



Schweizerische Eidgenossenschaft  
Confédération suisse  
Confederazione Svizzera  
Confederaziun svizra

Département fédéral des finances DFF  
**Administration fédérale des contributions AFC**  
Politique fiscale

Berne, le 28 juin 2016

---

# **Vers une détection automatique des particularités locales de l'imposition suisse**

Une application de la géostatistique à l'analyse des données fiscales

---

**Rudi Peters**

Division Economie et statistique fiscale  
E-mail: [rudi.peters@estv.admin.ch](mailto:rudi.peters@estv.admin.ch) / Tél.: 058 462 73 87

Ce document ne reflète pas nécessairement la position officielle de l'Office, du Département ou du Conseil fédéral. Les thèses et les éventuelles inexactitudes contenues dans ce document n'engagent que son auteur.

## Résumé

Nous montrons dans ce document comment détecter aisément et automatiquement les particularités locales d'un territoire donné, c'est-à-dire comment trouver les lieux où, simultanément, la variable d'intérêt est plus grande ou plus petite que la moyenne du territoire et où elle est en contraste par rapport à son voisinage. Nous illustrons la méthode sur quatre exemples issus de la statistique fiscale.

Ce document sensibilise le lecteur au concept d'autocorrélation spatiale et à l'utilisation de diagramme de dispersion de Moran et de carte LISA pour étudier les associations spatiales locales. Il permet également au profane de comprendre la démarche rigoureuse du statisticien pour juger de la significativité des phénomènes observés.

L'approche de ce document est purement descriptive (statistique descriptive). Elle ne vise qu'à montrer les associations spatiales entre les observations, sans chercher à expliquer les phénomènes locaux. Dans un deuxième temps, on pourrait s'intéresser à développer des modèles économétriques explicatifs qui prennent en compte la dépendance spatiale des observations.

## Zusammenfassung

Wir zeigen in diesem Dokument, wie auf einfache Weise und automatisch lokale Besonderheiten eines bestimmten Territoriums aufgedeckt werden können, d.h. wie die Gebiete identifiziert werden können, in welchen die interessierende Variable gleichzeitig grösser oder kleiner ist als der Durchschnitt des betrachteten Territoriums und wo sich diese Gebiete deutlich von ihren Nachbargebieten abheben. Wir illustrieren die Anwendung dieser Methode mit vier Beispielen aus der Statistik zur Steuerbelastung.

Das Dokument sensibilisiert den Leser für das Konzept der räumlichen Autokorrelation und die Anwendung von Streudiagrammen für den Index von Moran sowie die Karten LISA zwecks Untersuchung von lokalen räumlichen Zusammenhängen. Gleichzeitig erlaubt es dem Nichtfachmann, die rigorose Vorgehensweise des Statistikers zur Beurteilung der Signifikanz von beobachteten Phänomenen zu verstehen.

Der Ansatz in diesem Papier ist rein deskriptiv (deskriptive Statistik). Es werden lediglich räumliche Korrelationen zwischen Beobachtungen aufgezeigt, ohne zu versuchen, diese lokalen Besonderheiten zu erklären. In einem nächsten Schritt könnten ökonomische Modelle entwickelt werden, welche die räumliche Abhängigkeit zwischen den Beobachtungen explizit berücksichtigen.

## Table des matières

<b>1</b>	<b>Introduction</b>	<b>6</b>
<b>2</b>	<b>L'autocorrélation spatiale globale</b>	<b>8</b>
2.1	L'indice de Moran .....	8
2.2.1	Définition .....	8
2.2.2	Exemple : l'autocorrélation spatiale de la charge fiscale dans les cantons.....	12
2.2	La significativité statistique de l'indice de Moran .....	15
2.2.1	Le principe du test statistique .....	15
2.2.2	L'estimation de la valeur-p.....	17
<b>3</b>	<b>Les associations spatiales locales</b>	<b>21</b>
3.1	L'indice LISA.....	21
3.1.1	Définition .....	21
3.1.2	Exemple : les indices LISA de la charge fiscale à Berne .....	23
3.2	Les types d'association spatiale .....	25
3.3	Le diagramme de dispersion de Moran.....	27
3.4	La significativité statistique de l'association spatiale .....	29
3.4.1	Le test statistique .....	29
3.4.2	Exemple : les associations locales significatives de la charge fiscale à Berne ...	30
<b>4</b>	<b>Quelques exemples d'analyse spatiale</b>	<b>36</b>
4.1	Les charges fiscales localement singulières, par canton.....	36
4.2	Les lieux de hauts ou de faibles revenus.....	38
4.3	Les lieux où préfèrent vivre les « gros contribuables » .....	40
4.4	Une exploitation des déductions fiscales différente sur le territoire .....	43
<b>5</b>	<b>Bibliographie</b>	<b>47</b>

## Abréviations des noms de canton

ZH	Zurich	SH	Schaffhouse
BE	Berne	AR	Appenzell R.-Ext.
LU	Lucerne	AI	Appenzell R.-Int.
UR	Uri	SG	Saint-Gall
SZ	Schwyz	GR	Grisons
OW	Obwald	AG	Argovie
NW	Nidwald	TG	Thurgovie
GL	Glaris	TI	Tessin
ZG	Zoug	VD	Vaud
FR	Fribourg	VS	Valais
SO	Soleure	NE	Neuchâtel
BS	Bâle-Ville	GE	Genève
BL	Bâle-Campagne	JU	Jura

## Remerciements

Nous remercions vivement M. Markus Eichenberger pour son appui au calcul et à la visualisation des résultats (utilisation du logiciel SAS).

Pour une bonne lecture du document, nous recommandons un visionnage en couleur des Graphiques.

# 1 Introduction

Dans ce document, nous montrons comment analyser la répartition spatiale d'une variable socio-économique sur un territoire donné (par exemple : la Suisse ou un canton du pays) et faire ressortir les particularités locales significatives d'un point de vue statistique.

Les valeurs des entités spatiales (par exemple : les communes de la Suisse ou du canton) se répartissent-elles de façon aléatoire sur le territoire ou interfèrent-elles en fonction de leur proximité géographique ? Les différences entre des entités proches sont-elles plus faibles (ou plus importantes) qu'entre des entités éloignées ? Y-a-t-il des zones géographiques (des « grappes » d'entités) à hautes/faibles valeurs (valeurs supérieures ou inférieures à la valeur moyenne) et en contraste par rapport aux régions avoisinantes ? C'est à ce type de questions que nous répondrons dans ce document.

Nous introduisons dans ce rapport le concept d'association spatiale (« l'autocorrélation spatiale ») de la variable d'étude. La mesure peut être globale (pour tout le territoire) ou locale (autour d'une entité) ; nous considérerons dans cette étude comme indicateurs global et local de l'association spatiale respectivement les indices de Moran (chapitre 2) et les indicateurs LISA (chapitre 3).

La théorie est illustrée par quatre exemples tirés de la statistique fiscale (chapitre 4). Dans chacune des applications, on fait ressortir les lieux où, simultanément, la variable d'intérêt est plus grande ou plus petite que la moyenne du territoire et où elle est en contraste par rapport au voisinage. L'approche est ici uniquement descriptive ; l'explication des phénomènes peut être recherchée dans un second temps, en utilisant par exemple des modèles économétriques qui prennent en compte les interactions (souvent positives) entre les régions.

## 1. La répartition de la charge fiscale au sein des cantons.

Les différences locales de charge fiscale entre les communes d'un même canton sont mises en évidence. Sont considérés à cet effet les taux d'imposition qui s'abattent sur les revenus (impôts cantonal, communal et paroissial) des couples mariés avec deux enfants, dont le revenu du travail constitue la seule source d'imposition et dont seulement un des conjoints exerce une activité lucrative. Il est de plus supposé que le contribuable dispose d'un revenu brut de 100'000 francs<sup>1</sup>, qu'il est soumis à la taxation dite « normale » (pas imposé à la source, à forfait ou à un taux spécial tenant compte des revenus déjà imposés à l'étranger) et qu'il applique les déductions forfaitaires permises à son lieu de domicile. Les données proviennent de la statistique 2014 de la charge fiscale.<sup>2</sup>

---

<sup>1</sup> La charge fiscale entre les communes d'un canton varie normalement par un facteur multiplicatif. Les indices calculés ne devraient dès lors pas ou peu être affectés par le montant du revenu brut choisi. De petites variations peuvent apparaître si un impôt personnel (montant fixe) est perçu ou si l'impôt paroissial n'est pas une fraction de l'impôt communal.

<sup>2</sup> Voir <http://www.estv.admin.ch/dokumentation/00075/00076/00720/index.html?lang=fr>.

## **2. Les différences régionales de capacité économique des contribuables.**

Sont montrées les différences locales de revenus en Suisse telles qu'elles ressortent de la dernière étude de l'AFC sur le bien-être en Suisse.<sup>3</sup> Est considéré ici le revenu équivalent net moyen par commune des personnes physiques soumises à une taxation normale (année 2010). Les différences de ségrégation sociale intercommunale sont également montrées au niveau des grandes régions.

## **3. La concentration régionale des gros contribuables**

Les gros contribuables participent pour une grande partie des entrées fiscales du pays. Nous montrons que ceux-ci se répartissent de façon non uniforme sur le territoire du pays (année 2012), ils sont attirés par certains lieux et repoussés par d'autres.

## **4. Les attitudes locales en matière de déductions fiscales.**

Sont mises en évidence les différences locales dans l'application de quelques déductions fiscales par les personnes physiques dans le cadre de l'impôt fédéral direct, à savoir :

- les déductions des frais d'entretien des immeubles par rapport à la valeur officielle de ces biens (canton de Berne, année 2011) ;
- les cotisations versées au pilier 3a (année 2012).

---

<sup>3</sup> Voir. [http://www.estv2.admin.ch/f/dokumentation/zahlen\\_fakten/karten/2010/grafiken\\_2010.php](http://www.estv2.admin.ch/f/dokumentation/zahlen_fakten/karten/2010/grafiken_2010.php).

## 2 L'autocorrélation spatiale globale

En statistique, il existe des indicateurs globaux mesurant l'association spatiale d'une variable donnée  $y$  sur l'entièreté d'un territoire géographique. L'indice de Moran en est le plus connu et c'est celui-ci qui sera calculé dans cette étude.

### 2.1 L'indice de Moran

L'indice de Moran est un indicateur couramment utilisé pour mesurer l'autocorrélation spatiale globale d'une variable sur un territoire. Un nombre suffisant d'entités régionales sur le territoire est cependant souhaité pour pouvoir interpréter sa significativité.

#### 2.2.1 Définition

L'indice de Moran se définit par l'expression mathématique suivante :

$$I = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}}{\sum_{i=1}^n (y_i - \bar{y})^2} n}$$

avec  $\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$  la moyenne de la variable  $y$  sur le territoire,  $w_{ij}$  le poids de proximité associé aux unités spatiales  $(i, j)$  et  $n$  le nombre d'entités différentes sur la surface.

L'indice de Moran correspond à la moyenne pondérée des « similarités »  $sim_{ij} = (y_i - \bar{y})(y_j - \bar{y})$  qui existent entre toutes les paires d'unités spatiales  $(i, j)$ , en donnant à chaque paire d'unités  $(i, j)$  un poids  $w_{ij}$  qui est fonction de leur proximité géographique respective (plus les entités sont « proches », plus le poids est élevé). L'indice est encore échelonné en divisant la moyenne pondérée par la moyenne des déviations à la moyenne au carré

$var(y) = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}$  (la variance de la variable).

L'indice de Moran peut encore s'écrire de façon plus concise comme :

$$I = \sum_{i=1}^n z_i \sum_{j=1}^n \tilde{w}_{ij} z_j,$$

avec :

- $z_i = \frac{y_i - \bar{y}}{\sqrt{\text{var}(y)}}$  la valeur spatialement standardisée de l'observation  $y_i$  par rapport à la

moyenne du territoire, avec  $\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$  la moyenne et  $\text{var}(y) = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}$  la variance de la variable d'intérêt,

- $\tilde{w}_{ij} = \frac{w_{ij}}{\sum_{l=1}^n \sum_{k=1}^n w_{lk}}$  le poids relatif des entités  $(i, j)$  (le poids spatial divisé par la somme des poids).

### La détermination de la matrice des poids

Différentes façons de calculer la matrice  $M = (w_{ij})$  des poids de proximité entre les entités sont proposées dans la littérature. Elle peut par exemple être définie sur la base de l'adjacence ou non des entités spatiales (par exemple,  $w_{ij} = 1$  si les entités  $(i, j)$  ont une frontière commune et  $w_{ij} = 0$  sinon), à partir des distances  $d_{ij}$  entre les entités (par exemple,  $w_{ij} = \frac{1}{d_{ij}}$ ,

$w_{ij} = \frac{1}{d_{ij}^2}$  ou tout autre fonction décroissante en  $d_{ij}$ ) ou d'une combinaison de critères d'adjacence et de distance (par exemple,  $w_{ij} = 1$  si l'entité  $j$  se situe dans un rayon  $d^*$  de l'entité  $i$  et  $w_{ij} = 0$  sinon).

Dans ce document, la distance  $d_{ij}$  entre les entités  $i$  et  $j$  est mesurée par le nombre de kilomètres en ligne droite entre les centres de ces entités.<sup>4</sup> En considérant la distance « à vol d'oiseau » plutôt que par route, on peut négliger quelque peu l'éloignement, surtout dans les régions du pays couvertes de montagnes ou de lacs.

<sup>4</sup> A partir de l'année 2013, l'Office fédéral de la statistique détermine pour chaque commune suisse un centre et met à disposition les coordonnées géo-spatiales de celui-ci. Nous utilisons ces informations pour le calcul des distances entre les communes. Pour les années antérieures, nous calculons les distances en considérant les points de gravité des surfaces polynomiales décrites par les communes.

## La standardisation de la matrice des poids

Afin de donner la même importance aux différentes entités quel que soit le nombre ou la proximité des entités voisines, les poids  $w_{ij}$  de chaque entité  $i$  sont en général standardisés par

une division avec la somme des poids des entités connectés à  $i$  :  $w_{ij} = \frac{w_{ij}}{\sum_{k=1}^n w_{ik}}$  (division de

chaque élément de la matrice  $M = (w_{ij})$  par la somme des poids de la ligne). Par cette stan-

dardisation,  $\sum_{j=1}^n w_{ij} = 1$  (la somme des poids de chaque ligne  $i$  de la matrice  $M$  est égale à

1) et  $\sum_{i=1}^n \sum_{j=1}^n w_{ij} = n$  (la somme des poids de la matrice  $M$  est égale à  $n$ ).

Avec les poids standardisés, l'expression de l'indice de Moran se simplifie :

$$I = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

ou encore :

$$I = \frac{1}{n} \sum_{i=1}^n z_i \sum_{j=1}^n w_{ij} z_j$$

avec  $z_i = \frac{y_i - \bar{y}}{\sqrt{\text{var}(y)}}$  la valeur spatialement standardisée de l'observation  $y_i$ .

L'indice de Moran correspond à la covariance (la moyenne du produit des écarts à la moyenne) entre la variable des valeurs (standardisées)  $z_i$  aux entités  $i$  et la moyenne pondérée

$\sum_{j=1}^n w_{ij} z_j$  des valeurs (standardisées) avoisinantes :  $I = \text{cov}(z_i, \sum_{j=1}^n w_{ij} z_j)$ .

## L'interprétation de l'indice

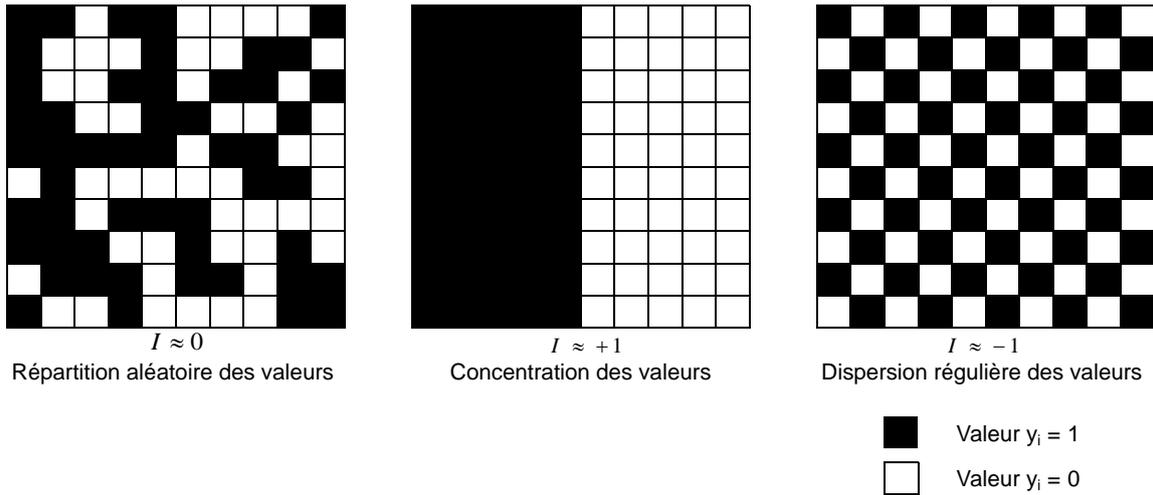
Avec la standardisation de la matrice des poids, l'indice d'autocorrélation spatiale va se situer entre -1 et +1 (comme la mesure de la corrélation de Pearson entre deux variables différentes). Le montant donne une indication de la façon dont les valeurs de la variable d'intérêt  $y$  se répartissent spatialement sur le territoire (voir le graphique 1 dans le cas d'une variable binaire).

