Chapitre 10 - Répartition des richesses, Inégalités

On s'intéressera dans tout ce chapitre à la répartition des salaires nets mensuels en France, en euros, en équivalent temps plein, pour l'année 2016.

Salaires mensuels nets (en euros)	Centre de la classe	Effectifs (en EQTP)	Effectifs cumulés croissants	Fréquences	Fréquences cumulées croissantes
Moins de 1 500		4 296 589			
De 1 500 à 2 000		4 993 421			
De 2 000 à 2 500		2 782 224			
De 2 500 à 3 000		1 521 004			
De 3 000 à 3 500		923 574			
De 3 500 à 4 000		574 509			
De 4 000 à 4 500		375 296			
De 4 500 à 5 000		253 239			
De 5 000 à 5 500		177 111			
De 5 500 à 6 000		126 844			
De 6 000 à 6 500		94 436			
De 6 500 à 7 000		70 258			
De 7 000 à 7 500		65 928			
De 7 500 à 8 000		41 529			

Compléter le tableau.

I - Statistiques descriptives

1. Médiane

Définition

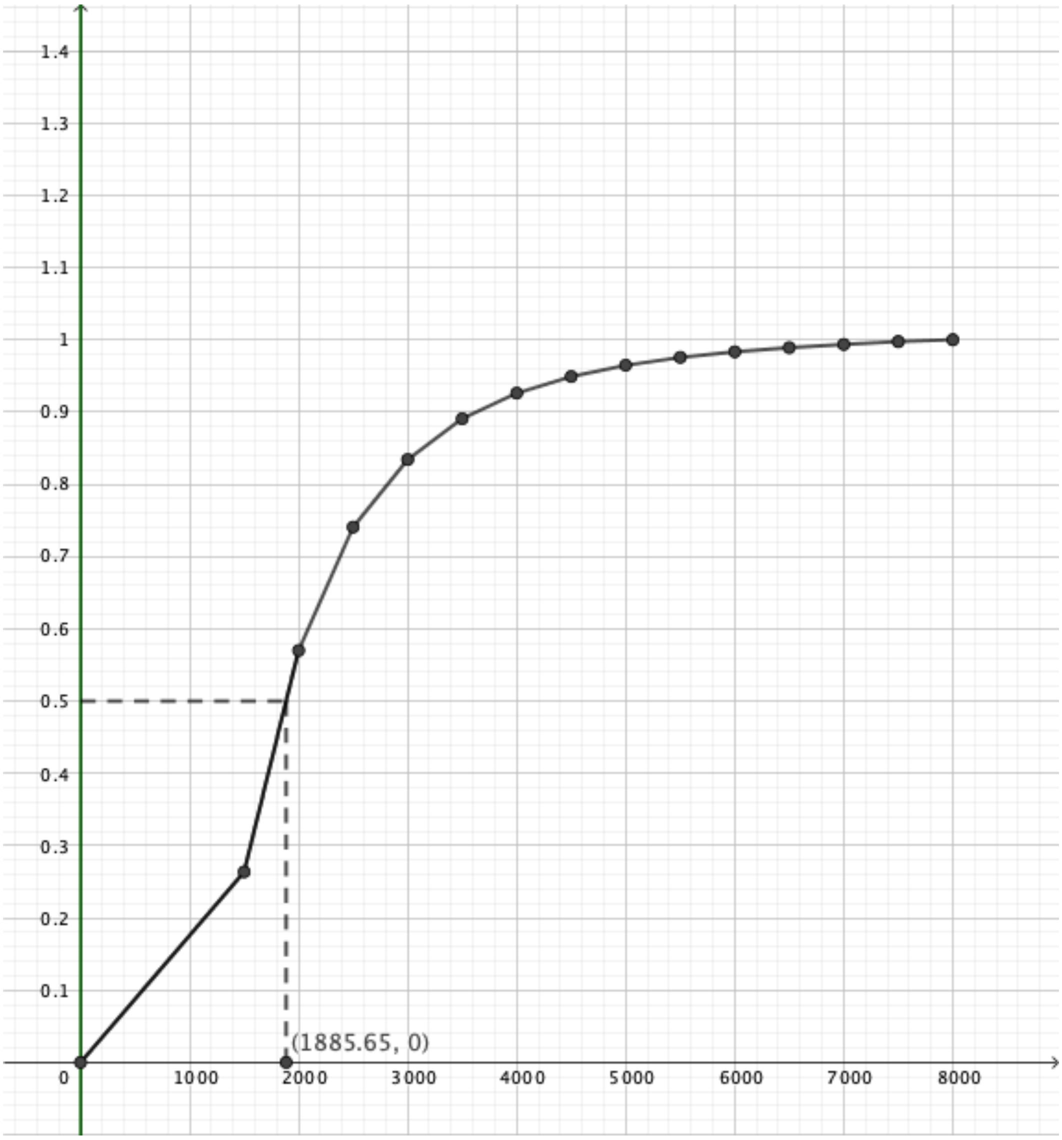
- Si la série est de taille impaire ($N = 2n + 1$), alors la médiane est la donnée de rang $n + 1$.
- Si la série est de taille paire ($N = 2n$), alors la médiane est la demi-somme des données de rang n et de rang $n + 1$.

