

Análisis estadístico del potencial empresarial de Andalucíaⁱ

Statistical Analysis of the Business Activity in Andalusia

Análise estatística do potencial de negócios da Andalusia

Manuel Ceballos González¹ y María Infante Bermejo²

Autores

¹Universidad Loyola Andalucía. Av. de las Universidades, 41704 Dos Hermanas, Sevilla (Spain). mceballos@uloyola.es

²Universidad Loyola Andalucía. Av. de las Universidades, s/n, 41704 Dos Hermanas, Sevilla (Spain). minfantebermejo@al.uloyola.es

Corresponding author: Manuel Ceballos González. Universidad Loyola Andalucía. mceballos@uloyola.es

Copyright: © 2022 Revista Dimensión Empresarial / Vol. 20 No. 4 (2022) / e-ISSN: 2322-956X

Tipo de artículo: Artículo de investigación / **Recibido:** 2/11/2022 **Aceptado:** 19/12/2022

JEL Classification: C10; C14; L10

Cómo citar:

Ceballos González, M. & Infante Bermejo, M. (2022). Análisis estadístico del potencial empresarial de Andalucía. *Revista Dimensión Empresarial*, 20(4), 74-97 DOI: 10.15665/dem.v20i4.3109

Resumen

La industria y la actividad empresarial en Andalucía han tenido una gran importancia a lo largo de la historia. Hoy día, Andalucía es la tercera comunidad autónoma de España que mayor crecimiento empresarial ha experimentado durante el último año. El objetivo del presente artículo es analizar el potencial empresarial de los distintos municipios de Andalucía durante el año 2020. Para ello, se ha llevado a cabo un estudio estadístico haciendo uso del programa Stata y una base de datos del Sistema de Información Multiterritorial de Andalucía (SIMA). Los resultados obtenidos permiten concluir que en Andalucía existen diferencias estadísticamente significativas con respecto al consumo tanto a nivel municipal como provincial. Esto da lugar a la presencia de datos atípicos. Tras varios test se ha concluido

que los datos de la variable potencial empresarial se ajustan a una distribución normal y que existe una correlación lineal fuerte con otras tres variables de la base de datos de SIMA.

Palabras clave: Análisis estadístico, Base de datos, Andalucía, Actividad empresarial.

Abstract

Industry and business activity in Andalusia have been very important through history. Nowadays, Andalusia is the third autonomous community in Spain that has experienced the highest business growth in the last year. The main goal of this paper is to analyse the business activity in every city and town of Andalusia during the year 2020. In order to do so, we have developed a statistic analysis by using Stata software and the database of the Mutiterritorial Information

ⁱ Artículo de investigación desarrollado en la Universidad Loyola Andalucía.

System of Andalusia. The results obtained allow us to conclude that in Andalusia there are statistically significant differences with respect to consumption both at the municipal and provincial level. This yields to the existence of atypical and highly atypical data. After various tests, it has been concluded that the data for the business potential variable fits a normal behaviour and that there exists a strong correlation between this variable and other three variables from the database.

Keywords: Statistical study, Database, Andalusia, Business activity.

Resumo

A indústria e a actividade empresarial na Andaluzia tiveram uma grande importância ao longo da história. Hoje, a Andaluzia é a terceira comunidade autónoma de Espanha que

registou o maior crescimento empresarial no último ano. O objetivo deste artigo é analisar o potencial de negócios dos diferentes municípios da Andaluzia durante o ano de 2020. Para isso, foi realizado um estudo estatístico usando o programa Stata e um banco de dados do Sistema de Informação Multiterritorial da Andaluzia. Os resultados obtidos permitem concluir que na Andaluzia existem diferenças estatisticamente significativas no consumo tanto a nível municipal como provincial. Isso resulta na presença de outliers. Após vários testes, concluiu-se que os dados da variável potencial de negócio obedecem a uma distribuição normal e que existe uma forte correlação linear com outras três variáveis da base de dados.

Palavras chave: Estudo estatístico, Base de dados, Andaluzia, Potencial de negócios.

1. Introducción

Andalucía es una de las comunidades autónomas más importantes de España. Destaca por su notable potencial empresarial, una sólida economía y una ubicación estratégica que la convierte en un lugar atractivo para el establecimiento y crecimiento de empresas. El potencial empresarial de Andalucía se ve respaldado por su amplia infraestructura de transporte, que facilita la conexión con otros mercados nacionales e internacionales. Esta accesibilidad fomenta el comercio y la exportación, creando oportunidades para empresas de diferentes sectores. Además, cuenta con una mano de obra altamente cualificada y diversificada. Por todo ello, ha sido la tercera co-

munidad que mayor crecimiento empresarial ha experimentado durante el año 2021 (IECA, 2021). Más concretamente, se crearon un total de 15.000 empresas nuevas. La que más destaca es el Grupo Bidafarma cuyo principal cometido es suministrar medicamentos a más de 12.500 farmacias al día.

No obstante, para comprender el potencial empresarial de Andalucía y tomar decisiones, resulta fundamental realizar un estudio estadístico. Este tipo de análisis permite conocer aspectos relevantes como la estructura económica, los sectores emergentes, las tendencias de mercado, etc. Además, permite a los emprendedores, inversores y responsables políticos tomar decisiones estratégicas

basadas en evidencia con el objetivo de aprovechar al máximo las oportunidades empresariales en la región.

En el siguiente artículo se pretende determinar, analizar y justificar a través de un estudio estadístico las características de las actividades empresariales en todos los municipios de Andalucía. Para ello, se hará uso del software Stata y de una base de datos recogida en la plataforma del Sistema de Información Multiterritorial de Andalucía (SIMA) de la web del Instituto de Estadística y Cartografía de dicha comunidad. El programa Stata es un software estadístico cuya primera versión data de 1985. Ofrece a los usuarios analizar, administrar y generar visualizaciones gráficas de los datos. El Instituto de Estadística y Cartografía de Andalucía está formado por varios organismos como la Comisión Andaluza de Estadística y Cartografía y la Comisión Técnica de Estadística (Álvaro et al., 2013). La plataforma SIMA se encuentra en la web de dicho instituto y es un banco de datos con información multitemática y multiterritorial de todos los municipios de Andalucía (SIMA, 2022). Entre estos datos podemos encontrar el entorno físico y medio ambiente, demografía y población, sociedad, economía, mercado del trabajo y hacienda.

El principal objetivo del presente artículo es analizar el potencial empresarial de los distintos municipios de Andalucía durante el año 2020. Para ello, se ha llevado a cabo un análisis estadístico de la variable potencial empresarial incluyen-

do estadísticos descriptivos, diagramas y estudio de los datos atípicos y extremadamente atípicos. El segundo objetivo es analizar el porcentaje de empresas por actividad económica. Un tercer objetivo es la clasificación del tipo de potencial empresarial (bajo, medio-bajo, medio-alto o alto) en cada municipio mediante un criterio basado en los cuartiles. El cuarto sería un análisis comparativo de medias de potencial empresarial provincial estableciendo aquellas parejas de provincias entre las que existen diferencias estadísticamente significativas. El quinto corresponde a la realización de un test de normalidad. El sexto y último objetivo es el desarrollo de un análisis bivalente con el resto de las variables de la base de datos de SIMA.

La estructura general de este artículo es la siguiente: tras esta breve introducción, la segunda sección tratará el marco teórico y antecedentes. En la tercera sección mostramos la metodología utilizada. A continuación, en la cuarta sección abordamos un estudio estadístico sobre la variable potencial empresarial con un sumario de estadísticos descriptivos y varios diagramas de representación: histograma, tallo-hojas y cajas-patas. También estudiamos la distribución del número de empresas por actividad económica. Además, analizamos el tipo de potencial empresarial en cada provincia andaluza mediante una clasificación basada en los cuartiles. Después, analizamos los datos atípicos y extremadamente atípicos a nivel autonómico y provincial. Para con-

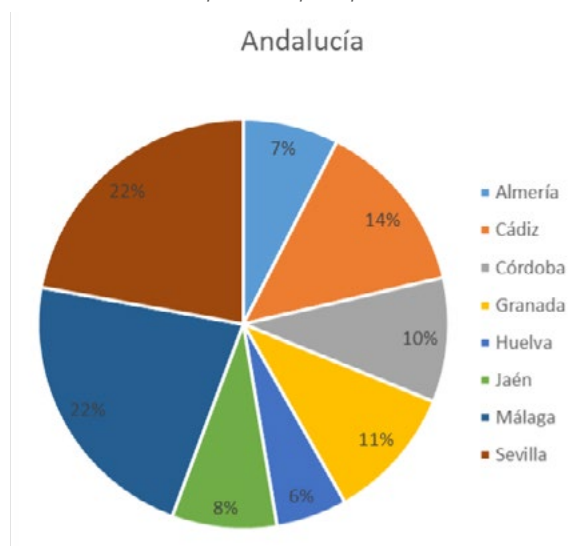
tinuar, realizamos un test de hipótesis comparando las medias provinciales para estudiar si existen diferencias estadísticamente significativas entre ellas, un test de normalidad y un análisis bivariante con el resto de variables de la base de datos SIMA. La quinta sección expone una discusión sobre los resultados obtenidos y limitaciones. Por último, una sección de conclusiones y futuras investigaciones y la bibliografía.

