

TD1: Statistique descriptive univariée

Pr. L. A. Allamy

Université Cadi Ayyad
Faculté des Sciences Semlalia Marrakech

6 février 2021

Exercice 1 :

Classer les variables ci-dessous selon leur type (nature) :
Poids, Couleur des yeux, Nombre de frères et soeurs, Marque de PC portable, Profession, Nationalité, Taille, Mention, Revenu mensuel, Langue maternelle.

Qualitatif nominal : Couleur des yeux, Marque de PC portable, Profession, Nationalité, Langue maternelle.

Qualitatif ordinal : Mention.

Quantitatif discret : Nombre de frères et soeurs.

Quantitatif continu : Taille, Revenu mensuel.

Exercice 2 :

On s'intéresse à la quantité de livres lus dans l'année, par un groupe d'étudiants

($A = \text{peu}$, $B = \text{moyen}$, $C = \text{beaucoup}$, $D = \text{exceptionnel}$).

Soit la liste des prénoms des étudiants de ce groupe, suivis de l'indication de leur appétit de lecture :

Zineb (C), Youssef (C), Ali (A), Yassir (B), Amine (A), Catherine (B), Jean (C), Loubna (B), Leila (B), Fatima (C), Taha (D), Farida (B), Bouchra (A), Malak (C), Naima (C), Abdellah (C), Akram (C), Souad (D), Taoufiq (C), Yahya (C).

1. Quelle est la nature de la variable "Appétit de lecture" ?
2. Construire le tableau représentatif de cette distribution.
3. Représenter cette distribution à l'aide des diagrammes adéquats.

Exercice 2 : Solution

1. Nature de la variable “appétit de lecture” : qualitative ordinale.
2. Tableau représentatif de la distribution :

Modalités	Effectifs	Fréquences
Peu (A)	3	$3/20 = 0.15$
Moyen (B)	5	$5/20 = 0.25$
Beaucoup (C)	10	$10/20 = 0.50$
Exceptionnel (D)	2	$2/20 = 0.10$
Total	20	1

Diagramme en barres

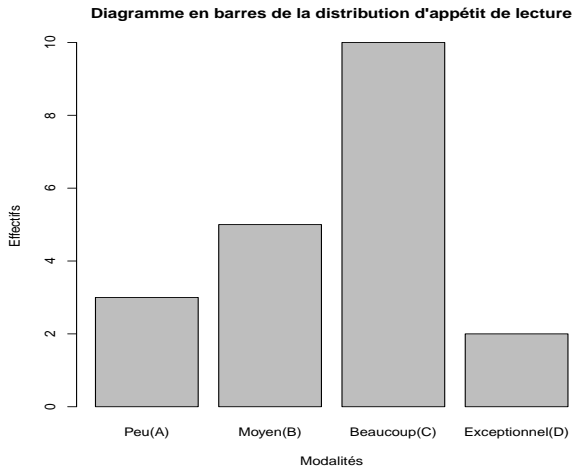
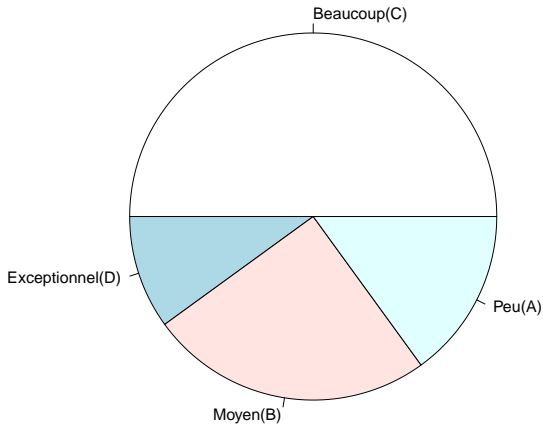


Figure 1 :

Diagramme en secteurs

Diagramme en secteurs de la distribution de l'appétit de lecture



Exercice 3 :

Exercice 3 : Dans une école, un questionnaire comprenant une question sur le nombre de frères et soeurs, a été rempli par une des classes. La collecte des données a fourni les résultats suivants :
1, 2, 1, 3, 1, 4, 2, 1, 3, 1, 2, 5, 2, 1, 1, 3, 2, 1, 2, 3, 1, 1, 1, 2, 4, 2, 1, 3.

