

FACULTAD DE CIENCIAS ECONÓMICAS

**TESIS EN OPCIÓN AL TÍTULO ACADÉMICO DE
MÁSTER EN MATEMÁTICA APLICADA**

**Análisis Exploratorio de Datos Espaciales para la identificación de
patrones socio-demográficos en estudios de población. Provincia
Cienfuegos**

Autor: Domingo Valladares Pérez

Tutores: DrC. Elia Natividad Cabrera Álvarez

MsC. Miguel Santana Justiz

Resumen

Impulsar el desarrollo de los territorios a partir de la estrategia del país, contribuye al fortalecimiento de los municipios y a la reducción de desproporciones. En tal empeño, para la Oficina Nacional de Estadística e Información (ONEI), es útil y necesario disponer de una herramienta capaz de procesar estadísticamente grandes volúmenes de datos sobre la dinámica poblacional de cada espacio del país.

Es propósito de este estudio, elaborar un procedimiento estadístico para la identificación de patrones de asociación espacial en los variables socio-demográficos de la población de Cienfuegos. Dicho procedimiento se distingue por la combinación del análisis estadístico clásico con el Análisis Exploratorio de Datos Espaciales desde la perspectiva de la econometría espacial, a partir de variables e indicadores asociados a la dinámica poblacional. Se enfatiza en la autocorrelación espacial sobre la base del GeoDa y la metodología propuesta por Luc Anselin con modificaciones aportadas en esta investigación.

Los resultados sintetizan las 3 etapas y 7 pasos del procedimiento, tomando como base informativa de referencia el Censo de Población y Viviendas 2012. Se identificaron los patrones de asociación de 8 variables sociodemográficas para la provincia Cienfuegos, representativas de los coeficientes de dependencia y vejez, las tasas de actividad y ocupación, el índice de masculinidad, el estado civil, la educación y el color de la piel.

La propuesta puede ser generalizada a otras provincias del país y constituye un resultado pertinente para la actualización de la estrategia de desarrollo de Cienfuegos.

Índice

Contenido

Introducción	1
Capítulo 1. Fundamentos teóricos de la econometría espacial. Aplicaciones a estudios de población.	7
1.1. Introducción.....	7
1.2 Efectos Espaciales	8
1.2.1 Dependencia Espacial	8
1.2.2. Heterogeneidad espacial	11
1.3. Incorporando la Dependencia Espacial en el Modelo Econométrico.....	13
1.3.1 Matriz de Contigüidad.	14
1.3.2. Detección Global de Dependencia Espacial Univariante.....	18
1.4. Análisis exploratorio de datos espaciales	21
1.5. Estudios de población. Surgimiento y desarrollo	23
1.5.1 Evolución de los estudios sociodemográficos.	25
1.5.2 Estadísticas y su vínculo con los estudios de población.....	25
1.5.3 Distribución territorial de la población.....	27
1.5.4 El tiempo en Demografía. El Diagrama de Lexis.	30
1.5.5 Razones, proporciones y tasas.	31
Capítulo 2 Procedimiento para la identificación de patrones de asociación espacial	33
2.1 Antecedentes metodológicos.....	33
2.1.1 Procedimiento para la identificación de patrones de asociación espacial.....	35
2.2 Etapa 1: Reconocimiento del Territorio.....	35
2.3 Etapa 2. Análisis Exploratorio de Datos Espaciales.....	42
2.4 Etapa 3. Análisis de Resultados	54
Capítulo 3. Aplicación del procedimiento para la identificación de patrones de asociación espacial.	56
3.1 Introducción.....	56
3.2 Etapa 1 Reconocimiento del territorio.....	56
3.3 Etapa 2. Análisis Exploratorio de Datos Espaciales.....	61
3.4 Etapa 3. Análisis de Resultados	72
Conclusiones	74
Recomendaciones	75
Referencias Bibliográficas	76
Anexos	81
Anexos 1 Mapa conceptual sobre el análisis exploratorio de datos.....	81
Anexo 2. Resumen de Estadísticos descriptivos de AED.	82

Introducción

Los estudios relacionados con la dimensión espacial del desarrollo contribuyen a solucionar problemáticas locales potenciando los recursos endógenos de la localidad, a la vez que favorecen su autonomía y con ello las condiciones de vida de la sociedad. En este proceso es esencial la participación de los habitantes como actores fundamentales del cambio y la transformación con la finalidad de dinamizar el sistema para el bien colectivo.

Según Rabelo (2006), el desarrollo en su visión territorial presenta dos enfoques. El primero evade el espacio geográfico y centra su análisis en el crecimiento de la economía teniendo en cuenta las leyes del mercado como el encargado de asignar eficientemente los recursos, siendo este más factible para países desarrollados. Contrapuesto a este se presenta la segunda visión acerca de las disparidades territoriales que requieren medidas económicas, políticas y sociales con la intervención directa del Estado, adecuándose más a las características de los países subdesarrollados considerando sus disímiles espacios geográficos. Se reconoce la importancia de analizar el desarrollo en los espacios locales dándose un cambio significativo en la teoría del desarrollo.

