# Chapitre V: STATISTIQUE DESCRIPTIVE

## Programme détaillé (EM)

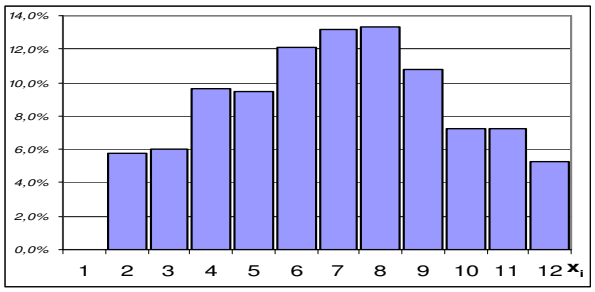
| | Contenu | Exclusions / inclusions |
|---|---|---|
| **6.1.** | Classification de données en données discrètes ou continues. | |
| **6.2.** | Données discrètes simples : tableaux statistiques ;<br>polygones statistiques. | |
| **6.3.** | Données groupées discrètes ou continues : tableaux statistiques ; valeurs centrales d'un intervalle ; limites inférieure et supérieure d'un intervalle.<br>Histogrammes des effectifs.<br>Diagrammes à tiges et feuilles. | *Un histogramme des effectifs utilise des intervalles de classes égales.* |
| **6.4.** | Tables de fréquences cumulées pour des données discrètes groupées et pour des données continues groupées ; courbes des fréquences cumulées.<br>Diagrammes à boîtes et moustache (tracés en boîte).<br>Centiles ; quartiles. | *Les élèves doivent utiliser les valeurs centrales d'un intervalle pour estimer la moyenne des données groupées. Il peuvent établir le lien entre la médiane, le 50e centile et la courbe des fréquences cumulées.*<br>*Les élèves doivent être familiers avec la notation sigma $\Sigma$.*<br>*Inclus : prise de conscience du concept de dispersion et compréhension de la signification de la valeur numérique de l'écart-type, $s_n$.. On s'attend à ce que les élèves utilisent une calculatrice à écran graphique pour calculer l'écart-type.*<br>*Les élèves doivent comprendre les concepts de population et d'échantillon. Ils doivent également être conscients qu'en général, la moyenne de la population , $\mu$, et l'écart-type de la population, $\sigma$, sont inconnus et que la moyenne de l'échantillon, $\bar{x}$ et l'écart-type de l'échantillon, $s_n$, ne sont que des estimateurs de ces quantités.* |
| **6.5.** | Mesures de tendance centrale.<br>Pour des données discrètes simples : moyenne ; médiane ; mode.<br>Pour des données continues et discrètes groupées : moyenne approchée ; classe modale ; 50e centile. | |
| **6.6.** | Mesures de dispersion : étendue ; intervalle interquartile ; écart-type. | |

## Formulaire

| | | |
|---|---|---|
| 6.4 | Intervalle interquartile | (delta) Q = $q_3 - q_1$<br>Les observations aberrantes sont les points qui ont une valeur inférieure à $q_1 - 1{,}5$ (delta) Q ou supérieure à $q_3 + 1{,}5$ (delta) Q |
| 6.5 | Moyenne | $\bar{x} = \frac{\sum_{i=1}^{k} f_i x_i}{n}$, avec $n = \sum_{i=1}^{k} f_i$ |
| 6.6 | Écart-type | $s_n = \sqrt{\frac{\sum_{i=1}^{k} f_i (x_i - \bar{x})^2}{n}}$, avec $n = \sum_{i=1}^{k} f_i$ |

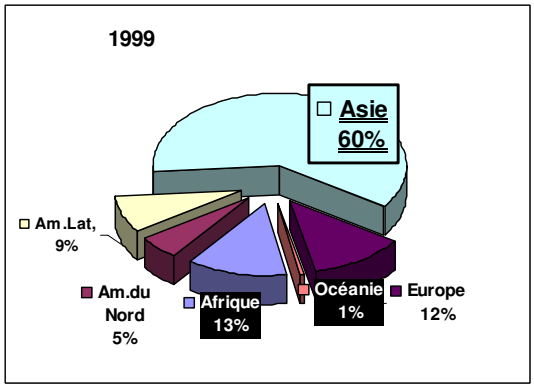
# V Statistiques

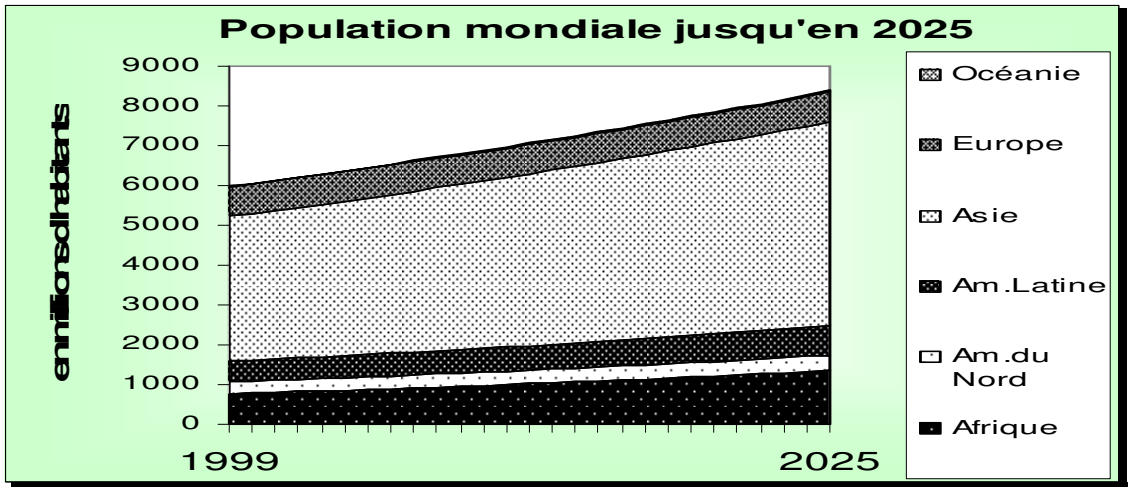
## *Introduction*





| | population en Mio | taux de croissance |
|---|---|---|
| Afrique | 771 | 2,20% |
| Amérique du Nord | 303 | 0,86% |
| Amérique Latine | 530 | 1,45% |
| Asie | 3637 | 1,33% |
| Europe | 728 | 0,04% |
| Océanie | 30 | 1,14% |




La statistique a pour but de recueillir et d'étudier des données sur des ensembles de même nature, par exemple des personnes, des animaux, des pays, des entreprises etc. Il ne faut pas confondre **la** statistique qui est la science et **une** statistique qui est un ensemble de données chiffrées sur un sujet précis. Les premières statistiques correctement élaborées ont été celles des recensements démographiques. Ainsi le vocabulaire statistique est essentiellement celui de la démographie.

A la base de toute étude statistique, il y a une ***population*** , formée d'***individus*** ou d'***unités statistiques*** . Une partie de la population est appelée un ***échantillon.*** Une étude portant sur un échantillon est appelée un ***sondage***.

Le problème qui se pose souvent dans la pratique est le suivant :
L'échantillon choisi est-il ***représentatif*** de la population ? Par exemple un sondage effectué auprès de 500 Luxembourgeois sur leur intention de vote concernant le référendum sur la constitution européenne permet-il de tirer des conclusions sur toute la population du pays ?

# V Statistiques

On distingue :

- **La statistique descriptive** qui comprend la collecte des données, leur regroupement, leur représentation sous forme de tableaux et de graphiques, le calcul de totaux, de pourcentages, de moyennes et d'autres grandeurs caractéristiques.
- **La statistique mathématique ou la statistique inférentielle** qui essaie, à partir d'observations faites sur un échantillon, de tirer des conclusions qui portent sur toute la population. La statistique inférentielle fait intervenir le calcul des probabilités.

L'étude statistique porte sur un ou plusieurs traits communs à toutes les unités statistiques : ces traits étudiés s'appellent ***caractères***.

| **Population** | **Caractère** |
|---|---|
| Éléments chimiques | Nombre d'isotopes |
| Galaxies | Nombre d'étoiles |
| Pays de l'UE. | Produit intérieur brut |
| Films …. | Recettes |
| Mois de l'année 2001 | Température moyenne |
| Élèves du BI 2005-2006 | Couleur des yeux |
| Habitants du Luxembourg au 31.10.2004 | État civil |

Un caractère est dit :

- ***qualitatif***, quand les valeurs ne peuvent être ni ordonnées ni ajoutées (groupe sanguin, couleur des yeux, état civil).
- ordinal, quand les valeurs peuvent être ordonnées mais pas ajoutées (p.ex. opinions exprimées sur une échelle de valeurs)
- ***quantitatif***, quand les valeurs sont numériques (mesures physiques, physiologiques, économiques).

Les valeurs que peut prendre un caractère s'appellent les ***modalités***. Par exemple les modalités du caractère « état civil » sont : célibataire, marié, divorcé, veuf. Pour des raisons de facilité de traitement informatique ou mathématique, on cherche à se ramener à des caractères quantitatifs par un codage. Si le caractère initial est qualitatif, le codage sera souvent binaire. Le cas le plus simple est celui d'un référendum, où il n'y a que deux modalités, codées 0 et 1. Il faut se souvenir que le codage est arbitraire, et que les résultats numériques que l'on obtient après codage peuvent dépendre de celui-ci. Des techniques spécifiques permettent de traiter plus particulièrement les caractères qualitatifs et ordinaux. Nous nous limiterons ici aux caractères quantitatifs.

Parmi les caractères quantitatifs, on distingue encore entre

- les ***caractères quantitatifs discrets*** qui ne prennent que peu de modalités isolées distinctes

Exemples : le nombre de petits par portée, le nombre d'élèves dans une classe.

- les ***caractères continus*** pour lesquels toutes les valeurs observées sont a priori différentes, toutes les valeurs réelles d'un certain intervalle sont à priori possibles.

Exemples : le poids, la taille, ….
La frontière entre continu et discret est beaucoup moins claire en pratique qu'en théorie.

Une fois recueillies, les données brutes se présentent souvent comme une liste de nombres peu lisible. Le ***traitement statistique*** consiste à compresser et à résumer les données par des quantités calculées et des représentations graphiques, afin d'***extraire l'information essentielle***. On ne traite pas un échantillon sans avoir une question précise à lui poser. Etant donné par exemple un échantillon de tailles de filles de 18 ans, le traitement ne sera pas le même selon que l'on sera un nutritionniste qui cherche à étudier l'influence du régime alimentaire sur la croissance, ou un fabriquant de vêtements qui cherche à dimensionner ses patrons.

# 1. Classification et présentation des données

## 1.1. Données discrètes

### 1.1.a. Premier exemple

Une enquête portant sur le nombre d'enfants par foyer, réalisée auprès des élèves de notre classe (BI 1 et BI 2) respectivement leurs familles. Nous relevons les résultats suivants :

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | | | | | |

Un premier traitement consiste à regrouper les modalités et à noter le nombre de fois que chaque modalité est apparue . Les modalités sont parfois notées $x_1$, $x_2$, ….., $x_i$, …,$x_k$ , leur ensemble est une ***variable statistique***, et le nombre de fois que les modalités apparaissent sont les ***effectifs*** (éventuellement fréquences ou ***fréquences absolues*** au BI) , et sont notées $f_1$, $f_2$, …, $f_i$, …., $f_k$ (éventuellement : $n_i$ ou $e_i$).

| Nombre d'enfants par famille ($x_i$) | | | | | | |
|---|---|---|---|---|---|---|
| dépouillement | | | | | | |
| Nombre de familles ($f_i$) | | | | | | |

***Remarque :*** ***L'échantillon choisi n'est certainement pas représentatif pour la population (p.ex. les familles au Luxembourg), et cela pour plusieurs raisons. Quelles sont les principales ?***

Nous venons de réaliser une ***série statistique***. En général, il s'agit de l'ensemble de tous les couples ($x_i$, $f_i$).

L'***effectif total*** ou ***la taille*** de l'échantillon est noté n (parfois N) Dans notre exemple : n =

En général, on a : $n = f_1 + f_2 + \ldots + f_k = \sum_{i+1}^{k} f_i$

Souvent on ajoute aux effectifs les ***fréquences (ou fréquences relatives)***. C'est le rapport entre l'effectif de la modalité et la taille de la population. Les fréquences sont surtout utiles pour comparer des séries statistiques.

| Nombre d'enfants par famille ($x_i$) | | | | | | | |
|---|---|---|---|---|---|---|---|
| Effectifs $f_i$ | | | | | | | |
| Fréquences relatives sous forme de fraction | | | | | | | |
| Fréquences relatives en écriture décimale | | | | | | | |
| Fréquences relatives en % | | | | | | | |

Remarque : Une fréquence est un nombre entre 0 et 1 (entre 0 % et 100 %). La somme des fréquences de toutes les modalités vaut 1 respectivement 100 %.

