# 2. Statistique descriptive

*« Citoyens ! Cessez de croire yeux fermés les statisticiens ! Apprenez à jauger … »*

*« Les corrélations qui vous montrent que plus il y a de médecins plus on meurt jeune ! »*

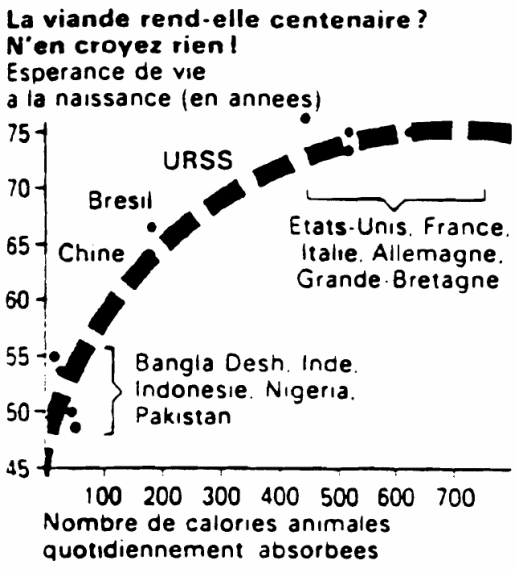

## § 2.1 Quelques erreurs courantes en statistique

Des corrélations qui s'entendent comme chèvre et chou

Si l'on écrit $A = 2 \cdot B$ on a une relation fonctionnelle. La connaissance de B détermine celle de A et réciproquement. Dans la pratique la liaison entre deux grandeurs est souvent moins étroite, quand B augmente A augmente également mais d'une façon moins régulière et avec des exceptions.

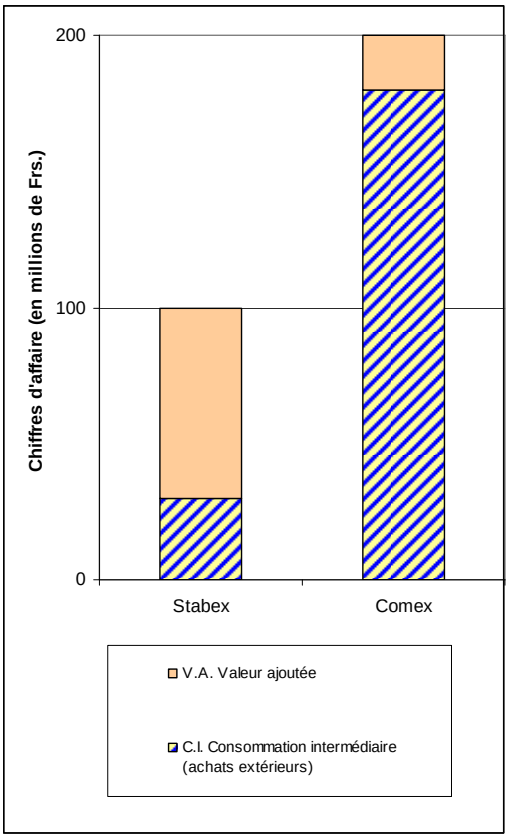
Ainsi il existe une corrélation entre la consommation d'alcool des automobilistes et la proportion d'accidents qu'ils provoquent. Dans ce cas, il s'agit bien d'une relation de cause à effet et l'on peut discerner la cause sans hésitation ; ce ne sont pas les accidents qui incitent les automobilistes à boire !

Mais les choses ne sont pas toujours aussi évidentes. Et parfois les deux grandeurs n'ont aucun rapport entre elles mais dépendent d'un troisième facteur. II en est souvent ainsi des séries chronologiques (l'évolution de phénomènes dans le temps). L'exemple de la corrélation entre l'espérance de vie et la consommation de viande est typique à cet égard. La représentation graphique de cette corrélation montre très bien que l'espérance de vie croit en même temps que la consommation de calories animales. Mais il n y a cependant pas de cause à effet entre les deux phénomènes contrairement à ce que fait croire la juxtaposition de ces chiffres pour la simple et bonne raison que le fait de ne manger que de la viande par exemple n'accroit pas 1'espérance de vie mais la réduit. Dans le cas présent l'espérance de vie et la consommation de viande évoluent de pair à cause d'un troisième facteur qui est le niveau de développement économique des différents pays.



Des classements d'entreprises ineptes

Une entreprise achète à l'extérieur des matières premières, énergie, produits semi-finis, etc. et les transforme en produits finis. Soit l'entreprise *Stabex* qui achète 30 millions de francs de produits à l'extérieur pour une production (chiffre d'affaires ou montant des ventes) égale à 100 millions. Sa contribution à la production nationale ou **« valeur ajoutée »** est égale à 100 – 30 = 70 millions de francs. Soit une autre entreprise, appelons la *Comex*, dont les achats s'élèvent à 180 millions pour un chiffre d'affaires égale à 200 millions (cela peut arriver si elle transforme peu les matières premières achetées). Sa valeur ajoutée est égale à 20 millions de francs.




Le classement d'après la valeur ajoutée, le seul rationnel pour apprécier l'importance économique des entreprises, donne les résultats suivants :
1. Stabex 70
2. Comex 20

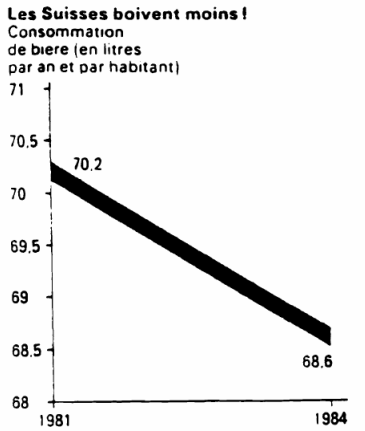
Mais si l'on considère le chiffre d'affaires on aboutit au classement suivant :
1. Comex 200
2. Stabex 100

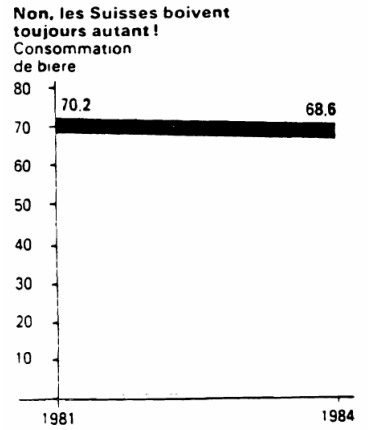
Conséquence avec un prix du pétrole élevé les sociétés de raffinage ont des chiffres d'affaires flatteurs. Que le prix du pétrole quadruple soudain et elles s'envoleront dans le classement des entreprises sans que leur contribution à l'économie nationale augmente pour autant.

Alors pourquoi classe-t-on les entreprises d'après leur chiffre d'affaires et non d'après leur valeur ajoutée ? Tout simplement parce que cette dernière est plus difficile à connaître. C'est pour la même raison que l'on classe les exploitations agricoles d'après leur surface alors qu'un hectare de pacage maigre n'est nullement comparable à un hectare de cultures florales sous serre !

Des graphiques vrais qui mentent comme des arracheurs de dents

Rien de plus facile que de mentir avec des chiffres vrais. II suffit de construire des graphiques dont l'ordonnée ne part pas de zéro pour accentuer une tendance « trop légère. » Ainsi, la légère baisse de la consommation de bière par habitant en Suisse entre 1981 et 1984 peut être interprétée de manière fort différente selon l'impression qu'on recherche.



Des taux de natalité et de mortalité qui se mordent la queue

Grâce à des taux de natalité élevés dans le passé la population des Pays-Bas est relativement jeune, plus que celle de la France. Elle compte de ce fait une assez bonne proportion de femmes en âge de procréer et une proportion relativement faible de personnes âgées. Il en résulte que le taux de natalité (12 ‰) dépasse le taux de mortalité (8 ‰) alors que le nombre d'enfants par femme, de l'ordre de 1,5 est très inferieur à celui qui serait nécessaire (2,1) pour assurer le renouvellement de la population dans le long terme. Les Pays Bas sont donc déjà en déclin démographique mais celui-ci est caché par des taux de natalité et de mortalité trompeurs, même s'ils sont « statistiquement vrais ».

*« Il y a trois sortes de mensonges : les mensonges simples, les mensonges affreux et les statistiques! » (Benjamin Disraeli[1])*

*« Je ne crois aux statistiques que lorsque je les ai moi-même falsifiées.» (Winston Churchill[2])*

*« Les statistiques, c'est comme le bikini. Ce qu'elles révèlent est suggestif. Ce qu'elles dissimulent est essentiel. » (Aaron Levenstein[3])*

## § 2.2 Introduction

Statistiques, enquêtes, sondages, moyennes, indices... sont diffusés à longueur de colonnes dans les journaux écrits et télévisés. Ces travaux sont souvent, mais pas toujours, scientifiquement rigoureux. Les médias s'en font l'écho sous des formes très discutables : les illustrations graphiques relèvent parfois de la pure fantaisie. L'usage de la statistique devient abusif. Le grand public reste perplexe et en conclut: "on fait dire ce que l'on veut aux chiffres". Un peu de culture mathématique permet d'éviter les pièges dûs à une interprétation hâtive de chiffres.

La **statistique** est l'activité qui consiste à recueillir, traiter et interpréter un ensemble de données d'informations. Le traitement des données consiste à produire des **statistiques**. Parmi les différentes branches que regroupe cette activité, il paraît nécessaire d'en distinguer trois principales :

- La **collecte** des données.
- Le **traitement des données** collectées est aussi appelé la **statistique descriptive.**
- L'**interprétation des données**, aussi appelée l'**inférence statistique**, s'appuie sur la théorie des sondages et la **statistique mathématique**.

