TUTORAT UE 4 2014-2015 – Biostatistiques
CORRECTION Séance n°5 – Semaine du 27/10/2014

Tests statistiques
M. Molinari

QCM n°1 : C, D, E

A. Faux. Ici, on étudie les 8 mêmes étudiants avant leur entrée en PACES puis après leur entrée en PACES donc les données sont appariées.
B. Faux. Le test du X² de Mac Nemar s'utilise pour des données qualitatives, or ici les données sont quantitatives continues (une durée). On utilise le test de Student pour données appariées.
C. **Vrai**. On construit d'abord une ligne pour les différences :

x_1-x_2	+4,2	-2,8	+0,5	-2,1	+0,7	-1,2	+4,4	+2

On calcule la moyenne de l'échantillon des différences en tenant compte des signes : m=0,7125.
On calcule la variance : $\sum(X_1-X_2)^2$=55,43 donc $s^2=\frac{55.43}{8}$ - $(0,7125)^2$=6,421.
On calcule l'écart type estimé dans la population : S= $\sqrt{\frac{8}{7} \times s^2}$ = 2,709.
On peut ensuite calculer $t_{obs}=\frac{0,7125-0}{\frac{2,709}{\sqrt{8}}}$ = 0,7439 soit 0,74 à 10^{-2} près.

D. **Vrai**. Si α=0,05 alors t_α=2,365 et si α=0,3 alors t_α=1,119. Dans les deux cas $t_{obs}<t_\alpha$, donc on ne peut pas rejeter H0.
E. **Vrai**. Pour les tests statistiques, augmenter la taille de l'échantillon permet d'augmenter la puissance du test.

QCM n°2 : A, D, E

A. **Vrai**. On compare une moyenne théorique à une moyenne observée dans un échantillon de grande taille (>30).
B. Faux. Etant donné que l'échantillon est de grande taille (>30), ce n'est pas nécessaire car la moyenne de la variable va converger vers une loi Normale. On doit supposer que la variable suit une loi normale si l'échantillon est inférieur à 30 sujets (pour le test de Student).
C. Faux. $tobs = \frac{2,01-1,92}{\frac{\sqrt{0,1024}}{\sqrt{72}}} = 2,39$.
D. **Vrai**. On lit t_α=1,96 dans la table de la loi normale (α=0,05), $t_\alpha < t_{obs}$ donc on peut rejeter H0.
E. **Vrai**. On fait une lecture inverse de la table de l'écart réduit et on cherche à encadrer t_{obs} : $2,326<t_{obs}<2,576$ donc la p-value est comprise entre 0,01 et 0,02.

QCM n°3 : **A, B, C, D**

A. **Vrai**. L'hypothèse composite ou complémentaire est l'hypothèse H1.

B. **Vrai**. Le test de Student permet de comparer deux moyennes sur deux échantillons indépendants. Il nécessite des conditions d'applications : normalité des valeurs (car petits échantillons) et égalité des variances. Il faut donc vérifier l'égalité des variances avec le test F.

Test F : $tobs = \frac{\frac{21}{20}\times 0{,}29^2}{\frac{16}{15}\times 0{,}21^2} = 1{,}88$ Et dans la table de Fisher pour α=0,05 on a t_α=2,33. t_{obs}<t_α donc on peut accepter l'hypothèse de l'égalité des variances : on peut appliquer le test de Student.

C. **Vrai**. Test de Student: on calcule d'abord $S^2 = \frac{(n1-1)S1^2+(n2-1)S2^2}{n1+n2-2} = \frac{20\times\frac{21}{20}\times 0{,}29^2+15\times\frac{16}{15}\times 0{,}21^2}{21+16-2} = 0{,}07062$ puis $t_{obs} = \frac{2{,}74-2{,}55}{\sqrt{\frac{S^2}{21}+\frac{S^2}{16}}}$ =2,15.

D. **Vrai**. On lit t_α dans la table de Student à 21+16-2=35 ddl, soit la dernière ligne (+∞, converge vers la loi Normale) : t_α=1,96. On rejette H0 au risque de 5%.

E. Faux. On a démontré ici une association statistique, mais il faut un faisceau d'arguments (incluant le lien statistique) pour aboutir à une conclusion clinique. Les tests ne permettent que d'établir une conclusion statistique.

QCM n°4 : **F**

A. Faux. Le test ANOVA permet la comparaison de moyennes de plusieurs échantillons indépendants.

B. Faux. Parmi les conditions d'application, il y a aussi l'égalité des variances des différents échantillons.

C. Faux. Le test ANOVA est un test global : l'hypothèse H1 suppose qu'au moins l'une (parfois plus) des moyennes (peu importe laquelle) des différents échantillons est différente d'une autre.

