

	<b>Analyse spatiale : formes et processus (M1)</b>
	<b>Cours 2 : DISTRIBUTION STATISTIQUE ET DISTRIBUTION SPATIALE</b>

Claude GRASLAND – Professeur de Géographie - Université Paris 7

### **Objectifs**

- 1- Méthodes statistiques élémentaires permettent de décrire la distribution d'un ensemble de lieux caractérisés par leurs position (x,y).
- 2- Modélisation des distributions spatiale en fonction d'hypothèses simples d'organisation de l'espace.

### **Plan de cours**

#### A.) LES VALEURS CENTRALES D'UNE DISTRIBUTION SPATIALE (x,y)

- Point moyen et distance-type
- Point médian et quantiles d'accessibilité
- Points modaux et pics de concentration

#### B) ANALYSE EMPIRIQUE DE LA DISPERSION SPATIALE

- Dispersion par rapport à un centre
- Dispersion par rapport à plusieurs centres
- Dispersion par rapport à un axe
- Dispersion par rapport à un réseau

#### C) ANALYSE THEORIQUE DES DISTRIBUTIONS SPATIALES

- La distribution aléatoire (loi de Poisson)
- Méthode des quadrats
- Méthode du plus proche voisin

### **Bibliographie**

Pour ceux qui ont la chance de pouvoir se le procurer, l'ouvrage de P.J. Taylor fournit sans doute la meilleure introduction à l'analyse des semis de points et comporte de nombreux exercices.

- **Pumain D., Saint-Julien T., 1997, *L'analyse spatiale (1) : Localisations dans l'espace*, Cursus-géographie, Armand-Colin, Paris, 167 p.**
- **Beguïn H., 1979, *Méthodes d'analyse géographique quantitative*, Litec, Paris, 252 p.**
- **Taylor P.J., 1977, *Quantitative methods in Geography*, Waveland Press Inc., Prospect Heights Illinois, pp. 133-172**

Exemples d'application intéressant dans les domaine de recherche à cheval sur l'archéologie et la géographie.  
Voir par exemple :

- **Archaeomedes, 1998, *Des oppida aux métropoles*, Paris, Anthropos, 280 p.**

## Document 1 : POINTS CENTRAUX ET PARAMÈTRES DE DISPERSION ASSOCIÉ

(a) Point moyen (non pondéré)

Le point moyen non pondéré est  $G(m_X, m_Y)$   $\left\{ \begin{array}{l} m_X = \frac{1}{N} \sum_{i=1}^N X_i \\ m_Y = \frac{1}{N} \sum_{i=1}^N Y_i \end{array} \right.$

Propriété :  $\sum_{i=1}^N (DE_{ij})^2$  est minimum si  $j=G$

Le paramètre de dispersion associé est la **distance-type**  $\sigma_D$  qui est la racine carrée de la moyenne du carré des distances, c'est-à-dire la racine carrée de la somme des variances de X et de Y.

$$\sigma_D = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - m_X)^2 + (Y_i - m_Y)^2} = \sqrt{\sigma_X^2 + \sigma_Y^2}$$

G est le barycentre des points, ce qui signifie qu'il **minimise la somme des distances euclidiennes élevées au carré à tous les points.**

(b) Point moyen (pondéré)

Le point moyen pondéré est  $G_P(m_{X_P}, m_{Y_P})$   $\left\{ \begin{array}{l} m_{X_P} = \frac{\sum_{i=1}^N P_i X_i}{\sum_{i=1}^N P_i} \\ m_{Y_P} = \frac{\sum_{i=1}^N P_i Y_i}{\sum_{i=1}^N P_i} \end{array} \right.$

Propriété :  $\frac{\sum_{i=1}^N P_i (DE_{ij})^2}{\sum_{i=1}^N P_i}$  est minimum si  $j=G$

Le paramètre de dispersion associé est la **distance-type pondérée**  $\sigma_{D,P}$  qui est égale à la racine carrée de la moyenne du carré des distances à tous les membres de la population P. Si on note  $\sigma_{X,P}$  l'écart type de X pondéré par P et  $\sigma_{Y,P}$  l'écart type de Y pondéré par P, on a :

$$\sigma_{D,P} = \sqrt{\frac{\sum_{i=1}^N P_i [(X_i - m_X)^2 + (Y_i - m_Y)^2]}{\sum_{i=1}^N P_i}} = \sqrt{(\sigma_{X,P})^2 + (\sigma_{Y,P})^2}$$

$G_P$  est le barycentre de la population P, ce qui signifie qu'il **minimise la somme des distances euclidiennes élevées au carré à tous les habitants de la population P.**

*(c) Point médian (non pondéré)*

**Le point médian non pondéré est le point M** minimisant la somme des distances à tous les points

$$\text{Propriété : } \sum_{i=1}^N D_{ij} \text{ est minimum si } j = M$$

Les paramètres de dispersion associés sont les **quantiles  $D_{M,\%}$**  de la distribution des points en fonction de la distance au point médian.

Ex. **La distance médiane  $D_{M,50\%}$**  est le rayon du cercle centré sur le point médian M permettant de rassembler 50% des points de la distribution.

**En distance rectilinéaire**, les coordonnées de M sont [médiane(X), médiane(Y)]. **En distance euclidienne**, elles ne peuvent être obtenues que par approximations successives.

*(d) Point médian (pondéré)*

**Le point médian pondéré est le point  $M_P$**  minimisant les distances à l'ensemble de la population P

$$\text{Propriété : } \frac{\sum_{i=1}^N P_i D_{ij}}{\sum_{i=1}^N P_i} \text{ est minimum si } j = M_P$$

Les paramètres de dispersion associés sont les quantiles  **$D_{M,P\%}$**  de la distribution de la population en fonction de la distance au point médian

Ex. **La distance médiane pondérée  $D_{M,P50\%}$**  est le rayon du cercle centré sur le point médian M permettant de rassembler 50% de la population P.

**En distance rectilinéaire** les coordonnées de  $M_P$  correspondent au point de croisement des deux lignes horizontales et verticales qui partagent le plan en quatre quartiers de population égales. **En distance euclidienne**, le point médian pondéré ne peut être obtenu que par approximations successives.

