S1.3 La dispersion statistique

Thérèse Saint-Julien

Dispersion statistique : définition

- On appelle *dispersion statistique*, la tendance qu'ont les valeurs de la distribution d'un caractère à s'étaler, à se disperser, de part et d'autre d'une valeur centrale.
- On distingue la **dispersion absolue** (mesurée dans l'unité de mesure du caractère), et la **dispersion relative** (mesurée par un nombre sans dimension).

1. Les mesures de la dispersion absolue

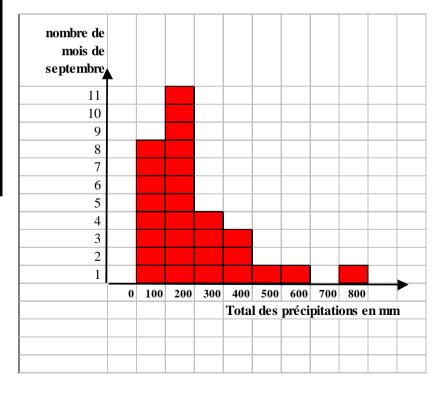
Arrondiss ements	Nombre d'habitants en 1999 (en milliers)	Part des cadres dans le total de la population active résidente (%)
Paris 1er	17	45
Paris 2e	20	39
Paris 3e	34	44
Paris 4e	31	46
Paris 5e	59	52
Paris 6e	45	53
Paris 7e	57	49
Paris 8e	39	43
Paris 9e	56	43
Paris 10e	90	33
Paris 11e	149	35
Paris 12e	137	37
Paris 13e	172	34
Paris 14e	133	43
Paris 15e	225	46
Paris 16e	162	45
Paris 17e	161	40
Paris 18e	185	28
Paris 19e	173	25
Paris 20e	183	26

- Les paramètres de dispersion absolue indiquent de combien les valeurs d'une distribution s'écartent en général de la valeur centrale de référence. Un paramètre de dispersion absolue s'exprime toujours dans l'unité de mesure.
- Les mesures de la dispersion absolue s'expriment dans l'unité de mesure de la variable considérée. Exemples:
 - Pour le nombre d'habitants en 1999, il s'agira de milliers d'habitants
 - Pour la part des cadres dans le total de la population active résidente, il s'agira de % de personnes actives occupant un emploi de cadre
- Les trois paramètres de dispersion absolue les plus courants sont l'étendue, l'intervalle inter quantile, et l'écart type

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
1925	61

La dispersion inter annuelle des précipitations de septembre au Mont Aigoual entre 1896 à 1925

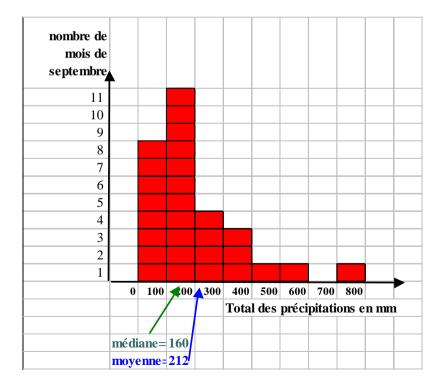
	Minimum	27 mm	
	Maximum	760 mm	
	Etendue	733 mm	
	1er quartile	90 mm	
Médiane	3ème quartile	252 mm	
	Intervalle interquartile	162 mm	
Moyenne	Variance	32246	
	Ecart-type	179,6 mm	



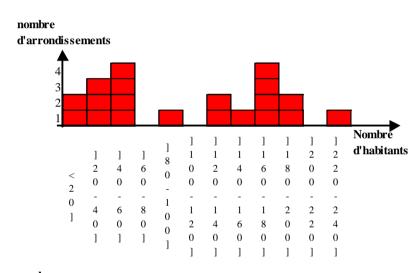
	Pluviosité du
	mois de
Année	
	septembre (en
	mm)
1896	119
1897	170
1898	31
1899	35
1900	583
1901	422
1902	27
1903	254
1904	153
1905	133
1906	27
1907	760
1908	49
1909	201
1910	70
1911	90
1912	252
1913	300
1914	200
1915	158
1916	382
1917	186
1918	150
1919	630
1920	211
1921	160
1922	182
1923	98
1924	112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

