Ch 3	Statistiques	2nde

I. Définir et représenter une série statistique

A. Vocabulaire

Définitions [1].

- La ***population*** d'une série statistique est l'ensemble des éléments appelés individus sur lesquels porte l'étude statistique (*On dit population même si ce ne sont pas des gens!*).
- Le ***caractère*** d'une série statistique est la propriété étudiée sur chaque individu.
- Un caractère est dit ***quantitatif***[1] lorsqu'il est possible de le mesurer en associant un nombre à chaque individu (âge, taille, note obtenue, salaire,…).
- Un caractère est dit ***qualitatif*** lorsqu'il n'est pas quantitatif (couleur des yeux, …).
- On dit qu'un caractère quantitatif (= *mesuré par un nombre*) est ***discret*** lorsqu'il ne peut prendre qu'un nombre fini de valeurs numériques (notes à un contrôle,…).
- On dit qu'un caractère quantitatif est **continu** lorsqu'il peut prendre un nombre infini de valeurs numériques (*taille, temps, distance…*).

Définitions [2].

- L'***effectif*** d'une valeur du caractère est le nombre de fois où cette valeur apparaît dans la série.
- L'***effectif cumulé croissant*** d'une valeur du caractère est la somme de tous les effectifs correspondants à des valeurs inférieures ou égales à cette valeur.
- La ***fréquence*** d'une valeur du caractère est le quotient de l'effectif de cette valeur par l'effectif total, càd $\text{fréquence} \overset{\text{def}}{=} \dfrac{\text{effectif}}{\text{effectif total}}$.

Remarques 3.

(a) La fréquence (tout court) est donc un nombre entre 0 et 1. Si les fréquences sont exprimée en pourcentage, on obtient alors des nombres entre 0 et 100 %.

(b) La somme des fréquences (tout court) vaut toujours 1 et la somme des fréquences exprimées en pourcentage vaut toujours 100 %.

(c) *Interprétation des ECC* : Lors d'un examen, si l'*effectif cumulé croissant* associé à la note 8 est 10 cela signifie que 10 élèves ont eu 8 ou moins à l'examen.

(d) *Pour vérifier vos calculs* : L'effectif cumulé croissant associé à la plus grande valeur de la série est toujours égal à l'effectif total. La fréquence cumulée croissant associé à la plus grande valeur de la série est toujours égale à 1 (ou à 100% si les fréquences sont exprimées en pourcentage).

B. Représentation d'une série statistique

Dans un ...	ce qui est proportionnel à l'effectif (et à la fréquence), c'est . . .
diagramme circulaire	les angles au centre
diagramme en bâtons	la hauteur des bâtons.
histogramme	l'aire des bâtons.

[1] « Quantitatif » vient de « quantité », et la quantité se mesure par un nombre.

II. Indicateurs statistiques

A. Médiane et quartiles

1. Qu'est-ce que c'est et comment les trouver ?

Définitions [4].

- La ***médiane Me*** est un nombre tel que au moins la moitié de l'effectif ait une valeur supérieure ou égale à la médiane et au moins la moitié de l'effectif ait une valeur inférieure ou égale à la médiane (l'idée est de séparer l'affectif en deux groupes de même taille mais ce n'est pas toujpurs possible)
  - Si l'effectif total N est un nombre impair, la médiane est le terme de rang $\frac{N+1}{2}$.
  - Si l'effectif total N est un nombre pair, la médiane est la moyenne des valeurs centrales, càd des termes de rang $\frac{N}{2}$ et $\frac{N}{2}+1$. *Dans ce cas, et c'est le seul, la médiane n'est PAS forcément une valeur de la série. Les quartiles, eux, sont toujours des valeurs de la série.*
- Le ***premier quartile Q_1*** est la plus petite valeur de la série telle qu'au moins 25 % des données soient inférieures ou égales à Q_1.
- Le ***troisième quartile Q_3*** est la plus petite valeur de la série telle qu'au moins 75 % des données soient inférieures ou égales à Q_3.

Savoir-Faire [5]. Comment trouver Q_1 et Q_3 ?

(a) On peut calculer $\frac{N}{4}$ où N est l'effectif total, et l'arrondir à la valeur supérieure (puisqu'on veut au moins 25%): On obtient le rang de la valeur Q_1. Idem Q_3 pour avec $\frac{3}{4}\times N$.