Pour faire parler les chiffres, on a très souvent recours à des diagrammes et des graphiques.

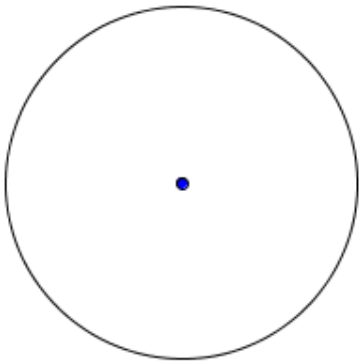
**Diagramme en bâtons (en barres)**

On peut représenter une variable statistique discrète par un diagramme en bâtons dont les bâtons (ou les bandes verticales) ont des hauteurs proportionnelles aux effectifs ou aux fréquences relatives.


**Diagramme circulaire (d. en secteurs, d. en quartiers de tarte, camembert)**

On partage le disque en secteurs angulaires dont les aires resp. les angles au centre sont proportionnels aux fréquences des valeurs.

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | |



### 1.1.b. Deuxième exemple

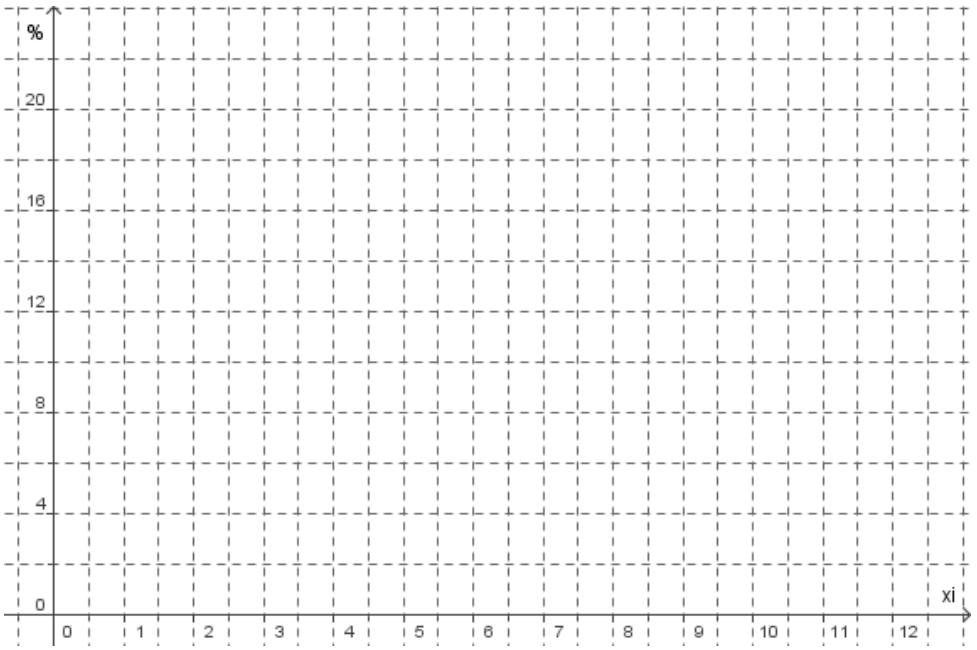
Chaque élève lance deux dés simultanément 100 fois de suite et note chaque fois la somme des points obtenus. Par exemple pour l'événement (3 ;5) il notera le résultat 8. Ensuite il notera le nombre de fois que chaque résultat est apparu (= effectif du résultat) et donne un tableau qui résume la situation.

| Résultat $x_i$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| dépouillement | | | | | | | | | | | |
| Effectif $f_i$ | | | | | | | | | | | |

En rassemblant ensuite les données de tous les élèves (ou plusieurs), nous obtenons le tableau suivant :

| **Résultat $x_i$** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| **Effectif $f_i$** | | | | | | | | | | | |
| **Fréquences en %** | | | | | | | | | | | |

# V Statistiques




***Exercice 711*** 4 candidats se présentent à une élection. Les résultats sont donnés dans le tableau :

| Candidat | A | B | C | D |
|---|---|---|---|---|
| Nombre de voix obtenues | 51 210 | 43 821 | 23 212 | 8 597 |
| Fréquence (en %) | | | | |

Compléter le tableau en donnant la fréquence en pourcentage pour chacun des candidats.
Représenter ces données par un diagramme ciculaire.

***Exercice 712***

Le tableau ci-dessous donne le nombre moyen de longs métrages réalisés par les dix plus gros producteurs mondiaux, entre 1990 et 1995 (source Unesco).

| Brésil | Chine | Etats-Unis | France | Inde | Italie | Japon | Philippines | Royaume-Uni | Thaïlande |
|---|---|---|---|---|---|---|---|---|---|
| 86 | 154 | 420 | 141 | 838 | 96 | 251 | 456 | 78 | 194 |

Représenter ces données par un diagramme à barres (ou diagramme en bâtons)

***Exercice 713***

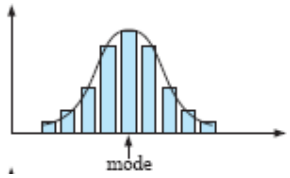
On a lancé 150 fois 3 dés et on a compté la somme des numéros obtenus à chaque lancer. Les résultats obtenus sont les suivants :

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 13 | 6 | 13 | 8 | 10 | 7 | 11 | 12 | 13 | 9 | 15 | 11 | 12 | 14 | 9 |
| 11 | 13 | 12 | 7 | 15 | 10 | 5 | 9 | 16 | 10 | 9 | 9 | 18 | 12 | 9 |
| 12 | 8 | 10 | 12 | 8 | 15 | 18 | 12 | 12 | 9 | 10 | 6 | 15 | 8 | 11 |
| 15 | 13 | 14 | 10 | 8 | 6 | 7 | 12 | 10 | 17 | 13 | 13 | 9 | 11 | 6 |
| 4 | 16 | 16 | 8 | 12 | 8 | 12 | 8 | 9 | 12 | 16 | 12 | 6 | 7 | 10 |
| 12 | 9 | 14 | 10 | 12 | 7 | 8 | 14 | 10 | 11 | 9 | 14 | 15 | 10 | 6 |
| 4 | 13 | 17 | 13 | 8 | 15 | 14 | 15 | 8 | 13 | 12 | 5 | 9 | 12 | 18 |
| 9 | 14 | 8 | 14 | 3 | 11 | 10 | 9 | 12 | 7 | 9 | 9 | 10 | 8 | 15 |
| 16 | 13 | 11 | 13 | 11 | 6 | 12 | 5 | 9 | 15 | 16 | 11 | 12 | 17 | 12 |
| 10 | 13 | 5 | 12 | 7 | 14 | 5 | 12 | 8 | 4 | 10 | 11 | 8 | 7 | 13 |

a. Classer les valeurs obtenues par ordre de grandeur croissante et calculer les effectifs et les fréquences correspondant à chaque valeur.
b. Donner un diagramme qui représente la série statistique.
c. Calculer la moyenne.
d. Quelle est la modalité qui a la plus grande fréquence ?

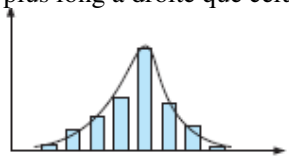
### 1.1.c. Distributions des données

Les histogrammes peuvent être **symétriques** ou **asymétriques.**




Un **histogramme** ( ou une distribution ) **symétrique** présente une longueur d'étalement à droite égale à celle de gauche.

Un **histogramme asymétrique** se présente sous deux formes
**Asymétrie positive** : étalement de droite est plus long à droite que celui de gauche.



**Asymétrie négative** :étalement de gauche est plus long à gauche que celui de droite.

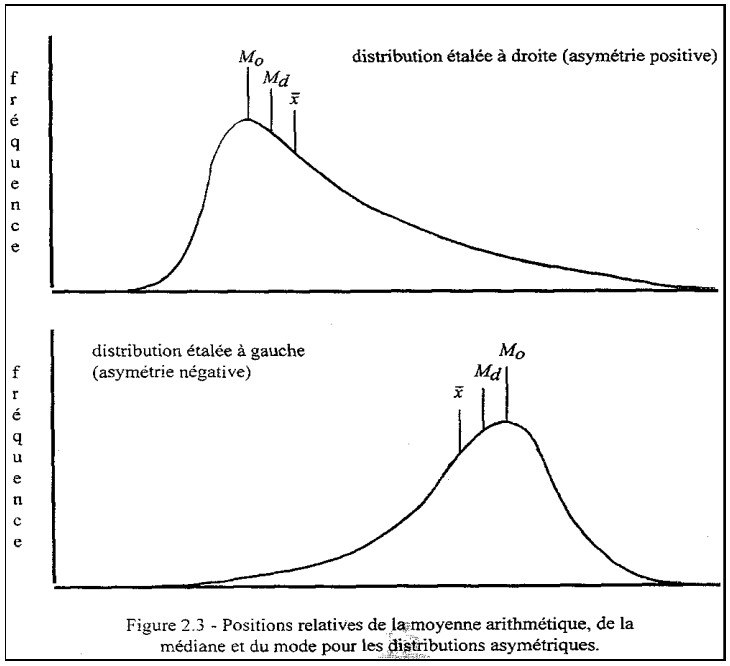
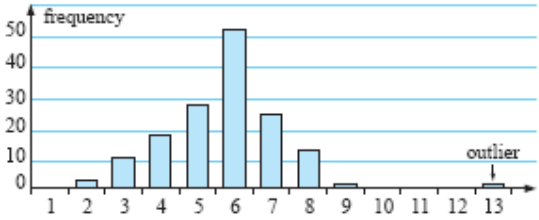



Figure 2.3 - Positions relatives de la moyenne arithmétique, de la médiane et du mode pour les distributions asymétriques.

**Valeurs aberrantes (outliers)**

Une valeur aberrante est une observation qui semble dévier de façon marquée par rapport à l'ensemble des autres membres de l'échantillon dans lequel il apparaît (voir aussi formulaire 6.4.). Les observations *non représentatives* ou *aberrantes* ont toujours été considérées comme une source de contamination, déformant l'information obtenue à partir des données brutes.




***Exercice 714***

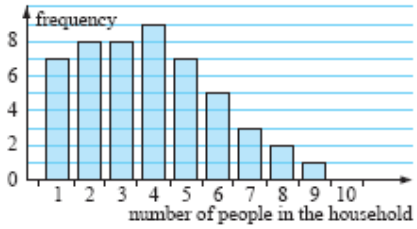
Le nombre d'allumettes par boîte est indiqué égal à 50. On a compté le nombre effectif d'allumettes dans un échantillon de 60 boîtes.

51 50 50 51 52 49 50 48 51 50 47 50 52 48 50 49 51 50 50 52
52 51 50 50 52 50 53 48 50 51 50 50 49 48 51 49 52 50 49 50
50 52 50 51 49 52 52 50 49 50 49 51 50 50 51 50 53 48 49 49

a. Quelle est la variable dans cette enquête ?
b. Les données sont-elles discrètes ou continues ?
c. Donnez un tableau des effectifs.
d. Construisez un histogramme.
e. Quel est le pourcentage des boîtes qui contiennent

***Exercice 715***

Pour un échantillon de foyers choisis au hasard, on a relevé le nombre de personnes qui vivent dans chaque foyer.
Les résultats trouvés dans cette enquête sont représentés par le diagramme ci-contre.




a. Combien de foyers figurent dans cette enquête ?
b. Combien d'entre eux ont seulement une ou deux personnes ?
c. Quel est le pourcentage des foyers à 5 personnes ou plus ?
d. Décrivez la distribution des données.

### 1.1.d. Regroupement des données

Nous venons déjà de voir que le traitement des données commence souvent par la transformation des données brutes en données groupées. Il existe une méthode pour présenter toutes les données sous une forme visuellement utile qui est appelée la ***représentation en tiges et feuilles***.

**Exemple :** Voici les résultats du dernier examen d'algèbre:
{13, 24, 46, 36, 14, 35, 24, 26, 36, 25, 24, 53, 42, 44, 33, 13, 11, 44, 23, 23, 24, 33, 48}
En ordre, nous avons : {11, 13, 13, 14, 23, 23, 24, 24, 24, 24, 25, 26, 33, 33, 33, 35, 36, 36, 42, 44, 44, 46, 48, 53}

**Diagramme à tige et à feuilles**

| | |
|---|---|
| 1 | 1,3,3,4 |
| 2 | 3,3,4,4,4,4,5,6 |
| 3 | 3,3,5,6,6 |
| 4 | 2,4,4,6,8 |
| 5 | 3 |

Le diagramme de tige et à feuilles est composé du couloir central (la tige) qui représente les dizaines. Chaque donnée est indiquée à droite de la tige (les feuilles) sur une des branches.

