

OUTILS POUR LES STATISTIQUES**1°) Vocabulaire.**

Une **série statistique** est une liste de **données** qui permet d'étudier **un caractère** chez une **population** précise. On appelle individu un élément de la population. Chaque **individu** de la population représente une **valeur** du caractère étudié. Le caractère peut être **qualitatif** (ex : couleur des yeux, code postal...), **quantitatif discret** (nombre de frères et sœurs, nombre d'animaux domestiques, nombre de téléviseurs au domicile...), ou **quantitatif continu** (temps de trajet en min pour venir à l'école, taille en cm des élèves de la classe, masse en grammes des œufs ramassés dans le poulailler...).

Pour étudier au mieux une série statistique, on compte le nombre de fois que chaque valeur apparaît, cette étape s'appelle le **dépouillement**. On regroupe souvent les informations dans un tableau. On appelle **effectif** le nombre de fois qu'une série prend une valeur, on appelle **effectif total** le nombre total d'individus dans la population. La valeur qui a le plus grand effectif s'appelle **le mode** de la série.

Pour transmettre une information, on peut représenter graphiquement une série statistique, ou calculer certains paramètres, de position ou de dispersion.

Nous allons rappeler ici les bases des statistiques descriptives au travers de l'étude d'exemples.

Applications directes :

- *Exercices de la page 167 et page 169.*
- *Exercices du 15 au 38 pages 178-180 du livre.*

2°) Etude d'une série statistique à caractère qualitatif.

On a une série à caractère qualitatif lorsqu'on ne peut pas calculer la moyenne : soit parce que les valeurs ne sont pas numériques (comme la couleur du véhicule, le nom du chanteur préféré, etc.) ; soit parce que les chiffres représentent un code (code postal, numéro de téléphone, etc.).

Couleur du véhicule x_i	gris	blanc	rouge	noir	vert	TOTAL
Effectif n_i	45	65	20	35	15	$N = 180$
Fréquence f_i	0,25	0,36	0,11	0,20	0,08	1
Fréquence en pourcentage	25	36	11	20	8	100
Angle au centre en degré α_i	90	130	40	70	30	360

Le mode de cette série est la valeur 'blanc'.

L'effectif total N correspond à la somme des effectifs. On utilise la notation suivante :

$$N = n_1 + n_2 + n_3 + n_4 + n_5 = \sum_{i=1}^{i=5} n_i$$

La fréquence permet de rendre compte de la répartition des données sur les différentes valeurs. La fréquence d'une donnée représente la proportion d'individus qui correspondent à cette valeur parmi la population. La fréquence est un nombre compris entre 0 et 1. On peut l'exprimer sous la forme d'un nombre décimal, ou sous la forme d'une fraction irréductible. La somme des fréquences doit être exactement égale à 1.

$$f_i = \frac{n_i}{N} = \frac{n_i}{\sum n_i}$$

On peut également exprimer la fréquence en pourcentage, tout simplement en multipliant les fréquences par 100. Dans ce cas la somme des fréquences doit être égale à 100.

Voir les 3^{ème} et 4^{ème} ligne du tableau pour la fréquence et la fréquence en pourcentage.

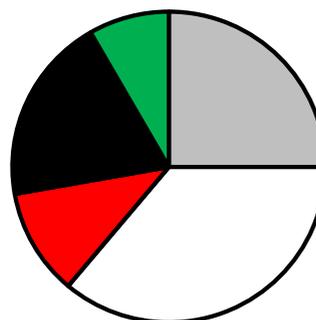
Les représentations les plus courantes pour ce type de série sont un diagramme en bâtons, un diagramme en tuyaux d'orgue, ou un diagramme circulaire. Nous allons montrer ici la construction du diagramme circulaire. Avant de tracer le diagramme, il faudra calculer la valeur de l'angle au centre de chaque secteur angulaire :

$$\alpha_i = f_i \times 360 = \frac{n_i}{N} \times 360$$

On a proportionnalité entre la valeur de l'angle au centre, la fréquence, et l'effectif.

Dans le diagramme circulaire, on ne fait pas apparaître la valeur des angles, et rarement les fréquences. En revanche il faut faire apparaître les valeurs, soit à l'intérieur soit par une légende, soit en mettant les couleurs correspondantes ; et ne pas oublier un titre.

Couleur du véhicule



Remarques : pour un diagramme semi-circulaire, on se référera à un angle au centre total de 180°.

On appelle écart interquartile le nombre $Q_3 - Q_1$, cette valeur correspond à l'écart correspondant entre les deux quartiles, c'est l'étendue de 50% des valeurs les plus centrales.

