

# Una Aplicación del Análisis Exploratorio de Datos Espaciales (AEDE) con GeoDa en el Fracaso Empresarial en España

José Horacio García Marí, Eva Tomaseti Solano, Javier Sánchez Vidal  
Facultad de Ciencias de la Empresa (Universidad Politécnica de Cartagena)  
C/ Real, 3, 30201 Cartagena (Murcia)  
Email: horacio@economitas.es

**Resumen.** Una de las consecuencias de la actual crisis económica es el incremento del fracaso empresarial en España y, por consiguiente, del número de concursos de acreedores. Sin embargo, su incidencia no ha sido homogénea en todo el territorio. Para el análisis de la distribución de los mismos se ha optado por la aplicación del análisis exploratorio de datos espaciales (AEDE), que es una disciplina de análisis diseñada para el tratamiento específico de los datos espaciales o geográficos, con la aplicación GeoDa. Los resultados obtenidos muestran una mayor incidencia en las zonas con mayor participación en el PIB nacional, al mismo tiempo que se produce un efecto contagio entre el número de concursos de una provincia y sus provincias vecinas.

## 1. Introducción

En la actualidad, nos encontramos ante una profunda recesión económica, altas tasas de paro, alto índice de morosidad, fuerte endeudamiento privado y público y un sistema financiero debilitado y con escasa liquidez para la concesión de créditos. En este contexto, las empresas se ven afectadas, en primer lugar, por no tener acceso a financiación ajena y, en segundo lugar, por la debilidad de la demanda interna, lo que ha llevado a un alarmante aumento del fracaso empresarial. De entre sus distintas manifestaciones, el presente estudio se centra en el análisis de los concursos de acreedores acaecidos en 2012 a lo largo de todo el territorio nacional.

## 2. La Econometría Espacial

Para el análisis de la distribución del número de concursos de acreedores en España se van a aplicar las técnicas de análisis exploratorio de datos espaciales (AEDE). Se trata de una disciplina del análisis exploratorio de datos, también denominado *data mining*, diseñada para el tratamiento específico de los datos espaciales o geográficos que relaciona el análisis estadístico con el gráfico, haciendo posible el estudio de las distribuciones espaciales y sus valores atípicos, esquemas de asociación espacial, agrupamientos espaciales y puntos calientes/fríos de negocio (*hot spot*) [1].

Sobre la base de lo anterior, se concluye que un buen método gráfico de AEDE es aquél capaz de analizar y representar dos características fundamentales en toda distribución espacial: tendencia y puntos atípicos.

El programa a utilizar es el GeoDa, ya que cuenta con las principales técnicas de análisis exploratorio de datos según los dos citados elementos de tendencia espacial o análisis global, que son de carácter geográfico y se refieren a todo el mapa (y no a una

parte del mismo) y puntos atípicos o análisis local. Este programa se concibe como un producto autosuficiente que funciona en cualquiera de los sistemas operativos Windows y Mac. Fue desarrollado por el profesor Luc Anselin de la Universidad de Illinois y tiene la ventaja de ser, hasta el momento, un producto de libre acceso en Internet.

## 3. Comportamiento del Concurso de Acreedores Español en 2012

Dada la naturaleza del análisis estadístico a desarrollar, la muestra que se utiliza está formada por toda la población de empresas concursadas en España en 2012. Los datos se han obtenido del INE.

Para un correcto análisis de la importancia de la incidencia de los concursos de acreedores en las diferentes provincias, resulta más apropiado analizar el impacto de los mismos en términos relativos sobre el total de empresas. Así, la variable a analizar es el número de empresas concursadas en cada provincia sobre el total de empresas de esa provincia. De este modo, se salva el sesgo que se produce por la asimétrica distribución del tejido empresarial en España.

### 3.1. Análisis Exploratorio

Para estudiar el comportamiento global del concurso de acreedores en España nos centramos en dos métodos de tendencia espacial, mapa de cuartiles e histograma de frecuencias; y el diagrama de caja (*box plot*) para el estudio de los atípicos.

El mapa de cuartiles es una representación cartográfica de la variable que mediante colores pone de manifiesto el valor de la misma en cada una de las provincias españolas en intervalos. En la figura 1 se puede apreciar una mayor concentración de concursos en la diagonal superior, con una especial incidencia en las provincias del este y noroeste, en

concreto, Palencia, Vizcaya, Guipúzcoa, Álava, La Rioja, Huesca, Zaragoza, Barcelona, Castellón, Valencia, Albacete, Murcia y Baleares.

Este resultado, a diferencia de lo que se podía esperar, no indica que la mayor incidencia en el número de concursos se produzca en las provincias que contribuyen menos al PIB y que están más duramente afectadas por la crisis económica, como es el caso de la mayoría de provincias andaluzas y las extremeñas, que se encuentran en el cuarto cuartil, sino sobre las que más contribuyen.

