CHAPITRE 1

STATISTIQUES DESCRIPTIVES À UNE VARIABLE

1.1 Introduction

1.1.1 Vocabulaire

Définition 1.1 *La statistique descriptive est l'ensemble des méthodes permettant le* **recueil** *de données, le* **classement** *et la* **réduction** *de ces données .*

Une étude statistique descriptive s'effectue sur une **population** (des étudiants, des électeurs, des animaux, des villes, des voitures...) dont les éléments sont des **individus** et consiste à observer et étudier un même aspect sur chaque individu, nommé **caractère** ou **variable** (nombre de modules validés, opinion politique, taille, nombre d'habitants, consommation...).
Il existe deux types de caractère :

1. **quantitatif** : c'est un caractère auquel on peut associer un nombre exprimant une quantité c'est-à-dire, pour simplifier, que l'on peut "mesurer". On distingue alors deux types de caractères quantitatifs :
   - **discret** : c'est un caractère quantitatif qui ne prend qu'un nombre fini de **valeurs isolées**. Par exemple le nombre d'enfants d'un couple et le nombre de modules validés par un étudiant.
   - **continu** : c'est un caractère quantitatif qui, théoriquement, peut prendre toutes les valeurs d'un intervalle de l'ensemble des nombres réels (ou bien entre deux valeurs quelconques prises par le caractère on peut toujours trouver une valeur intermédiaire prise par le caractère).
     Ses valeurs sont en général regroupées en **classes**. Par exemple la taille ou le poids d'un individu, le nombre d'heures passées devant la télévision.
2. **qualitatif** : lorsqu'il correspond à une catégorie ou une qualité comme la profession, l'opinion politique, la nationalité, le sexe, la taille vestimentaire, la mention et la couleur des yeux. Dans ce dernier cas, "Noir", "Vert", "Marron" etc... sont les **modalités** du caractère. On distingue des caractères qualitatifs **ordinaux** ou **nominaux**, selon que les modalités peuvent être naturellement ordonnées ou pas. La taille vestimentaire et la mention sont des caractères qualitatifs ordinaux, par contre la profession, l'opinion politique, la nationalité, la couleur des yeux, le sexe sont des caractères qualitatifs nominaux.

En général le nombre d'individus d'une population nommé **effectif total** est très élevé, donc une étude statistique descriptive de tous les éléments de la population est compliquée, coûteuse, longue à mener etc..., on est donc obligé de s'intéresser à une sous-population d'effectif très réduit par rapport à la population mère appelée **échantillon**. L'échantillon doit être **représentatif** (c'est-à-dire que son choix soit fait de telle façon que les proportions des différentes catégories de la population soient respectées) pour que l'étude sur l'échantillon reflète (ou soit significative par rapport à) la population mère.
Dans la suite, on fera des études statistiques sur des caractères quantitatifs.

(P)

1.1.2 Représentation d'une série statistique quantitative

1.1.2.1 Exemple discret n° 1

Une section composée de 30 étudiants a passé un examen.
La liste des 30 notes obtenues est notée dans un premier temps par :
– **Série représentée par une liste**
$(x_i)_{1\leqslant i\leqslant 30} = (8, 11, 5, 12, 2, 17, 7, 8, 19, 2, 10, 4, 7, 7, 11, 14, 7, 8, 12, 5,\ 10, 8, 5, 10, 11, 8, 12, 10, 14, 12)$.
On peut aussi regrouper les éléments de la série précédente dans un tableau où on ne considère que les 11 valeurs distinctes des x_i, notées encore $(x_i)_{1\leqslant i\leqslant 11}$ et les effectifs correspondants $(n_i)_{1\leqslant i\leqslant 11}$ (l'**effectif** n_i associé à une valeur x_i est le nombre de fois qu'apparaît cette valeur dans la série listée). La liste précédente peut ainsi être présentée dans le tableau suivant :
–**Série regroupée par valeurs dans un tableau**

Notes x_i	2	4	5	7	8	10	11	12	14	17	19
Effectifs n_i	2	1	3	4	5	4	3	4	2	1	1

On peut enfin regrouper les valeurs de cette même série en considérant qu'on peut répartir les notes dans 4 **classes** (intervalles d'intersections 2 à 2 vides et de réunion contenant l'ensemble des valeurs) d'**étendue** ou d'**amplitude** commune (c-à-d de longueur) égale à 5 et à chacune des 4 classes on associe le nombre n_i appelé **effectif** de la i-ème classe désignant le nombre de valeurs x_j, $1 \leqslant j \leqslant n$ contenues dans cette i-ème classe. On représente ces données dans le tableau suivant :
–**Série regroupée par classes dans un tableau**

Classes de X	$[0, 5[$	$[5, 10[$	$[10, 15[$	$[15, 20]$
Effectifs n_i	3	12	13	2

1.1.2.2 Exemple continu n° 2

A la question " Quelle quantité d'eau buvez-vous par jour ?", cinquante personnes interrogées ont donné des réponses qui sont représentées dans le tableau suivant (on notera X la quantité d'eau bue par jour en Litres (L)) :