- Représenter le tableau d'effectifs, associé à cette série.
- Calculer la moyenne et la médiane de cette série.
- Quels sont l'étendue et l'écart-type de cette distribution.
- Donner les trois quartiles Q_1 , Q_2 et Q_3 puis tracer le diagramme en boîte correspondant à cette série statistique.
- Existe-t-il des valeurs aberrantes ?

Solution de l'exercice 3

a. Tableau d'effectifs :

Valeurs	1	2	3	4	5	Total
Effectifs	12	8	5	2	1	28

b. La moyenne :

$$\bar{x} = \frac{12.1 + 8.2 + 5.3 + 2.4 + 1.5}{28} = \frac{56}{28} = 2.$$

La médiane :

$$Me = \frac{x_{(14)} + x_{(15)}}{2} = \frac{2 + 2}{2} = 2.$$

c. L'étendue :

$$e = 5 - 1 = 4.$$

La variance :

$$\frac{146}{28} - (2^2) = \frac{34}{28} = 1,21$$

d'où Ecart-type :

$$\sigma = \sqrt{1,21} = 1,10.$$

d. Calcul de Q_1 :

$$Q_1 = \frac{x_{(7)} + x_{(8)}}{2} = \frac{1 + 1}{2} = 1.$$

Calcul de Q_2 :

$$Q_2 = 2.$$

Calcul de Q_3 :

$$Q_3 = \frac{x_{(21)} + x_{(22)}}{2} = \frac{3 + 3}{2} = 3.$$

On déduit de ce qui précède les bornes suivantes :

Frontière basse= $Fb = Q_1 - 1,5(Q_3 - Q_1) = 1 - 1,5(3 - 1) = -2$
et

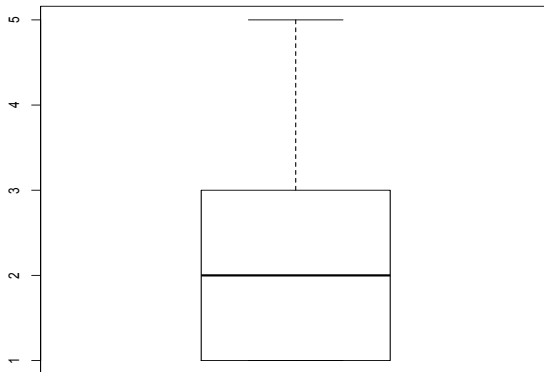
Frontière haute= $Fh = Q_3 + 1,5(Q_3 - Q_1) = 3 + 1,5(3 - 1) = 6$.

$b^- = \text{Inf}\{x_i/x_i \geq Fb\} = 1$.

$b^+ = \text{Sup}\{x_i/x_i \leq Fh\} = 5$.

e. Toutes les valeurs appartiennent à $[b^-; b^+]$ donc, **il n'y a aucune valeur aberrante.**

Diagramme en boîte du nombre de frères et soeurs



9. Indication sur le tracé, à la main, du diagramme en boîte :

On commence par tracer une droite horizontale avec une échelle allant de 0 à 20. La moyenne \bar{x} et les quartiles étant déjà calculés, on les place avec précision sur la droite .

La boîte sera le rectangle dont les deux côtés verticaux sont construits au niveau de Q_1 et Q_3 .

