

NOTE STATISTIQUE N° 27

Application du schéma de de Wijs à un phénomène démographique :
la répartition des densités de population à la surface de la France
continentale.

--:--:--:--:--:--:--

1/ Le formalisme de la géostatistique - et en particulier les lois de dispersion intrinsèque - n'intéresse pas seulement le champ des phénomènes géologiques ou miniers.

Il doit pouvoir s'appliquer à bien d'autres domaines. Nous pensons, en particulier, à certains phénomènes économiques, géographiques, sociologiques, etc... Les séries temporelles des économistes ou les densités de population des géographes sont des variables régionalisées. Il serait intéressant d'examiner si la répartition de grandeurs de cette nature peut être décrite par une loi de dispersion intrinsèque, et si, en particulier, la loi de dispersion logarithmique de de Wijs peut leur être appliquée.

A titre d'exemple, nous allons étudier la répartition des densités de population des départements de la France continentale (Corse exclue), et montrer que cette répartition obéit bien à un schéma de Wijsion, à condition de faire abstraction de l'agglomération parisienne, profondément étrangère à l'équilibre démographique du pays.

2/ Une densité de population est une variable régionalisée à deux dimensions. Sa définition suppose la donnée d'un champ et d'un support géométrique. Le champ, ou domaine homogène dans lequel on étudie la répartition de la variable, sera ici le territoire de la France continentale. La Corse a été exclue, en raison de son insularité.

Le support géométrique, sur lequel la densité de population est calculée, est ici le département. Il est certain que les différents départements n'ont ni la même taille, ni la même forme géométrique, et cela introduire une certaine imprécision dans les calculs. Cependant, malgré son arbitraire, le découpage administratif est relativement homogène, et il n'est pas absurde de raisonner sur un département moyen. La même difficulté se serait du reste présentée, si nous avions choisi d'étudier le phénomène à l'échelle de l'arrondissement ou de la commune : les derniers ne sont disponibles, de toutes façons, que dans le cadre du découpage administratif.

Les données utilisées sont présentées sur le tableau I. Elles ont été calculées à partir des valeurs de la surface et de la population de chaque département données par le petit Larousse (édition).

Les départements sont désignés par leur numéro minéralogique ; les densités sont exprimées en nombre d'habitants au km².

Tableau I									
Densité de population des départements français									
Dep	Densité	Dep	Densité	Dep	Densité	Dep	Densité	Dep	Densité
1	52.6	19	47.6	37	56.7	55	30.4	73	38.2
2	61.0	(20)	(30.7)	38	69.6	56	71.5	74	59.0
3	50.6	21	38.3	39	42.7	57	99.9	(75)	(9999.0)
4	12.0	22	73.0	40	25.2	58	36.0	76	133.8
5	15.1	23	33.7	41	37.7	59	332.0	77	68.8
6	120.0	24	41.9	42	131.7	60	67.5	(78)	(250.0)
7	46.0	25	56.6	43	45.7	61	44.5	79	52.3
8	46.6	26	40.8	44	95.5	62	173.0	80	70.4
9	29.8	27	52.5	45	50.9	63	59.8	81	51.6
10	39.0	28	43.5	46	29.7	64	54.0	82	45.1
11	42.5	29	103.2	47	49.3	65	44.6	83	61.6
12	35.2	30	64.8	48	17.4	66	55.4	84	69.9
13	185.5	31	80.3	49	68.7	67	141.0	85	56.4
14	71.4	32	30.3	50	67.8	68	134.3	86	44.6
15	32.2	33	79.6	51	47.0	69	322.0	87	60.5
16	52.4	34	74.0	52	29.1	70	37.8	88	57.8
17	57.5	35	82.8	53	49.2	71	58.8	89	35.6
18	39.2	36	36.6	54	100.1	72	66.3	90	143.2
: Seine (75) et S & O (78) groupés ensemble ont une densité									
: de 1010,0									

La Corse (20) n'est mentionnée que pour mémoire. Il n'en sera pas tenu compte. La valeur énorme de la densité de la Seine (75) - 10 000 habitants au km² - est due au caractère exclusivement urbain de ce minuscule département. Pour obtenir un département de taille comparable aux autres, il convient de diluer la Seine (75) dans la Seine et Oise (78). On obtient pour le groupe (75 & 78) une densité de 1010,0 moins aberrante que celle de la Seine, mais qui apparaît encore comme exceptionnellement élevée. Le caractère aberrant du monstre urbain se manifestera à la fois sur les histogrammes et sur les variogrammes.

