

En mathématiques, la notion de statistique est utilisée depuis l'Antiquité afin de faire des recensements de la population. Aujourd'hui elle joue un rôle important dans l'interprétation des résultats, et rend notamment lisible les valeurs d'un caractère, grâce à des calculs de paramètres.

Le premier recensement « moderne » au niveau national aurait ainsi été celui ordonné en 1694 par [Louis Phélypeaux](#), comte de Pontchartrain. Celui-ci sera suivi par divers recensements, dénombremements et enquêtes nationales conduits à intervalles irréguliers.

Le recensement de population de 1801 préparé par [Lucien Bonaparte](#) et [Jean-Antoine Chaptal](#) a été le point de départ d'une série de recensements effectués – avec plus ou moins de régularité - tous les cinq ans jusqu'en 1946. Depuis, les recensements ont été organisés par l'[Institut national de la statistique et des études économiques](#) (Insee) jusqu'en 1999, et sous une forme rénovée depuis 2004.

En 2013 en France, près de 350 articles de lois ou de codes se réfèrent au recensement, dont pour l'organisation des [élections municipales](#), la répartition de la [dotation globale de fonctionnement](#), la répartition des services de santé, certaines politique de prévention et gestion des risques. (#Wikipédia)

I. Pour prendre un bon départ : maîtriser le vocabulaire

Les statistiques utilisent un vocabulaire précis qu'il faut connaître.

1. Population / Individu :

On appelle **population** tout ensemble soumis à une étude statistique. Un **individu** est alors un élément de cet ensemble.

Le nombre d'individus constitue la **taille** de la population.

*Remarque : Si la population est de taille très importante, on peut ne faire porter l'étude statistique que sur une partie de la population ; dans ce cas on dit qu'on prend un **échantillon** de la population et on essaye, autant que faire se peut, que l'échantillon soit **représentatif** de la population. Dans le cas contraire, l'étude n'aurait pas d'intérêt.*

2. Caractère / Modalité :

L'aspect ou le trait sur lequel porte l'étude statistique est appelé **caractère** ou **variable statistique**.

Les différents traits observés sur ce caractère portent le nom de **modalités**.

Les modalités sont souvent notées x_i où i désigne un entier naturel.

3. Quantitatif / Qualitatif :

Un caractère est dit **quantitatif** si les modalités sont des valeurs numériques (mesures physiques, physiologiques, sociologiques, démographiques, économiques...).

Dans le cas contraire, c'est-à-dire lorsque les valeurs ne peuvent être ni ordonnées ni ajoutées, le caractère est dit **qualitatif**.

4. Discret / Continu :

Un caractère **quantitatif** est dit **discret** lorsqu'il ne prend qu'un nombre fini de valeurs numériques, isolées.

Dans le cas contraire, le caractère est **continu**, il prend une infinité de valeurs.

5. Classe / Centre :

Lorsque le caractère est **quantitatif et continu**, les modalités, *c'est-à-dire les différentes valeurs prises par le caractère* peuvent être regroupées en intervalles appelés **classes**. La valeur centrale de la classe est alors appelée **centre de la classe**.

6. Effectif / Effectif total / Série des effectifs :

L'**effectif** d'une modalité est le nombre d'individus possédant cette valeur du caractère.

L'**effectif total** est le nombre d'individus de la population. C'est la somme des effectifs de chaque modalité. *On le note N*

La **série statistique des effectifs** est la fonction qui à chaque valeur du caractère (modalité) associe l'effectif de cette modalité. Elle est définie le plus souvent à l'aide d'un tableau.

7. Fréquence / Pourcentage

La **fréquence** f_i d'une modalité est le **nombre compris entre 0 et 1** obtenu en divisant son effectif par l'effectif total. $f_i = \frac{n_i}{N}$ où n_i désigne l'effectif de la modalité x_i et N l'effectif total.

Remarques : On peut obtenir le pourcentage correspondant en multipliant la fréquence par 100.

La somme des fréquences est égale à 1. La somme des pourcentages est égale à 100.%

Exemple 1 :

Dans un village comptant 200 familles, on étudie le nombre d'enfants par famille.

