

Correction de TD N° 1 de statistique descriptive

Exercice 1 :

1. **Nom et prénom** : qualitatif nominal
2. **Sexe (1 pour masculin, 0 pour féminin)** : qualitatif nominal
3. **Age (ans)** : quantitatif continu
4. **Profession** : qualitatif ordinal
5. **Nombre d'incidents cardiaques antérieurs** : quantitatif discret
6. **Taille (en cm)** : quantitatif continu
7. **Poids (en kg)** : quantitatif continu
8. **Cholestérol (en g/l)** : quantitatif continu

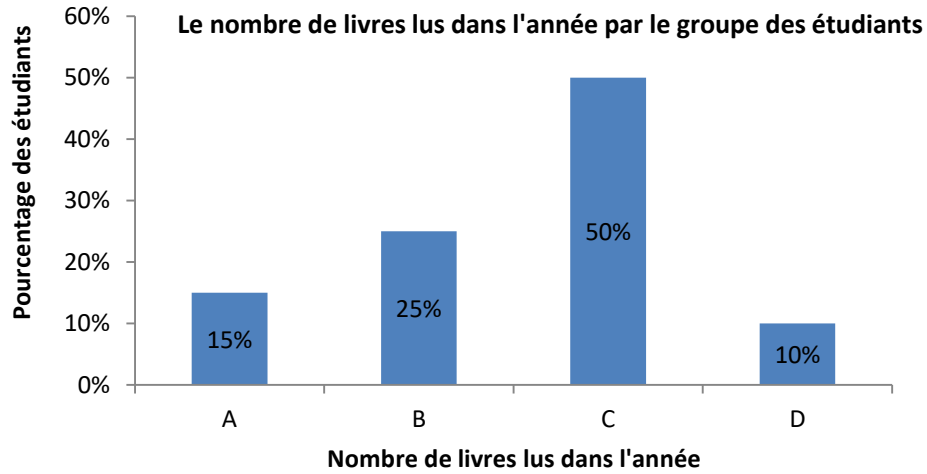
Exercice 2 :

1. **La population** c'est un groupe d'étudiants. **La taille de l'échantillon** c'est $n=20$. **Le caractère** c'est le nombre de livres lus dans l'année. **La nature de caractère** c'est qualitatif ordinal. **Les modalités de caractère** sont (A=peu, B=moyen, C=beaucoup, D=exceptionnel).
- 2.

Tableau représentatif de la distribution

| Le nombre de livres lus dans l'année | A | B | C | D | Total |
|--------------------------------------|-----|-----|-----|-----|-------|
| Effectif n_i | 3 | 5 | 10 | 2 | 20 |
| Fréquence en % | 15% | 25% | 50% | 10% | 100% |

3.



Exercice 3 :

1. **Population :** femmes de 40 ans de la république

Taille de la population : $n=100$

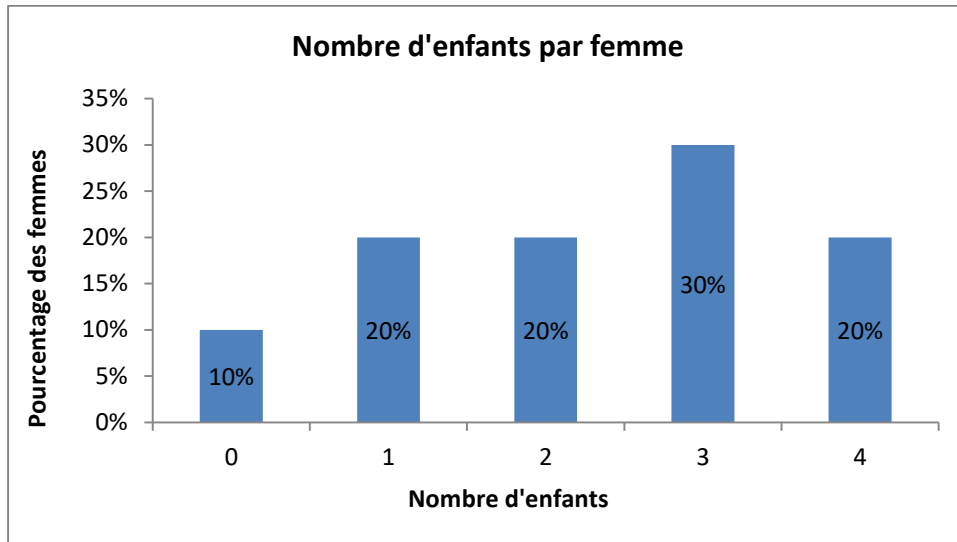
Le caractère : Nombre d'enfants par femme (x_i)