2. Marco teórico

Los primeros registros de empresas en Andalucía datan del año 1990. Según datos de la Seguridad Social (Morán Álvarez and Guzman Cuevas, 1991), Andalucía contaba con un total de 126.314 empresas. Actualmente triplica dicho número. La Figura 1 muestra la distribución de empresas por provincias de la comunidad autónoma en 1990.

Figura 1.

Número de empresas por provincia en 1990



Fuente: Elaboración propia

La mayor concentración de las empresas se encuentra en las provincias de Málaga y Sevilla. Sin embargo, en las provincias de Almería y Huelva, la actividad empresarial no superaba el 10%. En el año 2000, en Almería destacaban industria y servicios. En Cádiz y Córdoba destacaban el comercio y la reparación. En Granada, eran el comercio, reparación y hostelería. En Huelva, Jaén, Málaga y Sevilla, había 23.836, 33.113, 79.788 y 92.499 empresas, respectivamente (Demografía empresarial en Andalucía, 1996), destacando el comercio, hostelería, transporte y comunicaciones. En la Figura 2 se muestra la distribución de la actividad empresarial de 1996 al año 2000. El sector servicios es el que más evoluciona, mientras que la industria se mantiene casi invariante y la construcción crece ligeramente año a año.

En el período 2000-2013 las empresas españolas sufrieron la crisis económica. Hubo muchos estudios sobre el impacto de dicha crisis (Yserte et al., 2016). Años después y tras la COVID-19, los informes del Instituto de Estadística y Cartografía de Andalucía (IECA) sitúan actualmente a Andalucía como la tercera en el ranking de las comunidades autónomas con mayores crecimientos empresariales con un total de 15.000 nuevas empresas. Las primeras serían Madrid y Cataluña (Martínez, 2022). La empresa que más destaca en Andalucía es el Grupo Bidafarma, que en el año 2020 facturó 2.536 millones de euros (Espallargas, 2022). Destacar algunos estudios recientes como (Bellido-Ji-

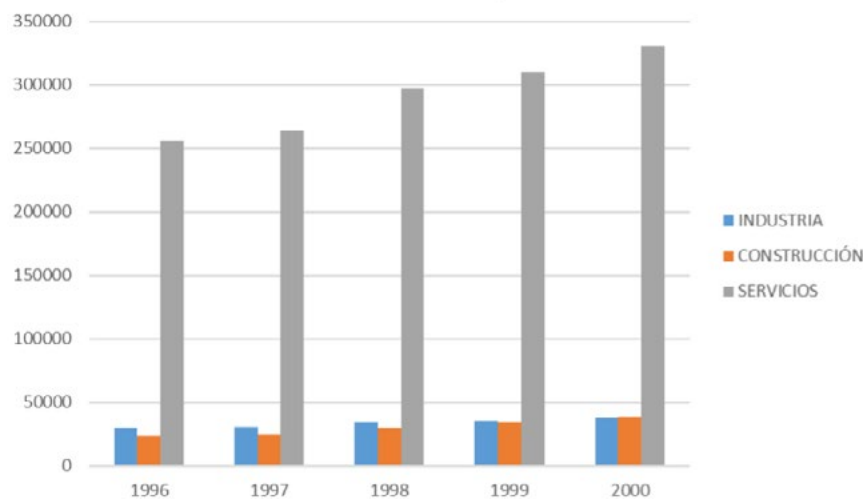
ménez et al., 2021), (CEA, 2020), (González, 2020) y (Andalucía en cifras de IECA, 2021) en los que se analiza la supervivencia de las empresas en Andalucía. Por último, señalar un reciente estudio que aborda el potencial empresarial de An-

dalucía desde una perspectiva de innovación, proporcionando información sobre los sectores clave y las oportunidades de negocio (Consejería de Economía, Conocimiento, Empresas y Universidad de la Junta de Andalucía, 2020).

Figura 2.

Actividad empresarial en Andalucía en 1996-2000

Distribución de la actividad empresarial en Andalucía



Fuente: Elaboración propia

Tras esta breve introducción histórica sobre la evolución empresarial de Andalucía, procedemos a mencionar la teoría de las Cadenas Globales de Valor (GVC en inglés), que busca comprender la forma en que las empresas y los países participan en la producción y distribución de bienes y servicios a nivel mundial. Esta teoría se centra en analizar cómo se organiza y coordina la cadena de actividades económicas a lo largo de diferentes etapas de producción, diseño, fabricación, comercialización y entrega final al consumidor (Kano, 2020).

Se han llevado a cabo estudios sobre la participación de Andalucía en las cadenas globales de valor dando una visión general de los sectores y actividades económicas en los que esta comunidad tiene una mayor presencia (IECA, 2017). También se ha analizado su relación con la competitividad y el desarrollo económico regional (Consejería de Economía, Conocimiento, Empresas y Universidad de la Junta de Andalucía, 2018).

Un estudio estadístico sobre el potencial empresarial de Andalucía puede ser de gran interés para la teoría de las Cade-

nas Globales de Valor pues permite identificar sectores y actividades económicas en los que Andalucía tiene una mayor presencia y participación. Esto ayudaría a comprender en qué áreas se encuentra el potencial empresarial y las oportunidades para el crecimiento económico en la región. Además, podría ayudar a revelar las interconexiones y los vínculos existentes entre las empresas andaluzas y las empresas a nivel nacional e internacional en las cadenas de valor. Esto proporcionaría información valiosa sobre las relaciones comerciales, la transferencia de conocimientos y la colaboración entre empresas, lo cual es crucial para mejorar la competitividad y la capacidad de innovación y aprovechar las oportunidades de crecimiento económico.

3. Metodología

En esta sección expondremos la metodología utilizada. Primeramente, veremos una introducción al análisis de datos. Después, analizaremos la fuente de datos utilizada y su tratamiento. Por último, comentaremos los métodos, técnicas y procedimientos aplicados.

3.1 Introducción al análisis de datos

Para llevar a cabo el análisis, se ha hecho uso de los datos proporcionados por la plataforma SIMA, que se encuentran recogido en la web del Instituto de Estadística y Cartografía de Andalucía. Los datos recogidos en dicha plataforma son muy diversos y analizan características de cada municipio andaluz como pueden

ser el entorno físico, medioambiental, demografía, población, sociedad, economía, etc. Entre todos ellos, se encuentra el número total de empresas por municipio en el último año. Los datos de SIMA fueron exportados a una hoja de cálculo Excel para su futura importación al programa Stata, que es el software estadístico que hemos utilizado para tratar con los datos y llevar a cabo el análisis del potencial empresarial en cada municipio de Andalucía.

El programa Stata ha tenido tres grandes precursores: BMDP (Biomedical Program), el SAS (Statistical Analysis System) y el SPSS (Statistical Programs for the Social Sciences). Tres son también las características más destacadas que han logrado que Stata sea una de las aplicaciones más usadas en estadística. En primer lugar, el empleo de instrucciones con un fácil lenguaje de modo interactivo. En segundo lugar, Stata está especializado en el análisis de regresiones ofreciendo una amplia gama de procedimientos que van desde la regresión simple hasta los modelos de ecuaciones estructurales. Por último, este software no solo puede empaquetar todas las instrucciones en un fichero, sino que también es posible para un técnico experto la elaboración de nuevas utilidades diferentes a las ya implementadas (Universidad Loyola, 2021). Stata destaca en análisis de muestras complejas, series temporales, análisis de datos históricos, datos de panel, gráficos de contornos y marginales y el análisis de modelos de ecuaciones estructurales (Bernardi, F. y Escobar Mercado, 2012).

3.2 fuentes y tratamiento de los datos

Para el análisis de datos, hemos utilizado los datos proporcionados por la plataforma SIMA y los hemos exportado a formato Excel e importado al editor de variables del programa Stata. A continuación, hemos seguido los pasos que se pueden encontrar en (Universidad Loyola, 2021) para realizar el análisis estadístico y obtener todos los resultados que se presentan en la cuarta sección. Todo ello referido a la variable potencial empresarial que se define como el cociente entre el número de empresas de un municipio y el número de habitantes del mismo. Para mejorar la calidad del análisis, se ha multiplicado el resultado por 100 para trabajar en porcentajes y evitar perder cifras decimales significativas en el análisis. Para esta variable, se han calculado en la Sección 4.1 los estadísticos descriptivos más importantes y diversas representaciones gráficas.