Dans notre exemple :

$$\frac{160}{4} = 40 \text{ donc } Q_1 \text{ est la } 40^{\text{ème}} \text{ valeur de la série, d'après les effectifs cumulés croissants, } Q_1 = 0.$$

$$160 \times \frac{3}{4} = 120 \text{ donc } Q_3 \text{ est la } 120^{\text{ème}} \text{ valeur de la série, d'après les effectifs cumulés croissants, } Q_3 = 2.$$

Remarque :

La médiane est un paramètre de position non sensible aux valeurs extrêmes. Afin de donner des informations plus précises, on donne habituellement la médiane accompagnée par l'intervalle interquartile, qui est un paramètre de dispersion.

La moyenne est un paramètre de position sensible aux valeurs extrêmes. Afin de donner des informations plus précises, on donne habituellement la moyenne accompagnée par l'étendue, qui est un paramètre de dispersion.

Représentations graphiques possibles d'une série à caractère quantitatif discret :

Diagramme bâtons :

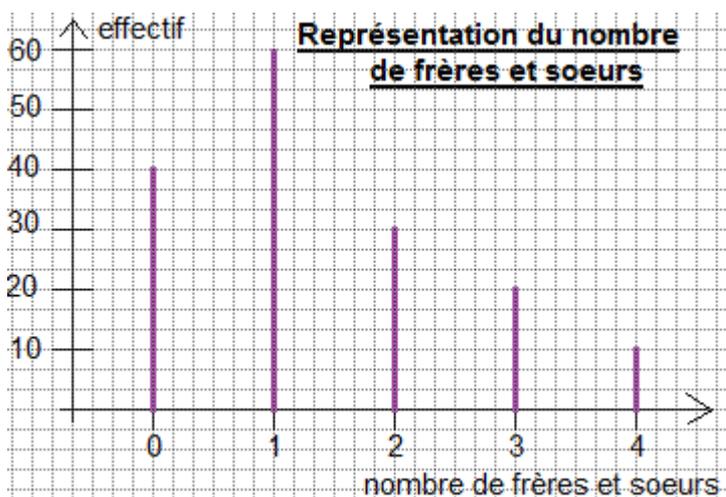
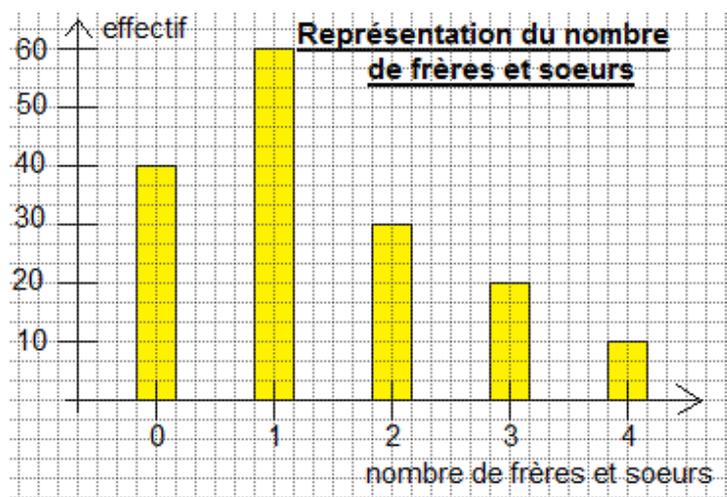


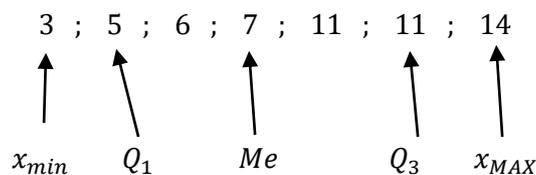
Diagramme en tuyaux d'orgue :



Les diagrammes circulaires sont également adaptés.

Complément : médiane, quartiles, valeurs extrêmes : comment synthétiser ces informations.