Nature de caractère : quantitatif ordinal

Les modalités de caractère : 0, 1, 2, 3 et 4

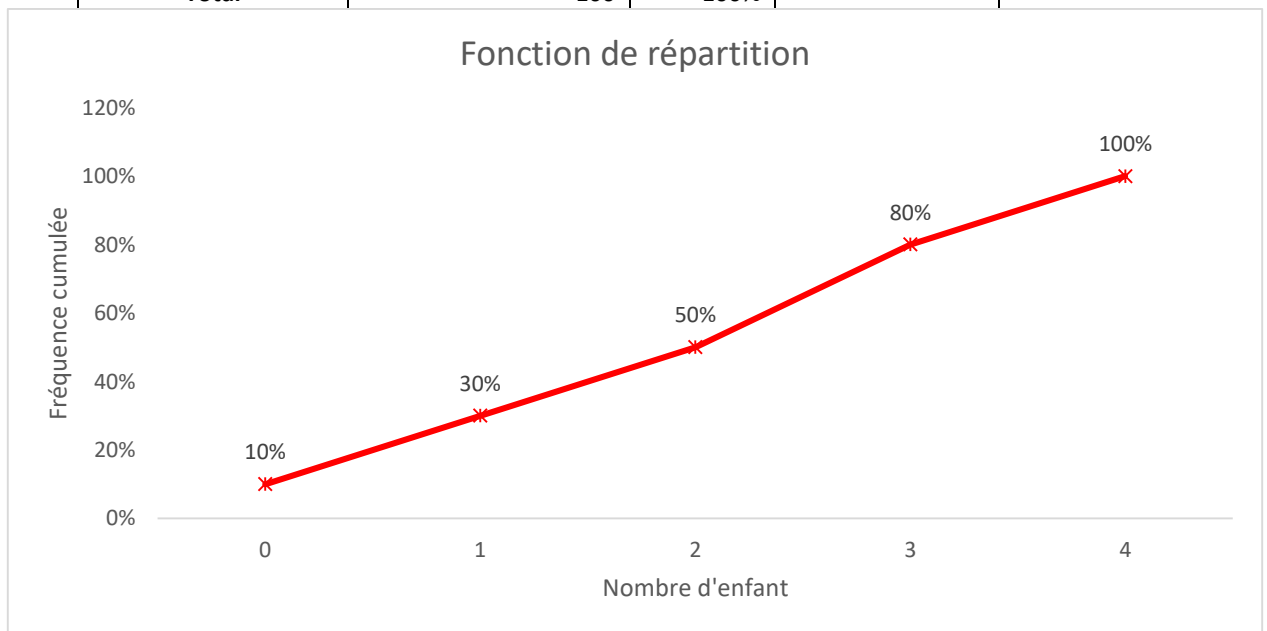
2.

| Nombre d'enfants (x_i) | Nombre de femmes (n_i) | Fréquence en % |
|----------------------------|----------------------------|----------------|
| 0 | 10 | 10% |
| 1 | 20 | 20% |
| 2 | 20 | 20% |
| 3 | 30 | 30% |
| 4 | 20 | 20% |
| Total | 100 | 100% |



3.

| Nombre d'enfants (x_i) | Nombre de femmes (n_i) | Fréquence en % | Fréquence cumulée |
|----------------------------|----------------------------|----------------|-------------------|
| 0 | 10 | 10% | 10% |
| 1 | 20 | 20% | 30% |
| 2 | 20 | 20% | 50% |
| 3 | 30 | 30% | 80% |
| 4 | 20 | 20% | 100% |
| Total | 100 | 100% | |



4. La proportion des femmes ayant moins de 4 enfants c'est 80%.

5.

La distribution des enfants selon le nombre de leurs frères et soeurs

| Catégories | Effectifs |
|---|-----------|
| Les enfants qui n'ont pas de frères et soeurs | 20 |
| Les enfants qui ont un frère ou une sœur | 40 |
| Les enfants qui ont deux frères et sœurs | 90 |
| Les enfants qui ont trois frères et sœurs | 80 |
| Les enfants qui ont plus de trois frères et sœurs | 0 |

Exercice 4 :

1.