Remarque

La médiane est un paramètre de tendance centrale.

Exemple

1. Donner la classe médiane des salires nets mensuels en 2016.
2. Déterminer le salaire médian français en 2016.

> 1. Il y a un effectif total de 16 295 962.
>   La classe médiane correspond à la classe de la 8 147 981 ème personne soit $[1500; 2000]$
> 2. On peut tracer la courbe des fréquences cumulées croissantes.
>
> 
>
>
> La médiane de cette série statistique est $1885,65€$.

2. Quartiles

Définition

Le premier quartile Q_1 de la série statistique est la plus petite donnée de la série qui partage la population en deux parties de telle sorte qu'**au moins 25%** des individus prennent une valeur inférieure ou égale à Q_1.

Définition

Le troisième quartile Q_3 de la série statistique est la plus petite donnée de la série partage la population en deux parties de telle sorte qu'**au moins 75%** des individus prennent une valeur inférieure ou égale à Q_3.

Définition

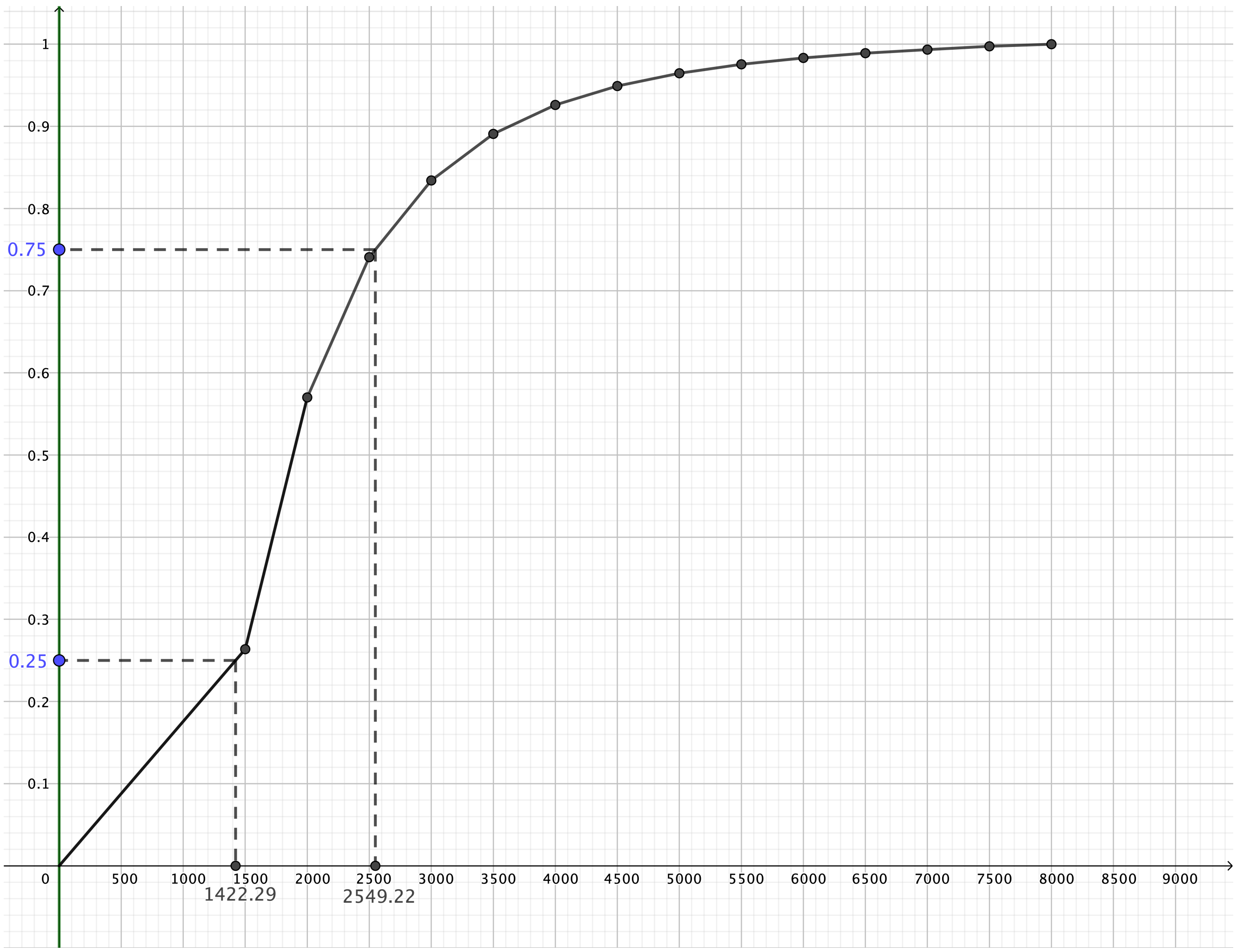
L'écart interquartile est le nombre $Q_3 - Q_1$.

Remarque

Les quartiles sont des paramètres de position et l'écart interquartile est un paramètre de dispersion.

