



ÉCONOMÉTRIE SPATIALE

Master 2 APE

Chapitre 2

Gilles de Truchis

gilles.detruchis@gmail.com

Site : www.varennnes-ecofin.com/



LES CHAPITRES DU COURS

1. Autocorrélation spatiale

1.1 Concept d'autocorrélation spatiale

1.2 Analyse de l'autocorrélation spatiale

1.3 Diagramme de Moran et l'autocorrélation

1.4 Local Indicators of Spatial Association (LISA)

AUTOCORRÉLATION SPATIALE

CONTENTS

1. Autocorrélation spatiale

1.1 Concept d'autocorrélation spatiale

1.2 Analyse de l'autocorrélation spatiale

1.3 Diagramme de Moran et l'autocorrélation

1.4 Local Indicators of Spatial Association (LISA)

VOISINAGE ET INTERACTION

- ▶ Dépendance spatiale ?
 - ▶ découle de la fonction de densité jointe (multivariée) des variables
 - ⇒ comment détecter une telle interaction ?

- ▶ Autocorrélation spatiale : interaction d'une unité spatiale avec ses voisins
 - ▶ Différent de l'autocorrélation temporelle car **multidirectionnel**
 - ⇒ la différence provient de la noncausalité

 - ▶ S'analyse à l'aide des moments de la fonction de densité multivariée jointe des variables

DANS LA PRATIQUE

- ▶ Les individus doivent être spatialement localisés
 - ▶ unités territoriales discrètes
 - ⇒ on va chercher des similarités ou dissimilarités entre voisins

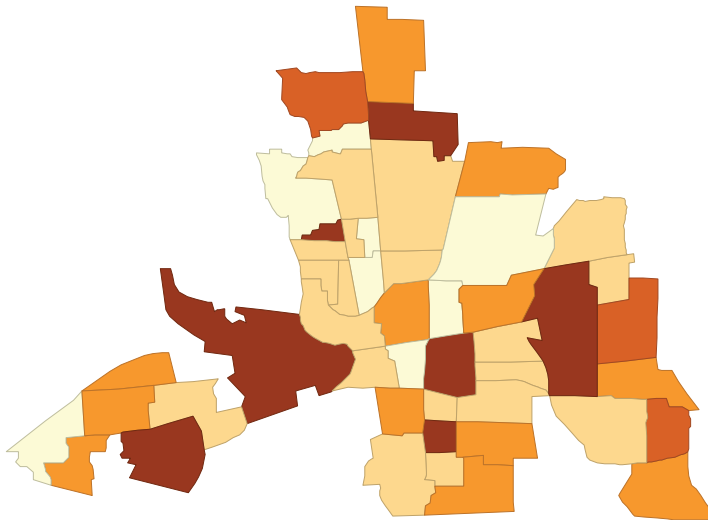
- ▶ Pourquoi une telle interaction ?
 - ▶ phénomène de contagion, débordement géographique, effet d'imitation
 - ⇒ la présence d'autocorrélation doit être testée

RÉPARTITION SPATIALE

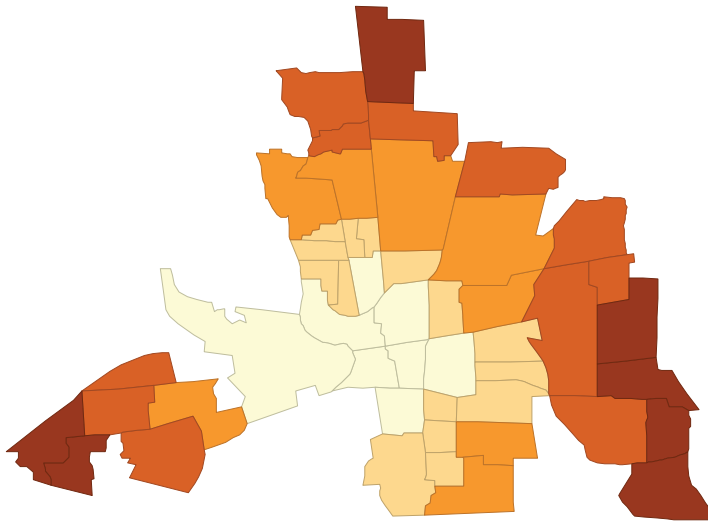
- ▶ La présence d'autocorrélation implique un schéma d'interaction particulier
 - ⇒ les unités spatiales ne sont pas disposées aléatoirement

- ▶ A l'inverse, l'absence d'autocorrélation implique :
 - ▶ une répartition aléatoire
 - ▶ une invariance par permutation aléatoire des individus
 - ⇒ une unité spatiale ne dépend pas de sa position par rapport à ses voisins

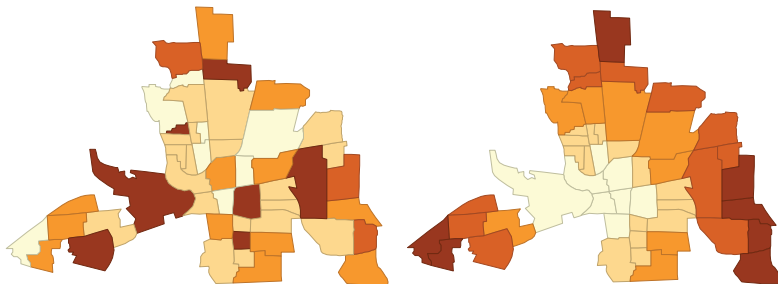
EXEMPLE AUTOCORRÉLATION SPATIALE 1



EXEMPLE AUTOCORRÉLATION SPATIALE 2



EXEMPLE AUTOCORRÉLATION SPATIALE 3



TYPE D'AUTOCORRÉLATION

► Négative

- Lorsqu'on regarde un échiquier, la règle "qui se ressemble s'éloigne prime"
⇒ la dissimilarité systématique avec son voisin révèle une autocorrélation négative