Nombre de pétales d'une fleur cueillie dans le jardin, pendant une semaine, série rangée par ordre croissant :



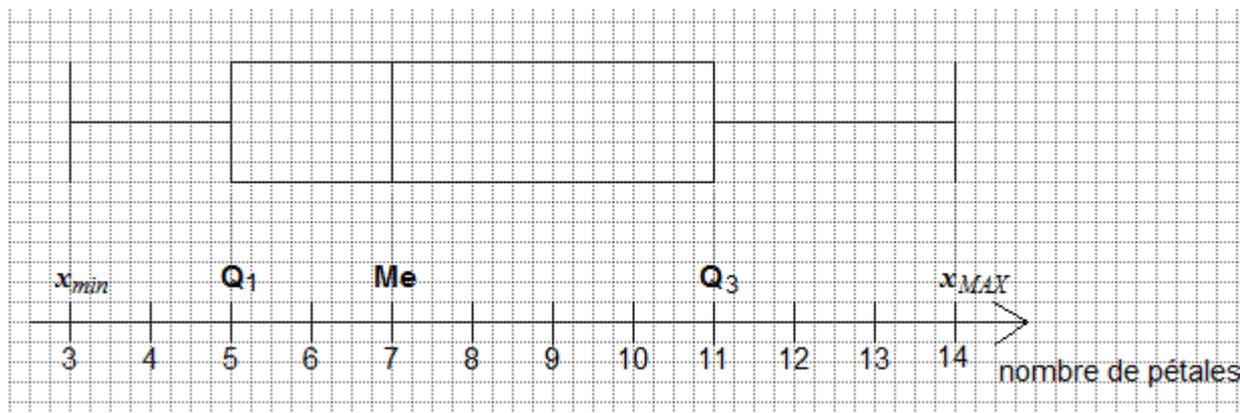
L'effectif total est 7.

$\frac{7+1}{2} = \frac{8}{2} = 4$ donc la médiane est la 4^{ème} valeur, soit 7.

$\frac{7}{4} = 1,75 \approx 2$ donc le premier quartile est la 2^{ème} valeur, soit 5.

$\frac{7 \times 3}{4} = 5,25 \approx 6$ donc le troisième quartile est la 6^{ème} valeur, soit 11.

Je trace un axe gradué représentant les valeurs, et je place sur l'axe les valeurs extrêmes, les quartiles, la médiane. Au-dessus, je construis un diagramme en boîte, aussi appelé boîte à moustaches :



Le diagramme en boîte permet d'avoir une représentation de la répartition des valeurs :

en effet 25% des valeurs sont comprises entre x_{min} et Q_1 ;

25% des valeurs sont comprises entre Q_1 et Me ;

25% des valeurs sont entre Me et Q_3 ;

et enfin 25% des valeurs sont comprises entre Q_3 et x_{MAX} .

4°) Etude d'une série statistique à caractère quantitatif continu.

On a une série statistique à caractère quantitatif continu lorsque l'on a des valeurs numériques non dénombrables. C'est-à-dire que trop de valeurs différentes peuvent être données. Exemple : masse en milligrammes d'une partie de téléviseur dans une chaîne de fabrication, longueur en mm d'un cheveu, temps de trajet en minutes, taille en cm...

Lors du dépouillement, on regroupera les valeurs dans des classes de valeurs, afin de rendre exploitable un tableau de valeurs. Les classes de valeurs sont des intervalles de la forme $[a ; b[$.

On appelle amplitude d'une classe la longueur de l'intervalle, c'est-à-dire le nombre $b - a$.

Pour calculer la moyenne, il est indispensable de calculer le centre des classes, par la formule $x_i = \frac{a+b}{2}$.

On appelle classe modale la classe qui a le plus grand effectif, et classe médiane la classe qui contient la médiane.

Pour trouver la médiane et les quartiles, on peut calculer les fréquences cumulées croissantes, et représenter le graphe des fréquences cumulées croissantes, aussi appelé polygone des fréquences cumulées croissantes. On procède ensuite à une lecture graphique.

La seule représentation graphique adaptée est l'histogramme. Dans un histogramme, la surface des rectangles est proportionnelle à l'effectif.

Exemple : étude du temps de trajet, en minutes, des élèves d'un lycée de Poitou-Charentes.

temps de trajet en minutes	[0;10[[10;20[[20;30[[30;40[[40;60[TOTAL
centre des classes x_i	5	15	25	35	50	
effectif n_i	34	72	116	53	25	$N = 300$
$x_i n_i$	170	1080	2900	1855	1250	$\sum x_i n_i = 7255$
fréquence en pourcentage f_i	11	24	39	18	8	100
fréquence cumulée croissante FCC	11	35	74	92	100	

Calcul de la moyenne :

$$\bar{x} = \frac{\sum x_i n_i}{N} = \frac{7\,255}{300} \approx 24,18$$

Le temps de trajet moyen est d'environ 24,18 minutes.

La classe modale est $[20 ; 30[$.

Les fréquences cumulées croissantes dépassent 50% à la classe médiane qui est $[20 ; 30[$.

Le polygone des effectifs cumulés croissants, ainsi que l'histogramme, sont sur la page suivante.

