Une série statistique peut contenir de très nombreuses données (parfois plusieurs milliers).

Il est donc nécessaire de trouver une façon de résumer ces données.

D'après des données de l'INSEE, *en 2008 un salarié français à temps complet du secteur privé et semi-public gagnait en moyenne 2 068 euros par mois, mais 10 % de ces salariés français à temps complet gagnait un salaire net mensuel inférieur à 1 124 euros*. On peut lire également que *la moitié des salariés a touché moins de 1 653 euros nets*. Ces éléments sont appelés des « indicateurs ».

On différencie les indicateurs de ***position*** et les indicateurs de ***dispersion***.

I. Indicateurs de position

I.1 Moyenne

C'est l'indicateur le plus répandu. On compare souvent une note à la moyenne de la classe.
En réalité, il existe plusieurs types de moyenne : on parlera ici seulement de *moyenne arithmétique*.

DÉFINITION

On considère une série statistique dont les valeurs du caractère sont $x_1, x_2, x_3, \dots, x_p$ et les effectifs associés : $n_1, n_2, n_3, \dots, n_p$. La ***moyenne*** de cette série statistique est :

$$\overline{x}=\frac{n_1\times x_1+n_2\times x_2+\dots+n_p\times x_p}{n_1+n_2+\dots+n_p}.$$

EXEMPLE C1

Un 12 et deux 14 donnent une moyenne de $\frac{1\times 12+2\times 14}{1+2}=\frac{40}{3}\approx 13{,}33$.

EXEMPLE C2

1. Déterminer la moyenne de l'élève qui a obtenu les notes ci-dessous :

Type	interro	interro	DS	DM	interro	DM	interro
Note	15 / 20	10,7 / 20	8 / 20	18 / 20	16 / 20	19 / 20	20 / 20
Coeff.	1	1	**3**	**0,5**	1	**0,5**	1

2. Faire de même avec les notes ci-dessous :

Type	interro	interro	DS	DM	interro	DM	interro
Note	9 / 12	8 / 15	8 / 20	18 / 20	8 / 10	19 / 20	5 / 5
Coeff.	1	1	**3**	**0,5**	1	**0,5**	1

3. Que constatez-vous ? Comment auriez-vous calculé la moyenne de cet élève ?

→ Correction : mathemathieu.fr/1707

REMARQUE : si on note f_i la fréquence de la valeur x_i , alors $\overline{x}=f_1x_1+f_2x_2+\dots+f_px_p$.
Par exemple, si 2 personnes ont 16 ans alors que 8 autres ont 15 ans, pour obtenir l'âge moyen du groupe on peut calculer $\frac{2}{10}\times 16+\frac{8}{10}\times 15$ au lieu de faire $\frac{2\times 16+8\times 15}{2+8}$.

I.2 Médiane, quartiles et déciles

DÉFINITIONS

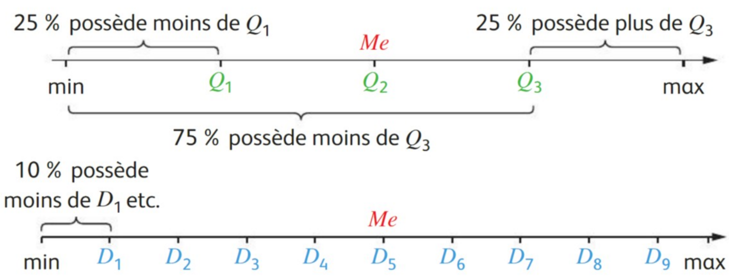
1. Une ***médiane*** d'une série statistique est un nombre, noté ***Me***, tel que :

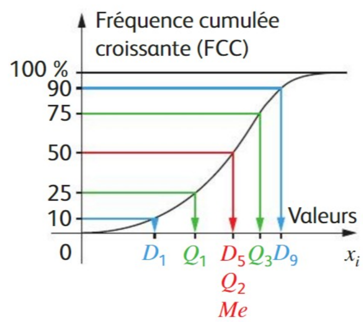
- au moins 25 % des individus ont une valeur du caractère inférieure ou égale à ***Me*** ;
- au moins 25 % des individus ont une valeur supérieure ou égale à ***Me***.

2. Le ***premier quartile*** d'une série statistique, noté Q_1, est la plus petite valeur telle qu'au moins 25 % des données lui soient inférieures ou égales.

Le ***troisième quartile*** d'une série statistique, noté Q_3, est la plus petite valeur telle qu'au moins 75 % des données lui soient inférieures ou égales.

