



Análisis de correspondencia múltiple entre el tamaño de las empresas y su participación económica en el Estado de México

Multiple correspondence analysis between the size of the companies and their economic participation in the State of Mexico

Armando Puebla Maldonado^{a*}

*Autor de correspondencia: armando.pueblama@uvmnet.edu - <https://orcid.org/0000-0002-5432-5789>

^a Universidad del Valle de México, División de Licenciatura, Área de Negocios y Sociales, Campus Lomas Verdes, Estado de México, México

INFORMACIÓN DEL ARTÍCULO

Presentado:
25/01/2023

Aceptado:
02/03/2022

Disponible online:
16/03/2023

CÓDIGO JEL:
C21, R12, R11

RESUMEN

La presente investigación tiene como propósito el análisis económico regional que coadyuve en el diagnóstico de la planeación sectorial del Estado de México. Para cumplir con este propósito se buscó determinar el nivel de asociación entre el tamaño de las unidades económicas y un conjunto de variables representativas de la actividad económica. Se partió de la metodología de minería de datos para aplicar de manera cuantitativa y descriptiva la metodología del Análisis de Correspondencia Múltiple (ACM). Los resultados sugieren, que hay un nivel de asociación positiva moderada alrededor del 50 % en dos de sus indicadores: El coeficiente de asociación de Pearson es de 0.4685 y el de Mathews con 0.5303. Sin embargo, en los indicadores V de Cramer y de Inercial Total fueron correlaciones positivas débiles respectivamente de 0.3062 y 0.2812. Ante esta situación se propuso como medida alternativa el análisis discriminante y el de varianza, los cuales arrojaron que las microempresas están más asociadas con el Personal Ocupado Total (POT) y la Producción Bruta Total (PBT). Por su parte, el Valor Agregado Censal Bruto (VACB) e Ingreso por Suministros de Bienes y Servicios (ISBB) con medianas y grandes empresas. Asimismo, la prueba Chi-Cuadrado, además de que comprobó que los datos son estadísticamente significativos al 95 %, también apoyo la hipótesis de una asociación entre el tamaño de las unidades económicas y las variables de Participación Económica. Los resultados abonan a que las políticas públicas deberían considerar la heterogeneidad de la planta productiva y su participación económica.

Palabras clave: Análisis de correspondencia, asociación, correlación, participación.

Como citar: Puebla Maldonado, A. (2023). Análisis de correspondencia múltiple entre el tamaño de las empresas y su participación económica en el Estado de México. *Economía & Negocios*, 5(1), 154-174. <https://www.doi.org/10.33326/27086062.2023.1.1653>



ABSTRACT

The present investigation has as purpose the regional economic analysis that contributes in the diagnosis of the sectorial planning of the State of Mexico. To fulfill this purpose, we sought to determine the level of association between the size of the economic units and a set of representative variables of economic activity. The data mining methodology was used to quantitatively and descriptively apply the Multiple Correspondence Analysis (MCA) methodology. The results suggest that there is a moderate positive association level around 50% in two of its indicators: Pearson's association coefficient is 0.4685 and Mathews's is 0.5303. However, in Cramer's V and Total Inertial indicators, there were weak positive correlations of 0.3062 and 0.2812, respectively. Given this situation, discriminant analysis and variance analysis were proposed as an alternative measure, which showed that microenterprises are more associated with Total Employed Personnel (POT) and Total Gross Production (PBT). For its part, the Gross Census Value Added (VACB) and Income from Supplies of Goods and Services (ISBB) with medium and large companies. Likewise, the Chi-Square test, in addition to verifying that the data is statistically significant at 95%, also supports the hypothesis of an association between the size of the economic units and the Economic Participation variables. The results support that public policies should consider the heterogeneity of the productive plant and its economic participation.

Keywords: Correspondence analysis, association, correlation, stake.

INTRODUCCIÓN

El Análisis de Correspondencia (AC) permite establecer una relación entre las variables y las observaciones, de tal manera que es posible señalar los conjuntos de observaciones semejantes y caracterizarlos de acuerdo con las variables utilizadas (Valverde et al., 1994, p. 135). En este sentido, Fernández (2002) señala que el AC es concebido como una técnica estadística diseñada para analizar Tablas de Contingencia, Tablas de Frecuencias, Tablas de Valoración y con otras Medidas de Correspondencia (p. 692). Por lo que se pretende hacer uso de las tablas de contingencia, ya que son de ampliamente recomendadas para hacer análisis en el área de Ciencias Sociales.

Una variante del AC es el Análisis de Correspondencia Múltiple (ACM) que no es tema nuevo dentro del análisis estadístico. Sin embargo, dentro del estudio de minería de datos, es posible que sea uno de los temas de aplicación relativamente nuevo. En el ámbito de la geografía probablemente sea una de las áreas que tenga mayor potencial de uso. Al respecto están los trabajos de Pérez Herver y Zapata Sánchez (2019), acerca de la correlación entre prácticas ambientales y desempeño de los hoteles en Quintana Roo o la investigación sobre la correlación entre los factores micro, meso y macroeconómicos y la evolución de empresas de base tecnológica en Colombia de Moreno Cuello y Ramos Camargo (2013), también el artículo sobre la contribución de los sistemas productivos en la generación de ingresos en familias cacaoteras de Sánchez Olaya et al. (2015). No está por demás señalar que, en minería de datos, la técnica de correspondencia forma parte de los modelos de asociación y dependencia que a su vez pertenecen a las técnicas de los modelos descriptivos (Leyva Jiménez, 2019).

Como se mencionó, las variantes del Análisis de Correspondencia (AC) son el Análisis de Correspondencia Simple (ACS) y el Múltiple (ACM), aplicados a datos categóricos multivariados con respecto a encuestados y el Análisis de Correspondencia Canónica. A diferencia de los métodos estadísticos tradicionales, como la regresión de mínimos cuadrados ordinarios (OLS), el ACS y ACM no hacen suposiciones sobre la estructura de los datos, sino que identifican la estructura de las relaciones entre las variables y entre los objetos (por ejemplo, individuos) en la investigación de encuestas, por lo general de los individuos (Blasius y Schmitz, 2015). Su objetivo es detectar patrones ocultos de asociación y facilitar la interpretación de datos. Los fundamentos matemáticos fueron desarrollados inicialmente en el área de estadística inferencial cuando se estudiaron los temas de asociación y correlación abordados por los matemáticos por Kal Person y Ronald Fisher a principios del siglo pasado. La técnica del AC como tal fue propuesta por Jean Paul Benzecri, quién la desarrolló en la década de los sesenta del siglo pasado (Alberti, 2017).