► Positive

- Correspond à l'expression inverse : "qui se ressemble s'assemble"
⇒ les observations similaires tendent à se concentrer dans l'espace

EXEMPLE AUTOCORRÉLATION NÉGATIVE

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25

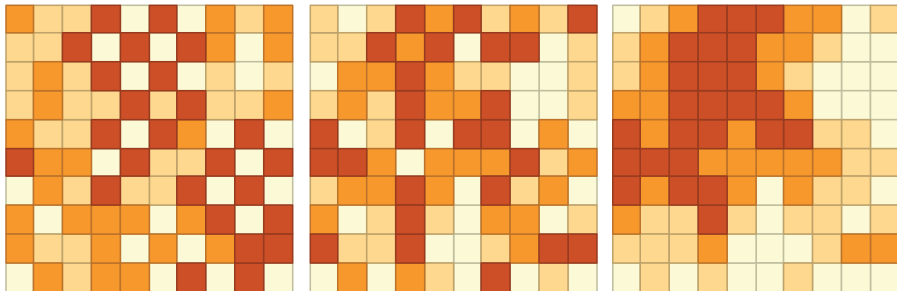
EXEMPLE AUTOCORRÉLATION POSITIVE

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25

EXEMPLE ABSENCE AUTOCORRÉLATION

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25

EXEMPLE SIMULÉ AVEC GEODA



ANALYSE GRAPHIQUE

Plusieurs types de diagramme permettent une analyse préliminaire de l'autocorrélation

- ▶ Diagramme du décalage spatial

- ▶ Il s'agit de comparer une observation à la moyenne pondérée de ses voisins

- ⇒ e.g. comparaison de x_i et $(Wx)_i$ avec W une matrice de pondération

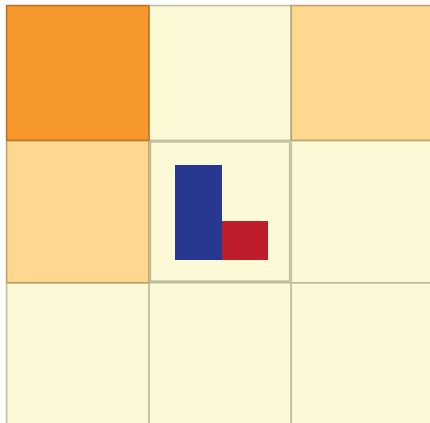
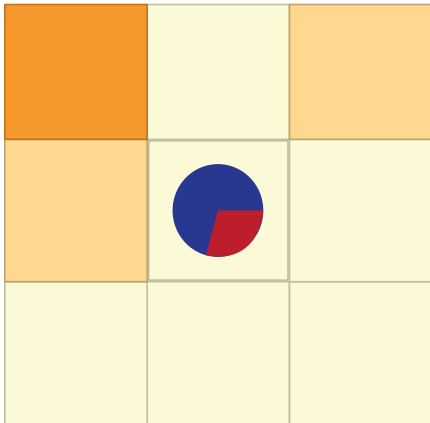
- ⇒ Diagramme en secteurs du décalage spatial

- ▶ x_i et $(Wx)_i$ sont représentés en part d'un disque

- ⇒ Diagramme en barres du décalage spatial

- ▶ x_i et $(Wx)_i$ sont représentés en barres

EXEMPLE DE DIAGRAMME



CONTENTS

1. Autocorrélation spatiale

1.1 Concept d'autocorrélation spatiale

1.2 Analyse de l'autocorrélation spatiale

1.3 Diagramme de Moran et l'autocorrélation

1.4 Local Indicators of Spatial Association (LISA)

OBJECTIF

- ▶ Comment tester la dépendance spatiale ?
 - ▶ Analyse de la nature spatialement aléatoire des unités
 - ⇒ Si aléatoire pas d'autocorrélation
 - ⇒ Sinon autocorrélation

- ▶ Les géostatisticiens ont développé des outils...
 - ▶ La statistique I de Moran
 - ▶ La statistique de Geary
 - ▶ Les indicateurs d'association spatiale
 - ▶ Les indicateurs locaux d'association spatiale (LISA)
 - ▶ ...

- ▶ ... et des méthodes de visualisation
 - ▶ Le diagramme de Moran
 - ▶ Les cartes LISA
 - ▶ ...

PRINCIPE DE L'INFÉRENCE PAR PERMUTATION

- ▶ Soit des réalisations $X_s = \{x_1, \dots, x_n\}$ présentant des **similarités**...
 - ▶ similarités en termes de localisations : W
 - ▶ similarités en termes de valeurs : V
 - ▶ On cherche à savoir si **l'association** de ces similarités reflète une allocation aléatoire
 - ▶ Si c'est le cas on devrait observer une invariance à une permutation
 - ▶ des localisations
 - ▶ ou des valeurs
 - ▶ En procédant à $n!$ permutation on obtient une **distribution permutationnelle**
 - ▶ de cette distribution on détermine des moments permutationnels
 - ▶ à partir de ces moments on construit une inférence par permutation
- ⇒ on peut tester si l'association observée diffère de l'association moyenne de la **distribution permutationnelle**