Pour pouvoir interpréter correctement un tel diagramme, il faut lui ajouter une légende.
P.ex. 2 | 4 signifie : 24

On peut aussi réaliser des diagrammes doubles, en particulier lorsqu'on veut comparer deux groupes :
Exemple : taille en cm des filles et des garçons de notre classe.

| **filles** | | **garçons** |
|---|---|---|
| | 14 | |
| | 15 | |
| | 16 | |
| | 17 | |
| | 18 | |
| | 19 | |

***Exercice 716***

Ci-contre on a représenté les données d'une enquête par un diagramme à tiges et feuilles. Donnez :

| | |
|---|---|
| 0 | 2 3 7 |
| 1 | 0 4 4 7 8 9 9 |
| 2 | 0 0 1 1 2 2 3 5 5 6 8 8 |
| 3 | 0 1 2 4 4 5 8 9 |
| 4 | 0 3 7 |
| 5 | 5 |
| 6 | 2 |

a. La valeur minimale.
b. La valeur maximale.
c. Le nombre de données supérieures à 25.
d. Le nombre de données qui ne dépassent pas 40.
e. Le pourcentage des données inférieures à 15.
f. Quelle est la description que vous donneriez de cette distribution ?

***Exercice 717*** 45 élèves ont passé un test sur 50 points dont voici les résultats :

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 25 | 28 | 35 | 42 | 44 | 28 | 24 | 49 | 29 | 33 | 33 | 34 | 38 | 28 | 26 |
| 32 | 34 | 39 | 41 | 46 | 35 | 35 | 43 | 45 | 50 | 30 | 22 | 20 | 35 | 48 |
| 36 | 25 | 20 | 18 | 9 | 40 | 32 | 33 | 28 | 33 | 34 | 34 | 36 | 25 | 42 |

a. Donnez un diagramme à tiges et feuilles en prenant 0,1,2,3,4,5 comme tige.
b. Refaites le diagramme en ordonnant maintenant les valeurs.
c. Quel est l'avantage de ce diagramme par rapport à un tableau des effectifs ?
d. Quelle est la note i) la plus élevée ii) la plus basse ?
e. On a attribué la note « A » aux élèves qui ont réussi au moins 40 points. Quel est le pourcentage des élèves qui ont obtenu « A » ?
f. Quel est la pourcentage de ceux qui ont obtenu moins que la moitié des points ?

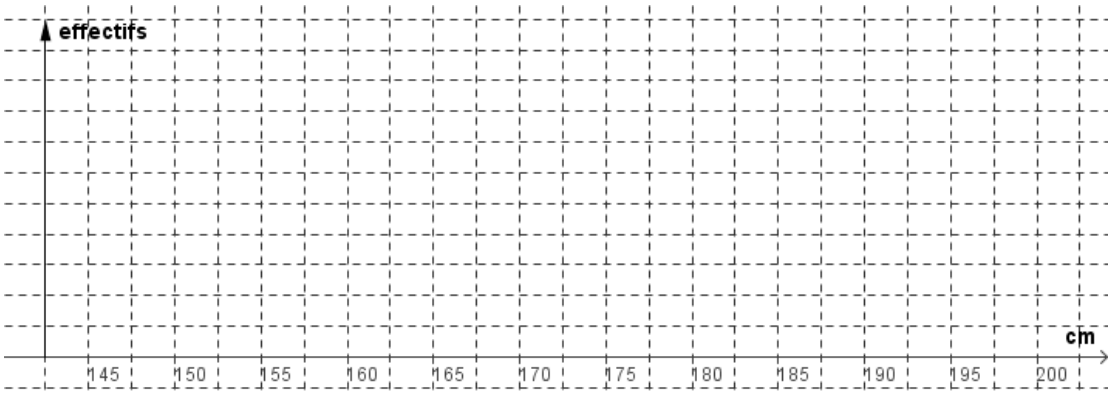
## 1.2. Données continues

### *1.2.a.* 1er exemple : *Taille en cm des élèves de notre classe*

Lorsque le caractère étudié prend trop de valeurs différentes (ce qui n'est pas encore tellement le cas ici en raison de la taille réduite de la population …), il devient fastidieux de traiter et de représenter les données comme on l'a fait dans les exemples précédents. On risquerait d'avoir des effectifs très petits et de se perdre dans les calculs trop détaillés. Au lieu de travailler avec toutes les modalités du caractère étudié, on les regroupe en un nombre raisonnable de **classes** ou d'intervalles. On ne compte plus le nombre de fois qu'on rencontre une valeur déterminée mais le nombre d'observations qui appartiennent à une classe donnée. Cette méthode permet de masquer le « désordre » aléatoire tout en préservant les tendances importantes qui caractérisent les données.

Le plus souvent on choisit des intervalles semi-ouverts du type [a,b[ définis par $a \leq x < b$.
Nous allons nous limiter au cas où les intervalles ont tous la même amplitude (= b – a).
Dans notre exemple, on peut choisir p.ex. des intervalles d'amplitude 5 cm.

| **Taille en cm** | **Effectifs** | **fréq. en %** | **Centre de classe (valeur centrale)** |
|---|---|---|---|
| [150,155[ | | | 152,5 |
| [155,160[ | | | 157,5 |
| [160,165[ | | | 162,5 |
| [165,170[ | | | 167,5 |
| [170,175[ | | | 172,5 |
| [175,180[ | | | 177,5 |
| [180,185[ | | | 182,5 |
| [185,190[ | | | 187,5 |
| [190,195[ | | | 192,5 |



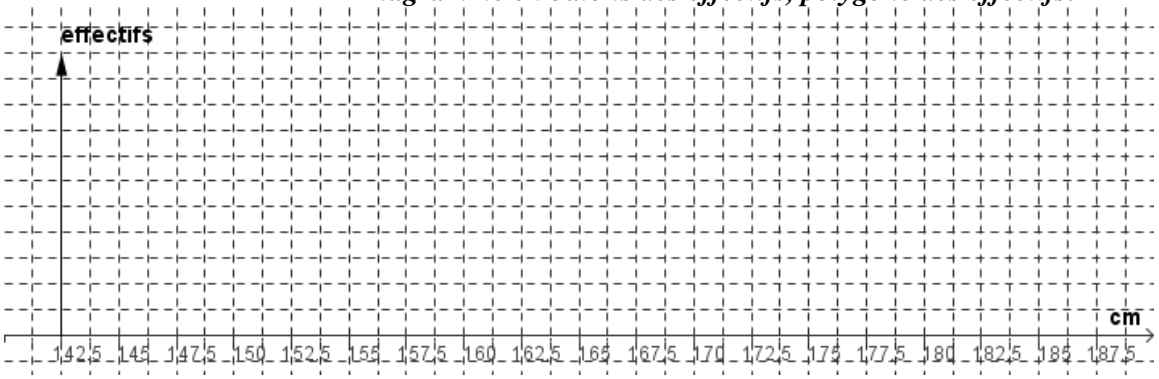
Pour les caractères quantitatifs continus, la représentation graphique la plus utilisée est **l'histogramme .** Sur les classes portés en abscisse on dessine des rectangles dont l'aire est proportionnelle à l'effectif (resp. la fréquence) de la classe. Dans le cas des intervalles de même amplitude, cela signifie aussi que les hauteurs des rectangles sont proportionnelles aux effectifs.

Dans le cas où toutes les classes ont la même amplitude, on peut également désigner chacune d'elles par son centre (valeur centrale, point milieu) qui est la moyenne arithmétique des limites supérieure et inférieure de la classe: $c = \frac{a+b}{2}$ pour l'intervalle [a,b[.

Les centres $c_i$ interviennent dans certains calculs (voir plus tard) ; on traite alors le caractère continu comme un caractère discret. On fait comme si l'effectif d'une classe [a,b[ était concentré sur le centre c de [a,b[. L'histogramme par contre suggère que cet effectif est réparti uniformément entre a et b. Dans les deux cas, on a perdu une partie de l'information initiale au bénéfice d'une simplification des calculs et d'une représentation plus efficace.

On peut également représenter un caractère continu par un diagramme en bâtons ou un polygone. Il suffit pour cela de porter sur l'axe des abscisses les centres de classes et de construire les points de coordonnées $(c_i, f_i)$.

***Diagramme en bâtons des effectifs, polygône des effectifs:***




***Exercice 718***

On a réalisé une enquête sur le temps mis par les élèves de 1re pour se rendre à l'école. Cela donne les résultas suivants, en minutes :

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 12 | 15 | 23 | 25 | 30 | 17 | 19 | 15 | 13 | 25 | 40 | 12 | 9 | 5 | 11 |
| 14 | 42 | 30 | 25 | 17 | 24 | 22 | 35 | 41 | 12 | 9 | 41 | 5 | 12 | 15 |
| 20 | 13 | 36 | 17 | 22 | 18 | 41 | 30 | 12 | 15 | 20 | 25 | 36 | 12 | 17 |
| 14 | 7 | 18 | 37 | 42 | 15 | 17 | 18 | 11 | 13 | 14 | 20 | 24 | 31 | 17 |

1) Dressez un tableau reprenant
   a. Les temps par classes de 10 minutes : [0,10[, [10,20[, …
   b. Les centres des classes  c. Les effectifs  d. Les fréquences
2) Répondez aux questions suivantes :
   a. Quel est l'effectif de la population sur laquelle porte l'enquête ?
   b. Combien y a–t-il d'élèves qui mettent plus de 20 minutes pour atteindre l'école ?
   c. Quel est le pourcentage des élèves arrivant à l'école en moins d'une demi-heure ?
3) Réalisez un graphique pour présenter les résultats.
4) Calculez la moyenne des temps en vous servant
   a. Des temps individuels  b. Des temps groupés par classe.
   Comparez les résultats et expliquez la différence.

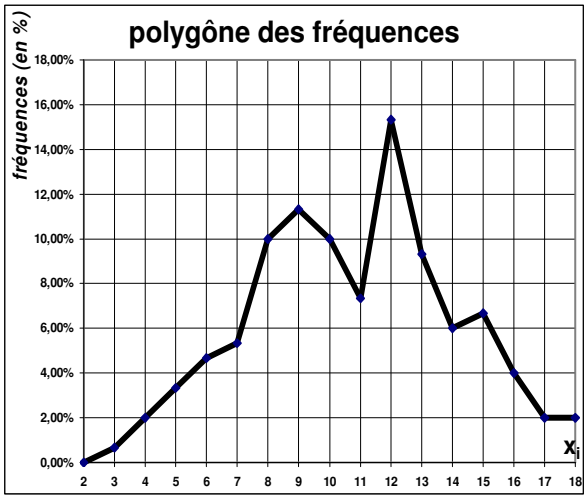
### *1.2.b.* 2e exemple

Les gens se demandent souvent quel est le nombre idéal d'intervalles à utiliser lorsqu'on groupe les données. On opte souvent pour une dizaine d'intervalles.

