Statistique descriptive
Bachelier en sciences informatiques
Vendredi 7 juin 2019 – Partie 1: théorie et QCM – 8h15-9h30

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

<u>Indications</u>
- 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

<u>Théorie</u>:

1. Soit $\{x_1, \ldots, x_n\}$ une série statistique quantitative univariée dont **l'effectif** n **est impair**.

   (a) Préciser, dans le cas spécifique de cette série brute, les formules permettant de calculer la moyenne $\bar{x}$ et la médiane $\tilde{x}$.

   (b) Soient $\mathbf{a} > \mathbf{0}$ et $b \in I\!R$. Déterminer (avec démonstration) l'effet d'un changement d'origine et d'échelle du type
   $$x_i \to x_i' = ax_i + b$$
   sur la moyenne et la médiane en tenant compte des informations données sur la série et sur la transformation.

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

   (a) Ecrire l'équation de la droite de régression de Y en X obtenue par la technique des moindres carrés en donnant précisément les formules des coefficients de pente et d'ordonnée à l'origine.

   (b) La Figure 1 correspond au diagramme de dispersion d'une série bivariée comptant 30 observations (le symbole plein représente le centre de gravité de ce nuage de points). Sans effectuer de calculs, dessiner une droite qui pourrait être le résultat de la recherche de la droite de régression de Y en X en exploitant la technique

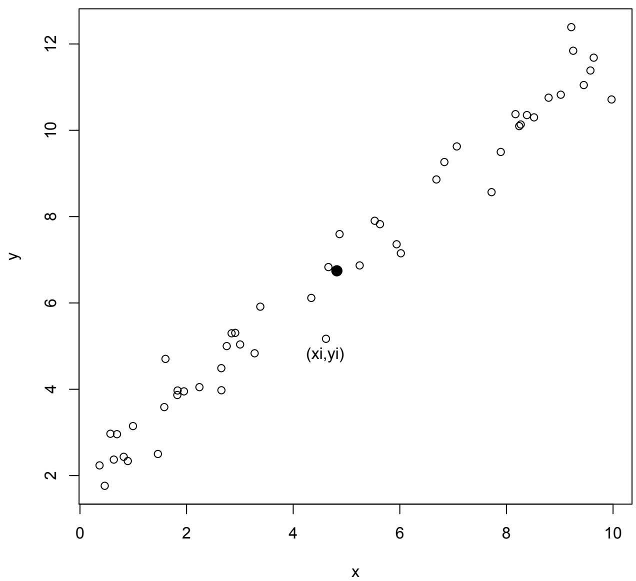

Figure 1: Diagramme de dispersion de Y en X

des moindres carrés. Illustrer, à l'aide de cette droite et sur le graphique, le résidu du couple bivarié (x_i, y_i) et indiquer où se situe la valeur ajustée correspondant à ce couple.

(c) Exprimer la variance de la série des résidus en fonction de r_{xy}^2 et s_y^2 où r_{xy}^2 est le coefficient de corrélation et s_y^2 la variance marginale de Y et démontrer cette formule.

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. Dans un club de sport il y a 60% de garçons. On sait que 40% des garçons de ce club font du tennis et 30% des filles de ce club font du tennis. Le pourcentage global de personnes pratiquant le tennis dans ce club est de

   ☐ 24 % ☐ 36 % ☐ 70 % ☐ On ne peut pas le calculer.

2. A la fin du mois d'avril, une tablette coûtait 999 EURO après avoir subi une baisse de 10% depuis le début du mois d'avril. Quel était le prix de cette tablette (arrondi à l'unité la plus proche) au début du mois d'avril?

   ☐ 1099 EURO ☐ 1110 EURO ☐ 899 EURO

3. Soit une variable quantitative mesurée en fonction d'une certaine unité de mesure (mètre, minute,...). En quelle unité l'écart interquartile s'exprime-t-il?

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

4. Lorsque l'on calcule le coefficient γ_1 de Fisher (moment centré d'ordre 3 divisé par l'écart-type au cube), quelle caractéristique de la série essaye-t-on de quantifier?

   ☐ la tendance centrale ☐ la dispersion ☐ la dissymétrie

5. Une enquête menée auprès d'un grand nombre de jeunes s'intéressait à leur loisir préféré (un seul choix possible entre les quatre modalités: `Shopping`, `Jeux Vidéos`, `Sport`, `Télé/Netflix`). Les résultats sont illustrés sur le diagramme en secteurs ci-dessous:

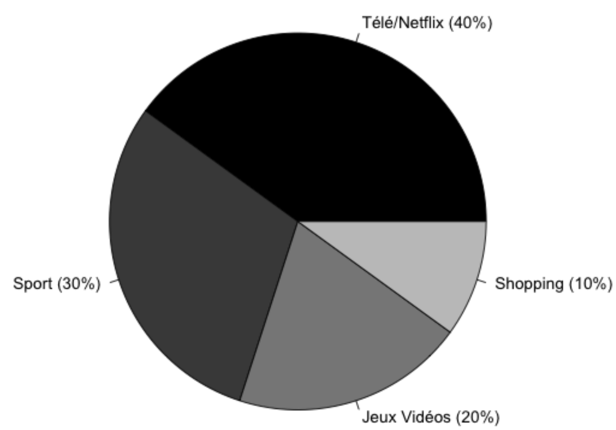

Figure 2: Diagramme en secteurs des loisirs d'un grand nombre de jeunes

   Quel est l'effectif du mode de cette série?

