HAI205X Probas stats - Correction

Assurez-vous de présenter vos réponses de manière claire et lisible. Les exercices peuvent être traités dans le désordre. La calculatrice est autorisée. Les valeurs numériques seront arrondies à deux décimales. Le barême est donné à titre indicatif. Vous avez 1 heure. Bon courage !

Exercice 1 (12.5 points)

1. (2.5 points) Le temps journalier devant les écrans pour 50 étudiants de licence d'informatique sont donnés dans le tableau suivant:

Temps journalier (heures)	$[0;2[$	$[2;4[$	$[4;6[$	$[6;8[$	$[8;10[$
Nombre d'étudiants	2	5	13	19	11

   (a) (0.5 points) Calculer le temps moyen passé, par jour, devant les écrants pour ces étudiants. **Solution:** $\bar{X} = \frac{1}{50}(2\times 1 + 5\times 3 + 13\times 5 + 19\times 7 + 11\times 9) = 6.28$

   (b) (1 point) Déterminer, par le calcul, la médiane. **Solution:** $Me = \frac{25-20}{39-20}\times 2 + 6 = 6.53$

Temps journalier (heures)	$[0;2[$	$[2;4[$	$[4;6[$	$[6;8[$	$[8;10[$
Nombre d'étudiants	2	5	13	19	11
Cumulé	2	7	20	39	50

   (c) (1 point) Déterminer, par le calcul, l'intervalle interquartile. **Solution:** $Q_1 = \frac{13-7}{20-7}\times 2 + 4 = 4.92$ **et** $Q_3 = \frac{38-20}{39-20}\times 2 + 6 = 7.89$ **donc** $IQR = Q_3 - Q_1 = 7.89 - 4.92 = 2.97$

2. (10 points) Le temps journalier devant les écrans pour 50 étudiants de licence de mathématiques sont donnés dans le tableau suivant:

Temps journalier (heures)	$[0;2[$	$[2;4[$	$[4;6[$	$[6;8[$	$[8;10[$
Nombre d'étudiants	6	12	21	8	3

   (a) (5.5 points) A l'aide d'un graphique, dont vous donnerez le nom, comparer les deux distributions. Commenter le résultat. **Solution:** **On calcule les effectifs cumulés (ou fréquences cumulées): On calcule la moyenne, la médiane,** Q_1**,** Q_3 **et l'intervalle interquartile.** $\bar{X} = \frac{1}{50}(6\times 1 + 12\times 3 + 21\times 5 + 8\times 7 + 3\times 9) = 4.60$**,** $Me = \frac{25-18}{39-18}\times 2 + 4 = 4.67$**,** $Q_1 = \frac{13-6}{18-6}\times 2 + 2 = 3.17$**,** $Q_3 = \frac{38-18}{39-18}\times 2 + 4 = 5.90$ **et** $IQR = 5.90 - 3.17 = 2.73$**. On regroupe les informations sous forme de boxplot.**

Temps journalier (heures)	[0; 2[	[2; 4[	[4; 6[	[6; 8[	[8; 10[
Nombre d'étudiants	6	12	21	8	3
Cumulé	6	18	39	47	50

(b) (4.5 points) Evaluer la dispersion relative du temps d'écran des deux groupes d'étudiant. En déduire le groupe le plus homogène **Solution:** **On calcule le coefficient de variation et le coefficient de dissymétrie de Pearson.** $C_v(info) = \frac{\sigma}{\bar{X}} = \frac{\sqrt{4.44}}{6.28} = \frac{2.11}{6.28} = 0.34$ **et** $C_v(math) = \frac{\sqrt{4.32}}{4.60} = \frac{2.08}{4.60} = 0.45$. **La dispersion chez les matheux est plus légèrement plus grande.** $S(info) = \frac{\bar{X}-Mode}{\sigma} = \frac{6.28-7}{2.11} = -0.34$ **et** $S(math) = \frac{\bar{X}-Mode}{\sigma} = \frac{4.60-5}{2.08} = -0.19$ **les séries sont étalées à gauche. On remarque que le temps d'écran est plus grand en moyenne chez les informaticiens, mais les valeurs sont plus dispersés chez les matheux.**

Exercice 2 (5.5 points)

On s'intéresse au développement de la myopie chez les étudiants. Le niveau de dioptrie (mesure allant de 0 à 20) est mesuré sur 8 étudiants pendant leur cursus universitaire.

Numéro i de l'étudiant	1	2	3	4	5	6	7	8
Dioptrie	0.50	0.25	2.10	1.90	2.90	3.50	4.30	5.90
Temps d'étude	0.50	1	2	2	2.50	3.50	3.50	4.50