*(e) Exemple de calcul*

i	$X_i$	$Y_i$	$P_i$
1	10	40	500
2	60	10	200
3	70	50	100
4	80	30	100
5	90	40	100

⇒ **en distance euclidienne,**

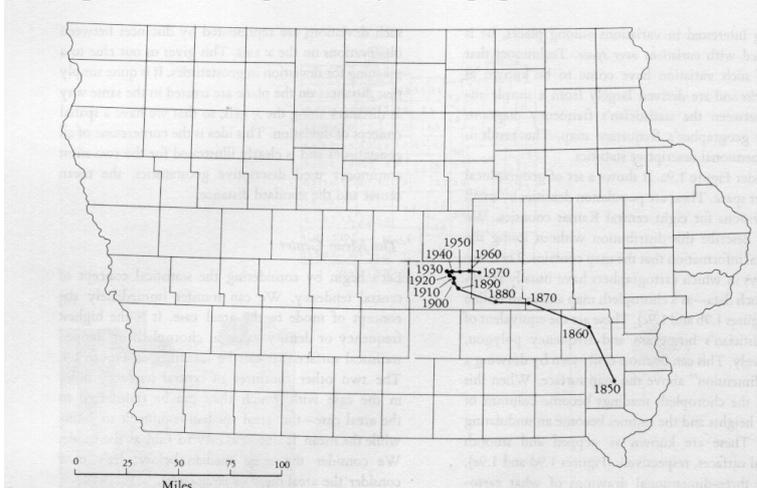
- le point moyen G a pour coordonnées **(62,34)** avec une distance-type de **31.0**
- le point moyen pondéré  $G_P$  a pour coordonnées **(41,34)** avec une distance-type de **34.5**

⇒ **en distance rectilinéaire**

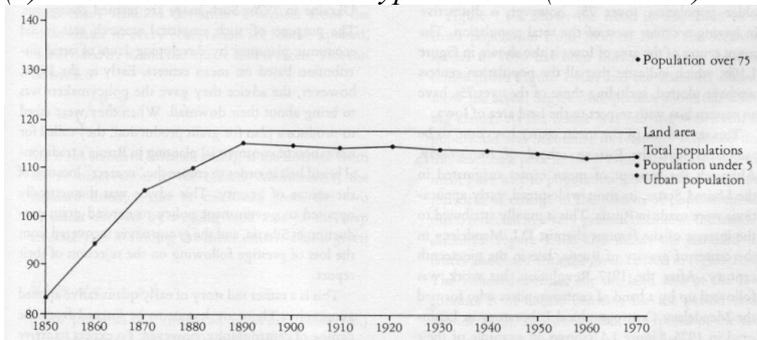
- le point médian M a pour coordonnées **(70,40)** avec une distance médiane de **40**
- le point médian pondéré  $M_P$  a pour coordonnées **(35,40)** avec une distance médiane de **30**

## Document 2 : EXEMPLES D'UTILISATION DU POINT MOYEN ET DE LA DISTANCE-TYPE

(a) Déplacement du point moyen de population en Iowa (1850-1970)



(b) Evolution de la distance-type en Iowa (1850-1970)



(c) Calcul des centres de gravité de différentes productions en Ukraine vers 1926

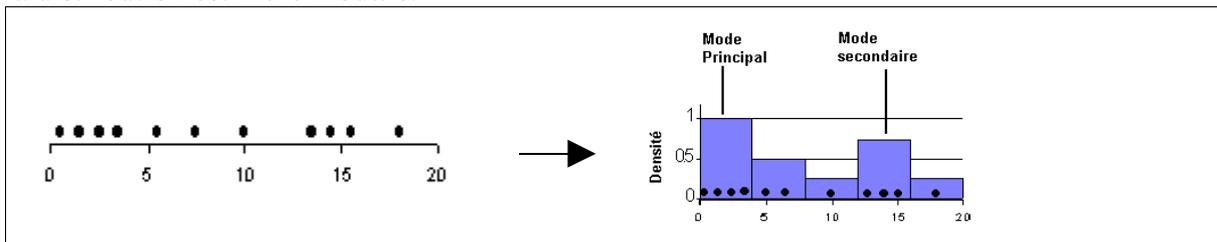


### DOCUMENT 3 : DETERMINATION DES PICS DE CONCENTRATION D'UNE DISTRIBUTION SPATIALE (points modaux)

#### (a) détermination des points modaux à l'aide d'un maillage territorial

##### *Le cas unidimensionnel*

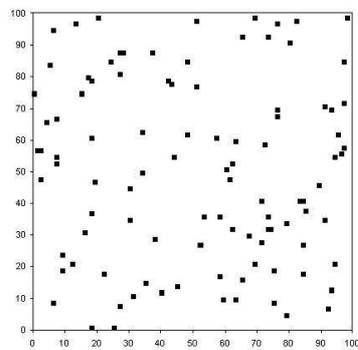
Dans le cas d'une **distribution statistique (ou d'une distribution spatiale unidimensionnelle)**, le mode correspond au centre de la **classe** qui a la **fréquence moyenne (effectif/ amplitude)** la plus élevée, c'est-à-dire à la zone de la distributions où les valeurs sont le plus **concentrées**. Si l'histogramme comporte plusieurs "pics" séparés par des "creux", la distribution est **multimodale**.



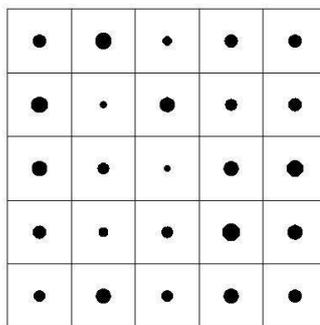
##### *Le cas bidimensionnel*

Dans le cas d'une **distribution spatiale (bidimensionnelle)**, les classes correspondent à des **unités territoriales**, l'amplitude correspond à la **surface** et la fréquence moyenne à la **densité spatiale (effectif/surface)**. Les **points modaux** correspondent alors aux **pics de densité** séparés par des espaces moins denses.

