

Chapitre 2

Détecter la dépendance spatiale (Moran, Geary, LISA)

Corrigé

Jaouad Madkour

Faculté d'Économie et de Gestion de Tanger
Exercices et corrigés · Économétrie spatiale (Master)

1 Exercices du chapitre

Correction — Exercice 1

- (a) $I = -1/3 \approx -0,3333$ (calcul détaillé identique à celui du cours).
- (b) Pour $y = (100, 1, 1, 1)$, on trouve à nouveau $I = -0,3333$, exactement.
- (c) Avec un graphe **complet** sur n unités (chaque unité voisine de toutes les autres, poids $1/(n-1)$), on a l'identité algébrique $\sum_{i \neq j} (y_i - \bar{y})(y_j - \bar{y}) = [\sum_i (y_i - \bar{y})]^2 - \sum_i (y_i - \bar{y})^2 = 0 - \sum_i (y_i - \bar{y})^2$ (la somme des écarts centrés est toujours nulle). En substituant dans la formule de I avec $S_0 = n$, tous les termes en y s'annulent et il ne reste que $I = -1/(n-1)$ — une identité **purement algébrique**, vraie pour *n'importe quelle* série y , dès lors que W est le graphe complet standardisé en ligne.
- (d) Un test de Moran n'est informatif que si W n'est **pas** un graphe complet : il faut soit un nombre d'unités n suffisamment grand pour qu'un graphe complet reste peu réaliste géographiquement, soit une structure de voisinage **creuse** (chaque unité n'ayant qu'un sous-ensemble de voisines, comme la contiguïté tour de l'exercice 2), pour que I puisse effectivement varier avec les données et révéler une dépendance réelle.

Correction — Exercice 2

- (a) $\bar{y} = (12 + 15 + 9 + 11)/4 = 11,75$. Écarts centrés : $y_A - \bar{y} = 0,25$; $y_B - \bar{y} = 3,25$; $y_C - \bar{y} = -2,75$; $y_D - \bar{y} = -0,75$.
- (b) $I = -0,0133$.
- (c) $C = 0,76$.
- (d) $I = -0,0133$ est très **éloigné** de $E[I] = -1/3 \approx -0,333$ (c'est-à-dire proche de 0) : ceci signale l'**absence** de structure spatiale marquée sous ce critère de voisinage, plutôt que la dépendance négative systématique qu'imposerait un graphe complet. $C = 0,76 < 1$ suggère au contraire une **légère autocorrélation positive** (valeurs voisines un peu plus semblables que la moyenne). Les deux indices ne racontent donc pas tout à fait la même histoire ici : le signal est faible et ambigu, ce qui est cohérent avec un échantillon de seulement $n = 4$ régions — bien trop petit pour qu'un test de significativité tranche la question avec fiabilité.

Correction — Exercice 3

- (a) $z_A \approx -0,5071$; $z_B \approx 0,1690$; $z_C \approx 1,5213$; $z_D \approx -1,1832$.
- (b) $(Wz)_A \approx 0,1690$ (positif, $z_A < 0$) : quadrant **Bas-Haut**. $(Wz)_B \approx -0,0563$ (négatif, $z_B > 0$) : quadrant **Haut-Bas**. $(Wz)_C \approx -0,5071$ (négatif, $z_C > 0$) : quadrant **Haut-Bas**. $(Wz)_D \approx 0,3944$ (positif, $z_D < 0$) : quadrant **Bas-Haut**.
- (c) $I_A = z_A(Wz)_A \approx -0,0857$; $I_B \approx -0,0095$; $I_C \approx -0,7714$; $I_D \approx -0,4667$.
- (d) $\sum_i I_i \approx -0,0857 - 0,0095 - 0,7714 - 0,4667 = -1,3333 = 4 \times (-1/3) = n \times I$. L'identité est vérifiée numériquement — cohérent avec le fait que toutes les régions se situent ici dans les quadrants « de contraste » (Haut-Bas, Bas-Haut), ce qui produit un I global négatif.