Cette distinction ne consiste pas à définir plusieurs domaines étanches. En effet, le traitement et l'interprétation des données ne peuvent se faire que lorsque celles-ci ont été récoltées. Réciproquement, la statistique mathématique précise les règles et les méthodes sur la collecte des données, pour que celles-ci puissent être correctement interprétées.

Statistique descriptive

Le but de la statistique est d'extraire des informations pertinentes d'une liste de nombres difficile à interpréter par une simple lecture. Nous allons traiter le sujet de la **statistique descriptive**. **Descriptive** signifie que l'on part de données existantes. **La statistique descriptive** a pour but de présenter, en un petit nombre de résultats, des données concernant une population trop nombreuse pour que la liste explicite de ses caractéristiques soit compréhensible: à quoi servirait en effet une liste alphabétique de tous les Suisses avec leur âge, leur profession et leur salaire ? La statistique descriptive proprement dite se propose de déterminer les caractéristiques de ces grandeurs (âge, profession ou salaire), sans passer par la liste exhaustive, en se basant sur des procédés de généralisations fiables.

[1] Premier ministre britannique 1804-1881
[2] Premier ministre britannique 1874 -1965
[3] Economiste américain 1901 - 1986

L'utilité de ces caractéristiques générales réside d'une part dans une **mise en évidence d'une certaine réalité**, et d'autre part dans la détermination des données nécessaires aux calculs des probabilités, par exemple pour l'établissement de prévisions, de conditions d'assurances ou pour la mise au point d'un système expert (critères de décision).

Description intrinsèque d'une distribution d'observations
Sans aucun apriori sur la question que l'on se pose, quelques statistiques simples permettent de la décrire:

- la moyenne $\overline{x}$
- la médiane $\tilde{x}$
- le mode
- la variance $v$
- l'écart-type $\sigma$
- intervalle semi-quartile $I$
- le maximum et le minimum

Les trois premiers (à gauche) sont souvent nommé « **critères de position** », et les autres entrent plutôt dans la catégorie des « **critères de dispersion.** »

## § 2.3 Collecte de l'information, dépouillement de l'information et vocabulaire

La collecte de l'information peut être :

- **directe**: sondages
- **indirecte**: on utilise des données existantes (bilans, ...)

Cette collecte doit être objective avec suffisamment de données mais sans excès pour rester utilisable.

L'ensemble examiné est appelé la **population** (pas nécessairement des gens). Par exemple : des personnes, des objets, des lieux, des moments, … **L'individu** est un l'élément de la population. On s'intéresse alors à son **caractère** qui est la propriété subjective qui nous intéresse (âge, sexe, nombre d'enfants, taille, …)

On distingue différents types de **caractère** :

- qualitatif (par exemple catégorie professionnelle, couleur d'un objet, parti politique)
- quantitatif discret[4] (par exemple nombre d'enfants ou note trimestrielle)
- quantitatif continu[5] (par exemple longueurs, vitesses des particules d'un gaz, ...)

Dans le cas quantitatif, on parle aussi de **variable statistique**, plutôt que de caractère.

On reporte ensuite les caractères suivant une **partition** (partage d'un ensemble en sous-ensembles disjoints et exhaustifs). II est bien sûr absolument indispensable que le caractère de chaque individu observé puisse être reporté de façon claire et univoque dans un et un seul sous-ensemble de la partition. Exemples de partitions: sexe, âge (en années), tranches de revenu etc.

La ***population*** est donc l'ensemble des ***individus*** sur lesquels porte une étude. Ses individus sont classés selon un ou plusieurs ***caractères*** suivant des partitions qui sont des **classes**.

Exemple 1 : La ***population*** est l'ensemble des vélos produits par CIPEDVELO en 2001.
L'étude porte sur le type de vélos (VTT, ville, course, ....).
Le ***caractère*** est le type de vélos et les ***classes*** sont : VTT, course, ville, ...
Un ***individu*** est un vélo.

[4] variable numérique
[5] il peut prendre toute valeur dans un certain intervalle réel.

Exemple 2 :
Lors d'une course de vitesse les 40 participants ont mis les temps suivants pour effectuer le parcours

| **Temps (Classes)** | **Centre des classes** $x_i$ | **Effectifs de chaque classe** $n_i$ |
|---|---|---|
| [43-45[ | 44 | 2 |
| [45-47[ | 46 | 3 |
| [47-49[ | 48 | 7 |
| [49-51[ | 50 | 11 |
| [51-53[ | 52 | 8 |
| [53-55[ | 54 | 6 |
| [55-57[ | 56 | 3 |

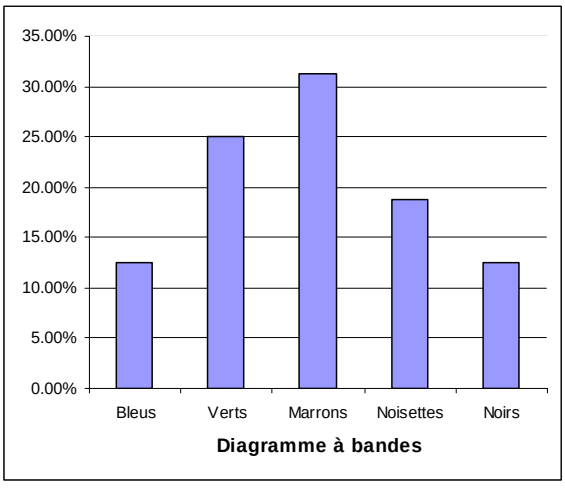
La **population** est : les 40 participants
Un **individu** est : un coureur
Le **caractère quantitatif continu** est : le temps *(variable statistique continue)*
Les **classes** sont : les intervalles des temps

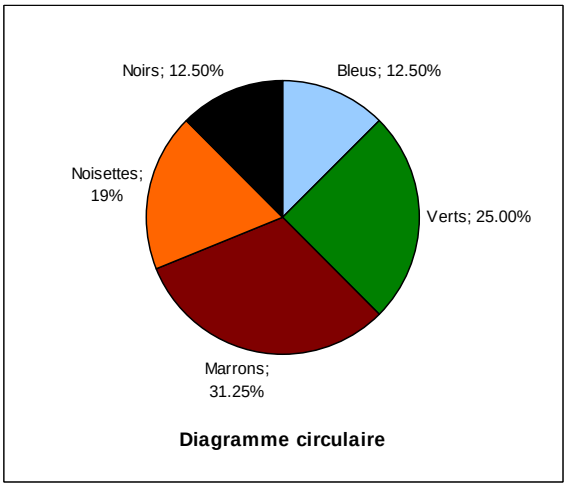
Les **variables statistiques continues** sont les différents temps pris dans les différents intervalles. On considère ces variable sous la **forme discrète en prenant le centre des classes,** cela nous permettra de simplifier l'étude : $x_1 = 44$, $x_2 = 46$, ...
Et dans chaque cas on a l'**effectif de chaque classe** : $n_1 = 2$, $n_2 = 3$, ...

*On passe du cas continu à un cas discret en utilisant les centres des classes*
*Le centre de classe est égal à la moyenne des extrémités de la classe.*

## § 2.4 Visualisation des données, effectifs et fréquences

Il est possible de passer directement à la partie calculatoire, mais il est néanmoins appréciable de visualiser ces données. La représentation graphique fait également partie de la statistique descriptive. Les données brutes sont préalablement regroupées[6] et mises sous forme de tableau triées en **classes exhaustives** (partition), d'amplitudes à choix, dont on répertorie **l'effectif** ou la **fréquence**. Les deux types de graphiques les plus courants sont :




[6] Le fait qu'il y a un choix implique nécessairement une certaine subjectivité.

Définition :

La ***fréquence*** (relative) d'une classe est définie par : $f_{classe} = \frac{Effectif\_de\_la\_classe}{Effectif\_total}$

exprimée généralement en %.

Exercice 1 :
18 élèves ont passé un examen. Voici, en vrac, leurs résultats :

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 5 | 3 | 5 | 5 | 4 | 4 | 4 | 5 | 4 |
| 2 | 3 | 3 | 4 | 4 | 6 | 5 | 6 | 4 |

Ranger ces résultats dans un tableau et les illustrer à l'aide d'un diagramme à bande.

**Activité 1 : atelier informatique.**

Remarque :
Dans le cas où le caractère est **quantitatif**, on parle alors de variable statistique discrète ou continue, il est alors possible de caractériser l'information. L'idée est de quantifier des impressions comme grand, plus, large, trouver des "thermomètres" numériques qui correspondent aux propriétés qualitatives structurelles d'un ensemble de données statistiques.
On distingue deux types principaux : les **mesures de centrage** et les **mesures de dispersion**

## § 2.5 Mesure de tendance centrale (critères de position) : moyenne, médiane, mode, …

Les graphiques donnent une bonne idée de la manière dont un caractère est distribué, mais on cherche souvent à illustrer cette distribution de manière beaucoup plus succincte par quelques nombres caractéristiques. Parmi ceux-ci les mesures de la tendance centrale jouent un rôle essentiel. La plus connue de ces mesures est la **moyenne**. Mais on utilise d'autres mesures encore comme : la **médiane**, le **mode**,…

**Moyenne arithmétique** : (événements isolés)
C'est la somme des données divisée par le nombre de ces données:

$$\overline{x} = \frac{x_1 + x_2 + ... + x_n}{n} \qquad \overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Exercice 2 :
Si les nombres suivants représentent des âges : 12 13 30 23 28 15 30
Calculer l'âge moyen de ces personnes.

