



Université Pierre et Marie Curie

Annales de Biostatistique

PCEM1 - PAES (UE4)

1998 - 2008

**J.L. Golmard
A. Mallet
V. Morice**

Mise à jour : 26 mars 2008

Liste des exercices

3	Liste des exercices
5	1 : Concours Nouméa 2008
13	2 : Concours 2007
15	3 : Concours Nouméa 2007
21	4 : Concours 2006
29	5 : Concours Nouméa 2006
33	6 : Concours 2005
43	7 : Concours Nouméa 2005
47	8 : Concours 2004
53	9 : Concours Nouméa 2004
59	10 : Concours 2003
65	11 : Concours Nouméa 2003
71	12 : Concours 2002
77	13 : Concours 2001
83	14 : Concours 2000
87	15 : Concours 1999
93	16 : Concours 1998

1

Concours Nouméa 2008

Durée 1 heure 30. Tout document autorisé.

L'épreuve comporte 20 QCM

Rappels :

- La somme de variables aléatoires distribuées normalement est une variable aléatoire distribuée normalement
- Si une variable est distribuée selon une loi de χ^2 à n degrés de liberté, son espérance est n et sa variance $2n$.

Exercice 1 (2 QCM)

1.1 On s'intéresse à la conformité des poches de sang utilisées dans un hôpital. On appelle p la probabilité pour une poche d'être non-conforme. Cette probabilité est supposée faible. On contrôle la conformité d'un nombre n de poches, assez grand (au moins 100).

- A. La probabilité de n'avoir aucune poche non-conforme parmi les n est $1-p$
- B. La probabilité de n'avoir aucune poche non-conforme parmi les n est p^n
- C. La probabilité de n'avoir aucune poche non-conforme parmi les n est $(1 - p)^n$
- D. En faisant l'approximation par la loi de Poisson, la probabilité de n'avoir aucune poche non-conforme parmi les n est e^{-p}
- E. En faisant l'approximation par la loi de Poisson, la probabilité de n'avoir aucune poche non-conforme parmi les n est e^{-np}

1.2 On calculera les probabilités en utilisant la loi de Poisson et avec un arrondi à 2 décimales. On dit que l'observation de zéro poche non-conforme parmi les n est incompatible avec la probabilité p si la probabilité d'observer zéro poche non-conforme est inférieure à 5%.

- A. Les valeurs $n=300$ et $p=0,08$ sont incompatibles avec l'observation de 0 poche non-conforme
- B. Avec $n=300$, il faut et il suffit que $p<0,05$ pour qu'il y ait compatibilité avec l'observation de 0 poche non-conforme
- C. Avec $n=300$, il faut et il suffit que $p<0,01$ pour qu'il y ait compatibilité avec l'observation de 0 poche non-conforme
- D. Il faut et il suffit que $p>3/n$ pour qu'il y ait compatibilité avec l'observation de 0 poche non-

conforme

- E. Il faut et il suffit que $p < 3/n$ pour qu'il y ait compatibilité avec l'observation de 0 poche non-conforme

Exercice 2 (1 QCM)

2.1 La densité de probabilité d'une variable aléatoire X est donnée par $f(x) = h x^2$, si $0 < x < 1$, et $f(x) = 0$ sinon. h est une constante à déterminer.

- A. $h = 2$
- B. $h = 3$
- C. $E(X) = 0,5$
- D. $E(X) = 0,75$
- E. $\Pr(X < 0,5) < 0,5$

Exercice 3 (1 QCM)

3.1 Pour détecter une certaine maladie, on dispose de deux examens notés A et B et dont les résultats peuvent être positifs ou négatifs et sont indépendants, que la maladie soit ou non présente. On connaît les sensibilités $Se(A)=0,90$ et $Se(B)=0,60$, ainsi que les spécificités $Sp(A)=0,80$ et $Sp(B)=0,80$. La démarche choisie conclut qu'un sujet est malade dès qu'un des deux résultats est positif et qu'il est non malade sinon.

- A. Si le sujet est malade, la probabilité d'être négatif pour A et positif pour B est de 0,10
- B. Si le sujet est malade, la probabilité d'être négatif pour A et positif pour B est de 0,20
- C. La sensibilité obtenue par la démarche choisie est 0,96
- D. La sensibilité obtenue par la démarche choisie est 0,54
- E. La spécificité obtenue par la démarche choisie est 0,64

Exercice 4 (1 QCM)

4.1 Dans une certaine population la prévalence d'une maladie est 0,7. On dispose d'un test diagnostique dont la sensibilité est 0,80 et la spécificité est 0,80

- A. La valeur prédictive positive est supérieure à 80%
- B. La valeur prédictive positive est inférieure à 20%
- C. La valeur prédictive négative est supérieure à 80%
- D. La valeur prédictive négative est inférieure à 20%
- E. La probabilité que le résultat du test diagnostique soit négatif est 0,38

Exercice 5 (1 QCM)

5.1 La courbe ROC d'un examen est définie par la fonction $y = \sqrt{x}$, avec (x compris entre 0 et 1)

- A. Lorsque la spécificité vaut 0,25, la sensibilité vaut 0,5
- B. Lorsque la spécificité vaut 0,75, la sensibilité vaut 0,5
- C. Lorsque la spécificité vaut 0, la sensibilité vaut 1
- D. Lorsque la spécificité vaut 1, la sensibilité vaut 1
- E. Lorsque la spécificité vaut 1/2, la sensibilité vaut $1/\sqrt{2}$

Exercice 6 (1 QCM)

6.1 Une certaine intervention chirurgicale se décompose habituellement en 3 phases : la phase de préparation, l'intervention proprement dite, et le réveil. Les durées de ces phases sont distribuées selon des lois normales indépendantes, avec :

- Pour la préparation : moyenne 3h, écart-type 1h
 - Pour l'intervention : moyenne 4h, écart-type 2h
 - Pour le réveil : moyenne 5h, écart-type 3h
- A. La durée totale moyenne de ce type d'intervention est de 12h
 - B. L'écart-type de la durée totale de l'intervention est 3h45' (à une minute près)
 - C. Il y a 95 chances sur 100 pour que la durée totale de l'intervention soit comprise entre 5,2h et 18,8h (bornes à 0,1 près)
 - D. Il y a 95 chances sur 100 pour que la durée totale de l'intervention soit comprise entre 4,6h et 19,4h (bornes à 0,1 près)
 - E. Il y a 95 chances sur 100 pour que la durée totale de l'intervention soit comprise entre 3,1h et 20,9h (bornes à 0,1 près)

Exercice 7 (2 QCM)

7.1 Dans une certaine pathologie, la durée d'hospitalisation est distribuée selon un χ^2 à 18 degrés de liberté

- A. L'écart-type de la durée d'hospitalisation est 6 jours
- B. 50% des sujets ont une durée d'hospitalisation supérieure à 18 jours
- C. Plus de 5% des patients restent hospitalisés au moins 4 semaines
- D. Plus de 10% des patients restent hospitalisés au moins 4 semaines
- E. Moins de 25% des patients restent hospitalisés entre 3 et 4 semaines

7.2 En réalité, la distribution de la durée d'hospitalisation est mal connue. Quel est le nombre n de sujets sur lesquels mesurer la durée d'hospitalisation pour pouvoir estimer la durée moyenne de séjour avec une précision de ± 1 jour au risque 5% (on prendra 6 comme valeur de l'écart-type)

- A. $10 \leq n \leq 50$
- B. $51 \leq n \leq 120$
- C. $121 \leq n \leq 150$
- D. $151 \leq n \leq 170$
- E. $171 \leq n \leq 200$

Exercice 8 (1 QCM)

8.1 En établissant une formule sanguine, on souhaite estimer la proportion de polynucléaires parmi les leucocytes (globules blancs). Parmi 100 leucocytes, on compte 80 polynucléaires. Dans cette question, on arrondit $u_{0,05}$ à 2.

- A. L'intervalle de confiance de niveau 95% de la proportion de polynucléaires est [0,76 ; 0,84]
- B. L'intervalle de confiance de niveau 95% de la proportion de polynucléaires est [0,74 ; 0,86]
- C. L'intervalle de confiance de niveau 95% de la proportion de polynucléaires est [0,72 ; 0,88]
- D. La probabilité que l'intervalle de confiance de niveau 80% contienne la proportion de polynucléaires est 0,20
- E. Pour obtenir une estimation deux fois plus précise, il faut doubler le nombre de leucocytes

Exercice 9 (1 QCM)

9.1 Une balance A indique le poids sans biais, mais avec une erreur de mesure dont l'écart-type est égal à 2. Une balance B a un biais systématique égal à une constante d , et un écart-type égal à 3. On ne dispose que de la balance B pour se peser, mais on est prêt à se peser plusieurs fois (n fois) et à estimer son poids par la moyenne arithmétique des pesées de manière à obtenir un résultat aussi bon que celui qu'on aurait obtenu avec la balance A en une seule pesée, en terme d'erreur quadratique moyenne.

- A. L'erreur quadratique moyenne de la balance A vaut 2
- B. L'erreur quadratique moyenne de la balance A vaut 4
- C. L'erreur quadratique moyenne après n pesées avec la balance B vaut d^2+9
- D. L'erreur quadratique moyenne après n pesées avec la balance B vaut $(d^2+9)/n$
- E. Avec $d = 1$, il faut 3 pesées avec la balance B pour obtenir un résultat de qualité égale à celle de la balance A en une seule pesée

Exercice 10 (1 QCM)

10.1 Pour estimer la tension artérielle diastolique moyenne dans une population on tire au sort 200 personnes qu'on répartit en deux échantillons de 100 personnes. On obtient les deux moyennes suivantes : 79 et 81 mmHg, et pour écart-type observé : 10 mmHg dans chaque échantillon. On construit un intervalle de confiance de niveau 0,9 sur chacun de ces échantillons et on construit l'intersection de ces deux intervalles

- A. La probabilité que les intervalles de confiance contiennent tous les deux la vraie valeur de la tension artérielle diastolique moyenne est 0,9
- B. La probabilité que les intervalles de confiance contiennent tous les deux la vraie valeur de la tension artérielle diastolique moyenne est 0,81
- C. La probabilité que les intervalles de confiance contiennent tous les deux la vraie valeur de la tension artérielle diastolique moyenne est 0,09
- D. La probabilité que l'intersection des deux intervalles contienne la vraie valeur de la tension artérielle diastolique moyenne est 0,81

- E. La probabilité que l'intersection des deux intervalles contienne la vraie valeur de la tension artérielle diastolique moyenne est 0,9

Exercice 11 (1 QCM)

11.1 Deux sérodiagnostics A et B sont utilisés pour dépister une certaine maladie. Ces deux tests ont été utilisés sur un ensemble d'individus. Dans 30 cas, les deux tests ont donné des conclusions discordantes : dans 20 cas, A a été négatif et B positif ; dans 10 cas, A a été positif et B négatif. Si les deux examens donnent la même proportion de résultats positifs, les probabilités de chacun des deux types de discordance sont égales à $1/2$.

On va donc comparer la proportion observée de discordances de type A négatif, B positif (20/30) avec la valeur théorique $1/2$.

- A. L'hypothèse alternative est que l'un des deux sérodiagnostics donne plus souvent que l'autre des résultats positifs
- B. L'hypothèse nulle est que l'un des deux sérodiagnostics donne plus souvent que l'autre des résultats négatifs
- C. Le paramètre calculé est inférieur à 2
- D. Le paramètre calculé est supérieur à 2,5
- E. On n'a pas mis en évidence que l'un des sérodiagnostics donne plus souvent que l'autre des résultats positifs

Exercice 12 (1 QCM)

12.1 On veut comparer les proportions moyennes de lymphocytes dans deux formes d'une maladie, qu'on notera formes A et B. On tire au sort deux groupes de 50 malades chacun, atteints respectivement des formes A et B. On obtient les résultats suivants : les variances des proportions sont de 0,05 et 0,1 dans les groupes A et B, et les proportions moyennes de lymphocytes sont respectivement de 0,5 et de 0,6.

- A. Pour comparer les proportions moyennes, on peut réaliser un test du χ^2
- B. Pour comparer les proportions moyennes, on peut réaliser un test de comparaison de 2 moyennes observées
- C. Après avoir effectué les calculs, on ne rejette pas l'hypothèse nulle
- D. Après avoir effectué les calculs, on rejette l'hypothèse nulle et $p < 0,05$
- E. Après avoir effectué les calculs, on rejette l'hypothèse nulle et $p < 0,001$

Exercice 13 (1 QCM)

13.1 On veut tester l'hypothèse d'égalité d'atteinte selon le sexe dans une maladie, c'est-à-dire l'hypothèse selon laquelle les proportions d'hommes et de femmes sont égales parmi les personnes atteintes de cette maladie. On tire au sort 100 personnes atteintes de cette maladie, et on observe m hommes. On rejettera l'égalité d'atteinte selon le sexe quand :

- A. $m \in [50 ; 53]$
- B. $m \in [55 ; 57]$
- C. $m \in [61 ; 65]$
- D. $m \in [70 ; 75]$
- E. $m \in [82 ; 83]$

Exercice 14 (1 QCM)

14.1 En vue de comparer deux traitements T1 et T2 d'une affection bénigne, on répartit entre ces deux traitements 250 malades par tirage au sort. Les résultats sont indiqués dans le tableau ci-dessous

Traitement	Etat du malade après 5 jours de traitement			Total
	Stationnaire	Amélioré	Guéri	
T1	15	70	35	120
T2	25	85	20	130

- A. Le test à effectuer est un χ^2 à 3 ddl
- B. L'hypothèse nulle testée est que les fréquences d'état stationnaire, amélioré, guéri sont identiques avec les 2 traitements
- C. L'hypothèse nulle testée est que les fréquences d'état stationnaire, amélioré, guéri sont différentes avec les 2 traitements
- D. Le test statistique montre que les 2 traitements ont des efficacités différentes
- E. Le degré de signification est inférieur à 1%

Exercice 15 (3 QCM)

15.1 On compare l'efficacité de 2 antalgiques A et B dans un essai randomisé sur 2 groupes de 50 sujets. L'efficacité est évaluée par la diminution de la douleur mesurée à l'aide d'une Echelle Visuelle Analogique (EVA). Les résultats obtenus sont indiqués dans le tableau ci-dessous

	effectif	moyenne observée	variance observée
A	50	23	70
B	50	20	90

- A. L'hypothèse nulle est que la diminution de la douleur dans la population est la même avec les deux traitements
- B. Pour effectuer le test de comparaison, on doit supposer que la distribution de la diminution

de la douleur est normale

- C. On ne peut pas rejeter l'hypothèse nulle
- D. Le degré de signification est $p < 0,05$
- E. A est plus efficace que B

15.2 Sachant qu'on s'intéresse à une variation de la diminution de la douleur de ± 3 selon le traitement, combien faudrait-il de sujets par groupe pour la mettre en évidence (si elle existe) en utilisant un test de puissance 90%. On utilisera les variances de la QCM précédente pour les calculs

- A. $n = 40$
- B. $40 < n \leq 100$
- C. $100 < n \leq 200$
- D. $200 < n \leq 600$
- E. $600 < n \leq 1000$

15.3 On dispose de 200 malades pour réaliser un essai thérapeutique sur deux groupes de 100 sujets. On s'intéresse toujours à la mise en évidence d'une variation de la diminution de la douleur de ± 3 selon le traitement. Quelle sera la puissance du test. On utilisera les variances de la QCM 15.1 pour les calculs

- A. $< 30\%$
- B. 30-50%
- C. 51-80%
- D. 81-90%
- E. $> 90\%$

Exercice 16 (1 QCM)

16.1 On suit une population de malades bénéficiant d'une intervention chirurgicale prise comme origine du temps ($t=0$)

Un an après l'intervention, il y a 90% de survivants

Deux ans après l'intervention, il y a 80% de survivants

Trois ans après l'intervention, il y a 60% de survivants

Quatre ans après l'intervention, il y a 40% de survivants

Cinq ans après l'intervention, il y a 20% de survivants

- A. La probabilité qu'un malade a de mourir entre 2 ans et 4 ans après l'intervention est de 40%
- B. La probabilité de survivre 4 ans après l'intervention quand on a déjà survécu 3 ans est $1/3$
- C. La probabilité de survivre 4 ans après l'intervention quand on a déjà survécu 2 ans est $1/2$
- D. La probabilité de vivre encore au moins 4 ans quand on a survécu 1 an après l'intervention est de 20%
- E. La probabilité de vivre encore au moins 4 ans quand on a survécu 1 an après l'intervention est de $2/9$

2

Concours 2007

Avertissement : le concours 2007 ne sera pas mis en ligne

3

Concours Nouméa 2007

*Durée 1 heure 30. Tout document autorisé.
L'épreuve comporte 20 QCM*

Exercice 1 (1 QCM)

1.1 La colposcopie est un examen optique permettant de dépister les dysplasies du col de l'utérus en détectant des zones suspectes. Lorsque (et seulement lorsque) une zone apparaît suspecte au colposcopiste, une biopsie (prélèvement superficiel de quelques mm^2) est réalisée et son histologie précisée grâce à un examen effectué par un anatomopathologiste. L'histologie donne le diagnostic.

On voudrait caractériser les propriétés métrologiques de la colposcopie. Avec les seules données ainsi recueillies chez un certain nombre de femmes on pourra estimer :

- A. La sensibilité de la colposcopie.
- B. La spécificité de la colposcopie.
- C. La valeur prédictive positive de la colposcopie.
- D. La valeur prédictive négative de la colposcopie.
- E. La proportion des faux positifs parmi les positifs.

Exercice 2 (1 QCM)

2.1 Dans une certaine maladie grave, la fonction de survie à compter du diagnostic est divisée par deux entre deux dates quelconques différant d'un an. Indiquez les propositions exactes :

- A. La probabilité de survivre au moins un an après le diagnostic est 0,5.
- B. Les patients décèdent sûrement dans les deux ans qui suivent le diagnostic.
- C. La probabilité de décéder dans les deux ans suivant le diagnostic est de 0,875.
- D. Si l'on a survécu un an, la probabilité de survivre une année supplémentaire est 0,5.
- E. Chaque jour, la probabilité de survivre une année supplémentaire est 0,5.

Exercice 3 (1 QCM)

3.1 Un(e) de vos collègues, peu familier(e) de la biostatistique s'intéresse à la fréquence d'une certaine maladie chronique (sa prévalence). Il (elle) pense que cette fréquence est de 20 % dans sa

clientèle. Par ailleurs il (elle) a consulté les fiches de 900 de ses patients et trouvé 90 patients atteints de cette maladie. Indiquez les propositions exactes.

- A. Après calcul vous pouvez lui dire que la prévalence n'est pas 20 % dans sa clientèle
- B. L'intervalle de confiance de la prévalence (au risque 5 %) est environ $[0,08 \ 0,12]$, à 0,01 près pour chaque borne.
- C. L'intervalle de confiance de la prévalence (au risque 5 %) est environ $[0,17 \ 0,23]$, à 0,01 près pour chaque borne.
- D. L'intervalle de pari de la prévalence (au risque 5 %) était environ $[0,08 \ 0,12]$ à 0,01 près pour chaque borne.
- E. L'intervalle de pari de la prévalence (au risque 5 %) était environ $[0,17 \ 0,23]$ à 0,01 près pour chaque borne.

Exercice 4 (1 QCM)

4.1 On a constaté qu'au service d'urgences d'un certain hôpital se présentaient en moyenne 3 malades chaque jour. Le nombre de patients se présentant chaque jour est supposé suivre une distribution de Poisson. Lorsque 3 malades au plus se présentent, un médecin peut assurer seul leur prise en charge. Lorsqu'il y a au moins quatre malades, un second médecin est mobilisé.

- A. La probabilité pour qu'aucun malade ne se présente de la journée est 0,1 environ (à 0,01 près).
- B. La probabilité pour qu'au plus un malade se présente dans la journée est 0,2 environ (à 0,01 près).
- C. Le nombre moyen de médecins mobilisés pour cette consultation est 1,35 (à 0,02 près).
- D. Le nombre moyen de médecins mobilisés pour cette consultation est 1,65 (à 0,02 près).
- E. Le nombre moyen de médecins mobilisés pour cette consultation est 1,92 (à 0,02 près).

Exercice 5 (10 QCM)

On s'intéresse à une maladie dont la prévalence dans la population étudiée est de 10 %. On notera M l'événement « présence de la maladie ».

On dispose d'un examen e_1 pour faire le diagnostic de cette maladie. On note E_1 l'événement « le résultat de e_1 est positif ». On décide de traiter les patients pour lesquels E_1 est présent et de ne pas traiter les autres.

Cet examen donne un résultat positif dans 50 % de la population, et 19 % des résultats positifs sont des malades.

5.1 La sensibilité de e_1 est

- A. 0,095
- B. 0,19
- C. 0,5
- D. 0,905
- E. 0,95

5.2 Sa valeur prédictive négative est

- A. 0,1
- B. 0,19
- C. 0,55
- D. 0,9
- E. 0,99

5.3 On donne maintenant les coûts (au sens large) des diverses actions

- coût de réalisation de l'examen $e_1 = 1$
- coût d'administration du traitement = 5
- l'administration à tort du traitement a des effets néfastes assimilables à un coût supplémentaire = 5
- ne pas traiter un malade a des conséquences graves assimilables à un coût supplémentaire = 100

On veut calculer le coût moyen associé à la stratégie choisie (traiter si et seulement si l'examen donne un résultat positif). La variable aléatoire coût dont on cherche l'espérance peut prendre 4 valeurs, associées aux 4 possibilités : ne pas traiter avec raison, ne pas traiter à tort, traiter avec raison, traiter à tort.

On s'intéresse à l'éventualité « ne pas traiter avec raison »

- A. Son coût est 1
- B. Son coût est 100
- C. Son coût est 101
- D. Sa probabilité est 0,405
- E. Sa probabilité est 0,495

5.4 On s'intéresse à l'éventualité « ne pas traiter à tort »

- A. Son coût est 1
- B. Son coût est 100
- C. Son coût est 101
- D. Sa probabilité est 0,005
- E. Sa probabilité est 0,095

5.5 Le coût moyen de la stratégie est environ

- A. 1
- B. 5
- C. 6
- D. 11
- E. 89

5.6 Si on décidait de ne plus faire l'examen e_1

- A. Le coût moyen de ne traiter personne serait 10
- B. Le coût moyen de ne traiter personne serait > 15
- C. Le coût moyen de traiter tout le monde serait 5
- D. Le coût moyen de traiter tout le monde serait > 8
- E. Ces résultats ne permettent pas d'affirmer qu'il est plus intéressant de faire l'examen avant de décider de traiter ou non

5.7 On dispose d'un autre examen e_2 pour détecter la présence de la maladie. L'événement « le résultat de e_2 est positif » est noté E_2 et est en faveur de la présence de la maladie. La sensibilité de cet examen est de 90 % et sa spécificité de 80 %. La probabilité que les deux examens donnent simultanément un résultat positif sur un même patient est de 0,25. La probabilité d'avoir la maladie et d'avoir les deux résultats positifs est de 0,085.