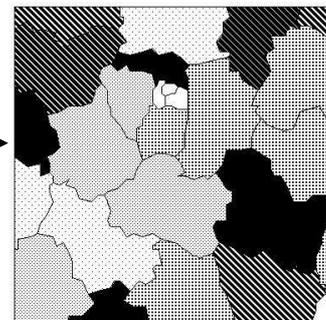
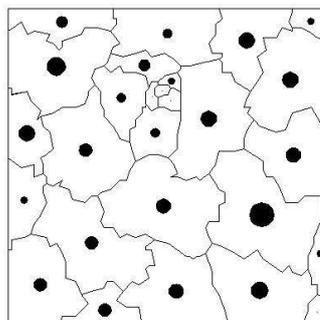
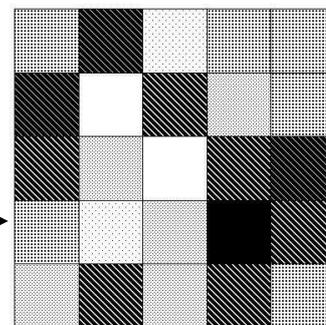
#### Distribution spatiale



#### Effectifs



#### Densité

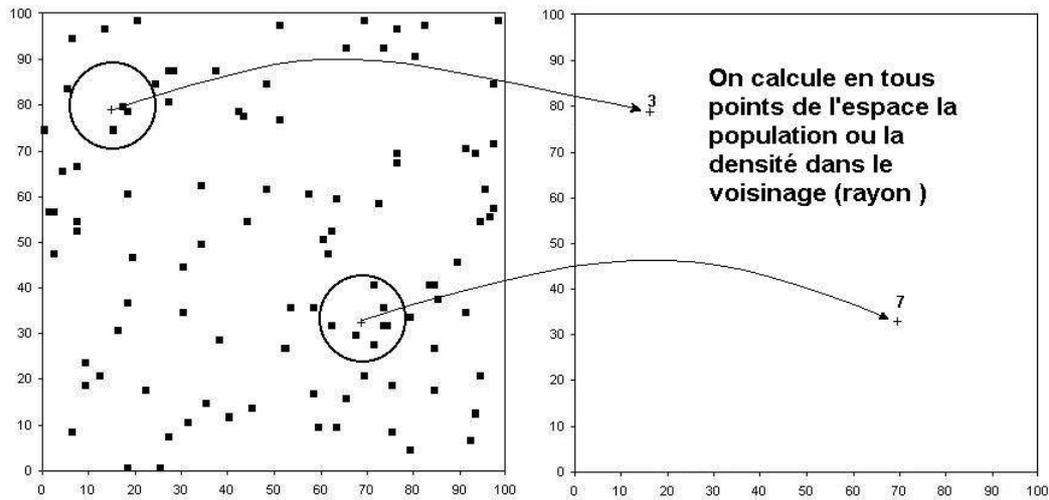


On remarque que le nombre de points modaux et leur position varie en fonction du découpage territorial

**(b) détermination des points modaux sans maillage territorial (lissage)**

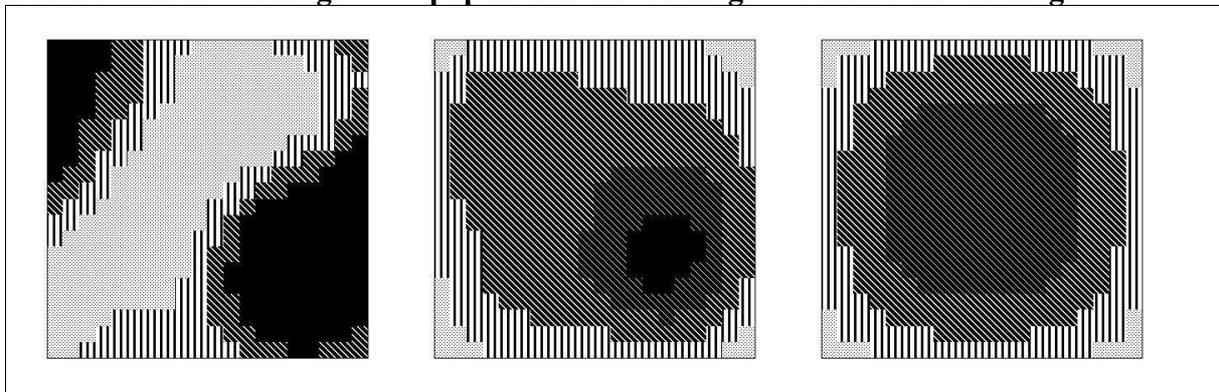
Pour éviter les inconvénients liés aux différences de forme ou de taille des unités spatiales, on peut recourir à des **méthodes de lissage** pour trouver les points nodaux. Ces méthodes sont très nombreuses.

La méthode la plus simple consiste à utiliser une **fenêtre mobile** avec un diamètre fixé par l'utilisateur :



On obtient alors une carte lissée de la **densité de population dans le voisinage** en corrigeant les **effets de bordure**

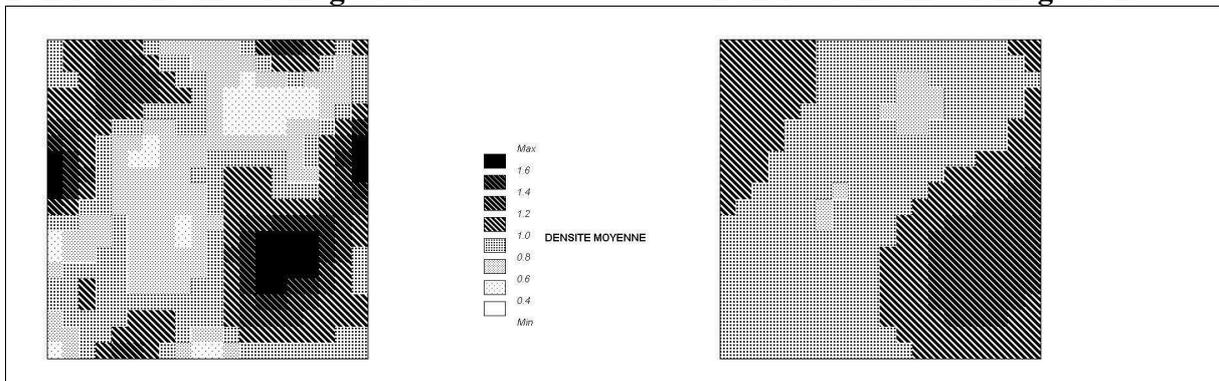
**Densité dans le voisinage = population du voisinage / surface du voisinage**