$\overline{x} =$ ans

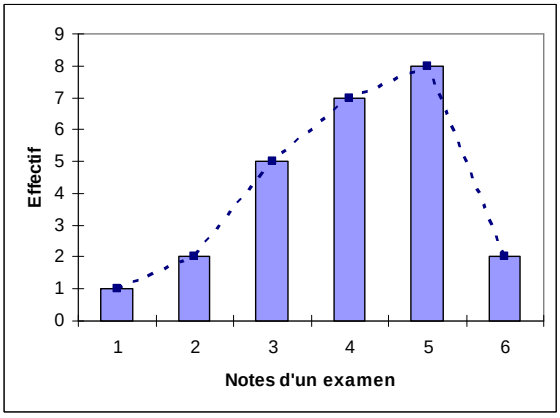
**Moyenne pondérée :** (événements regroupés ou pondérés)
Si on a $n_1$ fois la donnée $x_1$, $n_2$ fois la donnée $x_2$, ... avec respectivement les fréquences $f_1$, $f_2$,... et $n = n_1 + n_2 + ... + n_k$, alors la moyenne est donnée par :

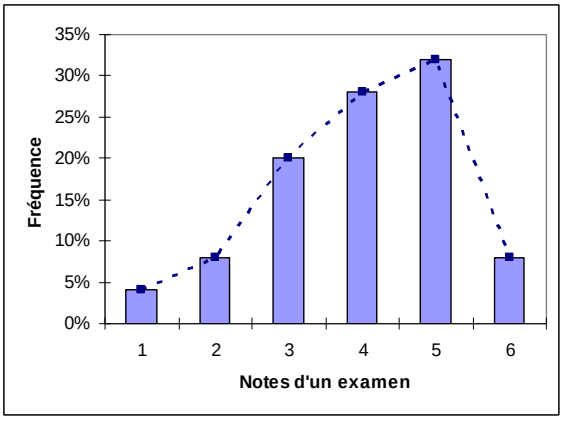
$$\overline{x} = \frac{n_1 x_1 + n_2 x_2 + ... + n_k x_k}{n} \qquad \overline{x} = \frac{\sum_{i=1}^{k} n_i x_i}{n}$$

Exemple 1 :
25 élèves ont passé un examen et on a relevé :

| **Note** $x_i$ | **Effectif** $n_i$ | **Note pondérée** $n_i x_i$ | **Fréquence** $f_i$ |
|---|---|---|---|
| 1 | 1 | 1 | 4 % |
| 2 | 2 | 4 | 8 % |
| 3 | 5 | 15 | 20 % |
| 4 | 7 | 28 | 28 % |
| 5 | 8 | 40 | 32 % |
| 6 | 2 | 12 | 8 % |
| *Sommes :* | $n = 25$ | $\sum_{i=1}^{k} n_i x_i = 100$ | *100 %* |

On a : $\overline{x} = \frac{1}{n} \cdot \sum_{i=1}^{k} n_i x_i = \frac{100}{25} = 4$



Remarque importante :
La moyenne est la plus familière et la plus utilisée des mesures de tendance centrale. Elle est influencée par toutes les valeurs de $x_i$ et $n_i$ observées et à ce titre **malheureusement très sensible aux valeurs extrêmes**, au point d'en perdre parfois une bonne partie de sa représentativité, surtout dans les échantillons de petite taille.

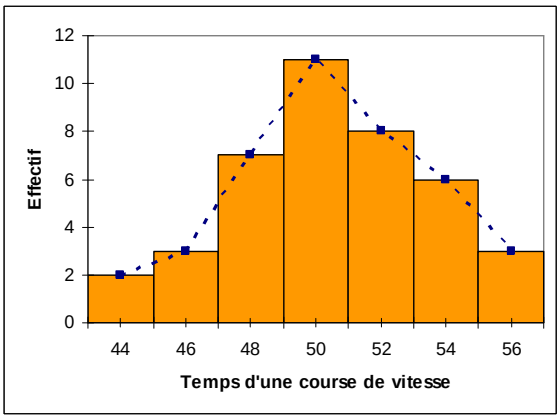
Exemple 2 : Voici six salaires mensuels :
3'500.- 4'200.- 4'600.- 5'000.- 6'200.- 36'500.-

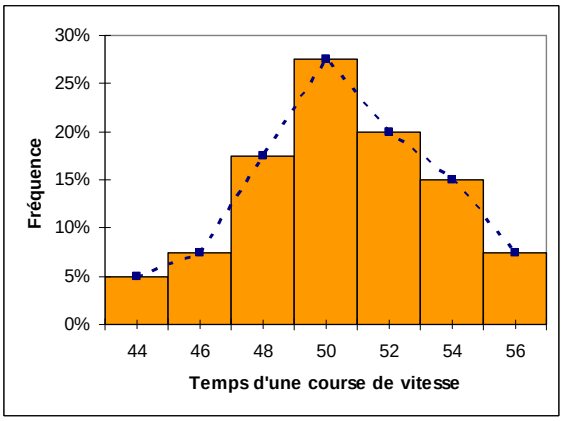
La moyenne est : $\bar{x} = 10'000.-$ Frs. !!!

Exemple 3 :
Lors d'une course de vitesse les 40 participants ont mis les temps suivants pour effectuer le parcours

| **Temps (secondes)** | **Centre** $x_i$ | **Effectif** $n_i$ | **Temps pondéré** $n_i x_i$ | **Fréquence** $f_i$ |
|---|---|---|---|---|
| [43-45[ | 44 | 2 | 88 | 5 % |
| [45-47[ | 46 | 3 | 138 | 7,5 % |
| [47-49[ | 48 | 7 | 336 | 17,5 % |
| [49-51[ | 50 | 11 | 550 | 27,5 % |
| [51-53[ | 52 | 8 | 416 | 20 % |
| [53-55[ | 54 | 6 | 324 | 15 % |
| [55-57[ | 56 | 3 | 168 | 7,5 % |
| | *Sommes :* | *n = 40* | $\sum_{i=1}^{k} n_i x_i = 2020$ | *100 %* |

On a : $\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{k} n_i x_i = \frac{2020}{40} = 50,5$




Exercice 3 :
Calculez la moyenne des notes de l'exercice 1

Exercice 4 :
On a classé un groupe de personnes selon la pointure de leurs pieds. Voici les résultats en vrac :
40, 36, 36, , 39, 39, 39, 39, 41, 41, 39, 39, 36, 36, 40, 40, 40, 32, 32, 32, 33, 37, 37, 37, 37, 37, 33, 33, , 38, 38, 38, 38, 34, 34, 34, 35, 40, 40, 40, 40, 35, 33, 33, 34, 34, 34, 34, 35, 35,35, 35, 35, 35, 35, 35, 36, 36, 36, 36, 36, 37, 37, 39, 39, 39, 39, 39, 39, 37, 38, 38, 38, 38, 38, 38, 38, 38, 37, 37, 40, 40, 40, 40, 37, 37, 37, 37, 38, 38, 38, 38

a) Calculez la fréquence de chaque pointure
b) Calculez la moyenne des pointures
c) Représentez ces résultats sur un diagramme à bandes

Exercice 5 :
Lors d'un festival de cinéma, les films ont été classés selon leur durée.
Voici les résultats obtenus :

| Durée (min) | effectif |
|---|---|
| [60;70[ | 3 |
| [70;80[ | 6 |
| [80;90[ | 9 |
| [90;100[ | 10 |
| [100;110[ | 7 |
| [110:120[ | 2 |

a) Quelle est la population étudiée ?
b) Qu'est-ce qu'un individu ?
c) Quel caractère étudie-t-on ?
d) Réalisez une étude complète :
Fréquences
Moyennes
Diagramme

*Un autre nombre important associés à un caractère est :*

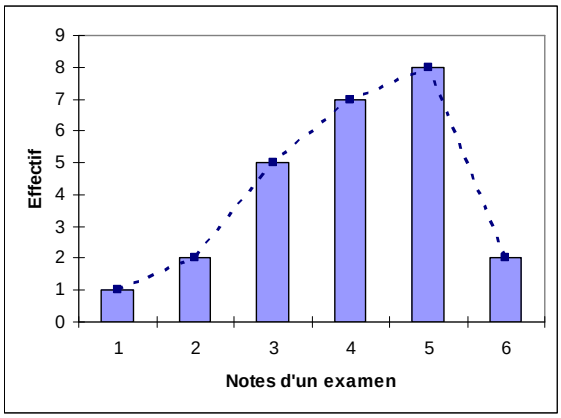
**Le mode**
Le ***mode*** est la classe ou les classes qui ont la plus grande fréquence (l'effectif plus grand).
C'est la valeur la plus fréquente dans une série de données.

Remarques :
1. Dans certaines distributions il y plusieurs modes (multimodales).
2. Le mode est insensible aux valeurs extrêmes
3. Il est moins utilisé que la moyenne ou la médiane.

