

# Les statistiques descriptives

## Première S

L'équipe des professeurs de mathématiques  
Lycée Stendhal

“Il est prouvé que fêter les anniversaires est bon pour la santé. Les statistiques montrent que les personnes qui en fêtent le plus deviennent les plus vieilles.”

Den Hartog (Professeur américain)

Année 2016-2017

Liste des savoirs et savoir-faire du chapitre :

CODE	INTITULE	Bilan		
		A	EA	NA
S0101	Déterminer moyenne, médiane et quartiles d'une série statistique			
S0102	Déterminer variance et écart-type d'une série statistique			
S0103	Dresser le diagramme en boîte d'une série statistique			
S0104	Décrire le diagramme en boîte d'une série statistique			
S0105	Comparer deux séries statistiques avec les outils appropriés			

Compétences dans tous les chapitres :

INTITULE	Bilan		
	A	EA	NA
<b>Chercher</b>			
<b>Modéliser</b>			
<b>Représenter</b>			
<b>Calculer</b>			
<b>Raisonner</b>			
<b>Communiquer</b>			

## Table des matières

<b>1</b>	<b>Notation</b>	<b>3</b>
<b>2</b>	<b>Quelques généralités et rappels</b>	<b>3</b>
2.1	Tri à plat . . . . .	3
2.2	Effectif total d'une série . . . . .	3
2.3	Fréquence d'apparition d'une valeur . . . . .	4
2.4	Tableau Standard statistique . . . . .	4
<b>3</b>	<b>Critère de position</b>	<b>4</b>
3.1	Médiane . . . . .	4
3.2	Quartiles . . . . .	5
3.3	Déciles . . . . .	5
3.4	Moyenne . . . . .	6
3.5	Modes . . . . .	6
<b>4</b>	<b>Critère de dispersion</b>	<b>7</b>
4.1	Etendue . . . . .	7
4.2	Ecart Inter-Quartiles . . . . .	7
4.3	Ecart Inter-Déciles . . . . .	7
4.4	Variance . . . . .	7
4.5	Ecart-Type . . . . .	9
<b>5</b>	<b>Représentation statistiques</b>	<b>9</b>
5.1	Diagramme en boîte . . . . .	9
5.2	Polygône des fréquences cumulées . . . . .	10
<b>6</b>	<b>Plages de normalité des distributions normales Gaussiennes</b>	<b>10</b>

## 1 Notation

La somme de  $n$  nombres numérotés de 1 à  $n$  peut s'écrire :

$$x_1 + x_2 + x_3 + x_4 + \cdots + x_{n-1} + x_n$$

mais cette écriture est longue et les pointillés ne sont pas satisfaisants.

On écrira, pour faire moins long et éviter les pointillés, cette somme à l'aide du symbole Sigma :

$$\sum_{i=1}^n x_i$$

Exemples :

$$1. \sum_{i=0}^n x_i =$$

$$2. \sum_{i=1}^{n-1} i =$$

$$3. \sum_{i=0}^n i^2 =$$

$$4. \sum_{i=0}^n i(i+1) =$$

## 2 Quelques généralités et rappels

### 2.1 Tri à plat

On note  $(x_i; n_i)_{i \in \mathbb{N}}$  la série statistique ci-dessous :

**Rappels** : L'effectif  $n_i$  est le nombre de fois où apparaît la valeur  $x_i$  dans la série.

valeurs $x_i$	$x_1$	$x_2$	$x_3$	$x_4$	...	...	...	$x_{k-2}$	$x_{k-1}$	$x_k$
Effectifs $n_i$	$n_1$	$n_2$	$n_3$	$n_4$	...	...	...	$n_{k-2}$	$n_{k-1}$	$n_k$

### 2.2 Effectif total d'une série

**Définition** :

L'effectif total  $N$  de la série statistique est la somme de tous les effectifs ou le nombre de valeurs total dans cette série :

$$N = \sum_{i=1}^k n_i = n_1 + n_2 + n_3 + \dots + n_{k-1} + n_k$$

## 2.3 Fréquence d'apparition d'une valeur

**Définition :**

La fréquence d'apparition d'une valeur  $x_i$  est la proportion de cette valeur par rapport à l'effectif total.

$$\text{Fréquence par rapport à 1 : } f_i = \frac{\text{Effectif de la valeur}}{\text{Effectif total}} = \frac{n_i}{N}$$

$$\text{Fréquence par rapport à 100 : } F_i = \frac{100n_i}{N}$$

**Propriétés :**

$$S_f = \sum_{i=1}^k f_i = f_1 + f_2 + f_3 + \dots + f_{k-1} + f_k = 1$$

$$S_F = \sum_{i=1}^k F_i = F_1 + F_2 + F_3 + \dots + F_{k-1} + F_k = 100$$

## 2.4 Tableau Standard statistique

Le tri à plat d'une série statistique est un tableau contenant les valeurs de la série, les effectifs, les effectifs cumulés croissants, les fréquences, les fréquences cumulées croissantes, les pourcentages et les pourcentages cumulés croissants.