Reprenons l'exemple de l'exercice 3 qu'on a traité comme un caractère discret. On peut aussi regrouper les modalités en classes de 3 valeurs p.ex. et suivre le traitement d'un caractère continu.

| Xi | fi | % | $x_i f_i$ |
|---|---|---|---|
| 2 | 0 | 0,00% | 0 |
| 3 | 1 | 0,67% | 3 |
| 4 | 3 | 2,00% | 12 |
| 5 | 5 | 3,33% | 25 |
| 6 | 7 | 4,67% | 42 |
| 7 | 8 | 5,33% | 56 |
| 8 | 15 | 10,00% | 120 |
| 9 | 17 | 11,33% | 153 |
| 10 | 15 | 10,00% | 150 |
| 11 | 11 | 7,33% | 121 |
| 12 | 23 | 15,33% | 276 |
| 13 | 14 | 9,33% | 182 |
| 14 | 9 | 6,00% | 126 |
| 15 | 10 | 6,67% | 150 |
| 16 | 6 | 4,00% | 96 |
| 17 | 3 | 2,00% | 51 |
| 18 | 3 | 2,00% | 54 |
| ☞ | **150** | **100,00%** | **1617** |

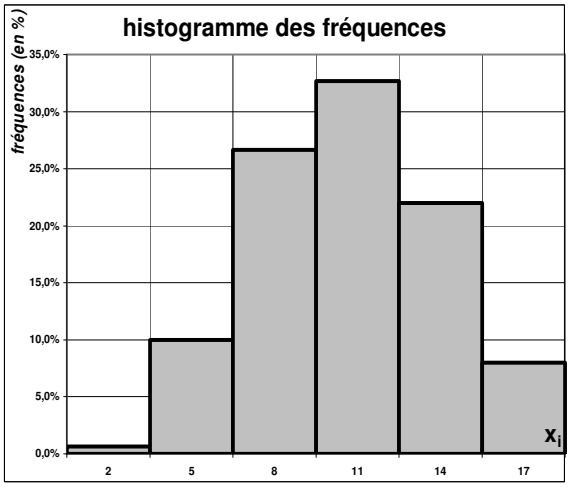
**c)** **moyenne= 10,78**




**d)** **12** **(le mode)** **e)** **47,33%**

| classes | $c_i$ | $f_i$ | % | $c_i f_i$ |
|---|---|---|---|---|
| [1,3] | 2 | 1 | 0,7% | 2 |
| [4,6] | 5 | 15 | 10,0% | 75 |
| [7,9] | 8 | 40 | 26,7% | 320 |
| [10,12] | 11 | 49 | 32,7% | 539 |
| [13,15] | 14 | 33 | 22,0% | 462 |
| [16,18] | 17 | 12 | 8,0% | 204 |
| | | 150 | 100% | 1602 |

**moyenne= 10,68**




***Exercice 719***

Des écoliers de 11 à 13 ans ayant été mesurés, leurs tailles ont été classées dans le tableau suivant :

| **Tailles en cm** | **Nombre d'élèves** |
|---|---|
| [130,134[ | 2 |
| [134,138[ | 6 |
| [138,142[ | 41 |
| [142,146[ | 37 |
| [146,150[ | 20 |
| [150,154[ | 4 |
| [154,158[ | 2 |
| [158,162[ | 1 |

1) Calculez les fréquences correspondant à chaque classe.
2) Déterminez les centres des classes.
3) Tracez l'histogramme et le polygône des effectifs.
4) On voudrait pouvoir indiquer pour chaque taille x le pourcentage des élèves qui ont une taille inférieure ou égale à x. Ce nombre est appelé la ***fréquence cumulée de x***.
   a. Ajoutez d'abord dans le tableau les fréquences cumulées des limites supérieures des intervalles.
   b. Construisez ensuite le polygone des fréquences cumulées qu'on obtient si on relie les points qui correspondent aux résultats trouvés en a.
   c. Quelle est la fréquence cumulée de 145 cm ?
   d. Quelle est la taille dont la fréquence cumulée est 0,5 cm ?

# 2. Effectifs et fréquences cumulés

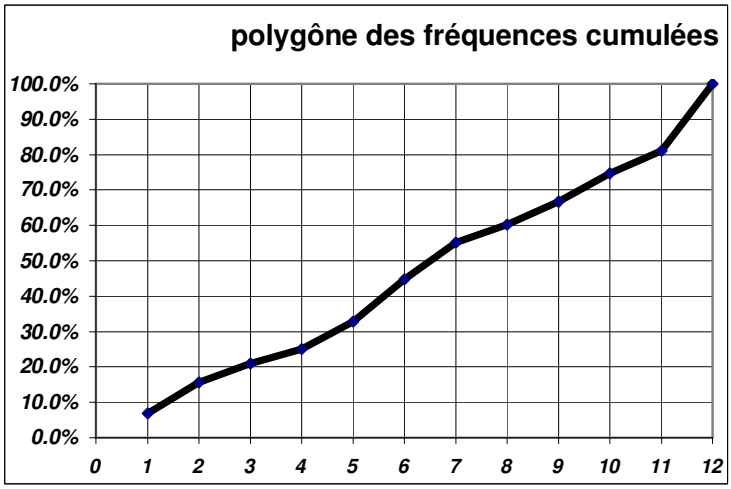
## 2.1. Données discrètes

Exemple : Un commerçant a réalisé les bénéfices suivants pendant les douze mois d'une année :

| janv | févr | mars | avr | mai | juin | juil | août | sept | oct | nov | déc |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 5600 | 7350 | 4500 | 3280 | 6600 | 9800 | 8750 | 4230 | 5340 | 6700 | 5320 | 15680 |

A la fin de chaque mois il veut calculer le bénéfice total de l'année jusqu'à ce jour, c.à.d. le bénéfice cumulé. Il doit alors calculer la somme des bénéfices de ce mois et de tous les mois qui précèdent. Il peut aussi exprimer ces résultats en pourcentages par rapport au bénéfice total de l'année (ceci bien sûr seulement à la fin de l'année).

| $x_i$ | $f_i$ | eff. cumulé | fréq. | fréq. cumulée |
|---|---|---|---|---|
| 1 | 5600 | 5600 | 6.7% | 6.7% |
| 2 | 7350 | 12950 | 8.8% | 15.6% |
| 3 | 4500 | 17450 | 5.4% | 21.0% |
| 4 | 3280 | 20730 | 3.9% | 24.9% |
| 5 | 6600 | 27330 | 7.9% | 32.9% |
| 6 | 9800 | 37130 | 11.8% | 44.7% |
| 7 | 8750 | 45880 | 10.5% | 55.2% |
| 8 | 4230 | 50110 | 5.1% | 60.3% |
| 9 | 5340 | 55450 | 6.4% | 66.7% |
| 10 | 6700 | 62150 | 8.1% | 74.7% |
| 11 | 5320 | 67470 | 6.4% | 81.1% |
| 12 | 15680 | 83150 | 18.9% | 100.0% |




De façon générale :

***L'effectif cumulé d'une valeur est la somme de l'effectif de cette valeur et de toutes celles qui précèdent.***

Par exemple l'effectif cumulé de $x_5$ est $f_1 + f_2 + f_3 + f_4 + f_5 = \sum_{j=1}^{5} f_j = \sum_{j \le i} f_j$ .

L'effectif cumulé de $x_i$ peut être noté formellement $\sum_{j \le i} f_j$ .

De façon analogue on définit les fréquences cumulées des valeurs du caractère.
L'effectif cumulé de la dernière valeur correspond à l'effectif total, et la fréquence cumulée de la dernière valeur doit donner 1 respectivement 100 %.
Le polygône des effectifs resp. des fréquences cumulés est la ligne polygônale qui relie les points dont les abscisses sont les valeurs du caractères, et les ordonnées les effectifs ou fréquences cumulés. Le graphique des fréquences cumulées est une ligne (brisée) qui monte nécessairement de gauche à droite.

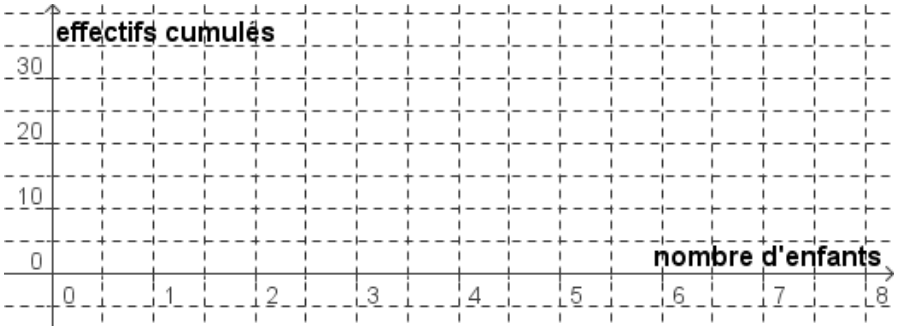
***Questions :***

- ***A quel moment de l'année le commerçant a-t-il réalisé la moitié du bénéfice de l'année ?***
- ***Comment peut-on retrouver les « meilleurs » mois de l'année sur le graphique ?***

*Remarque : Parfois on trace plutôt une* ***courbe en escalier*** *pour montrer que le caractère ne prend que des valeurs isolées, et qu'aucune observation ne s'ajoute entre ces valeurs isolées ; dans notre exemple, on peut cependant supposer que le bénéfice réalisé pendant un mois se répartit uniformément sur ce mois.*

*Reprenons l'exemple 1.1.a. de la page 4 :*

| Nombre d'enfants par famille ($x_i$) | | | | | | |
|---|---|---|---|---|---|---|
| Nombre de familles ($f_i$) | | | | | | |
| Effectifs cumulés | | | | | | |




***Exercice 720***

Des élèves de deux écoles différentes ont passé des tests cotés sur 20. Voici les résultats :

**A**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 14 | 16 | 8 | 12 | 13 | 8 | 10 | 12 |
| 11 | 17 | 16 | 18 | 9 | 4 | 12 | 16 |
| 13 | 12 | 9 | 16 | 10 | 14 | 17 | 14 |

**B**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 8 | 12 | 16 | 15 | 20 | 17 | 16 | 13 |
| 9 | 13 | 20 | 15 | 12 | 7 | 6 | 8 |
| 3 | 12 | 17 | 13 | | | | |

1) Pour chaque école, réalisez un tableau reprenant les résultats, les effectifs, les fréquences, les effectifs cumulés et les fréquences cumulées.
2) Comparez les résultats obtenus en réalisant le diagramme qui vous semble le plus adéquat.
3) Reprenez l'étude en groupant les résultats par classes convenablement choisies.

## 2.2. Données continues

Dans le cas d'un caractère continu, les effectifs cumulés sont associés aux classes $[a_i , a_{i+1}[$ .
Pour construire le polygone des effectifs cumulés, on construit les points dont les abscisses sont **les limites supérieures des classes (expliquez pourquoi !)**, et dont les ordonnées sont les effectifs cumulés. Le polygone des effectifs cumulés est la ligne polygonale joignant ces points. Pour construire le polygone des fréquences cumulées, on remplace les effectifs par les fréquences dans ce qui précède.
On admet qu'il y a linéarité, c'est-à-dire répartition uniforme des effectifs à l'intérieur de chaque classe.

***Exercice 721*** Un fabricant de bouteilles a enregistré le nombre suivant de bouteilles rejetées chaque jour, sur une période de 30 jours :

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 112 | 127 | 92 | 147 | 134 | 131 | 104 | 99 | 116 | 122 |
| 125 | 118 | 109 | 96 | 142 | 127 | 106 | 100 | 132 | 138 |
| 113 | 123 | 118 | 131 | 115 | 94 | 144 | 124 | 117 | 124 |

a. Classez les données en 6 intervalles, en commençant par 91 – 100, et en terminant par 141 – 150. Complétez ensuite un tableau comprenant les effectifs et les fréquences, ainsi que les effectifs et les fréquences cumulés.
b. Déduisez de votre tableau :
   i. le nombre de fois qu'il y avait moins de 120 bouteilles refusées
   ii. le pourcentage des jours où l'on avait entre 91 et 100 bouteilles refusées (inclusivement).
   iii. le pourcentage des jours avec moins de 130 bouteilles refusées.
c. Construisez un histogramme et un polygone des effectifs cumulés.

***Exercice 722*** Un institut de marketing a réalisé auprès de 100 personnes une étude sur la fréquence de leurs visites au restaurant sur une année.

Les résultats sont résumés dans le tableau ci-contre.

| Visites par année | nombre de pers. |
|---|---|
| 1 - 20 | 15 |
| 21 - 40 | 21 |
| 41 - 60 | 24 |
| 61 - 80 | 18 |
| 81 - 100 | 12 |
| 101 - 120 | 6 |
| 121 - 140 | 4 |

a. Ajoutez au tableau les fréquences cumulées.
b. Construisez le polygone des fréquences.
c. Construisez le polygone des fréquences cumulées et répondez aux questions suivantes :
   i. Quel est le pourcentage des personnes qui ont mangé moins de 81 fois au restaurant ?
   ii. Une personne se situe dans les derniers 25 % de la distribution. Quel est le nombre maximal de fois qu'elle est allée au restaurant ?

# 3. Indicateurs numériques

Le dernier niveau de description statistique est le résumé numérique d'une distribution statistique par des indicateurs numériques ou des paramètres caractéristiques.

## 3.1. Indicateurs de position de tendance centrale

Ces paramètres ont pour objectif des caractériser l'ordre de grandeur des observations. Les trois principales mesures de tendance centrale sont le mode, qui se base sur quelques données seulement, la médiane, qui fait abstraction de la plupart des données, et la moyenne (arithmétique), qui se calcule à partir de toutes les données.

### 3.1.a. Le mode (Mo)

***Pour un caractère discret, le mode est la valeur du caractère la plus fréquente dans l'échantillon.***

Le mode correspond donc à la modalité dont l'effectif est le plus élevé. Il n'est pas toujours unique (un mode → série unimodale, deux modes → série bimodale, plusieurs modes → plurimodale).

Pour un caractère continu, on appelle mode ou classe modale la classe correspondant au plus grand effectif.