   ☐ 60% ☐ 60 ☐ On n'a pas assez d'information pour le calculer.

6. Après avoir corrigé les examens de physique de sa classe de 25 élèves en attribuant des cotes sur 20 points (valeurs entières entre 0 et 20), un professeur se rend compte que la moyenne vaut 9, avec un écart-type égal à 3. Il décide, en vue de ne pas démoraliser ses élèves, de modifier ses cotes à l'aide de la transformation affine suivante: $x'_i = \frac{2}{3}x_i + 10$.

   - L'écart-type des cotes transformées
     ☐ subit la même transformation
     ☐ est multiplié par 2/3
     ☐ reste inchangé
     ☐ On n'a pas assez d'information pour conclure.
   - En considérant que tous les élèves doivent pouvoir bénéficier de cette transformation, quelle(s) contrainte(s) faut-il imposer sur les cotes initiales minimale et maximale pour que chaque cote transformée soit compris en 0 et 20?
     ☐ Pas de contrainte sur $x_{(1)}$ mais $x_{(25)} \leq 10$
     ☐ $x_{(1)} \geq 3$ et $x_{(25)} \leq 15$

☐ Pas de contrainte sur $x_{(1)}$ mais $x_{(25)} \leq 15$

☐ Aucune contrainte

- Avant la modification des cotes de ce professeur, la corrélation entre ses points et ceux attribués aux mêmes élèves par sa collègue enseignant les mathématiques était égale à 0.85. Que devient la corrélation après la transformation du professeur de physique, les cotes de mathématique restant inchangées?

  ☐ Elle est multipliée par 2/3 ☐ Elle reste inchangée ☐ On n'a pas assez d'information pour conclure.

7. Une enquête a été menée auprès de 1000 personnes afin de déterminer le temps consacré au sport par semaine (les données étaient enregistrées en minutes). L'ogive des fréquences cumulées construite après avoir regroupé les observations en 5 classes (dont les bornes s'expriment en heures) est représentée à la Figure 3.

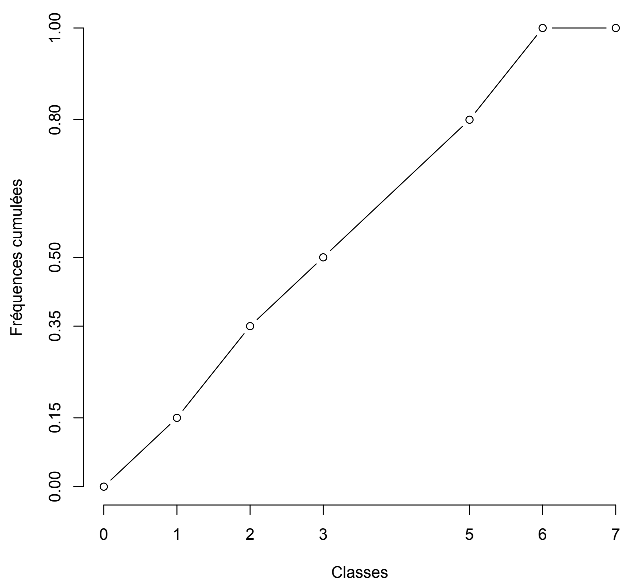

Figure 3: Ogive d'une série décomposée en 5 classes

- La proportion de personnes pratiquant entre 1h30 et 4h de sport par semaine est égale à

  ☐ 0.35 ☐ 0.40 ☐ 0.65 ☐ On ne peut pas la calculer.

- Sur l'histogramme d'aire unitaire représentant cette série groupée, la hauteur du rectangle correspondant à la classe $]3, 5]$ serait égale à

  ☐ 0.15 ☐ 0.30 ☐ 0.5 ☐ On n'a pas assez d'information pour la calculer.

8. Soit une série statistique quantitative bivariée constituée des observations $(x_i, y_i), 1 \leq i \leq n$, dont les moyennes marginales sont nulles et dont les variances marginales sont

égales à 1. On précise également que l'équation de la droite de régression de Y en X, estimée par la technique des moindres carrés, est donnée par $y = x/2$. Le coefficient de détermination associé à cet ajustement est égal à

☐ 1 ☐ 1/2 ☐ 1/4 ☐ On n'a pas assez d'information pour le calculer

9. La distribution du nombre d'écoles implantées dans 100 communes belges est décrite par la courbe cumulative des fréquences cumulées à la Figure 4.

