

Faculté de Médecine Denis Diderot - Université Paris 7

Année Universitaire 2007-2008

Lundi 28 janvier 2008

**Epreuve de Biostatistique**

Durée : 1 Heure

**Attention**

A lire attentivement avant de composer

Ce sujet comporte 14 pages

Penser à utiliser les tables numériques et statistiques en fin de sujet (de la page 5 à la page 14)

Le sujet comprend **12 questions (ou QCM) indépendantes**, de la page 2 à la page 4

Les questions à choix simple n'ont qu'**une seule** proposition vraie

Les questions à choix multiples ont **une ou plusieurs** proposition(s) vraie(s), **toujours entre 1 et 4**

La notation des questions à choix simple (**une seule** proposition vraie) sera **binaire**, celle des questions à choix multiples (**une ou plusieurs** proposition(s) vraie(s)) sera **pondérée**

**Les questions sans réponse seront considérées comme nulles**

Conventions

- Les décimales sont notées avec une virgule
- Les majuscules ont un sens et sont à respecter
- **On considérera toujours que  $1,96 \approx 2$**

**Exercice 1** Huit étudiants souhaitent organiser une fête de fin d'année. Ils décident d'envoyer des invitations, tout en réalisant l'organisation de cette fête en parallèle.

**QCM 1** Leur première décision est de choisir la date de la fête, un samedi soir. En supposant qu'ils tirent au hasard uniforme la date dans les 40 semaines de 2008 hors vacances universitaires (12 en hiver, 12 au printemps, 12 à l'automne et 4 en été), et qu'ils retiennent le numéro de la semaine sélectionnée, dire LA proposition vraie.

- a - Il s'agit d'une expérience aléatoire qualitative
- b - Il s'agit d'une expérience aléatoire discrète
- c - Il s'agit d'une expérience aléatoire continue
- d - La probabilité de choisir l'une des saisons possibles est constante
- e - La probabilité de choisir l'un des samedis possibles est nulle.

**QCM 2** Les huit étudiants aimeraient chacun que la fête ait lieu la semaine de leur anniversaire. On supposera que les huit étudiants ont chacun leur anniversaire une semaine différente dans les 40 semaines retenues, indépendamment de la saison. Soit S, le samedi tiré au sort. Dire la(les) proposition(s) vraie(s).

- a - La probabilité que S soit au printemps est 0,025
- b - La probabilité que S soit au printemps est 0,30
- c - La probabilité que S soit dans la semaine d'un anniversaire est 0,20
- d - La probabilité que S, tiré dans la semaine d'un anniversaire, soit au printemps est 0,06
- e - La probabilité que S, tiré dans la semaine d'un anniversaire, soit au printemps est 0,30

**QCM 3** Le samedi 29 mars est finalement retenu. Les étudiants décident maintenant d'établir le nombre d'invitations à envoyer. En supposant que le nombre d'invitations d'un étudiant est une variable gaussienne de moyenne 200 et d'écart-type 50, dire la(les) proposition(s) vraie(s).

- a - La probabilité pour qu'un étudiant invite moins de 150 étudiants est de 0,32 (à 0,01 près)
- b - La probabilité pour qu'un étudiant invite moins de 150 étudiants est de 0,16 (à 0,01 près)
- c - On peut parier à 95% que le nombre d'invités d'un étudiant sera compris entre 170 et 230
- d - On peut parier à 95% que le nombre d'invités d'un étudiant sera compris entre 100 et 300
- e - On peut parier à 95% que le nombre d'invités d'un étudiant sera supérieur à 120 (à 5 près)

**QCM 4** Si chaque étudiant adresse en moyenne 200 invitations, avec un écart-type de 50, et en supposant toujours les nombres d'invitations gaussiens, dire la(les) proposition(s) vraie(s).

- a - Le nombre total d'invitations est une variable gaussienne d'écart-type 140 (à 5 près)
- b - Le nombre total d'invitations est une variable gaussienne d'écart-type 400 (à 5 près)
- c - On a supposé les nombres d'invitations des 8 étudiants indépendantes
- d - On a appliqué le théorème limite central
- e - On ne peut faire ce calcul car il y a 8 étudiants et le théorème limite central ne s'applique pas

**QCM 5** Les huit étudiants décident d'envoyer 1200 invitations. Chacun met en moyenne 10 secondes pour mettre une invitation sous enveloppe et la cacheter. Sous l'hypothèse que la durée de mise sous enveloppe d'une invitation est exponentielle, dire LA proposition vraie.

