TUTORAT UE 4 2013-2014 – Biostatistiques
CORRECTION Séance n°4 – Semaine du 14/10/2013

Lois de Probabilités ; Tests statistiques
M.SABATIER ; M MOLINARI

Séance préparée par les tuteurs du TSN.

QCM n°1: **A, B, C**

A. **Vrai.**
B. **Vrai.** Avec les valeurs de l'énoncé on calcule :
La moyenne : (1,5+1,96+...+0,90+2,80)/ 10= 1,751
La variance observée avec la formule : $s^2=1/n \times \sum xi^2- m^2$ =0,3359 ainsi que l'écart type observé s=0,5796
La variance estimée grâce à la formule : $S^2=n/n-1 \times s^2$ =0,3732 et l'écart type estimé S=0,6109.
C. **Vrai.** On cherche l'intervalle au seuil de 95%.
n<30 donc on cherche t dans la table de Student à n-1 ddl pour alpha=5% => t=2,262
grâce à la formule $[m-t_{\alpha/2,n-1}*S/\sqrt{n} ; m+t_{\alpha/2,n-1}*S/\sqrt{n}]$, on trouve que l'IC est de [1,3140 ; 2,1880].
D. Faux. On cherche l'intervalle au seuil de 90%
n<30 donc on cherche t dans la table de Student à n-1 ddl pour alpha=10% => t=1,833.
grâce à la formule $[m-t_{\alpha/2,n-1}S/\sqrt{n} ; m+t_{\alpha/2,n-1}S/\sqrt{n}]$, on trouve que l'IC est de [1,397 ; 2,105] .
On trouve les valeurs de l'énoncé si l'on prend l'écart type observé et non celui estimé !
E. Faux. Si on avait eu 45 patients, alors n étant supérieur à 30, on n'aurait pas eu à supposer que la variable suit une loi normale. Le calcul de l'IC aurait utilisé le c de la table de la loi normale et non pas le t de la table de Student.
F. Faux

QCM n°2 : **A;C;D;E**

A. **Vrai.** Pour cela on utilise la formule suivante $S^2=n/n-1 \times s^2$ avec s^2=5,76.
B. Faux. n>30. On n'a donc pas besoin de supposer que la v.a.r suit une loi normale puisque l'effectif est assez grand et donc que le théorème central limite s'applique.
C. **Vrai**. On lit le c dans la table de l'écart réduit (loi normale centrée réduite)
D**. Vrai**. On prend c=1,96 et S=2,4310 et on calcule grâce à la formule suivante : $[m-c_{\alpha/2} * S/\sqrt{n} ; m+c_{\alpha/2} * S/\sqrt{n}]$
E. Vrai. Pour un seuil de confiance de 90% on lit dans le tableau que c=1,645. On utilise la même formule que ci dessus et on trouve bien [8,77 ; 10,03]
F. Faux

QCM n°3 : **A, B, E**

A**. Vrai.** On a en effet n<30, cette hypothèse est donc obligatoire.
B. **Vrai.** On a m=2, la variance observée= $0,5^2$ =0,25 et donc $S^2=n/n-1 \times s^2$ =0,2639.
On trouve t=2,101 dans la table de student à n-1=18.
On utilise la formule $[m-t_{\alpha/2,n-1}*S/\sqrt{n} ; m+t_{\alpha/2,n-1}*S/\sqrt{n}]$ et on trouve bien [1,752 ; 2,248].
C. Faux. Pour la variance on utilise la loi du chi deux à n-1 ddl.
D. Faux. C'est l'IC pour l'écart type !
En effet, on va utiliser la formule suivante : $[(n-1) \times S^2/b ; (n-1) \times S^2/a]$ qui nous donne l'intervalle de confiance pour la variance. On prend a=10,865 tel que $P((n-1)S^2/\sigma^2 \leq a)=\alpha/2$ et b=25,989 tel que $P((n-1)S^2/\sigma^2 \leq b)=1-\alpha/2$.

Pour la variance on obtient [0,1828 ; 0,4372] ≈ [0 ,18 ;0,44] et on a [0,4276 ; 0,6612] pour l'écart type (on prend la racine carrée).

E. Vrai.

F. Faux.

QCM n°4 : **A, D, E**

A. **Vrai.**

B. Faux. On vérifie si on peut approximer par une loi normale: n=80>30, n x p = 80 x 0,625 = 50>5
n x (1 - p) = 80 x 0,375 = 30 >5, on peut donc faire l'approximation.