Classes de X	[0 ; 0,5[	[0,5 ; 1[	[1 ; 1,5[	[1,5 ; 2[	[2 ; 3[
Effectifs n_i	12	21	9	6	2

1.1.2.3 Résumé des représentations

En résumé, une série statistique se présente sous la forme :
1• listée : $(x_i)_{1\leqslant i\leqslant n}$, 2• regroupée par valeurs : $(x_i, n_i)_{1\leqslant i\leqslant p}$, $p \in \mathbb{N} - \{0, 1\}$, où p désigne le nombre de valeurs distinctes, 3• regroupée par classes : à la i^{eme} $(1 \leqslant i \leqslant p)$ classe on associe n_i le nombre de valeurs de la série contenues dans cette classe où p désigne le nombre classes.
La deuxième présentation citée ci-dessus est valable en général pour les séries discrètes avec valeurs répétées. Par contre pour le cas continu ou discret avec un effectif total élevé (dépassant en pratique 40), on choisit en général un groupement par classes, car la 3^{eme} représentation s'avère plus simple, rapide et commode mais moins précise. Désormais, on désignera par p le nombre de valeurs distinctes ou le nombre de classes selon les cas (discret ou continu).

1.1.2.4 Définitions : fréquence, fréquence cumulée et effectif cumulé

Définition 1.2 — *On appelle **fréquence (relative)** associée à une valeur x_i ou à la i-ème classe d'un caractère X $(1 \leqslant i \leqslant p)$ le nombre f_i défini par*

$$\boxed{f_i = \frac{n_i}{n}}$$

elle désigne la proportion d'individus ayant comme valeur x_i ou appartenant à la i-ème classe.

— *On appelle **fréquence cumulée (croissante)** associée à une valeur x_k (les $x_i, 1 \leqslant i \leqslant p$ qu'on suppose ordonnés dans l'ordre croissant(dans le cas contraire il faut commencer par les ordonner !)) ou à la k-ième classe d'un caractère X le nombre*

$$\boxed{F_k = \sum_{i=1}^{k} f_i}$$

elle désigne la proportion d'individus ayant comme valeurs $x_i \leqslant x_k$ ou appartenant aux k premières classes.

— *On appelle **effectif cumulé (croissant)** associé à une valeur x_k (les $x_i, 1 \leqslant i \leqslant p$ qu'on suppose ordonnés dans l'ordre croissant(dans le cas contraire il faut commencer par les ordonner !)) ou à la k-ième classe d'un caractère X le nombre*

$$\boxed{N_k = \sum_{i=1}^{k} n_i}$$

il désigne le nombre d'individus ayant comme valeurs $x_i \leqslant x_k$ ou appartenant aux k premières classes.

1.1.2.5 Propriétés importantes et conventions

— L'effectif total n d'une série est défini par :

$$\boxed{n = \sum_{i=1}^{p} n_i = n_1 + n_2 + \ldots + n_p}$$

— Les fréquences d'une série vérifient :

$$\boxed{0 \leqslant f_i \leqslant 1, \ \sum_{i=1}^{p} f_i = f_1 + f_2 + \ldots + f_p = 1}$$

— Dans la suite de ce cours et lorsqu'il s'agit d'une série continue on utilisera l'hypothèse de **répartition uniforme** c-à-d que les valeurs sont supposées équiréparties dans chaque classe.

1.2 Tableaux statistiques et graphes

Afin de représenter d'une façon graphique une série ou déterminer ses paramètres statistiques, on aura souvent besoin de dresser des tableaux statistiques qui contiennent un maximum d'informations nécessaires et utiles. L'intérêt de tels tableaux sera expliqué ultérieurement.

1.2.1 Tableaux statistiques

1.2.1.1 Tableaux : Exemples typiques

1. **Tableau statistique N° 1 associé au cas discret : n° 1**

	i=1	i=2	i=3	i=4	i=5	i=6	i=7	i=8	i=9	i=10	i=11	Sommes
x_i	2	4	5	7	8	10	11	12	14	17	19	*
n_i	2	1	3	4	5	4	3	4	2	1	1	30
N_i	2	3	6	10	15	19	22	26	28	29	30	*
f_i	2/30											1
F_i												*
$n_i x_i$												276
$n_i x_i^2$												2996

2. **Tableau statistique N° 2 associé au cas continu : n° 2**

	i=1	i=2	i=3	i=4	i=5	Sommes
Classes de X	$[0\ ;0.5[$	$[0.5\ ;1[$	$[1\ ;1.5[$	$[1.5\ ;2[$	$[2\ ;3[$	*
Centres x_i	0.25	0.75	1.25	1.75	2.5	*
n_i	12	21	9	6	2	50
N_i	12	33	42	48	50	*
f_i	12/50	21/50	9/50	6/50	2/50	1
F_i	12/50	33/50	42/50	48/50	1 ~~50~~	*
$n_i x_i$	3	15.75	11,25	10,5	5	45.5
$n_i x_i^2$	0,75	11,81	14,06	18,37	12,5	57.5