(b) Utilisation des FCC[2] pour trouver Q_1 et Q_3 : Le premier quartile Q_1 est la valeur de la série qui permet d'obtenir pour la première fois une fréquence cumulée croissante supérieure ou égale à 25 %. De même, Q_3 est la valeur de la série qui permet d'obtenir pour la première fois une fréquence cumulée croissante supérieure ou égale à 75 %.

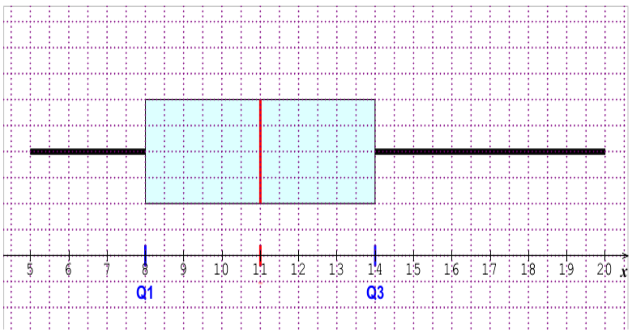
2. Résumer une série statique par un diagramme en boîte ou un diagramme en boîtes à moustaches

♠ Exemple 1. Notes obtenues lors d'un devoir. Les notes obtenues sont rangées dans l'ordre croissant.
5-6-6-6-7-7-7-8-8-9-9-10-10-10-10-11-12-12-13-13-13-13-14-14-14-15-15-15-17-18-20
L'effectif total N est égal à …….. donc Me est le terme de rang ……… de la série statistique rangée dans l'ordre croissant. On a alors Me = …….

$\frac{N}{4}$ =…….. Ce n'est pas un entier donc le premier quartile est le terme de rang ……….. soit Q1 = ……

$\frac{3N}{4}$ =…….. donc le troisième quartile est le terme de rang …….. soit Q3 = ……

Représentation des résultats sur un diagramme en boîte à moustache, ou de Tuckey :



Règles de construction des boîtes à moustache :
La boite démarre à Q_1.
La boite finit à Q_3.
Le trait vertical dans la boite correspond à la médiane.
L'extrémité de la moustache de gauche donne la plus petite valeur de la série (parfois le premier décile).
L'extrémité de la moustache de droite donne la plus grande valeur de la série (parfois le dernier décile).
On met parfois dans la boite un point ou un losange correspondant à la moyenne.

[2] *FCC = fréquence(s) cumulée(s) croissante(s)*

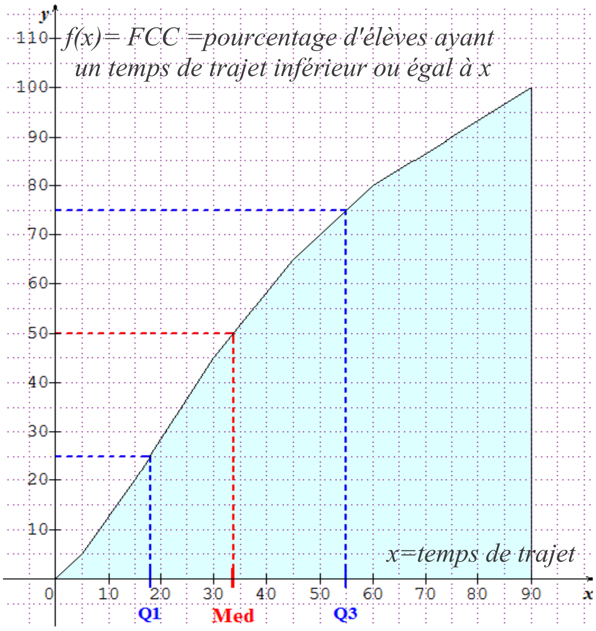
x=temps de trajet

♠ Exemple 2. Caractère quantitatif continu

On a relevé les temps de trajet moyens de 100 élèves d'un lycée. Les résultats sont regroupés dans le tableau suivant.

Temps en minutes	[0 ; 5[	[5 ; 15[	[15 ; 30[	[30 ; 45[	[45 ; 60[	[60 ; 90[
Effectifs	5	15	25	20	15	20
Effectifs cumulés croissants (*A vous!*)						

En supposant que la répartition à l'intérieur des classes[3] est uniforme (c'est à dire que sur le diagramme des fréquences cumulées croissantes, les points sont joints par des segments), calculer sa médiane Me et ses quartiles Q_1 et Q_3 .