La **population** est constituée des 200 familles, sa taille est $N=200$. Le **caractère étudié** est le nombre d'enfants, c'est un **caractère quantitatif**. Si au terme de l'étude on obtient des valeurs 0, 1, 2, 3, 4, 5 ou 6 le caractère possède ces **7 modalités**. On peut établir un tableau de résultats :

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	
modalités	0	1	2	3	4	5	6	TOTAL
Effectifs	48	62	35	26	15	9	5	$N=200$
Fréquences	$\frac{48}{200} = 0,24$	0,31	0,175	0,13	0,075	0,045	0,025	1

distribution des effectifs ou série des effectifs.

Exemple 2 :

On étudie la couleur des yeux chez 40 élèves d'un lycée.

La **population** est constituée par les 40 élèves, sa taille est $N = 40$. Le **caractère étudié** est la couleur des yeux, c'est un **caractère qualitatif**. Si au terme de l'étude on obtient des valeurs bleu, marron, noir, ou vert, le caractère possède ces **4 modalités**.

On peut établir un tableau de résultats :

	x_1	x_2	x_3	x_4	
modalités x_i	Bleu	Marron	Vert	Noir	TOTAL
effectifs n_i	15	12	8	5	$N=40$
Fréquences f_i	$\frac{15}{40} = 0,375$	0,3	0,2	0,125	1
en %	37,5%	30%	20%	12,5%	100

II. Effectifs et fréquences cumulé(e)s

1. Définitions

On note x_i la $i^{ème}$ modalité prise par un caractère quantitatif.

L'effectif (fréquence) cumulé(e) croissant(e) (ECC, resp. FCC) de x_i est la somme des effectifs (fréquences) des valeurs inférieures ou égales à x_i .

L'effectif (fréquence) cumulé(e) décroissant(e) (ECD, resp. FCD) de x_i est la somme des effectifs (fréquences) des valeurs supérieures ou égales à x_i .

2. Exemples

Exemple 1 : enquête sur le nombre d'enfants par famille dans un village

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	
Modalités	0	1	2	3	4	5	6	TOTAL
Effectifs	48	62	35	26	15	9	5	200
ECC	48	110	145	171	186	195	200	

prise d'initiative

Point méthode : On calcule les effectifs cumulés croissants (ECC) de la *gauche* vers la *droite*.

On peut lire par exemple que *145* familles ont *au plus* 2 enfant (soit 0, 1 ou 2 enfants).

Exemple 3 : enquête sur la masse des nouveaux nés dans un hôpital.

	x_1	x_2	x_3	x_4	
Masses (kg)	$[2,5 ; 3[$	$[3 ; 3,5[$	$[3,5 ; 4[$	$[4 ; 4,5[$	TOTAL
Effectifs	17	8	5	12	42
Fréquences	0,40	0,19	0,12	0,29	1
FCD	1	0,60	0,41	0,29	

$$f_1 = \frac{n_1}{N} = \frac{17}{42}$$

Point méthode : les données sont regroupées en intervalles semi-fermés appelés classes.

On calcule les fréquences cumulées décroissantes de la *droite* vers la *gauche*.

On peut lire par exemple que *0,41* des nouveaux nés ont une masse *supérieure ou égale* à *3,5 kg*.

III. Représentations graphiques d'une série statistique

Le choix de la représentation graphique dépend de la série statistique et du type d'interprétation souhaité.

1. Caractère qualitatif

Pour un caractère qualitatif, on peut faire :

- un diagramme à bâtons (ou en barres) ;
- un diagramme circulaire.

Exemple 2 : enquête sur la couleur des yeux dans un groupe de 40 élèves d'un lycée

Modalités	Bleu	Marron	Vert	Noir	TOTAL
Effectifs	15	12	8	5	40
angle	135	108	72	45	360

Point méthode : l'angle au centre est proportionnel à l'effectif de la modalité. On peut donc le calculer par les techniques usuelles sur les tableaux de proportionnalité (addition ou soustraction, multiplication ou division, ou encore par produit en croix).

