Chapitre 2.

Représentation numérique des données.

1- Caractéristiques de position

1-1 Moyenne arithmétique pondérée

Définition

Dans le cas d'une variable discrète, la moyenne arithmétique est donnée par :

$$\bar{x}=\frac{1}{N}\sum_{i=1}^{p} n_i x_i \text{ ou } \bar{x}=\sum_{i=1}^{p} f_i x_i$$

n_i est l'effectif pour la modalité x_i

La moyenne arithmétique de données groupées

Autant que faire se peut, ce type de calcul est à éviter car source d'imprécision et d'erreur trop importantes. Cependant, on peut être confronter à une situation où seules des données groupées sont disponibles. Dans ce cas, et seulement dans celui-là, on peut être autorisé à calculer une moyenne à partir de classes. On agit alors comme si tous les résultats d'une classe se trouvaient au centre de celle-ci. La moyenne de la distribution est alors calculée à partir des valeurs centrales des classes pondérées par leurs effectifs respectifs.

$$\bar{x}=\frac{1}{N}\sum_{i=1}^{p} n_i c_i$$

Le résultat est au final assez peu différent de celui obtenu par la moyenne arithmétique simple car la moyenne arithmétique simple, vu le nombre important de valeurs et compte tenu de **la structure** de l'échantillon, tient compte, de façon presque naturelle, du poids des individus en attribuant implicitement à chaque individus le poids de sa catégorie.

Application

Soit la série statistique X tels que X={x1, x2, ….,xk } sont les valeurs du caractère étudié et n1, n2, ……,np les effectifs correspondants, la moyenne de la série statistique noté $\bar{x}$ est :

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \ldots\ldots + n_px_p}{N} = \frac{\Sigma n_i x_i}{N} = \Sigma\, f_i x_i \qquad \text{car} \qquad \frac{n_i}{N} = f_i$$

Exemple 1.1:

- cas variable statistique discontinue « discrète »

un élève a obtenu en mathématiques les 5 notes suivantes : 5 ; 9 ; 12 ; 15 ; 19 .

Question : Quelle est sa moyenne $\bar{x}$?

$$\bar{x} = \frac{1}{5}[5+9+12+15+19] \Rightarrow \bar{x} = 12. car\ n_i = 1. \forall i \in \{1,2,3,4,5\}$$

cas variable statistique continue :

Soit le tableau statistique suivant :

Moyenne	ci	Effectif
[7 ; 9 [	8	7
[9 ; 11 [	10	9
[11; 13 [	12	3
[13 ; 15 [	14	6
[15 ; 17[	16	5

Centre de la classe [a-b[est $c_i = \frac{a+b}{2}$.

Donc la moyenne de cette série est :

$$\bar{x} = \frac{1}{30}\sum_{i=1}^{5} n_i\, c_i = \frac{1}{30}[8 \times 7 + 10 \times 9 + 12 \times 3 + 14 \times 6 + 16 \times 5]$$

$$\Rightarrow \bar{x} = 11.533$$

1.2 . La variance

Définition

La variance de la série statistique se note σ_x^2 (ou encore $V(x)$) et se définit comme suit :

$$\mathrm{V}(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Elle correspond à la moyenne des carrés des différences entre les observations et leur moyenne $\bar{x}$.

Dans le cas de n observations, ordonnées dans un tableau statistique (x_i, n_i) , présentant r modalités.

$$\mathrm{V}(X) = \frac{1}{n}\sum_{i=1}^{n} n_i(x_i - \bar{x})^2$$

La variance (ou fluctuation) est la moyenne arithmétique des carrés des écarts à la moyenne.

Définition

L'écart-type, noté σ_x , est la racine carrée de la variance.

$$\sigma_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n} n_i \cdot (x_i - \bar{x})^2} \quad \textbf{Ou} \quad \sigma_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

1-3 La médiane

Définition

La médiane d'une série statistique rangée dans l'ordre croissant (x_1 , x_2 , x_3 , x_4 , ..., x_n)noté Me .
défini de la façon suivante :
Si n = 2p est pair , Me est le centre de l'intervalle [x_p ; x_{p+1}].
Si n est impair, Me est le nombre x_p où p = (n + 1)/2 .

