Statistique descriptive

Bachelier en sciences informatiques

Mardi 3 septembre 2019 – Partie 1: Théorie et QCM – 14h-15h15

NOM: .... PRENOM: ....

Indications

- Cette partie de l'examen dure 1h15.
- Il n'est pas permis d'utiliser une machine à calculer ni un ordinateur.
- Le tableau ci-dessous précise la répartition des points entre les différentes questions. Il n'est pas obligatoire de répondre aux questions dans l'ordre. Cependant, pour faciliter la correction et éviter les erreurs, vous êtes priés, à la fin de l'examen, de préciser pour chaque question si vous l'avez résolue (même partiellement) ou non en entourant soit OUI soit NON dans le tableau ci-dessous:

Théorie		QCM
Q1	Q2	
OUI NON	OUI NON	OUI NON
/11	/14	/15

TOTAL
/40

Théorie:

1. Soit $S = \{x_1, \ldots, x_n\}$ une série quantitative univariée de moyenne arithmétique $\bar{x}$.

   (a) Définir la série centrée associée et démontrer que sa moyenne est nulle.

   (b) Démontrer le résultat ci-dessous:

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 \leq \sum_{i=1}^{n}(x_i - a)^2 \text{ pour tout } a \in \mathbb{R}.$$

2. Soit $S = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ une série statistique quantitative bivariée obtenue en observant les variables X et Y sur n individus.

   (a) Donner la formule permettant de calculer la covariance de la série.

   (b) Déterminer (avec démonstration) l'effet d'un changement d'origine et d'échelle du type $x_i \longrightarrow x'_i = ax_i + b$ et $y_i \longrightarrow y'_i = cy_i + d$ sur ce paramètre statistique.

   (c) En exploitant la propriété démontrée ci-dessus (et **sans refaire** les développements), déterminer la covariance de la série bivariée *standardisée* correspondant à la transformation spécifique suivante:

$$x_i \longrightarrow x'_i = \frac{x_i - \bar{x}}{s_x} \text{ et } y_i \longrightarrow y'_i = \frac{y_i - \bar{y}}{s_y},$$

   avec s_x et s_y les écarts-types marginaux supposés non nuls.

QCM: Pour chaque question à choix multiples, veuillez choisir une et une seule réponse possible. En cas de réponse correcte, 1 point est acquis; en cas de réponse incorrecte, 0.25 points sont retranchés et si aucune réponse n'est cochée, aucun point n'est gagné ni perdu.

1. Le prix (au litre) du vin préféré de Monsieur et Madame Lambert a augmenté de 50% en un an. De quel pourcentage ce couple doit-il réduire sa consommation de ce vin de manière à garder le coût de son achat inchangé?

   ☐ 50% ☐ 33% ☐ 25% ☐ Aucune de ces réponses

2. Soit une variable quantitative mesurée en fonction d'une certaine unité de mesure (mètre, minute,...). En quelle unité le coefficient de dissymétrie de Yule et Kendall (défini en terme des quartiles) s'exprime-t-il?

   ☐ unité ☐ nombre pur (sans unité) ☐ unité au carré

3. Les tailles de 100 étudiants de BLOC 1 ont été mesurées en cm et sont décrites par l'ogive des fréquences cumulées représentée à la Figure 1.

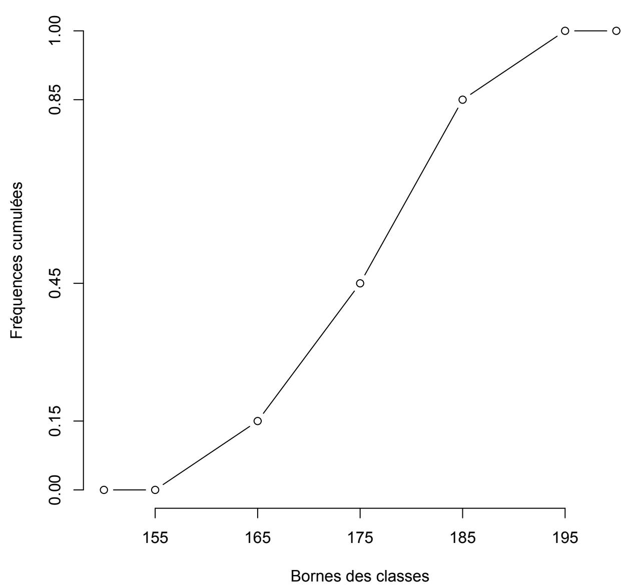

Figure 1: Ogive des fréquences cumulées des tailles de 100 personnes

   (a) Le 7ème décile de cette distribution estimé à l'aide de l'ogive vaut:

   ☐ 0.7 ☐ 175 ☐ 181.25 ☐ On n'a pas suffisamment d'information pour le calculer

   (b) Quel est l'effectif de la classe modale de cette série?

   ☐ 0.4 ☐ 40 ☐ 45 ☐ 85 ☐ On n'a pas suffisamment d'information pour le calculer

4. L'intensité des tremblements de terre est mesurée à l'aide d'un sismographe. Un particulier possède un tel appareil, mesurant précisément l'intensité lorsque celle-ci se situe entre 4 et 9 sur l'échelle ouverte de Richter. Ce particulier a enregistré les données suivantes lors des 10 derniers tremblements de terre que sa machine a ressenti:

   4.5. – L – 5.5 – H – 8.7 – 8.9 – 6.0 – H – 5.2 – 7.2

   où L indique que le tremblement de terre correspondant a une intensité inférieure à 4 et H précise que l'intensité est supérieure à 9.

