ENSA d'Al-Hoceima | Année 2019/2020
AP-II, Deuxième Année, | Semestre 4
Probabilité & Statistique, | TD0

Exercice avec solution : Statistique descriptive

Exercice 1

On a mesuré la taille d'un groupe d'élèves en arrondissant les résultats à 5 cm

TABLE 1 – Tableau de données

Taille arrondie (cm)	Nombre d'élèves	Fréquence	Fréquence cumulée
155	6		
160	9		
165	5		
170	3		
175	1		

1. S'agit-il d'une série statistique continue où discrète ? **Justifier**
2. Remplir le tableau par les fréquences et fréquences cumulées. Représentez, dans le même graphique, le diagramme de fréences et de fréquences cumulées.
3. Calculer la moyenne, la médiane et l'écart-type de cette série statistique.
4. En comparant la moyenne et la médiane de cette série, que dire de cette distribution.
5. On compare la distribution précédente à celle d'un autre groupe dont la moyenne et l'écart-type sont respectivement $M_2 = 165$ et $\sigma_2 = 5.6$.
   Laquelle des 2 distributions est plus homogène ?

Solution :

1. La série statistique continue, en effet, la phrase **la taille d'un groupe d'élèves en arrondissant les résultats à 5 cm** veut dire que la taille du groupe d'élèves est majorée à la plus grande sur une amplitude de 5cm, c'est à dire le chiffre affiché pour chaque classe, dans le tableau, représente le majorant des tailles sur un écart de 5cm. **Par exemple :** pour la première classe ou la taille est 155cm avec un effectif de 6 élèves, alors la taille la plus petite de ce groupe est 150cm, donc les tailles de cette classe de 6 élèves peuvent aller de 150cm à 155cm ; d'où il s'agit bien d'un intervalle de taille allant de 150cm à 155cm, soit $[150, 155[$. Ainsi, de suite pour les autres classes.
2. Les valeurs du tableau par les fréquences et fréquences cumulées : l'effectif total des élèves est $6 + 9 + 5 + 3 + 1 = 24$ élèves.

   Représentation graphique : le diagramme de freqences et de fréquences cumulées
3. Calculons la moyenne, la médiane et l'écart-type de cette série statistique :
   Moyenne :

$$M_1 = \overline{x} = \frac{1}{24}(6 \times 155 + 9 \times 160 + 5 \times 165 + 3 \times 170 + 1 \times 175) = \frac{3880}{24} = \frac{485}{3} = 161.6667$$

TABLE 2 – Tableau de données

Taille arrondie (cm)	Nombre d'élèves	Intervalle de tailles	centre	Fréquence	Fréquence cumulée
150	0	[150,155[	152.5	0	0
155	6	[155,160[	157.5	$\frac{6}{24} = \frac{1}{4}$	$0 + \frac{1}{4} = \frac{1}{4}$
160	9	[160,165[	162.5	$\frac{9}{24} = \frac{3}{8}$	$\frac{9}{24} + \frac{6}{24} = \frac{15}{24} = \frac{5}{8}$
165	5	[165,170[	167.5	$\frac{5}{24}$	$\frac{5}{24} + \frac{15}{24} = \frac{20}{24} = \frac{5}{6}$
170	3	[170,175[	172.5	$\frac{3}{24} = \frac{1}{8}$	$\frac{3}{24} + \frac{20}{24} = \frac{23}{24}$
175	1	[175,180[	177.5	$\frac{1}{24}$	$\frac{1}{24} + \frac{23}{24} = 1$

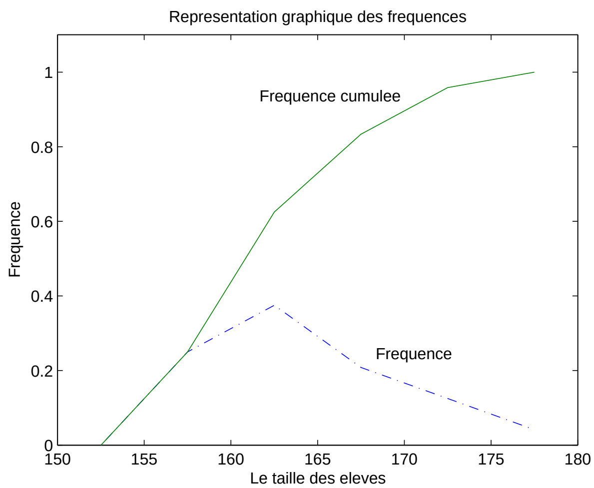

FIGURE 1 – Le diagramme de fréquences et fréquences cumulées.

Médiane : l'effectif total 24 est un nombre pair, alors on utilise la formule

$$M_e = \frac{X_{24/2} + X_{24/2+1}}{2} = \frac{X_{12} + X_{13}}{2} = \frac{160 + 160}{2} = 160$$

On remarque aussi que le milieu 12 sur l'effectif total se trouve entre les milieux de tailles 157.5 et 162.5 ; alors la médiane doit appartenir à cet intervalle, soit $M_e \in [157.5; 162.5]$; donc on pourra utiliser cette formule de type Thalès

$$\frac{\frac{15}{24} - \frac{6}{24}}{162.5 - 157.5} = \frac{\frac{12}{24} - \frac{6}{24}}{M_e - 157.5}$$

il s'agit bien d'une équation dont l'inconnu est la médiane Me, et après la résolution on obtient $M_e = 160.8333$ et donc on se limite à la médiane $M_e = 160$.

Variance : On utilise la formule de Hygens

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{p} n_i X_i^2 - M_1^2$$

avec $n = 24$ et $p = 6$ classes, alors

$$\sigma^2 = \frac{1}{24} \left(6 \times 155^2 + 9 \times 160^2 + 5 \times 165^2 + 3 \times 170^2 + 1 \times 175^2\right) - 161.6667^2 = 30.5448$$

or l'écart-type est la racine carrée positive de la variance, alors

$$\sigma_1 = \sqrt{\sigma^2} = \sqrt{30.5448} = 5.5267$$

4. En comparant la moyenne et la médiane de cette série, on remarque que

$$\overline{X} = M_1 = 161.6667 > 160.8333 = M_e$$

alors on peut dire que la série statistique a un biais cognitif (série biaisée) vers le côté droit de la médiane, donc on conclut que la distribution de la série ne suit pas une loi normale.

5. On compare la distribution précédente à celle d'un autre groupe dont la moyenne et l'écart-type sont respectivement $M_2 = 165$ et $\sigma_2 = 5.6$.
D'abord, on remarque que $\sigma_1 = 5.5267 < 5.6 = \sigma_2$, alors dans la **globalité = sur la totalité des individus de la série**, la série 1 est dispersée d'une manière plus homogène que la deuxième car lorsque l'écart-type est petit, alors la série a une meilleure dispersion autour de sa moyenne.
Mais : comme $M_2 = 165 > 161.6667 = M_1$, alors les effectifs partiels changent, donc il se peut que si on fait une mesure par prise d'échantillon (division de la série sur des échantillons), dans ce cas on pourra dire que la deuxième est meilleure. Le deuxième cas reste relatif car on n'a plus de valeurs d'effectifs partiels pour savoir plus sur la série 2, c'est à dire qu'aucun signe de réalité derière.

$\square$