1.1 L'étendue ou amplitude

- l'étendue ou amplitude d'une distribution est égale à la différence entre la plus grande et la plus petite valeur de la distribution :
- Etendue de $X = X_{max} X_{min}$



1.2 La mesure de la dispersion statistique et les valeurs centrales



la mesure de l'étendue exceptée, chacun des paramètres de dispersion statistique caractérise le degré de dispersion des valeurs de la distribution statistique de part et d'autre d'une valeur centrale de référence

intervalle interquantile par rapport à la médiane

variance et écart-type par rapport à la moyenne

nombre							
d'arrondissen	nents						
5							_
3							
2							
1							
•]]]]]]	Part des emplois de
<	2	3	3	4	4	5	cadres dans la
2	4	0	5	0	5	0	population active
4	-	-	-	-	-	-	résidente
1	3	3	4	4	5	5	Coracine
1	0	5	0	5	0	5	
]]]]]]	

Arrondissements	Part des cadres dans la population active %	Nombre d'habitants en 1999 (en milliers)
Médiane	43,0	111
Moyenne	40,3	106

1. Mesures de la dispersion autour de la médiane

Année	Pluviosité du mois de septembr e (en mm)	Nombre de mois de septembre
1902	27	
1906	27	
1898	31	
1899	35	
1908	49	
1910	70	
1911	90 —	1er quartile=90 mm
1923	98	
1924	112	
1896	119	
1905	133	
1918	150	
1904	153	
1915	158	➤ médiane=159 mm
1921	160	inediane—137 mm
1897	170	
1922	182	
1917	186	
1914	200	
1909	201	
1920	211	
1912	252 —	3e quartile=252 mm
1903	254	
1913	300	
1916	382	
1901	422	
1900	583	
1919	630	
1907	760	

≻Quantiles

1. **Quantiles** : les quantiles sont les valeurs du caractère qui définissent les bornes d'une partition en classes d'effectifs égaux.

Ces particuliers

Les *quartiles* sont les trois valeurs qui permettent de découper la distribution en quatre classes d'effectifs égaux. On les note X_{q1} , X_{q2} et X_{q3} .

Partition du caractère
$$X_{\min} \rightarrow X_{q1}$$

 $\rightarrow X_{q2} \rightarrow X_{q3} \rightarrow X_{\max}$

fréquence des éléments: 25% 25% 25%

Remarque : X_{q2} est égal à la médiane.

Intervalle	162 mm
interquartile	102 111111

Quantiles

Précipitations en mm (déciles)

33	1er décile
65,5	2e décile
105	3e décile
141,5	4e décile
159	5e décile
184	6e décile
206	7e décile
277	8e décile
502,5	9e décile

2. Déciles - Les déciles sont les neufs valeurs de X qui permettent de découper la distribution en dix classes d'effectifs égaux. On les note X_{d1}...X_{d9}.

Etc..

➤Intervalle interquantile

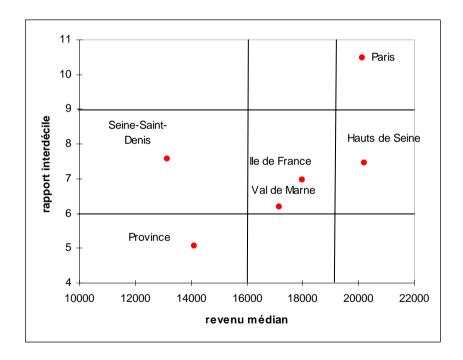
- L'intervalle interquartile est l'étendue de la distribution sur laquelle se trouvent concentrée la moitié des éléments dont les valeurs de X sont les plus proches de la médiane. On exclut alors de la distribution les 25% des valeurs les plus faibles et les 25% des valeurs les plus fortes de X. Cet intervalle se note: $(X_{q3}-X_{q1})$.
- L'intervalle interdécile est *l'étendue de la distribution sur laquelle se trouvent concentrés 80% des éléments* dont les valeurs de X sont les moins différentes de la médiane. On exclut alors de la distribution les 10 % des valeurs les plus faibles et les 10% des valeurs les plus fortes. Il se note $(X_{d9}-X_{d1})$.