Population : 40 entreprises industrielles

Unité statistique (individu): une entreprise industrielle

Le caractère : le nombre de salariés

La nature de caractère : quantitatif discret

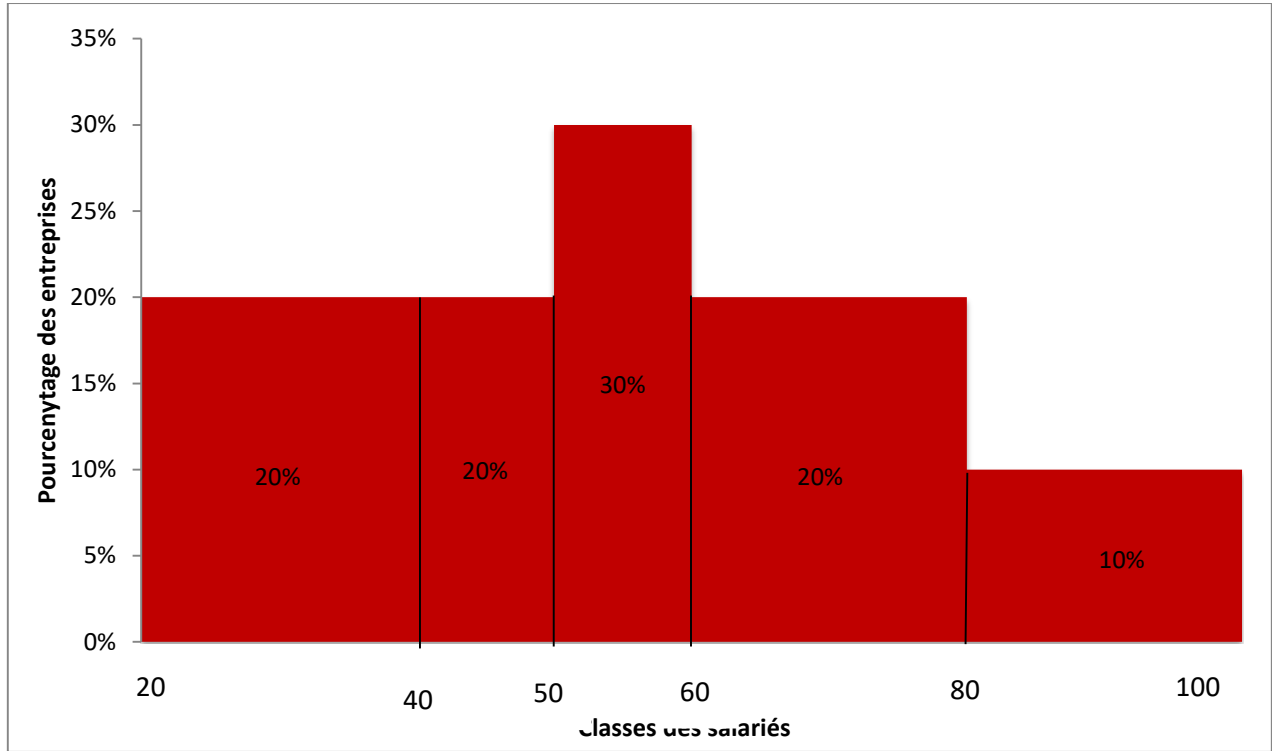
Les modalités de caractère : tous les entiers naturels

2.

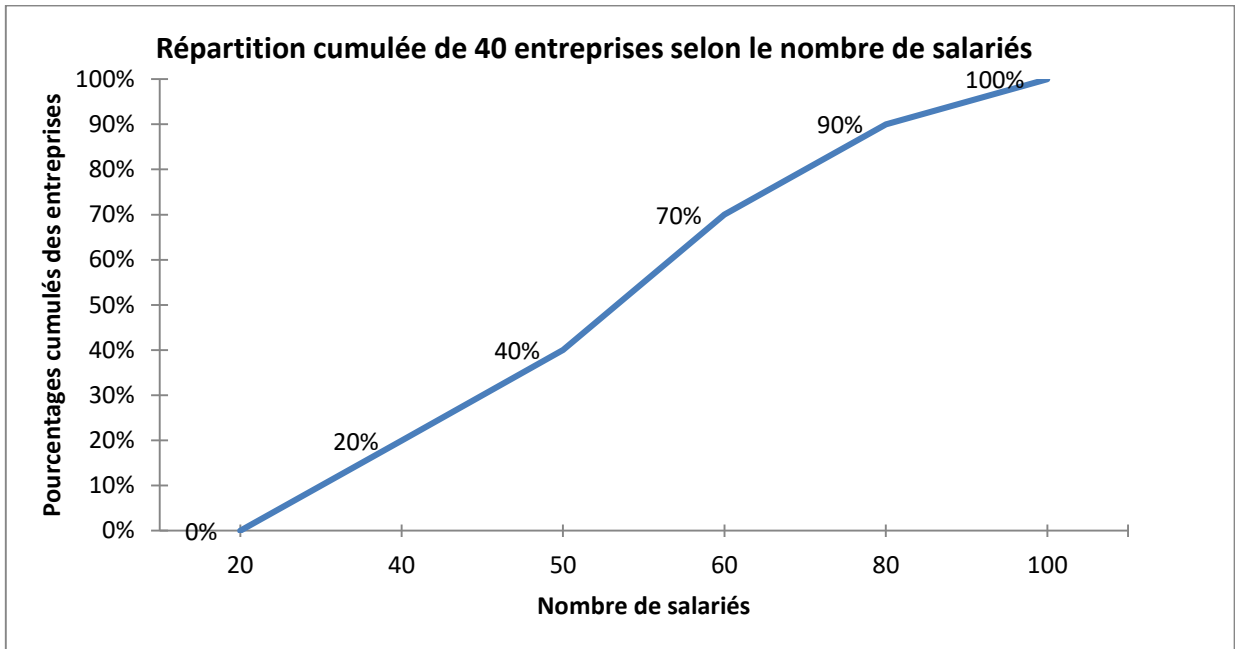
| Classes | [20,40[| [40,50[| [50,60[| [60,80[| [80,100] | Total |
|--------------------------------|---------|---------|---------|---------|----------|-------|
| Effectifs n_i | 8 | 8 | 12 | 8 | 4 | 40 |
| Fréquence f_i | 20% | 20% | 30% | 20% | 10% | 100% |
| Fréquence cumulé F_i en % | 20% | 40% | 70% | 90% | 100% | |

3.

Histogramme des fréquences de la distribution des classes des salariés



4.



