**Probabilité et statistique I (partim statistique descriptive)**

Bachelier en sciences informatiques

Mardi 20 juin 2017 – Partie 1: théorie et QCM – 9h-10h

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

<u>**Indications**</u>
- Cette partie de l'examen dure 1h.
- Il n'est pas permis d'utiliser une machine à calculer.
- 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 |
| /8 | /12 | /10 |

| TOTAL |
|---|
| /30 |

<u>**Théorie**</u>:

1. Soit $\{x_1, \ldots, x_n\}$ une série statistique quantitative univariée de moyenne $\bar{x}$ et de variance $s^2$.

   (a) Déterminer (avec démonstration) l'effet d'un changement d'origine et d'échelle sur ces deux paramètres statistiques.

   (b) **Déduire** de ces propriétés la moyenne et la variance de la nouvelle série définie comme suit:
   $$x_i' = \frac{x_i - \bar{x}}{s}$$

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 (la croix représente le centre de gravité de ce nuage de points). A main levée (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 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.

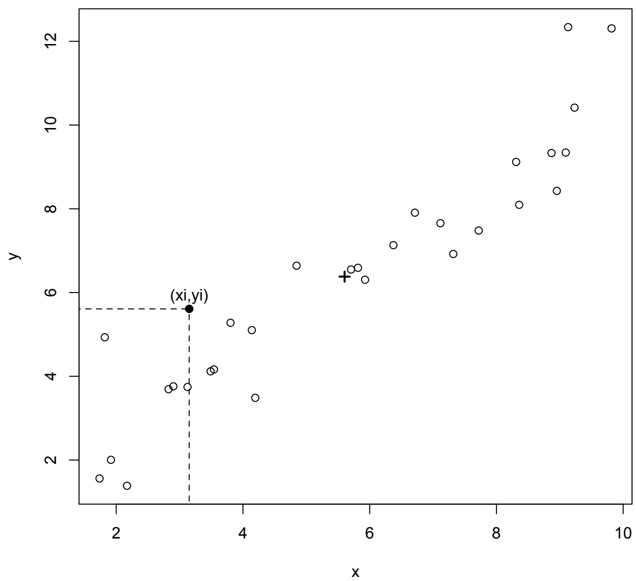


Figure 1: Diagramme de dispersion de $Y$ en $X$

(c) Exprimer la moyenne et la variance de la série des valeurs ajustées en fonction de la moyenne et de la variance de la série $Y$ de départ et démontrer ces formules.

**QCM**: Pour chaque question, veuillez choisir une et une seule réponse possible pour chaque choix multiple. 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. Lors d'un référendum local, 40% des femmes et 70% des hommes ont répondu oui à la question posée (réponses possibles: oui – non). Sachant que l'électorat contient 65% de femmes et qu'il n'y a eu aucun vote blanc, déterminer si les assertions suivantes sont vraies ou fausses:

   - Il y a plus d'hommes ayant voté oui que de femmes ayant voté oui : ☐ Vrai – ☐ Faux
   - Le vote oui l'a remporté : ☐ Vrai – ☐ Faux

2. Les nombres d'écoles implantées dans 50 communes de la province de Liège sont décrits dans le tableau ci-dessous:

| Nombres d'écoles | Effectifs |
|---|---|
| 2 | 7 |
| 3 | 6 |
| 4 | 8 |
| 5 | 7 |
| 6 | 6 |
| 7 | 6 |
| 8 | 4 |
| 9 | 3 |
| 10 | 3 |

Table 1: Tableau statistique des nombres d'écoles dans 50 communes

   - Le mode de cette série est égal à 8: ☐ Vrai – ☐ Faux
   - Le 9ème décile de cette série vaut 9: ☐ Vrai – ☐ Faux
   - Le graphique le plus adéquat pour représenter la distribution des effectifs cumulés de cette série est
     ☐ le graphique de gauche de la Figure 2
     ☐ le graphique de droite de la Figure 2
     ☐ aucun de ces deux graphiques.

