TUTORAT UE 4 2013-2014 – Biostatistiques
CORRECTION Séance n°7 – Semaine du 18/11/2013

Séance de révisions générales

<u>QCM n°1 :</u> **A, D**

A. **Vrai**. On traduit l'énoncé comme ceci : $P(\bar{M}/E) = 0{,}2;$ $P(M/\bar{E}) = 0{,}2,$ où $M: Diarrhee\ et\ E = exposition\ a\ l'alcool$.
$RR = \frac{P(M/E)}{P(M/\bar{E})} = \frac{1-P(\bar{M}/E)}{P(M/\bar{E})} = \frac{1-0.2}{0{,}2} = 4$

B. Faux. Comme le risque relatif est significativement supérieur à 1, l'exposition à l'alcool est un facteur de risque.

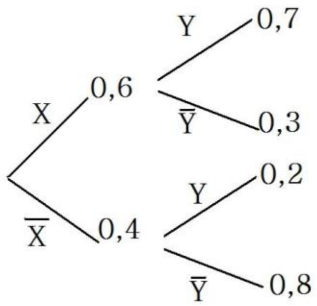
C. Faux. Le risque relatif à lui seul ne peut que déterminer l'existence d'une association, mais pas un lien de causalité. Un faisceau d'arguments est nécessaire pour établir ce lien causal.

D. **Vrai**. Ici, c'est la relation dose-effet qui est illustrée. Elle fait bien partie de la liste d'arguments nécessaires à l'établissement du lien causal.

E. Faux. La variabilité inter-sujets peut en effet induire un biais dans cette étude mais elle est une variabilité biologique et non pas analytique.

<u>QCM n°2 :</u> **A, C, D, E**

On fait donc un arbre de probabilité ou Y représente la maladie et X, le gène muté.



A. **Vrai**. $P(Y) = P(Y \cap X) + P(Y \cap \bar{X}) = P(Y/X) x P(X) + P(Y/\bar{X}) x\, P(\bar{X}) = 0{,}7x0{,}6 + 0{,}2x0{,}4 = 0{,}5$.

B. Faux. *Cf. item A.*

C. **Vrai**. On le déduit de l'arbre : $P(Y/\bar{X}) = 0{,}2$.

D. **Vrai**. $P(Y \cap X) = P(Y/X) x P(X) = 0{,}7x0{,}6 = 0{,}42$.

E. **Vrai**. $P(\bar{Y} \cap \bar{X}) = P(\bar{Y}/\bar{X}) x\, P(\bar{X}) = 0{,}8x0{,}4 = 0.32$.

<u>QCM n°3 :</u> **A, C**

A. **Vrai**. $P(\ 2\ M\&M) = \frac{C_3^2}{C_9^2} = 0{,}083$.

B. Faux. *Cf. item A.*

C. **Vrai**. $P\ (1\ KitKat\ et\ 1\ Smarties) = \frac{C_2^1\ x\ C_4^1}{C_9^2} = 0{,}222$.

D. Faux. $P(4\ Smarties\ et\ 1\ M\&M\ et\ 1 KitKat) = \frac{C_4^4\ x\ C_3^1 x\ C_2^1}{C_9^6} = 0{,}071$.

E. Faux. $P(\ 2\ amuse-gueules) = \frac{C_9^2}{C_9^2} = 1$.

QCM n°4 : **B, D**

A. Faux. Traduction de l'énoncé : $P(M) = 0{,}25;\ P(D/M\ ) = 0{,}2\ et\ P(M/D\ ) = 0{,}4$ où D représente la dépression et M, la méthadone. On peut donc déduire que $P(\bar{M}) = 0{,}75$.
On utilise le Théorème de Bayes : $P(M/D\ ) = \frac{P(D/M\ )x\ P(M)}{P(D/M\ )x\ P(M)+\ P(D/\bar{M}\ )xP(\bar{M})}$

$$0{,}4 = \frac{0{,}2x\ 0{,}25}{0{,}2x\ 0{,}25 + \ P(D/\bar{M}\ )x(0{,}75)}$$
$$P(D/\bar{M}\ ) = 0{,}1$$

B. **Vrai**. *Cf. item A.*
C. Faux. *Cf. item A.*
D. **Vrai**. $P(\bar{M}/D\ ) = 1 - P(M/D\ ) = 1 - 0{,}4 = 0{,}6$.
E. Faux. *Cf. item D.*

QCM n°5 : **A, B, E**

A. **Vrai**. Traduction de l'énoncé par l'intermédiaire d'un tableau ou S représente la scoliose et T, l'élévation de la température.