Exemple 1 :
25 élèves ont passé un examen et on a relevé :

| **Note** $x_i$ | **Effectif** $n_i$ | **Fréquence** $f_i$ |
|---|---|---|
| 1 | 1 | 4 % |
| 2 | 2 | 8 % |
| 3 | 5 | 20 % |
| 4 | 7 | 28 % |
| 5 | 8 | 32 % |
| 6 | 2 | 8 % |

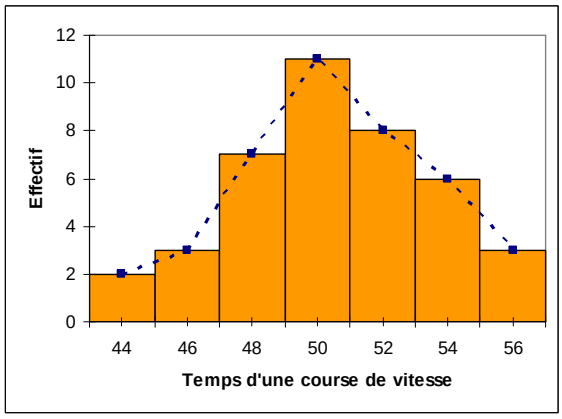
Le mode vaut 5. La note 5 est la plus fréquente




Exemple 2 :
Lors d'une course de vitesse les 40 participants ont mis les temps suivants pour effectuer le parcours

| **Temps (secondes)** | **Effectif** $n_i$ | **Fréquence** $f_i$ |
|---|---|---|
| [43-45[ | 2 | 5 % |
| [45-48[ | 3 | 7,5 % |
| [48-49[ | 7 | 17,5 % |
| [49-51[ | 11 | 27,5 % |
| [51-53[ | 8 | 20 % |
| [53-55[ | 6 | 15 % |
| [55-58[ | 3 | 7,5 % |

Le mode[7] vaut [49-51[.
Le temps le plus fréquent est dans la classe [49-51[.




[7] On donne parfois la valeur du centre de la classe ici le mode vaudrait alors 50.

*Si l'inconvénient majeur de la moyenne est sa sensibilité aux valeurs extrêmes, il existe un autre indicateur important, cette fois insensible aux valeurs extrêmes, c'est :*

**La médiane**
La médiane $\tilde{x}$ est une valeur telle que la moitié des valeurs $x_i$ de l'échantillon lui sont inférieures ou égales et l'autre moitié supérieures ou égales.

La ***médiane*** est la valeur qui sépare la population en deux groupes égaux. C'est à dire que 50% de la population est au dessous de la médiane et l'autre 50% est au dessus.

Exemple : Voici six salaires mensuels :
3'500.- 4'200.- 4'600.- 5'000.- 6'200.- 36'500.-

La moyenne est : $\bar{x} = 10'000.- \text{Frs.}$

Alors que la médiane est : $\tilde{x} = \frac{4600+5000}{2} = 4'800$

**Remarque:**
**La médiane n'est pas affectée pas les valeurs extrêmes de la distribution.**

***Dans les distributions asymétriques, la médiane donne une idée plus « équilibrée » de la tendance centrale que la moyenne.***

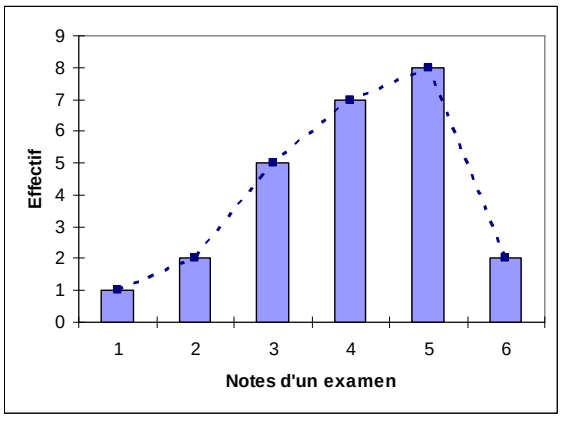
**a) La médiane pour un échantillon discret**
Dans le cas d'un échantillon discret de *n* valeurs de $x_i$ rangées dans l'ordre croissant :
$x_1 \le x_2 \le x_3 \le .... \le x_n$ la **médiane** $\tilde{x}$ est la valeur du milieu.

Si *n* impair alors $\tilde{x} = x_{\frac{n+1}{2}}$ si n pair alors $\tilde{x} = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$ (moyenne des deux valeurs centrales)

Exemple :
25 élèves ont passé un examen et on a relevé :

| **Note** $x_i$ | **Effectif** $n_i$ | **Fréquence** $f_i$ |
|---|---|---|
| 1 | 1 | 4 % |
| 2 | 2 | 8 % |
| 3 | 5 | 20 % |
| 4 | 7 | 28 % |
| 5 | 8 | 32 % |
| 6 | 2 | 8 % |

1,2,2,3,3,3,3,3,3,4,4,4,**4**,4,4,4,5,5,5,5,5,5,5,5,5,6,6




La médiane vaut 4 car c'est la 13ème note sur les 25. $\tilde{x} = 4$

Exercice 6 : Pour chaque série ci-dessous, déterminer la médiane.

a) 1,1,1,2,2,2,3,3,4,4,5,5,6,6,6,6,8,100,110

b) 1,1,1,2,2,2,3,3,4,4,5,5,6,6,6,6,8,100,110,200

c) 1,1,1,2,2,2,3,3,4,4,5,5,6,6,6,6,8,100,110,50000

d) 1,1,1,2,2,2,3,3,4,4,5,5,6,6,6,6,8,100,11000,500000

Exercice 7 : Pour chaque série (identique à l'exercice 6), calculer la moyenne.

a) 1,1,1,2,2,2,3,3,4,4,5,5,6,6,6,6,8,100,110

b) 1,1,1,2,2,2,3,3,4,4,5,5,6,6,6,6,8,100,110,200

c) 1,1,1,2,2,2,3,3,4,4,5,5,6,6,6,6,8,100,110,50000

d) 1,1,1,2,2,2,3,3,4,4,5,5,6,6,6,6,8,100,11000,500000

Que remarque-t-on ?

Exercice 8 : Déterminer la médiane de chaque statistique.

| Note | Effectif | Effectif cumulé |
|---|---|---|
| 1 | 1 | |
| 2 | 2 | |
| 3 | 3 | |
| 4 | 6 | |
| 5 | 4 | |
| 6 | 1 | |
| | 17 | |

| Note | Effectif | Effectif cumulé |
|---|---|---|
| 1 | 1 | |
| 2 | 2 | |
| 3 | 6 | |
| 4 | 3 | |
| 5 | 4 | |
| 6 | 1 | |
| | 17 | |

| Note | Effectif | Effectif cumulé |
|---|---|---|
| 1 | 1 | |
| 2 | 2 | |
| 3 | 6 | |
| 4 | 3 | |
| 5 | 4 | |
| 6 | 2 | |
| | 18 | |

Exercice 9: (cas avec une valeur extrême)

L'entreprise de Giles Baytes est composée de 17 personnes.
Ci-dessous, la paye mensuelle de chaque employé (en U)

| | | | | |
|---|---|---|---|---|
| 2000 U | 3000 U | 2000 U | 2000 U | 2000 U |
| 2000 U | 1000 U | 1000 U | 100000 U | |
| 1000 U | 3000 U | 1000 U | 3000 U | |
| 2000 U | 2000 U | 1000 U | 3000 U | |

a) classer ces résultats dans l'ordre croissant
b) calculer leur moyenne
c) calculer leur médiane
d) représenter ces résultats sur un diagramme à bande
e) indiquer la moyenne et la médiane

Exercice 10 : (cas sans valeur extrême)

L'entreprise *Equity* est composée de 17 personnes.
Ci-dessous, la paye mensuelle de chaque employé (en U)

| | | | | |
|---|---|---|---|---|
| 2000 U | 1000 U | 1000 U | 3000 U | 3000 U |
| 1000 U | 2000 U | 2000 U | 2000 U | |
| 1000 U | 1000 U | 3000 U | 2000 U | |
| 3000 U | 1000 U | 1000 U | 3000 U | |

a) classer ces résultats dans l'ordre croissant
b) calculer leur moyenne
c) calculer leur médiane
d) représenter ces résultats sur un diagramme à bande
e) indiquer la moyenne et la médiane

**b) La médiane pour un échantillon continu**
Dans le cas d'un échantillon continu pour estimer la ***médiane*** il faut passer par la ***fonction cumul*** (cumul des %). La médiane sera alors la valeur de $x$ pour laquelle la fonction vaut 50%

Les fréquences relatives sont : $f_i = \frac{n_i}{n} \qquad i = 1, 2, ...., k$

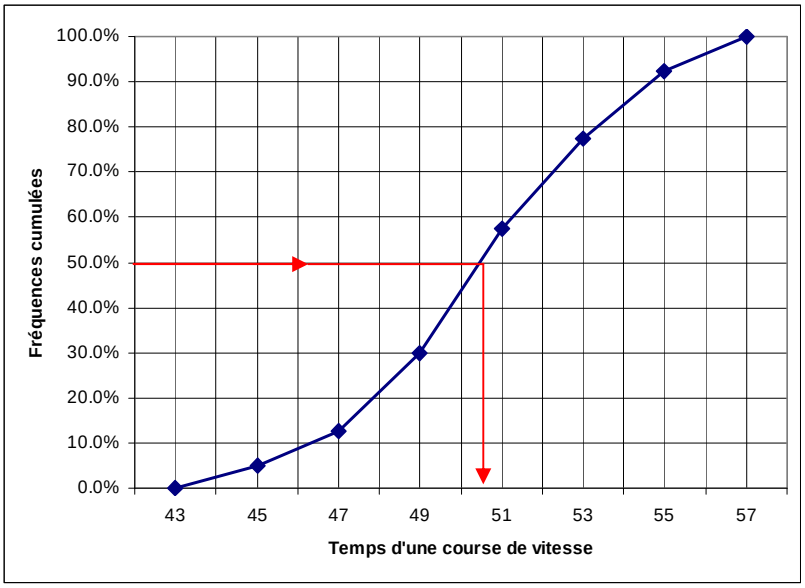
**Fréquences cumulées (fonction cumul) :**
Pour connaître la proportion $F(x)$ (dite fréquence cumulée) des individus qui présentent des valeurs $x_i$ du caractère inférieur ou égale à $x$, on additionne toues les fréquences $f_i$ qui correspondent aux $x_i$ tels que $x_i < x$.