C. Faux. [p0 – C $_{\alpha/2}$ $\sqrt{(\mathrm{p0}\,(1-\mathrm{p0})/\,\mathrm{n})}$; p0 + C $_{\alpha/2}$ $\sqrt{(\mathrm{p0}\,(1-\mathrm{p0})/\,\mathrm{n})}$]
On lit dans la table de l'écart réduit C $_{\alpha/2}$ = 1,96 d'où l'intervalle [0,5189 ; 0,731].

D. **Vrai.** On multiplie chaque borne de l'intervalle par n = 80 d'où l'intervalle [41,5 ; 58, 5].

E. **Vrai**. Plus on augmente la taille de l'intervalle, plus le risque de tomber à l'extérieur diminue : Ex: pour α= 3%, on a [0, 508 ; 0, 742]

F. Faux.

QCM 5: **B, E**

A. Faux. H1: m0≠µ. H0 est ce que l'on veut rejeter donc H0: µ=42.4.

B. **Vrai.** On a 80 ≥ 30 donc on fait une approximation par le test de l'écart réduit, mais on pourrait aussi utiliser le test de Student.

C. Faux. T_{obs} = (43,2 - 42,4) / (3,3 /$\sqrt{80}$) = 2,168.

D. Faux. On rejette bien H0 pour α= 5% car T_{obs} ≥ 1,96
On ne rejette pas pour α= 2% car T_{obs} ≤ 2,326
On n'accepte JAMAIS H0.

E. **Vrai.** Dans la table on lit T_α= 2,17 pour α= 0,03.

F. Faux

QCM n°6 **: B, E**

A. Faux. On utilise un test de student mais pour données appariées donc avec 9 degrés de liberté.

B. **Vrai.** Calcul de la variable (x1 - y1): 5 ; 2 ; -4 ; 18 ; -3 ; 0 ; -1 ; 6 ; -3 ; -1
On trouve une moyenne de 1, 9.

C. Faux. Variance = [5^2 + 2^2 + $(-4)^2$...] x 1/9 -10/9 $\underline{x}^2$= 47,2-4.0111=43.2111
Donc l'écart type = 6, 57

D. Faux. T_{obs} = 1,9/ (6,87 /$\sqrt{10}$) = 0,915 or Tα = 2,262 donc on ne rejette pas H0.

E. **Vrai**. Le fait d'avoir des données appariées ne changent rien, on crée donc une nouvelle variable qu'on compare à une valeur de référence donc on peut utiliser le test de student ou de l'écart réduit pour n>30.

F. Faux.

QCM n°7: **A, D, E**

A. **Vrai.** (cf. cours)

B. Faux. on utilise des estimateurs qui peuvent être ponctuels ou par intervalle

C. Faux ; Un intervalle de confiance permet d'estimer un paramètre avec un risque alpha de première espèce connu.

D. Vrai.

E. Vrai.

F. Faux.

QCM n°8: **B**

A. Faux: Pour les populations on utilise les lettres grecques (µ) et non m_1 et m_2!!!

B. **Vrai.**

C. Faux: on doit vérifier l'égalité des variances (grâce au test de Fisher) et la variable doit suivre une loi normale dans chaque population. n doit être inférieur à 30.

D. Faux. Il n'y a pas de degré de liberté avec le test de l'écart réduit.

E. Faux. On les utilise aussi pour les échantillons appariés mais en travaillant sur les différences.
F. Faux.

QCM n°9: **D**

A. Faux. Le test de Fisher permet de comparer des variances, le test exact de Fisher permet de comparer 2 pourcentages observés.
B. Faux. On veut vérifier les conditions d'application de certains tests paramétriques qui exigent que les variances soient identiques (test de student)
C. Faux. T_{obs} est un rapport de variances et non d'écart type!!
D. **Vrai**. Le test F permet de vérifier les conditions d'utilisation du test de Student pour comparer 2 moyennes observées sur des échantillons indépendants.
E. Faux. Attention car ANOVA permet de comparer des moyennes par l'analyse de la variance !
F. Faux.

QCM n°10: **D, E**

A. Faux. Ce sont les effectifs attendus (théoriques) qui doivent être supérieurs à 5. Ici c'est le cas, on peut donc appliquer le test du X^2.
B. Faux. Le test du chi2 d'homogénéité et de l'écart-réduit sont permis, on préfère donc les utiliser. Le test exact de Fisher est aussi permis mais sera plutôt employé pour de petits échantillons
C. Faux. On réalise d'abord le tableau des effectifs théoriques. Chaque case est le résultat du produit de la ligne * le produit de la colonne divisé par le nombre total de l'échantillon.