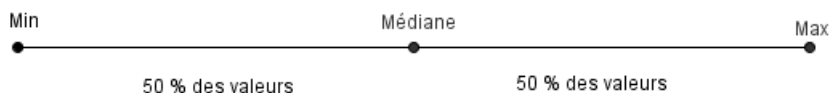
valeurs $x_i$	$x_1$	$x_2$	$x_3$	...	...	...	$x_{k-1}$	$x_k$
Effectifs $n_i$	$n_1$	$n_2$	$n_3$	...	...	...	$n_{k-1}$	$n_k$
Effectifs Cum Croi $N_i$	$n_1$	$N_1 + n_2$	$N_2 + n_3$	...	...	...	$N_{k-2} + n_{k-1}$	$N$
Fréquences $f_i$	$f_1$	$f_2$	$f_3$	...	...	...	$f_{k-1}$	$f_k$
Fréquences Cum Croi $F_i$	$f_1$	$F_1 + f_2$	$F_2 + f_3$	...	...	...	$F_{k-2} + f_{k-1}$	1
Pourcentages $p_i$	$p_1$	$p_2$	$p_3$	...	...	...	$p_{k-1}$	$p_k$
Pourcentages Cum Croi $P_i$	$p_1$	$P_1 + p_2$	$P_2 + p_3$	...	...	...	$P_{k-2} + p_{k-1}$	100

## 3 Critère de position

### 3.1 Médiane

**Définition :**

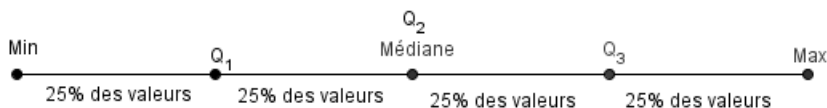
La médiane d'une série statistique est la valeur qui partage cette série en deux séries de même effectif.



### 3.2 Quartiles

#### Définition :

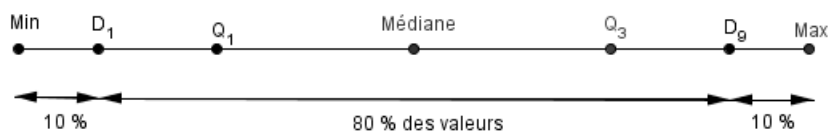
Les quartiles d'une série statistique sont les valeurs qui partagent cette série en quatre séries de même effectif.



### 3.3 Déciles

#### Définition :

Les déciles d'une série statistique sont les valeurs qui partagent cette série en dix séries de même effectif.



Si  $D_1$  est le premier décile et  $D_9$  le neuvième de la série statistique, alors :

10 % des valeurs de la série sont dans  $[Min, D_1]$

80 % des valeurs de la série sont dans  $[D_1, D_9]$

10 % des valeurs de la série sont dans  $[D_9, Max]$

#### Méthode pour trouver les déciles :

Il faut commencer par classer la série dans l'ordre croissant.

On utilisera une méthode approximative mais qui donnera des résultats significatifs pour des séries à grands effectifs.

Calculer  $\frac{N}{10}$  et on note  $a$  l'entier supérieur à  $\frac{N}{10}$ .

Calculer  $\frac{9N}{10}$  et on note  $b$  l'entier supérieur à  $\frac{9N}{10}$ .

▷  $D_1$  est la  $a$  ième valeur de la série statistique.

▷  $D_9$  est la  $b$  ième valeur de la série statistique.

### 3.4 Moyenne

Définition :

La moyenne arithmétique de la série statistique est le nombre :

$$\bar{x} = \frac{1}{N} \sum_{i=1}^k x_i \times n_i = \frac{x_1 n_1 + x_2 n_2 + x_3 n_3 + \dots + x_{k-1} n_{k-1} + x_k n_k}{N}$$

ou

$$\bar{x} = \sum_{i=1}^k x_i \times f_i = x_1 f_1 + x_2 f_2 + x_3 f_3 + \dots + x_k f_k$$

ou

$$\bar{x} = \frac{1}{100} \sum_{i=1}^k p_i \times n_i = \frac{p_1 n_1 + p_2 n_2 + p_3 n_3 + \dots + p_{k-1} n_{k-1} + p_k n_k}{100}$$

Propriétés de la moyenne :

1. Si  $\bar{x}$  est la moyenne d'un groupe d'effectif  $N_1$  et  $\bar{y}$  la moyenne d'un groupe d'effectif  $N_2$  alors la moyenne  $\bar{z}$  de la série constituée de l'ensemble des deux groupes est :

$$\bar{z} = \frac{N_1 \bar{x} + N_2 \bar{y}}{N_1 + N_2}$$

2. Si  $\bar{x}$  est la moyenne d'une série  $(x_i, n_i)$  alors la moyenne de la série  $(ax_i + b, n_i)$  est :

$$\bar{y} = a\bar{x} + b$$

3. **Moyenne élaguée :**

Quand une valeur aberrante, correspondant à une erreur de mesure ou à une situation exceptionnelle, est présente dans une série, elle influence considérablement la valeur moyenne. Une moyenne calculée après avoir enlevé certaines valeurs est appelée **Moyenne élaguée**.