Le degré de généralisation est plus ou moins important selon le rayon de lissage choisi :

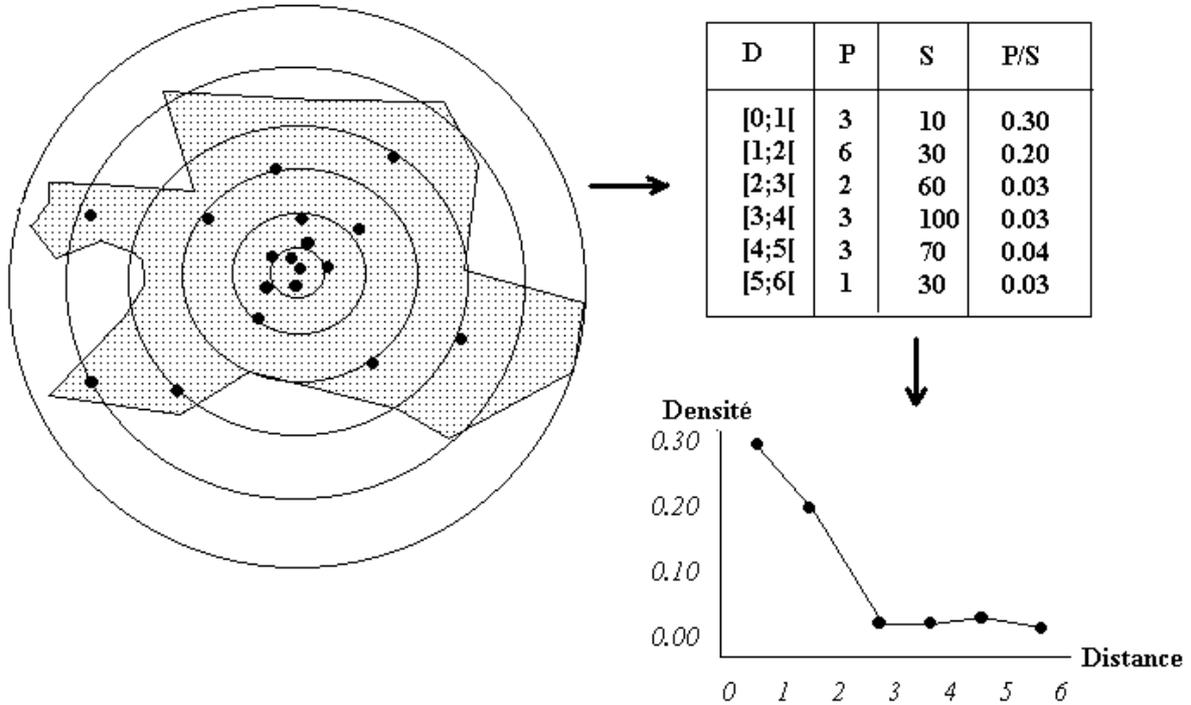
**Densité dans un voisinage de 10**

**Densité dans un voisinage de 20**

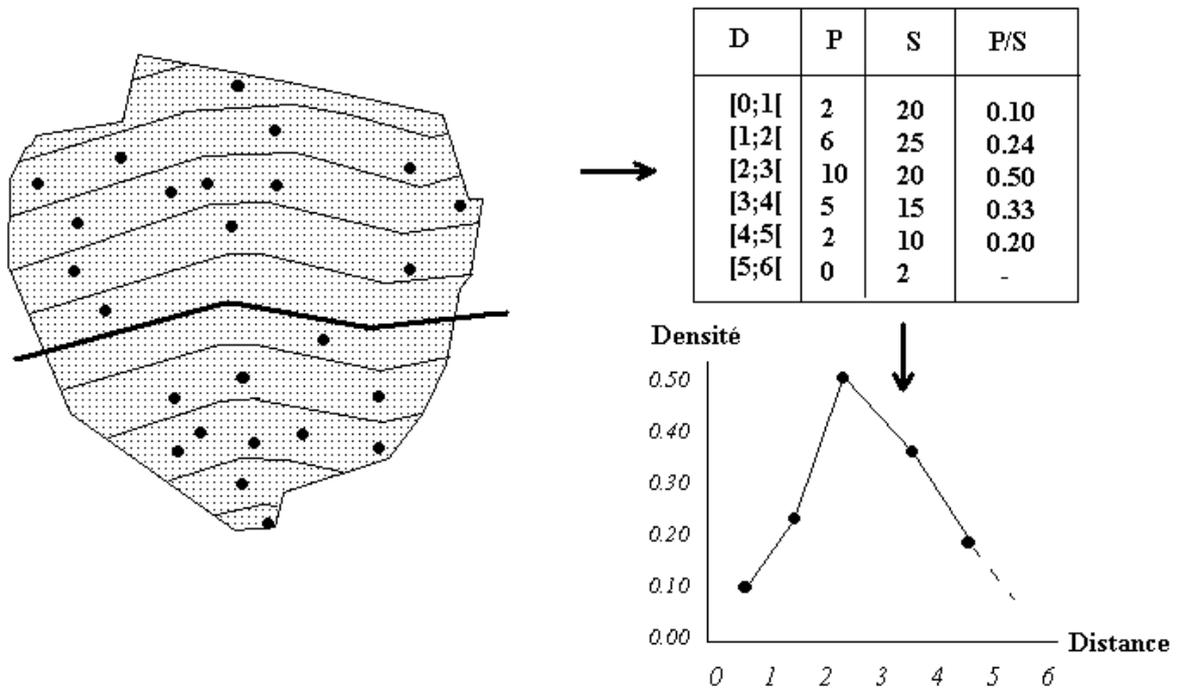


## Document 4 : MÉTHODE D'ANALYSE DES DENSITÉS

### 4.1 Densités par rapport à un point



### 4.2 Densités par rapport à un axe



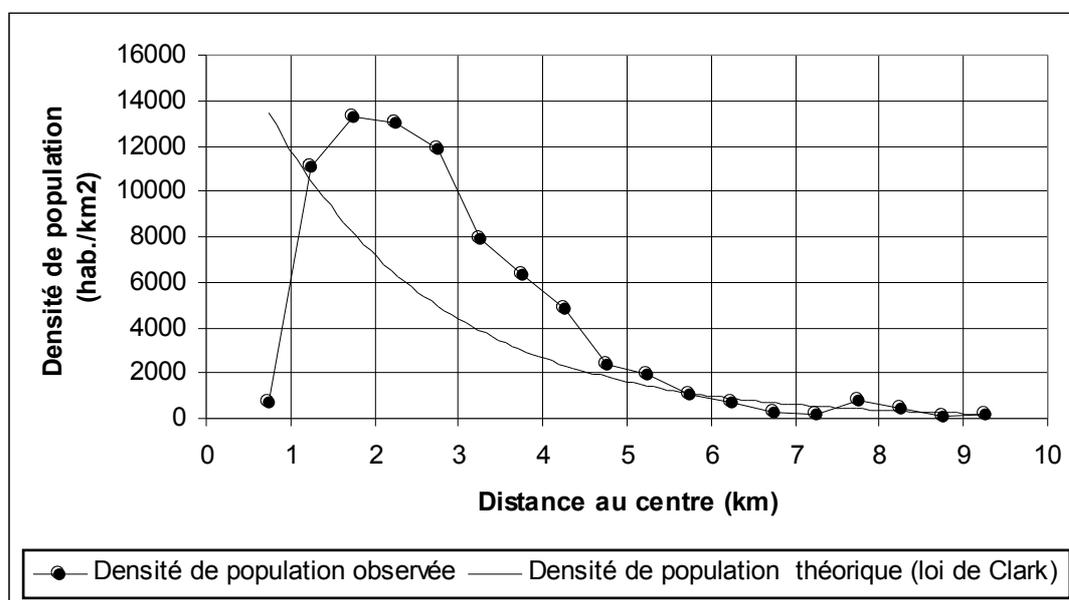
## Document 5 : APPLICATION DE LA LOI DE CLARK À LA POPULATION DE YAOUNDÉ EN 1987

(a) Distribution de la population de Yaoundé en fonction de la distance au centre en 1987

Distance au centre (km)	Surface		Population en 1987		Densité	
	km <sup>2</sup>	%	hab.	%	hab./km <sup>2</sup>	indice 100
[0;1[	3.8	1%	2700	0%	703	29
[1;1.5[	2.5	1%	28000	4%	11111	461
[1.5;2[	6.4	2%	84900	14%	13245	549
[2;2.5[	5.7	2%	74600	12%	12997	539
[2.5;3[	5.7	2%	67700	11%	11877	492
[3;3.5[	18.2	7%	144600	23%	7941	329
[3.5;4[	9.0	3%	57300	9%	6346	263
[4;4.5[	4.2	2%	20200	3%	4798	199
[4.5;5[	15.3	6%	36400	6%	2373	98
[5;5.5[	19.1	7%	37700	6%	1974	82
[5.5;6[	26.5	10%	27300	4%	1032	43
[6;6.5[	14.5	6%	10000	2%	690	29
[6.5;7[	26.8	10%	7600	1%	283	12
[7;7.5[	6.9	3%	1300	0%	190	8
[7.5;8[	3.3	1%	2700	0%	828	34
[8;8.5[	23.2	9%	9500	2%	410	17
[8.5;9[	9.6	4%	1200	0%	125	5
[9;9.5[	58.3	23%	11200	2%	192	8