Los recursos humanos son fundamentales en el desarrollo de cualquier actividad, por eso es importante tener en cuenta la dinámica poblacional, en el caso de la cubana esta ha tenido lugar en el contexto de una dinámica y estructura demográfica particular y abundante en desafíos, caracterizada por un “creciente proceso de envejecimiento de su población, resultado de un contexto donde los niveles de fecundidad están debajo del reemplazo durante más de tres décadas, un saldo migratorio externo sostenidamente negativo, así como niveles bajos de mortalidad” (Aja Díaz, 2015).

En la actualidad resulta necesario que los responsables de la formulación de políticas fundamenten su pensamiento estratégico en el conocimiento de la Demografía (Araujo González, 2016). Otras experiencias en el contexto internacional como es el caso de la expuesta por Burriel en España, hacen referencia al uso de la Demografía en la planificación territorial, o la propuesta de observatorios locales para relacionar la información sociodemográfica con la planificación y la participación ciudadana (de Orueta, 2003) .

Para el escenario demográfico de Cuba, no se trata solo de querer utilizar información demográfica en la gestión del desarrollo local, sino que es esencial y necesario que las políticas públicas accionen sobre esa realidad y mitiguen efectos negativos para recuperar tendencias más favorables (Aja Díaz, 2015). Gestionar el desarrollo local al margen de lo que la situación sociodemográfica sería como andar a ciegas sin comprender las tendencias de desarrollo de los territorios y la multiplicidad de hechos demográficos muchas veces inadvertidos en el maremágnum de procesos sociales que tienen lugar (Araujo González,

2016). Otros puntos de vista hacen alusión al escaso tiempo dedicado por los gestores locales a las proyecciones poblacionales y a los recursos invertidos, basando sus estimaciones en datos incompletos o construyendo indicadores con datos censales desactualizados (Torres, 2006).

Se refiere al conocimiento suficiente a nivel de los decisores de problemas, tales como el nivel del embarazo adolescente, las tendencias del índice de dependencia o los impactos que el proceso de envejecimiento tiene y los modos de abordaje adecuados en su desenvolvimiento. Del mismo modo, los datos demográficos son importantes en el desarrollo y proyección de planes educativos o sanitarios (de Orueta, 2003), así como en la proyección de proyectos productivos o de servicios sociales. La utilización eficiente y eficaz de la información sociodemográfica requiere la preparación del personal especializado (Phélan, 2007) capaz de ejecutar y realizar los análisis demográficos necesarios para la interpretación y traducción de estos en acciones y proyecciones necesarias para el desarrollo local. A ello ha contribuido considerablemente en el caso cubano, el papel del Centro de Estudios Demográficos de Estudios Demográficos de la Universidad de La Habana (CEDEM-UH) (Barros & Aroca, 2014).

Los datos sociodemográficos nacionales que se obtienen en los censos de población y vivienda, resultan muchas veces insuficientes cuando se intenta su extrapolación a la escala local. Los indicadores a nivel local en varias ocasiones tienen que ser calculados, con la dificultad añadida que representa la necesidad de una base de cálculo diferenciada. Estas bases de cálculos nacionales se basan en denominadores mayores que las áreas locales. Tal es el caso del indicador mortalidad infantil calculado sobre mil nacidos vivos en un municipio donde no nacen mil niños o el de la mortalidad materna, que se calcula sobre cien mil nacidos vivos en un territorio donde tampoco nacen cien mil.

En esos casos, los resultados que se obtienen tienen un sesgo, siendo necesario aplicar otros cálculos propios de la metodología utilizada por la Demografía para áreas pequeñas. Otra problemática radica en determinar el tipo de indicadores que es necesario levantar o calcular cuando en determinados hechos demográficos se mezclan elementos demográficos y administrativos, como pueden ser, por ejemplo, la población en edad laboral por grupos de edades y las necesidades laborales por sectores o la proyección de niños de las primeras edades en correspondencia con la fuente de personal pedagógico calificado para el proceso educativo y las disponibilidades materiales en las escuelas.

Ante esta problemática, es aconsejable la elaboración de cartografías de las áreas pequeñas en Sistemas de Información Geográfica (SIG), de manera que permitan establecer en un mismo sistema de información, datos de diferentes orígenes (Araujo, 2017). En última instancia, de lo que se trata es de identificar de acuerdo a las características propias de las

áreas locales, los elementos que han de constituirse en indicadores específicos del nivel local y determinar sus bases de cálculo para la construcción de tales indicadores. Esta tarea permitirá erigir la base sobre la que se fundamente la construcción de los sistemas de información a nivel local y territorial, que al mismo tiempo constituyan el fundamento para la participación ciudadana de decisores, instituciones e individuos que formen parte del territorio correspondiente.