Exemple

Déterminer les 1er et 3ème quartiles, ainsi que l'écart interquartile.

> On peut tracer la courbe des fréquences cumulées croissantes.
>
> 
>
>
> Le premier quartile de cette série statistique est $Q_1 = 1422,29€$ et le troisième quartile est $Q_3 = 2549,22$. (on peut trouver ces résultats par le calcul: on trouve les équations des "segments" correspondants à Q_1 et Q_3 puis on trouve les antécédants de $0,25$ et $0,75$)
>
> L'écart interquartile est $Q_3 - Q_1 = 2549,22 - 1422,29 = 1126,93$.

3. Déciles

Définition

Soit $k \in [\![1;9]\!]$. Le k-ième décile D_k de la série statistique est la plus petite donnée de la série qui partage la population en deux parties de telle sorte qu'**au moins** $k \times 10$ des individus prennent une valeur inférieure ou égale à D_k.

Définition

L'écart interquartile est le nombre $D9 - D_1$.

Définition

Le rapport interdécile est le nombre $\frac{D_9}{D_1}$.

Remarque

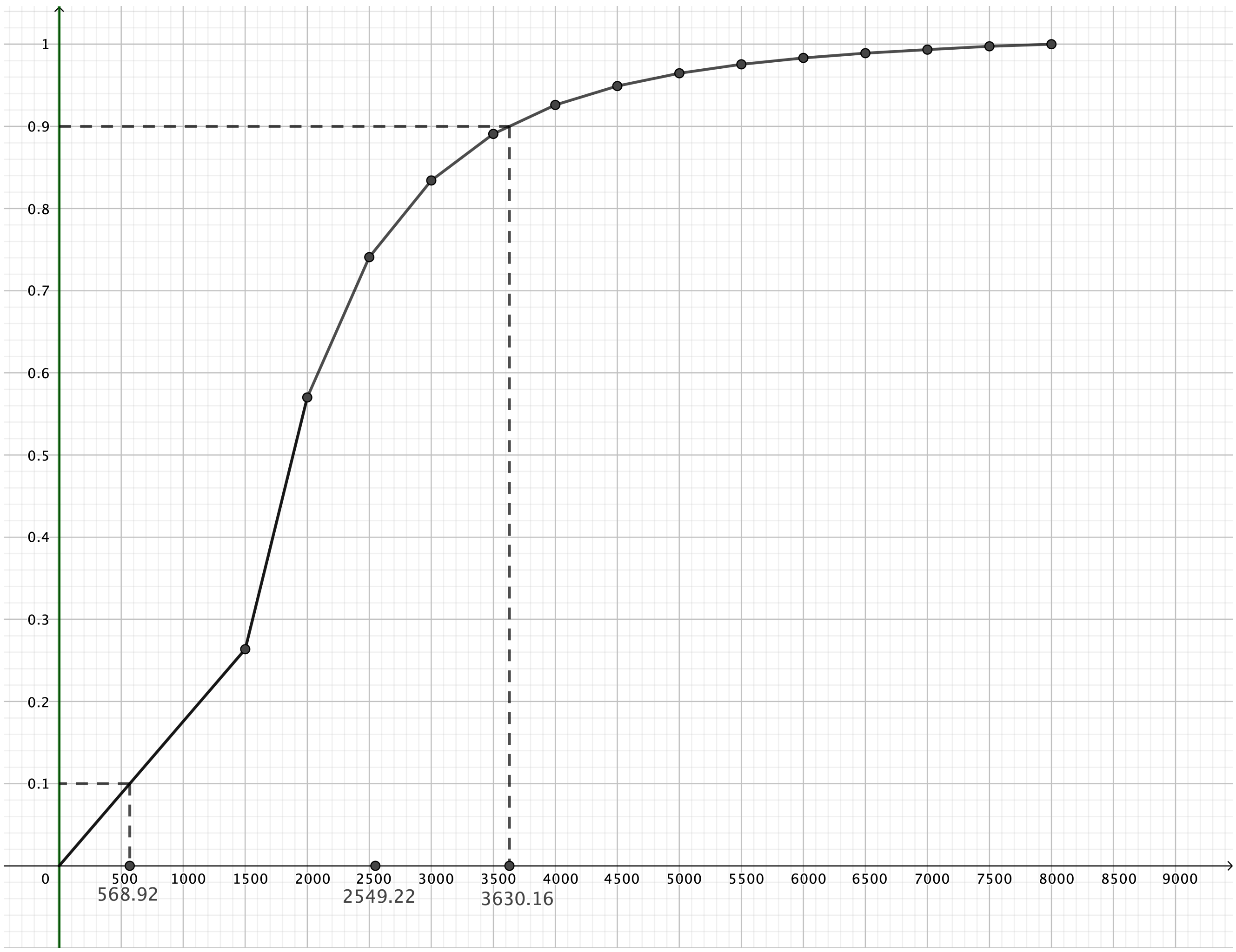
Les déciles sont des paramètres de position et l'écart interdécile est un paramètre de dispersion.