Un montant avoisinant 0 reflète **une répartition aléatoire** des valeurs sur le territoire (les produits croisés positifs et négatifs de la formule se compensent).<sup>5</sup>

Plus le montant s'éloigne de 0, plus cela témoigne d'un certain arrangement spatial des valeurs :

<sup>5</sup> On verra au chapitre 2.2.2 qu'en cas d'indépendance spatiale des valeurs, l'espérance de l'indice (la valeur attendue) est égale à  $E[I] = -\frac{1}{n-1}$  et tend vers 0 quand le nombre  $n$  d'unités devient élevé.

- un indice se rapprochant de la borne +1 indique **une concentration** de valeurs similaires sur le territoire (des « grappes » ou « clusters » en anglais) : les valeurs des entités proches ont tendance à être soit plus grandes (« hot spots ») soit plus petites (« cold spots ») que la moyenne du territoire (présence d'unités avec des voisins similaires) ;
- un indice se rapprochant de la borne -1 indique **une dispersion régulière** des valeurs sur le territoire : si la valeur d'une entité est supérieure à la moyenne du territoire, les valeurs des entités proches ont tendance à être inférieures à la moyenne, et vice-versa (présence d'unités avec des voisins non similaires).



**Graphique 1 : les types de répartition spatiale d'une variable binaire.**

## 2.2.2 Exemple : l'autocorrélation spatiale de la charge fiscale dans les cantons

L'indice de Moran de la charge fiscale (impôts cantonal, communal et paroissial) entre les communes de chaque canton est calculé ci-après (voir les graphiques 2a et 2b) à titre d'exemple pour les couples mariés avec deux enfants (année 2014), en utilisant les poids d'adjacence valant 1 si les communes ont une frontière commune et 0 autrement. Ne sont pas pris en considération les communes des cantons enclavés dans un autre canton.

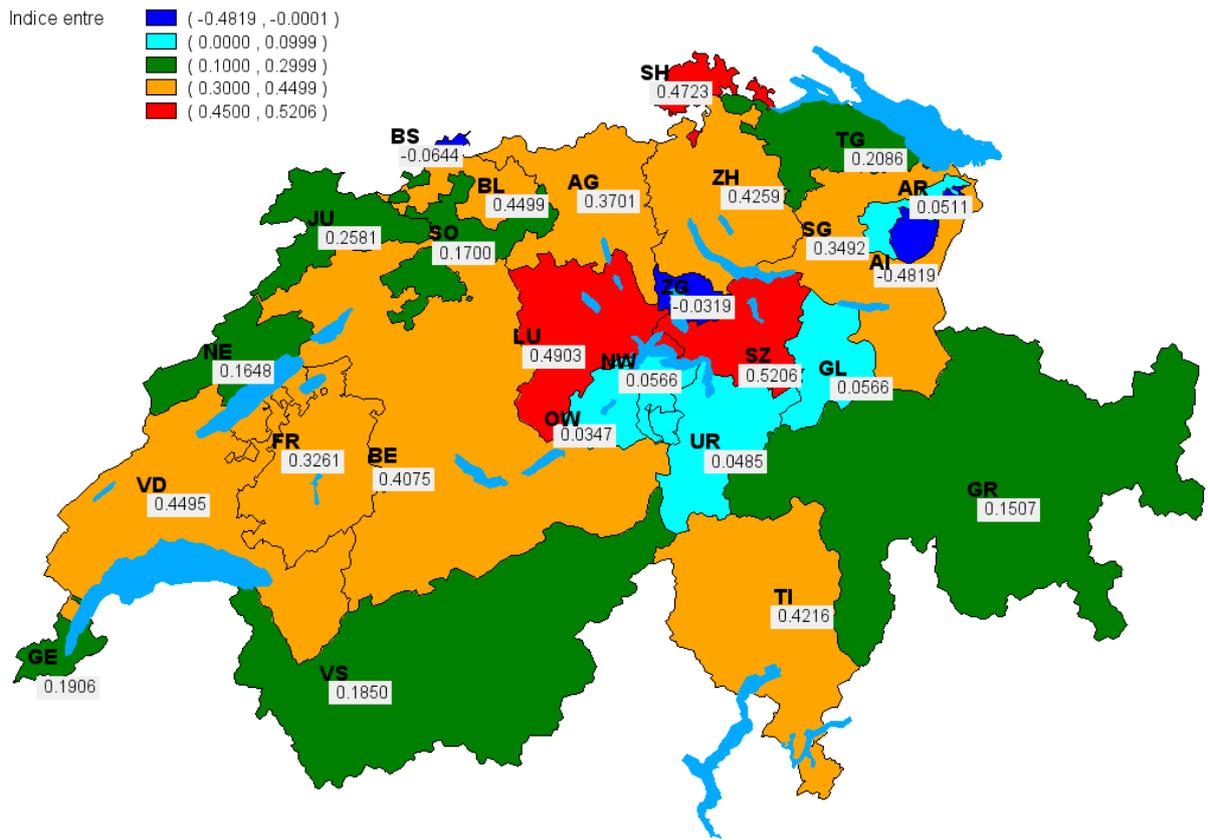
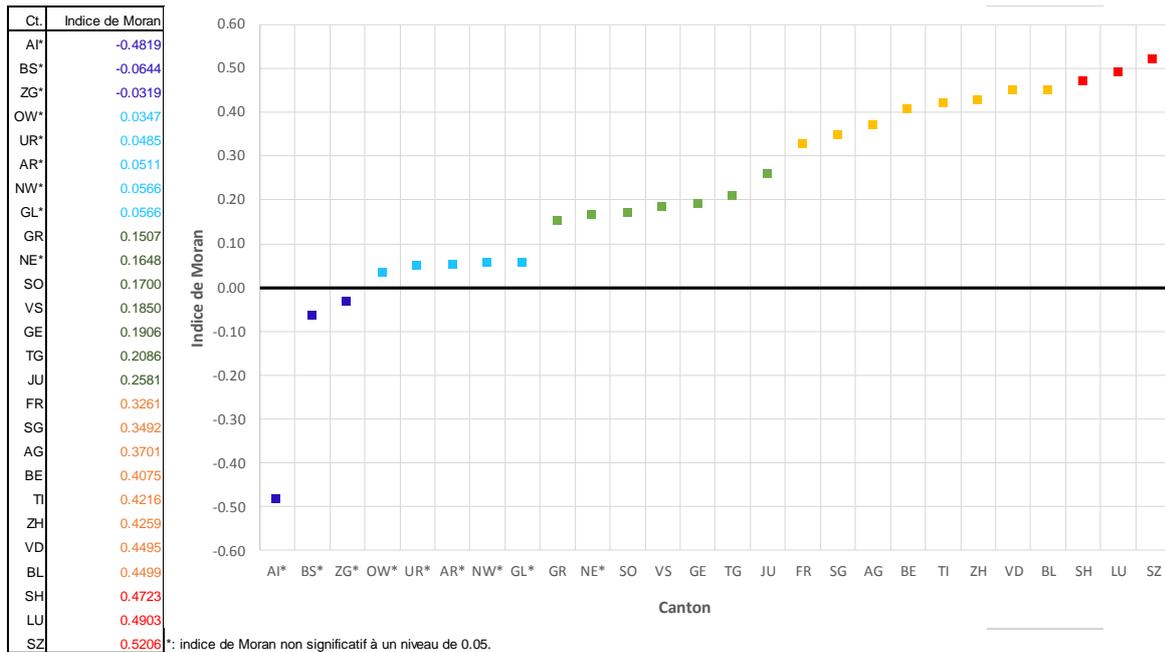
L'indice généralement positif de l'indice nous apprend que la charge fiscale est généralement plus similaire entre les communes voisines qu'entre les communes éloignées. Les communes des cantons de Schwyz, de Lucerne et de Schaffhausen en sont les exemples les plus marquants (indice valant respectivement 0.5206, 0.4903 et 0.4723).

L'autocorrélation spatiale (l'indice différent de zéro) n'est cependant pas significative d'un point de vue statistique (à un niveau de 0.05, estimation asymptotique) dans neuf cantons, à savoir à Appenzell R.-Int., à Bâle-Ville, à Zoug, à Obwald, à Uri, à Appenzell R.-Ext., à Nidwald, à Glaris et à Neuchâtel (indice faible et/ou pas assez de communes dans le canton, voir le chapitre suivant).

L'indice de Moran de la charge fiscale peut être calculé avec d'autres poids  $w_{ij}$  de proximité que les poids d'adjacence (poids 1). Si on dispose de la distance  $d_{ij}$  entre les entités  $(i, j)$ , on peut par exemple définir les poids  $w_{ij}$  :

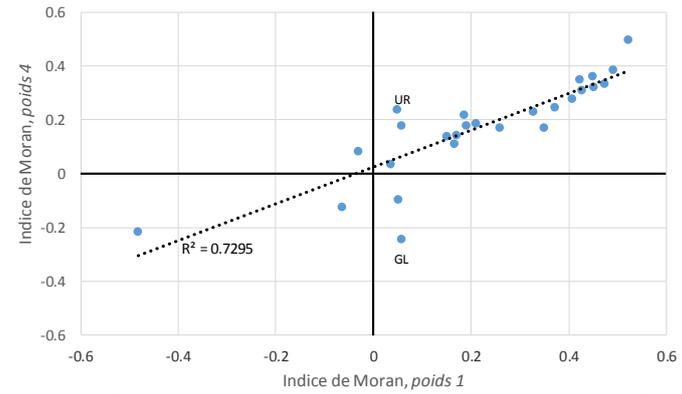
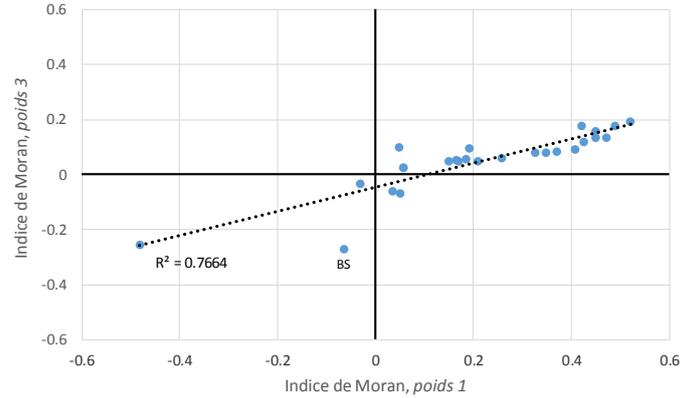
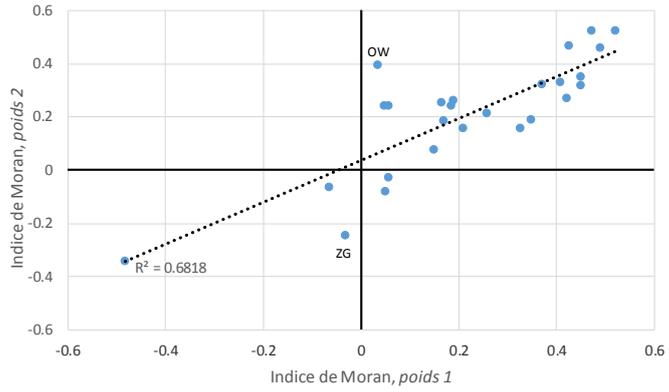
- comme la grandeur binaire valant  $w_{ij} = 1$  si l'entité  $j$  se situe dans un rayon  $d^*$  de l'entité  $i$  et  $w_{ij} = 0$  sinon (pour assurer au moins une valeur  $w_{ij} = 1$  à chaque unité  $i$  ( $i = 1, \dots, n$ ), le rayon  $d^*$  peut par exemple être choisi en prenant la plus grande valeur des  $n$  distances  $d_{ij^*}$  qui séparent chaque unité  $i$  à son unité  $j^*$  la plus proche ; poids 2) ;
- comme une fonction décroissante de la distance entre les entités  $(i, j)$ , telle que l'inverse de la distance ( $w_{ij} = \frac{1}{d_{ij}}$  ; poids 3) ou l'inverse du carré de la distance ( $w_{ij} = \frac{1}{d_{ij}^2}$ , poids 4).

Une comparaison des valeurs de l'indice est faite suivant le type de poids considéré (voir les graphiques 3a, 3b et 3c). On remarque que, suivant les poids, la hauteur de l'indice peut changer mais qu'une forte cohérence existe quant à l'importance relative des valeurs entre les cantons. Des différences apparaissent surtout pour les petits cantons avec peu de communes, tels qu'Obwald, Zoug, Bâle-Ville, Uri et Glaris.



Graphiques 2a et 2b : l'indice de Moran par commune de la charge fiscale des couples mariés avec deux enfants (année 2014).

Ct.	Indice de Moran, calculé avec différents types de poids			
	Poids 1	Poids 2	Poids 3	Poids 4
ZH	0.4259	0.4675	0.1194	0.3118
BE	0.4075	0.3283	0.0916	0.2801
LU	0.4903	0.4590	0.1754	0.3852
UR	0.0485	0.2424	0.0987	0.2373
SZ	0.5206	0.5247	0.1911	0.4963
OW	0.0347	0.3961	-0.0604	0.0354
NW	0.0566	0.2395	0.0269	0.1777
GL	0.0566	-0.0277	0.0269	-0.2441
ZG	-0.0319	-0.2444	-0.0335	0.0832
FR	0.3261	0.1558	0.0780	0.2308
SO	0.1700	0.1841	0.0484	0.1432
BS	-0.0644	-0.0644	-0.2729	-0.1249
BL	0.4499	0.3165	0.1322	0.3237
SH	0.4723	0.5243	0.1340	0.3323
AR	0.0511	-0.0811	-0.0666	-0.0954
AI	-0.4819	-0.3423	-0.2572	-0.2142
SG	0.3492	0.1891	0.0778	0.1710
GR	0.1507	0.0781	0.0478	0.1397
AG	0.3701	0.3238	0.0826	0.2459
TG	0.2086	0.1574	0.0484	0.1888
TI	0.4216	0.2687	0.1778	0.3519
VD	0.4495	0.3497	0.1565	0.3635
VS	0.1850	0.2434	0.0577	0.2192
NE	0.1648	0.2539	0.0509	0.1111
GE	0.1906	0.2614	0.0940	0.1784
JU	0.2581	0.2129	0.0604	0.1724



**Tableau 1 et graphiques 3a, 3b et 3c : une comparaison des indices de Moran de la charge fiscale des couples mariés avec deux enfants suivant le type de poids choisi (année 2014).**

## 2.2 La significativité statistique de l'indice de Moran

D'une valeur  $I_{obs}$  observée pour l'indice non nulle, peut-on conclure à une association spatiale  $I$  statistiquement significative? Si on considère les différentes valeurs  $y_i$  de la variable d'intérêt comme issues d'un tirage aléatoire, une faible variation par rapport à zéro peut en effet être attribuée au simple fait du hasard par un nombre trop faible d'observations. Nous allons montrer dans ce chapitre comment reconnaître la présence d'une association spatiale significative.

### 2.2.1 Le principe du test statistique

Pour savoir si l'association spatiale est significative, un test statistique va être effectué. En fonction du type de conclusion que l'on cherche à obtenir, le test sera soit unilatéral (l'association spatiale est-elle significativement strictement positive, respectivement négative ?) soit bilatéral (l'association spatiale est-elle significativement non nulle ?).

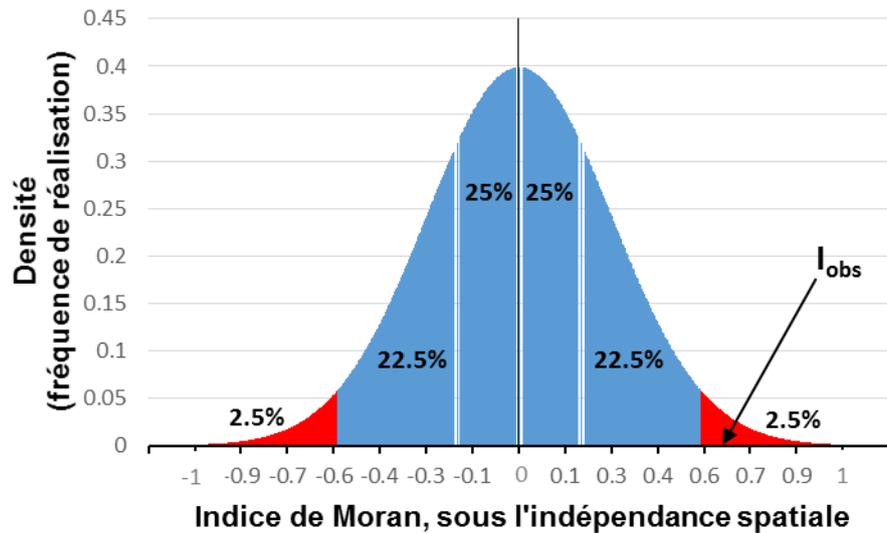
L'hypothèse nulle du test est l'indépendance spatiale ( $H_0 : I = 0$ ), c'est-à-dire une répartition spatiale aléatoire des valeurs  $y_i$ . Dans le test unilatéral (le « one-sided test » en anglais), l'hypothèse alternative du test est la considération d'une association spatiale strictement positive ( $H_1 : I > 0$ ), respectivement négative ( $H_1 : I < 0$ ), tandis que dans le test bilatéral (le « two-sided test » en anglais), l'hypothèse alternative est l'acceptation d'une corrélation spatiale non nulle ( $H_1 : I \neq 0$ ), sans imposer a priori le caractère positif ou négatif de l'association (ce sont les résultats observés qui permettront de conclure à l'une ou à l'autre option).