	M	$\bar{M}$	$Totaux$
T	7	18	25
$\bar{T}$	3	72	75
$Totaux$	10	90	100

On déduit du tableau que 72 personnes ne présentent ni scoliose ni élévation de la température.

B. **Vrai**. Spécificité= $P(\bar{T}/\bar{M}) = \frac{72}{90} = 0.8$.
C. Faux. Sensibilité= $P(T/M) = \frac{7}{10} = 0{,}7$.
D. Faux. On conclut cela via la VPP, qui est mesure la probabilité qu'un patient présentant le signe (élévation de température) soit atteint de la maladie (scoliose). VPP= $P(M/T) = \frac{7}{25} = 0{,}28$. On ne peut dire qu'à une certitude de 28% qu'un patient présentant une élévation de température est atteint de scoliose.
E. **Vrai**. On conclut cela via la VPN qui mesure la probabilité qu'un patient ne présentant pas le signe (élévation de température) ne soit pas atteint de la maladie (scoliose). VPP= $P(\bar{M}/\bar{T})$= $\frac{72}{75}$ =0,96. On peut donc dire avec une certitude de 96% qu'un patient ne présentant pas d'élévation de température n'est pas atteint de scoliose.

QCM n°6 : **B, E**

A. Faux. X suit une loi binomiale de paramètres n=70 et p= $\frac{E(X)}{n} = \frac{45}{70}$.
B. **Vrai**. Ecart-type= $\sqrt{n \times p \times (1-p)}$ =4,01.
C. Faux. L'approximation par la loi de Poisson est impossible car p>0,5.
D. Faux. Avant toute chose, il faut s'assurer que les conditions d'une approximation par la loi Normale sont réunies. Pour rappel, ces conditions sont les suivantes :
 – n>30 : ici, n=70, donc la condition est validée.
 – np>5 : ici, np=E(X)=45, donc la condition est validée.
 – n(1-p)>5 : ici, n(1-p)=nx(1- $\frac{45}{70}$)=25, donc la condition est validée.

De fait, par approximation, X suit une loi Normale de paramètres µ=nxp=70x $\frac{45}{70}$ et σ= $\sqrt{n \times p \times (1-p)} = \sqrt{70 \times \frac{45}{70} \times (1 - \frac{45}{70})}$. Donc P(X>35) $\rightarrow$ P(X>35,5)=1-P(X≤35,5)=1-P(U<-2,37)=1-π(-2,37)=1-0,0089=0,9911. *Remarque* : ne pas oublier la correction de continuité.
E. **Vrai**. *Cf. item D.*

QCM n°7 : **A, B, E**

A. **Vrai**. P(X≥1)=1-P(X=0)=1-15,3x10^{-12}≈1.

B. **Vrai**. D'après l'énoncé : P(X=0)=15,3x10^{-12} $\Leftrightarrow$ e$^{-\lambda}$=15,3x10^{-12} $\Leftrightarrow$ λ=-ln(15,3x10^{-12})=24,9.

C. Faux. On ne peut pas approximer la loi de Poisson par la loi Binomiale, seul l'inverse est possible.

D. Faux. λ>20, l'approximation par la loi Normale est donc possible. Ainsi par approximation X suit une loi Normale de paramètres μ=λ et σ= $\sqrt{\lambda}$. Donc P(10≤X≤20) $\rightarrow$ P(9,5≤X≤20,5)=P(-3,09≤U≤-0,88)=π(-0,88)-π(-3,09)=0,1894-0,00100=0,1884. *Remarque* : ne pas oublier la correction de continuité.

E. **Vrai**. *Cf. item D*.

QCM n°8 : **D, E**

A. Faux. D'après l'énoncé : P(X<20)=0,3264 et P(X>35)=0,8997, donc par lecture inverse de la table : $\frac{20-\mu}{\sigma}=-0{,}45$ et $\frac{35-\mu}{\sigma}=1{,}28$. On obtient donc un système de deux équations à deux inconnues, ainsi par résolution du système on trouve : μ=23,902 et σ=8,6705.