Le rapport interdécile met en évidence le rapport entre le haut et le bas de la distribution. C'est un indicateur d'inégalité de répartition. Plus le rapport est éloigné de 1 plus la répartition est inégale. Plus le rapport se rapproche de 1, plus la répartition est homogène.

Exemple

Déterminer les 1er et 9ème déciles,, approchée par défaut à l'euro, ainsi que l'écart interdécile et le rapport interdécile.

> > On peut tracer la courbe des fréquences cumulées croissantes.



Le premier décile de cette série statistique est $D_1 = 568€$ et le neuvième décile est $D_9 = 3630.$

L'écart interdécile est $D_9 - D_1 = 3630 - 568 = 3062.$

Le rapport interdécile est $\frac{D_9}{D_1} = \frac{3630}{568} \approx 6,39$

On peut déduire du rapport interdécile que la répartition salariale est inégale.

4. Diagramme en boîte

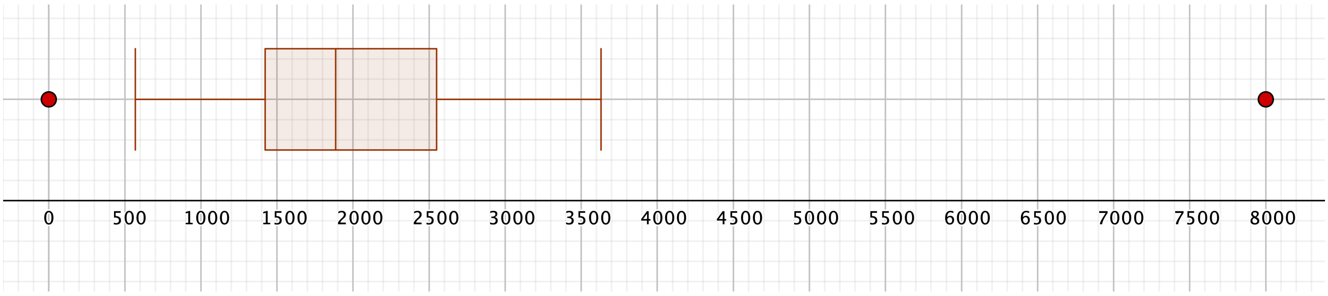
On peut donc résumer une série statistique à caractères quantitatifs par la médiane, les quartiles (et donc l'écart interquartile).

Tout ceci peut se traduire par un diagramme en boîte.
Pour tracer un diagramme en boîte, il faut :

- Tracer un axe horizontal.
- Placer les valeurs maximale et minimale de la série. Elles sont représentées par un point.
- Placer Q*1 et Q*3. Ils sont représentés par un grand trait vertical.
- Placer D*1 et D*3. Ils sont représentés par un petit trait vertical.
- On trace le rectangle dont les petits côtés sont les quartiles.
- On trace les segments qui joignent les petits côté du rectangle et les déciles.
- Placer la médiane Me. Elle est représentée par un grand trait vertical à l'intérieur du rectangle.

Exemple

Tracer le diagramme en boîte de la série statistique.



5. Moyenne

Définition

Soit la série statistique ci-dessous :

Valeurs	x_1	x_2	**...**	x_p
Effectifs	n_1	n_2	...	n_p

La moyenne arithmétique de cette série statistique est le nombre :

$$\overline{x} = \frac{n_1x_1+n_2x_2+\cdots+n_px_p}{n_1+n_2+\cdots+n_p} = \frac{\sum_{i=1}^{p} n_ix_i}{\sum_{i=1}^{p} n_i}$$

Remarque

La moyenne est un paramètre de tendance centrale.

Exemple

Calculer la moyenne de la série statistique.

> Pour calculer la moyenne, on utilise le centre de chaque classe.
>
> $\overline{x} \approx 2079,11$

6. Variance et écart type

Définition

Soit la série statistique $(x_1, x_2, \dots, x_p)$ et les effectifs correspondant $(n_1, n_2, \dots, n_p)$ de moyenne $\overline{x}$.

La variance V de la série statistique est la moyenne des carrés des écarts entre chaque valeur et la moyenne $\overline{x}$.

$$V=\frac{n_1\left(x_1-\bar{x}\right)^2+n_2\left(x_2-\bar{x}\right)^2+\ldots+n_p\left(x_p-\bar{x}\right)^2}{n_1+n_2+\ldots+n_p}=\frac{\sum_{i=1}^{p} n_i\left(x_i-\bar{x}\right)^2}{\sum_{i=1}^{p} n_i}$$

Remarque

On a aussi $V=\frac{n_1x_1^2+n_2x_2^2+\ldots+n_px_p^2}{n_1+n_2+\ldots+n_p}-\bar{x}^2=\frac{\sum_{i=1}^{p} n_ix_i^2}{\sum_{i=1}^{p} n_i}-\bar{x}^2$

Démonstration

Pour nous simplifier les calculs, posons N l'effectif total. $N=\sum_{i=1}^{p} n_i=n_1+n_2+\ldots+n_p$