### 3.5 Modes

Définition :

Les modes d'une série sont les valeurs ayant le plus grand effectif.

## 4 Critère de dispersion

### 4.1 Etendue

Définition :

L'étendue d'une série statistique est la différence entre la plus grande valeur et la plus petite, de la série.

$$Et = Max - Min$$

### 4.2 Ecart Inter-Quartiles

Définition :

L'écart inter-quartiles est la différence entre  $Q_3$  et  $Q_1$

$$E_Q = Q_3 - Q_1$$

L'intervalle inter-quartiles est l'intervalle entre  $Q_1$  et  $Q_3$

$$I_Q = [Q_1, Q_3]$$

### 4.3 Ecart Inter-Déciles

Définition :

L'écart inter-déciles est la différence entre  $D_9$  et  $D_1$

$$E_D = D_9 - D_1$$

L'intervalle inter-déciles est l'intervalle entre  $D_1$  et  $D_9$

$$I_D = [D_1, D_9]$$

### 4.4 Variance

Certaines séries statistiques peuvent avoir les mêmes critères de position comme la médiane et la moyenne.

Pour les différencier on va utiliser un nouvel outil qui va mesurer la dispersion de la série autour de la moyenne. On souhaite trouver une mesure de l'écart entre les valeurs de la série et sa moyenne. Si cet écart est grand alors la série est très hétérogène et les valeurs sont éloignées de la moyenne sinon si cet écart est petit la série est homogène et les valeurs rapprochées autour de la moyenne.

On pourrait calculer la moyenne des écarts à la moyenne mais celle-ci donne toujours 0 à cause des écarts qui sont opposés.

Démonstration :

$$\overline{(\bar{x} - x)} = \frac{1}{N} \sum_{i=1}^k (\bar{x} - x_i) n_i$$

Pour éviter ce problème, on va faire la moyenne des carrés des écarts à la moyenne. On note ce résultat, la **variance** de la série.

**Définition :**

**La variance d'une série statistique est la moyenne des carrés des écarts à la moyenne de chacune des valeurs.**

valeurs $x_i$	$x_1$	$x_2$	$x_3$	...	$x_{k-1}$	$x_k$
Effectifs $n_i$	$n_1$	$n_2$	$n_3$	...	$n_{k-1}$	$n_k$
$(\bar{x} - x_i)^2$	$(\bar{x} - x_1)^2$	$(\bar{x} - x_2)^2$	$(\bar{x} - x_3)^2$	...	$(\bar{x} - x_{k-1})^2$	$(\bar{x} - x_k)^2$

La variance de la série est donc la moyenne de la dernière ligne du tableau ci-dessus :

$$V(x) = \frac{1}{N} \sum_{i=1}^k (x - x_i)^2 n_i$$

ou

$$V(x) = \sum_{i=1}^k (x - x_i)^2 f_i$$

**Propriété :**

$$V(ax) = \dots\dots\dots$$

$$V(x + b) = \dots\dots\dots$$

donc

$$V(ax + b) = \dots\dots\dots$$

démonstration :



## 4.5 Ecart-Type

**Définition :**

L'écart-type  $\sigma$  est la racine carrée de la variance pour revenir aux même unités que les valeurs de la série statistique.

$$\sigma = \sqrt{V(x)}$$

**Propriété :**

Si on a deux séries  $S_1$  et  $S_2$  d'écart-type respectifs  $\sigma_1$  et  $\sigma_2$

Si  $\sigma_1 < \sigma_2$  alors la série  $S_1$  est plus homogène que la série  $S_2$  ou la série  $S_2$  est plus hétérogène que la série  $S_1$ .

