

LE RECOURS AU MODELE D'ANALYSE DE VARIANCE DANS LES TRAVAUX GEOGRAPHIQUES

HELLE Cécile

Laboratoire Structures et Dynamiques Spatiales
Université d'Avignon et des Pays de Vaucluse
Faculté des lettres et sciences humaines
BP 2013
84000 Avignon
Tél : 90 85 28 50 poste 31

Résumé

Dans le cadre de l'analyse du marché des terrains à bâtir dans l'espace vauclusien, on se propose de tester les éventuels apports (notamment par rapport aux méthodes classiquement mobilisées dans ce type d'études : modèle de régression linéaire multivariée...) de l'analyse de la variance. Deux axes de recherche sont privilégiés :

- le modèle de variance comme outil explicatif de la différenciation spatiale des prix des terrains à bâtir (grâce entre autres à une meilleure prise en compte des discontinuités spatiales),
- le modèle de variance comme outil de vérification, faisant suite à une démarche inductive de traitements des données ayant abouti à la projection d'un modèle d'organisation spatiale sur une structure observée dans la réalité.

Mots Clés

Analyse de la variance - Qualification de l'espace - Raisonnement inductif - Raisonnement déductif

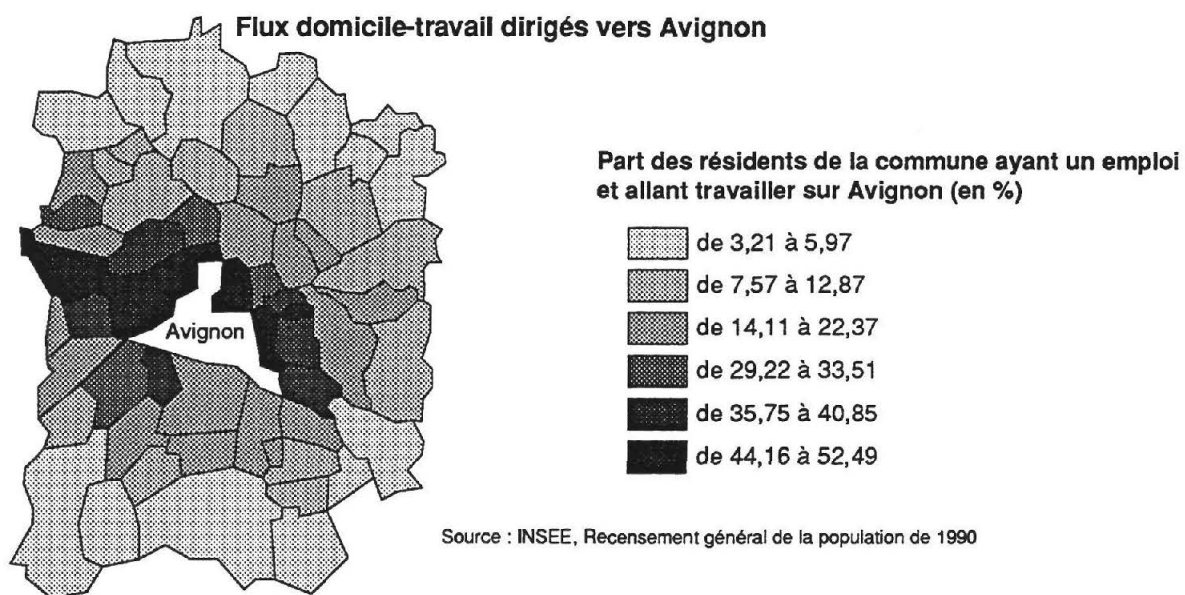
En Vaucluse, le processus d'urbanisation impulsé par la ville centre, Avignon, ainsi que par les centres urbains de taille moyenne gravitant autour de cette dernière entraîne de profondes modifications dans l'organisation du territoire et dans les paysages traditionnels. Or, cet espace, déjà densément occupé, est soumis à de multiples autres contraintes concurrentielles du point de vue de l'affectation du sol : agriculture à forte valeur ajoutée, villégiature touristique, grands axes de communication, protection du patrimoine... Aussi, l'explication des principaux mécanismes d'extension du bâti au regard de ces contraintes et de leurs interactions est apparue judicieuse. On a choisi de fonder cette analyse sur une étude préalable du marché des terrains à bâtir, se focalisant sur les différences spatiales du prix du m² de terrain constructible.