Cabe mencionar que la vinculación entre economía y espacio geográfico en las últimas décadas ha sido analizada desde diferentes puntos de vista, utilizando para ello el ACM, como el trabajo de Ghani et al. (2013), que lo abordan desde el punto de vista de los factores de emprendimiento, y como el rol de la educación y la infraestructura son importantes para generar el desarrollo local. Asimismo, el trabajo sobre las relaciones entre los capitales de salud, educativo, financiero, social y físico y el éxito de las microempresas en el área Metropolitana de Bucaramanga (Rangel Quiñones y Yañez Canal, 2018). En este sentido, Gaxiola Robles Linares et al. (2017) explican la correlación múltiple entre seguridad social y las características sociodemográficas y laborales de los trabajadores en el estado de México en el año 2015. También está la investigación sobre el análisis de correlación entre calidad del empleo y las 13 principales áreas metropolitanas de Colombia, por parte de Jiménez Restrepo y Páez Cortés (2014).

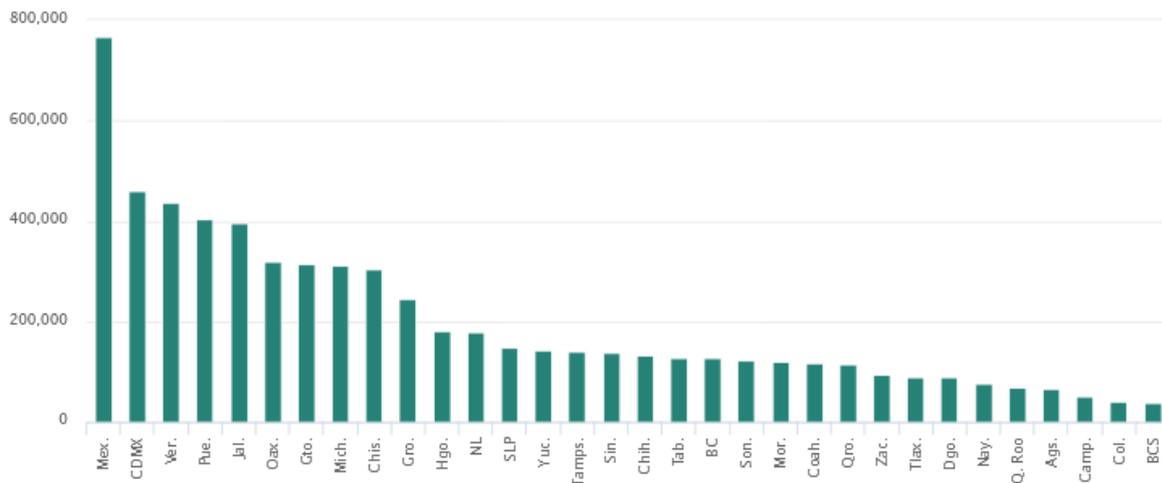
Las múltiples aplicaciones de la técnica también se encuentran en otras áreas como la Lingüística, Ciencias de la Educación, Psicología, Arqueología, Sociología, Economía, Mercadotecnia Y Salud; ya que el ACM explora las asociaciones inherentes a preguntas de opción múltiple o bien de datos cuantitativos. Al respecto están los trabajos de sustentabilidad publicados por varios autores. Por ejemplo, Aquino Zacarías et al. (2022) aplicaron el ACM en sustentabilidad del *tarwi* en la agricultura. También está el uso del ACM en la evaluación de la sostenibilidad financiera en el crédito del desarrollo humano (Valencia Núñez et al., 2022); la investigación sobre estructura de mercado de las sociedades productoras y comercializadoras de calzado por Benítez Gaibor et al. (2020). Incluso a nivel empresarial encontramos algunas aplicaciones del ACM, como, por ejemplo, el trabajo sobre sostenibilidad, *stakeholders* y crisis de empresa de Contreras-Pacheco et al. (2020). Con un enfoque más económico está el trabajo sobre desigualdad en hogares con adultos mayores en México en el año 2015, por Gaxiola-Robles Linares et al. (2020). Como lo menciona Colmenares (2017), el AC resulta sumamente útil a la hora de estudiar fenómenos o problemas de naturaleza multidimensional y tan complejos como son la pobreza, así como muchos otros fenómenos y problemas geográficos y espaciales (p. 45). En la ciencia económica en general sobresalen artículos en el ámbito de la correlación, la regresión y la econometría y poco se trabaja en el ACM aplicado a la geografía económica macroeconómica desde una perspectiva de minería de datos.

De ahí la importancia de aplicar la técnica a nivel regional con fines de política pública que coadyuven al desarrollo de la región y sobre todo en el caso del Estado de México, ya que del total

de establecimientos registrados en el Censo económico del 2019 que alcanzó la cifra de 6 373 169, el 12 % (764 780.280) correspondió al Estado de México, ubicándolo en el primer lugar a nivel nacional. Además, que se encuentra, junto con la zona Norte y Centro del país, entre los estados con la mayor actividad, ya que genera el más alto valor agregado por su actividad industrial, a diferencia del sur que se destaca por la generación de servicios, como se observa en la Figura 1.

Figura 1

Número de establecimientos por entidad federativa, 2019



Nota. Tomado de la página <https://www.inegi.org.mx/programas/ce/2019/>

Algunas investigaciones incluso señalan que la actividad productiva en la región, el crecimiento promedio, ha sido superior al nacional, además de ser poco dinámica y errática (Trejo Nieto y Negrete, 2018, p. 32). Tomando en consideración esta situación se plantea la siguiente pregunta: ¿Cuál es el nivel de asociación entre el Tamaño de los Establecimientos y su Participación Económica en el Estado de México? Para responder a este cuestionamiento se procedió a clasificar a las variables económicas en dos dimensiones de análisis:

Dimensión 1: El Tamaño de las Unidades Económicas (TUE). Es la variable independiente y se refiere al tamaño de las unidades económicas que está en función del número de trabajadores y que son: Microempresas (0-10), Pequeñas Empresas (11-50), Medianas Empresas (51-250) y Grandes Empresas (250 y más).

Dimensión 2: La Participación Económica. Representa la variable independiente y son aquellas variables económicas que son representativas de la actividad económica: Personal Ocupado Total (POT), las Remuneraciones (R), el Gasto por Consumo de Bienes y Servicios (GCBS), los Ingresos por Suministro de Bienes y Servicios (ISBB), la Producción Bruta Total (PBT) y el Valor Agregado Censal Bruto (VACB).

A partir de lo expuesto, y para probar cuáles son las variables económicas de la categoría de Participación Económica que tienen mayor peso con relación al tamaño de las empresas, se plantean 2 hipótesis estadísticas: Hipótesis Nula (H_0) que se busca contrastar y la Hipótesis Alternativa (H_a) que se afirma:

H_0 = Las variables son independientes: el Tamaño de las Unidades Económicas no está relacionado con la Participación Económica en el Estado de México.

H_a = Existe una relación de dependencia entre las variables Tamaño de las Unidades Económicas y Participación Económica en el Estado de México.