Exercice 5 :

| | Etude 1 | Etude 2 | Etude 3 |
|---|--|---|-------------------------|
| Population : | Les ménages | Les citoyens marocains | Les étudiants marocains |
| Caractère : | La consommation annuelle d'électricité | L'âge des citoyens | le goût musical |
| Type de caractère : | Quantitatif continue | Quantitatif continue | Qualitatif nominal |
| Echantillon prélevé pour effectuer l'étude : | 5000 ménages | Tous les citoyens car Il s'agit d'un recensement. | 100 étudiants marocains |

Exercice 6 :

1. Oui par un diagramme en bâton

| | | | | | | |
|----------------------------------|---|---|----|----|----|----|
| Note | 8 | 9 | 10 | 13 | 14 | 20 |
| Effectif n_i | 3 | 2 | 5 | 3 | 2 | 1 |

Oui on peut aussi utiliser un diagramme en secteurs circulaires ou le graphique en tuyaux d'orgue.

2.

| | | | | | | |
|--|--------|--------|--------|--------|--------|---------|
| Note | 8 | 9 | 10 | 13 | 14 | 20 |
| Effectif n_i | 3 | 2 | 5 | 3 | 2 | 1 |
| Effectif Cumulé N_i | 3 | 5 | 10 | 13 | 15 | 16 |
| Fréquence f_i en % | 18,75% | 12,50% | 31,25% | 18,75% | 12,50% | 6,25% |
| Fréquence cumulée F_i en % | 18,75% | 31,25% | 62,50% | 81,25% | 93,75% | 100,00% |

La proportion des candidats ayant au plus 13 est 81,25%.

3. La proportion des candidats ayant plus de 9 est :

$$P(X \geq 9) = 100\% - P(X < 9) = 100\% - 18,75\% = 81,25\%.$$

4. Le mode de cette distribution est 10. La médiane c'est 10. La moyenne c'est

$$m = \frac{8 \times 3 + 9 \times 2 + 10 \times 5 + 13 \times 3 + 14 \times 2 + 20 \times 1}{16} = 11,19$$

5. Si on supprime l'observation 20

| | | | | | |
|--|--------|--------|--------|--------|---------|
| Note | 8 | 9 | 10 | 13 | 14 |
| Effectif n_i | 3 | 2 | 5 | 3 | 2 |
| Effectif Cumulé N_i | 3 | 5 | 10 | 13 | 15 |
| Fréquence f_i en % | 20,00% | 13,33% | 33,33% | 20,00% | 13,33% |
| Fréquence cumulée F_i en % | 20,00% | 33,33% | 66,67% | 86,67% | 100,00% |

Le mode de la nouvelle distribution est 10. La médiane est 10. La moyenne c'est

$$m = \frac{8 \times 3 + 9 \times 2 + 10 \times 5 + 13 \times 3 + 14 \times 2}{15} = 10,6$$

Commentaire : La médiane et le mode ne sont pas influencés par les observations extrêmes. Alors qu'ils y en a des influences par les observations extrêmes dans les cas de la moyenne.

Exercice 7

1.

Population : Les élèves ayant plus de 18 ans d'une certaine école

Echantillon : 200 élèves ayant plus de 18 ans de l'école

Taille de l'échantillon : $n=200$ élèves

Le caractère : La taille des élèves

La nature de caractère : quantitatif discret

2. Distribution en terme d'effectif

| | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 135 | 143 | 145 | 146 | 147 | 148 | 149 | 150 | 151 | 152 | 153 | 154 | 155 |
| 1 | 2 | 2 | 3 | 2 | 3 | 4 | 2 | 5 | 2 | 4 | 7 | 9 |
| 156 | 157 | 158 | 159 | 160 | 161 | 162 | 163 | 164 | 165 | 166 | 167 | 168 |
| 4 | 2 | 7 | 3 | 3 | 1 | 11 | 8 | 2 | 10 | 12 | 7 | 8 |
| 169 | 170 | 171 | 172 | 173 | 174 | 175 | 176 | 177 | 178 | 179 | 180 | 181 |
| 8 | 10 | 6 | 6 | 7 | 3 | 2 | 9 | 4 | 4 | 1 | 8 | 1 |
| 183 | 184 | 186 | 192 | | | | | | | | | |
| 1 | 3 | 2 | 1 | | | | | | | | | |

3.