MESURES DE SIMILARITÉ

- ▶ Dans la pratique, il existe différentes mesures de W et V
- ▶ Concernant la localisation
 - ▶ la matrice W doit rendre compte de la proximité géographique
⇒ nous verrons par la suite les multiples possibilités pour construire W
- ▶ Concernant les valeurs
 - ▶ la matrice V doit rendre compte de l'écart entre les valeurs
 - ▶ mesure de type Geary : $(x_i - x_j)^2$
 - ▶ mesure de type Moran : $x_i x_j$
 - ▶ mesure de type RAS (Royaltey, Astrachan, Sokal 1975) : $|x_i - x_j|$
- ▶ Association des similarités :
 - ▶ Soit $W = w_{ij}$ et $V = v_{ij}$ des mesures de similarité :

$$\Gamma_{LV} = \sum_i \sum_j w_{ij} v_{ij}$$

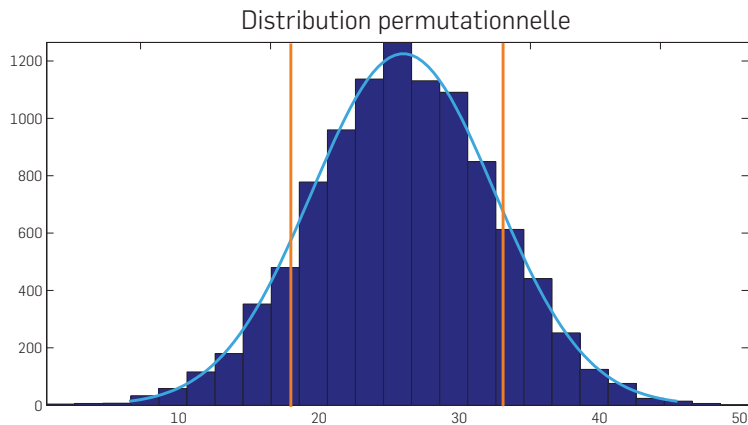
⇒ La statistique Γ mesure l'association entre les similarités W et V

INFÉRENCE PAR PERMUTATION

- ▶ Sous l'hypothèse nulle on a une allocation aléatoire
 - ▶ La référence est donc donnée par la **distribution permutationnelle**
- ▶ On distingue alors
 - ▶ L'association observée : $\Gamma_{LV}^{(o)}$
 - ▶ Les associations tirées des permutations : $\Gamma_{LV}^{(p)}$
- ▶ On peut ainsi créer une **probabilité permutationnelle critique**
 - ▶ Avec M le nombre de permutation
 - ▶ T le nombre de fois où $\Gamma_{LV}^{(p)} \geq \Gamma_{LV}^{(o)}$

$$p^{(p)} = \frac{T + 1}{M + 1}$$

INTUITION GRAPHIQUE



LA STATISTIQUE JOIN COUNT : SIMILARITÉ

- ▶ Valide en présence de variables qualitatives dichotomiques
⇒ variables prenant pour valeurs $\{0, 1\}$
- ▶ Mesures de similarité en termes de localisation
 - ▶ Pour la matrice W , considérons une **contiguïté** à l'ordre 1
⇒ $w_{ij} = 1$ si i et j sont contigus et 0 sinon
- ▶ Mesures de similarité en termes de valeurs
 - ▶ Une mesure de l'effet croisé capturera la similarité unitaire
⇒ $v_{ij}^{(1)} = x_i x_j = 1$ ssi $i = j = 1$ (la réciproque n'est pas vrai)
 - ▶ Une mesure du carré des écart capturera la dissimilarité
⇒ $v_{ij}^{(2)} = (x_i - x_j)^2 = 1$ ssi $i \neq j$ (la réciproque n'est pas vrai)
 - ▶ On introduit une nouvelle mesure pour la similarité nulle
⇒ $v_{ij}^{(3)} = (1 - x_i)(1 - x_j) = 1$ ssi $i = j = 0$ (la réciproque n'est pas vrai)

LA STATISTIQUE JOIN COUNT : ASSOCIATION

- ▶ $v_{ij}^{(1)}$ et $v_{ij}^{(2)}$ fournissent une information complète
 ⇒ l'information de $v_{ij}^{(3)}$ est superflue car déductible de $v_{ij}^{(1)}$ et $v_{ij}^{(2)}$

- ▶ La statistique d'association sera donc
 - ▶ la somme des associations de w_{ij} et $v_{ij}^{(1)}$ **et** w_{ij} et $v_{ij}^{(2)}$

- ▶ En appliquant la formule d'association on obtient la statistique :

$$\Gamma_{LV}^{(o)} = \frac{1}{2} \sum_i \sum_j w_{ij} v_{ij}^{(1)} + \frac{1}{2} \sum_i \sum_j w_{ij} v_{ij}^{(2)}$$

- ▶ En procédant à $n!$ permutation et en recalculant Γ_{LV}
 - ▶ on obtient $n!$ statistiques $\Gamma_{LV}^{(p)}$
 - ⇒ on peut alors construire une distribution de référence

LA STATISTIQUE JOIN COUNT : INFÉRENCE

- ▶ La statistique de test Γ_{LV}
 - ▶ est déterminée sous H_0 : absence d'autocorrélation spatiale
 - ▶ peut s'approximer par une loi normale

$$\Gamma_{LV} \sim \mathcal{N}(\mathbb{E}(\Gamma_{LV}), \mathbb{V}(\Gamma_{LV}))$$

- ▶ On peut donc calculer la z -value

$$z = \frac{\Gamma_{LV} - \mathbb{E}(\Gamma_{LV})}{\sqrt{\mathbb{V}(\Gamma_{LV})}} \sim \mathcal{N}(0, 1)$$

⇒ z mesure la distance (en écart-type) entre une statistique et l'espérance de sa distribution théorique (supposée)

- ▶ Dans la pratique on utilisera la **distribution de référence** issue des permutations

