

# Statistiques à une variable

## Table des matières

<b>1 Vocabulaire de la statistique</b>	<b>1</b>
Exemples . . . . .	1
Notations . . . . .	2
<b>2 Caractère quantitatif discret</b>	<b>2</b>
Représentations . . . . .	2
Paramètres de position . . . . .	3
La moyenne . . . . .	3
La médiane . . . . .	3
Le mode . . . . .	3
Comment annoncer cela à mes parents ? . . . . .	3
Paramètres de dispersion . . . . .	3
L'étendue . . . . .	3
Les quartiles . . . . .	3
L'écart-type . . . . .	4
Conclusion . . . . .	4
<b>3 Caractère quantitatif continu</b>	<b>4</b>
Représentation . . . . .	4
Caractéristiques . . . . .	5

## 1 Vocabulaire de la statistique

Le mot "statistique" vient du latin "status" qui signifie état. Une étude statistique consiste à observer et à étudier une particularité commune chez un groupe de personnes ou de choses. Les personnes ou les choses sont ce que l'on appelle des **individus**. Le groupe formé par ces individus est dénommé **population**. Cette particularité est elle appelée caractère.

Par exemple, on peut étudier les notes obtenues par les élèves d'une classe à un devoir. La population considérée est celle de la classe. Tous les individus de cette population présentent la particularité d'avoir une note. C'est le caractère.

L'étude statistique commence par un recueil de données. Par exemple dans le cadre des notes obtenues par une classe, on a que l'individu X a obtenu par exemple la note de 10. Les différentes valeurs pouvant être prises par un caractère portent le nom de **modalité**. Ainsi par exemple, 10 est une modalité que peut prendre la caractère note. 15 est une autre modalité possible de ce même caractère.

Mais savoir que tel individu présente telle modalité est peu parlant d'un point de vue global. Par contre savoir combien d'individus présentent cette modalité est nettement plus intéressant. Par modalité, on compte est le nombre d'individus qui la présentent (**effectif**).

Ainsi obtient-on une série statistique à une variable (sous-entendu un seul caractère est pris en compte et étudié).

### Exemples

- Série statistique 1** : Etude de la population formée par les députés nouvellement élus selon leur appartenance politique.

Parti	PCF	PS	PRS	DVG	Ecolo.	Divers	UDF	RPR	DVD	FN
Nombre	38	241	12	21	7	1	108	134	14	1

PRS : Parti Radical Socialiste DVG : Divers Gauche DVD : Divers Droite.

Le caractère "appartenance politique" est un **caractère qualitatif**.

Les modalités présentées par les individus sont PCF, PS, PRS Ce sont des valeurs caractéristiques.

On les qualifie de qualité. D'où le nom de caractère qualitatif !

2. **Série statistique 2** : Etude de la population formé par la classe de première ES 1 selon leur moyenne en maths au troisième trimestre.

Note	3	5	6	7	8	9	10	12	13	14	15	16
Nombre	1	2	3	3	7	3	2	2	3	3	2	2

Nombre d'élèves ayant obtenu la note.

Le caractère "moyenne en maths" est un **caractère quantitatif discret**. Quantitatif car la moyenne de chaque individu est un nombre. Discret car la moyenne ne peut être égale qu'à un nombre limité de valeurs.

3. **Série statistique 3** : Etude de la population française selon le temps passé quotidiennement devant le petit écran.

Temps	moins de 1h.	Entre 1 et 2h.	Entre 2 et 3h.	Entre 3 et 4h.	Plus de 4 h.
	[0; 1[	[1; 2[	[2; 3[	[3; 4[	[4; 5[
Pourcentage	10,1	15,1	17,1	16	41,7

Une personne peut regarder la télévision pendant 2 heures 38 alors qu'une autre peut ne la regarder que durant 2 heures 40. Pour un responsable publicitaire, cette différence de deux minutes n'est pas très parlante. De plus, connaître le temps exact que l'on passe devant le petit écran n'est pas une chose aisée. On le sait toujours à peu près mais jamais exactement. C'est pour cette raison que l'on regroupe les modalités (2 heures 38 par exemple) en ce que l'on appelle des classes qui sont en fait des intervalles de modalités (par exemple l'intervalle [2 heures ; 3 heures]). Ce caractère est dit **quantitatif** (car on peut le compter ou le mesurer) **continu** car un grand nombre de modalités pouvant être prises, on les regroupe en plages continues ou en intervalles de valeurs (classes).

## Notations

Dans tous les cas, il est plus simple de réunir les résultats dans un tableau :

Caractère $x_i$	$x_1$	$x_2$	...	$x_p$	Total
Effectif $n_i$	$n_1$	$n_2$	...	$n_p$	$N$

Le caractère peut prendre  $p$  valeurs différentes. L'effectif total est  $N = \sum_{i=1}^p n_i$ .