Pour effectuer le test, il est généralement calculé la *valeur-p* (« *p-value* » en anglais) associée à l'observation  $I_{obs}$  de l'indice, c'est-à-dire la probabilité que, sous l'hypothèse nulle d'indépendance spatiale, l'indice  $I$  soit égal à  $I_{obs}$  ou à un montant plus « extrême » (un montant plus éloigné de 0 que  $I_{obs}$ ).

Le principe du test statistique est montré au graphique 4.

Si la *valeur-p* est inférieure au seuil  $\alpha$  de significativité (une valeur préalablement définie, souvent fixée à  $\alpha = 0.05$ ), on peut rejeter l'hypothèse nulle en faveur de l'hypothèse alternative avec une probabilité d'erreur  $\alpha$  (l'erreur de « première espèce »). La dépendance spatiale est alors dite « statistiquement significative » : la similarité ou la dissimilitude des entités proches comparativement aux entités éloignées est alors assez nette pour qu'il soit improbable que l'indice de Moran soit non nul (test bilatéral), respectivement strictement positif ou négatif (test unilatéral), par le seul effet du hasard.

Par contre, si la *valeur-p* est supérieure au seuil  $\alpha$ , l'observation  $I_{obs}$  de l'indice n'est pas contradictoire avec l'hypothèse nulle (l'indépendance spatiale) et on ne rejette pas celle-ci. Remarquons la prudence du langage : on n'affirme pas que l'hypothèse nulle est vérifiée. La probabilité d'erreur  $\beta$  d'accepter à tort l'hypothèse nulle (l'erreur de « deuxième espèce ») peut être non négligeable.



**Graphique 4 : le principe du test statistique uni- et bilatéral (le rejet de l'indépendance spatiale, au seuil de significativité de respectivement 2.5% et 5%).**

On calculera dans ce document les *valeurs-p* associées au test bilatéral, n'imposant à priori aucune contrainte de signe à l'indice (positif ou négatif) et plus conservateur que le test unilatéral (une autocorrélation spatiale non significative dans un test bilatéral peut s'avérer significative dans un test unilatéral).

## 2.2.2 L'estimation de la valeur-p

Nous présentons ici deux manières différentes d'estimer la *valeur-p*.

### L'estimation asymptotique (normal test ou t-test)

Soit la statistique  $z = \frac{I - E[I]}{\sqrt{\text{var}[I]}}$  où  $E[I]$  et  $\text{var}[I]$  représentent l'espérance (la valeur attendue)

et la variance de l'indice. Sous l'hypothèse d'indépendance spatiale (les valeurs  $y_i$  se répartissent aléatoirement entre les  $n$  localisations), la statistique  $z$  tend asymptotiquement (quand le nombre  $n$  d'observations tend vers l'infini) vers une distribution normale réduite ( $z \rightarrow N(0,1)$  quand  $n \rightarrow +\infty$ ) et l'espérance et la variance de l'indice peuvent être estimées par les expressions suivantes (voir Cliff et Ord (1981)) :

- $\mu_0 = -\frac{1}{n-1}$  (montant indépendant des poids) ;
- $\sigma_0^2 = \frac{ns_1 - s_2s_3}{(n-1)(n-2)(n-3)\left(\sum_{i=1}^n \sum_{j=1}^n w_{ij}\right)^2}$  avec :
  - $s_1 = (n^2 - 3n + 3)\left(0.5 \sum_{i=1}^n \sum_{j=1}^n (w_{ij} + w_{ji})^2\right) - n\left(\sum_{i=1}^n (\sum_{j=1}^n w_{ij} + \sum_{j=1}^n w_{ji})^2\right) + 3\left(\sum_{i=1}^n \sum_{j=1}^n w_{ij}\right)^2$ ,
  - $s_2 = \frac{n^{-1} \sum_{i=1}^n (y_i - \bar{y})^4}{\left(n^{-1} \sum_{i=1}^n (y_i - \bar{y})^2\right)^2}$ ,
  - $s_3 = 0.5 \sum_{i=1}^n \sum_{j=1}^n (w_{ij} + w_{ji})^2 - 2n\left(0.5 \sum_{i=1}^n \sum_{j=1}^n (w_{ij} + w_{ji})^2\right) + 6\left(\sum_{i=1}^n \sum_{j=1}^n w_{ij}\right)^2$ .

Dès lors, si le nombre  $n$  d'observations est suffisamment élevé, on peut estimer la *valeur-p* en considérant celle associée à la valeur  $z_{score} = \frac{I_{obs} - \mu_0}{\sigma_0}$  d'une loi normale réduite (la surface formée par la queue de la cloche gaussienne, multipliée par deux dans le cas du test bilatéral).

A titre d'exemple (voir le tableau 2), nous estimons par cette approche la *valeur-p* de l'indice de Moran de la charge fiscale des couples mariés avec deux enfants pour les différents cantons, en utilisant les poids d'adjacence (voir les valeurs de l'indice du chapitre 2.2.2, avec *poids* 1). Pour les cantons avec peu de communes, l'estimation doit être prise avec toute la réserve qui se doit.

Canton	Nombre de communes <i>n</i>	Indice de Moran				z-score $z_{score} = \frac{I_{obs} - \mu_0}{\sigma_0}$	Valeur-p (loi normale) <i>valeur-p</i>
		Valeur observée <i>I<sub>obs</sub></i>	Valeur attendue, si indépend. spatiale				
			Espérance $\mu_0 = -1/(n-1)$	Variance $\sigma_0^2$			
ZH	170	0.42595	-0.00592	0.00247	8.68763	0.00000	
BE	360	0.40749	-0.00279	0.00113	12.18262	0.00000	
LU	83	0.49027	-0.01220	0.00543	6.82029	0.00000	
UR	20	0.04847	-0.05263	0.02086	0.70003	0.48391	
SZ	30	0.52056	-0.03448	0.01500	4.53153	0.00001	
OW	6	0.03469	-0.20000	0.03194	1.31310	0.18915	
NW	10	0.05661	-0.11111	0.03755	0.86560	0.38671	
GL	3	0.05661	-0.11111	0.03755	0.86560	0.38671	
ZG	11	-0.03186	-0.10000	0.04770	0.31197	0.75506	
FR	162	0.32610	-0.00621	0.00298	6.08571	0.00000	
SO	108	0.17002	-0.00935	0.00453	2.66428	0.00772	
BS	3	-0.06441	-0.50000	0.08750	1.47256	0.14087	
BL	86	0.44991	-0.01176	0.00541	6.27474	0.00000	
SH	26	0.47234	-0.04000	0.02685	3.12696	0.00177	
AR	20	0.05115	-0.05263	0.02549	0.65005	0.51566	
AI	5	-0.48193	-0.25000	0.03677	-1.20950	0.22647	
SG	77	0.34922	-0.01316	0.00635	4.54637	0.00001	
GR	146	0.15073	-0.00690	0.00280	2.97625	0.00292	
AG	213	0.37009	-0.00472	0.00194	8.52006	0.00000	
TG	79	0.20862	-0.01282	0.00585	2.89600	0.00378	
TI	135	0.42162	-0.00746	0.00309	7.71354	0.00000	
VD	318	0.44951	-0.00315	0.00135	12.34246	0.00000	
VS	134	0.18502	-0.00752	0.00288	3.58773	0.00033	
NE	37	0.16479	-0.02778	0.01223	1.74119	0.08165	
GE	44	0.19064	-0.02326	0.00925	2.22399	0.02615	
JU	57	0.25807	-0.01786	0.00816	3.05423	0.00226	

**Tableau 2 : la significativité par canton de l'indice de Moran de la charge fiscale des couples mariés avec deux enfants (année 2014), l'estimation asymptotique.**

## L'approche itérative (test de permutations)

Dans cette approche, on génère itérativement une série d'indices  $I_k$  sous l'hypothèse d'indépendance spatiale. A chaque itération, les valeurs  $y_i$  observées sont réallouées aléatoirement aux différentes entités géographiques et un indice de Moran  $I_k$  est recalculé.

Après avoir fait un nombre  $r$  suffisamment élevé de réarrangements de valeurs (par exemple,  $r = 999$  permutations), on peut estimer la *valeur-p* à partir de la distribution empirique des  $I_k$  ( $k = 1, \dots, r$ ). Pour le test d'autocorrélation positif (ou négatif), il suffit de compter, dans l'échantillon d'indices ainsi construit, la proportion d'indices supérieurs ou égaux (respectivement inférieurs ou égaux) à la valeur observée  $I_{obs}$  :

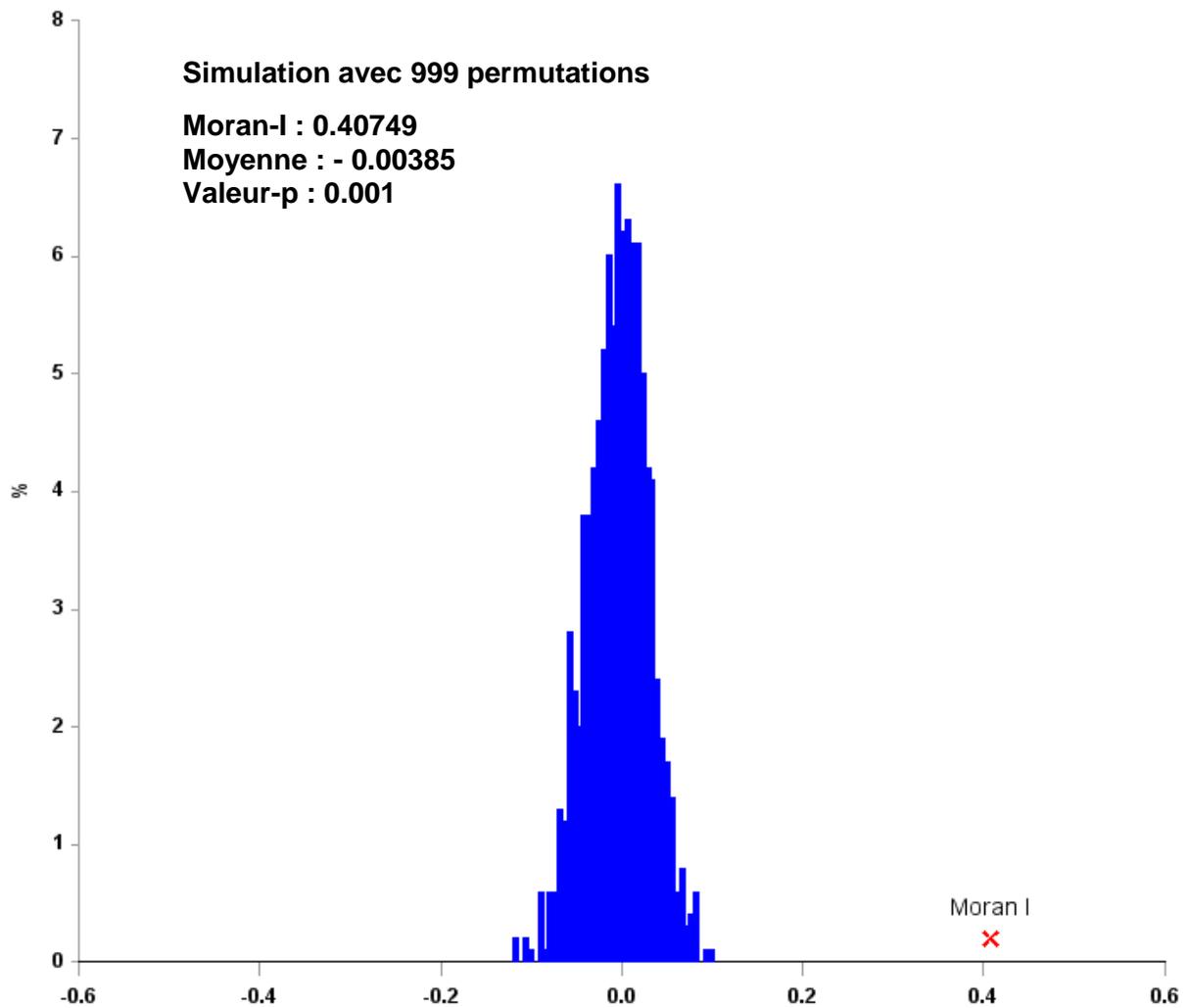
$$valeur - p = \frac{m + 1}{r + 1},$$

où  $m$  est le nombre d'indices générés  $I_k$  supérieurs ou égaux à  $I_{obs}$  (respectivement inférieurs et égaux à  $I_{obs}$ ). L'indice observé  $I_{obs}$  est généralement rajouté à l'échantillon, d'où l'addition au numérateur et au dénominateur d'une unité supplémentaire. Pour le test bilatéral, le même calcul peut être mené si ce n'est qu'il faut préalablement centrer les valeurs (en soustrayant, à l'indice observé  $I_{obs}$  et aux indices  $I_k$  générés, la moyenne  $\bar{I}$  des indices  $I_k$  de tous les réarrangements) et considérer pour  $m$  le nombre de valeurs  $I_k - \bar{I}$  au moins aussi distant de la valeur nulle que  $I_{obs} - \bar{I}$ . Notons que si l'indice observé est nettement significatif, la valeur  $m$  sera nulle ou très faible, de sorte que la *valeur - p*  $= \frac{m + 1}{r + 1} \approx \frac{1}{r + 1}$  dépendra essentiellement du nombre  $r$  de permutations effectuées ; on parle aussi de « *pseudo valeur-p* » pour la *valeur-p* estimée par simulation.

Cette approche, plus coûteuse en temps calcul, a l'avantage de ne pas devoir supposer une distribution gaussienne (la loi normale) à la mesure de l'indice.

A titre d'exemple (voir le graphique 5), nous appliquons cette approche à la vérification de la présence d'une autocorrélation spatiale significative entre les communes du canton de Berne relative à la charge fiscale des couples mariés avec deux enfants (avec comme précédemment les poids d'adjacence *poids* 1). Nous générons 999 permutations dont l'indice moyen se situe à -0.00385. Aucune permutation n'a donné un indice aussi éloigné de la valeur moyenne que

l'indice observé  $I_{obs} = 0.40749$  et la *valeur - p*  $= \frac{m + 1}{r + 1} = \frac{0 + 1}{999 + 1} = 0.001$ .



**Graphique 5 : la significativité de l'indice de Moran de la charge fiscale des couples mariés avec deux enfants par l'approche itérative avec 999 permutations, canton de Berne et année 2014.**

### 3 Les associations spatiales locales

Les indicateurs globaux d'association spatiale donnent une indication « moyenne » de l'auto-corrélation spatiale de la variable d'intérêt sur l'aire géographique étudiée. S'il n'apparaît pas d'autocorrélation spatiale sur l'entière du territoire, des « grappes » de valeurs peuvent toutefois survenir à un niveau local. Les indicateurs locaux d'association spatiale permettent de les faire ressortir.

#### 3.1 L'indice LISA

##### 3.1.1 Définition

En définissant comme précédemment :

- la valeur spatialement standardisée  $z_i = \frac{y_i - \bar{y}}{\sqrt{\text{var}(y)}}$  de l'observation  $y_i$  (avec  $\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$

la moyenne et  $\text{var}(y) = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}$  la variance de la variable  $y$  sur le territoire)

- le poids  $w_{ij}$  standardisé de proximité associé aux unités spatiales  $(i, j)$ ,

on peut définir une version locale de l'indice de Moran en considérant la composante individuelle  $I_i$  de chaque entité  $i$  ( $i = 1, \dots, n$ ) :

$$I_i = z_i \sum_{j=1}^n w_{ij} z_j$$

Cet indice est appelé l'indice local d'association spatiale d'Anselin ou tout simplement l'indice LISA<sup>6</sup>. Un indice  $I_i$  positif révèle une entité  $i$  proche d'entités de caractéristique similaire (une « grappe » de hautes ou de faibles valeurs). Un indice  $I_i$  négatif marque par contre une entité  $i$  entourée d'entités de profil différent (un « outlier » local).

---

<sup>6</sup> Le terme LISA est l'abréviation de l'anglais « Local Indicator of Spatial Association ».

Cet indice  $I_i$  peut aussi être vu comme une mesure de l'association spatiale (la corrélation statistique) entre la valeur de la variable  $y$  à une localisation donnée  $i$  et les valeurs de la même variable aux locations avoisinantes  $j$  (l'autocorrélation spatiale de la variable).