Después, en la Sección 4.3, se lleva a cabo una clasificación de la tipología de potencial empresarial en cada municipio dependiendo de los valores de los cuartiles a nivel autonómico. Los valores de los cuartiles obtenidos con Stata fueron los siguientes:

$$Q1= 4.025045, Q2= 4.934255 \\ \text{y } Q3= 5.852754$$

Esto nos permite obtener una clasificación de los datos. Así, un municipio tiene un potencial empresarial Bajo si éste se encuentra por debajo del 25% (valores menores que Q1). Un municipio tendrá un

potencial Medio-Bajo si se encuentra entre el 25 y 50% (valores mayores que Q1, pero menores Q2). Análogamente el potencial será Medio-Alto o Alto si éste se sitúa entre el 50 y 75% (valores mayores que Q2 pero menores que Q3) o bien está por encima del 75 % (valores mayores que Q3), respectivamente. A partir del valor de los cuartiles, se obtuvo el valor del rango intercuartílico:

$$RIC = Q3 - Q1= 1.82771$$

Esto se utiliza en la Sección 4.4 para estudiar aquellas observaciones cuyos valores difieren del resto de los datos. Así, los datos atípicos vienen determinados por aquellas observaciones que quedan por debajo de $Q1 - 1,5 \times RIC$ o bien por encima de $Q3 + 1,5 \times RIC$. Los datos extremadamente atípicos están formados por aquellos valores que se encuentran por debajo de $Q1 - 3 \times RIC$ o bien por encima de $Q3 + 3 \times RIC$. Además de obtener estos valores, se estudió cómo se distribuyen entre las distintas provincias andaluzas.

En la Sección 4.5, se aborda un estudio de comparación de medias provinciales. Para ello hemos usado la media, desviación típica y el número de municipios de cada provincia y hemos aplicado el test de la distribución T de Student para comparar medias y ver si existen diferencias estadísticamente significativas entre las provincias. Este tipo de distribución se utiliza para estimar el valor de la media poblacional en una distribución normal. En la Sección 4.6 veremos cómo efectivamente nuestra variable potencial

empresarial sigue una distribución normal. El test mencionado se aplica cuando el tamaño muestral es demasiado pequeño como para que el estadístico en el que está basada la inferencia esté normalmente distribuido. En cuanto a la formulación del test, comienza considerando dos provincias A y B y realiza el contraste de hipótesis

$$H_0: \mu_A = \mu_B, \quad H_A: \mu_A \neq \mu_B$$

El estadístico del contraste viene dado

$$t = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}}$$

donde \bar{X}, S^2, n son la media, desviación típica y número de municipios para cada provincia. Tal y como podemos observar, se define como la diferencia de las medias dividido entre la raíz cuadrada de la suma de los cuadrados de las desviaciones entre el tamaño de su muestra correspondiente.

La Sección 4.6 tiene como objetivo realizar el test de normalidad de Kolmogorov-Smirnov para contrastar si la variable potencial empresarial se ajusta o no a una distribución normal. Estos tests están basados en un contraste de hipótesis y se realizan mediante el programa Stata con el comando `ksmirnov`.

Por último, en la sección 4.7 se lleva a cabo un análisis bivalente comparando el potencial empresarial con el resto de variables de la base de datos de la plataforma SIMA. Mediante Stata se ha generado un diagrama de dispersión y una tabla de correlación. En dicha sección

mostramos los resultados para las tres variables con las que se ha obtenido una correlación fuerte.

4. Resultados

Los resultados se presentarán en siete apartados distintos acorde a los objetivos que se indicaron en la Introducción. Trataremos con la variable potencial empresarial que viene definida como el cociente entre el número total de empresas de un municipio y el número de habitantes del mismo. Desarrollaremos un análisis estadístico utilizando el programa Stata. En primer lugar, se realizará un estudio calculando los estadísticos descriptivos y diagramas representativos más importantes. En segundo lugar, se estudiará la distribución del número de empresas por actividad económica. Después, analizaremos el tipo de potencial empresarial de cada municipio acorde a los criterios que se indicaron en la sección anterior. A continuación, se estudiarán los datos atípicos y extremadamente atípicos a partir de los estadísticos calculados en los dos primeros apartados. En el cuarto apartado, se lleva a cabo una comparación de medias provinciales con respecto al potencial empresarial aplicando el test de la distribución T de Student. Después, aplicaremos los Test de Kolmogorov-Smirnov y de Shapiro-Wilk para contrastar si los datos de nuestra variable se ajustan o no a una distribución normal. Por último, en el sexto apartado se realizará un estudio bivalente analizando si existe correlación entre nuestra variable y las demás

recogidas en la base de datos de la plataforma SIMA.

4.1. Estadísticos descriptivos y diagramas

Se ha obtenido un resumen de estadísticos descriptivos para la variable potencial empresarial (ver Tabla 1). Para ello, se han analizado un total de 794 municipios andaluces, obteniendo una media de potencial empresarial de 4.942655 y una desviación típica de 1.587743. También se ha obtenido el valor de los percentiles y de varios coeficientes como el de asimetría y el de curtosis. Destacar que la mediana es de 4.934255, muy parecida a la media. Por lo tanto, la distribución será simétrica. El mayor potencial empresarial per cápita registrado es de 14.0940, dato que corresponde al municipio de Bubión (Granada). Esto es debido a que este municipio tiene tan solo 298 habitantes, una extensión bastante grande y muchas empresas con sede en tal municipio.

Después, hemos obtenido los diagramas o representaciones gráficas más im-

portantes para dicha variable. Estos son el histograma, el diagrama de tallos y hojas y el de cajas y patas (ver Figuras 3-5). En los dos primeros se puede observar la simetría comentada anteriormente y cómo se distribuye la variable potencial empresarial, donde la mayoría de los valores se concentran en torno a la media.

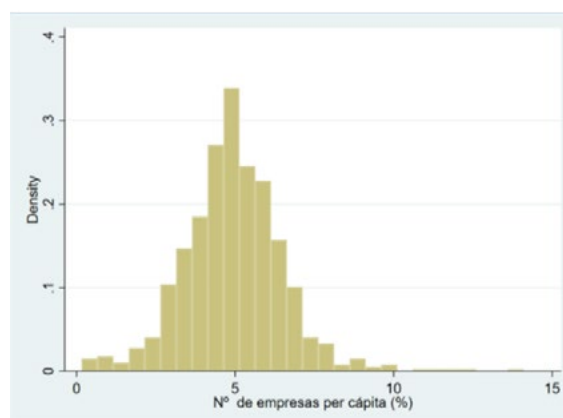
En el diagrama de cajas y patas (Figura 5), se ha señalado los números 1, 2 y 3, que corresponden al primer, segundo y tercer cuartil (Q_1 , Q_2 y Q_3), respectivamente. Las medidas de la caja están definidas por la medida del rango intercuartílico. Los números 4 y 5 corresponden a los bigotes y determinan el límite de la aparición de datos atípicos. Se observa la existencia de datos atípicos y extremadamente atípicos por encima tal y como veremos en la sección 4.4. Las barreras de la izquierda y derecha (puntos 4 y 5) se calcula, respectivamente como $Q_1 - (1.5 \times RIC) = 4.025045 - (1.5 \times 1.82771) = 1.28348$. $Q_3 + (1.5 \times RIC) = 5.852754 + (1.5 \times 1.82771) = 8.594319$.

Tabla 1.
Estadísticos descriptivos

. summarize, detail				
Nº de empresas per cápita (%)				
	Percentiles	Smallest		
1%	.8333333	.1605136		
5%	2.460457	.3192338		
10%	3.078556	.3414426	Obs	794
25%	4.025045	.3416856	Sum of Wgt.	794
50%	4.934255		Mean	4.942655
75%	5.852754	Largest	Std. Dev.	1.587743
		11.40271		
90%	6.71277	12.07815	Variance	2.520927
95%	7.37611	12.24726	Skewness	.5213226
99%	9.62963	14.09396	Kurtosis	5.978265

Fuente: Elaboración propia en Stata

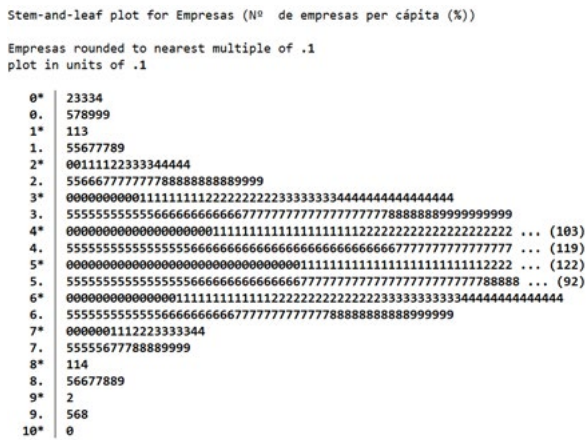
Figura 3.
Histograma



Fuente: Elaboración propia en Stata

Figura 4.