On détermine par le graphique de $F(x)$ facilement la médiane car $F(\tilde{x}) = 50\,\%$

Exemple :
Lors d'une course de vitesse les 40 participants ont mis les temps suivants pour effectuer le parcours

| **Temps (secondes)** | **Centre** $x_i$ | **Effectif** $n_i$ | **Fréquence** $f_i$ | **Fréquence cumulée** $f_i x_i$ |
|---|---|---|---|---|
| [43-45[ | 44 | 2 | 5,0 % | 5,0 % |
| [45-47[ | 46 | 3 | 7,5 % | 12,5 % |
| [47-49[ | 48 | 7 | 17,5 % | 30,0 % |
| [49-51[ | 50 | 11 | 27,5 % | 57,5 % |
| [51-53[ | 52 | 8 | 20,0 % | 77,5 % |
| [53-55[ | 54 | 6 | 15,0 % | 92,5 % |
| [55-57[ | 56 | 3 | 7,5 % | 100,0 % |

| Temps | Fréquences cumulées |
|---|---|
| 43 | 0 % |
| 45 | 5 % |
| 47 | 12.5 % |
| 49 | 30 % |
| 51 | 57.5 % |
| 53 | 77,5 % |
| 55 | 92,5 % |
| 57 | 100 % |




La médiane vaut :

$$\tilde{x} = 50,5$$

Exercice 11 :
L'étude de la taille des élèves d'une classe a donné les résultats suivants :

| Taille (cm) | Effectif |
|---|---|
| [ 140 ; 150 [ | 1 |
| [ 150 ; 160 [ | 5 |
| [ 160 ; 170 [ | 8 |
| [ 170 ; 180 [ | 6 |
| [ 180 ; 190 [ | 4 |

Faire l'étude complète de ce caractère : fréquences, moyenne, mode, histogramme et médiane

Exercice 12 :
L'étude de l'âge des habitants d'un immeuble a donné les résultats suivants :

| Âge | effectif |
|---|---|
| [ 0 ; 18 [ | 20 |
| [ 18 ; 36 [ | 36 |
| [ 36 ; 54 [ | 20 |
| [ 54 ; 82 [ | 15 |
| [ 82 ; 90 [ | 9 |

Faire l'étude complète de ce caractère : fréquences, moyenne, mode, histogramme et médiane

**Activité 2 : atelier informatique (avec Excel ou Calc) .**

Activité 2A :
Considérer des longueurs en cm de 40 boas d'un zoo.

| Classes | Effectif |
|---|---|
| [380;**390**[ | |
| [390;**400**[ | |
| [400;**410**[ | |
| [410;**420**[ | |
| [420;**430**[ | |
| [430;**440**[ | |
| [440;**450**[ | |
| | 40 |

Compléter librement ce tableau et faire une étude complète.
Varier ensuite les effectifs dans chaque classe et observer les changements.

Activité 2B :
Soit une série qui donne le nombre d'enfants de 20 femmes d'un échantillon de la population.
Nombre d'enfants = {4, 0, 1, 1, 2, 2, 2, 3, 3, 4, 2, 3, 4, 5, 2, 1, 3, 3, 4, 5}

Faire une étude complète.

Activité 2C :
Faire une étude complète des ménages Français en 1999 :

| **Nombre de personnes dans le ménage** | **Effectif de la population en 1999** |
|---|---|
| 1 | 8'089'434 |
| 2 | 8'086'664 |
| 3 | 3'619'655 |
| 4 | 3'058'684 |
| 5 | 1'182'235 |
| 6 ou plus | 408'959 |

N.B. On considère 6 ou plus, comme valant 6.

Activité 2D :
Faire une étude complète de la population française active par âge en 1999 :

| **Age** | **Effectif** |
|---|---|
| 15-24 | 2'289'542 |
| 25-29 | 3'628'502 |
| 30-34 | 3'881'554 |
| 35-39 | 3'865'252 |
| 40-44 | 3'880'300 |
| 45-49 | 3'696'642 |
| 50-54 | 3'305'288 |
| 55 et + | 2'225'411 |

## § 2.6 Mesures de dispersion (critères de dispersion) : variance, écart-type, quartile, …

Les mesures de tendance centrale vues au paragraphe précédent, aussi importantes qu'elles soient, ne sauraient donner une idée de la manière dont les valeurs sont distribuées au voisinage de ces valeurs centrales. Aussi est-il utile d'introduire une mesure pour rendre compte de cette dispersion.

Pour des questions théoriques, on introduit la moyenne du carré des écarts qui est :

**La variance statistique**

$$v = \frac{\sum n_i \cdot (\overline{x} - x_i)^2}{n}$$

*Grandeur théorique sans unité !*

Pour revenir à l'unité de l'échantillon initial, on définit finalement :

**L'écart-type :**

$$\sigma = \sqrt{v}$$

Géométriquement on peut caractériser de manière assez générale **l'écart-type comme étant le rayon** (autour de la moyenne) de la « cloche » de la distribution des résultats **englobant environ les 2/3 des données.**

Ainsi si on utilise la moyenne pour mesurer la tendance centrale, on lui associera tout naturellement l'écart-type pour mesurer la dispersion (par rapport à la moyenne).

**Distribution normale :**
On peut montrer que lorsque la population a une distribution normale alors :

- 68,3 % des valeurs sont situées entre $\overline{x} - \sigma$ et $\overline{x} + \sigma$
- 95,4 % des valeurs sont situées entre $\overline{x} - 2\sigma$ et $\overline{x} + 2\sigma$
- 99,8 % des valeurs sont situées entre $\overline{x} - 3\sigma$ et $\overline{x} + 3\sigma$

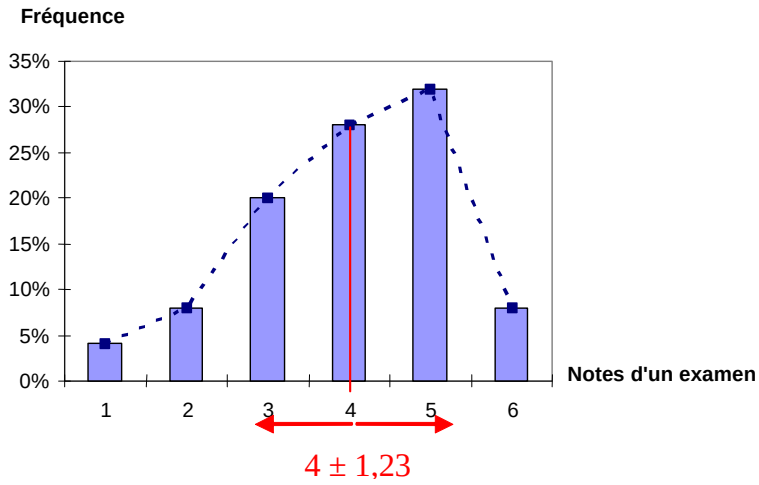
Exemple 1 :
25 élèves ont passé un examen et on a relevé :

| **Note** $x_i$ | **Effectif** $n_i$ | **Note pondérée** $n_i x_i$ | **Carré de l'écart** $(\overline{x} - x_i)^2$ | **Carré de l'écart pondéré** $n_i \cdot (\overline{x} - x_i)^2$ |
|---|---|---|---|---|
| 1 | 1 | 1 | 9 | 9 |
| 2 | 2 | 4 | 4 | 8 |
| 3 | 5 | 15 | 1 | 5 |
| 4 | 7 | 28 | 0 | 0 |
| 5 | 8 | 40 | 1 | 8 |
| 6 | 2 | 12 | 4 | 8 |
| | *n = 25* | *100* | | *38* |

$$\sigma = \sqrt{1,52} = 1,23$$

$$\overline{x} = \frac{100}{25} = 4$$

$$v = \frac{38}{25} = 1,52$$




*Si l'on suit une loi normale théoriquement 2/3 des notes sont comprises entre 2,88 et 5,23.*

On remarquera dans notre exemple qu'en réalité 20 notes sur 25 soit 80 % sont situées entre 2,88 et 5,23 et que 24 sur 25 soit 96 % sont situées entre 1,54 et 6,46.

Le plus utilisé dans la pratique pour calculer la variance est le :

**Théorème de König-Huyghens :**

On montre facilement que :

$$v = \frac{\sum n_i x_i^2}{n} - \overline{x}^2 \qquad \text{et} \qquad v = \sum f_i x_i^2 - \overline{x}^2$$

Il permet de simplifier grandement les calculs. Pour s'en convaincre, reprenons l'exemple précédent.

