

# Cours d'introduction à l'analyse statistique

## 3

### Paramètres de position d'une distribution

Les paramètres de position (ou valeurs centrales) sont des valeurs numériques qui « résument » une série statistique en caractérisant l'ordre de grandeur des observations. Ils s'expriment dans la même unité que les observations. Les paramètres de position permettent de situer la position de plusieurs séries comparables. Lorsque la distribution est parfaitement symétrique, mode, moyenne et médiane sont confondues.

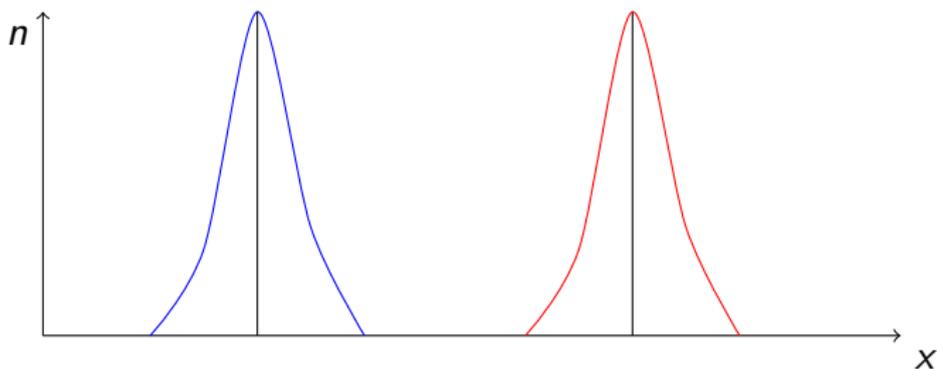


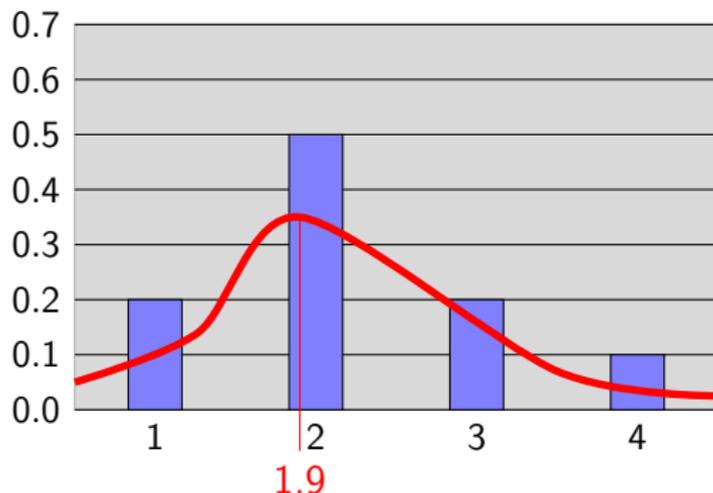
FIGURE: Les deux courbes ont la même allure, mais ne se positionnent pas du tout au même endroit sur l'axe des valeurs (des modalités). Les paramètres de position le mettent clairement en évidence.

# PLAN

- Mode
- Quantiles, quartiles et médiane
- Moyenne

## Le mode, défini pour toute variable aléatoire

Le mode d'une variable qualitative ou quantitative discrète : modalité dont la fréquence (absolue ou relative) est la plus élevée. Dans le cas où une variable continue a été regroupée en classes, le mode est la classe dont la fréquence est la plus élevée.



Dans l'exemple ci-dessus, le mode de la variable discrète est 2, celui de la variable continue, 1.9.

## Cas particulier d'un regroupement de classes

Lorsqu'une distribution est regroupée en classes, Il est fréquent de préférer donner un chiffre à un résultat qualitatif comme un intervalle.

Ainsi, si la distribution est uniforme, on préférera dire que le mode est  $\frac{a_i + b_i}{2}$  plutôt de dire que c'est l'intervalle  $[a_i, b_i]$ . On parle alors de classe modale et de mode.