Après la saisie de l'information initiale grâce au dépouillement systématique des extraits d'actes notariés archivés aux services fiscaux départementaux et la présentation cartographique des caractéristiques essentielles du fonctionnement du marché vauclusien des terrains à bâtir, le choix méthodologique de la modélisation statistique à adopter pour expliquer la différenciation spatiale du prix du m² s'est posé.

Au regard des études récentes tentant d'explicitier cette différenciation dans un espace régional, ce choix peut se porter sur le modèle de régression linéaire multivariée en raison des multiples avantages qu'il présente. Ce dernier assure en effet des résultats globalement satisfaisants¹ ; de plus, il permet l'intégration, au côté de la distance, de plusieurs autres facteurs explicatifs ; enfin, il autorise le classement éventuel des différentes variables explicatives par ordre d'importance grâce à une méthode d'analyse itérative qui permet de juger la qualité de la réduction de l'information résiduelle lors de l'intégration d'un nouveau facteur explicatif.

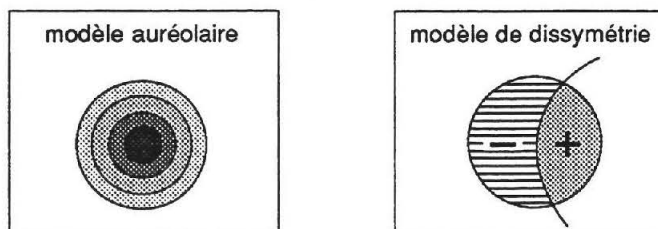
Néanmoins, ce type de modèle ne parvient pas à appréhender de façon satisfaisante l'information qualitative.

Première phase : La représentation cartographique



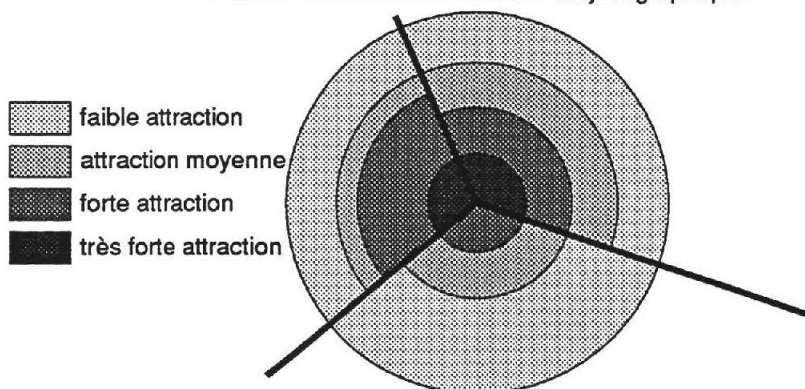
Deuxième phase : La modélisation graphique

Les deux chorèmes utilisés



Avignon, centre d'emploi attractif

Une structure spatiale de type auréolaire, complexifiée par un phénomène de dissymétrie dû à une frontière administrative et hydrographique



Ce problème apparaît crucial puisque les ouvrages qui procèdent au recensement méthodique des paramètres jouant sur la formation des prix des terrains à bâtir révèlent que nombre de ces derniers sont des variables qualitatives : les variables d'agrément sur l'environnement (existence d'un boisé, vue particulière...) ou les critères juridiques de constructibilité (mise en place éventuelle d'un POS, règlements architecturaux restrictifs...) en sont autant d'exemples.