Exemple 1 :
25 élèves ont passé un examen et on a relevé :

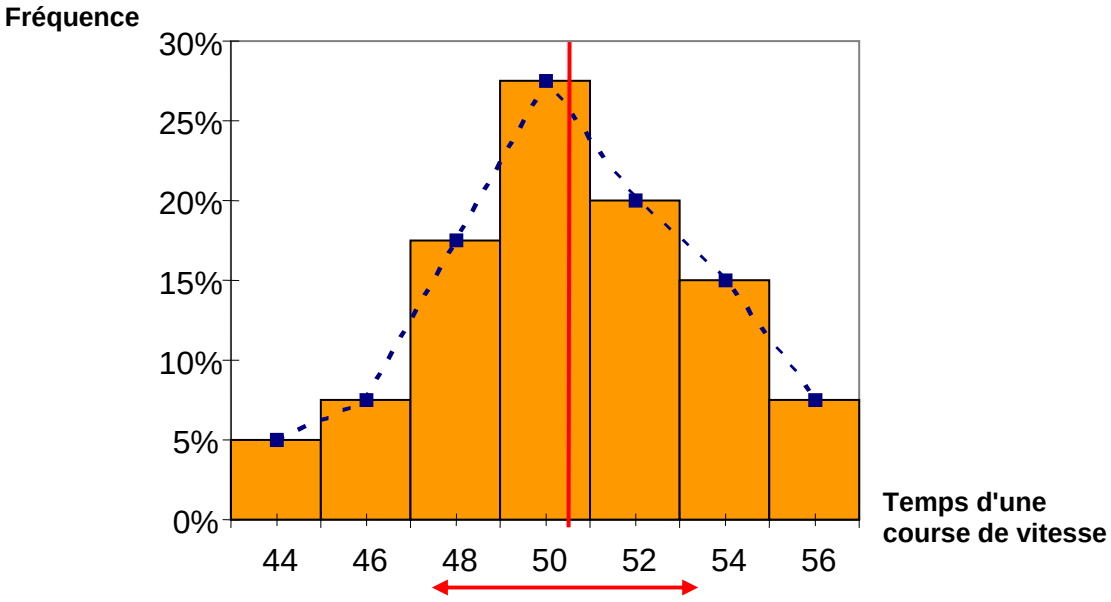
| **Note** $x_i$ | **Effectif** $n_i$ | **Note pondérée** $n_i x_i$ | $x_i^2$ | $n_i x_i^2$ |
|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 |
| 2 | 2 | 4 | 4 | 8 |
| 3 | 5 | 15 | 9 | 45 |
| 4 | 7 | 28 | 16 | 112 |
| 5 | 8 | 40 | 25 | 200 |
| 6 | 2 | 12 | 36 | 82 |
| | *n = 25* | *100* | | *438* |

$$\overline{x} = \frac{100}{25} = 4 \qquad v = \frac{\sum n_i x_i^2}{n} - \overline{x}^2 = \frac{438}{25} - 4^2 = 1,52 \qquad \sigma = \sqrt{1,52} = 1,23$$

Exemple 2 :
Lors d'une course de vitesse les 40 participants ont mis les temps suivants pour effectuer le parcours

| **Temps (secondes)** | **Centre** $x_i$ | **Effectif** $n_i$ | **Temps pondéré** $n_i x_i$ | $x_i^2$ | $n_i x_i^2$ |
|---|---|---|---|---|---|
| [43-45[ | 44 | 2 | 88 | 1936 | 3882 |
| [45-47[ | 46 | 3 | 138 | 2116 | 6348 |
| [48-49[ | 48 | 7 | 336 | 2304 | 16128 |
| [49-51[ | 50 | 11 | 550 | 2500 | 28500 |
| [51-53[ | 52 | 8 | 416 | 2804 | 21632 |
| [53-55[ | 54 | 6 | 324 | 2916 | 18496 |
| [55-57[ | 56 | 3 | 168 | 3136 | 9408 |
| | *Sommes :* | *n = 40* | *2020* | | 102384 |

$$\bar{x} = \frac{2020}{40} = 50,5 \qquad e = \frac{99}{40} = 2,48 \qquad v = \frac{102384}{40} - 50,5^2 = 9,35 \qquad \sigma = \sqrt{9,35} = 3,06$$




*Si l'on suit une loi normale 2/3 des temps sont compris entre 48,44 et 53,56 secondes.*

Propriétés de l'écart-type :

- On utilise l'écart-type que pour mesurer la dispersion autour de la moyenne d'un ensemble de données.
- L'écart-type n'est jamais négatif.
- L'écart-type est sensible aux valeurs aberrantes. Une seule valeur aberrante peut accroître l'écart-type et, par le fait même, déformer le portrait de la dispersion.
- Dans le cas des données ayant approximativement la même moyenne, plus la dispersion est grande, plus l'écart-type est grand.
- L'écart-type est zéro si toutes les valeurs d'un ensemble de données sont les mêmes (parce que chaque valeur est égale à la moyenne).

Remarque :
Quand on groupe une variable par intervalle de classe, on suppose que toutes les observations à l'intérieur de chaque intervalle sont égales au point milieu de l'intervalle. Ainsi, on ne tient pas compte de la dispersion des observations à l'intérieur de chaque intervalle, ce qui fait que l'écart-type est toujours inférieur à la valeur réelle. On devrait donc le considérer comme une approximation.

Exercice 13 :
Voici les résultats d' une enquête sur le poids (en kg) des bagages de 25 personnes

| | | | | |
|---|---|---|---|---|
| 35 | 25 | 30 | 25 | 30 |
| 25 | 20 | 30 | 40 | 35 |
| 35 | 35 | 30 | 25 | 30 |
| 35 | 35 | 35 | 30 | 35 |
| 40 | 30 | 25 | 35 | 40 |

a) Ranger ces résultats dans un tableau, puis calculer : la moyenne et l'écart-type
b) Représenter ces résultats à l'aide d'un histogramme, sur lequel vous indiquerez la moyenne et l'écart -type.

Exercice 14 :
Voici un tableau qui résume la durée de 50 films d'une vidéothèque.

a) Compléter le tableau suivant :

| durée (min) | Effectifs $n_i$ | $x_i$ |
|---|---|---|
| [100;120[ | **2** | |
| [120;140[ | **3** | |
| [140;160[ | **12** | |
| [160;180[ | **15** | |
| [180;200[ | **11** | |
| [200;220[ | **4** | |
| [220;240[ | **2** | |
| [240;260[ | **1** | |
| | | |

a) Compléter ce tableau, puis calculer : la moyenne et l'écart-type
b) Représenter ces résultats à l'aide d'un histogramme, sur lequel vous indiquerez la moyenne et l'écart -type.

Exercice 15 :

L'île ALPHA est habitée par 100 personnes, 50 hommes et 50 femmes.
Tous les hommes chaussent du 42 et toutes les femmes du 34.

a) Calculer la pointure moyenne des habitants de cette île.
b) Calculer l'écart type à cette moyenne.

L'île OMEGA est habitée par 100 personnes (aussi!), 50 hommes et 50 femmes.
Tous les hommes chaussent du 40 et toutes les femmes du 36.

c) Calculer la pointure moyenne des habitants de cette île.
d) Calculer l'écart type à cette moyenne.

L'île GAMMA est habitée par 100 personnes (décidément !), 50 hommes et 50 femmes.
Tous les hommes et toutes les femmes chaussent du 38

e) Calculer la pointure moyenne des habitants de cette île.
f) Calculer l'écart type à cette moyenne.

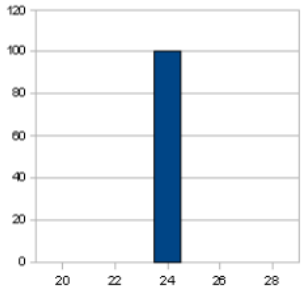
g) Pour chaque île, représenter graphiquement les données
h) Expliquer, les conséquences d'un « fort » écart type sur l'allure d'un graphique.
i) Expliquer l'utilité de l'écart type

Exercice 16 :
Pour chaque écart-type ci-dessous, retrouvez le diagramme correspondant :

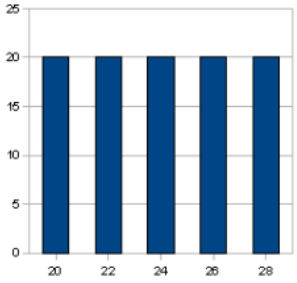
| | |
|---|---|
| Écart-type de : 0 | Diagramme : .... |
| Écart-type de : 2,21 | Diagramme : .... |
| Écart-type de : 2,31 | Diagramme : .... |
| Écart-type de : 2,83 | Diagramme : .... |
| Écart-type de : 3,29 | Diagramme : .... |

Toutes ces statistiques ont la même moyenne (24).

