

# Statistiques - cours - 1 STG

F.Gaudon

12 février 2008

## Table des matières

<b>1</b>	<b>Vocabulaire des statistiques</b>	<b>2</b>
<b>2</b>	<b>Représentations graphiques</b>	<b>2</b>
2.1	Caractères qualitatifs ou quantitatifs discrets . . . . .	2
2.2	Caractères quantitatifs continus . . . . .	3
2.2.1	Histogramme . . . . .	3
2.2.2	Polygone des effectifs cumulés . . . . .	3
<b>3</b>	<b>Caractéristiques d'une série statistique</b>	<b>4</b>
3.1	Caractéristiques de position . . . . .	4
3.1.1	mode . . . . .	4
3.1.2	Médiane . . . . .	5
3.1.3	Première définition des quantiles . . . . .	6
3.1.4	Deuxième définition des quantiles . . . . .	7
3.1.5	Moyenne . . . . .	7
3.2	Caractéristiques de dispersion . . . . .	8
3.2.1	Étendue . . . . .	8
3.2.2	Écart type . . . . .	9
3.2.3	Écart inter quartiles . . . . .	10
<b>4</b>	<b>Diagrammes en boîte</b>	<b>10</b>
<b>5</b>	<b>Tableaux croisés</b>	<b>11</b>

# 1 Vocabulaire des statistiques

## Définitions :

- *Population* : ensemble étudié ;
- *individu* : élément de la *population* ;
- *caractère* : propriété étudiée.  
Un *caractère* peut être :
  - *qualitatif* : les valeurs du *caractère* ne sont pas numériques (ex. : ville d'origine) ;
  - *quantitatif* : les valeurs du *caractère* sont numériques (ex. taille).  
Un *caractère quantitatif* peut être :
    - *discret* : valeurs peu nombreuses (ex. : nombre d'enfants) ;
    - *continue* : valeurs infinies ou très nombreuses (ex. : taille).Les valeurs d'un caractère sont notées  $x_i$ , on supposera qu'elles sont au nombre de  $k$  dans la série.
- *classe* : partie de la *population* de valeurs de *caractères* proches ;
- *effectif* : nombre d'individus d'une classe (noté  $n_i$ ) ou d'une population (effectif total noté  $N$ ).
- *fréquence* :  $f_i = \frac{n_i}{N}$  ou en pourcentage  $f_i = \frac{n_i}{N} \times 100$ .

## Remarques :

- $n_1 + n_2 + n_3 + \dots + n_k = N$  ce que l'on note  $\sum_{i=1}^k n_i = N$
- $f_1 + f_2 + \dots + f_k = 1$  ce que l'on note  $\sum_{i=1}^k f_i = 1$

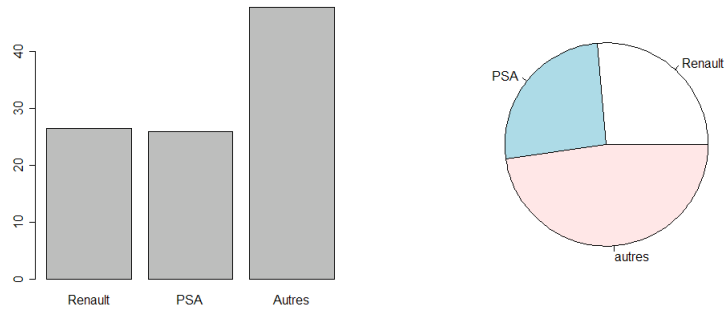
## 2 Représentations graphiques

### 2.1 Caractères qualitatifs ou quantitatifs discrets

On utilise un diagramme *en bâtons* ou *en barres* ou *circulaire*.

#### Exemple :

Ventes de voitures neuves en août 1996 :		
Renault	PSA	étrangères
26,4%	25,9 %	47,7%

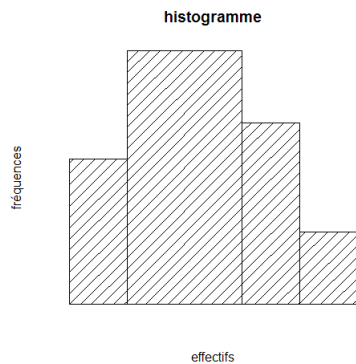


## 2.2 Caractères quantitatifs continus

### 2.2.1 Histogramme

Propriété :

Dans un *histogramme* l'effectif d'une *classe* est proportionnel à *l'aire des rectangles* et pas à *hauteur*.