Par ailleurs, l'étude statistique et cartographique des principaux processus de fonctionnement du marché vaclusien des terrains à bâtir confirme l'acuité du problème en dévoilant l'existence non pas d'un espace foncier vaclusien mais de sous-espaces fonciers. La différenciation spatiale du prix du m² de terrain à bâtir semble donc se définir dans le département de Vaucluse plus par des discontinuités que par des transitions : dans la périphérie d'Avignon, même si l'on observe un échelonnement des prix avec la distance au centre urbain, cette organisation auréolaire générale² masque une réalité plus complexe puisque, d'une commune à l'autre, située pourtant à égale distance d'Avignon, il existe une variation des prix qui témoigne de l'intervention, dans la logique d'organisation du marché foncier, de facteurs plus subjectifs comme le cadre de vie.

Ces deux constatations attestent de l'opportunité de travailler en géographie foncière sur de l'information qualitative plutôt que sur des données quantitatives. Ainsi, à titre d'exemple, qualifier les communes selon le type d'espace auquel elles appartiennent semble aussi justifié que de déterminer leur niveau d'équipement en services aux particuliers.

Parmi toute la statistique des variables qualitatives, le modèle d'analyse de variance est un recours possible puisqu'il permet l'explication d'une variable quantitative Y par un ensemble de variables explicatives qualitatives (Xi avec $i = 1...n$).

Paradoxalement, bien que le concept d'analyse de variance soit très usité en géographie, les géographes utilisent peu fréquemment le modèle d'analyse de variance. Pourtant la logique de ce modèle qui permet d'observer la structure spatiale de Y en comparant la distribution totale de Y aux distributions de Y dans différents sous-espaces³, rend possible d'une part, le test d'hypothèses sur les répartitions spatiales (donc la prise en considération de l'espace géographique autrement que comme simple support)⁴ et d'autre part, l'introduction dans un processus de modélisation de variables qualitatives⁵.

Seul le difficile respect des trois contraintes assurant la validité du modèle (groupes d'effectifs voisins, distribution gaussienne de l'ensemble des valeurs et des valeurs de chaque groupe, variance des groupes de même ordre de grandeur) peut expliquer cette attitude frileuse. Ainsi, une description adéquate de l'espace foncier nécessite-t-elle la juxtaposition de multiples qualités. Or, cette multiplicité a pour corollaire la multiplication des sous-espaces et par conséquent la diminution du nombre des individus dans chaque entité spatiale, critère intervenant sur la significativité des résultats obtenus. Par ailleurs, plus le nombre de paramètres explicatifs est grand, plus difficile est l'interprétation des résultats car les variations liées aux interactions entre les différents facteurs sont d'autant plus nombreuses. A titre d'exemple, une analyse de variance à quatre facteurs engendre six interactions de premier ordre (intervention de deux facteurs), quatre interactions de deuxième ordre et une interaction provenant simultanément des quatre variables.

Un exemple va nous permettre de montrer de façon plus concrète l'optique dans laquelle on souhaite utiliser le modèle d'analyse de variance.