<b>Total</b>	<b>259.1</b>	<b>100%</b>	<b>624900</b>	<b>100%</b>	<b>2412</b>	<b>100</b>

N.B. Les calculs sont effectués à l'aide du centroïde des chefferies : c'est pourquoi la croissance des surfaces n'est pas régulière.

(b) Forme de la décroissance de la densité de population en fonction de la distance au centre à Yaoundé en 1987



Source : Bopda A., 1997, *Yaoundé dans la construction nationale au Cameroun : territoire urbain et intégration*, Thèse, Université Paris 1, 2 vol.;

## Document 6 : Test de la forme d'une distribution

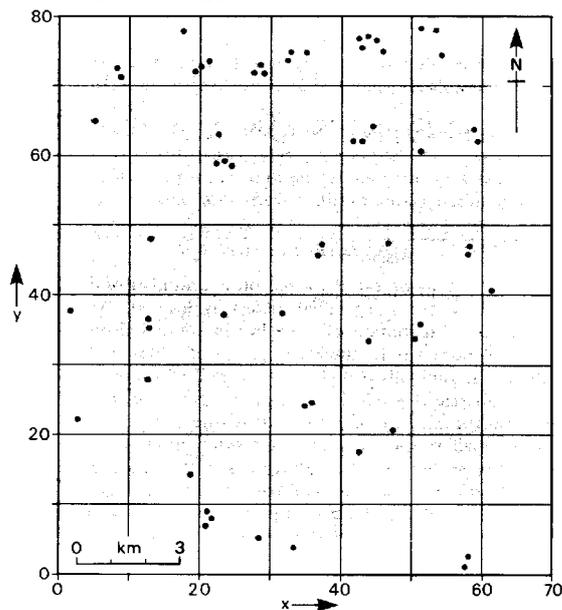
### a) Formes théoriques d'une distribution de points

- ⇒ une distribution est **aléatoire (distribution de Poisson)** si
- (1) tous les emplacements de l'espace ont la même probabilité d'accueillir un point
  - (2) la position d'un point nouveau est indépendante de la position des points précédents
- ⇒ une distribution aura tendance à être **concentrée** si
- (1) certains emplacements de l'espace ont plus de chances d'accueillir un point
  - (2) la localisation d'un premier point favorise l'apparition d'autres points à proximité
- ⇒ une distribution aura tendance à être **régulière** si
- (1) tous les emplacements de l'espace ont la même probabilité d'accueillir un point
  - (2) la localisation d'un premier point défavorise l'apparition d'autres points à proximité