- a - Il y a au plus une chance sur 3 qu'un étudiant ait fini une invitation en moins de 10 secondes
- b - Il y a au plus une chance sur 6 qu'un étudiant ait fini une invitation en moins de 10 secondes
- c - Il y a au plus une chance sur 10 qu'un étudiant ait fini une invitation en moins d'une minute
- d - Il y a au plus une chance sur 100 qu'un étudiant ait fini une invitation en moins d'une minute
- e - Il y a au plus une chance sur 100 qu'un étudiant n'ait pas fini une invitation en moins d'une minute

**QCM 6** Dans la mise sous enveloppes des 1200 invitations, quelques erreurs ont été commises. En supposant que les erreurs sont indépendantes et qu'elles ont été effectuées pour une invitation sur 1000, dire la(les) proposition(s) vraie(s).

- a - La probabilité qu'il n'y ait aucune erreur est de 0,30 (à 0,002 près)
- b - La probabilité qu'il n'y ait aucune erreur est de 0,37 (à 0,002 près)
- c - On a assimilé la loi du nombre d'erreurs à une loi de Poisson
- d - On a assimilé la loi du nombre d'erreurs à une loi exponentielle
- e - L'approximation de la loi du nombre d'erreurs par une loi de Gauss n'est pas valide ici

**Exercice 2** On s'intéresse à la survenue de bactéries résistantes dans les selles chez les patients recevant un traitement antibiotique de 10 jours. Une large étude réalisée en 2007 a retrouvé une résistance chez 80% des patients avec l'antibiotique A. On réalise une étude en janvier 2008 chez 100 patients qui reçoivent l'antibiotique B et on trouve 60 patients présentant une résistance.

**QCM 7** On cherche à tester si la proportion de résistance observée dans cette étude avec l'antibiotique B est différente de celle trouvée en 2007 avec l'antibiotique A. On arrondira les pourcentages à 1% près. Parmi les propositions suivantes, dire la(les) proposition(s) vraie(s).

- a - L'intervalle de pari à 95% de la proportion observée de résistance avec l'antibiotique B sur un échantillon de 100 patients vaut [50%; 70%]
- b - L'intervalle de pari à 95% de la proportion observée de résistance avec l'antibiotique A sur un échantillon de 100 patients vaut [72%; 88%]
- c - Pour faire le test on peut utiliser un test du  $\chi^2$  à 2 degrés de liberté
- d - Pour faire le test on peut utiliser l'intervalle de pari à 95% de la proportion observée de résistance avec l'antibiotique A sur un échantillon de 100 patients
- e - L'hypothèse nulle de ce test s'écrit  $p = 80\%$  où  $p$  est la proportion observée de résistance avec l'antibiotique B.

**QCM 8** On teste si la proportion de résistance observée dans cette étude avec l'antibiotique B est différente de celle trouvée en 2007 avec l'antibiotique A et on utilise le test de l'écart-réduit. Parmi les propositions suivantes, dire LA proposition vraie.

- a - On ne rejette pas l'hypothèse nulle au risque 5%
- b - Le degré de signification du test est inférieur à  $10^{-6}$
- c - Le degré de signification du test est inférieur à  $10^{-7}$
- d - On conclut à la responsabilité de l'antibiotique B dans la diminution de la résistance (au risque 5%).
- e - On conclut que l'antibiotique B donne autant de résistance que l'antibiotique A (au risque 5%).

**Exercice 3** On décide de réaliser une étude clinique pour comparer les proportions de survenue de résistance chez deux groupes de patients de taille identique, un groupe recevant l'antibiotique A, l'autre recevant l'antibiotique B. On souhaite un risque de seconde espèce de 10% et de première espèce de 5%. On suppose que la proportion de référence est de 80% et la différence attendue est de 20%.

