

S1-2 Distribution statistique et valeurs centrales

Thérèse Saint-Julien

1. La notion de valeur centrale

- **1.1 La distribution statistique**

Ensemble ordonné des valeurs prises par le caractère X pour les éléments de l'ensemble E

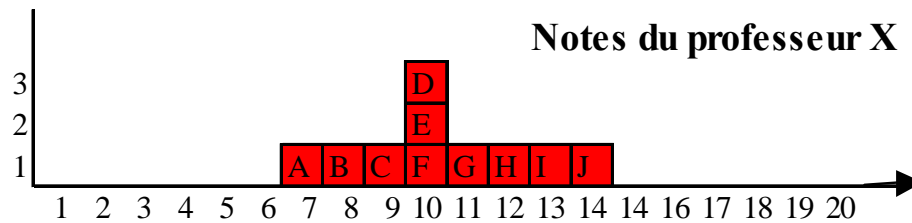
Image d'une distribution statistique: le diagramme de distribution

- **1.2 Qu'appelle t-on valeur centrale d'une distribution?**

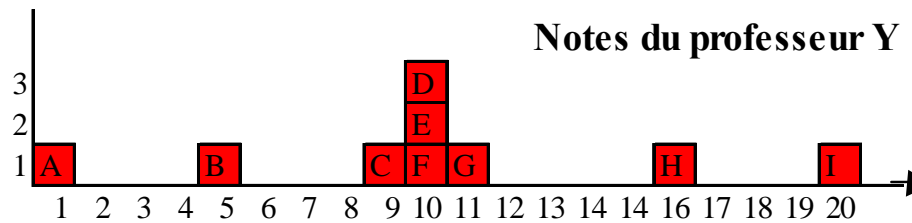
une valeur centrale est un résumé, par une seule valeur, de l'ensemble des valeurs d'une distribution statistique.

La distribution statistique: exemple de la notation de deux professeurs

nombre d'élèves

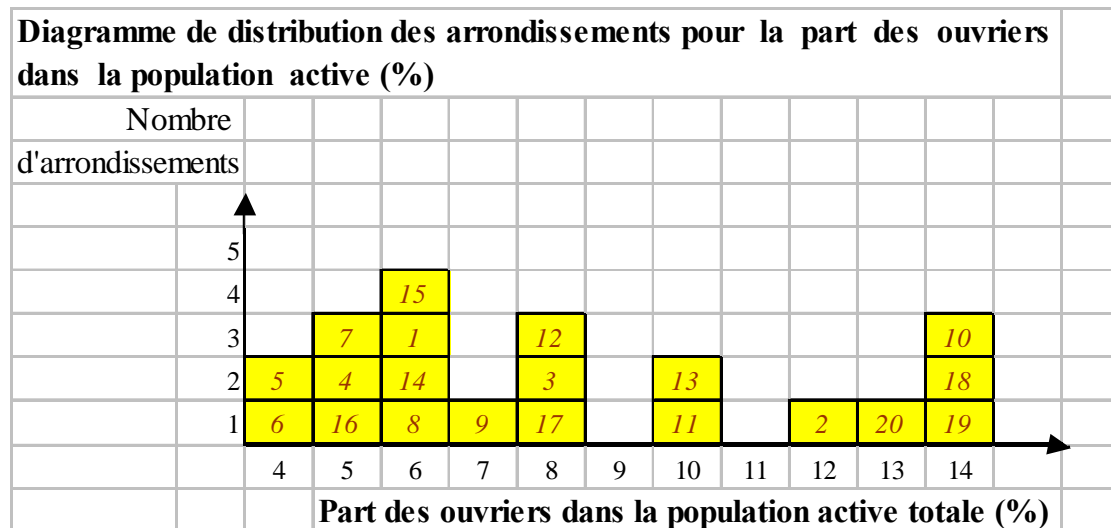


nombre d'élèves



Elèves	notes du professeur X	notes du professeur Y
	/20	/20
A	7	0
B	8	5
C	9	9
D	10	10
E	10	10
F	10	10
G	11	11
H	12	15
I	13	20
moyenne	10	10
médiane	10	10

La part des ouvriers dans la population active résidente des arrondissements parisiens (%)