   - Selon la convention imposée dans le cours, le 9ème décile de cette série vaut :
     ☐ 4 ☐ 4.5 ☐ 0.9 ☐ Il n'est pas défini
   - La moyenne tronquée de paramètre $\alpha = 0.25$ de la série vaut
     ☐ 2.9 ☐ 3 ☐ Il manque des informations pour calculer cette moyenne

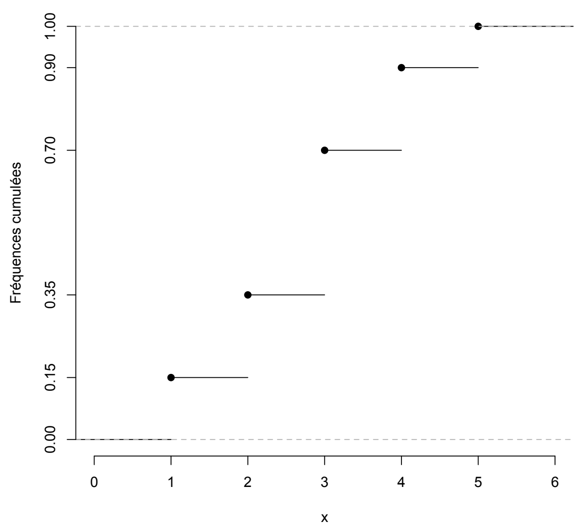

Figure 4: Représentation de la distribution des fréquences cumulées

10. Une droite de régression estimée par la technique des moindres carrés a été ajustée au nuage de points de la Figure 5.

    Des graphiques de résidus représentés en fonction de la variable explicative X sont disponibles à la Figure 6. Quel graphique des résidus représente effectivement les résidus de la droite ajustée sur les données de la Figure 5?

    ☐ Le graphique (a) ☐ Le graphique (b) ☐ Le graphique (c)

    ☐ Aucun des trois graphiques

11. Que vaut la valeur pivot supérieure (ou de droite) de la série représentée par la boîte à moustaches de base (les moustaches vont jusqu'aux valeurs minimale et maximale) de la Figure 7?

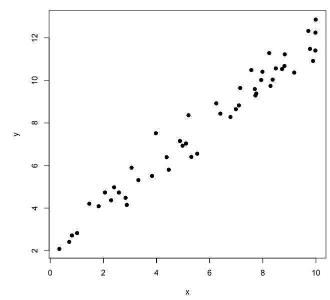

Figure 5: Diagramme de dispersion construit sur les deux variables X et Y

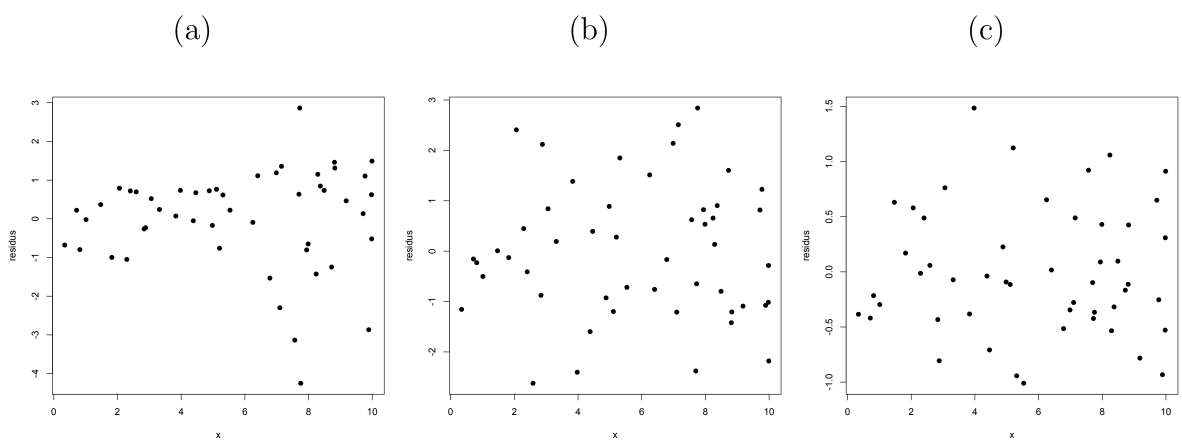

Figure 6: Graphiques de résidus représentés en fonction des valeurs de X

☐ 7 ☐ 11.5 ☐ 14 ☐ Il manque des informations pour calculer cette valeur

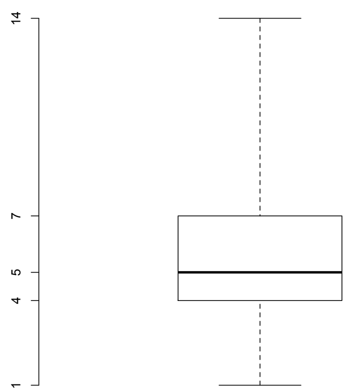

Figure 7: Boîte à moustaches d’une série quelconque