Le rapport interquantile : $X_{Q3}/X_{Q1;}$ ou interdéciles : X_{D1}/X_{D9}

Revenus fiscaux déclarées au titre de l'année 2001

	revenu médian	1er décile	9e décile	rapport interdécile
Paris	20147	4864	50961	10,5
Hauts de Seine	20195	6138	45716	7,4
Seine-Saint-Denis	13155	3658	27740	7,6
Val de Marne	17181	5841	36129	6,2
lle de France	17982	5581	38912	7,0
Province	14103	5446	27637	5,1

source DGI: revenus fiscaux localisés



Application du rapport inter déciles aux revenus des ménages en Ile-de-France

• La position de chaque point est définie en fonction de la valaur prise par le revenu médian (abscisse) et par le rapport inter déciles (ordonnée).

2. Mesures de la dispersion autour de la moyenne

1.4 Variance, écart-type et moyenne

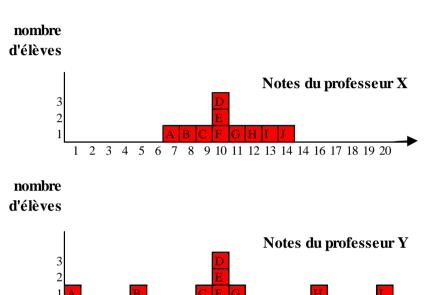
$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{n} (X_i - \overline{X})^2}$$

- La variance de X, notée σ² cest une mesure globale de la variation d'un caractère de part et d'autre de la moyenne arithmétique (quantité d'information). Elle exprime la dispersion dans une unité de l'ordre du carré de l'unité de mesure du caractère.
- Pour obtenir un paramètre de dispersion absolue, on calcule la racine carrée de la variance
- L'écart type, noté σ_x est la racine carré de la moyenne des carrés des écarts à la moyenne, c'est à dire la racine carrée de la variance.

Exemple de calcul de la variance et de l'écart-type

	Note	es du Profe	esseur X	Notes du Professeur Y			
Elèves	/20	écarts algébrique s à la moyenne	carré des écarts algébriques	/20	écarts algébrique s à la moyenne	carré des écarts algébriques	
		$(X_i - \overline{X})$	$(X_i - \overline{X})^2$		$(X_i - \overline{X})$	$\left[\left(\chi_{i}-\overline{\chi}\right)^{2}\right]$	
A	7	-3	9	0	-10	100	
В	8	-2	4	5	-5	25	
C	9	-1	1	9	-1	1	
D	10	0	0	10	0	0	
E	10	0	0	10	0	0	
\mathbf{F}	10	0	0	10	0	0	
G	11	1	1	11	1	1	
H	12	2	4	15	5	25	
I	13	3	9	20	10	100	
moyenne	10	0		10	0		
somme		0	28		0	252	
écart-type	1,8	1,8		5,3	5,3		



1 2 3 4 5 6 7 8 9 10 11 12 13 14 14 16 17 18 19 20

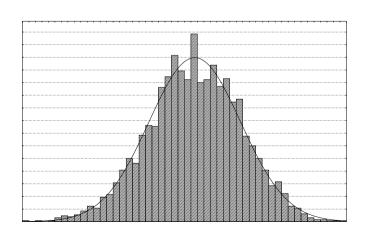
Valeurs centrales et paramètres de dispersion pour quelques indicateurs décrivant les arrondissements de Paris

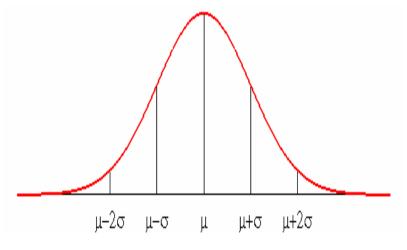
Paramères de la distribution	revenu moyen communal par unité de consommation en milliers d'Euros	Part chômeurs dans la population active %	Part des cadres dans la population active %	Part des ouvriers dans la population active %	Part des étrangers dans la population %	Nombre d'habitants en milliers en 1999
Nbr. de valeurs utilisées	20	20	20	20	20	20
Minimum	12807	8	25	4	11	17
Maximum	45460	17	53	14	28	225
Etendue	32652	9	28	10	17	208
1er quartile	15142	9	35	5	14	42
Médiane	20033	11	43	7	17	111
3ème quartile	25398	14	45	11	21	167
Intervalle interquartile	10256	4	11	6	8	125
intervalle interquartile relatif	0,51	0,41	0,25	0,78	0,48	1,12
Moyenne	22727	11	40	8	18	106
Variance	94804619	7	63	11	24	4327214
Ecart-type	9737	3	8	3	5	66
Ecart absolu moyen	7687	2	7	3	4	62
CV (écart-type/moyenne)	0,44	0,24	0,20	0,42	0,29	0,64