On considère les résultats des deux examens pour un même patient

- A. Ils sont indépendants, car ce n'est pas parce qu'on fait e_1 qu'on est obligé de faire e_2
- B. S'ils n'étaient pas indépendants, le résultat de e_2 n'aurait pas de lien avec celui de e_1
- C. S'ils étaient indépendants, le résultat d'un des examens n'apporterait aucune information sur le résultat de l'autre
- D. Si $Pr(E_2)$ était égal à 0,5 ils seraient indépendants
- E. Ils ne sont pas indépendants et $Pr(E_2) = 0,27$

5.8 La stratégie qu'on utilise maintenant est la suivante : on commence par faire l'examen e_1 ; si le résultat est négatif, on ne traite pas ; si le résultat est positif, on fait l'examen e_2 puis on choisit de traiter ou non selon que le résultat de e_2 est positif ou négatif.

Pour chaque patient, la probabilité de le traiter est

- A. La probabilité que les deux examens donnent des résultats positifs
- B. La probabilité que les deux examens donnent des résultats positifs et que le patient soit malade
- C. La probabilité que les deux examens donnent des résultats positifs sachant que le patient est malade
- D. La probabilité que e_2 donne un résultat positif, sachant que e_1 a donné un résultat positif
- E. La probabilité que e_2 donne un résultat positif, sachant que e_1 a donné un résultat positif et que le patient est malade

5.9 La stratégie précédente définit un examen composite e dont le résultat est positif si et seulement si e_1 fournit un résultat positif, puis e_2 fournit un résultat positif

- A. Sans calcul, on peut affirmer que la sensibilité de e est inférieure ou égale aux sensibilités de e_1 et e_2
- B. Sans calcul, on peut affirmer que la sensibilité de e est supérieure ou égale aux sensibilités de e_1 et e_2
- C. Sans calcul, on ne peut pas classer la sensibilité de e par rapport à celles de e_1 et e_2
- D. La sensibilité de e est de 85 %
- E. La sensibilité de e est de 95 %

5.10 On peut vérifier aisément que la spécificité de l'examen composite e est environ 82 %. On suppose que le coût de l'examen e_2 est égal à 2, les autres coûts étant ceux définis pour la question 3

- A. Il est raisonnable de préférer la nouvelle stratégie à celle ne faisant intervenir que l'examen e_1 si son coût moyen est inférieur à celui calculé question 5.
- B. Il ne faut adopter cette nouvelle stratégie que si la sensibilité et la spécificité sont meilleures que celles de e_1 seul
- C. Le coût moyen de la nouvelle stratégie est 3
- D. Le coût moyen de la nouvelle stratégie est environ 5,6
- E. Le coût moyen de la nouvelle stratégie est environ 8,2

Exercice 6 (6 QCM)

On réfléchit à une mesure de prévention innovante qui pourrait réduire le risque de complications après une intervention chirurgicale. On sait que le taux de complications de cette intervention est de 8 %. On espère grâce à cette mesure le ramener à 4 %.

Pour faire la preuve de l'intérêt de cette nouvelle mesure on s'apprête à réaliser un essai clinique auquel participeront des patients dont une moitié bénéficiera de la mesure, l'autre moitié n'en bénéficiant pas. On veut assurer à cette étude une puissance de 0,8.

6.1 La puissance de l'étude est :

- A. La probabilité de ne pas voir l'effet de la mesure alors qu'il y en a un.
- B. La probabilité de voir l'effet de la mesure alors qu'il y en a un.
- C. La probabilité de voir l'effet de la mesure alors qu'il n'y en a pas.
- D. La probabilité de conclure à juste raison au rejet de l'hypothèse nulle.
- E. La probabilité de rejeter l'hypothèse nulle.

6.2 Pour assurer la puissance requise il faut :

- A. Faire une étude randomisée.
- B. Inclure un total d'au moins 396 patients (à 2 près).
- C. Inclure un total d'au moins 542 patients (à 2 près).
- D. Inclure un total d'au moins 1108 patients (à 2 près).
- E. Inclure un total d'au moins 1482 patients (à 2 près).

6.3 Si on choisit d'inclure un total de 600 patients la puissance de l'étude sera :

- A. 0,18 (à 0,01 près)
- B. 0,32 (à 0,01 près)
- C. 0,46 (à 0,01 près)
- D. 0,54 (à 0,01 près)
- E. 0,82 (à 0,01 près)

6.4 Il a été finalement décidé de réaliser une étude randomisée en incluant 412 patients. Les résul-

tats obtenus sont les suivants. Sur les 206 patients ayant bénéficié de la mesure, 14 ont présenté une complication ; cela a été le cas chez 16 patients n'ayant pas bénéficié de la mesure. Indiquez les propositions exactes :

- A. Les proportions observées de complications diffèrent.
- B. On peut conclure que les risques de complications sont différents.
- C. On ne peut conclure que les risques de complications sont différents.
- D. On peut conclure que la mesure prévient les risques de complications.
- E. On peut conclure que la mesure ne prévient pas les risques de complications.

6.5 En réalité, on a noté au cours de l'étude le degré de sévérité (grave ou peu grave) des complications présentées par les patients ; les résultats plus détaillés sont présentés dans le tableau de contingence ci-dessous.

	Pas de complication	Complication peu grave	Complication grave
Patients ayant bénéficié de la mesure de prévention	190	4	12
Patients n'ayant pas bénéficié de la mesure de prévention	192	12	2

Ils permettent de répondre à une nouvelle question : la mesure de prévention a-t-elle une action sur les complications ?

Pour répondre à cette question :

- A. On doit réaliser un test d'ajustement.
- B. On doit réaliser un test d'homogénéité.
- C. On doit réaliser un test d'indépendance.
- D. L'hypothèse nulle sera : « la mesure n'a pas d'action ».
- E. L'hypothèse nulle sera : « la mesure a une action ».

6.6 A partir des données ci-dessus, on peut dire :

- A. On peut conclure que la mesure de prévention a une action sur les complications.
- B. On ne peut conclure à une action de la mesure sur les complications.
- C. Le paramètre du test mis en œuvre pour répondre à la question donne une valeur numérique comprise entre 1,7 et 2.
- D. Pour s'exprimer sur le degré de signification on doit écrire : $p < 0,01$.
- E. Pour s'exprimer sur le degré de signification on doit écrire : $p < 0,001$.

4

Concours 2006

Durée 1heure et 45 minutes. Tout document autorisé.

Exercice 1 (1 QCM)

1.1 On suppose que le nombre de sujets se présentant chaque jour à la consultation d'un médecin entre 11h et 12h suit une loi de Poisson de moyenne 3.

- A. La probabilité qu'au cours d'une consultation de 11h à 12h, aucun malade ne se présente est comprise entre 4 % et 6 % ;
- B. La probabilité qu'au cours d'une consultation de 11h à 12h, aucun malade ne se présente est comprise entre 94 % et 96 % ;
- C. La probabilité qu'au cours d'une consultation de 11h à 12h, un seul malade se présente est comprise entre 10 et 20 % ;
- D. La probabilité qu'au cours d'une consultation de 11h à 12h, un seul malade se présente est supérieure à 20 % ;
- E. La probabilité que 4 malades ou plus se présentent à la consultation est supérieure à 10 %

Exercice 2 (3 QCM)

2.1 Un enfant naît à terme s'il naît au bout de 38 semaines d'aménorrhée au moins (événement T) ; il naît avant terme sinon. L'événement « naître avant terme » est noté AT.

Une grossesse est dite « unique » si la mère ne porte qu'un seul fœtus (événement noté U). Sinon, la grossesse est dite « multiple » (événement M).

La probabilité de naissance avant terme est de 5 % pour une grossesse unique et de 45 % pour une grossesse multiple.

Dans les QCM 1et 2 de l'exercice, on s'intéresse à une population dans laquelle le taux de grossesses multiples est de 10 %.

- A. La probabilité 45 % donnée ci-dessus est $\Pr(AT \cap M)$
- B. La probabilité 45 % donnée ci-dessus est $\Pr(AT / M)$
- C. La probabilité 45 % donnée ci-dessus est $\Pr(M / AT)$
- D. La probabilité de naissance avant terme est 7 %
- E. La probabilité de naissance avant terme est 9 %

2.2 *Les réponses ne sont pas indépendantes de la QCM précédente.*

- A. La probabilité qu'il s'agisse d'une grossesse multiple en cas de naissance avant terme est comprise entre 0,45 et 0,55
- B. La probabilité qu'il s'agisse d'une grossesse multiple en cas de naissance avant terme est comprise entre 0,25 et 0,35
- C. La probabilité qu'il s'agisse d'une grossesse multiple en cas de naissance à terme est comprise entre 0,02 et 0,04
- D. La probabilité qu'il s'agisse d'une grossesse multiple en cas de naissance à terme est comprise entre 0,05 et 0,07
- E. La probabilité qu'il s'agisse d'une grossesse multiple en cas de naissance à terme est comprise entre 0,08 et 0,10

2.3 *Les réponses sont indépendantes de celles des QCM précédentes.*

On suppose maintenant que dans la population générale la probabilité de naissance avant terme est de 0,07. On veut recalculer le taux de grossesses multiples dans cette population.

- A. $\Pr(M) \geq 0,15$
- B. $\Pr(M) = 0,10$
- C. $\Pr(M) = 0,05$
- D. $\Pr(M) = 0,02$
- E. $\Pr(M) < 0,01$

Exercice 3 (2 QCM)

3.1 On considère une variable X_1 distribuée selon une loi normale d'espérance 0 et de variance 4, et une variable X_2 distribuée selon une loi normale centrée réduite.

- A. Si X_1 et X_2 sont indépendantes, $\text{var}(X_1 + X_2) = 9$
- B. Si X_1 et X_2 sont indépendantes (donc $X_1/2$ et X_2 aussi), $\text{var}(X_1/2 + X_2) = 2$
- C. Si la corrélation $\rho(X_1, X_2) = 0,5$, $\text{var}(X_1 + X_2) = 6$
- D. Si $\rho(X_1, X_2) = 0,5$, $\text{var}(X_1 + X_2) = 7$
- E. Quelle que soit la valeur de $\rho(X_1, X_2)$, on a toujours $\text{var}(X_1 + X_2) \leq 9$

3.2 *Les réponses sont indépendantes de la QCM précédente.*

- A. $E(X_1^2) = 0$
- B. $E(X_1^2) = 4$
- C. $E(X_1^2 + X_2^2) = 4$
- D. Si X_1 et X_2 sont indépendantes, $X_1^2 + X_2^2$ est distribuée selon un χ^2 à 2 degrés de liberté
- E. Si X_1 et X_2 sont indépendantes, $X_1^2/4 + X_2^2$ est distribuée selon un χ^2 à 2 degrés de liberté

Exercice 4 (5 QCM)

4.1 Dans une population comportant 10 % de personnes atteintes d'une certaine maladie, on

constitue 2 échantillons indépendants, l'un de 144 personnes atteintes (groupe M), l'autre de 256 non atteintes (groupe N).

Parmi les malades on observe 130 personnes possédant un signe S, mais seulement 51 ont ce signe chez les non malades. On veut évaluer l'intérêt de ce signe dans le diagnostic de la maladie.

- A. Il y a 144 vrais positifs observés
- B. Il y a 130 vrais positifs observés
- C. Il y a 181 vrais positifs observés
- D. On ne peut pas calculer la sensibilité
- E. En notant VP (respectivement FP) le nombre de vrais (respectivement faux) positifs observés la valeur prédictive positive peut s'estimer par $\frac{VP}{VP + FP}$

4.2 Les réponses sont indépendantes de la QCM précédente.

Les résultats sont calculés à 0,001 près

- A. La sensibilité varie avec la prévalence de la maladie
- B. La valeur prédictive positive varie avec la prévalence de la maladie
- C. La sensibilité est estimée à 0,903
- D. La valeur prédictive positive est estimée à 0,335
- E. La valeur prédictive positive est estimée à 0,718

4.3 Les réponses ne sont pas indépendantes de la QCM précédente.

Les résultats sont calculés à 0,001 près

- A. [0,849 ; 0,957] est un intervalle de confiance de niveau 0,94 pour la sensibilité
- B. [0,849 ; 0,957] est un intervalle de confiance de niveau 0,95 pour la sensibilité
- C. [0,849 ; 0,957] est un intervalle de confiance de niveau 0,97 pour la sensibilité
- D. Au niveau 98 %, les conditions de validité du calcul de l'intervalle de confiance de la sensibilité ne sont pas vérifiées
- E. Ces conditions ne sont pas vérifiées au niveau 99 %

4.4 Les réponses sont indépendantes des réponses aux QCM précédentes.

On note $ICSe_{1-\alpha}$ l'intervalle de confiance de niveau $1 - \alpha$ pour la sensibilité et $ICSp_{1-\alpha}$ l'intervalle de confiance de niveau $1 - \alpha$ pour la spécificité

- A. $\Pr(Se \in ICSe_{1-\alpha}) = \alpha$
- B. $\Pr(Se \in ICSe_{1-\alpha}) = 1 - \alpha$
- C. $\Pr[(Se \in ICSe_{1-\alpha}) \cap (Sp \in ICSp_{1-\alpha})] = \alpha$
- D. $\Pr[(Se \in ICSe_{1-\alpha}) \cap (Sp \in ICSp_{1-\alpha})] = (1 - \alpha)^2$
- E. $\Pr[(Se \in ICSe_{1-\alpha}) \cap (Sp \in ICSp_{1-\alpha})] = 1 - \alpha^2$

4.5 Les réponses ne sont pas indépendantes des réponses aux QCM précédentes.

On veut calculer un intervalle de confiance ICVPP pour la valeur prédictive positive à partir des intervalles de confiance $ICSe_{1-\alpha}$ et $ICSp_{1-\alpha}$. On admettra que la borne inférieure (resp. supérieure) de ICVPP se calcule à partir des bornes inférieures (resp. supérieures) de $ICSe_{1-\alpha}$ et $ICSp_{1-\alpha}$.

On veut que la VPP ait une probabilité de 0,94 de se trouver dans l'intervalle ICVPP.

Les résultats sont calculés à 0,001 près.

- A. Compte tenu de la précision de la table, il faut utiliser les intervalles $ICSe_{1-\alpha}$ et $ICSp_{1-\alpha}$ calculés avec $\alpha = 0,06$
- B. Compte tenu de la précision de la table, il faut utiliser les intervalles $ICSe_{1-\alpha}$ et $ICSp_{1-\alpha}$ calculés avec $\alpha = 0,03$
- C. La borne inférieure de ICVPP est 0,272
- D. La borne supérieure de ICVPP est 0,398
- E. Il y a moins de 6 chances sur cent pour qu'un patient présentant le signe ait un risque d'être atteint supérieur à 0,423

Exercice 5 (1 QCM)

5.1 Une étude clinique visant à comparer les proportions observées de patients qui sont encore douloureux une heure après l'administration d'un antalgique selon la dose qu'ils ont reçue (faible ou forte dose) a été dimensionnée (on a calculé un nombre de sujets nécessaires) de façon à assurer une puissance de 80 %.

- A. Cela veut dire qu'il y a 8 chances sur 10 pour que la forte dose soit plus efficace que la faible dose
- B. Cela veut dire que si les proportions des patients encore douloureux ne dépendent pas de la dose reçue, il y a 8 chances sur 10 pour qu'on le voie
- C. Cela veut dire que l'on pourra déceler une différence de $80/2 = 40$ % entre les deux proportions
- D. Pour faire le calcul il suffit de se donner la proportion supposée commune des patients qui resteraient douloureux quelle que soit la dose
- E. Pour faire le calcul on a dû se donner les deux proportions, différentes, de patients qui resteraient douloureux à chaque dose

Exercice 6 (1 QCM)

6.1 Un biologiste s'est engagé dans la comparaison du contenu en phytostérols de plaques d'athérome entre deux populations de patients : les patients recevant un régime supplémenté en phytostérols (patients S), dont il constitue un échantillon de 6 patients, les patients ne recevant pas un tel régime (patients NS), dont il constitue un échantillon de 8 patients. On lui a recommandé d'utiliser un test de Mann-Whitney-Wilcoxon. Il a obtenu une valeur calculée du paramètre du test égale à 2,6.

- A. Compte tenu de la valeur obtenue, il peut conclure que les proportions de plaques contenant des phytostérols diffèrent dans les deux populations
- B. Compte tenu de la valeur obtenue, il peut conclure qu'il y a plutôt plus de phytostérols dans les plaques des patients supplémentés
- C. Compte tenu de la valeur observée il ne peut pas conclure
- D. Le degré de signification du test doit être donné par $p < 0,05$
- E. Le degré de signification du test doit être donné par $p < 0,01$

Exercice 7 (2 QCM)

7.1 On cherche à étudier les relations entre vaccination contre la grippe et emploi. On a demandé à 350 personnes choisies au hasard dans une population, d'une part si elles s'étaient faites vacciner contre la grippe entre septembre et décembre 2005, et d'autre part leur statut socioprofessionnel, en distinguant les trois catégories : sans activité professionnelle (SP), profession de responsabilité élevée (PE), profession de responsabilité faible (PF). Les résultats sont indiqués sur le tableau de contingence ci-dessous.

	SP	PE	PF	Totaux
S'est fait vacciner	6	10	19	35
Ne s'est pas fait vacciner	123	90	102	315
Totaux	129	100	121	350

- A. Pour poursuivre l'étude il faut pratiquer un test de χ^2 à 6 degrés de liberté
- B. L'hypothèse nulle testée est que les variables vaccination et statut socioprofessionnel sont indépendantes
- C. L'hypothèse nulle testée est que les variables vaccination et statut socioprofessionnel sont liées
- D. L'hypothèse nulle testée est qu'en moyenne il y a autant de personnes vaccinées que non vaccinées dans la population
- E. Si tous les effectifs du tableau de contingence n'avaient pas été au moins égaux à 5, le test n'aurait pas pu être réalisé

7.2 Les réponses sont indépendantes de la QCM précédente.
Les valeurs sont calculées à 0,01 près.

- A. Le paramètre calculé du test vaut 4,12
- B. Le paramètre calculé du test vaut 8,47
- C. Le paramètre calculé du test vaut 10,64
- D. Les données permettent de rejeter l'hypothèse nulle
- E. Le degré de signification du test doit être donné par $p < 0,01$

Exercice 8 (3 QCM)

8.1 On s'intéresse à la durée T des symptômes rhumatologiques associés à l'infection par le Chikungunya dans l'île de la Réunion. On appellera S la fonction de survie associée à cette variable T , délai de survenue de l'événement « disparition des symptômes ». La durée est exprimée en jours.

On suppose que S prend notamment les valeurs suivantes :
 $S(2) = 0,9$; $S(4) = 0,81$; $S(12) = 0,5$; $S(31) = 0,3$.

- A. La probabilité de présenter encore des symptômes rhumatologiques à 5 jours est inférieure à 0,5
- B. La proportion « vraie » des patients ne présentant plus de symptômes rhumatologiques à 12 jours est 0,5
- C. 90 pour cent des patients présentant encore des symptômes rhumatologiques à 2 jours les présentent toujours à 4 jours
- D. Plus de la moitié des patients qui présentent encore des symptômes rhumatologiques à 12 jours les présentent toujours à 31 jours
- E. Plus du tiers des patients voient leurs symptômes durer plus de 31 jours

8.2 *Les réponses sont indépendantes de la QCM précédente.*

On suppose maintenant que $S(t) = \exp(-t/18)$; alors $E(T) = 18$ jours.

La probabilité de présenter encore des symptômes rhumatologiques à $(4+t)$ jours alors qu'on les présentait encore à 4 jours est la probabilité de voir ses symptômes durer encore au moins t jours alors qu'ils durent depuis 4 jours. On peut montrer que cette probabilité est égale à la probabilité qu'avaient à leur début les symptômes de durer au moins t jours. La phrase précédente reste juste si on remplace 4 par une valeur quelconque.

- A. Des trois phrases précédentes on déduit que la probabilité de voir les symptômes durer encore t jours alors qu'ils durent depuis x jours ne dépend pas de x
- B. Si l'on présente encore des symptômes rhumatologiques à 5 jours, l'espérance de la durée restante des symptômes est 13 jours
- C. Si l'on présente encore des symptômes rhumatologiques à 5 jours, l'espérance de la durée restante des symptômes est strictement comprise entre 13 et 18 jours
- D. Si l'on présente encore des symptômes rhumatologiques à 5 jours, l'espérance de la durée restante des symptômes est 18 jours
- E. Si l'on présente encore des symptômes rhumatologiques à 5 jours, l'espérance de la durée restante des symptômes est supérieure strictement à 18 jours

8.3 *Les réponses sont indépendantes de celles des QCM précédentes.*

Pour estimer cette fonction de survie S qui en réalité n'est pas connue, on recueille des données auprès d'un échantillon de sujets. On prévoit d'analyser les données de l'enquête le 1^{er} octobre 2006.

- A. Dans cet échantillon il faut inclure des sujets non infectés par le Chikungunya
- B. Individuellement, la date d'origine est la date de survenue des premiers symptômes rhumatologiques
- C. La date d'origine est la date de début de l'infection sur l'île
- D. La date de point est le 1^{er} octobre 2006
- E. La date de point est postérieure au 1^{er} octobre 2006

Exercice 9 (2 QCM)

9.1 Un nouveau médicament régulateur de la circulation rénale fait l'objet d'un essai.

Chez 49 patients on a dosé l'albuminurie (concentration d'albumine dans les urines) avant et après l'administration du médicament. Pour savoir si on constate une variation de l'albuminurie

après administration du médicament, on effectue un test d'hypothèses. Les données recueillies ont permis de calculer les renseignements suivants :

Moyenne observée avant administration :	80 mg/l
Ecart-type observé avant administration :	4 mg/l
Moyenne observée après administration :	79 mg/l
Ecart-type observé après administration :	3 mg/l
Moyenne observée des différences avant-après :	1 mg/l
Ecart-type observé des différences avant-après :	3,2 mg/l

- A. Le paramètre du test est compris entre 0,35 et 0,45
- B. Le paramètre du test est compris entre 1,35 et 1,45
- C. Le paramètre du test est compris entre 2,15 et 2,25
- D. Le degré de signification du test doit être donné par $p < 0,05$
- E. Le degré de signification du test doit être donné par $p < 0,03$

9.2 *Les réponses sont indépendantes de la QCM précédente.*

- A. La corrélation observée entre les valeurs après-avant administration est comprise entre 0,49 et 0,51
- B. La corrélation observée entre les valeurs après-avant administration est comprise entre 0,52 et 0,54
- C. La corrélation observée entre les valeurs après-avant administration est comprise entre 0,60 et 0,62
- D. On peut estimer la moyenne de la baisse de l'albuminurie par l'intervalle à 95 % $[1 \pm x]$ avec $x = 0,4$ (à 0,1 près)
- E. On peut estimer la moyenne de la baisse de l'albuminurie par l'intervalle à 95 % $[1 \pm x]$ avec $x = 0,9$ (à 0,1 près)

5

Concours Nouméa 2006

Durée 1 heure. Tout document autorisé.

Exercice 1 (questions rédactionnelles - 9 points)

Soit la fonction $f(x) = ce^{-\lambda x}$ avec $\lambda > 0$ si $x \geq 0$ et $f(x) = 0$ si $x < 0$.

Rappels :

Une primitive de $f(x)$ est $F(x) = \int_{-\infty}^x f(t)dt = \frac{c}{\lambda}(1 - e^{-\lambda x})$ si $x \geq 0$ et $F(x) = 0$ si $x < 0$.