1.2.2 Graphes

1.2.2.1 Diagramme et polygone en bâtons

Définition 1.3 *Dans le cas d'une variable discontinue X représentée par une série groupée par valeurs, on utilise le diagramme en bâtons construit de la manière suivante :*

1. *on commence par placer les valeurs x_i de la variable X sur l'axe des abscisses,*
2. *on place sur l'axe des ordonnées les effectifs n_i (resp. effectifs cumulés N_i, fréquences f_i, fréquences cumulées F_i) correspondants,*
3. *on trace pour chaque valeur x_i un segment(trait plein) parallèle à l'axe des ordonnées d'extrémités les points de coordonnées $(x_i, 0)$ et (x_i, n_i) (resp. (x_i, N_i), (x_i, f_i), (x_i, F_i)).*

L'ensemble de ces segments est appelé diagramme en bâtons des effectifs (resp. effectifs cumulés, fréquences, fréquences cumulées). En joignant les sommets des bâtons déjà construits on obtient alors le polygone des effectifs (resp. effectifs cumulés, fréquences, fréquences cumulées).

Construisons sur deux figures différentes le diagramme en bâtons des effectifs et celui des effectifs cumulés associés au tableau statistique N° 1.(voir FIG.5.1 et FIG.5.2)

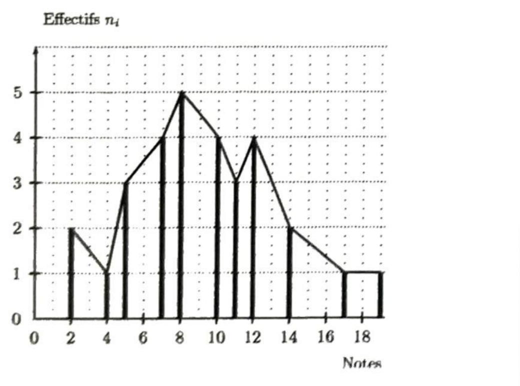

FIGURE 1.1 – Diagramme et polygone en bâtons des n_i

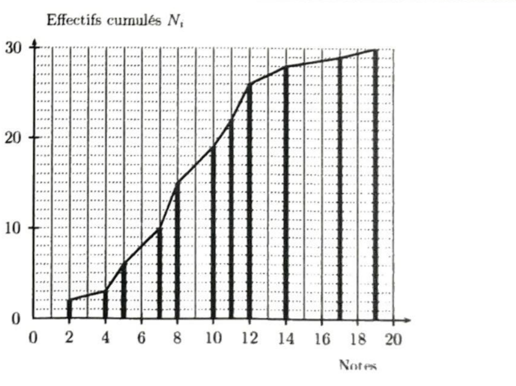

FIGURE 1.2 – Diagramme et polygone en bâtons des N_i

1.2.2.2 Histogramme et polygone des effectifs ou fréquences

Définition 1.4
1. *Dans le cas d'une variable continue, l'histogramme des effectifs (resp. des fréquences) est formé par les rectangles dont les bases inférieures sont les intervalles des classes et dont les aires sont proportionnelles aux effectifs (resp. les fréquences correspondant(e)s).*
2. *La courbe polygonale (ou polygone) des effectifs est obtenue en joignant par des segments les centres des bases supérieures des rectangles de l'histogramme des effectifs. Cette courbe est complétée par deux segments afin que l'aire du polygone soit égale à l'aire de l'histogramme (voir FIG.5.3).*

Construisons sur la même figure l'histogramme des effectifs et le polygone des effectifs associés au tableau statistique N° 2.(voir FIG.5.3)