Remarquons que l'indice (global) de Moran est égal à la moyenne pondérée des indices locaux  $I_i$  ( $i = 1, \dots, n$ ) :

$$I = \frac{1}{W} \sum_{i=1}^n W_i I_i,$$

où :

- $W_i = \sum_{j=1}^n w_{ij}$  est la somme des poids des entités connectés à  $i$  (somme des éléments de la ligne  $i$  de la matrice  $M$ ) ;
- $W = \sum_{i=1}^n W_i = \sum_{i=1}^n \sum_{j=1}^n w_{ij}$  est la somme de tous les poids (somme de tous les éléments de la matrice  $M$ ).

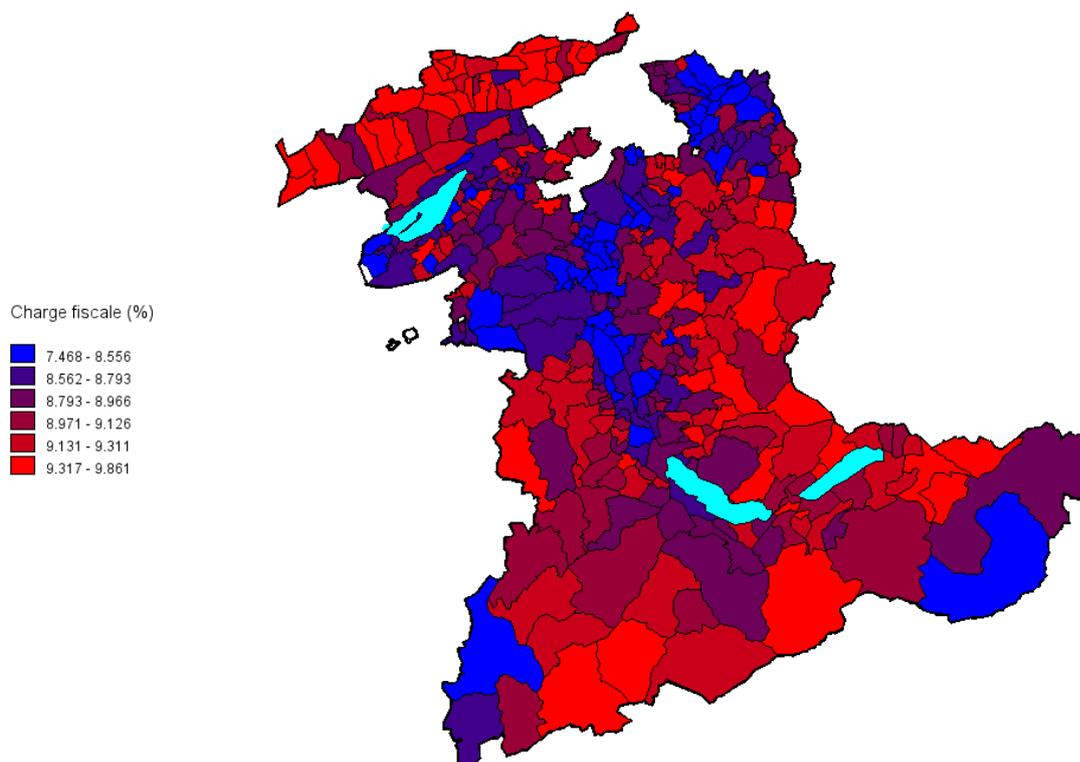
Si les poids sont standardisés,  $W_i = 1$  et  $W = n$  (chaque élément de la matrice  $M = (w_{ij})$  est divisé par la somme des éléments de la ligne). L'indice de Moran s'exprime alors simplement comme la moyenne arithmétique des indices locaux  $I_i$  ( $i = 1, \dots, n$ ) :

$$I = \frac{1}{n} \sum_{i=1}^n I_i.$$

### 3.1.2 Exemple : les indices LISA de la charge fiscale à Berne

Reprenons l'exemple de la répartition spatiale de la charge fiscale (impôts cantonal, communal et paroissial) des couples mariés avec deux enfants et concentrons-nous sur la situation du canton de Berne (excepté les communes de Clavaleyres et de Münchenwiler, enclavées dans le canton de Fribourg).

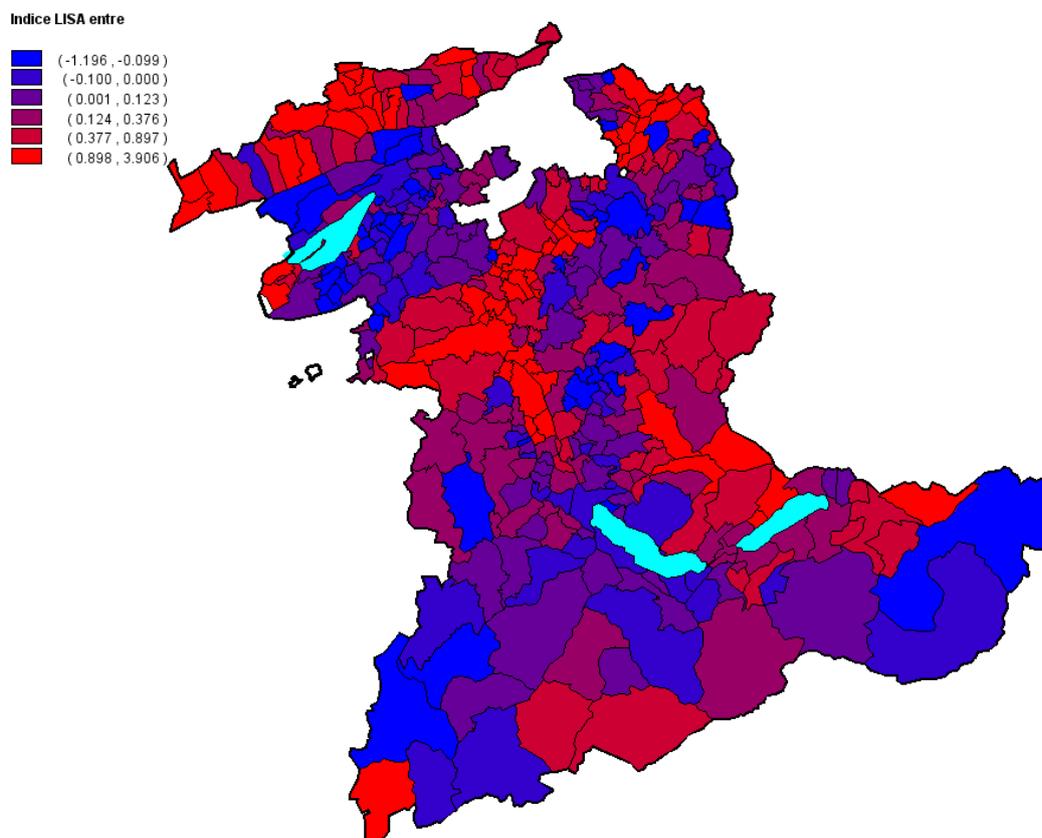
Si nous considérons les ménages avec un revenu brut de 100'000 francs, la charge fiscale varie dans le canton entre 7.468 et 9.861%, la valeur moyenne s'élevant à 8.947% (moyenne des charges fiscales des 360 communes bernoises). Au graphique 5, les communes du canton sont réparties entre six groupes de même taille (même nombre de communes dans chaque groupe), en fonction de la charge fiscale qui y règne (délimitation de l'intervalle des valeurs par les 20<sup>e</sup>, 40<sup>e</sup>, 60<sup>e</sup> et 80<sup>e</sup> centiles). Chaque groupe est représenté sur le dessin par une couleur, allant du bleu azur (faible charge entre 7.468 et 8.556%) au rouge clair (forte charge entre 9.317 et 9.861%). Le graphique laisse percevoir la présence de grappes dans la répartition de la charge fiscale (cf. les communes avoisinantes de même couleur). L'indice (global) de Moran est positif sur l'entièreté du territoire ( $I = 0.407$ ).



**Graphique 5 : la charge fiscale des couples mariés avec deux enfants et avec un revenu brut de 100'000 francs, canton de Berne et année 2014.**

Le graphique 6 représente les associations locales  $I_i$  des communes, en mesurant la proximité de chaque paire  $(i, j)$  de communes par la variable binaire  $w_{ij}$  valant 1 si les communes  $i$  et  $j$  ont une frontière commune et 0 autrement (poids d'adjacence).

Au total 76% des communes du canton (communes partagées en quatre classes de taille identique sur le graphique, voir les communes colorées dans une teinte rouge) ont une autocorrélation spatialement positive, c'est-à-dire une concentration au niveau local de charges fiscales supérieures ou inférieures à la moyenne. Autrement 24% des communes (communes partagées en deux classes de taille identique sur le graphique, voir les communes colorées en bleu foncé et en bleu azur) présentent localement une autocorrélation négative, c'est-à-dire une alternance au niveau local de charges fiscales supérieures ou inférieures à la moyenne.



**Graphique 6 : les indices LISA de la charge fiscale des couples mariés avec deux enfants, canton de Berne et année 2014.**

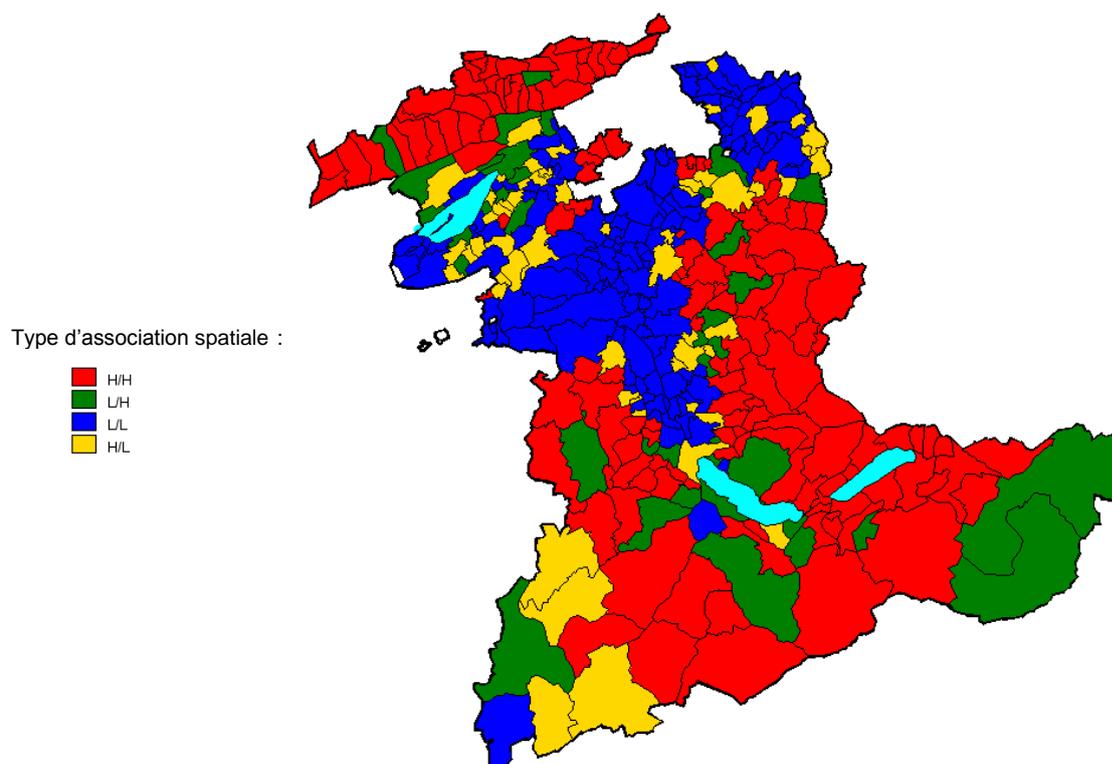
## 3.2 Les types d'association spatiale

L'indice LISA  $I_i$  de chaque entité est le produit des deux facteurs  $z_i$  et  $\sum_{j=1}^n w_{ij}z_j$ . Suivant le signe que prend chacun de ces facteurs, quatre types d'association spatiale peuvent être distingués :

- **une association spatiale High/High** où les valeurs  $z_i$  et  $\sum_{j=1}^n w_{ij}z_j$  sont strictement positives : l'entité fait partie d'une « grappe » (« clusters » en anglais) de hautes charges fiscales, nommée parfois « **hot spot** » ;
- **une association spatiale Low/Low** où les valeurs  $z_i$  et  $\sum_{j=1}^n w_{ij}z_j$  sont strictement négatives : l'entité fait partie d'une « grappe » (« clusters » en anglais) de faibles charges fiscales, nommée parfois « **cold spot** » ;
- **une association spatiale High/Low** où la valeur  $z_i$  est strictement positive et la valeur  $\sum_{j=1}^n w_{ij}z_j$  strictement négative : l'entité a une haute charge fiscale dans une aire de faibles charges, elle apparaît comme un « **outlier** » spatial de haute charge ;
- **une association spatiale Low/High** où la valeur  $z_i$  est strictement négative et la valeur  $\sum_{j=1}^n w_{ij}z_j$  strictement positive : l'entité est de faible charge fiscale dans une aire de hautes charges, elle apparaît comme un « **outlier** » spatial de faible charge.

Les associations High/High et Low/Low indiquent une autocorrélation spatiale positive entre la valeur du lieu et les valeurs avoisinantes; les associations High/Low et Low/High une autocorrélation spatiale négative.

A titre illustratif, les quatre types d'association spatiale relatifs à la charge fiscale à Berne des couples mariés avec deux enfants (l'exemple du chapitre 2.2.2) sont reportés sur une carte thématique (voir le graphique 7).



**Graphique 7 : les types d'association spatiale de la charge fiscale des couples mariés avec deux enfants, canton de Berne et année 2014.**

### 3.3 Le diagramme de dispersion de Moran

La dispersion spatiale des valeurs peut encore être visualisée sur un diagramme de dispersion, dite de Moran (le « Moran scatter plot » en anglais). Ce graphique met en évidence le type et la force de l'autocorrélation spatiale locale des données.

Le diagramme de dispersion de Moran se présente comme un nuage de points dans un plan bidimensionnel à deux axes, où en abscisse sont reprises les valeurs standardisées  $z_i$  de

chaque entité  $i$  et en ordonnée la moyenne pondérée  $\sum_{j=1}^n w_{ij}z_j$  des valeurs avoisinantes.

Les axes sont placés au centre (0,0) du nuage de points (les moyennes des  $z_i$  et des

$\sum_{j=1}^n w_{ij}z_j$  sont nulles). Ils définissent ainsi quatre cadrans, chaque cadran correspondant à un des types d'association spatiale évoquées précédemment, à savoir (de haut à droite et dans le sens des aiguilles d'une montre) l'association spatiale High/High, High/Low, Low/Low ou Low/High.

A titre illustratif, nous montrons au graphique 8 le diagramme de dispersion de Moran relatif à la répartition de la charge fiscale des couples mariés avec deux enfants dans le canton de Berne (voir le chapitre 3.1.2). Les cadrans définissent les mêmes « clusters » et « outliers » que sur la carte thématique du graphique 7 (les quatre types d'association sont représentés avec les mêmes couleurs que sur la carte thématique).

Nous limitons le diagramme aux valeurs  $z_i$  et  $\sum_{j=1}^n w_{ij}z_j$  comprises entre  $-3$  et  $+3$ . Les va-

leurs tombant en dehors de l'intervalle  $[-3,+3]$  réfèrent à des montants extrêmes (concerne moins de 1% des observations, en cas de répartition normale des valeurs). Dans notre

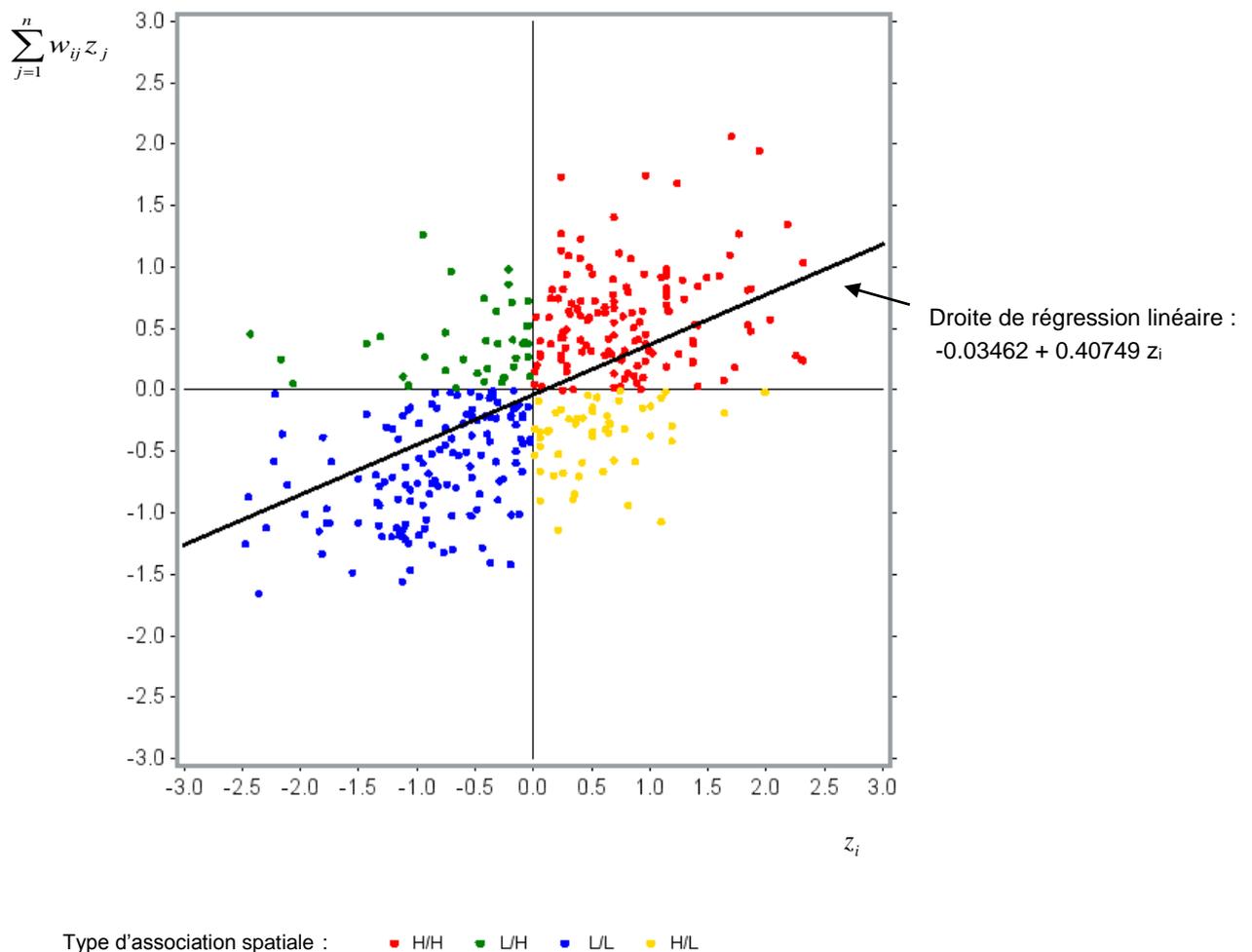
exemple, les communes Rumendingen ( $z_i = -3.20760$  et  $\sum_{j=1}^n w_{ij}z_j = 0.26905$ ) et Deisswil bei

Müchenbuchsee ( $z_i = -3.75696$  et  $\sum_{j=1}^n w_{ij}z_j = -1.03955$ ) ne sont pas représentées.