**Propriété :**

$$\sigma(ax) = \dots\dots\dots$$

$$\sigma(x + b) = \dots\dots\dots$$

donc

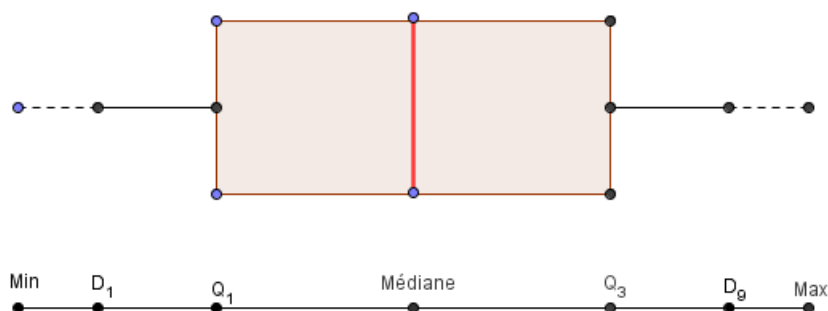
$$\sigma(ax + b) = \dots\dots\dots$$

démonstration :

## 5 Représentation statistiques

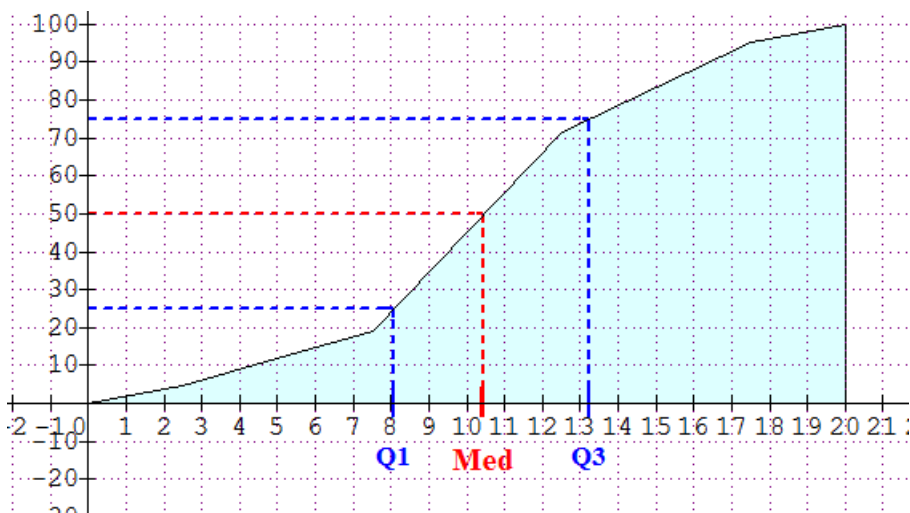
### 5.1 Diagramme en boîte

Les diagrammes en boîte, ou boîtes à moustaches, sont des diagrammes permettant de comparer rapidement des séries statistiques.



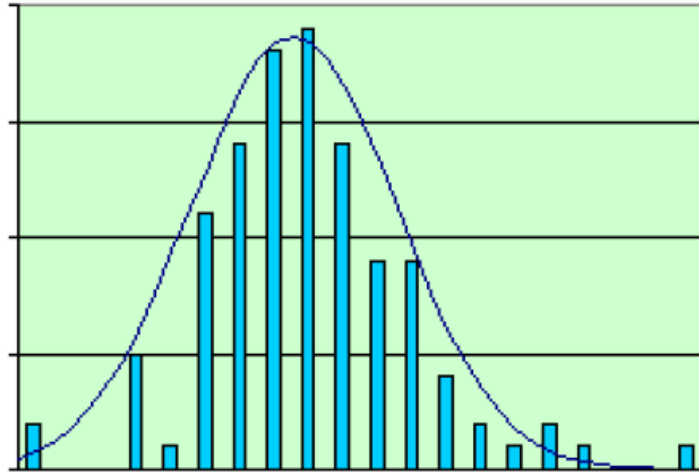
### 5.2 Polygone des fréquences cumulées

Le polygone des fréquences cumulées (en pourcentages) permet de lire rapidement la médiane et les quartiles d'une série statistique :



## 6 Plages de normalité des distributions normales Gaussiennes

Lorsque l'on fait des statistiques sur une grande quantité de valeurs, il arrive souvent que l'on obtienne des diagrammes ayant sensiblement la même forme dite EN CLOCHE ou COURBE DE GAUSS, comme ci-dessous :



Lorsque la série statistique donne une représentation graphique de la forme d'une courbe de Gauss, les données sont qualifiées de données Gaussiennes.

### Propriétés (Plages de normalité) :

On note  $\bar{x}$  la moyenne de la série et  $\sigma$  l'écart-type de la série.

1. Environ 68 % des données se trouvent dans l'intervalle  $[\bar{x} - \sigma, \bar{x} + \sigma]$   
On nomme cet intervalle **la plage de normalité pour le niveau de confiance 0.68**
2. Environ 95 % des données se trouvent dans l'intervalle  $[\bar{x} - 2\sigma, \bar{x} + 2\sigma]$   
On nomme cet intervalle **la plage de normalité pour le niveau de confiance 0.95**
3. Environ 99 % des données se trouvent dans l'intervalle  $[\bar{x} - 3\sigma, \bar{x} + 3\sigma]$   
On nomme cet intervalle **la plage de normalité pour le niveau de confiance 0.99**