Ex.1 : Donnéés brutes 1 2 2 3 3 3 3 4 4 4 4 5 6 deux modes : 3 et 4

Ex. 2 : Données groupées $x_i$ 1 2 3 4 5 6 7
$f_i$ 0 1 2 5 2 3 0 un mode : 4

***Détermination graphique :***
*Le mode est l'abscisse du point d'ordonnée maximum du diagramme en bâtons, du polygone des effectifs ou du polygone des fréquences.*

### 3.1.b. La médiane (Me)

***On appelle médiane d'une série statistique la valeur telle que l'effectif des valeurs inférieures à cette valeur est égal à l'effectif des valeurs supérieures.***

C'est donc la valeur qui partage l'échantillon en deux groupes de même taille.
Ou encore : la valeur dont la fréquence cumulée est 0,5 (50 %). On l'appelle également **50$^{e}$ centile** ou **2$^{e}$ quartile**.

Ex.1 : Données brutes (13) 1 2 2 3 3 3 **3** 4 4 4 4 5 6 la médiane est 3 (la 7$^{e}$ valeur)

Ex.2 : Données brutes (8) 4 5 6 **7 8** 9 9 10 médiane entre 7 et 8
On prend la moyenne : 7,5
(Ou l'intervalle médian ]7,8[ )

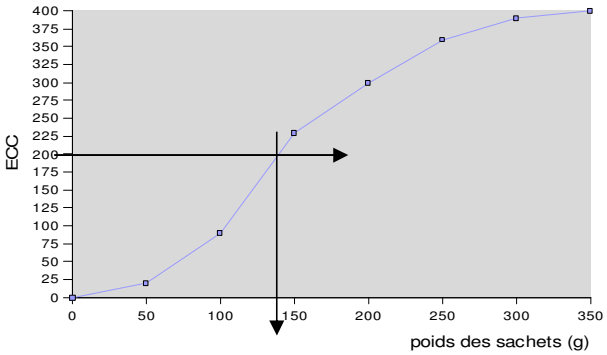
*Si la série comporte un nombre pair 2n de termes, la **médiane** de cette série est la demi-somme de la valeur du terme de rang n et de la valeur du terme de rang n+1.*
*Si la série comporte un nombre impair 2n+1 de termes, la **médiane** de cette série est la valeur du terme de rang n+1 (c'est-à-dire le terme partageant la série en deux groupes de même effectif).*

Ex.3 : Exemple 1.1.a. p 4

| | | | | | | |
|---|---|---|---|---|---|---|
| Nombre d'enfants par famille ($x_i$) | | | | | | |
| Nombre de familles ($f_i$) | | | | | | |
| Effectifs cumulés | | | | | | |

Ex. 4 : caractère continu

| Poids des sachets (en g) | nombre de sachets | Effectif cumulé |
|---|---|---|
| ]0 ; 50] | 20 | 20 |
| ]50 ; 100] | 70 | 90 |
| ]100 ; 150] | 140 | 230 |
| ]150 ; 200] | 70 | 300 |
| ]200 ; 250] | 60 | 360 |
| ]250 ; 300] | 30 | 390 |
| ]300 ; 350] | 10 | 400 |
| | 400 | |




Le rang de la médiane est N/2 = 400/2 = 200. La droite (*AB*) d'ordonnée N/2 et parallèle à l'axe des abscisses coupe la courbe en M. L'abscisse du point M, égale à 139 est la valeur médiane de la série statistique.
Sur le polygone des fréquences cumulées, il suffit de chercher l'abscisse du point d'ordonnée 0,5.

***Exercice 723*** On a enregistré les temps de jeunes athlètes courant les 200 m :

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 21.1 | 28 | 26.9 | 31.9 | 23.7 | 28.8 | 27.9 | 31.3 |
| 21.5 | 26.8 | 27.4 | 31.2 | 21.4 | 29.9 | 29.4 | 31.5 |
| 20.4 | 25.1 | 25.8 | 33.6 | 23.7 | 25.6 | 29.1 | 30.3 |
| 21.5 | 28.2 | 28.2 | 31.3 | 22.4 | 25.7 | 25.1 | 30.3 |
| 21.9 | 29.1 | 28.7 | 30.1 | 21.8 | 27.8 | 29.1 | 34.3 |
| 22.5 | 25.2 | 25.5 | 32.9 | 22.3 | 29 | 27.2 | 33.3 |

a) Construisez un diagramme en tiges et feuilles pour représenter les données.

b) Utilisez ce diagramme pour identifier la médiane.

### 3.1.c. La moyenne (arithmétique) ($\bar{x}$)

La moyenne est la mesure la plus commune de tendance centrale.

Pour les données non groupées :

**Soit un échantillon de *n* valeurs observées $x_1, x_2, \ldots, x_i, \ldots, x_n$ d'un caractère quantitatif *X*, on définit sa moyenne observée $\bar{x}$ comme la moyenne arithmétique des *n* valeurs :**

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

**Si les données observées sont regroupées en k classes d'effectifs $f_i$ (caractère continu ou discret), il faut les pondérer par les effectifs correspondants :**

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} f_i x_i \quad \text{avec } n = \sum_{i=1}^{k} f_i$$

Pour un caractère continu, les $x_i$ représentent ici les centres des classes.

Exemple :
On a relevé le prix de vente d'un CD et le nombre de CD vendus chez différents fournisseurs.
Les résultats sont donnés dans le tableau suivant :

| Prix de vente en euros | 15 | 16 | 17 | 18 | 19 |
|---|---|---|---|---|---|
| Nombre de CD vendus | 97 | 34 | 43 | 20 | 6 |

Donner le prix de vente moyen p de ce CD.

## Quelques avantages et inconvénients des caractéristiques de position

| | Avantages | Inconvénients |
|---|---|---|
| **Moyenne arithmétique** | - Facile à calculer,<br>- Répond au principe des moindres carrés. | - Fortement influencée par les valeurs extrêmes de la série<br>- Représente mal une population hétérogène (polymodale). |
| **Médiane** | - Pas influencée par les valeurs extrêmes<br>- Peu sensible aux variations d'amplitude des classes,<br>- Calculable sur des caractères cycliques (saison, etc.) où la moyenne a peu de signification. | - Se prête mal aux calculs statistiques,<br>- Suppose l'équi-répartition des données<br>- Ne représente que la valeur qui sépare l'échantillon en 2 parties égales. |
| **Mode** | - Pas influencée par les valeurs extrêmes<br>- Calculable sur des caractères cycliques (saison, etc.) où la moyenne a peu de signification,<br>- Bon indicateur de population hétérogène. | - Se prête mal aux calculs statistiques,<br>- Très sensible aux variations d'amplitude des classes,<br>- Son calcul ne tient compte que des individus dont les valeurs se rapprochent de la classe modale. |

***Exercice 724***
Trouvez la médiane pour les données suivantes :

a 21, 23, 24, 25, 29, 31, 34, 37, 41
b 105, 106, 107, 107, 107, 107, 109, 120, 124, 132
c 173, 146, 128, 132, 116, 129, 141, 163, 187, 153, 162, 184

***Exercice 725***
Trouvez la moyenne, le mode et la médiane pour les distributions suivantes :

a 3, 6, 5, 6, 4, 5, 5, 6, 7 b 13, 12, 15, 13, 18, 14, 16, 15, 15, 17.

***Exercice 726***
Trouvez x sachant que les valeurs 5, 9, 11, 12, 13, 14, 17 et x ont une moyenne de 12.

***Exercice 727***
Vers la fin d'une saison, un joueur de basket-ball a joué 14 matches avec une moyenne de 16,5 points par match. Au cours des deux derniers matches, il réalise 21 points respectivement 24 points. Quelle est finalement sa moyenne sur toute la saison ?

***Exercice 728***
Un échantillon de 12 mesures a une moyenne de 16,5 , et un échantillon de 15 mesures a une moyenne de 18,6. Quelle est la moyenne de 27 mesures ?

***Exercice 729***
15 mesures sur 31 sont au-dessous de 10 cm, et 12 mesures sont au-dessus de 11 cm. Trouvez la médiane sachant que les 4 autres mesures sont 10,1 cm, 10,4 cm, 10,7 cm et 10,9 cm.

# V Statistiques

***Exercice 730***

Une série de 9 mesures ont une médiane et une moyenne égales à 12. On sait que 7 des mesures sont 7, 9, 11, 13, 14, 17 et 19. Quelles sont les deux autres mesures ?

***Exercice 731***

Les 7 valeurs d'une échantillon sont 2, 7, 3, 8, 4, a et b, avec a < b. La moyenne est 6, et la médiane est 5. Trouvez :

a. a et b. b. le mode

***Exercice 732***

On a vérifié 51 paquets de chocolats et compté le contenu. Le tableau ci-contre indique le nombre de chocolats par paquet . Trouvez la moyenne, le mode et la médiane pour cette distribution.

| Nombre par paquet | effectifs |
|---|---|
| 32 | 6 |
| 33 | 8 |
| 34 | 9 |
| 35 | 13 |
| 36 | 10 |
| 37 | 3 |
| 38 | 2 |

***Exercice 733***

50 étudiants ont passé un test en mathématiques. Les résultats sont les suivants :

| Note | 0-9 | 10-19 | 20 - 29 | 30 - 39 | 40 - 49 |
|---|---|---|---|---|---|
| Effectif | 2 | 5 | 7 | 27 | 9 |

Trouvez le résultat moyen.

***Exercice 734***

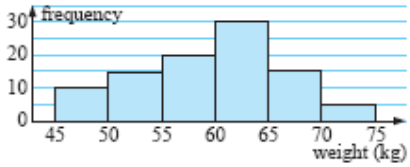
Voici les scores réalisés par Chloé dans ses matches de basket au cours de la dernière saison :

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 15 | 8 | 6 | 10 | 0 | 9 | 2 | 16 | 11 | 23 | 14 | 13 | 17 | 16 | 20 | 12 | 13 |
| 12 | 10 | 3 | 13 | 5 | 18 | 14 | 19 | 4 | 15 | 15 | 19 | 19 | 14 | 6 | 11 | 29 |
| 8 | 9 | 3 | 20 | 9 | 25 | 7 | 15 | 19 | 21 | 23 | 12 | 17 | 22 | 14 | 26 | |

a. Trouvez son score moyen.
b. Trouvez le score moyen en groupant les valeurs en intervalles
   i. 0-4, 5-9, 10-14 etc
   ii. 0-8, 9-16, 17-24, 25-30.
c. Commentez les résultats trouvés en a et b.

***Exercice 735***

L'histogramme ci-contre montre les poids mesurés en kg pour un groupe d'élèves âgés de 10 ans dans une école.




**a.** Quel est le nombre d'élèves dans l'échantillon ?
**b.** Calculez le poids moyen.
**c.** Combien d'élèves ont un poids inférieur à 56 kg ?
**d.** Quel est le pourcentage des élèves avec un poids entre 50 et 60 kg ?
**e.** Si un élève est choisi au hasard, quelle est alors la probabilité qu'il pèse moins de 60 kg ?

***Exercice 736***

Le tableau ci-contre indique l'âge des conducteurs impliqués dans des accidents de circulation en ville au cours d'une année. Tracez un diagramme des fréquences cumulées et utilisez-le pour trouver

a. l'âge médian des conducteurs impliqués dans les accidents
b. le pourcentage des conducteurs au-dessous de 23 ans.
c. Estimez la probabilité pour que l'âge d'un conducteur impliqué dans un accident soit
   i. inférieur ou égal à 27 ans ii. égal à 27 ans.

| Age (en années) | nombre d'accidents |
|---|---|
| $16 \leqslant x < 20$ | 59 |
| $20 \leqslant x < 25$ | 82 |
| $25 \leqslant x < 30$ | 43 |
| $30 \leqslant x < 35$ | 21 |
| $35 \leqslant x < 40$ | 19 |
| $40 \leqslant x < 50$ | 11 |
| $50 \leqslant x < 60$ | 24 |
| $60 \leqslant x < 80$ | 41 |

***Exercice 737***

Le tableau indique le nombre d'erreurs trouvées sur des pages choisies au hasard dans un annuaire téléphonique. Trouvez le nombre médian d'erreurs.

| Nombre d'erreurs | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| fréquence | 67 | 35 | 17 | 8 | 11 | 2 | 1 |

***Exercice 738*** Jana a passé 7 tests (sur un maximum de 12 points), mais peut seulement retrouver 5 des 7 résultats : 9,5,7,9 et 10. Elle demande à son professeur pour connaître les autres résultats. Celui-ci lui répond que le mode de ses résultats est 9 et que la moyenne est 8. Quels sont les deux autres résultats, sachant d'autre part que 5 était le plus mauvais résultat ?