Notons qu'avec les poids standardisés, la droite  $y = a + bx$  de la régression linéaire passant à travers le nuage de points  $(z_i, \sum_{j=1}^n w_{ij}z_j)$  a une pente égale à l'indice  $I$  de Moran :

$$b = \frac{\text{cov}(z_i, \sum_{j=1}^n w_{ij}z_j)}{\text{var}(z_i)} = \frac{1}{n} \sum_{i=1}^n z_i (\sum_{j=1}^n w_{ij}z_j).$$
 C'est ainsi que, dans notre exemple, la droite de la régression linéaire est inclinée à  $I = 0.40749$ .

Soulignons enfin que les diagrammes de dispersion de Moran sont dépendant des poids utilisés et que seuls les diagrammes exploitant les mêmes bases de poids peuvent être comparés entre elles.



**Graphique 8 : le diagramme de dispersion de Moran relatif à la charge fiscale des couples mariés avec deux enfants et la droite de régression linéaire, canton de Berne et année 2014.**

## 3.4 La significativité statistique de l'association spatiale

Comme pour l'indice de Moran, on veut déterminer les indices LISA qui définissent une relation spatiale au niveau local statistiquement significative (à un seuil  $\alpha$ , par exemple 0.05).

### 3.4.1 Le test statistique

A l'instar de ce qui a été fait au chapitre 2.2, on peut tester la significativité de l'association locale  $I_i$  :

1. en calculant le  $z_{i,score}$  de la statistique  $z_i = \frac{I_i - E[I_i]}{\sqrt{\text{var}[I_i]}}$  à la valeur observée  $I_{i,obs}$  de l'indice, où  $E[I_i]$  et  $\text{var}[I_i]$  représentent respectivement l'espérance et la variance de l'indice  $I_i$  sous l'hypothèse de l'indépendance spatiale locale ;
2. en estimant la *valeur-p* associée à la valeur  $z_{i,score} = \frac{I_{i,obs} - E[I_i]}{\sqrt{\text{var}[I_i]}}$ .

L'indépendance spatiale est rejetée localement avec une probabilité d'erreur  $\alpha$  si la *valeur-p* est inférieure à  $\alpha$ . Si la *valeur-p* est supérieure à  $\alpha$ , l'indice local ne peut pas être considéré comme significativement non nul et on ne peut rejeter l'hypothèse de l'indépendance spatiale au niveau local (au seuil de signification  $\alpha$ ). Nous calculerons dans ce document, comme pour l'indice de Moran, les *valeurs-p* associées au test bilatéral (voir le chapitre 2.2.1).

La *valeur-p* peut être calculée en supposant une distribution normale réduite à la statistique  $z_i$ . Sous l'hypothèse de l'indépendance spatiale locale, l'espérance et la variance de l'indice  $I_i$  peuvent être estimées de la manière suivante (voir Anselin (1995)) :

$$\bullet \quad E[I_i] = -\frac{\sum_{j=1}^n w_{ij}}{n-1} ;$$

$$\bullet \quad \text{var}[I_i] = \frac{(n-r) \sum_{j=1}^n w_{ij}^2}{n-1} + \frac{(2r-n) \left( \left( \sum_{j=1}^n w_{ij} \right)^2 - \sum_{j=1}^n w_{ij}^2 \right)}{(n-1)(n-2)} - \left( \frac{\sum_{j=1}^n w_{ij}}{n-1} \right)^2$$

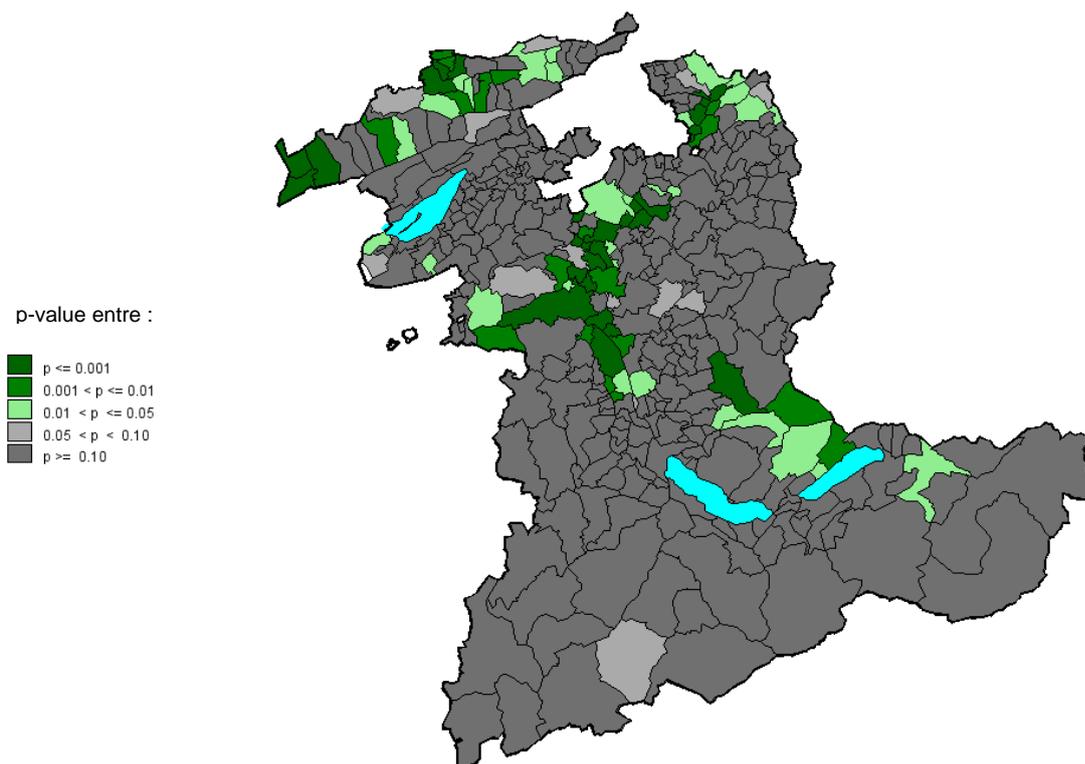
$$\text{avec } r = \frac{n \sum_{i=1}^n (y_i - \bar{y})^4}{\left( \sum_{i=1}^n (y_i - \bar{y})^2 \right)^2}.$$

Comme pour l'indice de Moran (voir le chapitre 2.2.1), la *valeur-p* peut aussi être estimée par permutations de façon itérative, sans devoir accepter l'hypothèse de la distribution normale. Deux façons de procéder sont proposées dans la littérature : une permutation aléatoire des valeurs de la variable d'étude (les valeurs  $y_i$  observées sont réallouées aléatoirement aux différentes entités géographiques) ou une permutation aléatoire des lignes de la matrice des poids (les distances entre les entités  $(i, j)$  sont interchangées aléatoirement).

### 3.4.2 Exemple : les associations locales significatives de la charge fiscale à Berne

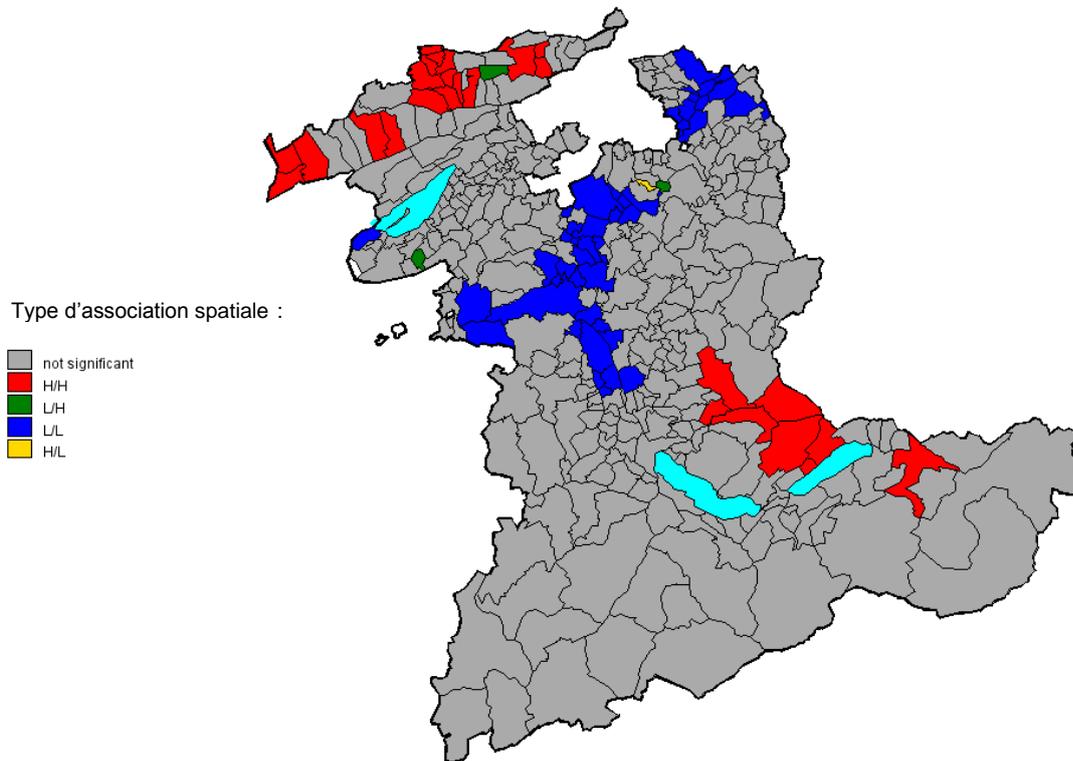
Reprenons l'exemple du chapitre 3.1.2 de la répartition de la charge fiscale des couples mariés avec deux enfants entre les communes du canton de Berne. Toutes les associations spatiales détectées ne sont pas significativement différentes de celles d'une répartition spatiale aléatoire. Au graphique 9, les degrés de significativité sont montrés. Ils sont calculés en utilisant l'approche asymptotique supposant une distribution normale. Moins élevé est la *valeur-p* de l'indice LISA d'une commune, plus significatif est l'indicateur de l'association. Les seuils suivants sont considérés :

- $p \leq 0.001$  : quasi-certitude de pouvoir rejeter l'indépendance spatiale ;
- $0.001 < p \leq 0.01$  : très forte présomption contre l'indépendance spatiale ;
- $0.01 < p \leq 0.05$  : forte présomption contre l'indépendance spatiale ;
- $0.05 < p \leq 0.1$  : faible présomption contre l'indépendance spatiale ;
- $p \geq 0.1$  : pas de présomption contre l'indépendance spatiale.



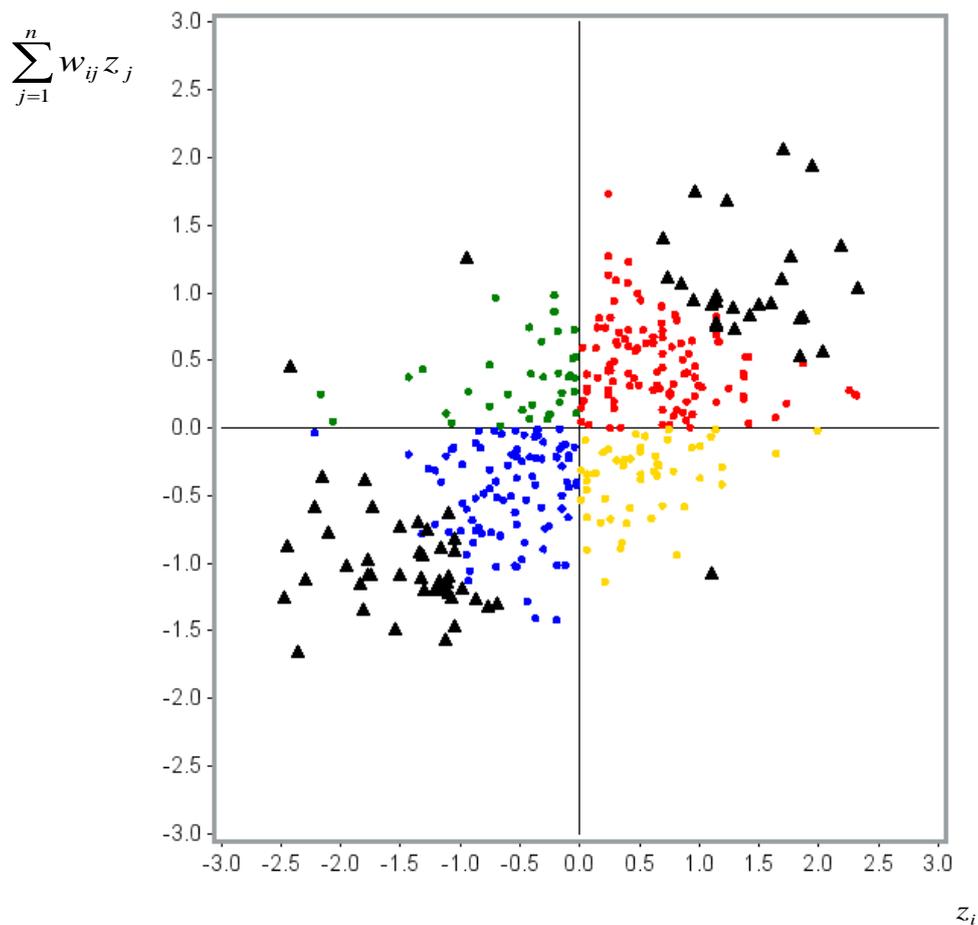
Graphique 9 : les valeurs-p des indices LISA  $I_i$  de la charge fiscale des couples mariés avec deux enfants, canton de Berne et année 2014.

Au graphique 10, seules les associations spatiales significatives au seuil de 0.05 sont reprises ( $p \leq 0.05$ ). Ce graphique, appelé carte LISA (« LISA map » en anglais), fait ressortir les grappes et les contrastes locaux qui sont réellement significatifs pour le canton (au seuil de 0.05).



**Graphique 10 : la carte LISA des associations spatiales significatives (à un seuil de 0.05) de la charge fiscale des couples mariés avec deux enfants, canton de Berne et année 2014.**

Les associations spatiales significatives sont encore pointées sur le diagramme de dispersion de Moran (voir le graphique 11). Au total, 73 des 360 communes bernoises apparaissent impliquées dans des relations spatiales significatives en estimant les *valeurs-p* sous l'hypothèse d'une loi normale. Des résultats légèrement plus conservateurs ressortent par l'approche itérative (permutations de la variable d'analyse) : environ 64 des 360 communes bernoises semblent impliquées dans des relations spatiales significatives par cette approche, le nombre variant légèrement suivant le choix (aléatoire) des permutations effectuées (en répétant la procédure trois fois, on obtient deux fois 64 communes et une fois 65 communes).



