Chapitre 13

Les statistiques

***

I. Médiane et quartiles

1. La médiane

Définition.

La ***médiane*** d'une série de N **valeurs rangées par ordre croissant** est le nombre Me tel que :

1. si N est impair, Me est ***la valeur centrale*** ;
2. si N est pair, Me est la ***demi-somme*** des deux valeurs centrales.

Remarque. Au moins 50 % des valeurs de la série sont inférieures ou égales à la médiane et au moins 50 % des valeurs lui sont supérieures ou égales.

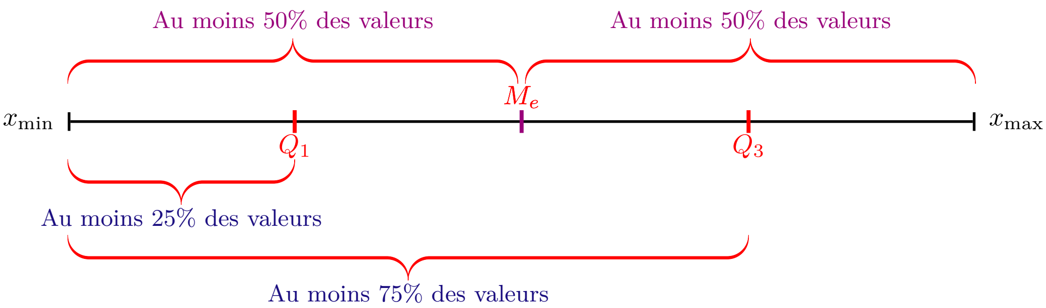
Exercice 1.13. On considère la série statique suivante : 2 ; 5 ; 6 ; 7 ; 8 ; 9 ; 10 ; 12 ; 14 ; 16 ; 17 ; 20.
Calculer la médiane de cette série.

2. Les quartiles

Définition.

On considère une série statistique à une variable x_i ayant N valeurs, ces N valeurs étant rangées en **ordre croissant**.

1. Les **quartiles** partagent la série ordonnée en quatre groupes de même effectif.
2. Le **premier quartile** noté Q_1, est la plus petite **valeur de la série** telle qu'au moins 25 % des valeurs lui sont inférieures ou égales.
3. Le **troisième quartile** noté Q_3, est la plus petite **valeur de la série** telle qu'au moins 75 % des valeurs lui sont inférieures ou égales.



3. Caractéristiques de dispersion

Définition.

1. **L'étendue** est la différence entre la plus grande et la plus petite valeur d'une série statistique.
2. **L'écart inter-quartile** est égal à la différence entre le troisième et le premier quartile.

4. Diagramme en boîte

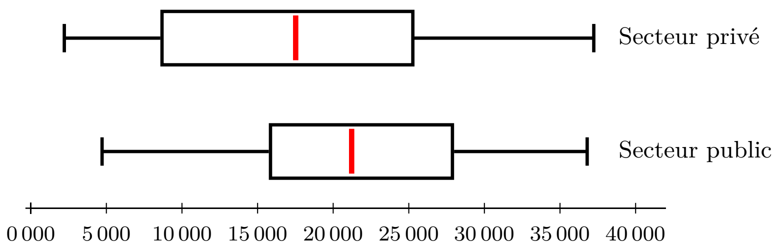
La représentation graphique de la dispersion d'une série statistique se fait à l'aide de diagramme en boîte appelés aussi « boîte à moustaches ».

Pour une catégorie donnée, on construit, en face d'un axe permettant de repérer les quantiles de la variable étudiée, un rectangle dont la longueur est égale à l'écart inter-quartile $Q_3 - Q_1$, la médiane est représentée par un trait. On ajoute alors des segments aux extrémités menant jusqu'aux valeurs extrêmes.

Exemple. Le tableau suivant donne la distribution du revenu salarial par secteur d'activité en France en 2014.

	D1	Q1	Médiane	Q3	D9
Secteur privé	2 218	8 570	17 520	25 377	37 234
Secteur public	4 716	15 744	21 221	27 996	36 797

Source : INSEE



II. Moyenne et écart-type

1. Moyenne

Définition.

Soit une série statistique dont les valeurs du caractère sont $x_1, x_2, \cdots, x_k$ et $n_1, n_2, \cdots, n_k$ effectifs associés.
La moyenne de la série statistique, notée $\bar{x}$, a pour valeur :

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \cdots + n_kx_k}{n_1 + n_2 + \cdots + n_k}$$

Remarque. Lorsqu'on présente la série statistique en ne donnant que la liste des valeurs, alors la moyenne est :

$$\boxed{\frac{x_1 + x_2 + \cdots + x_k}{k}}.$$

2. Variance et écart-type

Considérons deux groupes d'élèves, l'un de dix élèves et l'autre de huit élèves ; leurs notes de mathématiques à un contrôle sont :

Première série

note x_i	1	2	3	17	20
effectif n_i	3	1	1	1	4

Deuxième série

note x_i	8	10	11	12
effectif n_i	1	2	4	1

La moyenne de la première série est : $\dfrac{n_1x_1+\cdots+n_5x_5}{n_1+\cdots+n_5}=\dfrac{105}{10}=10,5.$

La moyenne de la deuxième série est : $\dfrac{84}{8}=10,5.$

Les deux moyennes sont égales ; pourtant, la répartition des notes n'est pas du tout la même.
Il faut donc trouver un moyen de mesurer la dispersion des nombres autour de la moyenne.
Pour cela, on utilise l'écart type.

Définition.

Soit une série statistique donnée par le tableau :

Valeur du caractère	x_1	x_2	$\cdots$	x_p	Total
Effectif	n_1	n_2	$\cdots$	n_p	N

La variance est : $V=\dfrac{n_1x_1^2+n_2x_2^2+\cdots+n_px_p^2}{N}-\overline{x}^2$

L'écart-type est $\sigma=\sqrt{V}$

3. Utilisation du couple moyenne-écart-type

Propriété.

- **La moyenne et l'écart-type** prennent en compte toutes les valeurs de la série et sont, de ce fait, influencés par les valeurs extrêmes.
- **L'écart-type** mesure la dispersion des valeurs autour de la moyenne : plus il est grand, plus les valeurs sont dispersées et moins la moyenne représente de façon significative la série.

Exercice 2.13. Dans le tableau ci-dessous se trouvent les salaires mensuels nets, en milliers d'euros, des employés de deux PME comptant chacune dix salariés :

PME A	1,2	1,2	1,2	1,3	1,4	1,8	2,1	2,2	4	6,6
PME B	1,4	1,5	1,6	1,65	1,9	2,1	2,25	2,7	3	4,9

1. Calculer le salaire moyen mensuel net pour chacune de ces PME. Que constate-t-on ?
2. Calculer l'écart-type de chacune de ces deux séries.
3. Comparer la répartition des salaires dans ces deux entreprises.