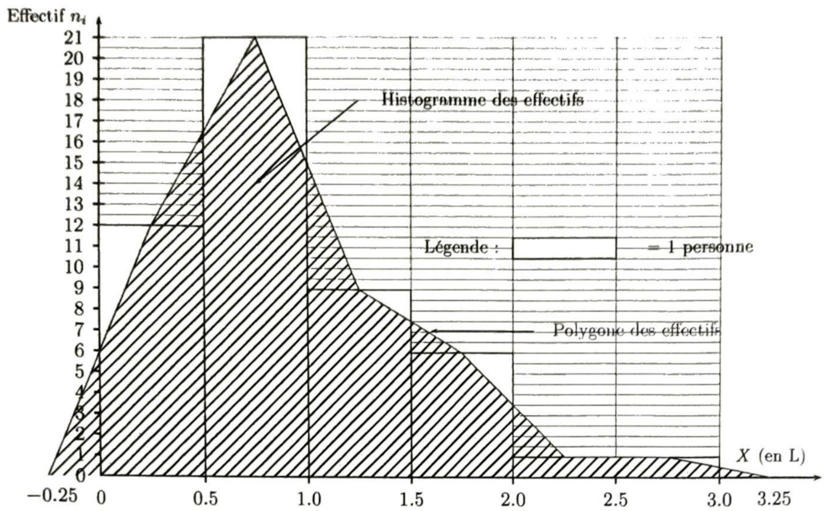

FIGURE 1.3 – Histogramme et polygone des effectifs

1.2.2.3 Histogramme et polygone des effectifs cumulés ou fréquences cumulées

Définition 1.5 **1.** *Dans le cas d'une variable continue, l'histogramme des effectifs cumulés est formé par les rectangles dont les bases inférieures sont les intervalles des classes et dont les hauteurs sont égales aux effectifs cumulés correspondants.*

2. *La courbe polygonale(ou polygone) des effectifs cumulés est obtenue en joignant par des segments les points dont l'abscisse est une borne supérieure d'une classe de la série et dont l'ordonnée est l'effectif cumulé correspondant à cette classe.*

Construisons sur la même figure l'histogramme des effectifs cumulés et le polygone des effectifs cumulés associés au tableau statistique N° 2.(voir FIG.5.4)

Remarque 1.1 *Pour le polygone des effectifs cumulés on joint les extrémités de droite (et surtout pas les centres) des bases supérieures, car par exemple* $N_3 = 42$ *signifie qu'il y a 42 personnes parmi celles interrogees buvant moins d'un litre et demi et* $N_4 = 48$ *signifie qu'il y a 48 personnes parmi celles interrogees buvant moins de deux litres.*

1.2.3 Paramètres ou caractéristiques des séries

Lorsqu'une série comporte un grand nombre de valeurs, on cherche à la résumer, si possible, à l'aide de quelques nombres significatifs appelés **paramètres** ou **caractéristiques**.
On va commencer par définir les notions de moyenne arithmétique (à rapprocher de la notion de barycentre), médiane, mode et les quartiles qui sont des paramètres de **position** et puis définir d'autres paramètres dits de **dispersion** : l'étendue, l'écart-interquartile, la variance et l'écart-type.

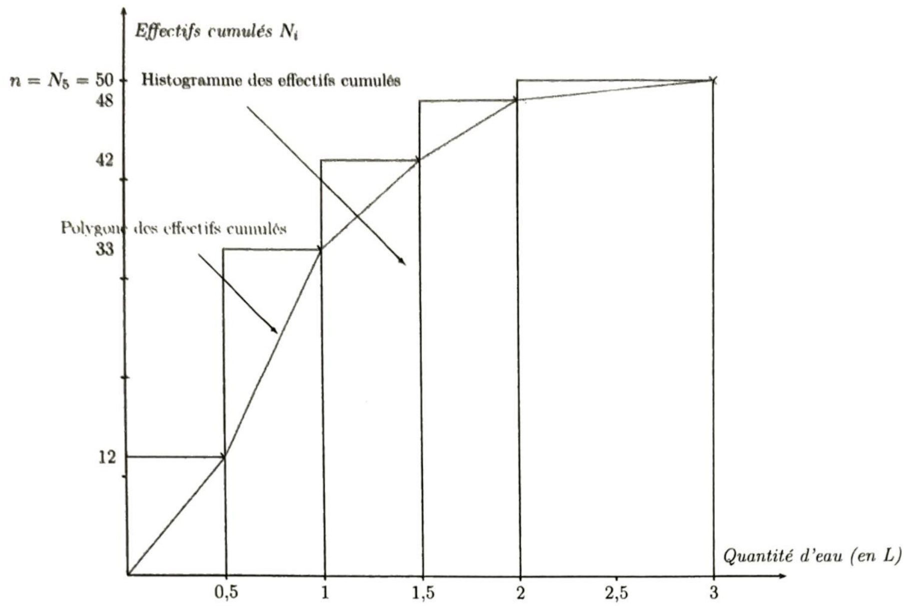

FIGURE 1.4 – Histogramme et polygone des effectifs cumulés

1.3 Paramètres de position d'une série statistique

1.3.1 Moyenne arithmétique d'une série statistique

Définition 1.6 1. **cas discret :** *Soit $(x_i)_{1 \leqslant i \leqslant n}$ une série discrète listée, On appelle* ***moyenne arithmétique*** *de cette série le nombre noté $\overline{x}$ ou $\overline{X}$ défini par :*

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

Soit une série discrète $(x_i, n_i)_{1 \leqslant i \leqslant p}$. La moyenne arithmétique de cette série s'écrit alors :

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{p} n_i x_i = \sum_{i=1}^{p} f_i x_i$$

2. **cas continu :** *Soit une série groupée en classes, on considère alors la série associée $(x_i, n_i)_{1 \leqslant i \leqslant p}$ où x_i représente le centre de la i^{eme} classe pour tout $1 \leqslant i \leqslant p$. On appelle* ***moyenne arithmétique*** *de la série de départ le nombre noté $\overline{x}$ ou $\overline{X}$ défini par :*