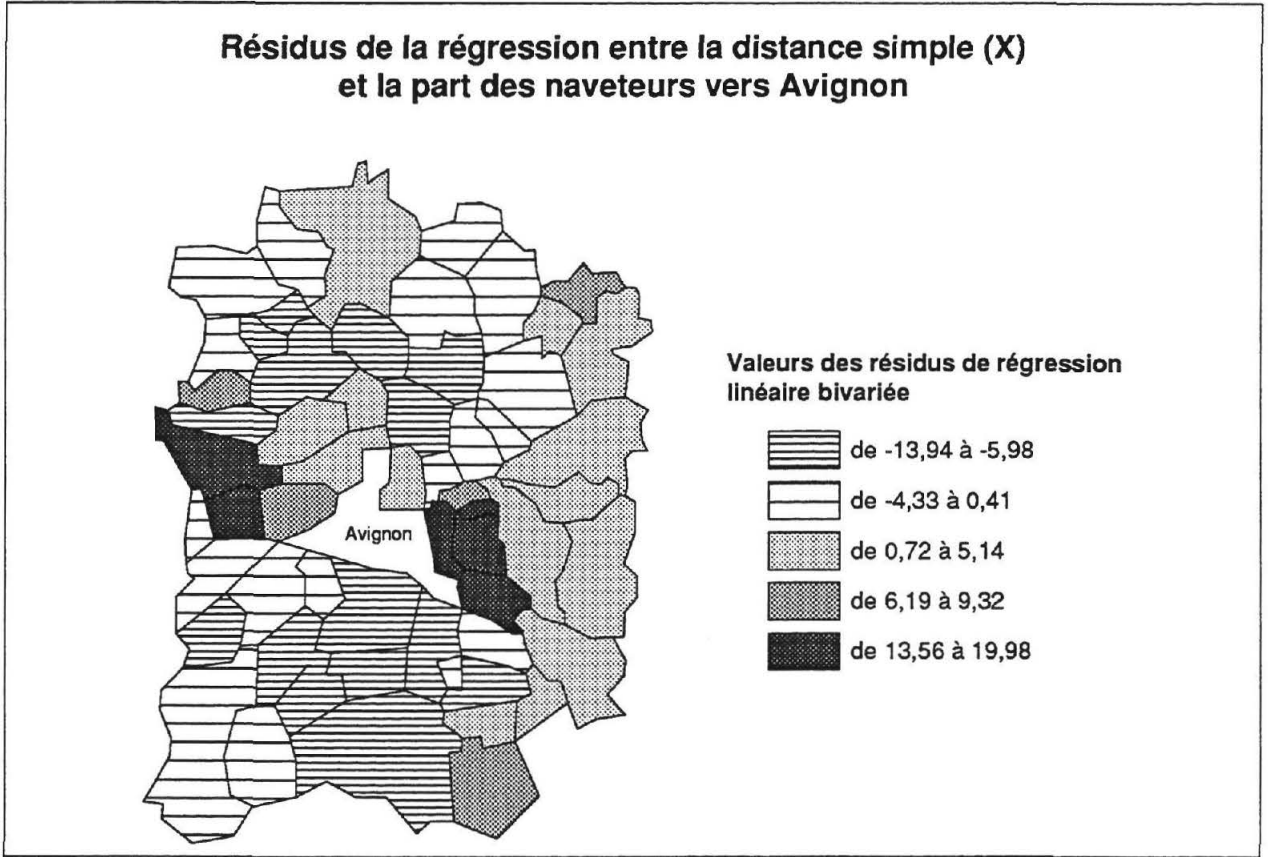
Bien que ne concernant pas directement notre sujet de recherche⁶, cet exemple affecte la région avignonnaise et étudie un indicateur du processus d'urbanisation : les migrations pendulaires en direction d'Avignon. Ces dernières autorisent en effet, l'appréhension du degré de dépendance des communes de l'espace avignonnais vis-à-vis de la ville centre, pôle d'emploi.

La première étape de l'analyse a consisté à représenter cartographiquement des flux domicile-travail dirigés vers Avignon : aux valeurs brutes (nombre de naveteurs vers Avignon), on a préféré un paramètre relatif (la part des résidents d'une commune ayant un emploi allant travailler sur Avignon).

La cartographie de cet indicateur révèle une structure spatiale de type auréolaire s'organisant autour d'Avignon. Cette structure est complexifiée dans les zones occidentale et méridionale de la périphérie par un phénomène de dissymétrie dû à une barrière administrative et hydrographique : l'agglomération d'Avignon est à cheval sur trois départements⁷ séparés les uns des autres par des axes fluviaux (Durance et Rhône).

Dans un deuxième temps, le modèle de régression linéaire bivariée est mobilisé afin d'expliquer la différenciation spatiale du phénomène "naveteurs vers Avignon" par le facteur distance à la ville centre⁸.

Les résultats obtenus sont satisfaisants puisque 69 % de la variation totale se trouvent expliqués par la plus ou moins grande proximité avec Avignon.