***Exercice 739*** Le tableau suivant montre le nombre d'appels téléphoniques effectués par jour par un groupe de 50 jeunes filles agées de 15 ans :

| **Nombre d'appels** | **effectifs** |
|---|---|
| 0 | 5 |
| 1 | 8 |
| 2 | 13 |
| 3 | 8 |
| 4 | 6 |
| 5 | 3 |
| 6 | 3 |
| 7 | 2 |
| 8 | 1 |
| 11 | 1 |

a) Trouvez la moyenne, la médiane et le mode.
b) Construisez un diagramme à barres et montrez les positions des 3 mesures centrales sur l'axe des abscisses.
c) Décrivez la distribution des données.
d) Pourquoi la moyenne est-elle plus élevée que la médiane pour ces données ?
e) Quelle est la mesure centrale qui convient le mieux pour cette série ?

***Exercice 740*** Les séries statistiques suivantes sont données par leur diagramme à tige et feuilles. Déterminez i) la moyenne ii) la médiane iii) le mode.

a)

| tige | feuilles |
|---|---|
| 5 | 356 |
| 6 | 0124679 |
| 7 | 3368 |
| 8 | 47 |
| 9 | 1 |

5/3 signifie 53

b)

| tige | feuilles |
|---|---|
| 3 | 7 |
| 4 | 0488 |
| 5 | 00136789 |
| 6 | 036777 |
| 7 | 069 |
| 8 | 1 |

3/7 signifie 3,7

## 3.2. Indicateurs de position de tendance non centrale

### 3.2.a. Les quartiles

On a vu que la médiane partage une série de données ordonnées en deux groupes égaux. Les quartiles font un partage en quatre groupes égaux.

- **Le premier quartile $Q_1$** (quartile inférieur) d'une série statistique est la valeur de la série qui correspond à la fréquence cumulée 0,25.
- **Le troisième quartile $Q_3$** (quartile supérieur) d'une série statistique est la valeur de la série qui corrspond à la fréquence cumulée 0,75.
- **Le deuxième quartile $Q_2$** est la médiane.

**Remarques :**

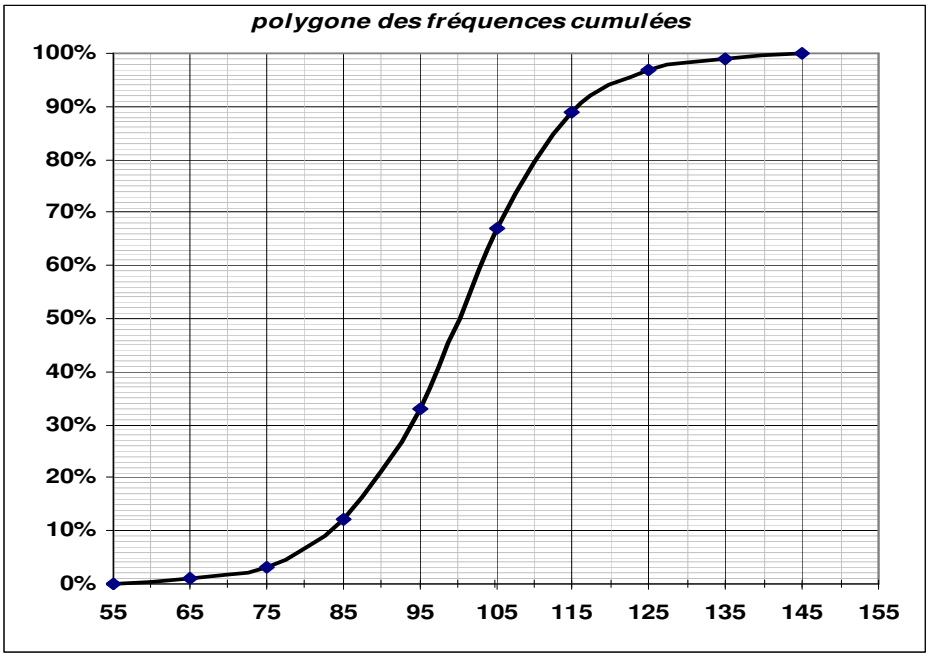
- On rencontre des petites différences quant à la manière de déterminer les quartiles pour des séries discrètes. Dans la pratique, ces différences ont généralement peu d'importance vu la taille des séries.
- Nous considérons ici les quartiles comme les médianes des deux séries obtenues après avoir partagé la série initiale par sa médiane.

Ex : 2 3 3 3 4 4 5 5 **5** 5 6 6 6 7 7 8 9 (N = 17)
Me = $Q_2$ = 5 (9e valeur)
2 3 3 **3 4** 4 5 5 5 6 6 **6 7** 7 8 9 $Q_1$ = 3,5 $Q_3$ = 6,5

Autre possibilité :
2 3 3 3 **4** 4 5 5 **5** 5 6 6 **6** 7 7 8 9
$Q_1$ = 4 (précédée de 4, suivie de 12 valeurs)
$Q_3$ = 6 (précédée de 12, suivie de 4 valeurs)

Ex : 2 2 3 4 4 5 5 6 **6 7** 7 8 8 8 9 10 11 13 (N = 18)
Me = $Q_2$ = 6,5
2 2 3 4 **4** 5 5 6 **6** **7** 7 8 8 **8** 9 10 11 13 $Q_1$ = 4 $Q_3$ = 8

Exemple (caractère continu) :




### 3.2.b. Les centiles

Les centiles sont les valeurs qui divisent la population en centièmes.
***Le n-ième centile est la valeur qui est précédée de n % des données.***
Ou encore :
***Le n-ième centile est la valeur dont la fréquence cumulée est n %.***

En particulier, le 25e centile correspond au 1er quartile, le 50e centile à la médiane, le 75e centile au 3e quartile.

Ex : Déterminez pour l'exemple au-dessus :
- Le 23e centile
- Le 78e centile
- L'intervalle des valeurs qui comprend les 80 % de la population qui restent lorsqu'on élimine les 10 % aux extrémités.

***Exercice 741*** Un enseignant, donnant des cours parallèles, a effectué un test dans ses classes. En voici les résultats :

| Notes | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cours A | 0 | 0 | 2 | 0 | 1 | 2 | 6 | 11 | 16 | 19 | 29 | 18 | 13 | 10 | 7 | 2 | 2 | 1 | 0 | 1 | 0 |
| Cours B | 0 | 0 | 0 | 1 | 1 | 2 | 2 | 8 | 18 | 21 | 37 | 32 | 16 | 5 | 4 | 1 | 2 | 0 | 0 | 0 | 0 |

a) Calculez la moyenne, la médiane et le mode de chacune des classes. Déterminez aussi les trois quartiles dans chaque classe.
b) Le professeur a le sentiment que les deux groupes n'ont pas le même profil, bien que la moyenne soit la même. Il commence à s'intéresser à la **dispersion** des notes autour de la moyenne. Pour cela, il compare le nombre $Q_3 - Q_1$ , appelé **écart interquartile**, qu'il obtient dans les deux classes. Quelle est la classe la plus homogène ?
c) Imaginez d'autres possibilités pour caractériser la façon dont les valeurs se répartissent autour de la moyenne ou de la médiane.

## 3.3. Indicateurs de dispersion (mesures de variabilité)

Si les mesures de tendance centrale nous informent sur une dimension importante d'une distribution, elles sont cependant souvent incomplètes. Elles peuvent même, dans certains cas, induire une représentation tronquée de la distribution observée. La dispersion des données autour de la tendance centrale est au moins aussi importante que la tendance centrale elle-même. On peut approcher cette dispersion principalement de trois manières différentes :

- On peut simplement regarder l'écart entre la valeur observée la plus basse et la plus élevée : C'est ce qu'on appelle l'***étendue*** de la distribution.
- On peut déterminer l'intervalle centré autour de la médiane comprenant la moitié (p.ex.) de la population, c.à.d. l'intervalle qui comprend la moitié centrale de la distribution. Cet intervalle est appelé ***intervalle interquartile***, et l'écart entre ses limites est l'***écart interquartile***.
- On peut mesurer les écarts entre les valeurs observées et la moyenne de ces valeurs, et calculer une moyenne de ses écarts. I s'agit de ***l'écart-moyen***. Nous allons cependant considérer ici des msesures un peu différentes : la ***variance*** et l'***écart-type***.

### 3.3.a. L'étendue

***On appelle étendue d'une série statistique la différence entre la plus grande et la plus petite valeur observée.***

L'étendue est facile à calculer et intuitive, mais elle s'appuie uniquement sur les valeurs extrêmes qui souvent sont accidentelles.

Exemple 1 :

| $x_i$ | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|
| $f_i$ | 0 | 2 | 4 | 3 | 0 | 2 | 7 | 0 |

étendue =

Exemple 2 :

| Tailles en cm | Nombre d'élèves |
|---|---|
| [130,134[ | 2 |
| [134,138[ | 6 |
| [138,142[ | 41 |
| [142,146[ | 37 |
| [146,150[ | 20 |
| [150,154[ | 4 |

Etendue =

### 3.3.b. L'écart interquartile

L'écart interquartile est une tentative visant à contourner le problème de l'énorme dépendance de l'étendue vis-à-vis des scores extrêmes.

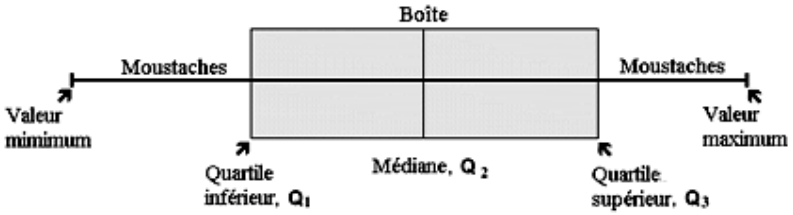
***La différence $Q_3 - Q_1$ est appelée écart interquartile.***
***L'intervalle $[Q_1, Q_3]$ compris entre le $1^{er}$ et le $3^e$ quartile porte le nom d'intervalle interquartile.***

L'écart interquartile est l'étendue restante si on élimine les 25 % supérieurs et inférieurs de la distribution. C'est donc l'étendue des 50 % du milieu des observations.

### 3.3.c. Diagramme en boîte à moustaches (à pattes)

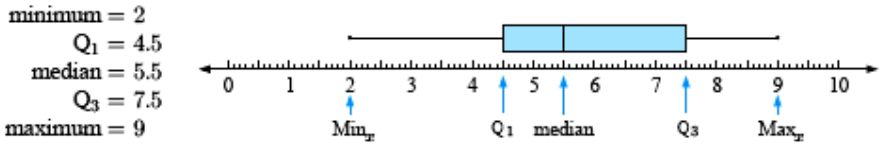
Il s'agit d'une visualisation du ***résumé par 5 nombres*** de la série statistique :

- Les deux valeurs extrêmes (donc également l'étendue)
- Les quartiles $Q_1$ et $Q_3$ (donc également l'écart interquartile $Q_3 - Q_1$)
- La médiane.




**Un diagramme en boite** est un rectangle délimité par le premier quartile et le troisième quartile .
Pour l'obtenir, on trace un axe horizontal ( ou vertical ) sur lequel on place les valeurs de $Q_1$ , $Q_3$ et Me . L'un des côtés du rectangle a pour longueur l'écart interquartile, l'autre est quelconque. On complète ce diagramme en traçant deux traits horizontaux : l'un joignant $Q_1$ au minimun de la série et l'autre joigant $Q_3$ au maximun de la série. Ce sont les moustaches parfois appelées pattes.

**Exemple**




### Traitement des valeurs aberrantes (atypiques) :

Les valeurs dites extrêmes, atypiques, exceptionnelles, *(outliers)* situées au-delà des valeurs adjacentes sont individualisées. Elles sont représentées par des marqueurs (carré, ou étoile, etc.).
On peut définir les valeurs atypiques comme étant des valeurs qui sont

- ***supérieures à la limite supérieure qui est*** $Q_3 + 1{,}5 \bullet (Q_3 - Q_1)$

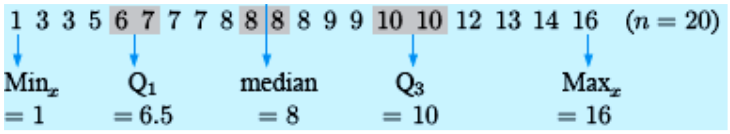
***ou bien***

- ***inférieures à la borne inférieure qui est*** $Q_1 - 1{,}5 \bullet (Q_3 - Q_1)$

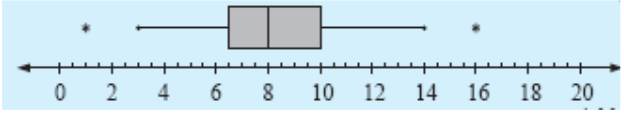
**Exemple :** On donne les valeurs

3, 7, 8, 8, 5, 9, 10, 12, 14, 7, 1, 3, 8, 16, 8, 6, 9, 10, 13, 7

Les valeurs ordonnées sont




EIQ = $Q_3 - Q_1 = 10 - 6{,}5 = 3{,}5$
Borne supérieure = $10 + 1{,}5 \bullet 3{,}5 = 15{,}25$ et Borne inférieure = $6{,}5 - 1{,}5 \bullet 3{,}5 = 1{,}25$
Les valeurs 1 et 16 sont donc à considérer comme atypiques.
On obtient alors le diagramme suivant :




Les moustaches sont tracées jusqu'aux dernières valeurs qui ne sont pas atypiques.