3. Soit une série quantitative dont on souhaite résumer les caractéristiques principales. Déterminer de quel type (paramètre de tendance centrale, de dispersion ou de dissymétrie) les paramètres suivants sont:

   - L'écart interquartile: ☐ Tendance centrale – ☐ Dispersion – ☐ Dissymétrie

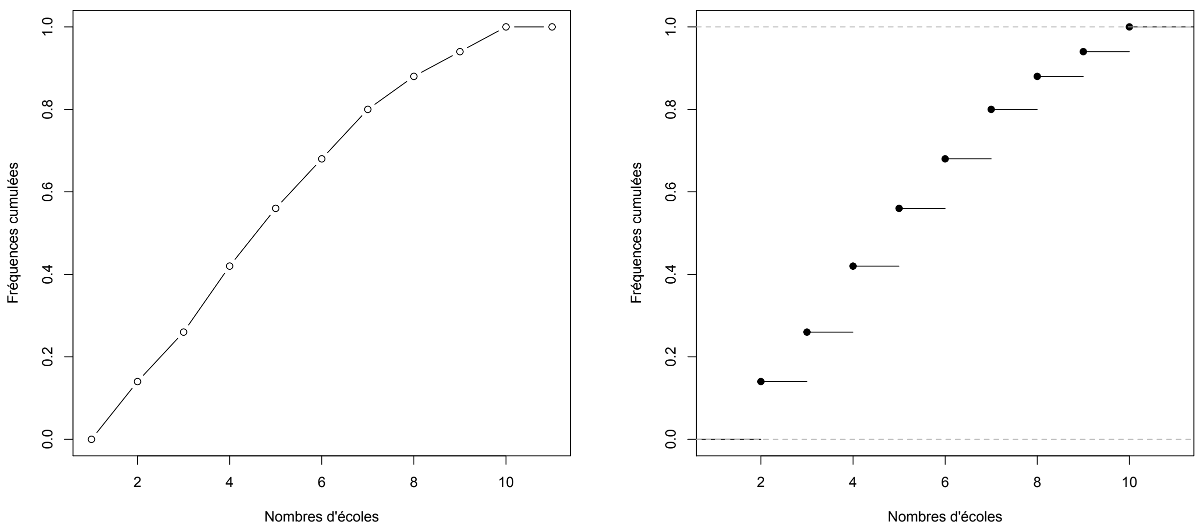


Figure 2: Représentations de la distribution des fréquences cumulées

- Le coefficient $\gamma_1 = \frac{m_3}{s^3}$ où $m_3$ est le moment centré d'ordre 3 et $s$ l'écart-type:
  ☐ Tendance centrale – ☐ Dispersion – ☐ Dissymétrie

4. L'Institut Maurice-Lamontagne du Québec est spécialisé dans l'étude de la pêche et des océans. Dans ce contexte, les membres de cet Institut mesurent un certain nombre de variables météorologiques chaque jour. Notamment, ils mesurent la température (variable $X$, en degrés celsius) ainsi que le nombre de millimètres de pluie tombée par mètre carré (variable $Y$). Pour une période temporelle fixée, ils ont obtenu une covariance entre les deux variables égale à $s_{xy}$ tandis que la corrélation vaut $r_{xy}$.

   - Si la série des températures est transformée en degrés Farhenheit (tranformation $1.8x_i + 32$), le coefficient de corrélation est multiplié par 1.8: ☐ Vrai – ☐ Faux
   - Si la série des précipitations est transformée en mètres, la covariance est multipliée par 1000: ☐ Vrai – ☐ Faux
   - Si les deux séries sont transformées comme expliqué ci-dessus, la pente de la droite de régression (calculée par la technique des moindres carrés) de $Y$ en $X$ est multipliée par 1800: ☐ Vrai – ☐ Faux