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{p} n_i x_i = \sum_{i=1}^{p} f_i x_i$$

Remarque 1.2 — *Ne jamais confondre dans la définition de la moyenne les paramètres n désignant l'effectif total (par lequel on divise) et le paramètre p (qui représente la borne supérieure pour l'indice de sommation) désignant le nombre de modalités "valeurs distinctes" ou nombre de classes.*

— *Par commodité la moyenne arithmétique sera appelée, s'il n'y a pas d'ambiguïté, la moyenne.*

•Exemples de calcul de la moyenne d'une série statistique

On reprend l'exemple discret n° 1 cité ci-dessus.

- Si on considère la série listée,

on calcule la moyenne par :

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{30}\sum_{i=1}^{30} x_i = \frac{8+11+5+\ldots+10+14+12}{30} = \frac{276}{30} = 9,2$$

- Si on considère la série regroupée par valeurs,
  la moyenne est alors **une moyenne pondérée** par les effectifs ; on la calcule en multipliant chaque valeur par l'effectif correspondant, en faisant la somme de ces produits et en divisant par l'effectif **total**, soit :

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{p} n_i x_i = \frac{1}{30}\sum_{i=1}^{11} n_i x_i = \frac{276}{30} = 9,2$$

  La dernière somme peut être lue à partir du tableau statistique N° 1.
- Si on considère la série regroupée par classes,
  en faisant un groupement en quatre classes d'amplitude cinq chacune, on considère alors que le centre de chaque classe représente la classe, on obtient le tableau statistique suivant :

Classes de X	$[0,5[$	$[5,10[$	$[10,15[$	$[15,20]$	TOTAL
Centres de classes x_i	2,5	7,5	12,5	17,5	
Effectifs n_i	3	12	13	2	30

  la moyenne de la série ainsi regroupée en classes est égale à :

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{p} n_i x_i = \frac{1}{30}\sum_{i=1}^{4} n_i x_i = \frac{3\times 2,5 + 12\times 7,5 + 13\times 12,5 + 2\times 17,5}{30} = \frac{295}{30} \approx 9,83.$$

Remarque 1.3 — *La moyenne et la moyenne pondérée par les effectifs sont égales : on ne modifie en rien la série sauf qu'on la représente différemment .*

— *Le regroupement en classes permet des calculs plus rapides mais ne permet pas d'obtenir la valeur exacte de la moyenne : en effet dans ce cas on suppose que le centre de la classe représente la classe mais ceci n'est qu'une approximation (entachée d'erreur) mais très utile.*

Exercice n° 1 : On reprend l'exemple continu n° 2 cité ci-dessus. Vérifier que $\overline{x} = 0,91$ en L.
écrire le résultat sous forme de phrase.

1.3.2 Médiane d'une série statistique

Définition 1.7 *La **médiane** $\mathcal{M}$ d'une série statistique est la valeur qui partage le groupe étudié en deux sous-groupes de même effectif chacun tels que :*

— *tous les éléments du premier groupe ont des valeurs inférieures ou égales à $\mathcal{M}$;*

— *tous les éléments du deuxième groupe ont des valeurs supérieures ou égales à $\mathcal{M}$.*

Détermination pratique de la médiane

1. **Cas discret** :
   Soit une série listée $(x_i)_{1\leqslant i\leqslant n}$, on commence d'abord par **ordonner** les valeurs de cette série dans l'ordre croissant (ou décroissant), on obtient alors une nouvelle série listée et dont les éléments sont ordonnés notée $(x'_i)_{1\leqslant i\leqslant n}$, la médiane se calcule selon la parité de n par les formules suivantes :
   — Cas où n est impair : $\mathcal{M} = x'_{\frac{n+1}{2}}$ = la valeur centrale de la série $(x'_i)_{1\leqslant i\leqslant n}$,
   — Cas où n est pair : $\mathcal{M} = \frac{1}{2}(x'_{\frac{n}{2}} + x'_{(\frac{n}{2}+1)})$ = la demi-somme des deux valeurs centrales de la série $(x'_i)_{1\leqslant i\leqslant n}$.
2. **Cas continu** :
   La médiane est l'abscise du point de rencontre de la courbe polygonale des effectifs (resp. fré-

quences) cumulé(e)s et la droite d'équation $y = \frac{n}{2}$ (resp. $y = \frac{1}{2} = 0,5$), on peut la déterminer graphiquement ou analytiquement par la méthode d'interpolation linéaire.

•Exemples de détermination de la médiane d'une série statistique

— La série d'entiers $(\overbrace{7,9,10,10},10,\overbrace{11,12,13,14})$ a pour médiane 10 $(n = 9, \mathcal{M} = x_5 = 10)$.

