

Statistiques

I) Introduction.

Lorsque l'on étudie une série statistique, on étudie une population composée d'individus. Les données peuvent avoir plusieurs caractères :

- caractère quantitatif discret : lorsque la série prend des valeurs isolées entières 0 ; 1 ; 2 ; ... (par exemple le nombre d'enfant dans une famille ou l'âge des individus).
- caractère quantitatif continu : lorsque la série peut prendre toutes les valeurs comprises entre deux bornes (par exemple la taille des individus ou le temps passé devant le télé).
- caractère qualitatif : lorsque la série ne prend pas des valeurs numériques (par exemple la couleur préférée ou le sport pratiqué comme loisir).

Voyons pour chacun des cas, sur des exemples, comment organiser les données, comment les représenter graphiquement et quels calculs l'on peut éventuellement faire.

II) Gestion des données à caractère quantitatif discret.

On considère la série des notes obtenues à un contrôle dans une classe de 4^{ème} :

8 – 9 – 14 – 8 – 12 – 9 – 7 – 12 – 9 – 13 – 9 – 11 – 12 – 7 – 9 – 8 – 11 – 8 – 8 – 15 – 10 – 14 – 8 – 13 – 7

1) Organisation.

Notes	7	8	9	10	11	12	13	14	15	Total
Effectifs	3	6	5	1	2	3	2	2	1	25
Effectifs Cumulés	3	9	14	15	17	20	22	24	25	
Fréquences	0,12	0,24	0,2	0,04	0,08	0,12	0,08	0,08	0,04	1
Fréquences (%)	12	24	20	4	8	12	8	8	4	100
Fréquences Cumulées (%)	12	36	56	60	68	80	88	96	100	

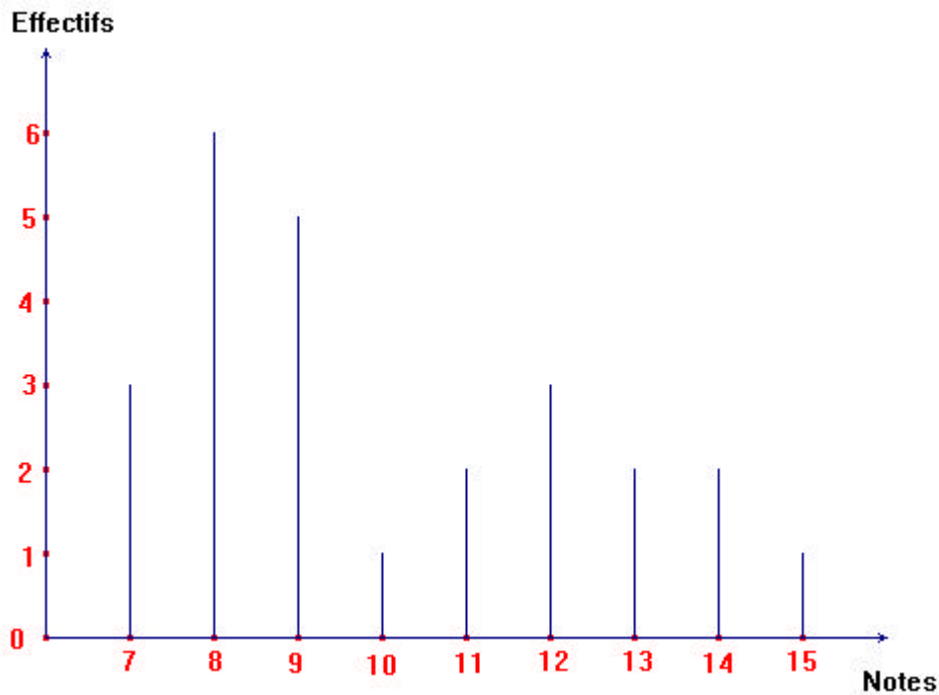
- L'effectif d'une note est le nombre d'individus ayant obtenu cette note.
- L'effectif total est la somme de tous les effectifs, c'est à dire le nombre d'éléments de la population.
- L'effectif cumulé d'une note N est le nombre d'individus ayant obtenu une note inférieure ou égale à N.

- Fréquence = $\frac{\text{Effectif}}{\text{Effectif total}}$; la somme des fréquences vaut toujours 1.

- Fréquence (%) = $\text{Fréquence} \times 100 = \frac{\text{Effectif}}{\text{Effectif total}} \times 100$; la somme des fréquences exprimées en pourcentages vaut toujours 100.

2) Représentation graphique des données.