$$\text{Amplitude de classe} = (193-130)/9=7$$

Regroupement des données par classes de même amplitude

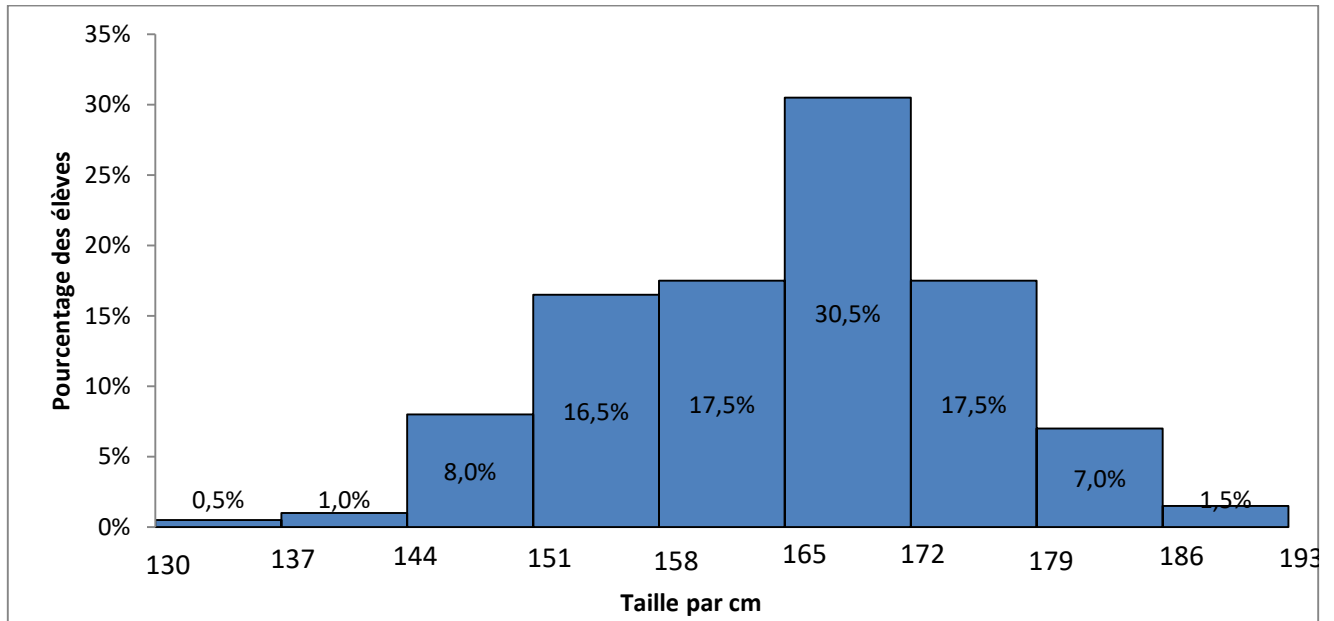
| | | | | | | | | |
|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| [130,137[| [137,144[| [144,151[| [151,158[| [158,165[| [165,172[| [172,179[| [179,186[| [186,193[|
| 1 | 2 | 16 | 33 | 35 | 61 | 35 | 14 | 3 |

4.

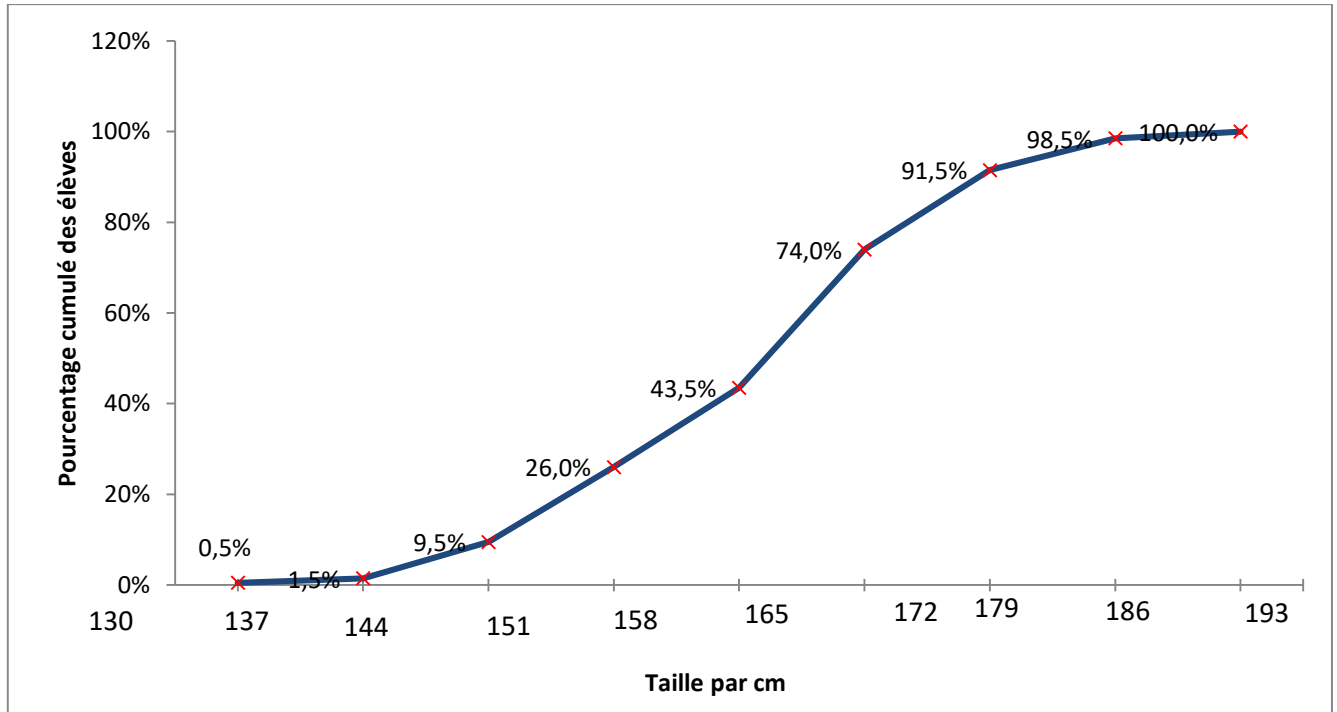
Tableau des fréquences et fréquence cumulée