Minimum	3,9
Maximum	14,3
Médiane	7,2
Moyenne	8,2

Arrondissements	Part des ouvriers dans la population active %
Paris 5e	3,9
Paris 6e	3,9
Paris 7e	4,8
Paris 4e	5,1
Paris 16e	5,4
Paris 15e	5,6
Paris 1er	5,7
Paris 14e	6,3
Paris 8e	6,4
Paris 9e	6,8
Paris 12e	7,6
Paris 3e	7,9
Paris 17e	8,3
Paris 13e	9,6
Paris 11e	10,2
Paris 2e	12,0
Paris 20e	12,9
Paris 10e	13,7
Paris 18e	13,7
Paris 19e	14,3

1.1. Trois valeurs centrales

- **Le mode**
- **La médiane**
- **La moyenne arithmétique**

1.2 Les principes devant guider le choix d'une valeur centrale

Le statisticien Yule (XIX^e siècle) a défini six propriétés souhaitables pour les valeurs centrales. Le tableau ci-contre désigne les avantages et inconvénients de chacune des trois valeurs centrales

Les valeurs centrales doivent être:	Mode	Médiane	Moyenne
1. définies de façon objective	oui 😊	oui 😊	oui 😊
2. dépendre de toutes les valeurs observées	non ☹️	non ☹️	oui 😊
3. avoir une signification concrète	oui 😊	oui 😊	non ☹️
4. être simples à calculer	oui 😊	oui 😊	oui 😊
5. être peu sensibles aux fluctuations de l'échantillon	non ☹️	oui 😊	non ☹️
6. se prêter aux calculs algébriques	non ☹️	non ☹️	oui 😊

2. Le mode: définition

Le mode (ou **valeur dominante** pour un caractère quantitatif, et **modalité dominante** dans le cas d'un caractère qualitatif), est la valeur ou modalité la plus fréquente d'une distribution.

Le mode correspond à la valeur lue en abscisse du sommet du diagramme de distribution. Lorsque celui-ci présente deux pics séparés par un creux, on dit que la distribution est bimodale, etc..

*** caractère qualitatif et caractère discret**

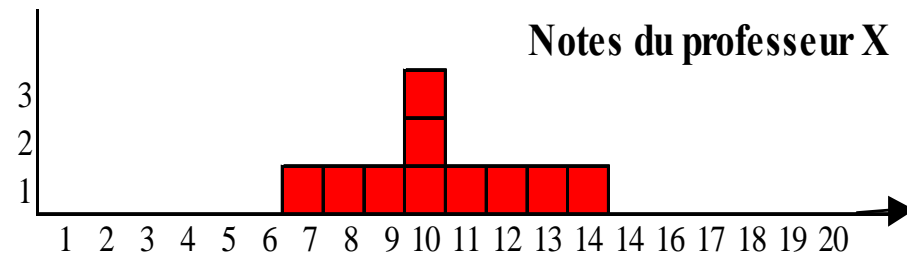
Pour un caractère qualitatif, ou pour un caractère quantitatif discret ayant un nombre de modalités inférieur au nombre d'éléments, le mode est la modalité ou la valeur qui a **la fréquence simple la plus élevée** (ou l'effectif le plus élevé, ce qui revient

*** Caractère quantitatif continu**

Les **valeurs possibles étant en nombre infini**, il est peu probable que deux éléments aient la même valeur. Dans ce cas, le mode ne peut pas être défini directement, il faut au préalable établir une **partition en classes**. Le mode est alors le **centre de la cl**

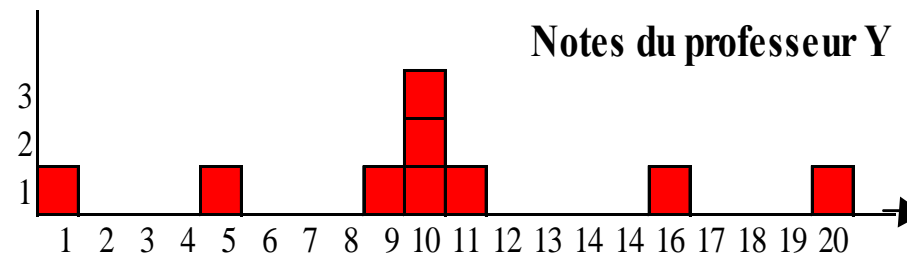
Le mode: exemple des notes données par les professeurs X et Y

