Chap 6 :

Statistiques

I. Vocabulaire général

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

1,5 - 3,75 - 3,75 - 3,75 - 4,5 - 5,5 - 6,25 - 7 - 8 - 8 - 8,25 - 8,5 - 9 - 10 - 10,75 - 11,25 - 11,75 - 12,75 - 14,5 - 16 - 16 - 17,75

Population : Ensemble sur lequel porte l'étude. (Ex : la classe de 2°1.)

Individus : Eléments qui composent la population. (Ex : les élèves de 2°1.)

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 votre chrono au 100 mètres au lieu des notes.)
    On les range alors par classes. par exemple $[1\,;2[$, $[2\,;3[\ldots$

Série statistique : Ensemble des valeurs collectées.
(On les représente graphiquement par un histogramme ou un diagramme en bâtons.)

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 $f_i = \dfrac{n_i}{N}$ et $f_1 + f_2 + \cdots = 1$.

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 <u>résumer</u> toutes ces valeurs à l'aide de quelques « outils pertinents ».

II. Outils de position et de dispersion

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, f_i la fréquence de la valeur x_i et N l'effectif total . On aura alors un tableau statistique du type suivant :

notess	x_1	x_2	$\dots$	x_p
effectifs	n_1	n_2	$\dots$	n_p

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

ou bien

notes	x_1	x_2	$\dots$	x_p
fréquences	f_1	f_2	$\dots$	f_p

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

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

⟶ Pour les notes du DM1 on a le tableau :

notes	1,5	3,75	4,5	5,5	6,25	7	8	8,25	8,5	9	10	10,75	11,25	11,75	12,75	14,5	16	17,75
effectifs	1	3	1	1	1	1	1	2	1	1	1	1	1	1	1	1	2	1

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.

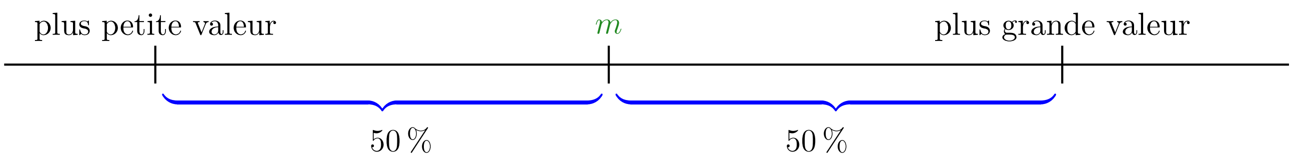
⟶ Pour les notes du DM1 c'est $17,75 - 1,5 = 16,25$.

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

2) Outils de position

1. Médiane

Définition 2 : La médiane d'une série statistique <u>ordonnée</u> 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.

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

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 $3-4-6-7-12-12$ on a $m=6,5$

⟶ Pour les notes du DM1 on a 22 notes il faut donc regarder la 11ème et la 12ème il s'agit de $8,25$ et 8,5. La médiane vaut donc 8,375.

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

2. Mode

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

⟶ Pour les notes du DM1 il n'y a que 3,75.

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

Remarque : Une série peut avoir plusieurs modes.
Le mode présente un intérêt surtout si son effectif est nettement supérieur aux autres.

3. 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 $\overline{x}$.

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

⟶ Pour les notes du DM1 c'est 9,02.

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.
⟶ A voir en TD.

III. Calculs de moyennes

Proposition 1 : (Linéarité de la moyenne)
Si une série de valeurs x_i a pour moyenne $\overline{x}$, la série de valeurs $a\,x_i + b$ (a et b étant deux réels) a pour moyenne $a\overline{x} + b$.

⟶ *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.

Proposition 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 $\overline{x}$ vaut :

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

$\longrightarrow$ *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 $\overline{x}$ et $\overline{y}$ alors la moyenne $\overline{X}$ de cette série est :

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

Exemple : La moyenne des 21 notes obtenues par les garçons au DM1 est égale à 9,07 sur 20, la moyenne des notes obtenues par les filles (ici qu'une seule) est 8 sur 20.
Calculer la moyenne de la classe.

On a $\overline{x} = \dfrac{21 \times 9,07 + 1 \times 8}{22} \approx 9,02.$

Remarque : *Attention* : $\overline{x} \neq \dfrac{9,07+8}{2}$.