CONCOURS OUVERTS LES 29 ET 30 SEPTEMBRE ET LES 1er ET 02 OCTOBRE 2020 POUR L'ADMISSION AU CYCLE DE FORMATION DES ELEVES DIRECTEURS D'HÔPITAL

CONCOURS EXTERNE – INTERNE ET TROISIÈME CONCOURS

VENDREDI 02 OCTOBRE 2020

4ème Épreuve écrite d'admissibilité

Durée : 4 heures – Coefficient : 3

STATISTIQUES

SUJET : pages 1 à 4

Examen de Statistique 2020

Le barème est donné à titre indicatif.
Le candidat est invité à lire le sujet dans son intégralité.

Notations et Quantiles

- soit X une variable aléatoire : $\mathbb{E}(X)$ et $\mathbb{V}(X)$ désignent respectivement l'espérance et la variance de X
- le quantile d'ordre 0.975 d'une loi de Student à 24 degrés de liberté est égal à 2.06
- le quantile d'ordre 0.95 d'une loi du χ^2 à 2 degrés de liberté est égal à 5.99

Exercice 1 Probabilités conditionnelles (1 pt)

Dans une étude de santé publique sur une cohorte comprenant 3/4 de femmes, on s'intéresse à la présence d'anticorps dans des prélèvements sanguins. La moitié des femmes de l'étude ont les anticorps et c'est le cas de 2/3 des hommes. On tire au hasard un individu dans la cohorte, il a les anticorps, quelle est la probabilité qu'il s'agisse d'un homme ?

Exercice 2 Lois discrètes (1 pt)

Nous considérons les enfants de parents hétérozygotes de génotype Aa. La distribution des enfants est

$$\mathbb{P}(AA) = 1/4 \quad \mathbb{P}(Aa) = 1/2 \quad \mathbb{P}(aa) = 1/4 .$$

Nous choisissons aléatoirement 240 de ces enfants. Nous définissons N_1, N_2, N_3 les nombres d'enfants de génotype AA, Aa et aa respectivement.

1 (0.5 pt) Quel est le loi de N_1 ? Calculer son espérance et sa variance.

2 (0.5 pt) Quel est le lien entre les variables N_1, N_2 et N_3 ? Vérifier que

$$\mathbb{P}(N_1 = k_1, N_2 = k_2, N_3 = k_3) \neq \mathbb{P}(N_1 = k_1)\mathbb{P}(N_2 = k_2)\mathbb{P}(N_3 = k_3) .$$

Exercice 3 Lois continues (2 pt)

Soit X une variable aléatoire de densité

$$f_X(x) = \begin{cases} x \exp\left(-\frac{x^2}{2}\right) & \text{si} \quad x > 0 \\ 0 & \text{sinon} \end{cases} .$$

1 (1 pt) Vérifier que $f_X(x)$ est bien une densité de probabilité.

2 (1 pt) Donner la loi et l'espérance de $Y = X^2$.

Exercice 4 Couple de variables aléatoires (3 pts)

Nous considérons un couple (X, Y) de variables aléatoires continues réelles de densité

$$f_{(X,Y)}(x,y) = \begin{cases} k\left(\frac{1}{x^2} + y^2\right) & \text{si} \quad 1 < x < 5, \quad -1 < y < 1 \\ 0 & \text{sinon} \end{cases}.$$

1 (0.5 pt) Pour quelle valeur de k la fonction f est-elle une densité ?

2 (1 pt) Calculer les densités marginales de X et de Y.

3 (0.5 pt) Les variables aléatoires X et Y sont-elles indépendantes ?

4 (1 pt) Déterminer la loi conditionnelle de X sachant que $Y = 0$.

Exercice 5 Estimation (7 pts)

Soit un n-échantillon $(X_1, \dots, X_n)$ de X admettant la densité

$$f_X(x;\theta) = \begin{cases} \frac{3}{(x-\theta)^4} & \text{si} \quad x > \theta + 1 \\ 0 & \text{sinon} \end{cases}$$

où $\theta > 0$ est un paramètre inconnu.

1 (1 pt) Calculer $\mathbb{E}(X)$ et $\mathbb{V}(X)$ (on pourra calculer $\mathbb{E}((X-\theta))$ et $\mathbb{V}((X-\theta))$).

2 (2 pts) Donner l'estimateur $\tilde{\theta}_n$ de θ par la méthode des moments basée sur $\mathbb{E}(X)$. Calculer $\mathbb{E}\left(\tilde{\theta}_n\right)$ et $\mathbb{V}\left(\tilde{\theta}_n\right)$.