Revenus	$n_i$	$f_i$
]0,100]	3	0.3
]100,200]	5	0.5
]200,300]	2	0.2
Total	10	1

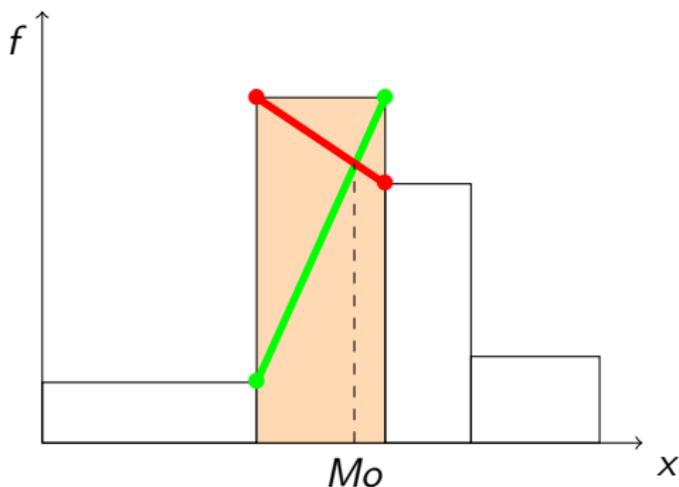
Classe modale ]100,200]  
mode : 150

Revenus	Amplitude	$n_i$	$f_i$
]0,100]	100	3	0.3
]100,150]	50	2	0.2
]150 ;200]	50	3	0.3
]200,300]	100	2	0.2
]Total		10	1

Classe modale ]150,200]  
mode :175

# Détermination graphique du mode d'une classe modale

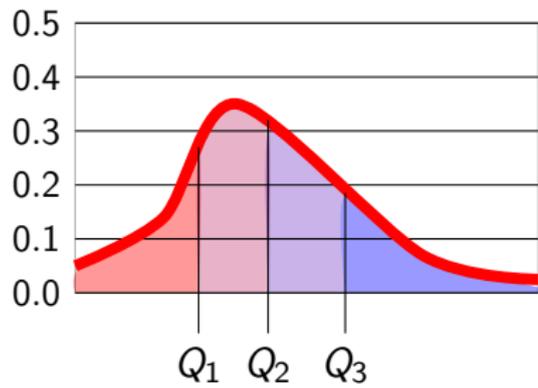
Si on ne suppose pas que la distribution est uniforme dans une classe modale, alors, on peut par convention évaluer graphiquement le mode en joignant les bornes supérieures et les bornes inférieures des classes qui précèdent et succèdent la classe modale, comme le montre le schéma ci-après.



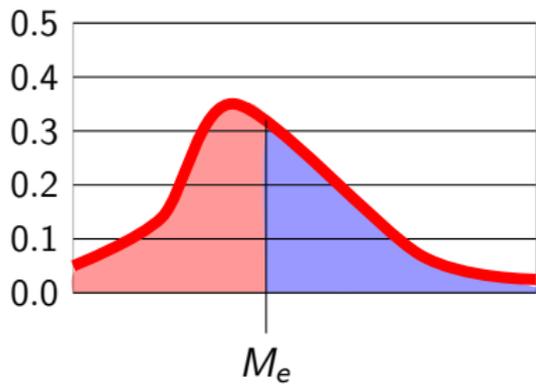
## Les quantiles : séparer une distribution en parts égales

Lorsque la variable est ordonnée, si elle est continue, et parfois même quand elle est discrète ordonnée, on cherche à représenter les différentes parties d'une distribution. On nomme *quantiles* les valeurs qui permettent de séparer la distribution en parts égales. L'opération varie avec le nombre de parts.

Dans le cas d'une séparation en quatre, les *quartiles* sont les valeurs qui partagent la distribution en 4 parties de 25%.



Dans le cas d'une séparation en deux, la *médiane* est la valeur qui partage la distribution en 2.



# Le quantile, défini pour les variable ordonnées

## Definition

les quantiles sont les valeurs de la variable partageant la série classée par ordre croissant de la variable en  $k$  sous-ensembles égaux.

