

Fiche méthode statistique / Observatoire des rémunérations 2018

1. Indicateur de position

En termes statistiques, un indicateur de position permet de situer, entre elles, les valeurs d'une série statistique. Il en existe plusieurs types :

A. Moyenne

Définition

La **moyenne** d'une série statistique est égale au quotient de la somme de toutes les valeurs de cette série par l'effectif total.

n = effectif total

x_i = valeur de la série

n_i = effectif des valeurs de la série

\bar{x} = moyenne des valeurs de la série

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_px_p}{n} = \frac{1}{n} \sum_{i=1}^p n_i x_i$$

Limite

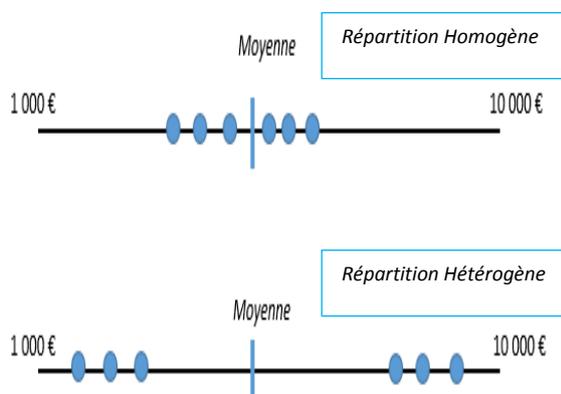


Figure 1: Schéma Répartition des valeurs d'une série statistique

La moyenne donne des informations sur la **tendance générale**.

Elle est influencée par les **valeurs extrêmes** de la série.

En revanche, elle ne donne aucune information sur la répartition des résultats : sont-ils rassemblés autour de la moyenne, ou très dispersés ? Pour deux situations très différentes, la moyenne peut rester la même, comme le montre la *Figure 1*.

B. Médiane

Définition

La **médiane** d'une série statistique dont les valeurs sont rangées suivant l'ordre croissant est la valeur qui partage cette série en deux séries de même effectif.

- Au moins **50 % des individus** de la population ont **une valeur inférieure ou égale à la médiane**
- Au moins **50 % des individus** de la population ont **une valeur supérieure ou égale à la médiane**

Individu	N°1	N°2	N°3	N°4	N°5	N°6
Valeur (Effectif pair)	4	4	7	9	10	
Médiane	7					
Valeur (Effectif impair)	8	9	9	11	12	15
Médiane	$\frac{9 + 11}{2} = \mathbf{10}$					

Tableau 1 : Explication médiane

Soit la série statistique d'effectif total n dont les valeurs sont rangées par ordre croissant.

- Si n est impair, la médiane est la valeur de rang $\frac{n+1}{2}$ de la série ordonnée. C'est la valeur centrale de la série ordonnée. (Voir Tableau 1)
- Si n est pair, la médiane est la demi-somme des valeurs des termes de rangs $\frac{n}{2}$ et $\frac{n}{2} + 1$. (Voir Tableau 1)

Limite

Comme la moyenne, la médiane donne une information de **tendance globale**. Cependant, contrairement à la moyenne, la médiane n'est pas influencée par les valeurs extrêmes.

Dans certains cas, c'est un avantage car elle **écarter les valeurs extrêmes parfois non représentatives** ; dans d'autres cas cela, peut constituer un inconvénient.

C. Les quantiles

Définition

En statistiques, les quantiles sont les valeurs qui divisent un jeu de données en intervalles contenant le même nombre de données. Il y a donc un quantile de moins que le nombre de groupes créés. Il existe plusieurs types de quantile :

- **Le Quartile** est chacune des **3 valeurs qui divisent un jeu de données**, triées

2. Indicateur de Dispersion

Rapport Interquartile et écart Interquartile

Le rapport interquartile est le **rapport des quantiles Q3/Q1**. C'est le nombre sans

selon une relation d'ordre, **en 4 parts égales**, de sorte que chaque partie représente 1/4 de l'échantillon de population. Ainsi, le premier quartile Q1 est la plus petite valeur de la série telle qu'au moins 25 % des valeurs de la série lui soient inférieures. Le troisième quartile Q3 est la plus petite valeur de la série telle qu'au moins 75 % des valeurs de la série lui soient inférieures. La médiane se situe elle entre Q2 et Q3.

- **Le Décile** est chacune des **9 valeurs qui divisent un jeu de données**, triées selon une relation d'ordre, **en 10 parts égales**, de sorte que chaque partie représente 1/10 de l'échantillon de population. Ainsi, le premier décile D1 est la plus petite valeur de la série telle qu'au moins 10 % des valeurs de la série lui soient inférieures. Le neuvième décile D9 est la plus petite valeur de la série telle qu'au moins 90 % des valeurs de la série lui soient inférieures. Le cinquième décile, Q5, correspond à la médiane.

D'autres quantiles existent tels **le centile** (1/100), **le quintile** (1/5) ou **le tercile** (1/3).

L'ensemble de ces quantiles se calcul de la même façon que la médiane

dimension qui donne une mesure relative des écarts entre les 25 % de la distribution les plus bas et les 25 % de la distribution les plus élevés. L'écart interquartiles est lui la simple distance entre le 1^{er} et le 3^{ème} quartile.