$e^{-\infty} = 0$; $\frac{de^{ax}}{dx} = ae^{ax}$; $e^x + ae^{bx} = e^x(1 + ae^{(b-1)x})$; $a \text{Ln}(b) = \text{Ln}(b^a)$

I- Généralités (1,5 points)

1.1 Montrer qu'il faut que $c = \lambda$ pour que $f(x)$ soit une densité de probabilité

1.2 Soit X une variable aléatoire distribuée selon la densité de probabilité $f(x)$. On s'intéresse à la probabilité que $X \leq x$. Comment s'appelle cette fonction de x ? Donner son expression.

La loi de X est appelée **loi exponentielle de paramètre λ** . On admettra que $E[X] = 1/\lambda$

II- Un test diagnostique (4,5 points)

On suppose qu'un certain produit injecté dans le sang disparaît après un temps T , variable aléatoire distribuée selon une loi exponentielle de paramètre 1.

Pour les personnes atteintes d'une certaine maladie, touchant 1 % de la population, le temps de disparition du produit dans le sang est distribué selon une loi exponentielle de paramètre $1/k$. Les personnes malades sont notées M . Celles non atteintes sont notées N .

1.3 On sait que le temps moyen de disparition du produit chez les malades est supérieur au temps moyen de disparition chez les non malades. Quelle contrainte cela impose-t-il à k ?

1.4 On veut choisir un seuil s tel que si pour une personne le délai de disparition du produit est supérieur au seuil, elle est déclarée porteuse du signe pathologique S conduisant à conclure à la présence de la maladie ; si ce délai est inférieur au seuil, elle est déclarée non porteuse.

Montrer que la sensibilité vaut $Se = e^{-\frac{s}{k}}$ et que la spécificité vaut $Sp = 1 - e^{-s}$

1.5 Supposons que $k = 3$ et qu'on ait choisi $s = 1,65$.

Calculer la sensibilité et la spécificité. Quel est le risque de ne pas détecter un malade ?

1.6 Calculer les valeurs prédictives

1.7 Les résultats précédents montrent que le test diagnostique n'est pas très bon. On peut penser que le seuil a été mal choisi. On a toujours $k = 3$.

On considère que le meilleur seuil est celui qui rend la somme $Se + Sp$ la plus grande possible. Déterminer ce seuil en utilisant les formules de la question 2 (on se contentera de trouver la valeur du seuil qui annule la dérivée de $Se + Sp$, sans vérifier qu'elle correspond à un maximum). Comparer le résultat avec la valeur précédemment utilisée ($s = 1,65$). Peut-on alors espérer obtenir un meilleur test diagnostique que le précédent ?

III- La surface sous la courbe ROC comme indicateur de la qualité d'un examen (3 points)

Si on dispose d'un examen parfait, il existe un seuil tel que $Se = Sp = 1$.

On montre que la surface sous la courbe ROC correspondante (pour $1 - Sp$ compris entre 0 et 1) est égale à 1. Pour les autres tests (imparfaits), cette surface est inférieure à 1.

1.8 Un examen est inutile si son résultat ne dépend pas du fait d'être malade ou non, c'est à dire que la présence du signe S est indépendante de la présence de la maladie M . Une définition de l'indépendance est $\Pr(S / M) = \Pr(S)$.

1.8.1 En exprimant $\Pr(S)$ à l'aide du théorème des probabilités totales, montrer que l'indépendance s'écrit aussi $\Pr(S/N) = \Pr(S)$, où N représente l'absence de la maladie.

1.8.2 En déduire la relation qui lie Se à $1 - Sp$ en cas d'indépendance

1.8.3 Sans calcul, quelle est la surface sous la courbe ROC correspondante

1.9 La surface sous la courbe ROC est donc comprise entre la valeur précédente et 1. Plus elle est proche de 1, meilleur est l'examen.

On suppose qu'il existe plusieurs formes du produit introduit dans la partie II, correspondant à plusieurs valeurs du nombre k .

On montre alors que la surface sous la courbe ROC (pour $1 - Sp$ compris entre 0 et 1) est $\frac{k}{k+1}$

1.9.1 A partir de quelle valeur de k cette surface est-elle au moins égale à 0,9

1.9.2 Avec cette valeur de k , quel est le meilleur seuil (tel que $Se + Sp$ soit le plus grand possible)

1.9.3 Avec cette valeur de k et ce seuil, que deviennent la sensibilité et la spécificité

Exercice 2 (questions rédactionnelles - 7 points)

On cherche à montrer qu'enrichir un greffon de cellules hématopoïétiques en progéniteurs hématopoïétiques est susceptible de diminuer la profondeur de la thrombopénie (déficit en plaquettes) chez des patients atteints d'un lymphome non hodgkinien bénéficiant d'une autogreffe.

Pour cela on constitue au hasard deux groupes de patients :

- le groupe de patients recevant les cellules enrichies (groupe E)
- le groupe de patients recevant les cellules non enrichies (groupe NE)

Les plaquettes de ces patients sont dosées chaque jour suivant la greffe pendant une période suffisante pour que chez chaque sujet on obtienne la valeur la plus basse post-greffe. Dans la suite les concentrations de plaquettes sont en $10^9/l$.

L'étude a été réalisée chez 8 patients du groupe E et 7 patients du groupe NE. Les valeurs minimum des plaquettes sont les suivantes :

- 2,8 ; 3 ; 3,4 ; 4,2 ; 4,4 ; 5,9 ; 7,1 chez les patients NE
- 4,5 ; 5,1 ; 5,7 ; 7,6 ; 9,1 ; 9,3 ; 9,5 ; 9,6 chez les patients E

2.1 (1 point)

On lit dans le rapport de l'étude :

« 7 patients du groupe NE et 8 patients du groupe E sont nécessaires pour assurer une puissance de 0,9 »

Expliquez le sens de cette phrase

2.2 (1 point)

On lit dans le rapport de l'étude :

« l'analyse sera faite en intention de traiter »

Expliquez le sens de cette phrase.

2.3 (5 points)

On lit dans le rapport de l'étude :

« les niveaux de plaquettes seront comparées par un test d'hypothèses »

Réalisez le test approprié. On veillera à préciser les étapes de mise en œuvre, et le degré de signification le cas échéant. On ira aussi loin que possible dans l'interprétation.

Exercice 3 (QCM - 3 points)

Dans cet exercice, les valeurs sont données avec une précision maximale de 3 chiffres après la virgule.

On s'intéresse à une population de femmes atteintes d'un cancer du sein. Le taux de survie 5 ans après la découverte du cancer est de 65 %.

Lors de la découverte du cancer, on peut définir la gravité du cancer par son stade (1 à 4). 45 % des femmes sont de stade 1, 30 % de stade 2, 15 % de stade 3, et 10 % de stade 4.

3.1 La probabilité qu'une femme de cette population soit de stade 4 et survive au moins 5 ans est 0,03.

- A. La probabilité de survivre au moins 5 ans pour une femme de stade 4 est 0,03
- B. La probabilité de survivre au moins 5 ans pour une femme de stade 4 est 0,065
- C. 70 % des femmes de stade 4 décèdent dans les 5 ans après la découverte de leur cancer
- D. En cas de décès dans les 5 ans, la probabilité que la femme ait été de stade 4 est 0,20
- E. En cas de décès dans les 5 ans, la probabilité que la femme ait été de stade 4 est 0,954

3.2 Le taux de survie à 5 ans est de 50 % chez les stades 3

- A. La probabilité qu'un cancer soit de stade 3 avec une survie d'au moins 5 ans est 0,5
- B. La probabilité qu'un cancer soit de stade 3 avec une survie d'au moins 5 ans est 0,075
- C. La probabilité qu'un cancer soit de stade 3 ou 4 avec une survie d'au moins 5 ans est 0,105
- D. La probabilité de survivre au moins 5 ans pour une femme de stade 3 ou 4 est 0,105
- E. La probabilité de survivre au moins 5 ans pour une femme de stade 3 ou 4 est 0,42

3.3 Le taux de survie à 5 ans est de 80 % chez les stades 1

- A. La probabilité qu'une femme soit de stade 2 et survive au moins 5 ans est 0,185
- B. La probabilité qu'une femme soit de stade 2 et survive au moins 5 ans est 0,195
- C. La probabilité de survivre au moins 5 ans pour une femme de stade 2 est 0,195
- D. La probabilité de survivre au moins 5 ans pour une femme de stade 2 est 0,617
- E. En cas de décès dans les 5 ans, la probabilité que la femme ait été de stade 2 est comprise entre 0,41 et 0,42

Exercice 4 (QCM - 1 point)

4.1 Un sondage a été effectué auprès de 900 patients récemment hospitalisés. On leur a demandé s'ils étaient satisfaits ou non de leur séjour hospitalier. La proportion de satisfaits a été publiée sous forme de la fourchette (intervalle de confiance) suivante : [0,78 0,82].

- A. La proportion observée de patients satisfaits est 0,815
- B. La proportion observée de patients satisfaits est 0,785
- C. L'intervalle de confiance présenté est de niveau 0,95
- D. L'intervalle de confiance présenté est de niveau 0,87
- E. L'intervalle de confiance présenté est de niveau 0,74

6

Concours 2005

Il faut ajouter aux auteurs **A.J. Valleron** et les enseignants de Saint-Antoine qui ont rédigé la moitié des questions de cette épreuve.

L'épreuve comporte deux parties de 3/4 heures chacune.

La première est à réaliser sans document ni calculatrice. Documents et calculatrice sont utilisables dans la seconde.

Exercice 1 (11 QCM - 10 points - sans document ni calculatrice)

1. Dans toute la partie sans document, on arrondira 1,96 à 2 dans les calculs.
2. Les dénominations « moyenne » et « moyenne vraie » sont synonymes.
3. On rappelle la formule de comparaison de moyennes observées (groupes A et B) :

$$\frac{m_A - m_B}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}}$$

4. On rappelle la formule d'estimation par intervalle d'une moyenne :

$$\left[m \mp u_\alpha \frac{s}{\sqrt{n}} \right] \text{ (Pitié-Salpêtrière)}$$

$$\left[m \mp |z|_\alpha \frac{s}{\sqrt{n}} \right] \text{ (St Antoine)}$$

5. On rappelle l'expression de la variance d'une somme de variables aléatoires indépendantes entre elles :

$$\mathbf{var(X1 + X2 + \dots + Xn)} = \mathbf{var(X1)} + \mathbf{var(X2)} + \dots + \mathbf{var(Xn)}$$

6. Distributions usuelles :

$$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) ; \frac{\lambda^n}{n!} \exp(-\lambda) ; C_n^k \Pi^k (1 - \Pi)^{n-k}$$

1.1 QCM 1

On se demande si un nouvel hypotenseur H est plus efficace que le traitement traditionnel T. On réalise un essai thérapeutique en double aveugle où la moitié des patients (tirés au sort) reçoit H, et l'autre reçoit T. Ces deux groupes de patients sont notés H et T. Malheureusement, la pharmacie s'est trompée dans la distribution des médicaments : tous les malades (T et H) ont reçu T, ce dont ni les malades ni les médecins ne se sont aperçus car l'essai est en double aveugle. Le sta-

tisticien effectue le test de comparaison entre les groupes H et T. Lui non plus ne connaît pas l'erreur de la pharmacie. Le test est effectué, comme toujours, au risque de première espèce 5 %. La puissance de l'étude a été établie à 90 %.

- A. Il y a 5 chances sur 100 pour que le test statistique final conduise à déclarer que T et H ont des effets différents
- B. Il y a 95 chances sur 100 pour que le test statistique final conduise à déclarer que T et H ont des effets différents
- C. Il y a 10 chances sur 100 pour que le test statistique final conduise à déclarer que T et H ont des effets différents
- D. Il y a 90 chances sur 100 pour que le test statistique final conduise à déclarer que T et H ont des effets différents
- E. Il y a 0 chance sur 100 pour que le test statistique final conduise à déclarer que T et H ont des effets différents

1.2 QCM 2

Chaque année, il y a une épidémie de syndromes grippaux. On suppose qu'elle touche en moyenne 15 % de la population, et que le fait d'avoir eu un syndrome grippal une année ne modifie pas le risque d'en avoir un l'année suivante (« pas d'immunité acquise »)

- A. L'énoncé indique que les événements « avoir un syndrome grippal en 2004 » et « avoir un syndrome grippal en 2003 » sont des événements indépendants
- B. La probabilité pour un sujet d'avoir eu un syndrome grippal en 2003 et à nouveau un syndrome grippal en 2004 est 30 % (= 0,30)
- C. La probabilité pour un sujet d'avoir eu un syndrome grippal en 2003 ou en 2004 est 30 %
- D. La probabilité pour un sujet d'avoir eu un syndrome grippal en 2003 ou en 2004 est 15 %
- E. Les sujets ayant eu un syndrome grippal en 2003 avaient 85 chances sur 100 de ne pas en avoir en 2004

1.3 QCM 3

On se demande si un traitement T modifie la glycémie des malades qui le reçoivent. On mesure la glycémie des sujets de deux groupes de 49 patients : les patients du premier groupe (groupe T) sont traités par T ; ceux de l'autre groupe ne sont pas traités (groupe NT). Les groupes sont constitués par tirage au sort, et on compare leur moyenne.

Les moyennes et variances observées dans les deux groupes sont : $m_T = 5,9$ mmol/ml, $m_{NT} = 5,5$ mmol/ml, $s^2_T = 0,4$, $s^2_{NT} = 0,6$.

- A. L'hypothèse nulle testée est que les moyennes observées m_T et m_{NT} sont identiques
- B. Au risque 5 %, les moyennes observées m_T et m_{NT} diffèrent significativement
- C. Le degré de signification est compris entre 1 % et 1 ‰
- D. Le degré de signification est compris entre 5 % et 1 %
- E. Le test est valide car les tailles des groupes sont suffisantes

1.4 QCM 4

Après avoir pratiqué un nombre très élevé de scanners dans la population des fumeurs de 50 à 55 ans, on considère que l'on dispose des probabilités suivantes :

- 30 % des fumeurs présentent une anomalie au scanner (avoir une anomalie au scanner est noté A).
 - 20 % des fumeurs indiquent tousser régulièrement (tousser régulièrement est noté T).
 - 10 % des fumeurs présentent une anomalie (A) et indiquent tousser régulièrement (T).
- A. Les événements A et T sont indépendants
 B. Les événements A et T sont incompatibles
 C. La moitié des sujets indiquant tousser régulièrement ont une anomalie au scanner
 D. Le tiers des sujets indiquant tousser régulièrement ont une anomalie au scanner
 E. Le dixième des sujets indiquant tousser régulièrement ont une anomalie au scanner

1.5 QCM 5

On note G_1 , S, G_2 , et M les 4 phases du cycle cellulaire (G_1 , S, et G_2 , sont respectivement la présynthèse, la synthèse et la post synthèse de l'ADN. M est la mitose).

On suppose que les durées de ces phases sont des variables aléatoires indépendantes entre elles. Les moyennes et écarts types des durées de G_1 sont respectivement de 10h et 5h ; de S : 7h et 3h ; de G_2 : 2h et 1h ; de M 1h et 1h. Les distributions des durées de G_1 , S, G_2 , et M ne sont pas gaussiennes.

- A. La durée moyenne du cycle cellulaire est de 20h
 B. L'écart type de la durée du cycle cellulaire est de 10h
 C. L'écart type de la durée du cycle cellulaire est de 6h
 D. L'écart type de la durée cellulaire est $< 6h$
 E. L'écart type de la durée cellulaire est $> 10h$

1.6 QCM 6

On considère 2 événements A et B sur un même ensemble fondamental E, tels que la probabilité de A ne change pas si on apprend que B ne s'est pas produit.

- A. A et B sont indépendants
 B. A et B sont incompatibles
 C. $\Pr(\underline{A}/B) = \Pr(\underline{A}/\bar{B})$
 D. $\Pr(A) = \Pr(A/B)$
 E. $\Pr(A \text{ ou } B) = \Pr(A) + \Pr(B)$

1.7 QCM 7

On considère deux variables aléatoires X et Y et $\rho(X,Y)$ leur coefficient de corrélation

- A. Si $\rho(X,Y) \neq 0$, on sait que X et Y ne sont pas indépendantes
 B. Si $\rho(X,Y) \neq 0$, on ne sait pas si X et Y sont ou non indépendantes
 C. Si X et Y sont indépendantes, on sait que $\rho(X,Y) = 0$
 D. Si X et Y sont indépendantes, on n'est pas sûr que $\rho(X,Y) = 0$
 E. Si X et Y sont liées, on sait que $\rho(X,Y) \neq 0$

1.8 QCM 8

Si le degré de signification d'un test ou d'une étude est inférieur à 0,01, alors c'est :

- A. que le risque de première espèce a été choisi à 0,01
- B. que l'on se trompe avec une probabilité inférieure à 0,01 si on ne rejette pas H_0
- C. que même si le risque de première espèce avait été choisi à 0,01, H_0 aurait été rejetée à ce risque
- D. qu'il aurait fallu choisir un risque de première espèce au moins égal à 0,99 pour rejeter H_0
- E. que la probabilité qu'avait, sous H_0 , le paramètre du test d'indiquer un écart à H_0 plus grand que ne l'a fait la valeur observée du paramètre est inférieure à 0,01.

1.9 QCM 9

M_n désigne comme dans le cours la variable aléatoire moyenne arithmétique. On note X la variable aléatoire d'intérêt, $E(X)$ son espérance et $\text{var}(X)$ sa variance.

- A. Le théorème central limite ne s'applique pas si X est une variable de Bernoulli
- B. Le théorème central limite concerne toutes les variables aléatoires quantitatives
- C. Le théorème central limite exprime que sous certaines conditions M_n a (à peu près) une distribution normale de moyenne $E(X)$, de variance $\text{var}(X)/n$
- D. Le théorème central limite exprime que sous certaines conditions M_n a (à peu près) une distribution normale centré réduite
- E. Le théorème central limite exprime que sous certaines conditions toutes les variables $(M_n - E(X))/\sqrt{\text{var}(X)}$ ont (à peu près) même distribution.

1.10 QCM 10

On cherche à estimer la moyenne de la durée de sommeil dans la population des patients traités par un somnifère. Pour cela on a observé chez 100 de ces patients leur durée de sommeil lors d'un enregistrement nocturne. On a obtenu : $\sum x_i = 670$ heures, et pour variance observée : $s^2 = 4$.

- A. L'estimation ponctuelle standard de la moyenne sera 6,7 heures
- B. L'estimation par intervalle de confiance de niveau 95 % sera environ [6,3 7,1]
- C. L'estimation par intervalle de confiance de niveau 95 % sera environ [5,9 7,5]
- D. Pour obtenir un intervalle de confiance de niveau 95 % de largeur totale 0,2 il aurait fallu observer 1600 patients
- E. Pour obtenir un intervalle de confiance de niveau 95 % de largeur totale 0,2 il aurait fallu observer 400 patients

1.11 QCM 11

On suppose que, habituellement, le nombre de sujets se présentant à la consultation d'un médecin entre 11h et 12h suit une loi de Poisson de moyenne 3. Pour les calculs, on utilisera les valeurs arrondies de la table suivante :

x	0	1	2	3	4	5
$\exp(-x)$	1	0,37	0,14	0,05	0,02	0,007

- A. La probabilité qu'aucun malade ne se présente au cours d'une consultation de 11h à 12h est 0,05
- B. La probabilité qu'aucun malade ne se présente au cours d'une consultation de 11h à 12h est

- 0,95
- C. La probabilité qu'un seul malade se présente au cours d'une consultation de 11h à 12h est 0,37
- D. La probabilité qu'un seul malade se présente au cours d'une consultation de 11h à 12h est 0,15
- E. La probabilité que 4 malades ou plus se présentent au cours d'une consultation de 11h à 12h est 0,027

Exercice 2 (11 QCM - 10 points - avec documents et calculatrice)

2.1 QCM 1

Un nouvel examen biologique permet de distinguer trois formes différentes, A, B et C, d'une maladie intestinale connue depuis longtemps et que l'on diagnostique facilement.

On se place dans la population des sujets présentant cette maladie ; on les appelle malades.

On voudrait montrer que cette classification en trois formes est liée à la douleur (présente-absente) que présentent certains de ces malades.

- A. Cela revient à montrer que la proportion observée de malades présentant des douleurs dépend de la forme de la maladie
- B. Cela revient à montrer que forme de la maladie (variable à trois classes, ou modalités : A, B, C) et douleur (variable à deux classes, ou modalités : présente, absente) ne sont pas indépendantes
- C. Pour le démontrer on peut constituer au hasard deux groupes de malades :
 — dans le premier on recueillera la présence ou l'absence de douleur
 — dans le second on recueillera la forme de la maladie
- D. Pour le démontrer on doit disposer de deux informations sur chaque malade : la forme de la maladie et le statut par rapport à la douleur
- E. Cela revient à montrer que connaître la forme de la maladie n'apporte pas d'information sur le risque de présence de douleur

2.2 QCM 2

Même énoncé que pour le QCM 1. Question indépendante des réponses au QCM 1.

On considère 219 malades. Après avoir recherché la forme de la maladie et interrogé ces malades sur leurs douleurs, vous disposez du tableau d'effectifs suivant :

	A	B	C
Présence de douleurs	38	50	56
Absence de douleurs	34	22	19

Pour répondre à la question précédente, on réalise un test d'hypothèses.

- A. L'hypothèse nulle peut s'exprimer ainsi : forme de la maladie et douleur sont liées

- B. L'hypothèse nulle peut s'exprimer ainsi : forme de la maladie et douleur sont indépendantes
- C. L'égalité $\Pr(\text{forme de la maladie}=A \text{ et douleur=présente}) = \Pr(\text{forme de la maladie}=A)$ est toujours juste sous H_0
- D. Le paramètre du test suit, sous H_0 , un χ^2 à 2 degrés de liberté
- E. On sait déjà que le test peut être réalisé car tous les effectifs du tableau sont supérieurs à 5.

2.3 QCM 3

Même énoncé que pour le QCM 1. Question indépendante des réponses aux QCM 1 et QCM 2 précédents.

Vous réalisez le test.

- A. Le paramètre calculé vaut environ 2,7
- B. Le test ne permet pas de démontrer la liaison entre les variables étudiées
- C. Le degré de signification est compris entre 0,05 et 0,01
- D. Le degré de signification est compris entre 0,01 et 0,001
- E. Dans ce cas on ne doit pas calculer de degré de signification.

2.4 QCM 4

On a suivi le devenir d'un grand groupe de malades atteints d'une maladie M, à partir de la date de diagnostic. On considère alors que l'on dispose des probabilités suivantes : au bout d'un an, 20 % des malades sont morts ; au bout de 2 ans, 50 % des malades sont morts ; au bout de 3 ans, 70 % des malades sont morts ; au bout de 4 ans, 80 % des malades sont morts ; au bout de 5 ans, 80 % des malades sont morts.