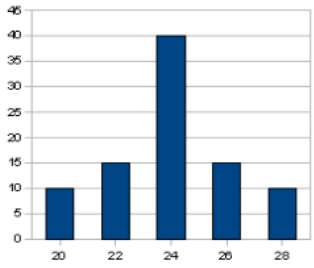
D1



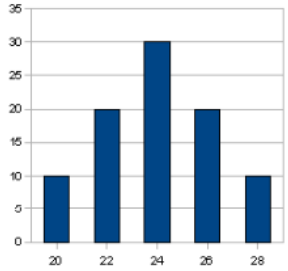

D2



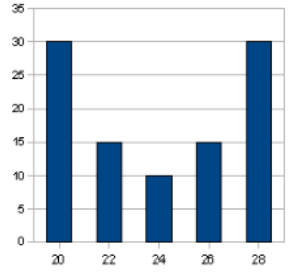

D3



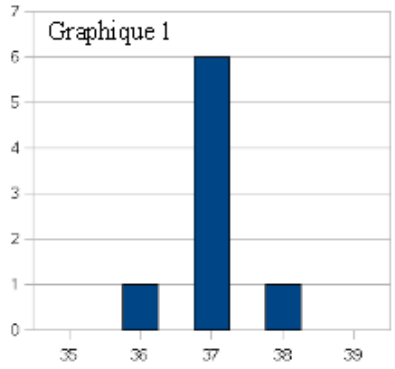

D4

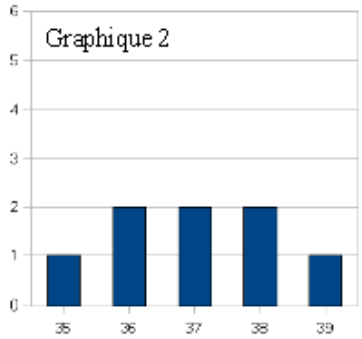


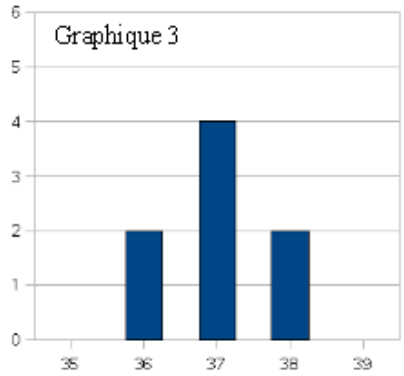

D5



Exercice 17 :
Retrouver, pour chaque graphique, la série de valeurs qui lui correspond




**Valeurs 1**
Moyenne = 37

**Valeurs 2**
Moyenne = 37

**Valeurs 3**
Moyenne = 37

* * *

*Si à la moyenne on associe l'écart-type, à la **médiane** on associera l'**intervalle semi-interquartile**. Il faut au préalable définir les **quartiles**.*

**Les quartiles :**
Soit F la fonction représentative du polygone des **fréquences cumulées**, on appelle respectivement $1^{er}$, $2^{ème}$ et $3^{ème}$ quartiles les valeurs :

$$F(Q_1)=\frac{1}{4}=25\% \qquad F(Q_2)=\frac{2}{4}=50\% \qquad F(Q_3)=\frac{3}{4}=75\%$$

On remarque que $Q_2$ n'est rien d'autre que la médiane $\tilde{x}$ déjà définie. On voit par ailleurs que l'intervalle $[Q_1;Q_3]$ contient 50 % des valeurs de l'échantillon (écart interquartile).

L'**intervalle semi-interquartile** (écart semi-quartile) est moitié de la longueur de cet intervalle :

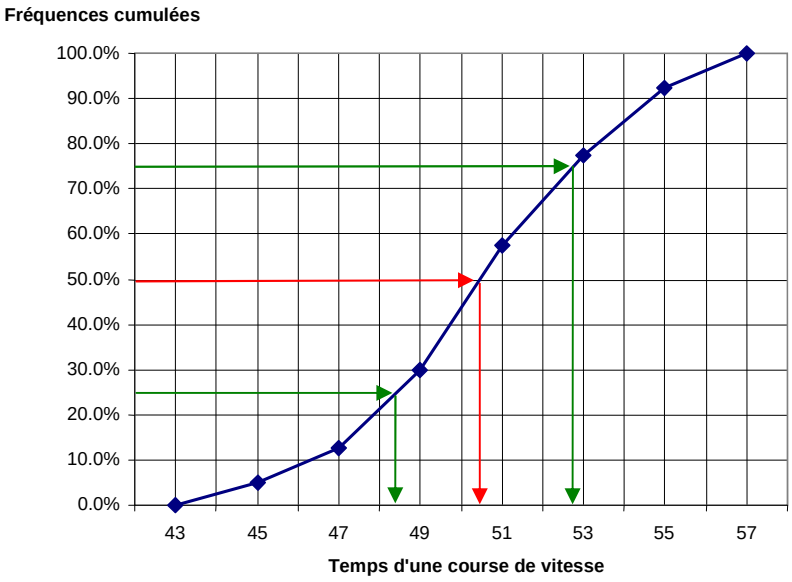
**L'intervalle semi-interquartile :**
$$I=\frac{Q_3-Q_1}{2}$$

*Cette mesure est associée à la médiane.*

L'écart semi-quartile n'est guère influencé par des valeurs plus élevées; c'est donc une bonne mesure de dispersion pour les distributions asymétriques. On utilise rarement des écarts semi-quartiles pour des ensembles de données dont les distributions sont normales. Lorsqu'un ensemble de données comporte une distribution normale, on a plutôt recours à l'écart-type.

Exemple : (cas continu)
*Voir l'énoncé en page 12.* Pour cette course de vitesse avec 40 participants, on avait obtenu le tableau suivant :

| Temps (secondes) | Fréquences cumulées |
|---|---|
| 43 | 0 % |
| 45 | 5 % |
| 47 | 12.5% |
| 49 | 30 % |
| 51 | 57,5 % |
| 53 | 77,5 % |
| 55 | 92,5 % |
| 57 | 100 % |




La médiane vaut : $\tilde{x} = 50,5$

$Q_1 = 48,4 \quad Q_3 = 52,8$

$$I = \frac{52,7 - 48,4}{2} = 2,15$$

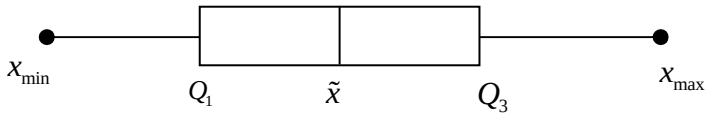
Cela signifie que dans un rayon moyen d'environ 2,15 secondes autour de la médiane on a 50% des échantillons.

**L'Étendue :** (Fourchette )

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

C'est la différence entre la valeur la plus élevée $x_{max}$ et la valeur la moins élevée $x_{min}$ d'un ensemble de données.

**Boîte à moustache**
On peut représenter un résumé en cinq nombres à l'intérieur d'un diagramme appelé **un tracé en rectangle et moustaches.**




Cette représentation est particulièrement indiquée pour une distribution asymétrique et s'il y a des observations inhabituelles (des valeurs aberrantes) dans l'ensemble de données. Les tracés en rectangle et moustaches sont idéals pour comparer des distributions, parce qu'ils font apparaître immédiatement le centre, la dispersion et l'étendue globale.

Exemple : (cas discret)
Gabrielle a commencé à travailler dans une boutique d'informatique il y a un an. Son superviseur lui a demandé de tenir un dossier du nombre d'ordinateurs qu'elle a vendus chaque mois.
L'ensemble de données qui suit indique le nombre d'ordinateurs qu'elle a vendus mensuellement au cours des 12 derniers mois : 34, 48, 1, 15, 58, 24, 20, 11, 19, 50, 28, 38.

On cherche :
a) la médiane
b) l'étendue
c) les quartiles supérieur et inférieur
d) l'écart interquartile et l'intervalle semi-interquartile

Résolution :

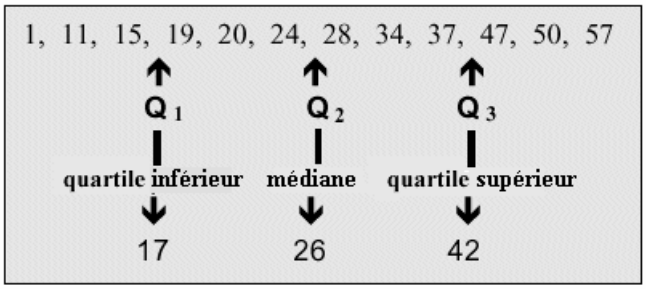
a) Les valeurs dans l'ordre ascendant sont : 1, 11, 15, 19, 20, **24, 28,** 34, 38, 48, 50, **57**.
Médiane = (6e + 8e observations) ÷ 2 = (24 + 28) ÷ 2 = **26**

b) Étendue = différence entre la valeur la plus élevée et la valeur la plus faible = 57 – 1 = **56**

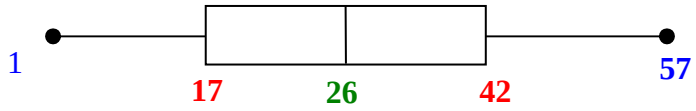
c) $Q_1$ = Quartile inférieur = valeur du milieu de la première moitié des données
= la médiane de 1, 11, 15, 19, 20, 24 = (3e + 4e observations) ÷ 2 = (15 + 19) ÷ 2 = **17**
$Q_3$ = Quartile supérieur = valeur du milieu de la seconde moitié des données
= la médiane de 28, 34, 38, 48, 50, 57 = (3e + 4e observations) ÷ 2 = (38 + 48) ÷ 2 = **42**

d) Écart interquartile = $Q_3 - Q_1 = 42 - 17 = 25$ et $I = \frac{25}{2} = 12,5$

On peut résumer ces résultats en cinq nombres : 1, 17, **26**, 42, **57**.




C'est-à-dire :




Exercice 18 :

Les températures énumérées ci-dessous sont les températures quotidiennes maximales (en degrés Celsius) enregistrées du 21 juin au 3 juillet :
29,3 ; 29,1; 28,2 ; 19,1 ; 18,8 ; 22,4 ; 18,4 ; 18,0 ; 20,2 ; 25,0 ; 25,8 ; 24,1 ; 22,1.
Calculer le résumé en cinq nombres et dessiner un tracé en rectangle et moustaches pour ces données.

