Méthodes Statistiques pour l'Ingénieur (MSI) : examen final 2016
Durée : 1h30

Barème :

- Exercice 1 : 5 points
- Exercice 2 : 6 points
- Exercice 3 : 5 points
- Exercice 4 : 4 points

Exercice 1 (5 points)

Les réponses sont à renseigner sur la feuille jointe au sujet et à remettre dans la copie.

Plusieurs *réponses ou affirmations peuvent être vraies pour chaque question. Une question est dite juste si et seulement si toutes les affirmations vraies sont cochées et aucune affirmation fausse n'est cochée.*

- *1 question juste : +0.5 point*
- *1 question fausse : -0.5 point*
- *1 question sans réponse : 0 point*

Exemples :

- *8 questions justes et 2 sans réponse = 4 points*
- *10 questions justes = 5 points*
- *5 questions justes, 5 questions fausses = 0 point*

Question 1 :

Un intervalle de confiance au seuil (ou risque) de 1% est :

A. Plus petit que le même intervalle de confiance au niveau de risque 5%.
B. Plus grand que le même intervalle de confiance au niveau de risque 5%.
C. Aucune de ces affirmations n'est vraie.

Question 2 :

Un enquêteur se place à l'accueil d'une station de métro et observe les arrivées. En fonction de son allure (sportive ou en costard-cravate, énervée ou calme, etc.) l'enquêteur interroge ou non la personne. Dès qu'il obtient 150 réponses, il s'arrête et donne son estimation.

A. La méthode d'échantillonnage est une méthode aléatoire simple.
B. La méthode d'échantillonnage est une méthode d'échantillonnage stratifié.
C. La méthode d'échantillonnage est une méthode aléatoire systématique.
D. La méthode d'échantillonnage est une méthode au jugé.
E. Aucune de ces affirmations n'est vraie.

Question 3 :

Dans une grande entreprise, on souhaite connaître la proportion des employés qui ont un abonnement aux transports en commun de l'agglomération. Pour chaque individu de la base de sondage, on dispose de la valeur de son revenu. On part de l'idée *a priori* que les revenus ont peut-être un lien avec le fait d'être abonné ou non aux transports en commun. On décide alors de constituer 3 strates dans la population : individus à faibles revenus (strate 1 : 3 800 employés) ; individus à revenus moyens (strate 2 : 1 000 employés) et individus à revenus élevés (strate 3 : 1 200 employés). Par sondage aléatoire simple, on sélectionne 50 employés par strate.

A. Un individu de la strate 2 a un poids égal à 10.
B. Un individu de la strate 2 a un poids égal à 20.
C. Un individu de la strate 2 a un poids égal à 30.
D. Aucune de ces affirmations n'est vraie.

Question 4 :

Pour contrôler la qualité d'une chaîne de production de rétroviseurs de véhicules Peugeot, on contrôle un rétroviseur sur 500.

A. La méthode d'échantillonnage est un échantillonnage au jugé.
B. La méthode d'échantillonnage est une méthode d'échantillonnage stratifié.
C. La méthode d'échantillonnage est une méthode aléatoire systématique.
D. La méthode d'échantillonnage est une méthode des quotas.
E. Aucune de ces affirmations n'est vraie.

Question 5 :

Un test statistique effectué au seuil de signification de 5% conduit à rejeter l'hypothèse nulle.

A. Le risque d'erreur de première espèce est 95%.
B. Le risque d'erreur de première espèce est 5%.
C. Le risque d'erreur de deuxième espèce est 5%.
D. Aucune de ces affirmations n'est vraie.

Question 6 :

A. Quand deux variables sont corrélées, il y a forcément un lien de cause à effet.
B. Quand deux variables sont corrélées, il n'y a pas forcément un lien de cause à effet.
C. Quand deux variables sont corrélées, il n'y a jamais de lien de cause à effet.
D. Aucune de ces réponses n'est vraie.

Question 7 :

On veut mesurer la prévalence du tabagisme chez les jeunes étudiants du département du Rhône. On sait que la consommation varie selon l'âge. Aussi, on va tirer au sort les individus de façon séparée au sein de deux groupes : les collégiens et les lycéens. De plus, la méthode de tirage accordera plus d'importance au groupe le plus hétérogène.

A. La méthode d'échantillonnage est une méthode aléatoire simple.

Exercice 2 (6 points)

On s'intéresse aux résultats de 10 pesées successives et indépendantes d'une même pièce mécanique, que l'on rapporte dans le tableau suivant (la masse est donnée en grammes) :

72.2	72.24	72.26	72.3	72.36	72.39	72.42	72.48	72.5	72.54

On admet que la variable aléatoire X, à laquelle on associe le résultat de la pesée de la pièce mécanique, suit une loi normale $X \sim N(m, \sigma^2)$.

1. A partir de l'échantillon obtenu, donnez des estimations ponctuelles de la moyenne m et de l'écart type σ de la population (2 points).

2. A partir de nombreuses études antérieures, on sait qu'en réalité, l'écart type de notre appareil de pesée, pour une pièce de cette masse, est de 0,08.

    a. Donnez un intervalle de confiance au risque de 5% pour la moyenne (2 points).
    b. Déterminez le niveau de risque α qu'il faudrait fixer pour que l'intervalle de confiance de la moyenne soit [72.31 ; 72.43] (2 points).

Exercice 3 (5 points)

On considère que le meilleur joueur d'une équipe de basket universitaire marque généralement 30 points par match. On veut décider si cette caractéristique se retrouve chez le meilleur joueur de l'équipe de l'ENTPE ou si celui-ci marque significativement moins de 30 points par match.

1. Formulez le problème sous la forme d'un test unilatéral en supposant une hypothèse Gaussienne (*i.e.* loi normale) pour la population (1 point).

2. On fait l'hypothèse que l'écart-type est connu et vaut 3 points par match. On observe les performances du meilleur marqueur de l'ENTPE pour les dix premiers matchs de la saison. Sur ces dix premiers matchs, celui-ci a marqué en moyenne 28.7 points par match. Que peut-on conclure, au risque de première espèce de 5% (2 points) ?

3. En réalité, le meilleur joueur de l'ENTPE marque exactement 29 points par match. Quel est le nom de l'erreur que l'on commet si l'on décide de ne pas rejeter (H_0) ? Que vaut cette erreur dans notre cas ? (2 points)

Exercice 4 (4 points)

Soit la variable aléatoire X de densité

$$f(x) = \frac{\lambda^4}{6} x^3 e^{-\lambda x}$$

qui est définie sur $\mathbb{R}^+$, avec $\lambda > 0$. On dispose d'un échantillon i.i.d $(X_1, \dots, X_n)$ de variable aléatoire parente X.

1. Donnez l'estimateur du maximum de vraisemblance du paramètre λ (2 points).

2. Calculez l'information de Fisher et la borne de Cramer-Rao. En déduire une loi approchée pour l'estimateur du maximum de vraisemblance (1 point).

3. Expliquez de façon générale le lien entre les notions précédentes et le concept d'efficacité d'un estimateur. Comment interviennent les statistiques exhaustives dans ce cadre théorique (1 point) ?