

# Les statistiques descriptives

## Classe de seconde

### L'équipe des professeurs de mathématiques

Lycée Stendhal

J'aimais et j'aime encore les mathématiques pour elles-mêmes comme n'admettant pas l'hypocrisie et le vague, mes deux bêtes d'aversion.

### Stendhal

Année 2011-2012

### Liste des savoirs et savoir-faire du chapitre :

CODE	INTITULE	Bilan		
		A	EA	NA
S0101	Savoir déterminer l'effectif d'une valeur de la série.			
S0102	Savoir déterminer la fréquence d'une valeur de la série.			
S0103	Savoir dresser le tableau des effectifs cumulés.			
S0104	Savoir dresser le tableau des fréquences cumulées.			
S0105	Savoir représenter une série par son nuage de points.			
S0106	Savoir représenter une série par un histogramme.			
S0107	Savoir représenter une série par sa fonction de répartition.			
S0108	Savoir déterminer la médiane d'une série par classement.			
S0109	Savoir déterminer la médiane d'une série graphiquement.			
S0110	Savoir déterminer les quartiles d'une série par classement.			
S0111	Savoir déterminer les quartiles d'une série graphiquement.			
S0112	Savoir déterminer la moyenne d'une série avec les effectifs.			
S0113	Savoir déterminer la moyenne d'une série avec les fréquences.			
S0114	Savoir comparer deux séries.			

# Statistiques Descriptives

## ( En Seconde )

Dernière mise à jour : Dimanche 31 Décembre 2011

---

Vincent OBATON, Enseignant au lycée Stendhal de Grenoble (Année 2011-2012)

---

J'aimais et j'aime encore les mathématiques pour elles-mêmes comme n'admettant pas l'hypocrisie et le vague, mes deux bêtes d'aversion.

**Stendhal**

## Table des matières

<b>1</b>	<b>Quelques généralités et rappels</b>	<b>5</b>
1.1	Tri à plat . . . . .	5
1.2	Effectif total d'une série . . . . .	5
1.3	Fréquence d'apparition d'une valeur . . . . .	5
1.4	Tableau Standard statistique . . . . .	5
<b>2</b>	<b>Critère de position</b>	<b>6</b>
2.1	Médiane . . . . .	6
2.2	Quartiles . . . . .	6
2.3	Déciles . . . . .	7
2.4	Moyenne . . . . .	8
2.5	Modes . . . . .	8
<b>3</b>	<b>Critère de dispersion</b>	<b>8</b>
3.1	Etendue . . . . .	8
3.2	Ecart Inter-Quartiles . . . . .	9
3.3	Ecart Inter-Déciles . . . . .	9
<b>4</b>	<b>Représentation statistiques</b>	<b>9</b>
4.1	Diagramme en boîte . . . . .	9
4.2	Autres diagrammes . . . . .	9
4.3	Polygône des fréquences cumulées . . . . .	11

## 1 Quelques généralités et rappels

### 1.1 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 $n_i$	$n_1$	$n_2$	$n_3$	$n_4$	...	...	...	$n_{k-2}$	$n_{k-1}$	$n_k$

### 1.2 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 :

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

### 1.3 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.

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

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

**Propriétés** :

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

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

### 1.4 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.

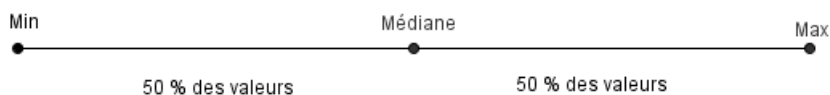
valeurs $x_i$	$x_1$	$x_2$	$x_3$	...	...	...	$x_{k-1}$	$x_k$
Effectifs $n_i$	$n_1$	$n_2$	$n_3$	...	...	...	$n_{k-1}$	$n_k$
Effectifs Cum Croi $N_i$	$n_1$	$N_1 + n_2$	$N_2 + n_3$	...	...	...	$N_{k-2} + n_{k-1}$	$N$
Fréquences $f_i$	$f_1$	$f_2$	$f_3$	...	...	...	$f_{k-1}$	$f_k$
Fréquences Cum Croi $F_i$	$f_1$	$F_1 + f_2$	$F_2 + f_3$	...	...	...	$F_{k-2} + f_{k-1}$	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

## 2 Critère de position

### 2.1 Médiane

#### Définition :

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



Si  $M_e$  est la médiane de la série statistique, alors :