Caractéristiques de la distribution normale

Caractéristiques de la courbe Normale :

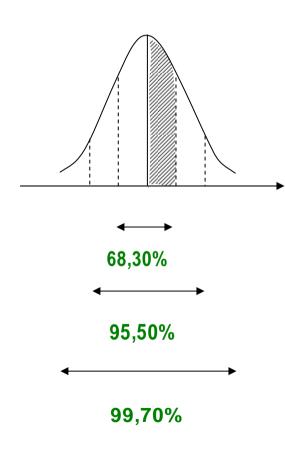
- 1. La variable x varie de $-\infty$ à $+\infty$
- 2. La fonction est toujours > 0
- 3. L'aire sous la courbe vaut 1
- 4. Elle est symétrique
- 5. Elle atteint son maximum au point \overline{X} (moyenne arithmétique)
- 6. Elle a une forme en « cloche » : plus on s'écarte à gauche et à droite de la moyenne arithmétique, plus la hauteur de la courbe diminue avec l'axe des abscisses comme asymptote.





La signification probabiliste de l'écart-type

Distribution normale



L'écart-type a l'avantage d'avoir une signification probabiliste. La théorie des probabilités permet en effet d'estimer la chance qu'a une valeur d'être éloignée de la moyenne de plus d'un certain nombre d'écart-types.

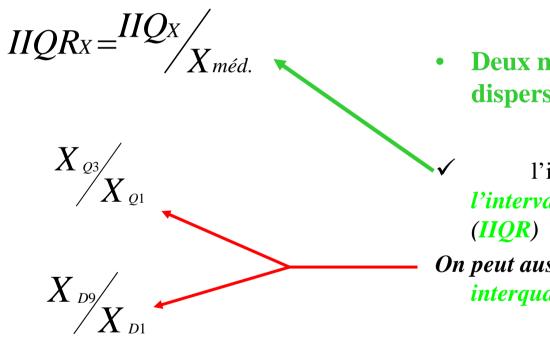
Lorsqu'une distribution est **gaussienne** (on dit aussi "**normale**") les probabilités de trouver les valeurs a une distance donnée de la moyenne sont les suivantes :

68,3% des valeurs sont entre
$$\overline{X_X} - 1\sigma_X$$
 et $\overline{X_X} + 1\sigma_X$

95,5% des valeurs sont entre
$$\overline{X_X}$$
 -2 σ_X et $\overline{X_X}$ + 2 σ_X 99,7% des valeurs sont entre $\overline{X_X}$ -3 σ_X

et
$$\overline{X}_x + 3\sigma_x$$

2. Les mesures de la dispersion relative



• Deux mesures usuelles de la dispersion relative à partir de:

Eliminer l'effet de l'unité de

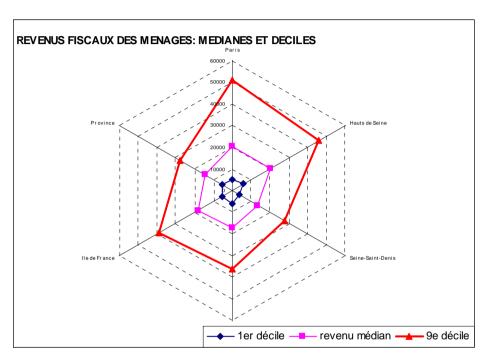
l'intervalle interquantile: l'intervalle interquantile relatif (IIQR)

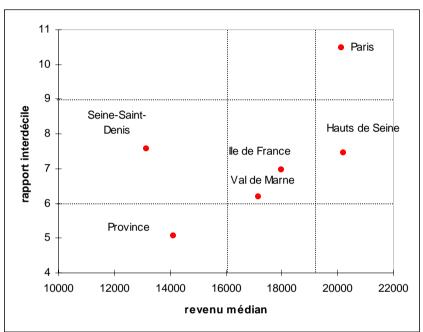
On peut aussi utiliser le rapport interquantile, par exemple:

$$CV = \sigma X X$$

l'écart-type: *le coefficient de variation (CV)*

La dispersion statistique des revenus des ménages à Paris et dans les départements de la petite couronne

