TUTORAT UE 4 2014-2015 – Biostatistiques
CORRECTION Colle n°2 – Semaine du 17/11/2014

Lois de probabilités 2 – Tests statistiques - Epidémiologie
M. Sabatier – M. Molinari – Mme Fabbro-Peray

Mardi 18 et 25 Novembre
ELECTIONS ETUDIANTES

Pensez à vous munir de votre carte étudiante pour aller voter !

QCM n°1 : **A, C, D, E**

A. **Vrai**. La proportion est égale à p=$\frac{15}{45}=\frac{1}{3}$. Au risque 10%, en lisant dans la table de la loi Normale car n>30, np>5 et nq>5, c=1,645. $\left[\frac{1}{3}-1{,}645\times\sqrt{\frac{\frac{1}{3}\times\left(1-\frac{1}{3}\right)}{45}};\frac{1}{3}+1{,}645\times\sqrt{\frac{\frac{1}{3}\times\left(1-\frac{1}{3}\right)}{45}}\right]=[0{,}2177;0{,}4489]$.

B. Faux. *Cf. item A*. *Remarque :« Au seuil de 90% » est une autre façon de dire « au risque de 10% ».*

C. **Vrai**. Même formule qu'à l'item A, avec un c toujours lu dans la table de la loi Normale mais au risque 5% donc c=1,96.

D. **Vrai**. Même formule que précédemment avec un c toujours lu dans la table de la loi Normale mais au risque 2% donc c=2,326.

E. **Vrai**.

QCM n°2 : **A, B**

A. **Vrai**. Car n<30.

B. **Vrai**. Comme n<30, on commence d'abord par calculer notre écart-type estimé :

S = $\sqrt{\frac{n}{n-1} \times s^2}$ $\sqrt{\frac{15}{14} \times 9}$ =3,1053.

IC à α=5% = $\left[m - t_\alpha \times \frac{S}{\sqrt{n}} ; m + t_\alpha \times \frac{S}{\sqrt{n}} \right]$ = $\left[4,5 - 2,145 \times \frac{3,11}{\sqrt{15}} ; 4,5 + 2,145 \times \frac{3,11}{\sqrt{15}} \right]$ = [2,780 ; 6,220].

On lit le t_α dans la table de Student à n-1 ddl, soit ici à 14 ddl.

C. Faux. Comme n>30, on n'a plus besoin d'émettre une hypothèse de normalité pour calculer l'IC de la moyenne, car le Théorème Central Limite s'applique.

D. Faux. IC à α=10% = $\left[m - t_\alpha \times \frac{S}{\sqrt{n}} ; m + t_\alpha \times \frac{S}{\sqrt{n}} \right]$ = $\left[8,5 - 1,645 \times \frac{7}{\sqrt{80}} ; 8,5 + 1,645 \times \frac{7}{\sqrt{80}} \right]$ = [7,213 ; 9,787].

<u>Remarque</u> : on considère ici que s ≈ S car n>30.

E. Faux. *Cf. item D*.

QCM n°3 : **B**

A. Faux. Sur une population.

B. **Vrai**.

C. Faux. La p-value peut également se trouver dans la table de Student, du Chi-deux ou autre, en fonction du test auquel elle se réfère.

D. Faux. En rejetant H0.

E. Faux.

QCM n°4 : **B, D, E**

A. Faux. Les échantillons ont des effectifs inferieurs à 30 et les variances des deux populations sont égales. On utilise le test de Student.

B. **Vrai**. *Cf. item A*.

C. Faux. Les variances considérées sont égales, donc inutile de faire ce test.

D. **Vrai**. s_1^2=0,0424 et s_2^2=0,0852

Donc $S_1^2 = \frac{10}{9} \times 0,0424$ =0,0471 et $S_2^2 = \frac{12}{11} \times 0,0852$ =0,09295.

$$S^2 = \frac{9 * (\frac{10}{9} * 0,206^2) + 11 * (\frac{12}{11} * 0,292^2)}{10 + 12 - 2} = 0,0724$$

$t_{obs} = \frac{|1,3 - 0,80|}{\sqrt{\frac{0,0724}{10} + \frac{0,0724}{12}}}$ =4,34.

On lit $t\alpha$ à 12+10-2 ddl.

t_α=2,086 $t_{obs} > t_\alpha$, on rejette H0.