3. Pour k allant de 1 à 9, le ***k*-ième décile**, noté D_k, est la plus petite valeur telle qu'au moins $10k$ % lui soient inférieures ou égales.





II. Indicateurs de dispersion

DÉFINITIONS

On appelle $[Q_1 ; Q_3]$ l'***intervalle interquartile***, et $Q_3 - Q_1$ est l'***écart interquartile***.

On appelle ***rapport interdécile*** le rapport entre le 9^e^ et le 1^er^ décile : $\dfrac{D_9}{D_1}$.

DÉFINITIONS

On considère une série statistique dont les valeurs du caractère sont $x_1, x_2, x_3, \dots, x_p$ et les effectifs associés : $n_1, n_2, n_3, \dots, n_p$.

La ***variance*** est la moyenne des carrés des écarts des valeurs à la moyenne $\overline{x}$:

$$V = \frac{n_1(x_1 - \overline{x})^2 + n_2(x_2 - \overline{x})^2 + \dots + n_p(x_2 - \overline{x})^2}{n_1 + n_2 + \dots + n_p} .$$

L'***écart-type*** est : $\sigma = \sqrt{V}$.

Dans la pratique, on préfère l'écart-type à la variance car il peut être comparé à l'ordre de grandeur des valeurs, ce qui n'est pas le cas de la variance.

L'écart-type tient compte des écarts de toutes les valeurs à la moyenne, donnant ainsi beaucoup de poids aux valeurs extrêmes.

Plus un écart-type est grand, plus les valeurs sont dispersées autour de la moyenne.

PROPRIÉTÉ

La variance est la moyenne des carrés diminuée du carré de la moyenne :

$$V=\frac{n_1 x_1^2+n_2 x_2^2+\ldots+n_p x_p^2}{n_1+n_2+\ldots+n_p}-\overline{x}^2 .$$

Démonstration :

EXEMPLE C3

Les 31 élèves d'une classe ont obtenu les notes suivantes à un contrôle de mathématiques :

Notes : x_i	7	8	9	10	11	12	13	14
Effectif : n_i	1	5	4	12	5	3	0	1

Calculer la moyenne $\overline{x}$ de cette série, puis sa variance V et son écart-type σ.
Déterminer également l'intervalle interquartile et le rapport interdécile.

EXEMPLE A1

p. 157 capacité 1

On s'intéresse à la répartition par tranche d'âges de la population en France et en Allemagne en 2019.

1. On donne ci-dessous le tableau de la répartition de la population allemande.

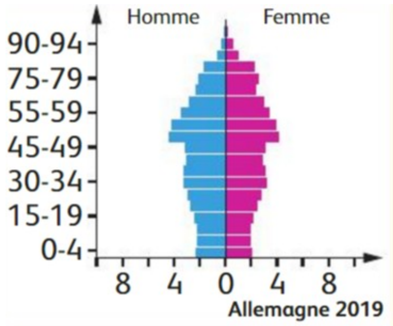
Âge	[0 ; 10[	[10 ; 20[	[20 ; 30[	[30 ; 40[	[40 ; 50[	[50 ; 60[
Effectif	6 886 516	7 177 289	8 825 857	10 416 224	9 866 297	13 574 955
Âge	[60 ; 70[	[70 ; 80[	[80 ; 90[	[90 ; 100[	[100 ; +∞[	TOTAL
Effectif	10 445 254	7 683 008	4 688 273	897 647	13 755	80 475 075

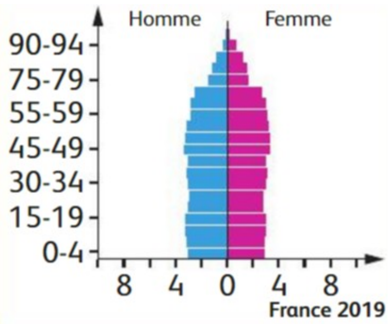
source : population.pyramid.net

a. Déterminer l'âge moyen $\overline{x_A}$ de la population allemande.

b. A l'aide de la calculatrice, déterminer l'écart-type σ_A de cette série (arrondir à 10^{-1}).

2. La population française est estimée à 65 466 334 habitants au 1er janvier 2019. L'âge moyen $\overline{x_F}$ de la population française est estimé à environ 42 ans avec un écart-type $\sigma_F \approx 24{,}5$. On donne de plus les pyramides des âges de chacune des deux populations en 2019.