34*127/200= 21,59	166*127/200 = 105,41
34*73/200 = 12,41	166*73/200 = 60,59

On calcule ensuite la "distance" entre les deux tableaux. Oij sont les effectifs calculés et Eij ont les effectifs théoriques.
$Tobs= \Sigma(Oij- Eij)^2/Eij$
$= (18-21,59)^2/21,59 + (109-105,41)^2/105,41 + (16-12,41)^2/12,41 + (57-60,59)^2/60,59$
$= 1,97$

D. **Vrai.**
E. **Vrai.** $tobs<\chi^2$, on ne rejette pas H0.
F. Faux.

QCM n°11: **A, B, C**

A. **Vrai.** p=13/75= 17,33%
B. **Vrai.** Les conditions d'applications sont np>5 et n(1-p)>5
C. **Vrai.** $T_{obs} = (0,1733 – 0,09)/ \sqrt{(0,1733 * (1 - 0,1733)/75)} = 1,91$
D. Faux. La formulation de l'hypothèse H0 impose de faire un test en bilatéral, on n'a pas d'a priori sur le sens de la différence.

On a Tobs = 1,91<1,96, on ne rejette donc pas H0. Attention car même si on avait pu rejeter H0, on n'aurait pas pu conclure que les PACES sont plus susceptibles de déprimer ; on aurait seulement pu conclure à une différence significative entre le nombre de déprimés dans la population générale et chez les PACES.

E. Faux. $T_{obs}= (0,1733 - 0,09)/\sqrt{(0,1733 * (1 - 0,1733)/80)} = 1,968$ qui est supérieur à 1,96.Cela montre qu'avec un plus grand échantillon, une différence est plus facile à montrer.
F. Faux.

QCM n°12 : **D, E**

A. Faux. Ils doivent être supérieurs ou égaux à 5.

B. Faux. Le χ^2 suit une loi à (r-1)ddl avec r le nombre de classes. Donc 3 ddl pour cet exemple.
C. Faux. D'abord il faut calculer les effectifs attendus d'après les fréquences de la population française.

A	0,44*2369 = 1042,36
O	0,45*2369 = 1066,05
B	0,08*2369 = 189,52
AB	0,03*2369 = 71,07

$X^2obs = (1018-1042{,}36)^2/1042{,}36 + (1113-1066{,}05)^2/1066{,}05 + ... = 23{,}30$
Attention, 23.55 c'est ce que l'on trouve lorsqu'on utilise les valeurs approchées, donc attention à vos arrondis.
D. **Vrai.** On lit dans la table à α=5% et 3ddl. 23,30>7,815
E. **Vrai.** On lit dans la table à α=0,1% et 3ddl. 23,30>16,266.
F. Faux.

QCM 13 : **A, B, C, E**
A. **Vrai.**
B. **Vrai.** $\chi^2\text{obs} = \frac{(f-g)^2}{f+g} = \frac{(50-30)^2}{50+30} = 5$
C. **Vrai.** χ^2_α à 1 ddl et à 5% est 3.841
D. Faux. $\chi^2_{obs} > \chi^2_\alpha$ donc on rejette H_0
E. **Vrai.** $X^2_\alpha = 5{,}412$ donc $\chi^2_{obs} < \chi^2_\alpha$ donc on ne rejette pas H_0
F. Faux.

QCM n°14 : **A, C, D**
A. **Vrai.** On cherche un lien
B. Faux. Tobs= $\frac{\sum(Oij-Eij)^2}{Eij}$
C. **Vrai.** (r-1)(k-1) ddl= (3-1)(2-1)=2 ddl
D. **Vrai.** $\chi^2\alpha$ à 2 ddl et α=5% est de 5.991 donc $\chi^2_{obs} > \chi^2_\alpha$ donc on rejette H_0 donc on met en évidence un lien statistique entre le système rhésus et la nature de la maladie
E. Faux. $\chi^2\alpha$ à 2 ddl et α=1% est de 9.210 donc $\chi^2_{obs} < \chi^2_\alpha$ donc on ne rejette pas H_0 donc on ne met pas en évidence un lien statistique entre le système rhésus et la nature de la maladie
F. Faux.