B. Faux. *Cf. item A*.

C. Faux. *Cf. item A*.

D. **Vrai**. *Cf. item A.*

E. **Vrai**. P(10<X<23)=P(-1,60≤U≤-0,10)=π(-0,10)-π(-1,60)=0,4602-0,0548=0,4054.

QCM n°9 : **A, D**

A. **Vrai.** X suit une loi binomiale, le succès correspondant au fait de suivre correctement sa prescription (n=100 et p=$\frac{17}{100}$).

B. Faux. On calcule les paramètres de la population, pas de l'échantillon.

C. Faux. *Cf. item D.*

D. **Vrai**. Avant de réaliser le calculer de l'intervalle de confiance, il faut s'assurer que les conditions d'approximation d'une loi Binomiale par une loi Normale sont réunies. Ici, n=100 (condition : n>30), np=17 (condition : np>5) et n(1-p)=83 (condition : n(1-p)>5). Les trois conditions étant réunies, ont peut alors calculer l'intervalle de confiance de la proportion de personnes suivant correctement leur prescription, tel que : IC=$\left[p \pm c_{\alpha/2} \times \sqrt{\frac{p\times(1-p)}{n}}\right]=[0{,}0964;0{,}2436]$. *Remarque* : $c_{\alpha/2}$=1,96.

E. Faux. *Cf. item D*.

QCM n°10 : **B, D**

A. Faux. La taille de l'échantillon étant inférieure à 30, l'hypothèse de normalité est nécessaire pour le calcul de l'IC de l'espérance.

B. **Vrai**. s²=$\frac{1}{n}\sum x_i{}^2-\bar{x}^2$=2,9775.

C. Faux. S²=$\frac{n}{n-1}s^2$=3,1902.

D. **Vrai**. IC=$\left[\bar{x} \pm t_{n-1;\alpha/2} \times \frac{S}{\sqrt{n}}\right]=[0{,}8379;2{,}4621]$, avec $t_{n-1;\alpha/2}$=1,761.

E. Faux. *Cf. item D*.

QCM n°11 : **F**

A. Faux. On considère en effet le risque α comme plus lourd de conséquence mais il correspond au rejet à tort de H0 ! (ou non rejet à tort de H1).
B. Faux. Pour un test paramétrique réalisé en unilatéral, le risque α est multiplié par 2. Ainsi, t_{obs} est inchangé, mais la valeur seuil t_α diminue. Donc si on rejette déjà en bilatéral, on rejettera aussi en unilatéral.
C. Faux. Ici, il est question de 2 moyennes observées, sur 2 séries appariées (car on demande aux mêmes étudiants en PACES quand ils étaient primants puis à présent doublants), de la même variable (distance domicile-faculté) . Il s'agit donc d'un test paramétrique (car on pose l'hypothèse de normalité de la variable), quantitatif (car on note les 15 valeurs avant et après), pour séries appariées et n<30 donc : Test de Student pour séries appariées avec lecture dans la table de Student à n-1 ddl.
Remarque :-Sans hypothèse de normalité on aurait pu utiliser Wilcoxon.
-Sans valeurs précises mais avec seulement des tendances (ici rapprochement/éloignement de la fac), on aurait pu utiliser un X^2 de MacNemar
-Sans hypothèse de normalité et sans valeurs précises, on aurait pu utiliser un test des signes.
D. Faux. Le test F s'intéresse à l'égalité des variances, tel que H0 : les variances des deux populations sont identiques. Si l'hypothèse H0 est conservée, cela signifie que les variances sont égales, on peut alors appliquer un test de Student ou de l'écart-réduit. En revanche, si l'hypothèse H0 est rejetée, cela signifie que les variances des deux populations ne sont pas égales ; on rentre alors dans le cadre de la réalisation d'un test d'Aspin-Welch.
E. Faux. Nous sommes dans le cadre de la comparaison de deux chewing-gums différents, donc nous travaillons sur échantillons indépendants. En effet, bien que ce soit une seule et même personne qui souffle dans chacun des deux chewing-gums, la réalisation de ce test porte sur l'aptitude du chewing-gum à former une bulle, et non pas sur la capacité de la personne à en former une. Nous sommes donc dans le cadre d'un test de Mann-Whitney.
F. **Vrai**.

QCM n°12: **A, D**

A. **Vrai**.
B. Faux. Attention, ici, on ne fait pas d'hypothèse sur la loi de la variable donc, par principe, on ne peut pas utiliser de tests paramétriques, de plus : f+g=9 <10.
C. Faux. Nous ne disposons que de tendances entre l'augmentation ou la diminution des téléchargements, et non du nombre précis de téléchargements pour les 23 élèves donc il faut utiliser le test de signes.