- A. La probabilité qu'un malade ayant déjà survécu 2 ans survive moins de 3 ans est 40 %
- B. La probabilité qu'un malade ayant déjà survécu 2 ans survive moins de 3 ans est 60 %
- C. La probabilité qu'un malade ayant déjà survécu 2 ans survive au moins 4 ans est 20 %
- D. La probabilité qu'un malade ayant déjà survécu 2 ans survive au moins 4 ans est 40 %
- E. La probabilité qu'un malade ayant déjà survécu 2 ans survive au moins 4 ans est 60 %

2.5 QCM 5

On sait d'avance que le taux sanguin T (en u/ml) d'une substance est plus élevé en moyenne chez les malades atteints de la maladie (notés : M) que chez les non malades (notés : NM). On veut se servir de la mesure de T pour aider au diagnostic de M. On suppose d'abord connaître parfaitement la distribution de T chez les malades M et les non malades NM.

Malades M	Distribution de T : gaussienne (normale)	$\mu_M = 110$	$\sigma_M = 10$
Non malades NM	Distribution de T : gaussienne (normale)	$\mu_{NM} = 100$	$\sigma_{NM} = 10$

- A. Il y a moins de 5 % des NM dont le taux est >110
- B. Il y a entre 10 et 15 % des NM dont le taux est >110
- C. Il y a plus de 15 % des NM dont le taux est >110
- D. La valeur du taux de T dépassée par 10 % des sujets NM est comprise entre 125 et 130

E. La valeur du taux de T dépassée par 10 % des sujets NM est supérieure à 130

2.6 QCM 6

Même énoncé que pour le QCM 5. Question indépendante des réponses au QCM 5.

On décide de fixer le seuil diagnostique S, au dessus duquel on déclare le sujet malade, à 100 u/ml.

- A. La sensibilité est comprise entre 15 % et 20 %
- B. La sensibilité est de 50 %
- C. La spécificité est comprise entre 15 % et 20 %
- D. La spécificité est de 50 %
- E. La spécificité est comprise entre 80 % et 85 %

2.7 QCM 7

Même énoncé que pour le QCM 5. Les réponses ne sont pas indépendantes de celles du QCM précédent.

On suppose en plus que la prévalence de la maladie M est de 10 %

- A. La valeur prédictive positive du test (avec le seuil 100) est comprise entre 30 et 50 %
- B. La valeur prédictive positive du test (avec le seuil 100) est comprise entre 50 et 70 %
- C. La valeur prédictive positive du test (avec le seuil 100) est comprise entre 70 et 90 %
- D. La valeur prédictive positive du test (avec le seuil 100) est inférieure à 30 %
- E. La valeur prédictive positive du test (avec le seuil 100) est supérieure à 70 %

2.8 QCM 8

Même énoncé que pour le QCM 5. Réponses indépendantes des réponses aux QCM précédents.

En réalité, on s'aperçoit qu'on ne connaît pas exactement la valeur de μ_M (qu'on avait supposée égale à 110 u/ml), et on veut l'estimer à partir d'un échantillon de malades. On note M_n la variable aléatoire moyenne arithmétique de T sur un échantillon de taille n.

On cherche la valeur de n telle que M_n ait 99 chances sur 100 d'être à moins de 2,5 u/ml de la vraie valeur de μ_M . Pour les calculs, on considèrera que l'écart type σ de T vaut 10.

- A. La variance de M_n sera σ^2/n
- B. La variance de M_n sera $n\sigma^2$
- C. La valeur de n est comprise entre 90 et 160
- D. La valeur de n est comprise entre 160 et 240
- E. La valeur de n est comprise entre 240 et 300

2.9 QCM 9

On s'intéresse ici à la durée de séjour des patients d'un service spécialisé donné, à l'hôpital. L'estimation de la moyenne et de la variance de la durée de séjour D s'obtient à partir des observations faites pendant une longue période (par exemple un an). Toutefois on pourrait penser qu'il suffit de venir un jour donné à l'hôpital et d'utiliser les durées de séjour (qui seront connues à la

sortie du patient) des seuls patients présents ce jour là pour obtenir les estimations de la moyenne et de la variance de D. Cette attitude conduit cependant à des résultats faux, la variable J (durée de séjour à un jour donné) étant différente de la variable D (durée de séjour).

Cet exercice vise à préciser certains liens entre J et D.

On notera μ_J et μ_D les espérances mathématiques de ces 2 variables ; σ_J^2 et σ_D^2 leurs variances. On considère 100 patients qui ont occupé successivement, sans interruption, et dans un ordre indéterminé, un lit donné du service spécialisé. Les durées de séjour de ces patients sont données par la table suivante :

Durée D	1 jour	2 jours	3 jours	4 jours	5 jours
Proportion des patients	0,1	0,4	0,3	0,1	0,1

- La durée totale d'occupation du lit par les 100 patients est de 270 jours.
- La durée totale d'occupation du lit par les patients qui restent 2 jours est de 40 jours.
- Durant la période d'occupation du lit, la probabilité que le patient présent dans le lit un jour donné ait une durée de séjour de 2 jours est de 0,4
- Durant la période d'occupation du lit, la probabilité que le patient présent dans le lit un jour donné ait une durée de séjour de 2 jours est de 0,296
- Durant la période d'occupation du lit, la probabilité que le patient présent dans le lit un jour donné ait une durée de séjour de 3 jours est de 0,333.

2.10 QCM 10

Même énoncé que pour le QCM 9. Les réponses sont indépendantes de celles du QCM précédent.

Dans la suite, on considère que les proportions données dans le tableau sont en réalité des probabilités et donc que les probabilités calculées à la question précédente sont valides dans la population générale.

- $\mu_D = 3$
- $\mu_D = 2,7$
- $\sigma_D^2 = 1,21$
- $\sigma_D^2 = 1,87$
- $\sigma_D^2 = 8,50$

2.11 QCM 11

Même énoncé que pour les QCM 9 et QCM 10. Les réponses ne sont pas indépendantes de celles du QCM 9.

On admettra que $\sigma_{J^2} = 1,31 = 1,142$.

- $\mu_J = 3,15$
- $\mu_J = 3$
- La probabilité que le patient présent dans le lit un jour donné ait une durée de séjour comprise entre 2 et 5 jours inclus est de 96,3 %

- D. Pour 100 patients, l'intervalle [2,9 jours ; 3,4 jours] est un intervalle de pari (ou de fluctuation) de niveau 95 % pour la durée moyenne de séjour calculée un jour donné (les bornes sont arrondies à la première décimale)
- E. Pour 100 patients, l'intervalle [2,9 jours ; 3,4 jours] est un intervalle de confiance de niveau 95 % pour la durée moyenne de séjour calculée un jour donné (les bornes sont arrondies à la première décimale)

7

Concours Nouméa 2005

Durée 45 minutes. Tout document autorisé.

La correction proposée en ligne est trop détaillée par rapport à ce qui est demandé pour le concours puisque les réponses aux QCM ne doivent pas être justifiées.

Exercice 1 (QCM - 10 points)

On admet dans cet exercice que la glycémie à jeun, notée G et exprimée en mmoles/l , est distribuée selon une loi normale de moyenne 10 et de variance 4 chez les diabétiques (notés D), alors qu'elle est distribuée selon une loi normale de moyenne 5 et de variance 1 dans le reste de la population.

1.1 On veut définir un intervalle, symétrique autour de 10, contenant la valeur de la glycémie de 95 % des diabétiques

- A. C'est un intervalle de pari d'une variable aléatoire
- B. C'est un intervalle de confiance
- C. Pour le calculer, on doit supposer qu'il y a au moins 30 individus diabétiques
- D. L'intervalle est approximativement $[6 \ 14]$
- E. L'intervalle est approximativement $[9,6 \ 10,4]$

1.2 On observe sur un individu une glycémie à jeun de 7 mmoles/l

- A. La probabilité de cette valeur est plus forte chez les diabétiques que chez les non diabétiques
- B. Cette valeur a la même probabilité chez les diabétiques que chez les non diabétiques
- C. La probabilité d'avoir une glycémie comprise entre 6,5 et 7,5 est de 0,065 chez les diabétiques
- D. La probabilité d'avoir une glycémie comprise entre 6,5 et 7,5 est 1,5 fois plus élevée chez les diabétiques que chez les non diabétiques
- E. Les individus dont la glycémie est comprise entre 6,5 et 7,5 sont appelés les vrais positifs

1.3 On veut définir un seuil g de la glycémie tel qu'on déclare diabétique un individu dont la glycémie à jeun est supérieure ou égale au seuil, et non diabétique un individu dont la glycémie est inférieure au seuil.

On choisit d'abord le seuil $g = 7$. On admet alors les résultats suivants :

sensibilité = $Se = 0,935$

spécificité = $Sp = 0,975$

- A. $Se = Pr(G = 7 / D)$
- B. $Se = Pr(G = 7 / \bar{D})$
- C. Si on choisissait un seuil égal à 6, on peut affirmer sans calcul que la sensibilité augmenterait
- D. Si on choisissait un seuil égal à 6, on peut affirmer sans calcul que la spécificité augmenterait
- E. Si on choisissait un seuil égal à 6, on peut affirmer sans calcul que la spécificité diminuerait

1.4 On choisit maintenant le seuil $g = 7,4$. Les valeurs de sensibilité et spécificité deviennent :

- A. $Se = 0,895$
- B. $Se = 0,905$
- C. $Se = 0,999$
- D. $Sp = 0,990$
- E. $Sp = 0,995$

1.5 On s'intéresse au risque de déclarer diabétique une personne qui ne l'est pas

- A. Ce risque est $1 - Se$
- B. Ce risque est $1 - Sp$
- C. Ce risque est $1 - VPP$
- D. Ce risque est $1 - VPN$
- E. Ce risque dépend de la prévalence du diabète

1.6 On considère la variable aléatoire C , coût (au sens large) des erreurs de diagnostic. Cette variable prend la valeur 1 pour l'erreur « déclarer non diabétique une personne diabétique », et la valeur k pour l'erreur « déclarer diabétique une personne qui ne l'est pas ». On veut calculer le coût moyen (espérance du coût) des erreurs de diagnostic.

- A. $E(C) = k + 1$
- B. $E(C) = (k + 1) / 2$
- C. $E(C) = 1 - Se + k (1 - Sp)$
- D. $E(C) = 1 - Sp + k (1 - Se)$
- E. $E(C) = 1 - VPP + k (1 - VPN)$

1.7 Pour choisir le meilleur seuil de séparation entre les diabétiques et les non diabétiques, on sélectionne, parmi $g = 7$ et $g = 7,4$, celui qui conduit au coût moyen le plus faible.

- A. Si $k = 1/2$ il faut choisir $g = 7$
- B. Si $k = 1/2$ il faut choisir $g = 7,4$
- C. Si $k = 3$ il faut choisir $g = 7$
- D. Si $k = 3$ il faut choisir $g = 7,4$
- E. Il existe une valeur $k > 1$ telle que les 2 seuils conduisent au même coût moyen

1.8 Dans la suite, on considère que le seuil choisi est $g = 7$
On donne la valeur prédictive $VPP = 0,59$

- A. La VPP peut s'exprimer en fonction de la prévalence du diabète
- B. La prévalence du diabète est environ 1,6 %
- C. La prévalence du diabète est environ 3,7 %
- D. La VPN est environ 0,997
- E. La VPN est environ 0,998

1.9 Les diabétiques se divisent en 10 % d'insulino-dépendants (DID) et 90 % de non insulino-dépendants (DNID).

10 % des diabétiques seront amputés du pied ; parmi eux, les 4/5 sont des DNID.

Dans cette question, on se restreint à la sous-population des diabétiques

- A. $Pr(\text{Amputation} \cap \text{DID}) = 0,010$
- B. $Pr(\text{Amputation} / \text{DID}) + 9Pr(\text{Amputation} / \text{DNID}) = 1$
- C. $Pr(\text{Amputation} / \text{DID}) = 0,020$
- D. $Pr(\text{Amputation} / \text{DID}) = 0,200$
- E. $Pr(\text{Amputation} / \text{DNID}) = 0,080$

1.10 On suppose qu'une amputation du pied ne peut être due qu'au diabète et que le risque d'amputation ne dépend pas du niveau de la glycémie :

$$Pr[\text{Amputation} / (D \cap (G \geq 7))] = Pr[\text{Amputation} / D]$$

On admettra la relation suivante, valable pour 3 événements A, B et C quelconques :

$$Pr[(A \cap B) / C] = Pr[A / (B \cap C)] Pr(B / C)$$

- A. La probabilité d'amputation pour un individu de glycémie supérieure ou égale à 7 est de 0,1
- B. La probabilité d'amputation pour un individu de glycémie supérieure ou égale à 7 est de 0,059
- C. La probabilité d'amputation pour un individu de glycémie supérieure ou égale à 7 est de 0,0037
- D. La probabilité d'amputation pour un individu de glycémie inférieure à 7 est environ de 0,0003
- E. La probabilité d'amputation pour un individu de glycémie inférieure à 7 est environ de 0,0002

Exercice 2 (10 points)

Fin 1997, aux Etats-Unis, on a interrogé tous les adultes habitant un même pâté de maisons (block) sur leur état de santé ressenti au cours de l'année. Au total 885 hommes et 959 femmes ont été interrogés.

Parmi eux, 704 hommes et 878 femmes ont déclaré avoir ressenti au moins un épisode de maladie aiguë ou chronique durant l'année. Ces derniers sujets ont en outre été interrogés sur leur comportement lors de cet épisode : ont-ils attendu la fin de l'épisode (comportement « attente »), ont-ils consulté un médecin de ville (comportement « ville »), ont-ils consulté à l'hôpital (comportement « hôpital ») ?

On a obtenu les résultats suivants :

	Attente	Ville	Hôpital
Femmes	191	600	87
Hommes	154	468	82

On veut savoir si le genre (femme-homme) et le comportement en cas d'épisode de maladie sont liés.

2.1 (1 point)

Pour répondre à la question on doit effectuer un test d'hypothèses. Pourquoi ?

2.2 (5 points)

Répondre à cette question en précisant les étapes de mise en œuvre du test. Le cas échéant, vous calculerez le degré de signification.

2.3 (1 point)

Si on voulait savoir si le comportement en cas d'épisode de maladie diffère entre les hommes et les femmes, serait-il nécessaire de changer les modalités de l'expérience ? Discutez.

2.4 (3 points)

La survenue d'au moins un épisode de maladie dans l'année est-elle liée au genre ? Limitez-vous à :

- exprimer le paramètre du test avant d'effectuer le calcul
- donner sa valeur calculée
- formuler clairement la décision

8

Concours 2004

Durée 45 minutes. Tout document autorisé.

La correction proposée en ligne est trop détaillée par rapport à ce qui est demandé pour le concours. En particulier, les réponses aux QCM ne doivent pas être justifiées.

Exercice 1 (6 points)

Un examen d'exploration fonctionnelle respiratoire requérant la participation active du patient est déconseillé chez les patients de plus de 80 ans, appelés sujets âgés, car sa réalisation est jugée trop difficile à réussir chez de tels sujets. Malgré tout et pour vous convaincre de la justesse de cette habitude, vous entreprenez une étude clinique concernant 200 patients : cent sujets âgés ($\text{âge} > 80$ ans), et cent sujets non âgés ($\text{âge} \leq 80$ ans). Parmi plusieurs autres caractéristiques vous notez leur résultat à une épreuve cognitive permettant d'attribuer au patient le statut « état mental normal » ou le statut « état mental altéré ».

A chacun de ces 200 patients vous avez tenté de faire passer l'épreuve fonctionnelle ; cela a pu conduire à l'échec ou au succès.

Les résultats expérimentaux obtenus sont consignés dans le tableau ci-dessous :

	Non âgé		Agé	
	Succès	Echec	Succès	Echec
Etat mental normal	72	18	48	12
Etat mental altéré	4	6	16	24

1.1 (4 points)

Les probabilités de succès diffèrent-elles selon que le patient présente un état mental normal ou un état mental altéré ?

Pour répondre à cette question on réalisera un test d'hypothèses dont on détaillera les étapes de mise en œuvre.

1.2 (0,5 point)

A quelle valeur estimez-vous la probabilité de succès chez les patients âgés présentant un état mental normal ?

1.3 (0,5 point)

A quelle valeur estimez-vous la probabilité de succès chez les patients non âgés présentant un état mental normal ?

1.4 (1 point)

Est-il raisonnable de déconseiller cet examen chez les patients âgés ?

Si non quelles seraient vos recommandations et pourquoi ?

Exercice 2 (QCM - 8,5 points)

Lors des interventions, les services de chirurgie d'un hôpital utilisent des boîtes de compresses. Chaque boîte contient normalement 10 compresses. A la fin d'une intervention, les compresses utilisées sont comptabilisées avant d'être jetées et le nombre obtenu est comparé avec le nombre de boîtes vidées pour s'assurer qu'aucune compresse n'est oubliée dans le corps du patient. Malheureusement certaines boîtes contiennent 9 compresses et d'autres 11.

Dans cet hôpital **60 %** des boîtes de compresses proviennent du fournisseur f_1 , et les **40 %** restantes viennent du fournisseur f_2 .

Le fournisseur f_1 indique que **3 %** de ses boîtes ont 9 compresses, et **0,01 %** en ont 11.

Le fournisseur f_2 indique seulement que **2 %** de ses boîtes ont 9 compresses.

Partie I

On notera 9C, 10C et 11C le fait qu'une boîte contient respectivement 9, 10 ou 11 compresses.

2.1 Dans l'hôpital, sans faire aucun calcul, les données précédentes permettent d'écrire

- A. $P(f_1) = 0,6$
- B. $P(f_1 \cap 9C) = 0,03$
- C. $P(f_1 / 9C) = 0,03$
- D. $P(9C / f_1) = 0,03$
- E. $P(9C)$ est au plus égal à 0,03
- F. $P(11C)$ est au plus égal à 0,0001

2.2 Quelle(s) est(sont) la(les) probabilité(s) correcte(s) (arrondies à 3 décimales) ?

- A. $P(9C) = 0,020$
- B. $P(9C) = 0,025$
- C. $P(9C) = 0,030$
- D. $P(f_1 \cap 9C) = 0,018$
- E. $P(9C / f_1) = 0,026$
- F. $P(f_1 / 9C) = 0,692$

2.3 Parmi les affirmations suivantes concernant f_2 , la(les)quelle(s) est(sont) exacte(s) ?

- A. Lorsque l'hôpital utilise une boîte de compresses, il y a une chance sur deux qu'elle provienne de f_2 .

- B. Si la boîte utilisée ne comporte que 9 compresses, il y a 2 % de chance qu'elle provienne de f_2 .
- C. Le risque d'utiliser une boîte provenant de f_2 et ne comportant que 9 compresses est de 0,8 %
- D. La probabilité qu'une boîte de 9 compresses provienne de f_2 vaut 0,308 (arrondi à 3 décimales)
- E. La probabilité qu'une boîte provienne de f_2 , conditionnellement au fait qu'elle ne contienne que 9 compresses, vaut 0,40
- F. La probabilité définie en E n'est pas calculable, puisque le fournisseur f_2 n'a pas indiqué son taux de boîtes de 11 compresses.

Partie II

On s'intéresse au nombre de boîtes de 9 compresses qu'on trouvera dans l'hôpital parmi un ensemble de n boîtes de compresses.

2.4 On sait que la proportion observée p_o du nombre de boîtes de 9 compresses est une variable aléatoire.

- A. L'espérance mathématique de p_o est 0,026
- B. L'espérance mathématique de p_o est 0,03
- C. La variance de p_o est $0,0291/n$ (arrondi à 3 chiffres significatifs)
- D. La variance de p_o est $0,0253/n$ (arrondi à 3 chiffres significatifs)
- E. p_o est distribuée selon une loi normale dès que $n \geq 30$
- F. p_o est distribuée selon une loi normale dès que $n \geq 193$

2.5 Lorsqu'on a deux variables aléatoires X et Y , telles que $Y = aX$, a étant une constante, on sait que si X est distribuée selon une loi normale, Y aussi. Que sait-on aussi ?

- A. $E(Y) = E(X)$
- B. $E(Y) = aE(X)$
- C. $E(Y) = a^2E(X)$
- D. $\text{var}(Y) = \text{var}(X)$
- E. $\text{var}(Y) = a\text{var}(X)$
- F. $\text{var}(Y) = a^2\text{var}(X)$

2.6 On étudie $n = 500$ boîtes de compresses. On veut définir un intervalle contenant le nombre de boîtes de 9 compresses qu'on va y trouver, avec un risque d'erreur de 10 %

- A. On va construire un intervalle de confiance
- B. On va construire un intervalle de pari
- C. Au risque 10 %, p_o se trouvera dans l'intervalle $[0,0143 ; 0,0377]$ (valeurs arrondies à 3 chiffres significatifs)
- D. Au risque 10 %, p_o se trouvera dans l'intervalle $[0,0174 ; 0,0426]$ (valeurs arrondies à 3 chiffres significatifs)
- E. Au risque 10 %, le nombre de boîtes de 9 compresses sera compris entre 9 et 21 (valeurs arrondies aux entiers les plus proches)
- F. Au risque 10 %, le nombre de boîtes de 9 compresses sera compris entre 7 et 19 (valeurs

arrondies aux entiers les plus proches)

Partie III

La présence de boîtes de 11 compresses peut conduire à oublier une compresse dans le corps du patient, donc à une réintervention. On aimerait donc connaître la probabilité d'obtenir des boîtes de 11 compresses avec le fournisseur f_2 .

Comme précédemment, on notera 11C le fait qu'une boîte contienne 11 compresses.

En utilisant le théorème de la multiplication, on montre que $P(11C) = \frac{P(11C/f_1)P(f_1)}{P(f_1/11C)}$

On peut alors en déduire que $P(11C/f_2) = P(11C/f_1) \frac{P(f_1)}{P(f_2)} \left(\frac{1}{1 - P(f_2/11C)} - 1 \right)$ et on peut vérifier qu'il s'agit d'une fonction croissante de $P(f_2/11C)$

2.7 Sur un ensemble de **30** boîtes de 11 compresses, on en dénombre **18** provenant de f_2 .

- A. La probabilité qu'une boîte de 11 compresses provienne de f_2 est alors 18/30
- B. 0,6 est une estimation ponctuelle de la probabilité conditionnelle $P(f_2/11C)$
- C. 0,6 est une estimation ponctuelle de la probabilité conjointe $P(f_2 \cap 11C)$
- D. Au risque 1 %, l'estimation par intervalle de confiance de la probabilité dont 0,6 est une estimation ponctuelle est [0,425 ; 0,775] (bornes arrondies à 3 décimales)
- E. Au risque 1 %, l'estimation par intervalle de confiance de la probabilité dont 0,6 est une estimation ponctuelle est [0,370 ; 0,830] (bornes arrondies à 3 décimales)
- F. La plus sévère des conditions de validité de l'intervalle de confiance consiste à vérifier que $11,10 > 5$ (valeur arrondie à 2 chiffres après la virgule)

2.8 En utilisant les informations et résultats précédents, on obtient donc

- A. L'estimation ponctuelle de $P(11C)$ est $1,5 \times 10^{-4}$
- B. L'estimation ponctuelle de $P(11C)$ est 10^{-4}
- C. L'estimation ponctuelle de $P(11C/f_2)$ est $1,5 \times 10^{-4}$
- D. L'estimation ponctuelle de $P(11C/f_2)$ est $2,25 \times 10^{-4}$
- E. L'estimation par intervalle de confiance de niveau 99 % de $P(11C/f_2)$ est $[1,11 \times 10^{-4} ; 5,17 \times 10^{-4}]$ (bornes arrondies à 2 décimales)
- F. L'estimation par intervalle de confiance de niveau 99 % de $P(11C/f_2)$ est $[0,88 \times 10^{-4} ; 7,32 \times 10^{-4}]$ (bornes arrondies à 2 décimales)

2.9 On finit par apprendre que f_2 produit **0,03 %** de boîtes contenant 11 compresses. On veut savoir s'il ne serait pas préférable de s'adresser à un seul des 2 fournisseurs f_1 et f_2 .