— La série $(\overbrace{7,9,10,10},10,11,\overbrace{11,12,13,14})$ a pour médiane $\frac{10+11}{2} = 10.5$ $\left(n = 10, \mathcal{M} = \frac{x_5 + x_6}{2} = 10.\right.$

— **Cas discret avec groupement par valeurs**
On reprend l'exemple discret n° 1 lorsque la série est groupée par valeurs. On réduit le tableau statistique N° 1 associé au tableau suivant :

	i=1	i=2	i=3	i=4	i=5	i=6	i=7	i=8	i=9	i=10	i=11
x_i	2	4	5	7	8	10	11	12	14	17	19
n_i	2	1	3	4	5	4	3	4	2	1	1
N_i	2	3	6	10	15	19	22	26	28	29	30

Les notes étant rangées par notes croissantes, $n = 30$ donc n est pair, or $30 = 14 + 1 + 1 + 14$ donc la médiane est la demi-somme des deux notes centrales : la valeur de la 15$^{\text{ème}}$ note et la valeur de la 16$^{\text{ème}}$ note c'est-à-dire $\frac{8+10}{2} = 9$.

$$\begin{array}{lcccccccccccc} \text{Rang :} & 1^{re} & 2^e & \ldots & 13^e & 14^e & 15^e & 16^e & 17^e & 18^e & \ldots & 29^e & 30^e \\ \text{Note :} & 2 & 2 & \ldots & 8 & 8 & 8 & 10 & 10 & 10 & \ldots & 17 & 19 \end{array}$$

$\underbrace{\quad 2 \; 2 \ldots 8 \; 8 \quad}_{\text{14 notes}}$ $\qquad$ $\underbrace{\quad 10 \; 10 \ldots 17 \; 19 \quad}_{\text{14 notes}}$

D'autre part, la note moyenne de ces 30 notes est $\overline{x} = 9,2$, donc la médiane et la moyenne sont (en général) **différentes**.

— **Cas continu**
On reprend l'exemple continu n° 1 lorsque la série est groupée par classes. On réduit le tableau statistique N° 2 associé au tableau suivant :

	i=1	i=2	i=3	i=4	i=5
Classes de X	[0 ; 0.5[	[0.5 ; 1[	[1 ; 1.5[	[1.5 ; 2[	[2 ; 3[
n_i	12	21	9	6	2
N_i	12	33	42	48	50

Ce dernier tableau nous permet de tracer le polygone des effectifs cumulés suivant :

Graphiquement, on peut estimer que $\mathcal{M}$ est environ égale à 0,8 L : ce qui veut dire que la moitié des personnes interrogées consomme moins de 0,8 L par jour (ou la moitié des personnes interrogées consomme plus de 0,8 L par jour).
Analytiquement, en utilisant la méthode d'interpolation linéaire on a :
$\mathcal{M} = 0,5 + (\frac{25-12}{33-12}) \times (1 - 0,5) \approx 0,8095$ c-à-d $\mathcal{M} \approx 0,8095$ Litres par jour

$\left[x = x_A + \frac{y - y_A}{y_B - y_A}(x_B - x_A) \right)$

1.3.3 Les quartiles

Définition 1.8 *Comme pour la médiane qui permet de partager le groupe étudié en deux sous-groupes d'effectifs égaux, intuitivement, les quartiles sont des nombres qui partagent le groupe étudié en quatre sous-groupes qui ont tous "sensiblement" le même nombre de termes, c'est-à-dire 25% de l'effectif total.*

— *Le premier quartile Q_1 est la plus petite valeur du caractère telle qu'au moins 25% des termes de la série aient une valeur du caractère qui lui soit inférieure ou égale.*

— *Le troisième quartile Q_3 est la plus petite valeur du caractère telle qu'au moins 75% des termes de la série aient une valeur du caractère qui lui soit inférieure ou égale.*