Associations spatiales non significatives / significatives (seuil 0.05)    ■ H/H    ■ L/H    ■ L/L    ■ H/L    ▲ signif

**Graphique 11 : les associations spatiales significatives (à un seuil de 0.05) de la charge fiscale des couples mariés avec deux enfants reportées sur le diagramme de dispersion de Moran, canton de Berne et année 2014.**

Avec l'estimation des *valeurs-p* sous l'hypothèse d'une loi normale, nous remarquons la présence :

- de six « hot spots » (en rouge sur la carte), des grappes de communes à charge fiscale plus élevée que la moyenne du canton (8.947%) :

Hot spots du canton de Berne							
Grappe	Commune	Charge	LISA	Variance	z-score	p-value	Association
1	435 La Ferrière	9.617	3.515	0.495	4.998	0.000001	High/High
1	441 Renan (BE)	9.712	3.777	0.495	5.370	0.000000	High/High
1	445 Sonvilier	9.808	2.952	0.329	5.148	0.000000	High/High
2	433 Cortébert	9.673	0.984	0.246	1.988	0.046793	High/High
2	434 Courtelary	9.673	1.505	0.246	3.037	0.002386	High/High
3	684 Châtelat	9.434	2.085	0.246	4.207	0.000026	High/High
3	696 Loveresse	9.398	1.132	0.197	2.560	0.010459	High/High
3	697 Malleray	9.749	1.167	0.140	3.132	0.001735	High/High
3	699 Monible	9.327	1.690	0.329	2.949	0.003191	High/High
3	703 Reconvilier	9.452	1.150	0.140	3.085	0.002038	High/High
3	706 Saicourt	9.642	2.247	0.163	5.568	0.000000	High/High
3	707 Saules (BE)	9.220	0.976	0.197	2.209	0.027173	High/High
3	710 Sornetan	9.861	2.404	0.197	5.430	0.000000	High/High
3	713 Tavannes	9.322	0.902	0.163	2.240	0.025107	High/High
3	715 Rebévelier	9.612	1.860	0.495	2.647	0.008119	High/High
4	681 Belprahon	9.381	1.014	0.246	2.048	0.040529	High/High
4	692 Eschert	9.398	1.072	0.246	2.166	0.030302	High/High
4	700 Moutier	9.398	0.891	0.163	2.212	0.026935	High/High
5	579 Habkern	9.238	0.825	0.140	2.217	0.026642	High/High
5	588 Niederried bei Interlaken	9.536	1.369	0.329	2.391	0.016821	High/High
5	589 Oberried am Brienersee	9.577	1.489	0.329	2.599	0.009362	High/High
5	904 Röthenbach im Emmental	9.505	1.196	0.108	3.649	0.000263	High/High
5	906 Schangnau	9.683	1.546	0.246	3.120	0.001806	High/High
5	924 Eriz	9.279	0.902	0.140	2.422	0.015422	High/High
5	935 Oberlangenegg	9.457	0.831	0.140	2.232	0.025593	High/High
6	785 Meiringen	9.398	0.874	0.163	2.170	0.030030	High/High

**Tableau 3 : les « hot spots » de la charge fiscale dans le canton de Berne, année 2014.**

- de trois « cold spots » (en bleu sur la carte), des grappes de communes à charge fiscale moins élevée que la moyenne du canton (8.947%) :

Cold spots du canton de Berne							
Grappe	Commune	Charge	LISA	Variance	z-score	p-value	Association
1	351 Bern	8.604	1.079	0.080	3.819	0.000134	Low/Low
1	352 Bolligen	8.533	0.846	0.097	2.726	0.006404	Low/Low
1	353 Bremgarten bei Bern	8.515	1.332	0.329	2.326	0.020043	Low/Low
1	354 Kirchlindach	8.444	0.949	0.122	2.727	0.006392	Low/Low
1	356 Muri bei Bern	7.985	2.121	0.108	6.464	0.000000	Low/Low
1	361 Zollikofen	8.357	1.613	0.197	3.646	0.000267	Low/Low
1	362 Ittigen	8.248	1.712	0.163	4.244	0.000022	Low/Low
1	401 Aeffligen	8.490	1.026	0.197	2.320	0.020323	Low/Low
1	411 Kernenried	8.259	1.884	0.246	3.802	0.000144	Low/Low
1	412 Kirchberg (BE)	8.562	1.152	0.122	3.309	0.000935	Low/Low
1	415 Lyssach	8.117	1.625	0.122	4.665	0.000003	Low/Low
1	420 Rütligen-Alchenflüh	8.223	2.108	0.197	4.761	0.000002	Low/Low
1	535 Deisswil bei Münchenbuchsee	7.468	3.906	0.197	8.816	0.000000	Low/Low
1	538 Fraubrunnen	8.675	0.897	0.140	2.407	0.016076	Low/Low
1	540 Jegenstorf	8.469	1.448	0.088	4.896	0.000001	Low/Low
1	543 Mattstetten	8.416	0.929	0.163	2.305	0.021151	Low/Low
1	544 Moosseedorf	8.233	2.418	0.197	5.461	0.000000	Low/Low
1	551 Urtenen-Schönbühl	8.505	1.748	0.197	3.950	0.000078	Low/Low
1	553 Wiggiswil	8.020	3.897	0.197	8.797	0.000000	Low/Low
1	557 Zuzwil (BE)	8.487	1.310	0.197	2.961	0.003069	Low/Low
1	623 Rubigen	8.426	1.455	0.246	2.937	0.003310	Low/Low
1	630 Allmendingen	8.337	2.295	0.246	4.631	0.000004	Low/Low
1	632 Wichtrach	8.515	0.682	0.097	2.200	0.027806	Low/Low
1	668 Mühleberg	8.099	0.766	0.108	2.341	0.019226	Low/Low
1	670 Neuenegg	8.495	1.349	0.246	2.723	0.006476	Low/Low
1	861 Belp	8.248	1.915	0.108	5.838	0.000000	Low/Low
1	865 Gelterfingen	8.357	1.077	0.140	2.891	0.003845	Low/Low
1	866 Gerzensee	8.645	1.011	0.163	2.509	0.012117	Low/Low
1	870 Kehrsatz	8.526	1.250	0.197	2.826	0.004716	Low/Low
1	875 Mühledorf (BE)	8.072	1.280	0.197	2.894	0.003807	Low/Low
2	321 Aarwangen	8.533	0.946	0.140	2.540	0.011089	Low/Low
2	323 Bannwil	8.533	1.535	0.140	4.117	0.000038	Low/Low
2	329 Langenthal	8.238	0.689	0.122	1.981	0.047589	Low/Low
2	341 Schwarzhäusern	8.515	1.194	0.246	2.411	0.015903	Low/Low
2	972 Berken	8.045	2.565	0.197	5.791	0.000000	Low/Low
2	976 Graben	8.508	1.265	0.163	3.137	0.001708	Low/Low
2	977 Heimenhausen	8.437	1.542	0.140	4.136	0.000035	Low/Low
2	979 Herzogenbuchsee	8.419	1.221	0.140	3.277	0.001051	Low/Low
2	980 Inkwil	8.508	1.333	0.329	2.327	0.019952	Low/Low
2	981 Niederbipp	8.266	1.004	0.246	2.028	0.042527	Low/Low
2	982 Niederönz	7.974	3.088	0.329	5.385	0.000000	Low/Low
2	990 Walliswil bei Niederbipp	8.177	1.978	0.197	4.467	0.000008	Low/Low
3	494 Gals	8.428	1.232	0.329	2.151	0.031474	Low/Low

**Tableau 4 : les « cold spots » de la charge fiscale dans le canton de Berne, année 2014.**

- de quatre « spatial outliers », des communes avec un profil de charge fiscale complètement différent de celui de leurs communes avoisinantes, à savoir :
  - les communes de Rumendingen, de Treiten, de Champoz (en vert sur la carte) qui ont une charge fiscale inférieure à la moyenne du canton (8.947%) et qui sont plutôt entourées de communes à charge fiscale supérieure à cette moyenne ;
  - la commune d'Oberösch (en jaune sur la carte) qui a une charge fiscale supérieure à la moyenne du canton (8.947%) et qui est plutôt entourée de communes à charge fiscale inférieure à cette moyenne ;

Outliers locaux du canton de Berne						
Commune	Charge	LISA	Variance	z-score	p-value	Association
421 Rumendingen	7.684	-0.863	0.163	-2.129	0.033282	Low/High
500 Treiten	7.992	-1.111	0.246	-2.233	0.025519	Low/High
683 Champoz	8.574	-1.196	0.140	-3.193	0.001408	Low/High
419 Oberösch	9.381	-1.181	0.246	-2.374	0.017599	High/Low

**Tableau 5 : les « spatial outliers » de la charge fiscale dans le canton de Berne, année 2014.**

## 4 Quelques exemples d'analyse spatiale

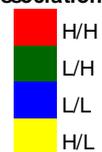
Les mesures des associations spatiales locales permettent de faire ressortir les particularités locales d'un territoire donné, là où, simultanément, la variable d'intérêt est plus grande ou plus petite que la moyenne du territoire et où elle est en contraste par rapport à son voisinage. Nous montrons ici quelques particularités locales au niveau des communes (entité spatiale) par rapport au canton (le territoire de référence) ou à la grande région à laquelle elle appartient, voire par rapport à l'ensemble du pays. Les exemples sont empruntés à la statistique fiscale. Sont exclues des calculs des cantons et des grandes régions les communes enclavées spatialement sur un autre territoire. Les p-values sont estimées par l'approche asymptotique sous l'hypothèse de normalité. Le seuil de significativité est placé à 0.05.

### 4.1 Les charges fiscales localement singulières, par canton

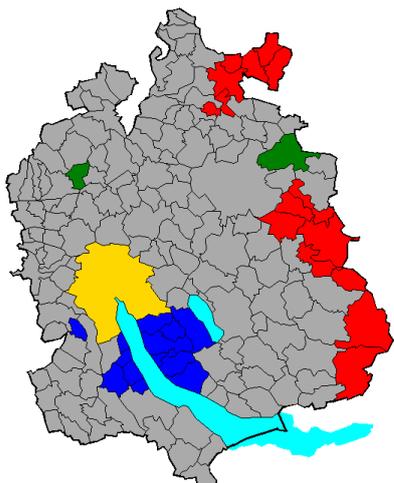
A l'instar de ce qui a été fait pour Berne au chapitre 3, nous calculons les particularités locales significatives de charge fiscale d'autres cantons, pour autant que ceux-ci comprennent un nombre suffisant de communes. Les calculs reposent sur les données de la charge fiscale des couples mariés avec deux enfants et sur une matrice de poids indiquant la proximité entre les différentes communes du canton, le poids valant 1 si les communes sont adjacentes et 0 autrement.

En appliquant le principe que « une image vaut mille mots », nous montrons ci-après les associations locales significatives des cantons de Zurich (autocorrélation spatiale  $I = 0.4259$ ), de Berne ( $I = 0.4075$ ), de Lucerne ( $I = 0.4903$ ), d'Argovie ( $I = 0.3701$ ), de Vaud ( $I = 0.4495$ ) et du Tessin ( $I = 0.4216$ ) sur la base de cartes thématiques (voir les graphiques 12a à 12f). Les résultats sont purement descriptifs ; nous ne cherchons pas, dans le cadre de cette étude, à donner une explication aux différences entre les communes qui ressortent de ces calculs.

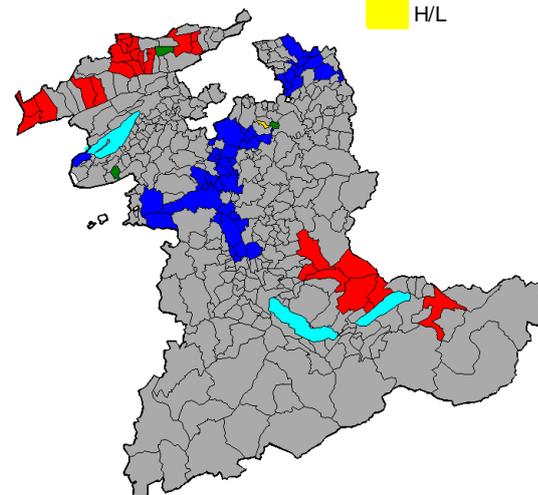
Type d'association spatiale :



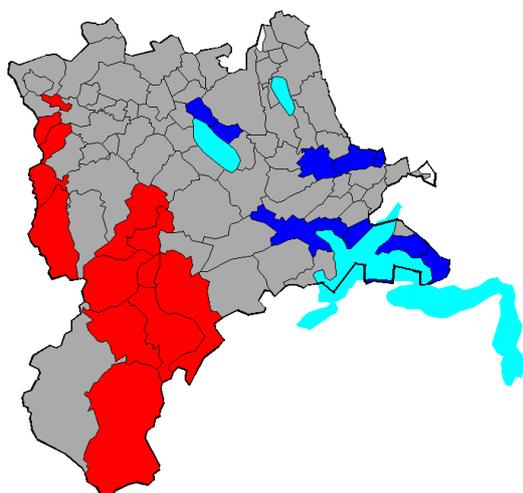
Zurich



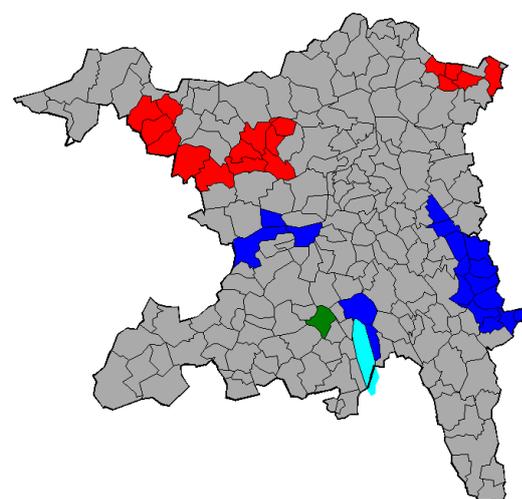
Berne



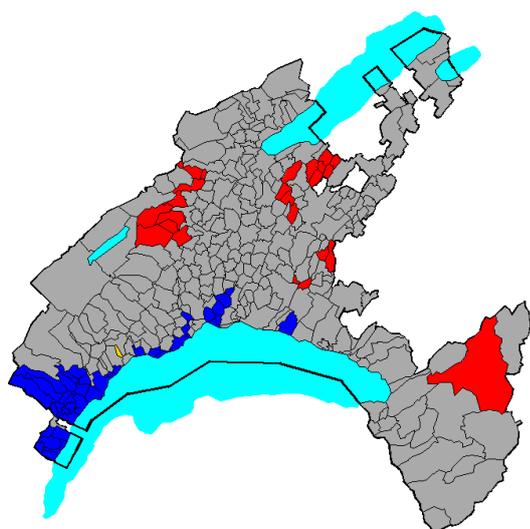
Lucerne



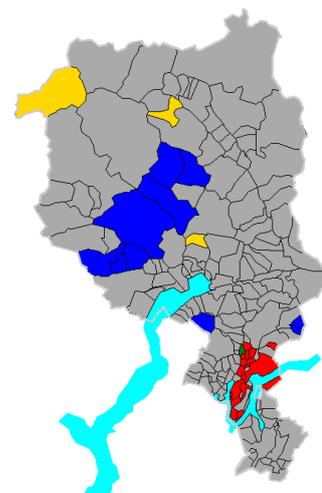
Argovie



Vaud



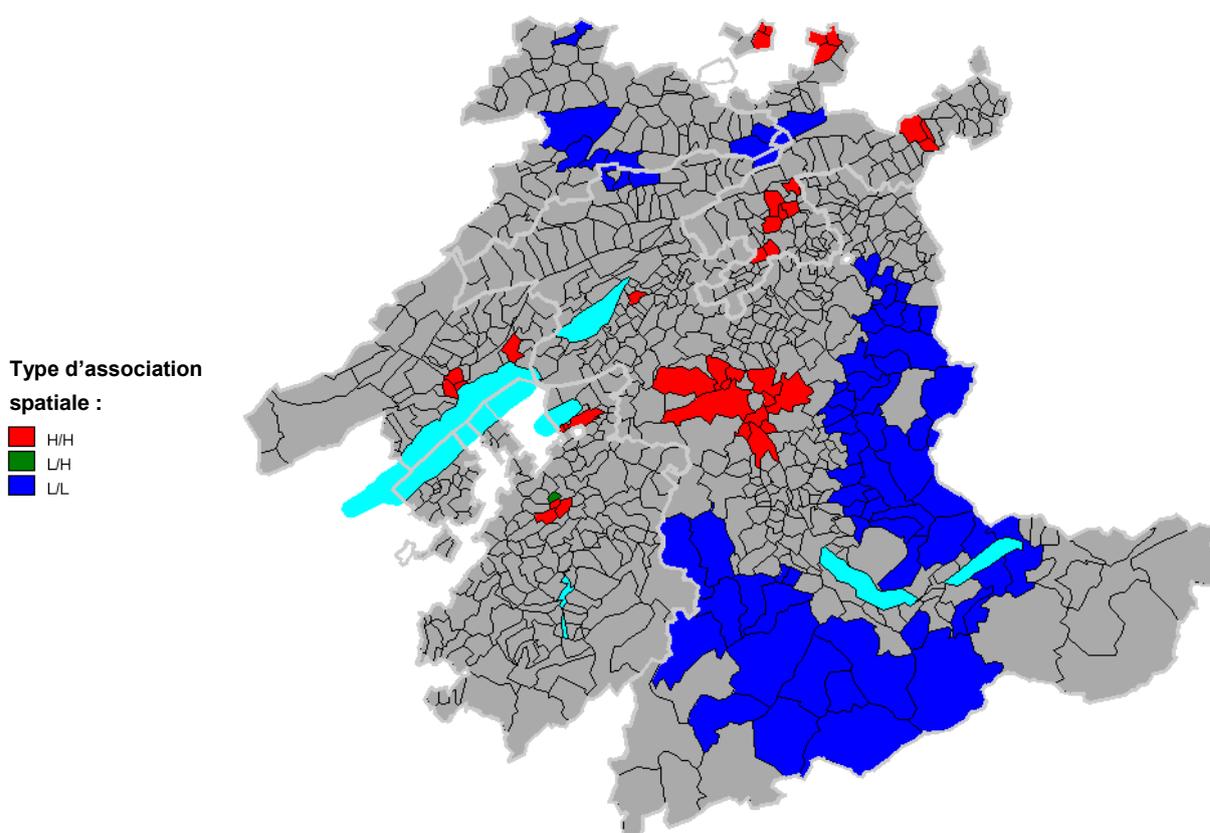
Tessin



Graphiques 12a à 12f : les associations spatiales significatives au sein de six cantons relatives à la charge fiscale des couples mariés avec deux enfants, année 2014.