Pour la compléter, on doit calculer les frontières basse (Fb) et haute (Fh) dont les formules sont données par :

$$Fb = Q_1 - 1,5(IQ) = -2 \text{ et } Fh = Q_3 + 1,5(IQ) = 6.$$

On déduit alors les bornes b^- et b^+ des moustaches, données par :

$$b^- = \inf\{x_i / x_i \geq Fb\} = 1.$$

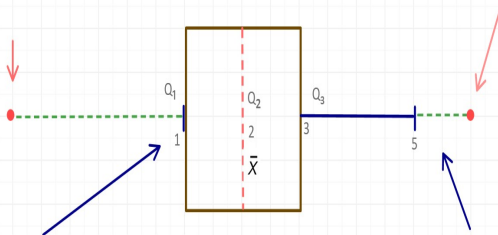
$$b^+ = \sup\{x_i / x_i \leq Fh\} = 5.$$

On place alors les moustaches de la boîte qui vont de b^- à Q_1 et de Q_3 à b^+ .

Boîte à moustaches réalisée à la main

$$Fb = Q_1 - 1.5 \times (Q_3 - Q_1) = Q_1 - 1.5(IQ)$$

$$Fh = Q_3 + 1.5 \times (Q_3 - Q_1) = Q_3 + 1.5(IQ)$$



$$b^- = \inf\{x_i \mid x_i \geq Fb\} = 1$$

$$b^+ = \sup\{x_i \mid x_i \leq Fh\} = 5.$$

Exercice 4 :

La taille en centimètres (cm) de 60 étudiants d'une école d'ingénieurs a été relevée. La collecte des données a fourni les résultats suivants :

167	169	171	172	167	171	169	171	172	165	173
174	168	164	170	175	173	169	173	170	171	176
172	163	171	171	174	168	163	165	168	168	173
170	168	171	170	164	170	172	172	167	170	169
167	171	173	168	167	177	171	172	159	169	165

1. Quelle est la population concernée par l'étude ? Quel est l'échantillon ? Quelle est la variable étudiée ? Donner sa nature.
2. Remplissez le tableau suivant en utilisant 5 classes de même amplitude (la borne inférieure de la 1^{ère} classe sera prise égale 155 et la borne supérieure de la dernière classe égale à 180).

X	n_i	N_i	c_i	a_i	h_i
[.....;[
[.....;[
[.....;[
[.....;[
[.....;[
[.....;[
Total					

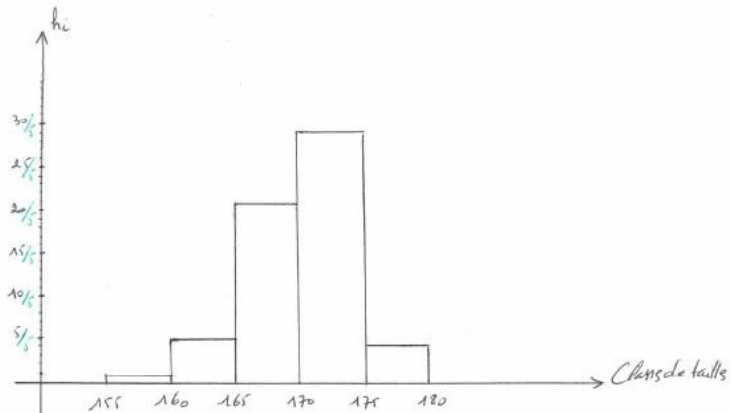
- ❶ Quelle est l'étendue de cette distribution ?
- ❷ Tracer l'histogramme correspondant à ces données. Y a-t-il une classe modale ?
- ❸ Quelle est la valeur approchée de la moyenne des tailles ?
- ❹ Quel est l'écart-type des tailles ?
- ❺ Quelle est la médiane des tailles ?
- ❻ Donner des valeurs approchées des deux autres quartiles. En déduire l'écart interquartile ?
- ❼ Dessiner le diagramme en boîte correspondant à cette distribution.
- ❽ Y a-t-il des valeurs aberrantes ?