nombre
d'élèves



	Mode
Professeur X	10
Professeur Y	10

nombre
d'élèves



Le mode: exemple des ouvriers dans les arrondissements parisiens

Arrondissements	Part des ouvriers dans la population active % (valeurs arrondies à l'unité	Part des ouvriers dans la population active (%) valeurs arrondies au 1/10
Paris 5e	4	3,9
Paris 6e	4	3,9
Paris 7e	5	4,8
Paris 4e	5	5,1
Paris 16e	5	5,4
Paris 15e	6	5,6
Paris 1er	6	5,7
Paris 14e	6	6,3
Paris 8e	6	6,4
Paris 9e	7	6,8
Paris 12e	8	7,6
Paris 3e	8	7,9
Paris 17e	8	8,3
Paris 13e	10	9,6
Paris 11e	10	10,2
Paris 2e	12	12,0
Paris 20e	13	12,9
Paris 10e	14	13,7
Paris 18e	14	13,7
Paris 19e	14	14,3

Mode Sur les valeurs
arrondies à l'unité,
distribution unimodale
mode= 6%

Sur les valeurs arrondies
au 1/10, distribution
bi-modale
mode= 3,9% et 13,7%

2.1 Les limites du mode

- **Le mode ne dépend pas de toutes les valeurs observées**
- **Le mode est instable, en particulier pour les distributions décrivant des ensembles d'éléments peu nombreux**
- **Le mode ne se prête pas aux calculs algébriques**

2.2 Les avantages du mode

- **Le mode a une signification concrète, qui rend son usage nécessaire dans un certain nombre de problèmes appliqués**

exemples:

La connaissance des heures de pointe dans la gestion des transports urbains

Les jours d'affluence pour les commerces

Les routes les plus fréquentées (points noirs dans le réseau)

etc..

3. La médiane

Les valeurs du caractère X étant classées par ordre croissant, la médiane, $X_{\text{méd}}$ est la valeur du caractère qui partage l'ensemble E décrit par X en deux sous ensembles d'effectifs égaux :

50 % des éléments ont des valeurs de X supérieures à $X_{\text{méd}}$ et 50% ont des valeurs inférieures.

La médiane ne peut être calculée que pour les caractères quantitatifs

3.1 Comment calculer la médiane à partir du tableau élémentaire

- On ordonne le tableau de dénombrement
- On repère l'élément qui partage la distribution en deux parties égales: soit l'élément qui a le rang $(N+1)/2$ pour le caractère X.
 - Si la distribution a un nombre impair d'éléments on trouve une valeur unique qui est la médiane,
 - si la distribution a un nombre pair d'éléments, on trouve deux valeurs qui déterminent un intervalle médian : on prend alors pour médiane le centre de cet intervalle médian.

La médiane, exemple de la répartition des ouvriers à Paris

Minimum	3,9
Maximum	14,3
Médiane	7,2
Moyenne	8,2

Nombre pair d'éléments:
*repérage de l'intervalle médian
arrondissement occupant la 10 ^e position dans l'ensemble ordonné des valeurs de X: Paris 9 ^e , valeur de X=6,8
arrondissement occupant la 11 ^e position dans l'ensemble ordonné des valeurs de X: Paris 12 ^e , valeur de X=7,6
*calcul de la médiane: $X_{\text{méd}} = (7,6 - 6,8) / 2 = 7,2$

Arrondissements, E=20	Part des ouvriers dans la population active %
Paris 5e	3,9
Paris 6e	3,9
Paris 7e	4,8
Paris 4e	5,1
Paris 16e	5,4
Paris 15e	5,6
Paris 1er	5,7
Paris 14e	6,3
Paris 8e	6,4
Paris 9e	6,8
Paris 12e	7,6
Paris 3e	7,9
Paris 17e	8,3
Paris 13e	9,6
Paris 11e	10,2
Paris 2e	12,0
Paris 20e	12,9
Paris 10e	13,7
Paris 18e	13,7
Paris 19e	14,3

7,2
%

tableau élémentaire avec
E=29 éléments

Année	Pluviosité du mois de septembre (en mm)
1896	119
1897	170
1898	31
1899	35
1900	583
1901	422
1902	27
1903	254
1904	153
1905	133
1906	27
1907	760
1908	49
1909	201
1910	70
1911	90
1912	252
1913	300
1914	200
1915	158
1916	382
1917	186
1918	150
1919	630
1920	211
1921	160
1922	182
1923	98
1924	112

tableau élémentaire ordonné en fonction de

Année	Pluviosité du mois de septembre (en mm)
1902	27
1906	27
1898	31
1899	35
1908	49
1910	70
1911	90
1923	98
1924	112
1896	119
1905	133
1918	150
1904	153
1915	158
1921	160
1897	170
1922	182
1917	186
1914	200
1909	201
1920	211
1912	252
1903	254
1913	300
1916	382
1901	422
1900	583
1919	630
1907	760