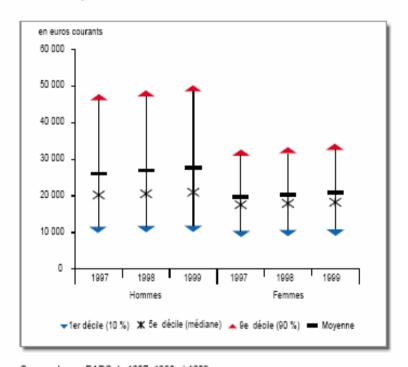




Revenus fiscaux (en euros) déclarées au titre de l'année 2001

Zones géographiques	revenu médian	1er décile	9e décile	rapport interdécile
Paris	20147	4864	50961	10,5
Hauts de Seine	20195	6138	45716	7,4
Seine-Saint-Denis	13155	3658	27740	7,6
Val de Marne	17181	5841	36129	6,2
lle de France	17982	5581	38912	7,0
Province	14103	5446	27637	5,1

Figure 2 - Evolution de la dispersion des salaires annuels nets par sexe entre 1997 et 1999



Evolution des salaires annuels des hommes et des femmes en Ile-de-France: moyennes, médianes et déciles

Source: Insee, DADS de 1997, 1998 et 1999

Figure 1 - Distribution par déciles des salaires annuels en 1998 et 1999 en euros courants

	lle-de-France					Province		Ecart de salaire	
		Salaire net 1998 de prélèvements				Salaire net de prélèvements		net IDF / province(%)	
	1998	1999	Hommes	Femmes	Hommes	Femmes	1998	1999	1999
1er décile (10 %)	11 289	11 406	11 702	10 777	11 824	10 864	10 405	10 502	+ 8,6
2e décile (20 %)	13 446	13 618	14 025	12 702	14 201	12 862	11 694	11 825	+ 15,2
3e décile (30 %)	15 384	15 618	16 089	14 493	16 326	14 693	12 850	13 004	+ 20,1
4e décile (40 %)	17 298	17 594	18 207	16 182	18 506	16 466	14 062	14 231	+ 23,6
5e décile (médiane)	19 329	19 707	20 582	17 922	20 982	18 263	15 419	15 602	+ 26,3
6e décile (60 %)	21 780	22 247	23 588	19 779	24 075	20 201	16 984	17 216	+ 29,2
7e décile (70 %)	25 108	25 685	27 764	22 120	28 369	22 633	18 917	19 221	+ 33,6
8e décile (80 %)	30 212	31 002	34 289	25 462	35 092	26 108	21 814	22 206	+ 39,6
9e décile (90 %)	40 980	42 157	47 162	31 536	48 427	32 618	27 429	27 947	+ 50,8
9e déc./1er déc.	3,63	3,70	4,03	2,93	4,10	3,00	2,64	2,66	111

Note de lecture : en 1999, 40 % des hommes touchaient un salaire annuel inférieur à 18 506 euros (4e décile), 40 % des femmes percevaient un salaire inférieur à 16 466 euros et 40 % des salairés gagnalient un salaire inférieur à 17 594 euros (4e décile), 40 % des femmes percevaient un salaire inférieur à 16 466 euros et 40 % des salairés gagnalient un salaire inférieur à 16 466 euros et 40 % des Source : Insee, DADS de 1998, panel DADS au 1/25 pour la province