MATERIAL Y MÉTODOS

Para dar respuesta a la pregunta de investigación, en el presente estudio se aplicó una metodología de tipo cuantitativa, descriptiva y de corte transversal, porque corresponde a datos oficiales generados en el año 2019, año en que se levantó el censo económico, y para validar los resultados se utilizó el software SPSS IBM versión 26.0. Por su parte, el análisis de minería de datos, de acuerdo con Signorelli y Biffignandi (2018), puede ser útil para dos propósitos diferentes: operacionales y estadísticos. Así como sucede en el ámbito de la estadística tradicional, en el caso de minería de datos, se parte de una metodología más enfocada al análisis de datos, el cual contiene una serie de pasos a seguir:

1.- Definición de Objetivos. Es importante delimitar hasta donde se quiere llegar con la información y los resultados obtenidos.

2.- Preprocesamiento de los Datos. en esta etapa se busca sobre todo definir, delimitar, procesar, filtrar los datos con el objetivo de seleccionar aquello que necesitamos y cumpla con el paso anterior. Para cumplir con este paso se requiere llevar a cabo las siguientes fases (Tufféry, 2011):

- a) Recopilar los datos
- b) Explorar y preparar los datos
- c) Enumerar los datos existentes
- d) Segmentar o dividir los datos
- e) Construcción o elección del modelo
- f) Elección de la técnica de minería de datos
- g) Extracción de la información

3.-Evaluación y Análisis de los Resultados. Revisar si los resultados son confiables y coherentes con los objetivos y la metodología propuesta, a fin de validar la información obtenida.

RESULTADOS

Al analizar los resultados, generados por la actividad económica, es imposible no separarlo del aspecto geográfico como lo describe Krugman (2010) en su famoso artículo *The New Economic Geography, Now Middle-aged*. Como el trabajo está enfocado a la aplicación de minería de datos a la geografía cuantitativa, es importante vincularlo con los pasos y fases mencionadas anteriormente a la metodología en geografía tal y como se plantea en este ámbito de estudio (Montes Galbán, 2015), el cual consiste en las siguientes fases:

Fase Descriptiva

Responde a la pregunta: ¿Cómo es el objeto geográfico? Corresponde al paso 2 de preprocesamiento de los datos donde se revisan y observan los datos agrupados en variables estadísticas descriptivas a utilizar en el proyecto. Se procedió entonces con la metodología de minería de datos, empezando por descargar la información del Censo Económico del 2019 del Instituto Nacional de Estadística y Geografía (INEGI) de cada entidad federativa junto con sus respectivas variables de análisis. Posteriormente, se filtraron los datos utilizando SPSS y se seleccionaron las variables de estudio.

Así, de acuerdo con información sobre las Características de los Negocios (Instituto Nacional de Estadística y Geografía, 2019) generados a partir de los datos del Censo Económico del 2019, el Estado de México ocupa el primer lugar en número de establecimientos en el país, como se muestra en la Tabla 1, lo cual refleja la importancia que tiene el Estado en la generación de empleos, la derrama salarial y la generación de riqueza, no solo en el Centro del país, sino también a escala nacional.

Tabla 1

Número de establecimientos

Entidad	Establecimientos	%
Estado de México	764 780	27.4
Ciudad de México	459 651	7.2
Veracruz	438 214	6.9
Puebla	405 111	6.4
Jalisco	397 119	6.2
Oaxaca	318 595	5.0
Total	2 783 470	59.17

Nota. Elaboración propia.

De hecho, se observa que hay una diferencia importante en cuanto al número de establecimientos entre el Estado de México y la Ciudad de México, que es de 307 896. A pesar de esta situación, ocupa el segundo lugar en Personal Ocupado Total, como se alcanza a apreciar en la Tabla 2.

Tabla 2*Personal Ocupado Total*

Entidad	Personal Ocupado Total	%
Ciudad de México	5 203 505	14.3
Estado de México	3 356 905	2.2
Jalisco	2 441 989	6.7
Nuevo León	2 705 990	5.7
Guanajuato	1 731 726	4.8
Veracruz	1 705 572	4.7
Total	17 145 687	38.4

Nota. Elaboración propia.

Los datos arrojan aparentemente que no hay mucha vinculación entre número de establecimientos y Personal Ocupado Total, por lo que es necesario incluir más variables y elaborar tablas de contingencia para su análisis.

Fase Explicativa

Consiste en tres etapas: la de construcción de tablas de frecuencias absolutas y relativas, la de construcción de perfiles por renglón y columna y tabla de contingencia, y finalmente de la contrastación de hipótesis. Esta fase busca contestar la pregunta ¿En dónde está y porque se ubica ahí? Para responderla se hace uso de la estadística descriptiva e inferencial para explicar el comportamiento y estructura de los datos, con el propósito de someter a prueba los datos a un marco teórico para extraer conclusiones. En este sentido, y partiendo de la primera etapa, se construyeron las tablas de frecuencia y de perfiles.

Tabla 3*Frecuencia Absolutas entre Tamaño de las Unidades Económicas y Participación Económica*

Tamaño	TUE	POT	R	GCBS	ISBS	PBT	VACB	Total
0 a 10 personas	604 982	1 190 983	21 000	250 657	369 058	199 159	118 485	2 754 324
11 a 50 personas	14 809	312 448	19 292	260 664	360 548	179 340	98 949	1 246 050
51 a 250 personas	3667	401 405	35 141	521 297	715 964	431 399	194 655	2 303 529
251 y más personas	1014	622 444	77 433	725 337	1 084 526	1 021 653	357 417	3 889 824
Total	624 472	2 527 280	152 867	1 757 955	2 530 096	1 831 551	769 506	10 193 727

Nota. Elaboración propia.

Tabla 4

Frecuencia Relativas entre Tamaño de las Unidades Económicas y Participación Económica

Tamaño	TUE	POT	R	GCBS	ISBS	PBT	VACB	Total
0 a 10 personas	0.0593	0.1168	0.0021	0.0246	0.0362	0.0195	0.0116	0.2702
11 a 50 personas	0.0015	0.0307	0.0019	0.0256	0.0354	0.0176	0.0097	0.1222
51 a 250 personas	0.0004	0.0394	0.0034	0.0511	0.0702	0.0423	0.0191	0.2260
251 y más personas	0.0001	0.0611	0.0076	0.0712	0.1064	0.1002	0.0351	0.3816
Total	0.0613	0.2479	0.0150	0.1725	0.2482	0.1797	0.0755	1.0000

Nota. Elaboración propia.

Por lo que respecta a la Tabla 4, en términos porcentuales se observa que los Ingresos Sobre Bienes y Servicios y la Población Ocupada Total son las variables que tienen las ponderaciones más altas con 0.2482 y 0.2479 respectivamente, lo que indica el valor que tiene la generación de empleo y de los ingresos. Respecto a filas, sobresale el peso que tienen las grandes empresas (0.318) en la actividad económica sobre las microempresas (0.2702), a pesar de que estas últimas representan el 5.93 % en el Tamaño de Unidades Económicas y las grandes empresas apenas el 0.1 %.