3/ Faisant tout d'abord abstraction de la position géographique des départements, c'est-à-dire du caractère régionalisé de la répartition des densités, examinons-la distribution, au sens statistique, des valeurs des Densités figurant sur le tableau I.

L'histogramme de ces valeurs est présenté sur le tableau II. Nous avons utilisé des classes logarithmiques de raison 10/10. Chaque classe est désignée par sa borne inférieure.

TABLEAU II			
Histogramme des densités de population des 88 départements (Corse exclue, Seine et Seine et Oise groupés ensemble)			
Classe	Fréquence	Classe	Fréquence
10.0	1	126	5
12.6	1	158	2
15.8	1	199	0
19.9	0	251	0
25.1	6	316	2
31.6	12	398	0
39.8	16	501	0
50.1	20	631	0
63.1	14	794	0
79.4	4	1000	1
100.0	3	Total	88

Cette distribution est voisine d'une distribution log normale. On peut parler de distribution lognormaloïde. On voit cependant sur l'histogramme du tableau II, que, malgré l'emploi des classes logarithmiques, la courbe des fréquences conserve une certaine dissymétrie positive (courbe étalée vers la droite). La distribution réelle est encore plus dissymétrique que la courbe lognormale.

Ce point justifie a posteriori l'emploi de classes logarithmiques. En égard à l'échelle de variation du phénomène, il convient d'étudier les propriétés des logarithmes des densités, plutôt que des densités elles-mêmes. La moyenne géométrique $\bar{y}$ et la variance s^2 (variance du log neperien des densités) de la distribution prennent les valeurs suivantes, selon que l'on tient compte ou non de la valeur densité = 1010 du groupe (75 + 78) :

	Avec (75+78)	Sans (75+78)
Moyenne géométrique $\bar{y}$	58.4	56.4
Variance s^2	0.433	0.334

On note l'influence du groupe Seine + Seine et Oise sur la valeur numérique de la variance . La valeur IOIO anormalement élevée de la densité de ce groupe est responsable à elle seule du quart de la variance. La population obtenue en supprimant la dernière classe du tableau II, beaucoup moins dispersée, est aussi beaucoup plus symétrique - et donc plus proche de la lognormalité.

Pour mettre ce point en évidence, nous avons testé, au moyen d'un test χ^2 à quatre degrés de liberté, l'accord de la distribution complète avec la loi lognormale de paramètres $\gamma = 58.4$, $\sigma^2 = 0.433$, et de la distribution obtenue en supprimant (75+78) avec la loi lognormale de paramètres $\gamma = 56.4$, $\sigma^2 = 0.334$. On obtient les valeurs suivantes :

	Avec (75 + 78)	Sans (75 + 78)
χ^2	14.4	13.10
$P(\chi^2)$	0.007	0.011

En réalité, il n'est pas correct de donner une interprétation quantitative du test, sous la forme d'une probabilité $P(\chi^2)$, puisque les densités des 88 départements ne sont nullement indépendantes les unes des autres. Mais la signification qualitative du test demeure : la suppression du département (75 + 78) améliore sensiblement l'ajustement, mais l'accord avec la loi lognormale reste assez médiocre. En fait, même après cette suppression, la courbe des fréquences conserve une dissymétrie légèrement positive. Son allure suggère qu'à partir d'une certaine époque historique, un nombre relativement petit de départements privilégiés (Nord, Pas-de-calais) Rhône, etc..) s'est mis à évoluer et à progresser plus vite que le reste du pays, phénomène dont le groupe (75 + 78) ne serait qu'une manifestation particulièrement exagérée. Cependant, nous verrons que ces départements s'intègrent bien au réseau spatial des corrélations de Wijsiennes, tandis qu'il n'en est absolument pas de même du groupe Seine-Seine et Oise.

4/ Admettons maintenant que la répartition des densités (ou, plus exactement, de leurs logarithmes) soit du type de Wijsien, c'est-à-dire qu'il existe une fonction intrinsèque du type $3\sigma \log r$, 3σ étant la dispersion absolue.