D. Faux. *Cf. item C*.

E. Faux. Sous H0, la statistique de test suit une distribution de la loi de Fisher.

F. **Vrai**.

QCM n°5 : **D, E**

A. Faux. On souhaite comparer deux moyennes entre deux échantillons indépendants.

B. Faux. Le test de Student pour données indépendantes nécessite des conditions d'application : la variable doit suivre une loi normale, égalité des variances.
Pour tester l'égalité des variances, on fait le test F :
$t_{obs} = \frac{0{,}23^2}{0{,}12^2} = \frac{0{,}23^2}{0{,}12^2} = 3{,}67.$
t_α= 2,98 avec la table du test de Fisher. On a $t_\alpha < t_{obs}$ donc on doit rejeter l'hypothèse H0 : les variances ne sont pas égales, on ne peut pas appliquer le test de Student. Il faut utiliser le test d'Aspin Welch qui ne nécessite qu'une condition d'application : la variable suit la loi Normale.

C. Faux. Pour le test d'Aspin Welch, on lit t_α dans la table de la loi de Student.
Pour connaitre le degré de liberté m : on calcule $c = \frac{\frac{0{,}23^2}{11}}{\frac{0{,}23^2}{11}+\frac{0{,}12^2}{11}} = \frac{529}{673}$ puis $\frac{1}{m} = \frac{c^2}{10} + \frac{(1-c)^2}{10}$ soit m=15,06 donc le degré de liberté est 15.

D. **Vrai**. $t_{obs} = \frac{0{,}56-0{,}48}{\sqrt{\frac{0{,}23^2}{11}+\frac{0{,}12^2}{11}}}$=1,02.

E. **Vrai**. Pour α=0,05, t_α=2,131 donc on ne peut pas rejeter H0.

QCM n°6 : **B, C, E**

A. Faux. Les données sont indépendantes, on utilise un test de Mann Whitney.
B. **Vrai**.
C. **Vrai**.
D. Faux. « Les femmes et les hommes ont un seuil de douleur égal » (ce qu'on veut rejeter).
E. **Vrai**.

Valeurs	9	12	13	15	15	16	16	17	17	18
Rang	1	2	3	4,5	4,5	6,5	6,5	8,5	8,5	10

$S_{rg\ Hommes}$=31,5 et $S_{rg\ Femmes}$=23,5, on a donc $U_{Hommes}=5x5+\frac{5x6}{2}-31{,}5=8{,}5$ et $U_{Femmes}= 5x5+\frac{5x6}{2} - 23{,}5=16{,}5$.
U=8,5 et U_α=2 à 5% et 0 à 1% (lire à (0 ; 5)), soit $U>U_\alpha$, on ne rejette pas H0.

QCM n°7 : **C**

A. Faux. Les données sont appariées car on travaille sur les mêmes personnes, avant et après une campagne de prévention. On utilise donc le test de Wilcoxon
Remarque : on aurait pu utiliser le test des signes, mais Wilcoxon est plus puissant.
B. Faux. On teste l'hypothèse nulle « les deux distributions sont identiques » ou « la campagne de prévention n'apporte aucun changement ».
C. **Vrai**. On travaille sur la différence avant-après (ou l'inverse)

Valeurs	0	4	5	-2	3	6	-1	6
Rang	-	4	5	2	3	6,5	1	6,5

$S_{rang\ +}$=25 et $S_{Rang\ -}$ =3 donc S=3.
D. Faux. Comme m=effectif-différences nulles=7<30, on lit directement dans la table le S_{seuil}, qui est égal à 2 (à 5%) et 0 (à 2%), donc on ne rejette H0 ni au risque de 5%, ni au risque de 2%.
E. Faux. *Cf. item D*.

QCM n°8 : **B**

A. Faux. C'est l'inverse**.**
B. **Vrai**.
C. Faux. T=30 (la valeur minimale entre les différences positives et négatives), et m=100-20=80, donc m est grand. On fait alors l'approximation par la loi normale N (0,5m ; 0,5√m), soit N (40 ; $2\sqrt{5}$).
$Z=\frac{T+0{,}5-\mu}{\sigma} = \frac{30{,}5-40}{2\sqrt{5}}$ =2,12.
D. Faux. On lit dans la table un t_α=1,96 (à 5%) ou 2,576 (à 1%). Dans le premier cas, $|Z|>t_\alpha$, on rejette H0. Mais pour α=0,01, $|Z|<t_\alpha$ donc on ne rejette pas l'hypothèse H0.
E. Faux. *Cf. item D*.

QCM n°9 : **A**

A. **Vrai**. On compare deux proportions observées, on peut alors utiliser le test de l'écart réduit ou du X² (si les conditions sont respectées pour l'un, elles le sont pour l'autre obligatoirement). Ici, tous les effectifs théoriques sont supérieurs à 5, les conditions sont donc respectées
B. Faux. L'hypothèse H0 à tester est « le type d'opération n'a pas d'influence sur le risque de complications » (ce qu'on veut rejeter).
C. Faux. Sous l'hypothèse H0, les deux pourcentages théoriques obtenus à partir des échantillons observés sont égaux.
D. Faux. On va faire un test de l'écart réduit pour deux proportions observées :
$p=\frac{0{,}26x50+0{,}2x35}{35+50}$=4/17. Donc $t_{obs}=\frac{0{,}26-0{,}2}{\sqrt{(\frac{4}{17}x\frac{13}{17}x\left(\frac{1}{50}+\frac{1}{35}\right))}}$ = 0,642. C'est donc le t_{obs} qui est égal à 0,642 et non le t_α ! On lit le t_α dans la table de l'écart réduit, et on trouve, pour un risque de 5% une valeur de 1,96
E. Faux. Comme $t_{obs}< t_\alpha$, on ne rejette pas l'hypothèse nulle, on ne peut donc pas conclure sur l'éventuel lien entre la technique d'opération et le risque de complications. */!\ de plus, on réalise ici un test en bilatéral ; nous n'avons donc pas la possibilité de conclure sur le sens de la différence*.

