

Préparation à l'agrégation externe de Sciences Sociales

Statistique descriptive

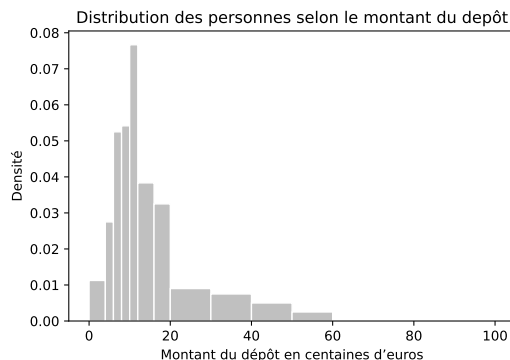
2021-2022

Exercice 1 (2014)

1. On représente graphiquement cette distribution en utilisant un histogramme. Pour le construire, il faut trouver la fréquence (f_i) de chaque classe et sa densité ($d_i = f_i/a_i$) :

x_i	n_i	f_i	$d_i = f_i/a_i$
[0; 4[270	0.045	0.01125
[4; 6[330	0.055	0.0275
[6; 8[630	0.105	0.0525
[8; 10[650	0.108	0.054
[10; 12[920	0.153	0.0765
[12; 16[920	0.153	0.038
[16; 20[780	0.13	0.0325
[20; 30[540	0.09	0.009
[30; 40[450	0.075	0.0075
[40; 50[300	0.05	0.0075
[50; 60[150	0.025	0.0025
[60; 100[60	0.01	0.00025

L'histogramme qu'on obtient est le suivant :



La classe modale est la classe avec la plus grande densité, dans ce cas il s'agit de la classe [10; 12[.

2. Pour calculer la moyenne on utilise la formule :

$$\bar{x} = \frac{1}{N} \sum_{i=1}^k n_i c_i,$$

où c_i est le centre de la classe. Alors,

$$\begin{aligned} \bar{x} &= \frac{1}{6000} [2 \times 270 + 5 \times 330 + 7 \times 630 + \dots + 55 \times 150 + 80 \times 60] \\ &= 17.55 \end{aligned}$$

Pour calculer l'écart type on utilise la formule :

$$\sigma_x = \sqrt{\frac{1}{N} \left(\sum_{i=1}^k n_i c_i^2 \right) - \bar{x}^2},$$

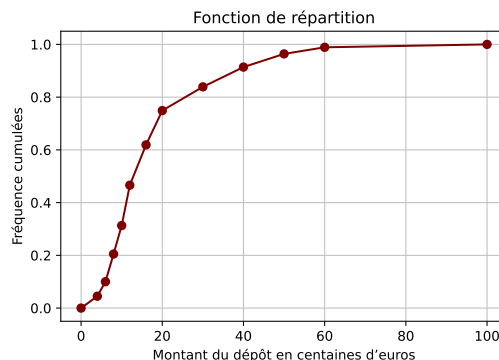
où c_i est le centre de la classe. Alors,

$$\begin{aligned} \sigma_x &= \sqrt{\frac{1}{6000} [2^2 \times 270 + 5^2 \times 330 + 7^2 \times 630 + \dots + 55^2 \times 150 + 80^2 \times 60] - 17.55^2} \\ &= 13.7 \end{aligned}$$

3. Pour tracer la fonction de répartition il faut trouver les fréquences cumulées (F_i) :

x_i	n_i	f_i	F_i
[0; 4[270	0.045	0.045
[4; 6[330	0.055	0.1
[6; 8[630	0.105	0.205
[8; 10[650	0.108	0.313
[10; 12[920	0.153	0.466
[12; 16[920	0.153	0.619
[16; 20[780	0.13	0.749
[20; 30[540	0.09	0.839
[30; 40[450	0.075	0.914
[40; 50[300	0.05	0.964
[50; 60[150	0.025	0.989
[60; 100[60	0.01	1

Alors, le graphique de la fonction de répartition est :



Pour trouver la médiane il faut noter qu'elle se trouve dans la classe [12; 16[. On calcule la médiane par interpolation linéaire à l'intérieure de la classe :

$$\frac{Me - 12}{16 - 12} = \frac{0.5 - 0.466}{0.619 - 0.466} \implies Me = 12.89.$$

L'intervalle inter décile est l'intervalle $[D_1; D_9]$. La valeur de D_1 est 6. Pour trouver D_9 il faut tout d'abord noter que D_9 se trouve dans la classe [30; 40[. On le calcule par interpolation linéaire à l'intérieure de la classe :

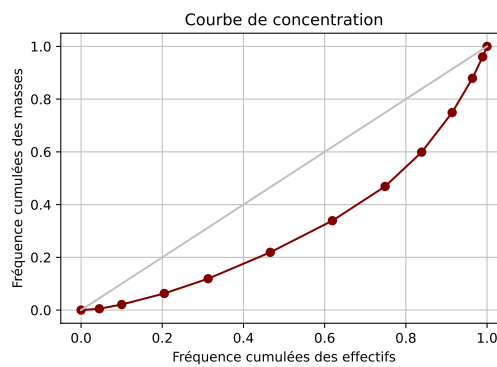
$$\frac{D_9 - 30}{40 - 30} = \frac{0.9 - 0.839}{0.914 - 0.839} \implies D_9 = 38.13.$$

Pourtant, l'intervalle inter décile est [6; 38.13]. Cela veut dire que le 80% des individus ont fait des dépôts d'un montant d'entre 6 et 38.13 centaines d'euros.

4. Pour construire la courbe de concentration il faut calculer la masse de chaque classe ($n_i c_i$), la fréquence des masses ($n_i c_i / m_{\text{totale}}$) et les fréquences cumulées des masses (Q_i) :

x_i	n_i	$n_i c_i$	$n_i c_i / m_{\text{totale}}$	Q_i
$[0; 4[$	270	540	0.005	0.005
$[4; 6[$	330	1650	0.016	0.021
$[6; 8[$	630	4410	0.042	0.063
$[8; 10[$	650	5850	0.056	0.119
$[10; 12[$	920	10120	0.1	0.219
$[12; 16[$	920	12880	0.12	0.339
$[16; 20[$	780	14040	0.13	0.469
$[20; 30[$	540	13500	0.13	0.599
$[30; 40[$	450	15750	0.15	0.749
$[40; 50[$	300	13500	0.13	0.879
$[50; 60[$	150	8250	0.08	0.96
$[60; 100[$	60	4800	0.05	1
$m_{\text{totale}} = 105290$				

La courbe de concentration ou courbe de Lorentz est alors :



On va calculer l'indice de concentration de Gini comme :

$$G = \frac{\int_0^1 l(t) dt}{1/2},$$

où l désigne la courbe de Lorentz. En utilisant la méthode des trapèzes, on obtient $\int_0^1 l(t) dt = 0.309$, donc $G = 0.618$.

Pour trouver la médiane il faut noter qu'elle se trouve dans la classe $[20; 30[$. On calcule la médiane par interpolation linéaire à l'intérieur de la classe :

$$\frac{Ml - 20}{30 - 20} = \frac{0.5 - 0.469}{0.599 - 0.469} \implies Ml = 22.38.$$

Le 50% de la masse totale est obtenue à partir des individus qui ont fait des dépôts d'un montant inférieure à 22.38 centaines d'euros.