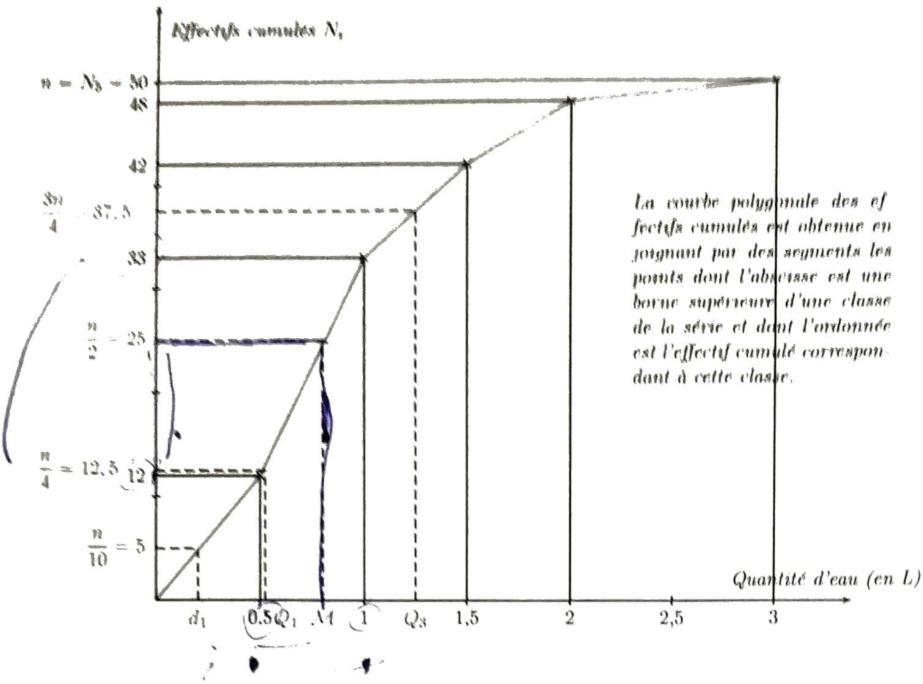

FIGURE 1.5 – Polygone des effectifs cumulés et quantiles

Détermination pratique des quartiles

1. **Cas discret** :
   Soit une série listée $(x_i)_{1\leqslant i\leqslant n}$, on commence d'abord par **ordonner** les valeurs de cette série dans l'ordre croissant (ou décroissant), on obtient alors une nouvelle série listée et dont les éléments sont ordonnés notée $(x'_i)_{1\leqslant i\leqslant n}$, les deux quartiles Q_1 (le premier) et Q_3 (le troisième) se calculent selon la convention suivante :

   Pour le premier quartile

   — Cas où n n'est pas divisible par 4 : $Q_1 = x'_{[\frac{n}{4}]+1}$,

   — Cas où n est divisible par 4 : $Q_1 = \dfrac{1}{2}(x'_{\frac{n}{4}} + x'_{\frac{n}{4}+1})$,

   Pour le troisième quartile

   — Cas où n n'est pas divisible par 4 : $Q_3 = x'_{3[\frac{n}{4}]+1}$,

   — Cas où n est divisible par 4 : $Q_3 = \dfrac{1}{2}(x'_{3\frac{n}{4}} + x'_{3\frac{n}{4}+1})$,

   où $[x]$ désigne la partie entière de x.

2. **Cas continu** :
   La procédure est la suivante :
   Comme pour la médiane, on commence par tracer le polygone des effectifs cumulés et on "adopte" les valeurs suivantes :
   Q_1 (resp. Q_3) est la valeur correspondant à l'effectif cumulé égal $\dfrac{n}{4}$ (resp. $\dfrac{3n}{4}$).
   La lecture des quartiles Q_1 et Q_3 peut se faire graphiquement sur le polygone des effectifs cumulés en considérant les abscisses des points d'ordonnées respectives $\dfrac{n}{4}$ et $\dfrac{3n}{4}$.
   le calcul des quartiles Q_1 et Q_3 se fait analytiquement par la méthode d'interpolation linéaire comme pour la médiane.

•Exemples de détermination des quartiles d'une série statistique
Exercice n° 2 : On reprend l'exemple continu n° 2 cité ci-dessus. Calculer analytiquement les deux quartiles Q_1 et Q_3.

1.3.4 Mode ou dominante d'une série statistique

Définition 1.9 *On appelle* **mode ou dominante** *(resp.* **classe modale ou classe dominante***) d'une série l'élément (resp. la classe) de la population correspondant au plus grand effectif. Dans le cas des série groupées en classes, le mode ou la dominante est le centre de la classe modale.*

Remarque 1.4 *En ce qui concerne le mode d'une série groupée en classes, il y a une définition plus intéressante et précise que le milieu de la classe modale. Cette remarque sera détaillée en classe.*

•Exemples de détermination des modes d'une série statistique
Exercice n° 3 : On reprend l' exemple discret n° 1 et l'exemple continu n° 2 cités ci-dessus. Calculer les deux modes associés .

1.4 Paramètres de dispersion

1.4.1 l'étendue

Définition 1.10 ***L'étendue*** *d'une série statistique $(x_i)_{1\leqslant i\leqslant n}$, notée e, est égale à la différence entre la plus grande et la plus petite valeur prises par cette série :*

$$e = x_{max} - x_{min}$$

où $x_{max} = max\{x_i,\ 1 \leqslant i \leqslant n\}$, $x_{min} = min\{x_i,\ 1 \leqslant i \leqslant n\}$.
Elle mesure une "sorte de dispersion" de cette série.

Les critères de dispersion indiquent de quelle façon les valeurs du caractère sont groupées (plus ou moins resserrées) autour des valeurs "centrales" : l'étendue est le plus simple de ces paramètres. En général, lorsque l'étendue est élevée, la dispersion est grande.
Exemple : pour la série avec les 30 notes (exemple discret n° 1), l'étendue est de $19 - 2 = 17$.