Con el objetivo de pasar a la etapa 2, se calculó primeramente la tabla de perfil promedio por renglón, con la idea de conocer la proporción de cada categoría de Participación Económica en el Tamaño de las Unidades Económicas y se determinó dividiendo la frecuencia de cada variable de la Participación Económica (n_{i+}) entre el Tamaño de las Unidades Económicas (n).

$$r_i = \frac{n_{i+}}{n} \tag{1}$$

Tabla 5

Perfiles Renglón entre Tamaño de las Unidades Económicas y Participación Económica

Tamaño	TUE	POT	R	GCBS	ISBS	PBT	VACB	Total
0 a 10 personas	0.2196	0.4324	0.0076	0.0910	0.1340	0.0723	0.0430	1.0000
11 a 50 personas	0.0119	0.2508	0.0155	0.2092	0.2894	0.1439	0.0794	1.0000
51 a 250 personas	0.0016	0.1743	0.0153	0.2263	0.3108	0.1873	0.0845	1.0000
251 y más personas	0.0003	0.1600	0.0199	0.1865	0.2788	0.2626	0.0919	1.0000
Perfil promedio del renglón	0.0583	0.2544	0.0146	0.1782	0.2532	0.1665	0.0747	1.0000

Nota. Elaboración propia, realizada a partir de la aplicación de la Ecuación 1.

En la Tabla 5 observa que las participaciones más altas las tienen las microempresas en la POT (0.4324), le sigue el ISBS de las medianas y pequeñas empresas (0.3108 y 0.2894). En promedio sobresalen el POT y el ISBS, lo que significa que son variables a considerar al momento de aplicar la fase de contrastación de hipótesis.

Asimismo, se hizo el cálculo del perfil por columna (Participación Económica). Se calcula dividiendo la frecuencia de cada variable de la Participación Económica (n_{+j}) entre el total de del Tamaño de las Unidades Económicas (n).

$$r_i = \frac{n_{+j}}{n} \quad (2)$$

Tabla 6

Perfiles Columnas entre Tamaño de las Unidades Económicas y Participación Económica

Tamaño	TUE	POT	R	GCBS	ISBS	PBT	VACB	Perfil
								promedio por columna
0 a 10 personas	0.9688	0.4713	0.1374	0.1426	0.1459	0.1087	0.1540	0.3041
11 a 50 personas	0.0237	0.1236	0.1262	0.1483	0.1425	0.0979	0.1286	0.1130
51 a 250 personas	0.0059	0.1588	0.2299	0.2965	0.2830	0.2355	0.2530	0.2089
251 y más personas	0.0016	0.2463	0.5065	0.4126	0.4287	0.5578	0.4645	0.3740
Total	1.0000							

Nota. Elaboración propia, realizada a partir de la aplicación de la Ecuación 2.

De la Tabla 6, los perfiles por columna muestran la proporción de cada tamaño de empresa en cada una de las variables de la Participación Económica. Se aprecia que las microempresas destacan con 0.9688, le siguen la PBT y R en grandes empresas (0.5578 y 0.5065). Esto significa la importancia que tiene las microempresas en la actividad económica del Estado y que las grandes empresas son las que más generan Valor Agregado.

Una vez obtenidas las tablas de perfiles, se pasó a la etapa 3, con el propósito de obtener la tabla de coeficientes de contingencia, que se obtiene dividiendo los valores resultantes de los perfiles renglón o columna por su respectivo promedio, el cual nos da por resultado los valores que se muestran en la Tabla 7. Se observa el peso que tienen el Tamaño de las Unidades Económicas en las microempresas, la PBT en las grandes empresas y los ISBS en las medianas empresas.

Tabla 7

Coefficiente de contingencia entre tamaño de las Unidades Económicas y Participación Económica

Tamaño	TUE	POT	R	GCBS	ISBS	PBT	VACB	Total
0 a 10 personas	3.76	1.70	0.52	0.51	0.53	0.43	0.58	8.04
11 a 50 personas	0.20	0.99	1.06	1.17	1.14	0.86	1.06	6.50
51 a 250 personas	0.03	0.69	1.05	1.27	1.23	1.12	1.13	6.51
251 y más personas	0.00	0.63	1.37	1.05	1.10	1.58	1.23	6.95
Total	4.00	28.00						

Nota. Elaboración propia.

Finalmente, en esta etapa se someten a prueba de homogeneidad los datos obtenidos y a un conjunto de pruebas de asociación y de correspondencia. Se busca probar la hipótesis a una comprobación estadística (prueba Chi-Cuadrada), con el objetivo de verificar si se acepta o se rechaza. Para eso se calculan y se comparan las frecuencias esperadas con las observadas y, si no hay diferencias entre ellas, se establece que no hay asociación entre ellas; el cual se lleva a cabo a través de los siguientes pasos:

Paso 1: El cálculo de los valores esperados (n_{ij}^e) se consigue a través de la siguiente fórmula:

$$n_{ij}^e = \frac{n_{i+} \cdot n_{+j}}{n} \quad (3)$$

Cuyos resultados se muestran en la Tabla 8.

Tabla 8

Frecuencia Relativas entre Tamaño de las Unidades Económicas y Participación Económica

Tamaño	TUE	POT	R	GCBS	ISBS	PBT	VACB	Total
0 a 10 personas	168 731.0	682 865.8	41 304.3	474 995.8	683 626.7	494 881.3	207 918.9	8.04
11 a 50 personas	76 333.5	308 927.0	18 686.0	214 887.0	309 271.2	223 883.2	94 062.1	6.50
51 a 250 personas	141 115.2	571 102.5	34 544.1	397 254.1	571 738.8	413 885.0	173 889.2	6.51
251 y más personas	238 292.3	964 384.7	58 332.5	670 818.0	965 459.3	698 901.5	293 635.8	6.95
Total	4.00	28.00						

Nota. Elaboración propia, realizada a partir de la aplicación de la Ecuación 3.

Paso 2: Una vez obtenidas las frecuencias esperadas se procede al cálculo del estadístico Chi-Cuadrado de Pearson que se obtiene de la siguiente fórmula:

$$X^2 f = \sum \frac{(f_{ij} - e_{ij})^2}{e_{ij}} \quad (4)$$

Donde:

f_{ij} = Valores observados

e_{ij} = Valores esperados

Haciendo las operaciones el resultado es de $X^2 f = 2\ 867\ 314.191$

Paso 3: Se le asigna un nivel de significancia del 95 %, cuyo valor en tabla viene dado por:

$$X^2_{gl(r-1)(c-1)} \alpha = X^2_{gl(4-1)(7-1)} 0.95 = 28.869$$

Paso 4: Criterio de decisión. Se comparan los valores de Chi-Cuadrado de fórmula respecto al Chi-Cuadrado de tabla, considerando el siguiente criterio:

Si $X^2 f < X^2 t$ no se rechaza H_0

Si $X^2 f > X^2 t$ se rechaza H_0

Paso 5: Toma de decisión. Como $2\ 867\ 314.25 > 28.869$ se rechaza H_0 . Para validarlo a continuación se presentan los resultados generados por SPSS 26.0 en la Tabla 9.

Tabla 9

Pruebas de Chi-Cuadrado

Prueba	Valor	gl	Significación asintótica (bilateral)
Chi-Cuadrado de Pearson	2 867 314.191	18	0.000
Razón de verosimilitud	2 817 825.482	18	0.000
N de casos válidos	10 193 726		

Nota. Elaboración propia.

