Statistique descriptive

Bachelier en sciences informatiques

Correction de l'examen écrit du 3 septembre 2019

QCM:

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

   Justification : Notons p_0 le prix du litre de vin l'année passée et p_1 le prix du litre de vin actuellement. On a

   $$p_1 = 1.5p_0.$$

   Avec une somme S, le couple pouvait, l'année passée, acheter une quantité

   $$q_0 = \frac{S}{p_0}$$

   de vin. Avec cette même somme, ils peuvent actuellement acheter une quantité

   $$q_1 = \frac{S}{p_1} = \frac{S}{1.5p_0}$$

   de vin. On a alors un pourcentage de variation donné par

   $$\frac{q_1 - q_0}{q_0} = \frac{\frac{S}{1.5p_0} - \frac{S}{p_0}}{\frac{S}{p_0}} = \frac{S - 1.5S}{1.5S} = \frac{-0.5}{1.5} = -0.33.$$

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.

   (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

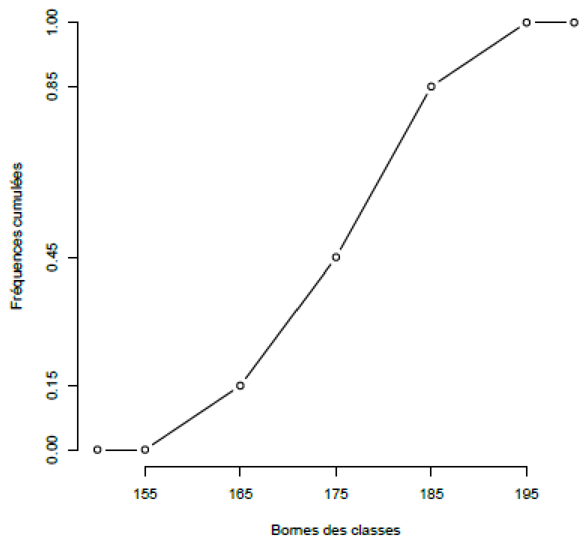

Figure 1: Ogive des fréquences cumulées des tailles de 100 personnes

<u>*Justification*</u> : Le 7ème décile, noté $q_{0.7}$, correspond à une fréquence cumulée de 0.7. Dès lors, sur l'ogive des fréquences cumulées, le point $(q_{0.7}, 0.7)$ est situé sur la droite définie par les points $(175, 0.45)$ et $(185, 0.85)$, ayant pour équation

$$y - 0.45 = \frac{0.85 - 0.45}{185 - 175}(x - 175).$$

En injectant les coordonnées $(q_{0.7}, 0.7)$ dans l'équation de cette droite, on obtient la valeur de $q_{0.7}$.

(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

<u>*Justification*</u> : La classe modale est la classe $]175, 185]$, ayant une fréquence égale à $0.85 - 0.45 = 0.4$ et donc un effectif égal à $0.4 \times 100 = 40$.

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.

Justification : La série comptant 10 observations, la médiane est donnée par la moyenne arithmétique entre les observations de rangs 5 et 6, i.e. $\frac{6+7.2}{2} = 6.6$.

- 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.

  Justification : Pour calculer la moyenne tronquée au seuil $\alpha = 0.2$, il faut retirer les $0.2 \times 10 = 2$ observations les plus petites et les 2 observations les plus grandes. La moyenne calculée sur les 6 observations restantes est donnée par 6.92.

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

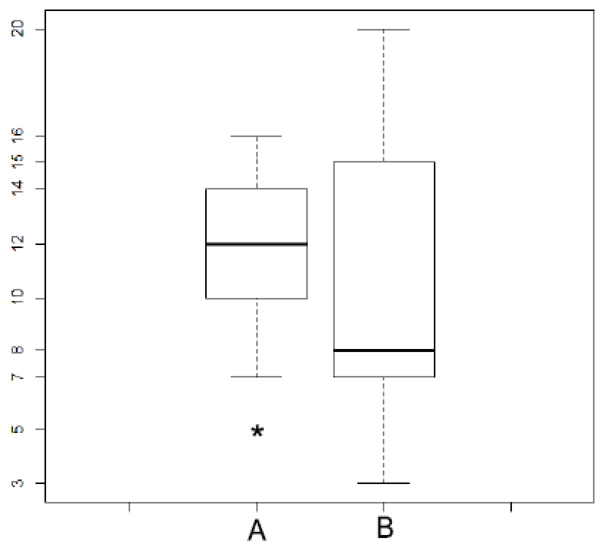

Figure 2: Boîtes à moustaches de deux séries quantitatives

- L'étendue de la série A vaut

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

  Justification : L'étendue étant définie par l'écart entre la plus grande et la plus petite observation, elle vaut $16 - 5 = 11$.

- 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

   Justification : La fréquence marginale de la modalité `Pas heureux` est donnée par $\frac{5+10+15}{200} = \frac{3}{20}$.

   (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

   Justification : La fréquence de la modalité `Pas heureux` conditionnellement au fait que le travailleur est en bonne santé est donnée par $\frac{15}{15+25+35} = \frac{1}{5}$.

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

   Justification : Puisque $\hat{a} = \dfrac{s_{xy}}{s_x^2}$, on a $r_{xy} = \dfrac{s_{xy}}{s_xs_y} = \hat{a}\dfrac{s_x}{s_y}$.

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

<u>*Justification*</u> : On a $\hat{a} = \dfrac{s_{xy}}{s_x^2} = \dfrac{r_{xy}s_xs_y}{s_x^2} = \dfrac{0.5 \times 1 \times 1}{1} = 0.5$, et $\hat{b} = \bar{y} - \hat{a}\bar{x}$, i.e. $\bar{y} = 2 + 0.5 \times 1 = 2.5$.

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.

    <u>*Justification*</u> : La pente de la droite de régression de Y en X correspond à l'augmentation de la variable dépendante lors d'une augmentation unitaire de la variable explicative.