| Classes | [130,137[| [137,144[| [144,151[| [151,158[| [158,165[| [165,172[| [172,179[| [179,186[| [186,193[|
|-------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| Effectif | 1 | 2 | 16 | 33 | 35 | 61 | 35 | 14 | 3 |
| Fréquence | 0,5% | 1,0% | 8,0% | 16,5% | 17,5% | 30,5% | 17,5% | 7,0% | 1,5% |
| Fréquence cumulée | 0,5% | 1,5% | 9,5% | 26,0% | 43,5% | 74,0% | 91,5% | 98,5% | 100,0% |

Distribution des élèves par classes de taille en cm



Le polygone de fréquence

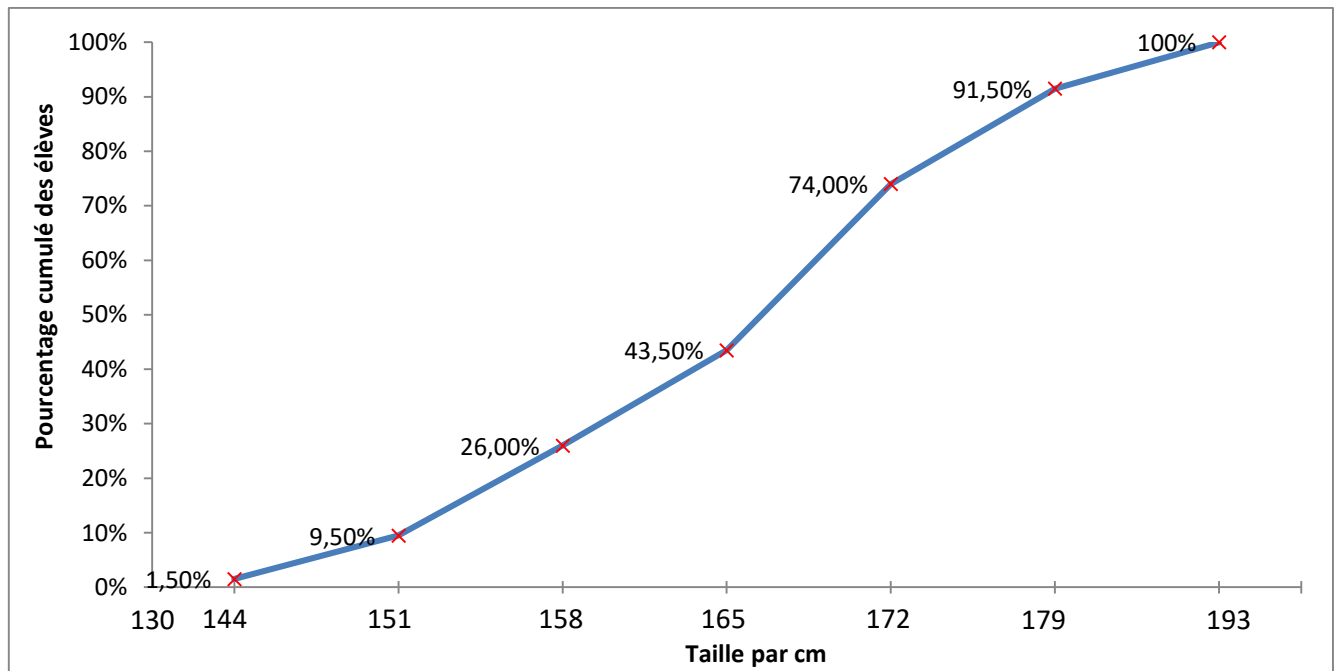


5.

Après regroupement des deux premières classes et les deux dernières on trouve le tableau suivant :

| Classes | [130,144[| [144,151[| [151,158[| [158,165[| [165,172[| [172,179[| [179,193] |
|---------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| Effectif | 3 | 16 | 33 | 35 | 61 | 35 | 17 |
| Fréquences | 1,5% | 8,0% | 16,5% | 17,5% | 30,5% | 17,5% | 8,5% |
| Fréquences cumulées | 1,5% | 9,5% | 26,0% | 43,5% | 74,0% | 91,5% | 100% |

Polygone de fréquence



6. À partir du tableau des fréquences cumulées

$$P(144 \leq X \leq 179) = 0,915 - 0,015 = 90\%$$

Ou bien

$$P(144 \leq X \leq 179) = \frac{\text{Effectif des élèves ayant une taille entre 144 et 179}}{\text{effectif total des élèves}} = \frac{16+33+35+61+35}{200} = 90\%$$

7.