- ▶  $k = 2$  c'est la *médiane*  $M_e$
- ▶  $k = 4$  c'est les *quartiles*  $Q_1, Q_2, Q_3$
- ▶  $k = 10$  c'est les *deciles*  $D_1, D_2, \dots, D_9$
- ▶  $k = 100$  c'est les *centiles*  $C_1, C_2, \dots, C_{99}$

## Calcul du $n^{\text{ième}}$ quantile ( $n < k$ )

- ▶ Classer les données en ordre croissant, et calculer les fréquences cumulées  $F(x)$
- ▶ Si  $\exists x_i / F(x_i) = n/k$  : le  $n^{\text{ième}}$  quantile est  $x_i$ .
- ▶ Si  $\exists x_{i-1}, x_i / F(x_{i-1}) < n/k < F(x_i)$  : le  $n^{\text{ième}}$  quantile est  $x_i$ . On peut parfois considérer l'intervalle  $]x_{i-1}, x_i]$  ou en faire la moyenne de  $x_{i-1}, x_i$  (dans le cas de la médiane, on parle d'intervalle médian).

# Exemple



Modalité	Effectifs	Fréquences	Fréq. cumulées	Qi
2	1	0.1	0.1	
3	3	0.3	0.4	0.25
4	4	0.4	0.8	0.5 ; 0.75
6	2	0.2	1	
Total	10	1		

Dans la pratique, il faut trouver les modalités dont la fréquence cumulée est “juste au-dessus” de 0.25, 0.5, 0.75



Prouver dans l'exemple suivant que le nombre d'enfants median est 2

nb enfants	0	1	2	3	+ de 3
Effectifs	2	1	4	2	1

## Calcul d'une médiane par interpolation

Dans le cas d'une répartition par classes, il arrive fréquemment qu'on ne puisse trouver qu'une classe médiane, c-a-d,

$$[x_{i-1}, x_i] \quad \text{telle que} \quad F(x_{i-1}) < \frac{1}{2} < F(x_i)$$

Deux analyses sont alors possibles

- ▶ soit dire, par convention, que la médiane est  $x_i$
- ▶ soit calculer, sous l'hypothèse que la population est uniformément distribuée dans la classe, la médiane par interpolation linéaire. On trouve alors

$$M_e = \frac{\frac{1}{2} - F(x_{i-1})}{F(x_i) - F(x_{i-1})} x_{i-1} + \frac{F(x_{i-1}) - \frac{1}{2}}{F(x_i) - F(x_{i-1})} x_i$$

**Exemple :** La classe médiane de revenus est la classe  $]100,200]$ .

En supposant que  $F(100) = p(X \leq 100) = 0.3$  et que

$F(200) = p(X \leq 200) = 0.8$ , montrer que l'interpolation linéaire de la médiane est  $M_e = 60$ .

# Moyenne arithmétique d'un ensemble de $N$ nombres

## Définition

La moyenne arithmétique de  $N$  nombres est égale à la somme de ces nombres divisée par leur nombre.

$$\bar{x} = \frac{1}{N} \sum_i x_i$$

## Exemple simple

3 individus, gagnent respectivement 10.000 euros, 20.000 euros et 30.000 euros. La moyenne de leur revenu est 20.000 euros.

## Remarque

La moyenne arithmétique est exactement la quantité qui pourrait être identiquement distribuée à chaque individu. En effet, la conséquence directe de la définition de  $\bar{x}$  est :  $N \bar{x} = \frac{1}{N} \sum_i x_i$ .

## Moyenne arithmétique d'une distribution

Dans le cas d'une distribution, il faut prendre en compte la fréquence d'apparition de chacune des réalisations.

