

---

# Agregación de datos espaciales para el análisis económico: nuevos métodos y aplicaciones

---

El problema de las unidades de  
área modificable (PUEM)

---

Luz Marina Contreras Saldarriaga

---

## Contenidos

Introducción.....	3
Justificación.....	4
Objetivos.....	5
Resultados.....	6
Conclusiones.....	9
Bibliografía.....	10

## Introducción

La validez de muchos estudios dependen del supuesto que las unidades espaciales utilizadas son dadas *a priori*; sin embargo, los sistemas de zona no son fijos o únicos; los mismos individuos pueden ser agrupados de diferentes formas generando así distintas representaciones de áreas (Openshaw & Taylor, 1981).

Los resultados de un análisis particular dependen de las unidades de área utilizadas en el estudio (Dudley, 1991); este efecto es conocido como el Problema de las Unidades de Área Modificable (PUEM) o en inglés como The Modifiable Areal Unit Problem (MAUP).

El PUEM tiene dos aspectos (Green & Flowerdew): el efecto de escala (scale effect) y el efecto de zona (zoning effect). El efecto de escala se refiere a las diferencias de los resultados que se obtienen cuando los datos son agrupados progresivamente en menos áreas cada vez más grandes; el efecto de zona a la variación de los resultados cuando se agrega en un mismo número de zonas pero de maneras distintas.

## Justificación

Muchos estudios utilizan datos de área agregados por razones como el costo que genera tener una cantidad grande de datos o la dificultad de tener datos para cada individuo (Prais & Aitchison, 1954). Por esta razón es esencial entender cómo se ven afectados los resultados de acuerdo a las áreas que se elijan.

La investigación pretende estudiar el Problema de las Unidades de Área Modificable con el fin de buscar nuevos métodos que se puedan aplicar a análisis económicos. Para esto se divide en dos grandes bloques:

La primera parte consiste en el diseño de un índice que mida la sensibilidad de una variable al PUEM ante cambios de agregación (asumiendo una escala fija).

La otra etapa busca crear un nuevo algoritmo de agregación espacial que incorpore dentro de su proceso iterativo información sobre el grado de autocorrelación espacial a medida que el número de áreas disminuye.

## Objetivos

Los objetivos de la primera etapa de la investigación son los siguientes:

1. Hacer una revisión de literatura centrada en estudios encaminados a evaluar la sensibilidad de los estadísticos univariados y bivariados frente a cambios de escala y agregación.
2. Hacer una revisión de literatura para conocer los efectos de la agregación de datos sobre la estimación de Mínimos Cuadrados Ordinarios (MCO) de un modelo de regresión lineal simple, considerando el caso clásico y el caso en el que existe autocorrelación espacial entre las unidades que se agregan.

## Resultados

De la revisión de literatura se encontró que la agregación (efecto de zona) afecta claramente algunos estadísticos, tales como el coeficiente de correlación, la varianza y los estimadores de los modelos de regresión lineal.

Del coeficiente de correlación se puede decir que cambia drásticamente con la agregación (Amrhein, 1993) y por esta razón no es un buen estimador de la correlación de la población (Prais & Aitchison, 1954).

Tanto para la población como para la muestra hay la varianza es ligeramente diferente bajo el efecto de zona (Amrhein, 1993); aparentemente las poblaciones con varianzas altas son más propensas a presentar efectos de zona que las poblaciones de varianzas bajas.

Los coeficientes de regresión cambian dramáticamente con la agregación (Amrhein, 1993), pero no se logró identificar un patrón claro en estos cambios.

La media resulta ser resistente a los efectos de agregación (Amrhein, 1993): no hay diferencia de tres cifras decimales entre la media de la población y la media de las muestras.

Por otro lado, se encontró que, en general, para analizar el Problema de las Unidades de Área Modificable y el efecto de agregación es indispensable tener en cuenta la autocorrelación espacial. Esto porque existen muchas diferencias entre zonas contiguas y no contiguas que se deben precisamente a este fenómeno (Openshaw & Taylor, 1979); se espera que las zonas contiguas tengan autocorrelación espacial.

La dependencia espacial juega un papel cuando se agrega por proximidad (Arbia, 1989); ésta se elimina si se agrega de manera aleatoria pues de esta forma las diferencias que se encuentran se deben a variables aleatorias asociadas con la generación de los datos (Steel & Holt, 1996).

Más específicamente, la autocorrelación espacial afecta:

- Variación: Hay menos pérdida de variabilidad en zonas contiguas que no contiguas (Openshaw & Taylor, 1979).
- Coeficiente de correlación: Si se agrega con grupos aleatorios, no hay impacto sistemático claro en el coeficiente de correlación, pero si se agrega por proximidad aumenta el coeficiente de correlación (Fotheringham & Wong, 1991).