1.4.2 écart interquartile

Définition 1.11 *L'* **intervalle interquartile** *est l'intervalle* $[Q_1; Q_3]$.
L' **écart interquartile** *est le nombre $Q_3 - Q_1$. C'est la longueur de l'intervalle interquartile.*

Remarque 1.5 *Contrairement à l'étendue, l'écart interquartile élimine les valeurs extrêmes : ce peut être un avantage. En revanche il ne prend en compte que 50% de l'effectif : ce peut être un inconvénient.*

1.4.3 Variance et écart-type

1.4.3.1 Introduction

Exercice n° 4 : On considère les deux séries suivantes :

x_i	2	3	4	5	8	8	10	10	10	12	12	15	16	17	18
y_i	5	6	7	8	9	9	10	10	10	11	11	12	13	14	15

1. Calculez les moyennes, les médianes et les modes de ces deux séries.
2. Complétez le tableau suivant :

$(x_i - \bar{x})^2$															
$(y_i - \bar{y})^2$															

3. Calculez dans chacun des cas la moyenne des carrés des écarts pris par rapport à la moyenne. Conclusion ?

1.4.3.2 Définitions et théorème

Définition 1.12
On appelle **variance** *d'une série quelconque à caractère quantitatif discret, $(x_i, n_i)_{1\leqslant i\leqslant p}$, le nombre :*

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

L'écart-type *de cette série est $\sigma = \sqrt{V}$.*
Si la série est regroupée en classes *ou si la caractère est quantitatif continu, avec l'hypothèse d'une* **répartition uniforme** *à l'intérieur de chaque classe, on remplace chaque classe par son centre. On est ainsi ramené à un cas discret.*

Remarque 1.6 — *On est amené à considérer la racine carrée de la variance pour avoir un résultat exprimé dans la même unité que le caractère étudié.*
— *Il existe un autre moyen de calculer V qui évite le calcul de $(x_i - \bar{x})^2$, le théorème suivant précise cette possibilité :*

Théorème 1.13 (Théorème de Koenig)

$$V = \frac{1}{n}\sum_{i=1}^{p} n_i x_i^2 - \bar{x}^2 = \overline{x^2} - \bar{x}^2, \quad \text{où} \quad \overline{x^2} = \frac{1}{n}\sum_{i=1}^{p} n_i x_i^2$$

Preuve.

Exercice n° 5 : Calculer la moyenne, la variance et l'écart-type associés au cas continue n° 2.

1.5 Théorème de changement de variable

Il arrive souvent (surtout dans le cas continu) d'avoir des séries "compliquées" dans le sens suivant : si on note $(x_i, n_i)_{1\leqslant i\leqslant p}$ une suite quelconque, le calcul des $n_i x_i$ et surtout $n_i(x_i)^2$ est coûteux en temps et en difficulté. Dans ce cas on cherche à faire une (pas unique!) transformation affine sur tous les termes x_i du type $y_i = \dfrac{x_i - b}{a}$ afin de "faciliter" les calculs avec la nouvelle série auxiliaire $(y_i, n_i)_{1\leqslant i\leqslant p}$. On détermine facilement alors la moyenne, la variance et l'écart-type de la nouvelle série puis on obtient la moyenne, la variance et l'écart-type de la première série par les formules données dans le théorème suivant :

Théorème 1.14
Soit $(x_i, n_i)_{1\leqslant i\leqslant p}$ une série statistique d'effectif n de moyenne $\bar{x}$, de variance V_x et d'écart-type σ_x. La série de même effectif $(y_i, n_i)_{1\leqslant i\leqslant p}$, telle que pour tout $1 \leqslant i \leqslant p$, $y_i = \dfrac{x_i - b}{a}$ $(a \in \mathbb{R}^*, b \in \mathbb{R})$ a pour moyenne $\bar{y}$, variance V_y et pour écart-type σ_y tels que :

$$\bar{y} = \frac{\bar{x} - b}{a} \quad , \quad V_y = \frac{V_x}{a^2} \quad , \quad \sigma_y = \frac{\sigma_x}{|a|}$$

ce qui équivaut à :

$$\bar{x} = a\bar{y} + b \quad , \quad V_x = a^2\, V_y \quad , \quad \sigma_x = |a|\sigma_y$$

Exercice n° 6 : Faire un changement de variable du type $y_i = \dfrac{x_i - b}{a}$ où a et b sont bien choisis, puis recalculer la moyenne, la variance et l'écart-type associé au cas continue n° 2. Pour cela compléter le tableau statistique suivant :

	i=1	i=2	i=3	i=4	i=5	Sommes
Classes de X	[0 ; 0.5[	[0.5 ; 1[	[1 ; 1.5[	[1.5 ; 2[	[2 ; 3[	*
Centres x_i	0.25	0.75	1.25	1.75	2.5	*
$y_i = \dfrac{x_i - b}{a}$						*
n_i	12	21	9	6	2	50
$n_i y_i$						
$n_i y_i^2$						