Pour calculer la valeur théorique de la variance des densités, il conviendrait en toute rigueur de déterminer l'équivalent linéaire moyen des départements et celui du territoire métropolitain

lui-même. En première approximation, on admettra que ces équivalents sont dans le même rapport que les racines carrées des surfaces,

1. S'agissant d'une répartition à 2 dimensions, on aurait pu adopter la notation $2x$. Nous avons préféré conserver $3x$, afin de permettre une comparaison directe avec les dispersions absolues des teneurs dans un gisement minier à 3 dimensions ou, si l'on veut, que le département "moyen" est équivalent à une figure géométriquement semblable au contour du territoire français. Dans ces conditions, la variance des densités est donnée par la formule classique de de Wijs :

$$\sigma^2 = \frac{3x}{2} \log_e \frac{S}{4} = \frac{3x}{2} \log_e 88$$

Des valeurs expérimentales de la variance σ^2 , on déduit ainsi les estimations suivantes de la dispersion absolue $3x$, selon que l'on conserve, ou non, le groupe $-(75+78)$;

	Avec (75 + 78)	Sans (75 + 78)
Dispersion absolue $3x$	0.194	0.150

Si l'hypothèse d'une loi de dispersion absolue de Wijsienne est vérifiée, l'étude du variogramme doit nous conduire à des estimations de $3x$ compatibles avec les précédentes.

5/ La détermination expérimentale du variogramme présente quelques difficultés, du fait que les différents départements n'ont ni la même taille, ni la même forme géométrique. Il est aisé d'attribuer à chacun d'entre eux un centre de gravité et d'admettre que la distance de deux départements est définie comme la distance de leurs centres de gravité. On pourrait penser qu'il suffirait alors de classer les différents couples possibles de départements d'après les valeurs de leurs distances r , de définir plusieurs classes ou intervalles de distances $r_1, r_2, r_3 \dots$ et de calculer les variances $\sigma_1^2, \sigma_2^2, \sigma_3^2$ du logarithme népérien $\log \frac{d_i}{d_j}$

du rapport des densités d_i et d_j des couples de départements contenus dans chaque classe. Un essai a été tenté, sans aboutir à des résultats intéressants. Cet échec se comprend si l'on fait intervenir le principe de similitude. Imaginons, en effet, que tous les départements soient des cercles de rayons différents. Deux cercles égaux et tangents apportent la même contribution au variogramme, quel que soit leur rayon R . Par exemple, sur la figure 1, il en est ainsi des deux couples A et B de ces cercles de rayons R et $R/10$, dont les centres sont distants respectivement de $2R$ et $\frac{R}{5}$, au contraire, le couple C, constitué de deux cercles