2 Exercice cumulatif (chapitre 2, chapitre 1 et introduction)

Correction — Exercice 4

- (a) **Non**. L'indice de Moran mesure uniquement l'existence d'une **corrélation statistique** entre la valeur d'une unité et celle de ses voisines, pondérée par W ; il ne dit rien sur le *mécanisme* qui la produit. Un I élevé et significatif est parfaitement compatible avec une corrélation nuisible (une variable omise, elle-même spatialement corrélée, comme un facteur régional commun) tout autant qu'avec une dépendance substantielle.
- (b) Le chercheur peut légitimement conclure que l'hypothèse d'**indépendance spatiale** des

observations, nécessaire à l'inférence usuelle par MCO, est rejetée par les données — sans pouvoir encore dire si cela biaiserait les coefficients (SAR) ou seulement les écarts-types (SEM).

(c) Si W était construite à partir de la variable de pauvreté elle-même (ou d'une variable corrélée avec elle), un I significatif deviendrait pratiquement automatique par construction, indépendamment de toute vraie dépendance spatiale sous-jacente — on retrouverait le même problème de confusion entre dépendance spatiale et corrélation entre régresseurs que la mise en garde d'Anselin (chapitre 1) visait à éviter.

(d) **Non.** L'indice de Moran est un test de *détection*, pas un outil de *diagnostic* : il ne distingue pas où loge la dépendance. Trancher entre SAR et SEM exige des outils supplémentaires (tests LM spécifiques, présentés au chapitre 5) qui vont au-delà du simple indice de Moran.

3 Applications en sciences économiques

Correction — Exercice 5

(a) $E[I] = -1/(30 - 1) = -1/29 \approx -0,0345$.

(b) $z = \frac{0,42 - (-0,0345)}{\sqrt{0,0087}} = \frac{0,4545}{0,0933} \approx 4,873$.

(c) **Oui** : $4,873 \gg 1,96$, l'hypothèse nulle d'absence d'autocorrélation spatiale est très largement rejetée.

(d) Avec seulement $n = 30$ quartiers, le cours recommande de privilégier le **test par permutation** plutôt que l'approximation normale asymptotique : celle-ci suppose une distribution sous-jacente (normalité, ou au moins un moment d'ordre 4 fini) que peu d'observations vérifient rarement de façon fiable, alors que le test par permutation construit sa propre référence directement à partir des données observées, sans hypothèse distributionnelle supplémentaire.

Correction — Exercice 6

(a) Une commune **Haut-Haut** présente un taux d'incidence élevé, **entourée** de communes également à taux élevé — un véritable foyer épidémique localisé, pas un cas isolé.

(b) Une commune **Bas-Haut** présente un taux d'incidence **faible**, alors que ses voisines ont un taux élevé — une commune qui « résiste » localement à une épidémie par ailleurs concentrée dans son voisinage immédiat, une configuration atypique qui mérite une attention particulière (mesures de prévention efficaces à identifier et généraliser, ou simplement sous-déclaration à vérifier).

(c) L'indice de Moran global aurait fourni un **unique verdict** sur l'ensemble des 120 communes (dépendance significative ou non), sans jamais indiquer *où* se situent les foyers. La carte LISA localise précisément les 15 communes en foyer actif et les 3 communes atypiques, une information directement actionnable qu'un indice global, par construction, ne peut pas fournir.

(d) Concentrer prioritairement les campagnes de vaccination sur les 15 communes Haut-Haut et leur voisinage immédiat (le foyer et sa zone de propagation la plus probable), plutôt que de répartir uniformément des ressources limitées sur les 120 communes — et examiner spécifiquement les 3 communes Bas-Haut pour comprendre si leur résistance apparente reflète une protection réelle (à répliquer ailleurs) ou une sous-déclaration (à corriger).