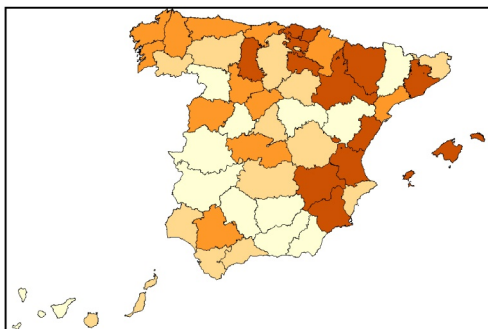


Fig. 1. Mapa Cuartiles de las Empresas Concursadas sobre el Total de Empresas en 2012

El histograma de frecuencias es una aproximación a la función de densidad de la variable que resulta de utilidad para detectar asimetrías y otras peculiaridades de la distribución. Resulta de especial utilidad ya que, en aquellos casos en los que la variable tiene una distribución muy alejada de la distribución normal, el mapa de cuartiles no resulta de utilidad [2]. En nuestro caso, la distribución de los concursos de acreedores en 2012 se encuentra normalmente distribuida a lo largo del territorio español (ver Figura 2).

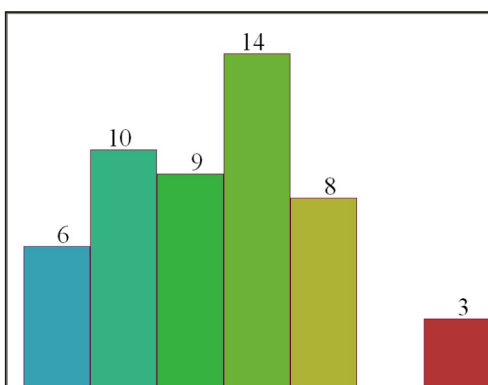


Fig. 2. Histograma de las Empresas Concursadas sobre el Total de Empresas en 2012

El diagrama de caja detecta los valores atípicos situándolos por encima o por debajo de las cotas resultantes de sumar/restar a la media el producto de los valores del tercer (primer) cuartil por 1,5 veces. La caja está formada por el valor inferior del primer cuartil hasta el valor superior del tercer cuartil, destacándose la mediana en la mitad de la caja [2]. En 2012 la no incidencia de casos atípicos evidencia

que, a pesar del distinto acaecimiento de los concursos de acreedores que evidencia la Figura 1, esta heterogeneidad se mueve dentro de unos valores razonables (ver figura 3).

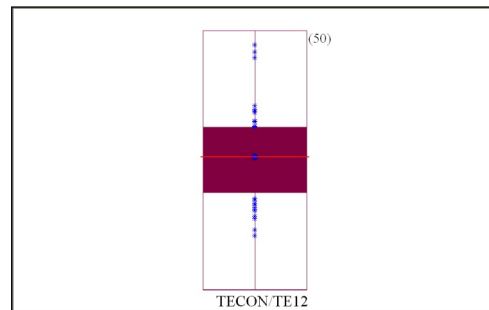


Fig. 3. Diagrama de Caja de las Empresas Concursadas sobre el Total de Empresas en 2012

### 3.2. Análisis de Autocorrelación Espacial

La autocorrelación espacial es la concentración o dispersión de los valores de una variable en un mapa. Refleja el grado en que la variable objeto de estudio de una unidad geográfica es similar a la de otras unidades geográficas próximas. Así, se puede afirmar que existe autocorrelación espacial en la medida en que existe una relación entre lo que sucede en un punto determinado del espacio y lo que acontece en otros puntos de ese mismo espacio [3].

Para verificar la existencia o ausencia de autocorrelación espacial, se pueden utilizar diferentes índices, siendo el índice I de Moran [4] el más conocido y utilizado en la práctica [3]. En el presente estudio, el I de Moran indica la intensidad de autocorrelación entre las provincias, es decir, para cada valor del estadístico es posible realizar una inferencia para evaluar el nivel de significatividad estadística de rechazo de la hipótesis nula de ausencia de similitud o disimilitud entre el número de concursos de las diferentes provincias españolas.

En el cálculo del I de Moran resulta determinante la definición de vecindad entre las unidades espaciales analizadas que se utilice, en este caso la vecindad de primer orden tipo reina.

En la Figura 4 se observa un I de Moran que es positivo y significativo al 95%, tratado a 999 permutaciones [5]. Se puede rechazar la hipótesis nula, es decir, existe una relación directa entre valores similares de la variable en una provincia y las cercanas entre sí, de forma que las provincias con valores elevados de la variable se encuentran rodeadas de otras provincias en las que la variable arroja también valores elevados, y viceversa. En este caso, en el que se produce una concentración del número de concursos en una zona geográfica concreta, se dice que existe “contagio”, lo que nos indica que el número de concursos de una zona geográfica es influido por el número de concursos de sus vecinas [3].