LA STATISTIQUE JOIN COUNT : INFÉRENCE

- Définissons S_0, S_1, S_2 et T_0, T_1, T_2 :

$$S_0 = \sum_{i \neq j} w_{ij}, \quad S_1 = \frac{1}{2} \sum_{i \neq j} (w_{ij} + w_{ji})^2, \quad S_2 = \sum_i \left(\sum_j w_{ij} + \sum_j w_{ji} \right)^2$$

$$T_0 = \sum_{i \neq j} v_{ij}, \quad T_1 = \frac{1}{2} \sum_{i \neq j} (v_{ij} + v_{ji})^2, \quad T_2 = \sum_i \left(\sum_j v_{ij} + \sum_j v_{ji} \right)^2$$

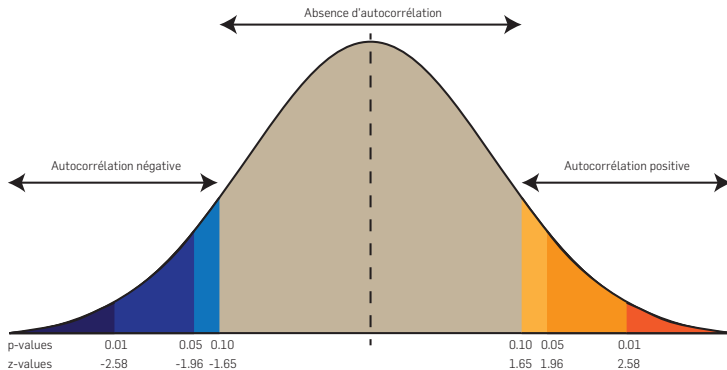
- Le calcul de $\mathbb{E}(\Gamma_{LV})$ est ainsi donné

$$\mathbb{E}(\Gamma_{LV}) = \frac{S_0 T_0}{n(n-1)}$$

- Le calcul de $\mathbb{V}(\Gamma_{LV})$ est ainsi donné

$$\begin{aligned} \mathbb{V}(\Gamma_{LV}) = & \frac{S_1 T_1}{2n(n-1)} + \frac{(S_2 - 2S_1)(T_2 - 2T_1)}{4n(n-1)(n-2)} \\ & + \frac{(S_0^2 + S_1 - S_2)(T_0^2 + T_1 - T_2)}{n(n-1)(n-2)(n-3)} - \mathbb{E}(\Gamma_{LV})^2 \end{aligned}$$

INTUITION GRAPHIQUE



LA STATISTIQUE DE MORAN : SIMILARITÉ

- ▶ Valide en présence de variables quantitatives
- ▶ Mesure de similarité en termes de localisation
 - ▶ Pour la matrice W , considérons une **contiguïté** à l'ordre 1
 $\Rightarrow w_{ij} = 1$ si i et j sont contiguës et 0 sinon
 - ▶ Pour W , on pourra également considérer une standardisation en ligne
 $\Rightarrow W^{(s)} = w_{ij}(\sum_j w_{ij})^{-1}$ ce qui implique que $\sum_j W^{(s)} = 1$
- ▶ Mesure de similarité en termes de valeurs
 - ▶ Considérons une mesure de l'effet croisé des **variables centrées**
 $\Rightarrow v_{ij} = (x_i - \bar{x})(x_j - \bar{x})$ avec \bar{x} la moyenne empirique

LA STATISTIQUE DE MORAN : ASSOCIATION

- En appliquant la formule d'association on obtient la statistique :

$$\Gamma_{LV}^{(o)} = \sum_i \sum_j w_{ij}^{(s)} v_{ij}$$

- La statistique de Moran, notée I , est alors ainsi donnée

$$\begin{aligned} I &= \frac{n}{\sum \sum w_{ij}^{(s)}} \frac{\sum_i \sum_j w_{ij}^{(s)} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \\ &= \frac{\Gamma_{LV}^{(o)}}{S_0} \times \frac{1}{n^{-1} \sum_i (x_i - \bar{x})^2} = \frac{1}{S_0} \Gamma_{LV}^{(o)} \times \frac{1}{\sigma_n^2} \end{aligned}$$

$$\text{car } S_0 = \sum \sum w_{ij}^{(s)}$$

- Notons que le premier terme est proche de $cov(x_i, x_j)$ si contiguë
 \Rightarrow en effet, la standardisation implique que $S_0 = n$
- L'indice I de Moran est donc proche d'un coefficient de corrélation

LA STATISTIQUE DE MORAN : INFÉRENCE

- L'inférence statistique est différente selon les hypothèses retenues :

- Si $X_s = (x_1, \dots, x_n)' \sim \mathcal{N}$

- Sous H_0 (absence d'autocorrélation), on a

$$\mathbb{E}(I) = \frac{-1}{n-1}, \quad \mathbb{V}(I) = \mathbb{E}(I^2) - \mathbb{E}(I)^2, \quad \mathbb{E}(I^2) = \frac{3S_0 + n^2 S_1 - nS_2}{S_0^2(n-1)}$$

- Si $X_s = (x_1, \dots, x_n)' \sim ?$

- Avec m_k le moment d'ordre k on a $\mathbb{E}(I) = \frac{-1}{n-1}$ et $\mathbb{V}(I) = \mathbb{E}(I^2) - \mathbb{E}(I)^2$

$$\mathbb{E}(I^2) = \frac{n[(n^2 - 3n + 3)S_1 - nS_2 + 3S_0^2] - m_4 m_2^{-2}[(n^2 - n)S_1 - 2nS_2 + 6S_0^2]}{n(n-1)(n-2)(n-3)S_0^2}$$

- Dans les deux cas, $I \sim \mathcal{N}(\mathbb{E}(I), \mathbb{V}(I))$ et la z -value est

$$z = \frac{I - \mathbb{E}(I)}{\sqrt{\mathbb{V}(I)}} \sim \mathcal{N}(0, 1)$$

- Dans la pratique on utilisera la **distribution de référence** issue des permutations

LA STATISTIQUE DE MORAN : INTERPRÉTATION

- Pour analyser le I de Moran, on le compare à

$$\mathbb{E}(I) = \frac{-1}{n-1}$$

- Sous l'hypothèse nulle d'absence d'autocorrélation...