3 (2 pts) Donner l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ de θ.

4 (2 pts) Calculer $\mathbb{E}\left(\hat{\theta}_n\right)$. En déduire un estimateur sans biais $\theta_n^{\#}$ de θ fonction de $\hat{\theta}_n$.

Exercice 6 Intervalle de confiance (1 pt)

Nous considérons un 25-échantillon $X_1, \ldots, X_{25}$ de $X \sim \mathcal{N}(\theta, \sigma^2)$ où les paramètres $\theta \in \mathbb{R}$ et $\sigma^2 > 0$ sont inconnus. Nous observons

$$\bar{x}_{25} = \frac{1}{25}\sum_{i=1}^{25} x_i = 0.1 \quad \text{et} \quad s_{25}^2 = \frac{1}{24}\sum_{i=1}^{25}(x_i - \bar{x}_{25})^2 = 4\,.$$

Donner un intervalle de confiance bilatéral symétrique de niveau (de confiance) 95% pour θ. Nous noterons la réalisation de cet intervalle $\mathrm{IC}(\theta)$.

Exercice 7 Test (2 pts)

Nous désirons savoir s'il existe, dans une population d'individus atteints du cancer de la peau, un lien entre l'âge de l'individu et ses chances de guérison. Nous menons une enquête sur 3865 individus répartis dans trois classes d'âge et obtenons les résultats ci-dessous

âge /statut	guéris	non guéris
50-60	1601	502
60-70	753	246
70-80	571	192

Étudier le lien entre les variables âge et statut, on pourra utiliser un test d'hypothèse associé à un risque de première espèce asymptotiquement égal à $\alpha = 0.05$.

Exercice 8 Régression linéaire (3 pts)

Nous disposons d'un échantillon de 30 sujets sur lesquels une étude est menée pour déterminer si la valeur de la Pression Artérielle Systolique (PAS) dépendait de l'âge.

Dans cet exercice, l'évaluation se fait sous forme d'un Questionnaire à Choix Multiples : à chaque question ci-dessous correspond une ou plusieurs réponses possibles dans les propositions données. Une réponse erronée n'engendre aucun retrait de point.

QCM1 (0.5 pt) Pour déterminer s'il existe une liaison entre l'âge et la PAS, il est possible d'utiliser

A) un test de comparaison de 2 moyennes observées sur 2 échantillons appariés
B) un test du Khi-Deux
C) un test du coefficient de corrélation
D) un test de comparaison de 2 moyennes observées sur 2 échantillons indépendants
E) les propositions A, B, C, D sont fausses

QCM2 (0.5 pt) Les conditions d'application à vérifier avant d'estimer la pente et l'ordonnée à l'origine de la droite de régression linéaire de la PAS en fonction de l'âge sont

A) un degré de signification inférieur à 5%
B) l'indépendance des observations
C) la liaison linéaire entre la pression artérielle systolique et l'âge
D) les effectifs théoriques attendus sous l'hypothèse nulle H_0 sont tous supérieurs ou égaux à 5
E) les propositions A, B, C, D sont fausses

QCM3 (0.5 pt) Dans la droite de régression de la PAS en fonction de l'âge, dont l'équation est

$$\text{PAS} = \alpha + \beta \text{age}$$

A) l'âge est la variable dépendante
B) l'âge est la variable explicative
C) la PAS est la variable indépendante
D) la PAS est la variable dépendante
E) les propositions A, B, C, D sont fausses

QCM4 (0.5 pt) Le degré de signification (P-value) associé au test de nullité du coefficient de la pente de la droite de régression est inférieur à 0.001. Comment interpréter cette information ?

A) la pente de la droite de régression est égale à 0
B) la PAS moyenne diffère significativement de l'âge moyen
C) la pente de la droite de régression diffère significativement de 0
D) la pente de la droite de régression est significativement inférieure à 0.001
E) les propositions A, B, C, D sont fausses

QCM5 (1 pt) Sur les 30 données disponibles, l'estimation du coefficient de la pente de la droite de regression est égal à 1, la valeur moyenne de la PAS est égale à 145 millimètres de mercure et l'âge moyen est de 45 ans. L'estimation du coefficient de l'ordonnée à l'origine de la droite de régression est égale à

A) 45/145
B) 145/45
C) 45
D) 100
E) les propositions A, B, C, D sont fausses