En los últimos años, los avances en los medios de cómputo han permitido el desarrollo y la implementación de la Econometría Espacial, que proporciona las técnicas de contrastación y de estimación necesarias para trabajar con datos que presentan problemas de heterogeneidad y/o dependencia espacial, mientras los SIG, permiten un amplio abanico de operaciones basadas en su aproximación a la presentación.

La Oficina Nacional de Estadística e Información (ONEI), dispone de un gran volumen de información recopilada en múltiples variables e indicadores relacionados con diversas características de cada espacio del país. Estos datos permiten la realización de una caracterización adecuada de la población y su integración a un SIG. Su estudio tributa a la posible implementación de políticas atemperadas al contexto territorial.

En la provincia Cienfuegos se emprenden múltiples proyectos de desarrollo local que repercuten de forma efectiva en el sector de la Educación, la Salud, la Agricultura y el Turismo, entre otros. Por tal razón, la Universidad de Cienfuegos, ha venido incursionando en la temática de la dinámica de población en aras del desarrollo local, realizando diferentes investigaciones en los ocho municipios que conforman la provincia. Disímiles enfoques han predominado en estas investigaciones. Propiamente para la dimensión demográfica, se han realizado estudios para el espacio geográfico de municipios de Cienfuegos, donde se construye una conciencia socioespacial como herramienta de trabajo para los gobiernos (Iglesias, et al., 2022). Adicionalmente, se ha enfatizado en el tratamiento de la información con particularidades para el análisis de indicadores sociodemográficos y el uso de métodos estadísticos multivariados en la construcción de un Índice de Desarrollo Municipal (Becerra, 2004), además se realizó un estudio sobre el bienestar de las familias en Cienfuegos (Mazaira, 2011) donde se determinó un índice sintético a partir de múltiples indicadores sociodemográficos teniendo como sustento el Análisis de componentes principales.

En el contexto montañoso de la provincia se han realizado estudios desde el área temática sociodemográfica para elaborar un índice de Desarrollo Sostenible, incursionado con métodos estadísticos de naturaleza descriptiva, (Díaz Gispert, 2011) y, quien propone modelos estadísticos matemáticos de naturaleza multivariada para evaluar el desarrollo sostenible por dimensiones. En el mismo contexto se han construido escenarios con horizonte 2025,

articulando la Prospectiva con los métodos estadísticos multivariados, a partir del Análisis Exploratorio de Datos para varias dimensiones del desarrollo sostenible, entre ellas, la dimensión social. (Cabrera, 2016).

Recientemente se concluyeron investigaciones que combinan los estadísticos globales y locales de autocorrelación espacial y técnicas estadísticas de clasificación multivariada, en el macizo Montañas de Guamuhaya a nivel de asentamientos (Conde, 2020). Adicionalmente se realizó un estudio sobre la base de técnicas multivariadas de clasificación para la identificación de patrones de asociación espacial en la población urbana envejecida del municipio de Cienfuegos (Rodríguez, 2020).

Sin embargo, ninguna de las investigaciones anteriores cuenta con una herramienta capaz de articular el análisis exploratorio de datos espaciales en profundidad, abarcando entre otros aspectos, la identificación de atípicos, la distribución espacial, y la detección de efectos especiales para cualquier unidad geográfica, las técnicas multivariadas de clasificación y con soporte en SIG

Ante esta situación, se enuncia el siguiente **problema científico**:

No son suficientes los estudios de econometría espacial para el análisis de indicadores sociodemográficos de la población Cienfuegos, lo que impide la identificación de patrones socioespaciales, de su comportamiento y limitan el desarrollo de políticas gubernamentales más efectivas para el tratamiento diferenciado a este sector.

Objeto de la investigación: Los métodos del análisis espacial.

Campo: Los métodos y técnicas del Análisis Exploratorio de Datos Espaciales (AEDES) para el análisis espacial en la provincia Cienfuegos.

La investigación persigue como **objetivo general**:

Elaborar un procedimiento estadístico para la identificación de patrones de asociación espacial en la dimensión socio-demográfica de la población de Cienfuegos, que contribuya con la actualización de su estrategia de desarrollo.

Objetivos Específicos:

1. Caracterizar la evolución de la econometría espacial, las técnicas y métodos para el análisis de datos espaciales y su aplicación en los estudios de población.
2. Identificar la viabilidad de la articulación entre los estadísticos globales y locales de asociación del AEDE y las técnicas estadísticas multivariadas para el reconocimiento de relaciones en los indicadores socio-demográficos de la provincia Cienfuegos.