Ainsi

$$V=\frac{\sum_{i=1}^{p} n_i\left(x_i-\bar{x}\right)^2}{N}=\frac{\sum_{i=1}^{p} n_ix_i^2-2n_ix_i\bar{x}+n_i\bar{x}^2}{N}=\frac{\sum_{i=1}^{p} n_ix_i^2-\sum_{i=1}^{p} 2n_ix_i\bar{x}+\sum_{i=1}^{p} n_i\bar{x}^2}{N}=\frac{\sum_{i=1}^{p} n_ix_i^2-2\bar{x}\sum_{i=1}^{p} n_ix_i+\bar{x}^2\sum_{i=1}^{p} n_i}{N}$$

$$V=\frac{\sum_{i=1}^{p} n_ix_i^2}{N}-2\bar{x}\frac{\sum_{i=1}^{p} n_ix_i}{N}+\bar{x}^2\frac{\sum_{i=1}^{p} n_i}{N}=\frac{\sum_{i=1}^{p} n_ix_i^2}{N}-2\bar{x}\times\bar{x}+\bar{x}^2\times\frac{N}{N}=\frac{\sum_{i=1}^{p} n_ix_i^2}{N}-2\bar{x}^2+\bar{x}^2=\frac{\sum_{i=1}^{p} n_ix_i^2}{\sum_{i=1}^{p} n_i}-\bar{x}^2$$

Remarque

Avec les fréquences, on a :

$V=f_1(x_1-\bar{x})^2+f_2(x_2-\bar{x})^2+\cdots+f_p(x_p-\bar{x})^2$ ou $V=f_1x_1^2+f_2x_2^2+\cdots+f_px_p^2-\bar{x}^2$

Définition

Afin de mesurer la dispersion avec la même unité que les valeurs de la série, on définit l'écart-type de la série par $\sigma=\sqrt{V}$.

Remarque

La variance et l'écart type sont des paramètres de dispersion.

Exemple

Calculer la variance et l'écart type de la série statistique.

> $V=0,26\times(750-2079,11)^2+0,31\times(1750-2079,11)^2+\ldots+0,0025\times(7750-2079,11)^2=1540637,47$

> > $\sigma = \sqrt{V} = 1241,22$

7. Résumé d'une série statistique

On résume souvent une série statistique par un paramètre de tendance centrale associée à un paramètre de dispersion (ou plus). Deux choix sont proposés :

- le couple (moyenne ; écart-type) qui a l'inconvénient d'associer deux paramètres sensibles aux valeurs extrêmes.
- le triplet (médiane ; écart interquartile; écart interdécile) qui n'a pas ce défaut mais dont la détermination est moins pratique.

Remarque

La calculatrice permet de voir la plupart des résultats de ce cours.

`Stats/Edit/Edite`

On rentre les valeurs dans L_1 et si besoin les coefficients correspondants dans L_2.

`Stats/Calc/Stat 1-Var` puis L_1 ,L_2

(on met `Stat 1-Var` L_1 s'il n'y a pas de coefficients dans L_2)

S'affichent alors les informations concernant la série statistique, à savoir, dans l'ordre :

- la moyenne
- la somme des termes
- la somme des carrés des termes
- moyenne des écart type des échantillons (inutile pour nous)
- l'écart type de la population
- l'effectif total
- la valeur minimale
- le premier quartile (peut être faux sur les TI-82)
- la médiane
- le troisième quartile (peut être faux sur les TI-82)
- la valeur maximale

II - Outils d'analyse statistique

1. Courbe de Lorenz

Définition

Une courbe de Lorenz est une représentation graphique qui met en relation la proportion $x\%$ d'une population détentrice d'une part d'une grandeur à la part $y\%$ de la grandeur détenue.