   - La médiane de ces données vaut

     ☐ 6.0 ☐ 8.8 ☐ 6.6 ☐ la médiane ne peut pas être calculée car toutes les valeurs ne sont pas connues.

   - La moyenne tronquée au seuil $\alpha = 0.2$ vaut

     ☐ 6.0 ☐ 6.6 ☐ 6.92 ☐ la moyenne tronquée ne peut pas être calculée car toutes les valeurs ne sont pas connues.

5. On considère les boîtes à moustaches de la Figure 2.

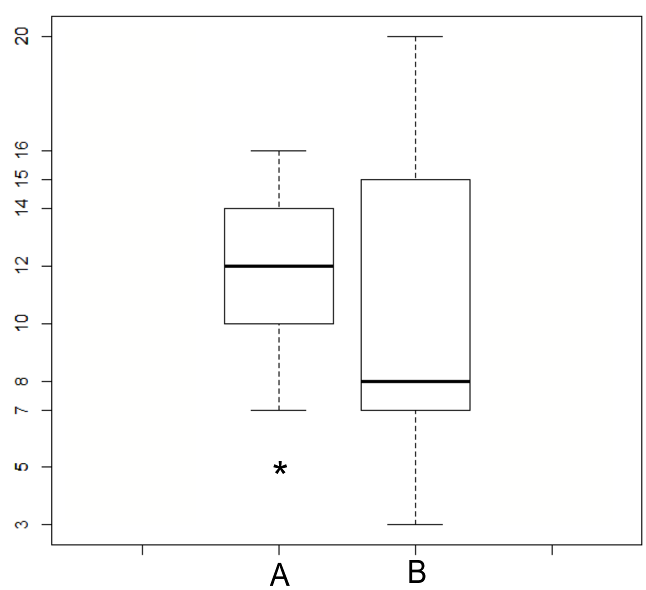

Figure 2: Boîtes à moustaches de deux séries quantitatives

   - L'étendue de la série A vaut

     ☐ 0 ☐ 4 ☐ 9 ☐ 11 ☐ 16 ☐ On n'a pas suffisamment d'information pour la calculer

   - La moyenne de la série B vaut

     ☐ 8 ☐ 10 ☐ 11 ☐ On n'a pas suffisamment d'information pour la calculer

6. Soit une variable binaire prenant uniquement les valeurs 0 et 1, observée sur n individus. La moyenne de cette variable est égale à la fréquence de la valeur 1:

   □ VRAI □ FAUX

7. Lorsque l'on calcule l'étendue d'une série statistique quantitative, quelle caractéristique de cette série essaye-t-on de quantifier?

   □ la tendance centrale □ la dispersion □ la dissymétrie

8. Une enquête menée auprès de 200 travailleurs se focalisait sur le lien entre l'état de santé (variable qualitative ordinale décrite par les trois modalités `Mauvais`, `Moyen` et `Bon`) et le niveau de bonheur (variable qualitative ordinale caractérisée par les trois modalités `Pas heureux`, `Plutôt heureux` et `Très heureux`). La tableau de contingence construit sur les données observées est disponible ci-dessous:

	Etat de santé		
Conditions	`Mauvais`	`Moyen`	`Bon`
`Pas heureux`	5	10	15
`Plutôt heureux`	15	30	25
`Très heureux`	10	55	35

Table 1: Tableau de contingence Etat de santé - Bonheur

   (a) La fréquence marginale de la modalité `Pas heureux` est égale à

   □ 30 □ 0.3 □ 0.2 □ 3/20

   (b) La fréquence de la modalité `Pas heureux` conditionnellement au fait que le travailleur est en bonne santé vaut:

   □ 15 □ 7/15 □ 3/40 □ 1/5

9. Soit $\hat{a}$ l'estimation obtenue par la technique des moindres carrés de la pente de la droite de régression linéaire de Y en X. Dès lors, le coefficient de corrélation entre X et Y vaut

   □ $\hat{a}s_xs_y$ □ $\frac{\hat{a}s_x}{s_y}$ □ $\frac{s_xs_y}{\hat{a}}$ □ 1

   □ On ne dispose pas des informations nécessaires pour répondre

10. Lors de l'observation de deux variables statistiques quantitatives sur n individus, on constate que les valeurs observées vérifient exactement la relation linéaire suivante: $y = ax + 2$ pour a une constante différente de 0. De plus, les variances marginales des deux variables sont égales à 1. Sachant que $\bar{x} = 1$ et que la corrélation vaut 0.5, que vaut la moyenne marginale $\bar{y}$?

    □ 0 □ 0.5 □ 1 □ 2.5 □ On n'a pas assez d'informations pour la calculer

11. On s'intéresse à la relation entre le nombre de bières (cannette de 33 cl) consommées (variable X) et le niveau d'alcoolémie dans le sang (variable Y, mesurée en une certaine unité). A l'aide de la technique des moindres carrés et à partir de données mesurées sur des personnes présentes à un festival, on a obtenu l'équation suivante pour la droite de régression de Y en X: $y = -0.0127 + 0.018x$.

    Deux amis, repris parmi les personnes testées, ont consommé respectivement x^* et $x^* + 1$ canettes de bière. Selon ce modèle, celui qui a consommé $x^* + 1$ canettes a un niveau attendu d'alcoolémie dans le sang

    □ inférieur de 0.0127 unité par rapport à celui de son ami,

    □ supérieur de 0.018 unité par rapport à celui de son ami,

    □ supérieur de $0.018 - 0.0127 = 0.0053$ unité par rapport à celui de son ami.