

Statistiques

Exemple d'étude d'une série statistique

Une compagnie de taxis a relevé les distances parcourues (en milliers de km) par ses véhicules avant qu'elle ne s'en sépare. Les résultats sont consignés dans le tableau suivant :

Distance	[60;70[[70;80[[80;90[[90;100[[100;110[[110;120[[120;130[[130;140[[140;150[[150;160[
Effectif	6	10	15	21	25	38	18	12	4	1

Nous allons envisager deux méthodes d'étude statistique de ces résultats, c'est à dire deux façons de les résumer.

A. Médiane et quartiles

1- Définitions

La médiane est la valeur du caractère étudié pour laquelle 50% de la population a une valeur inférieure et 50% de la population a une valeur supérieure.

Le 1^{er} quartile est la valeur du caractère étudié pour laquelle 25% de la population a une valeur inférieure et 75% de la population a une valeur supérieure.

Le 3^{ème} quartile est la valeur du caractère étudié pour laquelle 75% de la population a une valeur inférieure et 25% de la population a une valeur supérieure.

2- Méthode de détermination

On commence par calculer les effectifs cumulés, puis les fréquences cumulées.

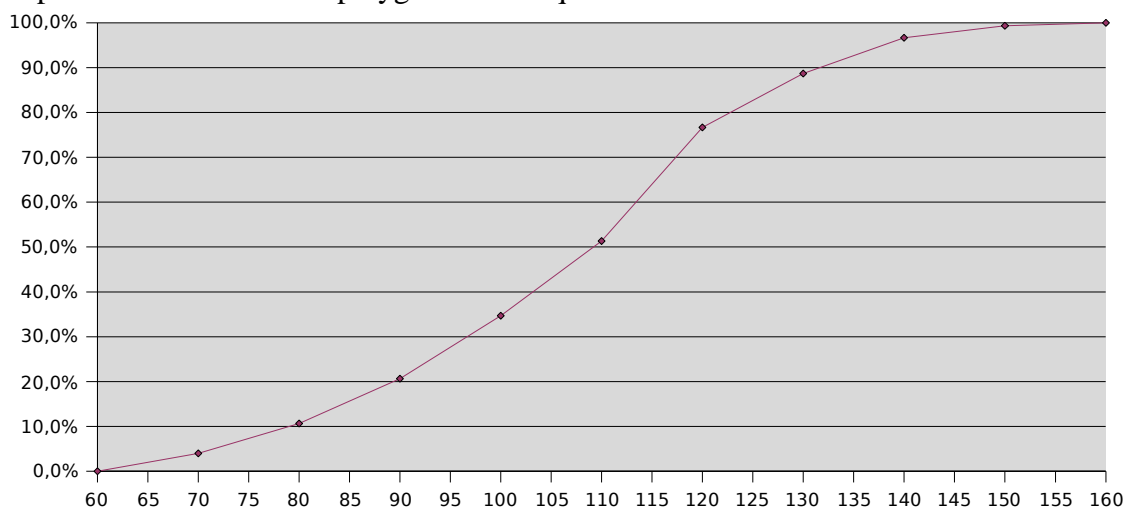
L'effectif cumulé correspondant à une valeur du caractère est l'effectif de la population qui a une valeur inférieure.

On obtient :

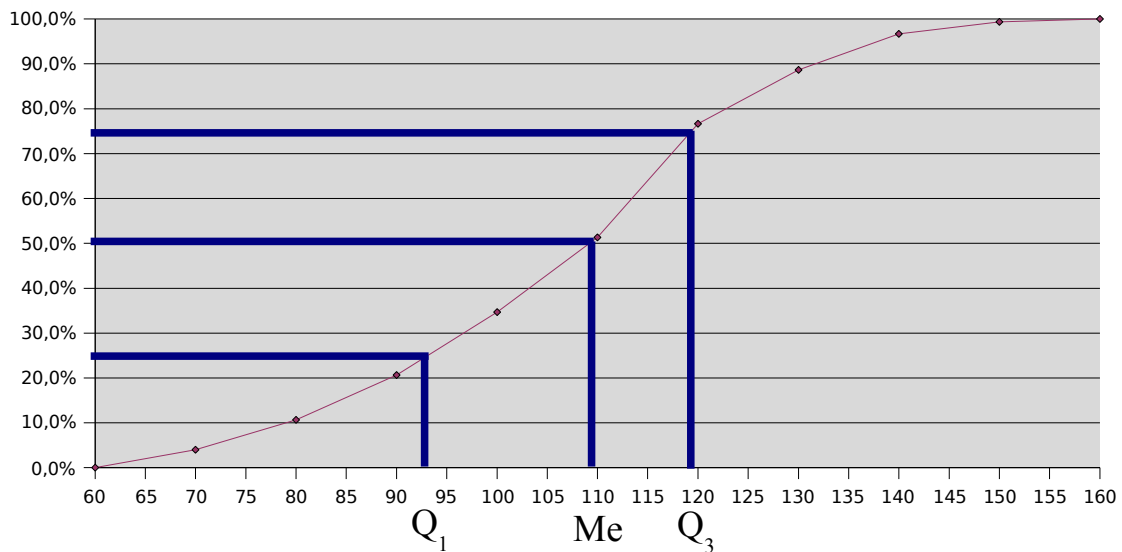
Distance	60	70	80	90	100	110	120	130	140	150	160
Eff. Cum.	0	6	16	31	52	77	115	133	145	149	150
Fréq. Cum.	0,0%	4,0%	10,7%	20,7%	34,7%	51,3%	76,7%	88,7%	96,7%	99,3%	100,0%