E. **Vrai**. Pour un α=0,1%, t_α=3,850. $t_{obs} > t_\alpha$, on rejette H0.

QCM n°5 : **C**

A. Faux. L'hypothèse nulle est : la population de personnes stressées (dont est issue l'échantillon) perd autant de cheveux que la population en général. L'hypothèse porte toujours sur la population.
B. Faux. L'effectif est inférieur à 30, donc on utilise le test de Student pour comparer une moyenne observée à une moyenne théorique.
C. **Vrai**. On calcule $t_{obs}=\frac{66,4-60,3}{\frac{\sqrt{\frac{18}{17}\times 10,3^2}}{\sqrt{18}}}$=2,44 et on lit t_α=1,740 dans la table de Student (ddl=n-1=17 et α=0,1). Comme $t_{obs} > t_\alpha$, on rejette H0.
D. Faux. On lit $t_\alpha = 2,567 > t_{obs}$. Dans ce cas, on ne peut pas rejeter H0, mais cela ne veut pas dire qu'on l'accepte : on ne peut jamais accepter H0, on peut seulement rejeter/ne pas rejeter H0.
E. Faux. Pour estimer la p-value, on fait une lecture inverse de la table de Student (avec ddl=17). On cherche à encadrer t_{obs} : $2,110 < t_{obs} < 2,567$. Donc la p-value est comprise entre 2% et 5 %.

QCM n°6 : **A, D, E**

A. **Vrai**.
B. Faux. Ici, les hypothèses concernent la moyenne.
C. Faux. n_1 et n_2 >30, on réalise un écart réduit en lisant dans la table de l'écart réduit.
D. **Vrai**. $t_{obs}=\frac{|10,4-7,1|}{\sqrt{\frac{4,7}{88}+\frac{6,4}{85}}} = 2,4934$.
E. **Vrai**. On encadre la p-value par lecture inverse de la table. $1,96 < t_{obs} < 2,576$ d'où 0,01<p-value<0,05 et donc on peut rejeter H0 au risque de 5%.

QCM n°7 : **B, C**

A. Faux. χ^2 d'ajustement à (R-1)ddl (avec R=Nombre de classes)= 3-1=2 ddl.
B. **Vrai**.

	Effectifs attendus théoriques
Absents	0,25 x 210 = 52,5
Malades	0,45 x 210 = 94,5
Sobres	0,30 x 210 = 63

C. **Vrai**. $\chi^2_{obs}=\sum\frac{(O_{ij}-E_{ij})^2}{E_{ij}}=\frac{(49-52,5)^2}{52,5}+\frac{(90-94,5)^2}{94,5}+\frac{(71-63)^2}{63}$=1,46.
D. Faux. χ^2_α=5,991. On lit dans la table à 2ddl.
E. Faux. $\chi^2_\alpha > \chi^2_{obs}$, On ne rejette pas H0 à 5% : les distributions ne sont pas significativement différentes.

QCM n°8 : **C, D, E**

A. Faux. L'énoncé ne nous donne aucune hypothèse, on fera donc un test non paramétrique, avec des données appariées : test de Wilcoxon (plus puissant qu'un test des signes, qui pourrait également être employé).
B. Faux. H0 :«le passage en P2 n'a pas d'influence sur le nombre de soirées arrosées par mois dans cette <u>population</u> d'étudiants»
C. **Vrai**.
D. **Vrai**. Les sommes des rangs sont données, S= minimum (S_{rg+} ; S_{rg-}) = 1. On va lire le S_{seuil} dans la table de Wilcoxon en bilatéral à N=différences non nulles=8. On lit S_{seuil} =4 (à 5%), et comme $S<S_{seuil}$, on rejette H0 au risque de 5%.
E. **Vrai**. *Cf. item D.*

QCM n°9 : **A, C, D**

A. **Vrai**. Prévalence=cas existants au début de la période + cas apparus au cours de la période=cas existants à la fin de la période=$\frac{45}{65} \approx 0{,}69$.

B. Faux. *Cf. item A. /!\ attention à ne pas mettre au dénominateur la population à risque si on demande la prévalence en France.*

C. **Vrai**. La létalité correspond à la part des décès dus à la maladie ($\frac{2{,}5}{3} \approx 0{,}833$ millions) chez les personnes atteintes de la maladie (45 millions).

D. **Vrai**. **/!** attention à bien mettre au numérateur le nombre de décès dus à la maladie d'amour ($\frac{2{,}5}{3}$ millions) et au numérateur la population étudiée.