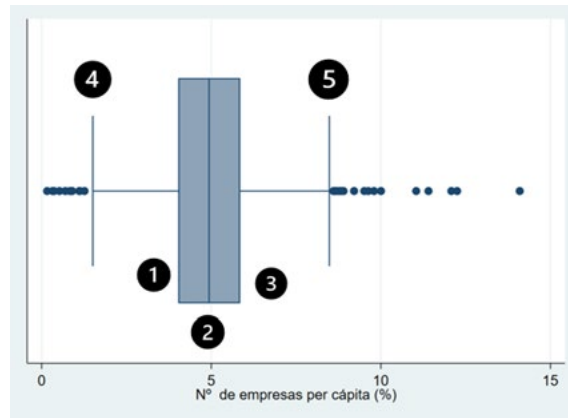
Diagrama de tallos y hojas



Fuente: Elaboración propia en Stata

Figura 4.

Diagrama de cajas y patas



Fuente: Elaboración propia en Stata

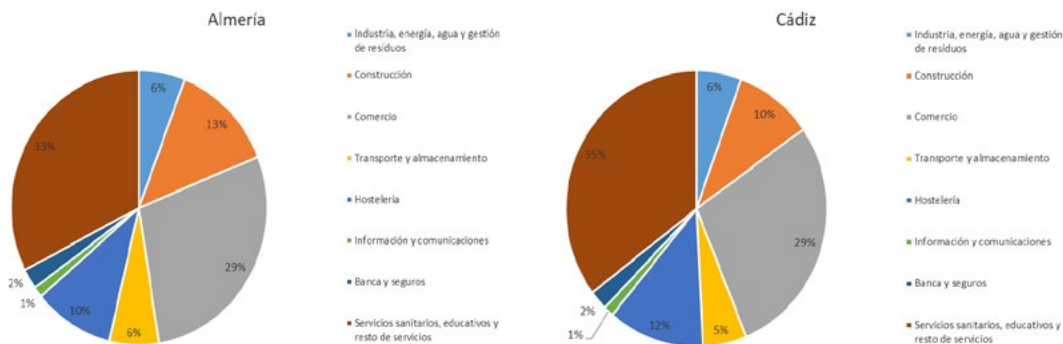
4.2. Número de empresas por actividad económica

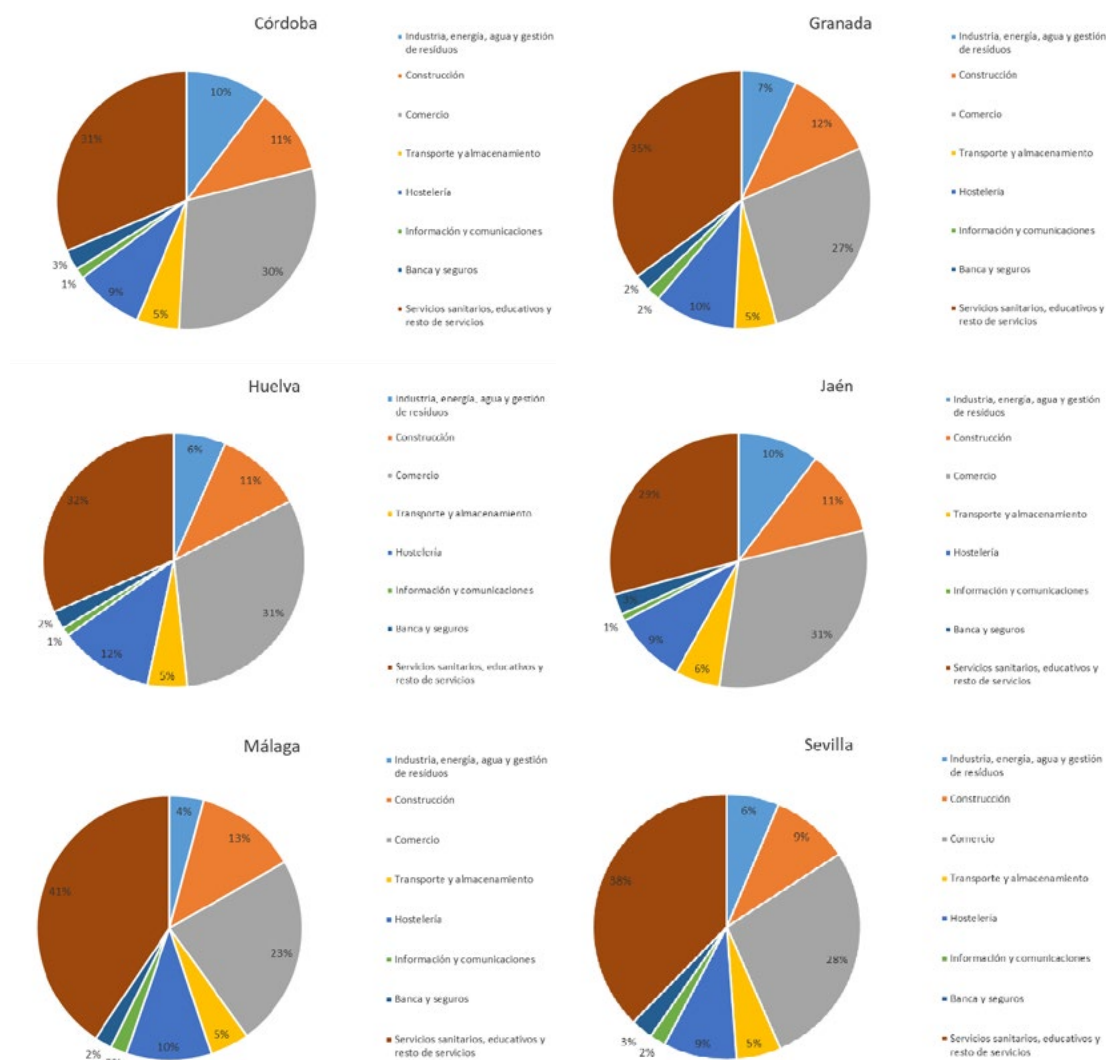
En esta subsección se muestra la distribución del número de empresas por actividad económica del año 2020 según la Clasificación Nacional de Actividades Económicas (CNAE 09) de Andalucía (INE, 2009). Así, las gráficas de la Figura 6 han sido elaboradas con los datos obtenidos del Instituto de Estadística y Cartografía de Andalucía. Se observa como el sector servicios destaca en todas las

provincias a excepción de Jaén, donde la actividad económica más importante corresponde al comercio. Señalar además que este sector es el segundo más predominante en el resto de provincias. La construcción es un sector que las provincias donde más se desarrolla es en Almería y Málaga. Los demás sectores tienen un desempeño similar en todas las provincias.

Figura 6.

Número de empresas por actividad económica en cada provincia de Andalucía





Fuente: Elaboración propia

4.3. Tipos de potencial empresarial

En este apartado analizaremos el tipo de potencial empresarial de cada municipio andaluz. Así, mostramos un diagrama sectorial (Figura 7) con la distribución de la tipología de potencial empresarial: bajo, medio-bajo, medio-alto o alto. Se observa como el de tipo alto es predominante en la provincia de Málaga, mientras que en Almería, Granada y Jaén predomina el tipo bajo. En las provincias de Cádiz y Córdoba, sin embargo, destaca el

potencial medio-alto y en las de Huelva y Sevilla el medio-bajo.