3. Elaborar las etapas y pasos del procedimiento para la identificación de patrones de asociación espacial en la dimensión sociodemográfica de Cienfuegos
4. Identificar los patrones de asociación espacial de los indicadores sociodemográficos al nivel de Consejos Populares de la Provincia Cienfuegos.

Hipótesis de la Investigación

Si se elabora un procedimiento estadístico para la dimensión socio-demográfica, donde se combinen los estadísticos globales y locales de asociación del AEDE y las técnicas estadísticas multivariadas, entonces podrán identificarse los patrones de asociación espacial de la provincia Cienfuegos, contribuyendo a la actualización de su estrategia de desarrollo.

Variables:

- **Variable independiente:** La articulación de los métodos estadísticos espaciales y los indicadores de la dimensión socio-demográfica.
- **Variable dependiente:** Patrones de asociación espacial

Aporte Práctico:

El procedimiento elaborado contribuye con la línea de estudios demográficos del Proyecto Asociado a Programa Nacional, “Desarrollo de nuevos modelos y métodos matemáticos para la toma de decisiones”, de la Facultad de Matemática y Computación, (Universidad de La Habana) donde también participa la Universidad de Cienfuegos. Específicamente contribuye con la tarea: Elaboración de algoritmos para caracterizar el comportamiento espacial y/o temporal de una población y la aplicación de los resultados para detectar segregación en la estructura de una ciudad. El procedimiento resultante es flexible y contribuye con la política para impulsar el desarrollo territorial y sus resultados son un importante aporte para la actualización de las estrategias de desarrollo a escala municipal y provincial.

Métodos utilizados en la investigación

Se utilizan métodos del nivel teórico tales como el Histórico-lógico para la investigación de los antecedentes en el contexto internacional, nacional y local, así como la evolución de la Econometría espacial en el desarrollo y la fundamentación del problema científico; Analítico-Sintético para el estudio del origen y evolución del objeto de estudio, el tratamiento de la información y su reducción; Inductivo-deductivo, para establecer las etapas y pasos del procedimiento; Sistémico, para fundamentar y establecer los nexos entre las etapas propuestas; y el de Modelación. De los métodos empíricos se utiliza el Análisis de documentos. Se emplean además métodos estadísticos y econométricos que contribuyen al

objetivo propuesto, entre ellos el Análisis Exploratorio de Datos Espaciales (AEDE); los estadísticos globales y locales de autocorrelación espacial y el Análisis de Clúster.

El procesamiento de la información se realiza con el software GeoDa.

Estructura de la Tesis.

La tesis se estructura en Introducción, tres capítulos, conclusiones y recomendaciones, bibliografía y anexos.

Capítulo 1: Fundamentos teóricos de la econometría espacial. Aplicaciones a estudios de población. Se exponen en este capítulo los principales aspectos teóricos relativos a la Econometría Espacial, las técnicas y métodos para el análisis de datos espaciales y su aplicación en los estudios de población.

Capítulo 2: Procedimiento para la identificación de patrones de asociación espacial. En este capítulo se desarrollan las etapas y pasos para la identificación de dichos patrones, articulando métodos estadísticos y los estadísticos globales.

Capítulo 3: Resultados de la identificación de patrones de asociación espacial en los Consejos Populares de la provincia Cienfuegos como unidad geográfica seleccionada para demostrar la viabilidad del procedimiento y los resultados obtenidos.

CONCLUSIONES Y RECOMENDACIONES

Las conclusiones y recomendaciones resultan de los principales resultados obtenidos en la investigación, la que se apoya en un grupo de anexos que son importantes para profundizar en los análisis realizados desde el punto de vista descriptivo y gráfico.

Capítulo 1. Fundamentos teóricos de la econometría espacial. Aplicaciones a estudios de población.

1.1. Introducción

La econometría espacial es una disciplina de la econometría general que incluye el conjunto de técnicas de especificación, estimación, contraste y predicción necesarias para el tratamiento de los datos espaciales. En palabras de Anselin (2001), podría definirse como “la parte de la econometría que se ocupa del tratamiento de la interacción espacial (dependencia espacial) y la estructura espacial (heterogeneidad espacial) en los modelos de regresión de corte transversal y de datos de panel”.