- ... on conclura à une autocorrélation spatiale positive si $I > \mathbb{E}(I)$
 - On regardera aussi le signe et la magnitude de la z -value ($z > 0$)
- ... on conclura à une autocorrélation spatiale négative si $I < \mathbb{E}(I)$
 - On regardera aussi le signe et la magnitude de la z -value ($z < 0$)

LA STATISTIQUE DE GEARY : SIMILARITÉ

- ▶ Valide en présence de variables quantitatives
- ▶ Mesure de similarité en termes de localisation
 - ▶ Pour la matrice W , considérons une **contiguïté** à l'ordre 1
 $\Rightarrow w_{ij} = 1$ si i et j sont contiguës et 0 sinon
 - ▶ Pour W , on pourra également considérer une standardisation en ligne
 $\Rightarrow W^{(s)} = w_{ij}(\sum_j w_{ij})^{-1}$ ce qui implique que $\sum_j W^{(s)} = 1$
- ▶ Mesure de similarité en termes de valeurs
 - ▶ Considérons une mesure de proximité quadratique
 $\Rightarrow v_{ij} = (x_i - x_j)^2$ avec \bar{x} la moyenne empirique

LA STATISTIQUE DE GEARY : ASSOCIATION

- En appliquant la formule d'association on obtient la statistique :

$$\Gamma_{LV}^{(o)} = \sum_i \sum_j w_{ij}^{(s)} v_{ij}$$

- La statistique de Geary, notée C , est alors ainsi donnée

$$\begin{aligned} C &= \frac{n-1}{2 \sum \sum w_{ij}^{(s)}} \frac{\sum_i \sum_j w_{ij}^{(s)} (x_i - x_j)^2}{\sum_i (x_i - \bar{x})^2} \\ &= \frac{\Gamma_{LV}^{(o)}}{2S_0} \times \frac{1}{(n-1)^{-1} \sum_i (x_i - \bar{x})^2} = \frac{\Gamma_{LV}^{(o)}}{2S_0 \sigma_{n-1}^2} \end{aligned}$$

$$\text{car } S_0 = \sum \sum w_{ij}^{(s)}$$

LA STATISTIQUE DE GEARY : INFÉRENCE

- L'inférence statistique est différente selon les hypothèses retenues :

- Si $X_s = (x_1, \dots, x_n)' \sim \mathcal{N}$

- Sous H_0 (absence d'autocorrélation), on a

$$\mathbb{E}(C) = 1, \quad \mathbb{V}(C) = \frac{(2S_1 + S_2)(n-1) - 4S_2}{2(n+1)S_0}$$

- Si $X_s = (x_1, \dots, x_n)' \sim ?$

- Sous H_0 (absence d'autocorrélation), on a $\mathbb{E}(C) = 1$ et

$$\mathbb{V}(C) = \frac{\left[(n-1)S_1(n^2 - 3n + 3 - (n-1)m_4m_2^{-2}) \right] + \left[S_0^2(n^2 - 3 - (n-1)^2m_4m_2^{-2}) \right]}{n(n-2)(n-3)S_0^2} - \frac{\left[\frac{1}{4}((n-1)S_2(n^2 + 3n - 6 - (n^2 - n + 2)m_4m_2^{-2})) \right]}{n(n-2)(n-3)S_0^2}$$

- Dans les deux cas, $C \sim \mathcal{N}(\mathbb{E}(C), \mathbb{V}(C))$ et $z = \frac{C - \mathbb{E}(C)}{\sqrt{\mathbb{V}(C)}} \sim \mathcal{N}(0, 1)$

- Dans la pratique on utilisera la **distribution de référence** issue des permutations

LA STATISTIQUE DE GEARY : INTERPRÉTATION

- Pour analyser le C de Geary, on le compare à

$$\mathbb{E}(C) = 1$$

- Notons que la plupart du temps $0 < C < 2$
- Sous l'hypothèse nulle d'absence d'autocorrélation...
 - ... on conclura à une autocorrélation spatiale positive si $0 < C < 1$
 - On regardera aussi le signe et la magnitude de la z -value ($z > 0$)
 - ... on conclura à une autocorrélation spatiale négative si $1 < C < 2$
 - On regardera aussi le signe et la magnitude de la z -value ($z < 0$)

CONTENTS

1. Autocorrélation spatiale

1.1 Concept d'autocorrélation spatiale

1.2 Analyse de l'autocorrélation spatiale

1.3 Diagramme de Moran et l'autocorrélation

1.4 Local Indicators of Spatial Association (LISA)

LE DIAGRAMME DE MORAN

- ▶ En forme matricielle, le I de Moran s'écrit

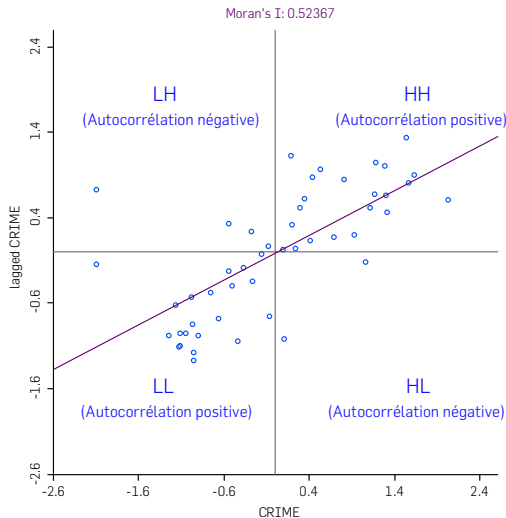
$$I = \frac{n}{S_0} \frac{Z' W Z}{Z' Z}$$