### b) Méthode des quadrats

- (1) Soit un semis de  $N$  points distribués sur un espace  $E$
- (2) On recouvre l'espace  $E$  d'un ensemble de  $K$  mailles de formes régulières. Le nombre moyen de points par maille est égale à  $D=N/K$
- (3) On associe à chaque maille  $i$  le nombre  $D_i$  de points qu'elle contient, puis on calcule la variance du nombre de points par maille  $V(D)$  et on en déduit l'indice de concentration  
 $IC=V(D)/D$   
 **$IC \approx 1$  : distribution aléatoire (loi de Poisson)**  
 **$IC > 1$  : distribution plutôt concentrée**  
 **$IC < 1$  : distribution plutôt régulière**
- (4) On dénombre les nombre de mailles en fonction du nombre de points qu'elles contiennent. On note  $K(n)$  le nombre de mailles comportant  $n$  points.
- (5) On détermine les effectifs théoriques  **$K^*(n)$  qui seraient obtenus si la distribution était aléatoire** à l'aide de la formule suivante :  
$$K^*(n) = K \frac{D^n}{n!} \exp(-D)$$
- (6) On teste l'égalité des deux distributions  $K(n)$  et  $K^*(n)$  à l'aide d'un **test du Chi-2**. En fonction des résultats du test **on conclue au caractère aléatoire ou non aléatoire de la distribution observée.**

c) Exemple d'application de la méthode des quadrats



Carte par point des églises et des chapelles dans le sud du comté de Leicester, UK. (d'après Unwin D., 1981, p. 30)

Nombre de points <b>n</b>	Nombre de quadrats <b>K</b>	Nombre de points <b>n.K</b>	Ecart à la moyenne	
			<b>(n-D)</b>	<b>K(n - D)<sup>2</sup></b>
0	25	0	-1.071	28.676
1	15	15	-0.071	0.076
2	8	16	0.929	6.904
3	5	15	1.929	18.605
4	1	4	2.929	8.579
5	2	10	3.929	30.874
Total	56	60		93.714

Densité moyenne **D** = nb. de points / nb. de quadrats = 60/56 = **1.071**

Variance **V(D)** = 93.714 / 55 = **1.704**

Indice de concentration **IC** = V(D)/D = **1.590**

n	Observé	Théorique	(O-E) <sup>2</sup> /E
0	25	19.2	1.762
1	15	20.6	1.502
2	8	11.0	0.823
3 et +	8	5.2	1.444
Total	56	56	5.531

Nombre de degrés de liberté = nb. de classes - 1 = 4-1=3

$\text{Chi}^2(3, 0.01) = 11.34$

$\text{Chi}^2(3, 0.05) = 7.82$

$\text{Chi}^2(3, 0.10) = 6.25$

$\text{Chi}^2(3, 0.20) = 4.64$

⇒ **Conclusion : La distribution semble concentrée (IC > 1) mais on ne peut pas rejeter l'hypothèse qu'elle soit aléatoire (risque d'erreur > 20%).**

*d) Méthode du plus proche voisin*

1) Soit un semis de  $N$  points distribués sur un espace de surface  $S$ . On note  $d$  la densité moyenne de points par unité de surface à l'intérieur de l'espace considéré ( $d=N/S$ )

(2) On calcule pour chaque point  $i$  la distance  $D_{\min}(i)$  qui le sépare de son voisin le plus proche.

(3) On calcule ensuite la moyenne des distances observées au plus proche voisin  $D_o$

(4) On détermine la distance théorique moyenne au plus proche voisin  $D_T$  dans le cas d'une distribution aléatoire à l'aide de la formule :

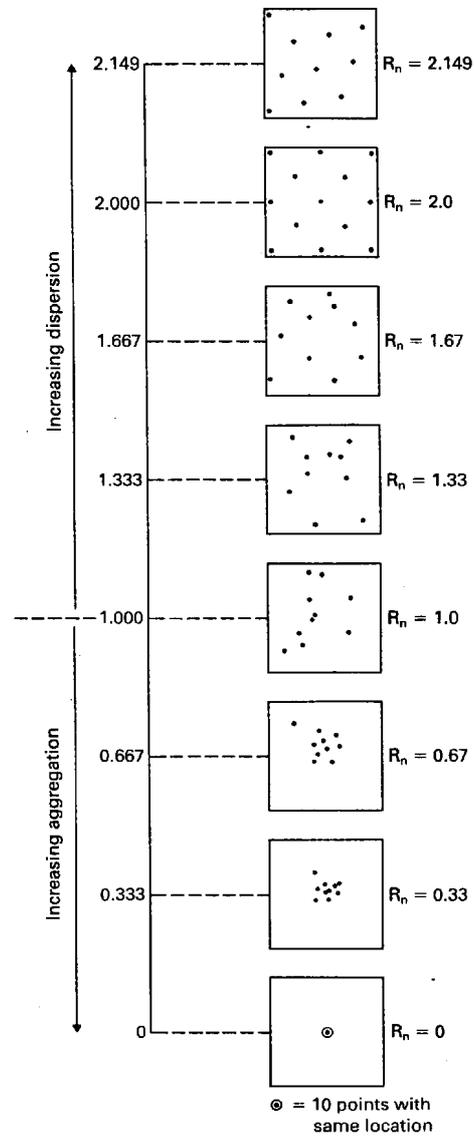
$$D_T = 0.5 / \sqrt{d}$$

(5) On calcule l'indice de dispersion qui est le rapport entre ces deux distances :

$$R = D_o / D_T$$

(6) On peut enfin tester le caractère aléatoire de la distribution à l'aide d'un test paramétrique (Cf. manuels d'analyse spatiale).