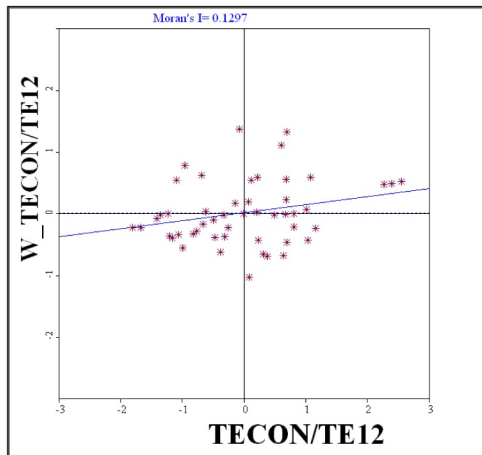


Fig. 4. Índice I de Moran de las Empresas Concursadas sobre el Total de Empresas en 2012

Sin embargo, el I de Moran que identifica el contagio a nivel global no identifica el patrón de estas relaciones espaciales, esto es, si el esquema de autocorrelación espacial detectado en todo el territorio en su conjunto se mantiene también a nivel local. Para ello, debemos acudir a los indicadores de autocorrelación local como el índice I de Moran local. En este sentido, para visualizar de forma clara y directa la presencia (o ausencia) de autocorrelación espacial local el AEDE dispone de diversas herramientas, la más utilizada es el denominado mapa LISA.

El mapa LISA localiza los agrupamientos espaciales, que se producen cuando una unidad espacial que registra un valor alto (bajo) de la variable se encuentra rodeada de unidades espaciales que también registran un valor alto (bajo) de dicha variable, y los atípicos espaciales, que son aquellos que surgen cuando una unidad espacial con un elevado valor de la variable se encuentra rodeada de unidades espaciales en las que la variable registra valores pequeños, o viceversa [3 y 6].

En la Figura 5 se muestran cuatro tipos de conglomerados espaciales a una significatividad del 95%: 1) Alto-Alto: Guipúzcoa y Navarra tienen un número de concurso de acreedores en relación con el número de empresas superior a la media nacional y, además, están rodeadas por provincias vecinas que también se encuentran sobre la media. Estas provincias corresponden a los denominados conglomerados calientes (hot spots). 2) Alto-Bajo: Salamanca, Segovia, Toledo y Sevilla presentan un número de concursos alto a pesar de estar rodeadas por provincias con valores por debajo de la media. 3) Bajo-Alto: representa a provincias cuyo número de concursos de acreedores es significativamente inferior al de sus provincias vecinas como es el caso de Burgos, Teruel, Alicante, Canarias y Baleares, en éste último caso, esto es así a pesar de pertenecer al primer cuartil. Ésto se explica por el hecho de que sus provincias vecinas Castellón y Valencia presentan los valores más altos de España junto con Álava.

Así, pertenecen a la séptima barra del histograma, mientras que Baleares se sitúa en la quinta, y son las que están más próximas a la cota superior del diagrama de caja. 4) El número de concursos del resto de provincias no tienen una diferencia significativa con el de sus provincias vecinas (ver Figura 5).

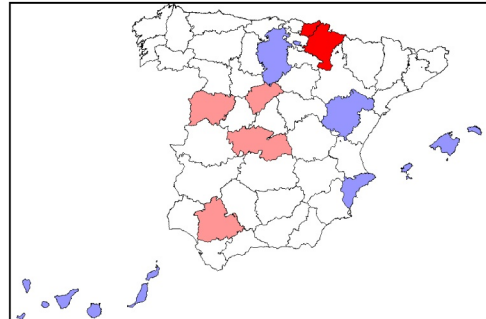


Fig. 5. Mapa LISA de las Empresas Concursadas sobre el Total de Empresas en 2012

## 4. Conclusiones

La distribución del concurso de acreedores en España tiene un comportamiento diagonal con una mayor incidencia en las zonas de mayor riqueza del país. Además, queda contrastado que, para el año 2012, existe contagio en la incidencia del número de concursos entre las distintas provincias españolas.

## Referencias

- [1] Chasco, C. (2001) Análisis Exploratorio de Datos Espaciales al Servicio del Geomarketing. III Seminario sobre Nuevas Tecnologías en la Investigación, el Marketing y la Comunicación. E-Participación.
- [2] Chasco, C. (2010) Análisis Exploratorio de Datos Espaciales (AEDE): Técnicas Básicas. Material Docente Curso Tratamiento Estadístico e Informático de Datos Georeferenciados: La Econometría Espacial. UPCT.
- [3] Sánchez, M. (2008) Análisis Espacial de Datos y Turismo: Nuevas Técnicas para el Análisis Turístico. Una Aplicación al Caso Extremeño. Revista de Estudios Empresariales, 2, 48-66.
- [4] Moran, J. (1950) Notes on continuous stochastic phenomena. Biometrika, 37, 1/2, 17-27.
- [5] Anselin, L. (2005) Exploring Espatial Data with GeoDa: A Workbook.
- [6] Naciones Unidas (2012) Guía para estimar la pobreza infantil.