Diagramme circulaire

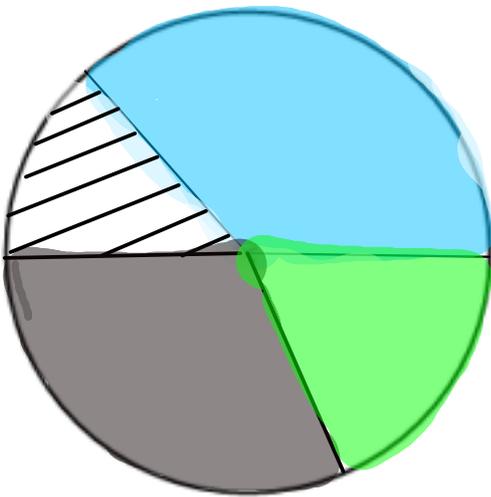
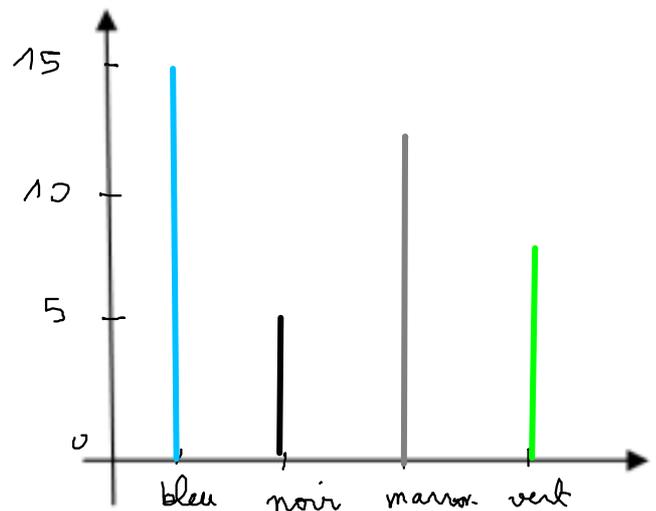


Diagramme en barres



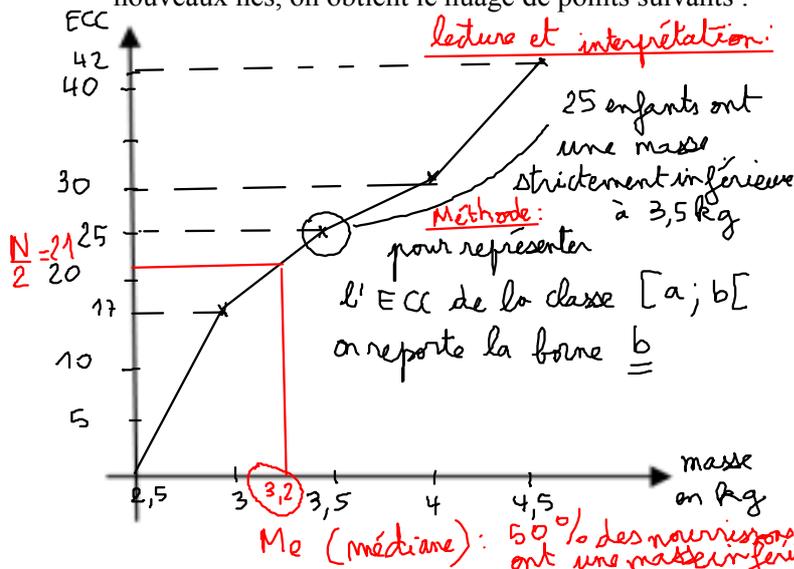
2. Caractère quantitatif

Pour un caractère quantitatif, on peut utiliser les représentations graphiques précédentes ou encore faire un nuage de points.

Point méthode : lorsque les données sont réparties en classe, on travaille avec le centre de la classe.

Exemple 3 :

Avec les effectifs cumulés croissants des masses des nouveaux nés, on obtient le nuage de points suivants :



Avec les fréquences cumulées croissantes des masses des nouveaux nés, on obtient le diagramme suivant :