Le signe  $\sum$  signifie que l'on fait une somme. Ici on ajoute les nombre  $n_i$  pour  $i$  allant de 1 jusqu'à  $p$ .

**Définition 1.1** La **fréquence** d'un caractère est le nombre d'individus possédant ce caractère divisé par l'effectif total de la population.  $f_i = \frac{n_i}{N}$ .

Le plus souvent, la fréquence est exprimée en pourcentage.

## 2 Caractère quantitatif discret

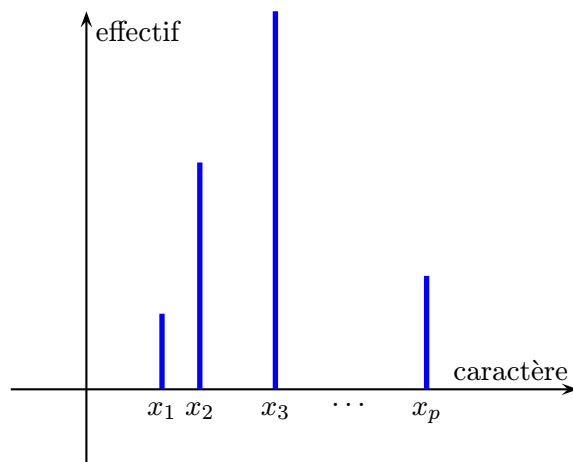
Caractère $x_i$	$x_1$	$x_2$	...	$x_p$	Total
Effectif $n_i$	$n_1$	$n_2$	...	$n_p$	$N$

Les caractères  $x_i$  sont des nombres et ils prennent  $p$  valeurs différentes. L'effectif total est  $N = \sum_{i=1}^p n_i$ .

## Représentations

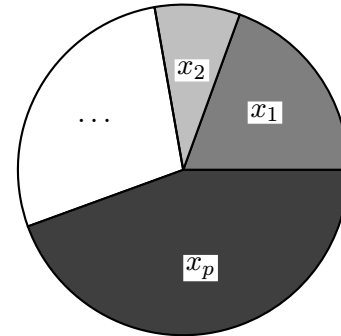
On représente les variables aléatoires discrètes sous forme :

de **diagramme en bâtons**. La hauteur des bâtons est proportionnelle aux effectifs.



ou de **diagramme circulaire**. L'angle au centre est proportionnel aux effectifs. Dans la pratique, on construit le tableau:

Caractère $x_i$	$x_1$	...	$x_p$	Total
Effectif $n_i$	$n_1$	...	$n_p$	$N$
Angle		...		360



Si l'on veut avoir une idée sur la population étudiée, il va falloir se donner des outils (paramètres) nous donnant l'"allure" de cette série.

## Paramètres de position

### La moyenne

**Définition 2.1** La moyenne est  $\bar{x} = \frac{1}{N} \sum_{i=1}^p n_i x_i$ .

### La médiane

**Définition 2.2** La médiane est le nombre qui partage la série statistique en deux séries de même effectif, l'une dont le caractère est plus petit que la médiane, l'autre dont le caractère est plus grand que la médiane.

Dans la pratique, on dresse le tableau des effectifs cumulés croissants (ou des fréquences cumulées croissantes) et on conclue (c'est le caractère qui correspond à une fréquence cumulée de 50%).

### Le mode

**Définition 2.3** Le mode est le caractère qui a le plus grand effectif.

### Comment annoncer cela à mes parents ?

L'interrogation de statistique n'a pas été terrible: 8/20. Comment annoncer cela à mes parents ?

Dans l'ensemble il faut dire que ce n'était pas fameux. Nous sommes 10 en classe et les résultats sont catastrophiques! Pensez donc. Le petit génie a bien sûr fait 19, mais à part cela il y avait un 10, quatre 9 et trois 2. D'accord, le mode est 9/20 et la médiane est également 9/20. Mais si je calcule la moyenne, je trouve 7,9/20. Je dirai donc à Papa que j'ai au-dessus de la moyenne.

Encore un 8. Mais cette fois les notes sont: 2, 3, 4, 5, 7, 8 (moi), 9, 9, 18 et 19 (le génie). J'ai calculé la moyenne, mais cette fois elle est de 8,4; je suis en dessous de la moyenne; et le mode est 9. Heureusement, il n'y en a que 4 qui ont mieux réussi que moi et les 5 autres sont après. Je dirai donc à Papa que je suis au-dessus de la médiane.