Pour l'hôpital, le coût d'une boîte de 10 compresses est égal à 1 (dans une certaine unité) quel que soit le fournisseur.

Une boîte de 9 compresses est considérée comme ayant un coût de 1,5, puisqu'elle conduit à des achats complémentaires et à des gênes lors de l'intervention.

Une boîte de 11 compresses a un coût bien plus élevé, noté $1+x$

- A. L'espérance mathématique du coût d'une boîte du fournisseur f_1 est $10^{-4}x + 1,015$
- B. L'espérance mathématique du coût d'une boîte du fournisseur f_1 est $6 \times 10^{-5}x + 0,609$

- C. Si $x = 10$, il est préférable de n'acheter qu'au fournisseur f_1 qui conduit au coût le plus faible (en espérance mathématique)
- D. Si $x = 10$, il est préférable de n'acheter qu'au fournisseur f_2 qui conduit au coût le plus faible (en espérance mathématique)
- E. On peut continuer à acheter aux deux fournisseurs s'ils conduisent aux mêmes coûts (en espérance mathématique), ce qui se produit lorsque $x = 25$
- F. Les deux fournisseurs ont des coûts identiques (en espérance) lorsque $x = 50$

Exercice 3 (QCM - 2 points)

Dans cet exercice, les valeurs sont arrondies au centième près.

3.1 Un examen a une sensibilité égale à 0,8 et une spécificité égale à 0,9 pour une maladie M dans une population A. La prévalence de M dans cette population est de 0,1.

- A. La valeur prédictive positive de l'examen vaut 0,09
- B. La valeur prédictive positive de l'examen vaut 0,89
- C. La valeur prédictive positive de l'examen vaut 0,47
- D. La valeur prédictive négative de l'examen vaut 0,82
- E. La valeur prédictive négative de l'examen vaut 0,98
- F. La valeur prédictive négative de l'examen vaut 0,89

3.2 Dans une population B, l'examen a les mêmes valeurs de sensibilité (0,8) et de spécificité (0,9) pour la maladie M, mais sa valeur prédictive positive est de $2/3$. Que peut-on en déduire ?

- A. La prévalence de la maladie M vaut 0,2
- B. La prévalence de la maladie M vaut 0,8
- C. La prévalence de la maladie M vaut 0,5
- D. La proportion d'individus avec un examen positif dans la population B vaut 0,33
- E. La proportion d'individus avec un examen positif dans la population B vaut 0,24
- F. La proportion d'individus avec un examen positif dans la population B vaut 0,55

Exercice 4 (QCM - 3,5 points)

On cherche à évaluer l'intérêt diagnostique d'un nouvel examen pour une maladie G dans l'ensemble de la population française. D'après des connaissances acquises sur des examens voisins, on s'attend à trouver une sensibilité voisine de 0,9 et une spécificité voisine de 0,85. On va réaliser l'expérience suivante : on va tirer au sort 2 échantillons, un parmi les malades atteints de G, et un parmi les sujets non malades, et estimer les sensibilité et spécificité à l'aide de ces échantillons.

4.1 Si on souhaite estimer les sensibilité et spécificité de l'examen par des intervalles de confiance de niveau 90 % et une précision (demi-longueur des intervalles) de 0,01 pour chaque paramètre, combien d'individus au minimum doit-on tirer au sort dans les populations de malades et de non malades ?

- A. 2436 malades et 4899 non malades

- B. 3458 malades et 4898 non malades
- C. 5973 malades et 8461 non malades
- D. 2436 malades et 3451 non malades
- E. 3458 malades et 3451 non malades
- F. Les conditions de validité de sont pas vérifiées

4.2 On imagine qu'on répète 20 fois l'expérience précédente dans 20 régions françaises différentes. On s'intéresse à la variable aléatoire « nombre d'intervalles de confiance de niveau 90 % contenant la vraie valeur de la sensibilité ».

Quelle est la moyenne de cette variable ?

- A. 15
- B. 16
- C. 17
- D. 18
- E. 19
- F. 20

4.3 Combien d'expériences en moyenne parmi les 20 fourniront un intervalle de confiance de niveau 90 % contenant la vraie valeur de la sensibilité et un intervalle de confiance de niveau 90 % contenant la vraie valeur de la spécificité ? (résultat arrondi à l'entier le plus proche).

- A. 15
- B. 16
- C. 17
- D. 18
- E. 19
- F. 20

4.4 Quelle est la probabilité pour qu'aucune des 20 expériences ne fournisse un intervalle de confiance de niveau 90 % contenant la vraie valeur de la sensibilité ?

- A. $0,9^{20}$
- B. $0,1^{20}$
- C. $0,9^{10}$
- D. $0,1^{10}$
- E. 0,88 (arrondi à 0,01 près)
- F. 0,12 (arrondi à 0,01 près)

9

Concours Nouméa 2004

Durée 45 minutes. Tout document autorisé.

La correction proposée en ligne est trop détaillée par rapport à ce qui est demandé pour le concours puisque les réponses aux QCM ne doivent pas être justifiées.

**Dans cet énoncé, les questions comportent une ou plusieurs bonnes réponses.
Les 2 exercices sont indépendants.**

Le saturnisme infantile (intoxication au plomb) est un problème de santé publique important. On définit le saturnisme par une plombémie supérieure à $100 \mu\text{g/l}$.

Dans la suite on parlera simplement de saturnisme pour dire saturnisme infantile.

- Une des causes connues de saturnisme est le fait d'habiter dans un immeuble dégradé et construit avant 1948. Dans la suite, on notera HD un tel immeuble.
- Une autre cause de plombémie élevée est le comportement de PICA (l'enfant met à la bouche tout ce qu'il ramasse).

Exercice 1 (QCM - 10 points)

Dans un certain arrondissement de Paris, on sait que **10 %** des enfants présentent une plombémie supérieure à $100 \mu\text{g/l}$

1.1 Parmi les enfants habitant un HD, **35 %** présentent une plombémie supérieure à $100 \mu\text{g/l}$. Que représente ce taux de 0,35 :

- A. La probabilité de saturnisme dans l'arrondissement
- B. La probabilité d'habiter un HD et d'être atteint simultanément de saturnisme
- C. La probabilité conditionnelle de saturnisme sachant que l'habitat est un HD
- D. La prévalence du saturnisme dans l'arrondissement
- E. La sensibilité de l'habitation dans un HD pour le saturnisme
- F. La valeur prédictive positive de l'habitation dans un HD pour le saturnisme

1.2 On veut savoir s'il y a indépendance entre les variables habitat (HD ou non) et saturnisme (présent ou absent)

- A. Il faudrait faire un test statistique, mais on ne connaît pas la taille de l'échantillon
- B. Il ne faut pas faire de test statistique puisqu'on connaît les probabilités permettant de répondre à la question
- C. Les variables sont indépendantes car ce n'est pas parce qu'on a une plombémie supérieure à 100 µg/l que l'habitat est un HD
- D. Les variables sont indépendantes car ce n'est pas parce que l'habitat est un HD qu'on a une plombémie supérieure à 100 µg/l (la probabilité n'est que de 0,35)
- E. Les variables ne sont pas indépendantes car la probabilité de saturnisme est modifiée par l'habitat HD
- F. Il faudrait connaître la probabilité que l'habitat soit HD pour pouvoir répondre.

1.3 Dans l'arrondissement considéré, **15 %** des enfants habitent un HD

- A. La probabilité qu'un enfant de l'arrondissement soit atteint de saturnisme et habite un HD vaut 0,0525
- B. La probabilité de l'événement défini dans la proposition A vaut 0,035
- C. La probabilité qu'un enfant de l'arrondissement habitant un HD soit atteint de saturnisme est 0,525
- D. La probabilité de l'événement défini dans la proposition C vaut 0,233 (arrondie à 3 décimales)
- E. La probabilité qu'un enfant de l'arrondissement atteint de saturnisme habite un HD est 0,525
- F. La probabilité de l'événement défini dans la proposition E vaut 0,233 (arrondie à 3 décimales)

1.4 Une famille de 5 enfants habite un HD de cet arrondissement. On s'intéresse à la probabilité que 2 enfants exactement de cette famille soient atteints de saturnisme

- A. Cette probabilité est comprise entre 0,072 et 0,073
- B. Cette probabilité est comprise entre 0,295 et 0,296
- C. Cette probabilité est comprise entre 0,336 et 0,337
- D. Cette probabilité vaut 0,4
- E. Il n'y a pas de condition à vérifier pour que la valeur obtenue soit valide
- F. Il faudrait que $0,4 \times 5 = 5$ pour que la valeur obtenue soit valide

1.5 Les enfants de l'arrondissement vivant dans des HD ont une probabilité de saturnisme de **0,49** s'ils ont un comportement de PICA, et de **0,05** sinon.

Si on restreint la population aux enfants de l'arrondissement vivant dans des HD :

- A. La probabilité de saturnisme pour ces enfants est le produit des probabilités 0,49 et 0,05
- B. La probabilité de saturnisme pour ces enfants est la somme de la probabilité de saturnisme chez ceux ayant un comportement de PICA et de la probabilité de saturnisme chez ceux n'ayant pas ce comportement
- C. La probabilité de saturnisme pour ces enfants est la somme de la probabilité de saturnisme et de comportement PICA, et de la probabilité de saturnisme et d'absence de comportement PICA
- D. La probabilité d'avoir un comportement PICA pour ces enfants est comprise entre 0,11 et

0,12

- E. La probabilité d'avoir un comportement PICA pour ces enfants est comprise entre 0,68 et 0,69
- F. La probabilité d'avoir un comportement PICA pour ces enfants est comprise entre 0,71 et 0,72

1.6 On considère une famille vivant dans un HD de l'arrondissement et comprenant 3 enfants ayant un comportement de PICA. On veut connaître la probabilité que 2 exactement de ces 3 enfants soient atteints de saturnisme.

- A. Cette probabilité est comprise entre 0,036 et 0,037
- B. Cette probabilité est comprise entre 0,122 et 0,123
- C. Cette probabilité est comprise entre 0,238 et 0,239
- D. Cette probabilité est comprise entre 0,367 et 0,368
- E. Cette probabilité est comprise entre 0,666 et 0,667
- F. Cette probabilité est égale à 0,980

1.7 On admet sans calcul les probabilités suivantes concernant les familles de l'arrondissement habitant un HD (ces valeurs sont calculables à partir des données des questions précédentes) :

- Si une famille a 3 enfants ayant un comportement de PICA
 - La probabilité qu'aucun d'eux ne soit atteint de saturnisme est 0,133
 - La probabilité qu'un et un seul d'entre eux soit atteint est 0,382
- Si une famille a 2 enfants n'ayant pas un comportement de PICA
 - La probabilité qu'aucun d'eux ne soit atteint de saturnisme est 0,903
 - La probabilité qu'un et un seul d'entre eux soit atteint est 0,095
- A. Dans une famille de 5 enfants habitant un HD, parmi lesquels 3 exactement ont un comportement de PICA, la probabilité qu'il y ait exactement 2 enfants atteints de saturnisme et qu'ils aient un comportement de PICA est comprise entre 0,032 et 0,034
- B. La probabilité de l'événement défini dans la proposition A est comprise entre 0,214 et 0,216
- C. La probabilité de l'événement défini dans la proposition A est comprise entre 0,331 et 0,333
- D. Dans une famille de 5 enfants habitant un HD, parmi lesquels 3 exactement ont un comportement de PICA, la probabilité qu'il y ait exactement 2 enfants atteints de saturnisme est comprise entre 0,068 et 0,071
- E. La probabilité de l'événement défini dans la proposition D est comprise entre 0,250 et 0,253
- F. La probabilité de l'événement défini dans la proposition D est comprise entre 0,367 et 0,370

Exercice 2 (QCM - 10 points)

Dans un arrondissement de Paris voisin (noté a_2), on tire au sort un échantillon de **200** enfants. On observe que **40** d'entre eux sont atteints de saturnisme. On se demande si le taux de saturnisme infantile dans cet arrondissement diffère de **10 %**, taux dans l'arrondissement étudié en première partie (noté a_1).

2.1 Pour répondre à cette question on envisage de faire un test statistique

- A. Cela n'est pas vraiment nécessaire, car $0,2 \neq 0,1$
- B. On choisit un test de comparaison de 2 proportions observées
- C. On choisit un test de comparaison d'une proportion observée à une proportion théorique
- D. Pour pouvoir effectuer le test en utilisant la loi normale, il faut que le nombre d'individus soit au moins 30 (ici on a bien $200 \geq 30$)
- E. Pour pouvoir effectuer le test en utilisant la loi normale, il faut que $200 \times 0,1$ et $200 \times 0,9$ soient supérieurs ou égaux à 5 (ce qui est vérifié ici)
- F. Pour pouvoir effectuer le test en utilisant la loi normale, il faut que $200 \times 0,2$ et $200 \times 0,8$ soient supérieurs ou égaux à 5 (ce qui est vérifié ici)

2.2 L'hypothèse nulle du test est

- A. $0,2 = 0,1$
- B. Les proportions observées de saturnisme sont les mêmes dans les 2 arrondissements
- C. Les probabilités de saturnisme sont les mêmes dans les 2 arrondissements
- D. La proportion vraie de saturnisme dans l'arrondissement a_2 est égale à 0,1
- E. Les taux de saturnisme ne sont pas les mêmes dans les deux arrondissements
- F. La proportion vraie de saturnisme dans l'arrondissement a_2 est différente de 0,1

2.3 Le risque de première espèce du test :

- A. C'est la probabilité de rejeter l'hypothèse nulle
- B. C'est la probabilité de rejet à tort de l'hypothèse nulle
- C. C'est la probabilité de ne pas rejeter l'hypothèse nulle alors qu'elle est fautive
- D. On le choisit à 95 %
- E. On le choisit à 5 %
- F. Le test serait plus puissant si on diminuait ce risque

2.4 La règle de décision

- A. Elle est basée sur un intervalle construit autour d'une valeur observée
- B. Elle est basée sur un intervalle construit autour d'une valeur théorique
- C. La probabilité que la valeur du paramètre calculé pour le test tombe dans cet intervalle, si l'hypothèse nulle est vraie, est le risque α .
- D. La probabilité que la valeur du paramètre calculé pour le test tombe en dehors de cet intervalle, si l'hypothèse nulle est vraie, est la puissance du test.
- E. Si la valeur calculée pour le paramètre tombe en dehors de l'intervalle, on doit rejeter l'hypothèse nulle
- F. Si la valeur calculée pour le paramètre tombe dans l'intervalle, on dit qu'on ne peut pas rejeter l'hypothèse nulle car on aurait un risque α de se tromper si on l'acceptait

2.5 Le paramètre z du test

- A. Sous l'hypothèse nulle, le paramètre z du test suit une loi normale centrée réduite
- B. Sous l'hypothèse nulle, le paramètre z du test suit une loi normale de moyenne 0,1

- C. Sous l'hypothèse alternative, le paramètre z du test suit une loi normale de moyenne 0,2
- D. L'intervalle de pari à utiliser est $[-1,96 ; 1,96]$
- E. La valeur calculée du paramètre z est comprise entre 3,5 et 3,6
- F. La valeur calculée du paramètre z est comprise entre 4,7 et 4,8

2.6 Conclusion

- A. On n'a pas montré de différence entre les 2 arrondissements
- B. On rejette l'hypothèse nulle au risque 5 %
- C. Le taux de saturnisme dans l'arrondissement a_2 est supérieur à celui de l'arrondissement a_1 .
- D. Il ne faut pas calculer de degré de signification
- E. Le degré de signification est inférieur à 10^{-5}
- F. Le degré de signification est inférieur à 10^{-3} (mais supérieur à 10^{-5})

2.7 On désire estimer la valeur du taux de saturnisme de l'arrondissement a_2 par un intervalle symétrique autour de la valeur observée, avec un risque d'erreur de 5 %. On admettra que toutes les conditions de validité nécessaires aux calculs sont vérifiées.

Les bornes des intervalles proposés sont arrondies à 3 décimales.

- A. L'intervalle est $[0,100 ; 0,300]$
- B. L'intervalle est $[0,145 ; 0,255]$
- C. L'intervalle est $[0,158 ; 0,242]$
- D. L'effectif de l'échantillon qui aurait permis d'obtenir un intervalle dont la demi-longueur serait de 1 % est compris entre 3450 et 3600 enfants
- E. L'effectif de l'échantillon qui aurait permis d'obtenir un intervalle dont la demi-longueur serait de 1 % est compris entre 6050 et 6400 enfants
- F. L'effectif de l'échantillon qui aurait permis d'obtenir un intervalle dont la demi-longueur serait de 1 % est compris entre 19950 et 20500 enfants

10

Concours 2003

Durée 45 minutes. Tout document autorisé.

La correction proposée en ligne est trop détaillée par rapport à ce qui est demandé pour le concours. En particulier, les réponses aux QCM ne doivent pas être justifiées.

Exercice 1 (7 points)

On souhaite savoir si un nouveau traitement médical a une efficacité préventive de la survenue d'occlusion dans l'année suivant une certaine intervention digestive. On hésite en outre entre deux doses d'administration. On envisage alors une étude dans laquelle vont être constitués trois groupes de patients d'effectifs égaux. Un groupe recevra un placebo (traitement sans efficacité biologique), un groupe le traitement à la dose D1, un groupe le traitement à la dose D2.

L'objectif de l'étude est de montrer que les probabilités de survenue d'occlusion ne sont globalement pas identiques selon que les patients reçoivent le placebo, le traitement à la dose D1 ou le traitement à la dose D2.

L'étude est conçue de sorte que la puissance du test statistique permettant de répondre à la question posée soit de 80 % lorsque l'on suppose que la probabilité d'occlusion est 0,03 aussi bien avec la dose D1 qu'avec la dose D2, alors qu'elle vaut 0,06 sous placebo. Les calculs montrent que dans ces conditions il convient d'inclure 617 patients dans chacun des trois groupes.

1.1 (0,8 point)

Rappeler en quelques mots, et dans le contexte de l'étude envisagée, ce que signifie cette puissance de 80 %.

1.2 (4,7 points)

Après avoir suivi les patients pendant un an après l'intervention digestive, on a obtenu des résultats consignés dans le tableau de contingence ci-dessous :

	Groupe placebo	Groupe Dose D1	Groupe Dose D2
Survenue d'occlusion	40	24	17
Non survenue d'occlusion	577	593	600

Y a-t-il une différence entre les trois groupes de patients vis-à-vis du risque d'occlusion à un an ?

Pour répondre à cette question on effectuera un test d'hypothèses dont on détaillera toutes les étapes.

1.3 (1 point)

On se demande si les résultats obtenus chez les patients recevant le placebo sont compatibles avec une probabilité de récurrence à un an de 0,06 (qui est la probabilité supposée de survenue d'occlusion chez les patients non traités). Pour répondre à cette question on décide de **faire un test autre qu'un χ^2** .

Parmi les propositions suivantes, choisissez la bonne réponse :

- A. Au signe près, le paramètre du test vaut environ 3,13 et la probabilité d'occlusion est donc supérieure à 0,06, $p < 10^{-3}$
- B. Au signe près, le paramètre du test vaut environ 3,13 et le résultat est donc compatible avec une probabilité d'occlusion de 0,06
- C. Au signe près le paramètre du test vaut environ 0,505 et la probabilité d'occlusion est donc supérieure à 0,06, $p < 10^{-3}$
- D. Au signe près, le paramètre du test vaut environ 0,505 et le résultat est donc compatible avec une probabilité d'occlusion de 0,06
- E. Au signe près, le paramètre du test vaut environ 3,416 et le résultat est donc compatible avec une probabilité d'occlusion de 0,06
- F. Au signe près le paramètre du test vaut environ 3,416 et la probabilité d'occlusion est donc supérieure à 0,06, $p < 10^{-3}$

1.4 (0,5 point)

Exprimer en quelques phrases les conditions expérimentales que l'on a dû respecter pour pouvoir attribuer au(x) traitement(s) les différences éventuellement trouvées aux questions 2 et 3.

Exercice 2 (QCM - 8,5 points)

Partie I

Après une certaine opération digestive, la probabilité de faire une occlusion dans l'année est de **0,044**. Le fait d'avoir ou non fait une occlusion pendant une année ne modifie pas le risque d'occlusion pour les années suivantes.

On admettra dans l'exercice qu'on peut faire au plus une occlusion par an.

Les valeurs numériques proposées dans l'exercice sont arrondies avec 3 chiffres significatifs.

Cocher la ou les bonnes réponses.

2.1 Considérons les événements « faire une occlusion la première année » et « faire une occlusion la deuxième année »

- A. Ils sont identiques.
- B. Le premier implique le second.
- C. Le second implique le premier.
- D. Ils sont indépendants.
- E. Ils sont exclusifs.

2.2 Quelle est la probabilité d'avoir fait au moins une occlusion au bout de 5 ans ?

- A. 165×10^{-9}
- B. 0,019
- C. 0,201
- D. 0,220
- E. 0,799

2.3 La loi de probabilité du nombre d'occlusions qu'a faites un patient au bout de 5 ans est :

- A. Binomiale.
- B. Normale (sans condition particulière).
- C. Normale si on observe un nombre suffisant de patients.
- D. Student (car $n < 30$).
- E. χ^2 à 5 degrés de liberté

2.4 L'espérance (μ) et l'écart type (σ) du nombre d'occlusions qu'a faites un patient au bout de 5 ans sont :

- A. $\mu = 0$ et $\sigma = 1$
- B. $\mu = 0,044$ et $\sigma = 0,00841$
- C. $\mu = 0,044$ et $\sigma = 0,0917$
- D. $\mu = 0,220$ et $\sigma = 0,210$
- E. $\mu = 0,220$ et $\sigma = 0,459$

Partie II

En fait l'intervention digestive est **soit** une intervention sur le tube digestif haut (dite **intervention TH**), **soit** une intervention sur le tube digestif bas (dite **intervention TB**). La probabilité annuelle d'occlusion est de **0,03** pour une intervention TH, et de **0,10** pour une intervention TB.

2.5 Considérons les événements « faire une occlusion la première année après intervention TH » et « faire une occlusion la première année après une intervention TB » :

- A. Ils sont identiques.
- B. Le premier implique le second.
- C. Le second implique le premier.
- D. Ils sont indépendants.
- E. Ils sont exclusifs.

2.6 Quelle est la probabilité d'avoir fait au moins une occlusion au bout de 5 ans après une intervention TB ?

- A. 10^{-5}
- B. 0,090
- C. 0,410
- D. 0,500
- E. 0,590

Partie III

On sait que **20 % des interventions ont lieu sur le tube digestif bas.**

2.7 Quelle est la probabilité qu'un patient ait une intervention TB et au moins une occlusion dans les 5 ans ?

- A. 0,020
- B. 0,082
- C. 0,100
- D. 0,138
- E. 0,200

2.8 Quelle est, pour un patient, la probabilité d'avoir fait au moins une occlusion dans les 5 ans, si on ne connaît pas le type de l'intervention digestive ?