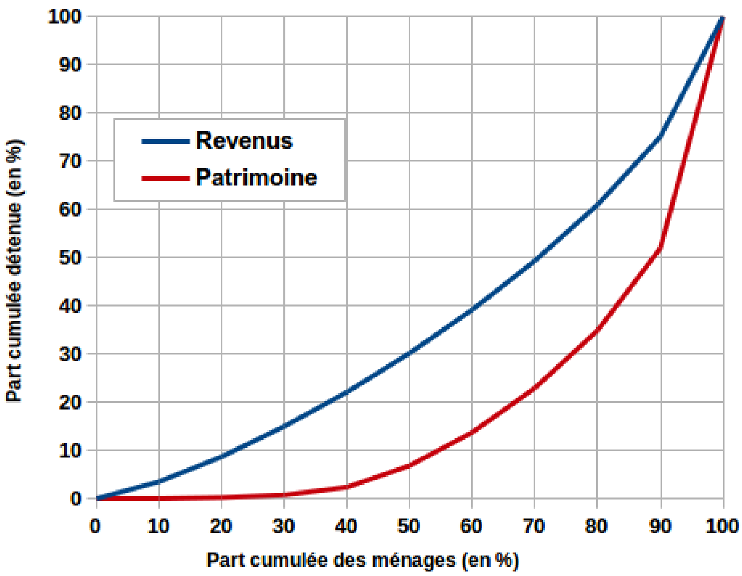


Comparer les deux séries.

III. Courbe de Lorenz et indice de Gini

En 1905, l'économiste américain **Max Otto Lorenz** (1876 - 1959) prépare son doctorat sur *La théorie économique des prix de chemin de fer* qu'il obtiendra un an plus tard, et publie un article qui décrit ce qu'on appellera plus tard ***une courbe de Lorenz*** : une représentation graphique d'une fonction qui, à la part x des détenteurs d'une part d'une grandeur, associe la part y de la grandeur détenue.

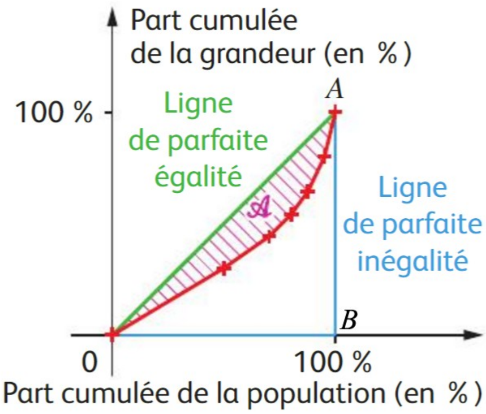
Par exemple :



← courbe de Lorenz des revenus et du patrimoine en France, en 2010

Mathématiquement, une courbe de Lorenz est la représentation graphique d'une fonction f vérifiant :

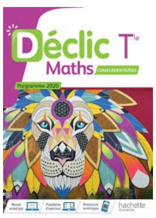
- f est définie sur $[0;1]$
- f est croissante et convexe sur $[0;1]$
- $f(0)=0$ et $f(1)=1$
- $\forall x \in [0;1],\ f(x) \leqslant x$.



La distribution des revenus est « parfaitement égalitaire » si tous les ménages reçoivent le même revenu : une répartition égalitaire est donc représentée par la première bissectrice du repère, la droite d'équation $y=x$.
De même, on parle de distribution « parfaitement inégalitaire » si un ménage accapare le revenu total : la fonction associée prend la valeur $y=0$ pour tout x < 100 %, et $y=100$ % quand x = 100 %.

Le ***coefficient de Gini***[1] (ou ***indice de Gini***) est égal au double de l'aire de la partie délimitée par la courbe de Lorenz et la première bissectrice.
Autrement dit, c'est aussi le rapport de cette aire et de l'aire du triangle OBA (d'aire 1/2).
Ce nombre varie donc entre 0 (égalité parfaite) et 1 (inégalité maximale), et permet de rendre compte de la répartition d'une variable (salaire, revenus, patrimoine) au sein d'une population, en mesurant le niveau d'inégalité de la répartition d'une variable dans la population.
Typiquement, il est utilisé pour mesurer l'inégalité des revenus dans un pays.
Attention cependant : cet indicateur ne fait pas de différence entre une inégalité dans les bas revenus et une inégalité dans les hauts revenus.

→ BILAN DU CHAPITRE & TRAVAIL EN AUTONOMIE ←



• Fiche bilan → p.160

• QCM 19 questions corrigées → p.161

1 Corrado Gini (1884 - 1965)