**Signification des valeurs de R**

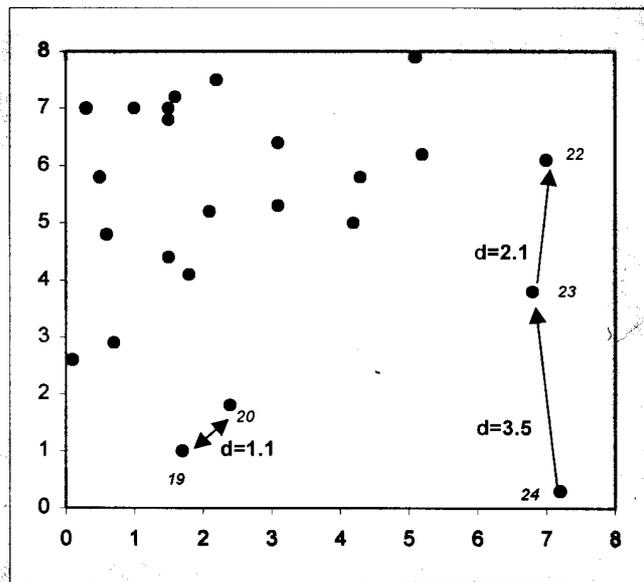


e) Exemple d'application de la méthode du plus proche voisin

(1) Tableau de données

i	Xi	Yi
1	1,5	7
2	1	7
3	1,5	6,8
4	0,5	5,8
5	2,2	7,5
6	0,3	7
7	0,6	4,8
8	1,8	4,1
9	2,1	5,2
10	4,3	5,8
11	1,6	7,2
12	3,1	6,4
13	0,7	2,9
14	0,1	2,6
15	1,5	4,4
16	3,1	5,3
17	5,2	6,2
18	5,1	7,9
19	1,7	1
20	2,4	1,8
21	4,2	5
22	7	6,1
23	6,8	3,8
24	7,2	0,3

(2) détermination des plus proche voisins



(3) calcul des distances au plus proche voisin

i	dmin
1	0,2
2	0,5
3	0,2
4	1,0
5	0,7
6	0,7
7	1,0
8	0,4
9	1,0
10	0,8
11	0,2
12	1,1
13	0,7
14	0,7
15	0,4
16	1,0
17	1,0
18	1,7
19	1,1
20	1,1
21	0,8
22	1,8
23	2,3
24	3,5

moyenne (Ro) 0,99  
écart-type (Eo) 0,74

(4) Calcul des paramètres théoriques de la distribution aléatoire

Surface 64  
effectif 24  
densité 0,375

distances théorique au plus proche voisin

moyenne (Rt) 0,816  
écart-type (Et) 0,087

(5) Calcul de l'indice R

$$R = R_o / R_t = 1,22$$

=> L'indice est supérieur à 1  
ce qui indique une tendance à la dispersion

=> La distribution observée est plus régulière qu'une distribution selon un processus de Poisson

(6) test d'adéquation entre distribution observée et distribution aléatoire

H0 : la distribution observée est aléatoire (processus de Poisson)