La econometría espacial es un sub-campo dentro de la econometría donde se incorpora el tratamiento de los efectos espaciales y los problemas que los mismos provocan en los modelos econométricos que utilizan datos de corte transversal y de panel. Esta disciplina ha tenido un importante crecimiento en las últimas décadas debido a tres principales razones (Anselin & Florax, 1995):

- La importancia del espacio geográfico dentro de la teoría económica para explicar aspectos como los rendimientos crecientes y las economías de aglomeración. Específicamente, la Nueva Geografía Económica (Fujita & Krugman, 2004) ha reconocido la importancia de la localización como factor influyente en la determinación de las variables económicas.
- La existencia de grandes bases de datos socio-económicos georeferenciados como REIS en Estados Unidos, sostenida por el departamento de Análisis Económico (BEA), y REGIO en Europa, sostenido por el departamento estadístico de la Comisión Europea (Eurostat).
- El avance tecnológico en el manejo de grandes bases de datos con información espacial: Sistemas de Información Geográfica y programas específicos para manipular y modelizar estos datos

Los mencionados avances han generado que la econometría espacial adquiera especial relevancia dentro de las principales líneas de investigación econométrica.

La econometría espacial es la parte de la econometría que se dedica al estudio de los fenómenos económicos espaciales. Los fenómenos económicos espaciales pueden definirse como aquellos en los que la variable espacio, entendida en sentido amplio y definida luego según una matriz de contacto, desempeña un papel importante que su exclusión podría provocar modelos econométricos con severos errores de especificación. Es decir, esta disciplina integra al estudio econométrico la influencia del espacio en el fenómeno de estudio. (Alonso, 2016), (Acevedo & Velázquez, 2008)

1.2 Efectos Espaciales

La información obtenida para uso de la ciencia regional posee características que provienen de su ubicación geográfica. Estas características se han denominado efectos espaciales, los cuales pueden ser divididos en dos tipos: dependencia espacial y heterogeneidad espacial. De los dos, la dependencia espacial, también denominada, autocorrelación espacial, es la más conocida. Esta dependencia puede ser expresada según la primera ley de la geografía de Tobler (1979), en la cual “todo está relacionado con todo lo demás, pero las cosas cercanas están más relacionadas que las cosas distantes”.

El segundo tipo de efecto espacial, la heterogeneidad espacial, está relacionada con la ausencia de estabilidad en el comportamiento o las relaciones bajo estudio. Más precisamente, esto implica que los parámetros y formas funcionales varían con la ubicación y no son homogéneos en los conjuntos de datos.

En resumen, el campo de la econometría espacial consiste de aquellos métodos y técnicas que, basados en una representación formal de la estructura espacial de dependencia y heterogeneidad, proveen los medios de realizar una especificación, estimación, prueba de hipótesis y predicción adecuados.

1.2.1 Dependencia Espacial

La dependencia espacial en una colección de observaciones de datos muestrales se refiere al hecho de que una observación asociada con una localización la cual se puede denominar i , depende de otras observaciones asociadas con localizaciones $j \neq i$. Esto puede ser formalmente expresado como:

$$y_i = f(y_j) \quad \text{para } i = 1; \dots; n; j \neq i \quad (1)$$

Donde cada observación de una variable y en $i \in S$ (con S como el conjunto que contiene todas las unidades espaciales de observación), está relacionada formalmente a través de la función f , con las magnitudes de la variable en otra unidad espacial en el sistema.

Anselin (2001) define este fenómeno como “coincidencia entre valores similares (de una variable) y localizaciones cercanas”. Aunque la dependencia espacial puede parecer similar a la dependencia en las series temporales, ya se ha indicado que este parecido sólo es real en parte, debido a la naturaleza multidireccional de la dependencia en el espacio frente a la clara situación unidireccional del tiempo.

Dos tipos de condiciones podrían conducir a una situación como la descrita.

- La primera puede ser producto de los errores de medición para las observaciones en unidades espaciales contiguas. Ejemplos de estos problemas son: La delimitación arbitraria de las unidades espaciales de observación, las cuales no recogen adecuadamente el proceso generador de los datos muestrales.
- La segunda está relacionada con problemas de agregación espacial, ya que, en muchas situaciones encontradas en la práctica, la información es recolectada sólo a una escala agregada, por lo tanto, podría haber muy poca correspondencia entre el alcance del fenómeno agregado bajo estudio y la delimitación de las unidades espaciales de observación.

La dependencia espacial ha sido un tema mucho más estudiado que la heterogeneidad espacial, desde los primeros trabajos estadísticos de Cliff y Ord (1973), referido en (Yrigoyen, 2003). Por su parte, Cliff y Ord (1981) presentan el análisis de la dependencia espacial como una técnica eficaz para el conocimiento de las causas y formas de propagación de epidemias y enfermedades. Además, este ha sido objeto de aplicaciones interesantes en otros campos, como: los recursos naturales y medio ambiente (Anselin, 2001), delincuencia (Baller, et al., 2006; Kubrin, 2003), actividad de I+D (Anselin, 2000; Varga, 2000).

Cuando una variable se distribuye de forma sistemática en un espacio geográfico se dice que el fenómeno en cuestión presenta dependencia, comúnmente denominada autocorrelación espacial, para lo que se han desarrollado diversas medidas o contrastes estadísticos que permiten determinar si la estructura espacial definida por una variable es o no significativa, es decir, si merece o no la pena su conocimiento y estudio.