Néanmoins, lorsque l'on cartographie les résidus de la régression menée précédemment, on observe une organisation spatiale très diversifiée. On parvient en effet à distinguer des groupes de communes auxquels semblent correspondre des logiques explicatives précises : ainsi, dans le sud de l'agglomération avignonnaise, les résidus fortement négatifs sont-ils dus au caractère encore fortement agricole de cette région (généralisation du modèle du COMTAT au nord des Bouches-du-Rhône) ; au contraire, à l'est et à l'ouest d'Avignon, les résidus fortement positifs proviennent du caractère très résidentiel de ce type de banlieue.

On est donc amené à mobiliser plusieurs facteurs explicatifs n'intervenant que pour certaines unités spatiales de l'espace étudié afin d'expliquer cette structure spatiale résiduelle caractérisée par ses discontinuités.

Pour diminuer cette part d'information résiduelle, il est possible de recourir au modèle de régression multivariée à 4 ou 5 facteurs. Néanmoins, cette démarche modélisatrice se caractérise par une plus grande complexité ; de plus, le recours au modèle de régression multivarié n'est pas optimal puisque les nouvelles variables explicatives introduites ne fonctionnent pas pour la plus grande majorité des résidus.

Devant ces deux limites et en tenant compte du fait que l'explication de l'information résiduelle semble autant résulter de qualités que de quantités précises, il paraît intéressant de mobiliser le modèle d'analyse de variance à 1 facteur.

A titre d'exemple, le paramètre explicatif "type de commune du point de vue de l'emploi" caractérisé par quatre modalités :

- commune à forte vocation agricole (plus du tiers des actifs travaille dans le secteur primaire),

- autre pôle d'activités,
- concurrence d'un autre centre d'emploi limitrophe (part des naveteurs vers ce pôle urbain plus importante que vers Avignon),
- commune à vocation résidentielle (plus de 70 % des actifs travaillent hors de leur commune de résidence)

permet d'expliquer près de 30 % de l'information résiduelle de la régression.

Ce petit exemple démontre qu'en travaillant sur de l'information qualitative, la démarche de modélisation statistique est simplifiée tout en offrant par ailleurs de bons résultats. De plus, lors de l'explicitation de la structure spatiale, nous avons mis en évidence un emboîtement d'échelles explicatives :

- la structure auréolaire concerne l'ensemble de l'espace d'étude ; le facteur explicatif majeur de la structure spatiale du phénomène naveteurs autour d'Avignon est la plus ou moins grande proximité au centre (69 % de l'information totale est explicitée),
- le type de commune du point de vue de l'emploi intervient à l'échelle de sous-espaces (30 % de l'information résiduelle laissée par le modèle de régression linéaire bivariée est expliquée),
- à la suite du modèle d'analyse de variance, l'information résiduelle qui se maintient est réellement résiduelle : elle ne peut être expliquée que par des spécificités locales.

La précision des données quantitatives n'est donc pas toujours nécessaire. Le recours successif à des techniques d'analyse de données quantitatives et qualitatives a permis une décomposition, un filtrage de l'information initiale : la régression a fait apparaître ce qui dans l'organisation des espaces relève de structures générales tandis que le modèle d'analyse de variance a mis en valeur ce qui dépend de particularités locales.