La médiane: exemple des précipitations au Mont-Aigoual au mois de septembre de 1896 à 1924

- Nombre impair d'éléments, E=29 éléments
- On repère **l'élément médian**. Dans cet exemple il s'agit du mois de septembre qui occupe la 15^e position pour les valeurs de X dans le tableau élémentaire ordonné, soit **le mois de septembre de l'année 1921**, qui est l'élément médian.
- Cet élément coupe la distribution en deux classes ayant le même nombre de mois de septembre: **La valeur de la médiane correspond au total des précipitations du mois de septembre 1921 soit: 160 mm.**
- 14 mois ont des valeurs inférieures à 121, et 14 autres des valeurs supérieures

3.2 Avantages et limites de la médiane (propriétés de Yule)

Qualités

- est définie de façon objective
- a une signification concrète
- est simple à calculer
- est peu sensibles aux fluctuations de l'échantillon

Limites

- ne dépend pas des valeurs observées mais de leur ordre
- ne se prête pas aux calculs algébriques

3.3 Propriétés de la médiane

$$\sum_{I=1}^N |X_i - A|$$

est minimum si

$$A = X_{\text{méd}}$$

□ • Propriétés de la médiane

La médiane est la valeur du caractère qui est la plus proche de toutes les autres valeurs de la distribution. C'est celle qui minimise les distances en valeur absolue au point $X_{\text{méd}}$

4. La moyenne arithmétique

- Elle est calculée pour les caractères quantitatifs
-
- Calculée à partir du tableau élémentaire, la moyenne arithmétique des valeurs prises par X pour les différents éléments i de E est la somme des valeurs de X_i divisée par le nombre d'éléments N

$$\bar{X} = \sum_{i=1}^n X_i / N$$

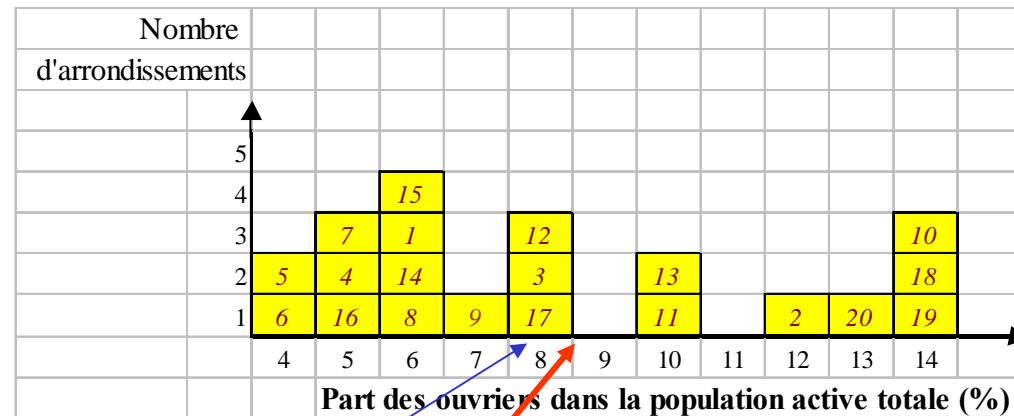
Ecritures

$$\bar{X} = \sum_{i=1}^n X_i / N$$

$$\bar{X} = (X_1 + X_2 + X_3 + \dots + X_i + \dots + X_n) / N$$

La moyenne des taux d'ouvriers dans la population active des arrondissements parisiens

Arrondissements	Part des ouvriers dans la population active (%)
Paris 5e	3,9
Paris 6e	3,9
Paris 7e	4,8
Paris 4e	5,1
Paris 16e	5,4
Paris 15e	5,6
Paris 1er	5,7
Paris 14e	6,3
Paris 8e	6,4
Paris 9e	6,8
Paris 12e	7,6
Paris 3e	7,9
Paris 17e	8,3
Paris 13e	9,6
Paris 11e	10,2
Paris 2e	12,0
Paris 20e	12,9
Paris 10e	13,7
Paris 18e	13,7
Paris 19e	14,3
moyenne	8,2

