

Un test sur l'indépendance de ces deux variables a été fait à l'aide du logiciel R, dont le résultat est fourni ci-dessous. Écrivez les hypothèses nulle et alternative, concluez en fonction du résultat et commentez. Le risque de première espèce est fixé à 10%.

```
        Pearson's Chi-squared test

data:  horoscope
X-squared = 10.6935, df = 6, p-value = 0.09833
```

Réponse:

a) (1pt) Quelles sont les hypothèses nulle et alternatives ?

H_0 : "Niveau d'étude" et "Fréquence de lecture" sont indépendantes

H_1 : Les deux variables ne sont pas indépendantes

b) (1 pt) Sous quelle condition ce test est-il valide ?
   Les cellules doivent avoir un effectif d'au moins 5

c) (1 pt) Quelle est votre conclusion ?
   La p-valeur est inférieure au seuil critique spécifié. Il y a suffisamment d'évidence pour conclure que les deux variables ne sont pas indépendantes, avec un niveau de confiance de 90%.

```
horoscope <- matrix(c(84,50,44,16,82,64,34,10,44,21,8,5),ncol=4,byrow=TRUE)
rownames(horoscope)<-c("Obligatoire","Secondaire","Universitaire")
colnames(horoscope)<-c("Jamais","Parfois","Souvent","Toujours")
horoscope <- as.table(horoscope)
horoscope
chisq.test(horoscope)
```

Problème 6 Âge des employés (5 points)

L'âge des employés d'une entreprise est considéré (variable âge, unité : année). Un test portant sur l'âge moyen a été réalisé à l'aide du logiciel R. Voici ci-dessous le résultat du test. Le risque de première espèce est fixé à 5%.

```
        One Sample t-test

data:  âge
t = -0.3771, df = 14, p-value = 0.3559
alternative hypothesis: true mean is less than 40
95 percent confidence interval:
     -Inf 42.24088
sample estimates:
mean of x
 39.38949
```

Réponse:

```
âge <- rnorm(15,mean=39,sd=5)
t.test(âge,mu=40, alternative="less")
```

a) (1 pt) Quelles sont les hypothèses nulle et alternative ?

$$
\begin{aligned}
H_0 &: \bar{\text{âge}} \geq 40 \\
H_1 &: \bar{\text{âge}} < 40
\end{aligned}
$$

b) (1 pt) Sous quelle condition ce test est-il valide ?
   La population doit suivre une loi normale.

c) (1 pt) Quelle est la taille de l'échantillon sélectionné ?
   15 car le nombre de degrés de liberté est de 14

d) (1 pt) Qu'est-ce que "$t = -0.3771$" ?
   La valeur de la statistique de test

e) (1 pt) Quelle doit être la conclusion du test ?
Il n'y a pas suffisamment d'évidence pour rejeter H_0, car la p-valeur est supérieure à 0.05 (=risque de première espèce). Donc l'âge moyen des employés de cette entreprise peut être considéré comme étant supérieur ou égal à 40 ans.