La dernière colonne de l'effectif cumulé donne l'effectif total de la population qui permet de calculer les fréquences cumulées.

On peut alors construire le polygone des fréquences cumulées croissantes :



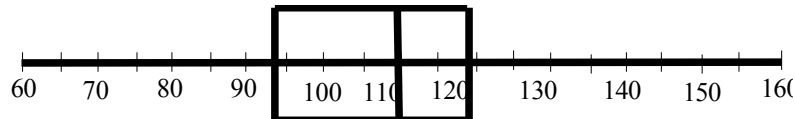
On utilise ce graphique pour déterminer le 1er quartile, la médiane et le 3ème quartile qui correspondent à des fréquences cumulées de 25%, 50% et 75%.



On lit ainsi :

- le 1^{er} quartile est $Q_1 \approx 93$
- la médiane est $Me \approx 110$
- le 3^{ème} quartile est $Q_3 \approx 119$

On peut résumer cette étude à l'aide d'un diagramme en moustache qui donne les valeurs extrêmes, les quartiles et la médiane.



B. Moyenne et écart-type

1- Définitions

On considère une série statistique x_i , et on appelle n_i l'effectif associé au caractère x_i .

a) Effectif total

L'effectif total de la population est $n = \sum n_i$

b) Moyenne

La moyenne est $\bar{x} = \frac{\sum n_i x_i}{n}$.

Si on représente les x_i par des points M_i sur un axe, la moyenne est l'abscisse du barycentre des points M_i affectés des coefficients n_i .

c) Variance

La variance sert à évaluer la dispersion des x_i autour de la moyenne. Elle donne la moyenne

des carrés des écarts à la moyenne, elle est égale à $V = \frac{\sum n_i (x_i - \bar{x})^2}{n}$.

On montre que la variance est égale à la différence entre la moyenne des carrés et le carré de la moyenne, soit $V = \frac{\sum n_i x_i^2}{n} - \bar{x}^2$.

d) Ecart-type

L'écart-type est la racine carrée de la variance; $\sigma = \sqrt{V}$. C'est une bonne unité pour mesurer l'écart à la moyenne.

2- Méthode de calcul

Dans les données étudiées le caractère est fourni sous forme d'intervalle. Nous commençons par remplacer chaque intervalle par une valeur unique, le milieu de l'intervalle.

Nous calculons ensuite les $n_i x_i$ et les $n_i x_i^2$.

x_i	65	75	85	95	105	115	125	135	145	155	Totaux
n_i	6	10	15	21	25	38	18	12	4	1	150
$n_i x_i$	390	750	1275	1995	2625	4370	2250	1620	580	155	16010
$n_i x_i^2$	25350	56250	108375	189525	275625	502550	281250	218700	84100	24025	1765750

La dernière colonne donne les totaux, c'est à dire $\sum n_i$, $\sum n_i x_i$ et $\sum n_i x_i^2$.

Nous pouvons en déduire :

L'effectif total : $n = \sum n_i = 150$

La moyenne : $\bar{x} = \frac{\sum n_i x_i}{n} = \frac{16010}{150} \approx 106,7$

La variance : $V = \frac{\sum n_i x_i^2}{n} - \bar{x}^2 = \frac{1765750}{150} - \left(\frac{16010}{150}\right)^2 \approx 379,7$

L'écart-type : $\sigma = \sqrt{V} \approx 19,5$