		Avant Hadopi		
		Téléch modéré	Téléch excessif	
Après Hadopi	Téléch modéré	1	3	4
	Téléch excessif	6	13	19
		7	16	23

D. **Vrai**. On a T+=6 et T-=3 donc la statistique de test T_{min} vaut 3. Or, on trouve dans la table avec m=9, une valeur p=0,254. En bilatéral, cela donne 0,508 donc p>α=0,05 ; on ne rejette pas H0.
E. Faux. Tout d'abord une hypothèse n'est jamais vraie ! (De plus en unilatéral, α diminue (=0,254) pour un test non paramétrique donc on ne risque pas de corroborer H1).

QCM n°13: **D, E**

A. Faux. On fait toujours les hypothèses dans la population.
B. Faux. On utilisera le test de MacNemar car on pose une hypothèse sur la loi de la variable ; de plus, f+g=57>10 et un test paramétrique est plus puissant que le test des signes.
C. Faux. Pour un test du X^2 de MacNemar, on lit dans la table du X^2 à 1ddl (lignes-1 x colonnes-1).

		Avant régime		
		<10kg	≥10kg	
Après régime	<10kg	35	14	49
	≥10kg	43	8	51
		78	22	100

D. **Vrai**. $t_{obs}=\frac{(f-g)^2}{f+g}=\frac{(43-14)^2}{43+14}\approx 14.75$ t_α=6,635 au risque 1% bilatéral donc $t_{obs}>t_\alpha$; on peut rejeter H0.
E. **Vrai**. t_α diminue car α augmente (on lit à 2% car on demande 1% en unilatéral, ce qui donne 5,412) donc on rejette aussi H0.

QCM n°14 : **B, D, E**

	f théoriques %	Oij	Eij
Forte consommation	40,4	1035	1010
Consommation modérée	26,16	682	654
Faible consommation	19,88	525	497
Aucune utilisation	13,56	258	339
TOTAL	100	2500	2500

A. Faux. On doit lire à r(=nombre de lignes)-1 ddl soit 3ddl.
B. **Vrai**. Tous les Eij sont >5. Attention, le carré dans la formule montre que le signe de la différence ne change pas !
C. Faux. On a $X^2_{obs}=\frac{(1035-1010)^2}{1010}+\frac{(682-654)^2}{654}+\frac{(525-497)^2}{497}+\frac{(258-339)^2}{339}=22.749$
D. **Vrai**. On trouve avec α=5% à 3 ddl un X^2_α=7,815. Donc on peut rejeter H0.
E. **Vrai**. On trouve avec α=0,1% un X^2_α=16,266. Donc on peut aussi rejeter H0.

QCM n°15 : **D**

A. Faux. Attention, cette même variable est distribuée une fois mais dans deux groupes différents, les séries sont donc indépendantes.
B. Faux. Pour obtenir une AMM, (autorisation de mise sur le marché) il vaut mieux que le nouveau médicament soit plus efficace que ceux actuellement utilisés, donc que la glycémie moyenne du groupe A soit inférieure à celle du groupe B, autrement dit un test unilatéral avec H1 : µA<µB. Toutefois, il faut noter que la réalisation des tests en unilatéral est rare, et que l'on peut autoriser deux médicaments à même efficacité pour d'autres raisons : concurrence, prix…
C. Faux. On sait que $t_{obs}=\frac{|m1-m2|}{\sqrt{\frac{S^2}{nA}+\frac{S^2}{nB}}}$ car n_A et n_B < 30. Il nous faut trouver S^2 et pour cela il nous faut les écarts types estimés S_A^2 et S_B^2. Avec la formule $S^2=(n/n-1)xs^2$, on trouve S_A^2=10,14 et S_B^2=27,63.
Puis, on utilise le formulaire avec $S^2=\frac{(nA-1)xSa^2+(nB-1)xSb^2}{nA+nB-2}$ = 18,88. On remplace les valeurs dans t_{obs} et on trouve t_{obs}= 0,187.
On cherche t_α dans la table de Student à nA+nB-2 ddl=48. Or, la loi de Student tend vers une loi Normale pour n>30 donc, en lisant à ddl=∞, on trouve 1,645 en unilatéral. Donc on ne rejette pas H0.
D. **Vrai.** On lit à 2% car on est en unilatéral on trouve un t_α=2,326, $t_{obs}<t_\alpha$; on ne rejette pas H0.
E. Faux. Attention, on ne fait jamais d'association clinique en test statistique, on établit toujours des liens statistiques.