Exercice 19 :

Le tableau suivant fournit un aperçu du nombre hypothétique de conflits de travail durant une période de dix ans.

| **Année** | **Nombre hypothétique de conflits de travail** |
|---|---|
| 1 | 266 |
| 2 | 231 |
| 3 | 223 |
| 4 | 262 |
| 5 | 260 |
| 6 | 230 |
| 8 | 191 |
| 8 | 182 |
| 9 | 165 |
| 10 | 153 |

Calculer le résumé en cinq nombres et dessiner la boîte à moustache

Exercice 20 :
Voici le nombre de parties de basket-ball auxquelles ont assisté 50 abonnés :
15, 10, 18, 11, 15, 12, 13, 16, 12, 14, 14, 16, 15, 18, 11, 16, 13, 18, 12, 16, 18, 15, 18, 15, 19, 13, 14, 18, 16, 15, 12, 11, 18, 16, 15, 10, 14, 15, 13, 16, 18, 15, 18, 11, 14, 18, 15, 14, 13, 16.
a) Compter les données.
b) Dessiner un diagramme à bandes.
c) Calculer la moyenne, la médiane et le mode.
d) Calculer la variance et l'écart-type
e) Calculer l'intervalle à l'intérieur duquel 95 % des observations devraient se situer.
f) Formuler un commentaire sur la dispersion des données.

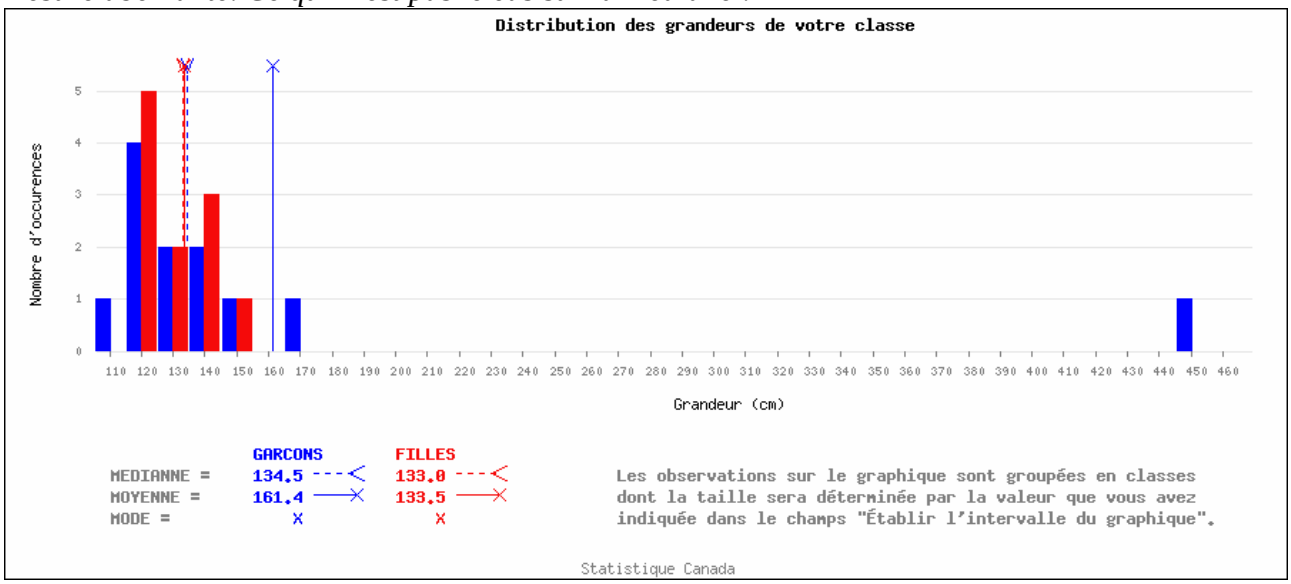
Exercice 21 : ***Tâches ménagères chez les hommes***
Une enquête aléatoire de 100 hommes mariés a donné la distribution suivante d'heures qu'ils consacraient par semaine à un travail ménager non rémunéré :

| Heure(s) | Nombre d'hommes |
|---|---|
| [0 ; 5[ | 1 |
| [5 ; 10[ | 18 |
| [10 ; 15[ | 24 |
| [15 ; 20[ | 25 |
| [20 ; 25[ | 18 |
| [25 ; 30[ | 12 |
| [30 ; 35[ | 1 |
| [35 ; 40[ | 1 |

a) Calculez les fréquences cumulées.
b) Dessinez l'ogive (ou la courbe de distribution) à l'aide de la fréquence cumulée
c) À partir de la courbe, trouvez une valeur médiane approximative.
Qu'est-ce que cette valeur indique ?
d) Trouver l'intervalle semi-interquartile.
e) Donnez le tracé en rectangle et moustaches.
f) Quel est le mode ?
g) Calculez la moyenne. Qu'est-ce que cette valeur indique ? Donner l'écart-type.
h) Décrivez brièvement la comparaison entre les valeurs moyenne, médiane et modale.
Comment détermineriez-vous si les femmes ont consacré plus d'heures par semaine que les hommes à un travail ménager non rémunéré ?

**Activité 3 : (sur le web)**
Quelle est la taille moyenne des élèves de votre classe ?
**http://www.statcan.ca/francais/kits/height_f.htm**

*Sur l'exemple ci-dessous on voit nettement l'influence sur la moyenne de la taille des garçons de la mesure aberrante. Ce qui n'est pas le cas sur la médiane !*




On trouvera d'autres activités sur la même page ou en particulier sur :
**http://www.statcan.ca/francais/edu/index_f.htm** ou **http://www.statcan.ca/start_f.html**

## Activité 4 : Calculatrice scientifique

TEXAS INSTRUMENTS

**TI-30X IIB**

et

**TI-30X IIS**

**Calculateur Scientifique**

**Stats** 2nd [STAT] [EXIT STAT] DATA STATVAR

1-VAR stats analyse les données à partir d'1 ensemble de données avec 1 variable mesurée, *x*. 2-VAR stats analyse les données couplées à partir de 2 ensembles de données avec 2 variables mesurées—*x*, la variable indépendante, et *y*, la variable dépendante. Vous pouvez entrer jusqu'à 42 ensembles de données.

Etapes pour définir les différentes valeurs de la variable.

1. Appuyez sur 2nd [STAT]. Sélectionnez 1-VAR ou 2-VAR. L'indicateur STAT s'affiche.
2. Appuyez sur DATA.
3. Entrez une valeur pour $x_1$. ENTER l'évalue et affiche la valeur.
4. Appuyez sur ⊙.
   - En mode Stat 1-VAR, entrez la fréquence d'occurrence (FRQ) correspondant aux différentes valeurs de la variable. FRQ par défaut=1. Si FRQ=0, les valeurs sont ignorées.
   - En mode Stat 2-VAR, entrez la valeur pour $y_1$ et appuyez sur ENTER.
5. Répétez les étapes 3 et 4 jusqu'à ce que toutes les valeurs des variables soient entrées. Vous devez appuyer sur ENTER ou ⊙ pour sauvegarder les valeurs ou la fréquence FRQ introduites en dernier. Si vous ajoutez ou supprimez des valeurs, la TI-30X II réordonne automatiquement la liste.
6. Quand toutes les valeurs et fréquences sont entrées :
   - Appuyez sur STATVAR pour afficher le menu des variables (voir le tableau pour les définitions) et leurs valeurs courantes, ou
   - Appuyez sur DATA pour revenir à l'écran STAT vierge. Vous pouvez faire les calculs avec les variables ($\bar{x}$, $\bar{y}$, etc.). Sélectionnez une variable dans le menu STATVAR et appuyez ensuite sur ENTER pour évaluer le calcul.
7. Ceci étant fait :
   - Appuyez sur 2nd [STAT] et sélectionnez CLRDATA pour effacer toutes les valeurs *sans* quitter le mode STAT, ou
   - Appuyez sur 2nd [EXIT STAT] pour effacer toutes les données statistiques et quitter le mode STAT (L'indicateur STAT s'éteint).

| Variables | Définition |
|---|---|
| **n** | Nombre total de valeurs *x* ou (*x*,*y*). |
| $\bar{x}$ ou $\bar{y}$ | Moyenne de toutes les valeurs *x* ou *y*. |
| **Sx** ou **Sy** | Ecart-type d'échantillon standard *x* ou *y*. |
| $\sigma x$ ou $\sigma y$ | Ecart-type de population standard *x* ou *y*. |
| $\Sigma x$ ou $\Sigma y$ | Somme de toutes les valeurs *x* ou *y*. |
| $\Sigma x^2$ ou $\Sigma y^2$ | Somme de toutes les valeurs $x^2$ ou $y^2$. |
| $\Sigma xy$ | Somme de ($x * y$) pour toutes les paires *xy*. |
| **a** | Pente de la droite de régression linéaire. |
| **b** | Ordonnée à l'origine de la droite de régression - interception *y*. |
| **r** | Coefficient de corrélation. |
| **x′** (2-VAR) | Utilise **a** et **b** pour calculer la valeur *x* prévue quand vous entrez une valeur *y*. |
| **y′** (2-VAR) | Utilise **a** et **b** pour calculer la valeur *y* prévue quand vous entrez une valeur *x*. |

Refaire l'exercice ci-dessous avec la calculatrice

Exercice 20 :
Voici le nombre de parties de basket-ball auxquelles ont assisté 50 abonnés :
15, 10, 18, 11, 15, 12, 13, 16, 12, 14, 14, 16, 15, 18, 11, 16, 13, 18, 12, 16, 18, 15, 18, 15, 19, 13, 14, 18, 16, 15, 12, 11, 18, 16, 15, 10, 14, 15, 13, 16, 18, 15, 18, 11, 14, 18, 15, 14, 13, 16.

Calculer la moyenne, la médiane et le mode.
Calculer la variance et l'écart-type

* * *

*La suite du chapitre est laissée à l'appréciation de l'enseignant(e).*