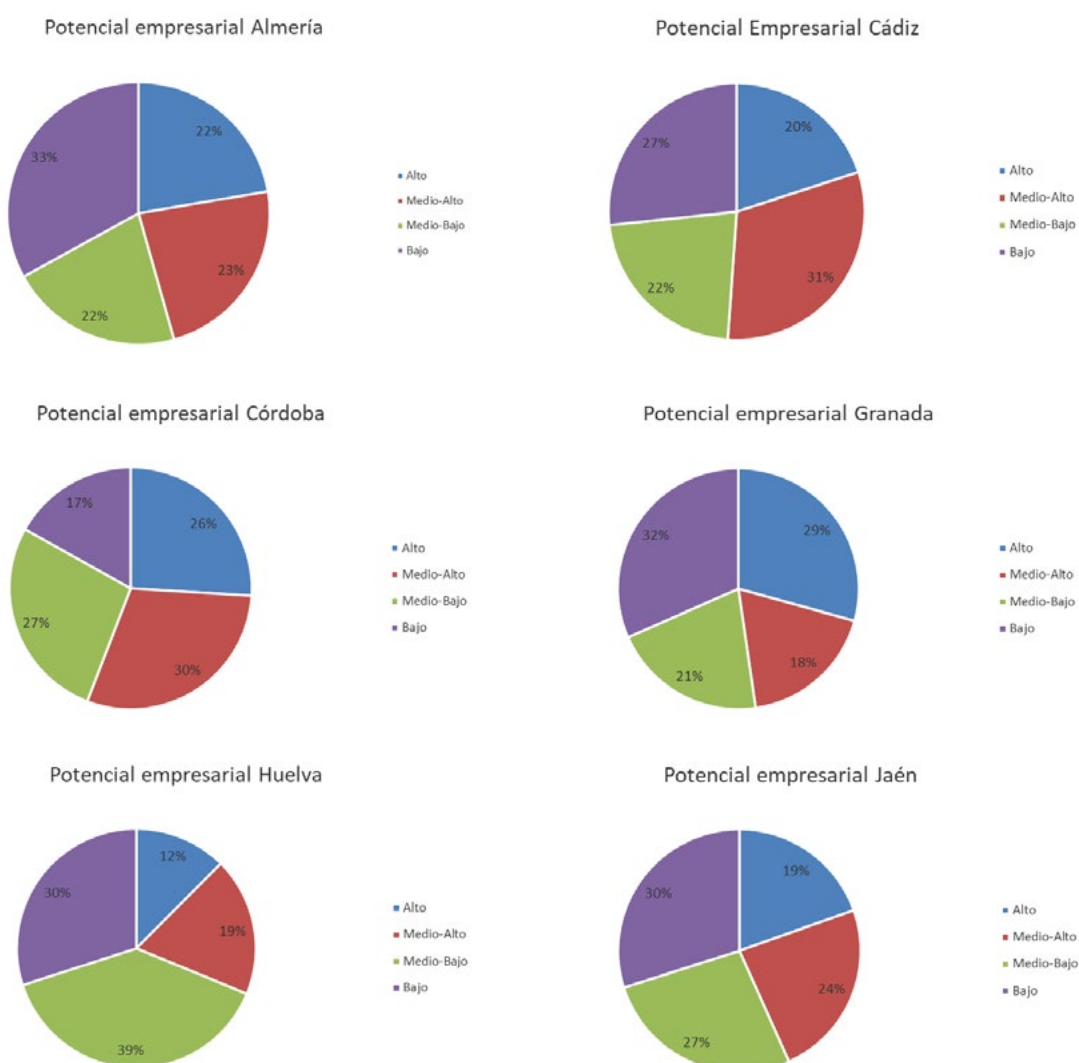
Después, en la Figura 8, se muestra cómo se distribuye el tipo de potencial empresarial en las ocho provincias. Estos gráficos han sido normalizados en proporción al número de municipios que posee cada provincia. Se observa que Málaga es la provincia que cuenta con mayor número de municipios de potencial empresarial alto. Muy seguida estaría

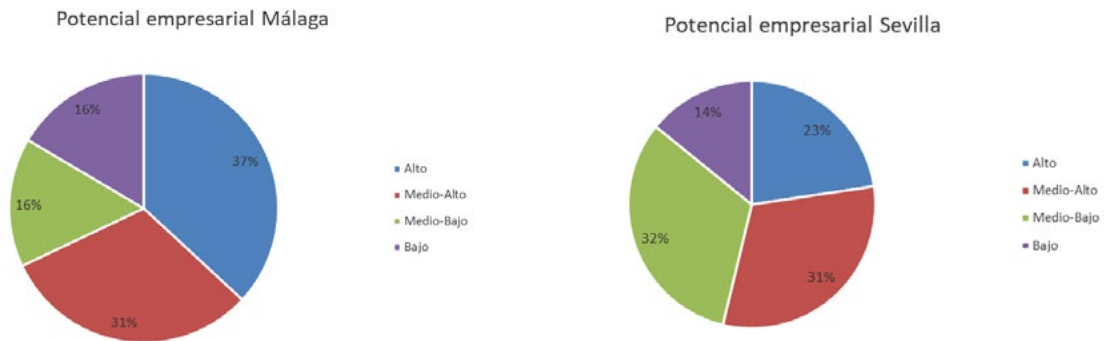
Granada y, por último, estaría Huelva. Las provincias con mayor número de empresas de potencial empresarial Medio-Alto serían Cádiz, Málaga y Sevilla y las que menos serían Huelva y Granada. La provincia con mayor número de municipios con Potencial empresarial Medio-Bajo es

Huelva. Las provincias con mayor número de municipios de Potencial empresarial Bajo son Almería y Granada. Por último, mostramos un mapa en la Figura 9 que recoge la distribución del tipo de potencial empresarial en toda Andalucía.

Figura 7.

Potencial empresarial en cada provincia de Andalucía.

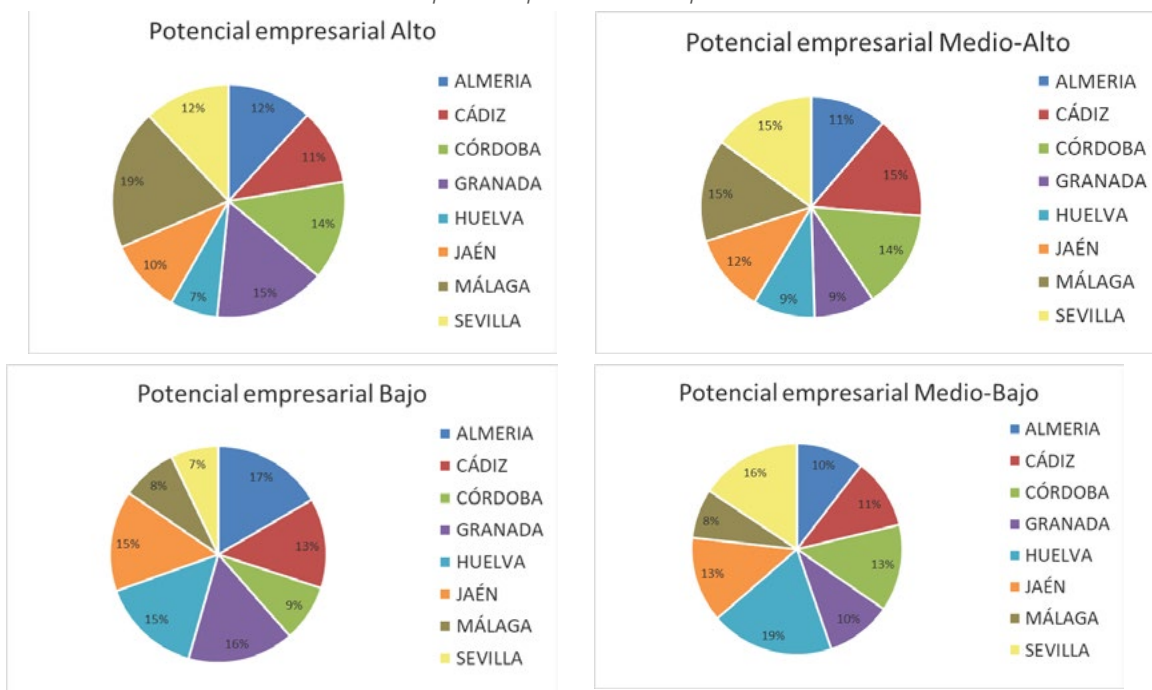




Fuente: Elaboración propia

Figura 8.

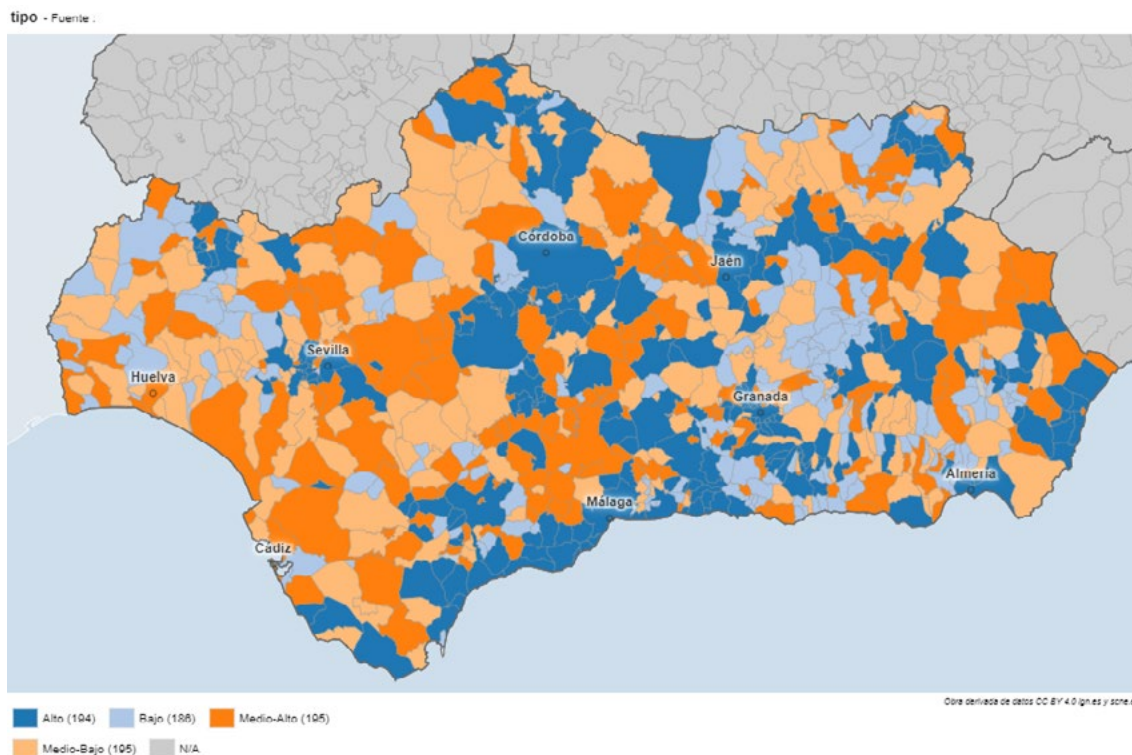
Distribución de los tipos de potencial empresarial en Andalucía