QCM n°16 : **A, E**

A. **Vrai**. La prévalence est un indicateur qu'on mesure à un moment donné, obtenue par des enquêtes transversales, donc figées dans le temps.
B. Faux. Il est vrai qu'on distingue trois niveaux de morbidité, mais la morbidité ressentie est la plus difficile à évaluer ; en effet, elle est subjective et sera donc propre à chaque individu.
C. Faux. Une augmentation de la prévalence n'est pas toujours signe d'une détérioration de la santé. Par exemple, la prévalence d'un cancer peut augmenter car les malades vivent plus longtemps avec leur maladie, donc ils meurent moins, grâce, au contraire, à de meilleurs soins médicaux.
D. Faux. La létalité est la part des décès dus à la maladie parmi les patients atteints de cette maladie.
E. **Vrai**. En effet, cela signifie qu'il y a plus de nouveaux malades.

QCM n°17 : **B, C, E**

A. Faux. On distingue bien les enquêtes expérimentales et les enquêtes d'observation mais ce sont les enquêtes expérimentales qui permettent l'imputation causale (les expérimentations vont permettre d'établir une relation de cause à effet).
B. **Vrai**.
C. **Vrai**. Exemple : enquête exposés-non exposés.
D. Faux. Elles présentent d'importants biais de sélection (choix de témoins représentatifs) et d'information (recueil des informations auprès des sujets).
E. **Vrai**.

QCM n°18 : **C**

A. Faux. Il s'agit bien d'une enquête observationnelle, type cas-témoins (on interroge les sujets sur leur exposition), rétrospective (consommation durant les 30 dernières années) mais elle n'est pas longitudinale car il n'y a pas de suivi dans le temps.
B. Faux. C'est le cas ici d'une maladie rare (2/10000), l'odds ratio est un bon estimateur du risque relatif.

	Malades	Non malades	
Exposés (alcool)	50	50	100
Non exposés (pas alcool)	100	350	450
	150	400	550

C. **Vrai**. D'après le tableau ci-dessous, l'odds ratio=(50*350)/(100*50)=3,5.
D. Faux. *Cf. item C.*
E. Faux. On peut dire que la consommation d'alcool est un facteur de risque car la valeur 1 n'appartient pas à l'intervalle, on peut donc établir un lien significatif mais pas un lien causal.

QCM n°19 : **C, D, E**

A. Faux. Attention, le biais de classement est aussi appelé biais d'information, il résulte d'une erreur de mesure sur l'exposition ou sur la maladie.
B. Faux. Le facteur de confusion ne doit pas être une conséquence de l'exposition ; par exemple, si l'on étudie la relation alcool-infarctus du myocarde, l'alcool étant fortement lié au tabac, ce dernier représente un facteur de confusion.
C. **Vrai**. Il y a un risque de biais de sélection important, notamment au moment de la sélection des groupes mais surtout en raison des perdus de vue car l'enquête est longue, ce qui peut conduire à l'abandon de certains sujets.
D. **Vrai**. C'est le principe du tirage au sort.
E. **Vrai**. Les patients peuvent ne pas respecter le protocole initial (par exemple, mal prendre leur traitement, ou encore l'arrêter…). Mais si on exclut ces patients de l'analyse, on risque de diminuer la comparabilité des groupes, entraînant ainsi un biais d'attrition, c'est pourquoi on choisit a priori d'analyser tous les patients, quelque soient leurs écarts au protocole (c'est l'analyse en intention de traiter).

QCM n°20 : **A, C, E**

A. **Vrai**.
B. Faux. En effet, le NSN dépend de la différence que l'on souhaite montrer, c'est donc pour ça qu'il est choisi a priori !!
C. **Vrai**.
D. Faux. Au contraire, le risque de se tromper est plus grand si le nombre de sujets est petit, pour montrer une différence entre deux groupes par exemple et ce avec un risque très faible, il nous faudra un NSN important.
E. **Vrai**. La protection des personnes (malades ou non) participant à l'essai est prioritaire, ainsi que leur consentement libre et éclairé.