La moyenne :

| Classes | [130,144[| [144,151[| [151,158[| [158,165[| [165,172[| [172,179[| [179,193] |
|------------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| Effectif | 3 | 16 | 33 | 35 | 61 | 35 | 17 |
| Centre de classe c_i | 138,5 | 147,5 | 154,5 | 161,5 | 168,5 | 175,5 | 186 |
| $C_i \cdot n_i$ | 415,5 | 2360 | 5098,5 | 5652,5 | 10278,5 | 6142,2 | 3162 |

$$m = \sum_{i=1}^n \frac{c_i \times n_i}{N}$$

Avec

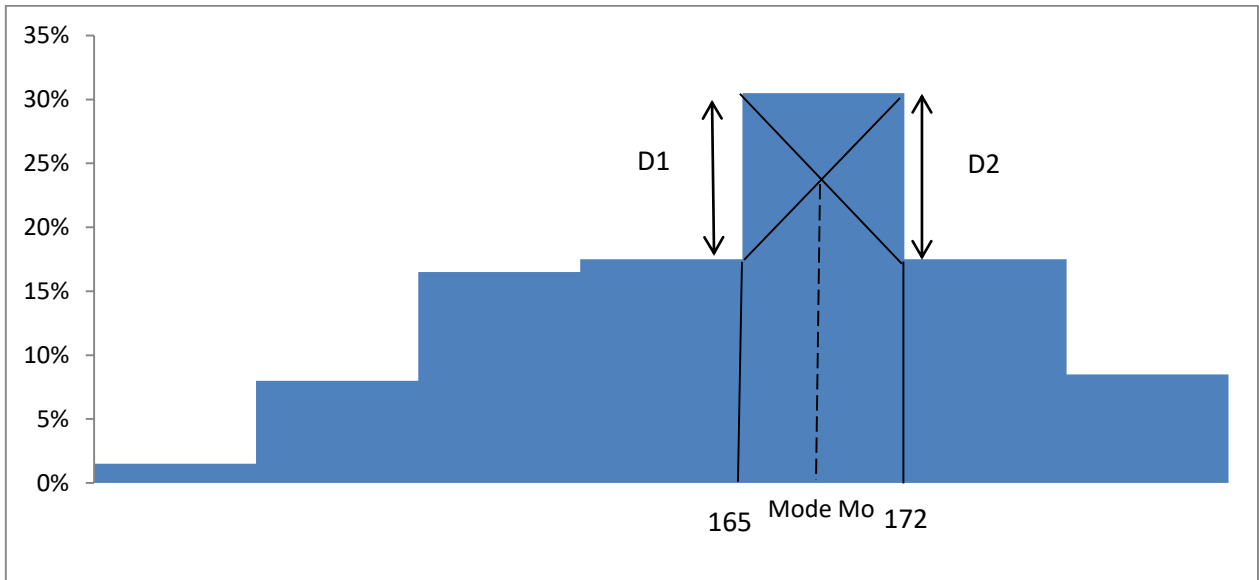
$$N = \sum_{i=1}^n n_i$$

$$m = \frac{415,5 + 2360 + 5098,5 + 5652,5 + 10278,5 + 6142,2 + 3162}{3 + 16 + 33 + 35 + 61 + 35 + 17} = 165,546$$

Le mode :

La classe modale c'est [165,172[

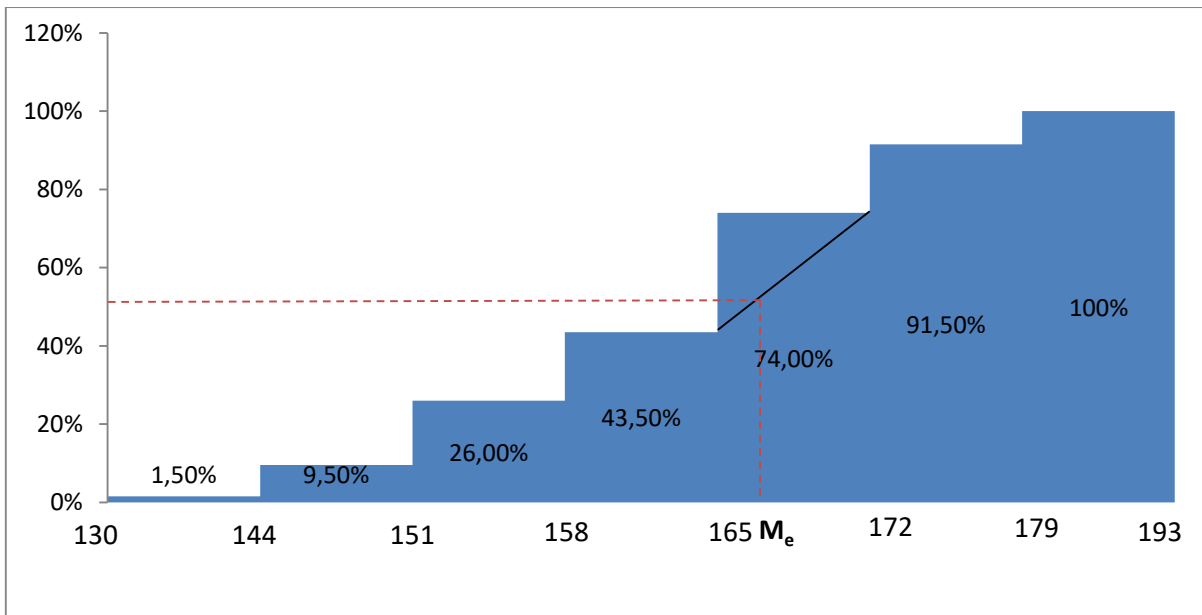
| Classes | [130,144[| [144,151[| [151,158[| [158,165[| [165,172[| [172,179[| [179,193] |
|---------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| Effectif | 3 | 16 | 33 | 35 | 61 | 35 | 17 |
| Fréquences | 1,5% | 8,0% | 16,5% | 17,5% | 30,5% | 17,5% | 8,5% |
| Fréquences cumulées | 1,5% | 9,5% | 26,0% | 43,5% | 74,0% | 91,5% | 100% |