- ① **Population** : Les étudiants d'une école d'ingénieurs.
- ② **Échantillon** : 60 étudiants de la même école d'ingénieurs.
- ③ **Caractère** : La taille.
- ④ **Nature** : Quantitative continue.

Remarque :

- la borne inférieure de la première classe étant 155, la borne supérieure de la dernière classe étant 180, L'étendue devient $180 - 155 = 25$, puisqu'on voudrait avoir 5 classes de même amplitude, l'amplitude serait : $\frac{25}{5} = 5$.

Moyenne semestrielles	Nombre d'étudiants n_i	N_i	a_i	c_i	$h_i = \frac{n_i}{a_i}$
[155; 160[1	1	5	157.5	1/5
[160; 165[5	6	5	162.5	1
[165; 170[21	27	5	167.5	21/5
[170; 175[29	56	5	172.5	29/5
[175; 180[4	60	5	177.5	4/5



On remarque que la classe modale est la classe $[170; 175[$, qui correspond à la hauteur la plus élevée.

5. $\bar{x} = 170$.

6. $s_X = ?$

$$\text{Variance}(X) = s_X^2 = 16.25 \Rightarrow s_X = \sqrt{16.25} \simeq 4.03.$$

7. Médiane ?

$n/2 = 60/2 = 30$. En regardant les effectifs cumulés, on trouve que :

$$Me \in [170; 175[.$$

$$\frac{Q_2 - 170}{30 - 27} = \frac{175 - 170}{56 - 27} \Rightarrow Q_2 = Me \simeq 170.517.$$

8. Q_1 et Q_3 ?

$Q_1 : \frac{60}{4} = 15$, 15 est atteint dans la classe $[165, 170[$

$\Rightarrow Q_1 \in [165, 170[$

$$\frac{Q_1 - 165}{15 - 6} = \frac{170 - 165}{27 - 6}$$

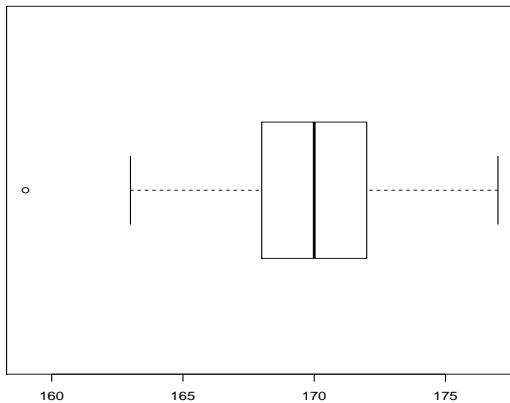
$Q_1 \simeq 167.143$

Q_3 : correspond à un effectif cumulé de $\frac{3n}{4} = 45$, 45 est atteint dans la classe $[170, 175[$

$$Q_3 = 170 + (45 - 27) \frac{175 - 170}{56 - 27} \simeq 173.103$$

L'écart-interquartile est $IQ = Q_3 - Q_1 = 173.103 - 167.43 \simeq 5,96$.

Diagramme en Boîte des tailles



Calcul des différentes bornes à la main

$$* Fb = Q_1 - 1,5(IQ) = 167.143 - 1.5(5.96) = 158.203,$$

$$* Fh = Q_3 + 1,5(IQ) = 173.103 + 1.5(5.96) = 182.043,$$

$$* b^- = \text{Inf}\{x_i / x_i \geq Fb\} = Fb = 158.203.$$

$$* b^+ = \text{Sup}\{x_i / x_i \leq Fh\} = 180.$$

Remarque :

Les valeurs entre 155 et 158,203 sont des valeurs aberrantes. !!!