Exemple :

prix	[0 ; 500[	[500 ; 1000[	[1000 ; 1500[	[1500 ; 2000[	[2000 ; 2500[
nombre d'articles	4	8	6	5	2

### 2.2.2 Polygone des effectifs cumulés

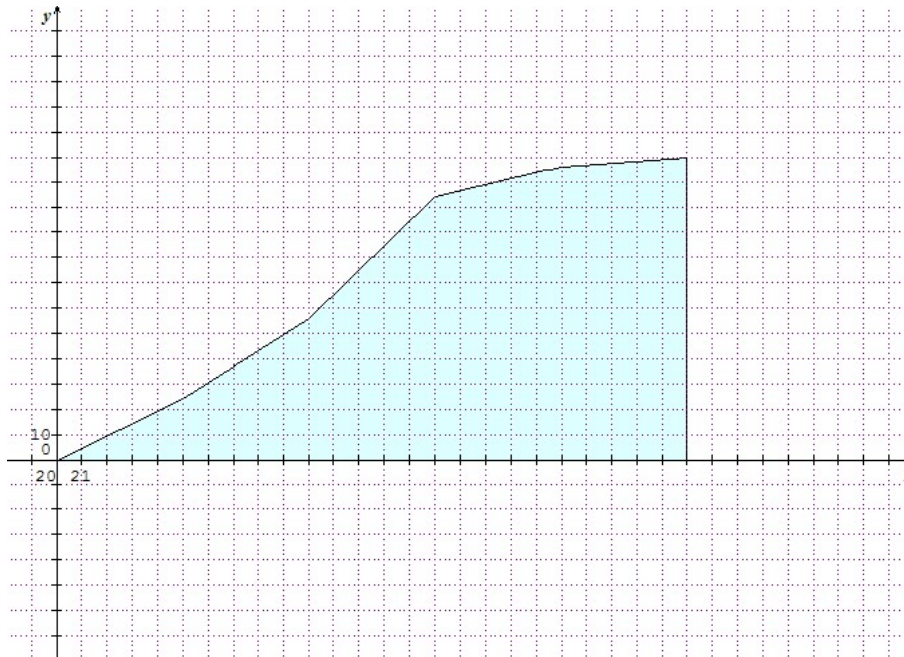
Méthode de construction :

Lorsque la série est regroupée en classes  $[a_i; b_i[$ , le polygone des effectifs cumulés croissants est formé des segments reliant les points ayant pour abscisse l'extrémité  $b_i$  de chaque classe et pour ordonnée l'effectif cumulé croissant de la classe correspondant.

**Exemple :**

Prix de vente moyen d'un article constaté dans 120 magasins.

Prix	[20; 25[	[25; 30[	[30; 35[	[35; 40[	[40; 45[
Effectif	24	32	48	12	4
E.C.C.	24	56	104	116	120



### 3 Caractéristiques d'une série statistique

#### 3.1 Caractéristiques de position

##### 3.1.1 mode

Définition :

Le *mode* (ou *classe modale*) est la classe ou la valeur du caractère de plus grand effectif.

**Remarque :**

Le *mode* ou la *classe modale* ne sont pas nécessairement unique pour une série statistique donnée.

### 3.1.2 Médiane

**Définition :**

La *médiane* est la valeur du caractère qui sépare la série statistique en deux groupes de même effectif.

**Exemples :**

- On considère la série : 1 ; 5 ; 6 ; 6 ; 7 ; 8 ; 9 ; 10 ; 10. L'effectif total est 9. La valeur centrale est donc la valeur de rang 5. La médiane est donc 7.
- On considère la série : 1 ; 5 ; 6 ; 6 ; 7 ; 8 ; 9 ; 10. L'effectif total est 8. Il n'y a donc pas une unique valeur centrale mais deux. On choisit donc pour médiane la moyenne entre ces deux valeurs centrales, soit les valeurs de rang 4 et 5. La médiane est donc  $\frac{6+7}{2} = 6,5$ .