Réponse : On lit sur le **diagramme des effectifs cumulés croissants** ci-contre la médiane et les quartiles, qui sont les abscisses respectives des points d'ordonnées 25%, 50% et 75% .

premier quartile : $Q_1=18$ minutes

médiane : $Me=33{,}75$ minutes

et troisième quartile $Q_3=55$ minutes.

B. Moyenne

Considérons la série statistique suivante, où chacune des valeurs x_i apparaît n_i fois :

Valeurs	x_1	x_2	.....	x_p
Effectifs	n_1	n_2	.....	n_p

Définition [6].

- La moyenne de la série est $\bar{x}=\dfrac{n_1x_1+n_2x_2+\ldots+n_px_p}{N}$ où $N=n_1+n_2+\ldots+n_p$ est l'effectif total. On dit que $\bar{x}$ est la ***moyenne pondérée*** des x_i , les ***coefficients*** ou ***poids*** étant les n_i .
- Avec les fréquences, la moyenne de la série est $\bar{x}=f_1x_1+f_2x_2+\ldots+f_px_p$.

Savoir-Faire [7]. Comment calculer la moyenne d'une série dont les valeurs ont été regroupées en classes (=intervalles)**?**

▪ Lorsqu'une série est répartie en classes, on peut considérer que la valeur au centre de la classe représente la classe dans le calcul.

▪ On calcule alors la moyenne pondérée avec cette valeur et on obtient une valeur approchée de la moyenne de la série.

♠ Exemple 3. *Taille d'un groupe d'individus*

Tailles en cm	[140 ; 150[	[150 ; 160[	[160 ; 170[	[170 ; 195[	total
Centre de la classe	145	155			
	48	397	913	642	

Calcul de la moyenne :

[3] « classe », dans cette phrase, est utilisé au sens « [0 ; 5[est une classe » et pas « La 205 est une classe ».

Propriétés de la moyenne [8].
- Si on ajoute un même nombre b à toutes les valeurs d'une série statistique, la moyenne est augmentée du même nombre b.
- Si on multiplie par un même nombre a toutes les valeurs d'une série statistique, la moyenne est multipliée par le même nombre a.

♠ Exemple 4. *Moyenne d'un groupe obtenu en fusionnant des groupes de moyenne connue*
La moyenne d'une classe de 21 élèves est 9,9. La moyenne d'une autre classe de 33 élèves est 10,7. Quelle est la moyenne pour l'ensemble des deux classes ?
Outil :

Utilisation des moyennes partielles [9]. Considérons deux séries statistiques d'effectifs n_1 et n_2 et de moyennes m_1 et m_2. La moyenne de la série obtenue en regroupant ces deux séries est la moyenne de m_1 et m_2 pondérées par les effectifs correspondants, càd $m=\dfrac{n_1 m_1+n_2 m_2}{n_1+n_2}$.

Réponse : m = $\dfrac{21\times9{,}9+33\times10{,}7}{21+33} = \dfrac{561}{54} \approx 10{,}39$

Remarque: On n'a pas de résultat de ce type pour la médiane : Connaissant les médianes des deux groupes on ne peut pas trouver celle du groupe obtenue en fusionnant les deux groupes. Il faut tout refaire ! L'existence de cette propriété est un avantage de la moyenne sur la médiane.

C. Paramètres de dispersion

Pour estimer la dispersion des valeurs d'une série statistique on peut utiliser son étendue ou, mieux, son écart inter-quartile.

Définitions [10].
- L'***étendue*** d'une série statistique est la différence entre la plus grande et la plus petite des valeurs du caractère étudié.
- L'***intervalle interquartile*** d'une série statistique est l'intervalle $[Q_1;Q_3]$.
- L'***écart interquartile*** d'une série statistique est la différence entre le 1er et le 3ème quartile, càd Q_3-Q_1.

♠ Exemple 5. *Avec les données de l'exemple 1 page 2*

1) L'étendue de la série ($=max-min$) est ……………………………………………

2) L'écart interquartile de la série est …………………………………………………

3) L'intervalle interquartile de la série est ……………………………………………

Sources : Labomath, cours de V. Degos, manuel Repères.

Table des matières