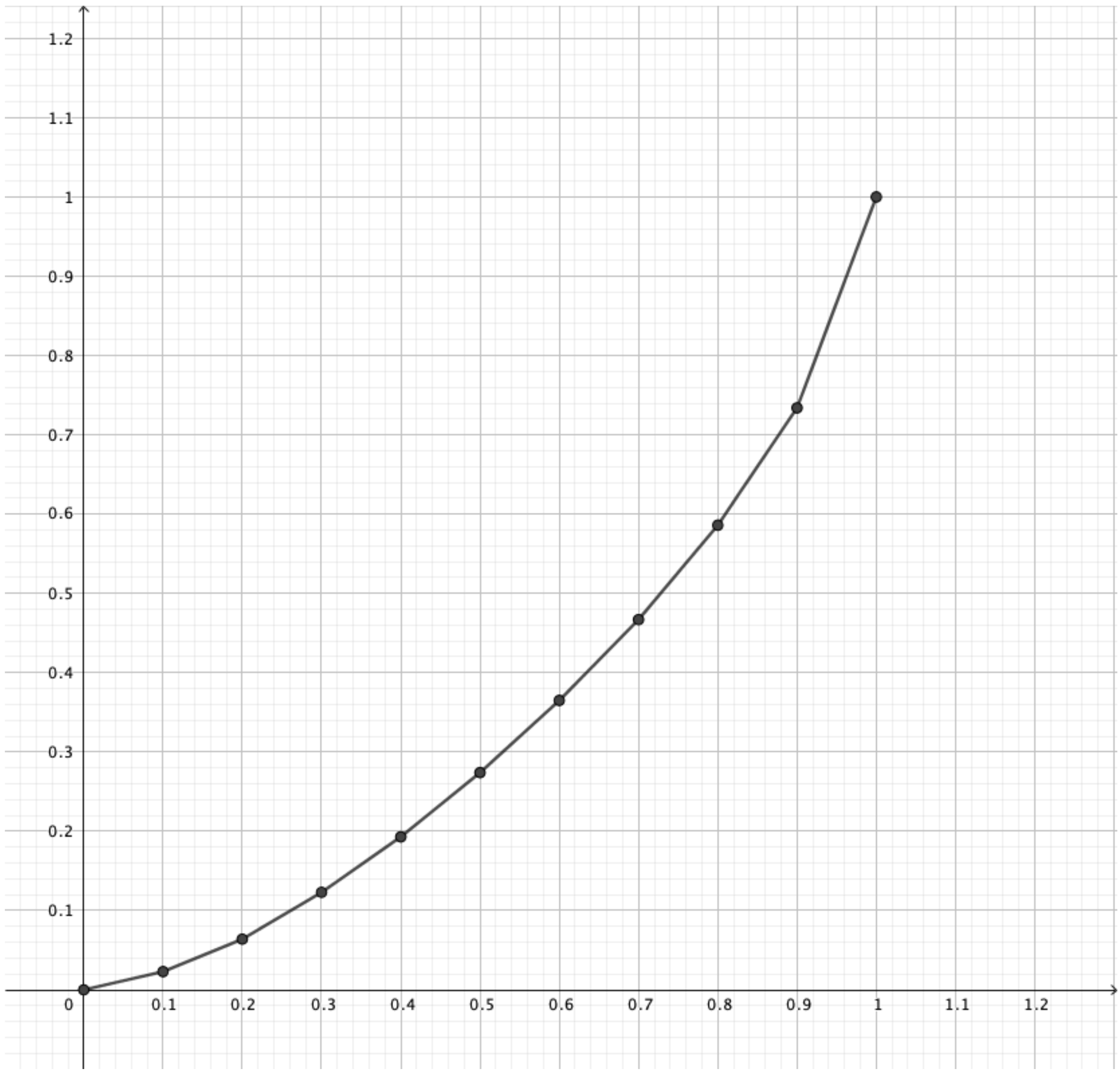
Exemple

1. Déterminer les 9 déciles de notre série statistique, approchée par défaut à l'euro.
2. Compléter le tableau ci-dessous. *Arrondir au dixième de pourcent.*

Tranche de revenu annuel disponible	Revenu annuel moyen	Part détenue (en%)	Part détenue cumulée (en%
Inf. à D_1	476		
De D_1 à D_2	857		
De D_2 à D_3	1208		
De D_3 à D_4	1465		
De D_4 à D_5	1675		
De D_5 à D_6	1881		
De D_6 à D_7	2130		
De D_7 à D_8	2473		
De D_8 à D_9	3051		
Sup. à D_9	5526		

1. Tracer la courbe de Lorenz correspondant à notre série statistique.

> 1. $D_1 = 5689$, $D_2 = 1137$, $D_3 = 1559$, $D_4 = 1722$, $D_5 = 1885$, $D_6 = 2087$, $D_7 = 2380$, $D_8 = 2817$, $D_9 = 3630$.
> 2. Voici le tableau complété:
>
> | Tranche de revenu annuel disponible | Revenu annuel moyen | Part détenue (en%) | Part détenue cumulée (en% |
> |---|---|---|---|
> | Inf. à D_1 | 476 | $2,3$ | $2,3$ |
> | De D_1 à D_2 | 857 | $4,1$ | $6,4$ |
> | De D_2 à D_3 | 1208 | $5,8$ | $12,3$ |
> | De D_3 à D_4 | 1465 | $7,1$ | $19,3$ |
> | De D_4 à D_5 | 1675 | $8,1$ | $27,4$ |
> | De D_5 à D_6 | 1881 | $9,1$ | $36,5$ |
> | De D_6 à D_7 | 2130 | $10,3$ | $46,7$ |
> | De D_7 à D_8 | 2473 | $11,9$ | $58,6$ |
> | De D_8 à D_9 | 3051 | $14,7$ | $73,4$ |
> | Sup. à D_9 | 5526 | $26,6$ | 100 |
>
> 3. Voici la courbe de Lorenz



2. Indice de Gini

Définition

L'indice de Gini, noté G, est égal au double de l'aire $\mathscr{A}$ de la partie délimitée par la courbe $\mathscr{C}$ de Lorenz et le segment $[OA]$ où O est l'origine du repère et A le point de corrdonnées $(1;1)$.

$G = 2\mathscr{A}$

Remarque

L'indice de Gini varie entre 0 (égalité parfaite) et 1(inégalité maximale).

Plus les inégalités sont importantes et plus la courbe de Lorenz s'éloigne du segment $[OA]$, et plus l'indice de Gini se rapproche de 1.