Definida de manera simple, la autocorrelación espacial es la concentración o dispersión de los valores de una variable en un mapa. Dicho de otra manera, la autocorrelación espacial refleja el grado en que objetos o actividades en una unidad geográfica son similares a otros objetos o actividades en unidades geográficas próximas (Goodchild, 1987).

El efecto de autocorrelación espacial puede ser de signo positivo o negativo, así como nulo.

- ✓ Se entiende por **autocorrelación espacial positiva** el fenómeno de asociación entre valores similares de una variable y localizaciones cercanas; es decir, cuando, en el espacio geográfico, los valores altos de una variable están rodeados por valores altos y viceversa. Éste sería el caso, del llamado efecto contagio o desbordamiento (“spillover”) que se produce en muchos fenómenos socioeconómicos de renta y desarrollo humano, en general, en los que su presencia en una región es causa de su extensión a regiones vecinas, favoreciendo la concentración del fenómeno en la zona.
- ✓ Por el contrario, existe **autocorrelación espacial negativa** en un espacio cuando los valores altos de una variable se encuentran rodeados por valores bajos de la misma, y

viceversa, como ocurre con la disposición de las casillas blancas y negras en un tablero de ajedrez. Esta configuración, en la que se produce una mayor disimilitud entre unidades geográficas cercanas que entre las lejanas es, por ejemplo, la que se produciría en fenómenos de jerarquías espaciales del tipo centro-periferia, en la que los centros de mayor actividad económica están rodeados de regiones periféricas próximas con baja actividad.

- ✓ Por último, se produce **ausencia de autocorrelación espacial** en una variable geográfica cuando ésta se distribuye de manera aleatoria sobre el espacio.

Sin embargo, no siempre que se produce un efecto de autocorrelación espacial en una variable será consecuencia de la existencia de una interacción espacial entre regiones, como los citados fenómenos de desbordamiento y jerarquías espaciales. Algunas veces, las variables espaciales pueden estar correlacionadas de forma espuria o falsa.

Según Anselin (1988), la aparición de dependencia espacial en una variable puede ser debida a una falsa causa, como ciertos errores de medida, y a verdaderos fenómenos de interacción espacial.

En muchas ocasiones, los datos son recogidos de la realidad en forma más o menos agregada, por lo que puede haber poca correspondencia entre el ámbito del fenómeno en estudio y el nivel de agregación de las unidades espaciales en observación. Es probable que, por este motivo, se produzcan ciertos errores de medida. Es más, estos errores tienden a ir más allá de los límites de las propias unidades espaciales afectadas; es decir, que los errores cometidos en el dato correspondiente a la región *i*, probablemente se transferirán a los errores de la región vecina *j*. Este trasvase espacial de los errores de medida es una de las causas obvias de existencia de dependencia espacial. En un contexto de análisis de regresión, esto puede fácilmente producir perturbaciones aleatorias no esféricas, así como problemas de error en las variables.

Ésta sería la verdadera causa de autocorrelación espacial, positiva o negativa, en los fenómenos propios de la ciencia regional y la geografía humana, en los que, como ya se ha indicado, la localización y la distancia son fundamentales a veces, porque dan lugar a una gran variedad de elaciones de interdependencia tanto en el espacio como en el tiempo. Las teorías de interacción espacial, los modelos de difusión y las jerarquías espaciales se han formulado para expresar matemáticamente estas relaciones de dependencia existentes entre diferentes puntos o regiones del espacio geográfico.

1.2.2. Heterogeneidad espacial

La heterogeneidad espacial surge cuando se trabaja con unidades espaciales (países, regiones, provincias, municipios, secciones censales) en las que un fenómeno se distribuye de manera distinta sobre el espacio. Por eso, este efecto espacial suele estar directamente relacionado con la localización geográfica, el área o cualquier otra característica de las unidades espaciales muestrales (Anselin, 1988a; Moreno & Vayá, 2002). El término heterogeneidad espacial se refiere a la variación en las relaciones sobre el espacio. Formalmente una relación lineal se representa como:

$$y_i = X_i\beta_i + \epsilon_i \quad i = 1, \dots, n \quad (2)$$

Según Anselin (2010), la heterogeneidad espacial puede ser definida como “inestabilidad estructural en forma de varianza no constante de los residuos de una regresión (heteroscedasticidad) o en los coeficientes del modelo, que es posible abordar mediante técnicas de econometría tradicional o con herramientas propias de econometría espacial”. Hay tres razones, como indica Anselin (2010), para analizar este efecto de heterogeneidad a través de técnicas propias de econometría espacial:

- ✓ En primer lugar, la estructura que subyace en la inestabilidad espacial es de carácter geográfico, en el sentido de que la localización de las observaciones es fundamental para determinar la forma o especificación de dicha variabilidad. Este sería, por ejemplo, el caso de la heteroscedasticidad de grupos (“groupwise”), que podría ser modelada a través de tantos valores de la varianza de la perturbación aleatoria como distintos grupos geográficos compactos puedan derivarse de los datos.
- ✓ En segundo lugar, dado que la estructura es espacial, la heterogeneidad suele producirse conjuntamente con el problema de autocorrelación espacial, siendo inadecuadas las técnicas de la econometría tradicional, dado que los contrastes habituales de heteroscedasticidad pueden estar sesgados en un contexto espacial.
- ✓ Por último, en un modelo de regresión de corte transversal, ambos efectos de autocorrelación y heterogeneidad espacial pueden ser, desde una óptica meramente observacional, totalmente equivalentes. Así, por ejemplo, un “cluster” o agrupamiento espacial (observado en localizaciones muy próximas) de los residuos con valores extremos podría ser interpretado como un problema de heterogeneidad espacial (heteroscedasticidad de grupos o “groupwise”), o también como un efecto de autocorrelación espacial. Por eso, deben estructurarse perfectamente ambos efectos espaciales para identificar correctamente los parámetros de un modelo con estos problemas y no considerar un aspecto independientemente del otro.

De la citada definición de Anselin (2010), se deduce que en un modelo de regresión lineal pueden distinguirse diversas especificaciones para el efecto de heterogeneidad espacial, según que se manifieste como heteroscedasticidad o como inestabilidad estructural paramétrica.

Especificación de la heteroscedasticidad espacial

La heteroscedasticidad consiste en la ausencia de estabilidad en la dispersión de un fenómeno, como sucede muchas veces con los residuos de una regresión y puede representarse como:

$$\text{Var}(u_1) = \sigma_i^2 \quad (3)$$

Donde σ_i^2 : indica que la varianza de la perturbación aleatoria es diferente para cada observación muestral i . Las causas de la existencia de heterogeneidad espacial en un modelo de regresión serían las siguientes:

- ✓ Utilización de datos procedentes de unidades espaciales irregulares, es decir, con diferente área o extensión territorial, como es el caso de las divisiones político administrativas (países, regiones, provincias, municipios, secciones censales u otras delimitaciones geopolíticas).
- ✓ Tratamiento de unidades geográficas en las que un fenómeno se distribuye de manera desigual en el espacio, sobre todo, cuando se utilizan datos de regiones extremas (centro-periferia, norte-sur, este-oeste) o cuando se trabaja con datos referidos tanto a antiguas áreas metropolitanas como zonas de nuevo asentamiento, provincias urbanas y provincias rurales, secciones censales del centro de una ciudad y secciones del extrarradio, etc. (Anselin, et al., 2008).
- ✓ Causas de tipo sociológico, como la existencia de diversos gustos o actitudes de la población, o político, cuando en la zona analizada se producen diferentes administraciones o políticas regionales que llevan a respuestas diferentes ante un mismo estímulo (Moreno & Vayá, 2002).
- ✓ Además, a estas situaciones hay que añadir las causas habituales del problema de heteroscedasticidad en los modelos de regresión lineal: omisión de variables relevantes u otro tipo de especificación errónea del modelo, que producen en el término de la perturbación aleatoria una varianza no constante.
- ✓ Por último, cabe señalar que algunas causas que provocan la heterogeneidad espacial pueden también originar la aparición de autocorrelación espacial, (especificaciones erróneas o errores de medida, sobre todo), siendo necesaria la contrastación de ambos efectos conjuntamente (Anselin, 1988b).

1.3. Incorporando la Dependencia Espacial en el Modelo Econométrico

Desde un punto de vista metodológico, la inclusión de la dependencia espacial en los modelos econométricos ha sido motivada por dos vías alternativas. Por un lado, el espacio se ha incorporado mediante fundamentos teóricos, es decir, siguiendo una especificación formal dentro de un modelo económico tal que refleje la interacción de los agentes y/o la interacción social (Brock & Durlauf, 2001). Por otro lado, desde los inicios ha existido una solución pragmática, introduciendo el espacio como una peculiaridad de los datos utilizados tales como la interdependencia espacial y la relevancia de los factores localizados en otros lugares (Paelinck, 1978).

Formalmente, la autocorrelación espacial puede definirse mediante la autocovarianza de la siguiente manera:

$$\text{Cov}(y_i, y_j) = E(y_i y_j) - E(y_i) \cdot E(y_j) \neq 0 \quad \forall i \neq j \quad (4)$$

Donde y_i e y_j son observaciones de una variable aleatoria localizada en i y j sobre el espacio.

