

Statistiques descriptives

I. Vocabulaire général

Afin d'illustrer toutes les définitions de vocabulaire que nous allons voir nous allons prendre un exemple : les notes au DM1 des élèves de la classe.

Population : Ensemble sur lequel porte l'étude. (Ex : la classe de 2^o1.)

Individus : Eléments qui composent la population. (Ex : les élèves de 2^o1.)

Caractère : Aspect que l'on observe chez les individus. (Ex : les notes du DM1.)

Il y a plusieurs types de caractères

- **Caractère qualitatif** : Le caractère n'est pas mesurable. (Ex : si on avait demandé la couleur des yeux au lieu des notes.)
- **Caractère quantitatif** : les valeurs prises par le caractère sont des nombres.
 - ★ **Quantitatif discret** : les valeurs sont isolées. (Ex : c'est le cas pour les notes.)
 - ★ **Quantitatif continu** : les valeurs constituent un intervalle. (Ex : si on avait demandé le poids au lieu des notes.)

Série statistique : Ensemble des valeurs collectées.

(En fonction du caractère étudié on parlera par exemple de série quantitative discrète ... etc.)

Effectif : Pour une valeur x_i du caractère c'est le nombre n_i d'individus de la population prenant cette valeur.

Effectifs cumulés croissants : Ils s'obtiennent en ajoutant au fur et à mesure les effectifs des valeurs précédentes.

Fréquence : Pour une valeur x_i c'est le quotient f_i de l'effectif n_i de la valeur par l'effectif total N . On a donc $f_i = \frac{n_i}{N}$ et $f_1 + f_2 + \dots = 1$.

Lorsque la série est connue par ses fréquences on parle de **distribution de fréquences**.

On parlera également de **fréquences cumulées croissantes** ... etc.

Lorsqu'on regarde une « liste » de données statistiques la quantité d'informations est généralement très grande et notre cerveau n'arrive pas à en retirer d'idées générales.

Le but des statistiques est de résumer toutes ces valeurs à l'aide de quelques « outils pertinents ».

II. Outils de position et de dispersion

- 1) Dans toute la suite **quand** on travaillera avec une série quantitative discrète, on notera x_i les valeurs du caractère, n_i les effectifs correspondants, N l'effectif total et f_i la fréquence de la valeur x_i . On aura alors un tableau statistique du type suivant :

valeurs	x_1	x_2	...	x_p
effectifs	n_1	n_2	...	n_p

et $n_1 + n_2 + \dots + n_p = N$

ou bien

valeurs	x_1	x_2	...	x_p
fréquences	f_1	f_2	...	f_p

et $f_1 + f_2 + \dots + f_p = 1$

2) La somme $x_1 + x_2 + \dots + x_n$ peut aussi s'écrire : $\sum_{i=1}^n x_i$ et se lit « somme de i égal 1 à n des x indice i ».

1) Outils de dispersion

Définition 1 : On appelle **étendue** d'une série statistique la différence entre la plus grande valeur du caractère et la plus petite.

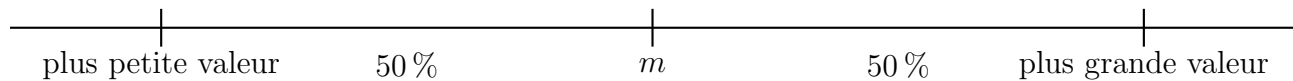
→ A donner pour les notes du DM1.

Remarque : L'étendue d'une série statistique traduit effectivement la « dispersion » de cette série.

2) Outils de position

a) Médiane

Définition 2 : La **médiane** d'une série statistique ordonnée est une valeur m telle qu'il y ait autant de valeurs supérieures ou égales à m que de valeurs inférieures ou égales à m .



Lorsque la série a un **effectif total impair** N :
on range les valeurs du caractère par ordre croissant et la médiane est le terme central.

Lorsque la série a un **effectif total pair** N :
on range les valeurs du caractère par ordre croissant et la médiane est la moyenne des deux valeurs centrales.

Exemple : Pour la série 3-4-6-7-12-12-13
on a $m = 7$

Exemple : Pour la série 3-4-6-7-12-12
on a $m = 6,5$

→ A donner pour les notes du DM1.

Remarque : La médiane est une sorte de « point d'équilibre » de la série.

b) Mode

Définition 3 : Dans le cas d'une série statistique (quantitative) discrète le **mode** est **une** valeur qui a le plus grand effectif associé.

Définition 4 : Dans le cas d'une série statistique (quantitative) continue, la **classe modale** est **une** classe correspondant au plus fort effectif.

→ A donner pour les notes du DM1.

Remarque : Une série peut avoir plusieurs modes.

Le mode présente un intérêt surtout si son effectif est nettement supérieur aux autres.

c) Moyenne

Définition 5 : La **moyenne** est la somme des valeurs du caractère pondérées par les effectifs et divisée par l'effectif total. Elle est notée \bar{x} .

$$\bar{x} = \frac{\sum_{i=1}^n n_i x_i}{N}$$

→ A donner pour les notes du DM1.

Remarque : Dans le cas d'une série quantitative continue ou d'un simple regroupement en classe on calcule une valeur approchée de la moyenne en choisissant comme valeurs du caractère les centres des classes et comme effectifs les effectifs des classes.

III. Calculs de moyennes

Propriété 1 : (Linéarité de la moyenne)

Si une série de valeurs x_i a pour moyenne \bar{x} , la série de valeurs $a x_i + b$ (a et b étant deux réels) a pour moyenne $a\bar{x} + b$.

→ Démonstration

Exemple : Si on enlève un point à tout le monde au DM1 on enlève 1 point à la moyenne.
Si on met ces notes sur 40 au lieu de 20 on double la moyenne.

Propriété 2 : (A partir d'une distribution de fréquences)

Si une série de valeurs x_i a pour distribution de fréquences les f_i , alors la moyenne \bar{x} vaut :

$$\bar{x} = \sum_{i=1}^n f_i x_i$$

→ Démonstration

Théorème 1 : (A partir de la moyenne de sous-groupes)

Si on répartit les éléments d'une série en deux sous-groupes disjoints d'effectifs P et Q et de moyennes respectives \bar{x} et \bar{y} alors la moyenne \bar{X} de cette série est :

$$\bar{X} = \frac{P\bar{x} + Q\bar{y}}{P + Q}$$

Exemple : La moyenne des 25 notes obtenues par les garçons à un devoir de français est égale à 11 sur 20, la moyenne des 2 notes obtenues par les filles est de 12,5 sur 20.
Calculons la moyenne de la classe.

$$\text{On a } \bar{x} = \frac{25 \times 11 + 2 \times 12,5}{31} \simeq 11,11.$$

Remarque : $\bar{x} \neq \frac{11 + 12,5}{2}$.