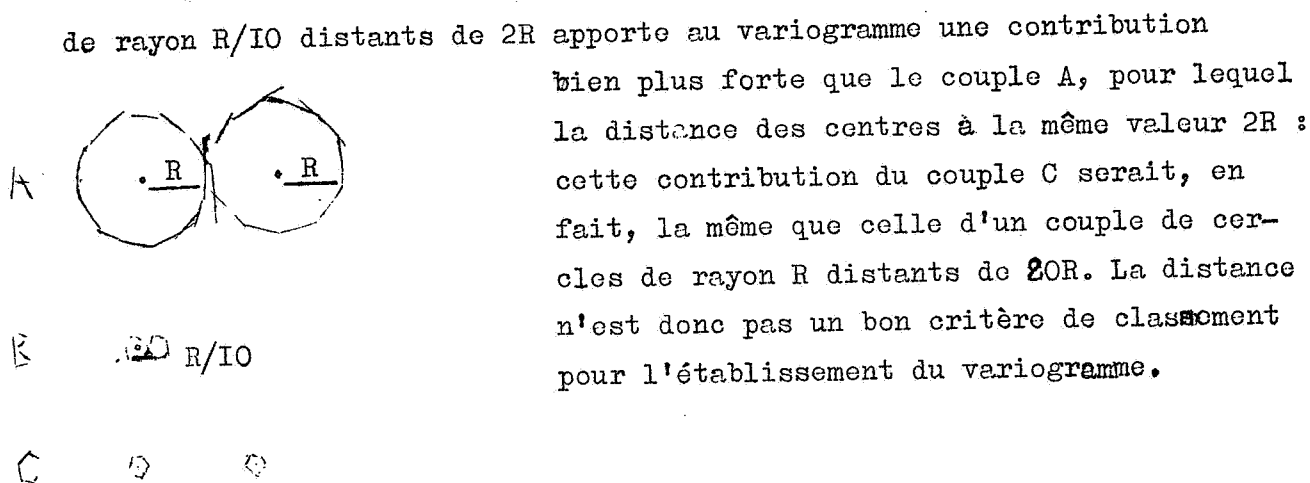


Fig 1

Pour rester en harmonie avec le principe de similitude, il convient d'adopter un critère topologique. Etant donné un département i , on considérera d'abord une première auréole, constituée par les départements J, k, l, m, n touchant le département i . On rangera dans une

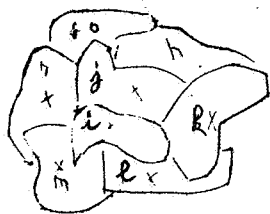


Fig 2

première classe tous les couples ij de première auréole, qui serviront à calculer un premier point du variogramme. Dans une deuxième auréole, on rangera les couples du type io ou ip , qui serviront à calculer un deuxième point du variogramme. On pourrait aller plus loin, et définir une troisième, une quatrième auréole, mais les difficultés matérielles deviennent considérables, et nous nous sommes limités aux deux premières.

Il est difficile de calculer les valeurs théoriques des variances $\overline{\sigma_1^2}$ et $\overline{\sigma_2^2}$ de première et deuxième auréole, puisqu'elles dépendent au plus haut point de la forme et de l'orientation mutuelle des départements intéressés. Mais, toutes choses égales d'ailleurs, ces facteurs de forme et d'orientation interviennent de façon semblable sur la première et la deuxième auréole : en passant du premier au deuxième point, le variogramme doit augmenter d'une quantité proportionnelle au log du rapport des rayons moyens des deux auréoles, soit :

$$\overline{\sigma_2^2} - \overline{\sigma_1^2} = 6\sigma^2 \log_e 2$$

puisque la deuxième auréole a un rayon moyen double de celui de la première. Plus précisément, on a trouvé que les rayons moyens des deux auréoles étaient dans le rapport de 1 à 1.94, de sorte que l'on devrait avoir :

$$\frac{2}{62} - \frac{2}{61} = 6 \times \log 1.94 = 1.32 \times 3 \times$$

On a calculé les valeurs expérimentales des variances $\frac{2}{61}$ et $\frac{2}{62}$ des logarithmes népériens des rapports des densités des couples de première et de deuxième auréole, calculé leur différence et divisé par 1.32, pour obtenir une nouvelle estimation de la dispersion absolue $3 \times$. Les résultats sont présentés sur le tableau suivant, où l'on distingue deux cas, selon que le groupe Seine + Seine et Oise a été ou non conservé.

	Avec (75 + 78)	Sans (75 + 78)
Première auréole : $\frac{2}{61}$	0.552	0.384
2ème auréole : $\frac{2}{62}$	0.725	0.585
$3 \times$	0.131	0.152

Si l'on conserve (75 + 78), on obtient une dispersion de 0.131 franchement faible vis à vis de celle que l'on a déduit de la variance globale, soit 0.194.

L'écart n'est pas admissible. Si au contraire on fait abstraction de la Seine et de la Seine et Oise, l'estimation déduite du variogramme - 0.152 - est en accord parfait avec l'estimation déduite de la variance globale (- 0.150 -).

Nous concluons que les densités de population des départements français peuvent être considérées comme réparties selon une loi de dispersion intrinsèque du type de Wijsien, avec une dispersion absolue logarithmique $3 \times = \frac{15}{100}$ comparable à celle que l'on observe dans certains gisements miniers, sous réserve toutefois de faire abstraction de la Seine et de la Seine et Oise. L'agglomération parisienne - le monstre urbain - échappe au réseau de corrélations spatiales régissant la répartition démographique à la surface du pays, et se comporte un peu comme un corps étranger.