Les données à caractère quantitatif discret se représentent sous forme d'un diagramme à bâtons. Ici, nous allons représenter le diagramme en bâtons des effectifs, la hauteur des bâtons sera donc proportionnelle aux effectifs. On prendra pour unités 1 carreau pour 1 sur les deux axes.



3) Calcul de la moyenne.

On peut calculer la moyenne des notes de deux façons différentes :

- soit en utilisant l'énoncé, en ajoutant toutes les valeurs et en divisant le résultat par l'effectif total :

Soit M la moyenne de cette série de notes

$$M = \frac{8+9+14+8+12+9+7+12+9+13+9+11+12+7+9+8+11+8+8+15+10+14+8+13+7}{25}$$

$$M = \frac{251}{25}$$

$$\boxed{M = 10,04}$$

La moyenne de cette série de notes est de 10,04.

- soit en utilisant le tableau des effectifs :

Soit M la moyenne de cette série de notes

$$M = \frac{3 \times 7 + 6 \times 8 + 5 \times 9 + 1 \times 10 + 2 \times 11 + 3 \times 12 + 2 \times 13 + 2 \times 14 + 1 \times 15}{25}$$

$$M = \frac{251}{25}$$

$$\boxed{M = 10,04}$$

La moyenne de cette série de notes est de 10,04.

III) Gestion des données à caractère quantitatif continu.

Une usine fabrique des vis de diamètres différents. Voici les diamètres (en millimètre) des vis passées entre les mains d'un employé en une heure :

18	21	24	12	14	16	13	17	14	18	21	22	18
14	16	12	18	21	22	14	18	21	16	12	21	16
11	13	15	13	17	18	21	22	24	24	20	16	15

1) Regroupement en classes.

On décide de regrouper les données en classes.

La classe $[a ; b[$ est l'ensemble des diamètres d tels que $a \leq d < b$.

L'amplitude de la classe $[a ; b[$ est : $b - a$.

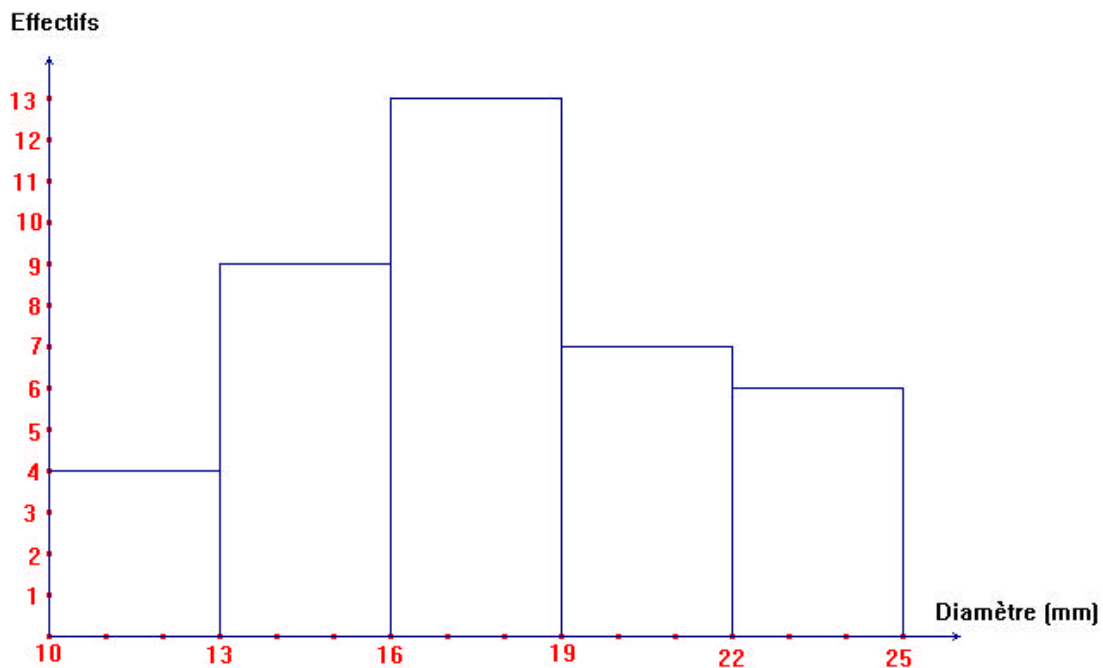
Le centre de la classe $[a ; b[$ est : $\frac{a+b}{2}$ (c'est la moyenne des deux bornes de la classe).

On décide de regrouper les données en classes d'amplitude 3 en commençant par la classe $[10 ; 13[$.