- A. 0,044
- B. 0,195
- C. 0,201
- D. 0,410
- E. 0,551

Partie IV

On se place dans la sous-population des patients ayant eu au moins une occlusion dans les 5 ans après l'intervention. On veut y déterminer la probabilité pour un patient d'avoir eu une intervention sur le tube digestif bas.

2.9 Sans faire de calcul, on sait que cette probabilité sera :

- A. Identique à celle d'avoir eu une intervention TB chez les patients n'ayant pas eu d'occlusion dans les 5 ans.
- B. Egale à la proportion d'intervention TB sur l'ensemble des patients opérés, diminuée de la probabilité d'intervention TB pour les patients n'ayant pas eu d'occlusion dans les 5 ans.
- C. La probabilité d'avoir eu une intervention TB et une occlusion dans les 5 ans.
- D. La probabilité d'avoir eu une intervention TB conditionnellement au fait d'avoir eu une occlusion dans les 5 ans.
- E. La probabilité d'avoir eu une occlusion dans les 5 ans conditionnellement au fait d'avoir eu une intervention TB.

2.10 La valeur de cette probabilité est :

- A. 0,082
- B. 0,200
- C. 0,408
- D. 0,421
- E. 0,590

Partie V

On suit pendant 1 an un échantillon de **195 patients** opérés. On calcule ensuite sur cet échantillon le pourcentage des individus ayant fait au moins une occlusion dans l'année.

2.11 La probabilité que ce pourcentage soit compris entre 0,03 et 0,1 est (arrondie avec 2 décimales seulement) :

- A. 0,17
- B. 0,34
- C. 0,50
- D. 0,66
- E. 0,83

2.12 Pour obtenir un intervalle de pari pour ce pourcentage, centré, de niveau 0,95, et de largeur 0,07, il faudrait envisager un échantillon de taille :

- A. 94
- B. 132
- C. 195
- D. 224
- E. 328

Exercice 3 (QCM - 3 points)

La thrombose des artères coronaires est un facteur de risque d'infarctus. Dans une population on cherche à dépister une thrombose des artères coronaires, de manière à pratiquer un pontage avant la survenue d'un infarctus chez les patients présentant une thrombose coronaire. La coronarographie est l'examen de référence pour diagnostiquer les thromboses coronaires : on supposera que sa sensibilité et sa spécificité sont de 100 % : une coronarographie positive (négative) est donc équivalente à la présence (à l'absence) d'une thrombose coronaire. Mais elle comporte un risque vital, aussi on décide de réaliser l'étude suivante : sur un groupe de sujets exposés on pratique dans un premier temps un examen moins informatif, mais non invasif, la scintigraphie. Chez les patients ayant une scintigraphie positive, on réalisera une coronarographie, tandis qu'en cas de négativité de la scintigraphie, on ne fera pas d'autre exploration.

3.1 Après cette étude, quel(s) paramètre(s) concernant l'intérêt de la scintigraphie pour diagnostiquer la thrombose coronaire pourrait-on estimer ?

- A. La sensibilité.
- B. La spécificité.
- C. La valeur prédictive positive.
- D. La valeur prédictive négative.
- E. La prévalence de la thrombose coronaire.

3.2 La scintigraphie est **positive chez 70 %** des patients de la population concernée. Quand on réalise une coronarographie chez les patients ayant une scintigraphie positive, on trouve en

moyenne 50 % de thrombose coronaire.

Avec les informations fournies, on peut affirmer que :

- A. La prévalence de la thrombose coronaire dans la population est de 35 %.
- B. La sensibilité de la scintigraphie est de 70 %.
- C. La sensibilité de la scintigraphie est de 50 %.
- D. La spécificité de la scintigraphie est de 50 %.
- E. La valeur prédictive positive de la scintigraphie est de 50 %.

3.3 On ajoute l'information suivante : en l'absence de dépistage, la proportion de thromboses coronaires dans la population concernée est de 40 %.

Parmi les propositions suivantes, lesquelles sont vraies (valeurs à 0,1 % près) ?

- A. La prévalence de la thrombose coronaire dans la population est de 40 %.
- B. La sensibilité de la scintigraphie est de 87,5 %.
- C. La spécificité de la scintigraphie est de 41,7 %.
- D. La proportion de faux négatifs dans la population est de 5 %.
- E. La valeur prédictive négative est de 83,3 %.

Exercice 4 (QCM - 1,5 points)

Après un traitement antibiotique, on cherche à estimer le risque de rechute d'une infection dans le mois qui suit le début du traitement. Pour estimer ce risque, on constitue par tirage au sort un échantillon de 100 malades auxquels on administre le traitement. Parmi ces malades, 20 rechutent dans le mois.

4.1 L'intervalle de confiance de niveau 90 % du risque de rechute est :

- A. [0,1216 ; 0,2784]
- B. [0,1969 ; 0,2031]
- C. [0,1342 ; 0,2658]
- D. [0,1974 ; 0,2026]
- E. On ne peut pas le calculer, car les conditions de validité ne sont pas remplies.

4.2 On appelle *intérêt* du traitement antibiotique la fonction inverse du risque de rechute de l'infection. (*intérêt* = $1/p$, où p est le risque de rechute). L'intervalle de confiance de niveau 95 % de l'intérêt du traitement antibiotique est (avec deux chiffres significatifs) :

- A. [0,72 ; 0,88]
- B. [3,6 ; 8,2]
- C. [0,74 ; 0,86]
- D. [3,8 ; 7,5]
- E. [4,0 ; 6,0]

11

Concours Nouméa 2003

Durée 45 minutes. Tout document autorisé.

La correction proposée en ligne est trop détaillée par rapport à ce qui est demandé pour le concours. En particulier, les réponses aux QCM ne doivent pas être justifiées.

Exercice 1 (7 points)

Tous les enfants d'une population de personnes porteuses d'une certaine anomalie génétique voient apparaître un déficit moteur dans leurs 6 premières années.

Soit X la variable aléatoire représentant la date d'apparition du déficit.

Sa distribution de probabilité est définie de la manière suivante :

- Pour tout intervalle x_1, x_2 dans la période 0 à 1 an :

$$P([x_1 < X \leq x_2]) = \frac{h}{2}(x_2^2 - x_1^2)$$

- Pour tout intervalle x_1, x_2 dans la période 1 à 3 ans :

$$P([x_1 < X \leq x_2]) = h(x_2 - x_1)$$

- Pour tout intervalle x_1, x_2 dans la période 3 à 6 ans :

$$P([x_1 < X \leq x_2]) = h \left(2x_2 - 2x_1 + \frac{x_1^2}{6} - \frac{x_2^2}{6} \right)$$

1.1 (1 point)

En remarquant que $P([0 < X \leq 6]) = 1$, montrer que $h = 1/4$.

1.2 (0,5 points)

Calculer la probabilité pour que le déficit apparaisse entre 2 et 4 ans.

1.3 (1,5 points)

Calculer cette probabilité pour les enfants chez qui le déficit n'est pas encore apparu à l'âge de 1 an.

1.4 (1,7 points)

On veut déterminer $F_X(x)$, la fonction de répartition de X . Cette fonction est évidemment nulle pour $x \leq 0$.

1.4.1 Montrer que pour $0 < x \leq 1$, on a $F_X(x) = x^2/8$

1.4.2 Montrer que pour $1 < x \leq 3$, on a $F_X(x) = x/4 - 1/8$

1.4.3 Calculer $F_X(x)$ pour $3 < x \leq 6$

1.5 (1 point)

A quel âge 50 % des personnes porteuses de l'anomalie génétique auront-elles développé un déficit moteur (c'est ce qu'on appelle l'âge médian) ?

1.6 (0,8 points)

On désire recoder l'âge d'apparition du déficit en classes. On introduit donc une nouvelle variable aléatoire Y définie par :

- $Y = 1/2 \Leftrightarrow 0 < X \leq 1$
- $Y = 2 \Leftrightarrow 1 < X \leq 3$
- $Y = 9/2 \Leftrightarrow 3 < X \leq 6$

Y ne peut prendre aucune autre valeur.

Définir la loi de probabilité de Y (c'est-à-dire les probabilités de chacune des valeurs possibles)

1.7 (0,5 points)

Calculer l'espérance mathématique de Y .

Exercice 2 (7 points)

Dans le but d'évaluer l'efficacité d'un nouveau traitement contre une infection virale, une étude clinique portant sur 180 patients a été menée. Ces patients ont été répartis en un groupe N de 100 patients bénéficiant du nouveau traitement et un groupe C de 80 patients bénéficiant du traitement classique.

Après douze semaines de traitement, la proportion de patients présentant une charge virale indétectable, ce qui correspond au succès du traitement, a été de 0,65 dans le groupe N et de 0,45 dans le groupe C .

2.1 (1 point)

Comment les patients ont-ils dû être répartis entre les deux groupes pour répondre à l'objectif initial ?

2.2 (0,5 point)

Quelle précaution supplémentaire a-t-on dû prendre pour recruter les 180 patients ?

2.3 (5,5 points)

En considérant que les précautions précédentes ont été prises, comparez les efficacités des deux traitements, l'efficacité étant définie par la probabilité de succès.

On répondra à cette question en réalisant un test d'hypothèses dont on détaillera toutes les étapes.

Exercice 3 (QCM - 6 points)**Partie I (2 points)**

Les sensibilités respectives de deux tests T_1 et T_2 de dépistage d'une infection virale sont Se_1 et Se_2 . Les résultats des deux tests sont indépendants dans la population des malades infectés.

3.1 Quelle serait la sensibilité d'un nouveau test T_3 considéré comme positif si T_1 et T_2 le sont en même temps ?

- A. $Se_1 + Se_2$
- B. $Se_1 \times Se_2$
- C. $Se_1 + Se_2 - Se_1 \times Se_2$
- D. $(1 - Se_1) \times (1 - Se_2)$
- E. On ne peut pas la calculer

3.2 Même question pour un test T_4 considéré comme positif si au moins un des tests T_1 et T_2 est positif.

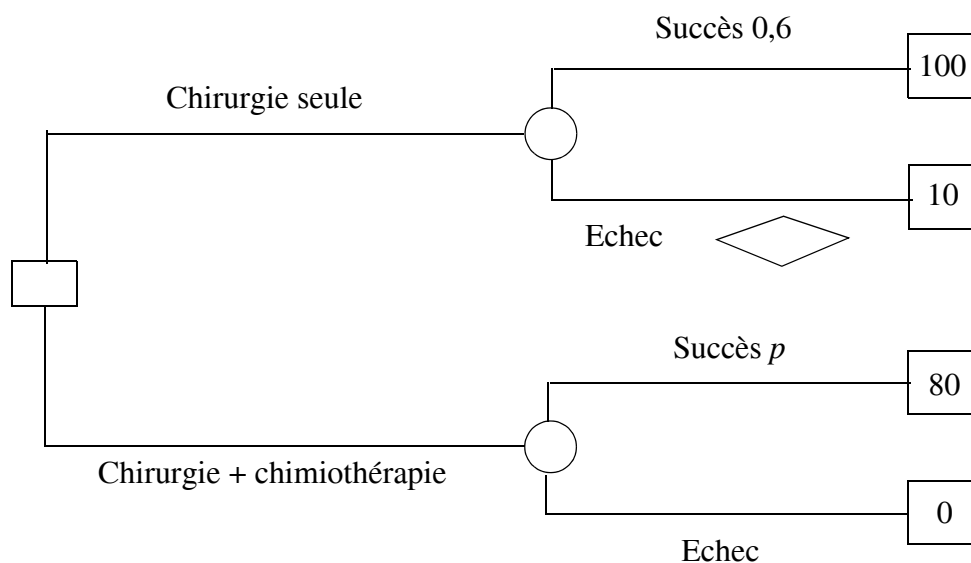
- A. $Se_1 + Se_2$
- B. $Se_1 \times Se_2$
- C. $Se_1 + Se_2 - Se_1 \times Se_2$
- D. $(1 - Se_1) \times (1 - Se_2)$
- E. On ne peut pas la calculer

3.3 Dans un contexte de dépistage, quel test choisiriez-vous parmi les quatre proposés ?

- A. T_1
- B. T_2
- C. T_3
- D. T_4
- E. On ne peut pas savoir

Partie II (2 points)

Dans le cadre du traitement d'un type de cancer, on a le choix entre deux modalités thérapeutiques : chirurgie seule ou chirurgie associée à une chimiothérapie. Chaque modalité peut conduire soit au succès, soit à un échec. L'ensemble du problème décisionnel est représenté par l'arbre ci-dessous :



3.4 Quelle est la valeur manquante dans le losange ?

- A. $1 - p$
- B. 90
- C. 0,5
- D. 0,4
- E. On ne peut pas la calculer

3.5 Le p de l'arbre indique :

- A. la proportion de succès après chirurgie et chimiothérapie
- B. la proportion de patients traités par chirurgie et chimiothérapie
- C. la proportion de succès
- D. la proportion de patients entrant dans l'étude
- E. le degré de signification du test

3.6 L'utilité moyenne de la chirurgie associée à la chimiothérapie est

- A. 80
- B. $80p$
- C. 50
- D. $1 - p$
- E. 0

3.7 Il sera préférable d'associer la chimiothérapie à la chirurgie

- A. en toute circonstance
- B. si p est supérieur à 0,5
- C. si p est inférieur à 0,6

- D. si p est supérieur à 0,8
- E. en aucune circonstance

Partie III (2 points)

On réalise une expérience pour étudier un nouveau test diagnostique. On constitue le groupe I en tirant au sort 100 individus dans la population des malades infectés et le groupe NI en tirant au sort 100 individus dans la population des sujets non infectés. Le test est positif chez 80 patients du groupe I et chez 10 sujets du groupe NI .

3.8 A partir de ces données, à quelle valeur pouvez-vous estimer la prévalence de l'infection ?

- A. 0,2
- B. 0,4
- C. 0,5
- D. 0,8
- E. On ne peut pas l'estimer

3.9 Quelle est la meilleure estimation ponctuelle de la sensibilité de ce nouveau test ?

- A. 0,2
- B. 0,4
- C. 0,45
- D. 0,8
- E. 0,89

3.10 En vous basant sur ces données, sur combien de personnes faudrait-il pratiquer le test pour estimer la spécificité par un intervalle de confiance de niveau 95 % de longueur 0,02 (la longueur est la différence entre les bornes supérieure et inférieure de l'intervalle). Choisir la valeur la plus proche.

- A. 59
- B. 79
- C. 1025
- D. 3457
- E. 6147

12

Concours 2002

Durée 45 minutes. Tout document autorisé.

La correction proposée en ligne est trop détaillée par rapport à ce qui est demandé pour le concours. En particulier, les réponses aux QCM ne doivent pas être justifiées.

Exercice 1 (7 points)

On pense qu'une certaine mutation génétique est liée à une augmentation de l'épaisseur du septum inter-ventriculaire. Afin d'étudier ce lien on procède à l'étude suivante. On génotype 90 sujets issus de la population générale et on mesure chez ces sujets l'épaisseur du septum. Parmi ces sujets, 40 sont trouvés porteurs de la mutation, 50 non porteurs.

Le tableau suivant indique des résumés des valeurs numériques obtenues :

Groupe	Moyenne expérimentale (mm)	Variance expérimentale (mm ²)
Porteurs	12,5	9,2
Non porteurs	11	9

1.1 (3,5 points)

Peut-on dire que les épaisseurs moyennes vraies du septum diffèrent entre les porteurs et les non porteurs ? Pour répondre à cette question on aura à mettre en œuvre un test d'hypothèses dont on détaillera avec soin les étapes, la conclusion, et, si nécessaire, le degré de signification.

1.2 (2 points)

On craint que l'épaisseur du septum, voire la nature du lien éventuel mutation-épaisseur du septum, ne dépende de l'âge. Aussi reprend-on l'analyse chez les sujets jeunes seuls (moins de 25 ans). Chez ces sujets jeunes, les valeurs obtenues étaient les suivantes.

Groupe	Effectifs	Moyenne expérimentale (mm)	Variance expérimentale (mm ²)
Porteurs jeunes	10	10	8,7
Non porteurs jeunes	13	8,5	8,9

Peut-on dire que chez les sujets jeunes les épaisseurs moyennes vraies du septum diffèrent entre les porteurs et les non porteurs ? On se contentera de :

- donner les conditions de validité
- donner l'expression du paramètre
- donner l'intervalle de pari
- formuler la conclusion

On donne la valeur calculée du paramètre (en valeur absolue) : 1,20

1.3 (0,9 points)

En observant les valeurs numériques et en particulier en remarquant que les différences des moyennes expérimentales sont les mêmes chez les sujets jeunes et chez tous les sujets, comment interprétez-vous globalement les résultats des questions 1 et 2 ?

1.4 (0,6 points)

Sous l'hypothèse nulle, quelle est la probabilité qu'avait la valeur absolue du paramètre d'excéder 1,2 chez les sujets jeunes ?

Exercice 2 (QCM - 6 points)

Dans cet exercice, les résultats sont donnés avec une précision de deux décimales.

Un patient suspect de présenter la maladie M se rend dans une consultation spécialisée. Il s'agit de décider s'il faut le traiter ou non. La décision optimale sera celle de plus grande utilité moyenne.

On dispose des informations suivantes :

- l'utilité d'être malade et traité est égale à 100,
- l'utilité d'être non malade et traité est égale à - 100,
- l'utilité d'être malade et non traité est égale à - 50,
- l'utilité d'être non malade et non traité est égale à 0.

Soit p la probabilité de la maladie M chez un patient donné.

2.1 L'utilité moyenne de traiter vaut :

- A. $200p - 100$

- B. $150p - 100$
- C. $50p - 50$
- D. $50p$
- E. $-50p$

2.2 L'utilité moyenne de ne pas traiter vaut :

- A. $200p - 100$
- B. $150p - 100$
- C. $50p - 50$
- D. $50p$
- E. $-50p$

2.3 En conséquence, il faut traiter dès que p est supérieure à :

- A. 0,33
- B. 0,40
- C. 0,50
- D. 0,60
- E. 0,67

2.4 Un examen complémentaire E1 a une sensibilité égale à 0,6 et une spécificité égale à 0,8 pour diagnostiquer la maladie M. On suppose que cette maladie a une prévalence égale à 0,1 dans la population des patients de cette consultation spécialisée.

Quelle est la valeur prédictive positive de l'examen E1 ?

- A. 0,08
- B. 0,20
- C. 0,25
- D. 0,43
- E. 0,75

2.5 Un deuxième examen E2 a la même sensibilité et la même spécificité que l'examen E1 pour la maladie M. On veut évaluer un test qui consiste à pratiquer les examens E1 et E2, pour un patient donné. On conclut à la positivité du test si les deux examens sont positifs, et à sa négativité si au moins un des deux examens est négatif. Les résultats des 2 examens sont supposés indépendants pour un même patient.

Pour le test ainsi défini, on a (plusieurs réponses sont possibles) :

- A. Sa sensibilité vaut 0,36
- B. Sa sensibilité vaut 0,64
- C. Sa spécificité vaut 0,64
- D. Sa spécificité vaut 0,96
- E. Sa valeur prédictive positive vaut 0,31
- F. Sa valeur prédictive négative vaut 0,93

2.6 En vous basant sur l'ensemble des informations fournies dans cet exercice (dans l'énoncé et vos réponses), indiquez, parmi les propositions suivantes, laquelle est (ou lesquelles sont) juste(s) :

- A. Il faut traiter la maladie M au moindre doute sans faire d'examen.
- B. Il faut traiter la maladie M dès qu'un examen est positif.
- C. Si le premier examen est négatif, le deuxième est inutile, car on ne traitera pas M quel que soit son résultat.
- D. Il faut toujours pratiquer deux examens, et ne traiter M que s'ils sont positifs tous les deux.
- E. Même si les deux examens sont positifs, il ne faut pas traiter la maladie M, car la valeur prédictive positive du test est trop faible.

Exercice 3 (QCM - 7 points)

Les trois parties peuvent être traitées indépendamment.

On considère un test d'hypothèses dont le paramètre z est, sous H_0 , distribué selon une loi normale centrée réduite.

On répète le test 5 fois sur la même population, de manière indépendante. Chaque test conduit à une valeur calculée du paramètre notée z_{ic} (i variant de 1 à 5). On se demande comment utiliser ces valeurs pour prendre une décision unique.

On choisit d'abord comme règle de décision de rejeter H_0 si au moins une des 5 valeurs z_{ic} est en dehors d'un intervalle de pari IP, à déterminer de manière à ce que le risque de première espèce α soit 0,05.

3.1

3.1.1 Cocher la ou les affirmations correctes, concernant le risque de première espèce α :

- A. α = probabilité de conclure H_1 alors que H_1 est vraie
- B. α = probabilité de conclure H_1 alors que H_0 est vraie
- C. α = probabilité de ne pas rejeter H_0 alors que H_1 est vraie
- D. α = probabilité d'obtenir au moins une des 5 valeurs du paramètre en dehors de IP alors que H_0 est vraie
- E. α = probabilité d'obtenir au moins une des 5 valeurs du paramètre en dehors de IP alors que H_1 est vraie

3.1.2 On suppose H_0 vraie, et on choisit comme intervalle de pari $IP_1 = [-1,96 ; 1,96]$
Quelle est la probabilité de rejeter H_0 , compte tenu de la règle de décision choisie :

- A. 0,05
- B. 0,95
- C. $5 \times 0,05 (= 0,25)$
- D. $0,05^5 (\approx 3,1 \times 10^{-7})$
- E. $0,95^5 (\approx 0,77)$
- F. $1 - 0,95^5 (\approx 0,23)$

3.2 Dans cette partie, on prend un intervalle de pari IP_2 tel que le risque de première espèce α soit 0,05 avec la règle de décision choisie.

3.2.1 On suppose H_0 vraie. On note a_2 la probabilité que le paramètre z n'appartienne pas à IP_2 . La probabilité d'obtenir **au moins une** des 5 valeurs z_{ic} en dehors de IP_2 est :

- A. a_2
- B. $5 a_2$
- C. a_2^5
- D. $1 - a_2^5$
- E. $(1 - a_2)^5$
- F. $1 - (1 - a_2)^5$

3.2.2 Cocher la ou les affirmations exactes lorsque $\alpha = 0,05$:

- A. $a_2 = 0,03$ (arrondi à deux décimales)
- B. $a_2 = 0,01$ (arrondi à deux décimales)
- C. $IP_2 = [-1,645 ; 1,645]$
- D. $IP_2 = [-1,514 ; 1,514]$
- E. $IP_2 = [-2,576 ; 2,576]$
- F. On ne doit pas calculer l'intervalle IP_2 , car les conditions de validité ne sont pas satisfaites ($5 < 30$)

3.2.3 Cocher la ou les affirmations correctes concernant le risque de deuxième espèce β , compte tenu de la règle de décision adoptée.