Références bibliographiques

- BELLIOT M. (1982) : "La formation des prix fonciers : déterminants sociaux et réglementaires : le cas des Yvelines". *Etudes foncières*, n° 17, pp. 39-44
- BELLIOT M. (1984) : "Les facteurs de formation du prix des sols". *Etudes foncières*, n° 22, pp. 28-33
- DUTAILLY J.C. (1971) : «Les valeurs foncières en région parisienne. Recherche d'un modèle». *Cahiers de l'I.A.U.R.J.F.*, vol. 25, 31 p.
- GRANELLE J.J. (1965) : *La valeur du sol et ses facteurs : sources statistiques et données bibliographiques*. Paris, Centre de recherche et de documentation sur la consommation, 117 p.
- GRANELLE J.J. (1970) : *Espace urbain et prix du sol*. Paris : Edition Sirey Collection recherches économiques et financières, 292 p.
- HERMAN J. (1986) : *L'analyse des données qualitatives. Tome 1 : Traitements d'enquêtes. Echantillons, répartitions, associations*. Paris, Masson, 178 p.
- HUBERMAN A.M., B. MILES M. (1991) : *Analyse des données qualitatives*. Bruxelles, De Boeck Université, 480 p.
- JOHNSTON R.J. (1978) : *Multivariate statistical analysis in geography*. London, Longman Group Limited, 280 p.
- MAYER R. (1965) : *Prix du sol et prix du temps. Essai sur la formation des prix fonciers*. Paris, Ministère de la construction, 42 p.
- MERLIN Pierre (1973) : *Méthodes quantitatives et espace urbain*. Paris, Masson, 190 p.
- RUFFY Victor (1989) : *Structure et dynamique économiques et géographiques d'un espace foncier*. Lausanne, Thèse Université de Lausanne, 320 p.

SCHEFFE H. (1959) : *The analysis of variance*. New York, John Wiley and Sons, Inc, 477 p.

SILK J. (1979) : "Analysis of covariance and comparison of regression lines". *Concepts and Techniques in Modern Geography*, n° 20, pp. 3-45

SILK J. () : "The analysis of variance". *Concepts and Techniques in Modern Geography*, n° 30, pp. 1- 57

Annexe

Formulation mathématique

$$\sum_i \sum_j (Y_{ij} - Y)^2 = \sum_i n_i (Y_i - Y)^2 + \sum_i \sum_j (Y_{ij} - Y_i)^2$$

variation totale variation intergroupe variation résiduelle

avec j : indice variant de 1 à n_i , effectif total du ième groupe,
 i : indice variant de 1 à k, nombre de groupes,
 n_i : effectif du groupe i,
 Y : moyenne de tous les individus,
 Y_i : moyenne de tous les individus appartenant au groupe i,
 Y_{ij} : valeur du j^{ème} individu du groupe i.

Test de fiabilité des résultats

Pour juger la qualité des résultats obtenus, un test d'indépendance significative est basé sur une statistique F qui intègre les deux critères jouant sur la fiabilité du résultat : le nombre de sous-espaces et le nombre d'individus dans ces sous-espaces.

$$F = \text{variation intergroupe} / (k - 1) / \text{variation intragroupe} / (N - k)$$

avec k : nombre de groupes
 N : nombre d'individus.

Cette quantité suit une loi de Snedecor : si pour un risque d'erreur donné, elle est supérieure au F lu dans la table de probabilité, l'indépendance entre X et Y est rejetée.

Notes

¹ Puisque le coefficient de détermination est compris généralement entre 0,5 et 0,65

² Dans l'espace périphérique avignonnais, de nombreux phénomènes : migrations pendulaires, attraction commerciale, évolution démographique... ont une organisation spatiale de type gradient auréolaire

³ Elle repose donc sur la comparaison de la variation intergroupe (mesurée par les différences entre les moyennes des sous-groupes) et de la variation intragroupe (traduisant la variation de chaque individu autour de la moyenne de son groupe)

⁴ L'expression statistique de l'espace respecte la notion de réalité hétérogène de sous-espaces séparés par des discontinuités

⁵ La possibilité de modéliser des phénomènes de gradation par paliers ne porte pas atteinte à la nature des données qualitatives

⁶ L'avancée de nos travaux, après un an de recherches, n'est en effet pas suffisante pour assurer la présentation d'un exemple lié directement au marché des terrains à bâtir de l'espace vaclusien

⁷ Ces trois départements sont le Vaucluse, le Gard et les Bouches-du-Rhône

⁸ Qui dit structure spatiale en anneaux concentriques dit en effet distance à un point entraînant des phénomènes assez régulièrement discontinus