Exemple

On peut approximer la courbe de Lorenz par la courbe représentative de la fonction f définie sur $[0; 1]$ par $f(x) = 0,65x^3 + 0,35x$

1. Calculer l'indice de Gini de notre série statistique.
2. Interpréter le résultat.

> 1. $G = 2\mathscr{A} = 2 \times \left[\frac{1}{2} - \left(\frac{0,023 \times 0,1}{2} + \frac{(0,023+0,064) \times 0,1}{2} + \ldots + \frac{(0,734+1) \times 0,1}{2}\right)\right]$
> $G = 1 - (0,023 \times 0,1 + (0,023 + 0,064) \times 0,1 + \ldots + (0,734 + 1) \times 0,1)$
> $G = 1 - 0,67 = 0,33$
> 2. L'indice de Gini est relativement faible donc les inégalités salariales ne sont pas trop importantes.

III - Quelques problèmes

1. Courbe de Lorenz et indice de Gini

L'Observatoire des inégalités considère que :

- la classe moyenne est constituée des personnes situées entre les 30 % les plus pauvres et les 20 % les plus riches
- une personne est qualifiée de« pauvre » lorsque son revenu est inférieur au seuil de pauvreté, égal à la moitié du revenu médian
- une personne est qualifiée de « riche » lorsque son revenu est supérieur au seuil de richesse, égal au double du revenu médian.

Le tableau suivant donne la répartition des revenus disponibles par ménage selon la tranche de revenu pour l'année 2015.

Tranche de revenu annuel disponible	Limite supérieure de tranche (décile)	Revenu annuel moyen	Part détenue (en%)
Inf. à D_1	13 630	10 030	2,8
De D_1 à D_2	17 470	15 630	4,3
De D_2 à D_3	21120	19 280	5,3
De D_3 à D_4	25 390	23 210	6,4
De D_4 à D_5	30 040	27 680	7,6
De D_5 à D_6	35 060	32 470	8,9
De D_6 à D_7	41 290	38 080	10,5
De D_7 à D_8	49 350	45 070	12,4
De D_8 à D_9	63 210	55 300	15,2
Sup. à D_9	///	96 240	26,5

La colonne « Part détenue (en %) » contient, pour chaque tranche, la part que représente la totalité des revenus disponibles de la tranche par rapport à l'ensemble de tous les revenus disponibles.

Partie A Analyse des données

1. Lire le revenu disponible annuel médian.
   En déduire les valeurs annuelles, puis mensuelles du seuil de pauvreté et du seuil de richesse, en euro.
2. Préciser la fourchette des revenus disponibles annuels, puis mensuels pour la classe moyenne.
3. Calculer le rapport interdécile. Interpréter.
4. a. Représenter les données en plaçant:

   - en abscisses, la part cumulée des ménages.
   - en ordonnées, le montant des déciles.

   b. En supposant que la répartition des revenus est régulière dans chaque tranche, estimer graphiquement :

   - la proportion de ménages situés en dessous du seuil de pauvreté.
   - la proportion de ménages situés au-dessus du seuil de richesse.

Partie B Courbe de Lorenz et indice de Gini

1. a. Calculer le revenu disponible annuel moyen de l'ensemble des ménages français en 2015.