Décidément, je n'ai pas de chance. Je suis abonné au 8/20. C'est sûrement la faute du prof ! Cette fois les questions étaient tellement dures qu'il y en a 3 qui ont eu 7/20 !. Les autres ont obtenu 19 (toujours

le même), 18,12,11,10 et 2 (c'est aussi toujours le même) J'ai calculé la moyenne; cela fait 10,1. Pas de chance, je suis en dessous. Et cette fois il y en a 5 qui ont plus que moi! Ça ne va plus l'histoire de la médiane !. Heureusement grâce aux trois copains, le mode est 7. Je dirai cette fois à Papa que je suis au-dessus du mode.

(J'espère qu'il ne comprend rien aux différences entre moyenne, médiane et mode !)

## Paramètres de dispersion

### L'étendue

L'étendue est simplement la différence entre la valeur maximale et la valeur minimale du caractère. Par exemple si les notes d'un devoir vont de 5 à 15 alors l'étendue est de 10.

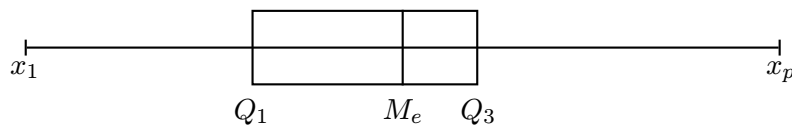
### Les quartiles

Rappelons comment est définie la médiane : après avoir ordonné tous les éléments de la population, on les répartit en deux parties de même effectif, que nous appellerons "partie basse" et "partie haute". La valeur théorique qui sépare ces deux parties s'appelle la médiane.

**Définition 2.4** On appelle **premier quartile**  $Q_1$  la médiane de la partie basse de la population. On appelle **troisième quartile**  $Q_3$  la médiane de la partie haute de la population.

Remarquons que le deuxième quartile est la médiane.

On peut alors représenter l'ensemble des valeurs par le schéma appelé diagramme de Turkey ou "boîte à moustaches" :



**Définition 2.5** La distance entre  $Q_1$  et  $Q_3$  s'appelle l'**écart interquartile**.

### L'écart-type

On veut mesurer la dispersion autour de la moyenne de la série statistique. On peut mesurer pour chaque caractère  $x_i$  son éloignement à la moyenne  $(x_i - \bar{x})^2$  (on prend le carré car on veut des nombres positifs). La variance est la moyenne de ces éloignements.

**Définition 2.6** La **variance** est  $V(x) = \frac{1}{N} \sum_{i=1}^p n_i (x_i - \bar{x})^2$ .

**Proposition 2.1** : La variance peut se calculer par une autre formule:  $V(x) = \left( \frac{1}{N} \sum_{i=1}^p n_i x_i^2 \right) - \bar{x}^2$ .

**Définition 2.7** L'**écart-type** est  $\sigma = \sigma_x = \sqrt{V(x)}$ .

Si un caractère possède une unité (par exemple le mètre) la variance est en l'unité au carré (en  $m^2$ ) donc l'écart-type a la même unité que le caractère.

### Conclusion

On peut résumer de façon satisfaisante une série statistique en donnant :

- soit sa médiane et son écart interquartile;
- soit sa moyenne et son écart-type.

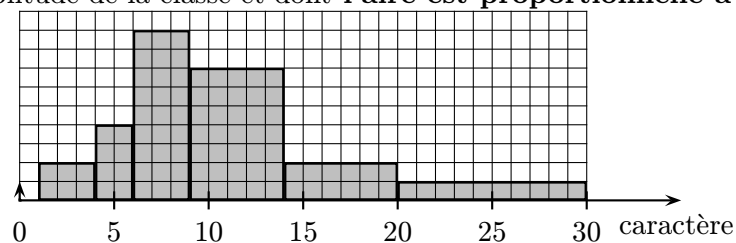
### 3 Caractère quantitatif continu

Les caractères sont des intervalles. Ceci est représenté dans le tableau cidessous :

Caractère	$[X_1; X_2[$	$[X_2; X_3[$	...	$[X_p; X_{p+1}[$	Total
centre des classes $x_i$	$x_1$	$x_2$	...	$x_p$	
Effectif $n_i$	$n_1$	$n_2$	...	$n_p$	$N$

#### Représentation

Pour leur représentation, on regroupe en général dans des classes adjacentes (qui se "touchent") d'amplitudes pas forcément égales. La représentation s'effectue alors grâce à un histogramme dont les rectangles sont de largeur l'amplitude de la classe et dont **l'aire est proportionnelle à l'effectif**.



#### Caractéristiques

Pour calculer moyenne et écart-type, on prend les formules connues avec les  $x_i$  centres des classes, c'est-à-dire  $x_i = \frac{X_i + X_{i+1}}{2}$ .