# V Statistiques

***Exercice 742*** Pour chacune des distributions suivantes, trouvez :

i.. le 1er quartile ii. le 3e quartile iii. l'écart interquartile iv. l'étendue

a. 2, 3, 4, 7, 8, 10, 11, 13, 14, 15, 15

b. 35, 41, 43, 48, 49, 50, 51, 52, 52, 52, 56

c.

| Tige | Feuilles |
|---|---|
| 1 | 3 5 7 7 9 |
| 2 | 0 1 3 4 6 7 8 9 |
| 3 | 0 1 2 7 |
| 4 | 2 6 |
| 5 | 1 |

4 | 2 signifie 42.

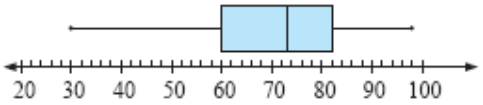
d.

| Résultat | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| effectif | 1 | 4 | 7 | 3 | 3 | 1 |

***Exercice 743*** Reprenez les données de l'exercice 30 et comparez les deux classes à l'aide des diagrammes à boîte et moustaches (utilisez aussi la TI 84).

***Exercice 744***

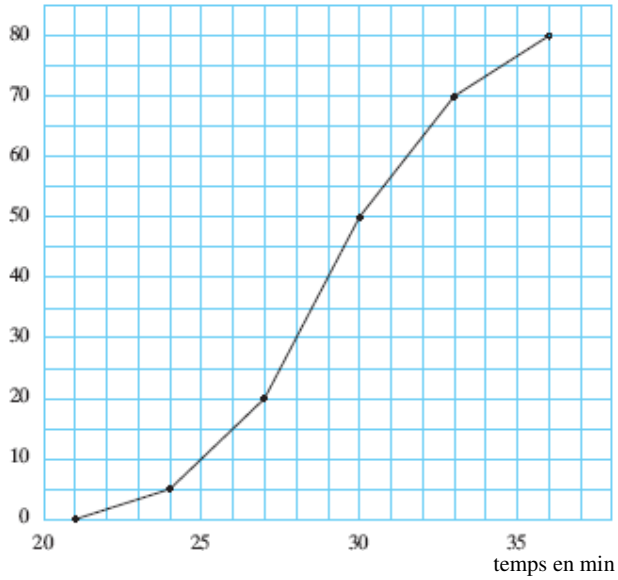
Le diagramme en boîte à moustaches ci-contre indique les résultats réalisés dans un test sur 100 points par un groupe d'élèves.




a. Quel est le score i) le plus élevé ii) le plus bas ?
b. Quel est le score médian ?
c. Quelle est l'étendue des scores réalisés ?
d. Quel est le pourcentage des élèves qui ont réalisés au moins 60 points ?
e. Donnez l'écart interquartile ?
f. Les premiers 25 % des élèves ont les résultats entre …. et …. Points.
g. Un élève qui a obtenu 70 points se trouve-t-il dans la première moitié de la classe ?
h. Commentez la symétrie de distribution.

***Exercice 745***

Le diagramme des fréquences cumulées suivant indique les temps en minutes réalisés par les 80 participants à une course de cross country. Trouvez :




a) $Q_1$
b) Le temps médian.
c) $Q_3$
d) L'écart interquartile.
e) Le 40e centile.

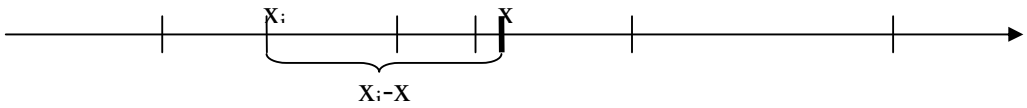
***Exercice 746*** Julie a examiné une nouvelle variété de pois et a compté les pois par gousse. Ses résultats sont :

5, 8, 10, 4, 2, 12, 6, 5, 7, 7, 5, 5, 5, 13, 9, 3, 4, 4, 7, 8, 9, 5, 5, 4, 3, 6, 6, 6, 6, 9, 8, 7, 6

a. Trouvez les trois quartiles.
b. Quel est l'écart interquartile.
c. Trouvez les bornes supérieure et inférieure pour les valeurs atypiques.
d. A-t-on des valeurs atypiques. Donnez le diagramme en boîte à moustaches.

## 3.3.d. Variance et écart-type

On peut penser que, pour mesurer le degré de dispersion des valeurs observées autour de la moyenne, le plus logique serait de considérer tous les écarts $x_i - \bar{x}$ et d'en calculer la moyenne. Mais le fait est que certaines de ces différences sont positives, d'autres négatives, et que finalement le résultat de ce calcul serait 0, car les écarts positifs et négatifs ce compenseraient exactement. En moyenne, la différence entre les valeurs observées et leur moyenne est zéro !




On peut alors être amené à prendre les valeurs absolues des écarts. Cette solution est parfaitement légitime, et la mesure qu'on obtient est appelée **écart (absolu) moyen** (on l'a peut-être calculée à l'exercice 30). Cette mesure de la dispersion ne figure cependant pas au programme du BI.

Au lieu de prendre la valeur absolue des écarts on peut aussi, pour éviter le signe négatif, élever les écarts au carré et en calculer ensuite la moyenne.

***On appelle variance la moyenne arithmétique des carrés des écarts entre les valeurs observées et leur moyenne. Notation : $s^2$ ou $s_n^2$ (ou V).***

Formule pour les données non groupées : $\boxed{s_n^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}$ n : effectif total

Formule pour les données groupées : $\boxed{s_n^2 = \frac{\sum_{i=1}^{k} f_i \cdot (x_i - \bar{x})^2}{n}}$ avec $\boxed{n = \sum_{i=1}^{k} f_i}$

***On appelle écart-type la racine carrée positive de la variance. Notation : $s$ ou $s_n$***

Pour les données groupées, on a donc : $\boxed{s_n = \sqrt{\frac{\sum_{i=1}^{k} f_i \cdot (x_i - \bar{x})^2}{n}}}$ avec $\boxed{n = \sum_{i=1}^{k} f_i}$

Remarques :

- *L'écart type est un paramètre plus fin que l'étendue, car il tient compte de la répartition de toutes les valeurs.*
- *L'écart type à la même unité que les valeurs de la série étudiée.*

- *L'écart type mesure la dispersion des valeurs de la série autour de la moyenne . Plus l'écart type est petit, plus les valeurs de la série sont concentrées autour de la moyenne.*
- *On peut résumer une série statistique par le couple : **( moyenne ; écart type )***

Exemple 1
Données non groupées : 2 3 3 4 4 4 5 6 6 7 (n = 10)

Exemple 2 (tableau de la page 9) :

| **Taille (cm)** | $f_i$ | **Centre de classe** $x_i$ | $f_i \cdot x_i$ | $(x_i - \bar{x})^2$ | $f_i \cdot (x_i - \bar{x})^2$ |
|---|---|---|---|---|---|
| [150,155[ | | 152,5 | | | |
| [155,160[ | | 157,5 | | | |
| [160,165[ | | 162,5 | | | |
| [165,170[ | | 167,5 | | | |
| [170,175[ | | 172,5 | | | |
| [175,180[ | | 177,5 | | | |
| [180,185[ | | 182,5 | | | |
| [185,190[ | | 187,5 | | | |
| [190,195[ | | 192,5 | | | |
| | n = | | | | |
| | | | $\bar{x}$ = | | $s_n^2$ = |

Il est recommandé d'organiser tous les calculs dans un tableau.

Au plus tard ici l'usage d'une calculatrice à écran graphique devient indispensable.

D'ailleurs on s'attend à ce que la plupart des calculs soient effectués (sans les calculs intermédiaires) à l'aide d'une calculatrice.

***Exercice 747*** L'institut national de la Santé et de la Recherche médicale (I.N.S.E.R.M.) a effectué une enquête sur les naissances en s'intéressant surtout aux poids des enfants à leur naissance.
On a relevé les résultats suivants :

| Poids en grammes | efectifs |
|---|---|
| [500,1000[ | 7 |
| [1000,1500[ | 12 |
| [1500,2000[ | 21 |
| [2000,2500[ | 52 |
| [2500,3000[ | 234 |
| [3000,3500[ | 532 |
| [3500,4000[ | 306 |
| [4000,4500[ | 75 |
| [4500,5000[ | 11 |

2. Calculez, en rassemblant tous les calculs dans un tableau :
a. La moyenne arithmétique.
b. La variance et l'écart-type.
c. Les fréquences en %.
d. Les fréquences cumulées en %.
3. Donnez le polygone des fréquences cumulées et déterminez les trois quartiles.
4. Quel est l'intervalle interquartile ?
   Que représente cet intervalle d'une façon générale ?
5. Déterminez le mode, la médiane et l'étendue de cette série statistique.

# V Statistiques

***Exercice 748*** Le tableau suivant donne les groupes d'âge des chauffeurs impliques dans les accidents de la circulation au cours d'une certaine année. Construisez le polygone des fréquences cumulées et utilisez-le pour déterminer·

1. l'âge médian des conducteurs impliqués dans les accidents
2. le pourcentage des conducteurs âgés de 23 ans ou moins
3. Estimez la probabilité pourqu'un conducteur impliqué dans un accident soit âgé
   - i) de 27 ans au plus
   - ii) de 27 ans .

| âge | Nombre d'accidents |
|---|---|
| $16 \leq x < 20$ | 59 |
| $20 \leq x < 25$ | 82 |
| $25 \leq x < 30$ | 43 |
| $30 \leq x < 35$ | 21 |
| $35 \leq x < 40$ | 19 |
| $40 \leq x < 50$ | 11 |
| $50 \leq x < 60$ | 24 |
| $60 \leq x < 80$ | 41 |

***Exercice 749*** Pour deux classes d'élèves qui étudient l'espagnol on a relevé les résultats suivants :

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Classe A | 64 | 69 | 74 | 67 | 78 | 88 | 76 | 90 | 89 | 84 | 83 | 87 | 78 | 80 | 95 | 75 | 55 | 78 | 81 |
| Classe B | 94 | 90 | 88 | 81 | 86 | 96 | 92 | 93 | 88 | 72 | 94 | 61 | 87 | 90 | 97 | 95 | 77 | 77 | 82 | 90 |

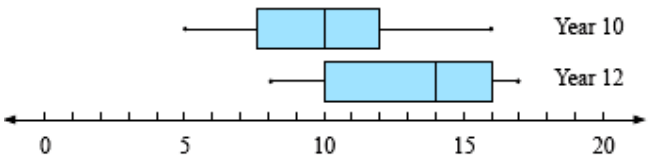
a. Déterminez dans chaque classe les 3 quartiles.
b. Donnez pour chaque classe le diagramme en boîte à moustaches en représentant aussi correctement les éventuelles valeurs atypiques.
c. Quelle est la classe à avoir réalisé les meilleurs résultats ? Pourquoi ?

***Exercice 750*** Les deux amis Shane et Brett ont comparé au cours d'une saison de cricket leurs scores réalisés :

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Shane: | 1 | 6 | 2 | 0 | 3 | 4 | 1 | 4 | 2 | 3 | 0 | 3 | 2 | 4 | 3 | 4 | 3 | 3 | |
| | 3 | 4 | 2 | 4 | 3 | 2 | 3 | 3 | 0 | 5 | 3 | 5 | 3 | 2 | 4 | 3 | 4 | 3 | |
| Brett: | 7 | 2 | 4 | 8 | 1 | 3 | 4 | 2 | 3 | 0 | 5 | 3 | 5 | 2 | 3 | 1 | 2 | 0 | |
| | 4 | 3 | 4 | 0 | 3 | 3 | 0 | 2 | 5 | 1 | 1 | 2 | 2 | 5 | 1 | 4 | 0 | 1 | |

a. Le caractère est-il discret ou continu ?
b. Entrez les données dans votre calculatrice.
c. Produisez un diagramme à bâtons pour chaque série de données.
d. Y a-t-il des valeurs aberrantes ? Faudrait-il les éliminer avant de commencer l'analyse ?
e. Décrivez l'allure de chaque distribution.
f. Comparez les mesures à tendance centrale des deux séries.
g. Comparer les mesures de variabilité des deux séries.
h. Produisez un diagramme avec les deux boîtes à moustaches côté à côté.
i. Imprimez si possible les graphes, diagrammes et autres résultats.
j. Quelle(s) conclusion(s) peut-on tirer ?