**Méthode de détermination :**

- Dans le cas d'un *caractère discret* à  $k = 2p + 1$  valeurs, la médiane est la  $p$ ème valeur ;
- dans le cas d'un *caractère discret* à  $k = 2p$  valeurs, la médiane est la moyenne entre la  $p$ ème et la  $p + 1$ ème valeur c'est à dire  $Me = \frac{x_p + x_{p+1}}{2}$  ;

**Exemple :**

On considère la série suivante :

0	1	2	3	4	5
0	2	5	12	8	3

L'effectif total est  $N = 30$ . Il est pair donc la médiane est la moyenne entre les valeurs de rang  $\frac{N}{2} = 15$  et  $\frac{N}{2} + 1 = 16$ . C'est donc  $\frac{3+3}{2} = 3$ .

**Remarque :**

Dans le cas de valeurs regroupées en classe, on parlera de classe médiane.

### 3.1.3 Première définition des quantiles

Plusieurs définitions des quartiles et déciles sont utilisées, la première présente l'avantage de faire apparaître la médiane comme l'un des quartiles, la deuxième présente l'avantage d'être plus aisée à calculer. L'une des deux est à connaître.

#### Définition :

On appelle *quantiles* de la série *ordonnée*, les nombres  $Q_1$  et  $Q_3$  tels que :

- 25% des valeurs sont inférieures à  $Q_1$  et 75% des valeurs sont supérieures à  $Q_1$  ;
- 25% des valeurs sont supérieures à  $Q_3$  et 75% des valeurs sont inférieures à  $Q_3$ .

#### Détermination pratique :

- $Q_1$  est la médiane de la série partielle constituée des valeurs inférieures à la médiane ;
- $Q_3$  est la médiane de la série partielle constituée des valeurs supérieures à la médiane.

#### Remarque :

On définit de même que les quartiles, les déciles d'une série statistiques qui partagent la série en 10 parties de même effectif.

#### Exemple :

$x_i$	30	45	50	60	61
$n_i$	2	3	2	2	2

L'effectif total est  $n = 11$ .

La médiane  $M_e$  est la  $\frac{11+1}{2} = 6^e$  valeur soit 30.

Le premier quartile  $Q_1$  est la  $\frac{11+1}{4} = 3^e$  valeur soit 45.

Le troisième quartile  $Q_3$  est la  $3 \frac{11+1}{4} = 9^e$  valeur soit 60.

$x_i$	30	45	50	60	61
$n_i$	2	3	2	2	3

L'effectif total est  $n = 12$ .

La médiane  $M_e$  est le centre de l'intervalle  $[30 ; 30]$  soit 50 (6e et 7e valeurs).

$Q_1$  est le centre de l'intervalle  $[30 ; 30]$  soit 45 (3e et 4e valeurs).

$Q_3$  est le centre de l'intervalle  $[60 ; 61]$  soit 60,5 (9e et 10e valeurs).

### 3.1.4 Deuxième définition des quantiles

**Définition :**

- Le *premier quartile* noté  $Q_1$  de la série statistique est la plus petite valeur telle qu'au moins 25% des valeurs lui soient inférieures ou égales ;
- Le *troisième quartile* noté  $Q_3$  de la série est la plus petite valeur telle qu'au moins 75% des valeurs lui soient inférieures ou égales.

**Détermination pratique :** On suppose la série ordonnée dans l'ordre croissant des valeurs du caractère. Soit  $N$  l'effectif total.

- Si  $\frac{N}{4}$  est un entier alors  $Q_1$  est la valeur de rang  $\frac{N}{4}$  et  $Q_3$  est la valeur de rang  $\frac{3N}{4}$  ;
- si  $\frac{N}{4}$  n'est pas entier, alors  $Q_1$  est la valeur dont le rang suit  $\frac{N}{4}$  et  $Q_3$  la valeur dont le rang suit  $\frac{3N}{4}$ .

**Définition :**

Le *n<sup>e</sup> décile* est la plus petite valeur telle que  $n \times 10 \%$  des effectifs au moins lui soient inférieurs ou égaux.

### 3.1.5 Moyenne

**Définition :**

La *moyenne* est donnée par :

$$\frac{n_1x_1 + n_2x_2 + \dots + n_kx_k}{N}$$

ou encore

$$\frac{\sum_{i=1}^k n_i x_i}{N}$$

On dit que la moyenne d'une série statistique est *élaguée* d'une ou de plusieurs valeurs lorsque la moyenne est calculée sans tenir compte de ces valeurs.

**Remarque :**

Dans le cas d'un *caractère continu* (donc de valeurs regroupées en *classes*), on prend pour  $x_i$  le centre  $\frac{a_i+b_i}{2}$  des classes  $[a_i; b_i[$ .