Diamètre	$[10 ; 13[$	$[13 ; 16[$	$[16 ; 19[$	$[19 ; 22[$	$[22 ; 25[$	Total
Centre des classes	11,5	14,5	17,5	20,5	23,5	
Effectifs	4	9	13	7	6	39
Effectifs cumulés	4	13	26	33	39	
Fréquences (%) arrondies à 0,1 près	10,3	23,1	33,3	17,9	15,4	100
Fréquences (%) cumulées arrondies à 0,1 près	10,3	33,4	66,7	84,6	100	

2) Représentation graphique des données.

Les données regroupées en classes se représentent sous forme d'un histogramme. Construisons ici l'histogramme des effectifs. (on prendra 1 carreau pour 1 mm en commençant à 10 sur l'axe des abscisses et 1 carreau pour 1 vis sur l'axe des ordonnées)



Dans un histogramme, l'aire des rectangles doit être proportionnelle aux effectifs. Comme ici les classes ont la même amplitude, la hauteur des rectangle est proportionnelle aux effectifs.

3) Calculs.

On peut calculer la moyenne de deux façons différentes :

- si on dispose des données initiales, il suffit d'ajouter toutes les valeurs et de diviser par l'effectif total :

Soit d le diamètre moyen d'une vis

$$d = \frac{18 + 21 + 24 + \dots + 20 + 16 + 15}{39}$$

$$d = \frac{678}{39}$$

$$\boxed{d \approx 17,4}$$

Le diamètre moyen d'une vis est d'environ 17,4 mm.

- si on ne dispose que des données regroupées en classes, il nous faut fixer des valeurs pour calculer une valeur approchée de la moyenne. Ici on utilisera les centres des classes (c'est à dire que l'on considérera qu'il y a 4 diamètres de 11,5 mm, 9 diamètres de 14,5 mm ...) mais on pourrait tout aussi bien utiliser les bornes inférieures ou les bornes supérieures des classes.

Soit d une valeur approchée du diamètre moyen des vis

$$d = \frac{4 \times 11,5 + 9 \times 14,5 + 13 \times 17,5 + 7 \times 20,5 + 6 \times 23,5}{39}$$

$$d = \frac{688,5}{39}$$

$$\boxed{d \approx 17,7}$$

Le diamètre moyen d'une vis est d'environ 17,7 mm.

IV) Gestion des données à caractère qualitatif.

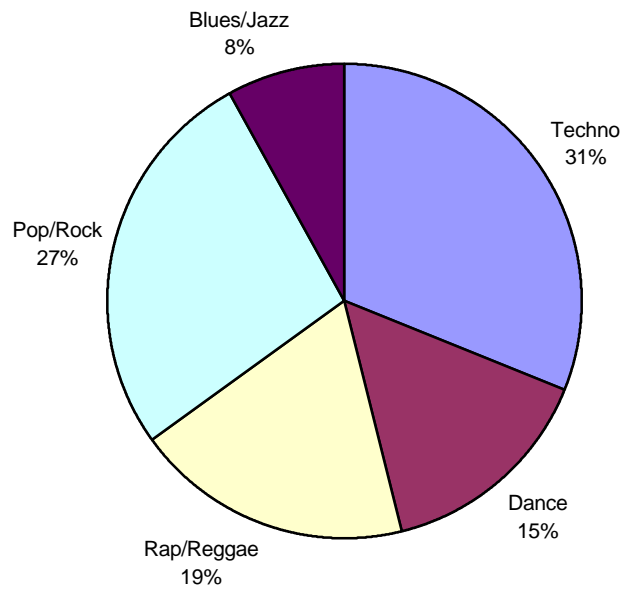
En avril 2001, des élèves de 4^{ème} d'un collège de Nîmes ont demandé à 300 jeunes de 13 à 15 ans le style de musique qu'ils préféreraient. 93 préfèrent la techno, 45 la dance, 57 le rap ou le reggae, 81 écoutent de la pop ou du rock et 24 aiment le blues ou le jazz.

1) Organisation.

Style	Techno	Dance	Rap/Reggae	Pop/Rock	Blues/Jazz	Total
Effectif	93	45	57	81	24	300
Fréquences (%)	31	15	19	27	8	100
Angles arrondis au degré près	112	54	68	97	29	360

2) Représentation graphique des données.

Des données à caractère qualitatif se représentent sous forme de diagramme circulaire (ou semi-circulaire) dans lequel les angles doivent être proportionnels aux effectifs.



Les données à caractère qualitatif prennent des valeurs non numériques, il est donc impossible de calculer une moyenne.