- A. β = probabilité de conclure H_1 alors que H_1 est vraie
- B. β = probabilité de conclure H_1 alors que H_0 est vraie
- C. β = probabilité de ne pas rejeter H_0 alors que H_1 est vraie
- D. β = probabilité d'obtenir au moins un des z_{ic} dans IP_2 alors que H_1 est vraie
- E. β = probabilité d'obtenir tous les z_{ic} dans IP_2 alors que H_1 est vraie
- F. β = probabilité d'obtenir au moins un des z_{ic} en dehors de IP_2 alors que H_1 est vraie

3.2.4 On désire calculer β dans le cas où, dans la réalité, le paramètre z est distribué selon une loi normale de moyenne 1 et de variance 1.

Dans ce cas, le paramètre z a une probabilité notée b_2 d'appartenir à IP_2 . On montre alors que $\beta = b_2^5$.

Cocher la ou les affirmations correctes :

- A. $b_2 = P(-1,645 \leq u \leq 1,645)$ avec $u \sim N(1,1)$
- B. $b_2 = P(-2,576 \leq u \leq 2,576)$ avec $u \sim N(1,1)$
- C. $b_2 = P(-2,645 \leq u \leq 0,645)$ avec $u \sim N(0,1)$
- D. $b_2 = P(-3,576 \leq u \leq 1,576)$ avec $u \sim N(0,1)$
- E. $\beta = 0,73$ (arrondi à deux décimales)
- F. $\beta = 0,22$ (arrondi à deux décimales)

3.3 Dans cette partie, on veut baser la règle de décision sur la variable \bar{z}_5 , moyenne expérimentale construite sur les 5 répétitions de z .

On suppose H_0 vraie (donc $z \sim N(0,1)$).

3.3.1 Parmi les affirmations suivantes concernant la loi de probabilité de \bar{z}_5 , cocher celles qui sont exactes :

- A. \bar{z}_5 suit une loi normale
- B. \bar{z}_5 suit une loi normale à condition de supposer certaines conditions supplémentaires
- C. Pour pouvoir continuer, on admettra que \bar{z}_5 suit une loi normale bien que les conditions de validité ne soient pas satisfaites
- D. $\bar{z}_5 \sim N(0,1)$
- E. $\bar{z}_5 \sim N(0,5)$
- F. $\bar{z}_5 \sim N(0,1/5)$

3.3.2 L'intervalle de pari IP_3 de niveau 0,95 est :

- A. $IP_3 = [-1,96 ; 1,96]$
- B. $IP_3 = [-2,576 ; 2,576]$
- C. $IP_3 = [-0,876 ; 0,876]$
- D. $IP_3 = [-0,392 ; 0,392]$
- E. $IP_3 = [-4,38 ; 4,38]$

13

Concours 2001

Durée 45 minutes. Tout document autorisé.

La correction proposée en ligne est trop détaillée par rapport à ce qui est demandé pour le concours. En particulier, les réponses aux QCM ne doivent pas être justifiées.

Exercice 1 (QCM - 6 points)

Dans cet exercice, vous devez cocher, pour chaque question, la ou les propositions ou les réponses qui sont exactes.

Dans une maladie M , on observe en général une élévation de la concentration d'une protéine P . On tire au sort, dans la population des patients atteints de M , un échantillon de 100 malades. Les moyenne et variance expérimentales de la concentration de P sont respectivement égales à 40 et 2500.

Soit μ_M la moyenne vraie de la concentration de P dans la population des malades atteints de M .

1.1

- A. L'intervalle de confiance de niveau 95 % de μ_M contient la valeur 32
- B. L'intervalle de confiance de niveau 60 % de μ_M est $40 \pm 0,524 \times 5$
- C. L'intervalle de confiance de niveau 99 % de μ_M est $40 \pm 2,576 \times 5$
- D. On ne peut pas calculer l'intervalle de confiance de niveau 99 % de μ_M , car une borne est inférieure à 30
- E. La probabilité que μ_M soit égale à 40 est 50 %

1.2 On tire au sort, dans la population des individus qui ne sont pas atteints de M , un échantillon de 50 personnes. On calcule les moyenne et variance expérimentales de la concentration de P , qui valent respectivement 20 et 2500.

Soit $\mu_{\bar{M}}$ la moyenne vraie de la concentration de P dans la population des individus non atteints de M .

On indique que :
$$\frac{40 - 20}{\sqrt{\frac{2500}{50} + \frac{2500}{100}}} = 2,3$$

- A. μ_M et $\mu_{\bar{M}}$ diffèrent au risque 5 %
- B. L'intervalle de confiance de $\mu_{\bar{M}}$ de niveau 95 % contient la valeur 32
- C. La probabilité pour que μ_M et $\mu_{\bar{M}}$ soient égales vaut 0,05
- D. Le degré de signification du test d'égalité des moyennes μ_M et $\mu_{\bar{M}}$ est inférieur à 0,01
- E. On ne peut pas rejeter l'hypothèse nulle d'égalité des moyennes μ_M et $\mu_{\bar{M}}$, peut-être par manque de puissance

1.3 On définit un seuil de concentration au-dessus duquel on considère qu'un individu est atteint de la maladie M . Si un individu a une concentration supérieure à ce seuil, on dit que l'examen est positif, et il est négatif dans le cas contraire.

On obtient alors les données suivantes : le nombre de vrais positifs est 50, celui des faux positifs est 5, le nombre de vrais négatifs est 45, et celui des faux négatifs est 50.

Quels sont les paramètres qu'on peut estimer à partir de ces données (**ne pas faire les calculs**) ?

- A. La sensibilité
- B. La spécificité
- C. La valeur prédictive positive
- D. La valeur prédictive négative
- E. La prévalence de M

1.4 Parmi les hypothèses nulles suivantes, lesquelles peut-on tester en utilisant les mêmes données (**ne faire ni calcul, ni test**) ?

- A. Les proportions d'examens positifs chez les patients atteints de M et chez les individus non atteints de M sont égales
- B. La sensibilité de l'examen est égale à 0,50
- C. La valeur prédictive positive de l'examen pour la maladie M est égale à 0,50
- D. La proportion d'individus atteints de M et ayant un examen positif dans la population générale est égale à 0,10
- E. Les proportions de malades et de non malades parmi les patients ayant un examen positif sont égales

1.5 On sait maintenant que la prévalence de la maladie M dans la population est égale à 0,30 ; la sensibilité est estimée à 0,5 ; la spécificité est estimée à 0,9.

La valeur prédictive positive de l'examen pour la maladie M peut être estimée par la valeur :

- A. 0,92
- B. 0,71
- C. 0,68
- D. 0,32
- E. 0,29

Exercice 2 (7 points)**Cet exercice comporte 4 questions principales indépendantes**

Après une certaine forme d'accident cardiaque, on considère qu'en l'absence de prise en charge, il y a 3 issues possibles : décès, complications donnant des séquelles irréversibles, état inchangé. Lorsqu'un patient a eu des complications, on considère qu'il ne peut plus revenir à l'état inchangé.

Au bout d'un temps $T_1 = 1$ après l'accident, on définit donc 3 événements possibles : D_1 = être décédé à T_1 (l'instant du décès est compris entre 0 et T_1) ; C_1 = être vivant, mais avec complications, à T_1 (les complications se sont produites entre 0 et T_1 , et il n'y a pas eu de décès ensuite) ; I_1 = être vivant dans l'état inchangé à T_1 .

Si on attend jusqu'au temps $T_2 = 2$ après l'accident, on définit 3 nouveaux événements : D_2 = être décédé à T_2 (l'instant du décès est entre 0 et T_2) ; C_2 = être vivant, mais avec complications, à T_2 ; I_2 = être dans l'état inchangé à T_2 .

2.1 (1,9 points)

Calcul des probabilités de décès à l'aide d'une fonction de répartition

On suppose que la densité de probabilité du décès à un instant t est donnée par $f(t) = \lambda e^{-\lambda t}$ avec $\lambda = 0,105$ et $t \geq 0$ ($t = 0$ correspond à l'instant de l'accident cardiaque ; pour $t < 0$, $f(t) = 0$).

2.1.1 Vérifier par dérivation que pour $t \geq 0$, la fonction de répartition est définie par $F(t) = 1 - e^{-\lambda t}$

2.1.2 En utilisant le fait que l'événement D_1 s'écrit aussi $[t \leq T_1]$, montrer que $P(D_1) = 0,1$. De même montrer que $P(D_2) = 0,19$.

2.1.3 Montrer que $P([T_1 < t \leq T_2] / [t > T_1])$ s'écrit $\frac{e^{-\lambda T_1} - e^{-\lambda T_2}}{e^{-\lambda T_1}}$
soit $1 - e^{-\lambda(T_2 - T_1)}$

2.1.4 En notant $\overline{D_1}$ l'événement « être vivant à T_1 », en déduire que $P(D_2 / \overline{D_1}) = 0,1$

2.2 (1,5 points)

Recalcul de $P(D_2) = 0,19$ sans utiliser la fonction de répartition

On suppose ne connaître que $P(D_1) = P(D_2 / \overline{D_1}) = 0,1$

2.2.1 Interpréter en une phrase l'écriture $P(D_2 / D_1)$. En déduire la valeur de cette probabilité.

2.2.2 En utilisant le théorème des probabilités totales, exprimer $P(D_2)$ en fonction des événements D_1 et $\overline{D_1}$.

2.2.3 En déduire la valeur de $P(D_2)$.

2.3 (1,8 points)

Calcul de $P(C_2) = 0,39$

Outre les probabilités données question 2, on suppose de plus que $P(D_2/\bar{D}_1) = P(D_2/I_1) = P(D_2/C_1)$

2.3.1 Interpréter en une phrase l'ensemble de ces égalités.

2.3.2 En déduire la valeur de $P(C_2 / C_1)$. On pourra s'aider des remarques suivantes :

- Une probabilité conditionnelle est une probabilité. En particulier pour 2 événements exclusifs A et B , on a $P((A \cup B) / C) = P(A / C) + P(B / C)$.
- Un patient qui est dans l'état C_1 à T_1 ne peut être que C_2 ou D_2 à T_2 .

2.3.3 On donne les probabilités suivantes :

$$P(C_1) = 0,3$$

$$P(C_2 / I_1) = 0,2$$

Calculer $P(C_2)$.

2.4 (1,8 points)

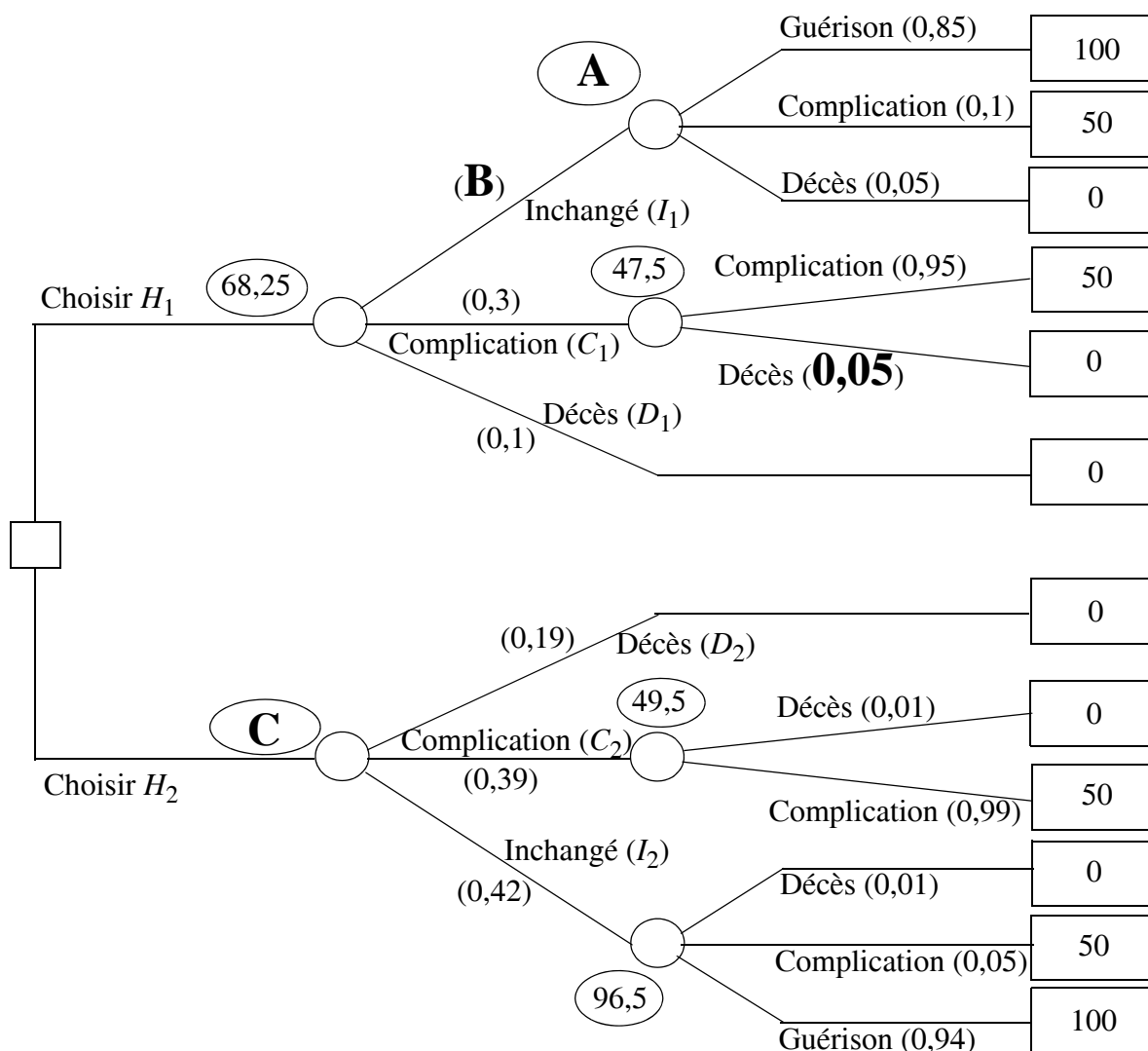
Choix entre 2 centres de soins

Après l'accident cardiaque, on a le choix entre 2 centres de soins H_1 et H_2 . Le premier permet une prise en charge à T_1 , le second à T_2 .

Après prise en charge, le patient peut être guéri (uniquement s'il était dans un état inchangé), avoir des complications (s'il en présentait déjà, ou si elles sont survenues après la prise en charge), ou être décédé (décès avant la prise en charge, ou après).

On choisit comme utilités 100 pour la guérison, 50 pour les complications, et 0 pour le décès.

L'ensemble des informations est reporté sur l'arbre de décision suivant



2.4.1 Interpréter la valeur 0,05 écrite en caractères gras dans l'arbre.

2.4.2 Calculer les 3 valeurs notées A, B et C sur l'arbre.

2.4.3 A quel centre de soins est-il préférable de s'adresser ?

Exercice 3 (7 points)

On pense que la concentration d'un facteur biologique F est liée à la vitesse d'élimination d'une substance S ; une trop faible vitesse d'élimination est génératrice de troubles.

- Le facteur F est étudié depuis longtemps, et on connaît la plage de ses valeurs normales. Cela permet de classer la valeur observée du facteur chez un sujet en

- « basse » (valeur inférieure à la valeur normale la plus basse) ;
- « normale » (dans la plage des valeurs normales) ;
- « élevée » (valeur supérieure à la valeur normale la plus élevée).
- Il existe un seuil de la vitesse d'élimination de S tel que
 - un sujet est atteint lorsqu'il est au dessous du seuil ;
 - le sujet est non atteint lorsqu'il est au dessus.

Pour mettre en évidence la liaison évoquée ci-dessus, on procède à une étude au cours de laquelle, chez 92 sujets issus de la population générale, on observe la concentration de F, qu'on classe en « basse », « normale », ou « élevée » et la vitesse d'élimination de S permettant de classer le sujet en « atteint » ou « non atteint ».

On peut alors dresser le tableau de contingence suivant :

Concentration du facteur F Etat	basse	normale	élevée
	atteint	9	6
non atteint	14	38	14

3.1 (0,6 point)

A quelle valeur estimez-vous la sensibilité du signe « concentration du facteur F élevée » ?

3.2 (4 points)

Concentration du facteur F (par classes) et vitesse d'élimination de S (par classes) sont-elles liées ? Le test effectué pour répondre à la question sera détaillé, incluant si nécessaire un degré de signification.

3.3 (1,5 points)

On vous fait remarquer qu'on pourrait traiter la même question en s'intéressant à la corrélation entre les variables quantitatives d'origine. Rappeler rapidement les hypothèses du test évoqué. Sa mise en œuvre donne une valeur calculée r_c égale à 0,12. Sans détailler la mise en œuvre du test, donner la conclusion en précisant, si nécessaire, le degré de signification.

3.4 (0,9 point)

Expliquer comment la réponse à la question 3 peut être cohérente avec la conclusion de la question 2.

14

Concours 2000

Durée 45 minutes. Tout document autorisé.

La correction proposée en ligne est trop détaillée par rapport à ce qui est demandé pour le concours. En particulier, les réponses aux QCM ne doivent pas être justifiées.

Exercice 1 (QCM - 6 points)

Dans une population de patients diabétiques, donc présentant une hyperglycémie, on tire au sort un échantillon de 50 malades. On évalue sur cet échantillon un nouveau test destiné à diagnostiquer une hyperglycémie. Sur l'ensemble de l'échantillon, le nombre de tests positifs est égal à 38. On réalise la même expérience sur une autre population de diabétiques, atteints d'un type de diabète différent, en tirant de nouveau au sort un échantillon de 50 patients, et on observe 32 tests positifs.

1.1 La sensibilité du test diagnostique est-elle égale dans les deux populations ?

- a. Non, car 0,76 est différent de 0,64.
- b. Oui.
- c. On ne peut répondre à la question, car on ne sait pas prouver une égalité stricte.
- d. Oui, car les sensibilités ne sont pas significativement différentes.
- e. Non, car les sensibilités sont différentes au risque $\alpha = 5\%$.

1.2 La sensibilité du test diagnostique est-elle différente dans les deux populations ?

- a. Oui, car 0,76 est différent de 0,64.
- b. Non, car après avoir réalisé le test statistique, on ne peut pas conclure à une différence significative.
- c. Oui, car les sensibilités diffèrent au risque $\alpha = 5\%$.
- d. Non, elles sont égales.
- e. Non, car la sensibilité d'un test diagnostique est toujours identique dans toutes les populations de malades.

1.3 Dans la suite de l'exercice, on suppose que la sensibilité du test diagnostique est égale dans les 2 populations de diabétiques.

Que vaut alors son meilleur estimateur ponctuel ?

- a. On ne peut pas le calculer
- b. 0,70
- c. 0,35
- d. 0,07
- e. 3,5

1.4 Cet estimateur est-il biaisé ?

- a. Oui
- b. Non
- c. On ne peut pas savoir

1.5 Que vaut exactement sa variance ?

- a. 0,21
- b. 0,046
- c. 0,0021
- d. 0,008256
- e. On ne peut pas la calculer exactement.

1.6 Peut-on estimer approximativement son erreur quadratique moyenne ?

- a. Oui, et elle vaut approximativement 0,21
- b. Oui, et elle vaut approximativement 0,046
- c. Oui, et elle vaut approximativement 0,0021
- d. Oui, et elle vaut approximativement 0,008256
- e. Non, on ne peut pas.

1.7 Que pensez-vous de la phrase suivante :

La probabilité pour que la sensibilité du test diagnostique soit comprise entre les valeurs 0,61 et 0,79 est égale à 0,95. (Nombres arrondis à 0,01 près).

- a. C'est toujours vrai
- b. C'est toujours faux
- c. C'est plausible, mais ça n'est pas certain
- d. On ne peut rien dire.

1.8 Que pensez-vous de la phrase suivante :

Si on tire au sort un individu dans une des deux populations de diabétiques, et qu'on réalise le test diagnostique, la probabilité pour que celui-ci soit positif est comprise entre 0,61 et 0,79. (Nombres arrondis à 0,01 près).

- a. C'est toujours vrai.
- b. C'est toujours faux.
- c. C'est plausible, mais ça n'est pas certain.
- d. On ne peut rien dire.

Exercice 2 (7 points)

Dans une certaine population, on considère une variable aléatoire X discrète (nombre d'accidents thrombo-emboliques) dont les valeurs possibles sont 1 ; 2 ; 3 et dont les probabilités correspondantes sont $p_1 = 0,3$; $p_2 = 0,5$; $p_3 = 0,2$.

2.1 (1,5 points)

Calculer l'espérance, la variance et l'écart-type de X

2.2 (3,2 points)

On considère une seconde variable Y (taux « normalisé » de glutathion), continue, liée à X de la manière suivante :

Si $X = 1$, $Y \sim N(1,1)$ (densité notée $f_1(y)$)

Si $X = 2$, $Y \sim N(0,1)$ (densité notée $f_2(y)$)

Si $X = 3$, $Y \sim N(-1,1)$ (densité notée $f_3(y)$)

On admettra que la densité de probabilité de y est définie par

$$f(y) = \sum_{i=1}^3 p_i f_i(y) = 0,3f_1(y) + 0,5f_2(y) + 0,2f_3(y)$$

2.2.1

2.2.1.1 Sans calcul, donner les valeurs de $\int_{-\infty}^{\infty} y f_1(y) dy$, $\int_{-\infty}^{\infty} y f_2(y) dy$, et $\int_{-\infty}^{\infty} y f_3(y) dy$

2.2.1.2 Montrer que l'espérance de Y est égale à 0,1 (on rappelle que l'intégrale d'une somme est la somme des intégrales).

2.2.2

2.2.2.1 En utilisant une des formules de calcul de la variance, déterminer les valeurs de

$$\int_{-\infty}^{\infty} y^2 f_1(y) dy, \int_{-\infty}^{\infty} y^2 f_2(y) dy, \text{ et } \int_{-\infty}^{\infty} y^2 f_3(y) dy$$

(N.B. aucun calcul d'intégrale n'est à faire)

2.2.2.2 Montrer que l'écart-type de Y est environ 1,22

2.2.3 On admettra que $E(XY) = -0,3$

Calculer la corrélation entre X et Y .

Interpréter la valeur obtenue.

2.3 (2,3 points)

2.3.1 Calculer les probabilités que $1 \leq Y \leq 2$ selon que $X = 1$, que $X = 2$, et que $X = 3$.

2.3.2 En déduire que la probabilité que $1 \leq Y \leq 2$, quelle que soit la valeur de X , est environ 0,175.

2.3.3 Si Y a une valeur comprise entre 1 et 2, trouver les probabilités que $X = 1$, que $X = 2$, et que $X = 3$

Exercice 3 (7 points)

Afin de faire la preuve de l'effet d'un médicament ancien sur le niveau d'une protéine circulante, on charge deux investigateurs de recruter chacun un groupe de 50 sujets.

L'un des investigateurs dosera la protéine chez les sujets de son groupe sans administrer le traitement, constituant ainsi le groupe NT de sujets.

L'autre investigateur administre le traitement aux 50 sujets de son groupe et dose la protéine sous traitement, après un délai suffisant, constituant le groupe T de sujets.

Les résultats obtenus sont les suivants :

	Groupe NT	Groupe T
Moyenne expérimentale	105	102
Variance expérimentale	106	102

3.1 (4 points)

Comparer les moyennes expérimentales dans les deux groupes. Pour cela il conviendra de mettre en œuvre un test d'hypothèses, dont on détaillera les étapes de mise en œuvre, que l'on interprétera et dont on calculera, le cas échéant, le degré de signification.