Mesures absolues et relatives de la dispersion statistique

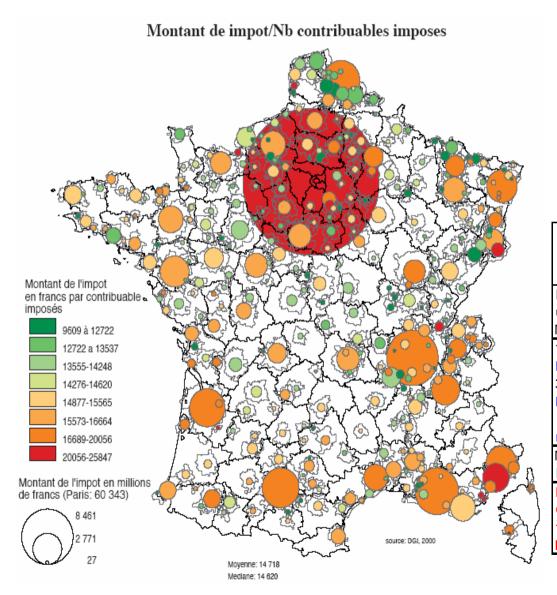
Arrondissements	Part des ouvriers dans la population active %	Part des cadres dans la population active %
Minimum	3,9	24,7
Maximum	14,3	53,0
Etendue	208	28,3
1er quartile	5,5	34,5
Médiane	7,2	43,0
3ème quartile	11,1	45,4
intervalle interquartile	5,6	11
intervalle interquartile relatif	0,78	0,25
Moyenne	8,2	40,3
Ecart-type	3,4	7,9
CV (écart-type/moyenne)	0,42	0,20

- Pour comparer les degré de dispersion de deux distributions on a recours aux mesures de dispersion relative, et non pas les mesures de dispersion absolue.
- Dans l'exemple ci-contre la comparaison directe des écart-types (ou des intervalles interquartiles) pourrait, à tort, laisser croire, qu'en moyenne, les disparités introduites entre les arrondissements par le caractère «Part des cadres dans la population active » sont supérieures à celles liées au caractère « Part des ouvriers dans la population active ». Il n'en est rien.
- Comme l'indique la comparaison des coefficients de variation les arrondissements parisiens sont, en moyenne, beaucoup plus différents les uns des autres du fait de la part des ouvriers (cv=0,42) que de la part des cadres(cv=contre 0,2).

3. Distribution statistique et répartition géographique

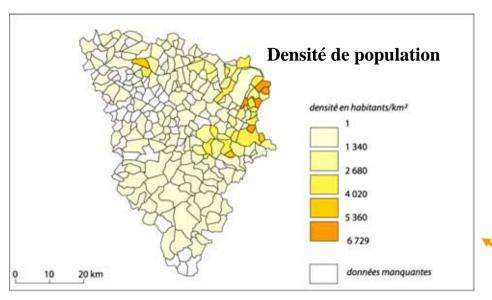
Montant moyen de l'impôt et nombre de contribuables dans les aires urbaines: distribution statistique et distribution

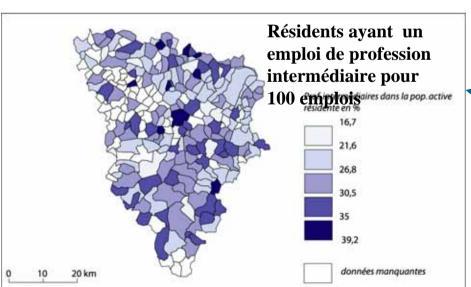




	montant de l'impôt par contribuable	nombre de contribuables imposés
Nbre d'aires urbaines	361	361
Minimum	9609	2656
1er quartile	13386	4832
Médiane	14620	8984
3ème quartile	15755	22354
Int. interquartile	2369	17522
Int. interquartile relatif	0,16	1,95
Maximum	25847	2745058
Etendue	16239	2742402
Moyenne	14718	30719
CV (écart- type/moyenne)	0,15	4,89
Ecart-type	2193	149915

Forme de la distribution, dispersion et distribution géographique (dépt des Yvelines)



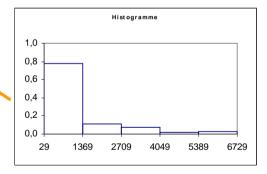


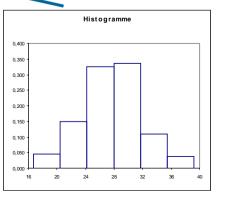
densité en
Paramètres des distributions 1999(habitan ts/km²)

prof.
Intermédiaire
s dans la
population
active
résidente en
1999 (%)

part des

Nbr. de valeurs utilisées	194	194
Minimum	29	17
1er quartile	99	25
Médiane	303	27
3ème quartile	1289	30
Maximum	6729	39
Etendue	6700	23
Moyenne	934	28
CV (écart-type/moyenne)	1,45	0,15
Ecart-type d'échantillon	1355	4
Ecart absolu moyen	993	3