Fuente: Elaboración propia

Figura 9.

Mapa de la distribución del potencial empresarial en Andalucía



Fuente: Elaboración propia

4.4 Datos atípicos y extremadamente atípicos

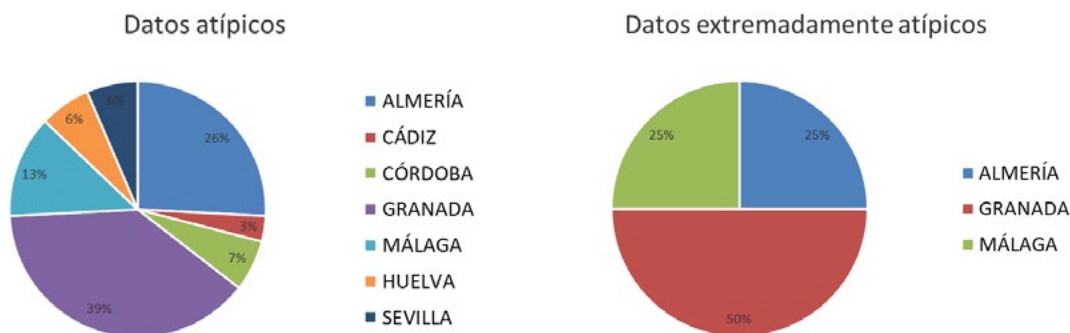
En el diagrama de cajas y patas que se expuso en la sección 4.1 se aprecia como existen datos atípicos y extremadamente atípicos en nuestra distribución. En la Figura 10 mostramos la distribución de datos atípicos y extremadamente atípicos por provincias. Se observa que la provincia que más datos atípicos contiene es Granada, mientras que Cádiz es la que alberga menos datos de este tipo. Esto concuerda con el hecho de que sea un municipio de Granada el de mayor potencial empresarial, como se vio en la Sección 4.1.

Municipios correspondientes a datos atípicos: Beires, Cóbdar, Olula de Castro, Antas, Fines, Láujar de Andarax, Mojácar, Viator (Almería); San Martín del Tesorillo (Cádiz); Fuente Carreteros, La Guijarrosa (Córdoba), Dehesas Viejas, Domingo Pérez de Granada, Fornes, Játar, Torrenueva Costa, Bubión, Busquístar, Capileira, Juviles, Pampaneira, Peligros, Trevélez (Granada); Cañaverall de León, La Zarza-Perrunal (Huelva); Benahavís, Frigiliana, Marbella, Nerja (Málaga); El Palmar de Troya, Salteras (Sevilla)

Municipios correspondientes a datos extremadamente atípicos: Antas (Almería); Bubión, Capileira (Granada); Marbella (Málaga)

Figura 10

Distribución de los datos atípicos por provincias



Fuente: Elaboración propia

4.5 Comparación de medias

En esta sección llevamos a cabo una comparación de medias provinciales aplicando el test de la distribución T de Student. Para ello, consideramos la Tabla 2 donde se han calculado la media y la desviación típica del potencial empresarial en cada provincia. Se indica además el número de municipios de cada una. Tal y como puede observarse a partir de los datos de dicha tabla, la provincia de Málaga es la que tiene mayor media con respecto al potencial empresarial. Le sigue la provincia de Sevilla. Estos datos concuerdan con el hecho de que estas son las dos provincias que tienen más municipios que corresponden a casos atípicos y extremadamente atípicos.

Para analizar más en profundidad en este estudio comparando las medias provinciales, hemos estudiado si existen diferencias estadísticamente significativas (con un nivel de significación del 95%) entre las distintas parejas de provincias andaluzas. Los resultados obtenidos son que las únicas parejas entre las que exis-

ten tales diferencias son: Almería y Málaga, Cádiz y Málaga, Córdoba y Málaga, Granada y Málaga, Huelva y Sevilla, Huelva y Málaga, Jaen y Málaga, Jaen y Sevilla.

En las Tablas 3-10 se puede observar que se obtiene un p-valor menor que 0.05. Por tanto, se acepta la hipótesis nula y existen diferencias significativas entre las provincias indicadas en cada caso. Los valores que se obtienen en dichas tablas son, respectivamente 0.0016, 0.0043, 0.0060, 0.0023, 0.0010, 0, 0, 0.0087

Tabla 2.

Media, desviación estándar y nº de municipios para la variable potencial empresarial per capita.

Provincia	Media	Desv. estándar	Nº municipios
Almería	4,7946	1,96049164	103
Cádiz	4,8282018	1,2153386	45
Córdoba	4,966148379	1,353392239	77
Granada	4,881494757	1,983911177	174
Huelva	4,531960661	1,228886544	80
Jaen	4,703447572	1,108539667	97
Málaga	5,53973872	1,598345234	103
Sevilla	5,092565856	1,197379093	106

Fuente: elaboración propia con Stata a partir de los datos de SIMA

Tabla 3.

Test t para las provincias de Almería (x) – Málaga (y).

	Obs	Mean	Std. err.	Std. dev.	[95% conf. interval]	
x	103	4.7946	.193173	1.960492	4.411442	5.177758
y	103	5.539739	.1574896	1.598345	5.227359	5.852119
Combined	206	5.167169	.1270081	1.822909	4.91676	5.417579
diff		-.7451387	.2492364		-1.236548	-.253729
diff = mean(x) - mean(y)					t =	-2.9897
H0: diff = 0					Degrees of freedom =	204
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		
Pr(T < t) = 0.0016		Pr(T > t) = 0.0031		Pr(T > t) = 0.9984		

Fuente: elaboración propia con Stata a partir de los datos de SIMA

Tabla 4.

Test t para las provincias de Cádiz (x) – Málaga (y).

	Obs	Mean	Std. err.	Std. dev.	[95% conf. interval]	
x	45	4.828202	.181172	1.215339	4.463074	5.19333
y	103	5.539739	.1574896	1.598345	5.227359	5.852119
Combined	148	5.323393	.1252734	1.524016	5.075824	5.570962
diff		-.7115369	.2668406		-1.238906	-.1841676
diff = mean(x) - mean(y)					t =	-2.6665
H0: diff = 0					Degrees of freedom =	146
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		
Pr(T < t) = 0.0043		Pr(T > t) = 0.0085		Pr(T > t) = 0.9957		

Fuente: elaboración propia con Stata a partir de los datos de SIMA

Tabla 5.

Test t para las provincias de Córdoba (x) – Málaga (y).

	Obs	Mean	Std. err.	Std. dev.	[95% conf. interval]	
x	77	4.966148	.1542334	1.353392	4.658966	5.273331
y	103	5.539739	.1574896	1.598345	5.227359	5.852119
Combined	180	5.29437	.1133929	1.521326	5.070611	5.518128
diff		-.5735903	.2257754		-1.019131	-.1280494

diff = mean(x) - mean(y) t = -2.5405
H0: diff = 0 Degrees of freedom = 178

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
Pr(T < t) = 0.0060 Pr(|T| > |t|) = 0.0119 Pr(T > t) = 0.9940

Fuente: elaboración propia con Stata a partir de los datos de SIMA

Tabla 6.

Test t para las provincias de Granada (x) – Málaga (y).