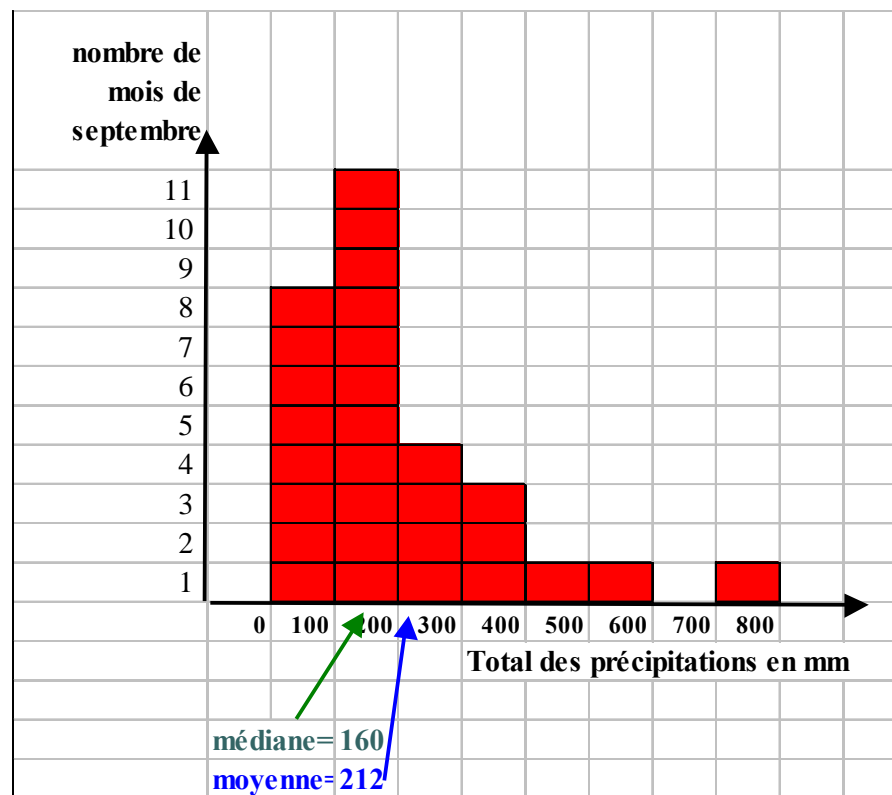


$X_{med.} = 7,2\%$

$$\bar{X} = \sum_{i=1}^n X_i / N = 8,2\%$$

Année	Pluviosité du mois de septembre (en mm)
1896	119
1897	170
1898	31
1899	35
1900	583
1901	422
1902	27
1903	254
1904	153
1905	133
1906	27
1907	760
1908	49
1909	201
1910	70
1911	90
1912	252
1913	300
1914	200
1915	158
1916	382
1917	186
1918	150
1919	630
1920	211
1921	160
1922	182
1923	98
1924	112

La moyenne des précipitations de septembre au Mont Aigoual de 1896 à 1924



4.1 Avantages et limites de la moyenne (propriétés de Yule)

Qualités

- est définie de façon objective
- dépend de toutes les observations
- est simple à calculer
- prête pas aux calculs algébriques

Limites

- n'a pas de signification concrète
- est sensible aux fluctuations de l'échantillon

