

Chapitre :

Statistiques



Fiche de cours des trois premières parties de ce chapitre.

I. Vocabulaire

Définition (Population) Une population est un ensemble de personnes ou d'objets sur lesquels on peut prélever des renseignements communs.

Un élément de la population est appelé individu.

L'effectif total de la population est le nombre d'individus.

Exemple les élèves d'une classe; les planches produites par une scierie; ...

Définition (Caractère) Un caractère est ce qui est étudié sur la population.

Exemple pour chaque des élèves on peut s'intéresser à : sa note à un devoir, son régime (P, DP, E), la couleur de ses yeux, ...

Exemple pour chaque planche de la scierie, on peut s'intéresser à : sa taille, le nombre de défauts, l'essence du bois, etc.

Une **étude statistique** est l'étude d'un ou plusieurs caractères précis sur une population.

Il y a deux types de caractères :

- les caractères **qualitatifs** sont ceux qui ne donnent pas de valeurs pour lesquelles une moyenne a un sens.
(exemples : régime, couleur des yeux, essence du bois).
- les caractères **quantitatifs** sont ceux que l'on peut mesurer, compter, ordonner.
(exemples : la note, le nombre de défauts, la taille).

Parmi ceux-ci on distingue deux types :

- ★ les caractères **continus** sont ceux dont les valeurs peuvent être pris sur des intervalles.
C'est souvent le cas des mesures (longueur, temps, ...).
(exemples : la taille peut prendre une infinité de valeurs si l'on mesure très précisément)
- ★ les caractères **discrets** sont ceux qui prennent un nombre restreint (voire fini) de valeurs.
(exemples : les notes, le nombre de défauts)

Pour les caractères qualitatifs il y a en général peu de choses à dire (la couleur d'yeux la plus fréquente). On s'intéresse ici par la suite aux caractères quantitatifs (on peut ordonner et faire la moyenne des notes entre autres).


II. Présentation par tableau

Une série statistique est la donnée brute de la valeur du caractère de chacun des individus composant la population étudiée.

Exemple on obtient une liste de notes : 15 ; 12 ; 08 ; 09 ; 12 ; 14 ; 09 ; 10 ; 17 ; 05.

Pour avoir une vision plus claire on préfère regrouper les valeurs dans un tableau, en les ordonnant. On obtient alors le tableau des effectifs :

Note	05	08	09	10	12	14	15	17
Effectif	1	1	2	1	2	1	1	1

 Bien identifier le caractère (note) et l'effectif, selon les tableaux ce n'est pas toujours évident : les deux lignes d'effectif et de caractère pouvant être appelée toutes deux « nombre de ... » ; c'est le cas par exemple si le caractère étudié est le nombre de pétales dans une population de fleurs.

On lit facilement **l'étendue de la série** : la valeur minimale (5) et la valeur maximale (17).

On peut alors éventuellement s'intéresser aux **effectifs cumulés croissants**.

Il s'agit de compter le nombre d'individus ayant leur caractère inférieur ou égal au caractère indiqué :

Note	05	08	09	10	12	14	15	17
Effectif cumulé croissant	1	2	4	5	7	8	9	10

L'effectif de la dernière colonne est donc l'effectif total. On peut alors voir directement le nombre de personnes ayant eu moins de 10 par exemple.

Au lieu des effectifs, on peut faire le même genre de tableau avec les **fréquences** (éventuellement cumulées aussi).

La fréquence est donnée par : $\frac{\text{effectif}}{\text{effectif total}}$.

On peut calculer la fréquence en pourcentage ; elle est donnée par : fréquence \times 100.

Voici le tableau donnant les fréquence cumulées croissantes en pourcentage :

Note	05	08	09	10	12	14	15	17
fréquences cumulées croissantes en %	10	22	40	50	70	80	90	100

On finit nécessairement par 100.

Dans le cas d'un caractère continu, on utilise des intervalles pour les valeurs (voir page 118).

III. Présentation par graphique

Il y a plusieurs représentations possibles.

Dans le cas d'un caractère discret, on peut utiliser des diagrammes en bâton (ou à barres) (voir page 17).

Dans le cas d'un caractère continu, on peut utiliser des histogrammes (voir page 27), pour lesquels l'**aire** des rectangles est **proportionnelle** aux effectifs (donc aux fréquences).

Il existe aussi les diagrammes circulaires ou semi-circulaires. Ici, c'est la mesure de l'angle qui est proportionnelle aux effectifs.

Pour les effectifs cumulés croissants on peut tracer une courbe (voir page 25), la **courbe représentative (ou polygone) des effectifs cumulés croissants**.

► **Exercices** : 9p15

► **Exercice** : fiche d'exercices (1)

IV. Moyenne, médiane et quartiles

Définition Supposons que le caractère prenne p valeurs x_1, \dots, x_p avec des effectifs respectifs n_1, \dots, n_p . On note N l'effectif total ($N = n_1 + \dots + n_p$).

La **moyenne** de la série est alors donnée par : $\bar{x} = \frac{n_1x_1 + \dots + n_px_p}{N}$.

Exemple Dans l'exemple du cours, la note moyenne est

$$\bar{x} = \frac{1 \times 5 + 1 \times 8 + 2 \times 9 + 1 \times 10 + 2 \times 12 + 1 \times 14 + 1 \times 15 + 1 \times 17}{10} = 11,1$$

Remarque on peut aussi utiliser les fréquences f_i au lieu des effectifs n_i : $\bar{x} = f_1x_1 + \dots + f_px_p$.

► **Exercices** : 25,26,28p18, 31p19

Quand les valeurs sont données sous formes de classes, c'est à dire d'intervalles de la forme $[a; b[$, on utilise le centre des classes comme valeurs pour les x_i , c'est à dire $x = \frac{a+b}{2}$ (c'est la valeur située « au milieu » de l'intervalle).

► **Exercices** : 29,30p19 (avec classes de valeurs)

Définition La **médiane** d'une série est la plus petite valeur du caractère notée Me telle que au moins 50% des individus ont une valeur du caractère inférieure ou égale à Me

La médiane ne se calcule pas, mais se détermine. Pour cela on peut s'aider du tableau des fréquences cumulées. La médiane est la valeur du caractère qui fait atteindre ou dépasser les 50%.

Exemple Dans notre exemple, à la quatrième note la fréquence cumulée atteint exactement 50%. La médiane est donc cette quatrième note, à savoir 10.

Définition Le **premier quartile** est la plus petite valeur du caractère notée Q_1 qui fait atteindre ou dépasser les 25% de fréquence cumulée. Le **troisième quartile** est la plus petite valeur du caractère notée Q_3 qui fait atteindre ou dépasser les 75% de fréquence cumulée.

Remarque la médiane pourrait être considérée comme un deuxième quartile.

Exemple Dans notre exemple, $Q_1 = 09$ et $Q_3 = 14$.

Remarque Lorsque l'on ne dispose pas des fréquences cumulées croissantes, il faut au moins disposer des effectifs cumulés croissants. Dans ce cas, en notant N l'effectif total :

- Pour obtenir Me on calcule dans un premier temps $\frac{N}{2}$; Me est alors la plus petite valeur pour laquelle l'effectif cumulé croissant dépasse $\frac{N}{2}$.
- Pour les quartiles le principe est le même, mais en calculant dans un premier temps $\frac{N}{4}$ et $\frac{3N}{4}$ respectivement.

Définition On appelle écart inter-quartile la différence $Q_3 - Q_1$.

► **Exercices** : 16,17,20p17 puis fiche d'exercices (2)

► **Exercice** : (en DM, expérience statistique)