3. Représentation Graphique

A. Boite à moustache

Définition

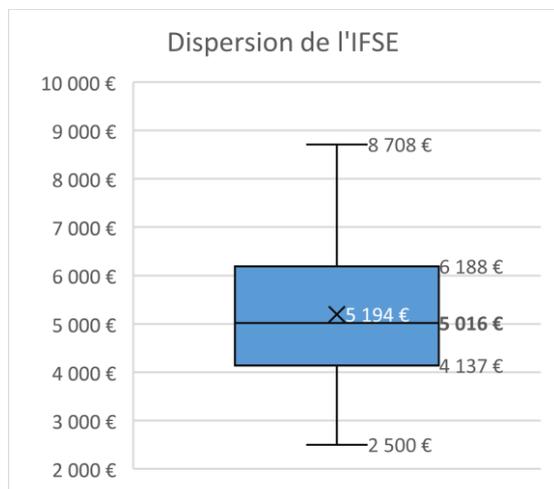


Figure 2 : Graphique de boîte à moustache simple

Une **boîte à moustaches** est, en terme statistique, une représentation graphique (**box plot, en anglais**). Elle est nommée ainsi parce que sa représentation fait penser à des moustaches.

Généralement, la boîte à moustaches permet, pour un jeu de données, de **représenter la médiane, les quartiles et les centiles ou les déciles**. Sont également souvent représentées les données extrêmes. On peut la trouver sous des formes différentes où on représentera le minimum et le maximum de l'échantillon à la place des centiles ou déciles.

Quand utilise-t-on une boîte à moustache ?

On utilise la boîte à moustache lorsque l'on veut représenter graphiquement la distribution des valeurs d'une série statistique ou observer un lien entre une variable quantitative et une variable qualitative.

Comment est-elle construite ?

C'est un graphique composé d'un rectangle duquel deux droites sortent afin de représenter certains éléments des données.

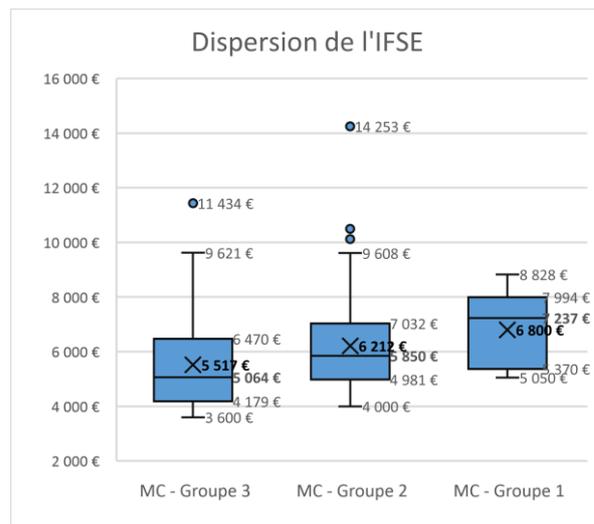


Figure 3: Graphique de boîte à moustache bivarié

- La **valeur centrale** du graphique (droite du milieu de la boîte) est la **médiane** (il existe autant de valeurs supérieures qu'inférieures à cette valeur dans l'échantillon).
- Les **bords du rectangle** sont les **quartiles** (Q1 pour le bord inférieur et Q3 pour le bord supérieur).
- Les **extrémités des moustaches** sont calculées en utilisant 1,5 fois l'**écart interquartile** (la distance entre le 1er et le 3ème quartile).
- Enfin, la **croix** dans la boîte représente la **moyenne** sur l'échantillon étudié.
- Les **valeurs à l'extérieur des moustaches** sont représentées par des points. On ne peut pas affirmer à 100%, qu'une observation à l'extérieur des moustaches, est une valeur aberrante. Par contre, cela indique qu'il faut étudier plus en détail cette observation.

On peut noter que 50% des observations se trouvent à l'intérieur de la boîte.

S'il n'y a pas de « moustache » à l'une ou l'autre des extrémités, cela veut dire qu'il n'y a pas de valeur au-dessus ou en dessous des quartiles Q3 et Q1. Sans oublier que si la boîte à moustache a la forme d'un trait horizontal, cela veut dire qu'il n'y a qu'une seule valeur

dans la série statistique. La boîte à moustache ne peut donc pas être sujet à interprétation.

B. Diagramme en Barres

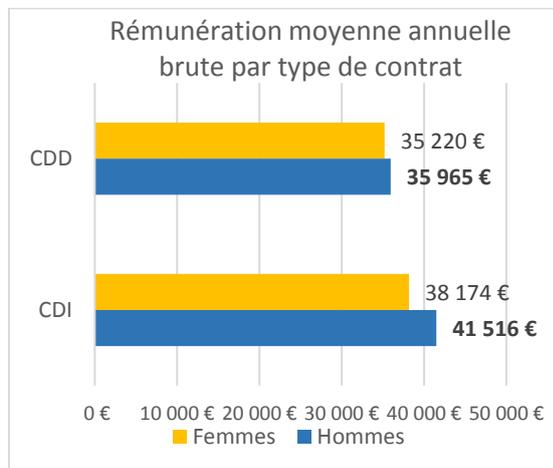


Figure 4 : Diagramme en barre groupé

Un **diagramme en barres** est un graphique qui présente des **variables discrètes** avec des barres rectangulaires avec des hauteurs ou des **longueurs proportionnelles aux valeurs** qu'elles représentent. Les barres peuvent être tracées **verticalement ou horizontalement**.

Selon la position des barres verticales ou horizontales, le diagramme montre, sur un axe, les catégories spécifiques (**variable qualitative**) comparées et sur l'autre axe une valeur mesurée (**variable quantitative discrète**). Certains diagrammes à barres présentent des barres regroupées, indiquant les valeurs de plusieurs variables mesurées (*voir Figure 4*).