avec $Z = (x_1 - \bar{x}, \dots, x_n - \bar{x})'$

- ▶ En supposant W standardisée, $S_0 = n$ et on a $I = Z' W Z (Z' Z)^{-1}$
 - ▶ Le I de Moran est alors une sorte de MCO sur les observations contiguës
- ▶ Avec un **scatter plot** on peut représenter les observations et la droite de régression dépendant de I
- ▶ On voit alors apparaître 4 quadrants :
 - ▶ LH : autocorrélation spatiale négative
 - ▶ HH : autocorrélation spatiale positive
 - ▶ LL : autocorrélation spatiale positive
 - ▶ HL : autocorrélation spatiale négative

INTUITION GRAPHIQUE

Crime data for 49 neighborhoods in Columbus, Ohio, 1980
Residential burglaries and vehicle thefts per 1000 households



INTUITION GRAPHIQUE

Crime data for 49 neighborhoods in Columbus, Ohio, 1980

Residential burglaries and vehicle thefts per 1000 households

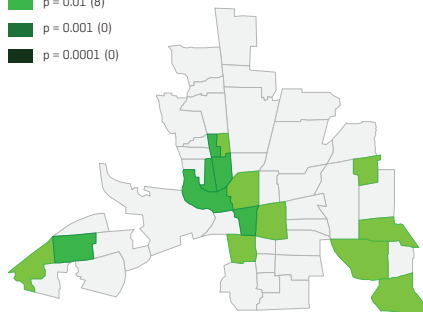
Not Significant (31)

$p = 0.05$ (10)

$p = 0.01$ (8)

$p = 0.001$ (0)

$p = 0.0001$ (0)



Moran's Significance Map

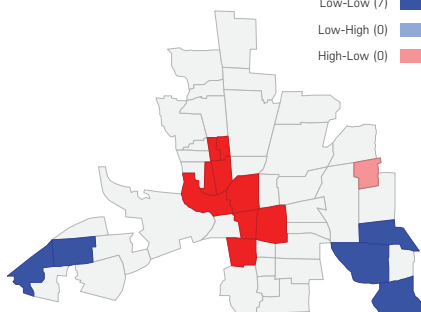
Not Significant (31)

High-High (11)

Low-Low (7)

Low-High (0)

High-Low (0)

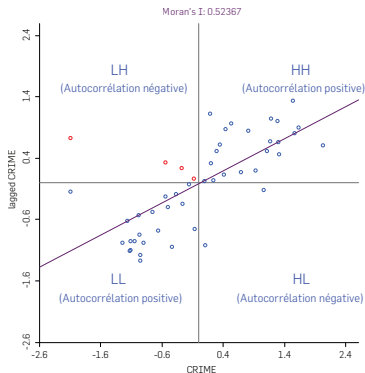


Moran's Cluster Map

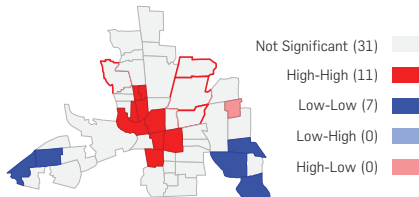
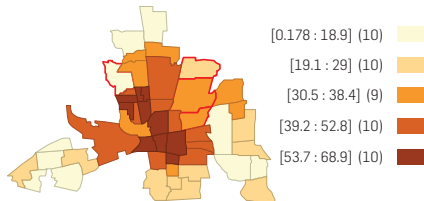
INTUITION GRAPHIQUE

Crime data for 49 neighborhoods in Columbus, Ohio, 1980

Residential burglaries and vehicle thefts per 1000 households



Crime Quantile Map

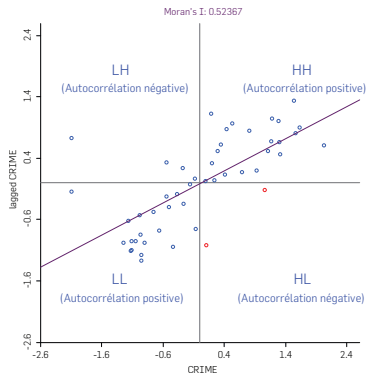


Moran's Cluster Map

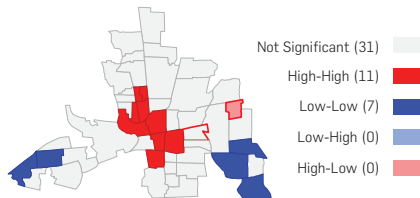
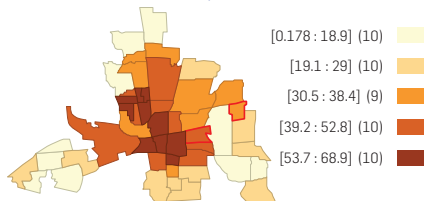
INTUITION GRAPHIQUE