On calcule la statistique z

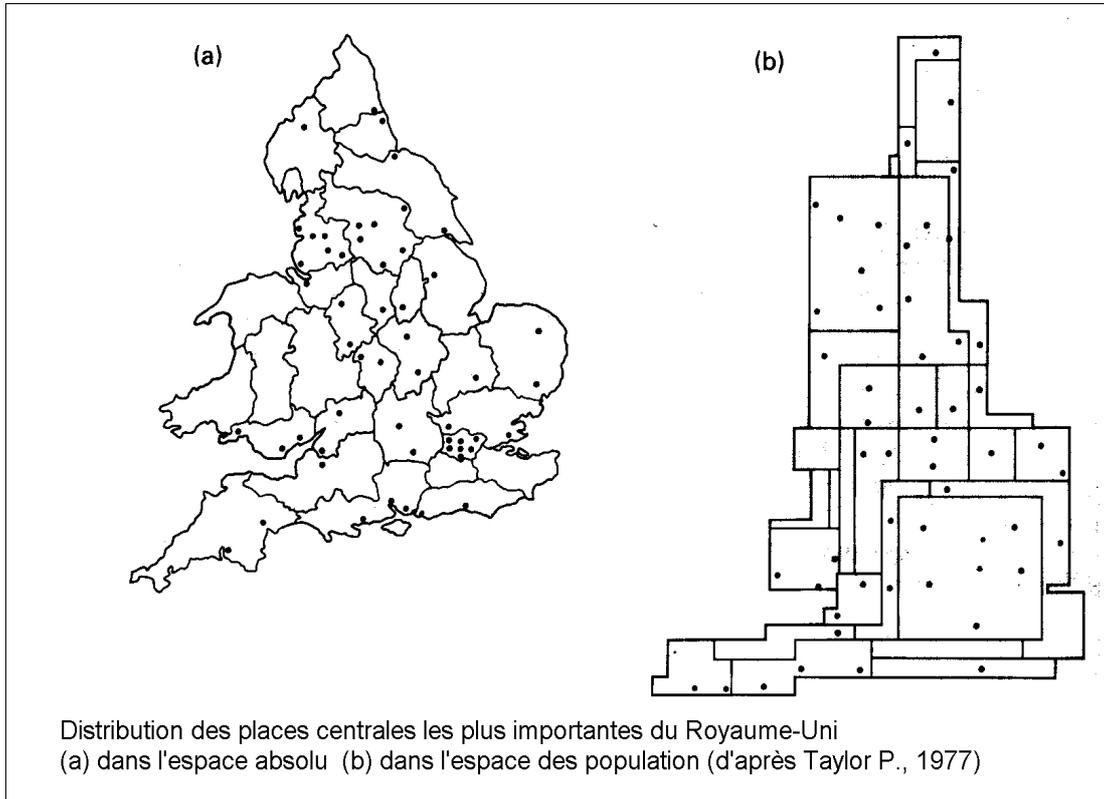
$$z = |R_o - R_t| / E_t$$

et on compare sa valeur à celle de la loi normale pour accepter ou rejeter H0

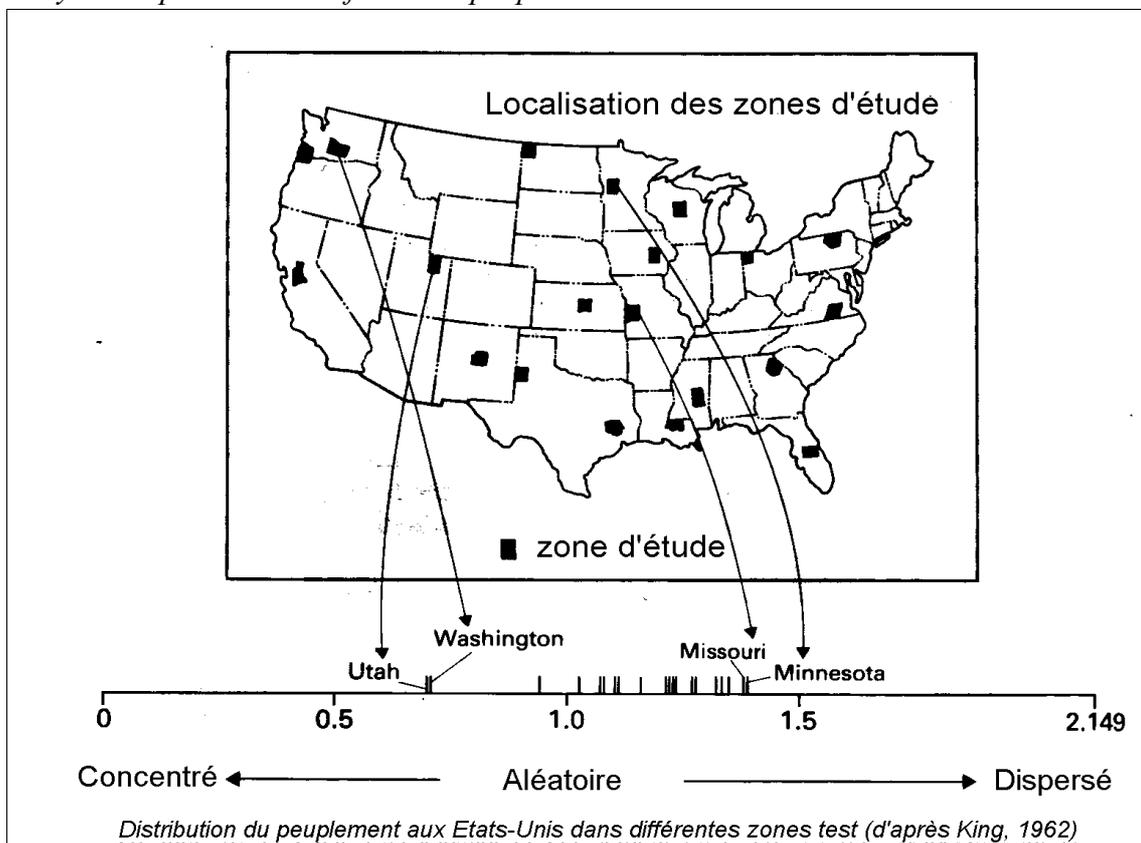
Conclusion : Contrairement à ce que laisserait penser la première impression visuelle, la distribution est plutôt dispersée. Mais son caractère aléatoire ne peut être rejeté.

## Document 7 : Applications des méthodes d'analyse des semis de points

### (a) localisation des centres commerciaux



### (b) Analyse comparative de la forme du peuplement



(c) Mise en évidence d'installations dangereuses

« Dans le Nord de l'Angleterre, certains se sont vivement émus d'une relation possible entre les fuites radioactives de la centrale nucléaire de Sellafield et l'apparition de la leucémie chez l'enfant. Lorsque l'on représentait les cas d'enfants leucémiques sur une carte, des groupements apparaissaient à la périphérie de la centrale. Après de nombreuses enquêtes (...), un verdict d'absence de preuve (...) fut prononcé. Les statisticiens témoignant pour le gouvernement et la centrale nucléaire firent remarquer qu'une répartition causée par un processus purement aléatoire pouvait produire des groupements de points. De plus, en prenant une petite zone de référence, la densité d'enfants leucémiques était très élevée et très au-dessus de la moyenne nationale. Le même nombre d'enfants dans une zone plus vaste donnerait des valeurs plus basses. Les groupements observés étaient simplement le résultat d'un processus aléatoire et du choix arbitraire des aires d'étude. Ils n'étaient, sur le plan statistique et donc (?) scientifique, pas significatifs mais seulement un artefact de la façon avec laquelle ils avaient été estimés dans l'espace géographique... estimèrent les statisticiens au service du gouvernement.

Cependant, cette conclusion ne satisfait pas le géographe Stan Openshaw (1988), l'un des chercheurs les plus créatifs pour fabriquer, grâce à la puissance des ordinateurs modernes, des béquilles à l'esprit humain. En utilisant la carte des cas de leucémie comme un système d'information géographique, il programma l'ordinateur pour parcourir la carte avec une petite grille circulaire, comme une machine à écrire se déplaçant sur une page caractère par caractère. A chaque point, l'ordinateur dénombrait les enfants dans une cellule de la grille et estimait la probabilité de trouver cette valeur par hasard d'après une distribution de Poisson. Si la probabilité était inférieure à 0,002, le traceur de l'ordinateur dessinait à cet endroit sur la carte une cellule de cette dimension et se déplaçait ensuite sur le point suivant, et ainsi de suite. A la fin d'un parcours complet, il repartait avec une grille aux mailles légèrement plus grandes, testait de nouveau chaque cellule, et ainsi de suite, inscrivant un cercle de la taille correspondante en chaque endroit de la carte où il était hautement improbable qu'un groupe de cas se constitue par hasard à cette échelle géographique. Ainsi, neuf millions « d'hypothèses spatiales » furent testées rigoureusement.

Les résultats étaient visibles : directement sur Sellafield, accumulés cellules après cellules, en une large tache noire, indiquant des groupements significatifs à presque toutes les échelles géographiques. Les regroupements n'étaient pas du tout un artefact mais étaient bien réels et dignes de susciter l'intérêt des scientifiques et des médecins (...). Mais l'analyse géographique produisit quelque chose de plus. Sur la côte Est de l'Angleterre septentrionale, il y avait sur la carte un autre pâté noir que personne n'avait imaginé. Au centre, il semblait y avoir un grand incinérateur urbain qui ne fonctionnait pas aux hautes températures pour lesquelles il avait été conçu, libérant ainsi des effluents plastiques et chimiques sur une vaste surface. »

Source : Gould P., 1992, « Epidémiologie et maladie », in Bailly, Ferras, Pumain (dir.), *Encyclopédie de Géographie*, Paris : Economica, pp. 958-959.