Vemos que la significación es de $p \leq 0.05$ ($0.000 < 0.005$), lo que significa que se acepta la Hipótesis Alternativa, en consecuencia, existe una asociación entre el Tamaño de las Unidades Económicas y las variables de Participación Económica.

Para respaldar los resultados obtenidos y apoyar la hipótesis se hicieron varias pruebas con medidas de asociación, empezando por el Coeficiente de Contingencia de Pearson cuyo valor es de

0.4685, el cual coincide con los resultados obtenidos en SPSS que también son significativos, ya que cumple con el criterio de $p \leq 0.05$ en todos los casos.

$$C_c = \sqrt{\frac{x_f^2}{x_f^2+n}} \quad (5)$$

$$C_c = \sqrt{\frac{2\ 867\ 314.25}{2\ 867\ 314.25+10\ 193\ 727}} = 0.469$$

Si el valor de este coeficiente es de 0 indica ausencia de asociación. Por el contrario, si el valor es de 1, indica asociación perfecta. Como el resultado de la Ecuación 5 es de 0.469 podemos interpretarlo como el 46.9 % de asociación entre ambas categorías: Tamaño de las Unidades Económicas respecto a la Participación Económica.

Posteriormente, se obtuvo la medida de V de Cramer, que es un procedimiento de corrección de la Chi-Cuadrado, lo cual indica obtener un valor máximo de 1 (máxima asociación) o un valor mínimo de 0 (mínima asociación), cuyo valor se obtuvo a partir de la fórmula:

$$V = \sqrt{\frac{x_f^2}{(N)(\min(f,c)-1)}} \quad (6)$$

Donde:

N = número total de observaciones

n = min (filas -1) (columnas -1)

$$V = \sqrt{\frac{2\ 867\ 314.25}{(10\ 193\ 727)(3)}} = 0.3062$$

Otro de los resultados obtenidos fue el del Coeficiente de Correlación de Mathews o de phi que, al igual que las anteriores medidas, mide la asociación entre dos variables nominales. Su resultado se obtiene a partir de la fórmula:

$$phi = \sqrt{\frac{x_f^2}{n}} \quad (7)$$

$$phi = \sqrt{\frac{2\ 867\ 314.25}{10\ 193\ 727}} = 0.5303$$

Cuyo rango de valor va de -1 a +1 y entre más se acerca a 1 mayor asociación. Estos resultados se pudieron corroborar con el software de SPSS 26.0, como se muestra en la Tabla 10.

Tabla 10*Medidas Simétricas*

		Valor	Error estándar asintótico	T aproximada	Significación aproximada
	Phi	0.530			0.000
Nominal por Nominal	V de Cramer	0.306			0.000
	Coefficiente de contingencia	0.469			0.000
Ordinal por ordinal	Tau-b de Kendall	0.314	0.000	1 246.644	0.000

Nota. Elaboración propia mediante SPSS 26.0.

* N.º de casos válidos: 10 193 726.

Finalmente, se calculó la Inercia Total. Esta calcula las distancias entre los perfiles de renglón o de columna y el perfil promedio, de tal manera que si la inercia es 0 significa que no hay correlación entre las variables, ya que los perfiles por renglón o columna son idénticos respecto al perfil promedio. Sucede lo contrario cuando se aproxima a 1. La fórmula del cálculo es:

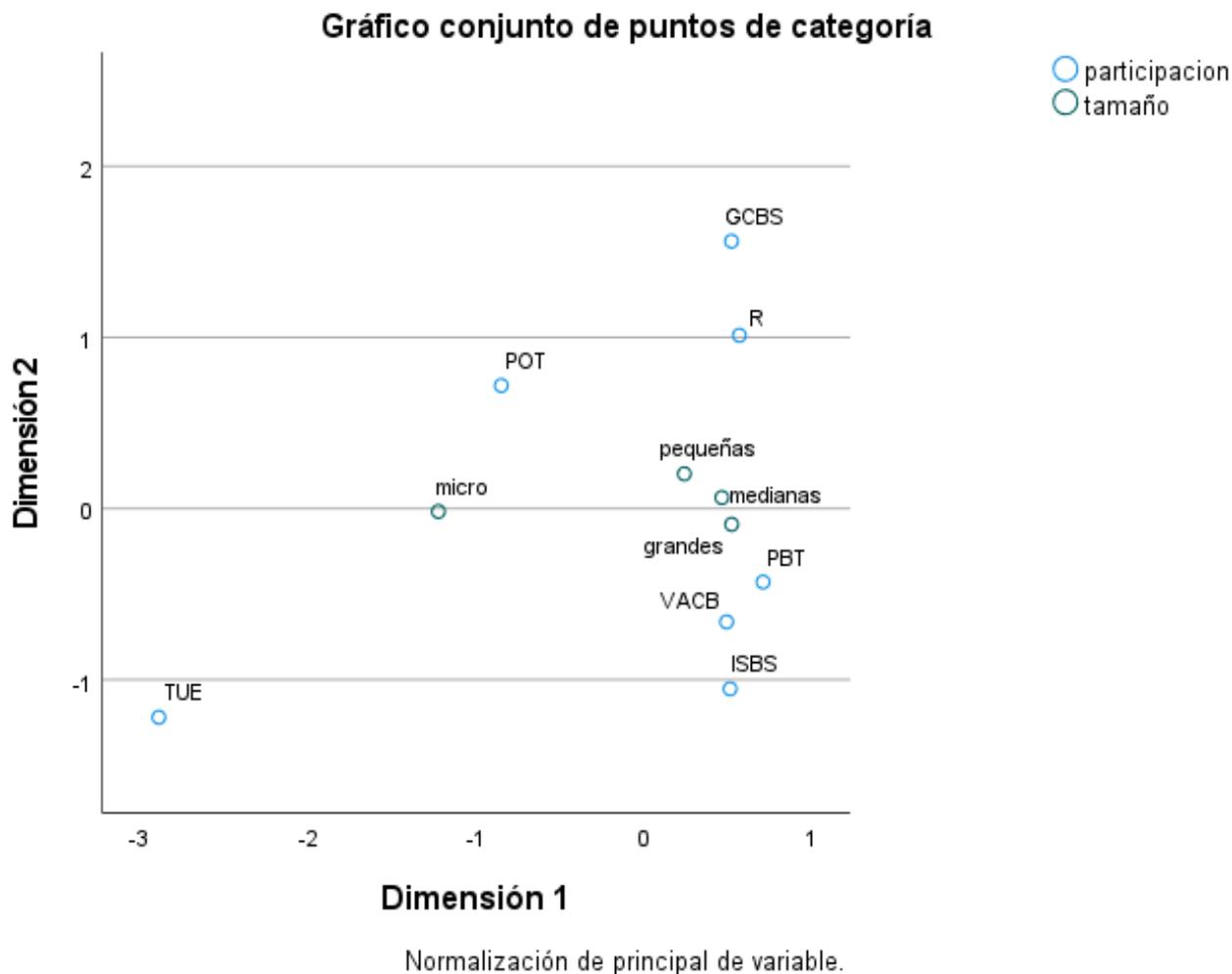
$$IT = \sum Total\ Renglon \frac{(perfil\ renglon\ observado - perfil\ renglon\ esperado)^2}{perfil\ renglon\ esperado} = \frac{2\ 867\ 314.191}{10\ 193\ 727} = 0.2812$$

Fase Aplicativa

Consiste en la evaluación y recomendaciones de los resultados obtenidos, con fines de planeación pública. Para ello nos vamos a apoyar en los resultados gráficos obtenidos por SPSS 26.0. Por ejemplo, en la Figura 2 de variables conjuntas se aprecia que existe una cercanía entre pequeñas, medianas y grandes empresas (Dimensión 1) con las variables de PBT, VACB, ISBYS (Dimensión 2), en contraste con la microempresa que está más alineada con la POT y el Tamaño de Unidades Económicas (TUE).