Exemple 2.1

a- **Médiane d'une série discrète**

on fait une étude statistique sur les 50 notes attribuées par un jury à un examen, voici les résultats obtenus en classant ces notes par ordre croissant (variable discrète).

X_i	Effectif	$N_i(c)$
2	5	5
5	7	12
7	3	15
9	7	22
10	8	30

n = 30 est pair , il faut donc prendre le centre de [7 ; 9]

Utilisons la colonne des effectifs cumulés pour déterminer la médiane : il y a 50 notes, la 15ème note est 7 et la 16ème : 9.

Dans le tableau il n'y a pas de valeur partageant la série statistique en deux groupe de même effectif, (l'effectif total est pair) dans ce cas l'intervalle médian est [7;9] et on prendre pour médiane le centre de cet intervalle : 08

b- Médiane d'une série continue

Si la variable est continue (regroupement par intervalle des résultats) le calcul de la médiane se fait autrement :

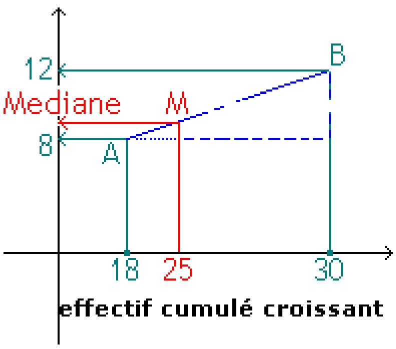
Notes	Effectifs	Effectifs cumulés
[0 ; 5[	10	10
[5 ; 8[	8	18
[8 ; 12[	12	30
[12 ; 15	11	41
[15 ; 20	9	50
	50	

Utilisons la colonne des effectifs cumulés pour déterminer la médiane : il y a 50 notes, 50 % de l'effectif total c'est 25, la médiane est ici la note

correspondant à l'effectif cumulé 25.
D'après la colonne "effectif cumulé" :

- 18 personnes ont moins de 8
- 30 personnes ont moins de 12

La médiane se trouve donc dans l'intervalle [8;12[(appelée classe médiane) on va la déterminer par interpolation linéaire.



Les points A, M, B sont alignés ce qui se traduit par les droites (AM) et (AB) ont même coefficient directeur (ou on utilise le théorème de Thalès dans le triangle bleu) :

$$\frac{Me-8}{25-18}=\frac{12-8}{30-18}$$

$$\frac{Me-8}{7}=\frac{4}{12}$$

$$Me-8=\frac{4}{12}\times 7$$

$$Me=8+\frac{4}{12}\times 7\approx 10{,}33$$

La médiane est environ 10,33

50 % environ des personnes ont eu moins de 10,33 et 50 % plus de 10,33 .

2-Caractéristiques de dispersion (faire intro avec moyenne de classe)

2-1 **Étendue**

Définition :

L'étendue d'une série statistique noté **E** est la différence entre la plus grande et la plus petite valeur.

$$E=X_{max}-X_{min}$$

Exemple :

a) Cas simple :

Quelle est l'étendue de cette série : 24 ; 7 ; 1 ; 9 ; 46 ; 15.

La plus grande valeur est 46 et la plus petite est 1.

46 - 1 = 45, donc l'étendue de la série est égale à 45.

b) Étendue à partir d'un tableau :

Quelle est l'étendue de la série ci-dessous:

Valeur	5	7	12	15
Effectif	2	7	9	25

La plus grande valeur est 15 et la plus petite est 5 donc l'étendue est égale à
15 - 5 = 10

2-3 Les quartiles

Définition

- Le premier quartile d'une série statistique est la plus petite valeur Q1 telle qu'au moins un quart des valeurs sont inférieures ou égales à Q1.

- Le troisième quartile d'une série statistique est la plus petite valeur Q3 telle qu'au moins trois quarts des valeurs sont inférieures ou égales à Q3.