- 50 % des valeurs de la série sont inférieures ou égales à  $M_e$
- 50 % des valeurs de la série sont supérieures ou égales à  $M_e$

#### Méthode pour trouver la médiane :

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

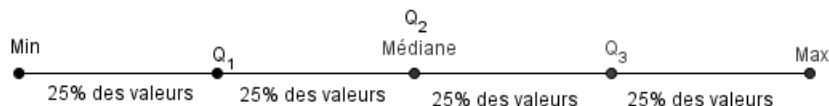
▷ Si  $\frac{N}{2} + 0,5 = d$  est entier alors la médiane est la  $d$  ième valeur de la série.

▷ Si  $\frac{N}{2} + 0,5 = d,5$  est décimale alors la médiane est entre la  $d$  ième et la  $d + 1$  ième valeur de la série.

### 2.2 Quartiles

#### Définition :

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



Si  $Q_1$  est le premier quartile et  $Q_3$  le troisième de la série statistique, alors :

25 % des valeurs de la série sont dans  $[Min, Q_1]$

50 % des valeurs de la série sont dans  $[Q_1, Q_3]$

25 % des valeurs de la série sont dans  $[Q_3, Max]$

### Méthode pour trouver les quartiles :

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. (Autement il suffit de couper en deux les deux séries  $[Min, M_e]$  et  $[M_e, Max]$ )

Calculer  $\frac{N}{4}$  et on note  $a$  l'entier supérieur à  $\frac{N}{4}$ .

Calculer  $\frac{3N}{4}$  et on note  $b$  l'entier supérieur à  $\frac{3N}{4}$ .

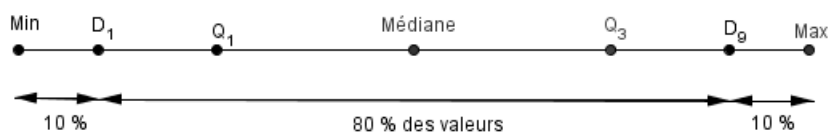
▷  $Q_1$  est la  $a$  ième valeur de la série statistique.

▷  $Q_3$  est la  $b$  ième valeur de la série statistique.

## 2.3 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.

## 2.4 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

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

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**.

## 2.5 Modes

Définition :

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

## 3 Critère de dispersion

### 3.1 Etendue

Définition :

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

$$Et = Max - Min$$

### 3.2 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  $Q_3$

$$I_Q = [Q_1, Q_3]$$

### 3.3 Ecart Inter-Déciles

Définition :

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

$$E_D = D_9 - D_1$$

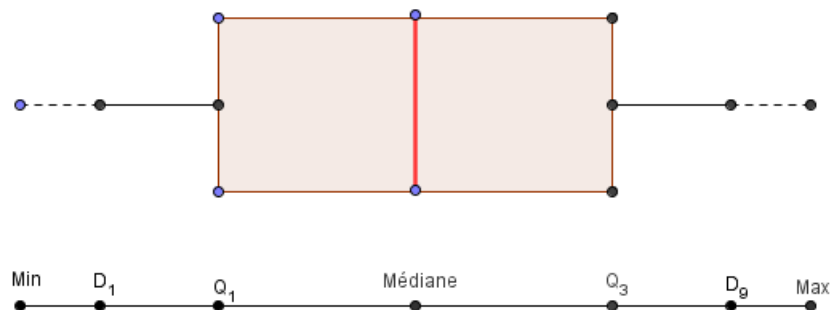
L'intervalle inter-déciles est l'intervalle entre  $D_1$  et  $D_9$

$$I_D = [D_1, D_9]$$

## 4 Représentation statistiques

### 4.1 Diagramme en boîte

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



### 4.2 Autres diagrammes

Diagramme en barres (Histogrammes) :