Otro factor que es importante tener en cuenta cuando hay agregación (zoning effect) es el patrón espacial debido a que los estadísticos dependen del patrón espacial de la asociación entre individuos (Steel & Holt, 1996).

En cuanto a la revisión de literatura sobre los efectos del PUEM sobre los modelos de regresión lineal se puede decir lo siguiente:

- El coeficiente de correlación aumenta a medida que disminuye el número de áreas con que se trabaja (Fotheringham & Wong, 1991).
- Las varianzas de los estimadores agrupados nunca puede ser mayor que la de los estimadores de las observaciones originales porque al agrupar no puede haber ganancia de información (Prais & Aitchison, 1954).
- No importa de qué manera se realice la agrupación, los estimadores siempre serán insesgados (Prais & Aitchison, 1954); sin embargo, si lo que se desea es tener un método de agrupación eficiente se sugiere agrupar observaciones lo más homogéneas posibles de acuerdo a la variable independiente.
- En cuanto a los coeficientes de regresión los autores difieren pues por un lado se encontró que cambian dramáticamente con la agregación (Amrhein, 1993) y por otro lado que la agregación no genera un impacto sistemático en la pendiente del parámetro (Fotheringham & Wong, 1991).
- Las propiedades de la regresión lineal  $Y = B_0 + B_1X + \varepsilon$  cambian cuando se estima por MCO en presencia de residuos correlacionados espacialmente (Moreno Serrano & Vayá Valcarce, 2000; Anselin, 1988). Estos cambios se presentan en el Cuadro 1. De estas diferencias es claro que no es adecuado estimar por MCO (Moreno Serrano & Vayá Valcarce, 2000; Anselin, 1988).

	<b>MCO sin presencia de autocorrelación espacial</b>	<b>MCO con presencia de autocorrelación espacial</b>
Estimaciones de los parámetros	Insesgados, consistentes y óptimos	Insesgados
Predicciones	Eficientes	Ineficientes
$R^2$	Correcto	Sobrevalorado
Inferencia basada en los test de significación de la t-student	Correcta	Sesgada
Test de inestabilidad estructural	Correcto	Especificaciones erróneas
Test de heterocedasticidad	Correcto	Especificaciones erróneas

Tabla 1. Cambios en la estimación de regresión lineal simple por MCO en presencia de autocorrelación espacial



## Conclusiones

- Una forma de evaluar la sensibilidad de una variable al Problema de las Unidades de Área Modificable ante cambios de agregación (efecto de zona) es tener en cuenta el patrón espacial y el nivel de autocorrelación espacial, donde distribuciones aleatorias y bajos niveles de dependencia espacial generaran baja sensibilidad, mientras que patrones espaciales o fuertes niveles de autocorrelación generaran sensibilidad alta.
- Una forma de contribuir al estudio del PUEM es implementando simulaciones que permitan determinar de manera clara y precisa sus efectos en los coeficientes de regresión lineal, dado que en la literatura se encuentran diferentes teorías.

## Bibliografía

Amrhein, C. G. (1993). Searching for the elusive aggregation effect: evidence from statistical simulations. *Environment and Planning* , 27 (1), 105-119.

Anselin, L. (1988). *Spatial Econometrics Methods and Models*. Springer.

Arbia, G. (1989). Theoretical problems motivation. En *Spatial Data Configuration in Statistical Analysis of Regional Economic and Related Problems*. Netherlands: Kluwer Academic Publishers.

Dudley, G. (1991). Scale, aggregation, and the modifiable area unit problem. *The Operational Geographer* , 9 (3), 28-33.

Fotheringham, A. S., & Wong, D. W. (1991). The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning* , 23 (7), 1025-1044.

Green, M., & Flowerdew, R. (1996). New evidence on the modifiable areal unit problem. En P. Longley, & M. Batty, *Spatial analysis: modelling in a GIS environment*. Nueva York: Cambridge: Geoinformation International.

Moreno Serrano, R., & Vayá Valcarce, E. (2000). *Técnicas econométricas para el tratamiento de datos espaciales*. Barcelona: EDICIONS DE LA UNIVERSITAT DE BARCELONA.

Openshaw, S., & Taylor, P. J. (1979). A million or so correlation coefficients, three experiments on the modifiable areal unit problem. En *Statistical Applications in the Spatial Science*. Londres.

Openshaw, S., & Taylor, P. (1981). The modifiable areal unit problem. En *Quantitative Geography*. Londres: Routledge & Keagan Paul.

Prais, J., & Aitchison, J. (1954). The Grouping of Observations in Regression Analysis. *Revue de l'Institut International de Statistique* , 22 (1), 1-22.

Steel, D., & Holt, D. (1996). Rules for random aggregation. *Environment and Planning* , 28 (6), 957-978.