## 4.2 Les lieux de hauts ou de faibles revenus

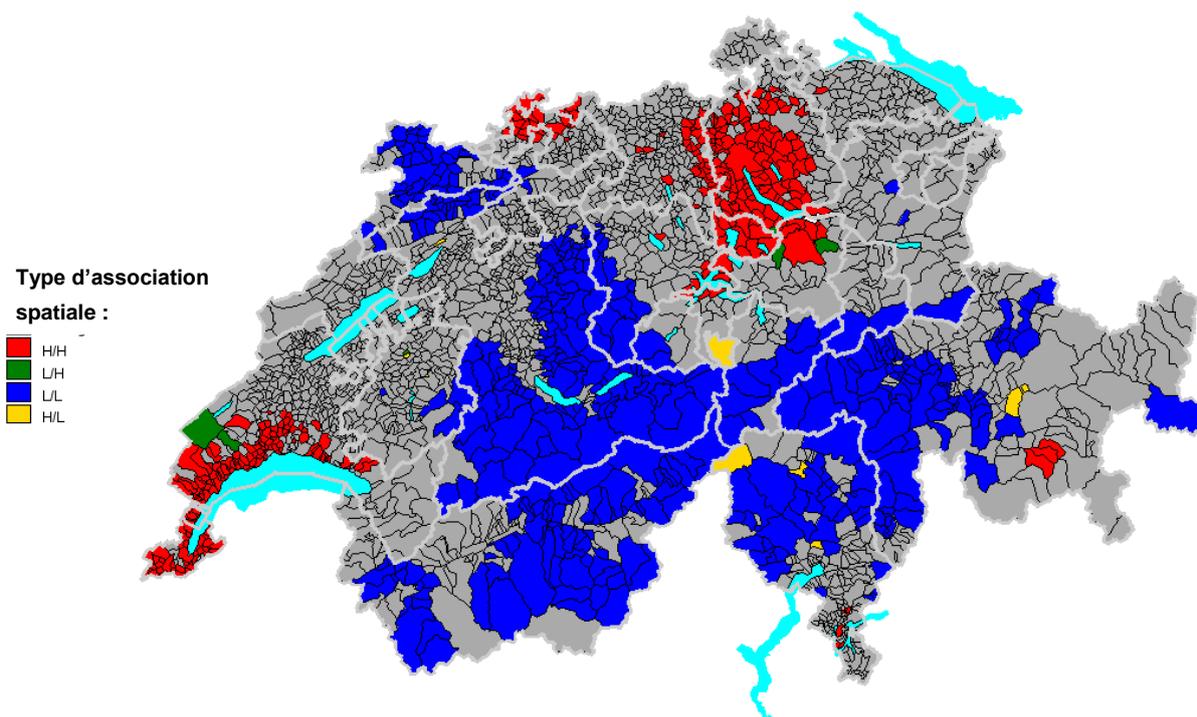
On peut aussi faire ressortir les différences locales significatives de revenus en Suisse. A titre d'exemple, nous montrons au graphique 13 les différences entre les communes de l'« Espace Mittelland » (grande région regroupant les cantons de Berne, de Fribourg, du Jura, de Neuchâtel et de Soleure). Il y apparaît des lieux où les revenus sont significativement plus élevés, respectivement moins élevés, que dans les communes avoisinantes. Les calculs reposent sur les dernières statistiques de l'Administration fédérale des contributions relatives aux revenus équivalents nets des personnes physiques soumises à une taxation normale (les valeurs moyennes par commune, année 2010) et sur la même notion de poids de proximité qu'au chapitre précédent (poids valant 1 si les communes sont adjacentes et 0 autrement).



**Graphique 13 : les associations spatiales significatives relatives au revenu équivalent net moyen par commune dans l'Espace Mittelland, année 2010.**

L'indice de Moran apparaît ici comme une mesure de la ségrégation sociale intercommunale, c'est-à-dire d'une certaine concentration géographique des contribuables de même niveau de revenus au sein de communes avoisinantes. Si on s'intéresse aux sept grandes régions de la Suisse, l'autocorrélation spatiale est la plus élevée en région lémanique ( $I = 0.663$ ) et en Suisse centrale ( $I = 0.625$ ), suivi ensuite par les régions de Zurich ( $I = 0.526$ ), de Suisse orientale ( $I = 0.482$ ), de la Suisse du Nord-Ouest ( $I = 0.468$ ), de l'espace Mittelland ( $I = 0.346$ ) et du Tessin ( $I = 0.282$ ).

Au graphique 14, nous faisons ressortir les différences locales significatives de revenus au niveau de toute la Suisse. Pour tenir compte de l'éloignement en kilomètres des communes, sont appliqués ici les poids « inverses des distances » dans le calcul des associations spatiales (les *poids3* du chapitre 2.2.2). Avec ces poids, l'indice de Moran au niveau de la Suisse s'élève à 0.389.



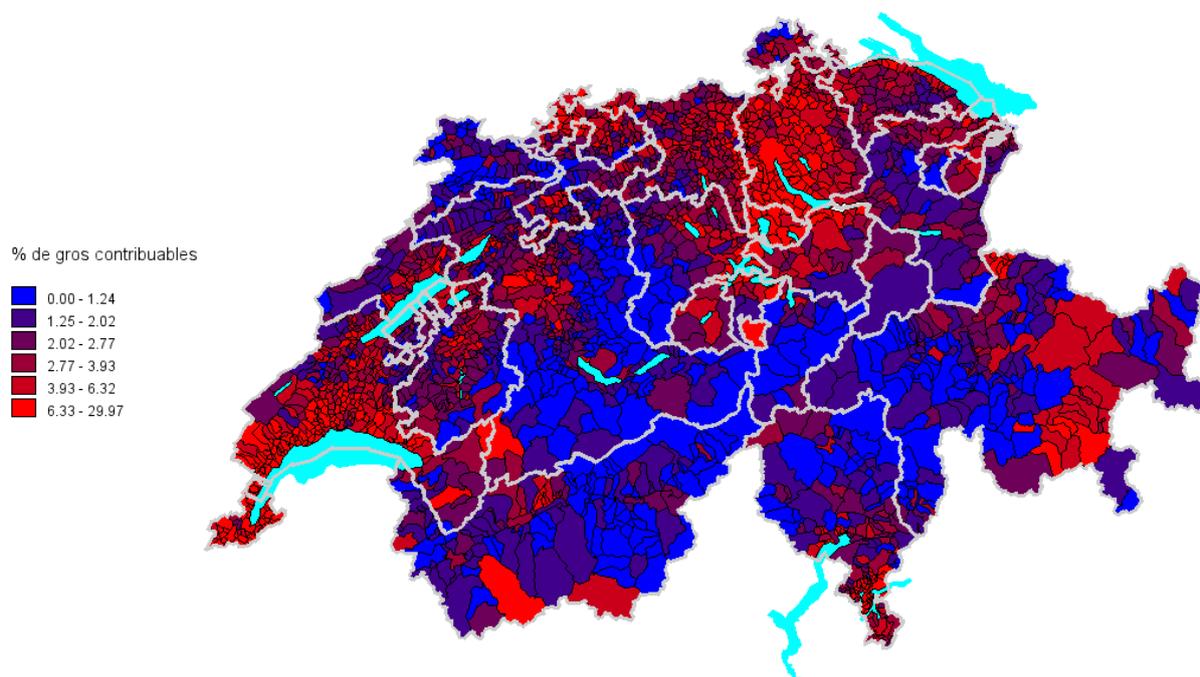
**Graphique 14 : les associations spatiales significatives relatives au revenu équivalent net moyen par commune en Suisse, année 2010.**

### 4.3 Les lieux où préfèrent vivre les « gros contribuables »

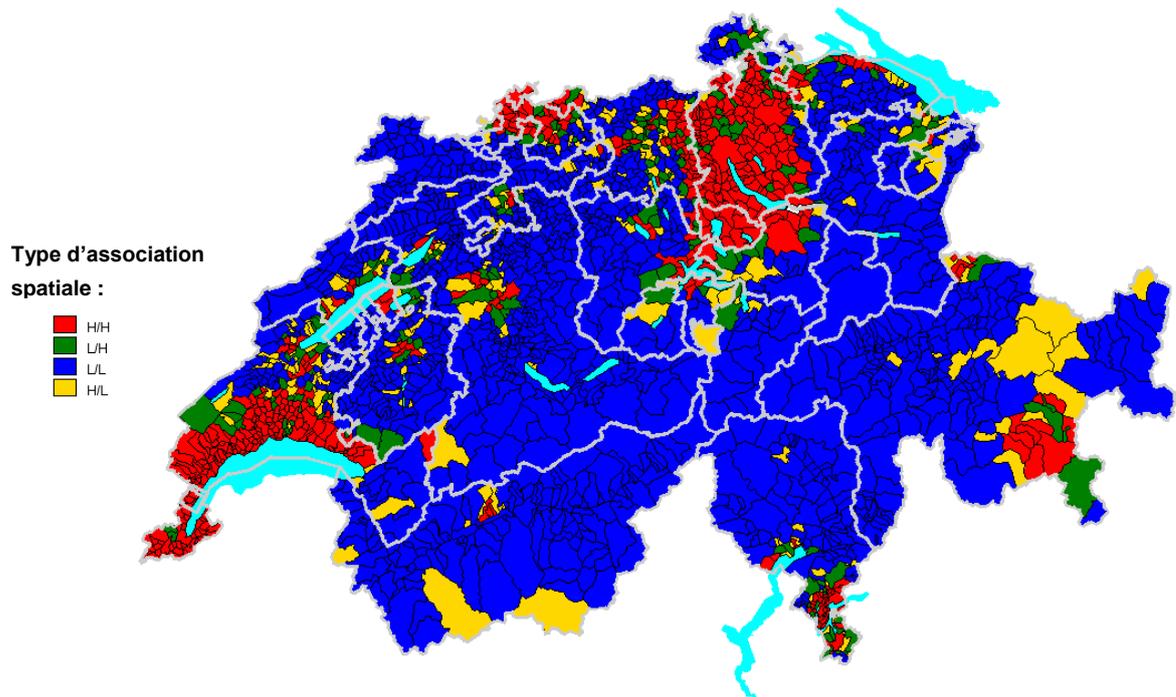
On peut également mettre en évidence par les indices LISA de la répartition géographique sur le territoire suisse des « gros contribuables ». Nous considérons ici comme « gros contribuables » les 5% de contribuables du pays payant le plus d'impôt fédéral direct, c'est-à-dire tous ceux qui acquittent un impôt fédéral direct d'au moins 5'755 francs ; ces contribuables assurent 68.4% des entrées de l'impôt sur les revenus pour la Confédération. Sont considérés pour cette étude les données de l'impôt fédéral direct, année 2012, de tous les contribuables, à l'exception de ceux soumis à une taxation à la source.

Si nous calculons la part de gros contribuables dans chaque commune suisse, nous constatons une forte concentration de « gros contribuables » en certains lieux de la Suisse. Un classement en six classes est présenté au graphique 15, les 5<sup>e</sup>, 10<sup>e</sup>, 50<sup>e</sup>, 90<sup>e</sup> et 95<sup>e</sup> centiles de pourcentage délimitant les six classes. Les différences de densité ressortent plus clairement au graphique 16 avec une représentation des associations spatiales **H/H** (High/High, une haute valeur entourée de hautes valeurs), **L/H** (Low/High, une basse valeur entourée de hautes valeurs), **L/L** (Low/Low, une basse valeur entourée de basses valeurs) et **H/L** (High/Low, une haute valeur entourée de basses valeurs). Comme au chapitre précédent, nous utilisons les poids « inverses des distances » dans le calcul des associations spatiales entre les communes du pays.

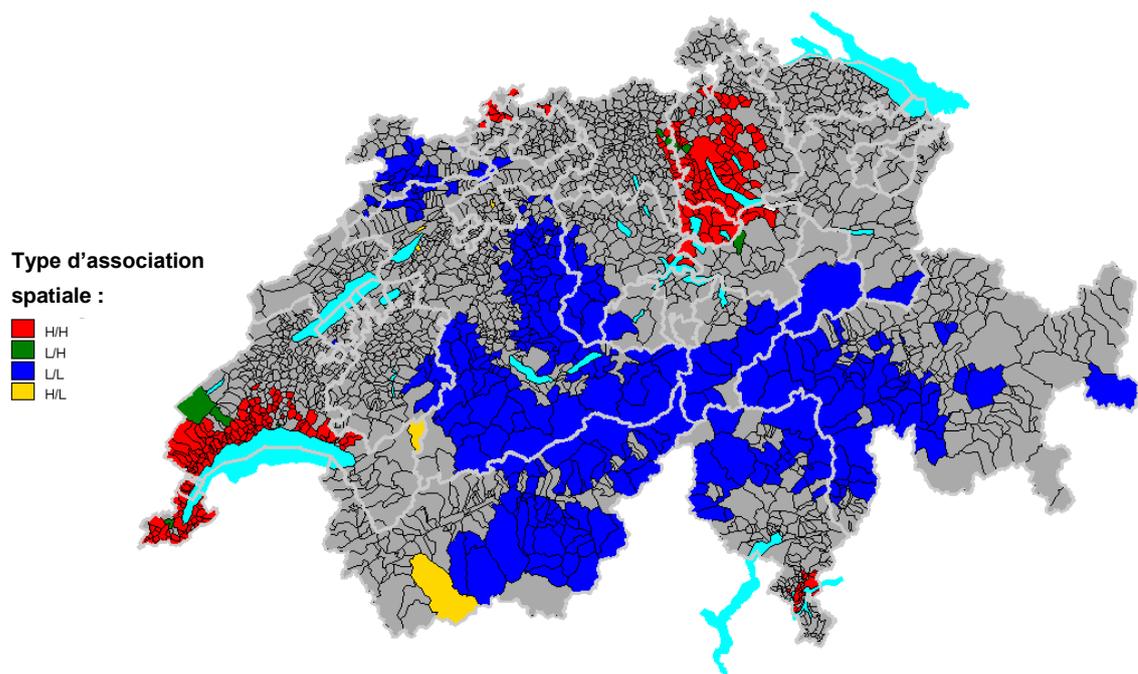
L'indice global d'autocorrélation spatiale s'élève à  $I = 0.499$ , indiquant une assez forte densité de « gros contribuables » en certains endroits du pays. Si on ne retient que les associations spatiales significatives (voir le graphique 17), on remarque que les gros contribuables se concentrent surtout au bord des lacs Léman, de Zurich, de Zoug, des quatre cantons et de Lugano. Par contre, certaines communes du Jura, de l'Oberland bernois, du Valais et des Grisons semblent localement davantage évitées par les gros contribuables.