Crime data for 49 neighborhoods in Columbus, Ohio, 1980

Residential burglaries and vehicle thefts per 1000 households



Crime Quantile Map

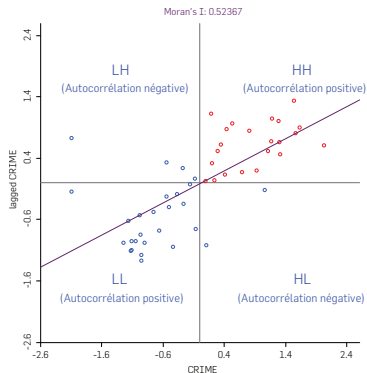


Moran's Cluster Map

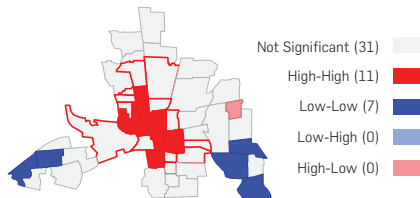
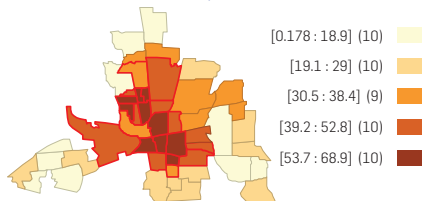
INTUITION GRAPHIQUE

Crime data for 49 neighborhoods in Columbus, Ohio, 1980

Residential burglaries and vehicle thefts per 1000 households



Crime Quantile Map

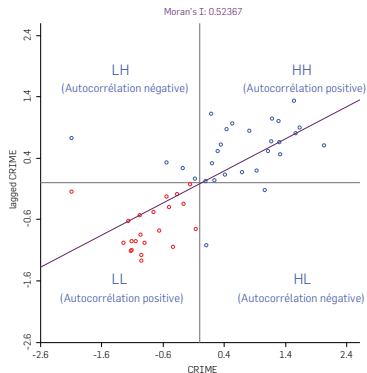


Moran's Cluster Map

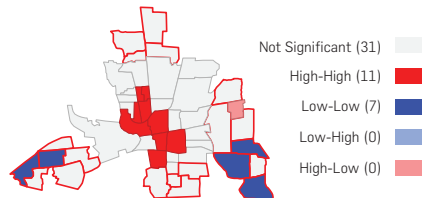
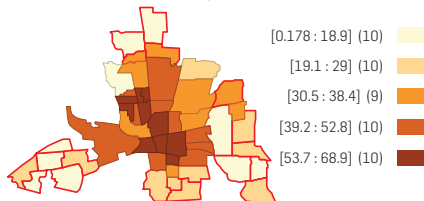
INTUITION GRAPHIQUE

Crime data for 49 neighborhoods in Columbus, Ohio, 1980

Residential burglaries and vehicle thefts per 1000 households



Crime Quantile Map



Moran's Cluster Map

CONTENTS

1. Autocorrélation spatiale

1.1 Concept d'autocorrélation spatiale

1.2 Analyse de l'autocorrélation spatiale

1.3 Diagramme de Moran et l'autocorrélation

1.4 Local Indicators of Spatial Association (LISA)

INTRODUCTION DES LISA

- ▶ Les indicateurs Γ , I et C sont des indicateurs globaux
 - ▶ la totalité de l'échantillon est considéré pour déduire le schéma d'autocorrélation
 - ⇒ l'échantillon est supposé homogène
- ▶ Pour relâcher cette contrainte, la littérature a proposé des indicateurs locaux
 - ⇒ Local Indicators of Spatial Association (LISA)
 - ▶ le schéma d'autocorrélation est évalué pour chaque observation
 - ⇒ l'hétérogénéité de l'échantillon peut apparaître

DÉFINITION DES LISA

- ▶ Anselin (1995, Geographical Analysis) propose une définition basée sur deux conditions
 - ▶ the Lisa for each observations gives an indication of the extent of significant spatial clustering of similar values **around that observation**
 - ▶ the sum of LISAs for all observations is proportional to a global indicator of spatial association
- ▶ Les statistiques LISA peuvent identifier deux types phénomènes :
 - ▶ des clusters locaux déviants du schéma d'autocorrélation globale
 - ▶ des phénomènes atypiques et extrêmes (outliers) d'association ou de dissociation

STATISTIQUE LOCALE DE MORAN

- Commençons par définir une version locale de la statistique Γ
 - pour une observation x_i , la version locale est $\Gamma_i = \sum_j w_{ij} v_{ij}$
- Il est alors possible de construire une statistique de Moran locale
 - Anselin (1995) montre qu'elle est de la forme

$$I_i = \frac{(x_i - \bar{x})}{m_2} \sum_j w_{ij} (x_j - \bar{x})$$

ce qui révèle que $\sum_i I_i = \gamma I$, $\gamma = S_0 m_2$

- $\mathbb{E}(I_i)$ et $\mathbb{V}(I_i)$ deviennent alors

$$\mathbb{E}(I_i) = - \sum_j w_{ij} (n-1)^{-1}$$

$$\mathbb{V}(I_i) = \frac{- \sum_j w_{ij}^2 (n - \frac{m_4}{m_2})}{n-1} + \frac{\sum_{k \neq i} \sum_{h \neq i} w_{ik} w_{ih} (2 \frac{m_4}{m_2} - n)}{(n-1)(n-2) - (\sum_j w_{ij})^2 (n-1)^{-2}}$$

- L'approximation normale n'est pas soutenable pour $z = \frac{C - \mathbb{E}(I_i)}{\sqrt{\mathbb{V}(I_i)}}$

STATISTIQUE LOCALE DE GEARY

- Il est également possible de construire une statistique de Geary locale
 - Anselin (1995) montre qu'elle est de la forme

$$C_i = \frac{1}{m_2} \sum_j w_{ij} (x_i - x_j)^2$$

- Le facteur de proportion avec l'indice global est le suivant :

$$\sum_i C_i = \gamma C, \quad \gamma = \frac{2nS_0}{n-1}$$

- $\mathbb{E}(C_i)$ et $\mathbb{V}(C_i)$ deviennent alors ?
 - Interprétation générale : selon le signe, le LISA va révéler
 - des clusters de similarité
 - des clusters de dissimilarité
- ⇒ en présence de nombreux clusters regroupé on parle de **hot-spot**