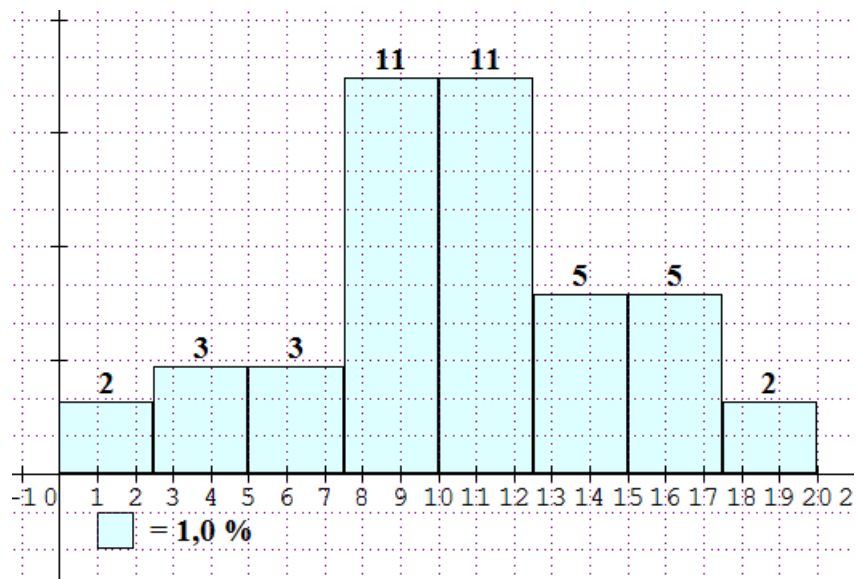
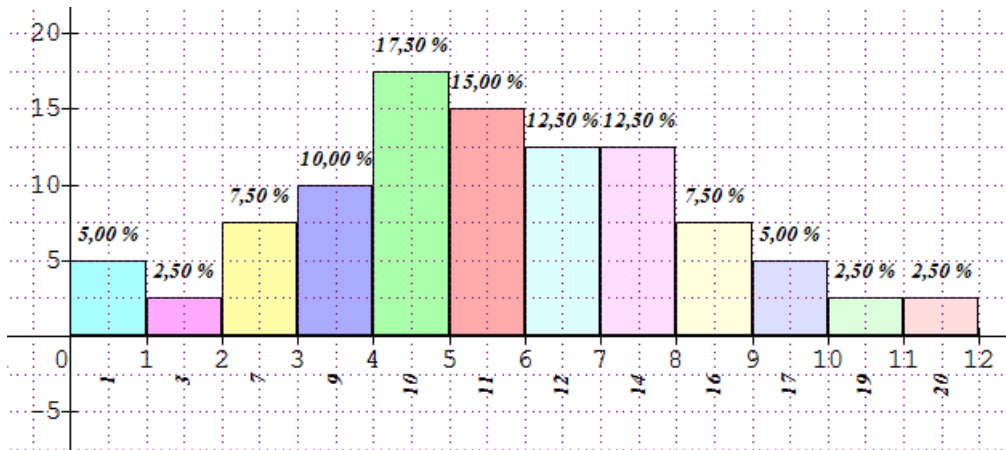


Diagramme en bâtons :

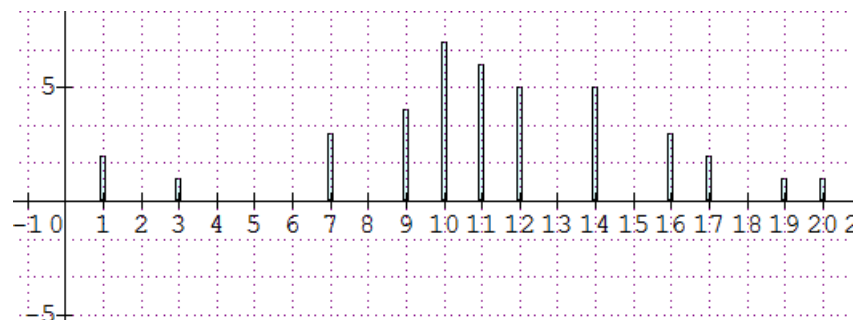
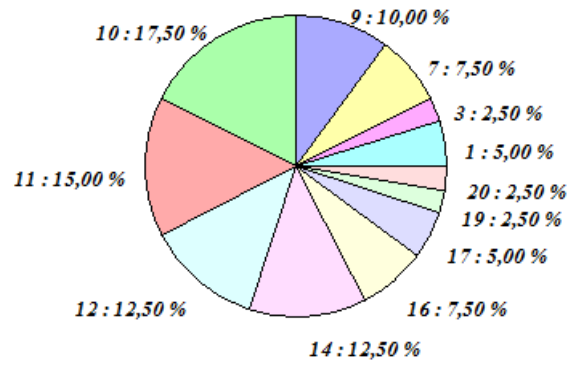


Diagramme en camembert :



### 4.3 Polygone des fréquences cumulées

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