	Obs	Mean	Std. err.	Std. dev.	[95% conf. interval]	
x	174	4.881495	.1503999	1.983911	4.58464	5.17835
y	103	5.539739	.1574896	1.598345	5.227359	5.852119
Combined	277	5.126257	.1126121	1.874239	4.904569	5.347945
diff		-.658244	.2300321		-1.111092	-.2053963

diff = mean(x) - mean(y) t = -2.8615
H0: diff = 0 Degrees of freedom = 275

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
Pr(T < t) = 0.0023 Pr(|T| > |t|) = 0.0045 Pr(T > t) = 0.9977

Fuente: elaboración propia con Stata a partir de los datos de SIMA

Tabla 7.

Test t para las provincias de Huelva (x) – Sevilla (y).

	Obs	Mean	Std. err.	Std. dev.	[95% conf. interval]	
x	80	4.531961	.1373937	1.228887	4.258485	4.805436
y	106	5.092566	.1162997	1.197379	4.861965	5.323167
Combined	186	4.851445	.0908757	1.23938	4.672159	5.030731
diff		-.5606052	.1793516		-.9144552	-.2067551

diff = mean(x) - mean(y) t = -3.1257
H0: diff = 0 Degrees of freedom = 184

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
Pr(T < t) = 0.0010 Pr(|T| > |t|) = 0.0021 Pr(T > t) = 0.9990

Fuente: elaboración propia con Stata a partir de los datos de SIMA

Tabla 8.

Test t para las provincias de Huelva (x) – Málaga (y).

	Obs	Mean	Std. err.	Std. dev.	[95% conf. interval]	
x	80	4.531961	.1373937	1.228887	4.258485	4.805436
y	103	5.539739	.1574896	1.598345	5.227359	5.852119
Combined	183	5.09918	.1130437	1.529227	4.876135	5.322225
diff		-1.007778	.2158975		-1.433778	-.5817784

diff = mean(x) - mean(y) t = -4.6679
H0: diff = 0 Degrees of freedom = 181

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
Pr(T < t) = 0.0000 Pr(|T| > |t|) = 0.0000 Pr(T > t) = 1.0000

Fuente: elaboración propia con Stata a partir de los datos de SIMA

Tabla 9.

Test t para las provincias de Jaen (x) – Málaga (y).

	Obs	Mean	Std. err.	Std. dev.	[95% conf. interval]	
x	97	4.703448	.1125551	1.10854	4.480027	4.926868
y	103	5.539739	.1574896	1.598345	5.227359	5.852119
Combined	200	5.134138	.1019272	1.441468	4.933141	5.335134
diff		-.8362911	.195632		-1.222081	-.4505014

diff = mean(x) - mean(y) t = -4.2748
H0: diff = 0 Degrees of freedom = 198

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
Pr(T < t) = 0.0000 Pr(|T| > |t|) = 0.0000 Pr(T > t) = 1.0000

Fuente: elaboración propia con Stata a partir de los datos de SIMA

Tabla 10.

Test t para las provincias de Jaen (x) – Sevilla (y).

	Obs	Mean	Std. err.	Std. dev.	[95% conf. interval]	
x	97	4.703448	.1125551	1.10854	4.480027	4.926868
y	106	5.092566	.1162997	1.197379	4.861965	5.323167
Combined	203	4.906632	.0820677	1.169285	4.744813	5.068452
diff		-.3891183	.1624023		-.709349	-.0688876

diff = mean(x) - mean(y) t = -2.3960
H0: diff = 0 Degrees of freedom = 201

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
Pr(T < t) = 0.0087 Pr(|T| > |t|) = 0.0175 Pr(T > t) = 0.9913

Fuente: elaboración propia con Stata a partir de los datos de SIMA

4.6 Test de normalidad

En esta sección vamos a estudiar si nuestra variable objeto de estudio se ajusta a una distribución normal. Para ello usaremos el Test de Kolmogorov-Smirnof. Este test se utiliza para contrastar si un conjunto de datos sigue una distribución normal o no y está basado en un contraste de hipótesis. Mostramos los resultados obtenidos con Stata en la Tabla

11. Como podemos comprobar, el p-valor asociado al test K-S es mayor que 0.05 (ya que en nuestro caso vale 0.053) y, por tanto, se acepta la hipótesis nula y se demuestra que la variable potencial empresarial se distribuye normalmente. Esto se puede comprobar a través del histograma añadiendo una línea de distribución normal tal y como se muestra en la Figura 11.

Tabla 11.

Test de Kolmogorov-Smirnof en Stata.

```
. ksmirnov Empresas=normal((Empresas-4.942655)/1.587743)

One-sample Kolmogorov-Smirnov test against theoretical distribution
normal((Empresas-4.942655)/1.587743)
```

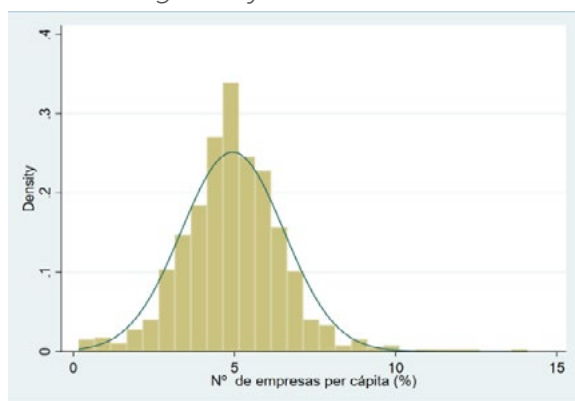
Smaller group	D	P-value
Empresas:	0.0430	0.053
Cumulative:	-0.0405	0.074
Combined K-S:	0.0430	0.106

Note: Ties exist in dataset;
there are 784 unique values out of 794 observations.

Fuente: elaboración propia con Stata

Figura 11.

Histograma y curva normalidad



Fuente: elaboración propia con Stata

4.7 Análisis bivariantes

En esta sección hemos analizado la correlación entre la variable número de

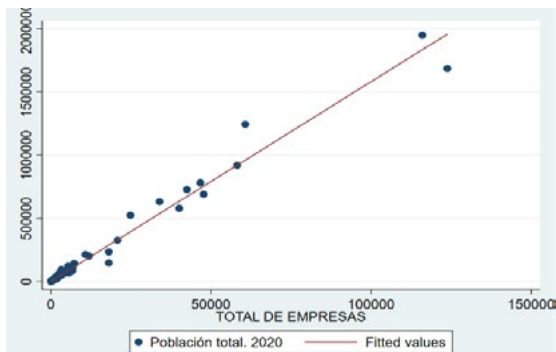
empresas y las demás variables recogidas en la base de datos de SIMA. De todas ellas, las únicas con las que se obtiene una correlación fuerte han sido: población total, número de extranjeros y número de habitantes no asalariados. Las Figuras 12-14 contienen los diagramas relativos a estos análisis bivariantes y el cálculo del coeficiente de correlación para cada pareja de variables. Podemos comprobar como dicho coeficiente es mayor que 0.5. De hecho, es mayor que 0.8 en todos los casos, así que podemos afirmar que existe una correlación lineal

fuerte entre la variable número de empresas y las demás variables enumeradas anteriormente. Añadir que la correlación entre número de empresas y población total resulta obvia a pesar de que existen datos atípicos y extremadamente atípicos que corresponden a valores que quedan fuera de la recta de regresión. Es

especialmente relevante que exista una correlación fuerte (aunque menor que en el caso de las otras dos variables) con el número de extranjeros. Esto podría significar que son ciudadanos que tratan de asentarse en municipios en los que hay empresas y oportunidades laborales.

Figura 12.

Población Total VS Número total de empresas

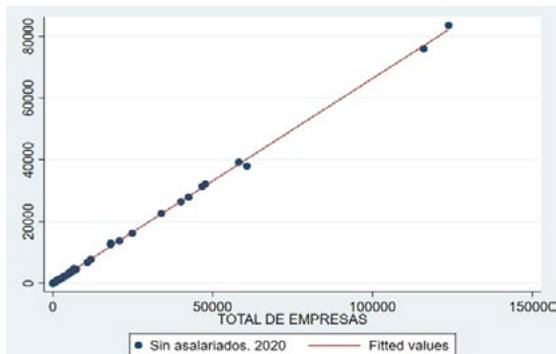


	Población Empresas	
Población	1.0000	
Empresas	0.9984	1.0000

Fuente: Elaboración propia

Figura 13.

Sin asalariados VS Número total de empresas

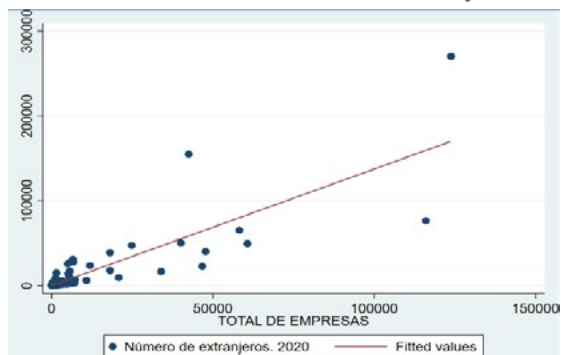


	Sin asalariados Empresas	
Sin asalariados	1.0000	
Empresas	0,997	1.0000

Fuente: Elaboración propia

Figura 14.