**Propriétés :**

Soit  $\bar{x}$  la *moyenne* d'une série d'effectif total  $N$ , de valeurs  $x_i$  et d'effectifs correspondants  $n_i$  pour  $i$  allant de 1 à  $k$ .

- Si  $\bar{x}_1$  et  $\bar{x}_2$  sont les *moyennes* de deux sous groupes de la série d'effectifs respectifs  $N_1$  et  $N_2$ , alors :

$$\bar{x} = \frac{N_1\bar{x}_1 + N_2\bar{x}_2}{N}$$

- La *moyenne* de la série de valeurs  $ax_i + b$  où  $a$  et  $b$  sont des nombres réels est égale à  $a\bar{x}_i + b$ .

## 3.2 Caractéristiques de dispersion

### 3.2.1 Étendue

**Définition :**

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



### 3.2.2 Écart type

**Définition :**

La *variance* de la série est le nombre noté  $V$  défini par :

$$V = \frac{n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \dots + n_k(x_k - \bar{x})^2}{N}$$

**Propriété :**

$$V = \frac{n_1x_1^2 + n_2x_2^2 + \dots + n_kx_k^2}{N} - \bar{x}^2$$

**Définition :**

L'*écart-type* de la série statistique est le nombre noté  $\sigma$  défini par :

$$\sigma = \sqrt{V}$$

**Propriété :**

Pour tout nombre réel  $x$ ,

$$(\bar{x} - x_1)^2 + \dots + (\bar{x} - x_p)^2 \leq (x - x_1)^2 + \dots + (x - x_p)^2$$

L'écart type minimise donc la distance des valeurs du caractère à la moyenne, c'est donc un indicateur approprié pour mesurer la dispersion des valeurs autour de la moyenne.

### 3.2.3 Écart inter quartiles

**Définition :**

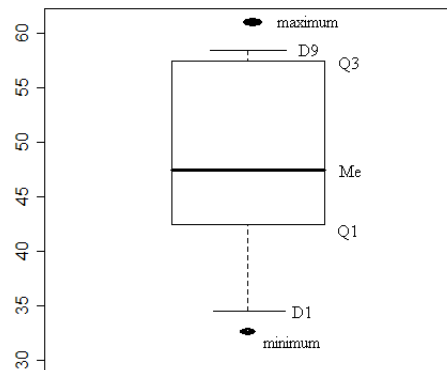
On appelle intervalle inter quartiles l'intervalle  $[Q_1; Q_3]$  et *écart inter quartiles* le nombre  $Q_3 - Q_1$ .

## 4 Diagrammes en boîte

**Définition :**

Un *diagramme* en boîte représente les caractéristiques statistiques suivantes :

- minimum et maximum ;
- médiane  $M_e$  ;
- le premier quartile  $Q_1$  et le troisième quartile  $Q_3$  ;
- le premier décile  $D_1$  et le neuvième décile  $D_9$  ;



**Remarque :**

On fait parfois figurer le minimum et le maximum de la série au lieu des premier et neuvième déciles.

## 5 Tableaux croisés

Utilisation :

Les *tableaux croisés* ou *tableaux à double entrée* permettent de représenter dans un unique tableau deux caractères étudiés d'une même série statistique.

Construction :

Soient  $A$  et  $B$  deux caractères étudiés d'une même série statistique et prenant les valeurs  $A_1, A_2, \dots$  pour  $A$  et  $B_1, B_2, \dots$  pour  $B$ .

	$A_1$	$A_2$	...	Total
$B_1$	$n_{1,1}$	$n_{1,2}$	...	$S'_1$
$B_2$	$n_{2,1}$	$n_{2,2}$	...	$S'_2$
...	...	...	...	...
Total	$S_1$	$S_2$	...	

Définition :

Dans le tableau ci-dessus :

- Les totaux par ligne  $S'_1, S'_2$ , etc. sont appelés *effectifs marginaux* pour le caractère  $B$  ;
- les totaux par colonne  $S_1, S_2$ , etc. sont appelés *effectifs marginaux* pour le caractère  $A$  ;
- la *fréquence conditionnelle* de  $A_1$  sachant  $B_1$  est le nombre noté  $f_{B_1}(A_1)$  défini par  $\frac{n_{1,1}}{S'_1}$  ;
- la *fréquence conditionnelle* de  $B_1$  sachant  $A_1$  est le nombre noté  $f_{A_1}(B_1)$  défini par  $\frac{n_{1,1}}{S_1}$  ;