Figura 2

Gráfica de correlación entre Dimensión 1 y 2 por categoría (Participación y Tamaño)

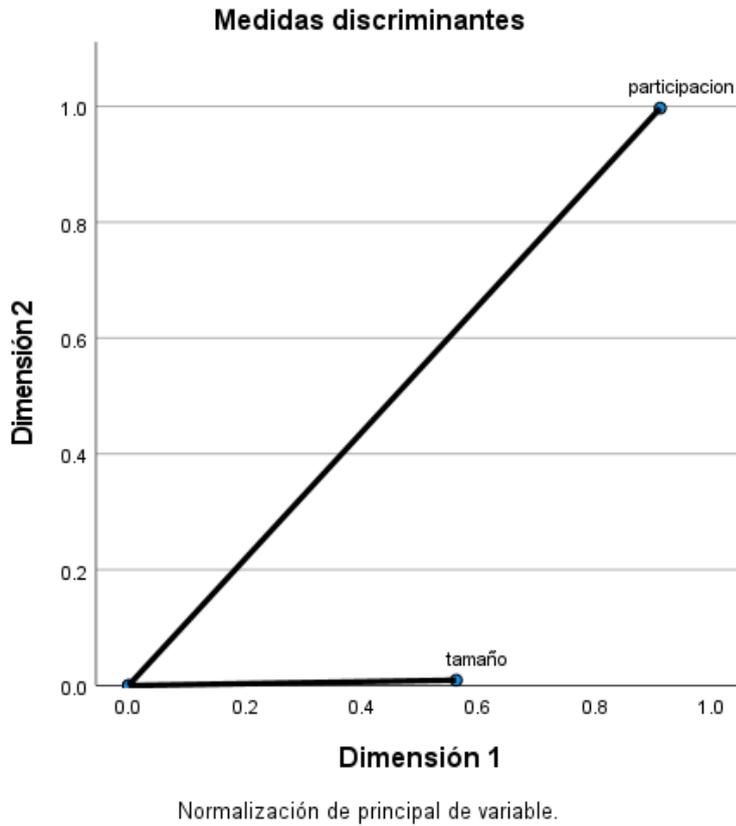


Nota. Elaboración propia, mediante SPSS 26.0.

Esto nos lleva a la explicación de que la presencia de un gran número de empresas y una alta población ocupada en el Estado de México se traduce en mayores niveles de valor, producción, ingresos y gastos de manera diferenciada las empresas de la región. Como se ve en la Figura 2 y en la Tabla 11 de medidas discriminantes, donde la participación presenta en ambas dimensiones valores superiores al 0.90 y en promedio de 0.954 sobre 0.286 del tamaño de los establecimientos, lo que indica que las variables de Participación Económica son significativas para explicar los resultados económicos que el tamaño de las empresas.

Figura 3

Gráfica de Medidas discriminantes entre Dimensión 1 y 2 por categoría (Participación y Tamaño)



Nota. Elaboración propia, mediante SPSS 26.0.

Tabla 11

Medidas discriminantes

	Ponderación de variable	Dimensión		Media
		1	2	
Tamaño	4	0.562	0.009	0.286
Participación	7	0.912	0.997	0.954
Total, Activo	11	8.631	7.017	7.824
% de varianza		78.466	63.788	71.127

Nota. Elaboración propia.

Finalmente, en la tabla del resumen del modelo se nota que la Dimensión 1 tiene una inercia mayor de 78.466 % que la Dimensión 2, lo que nos da una inercia total de 71.127 % de puntos en la nube, de manera que las variables de la Dimensión 1 aportan mayor valor de explicación que en los componentes de la Dimensión 2. Aun así, ambas están correlacionadas.

Tabla 12

Resumen del modelo

Dimensión	Varianza contabilizada para		
	Total (autovalor)	Inercia	% de varianza
1	8.631	0.785	78.466
2	7.017	0.638	63.788
Total	15.648	1.423	
Media	7.824	0.711	71.127

Nota. Elaboración propia.

DISCUSIÓN

Si bien se hicieron algunas pruebas para medir la correlación entre las variables estudiadas, los resultados difieren de coeficiente a coeficiente. Por ejemplo, de acuerdo al criterio de Chi-Cuadrado, el resultado indica que existe una asociación positiva moderada entre el tamaño de establecimientos y participación económica, ya que el coeficiente de asociación de Pearson dio como resultado 0.4685, mientras que el valor V de Cramer de 0.3062. Por otra parte, el resultado del coeficiente de correlación de Mathews es de 0.5303 y el de Inercia Total de 0.2812. Estos resultados sugieren solamente el grado de asociación mas no de causalidad, esto es, que el cambio de una variable no implica necesariamente el cambio en la otra variable (Lind et al., 2012, p. 469). ¿A qué le atribuimos estas diferencias? La realidad es que depende de múltiples factores, por ejemplo, Lagunes-Gómez et al. (2022) sostienen que la información georreferenciada puede presentar heterogeneidad y dependencia espacial (p. 4), lo que significa ausencia de estabilidad en el comportamiento de las variables y que los datos observados en un punto dependen de los valores observados en otro punto.