QCM n°10 : **A, B**

A. **Vrai**. Nos échantillons sont appariés, et les paires discordantes, f=12 et g=3 ont une somme supérieure à 10.
B. **Vrai**.
C. Faux. L'hypothèse H0 à tester est « les distributions sont équivalentes » (ce que l'on veut rejeter).
D. Faux. $X^2_{obs}=\frac{(f-g)^2}{f+g}=\frac{(12-3)^2}{15}=5{,}4$.
E. Faux. On lit la table du X^2 à 1ddl, on trouve un X^2_α=3,841 (à 5%) et X^2_α=5,41 (à 2%), donc on peut rejeter l'hypothèse nulle au risque de 5%. En revanche, on ne peut pas la rejeter au risque de 2%.

QCM n°11 : **B, E**

A. Faux. On est dans le cas d'une comparaison d'une distribution observée à une distribution théorique. Comme il y a plus de 2 proportions, on ne peut pas utiliser un test de l'écart réduit, on utilisera ici le X^2 (d'ajustement).
B. **Vrai**.
C. Faux*. Cf. item B*.
D. Faux. On peut tracer le tableau de X^2 suivant :

	R Blanc	R Marteau	R Baleine	R Pèlerin
Effectifs Obs	19	57	34	30
Effectifs Thé	28	49	21	42

$X^2_{obs}=\frac{(19-28)^2}{28}+\frac{(57-49)^2}{49}+\frac{(34-21)^2}{21}+\frac{(30-42)^2}{42}$=15,68
X^2_α (qu'on lit à 4-1=3ddl)= 7,815 (à 5%) et 11,345 (à 1%). Pour ces deux risques, on rejette H0.
E. **Vrai**.

QCM n°12 : **A, E**

A. **Vrai**. On va comparer des pourcentages théoriques des populations à partir des échantillons observés. De plus, tous les effectifs théoriques sont supérieurs à 5 (*cf. tableau item C*).
B. Faux. L'hypothèse H0 à tester est « les proportions d'apparition de cancer sont égales dans les deux **populations** ».
C. Faux. (*les nombres entre parenthèses sont les effectifs théoriques*)

	Traitement N	Traitement R	Total
Cancer	6 (5,25)	8 (8,75)	14
Pas de cancer	24 (24,75)	42 (41,25)	66
Total	30	50	80

Donc $X^2_{obs}=\frac{0{,}75^2}{5{,}25}+\frac{0{,}75^2}{8{,}75}+\frac{0{,}75^2}{24{,}75}+\frac{0{,}75^2}{41{,}25}$ = 0,208. On lit le X^2_α à (2-1)(2-1)ddl et on trouve 3,841 pour un risque de 5%.
D. Faux. Comme $X^2_{obs}< X^2_\alpha$, on ne peut pas rejeter H0.
E. **Vrai**.

QCM n°13 : **A, D**

A. **Vrai**.
B. Faux. Dans ce cas, on peut utiliser le test de Student ou le test de l'écart réduit (selon la taille des échantillons) : par exemple, si la variable qualitative est le sexe et la variable quantitative le poids. On peut également employer le test ANOVA.
Remarque : Pour tester l'indépendance de deux variables qualitatives, on utilise le X^2 d'indépendance.
C. Faux. La droite de régression permet de prédire la valeur moyenne de Y pour un X donné.
D. **Vrai**. Plus X et Y sont liés, plus ρ est proche de 1 donc (1- ρ^2) tend vers 0 : plus X et Y sont liés, plus la variance liée tend vers 0.
E. Faux. *Cf. item D.*

QCM n°14 : **A, C, E**

A. **Vrai**. Ici, on étudie comment varie le volume sanguin (Y) en fonction du poids de l'individu (X). Donc l'équation de la droite de régression est $y - 4{,}9 = 0{,}66 \times \frac{\sqrt{4}}{\sqrt{25}} \times (x - 68)$ = 0,264(x-68).
B. Faux. *Cf. item A.*
C. **Vrai**. La variance liée concerne Y, ici le volume sanguin. Elle vaut (1-0,66²) x 4=2,26.
D. Faux. *Cf. item C*.
E. **Vrai**. On utilise l'équation de la droite de régression pour calculer y sachant que x=72kg :
y=0,264(72-68) + 4,9=5,96L.