$$Fb = Q_1 - 1.5 \times (Q_3 - Q_1) = Q_1 - 1.5(IQ)$$

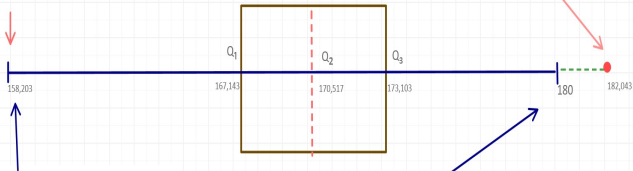


$$b^- = \inf\{x_i \mid x_i \geq Fb\}$$

$$Fh = Q_3 + 1.5 \times (Q_3 - Q_1) = Q_3 + 1.5(IQ)$$



$$b^+ = \sup\{x_i \mid x_i \leq Fh\}$$



Exercice 5

Dans le cadre d'une étude sur le temps X de mémorisation (en minutes), on a choisi au hasard, un échantillon de 28 élèves qu'on a soumis à un test. Les résultats obtenus sont donnés dans le tableau suivant :

X	n_i	N_i	c_i	a_i	h_i
[3; 4[3				
[4; 5[7				
[5; 6[5				
[6; 11[5				
[11; 13[6				
[13; 17[2				
Total					

Exercice 5

1. Quelle est la population étudiée ? Quelle est la taille de l'échantillon ? Préciser la variable étudiée et son type .
2. Compléter le tableau ci-dessus puis dessiner le diagramme correspondant à cette distribution. Que constate-t-on ?
3. Calculer la moyenne, la variance et l'écart-type de X .
4. Déterminer la médiane. Que représente cet indice ? Expliquer brièvement pourquoi il y a une différence non négligeable entre la moyenne et la médiane.

Solution de l'exercice 5

1. Population : Les élèves.

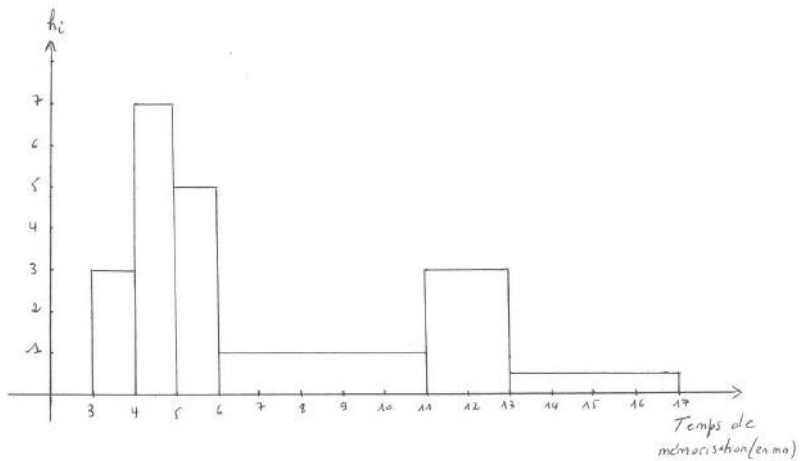
Taille de l'échantillon : 28 élèves.

Variable : Temps de mémorisation.

Type : Quantitatif continu.

2. Tableau complété :

X	n_i	N_i	c_i	a_i	$h_i = n_i/a_i$
$[3; 4[$	3	3	3,5	1	3
$[4; 5[$	7	10	4,5	1	7
$[5; 6[$	5	15	5,5	1	5
$[6; 11[$	5	20	8,5	5	1
$[11; 13[$	6	26	12	2	3
$[13; 17[$	2	28	15	4	0,5
Total	28				



On constate qu'il existe deux modes, un principal et relatif (secondaire).

3. $\bar{x} = 214/28 \simeq 7,64$.

$$s_X^2 = \frac{1}{n} \sum n_i c_i^2 - (\bar{x})^2 \simeq 13,194 \Rightarrow s_x \simeq 3,63.$$

4. $Me \in [5; 6[$. De la même manière on détermine $Me \simeq 5,8$

La médiane représente la valeur de la variable qui partage la série en 2 parties de même taille ou effectif.

BON COURAGE