**QCM 9** Parmi les propositions suivantes, dire LA proposition vraie.

- a - Il faut inclure au moins 40 patients par groupe
- b - Il faut inclure au moins 80 patients par groupe
- c - Il faut inclure au moins 40 patients en tout dans l'étude
- d - Il faut inclure au moins 80 patients en tout dans l'étude
- e - Les conditions de validité pour effectuer ce calcul ne sont pas vérifiées

**Exercice 4** On réalise une étude chez 100 patients présentant une pneumonie, la moitié recevant l'antibiotique A, l'autre l'antibiotique B. L'attribution de l'antibiotique est tirée au sort. On mesure le délai jusqu'à la disparition des symptômes. On observe dans le groupe recevant A une moyenne  $m_A$  de 10 jours et un écart type  $s_A$  de 10 jours. On observe dans le groupe recevant B une moyenne  $m_B$  de 5 jours et un écart type  $s_B$  de 10 jours. On compare les moyennes des délais entre les deux groupes.

**QCM 10** Parmi les propositions suivantes, dire la(les) proposition(s) vraie(s).

- a - Les moyennes sont significativement différentes entre les deux groupes (au risque 5%)
- b - Les moyennes sont significativement identiques entre les deux groupes (au risque 5%)
- c - Pour conclure que la différence des moyennes des délais de guérison est liée à l'antibiotique, il faut absolument que l'étude soit en double aveugle
- d - On n'a pas réalisé un essai thérapeutique randomisé
- e - On peut conclure que la différence des moyennes des délais de guérison est liée à l'antibiotique car on a réalisé un essai thérapeutique randomisé

**QCM 11** On calcule la puissance du test précédent, sous l'hypothèse alternative que la différence entre les moyennes est de 5 jours et en supposant les variances identiques à celles observées. Parmi les propositions suivantes, dire LA proposition vraie.

- a - La puissance est entre 25% et 35%
- b - La puissance est entre 35% et 50%
- c - La puissance est entre 50% et 65%
- d - La puissance est entre 65% et 75%
- e - La puissance est entre 75% et 85%

**QCM 12** On s'intéresse à la variable aléatoire  $D = M_A - M_B$  où  $M_A$  et  $M_B$  sont les variables aléatoires moyennes expérimentales du délai de guérison chez  $n_A$  patients recevant l'antibiotique A et  $n_B$  patient recevant l'antibiotique B. On définit :

$$Y = \frac{D}{\sqrt{\text{Var}(D)}}$$

Parmi les propositions suivantes, dire la(les) proposition(s) vraie(s).

- a - La variance de Y vaut 1
- b - Y suit toujours une loi normale
- c -  $\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}$  est un estimateur sans biais de  $\text{Var}(D)$
- d -  $\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}$  est une réalisation d'un estimateur sans biais de  $\text{Var}(D)$
- e -  $\frac{s_A^2}{n_A} - \frac{s_B^2}{n_B}$  est une réalisation d'un estimateur sans biais de  $\text{Var}(D)$

## CORRECTION DE BIOSTATISTIQUES 2007-2008

### QCM1 : réponse B

On peut tirer un chiffre entre 1 et 40 mais avec seulement des valeurs entières... donc variable discrète ! et c'est bien une expérience aléatoire vu que le résultat est imprévisible, reproductible ^^

### QCM2 : réponses B, C et E

Là c'est le vieux piège de la proba conditionnelle made in chevet 😊

$P(S) = 12/40 = 3/10 = 0,3$  aucune difficulté là

$P(S|A) = 8/40 = 0,2$  Ici le calcul était simple car chevet avait précisé que ça ne pouvait pas tomber sur deux semaines identiques (les anniversaires ^^)

Et enfin vla la difficulté ^^

$P(\text{printemps}|\text{anniversaire})$  : là fallait faire très gaffe a pas multiplier bêtement !

$P(P|A) = p(P \text{ inter } A)/p(A) = [p(A|P)*P(S)]/p(A)$

Et ça ça fait  $0,06/0,2 = 0,3$

### QCM 3 : B, D et E !

bon, je vais pas vous faire l'injure d'une loi de gauss bête comme le monde ^^

$P(Z < (150-200)/50) = P(Z > 1) = 0,159$

Ensuite on applique les mêmes méthodes pour les bornes 😊

$P(X < 170) = 0,274$  donc on peut rejeter C !

$P(X < 100) = 2,5\%$

$P(X > 300) = 2,5\%$

Chouette : une bonne réponse 🎯

$P(X < 120) = 0,055$  ce qui à 5 près marche 🎯

### QCM 4 : A et C

On fait une somme de variables (attention par d'écart-type)

$2500*8 = 20000$

On passe ensuite à l'écart type de la somme par simple racine ce qui fait 100 racine de 2 = 140 environ !