3.2 (0,8 point)

On s'aperçoit que le groupe de sujets NT est constitué de sujets apparemment plus jeunes (moyenne expérimentale des âges = 29 ans) que le groupe de sujets T (moyenne expérimentale des âges = 39 ans).

Or on sait que la concentration de cette protéine croît avec l'âge, en moyenne selon une progression de 0,2 unités par an. On décide de tenir compte de cet effet et de « ramener » tous les sujets à 35 ans. Ainsi, un sujet de 40 ans verra sa concentration de protéine diminuée de $0,2 \times 5 = 1$. La nouvelle valeur obtenue est appelée valeur ajustée.

Montrer que cela conduit à considérer pour nouvelles moyennes expérimentales les valeurs 106,2 et 101,2 respectivement pour les groupes NT et T .

3.3 (1 point)

Cet ajustement a également pour effet de réduire les écarts-types expérimentaux ; au total, lorsque l'on effectue le test de la question 1 avec ces nouvelles valeurs on obtient un paramètre z_c égal à 2,78. Que conclut-on alors ?

3.4 (0,7 point)

Sur un schéma sur lequel l'âge figure en abscisse, les concentrations en ordonnées, faites figurer les moyennes expérimentales des deux groupes NT et T avant et après ajustement. Utilisez ce schéma pour interpréter l'ensemble des résultats précédents.

3.5 (0,5 point)

Si vous aviez été consulté pour répondre à la question sur l'effet du traitement, quel schéma expérimental auriez-vous proposé ? Justifiez.

15

Concours 1999

Durée 45 minutes. Aucun document autorisé.

La correction proposée en ligne est trop détaillée par rapport à ce qui est demandé pour le concours. En particulier, les réponses aux QCM ne doivent pas être justifiées.

Exercice 1 (7 points)

Les trois parties de l'exercice peuvent être traitées séparément, à condition d'utiliser les informations données dans les parties précédentes.

Formules utiles :

- $P(F \cup G) = P(F) + P(G) - P(F \cap G)$
- Si $H = H_1 \cup H_2$ avec H_1 et H_2 disjoints, alors $P(F \cap H) = P(F \cap H_1) + P(F \cap H_2)$
- $P(\bar{F}/K) = 1 - P(F/K)$

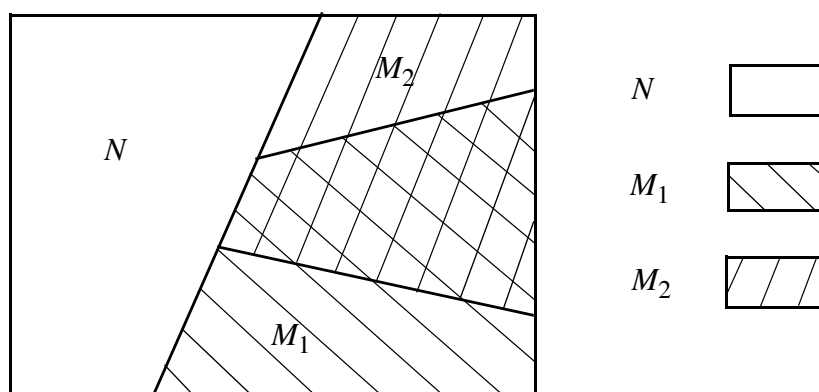
On considère 3 événements M_1 , M_2 et N , de probabilités respectives $P(M_1)$, $P(M_2)$ et $P(N)$, construits sur un ensemble fondamental E . Les événements M_1 et M_2 sont indépendants (donc M_1 et \bar{M}_2 aussi) ; N est incompatible avec M_1 et avec M_2 (donc avec $M_1 \cup M_2$).

1.1 (2 points)

1.1.1 Exprimer les probabilités des événements suivants en fonction de $P(M_1)$, $P(M_2)$ et $P(N)$:

- $A = M_1 \cap M_2$ (M_1 et M_2 en même temps)
- $M_1 \cap N$
- $B = M_1 \cap \bar{M}_2$ (M_1 et pas M_2)
- $N \cup (M_1 \cup M_2)$ (M_1 ou M_2 ou N)

1.1.2 Dans la suite de l'exercice, M_1 et M_2 correspondent à des états dont les probabilités sont $P(M_1) = 0,3$ et $P(M_2) = 0,2$. L'événement N correspond au fait d'être **ni dans M_1 , ni dans M_2** (donc : $N \cup M_1 \cup M_2 = E$). Ces événements sont représentés sur le schéma suivant où le rectangle délimite E .



Montrer sur un schéma analogue, faisant apparaître N , M_2 et B , que ces trois événements forment une partition de E .

1.1.3 Montrer que $P(N) = 0,56$.

1.2 (2,5 points)

Selon leur état, les individus présentent un signe S avec les probabilités suivantes :

- 0,20 pour ceux qui sont dans l'état M_1 ;
- 0,60 pour ceux qui sont dans l'état M_2 ;
- 0,11 pour ceux qui sont dans l'état N ;
- 0,83 pour ceux qui sont dans l'état A .

1.2.1 Quelle est la probabilité de présenter à la fois le signe S et l'état M_1 ?

Quelle est la probabilité de présenter à la fois le signe S et l'état A ?

1.2.2 En remarquant que A et B sont disjoints et que leur réunion est M_1 , réécrire la probabilité de présenter à la fois S et M_1 en fonction de la probabilité de présenter à la fois S et A , et de la probabilité de présenter à la fois S et B . En déduire que la probabilité de présenter à la fois S et B est environ 0,01.

1.2.3 En utilisant que B , M_2 et N forment une partition de E , montrer que la probabilité de présence de S est environ 0,19.

1.3 (2,5 points)

On désire utiliser la présence ou l'absence du signe S pour diagnostiquer l'état M_2 .

1.3.1 Calculer la sensibilité de S pour M_2 .

1.3.2 Calculer la valeur prédictive positive de S pour M_2 .

1.3.3 En utilisant que $\overline{M_2} = N \cup B$, N et B étant disjoints, calculer la spécificité de S pour M_2 .

Exercice 2 (8 points)

Afin d'étudier le lien possible entre l'exposition professionnelle à un agent chimique gazeux et un déficit neurologique léger, on a mesuré le degré d'exposition à cet agent dans un groupe de 35 travailleurs présentant ce déficit (sujets dits atteints dans la suite). La même chose a été accomplie auprès de 42 travailleurs ne présentant pas le déficit (sujets dits non atteints dans la suite). Le degré d'exposition est exprimé en mg/m^3 .

On a obtenu les résultats suivants :

	nombre	moyennes expérimentales du degré d'exposition	écarts-type expérimentaux du degré d'exposition
sujets atteints	35	5,3	2,1
sujets non atteints	42	4,0	2,2

2.1 (0,5 point)

En quelle unité les écarts-type sont-ils exprimés ?

2.2 (4,5 points)

Sujets atteints et non atteints sont-ils exposés à l'agent chimique au même degré moyen ? Pour répondre à la question on effectuera un test dont on énoncera précisément les hypothèses et les étapes de mise en œuvre, dont on donnera clairement la conclusion, et pour lequel on calculera le cas échéant le degré de signification.

2.3 (0,5 point)

Cette étude permet-elle de conclure que l'exposition à l'agent chimique cause le déficit ? Justifier brièvement votre réponse.

2.4 (2,5 points)

Avant de prendre connaissance des degrés d'exposition chez ces 35 sujets atteints et 42 sujets non atteints, on a entrepris des calculs non explicités ici, au terme desquels ont été obtenues trois densités de probabilité du paramètre du test de la question 2, sous trois hypothèses différentes.

Le graphique ci-dessous représente ces densités de probabilité ainsi que l'intervalle de pari construit dans le test de la question 2 ; précisément :

la courbe 1 correspond à l'hypothèse

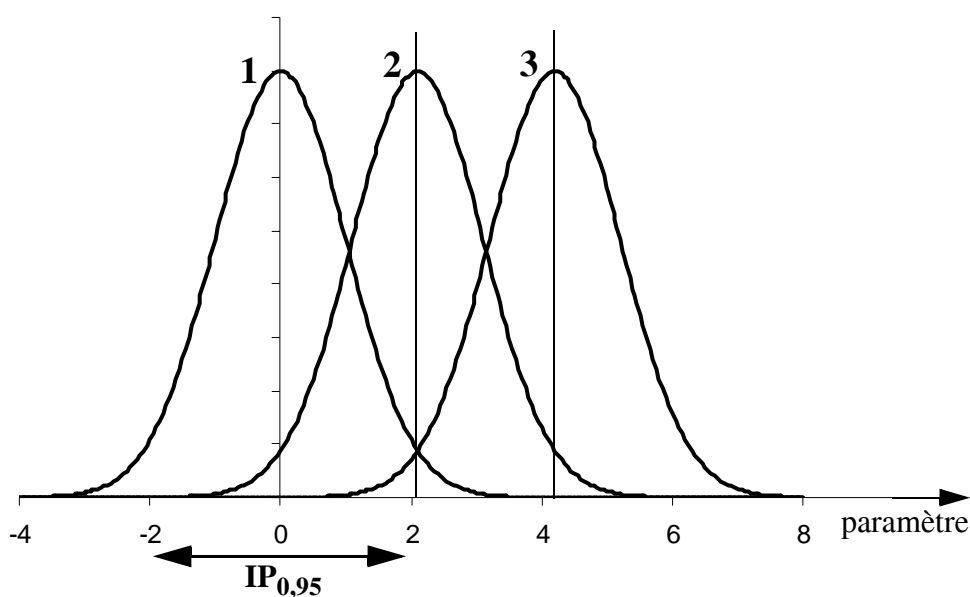
« la différence des moyennes vraies des degrés d'exposition est nulle » ;

la courbe 2 correspond à l'hypothèse

« la différence des moyennes vraies des degrés d'exposition est égale à 1 » ;

la courbe 3 correspond à l'hypothèse

« la différence des moyennes vraies des degrés d'exposition est égale à 2 ».



Sans faire de calcul et à partir du graphique ci-joint, répondre par **des valeurs approchées** aux questions suivantes. Les réponses seront justifiées et choisies parmi les valeurs suivantes : 0,01 ; 0,05 ; 0,25 ; 0,45 ; 0,55 ; 0,75 ; 0,95 ; 0,99.

- si la différence des moyennes vraies est égale à 1, quelle est la puissance du test effectué question 2 ?
- si la différence des moyennes vraies est égale à 2, quel est le risque de seconde espèce du test effectué question 2 ?

Exercice 3 (QCM - 5 points)

Pour toutes les questions de l'exercice, une et une seule réponse est juste.

On rappelle que si X et Y sont deux variables aléatoires indépendantes
 $var(aX + bY) = a^2 var(X) + b^2 var(Y)$.

On cherche à estimer la moyenne d'un dosage biologique dans une population.

3.1 A partir d'un échantillon A de 100 malades tirés au sort dans la population concernée, on obtient les données expérimentales suivantes : la moyenne expérimentale, notée m_A , est égale à 100, et la variance expérimentale est égale à 100.

Dans cette question, les réponses quantitatives sont arrondies à 0,1 près.

3.1.1 Déterminer un intervalle de confiance de niveau 0,95 de la moyenne du dosage dans la population.

- A. [80,4 119,6]

- B. [93,8 106,2]
- C. [98,0 102,0]
- D. [98,4 101,6]
- E. [98,8 99,7]

3.1.2 Quel est l'intervalle de confiance de niveau 0,999 de cette moyenne ?

- A. [67,1 132,9]
- B. [89,6 110,4]
- C. [96,1 103,9]
- D. [96,7 103,3]
- E. [98,4 101,6]

3.2 Un autre échantillon B du même dosage, indépendant de A , donne les résultats expérimentaux suivants : effectif 50 ; moyenne expérimentale, notée m_B , égale à 97 ; variance expérimentale 100. On pense que les deux échantillons proviennent de la même population.

Pour vérifier que les données ne contredisent pas cette hypothèse :

3.2.1 Suffit-il de vérifier que les deux intervalles de confiance de niveau 0,95 se recoupent ?

- A. oui
- B. non

3.2.2 Est-il nécessaire de faire un test ? si oui lequel ?

- A. non
- B. oui, le test des séries appariées
- C. oui, le test de comparaison de 2 moyennes observées
- D. oui, le test du chi carré
- E. oui, le test du coefficient de corrélation

3.2.3 Peut-on prouver l'égalité des moyennes dans les populations dont sont issus les échantillons A et B ?

- A. oui
- B. non

3.3 On admet dans la suite que les deux échantillons proviennent de la même population et que la variance vraie du dosage biologique est égale à 100. Les réponses quantitatives sont maintenant arrondies à 0,01 près.

On constitue un seul échantillon en regroupant les échantillons A et B .

3.3.1 Quelle est la valeur de la moyenne expérimentale du dosage, estimateur de la moyenne du dosage dans la population, noté e_1 ?

- A. 97,00

- B. 98,00
- C. 98,50
- D. 98,80
- E. 99,00

3.3.2 On peut remarquer, à partir de la réponse précédente, que e_1 est de la forme $e_1 = c m_A + (1 - c) m_B$. Quelle est la valeur de c ?

- A. 0,67
- B. 0,60
- C. 0,50
- D. 0,40
- E. 0,33

3.3.3 En admettant que **les moyennes expérimentales** m_A et m_B sont des variables aléatoires indépendantes, et en utilisant l'expression de e_1 donnée ci-dessus, déterminez la valeur de la variance de e_1 .

- A. 0,67
- B. 0,75
- C. 3,00
- D. 56,00
- E. 100,00

3.4 On étudie un autre estimateur de la moyenne du dosage. Ce nouvel estimateur, noté e_2 ,

est égal à $\left(\frac{1}{2}m_A + \frac{1}{2}m_B\right)$

3.4.1 Cet estimateur est-il sans biais ?

- A. oui
- B. non

3.4.2 Quelle est sa variance ?

- A. 0,50
- B. 0,67
- C. 0,75
- D. 3,00
- E. 100,00

3.4.3 Entre m_A , m_B , e_1 et e_2 , quel est le meilleur estimateur de la moyenne du dosage dans la population ?

- A. Celui qui a la plus grande valeur
- B. Celui qui a la plus petite valeur
- C. Celui qui a la plus grande variance
- D. Celui qui a la plus petite variance

16

Concours 1998

Durée 45 minutes. Aucun document autorisé.

La correction proposée en ligne est trop détaillée par rapport à ce qui est demandé pour le concours. En particulier, les réponses aux QCM ne doivent pas être justifiées.

Exercice 1 (7,5 points)

Dans le cadre de l'étude des propriétés d'un nouveau médicament, on examine les relations pouvant exister entre le délai d'apparition de l'effet de ce médicament et la concentration d'une protéine, mesurable chez les individus, dont on a de bonnes raisons de penser qu'elle interfère avec ce délai d'apparition.

Pour cela on constitue un échantillon de 102 patients ; chez chacun d'eux on mesure la concentration de la protéine (variable X), on administre le médicament et on mesure le délai d'apparition - en minutes - de son effet (variable D).

1.1 (2,2 points)

On s'intéresse d'abord à la corrélation entre les variables délai d'apparition de l'effet et concentration de la protéine. Les observations effectuées ont permis de calculer les valeurs suivantes :

- pour les délais : moyenne expérimentale $\bar{d} = 85,2$
 écart type expérimental $s_d = 41,8$
- pour les concentrations : moyenne expérimentale $\bar{x} = 20,1$
 écart type expérimental $s_x = 5,9$
- pour les produits : $\frac{1}{102} \sum d_i x_i = 1730,5$

Le coefficient de corrélation vrai entre ces variables est-il non nul ? On aura à effectuer un test d'hypothèses dont on détaillera les étapes de mise en œuvre.

1.2 (4,1 points)

Surpris par ce résultat, on décide de construire, à partir des variables étudiées, deux autres variables :

- La variable « importance du délai » qui prend deux valeurs :
 - « standard » si le délai d'apparition est inférieur à 82 heures ;
 - « élevée » si le délai d'apparition atteint ou excède cette valeur.

- La variable « niveau de concentration » qui prend trois valeurs :
 - « bas » si la concentration de la protéine est inférieure ou égale à 15 ;
 - « standard » si la concentration est comprise entre 15 et 25 ;
 - « élevé » si la concentration est supérieure ou égale à 25.

1.2.1 Quel est le type des variables nouvellement introduites ?

1.2.2 On désire savoir, à l'aide des seules données de la question 1, quelle est la proportion attendue de patients pour lesquels l'importance du délai est « standard ».

- Quelle estimation ponctuelle proposez-vous pour la moyenne vraie et l'écart type vrai du délai D ? Justifiez
- Si le délai est distribué selon une loi normale, quelle est cette proportion attendue ?

1.2.3 Les observations précédentes permettent de dresser le tableau de contingence suivant :

		Niveau de concentration		
		Bas	Standard	Elevé
Importance du délai	Standard	6	32	8
	Elevée	17	21	18

Le niveau de concentration et l'importance du délai sont-ils liés ? On effectuera un test d'hypothèses dont on indiquera les étapes de mise en œuvre.

1.2.4 Proposez une explication à la divergence des conclusions des deux tests effectués jusqu'ici.

1.3 (1,2 points)

Des considérations liées aux réactions enzymatiques conduisent à s'intéresser à une nouvelle variable X_2 définie par $X_2 = (X - 20)^2$.

Le coefficient de corrélation expérimental entre X_2 et le délai d'apparition D est égal à 0,72.

1.3.1 Peut-on dire que les variables X_2 et délai d'apparition sont corrélées ? *On se bornera à donner la conclusion du test, en la justifiant, mais sans détailler les étapes.*

1.3.2 Utilisez ces informations pour interpréter l'ensemble des résultats ; on pourra par exemple indiquer sur un dessin où figurerait la concentration en abscisse, le délai en ordonnée, l'allure des données expérimentales recueillies.

Exercice 2 (7,5 points)

Une source d'électrons permet de créer un point lumineux sur un écran. La position du point sur l'axe horizontal est une variable aléatoire X dont la densité de probabilité f_1 est :

$$f_1(x) = c(x^2 - 1) \quad \text{pour } -1 \leq x \leq 1$$

$$f_1(x) = 0 \quad \text{pour } x \leq -1 \text{ et pour } x \geq 1$$

2.1 (3,6 points)**2.1.1** Vérifier que $c = -3/4$ **2.1.2** Calculer l'espérance et la variance de X **2.1.3** Vérifier que la fonction de répartition de X est :

$$F_1(x) = \frac{1}{4}(x+1)^2(2-x) \quad \text{pour } -1 \leq x \leq 1$$

$$F_1(x) = 0 \quad \text{pour } x \leq -1$$

$$F_1(x) = 1 \quad \text{pour } x \geq 1$$

2.2 (2,5 points)

Le faisceau d'électrons peut en réalité emprunter deux trajets distincts. On hésite donc entre deux lois de probabilité pour la variable X . La première est la loi précédente et correspond au trajet 1. La seconde loi (trajet 2) est définie par la fonction de répartition F_2 :

$$F_2(x) = \frac{1}{32}(x+1)^2(5-x) \quad \text{pour } -1 \leq x \leq 3$$

$$F_2(x) = 0 \quad \text{pour } x \leq -1$$

$$F_2(x) = 1 \quad \text{pour } x \geq 3$$

Pour identifier le trajet suivi par le faisceau, on met en œuvre la démarche suivante, apparentée à un test d'hypothèses :

Hypothèses

H_0 : la fonction de répartition de X est $F_1(x)$ (le faisceau a suivi le trajet 1),

H_1 : la fonction de répartition de X est $F_2(x)$ (le faisceau a suivi le trajet 2).

Intervalle de pari pour X

Intervalle de la forme $[-1 ; a]$ où a est un nombre compris entre -1 et 1 .

Règle de décision

On observe une réalisation x de X et on adopte la règle suivante : si $x \in [-1 ; a]$, on choisit H_0 , sinon on choisit H_1 .

2.2.1 On prend $a = 0,73$. Calculer le risque de première α et celui de seconde espèce β .**2.2.2** Mêmes questions si $a = -0,46$.**2.2.3** En quoi la démarche proposée se distingue-t-elle des tests d'hypothèses habituels ?**2.2.4** On choisit la valeur de a (0,73 ou -0,46) pour laquelle le risque total $\alpha + \beta$ est minimal. Justifiez ce choix.**2.3 (1,4 points)**

On sait maintenant que le faisceau d'électrons emprunte le premier trajet (correspondant à la fonction de répartition F_1) avec une probabilité de 0,3 et le second trajet (fonction F_2) avec une probabilité de 0,7.

2.3.1 Si on observe un point lumineux dans l'intervalle $[-1 ; 0,73]$, quelle est la probabilité que le faisceau ait suivi le second trajet ?

Si le point est en dehors de cet intervalle, quelle est la probabilité que le faisceau ait suivi le premier trajet ?

2.3.2 Mêmes questions pour l'intervalle $[-1 ; -0,46]$.

2.3.3 En assimilant le fait que x soit dans l'intervalle $[-1 ; 0,73]$ à « présence d'un signe » et le fait que le faisceau ait suivi le premier trajet à « diagnostic vrai », calculer les sensibilité, spécificité et valeurs prédictives.

Que vous suggère la comparaison entre les valeurs prédictives ?

Exercice 3 (5 points)

Pour comparer 2 traitements A et B d'une même maladie, on constitue par tirage au sort 2 groupes de malades, l'un recevant A , l'autre B .

Le groupe A est constitué de 100 patients et le pourcentage observé de guérison est $p_{oA} = 0,5$. Le pourcentage observé de guérison dans le groupe B est p_{oB} .

Pour chacune des affirmations suivantes, indiquez **sur votre copie** le numéro de l'affirmation et si elle est :

- Vraie (c'est à dire *toujours* vraie)
- Fausse (c'est à dire *toujours* fausse)
- Autre (c'est à dire ni toujours vraie ni toujours fausse)

3.1

3.1.1 On déduit des observations que la probabilité de guérison avec le traitement A est de 0,5.

3.1.2 La probabilité de guérison avec le traitement A est dans l'intervalle $[0,40 ; 0,60]$ (valeurs arrondies).

3.1.3 L'intervalle de confiance de niveau 95 % de la probabilité de guérison avec le traitement A est $[0,40 ; 0,60]$ (valeurs arrondies).

3.1.4 L'intervalle de pari de la proportion observée de guérison avec le traitement A est $[0,40 ; 0,60]$ (valeurs arrondies).

3.2 Si le test de comparaison entre p_{oA} et p_{oB} ne montre pas de différence significative :

3.2.1 Cela implique nécessairement que les traitements ont les mêmes effets.

3.2.2 Si en fait les proportions vraies de guérison diffèrent, le résultat du test est dû à un manque de puissance.

3.3 Si le test de comparaison entre p_{oA} et p_{oB} montre une différence significative avec un degré de signification p :

3.3.1 p est inférieur à 0,05.

3.3.2 p est inférieur à 0,01.

3.3.3 Ce degré de signification veut dire que la probabilité pour que les deux traitements donnent la même proportion vraie de guérison est inférieure à p .

3.3.4 Ce degré de signification veut dire que si les deux traitements donnent la même proportion vraie de guérison, la probabilité que l'on avait d'obtenir une différence au moins égale à celle effectivement observée est inférieure à p .