1. (2 points) Calculer la covariance et le coefficient de corrélation. Commenter. **Solution:** **Soit X le temps d'étude et Y la mesure de dioptrie.** $\bar{X} = \frac{1}{8}(0.5 + 1 + 2 + 2 + 2.5 + 3.5 + 3.5 + 4.5) = 2.44$, $\bar{Y} = \frac{1}{8}(0.5 + 0.25 + 2.10 + 1.90 + 2.90 + 3.50 + 4.30 + 5.90) = 2.67$ **et** $\bar{XY} = \frac{1}{8}(0.5 \times 0.5 + 1 \times 0.25 + 2 \times 2.10 + 2 \times 1.90 + 2.5 \times 2.90 + 3.5 \times 3.50 + 3.5 \times 4.30 + 4.5 \times 5.90) = 8.70$ $Cov(X,Y) = \bar{XY} - \bar{X}\bar{Y} = 8.70 - 2.44 \times 2.67 = 2.18$. $\sigma_X = \sqrt{\frac{1}{8}(0.5^2 + 1^2 + 2^2 + 2^2 + 2.5^2 + 3.5^2 + 3.5^2 + 4.5^2) - 2.44^2} = \sqrt{1.58} = 1.26$ **et** $\sigma_Y = \sqrt{\frac{1}{8}(0.5^2 + 0.25^2 + 2.10^2 + 1.90^2 + 2.90^2 + 3.5^2 + 4.30^2 + 5.90^2) - 2.67^2} = \sqrt{3.16} = 1.78$, **le coefficient de corrélation est** $r_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{2.18}{1.26 \times 1.78} = 0.97$ **Le coefficient de corrélation de Pearson est proche de 1 donc la dioptrie est positivement corrélé linéaire avec le temps d'étude. Ici la régression optimale est donc linéaire.**

2. (2 points) Déterminer par la méthode des moindres carrés le coefficient de régression de la courbe. **Solution:** **On a** $Y = aX + b$. **On estime** $a = \frac{Cov(X,Y)}{V(X)} = \frac{2.18}{1.58} = 1.38$. **On peut maintenant estimer** $b = \bar{Y} - a\bar{X} = 2.67 - 1.38 \times 2.44 = -0.70$.

3. (1.5 points) Un étudiant rejoint l'étude après 4 année d'étude universitaire. Estimer sa dioptrie. Interpréter les résultats. (*indication : une dioptrie inférieur à 3.25 correspond à une myopie légère, inférieure à 6.75 à une myopie moyenne et au delà à un myopie forte.* **Solution:** **On a** $\hat{Y} = 1.38 \times 4 - 0.70 = 4.82$. **Après 4 année d'étude universitaire, l'étudiant aura, d'après le modèle posé, une dioptrie de 4.82, ce qui correspond à une myopie moyenne.**

Exercice 3 (4 points)
Il s'agit d'un questionnaire à choix multiple. La justification de certains calculs serait appréciée.

1. (1 point) Quand les amplitudes sont inégales, pour dessiner l'histogramme:
   (a) On calcule l'étendu
   (b) On corrige les effectifs
   (c) On calcule les effectifs cumulés
   (d) On calcule la densité des fréquences

   Solution: b, d

2. (1 point) La représentation graphique d'une variable quantitative discrète est
   (a) Un diagramme en bâton
   (b) Un histogramme
   (c) Une courbe de fréquences cumulées
   (d) Le polygone de fréquences

   Solution: a, c, d

3. (1 point) A partir de la Table 1, quelles affirmations sont vraies ?

X Y	$[0;3[$	$[3;6[$	$[6;9[$
$[0;20[$	18	2	0
$[20;40[$	8	10	2
$[40;60[$	4	10	6
$[60;80[$	2	8	10

Table 1: Répartition des mesures de dioptrie en fonction de l'âge.

   (a) Les fréquences marginales de Y sont $f_{[0;3[}(y) = 0.40$, $f_{[3;6[}(y) = 0.37$ et $f_{[6;9[}(y) = 0.23$.
   (b) Les fréquences marginales de X sachant $Y = [0;3[$ sont $f_{[0;20[}(x) = 0.56$, $f_{[20;40[}(x) = 0.25$, $f_{[40;60[}(x) = 0.12$ et $f_{[60;80[}(x) = 0.07$.
   (c) La moyenne marginale de X est 40.
   (d) La formule de la variance conditionnelle de Y sachant $X = [20;40[$ est $V_{X=[20;40[}(Y) = \frac{1}{20}(8 \times 0^2 + 10 \times 3^2 + 2 \times 6^2) - \bar{Y}^2_{X=[20;40[}$, où $\bar{Y}_{X=[20;40[}$ est la moyenne conditionnelle de Y sachant $X = [20;40[$.

   Solution: a, c $\bar{\bar{X}} = \frac{1}{80}(20 \times 10 + 20 \times 30 + 20 \times 50 + 20 \times 70) = 40$. **Pour la variance conditionnelle il faut utiliser les centres de classes.**

4. (1 point) A partir de la Table 1, quelles affirmations sont vraies ?
   (a) Il y a une corrélation positive entre âge et mesure de dioptrie.
   (b) Il y a une corrélation négative entre âge et mesure de dioptrie.
   (c) Le coefficient de corrélation linéaire est $r_{X,Y} = 0.81$.

(d) Le coefficient de corrélation linéaire est $r_{X,Y} = 0.31$.

$\boxed{\textbf{Solution:}}$ **a, d** $r_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$**.**
$Cov(X,Y) = \bar{XY} - \bar{X}Y = \frac{1}{80}(18 \times 10 \times 1.5 + \cdots + 10 \times 70 \times 7.5) - 40 \times 3.97 = 31.7$
$\sigma_X = \sqrt{500} = 22.36$ **et** $\sigma_Y = \sqrt{21.15} = 4.60$
donc $r_{X,Y} = \frac{31.7}{22.36 \times 4.60} = 0.31$**.**