**Graphique 15 : le pourcentage de « gros contribuables » dans les différentes communes du pays, année 2012.**



**Graphique 16 : les associations spatiales relatives à la densité de « gros contribuables » dans les communes, année 2012.**



**Graphique 17 : les associations spatiales significatives relatives à la densité de « gros contribuables » dans les communes, année 2012.**

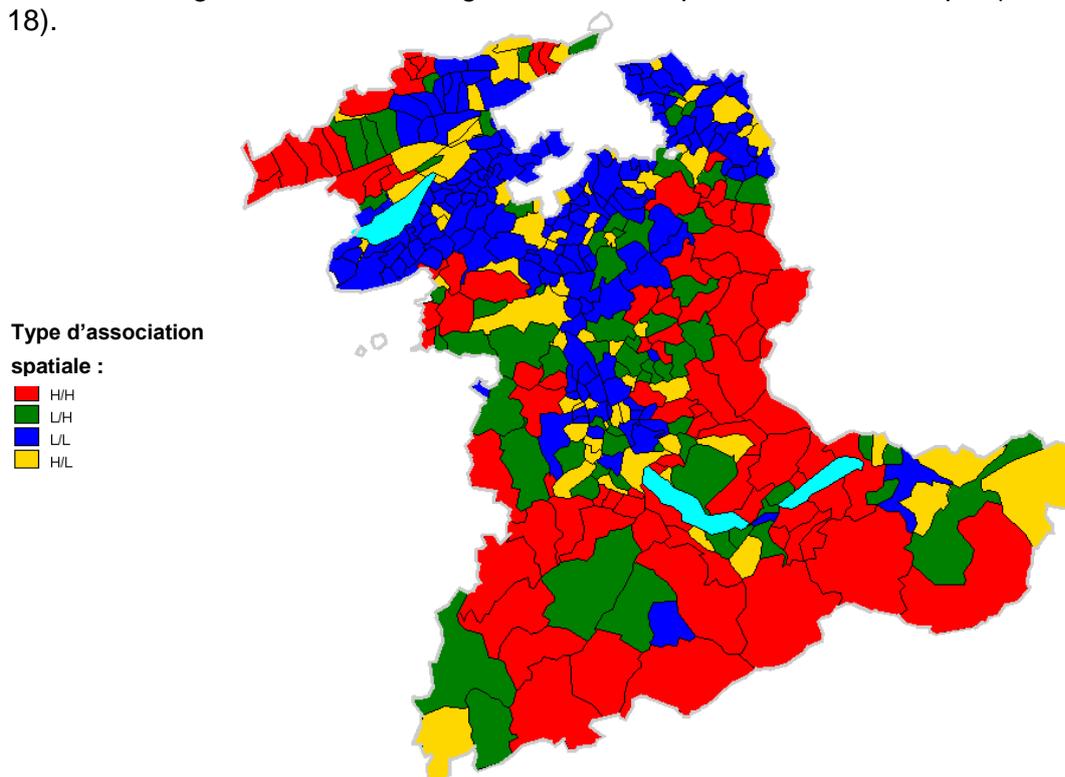
## 4.4 Une exploitation des déductions fiscales différente sur le territoire

Les déductions fiscales accordées dans l'imposition fédérale des revenus sont-elles exploitées par les contribuables de la même manière dans toutes les régions du pays ? L'analyse spatiale des données fiscales permet de répondre à cette question. A titre d'illustration, nous montrons la situation avec la déduction relative aux frais d'entretien des immeubles (uniquement pour le canton de Berne) et avec celle relative aux cotisations versées au pilier 3a.

### Les déductions sur les immeubles de particuliers

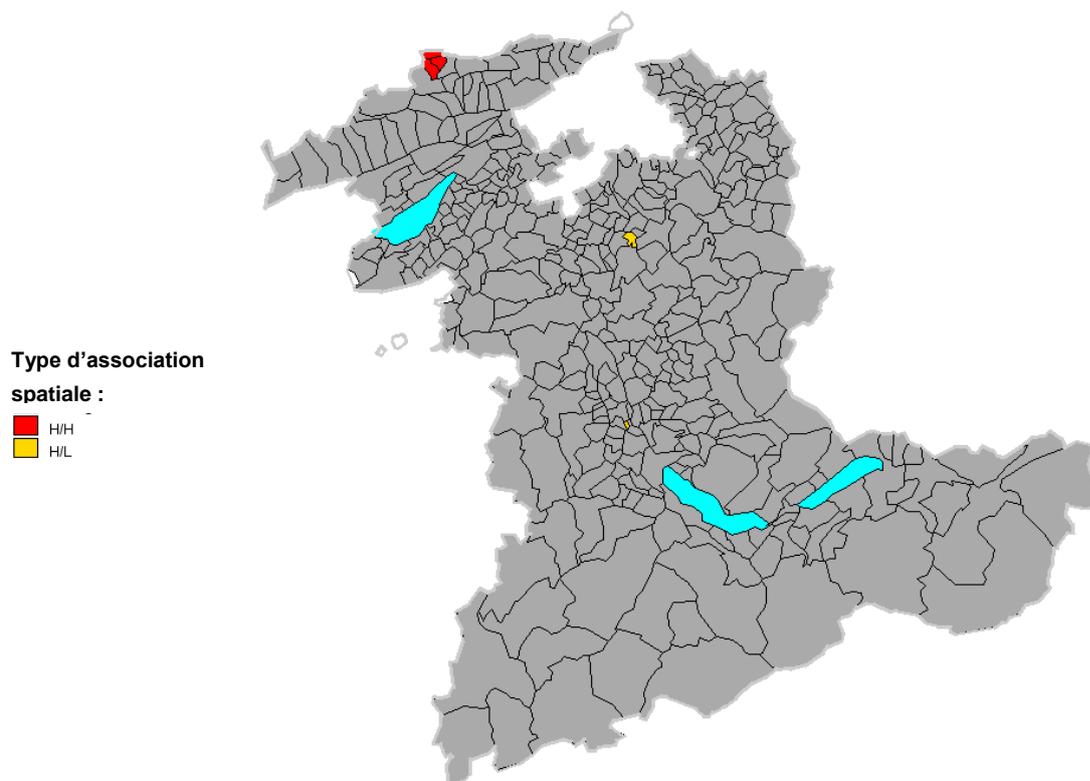
Sont considérées ici les données fiscales de 211'962 immeubles situées dans le canton et détenues par des personnes physiques (année 2011, après élimination de quelques données jugées non plausibles).<sup>7</sup> Est estimée pour chaque immeuble la part que constituent les frais d'entretien (réel ou forfaitaire), l'impôt foncier et les frais d'administration et de gestion par rapport à la valeur officielle de l'immeuble. Une moyenne par commune est calculée et la corrélation spatiale entre les communes mesurée en attribuant à chaque paire de communes un poids de proximité valant 1 quand les communes sont adjacentes et 0 autrement.

Globalement, la corrélation spatiale n'est pas très élevée au niveau cantonal ( $I = 0.137$ ). Les régions montagneuses ont tendance à déduire davantage de frais immobiliers que les communes urbaines ou de la plaine (voir le graphique 17). Cependant les différences locales sont très faibles et généralement non significatives d'un point de vue statistique (voir le graphique 18).



**Graphique 16 : les associations spatiales relatives aux déductions sur les immeubles dans les communes du canton de Berne, année 2011.**

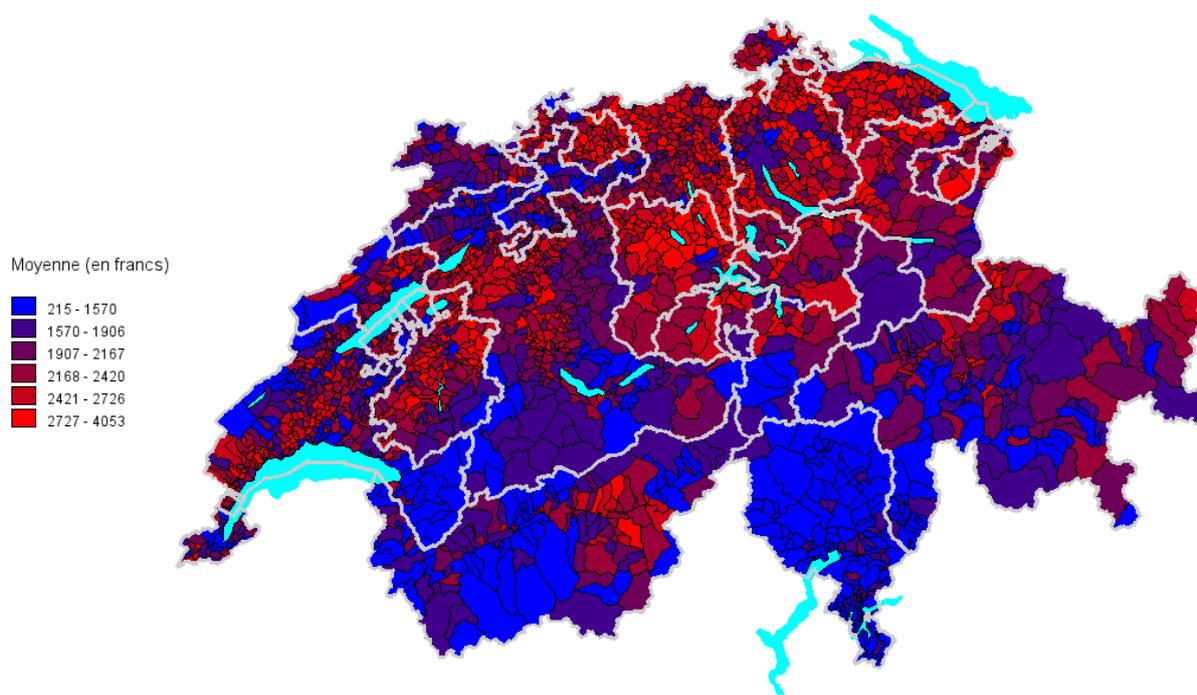
<sup>7</sup> Les communes de Gurbrü et Schelten ne sont pas considérées dans les calculs, aucune donnée étant récoltée pour ces communes.



**Graphique 17 : les associations spatiales significatives relatives aux déductions sur les immeubles dans les communes du canton de Berne, année 2011.**

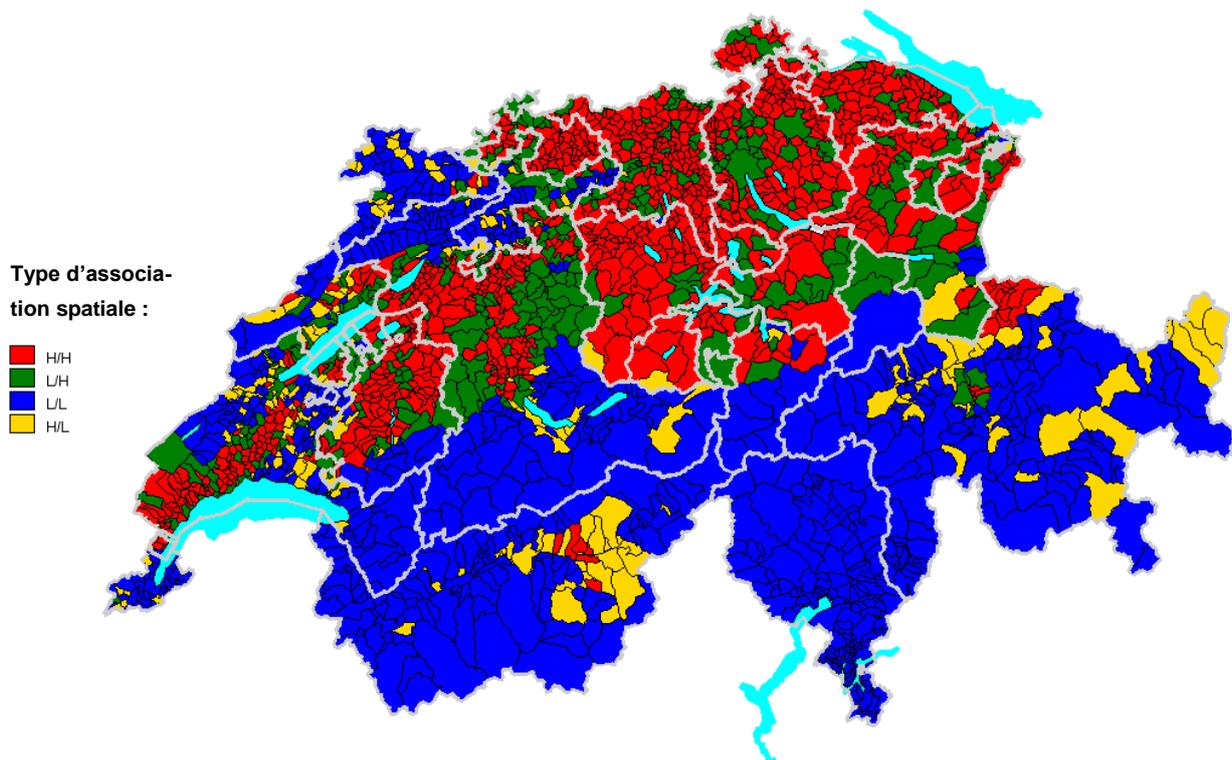
### Les cotisations versées au pilier 3a

Sont considérées ici au niveau de chaque commune suisse les montants moyens de cotisations versées au pilier 3a des contribuables (année 2012), à l'exception de ceux soumis à une taxation à la source. De grandes différences apparaissent entre les montants, les moyennes par commune variant entre 215 et 4'053 francs (voir le graphique 18).

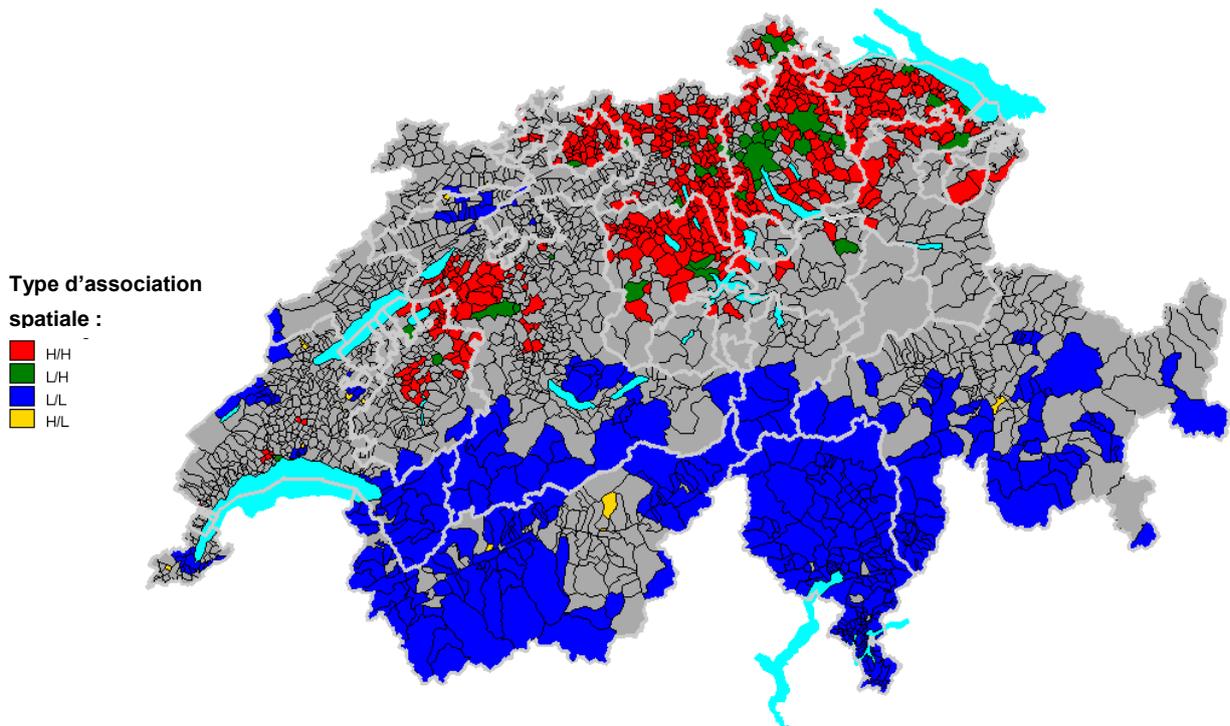


**Graphique 18 : la moyenne des déductions de cotisations versées au pilier 3a dans les différentes communes du pays, année 2012.**

A chaque paire de communes peut être attribué un poids de proximité correspondant à l'inverse de la distance entre leur centre de gravité. La corrélation spatiale apparaît globalement assez élevée en Suisse ( $I = 0.372$ ). Les déductions localement les plus importantes se retrouvent au nord du pays, les plus faibles au sud du pays et le long du massif jurassien (voir les graphiques 19 et 20).



**Graphique 19 : les associations spatiales relatives aux déductions moyennes des cotisations versées au pilier 3a dans les communes, année 2012.**



**Graphique 20 : les associations spatiales significatives relatives aux déductions moyennes des cotisations versées au pilier 3a dans les communes, année 2012.**

## 5 Bibliographie

- L. Anselin (1995). Local Indicators of Spatial Association-LISA, Geographical Analysis Vol. 27,1995, p 93-115.
- L. Anselin, I. Syabri et Y. Kho (2006). GeoDa: An Introduction to Spatial Data Analysis.
- L. Anselin et son équipe, divers travaux sous <https://geodacenter.asu.edu/>.
- R. Briggs (2010). Global Measures of Spatial Autocorrelation, Henan University.
- A.D. Cliff et J.K. Ord. (1981). Spatial processes: models and applications.
- M. Franklin (2013). Spatial Statistics. Course Syllabus.
- R. Kosfeld (2013), Universität Kassel, Assignment6\_result.pdf sous <http://www.uni-kassel.de/~rkosfeld/lehre/spatial/>.
- R. Kosfeld, H.-F. Eckey, M. Türck et M. A. Kassel. LISA (Local Indicators of Spatial Association).
- J. Lentz (2009). Global & Local Statistics References.
- G. Mack (2011) Der Blick zum Nachbarbetrieb: Produktionsballungen am Beispiel der Mutterkuhhaltung. Infotagung Agrarökonomie, 15.09.2011.
- M. S. Rosenberg & C. D. Anderson (2011), Center for Evolutionary Medicine and Informatics / Biodesign Institute School of Life Sciences, Arizona State University, <http://www.passagesoftware.net/manual.php>.
- W. Mitchell (2013). Introduction to Spatial Econometric Modelling.
- New York Census research Data Center (2008). Non-technical Overview of Geospatial Statistical Methods.
- A. Troy. Spatial Autocorrelation Basics.
- G. Upton et B. Fingleton (1985). Spatial data analysis by example, New York, Wiley.
- C. C. Yrigoyen (2007). Course on Spatial Econometrics with Applications.