4.2 Propriétés de la moyenne

1. Si A = moyenne de X

$$\sum_{i=1}^n X_i = n * \bar{X}$$

2. La somme des écarts à la moyenne est égale à zéro.

$$\sum_{i=1}^n (X_i - \bar{X}) = 0$$

—

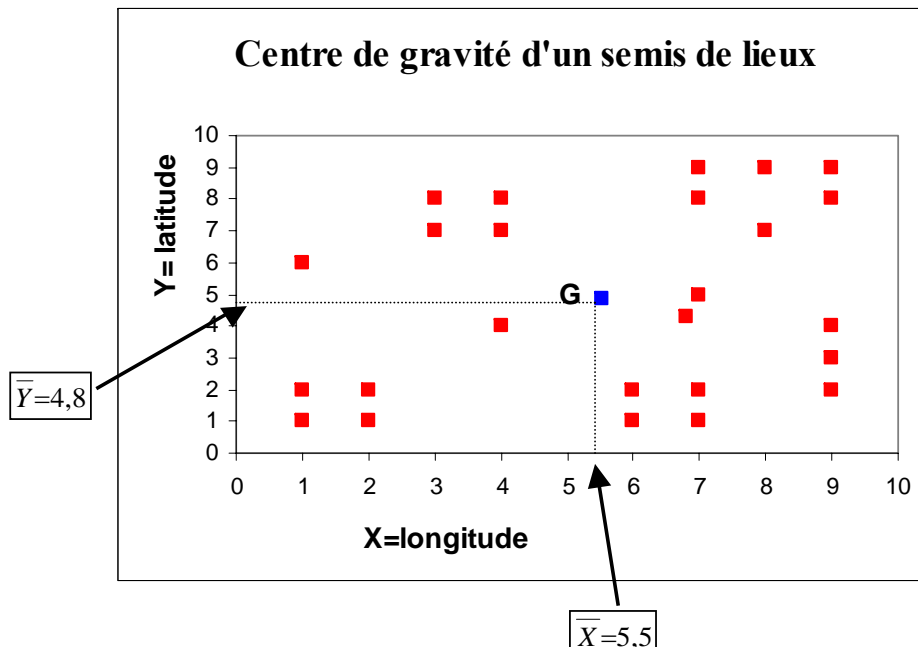
3. La moyenne minimise les distances au carré

$$\sum_{i=1}^N (X_i - A)^2$$

est minimum si ,et seulement si, A = moyenne du caractère X

$$A = \bar{X}$$

Exemple: le centre de gravité d'un semis de lieux



Lieux	$x=longitude$	$y=latitude$
A	1	1
D	2	1
K	6	1
M	7	1
B	1	2
E	2	2
L	6	2
N	7	2
T	9	2
U	9	3
H	4	4
V	9	4
O	7	5
F	3	7
I	4	7
R	8	7
G	3	8
J	4	8
P	7	8
W	9	8
Q	7	9
S	8	9
X	9	9
Centre de gravité		
moyenne	5,5	4,8

4.3 la moyenne pondérée

- **On recourt à la pondération lorsque les unités n'ont pas le même poids. Si chaque élément i de l'ensemble E est décrit par sa mesure X_i et son poids P_i , la moyenne pondérée est :**

$$\bar{X}_p = \frac{\sum_{i=1}^n (X_i * p_i)}{\sum_{i=1}^n p_i}$$

Arrond.	R_i =Revenu moyen de l'arrond. par habitant (en francs, 1999)	P_i =Nombre d'habitants en 1999	$R_i * P_i$
Paris 1er	167 906	16 888	2 835 596 528
Paris 2e	109 396	19 585	2 142 520 660
Paris 3e	134 496	34 248	4 606 219 008
Paris 4e	165 292	30 675	5 070 332 100
Paris 5e	153 080	58 849	9 008 604 920
Paris 6e	242 699	44 919	10 901 796 381
Paris 7e	298 195	56 985	16 992 642 075
Paris 8e	252 285	39 314	9 918 332 490
Paris 9e	128 315	55 838	7 164 852 970
Paris 10e	90 993	89 612	8 154 064 716
Paris 11e	98 519	149 102	14 689 379 938
Paris 12e	110 208	136 591	15 053 420 928
Paris 13e	100 127	171 533	17 175 084 691
Paris 14e	119 852	132 844	15 921 619 088
Paris 15e	136 470	225 362	30 755 152 140
Paris 16e	268 285	161 773	43 401 269 305
Paris 17e	147 961	160 860	23 801 006 460
Paris 18e	88 950	184 586	16 418 924 700
Paris 19e	84 504	172 730	14 596 375 920
Paris 20e	84 011	182 952	15 369 980 472
moyenne	149 077		
<i>somme</i>		2 125 246	283 977 175 490
moyenne pondérée			133 621

La moyenne pondérée: exemple du revenu moyen

Moyenne= moyenne des revenus moyens par habitant des arrondissements :

149 077 frs

Moyenne pondérée= revenu moyen par habitant à Paris

133 621 frs

Les arrondissements les plus peuplés sont aussi ceux qui en moyenne concentrent les ménages dont les revenus sont les moins élevés. D'où une moyenne pondérée plus faible que la moyenne par arrondissement

5. Valeurs centrales et forme de la distribution statistique