### Cas discret : à partir du tableau de fréquences

Une variable  $X$  prend les valeurs  $x_i$  avec la fréquence  $f_i$  pour  $i = 1, \dots, N$ . La moyenne de cette variable est

$$\bar{X} = \sum_i f_i x_i$$

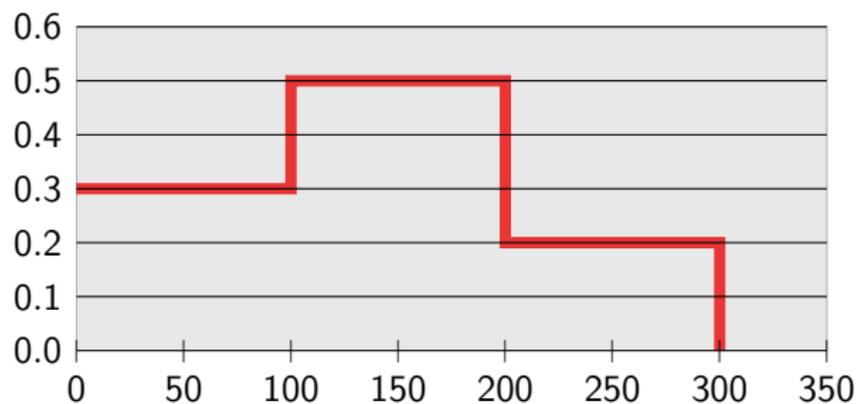
- ▶ la comparaison avec la formule du transparent précédent est immédiate.  $\frac{1}{N}$  est remplacé par la fréquence (individualisée) de chaque réalisation  $f_i$ .

### Cas continu : à partir de la fonction de distribution

Un variable  $X$  est définie par sa fonction de distribution  $f(x)$ , sa moyenne est

$$\bar{X} = \int_{-\infty}^{+\infty} x f(x) dx$$

## Exemple d'analyse d'une distribution.



1. Montrer que le mode est 150
2. Montrer que la médiane est 140
3. Montrer que la moyenne est 140

# Les boîtes à moustaches

La boîte à moustaches résume quelques caractéristiques de position du caractère étudié. Ce diagramme est utilisé principalement pour comparer un même caractère dans deux populations de tailles différentes.

Ce diagramme parfois appelé *diagramme à pattes* comporte 2 éléments :

1. un rectangle allant du premier quartile au troisième quartile et coupé par la médiane
2. des segments aux extrémités menant jusqu'aux valeurs extrêmes, ou jusqu'aux premier et neuvième déciles (D1 / D9)

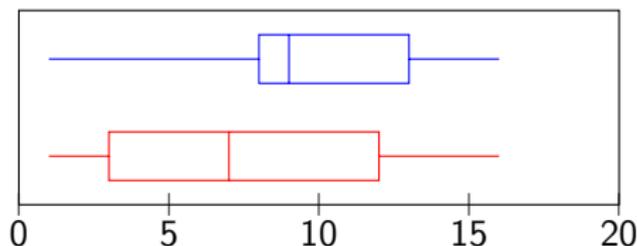


FIGURE: Comparaison de deux diagrammes en boîte à moustaches avec

- pour la boîte rouge :  $Q1 = 3$ ,  $M = 7$ ,  $Q3 = 12$ ,  $D1 = 1$ ,  $D9 = 16$
- pour la boîte bleue :  $Q1 = 7$ ,  $M = 9$ ,  $Q3 = 12$ ,  $D1 = 1$ ,  $D9 = 16$

## Boîtes à moustache et Excel

Dans une première étape, il est nécessaire que pour chacune des variables que vous voulez représenter (ici X,Y, Z), vous ayez fait une table avec les différents éléments constitutifs de la boîte à moustache, comme suit

	X	Y	Z
Minimum			
D1			
Q1			
Médiane			
Q3			
D3			
Maximum			

alors, avec l'assistant graphique, vous pourrez faire des boîtes à moustaches sur un même graphique.

La procédure est expliquée à l'adresse

<http://yallouz.arie.free.fr/excel/excelboxplot.php>

## Petits entraînements

Faire la boîte à moustache du tableau de données individuelles :

1	2	3	4	5	6	7	8	9	10	19	28
---	---	---	---	---	---	---	---	---	----	----	----

Faire la boîte à moustache de l'âge du tableau de données individuelles relatant l'âge des personnes hospitalisées dans un échantillon.

Classes	0-2	2-5	5-10	10-20	20-30	30-50
Effectifs	17	21	57	55	35	15

# Entraînements sur les bases de données

1. Sur FichierExemple1, faire une boîte à moustache des indemnisations au titre du pretium doloris pour chaque degré de pretium doloris (de 1 à 7).