Exemple :

Déterminer le premier et le troisième quartile de la série :

24 ; 7 ; 2 ; 9 ; 13 ; 5 ; 32 ; 8; 15.

→ On commence par ordonner la série : 2 ; 5 ; 7 ; 8 ; 9 ; 13 ; 15 ; 24 ; 32.

→ On calcule l'effectif total : 9.

→ 9 ÷ 4 = 2,25 , donc le premier quartile est la 3ème valeur : Q1 = 7

→ 3×(9 ÷ 4) = 6,75 , donc le premier quartile est la 7ème valeur : Q3 = 15

2 ; 5 ; 7 ; 8 ; 9 ; 13 ; 15 ; 24 ; 32.

Conclusion : Q1 = 7 et Q3 = 15.

2-4 Le Mode

a- Mode d'une série discrète

le mode est la valeur de la variable qui correspond à l’effectif maximal ou à la fréquence maximale.

Exemple :

Soit le tableau statistique suivant :

Valeur	5	7	12	15	N
Effectif	2	22	9	7	40

Le mode **Mo=7**. Car max $n_j=n_2=22$. Correspond valeur $X_2=7$

b- Médiane d'une série continue

Définition

Le **mode** est un complément à la moyenne et à la médiane. Il permet de donner un indicateur statistique de tendance centrale à un ensemble de données.

Le mode d'un ensemble d'observations est la valeur la plus fréquemment rencontrée.

Dans le cadre de distributions regroupées en classe, le calcul du mode se fait à partir de la classe où la fréquence est la plus élevée avec la formule ci-dessous :

Soit la classe modale $[e_i - e_{i+1}[$. Et $a_i = e_{i+1} - e_i$.

$$Mo = e_i + a_i \frac{\Delta_1}{\Delta_1 + \Delta_2}$$

e_i = Limite inférieur de la classe identifiée la plus fréquentale, dite modale. Δ_1 = Différence entre la fréquence de la classe modale et la classe précédente. Δ_2 = Différence entre la fréquence de la classe modale et la classe suivante.

a_i = Longueur de la classe modale.

Exemple :

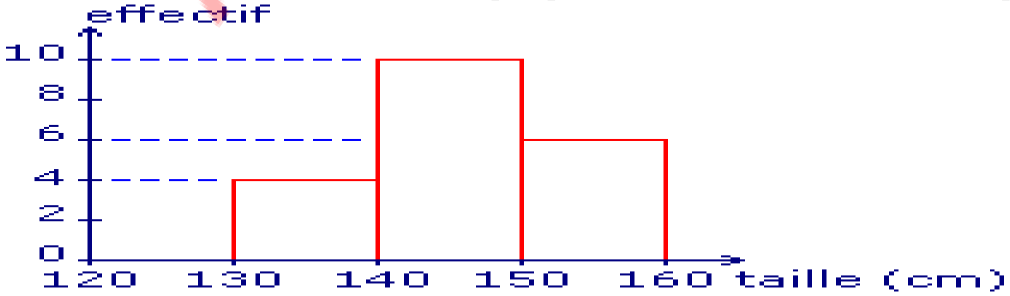
• Voici la répartition des tailles des enfants d'un club de sport.
Ainsi 4 enfants mesurent entre 130 cm et 140 cm.

Taille (en cm)	130-140	140-150	150-160
Effectif	4	10	6

• Les valeurs du caractère étudié (la taille) se présentent sous forme d'intervalles.
On construit un histogramme avec :

- sur l'axe horizontal, les tailles ;
- sur l'axe vertical, les effectifs.

• Les hauteurs des barres sont proportionnelles aux effectifs représentés.



On observe, sur le graphique construit, que la catégorie la plus nombreuse concerne les enfants qui mesurent entre 140 cm et 150 cm.

$$\boldsymbol{Mo} = \boldsymbol{e_i} + \boldsymbol{a_i}\frac{\Delta_1}{\Delta_1+\Delta_2} = 140 + 10\frac{10-4}{(10-4)+(\mathbf{10-6})} = 140 + 10\frac{6}{10} = 146$$

Donc

Le mode Mo = 146

………suite