Et on a évidemment supposé les 8 étudiants indépendants, principe de gauss et de l'étudiant tout court 🎯

### QCM 5 : réponse E

En fait Chevet c'est amusé, à juste titre à changer un peu la formulation pour faire peur ce qui est réussi !

En fait fallait pas se biler !

$P(X < 10) = 1 - 0,37$  (car  $1 - (e \text{ de } -1) = 0,63$ )

Bref 63% de chances de finir avant les 10 secondes... on rejette A et B avec pertes et fracas 🎯

On applique la même formule pour C et D avec cette fois  $x = 60$  ce qui donne  $P(X < 60) = 0,9975...$

De là facile de dire que  $P(X > 60) = 0,0025$  ce que demandait la E... qui est bonne du coup !

### QCM 6 : A C et E !

On applique une loi de poisson car n est géant et lambda ridicule, donc loi de poisson de 1,2  
De là on trouve grâce à la table que A est vraie et B fausse !  
Et la C... ben juste monsieur ^^ et donc la D au trou !  
Et quand à la E, elle est juste car on obtient  $np = 1,2 < 5$  donc pas valide

### QCM 7 : B et D

si je reprends un peu l'énoncé, on a retrouvé une résistance de 80% sur un large intervalle  $\implies \pi = 0.8$   
on a fait une expérience sur 100 personnes  $\implies p = 0.6$   
pour la A. on nous propose un intervalle de pari centré sur 0.6 c'est faux. On ne calcule pas d'intervalle de pari à partir d'une proportion observée.  
Pour la B. j'ai fait le calcul avec la formule du formulaire et on trouve bien ça. Ne pas oublier les conditions de validité. Ouf ça marche.  
Pour la C., on ne peut pas faire un test du chi 2 à 2 ddl, limite un ddl il me semble.  
Pour la D., j'ai mis oui. Si p n'appartient pas à l'intervalle de pari, on trouvera forcément une valeur absolue de z supérieur à 2 puisque l'intervalle de pari est à 95%.  
Pour la E. jamais d'hypothèse nulle avec les p, c'est observé ça...

### QCM 8 : B

on rejette on l'a vu plus haut.  
donc c'est pas la A. ni la E.  
Pour la D. ça marche pas, on nous dit pas que c'est un ETR.  
Donc c'est un des degré de signification. On calcule le z, je trouve 5.  
Je regarde dans ma table et je trouve inférieur à  $10^{-6}$

### QCM 9 : B

bah bon là calcul simple comme elle l'a dit  
on nous donne  $\beta = 10\%$  donc  $x_{\beta} = 1.28$   
 $\alpha = 5\%$  donc  $\epsilon_{\alpha} = 1.96$ , on regarde dans la table avec les carrés déjà tout fait et on voit que ça fait 10.  
 $\delta = 0.2$  on n'oublie pas le carré.  
 $\pi_0 = 0.8$   
Donc je trouve 80 et c'est par groupe  
Donc réponse B.

### QCM 10 : A et E

j'ai fait mon calcul de z et je trouve 2.5 donc réponse A  
et c'est un ETR donc réponse E.  
la C est fausse parce que le double aveugle est une condition souhaitable du double aveugle mais pas obligatoire.  
La D est fausse parce qu'on a tiré au sort.

QCM 11 : D

l'hypothèse alternative suit une loi normale de moyenne  $\delta/\sigma$  ce qui fait 2.5 et de variance 1

$P(Z' > 2)$  revient au calcul de la puissance, l'autre est négligeable  
changement de variable

$P(Z > -0.5)$  je trouve 69%

Donc réponse D

QCM 12 : A, B et D

on divise par l'écart type donc on réduit donc la variance est de 1.

$M_a$  et  $M_b$  suivent des lois normales donc leur différence aussi donc  $Y$  suit une loi normale (réduite)

La C est fausse, ce n'est pas l'estimateur mais une estimation.

Ensuite la d plutôt que la e parce que dans la formule c'est + et pas -.

*Ce document, ainsi que l'intégralité des cours de P1, sont disponibles gratuitement  
à l'adresse suivante : <http://coursplbichat-larib.weebly.com>*