$$M_0 = x_i + \frac{D1}{D1 + D2} (x_{i+1} - x_i)$$

$$M_0 = 165 + \frac{61 - 35}{(61 - 35) + (61 - 35)} (172 - 165) = 168,5$$

La médiane :



$$\frac{172 - 165}{74 - 43,5} = \frac{M_e - 165}{50 - 43,5}$$

$$\frac{7}{30,5} = \frac{M_e - 165}{6,5}$$

Alors

$$M_e = 166,49$$

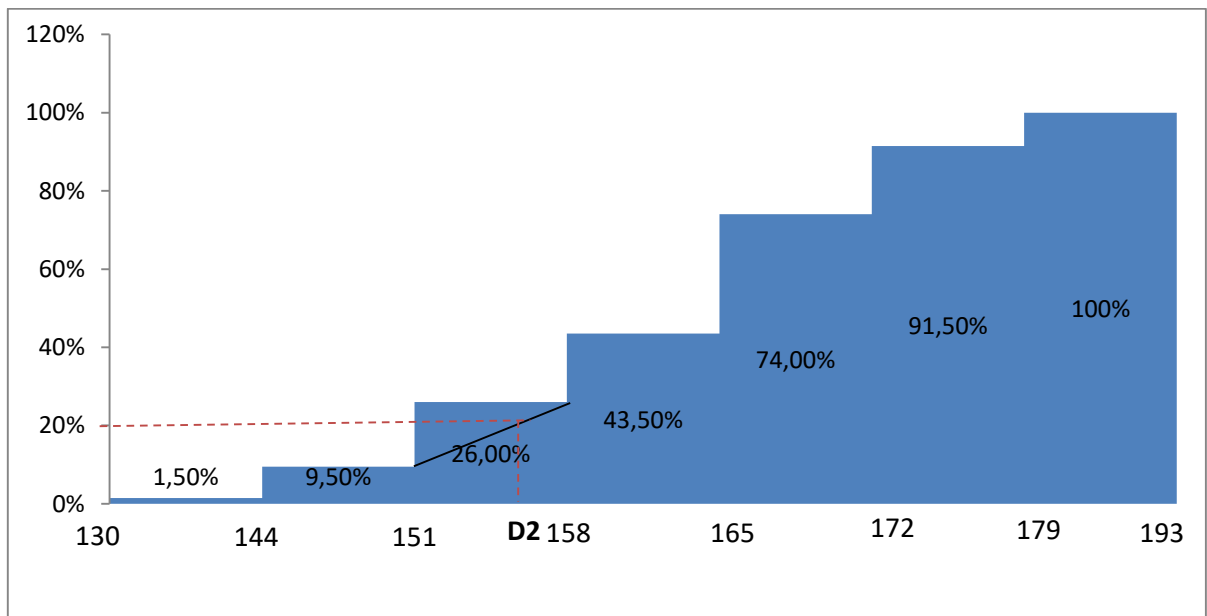
8.

- L'étendue

$$E = X_{\max} - X_{\min} = 192 - 135 = 57$$

- **Le deuxième décile**

Le cinquième décile c'est la médiane M_e . On calcule par, exemple, le deuxième décile D_2 .



$$\frac{158 - 151}{26 - 9,5} = \frac{D2 - 151}{20 - 9,5}$$

$$\frac{7}{16,5} = \frac{D2 - 151}{10,5}$$

Alors

$$D2 = 155,45$$

- **L'écart type :**

| Classes | [130,144[| [144,151[| [151,158[| [158,165[| [165,172[| [172,179[| [179,193] | Total |
|------------------------|-----------|-----------|-----------|-----------|------------|------------|-----------|------------|
| Effectif n_i | 3 | 16 | 33 | 35 | 61 | 35 | 17 | |
| Centre de classe c_i | 138,5 | 147,5 | 154,5 | 161,5 | 168,5 | 175,5 | 186 | |
| $c_i^2 * n_i$ | 57546,75 | 348100 | 787718,25 | 912878,25 | 1731927,25 | 1078008,75 | 588132 | 5504311,25 |

On a la variance est :

$$\begin{aligned}
 s_x^2 &= \frac{1}{n} \sum_{i=1}^q n_i (c_i - \bar{x})^2 \\
 &= \left(\frac{1}{n} \sum_{i=1}^q n_i c_i^2 \right) - \bar{x}^2 \\
 &= 27521,56 - 27405,48 = 116,08
 \end{aligned}$$

Alors, L'écart type c'est : $s_x = 10,77$