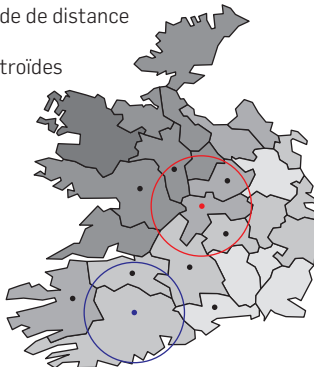
LIMITES DES STATISTIQUES LISA

- ▶ Les LISA présentent néanmoins des limites
 - ▶ Lorsqu'un **hot-spot** apparaît, il est difficile de comparer les LISA
 - ⇒ les statistiques sont corrélées entre elles
 - ▶ Les seuils de significativité permutationnelle ont tendance à être laxistes
 - ⇒ correction avec un ajustement de Bonferroni : α/m , où m est le nombre de tests
 - ▶ L'autocorrélation spatiale univariée pourrait venir d'une association multivariée
 - ⇒ tenir compte de l'effet de variables additionnelles
 - ⇒ nécessite d'introduire le concept de **régression spatiale**

LA STATISTIQUE GETIS-ORD

- Getis et Ord : statistiques de type **High/Low clustering**
 - mesure la concentration de valeurs hautes ou basses autour de i : **Hot Spot**
 - Cette mesure repose sur une matrice W de type **bandes de distance** et non de type contiguïté

Bande de distance
et
Centroïdes



LES STATISTIQUES GETIS-ORD

- Il s'agit d'un couple de statistiques notées G_i et G_i^*
- Getis et Ord (1992) proposent une première version :

$$G_i(d) = \frac{\sum_{j \neq i} w_{ij}(d) x_j}{\sum_{j \neq i} x_j}, \quad G_i^*(d) = \frac{\sum_j w_{ij}(d) x_j}{\sum_j x_j}$$

- Limites
 - W dépend du choix d'une distance seuil d
 - W est binaire : 1 ou 0 selon que la distance qui sépare i et j est inférieur ou supérieur à d (la distance de i à lui-même vaut 0 dans la matrice)
 - Les observations doivent être positives : $X > 0$

LES STATISTIQUES GETIS-ORD : INFÉRENCE

- Sous H_0 : absence d'autocorrélation, Getis et Ord (1992) montrent que (pour $G_i^*(d)$ également) :

$$\begin{aligned}\mathbb{E}(G_i) &= \frac{\sum_j w_{ij}(d)}{n-1} \\ \mathbb{V}(G_i) &= \frac{1}{(\sum_j x_j)^2} \left(\frac{\sum_j w_{ij}(d)(n-1 - \sum_j w_{ij}(d)) \sum_j x_j^2}{(n-1)(n-2)} \right) \\ &\quad + \frac{\sum_j w_{ij}(d)(\sum_j w_{ij}(d) - 1)}{(n-1)(n-2)} - \frac{(\sum_j w_{ij}(d))^2}{(n-1)^2}\end{aligned}$$

- Sous l'hypothèse nulle, $G_i \sim \mathcal{N}(\mathbb{E}(G_i), \mathbb{V}(G_i))$ et

$$z = \frac{G_i - \mathbb{E}(G_i)}{\sqrt{\mathbb{V}(G_i)}} \sim \mathcal{N}(0, 1)$$

LES STATISTIQUES GETIS-ORD GÉNÉRALISÉES

- Il s'agit d'un couple de statistiques notées G_i et G_i^*
- Ord et Getis (1995) proposent une généralisation (pour $G_i^*(d)$ également) :

$$G_i(d) = \frac{\sum_{j \neq i} w_{ij}(d) x_j - \bar{x} \sum_{j \neq i} w_{ij}(d)}{\sqrt{m_2} \sqrt{\left(n \sum_{j \neq i} w_{ij}^2 - \left(\sum_{j \neq i} w_{ij} \right)^2 \right) (n-1)^{-1}}}$$

avec $\sqrt{m_2} = \sqrt{n^{-1} \sum_j x_j^2 - \bar{x}^2}$

- Avantages
 - W n'est plus forcément binaire
 - X n'est plus forcément positive

LES STATISTIQUES GETIS-ORD : INTERPRÉTATION

- Sous l'hypothèse nulle, $G_i \sim \mathcal{N}(\mathbb{E}(G_i), \mathbb{V}(G_i))$ et

$$z = \frac{G_i - \mathbb{E}(G_i)}{\sqrt{\mathbb{V}(G_i)}} \sim \mathcal{N}(0, 1)$$

⇒ Selon le signe et la magnitude de z

- on conclura à une association spatiale positive ($z > 0$)

⇒ Un z positif élevé révélera que des valeurs élevées de x_j sont dans la bande de distance du point i

- on conclura à une association spatiale négative ($z < 0$)

⇒ Un z négatif élevé révélera que des valeurs faibles de x_j sont dans la bande de distance du point i

- Limites :

- Comment choisir d ?

- Comment comparer les statistiques (corrélées si éléments communs)

- Pour corriger (augmenter) la sévérité des tests on peut utiliser

⇒ ajustement de Bonferroni : α/m , où m est le nombre de tests et α le seuil de significativité retenu

⇒ ajustement de Šidák : $1 - (1 - \alpha)^{1/m}$, où m est le nombre de tests

INTUITION GRAPHIQUE

Crime data for 49 neighborhoods in Columbus, Ohio, 1980 Residential burglaries and vehicle thefts per 1000 households