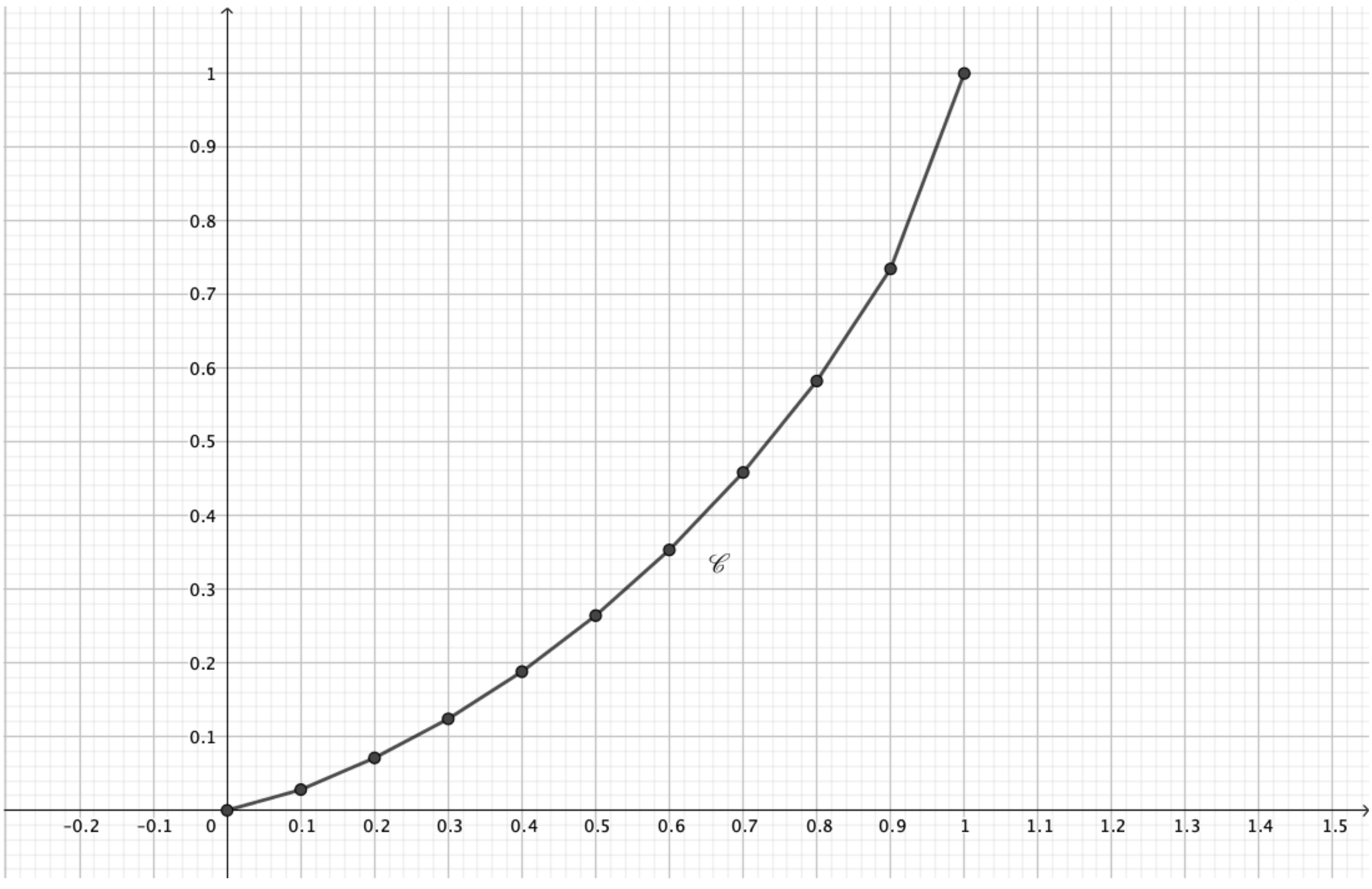
   b. En notant N l'effectif total des ménages en 2015, justifier que:

   - la somme de tous les revenus disponibles est égale à $36299 \times N$.
   - la somme de tous les revenus disponibles inférieurs à D_1 est égale à $10030 \times \frac{N}{10}$.

   c. En déduire que la masse des revenus détenus par la tranche « Inf. à D_1 » est égale à environ 2,8 %.

   d. Justifier de même que la masse des revenus détenus par la tranche « Sup. à D9 » est égale à environ 26,5 %.

2. La courbe $\mathscr{C}$ donnée ci-contre est la courbe de Lorenz des revenus disponibles pour l'année 2015.



Expliquer comment cette courbe a été construite.

3. La courbe $\mathscr{C}$ peut être approchée par la courbe représentant la fonction f définie sur $[0; 1]$ par :

$$f(x) = 0,7x^3 + 0,3x$$

On considère les points $O(0; 0)$ et $A(1; 1)$.

a. Justifier que la fonction f est croissante et convexe sur $[0; 1]$ puis que, $\forall x \in [0; 1], f(x) \leq x$.

b. Calculer l'intégrale $I = \int_0^1 f(x)\mathrm{d}x$.

c. En déduire la valeur de l'indice de Gini des revenus disponibles, égal à l'aire du domaine compris entre le segment $[OA]$ et la courbe $\mathscr{C}$.

d. Sur le site de l'INSEE, on peut lire que l'indice de Gini pour les revenus disponibles était égal à 0,348 en 2015.

Y a-t-il moins de 0,1 % d'écart avec la valeur obtenue à la question c. ?

2. La place de la France dans la zone Euro en 2018

Le tableau ci-dessous donne, dans la zone Euro pour l'année 2018:

- Le rapport interquintile $\frac{D_8}{D_2}$ de revenu disponible équivalent, c'est-à-dire le rapport entre le revenu total perçu par les 20% de la population ayant le revenu le plus élevé (quintile supérieur) et le revenu total perçu par les 20% de la population ayant le revenu le plus bas (quintile inférieur).
- Le coefficient de Gini G du revenu disponible équivalent.

Pays	D_8/D_2	G	**Pays**	D_8/D_2	G
Allemagne	5,07	0,311	Italie	6,09	0,334
Autriche	4	0,268	Lettonie	6,8	0,356
Belgique	3,78	0,256	Lituanie	7,09	0,369
Chypre	4,29	0,291	Luxembourg	5,72	0,332
Espagne	6,03	0,332	Malte	4,28	0,287
Estonie	6,07	0,306	Pays-Bas	4,05	0,274
Finlande	3,6	0,259	Portugal	5,22	0,321
France	4,23	0,285	Slovaquie	3,03	0,234
Grèce	5,51	0,323	Slovénie	3,4	0,234
Irlande	4,23	0,289	Total (Zone Euro)	5,07	0,306

Etudier la position de la France en Europe quant à la dispersion des revenus.

On calculera des caractéristiques statistiques (position et dispersion) et on construira un graphique pour argumenter.