Número de extranjeros VS Número total de empresas



	Extranjeros Empresas	
Extranjeros	1.0000	
Empresas	0,8472	1.0000

Fuente: Elaboración propia

5. Discusión y limitaciones

A pesar de no haber encontrado en la literatura otros trabajos en los que se realice un estudio estadístico similar al abordado en el presente artículo, nos encontramos con otros análisis estadísticos basados en la plataforma SIMA como (Cruces Pastor et al., 2009). Otros investigadores han desarrollado estudios donde se plantean modelos que permitan una reducción del consumo en Andalucía (Román-Collado and Carmona, 2021).

Tras el análisis realizado en la Sección 4.2, se puede concluir que, desde el año 2000 hasta hoy, Andalucía ha experimentado cambios significativos. El sector industrial se ha diversificado abarcando muchas actividades económicas. La construcción ha crecido durante los últimos 30 años pasando a suponer un 12%, a pesar de que este sector atravesó un declive durante la crisis de 2008. Por último, el sector servicios sigue siendo el más importante en todas las provincias a excepción de Jaén, donde destaca el

comercio. Este sector es el segundo más importante en Andalucía y le sigue la construcción.

En resumen, desde el año 2000 hasta hoy, Andalucía ha experimentado un crecimiento significativo en el sector servicios, especialmente en el turismo, mientras que el sector de la construcción ha enfrentado desafíos debido a la crisis financiera. Por su parte, la industria ha experimentado cambios y una mayor diversificación en busca de nuevas oportunidades económicas. Estos cambios reflejan la dinámica económica y los desafíos a los que se ha enfrentado Andalucía en su búsqueda por fortalecer su potencial empresarial y promover un desarrollo sostenible.

Para el estudio estadístico nos hemos encontrado con algunas limitaciones derivadas de la base de datos SIMA. La más importante es el hecho de que para algunos municipios no había datos registrados para ciertas variables. Por tanto, no se ha podido contar con algunos datos a

la hora de realizar determinados análisis bivariantes. Dichas ausencias de datos han sido muy pocas y por tanto despreciables. Otra limitación que nos hemos encontrado es que no se hay registros similares de otras comunidades autónomas, por tratar de hacer una comparativa con distintas regiones de España o incluso dar una visión a nivel nacional del potencial empresarial.

6. Conclusiones y perspectivas

Los resultados obtenidos permiten concluir que en Andalucía existen desequilibrios con respecto al potencial empresarial a nivel municipal y provincial. Se ha deducido que nuestra variable sigue una distribución normal. Se ha obtenido que el municipio que mayor potencial empresarial per cápita tiene es Bubión (Granada). Es un municipio de tan solo 298 habitantes, gran extensión y muchas empresas residen allí. Esto concuerda con que Granada es la provincia con mayor número de datos atípicos.

El estudio de número de empresas por actividad económica revela que el sector servicios destaca en todas las provincias a excepción de Jaén, donde predomina más el comercio. Esto es una diferencia importante entre Jaén y las restantes provincias. Al analizar la tipología, se ha concluido que las provincias de Almería, Granada y Jaén destacan por un potencial empresarial Bajo, Huelva y Sevilla por un potencial Medio-Bajo, en Cádiz y Córdoba destaca el Medio-Alto y, por último, tipo Alto en la provincia de Málaga.

Se han identificado 13 datos atípicos y 4 extremadamente atípicos. Es de vital importancia controlar este tipo de datos, pues influyen en el modelo de predicción que se utiliza para poder conjeturar o predecir valores futuros. Además, se ha llevado a cabo un estudio de comparación de medias que ha reflejado que el orden de las provincias de mayor a menor media es: Málaga, Sevilla, Córdoba, Granada, Cádiz, Almería, Jaén y Huelva. De acuerdo con el test de la distribución T de Student, existen diferencias estadísticamente significativas entre Málaga y todas las demás, entre Huelva y Jaén con Sevilla.

El análisis bivalente refleja como existen otras variables de la plataforma SIMA que están fuertemente correlacionadas linealmente con la variable consumo de energía eléctrica. Esto supone que podría establecerse un grado de dependencia o llegar incluso a describir relaciones de causa efecto entre dichas variables.

Como perspectiva de futuro, queda pendiente un análisis similar con posterioridad al año 2020. Tras la COVID-19, ha habido numerosos cambios a nivel social, empresarial y económico que pueden influir en los datos como el cierre de negocios y empresas, el aumento del teletrabajo y la necesidad de mantener una mayor limpieza y desinfección. Es nuestra intención realizar una comparativa posteriormente con los datos esperados y los reales que aparezcan recogidos en SIMA.

Otra posible futura línea de investigación sería ampliar el estudio de las variables que han resultado tener una correlación positiva con el número de empresas. Sería muy interesante analizar el número de extranjeros en cada municipio y tratar de establecer conclusiones sobre cómo la actividad empresarial de una región influye en la cantidad de extranjeros que allí residen. Esto daría lugar a nuevas estrategias sociales, económicas y políticas.

Por último, se espera poder continuar este análisis relacionándolo con la teoría de GVC para poder obtener más conclusiones teóricas. A su vez esto permitiría obtener información relevante para la toma de decisiones estratégicas en el impulso del desarrollo económico y la mejora de la competitividad de la comunidad autónoma de Andalucía.

7. Referencias

- Álvaro, C.F., Daniel, D. and Espinosa, I. (2013). El Instituto de Estadística y Cartografía de Andalucía. *XIV Congreso de Enseñanza y Aprendizaje de las Matemáticas*.
- Bellido-Jiménez, V., Martín-Martín, D. & Romero, I. (2021). The survival of new businesses in Andalusia (Spain): Impact of urbanisation, education, and gender. *Regional Science Policy & Practice*. 13. [10.1111/rsp3.12308](https://doi.org/10.1111/rsp3.12308)
- Bernardi, F. y Escobar Mercado, M. (2012). Análisis de datos con Stata (2a. ed.) (Colección Cuadernos Metodológicos, 45). Madrid, Spain - CIS.
- Confederación de Empresarios de Andalucía (CEA). (2020). Andalucía, Región Empresarial. España.
- Consejería de Economía, Conocimiento, Empresas y Universidad de la Junta de Andalucía. (2018). Análisis de las Cadenas Globales de Valor en Andalucía. España.
- Consejería de Economía, Conocimiento, Empresas y Universidad de la Junta de Andalucía. (2020). Estrategia de Innovación de Andalucía 2020. España.
- Cruces Pastor, E.M., de Haro García, J. & Sarrión Gavilán, M.D. (2009). Análisis estadístico de la realidad socioeconómica en Andalucía. Una aproximación a escala municipal. *Investigaciones regionales* (18), 107-138.
- González, M. and Palma, L. (2020). The “business–territory” relationship of cooperative societies as compared to the conventional business sector in the region of Andalusia. *Annals of Public and Cooperative Economics*. 91. [10.1111/apce.12282](https://doi.org/10.1111/apce.12282)
- Instituto de Estadística y Cartografía de Andalucía (IECA). (2017). Análisis de las Cadenas de Valor en Andalucía: Propuesta Metodológica y Estimación de la Participación Regional.
- Instituto de Estadística y Cartografía de Andalucía (IECA). (2021). Informe Económico de Andalucía 2021.
- Instituto de Estadística y Cartografía de Andalucía (IECA). (2021). Andalucía en Cifras.

- Instituto Nacional de Estadística (INE) (2009). CNAE-2009. https://www.ine.es/daco/daco42/clasificaciones/cnae09/notasex_cnae_09.pdf
- Kano, L., Tsang, E.W.K. and Yeung, H.Wc. (2020). Global value chains: A review of the multi-disciplinary literature. *J Int Bus Stud* 51, 577–622.
- Morán Álvarez, J.C. and Guzman Cuevas, J. (1991). Empresas y Empresarios en Andalucía. 16, 51–71.
- Román-Collado, R. & Carmona, M. (2021). Energy efficiency's key role in explaining the performance of energy consumption in Andalusia (Spain). *Environmental Science and Pollution Research* (28), 1-21. <https://doi.org/10.1007/s11356-020-11829-2>
- Sistema de Información Multiterritorial de Andalucía (SIMA) (2022) https://www.juntadeandalucia.es/instituto-deestadisticaycartografia/badea/informe/anual?CodOper=b3_151&idNodo=23204
- Universidad Loyola (2021). Guía de *stata*. yserte, R., Rivera, M. and Martínez, D. (2016). The economic crisis and the geography business network in Spain: 2000-2013. *Revista de Estudios Regionales* 165-195.