***Exercice 751*** Les diagrammes ci-contre comparent les heures passées par des élèves âgés de 10 resp. de 12 ans aux devoirs à domicile.




1. Trouvez pour chaque groupe d'élèves le résumé des 5 nombres.
2. Déterminez :

i) l'étendue ii) l'écart interquartile pour chaque groupe.

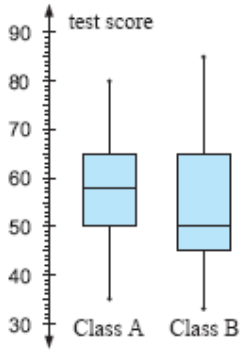
***Exercice 752*** Les tailles au cm près des garçons et des filles dans une école sont les suivantes :

| | |
|---|---|
| garçons | 164 168 175 169 172 171 171 180 168 168 166 168 170 165 171 173 187 179<br>181 175 174 165 167 163 160 169 167 172 174 177 188 177 185 167 160 |
| filles | 165 170 158 166 168 163 170 171 177 169 168 165 156 159 165 164 154 170<br>171 172 166 152 169 170 163 162 165 163 168 155 175 176 170 166 |

a. Trouvez le résumé par les 5 nombres pour chaque groupe. Vérifiez s'il y a des valeurs atypiques et donnez les diagrammes en boîtes à moustaches parallèles.
b. Comparez et commentez les distributions.

***Exercice 753***

Deux classes ont effectué le même test.
On a construit des diagrammes en boîtes à moustaches sur un même axe afin de pouvoir comparer les résultats.




a. Dans quelle classe avait-on
   i) le score le plus élevé ii) le score le plus bas
   iii) une dispersion plus grande des résultats ?
b. Trouvez :
   i) l'étendue des résultats pour la classe B
   ii) l'écart interquartile pour la classe A.
c. Le seuil pour passer le test était de 50 points. Quel est le pourcentage des élèves qui ont réussi
   i) en classe A ii) en classe B ?
d. Décrivez la distribution des résultats dans les deux classes.

***Exercice 754***

Sally et Joanne comparent leurs scores réalisés au cours des 8 derniers matches de basket-ball :

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Points de Sally | 23 | 17 | 31 | 25 | 25 | 19 | 28 | 32 |
| Points de Joanne | 9 | 29 | 41 | 26 | 14 | 44 | 38 | 43 |

a) Déterminez la moyenne et l'écart-type pour les deux séries.
b) Quelle est la mesure qui montre en particulier la régularité des joueuses ?

***Exercice 755***

Les poids en kg de 7 footballeurs sont : 79, 64, 59, 71, 68, 68 et 74.

1. Trouvez la moyenne et l'écart-type.
2. Curieusement le poids de chaque footballeur a augmenté de 10 kg 5 ans plus tard. Trouvez la moyenne et l'écart-type pour la nouvelle série.
3. Commentez vos résultats d'une manière plus générale.

***Exercice 756***

Les poids de 10 dindonneaux en kg sont : 0,8 1,1 1,2 0,9 1,2 1,2 0,9 0,7 1,0 1,1.

1. Trouvez la moyenne et l'écart-type.
2. Un mois plus tard, les poids ont doublé. Trouvez la moyenne et l'écart-type pour la nouvelle série.
3. Commentez vos résultats d'une manière plus générale.

## 3.4. Mesures de l'échantillon et estimateurs de la population

En général, les mesures calculées de tendance centrale et de dispersion se rapportent à des échantillons qu'on a extraits de populations. Le plus souvent, la moyenne et la variance de la population sont inconnues et peuvent seulement être estimées à partir des mesures calculées sur l'échantillon.

Dans les notations, on fait la différence entre les mesures de l'échantillon et les mesures de la population :

| échantillon | | population | | |
|---|---|---|---|---|
| $\bar{x}$ | *moyenne* | $\mu$ | *moyenne* | (mu) |
| $s_n^2$ | *variance* | $\sigma^2$ | *variance* | |
| $s_n$ | *écart-type* | $\sigma$ | *écart-type* | (sigma) |

Si l'échantillon est suffisamment grand, on peut prendre la moyenne $\bar{x}$ et la variance $s_n^2$ de l'échantillon comme estimateurs de la moyenne $\mu$ et de la variance $\sigma^2$ de la population.
Si on veut être plus précis, et en particulier pour des échantillons peu nombreux, un ***estimateur non biaisé*** (voir en bas) pour la variance $\sigma^2$ de la population est non pas $s_n^2 = \frac{1}{n}\sum_{i=1}^{k} f_i \cdot (x_i - \bar{x})^2$, mais plutôt

$$\boxed{s_{n-1}^2 = \frac{1}{n-1}\sum_{i=1}^{k} f_i \cdot (x_i - \bar{x})^2}$$ (on divise par n-1 au lieu de n).

On a donc la relation : $s_{n-1}^2 = \frac{n}{n-1} s_n^2$.

Entre la variance (estimée) de la population et la variance (calculée) de l'échantillon on a donc également la relation $\boxed{\sigma^2 = \frac{n}{n-1} s_n^2}$. De même on a pour les écarts-types : $\boxed{\sigma = \sqrt{\frac{n}{n-1}} s_n}$

Par contre, la moyenne $\bar{x}$ de l'échantillon est un estimateur non biaisé pour la moyenne $\mu$ de la population.

**Malheureusement , il y a des différences régionales dans les notations, et en particulier entre les notations du BI et celles de la TI 84 (voir en bas) !!!!!!!!!**

***Un estimateur non biaisé est un estimateur dont la valeur attendue à long terme est égale au paramètre à estimer. Par exemple, si nous pouvions continuer jusqu'à l'infini de prélever des échantillons et d'en calculer les moyennes $\bar{x}$, , nous nous rendrions compte que la moyenne des moyennes des échantillons serait finalement égale à la moyenne $\mu$ de la population.***

***Exercice 757***

Voici une statistique sur le nombre d'enfants par famille :

| Nombre d'enfants | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| Nombre de familles | 14 | 18 | 13 | 5 | 3 | 2 | 2 | 1 |

a) Calculez la moyenne et l'écart-type de l'échantillon.
b) Donnez des estimations non biaisées pour la moyenne et l'écart-type de la population dans laquelle on a choisi cet échantillon.

***Exercice 758***

Voici une statistique sur le nombre de cure-dents dans un lot de 48 boîtes :

| $x_i$ | 33 | 35 | 36 | 37 | 38 | 39 | 40 |
|---|---|---|---|---|---|---|---|
| $f_i$ | 1 | 5 | 7 | 13 | 12 | 8 | 2 |

a) Calculez la moyenne et l'écart-type de l'échantillon.
b) Donnez des estimations non biaisées pour la moyenne et l'écart-type de la population dans laquelle on a choisi cet échantillon.

***Exercice 759***

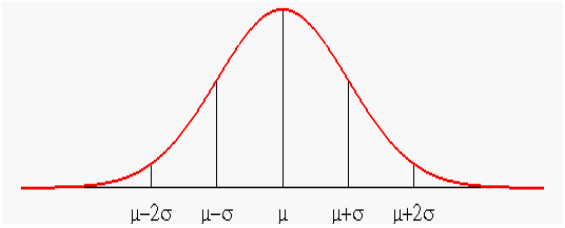
On a relevé les salaires hebdomadaires ( en dollars) de 200 ouvriers d'une aciérie :

a) Donnez des estimations pour le salaire moyen et l'écart-type des salaires.
b) Donnez des estimations non biaisées pour la moyenne et l'écart-type de la population dans laquelle on a choisi cet échantillon.

| Salaire (en $) | Nombre d'ouvriers |
|---|---|
| 360-369,99 | 17 |
| 370-379,99 | 38 |
| 380-389,99 | 47 |
| 390-399,99 | 57 |
| 400-409,99 | 18 |
| 410-419,99 | 10 |
| 420-429,99 | 10 |
| 430-439,99 | 3 |

## La signification de l'écart-type

Pour des populations de tailles suffisamment élevées, les distributions des séries statistiques se rapprochent souvent d'une distribution normale (voir plus tard) et prennent l'allure d'une courbe en cloche.




On peut montrer que pour une population normalement distribuée

- Environ 68 % de la population ont une mesure qui s'écarte de la moyenne d'au plus un écart-type vers le haut ou vers le bas , c.à.d. 68 % des mesures sont dans l'intervalle $[\mu-\sigma,\mu+\sigma]$.
- Environ 95 % des mesures sont dans l'intervalle $[\mu-2\sigma,\mu+2\sigma]$.
- Environ 99,7 % des mesures sont dans l'intervalle $[\mu-3\sigma,\mu+3\sigma]$.

***Exercice 760***

La taille moyenne des joueurs d'un championnat de basket est 184 cm. Si l'écart-type de la distribution des tailles est 5 cm, estimez le pourcentage des joueurs qui ont une taille

a. supérieure à 189 cm
b. supérieure à 179 cm
c. entre 174 et 199 cm
d. supérieure à 199 cm ?

***Exercice 761***

Les poids des nouveaux-nés de la maternité Prince Louis avaient l'année passée une moyenne de 3,0 kg avec un écart-type de 200 grammes. On enregistrait 545 naissances durant l'année. Estimez le nombre de nouveaux-nés qui avaient un poids

a. inféreur à 3,2 kg
b. entre 2,8 et 3,4 kg.

## Quelques exercices d'examens pour terminer

***Exercice 762*** (EM Mai 2002)

Dans les données ordonnées qui suivent, la moyenne est 6 et la médiane est 5 : 2,b,3,a,6,9,10,12

Déterminer : a) la valeur de a
b) la valeur de b.

***Exercice 763*** (EM Mai 2004)

Le tableau des effectifs cumulés ci-dessous donne les âges de 200 étudiants d'une faculté :

| Âge | Nombre d'étudiants | Effectifs cumulés |
|---|---|---|
| 17 | 3 | 3 |
| 18 | 72 | 75 |
| 19 | 62 | 137 |
| 20 | 31 | m |
| 21 | 12 | 180 |
| 22 | 9 | 189 |
| 23-25 | 5 | 194 |
| >25 | 6 | n |

1. Quelles sont les valeurs de m et n ?
2. Combien d'étudiants ont moins de 20 ans ?
3. Trouvez la valeur en années du premier quartile ?

***Exercice 764*** (NS Nov 2002)

Considérez les six nombres 2,3,6,9, a et b. La moyenne de ces nombres est 6 et la variance est 10. Trouvez a et b, avec $a < b$.

# V Statistiques

***Exercice 765*** (NM Mai 2002)

De janvier à septembre, le nombre moyen d'accidents de voiture par mois a été de 630. D'octobre à décembre, la moyenne a été de 810 accidents par mois.
Quel a été le nombre moyen d'accidents de voiture par mois pour l'année entière ?

***Exercice 766*** (NM Nov 2002)

Trois entiers positifs a, b et c, avec a < b < c, sont tels que leur médiane est 11, leur moyenne est 9 et leur étendue est 10. Trouvez la valeur de a.

***Exercice 767*** (EM, épreuve 2 2004)

**Le tableau ci-dessous présente les poids ($w$) et les nombres de poissons livrés un matin à un marché de quartier.**

| poids (kg) | effectifs | effectifs cumulés |
|---|---|---|
| $0{,}50 \le w < 0{,}70$ | 16 | 16 |
| $0{,}70 \le w < 0{,}90$ | 37 | 53 |
| $0{,}90 \le w < 1{,}10$ | 44 | $c$ |
| $1{,}10 \le w < 1{,}30$ | 23 | 120 |
| $1{,}30 \le w < 1{,}50$ | 10 | 130 |

**(a) (i)** Donnez la valeur de $c$. *[1 point]*

**(ii)** Sur du papier millimétré, tracez la *courbe des effectifs cumulés* pour ces données. Utilisez une échelle de 1 cm pour représenter 0,1 kg sur l'axe horizontal et 1 cm pour représenter 10 unités sur l'axe vertical. Légendez les axes clairement. *[4 points]*

**(iii)** Utilisez votre courbe pour montrer que le poids médian des poissons est 0,95 kg. *[1 point]*

**(b) (i)** Le zoo achète tous les poissons dont le poids est au-dessus du $90^{ème}$ centile. Combien de poissons le zoo achète-t-il ? *[2 points]*

**(ii)** Une entreprise de nourriture pour animaux domestiques achète tous les poissons du quartile inférieur. Quel est le poids maximum d'un poisson acheté par cette entreprise ? *[3 points]*

**(c)** Un restaurant achète tous les poissons qui sont à moins de 10 % du poids médian.

**(i)** Calculez les poids minimum et maximum des poissons achetés par le restaurant. *[2 points]*

**(ii)** Utilisez votre courbe pour déterminer combien de poissons vont être achetés par le restaurant. *[3 points]*