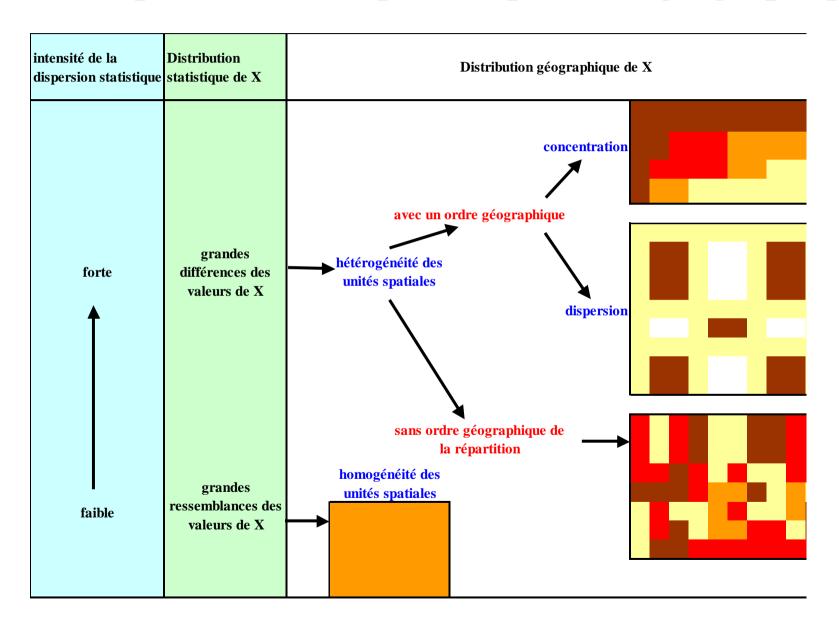


Montant moyen de l'impôt et nombre de contribuables dans les aires urbaines: distribution statistique et distribution géographique: synthèse

Caractéristiques des distributions géographiques	montant de l'impôt par contribuable	nombre de contribuables imposés		
Disparités entre les aires urbaines	Très faibles: les aires urbaines ont, en moyenne, tendance à se ressembler. Les valeurs des mesures de dispersion relative voisines de 0.	Très fortes: les aires urbaines sont en moyenne très dissemblables les unes des autres. Les mesures de dispersion relative très éloignées de 0		
Forme de la distribution	symétrique: méd.=moy. Les aires urbaines les plus nombreuses se concentrent autour de la moyenne	très dissymétrique: méd. < moy. Les aires urbaines les plus nombreuses se concentrent autour des valeurs les plus faibles		

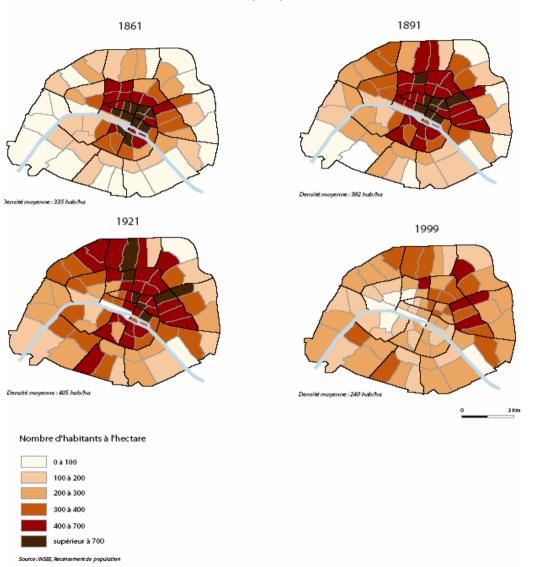
Dispersion statistique et répartition géographique

Dispersion statistique et répartition géographique



Les densités de population à Paris de 1861 à 1999

Evolution de la densité résidentielle brute par quartier.



La répartition en 1861

Forte dispersion statistique des valeurs= forte hétérogénéité spatiale

Dans ce cas, l'hétérogénéité est assortie d'un ordre spatial fort : les fortes densités concernent les quartiers des arrondissements centraux; les plus faibles concernent les quartiers les plus périphériques des arrondissements périphériques.

La répartition en 1999

La dispersion statistique est plus faible mais un ordre spatial demeure. Les fortes valeurs sont le fait des quartiers nord et est.

Dispersion statistique et répartition géographique dans l'exemple des arrondissements parisiens