E. Faux. Le calcul du nombre de cas incidents nécessite de prendre en compte les personnes ayant guéri et les personnes mortes au cours de l'année 2003 donc incidence = 45millions – 40millions + 2,5 millions +1,5 millions= 9 millions.

QCM n°10 : **F**

A. Faux. Il s'agit d'une étude cas-témoins rétrospective (absence de suivi dans le temps).

B. Faux. Nous sommes en effet dans le cadre d'une enquête cas/témoins, cependant, on ne peut pas distribuer au hasard les sujets dans le groupe cas et dans le groupe témoins. On doit nécessairement placer les cas d'un côté et les témoins de l'autre. Par contre, on essaie de prendre au sein des cas et des témoins des sujets représentatifs de la population.

C. Faux. Dans la mesure où je sélectionne les effectifs de malades et de non malades, je ne vais pas pouvoir calculer le risque d'être malade si je suis exposé. Je ne dispose pas d'informations suffisantes : je n'ai pas l'incidence de la maladie dans le groupe exposés et dans le groupe non exposés.

D. Faux. $P(E+/M+) = PE1 = \frac{5}{50} = 0{,}1$ et $P(E-/M-) = PE0 = \frac{50}{250} = 0{,}2$

On peut avec ces données calculer l'OR qui est de

$$OR = \frac{P(E/M) \big/ 1-P(E/M)}{P(E/\overline{M}) \big/ 1-P(\overline{E}/\overline{M})} = \frac{0{,}1 \big/ 1-0{,}1}{1-0{,}2 \big/ 0{,}2} = 0{,}0278.$$

E. Faux. On sait que l'OR a un intervalle de confiance significatif (il ne contient pas la valeur 1), cependant, nous sommes dans le cadre d'une maladie dont la prévalence est supérieure à 1%. Ce n'est donc pas une maladie suffisamment rare, l'OR ne sera pas une bonne approximation du RR.

F. **Vrai**.

QCM n°11 : **C, D**

A. Faux. L'étude est prospective lors d'un essai thérapeutique comparatif.

B. Faux. Au contraire, elle diminue les biais de sélection.

C. **Vrai**.

D. **Vrai**. Avant la randomisation, on ne doit pas pouvoir prédire quel traitement aura le patient. Or, si on inclut un patient allergique au miel, on sait qu'il aura forcément le traitement de référence.

E. Faux. On ne sait pas en quelle faveur va la différence. On peut simplement conclure que les 2 traitements ont une efficacité différente.

QCM n°12 : **D, E**

A. Faux. Il est fortement conseillé dans ces deux cas-là. Il est parfois impossible d'obtenir le double aveugle pour raisons logistiques ou autre, mais on fait l'essai malgré tout.
B. Faux. Un essai est dit ouvert si aucun aveugle n'y est pratiqué.
C. Faux. La randomisation est l'attribution au hasard d'un groupe et du type de traitement.
D. **Vrai**.
E. **Vrai**.

QCM n°13 : **A, B, D, E**

A. **Vrai**. $\mathrm{Se} = \frac{\mathrm{VP}}{\mathrm{VP}+\mathrm{FN}} = \frac{85}{85+15} = 0{,}85$
B. **Vrai**. $\mathrm{Sp} = \frac{\mathrm{VN}}{\mathrm{VN}+\mathrm{FP}} = \frac{300}{300+0} = 1$. Lorsque la spécificité est égale à 1, on dit que le test est pathognomonique de la maladie, c'est-à-dire qu'il n'est positif qu'en présence de la maladie.
C. Faux. $VPP = \frac{p(M \cap T+)}{p(T+)} = \frac{85}{85} = 1$.
D. **Vrai**.
E. **Vrai**.

QCM n°14 : **B**

A. Faux. Bêta est bien égal à 5%, mais ce risque est choisi a priori.
B. **Vrai**. Plus la différence est grande, moins il faut de sujets.
C. Faux. Pour un seuil de confiance à 90%, cela implique que le alpha a augmenté, passant de 5 à 10, le nombre de sujets nécessaires sera donc plus petit. Bêta diminue donc le NSN augmenterait. On ne peut donc pas savoir car ces informations se contredisent, ce ne sera pas forcément plus grand.
D. Faux. Quand on passe d'un test unilatéral à bilatéral, le NSN augmente.
E. Faux. C'est la variabilité du critère de jugement principal qui importe.