CONCLUSIONES

Los resultados indican que, si bien los coeficientes no fueron altos, al menos permite identificar el nivel asociación en función del tamaño de las unidades económicas. En el caso de las microempresas están más asociadas con la población ocupada y la producción, mientras que las variables Valor agregado e Ingresos con las medianas y grandes empresas. Estos resultados concuerdan con el planteamiento inicial de correspondencia múltiple y también con la información del Censo Económico, ya que, en México el 94.9 % son micronegocios y emplean el 40 % de la población ocupada, a diferencia de las medianas que emplean el 30.7 % y las grandes empresas el 32.1 %. ¿Cómo explicar esta asociación? Eriksson et al. (2014) hablan de la importancia del clima organizacional y el talento humano para incidir en el crecimiento de las ciudades y de las regiones. Lo que lleva a pensar que las medianas y grandes empresas ocupan mano de obra mejor calificada y disponen de mejores esquemas de organización, que se traduce en mayor valor agregado y de ingresos. Como mencionan Panzera y Postiglione (2022), estas desigualdades entre las

microempresas y la mediana y grandes empresas también se pueden explicar por la desigualdad en la distribución de las actividades económicas geográficamente (p. 687). Es importante considerar también el nivel de integración regional presente en el estado de México. Al respecto vale la pena mencionar la productividad como una forma de integración, como lo señalan Balanzo et al. (2020, p. 6), en el que se tiene vínculos económicos fuertes con una fuerte competencia entre las firmas, pero con fuertes presiones que pueden generar disputas por mano de obra calificada y de financiamiento. De allí la importancia del gobierno para darle mayor impulso al estado, como lo indican Solleiro-Rebolledo et al. (2021, p. 133). Es importante también señalar, como lo mencionan Masson-Guerra y Ortin-Ángel (2017), que el capital invertido se ve afectado por el número de empresas en la región y en el valor agregado generado (p. 2).

El tema no se agosta aquí, ya que queda abierta la posibilidad de investigar la correlación entre unidades económicas por sector de actividad y los resultados generados e incluso entre unidades económicas y principales problemas que se enfrentan, como altos impuestos, inseguridad, competencia desleal, exceso de trámites, etc., que seguramente será tema de otra investigación como el que presenta Carpio (2020, pp 127-128) acerca de la cooperación entre firmas, específicamente en las relaciones de confianza y reciprocidad que le imprimen fortaleza al desarrollo regional. También queda abierta la posibilidad de considerar los resultados del estudio sobre la estructura económica presentados por Cardona Reséndiz et al. (2018, p. 10) de los sectores clave, como el comercio y los servicios inmobiliarios y estratégicos como la manufactura y el transporte. Y también la investigación de Gaspar Sánchez y Orozco Hernández (2015, p. 57) sobre la nueva estructura económica encabezada por el sector comercial y de servicios en las zonas metropolitanas en México; así como las características geográficas de la economía como la discontinuidad, la inestabilidad y la multipolarización, señaladas por Propin Frejomil y Sánchez-Crispín (2001). Cabe destacar que, si se toman en cuenta las consideraciones señaladas anteriormente, las contribuciones a la riqueza del Estado se verán reflejadas en una compensación en la heterogeneidad entre los distintos tipos de tamaño de empresas.

CONTRIBUCIÓN DE LOS AUTORES

Armando Puebla Maldonado: Conceptualización, curación de datos, análisis formal, adquisición de fondos, investigación, metodología, administración del proyecto, recursos, software, supervisión, validación, visualización, redacción: borrador original y redacción: revisión y edición.

FUENTE DE FINANCIAMIENTO

Se utilizaron recursos propios para llevar a cabo el presente estudio.

CONFLICTOS DE INTERÉS

Los autores expresan que no existen conflictos de interés.

AGRADECIMIENTOS

No aplica.

PROCESO DE REVISIÓN

Este estudio ha sido revisado por pares externos en modalidad de doble ciego.

DECLARACIÓN DE DISPONIBILIDAD DE DATOS

Los datos se encuentran alojados junto a los demás archivos de este artículo, también se pueden pedir al autor de correspondencia.

REFERENCIAS

- Alberti, G. (2017). Correspondence Analysis. En S. López Varela (Ed.), *The SAS Encyclopedia of Archaeological Sciences*. Wiley-Blackwell. <https://doi.org/10.1002/9781119188230>
- Aquino Zacarías, V. C., Gómez Villanes, N. I., Porta Chupurgo, R. y Jiménez-Dávalos, J. (2022). Análisis de Correspondencia Múltiple de Sustentabilidad en Unidades Productoras de Tarwi (*Lupinus mutabilis* Sweet) en el valle del Mantaro, Perú. *Idesia (Arica)*, 40(1), 15-23. <https://dx.doi.org/10.4067/S0718-34292022000100015>
- Balanzo, A., Garavito, L., Rojas, H., Sobotova, L., Pérez, O., Guaquetá, D., Mojica, A., Pavajeau, J. y Sanabria, S. (2020). Typical Challenges of Governance for Sustainable Regional Development in Globalized Latin America: A Multidimensional Literature Review. *Sustainability*, 12(7), 1-21. <http://dx.doi.org/10.3390/su12072702>
- Benítez Gaibor, M. K., Martínez, J. P., Margalina, V. M. y Valencia Nuñez, E. R. (2020). Análisis de la estructura de mercado de las sociedades de la cadena productiva de calzado en Ecuador. *Economía: teoría y práctica*, (52), 99-120. <https://doi.org/10.24275/etpuam/ne/522020/benitez>
- Blasius, J. y Schmitz, A. (2015). Correspondence Analysis History of. En J. Wright. (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (Second Edition). Elsevier. <https://doi.org/10.1016/B978-0-08-097086-8.03215-3>
- Cardona Reséndiz, G. A., Cardenete Flores, M. A. y Martínez García, C. I. (2018). Estructura económica mexicana: sectores claves, estratégicos, impulsores e independientes 2012. *Revista de Economía*, 35(90), 9-50. <https://www.redalyc.org/articulo.oa?id=674070982001>
- Carpio, S. (2020). La cooperación entre firmas en el desarrollo regional: una revisión. *Retos Revista de Ciencias de la Administración y Economía*, 10(19), 117-133. <https://doi.org/10.17163/ret.n19.2020.07>
- Colmenares, S. (2017). Análisis Multivariable en Geografía Aplicado al Estudio de la Pobreza. *Terra. Nueva Etapa*, 23(53), 41-73. <https://www.redalyc.org/articulo.oa?id=72152384003>
- Contreras-Pacheco, O. E., Talero-Sarmiento, L. H. y Escobar-Rodríguez, L. Y. (2020). Sostenibilidad, stakeholders y crisis de empresa: un análisis estructurado de percepciones. *Suma de Negocios*, 11(24), 64-72. <https://doi.org/10.14349/sumneg/2020.v11.n24.a7>
- Eriksson, R. H., Hansen, H. K. y Lindgren, U. (2014). The Importance of Business Climate and People Climate on Regional Performance. *Regional Studies*, 48(6), 1135-1155. <https://doi.org/10.1080/00343404.2013.770140>
- Fernández, F. J. (2002). El uso del Análisis de Correspondencia Simple (ACS) como ayuda en la interpretación del dato en arqueología. Un caso de estudio. *Boletín Antropológico*, 20(55), 687-713. <https://www.redalyc.org/articulo.oa?id=71205505>

- Gaspar Sánchez, N. y Orozco Hernández, M. E. (2015). Configuración terciaria de la zona metropolitana de Toluca. *Bitácora Urbano Territorial*, 25(1), 57-66. <https://doi.org/10.15446/bitacora.v1n25.40184>
- Gaxiola Robles Linares, S. C., Lozano Keymolen, D., Montoya Arce, B. J. y Román Sánchez, Y. G. (2017). Situación de la seguridad social de los trabajadores asalariados de 50 a 59 años en el Estado de México, 2015. *Papeles de Población*, 23(94), 197-221. <https://doi.org/10.22185/24487147.2017.94.036>
- Gaxiola-Robles Linares, S. C., Lozano-Keymolen, D. y Montoya-Arce, B. J. (2020). Desigualdad en hogares con adultos mayores en México, 2015. *Papeles de Población*, 26(103), 233-265. <https://doi.org/10.22185/24487147.2020.103.09>
- Ghani, E., Kerr, W. R. y O'Connell, S. (2014). Spatial Determinants of Entrepreneurship in India, *Regional Studies*, 48(6), 1071-1089. <https://doi.org/10.1080/00343404.2013.839869>
- Instituto Nacional de Estadística y Geografía (INEGI) (2019). *Características de los Negocios, Censos Económicos 2019*. INEGI. https://www.inegi.org.mx/contenidos/programas/ce/2019/doc/frrdf_ce19.pdf
- Jiménez Restrepo, D. M. y Páez Cortés, J. N. (2014). Una metodología alternativa para medir la calidad del empleo en Colombia (2008-2012). *Sociedad y Economía*, (27), 129-154. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1657-63572014000200006&lng=en&tlng=es
- Jo, Y. y Lee, C. Y. (2014). Technological Capability, Agglomeration Economies and Firm Location Choice. *Regional Studies*, 48(8), 1337-1352. <https://doi.org/10.1080/00343404.2012.711946>
- Krugman, P. (2011). The New Economic Geography, Now Middle-aged, *Regional Studies*, 45(1), 1-7. <https://doi.org/10.1080/00343404.2011.537127>
- Lagunes-Gómez, I., Hernández-Orduña, M. G., Murrieta-Galindo, R., Hernández-Pitalua, D. y Mayorga-Cruz, D. (2022). Spatial Analysis of the Empirical Behavior of Municipal Institutional Capacity for the Formulation of Sustainable Growth Management Strategies with a Regional Focus: State of Veracruz, Mexico. *Sustainability*, 14(4), 1-21. <http://dx.doi.org/10.3390/su14042000>
- Leyva Jiménez, A. (2019). *Aplicación de la Minería de Datos en Geografía* [Tesis para obtener el grado de Licenciada, Universidad Nacional Autónoma de México]. Repositorio Institucional de la UNAM. https://ru.dgb.unam.mx/handle/DGB_UNAM/TES01000796601
- Lind, D.A., Marchal, W.G. y Wathen, S. A. (2012): *Estadística aplicada a los negocios y a la economía*. McGrawHill (15^a ed.). <http://up-rid2.up.ac.pa:8080/xmlui/handle/123456789/1542>
- Massón-Guerra, J. L. y Ortín-Ángel, P. (2017). Regional entrepreneurship capital and firm production. *Small Business Economics*, (49), 595-607. <http://dx.doi.org/10.1007/s11187-017-9851-0>

- Montes Galbán, E. (2015). Estructura Diacrónica de los Procesos de investigación Aplicada a la Geografía Cuantitativa. En G. Buzai, G. Cacace, L. Humacata y S. Lanzelotti, *Teoría y métodos de la geografía cuantitativa: Libro 1: Por una geografía de lo real*. MCA Libros. https://www.researchgate.net/profile/Gustavo-Buzai/publication/294572996_Teoria_y_metodos_de_la_Geografia_Cuantitativa_Libro_1_por_una_Geografia_de_lo_real/links/56c1bd3308ae2f498efc4a2c/Teoria-y-metodos-de-la-Geografia-Cuantitativa-Libro-1-por-una-Geografia-de-lo-real.pdf
- Moreno Cuello, J. L. y Ramos Camargo, J. L. (2013). Factores determinantes de la creación de empresas de base tecnológica en Colombia. *Revista de Economía del Caribe*, (12), 1-28. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S2011-21062013000200001&lng=en&tlng=es.
- Panzer, D. y Postiglione, P. (2022). The impact of regional inequality on economic growth: a spatial econometric approach. *Regional Studies*, 56(5), 687-702. <https://doi.org/10.1080/00343404.2021.1910228>
- Pérez Herver, M. J. y Zapata Sánchez, J. L. (2019). Análisis de las prácticas de responsabilidad social empresarial en las mipymes del sur de Quintana Roo a través de técnicas multivariantes. *Oikos Polis*, 4(1), 33-76. http://www.scielo.org.bo/scielo.php?script=sci_arttext&pid=S2415-22502019000100004&lng=es&tlng=es.
- Propin Frejomil, E. y Sánchez-Crispín, Á. (2001). Características básicas de la estructura territorial de la economía mexicana. *Investigaciones geográficas*, (46), 148-163. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0188-46112001000300011&lng=es&tlng=es.
- Rangel Quiñonez, H. S. y Yáñez Canal, G. (2018). Clasificación por capitales de una muestra de microempresarios del Área Metropolitana de Bucaramanga a partir del Análisis de Correspondencia Múltiple. *Ensayos de Economía*, 28(53), 165-192. <https://doi.org/10.15446/ede.v28n53.75018>
- Sánchez Olaya, D. M., Velandia Tibáquira, O. G. y Suárez Salazar, J. C. (2015). Contribución de sistemas productivos en la generación de ingresos en familias cacaoteras, departamento del Caquetá. *Revista de Ciencias Agrícolas*, 32(1), 37-54. <https://doi.org/10.22267/rcia.153201.23>
- Signorelli, S. y Biffignandi, S. (2018). From Big Data to Information: Statistical Issues Through a Case Study. En F. Mola, C. Conversano y M. Vichi (Eds.), *Classification, (Big) Data Analysis and Statistical Learning*. Springer International Publishing. https://doi.org/10.1007/978-3-319-55708-3_1
- Solleiro-Rebolledo, J. L., Castañón-Ibarra, R. y Martínez-Salvador, L. E. (2021). Governance and Critical Factors to Implement Smart Specialization Strategies in Developing Countries. The Case of the State of Mexico, México. *Revista De Economía Mundial*, (57), 119-136. <https://doi.org/10.33776/rem.v0i57.4589>

- Trejo Nieto, A. B. y Negrete, M. E. (2018). Dinámicas productivas y agrupamientos industriales en la región centro de México. *Región y sociedad*, 30(73), 1-35. <https://doi.org/10.22198/rys.2018.73.a905>
- Tufféry, S. (2011). *Data mining and statistics for decision making*. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470979174>
- Valencia Núñez, E. R., Valle Álvarez, A. T., Cruz Lascano, M. E. y Haro Sarango, A. F. (2022). Evaluación de la sostenibilidad financiera en el Crédito de Desarrollo Humano Asociativo en la provincia de Tungurahua, Ecuador. *Lecturas de Economía*, (97), 325-368. <https://doi.org/10.17533/udea.le.n97a346723>
- Valverde, C., Tamayo, L. M. y Vargas Chanes, D. (1994). Aplicación del análisis de correspondencias en el estudio de la vivienda. *Investigaciones geográficas*, (28), 129-150. https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0188-46111994000100004