Suponiendo que el conjunto de observaciones de un corte transversal es igual a N , en total existen $[N(N - 1)/2]$ autocovarianzas espaciales para todos los posibles pares de observaciones (considerando la simetría de la matriz de autocovarianzas). Dado que el número de parámetros a estimar excede al número de observaciones es imposible estimar cada una de las autocovarianzas. Adicionalmente, incrementar el tamaño muestral no soluciona el problema ya que el número de parámetros a estimar también crece, siendo este caso un ejemplo típico del problema de parámetros incidentales (Lancaster, 2000).

Desde una perspectiva univariante, la dependencia espacial puede ser representada mediante un modelo espacial autorregresivo no restringido tal como:

$$\left. \begin{aligned} y_i &= a_{ij}y_j + a_{ik}y_k + e_i \\ y_j &= a_{ji}y_i + a_{jk}y_k + \varepsilon_j \\ y_k &= a_{ki}y_i + a_{kj}y_j + \varepsilon_k \\ \varepsilon_i, \varepsilon_j, \varepsilon_k &\sim \text{i. i. d. } N(0; \sigma^2) \end{aligned} \right\} \rightarrow y = Ay + \varepsilon \quad (5)$$

Siendo A un matriz de coeficientes de interacción con la siguiente estructura

$$A = \begin{bmatrix} 0 & a_{ij} & a_{ik} \\ a_{jk} & 0 & a_{jk} \\ a_{ki} & a_{kj} & 0 \end{bmatrix}$$

El modelo (5) es de “(...) escasa utilidad práctica dado que esto resultará en un sistema con muchos más parámetros que observaciones (...). La solución al problema de la sobreparametrización, que surge cuando permitimos que cada relación de dependencia posea su propio parámetro, es imponer una estructura sobre las relaciones de dependencia espacial” (LeSage & Pace, 2009).

En otros términos, el modelo (5) se encuentra subidentificado y la solución clásica radica en imponer alguna estructura en la matriz A, parametrizando la interacción espacial como, por ejemplo: $A = \rho W$, donde ρ es un parámetro a estimar y W una matriz de contactos (también denominada matriz de contigüidades, pesos, ponderaciones, distancias o interacciones espaciales).

1.3.1 Matriz de Contigüidad.

Se denomina matriz de contigüidad o de conectividad al arreglo W donde tanto cada una de las filas como de las columnas representa una región en el espacio objeto de estudio. Esta matriz representa la relación que tiene cada una de las regiones con las demás regiones del espacio en estudio, tal como se vería en un mapa.

La necesidad de una matriz de contactos proviene del problema de subidentificación que afecta a la mayoría de los modelos espaciales. (Haining, 2003) menciona que “el primer paso para cuantificar la estructura de dependencia espacial en un conjunto de datos es definir, para el conjunto, la relación espacial existe entre ellos”.

Utilizando la estructura $A = \rho W$, el modelo (5) se transforma en un modelo restringido:

$$y = \rho W_y + \epsilon \tag{ 6}$$

Donde el término W_y es comúnmente denominado rezago espacial de la variable endógena

Suponiendo que el tamaño muestral es igual a N, la matriz W es de orden N x N y puede representarse de la siguiente manera:

$$W = \begin{bmatrix} 0 & w_{1,2} \dots & w_{1,j} \dots & w_{1,N} \\ w_{2,1} & 0 & \dots & w_{2,j} \dots & w_{2,N} \\ \dots & \dots & \dots & \dots & \dots \\ w_{i,1} & w_{i,2} \dots & 0 & \dots & w_{i,N} \\ \dots & \dots & \dots & \dots & \dots \\ w_{N,1} & w_{N,2} \dots & w_{N,j} \dots & \dots & 0 \end{bmatrix} \tag{ 7}$$

Donde las columnas y filas corresponden a las observaciones de corte transversal y los pesos w_{ij} ($i, j = 1, 2, \dots, N$) aproximan la relación entre dos localizaciones i y j . La diagonal principal está formada por ceros, estableciendo que ninguna observación puede estar relacionada consigo misma (no puede ser que la misma observación sea vecina de sí misma). Obsérvese que se ha intercambiado α_{ij} (parámetro a estimar dentro del modelo) por w_{ij} (hipótesis externa al modelo).

Llegado a este punto, el problema radica en cómo se construyen los pesos de la matriz W . Existen diferentes criterios sobre su construcción siguiendo, por ejemplo, alguna hipótesis de interacción. Cada hipótesis resultará en una matriz de ponderaciones diferentes llevando a un rezago espacial distinto. Es común que la representación del conjunto de vecinos obtenidos por algún criterio de selección se realice mediante una elección binaria, con $w_{i,j} = 1$ cuando i y j son vecinos, y $w_{i,j} = 0$ cuando no lo son.

En la práctica econométrica, la matriz de contactos es a menudo construida desde la geografía o geometría, usando los conceptos de contigüidad y distancia. Por ejemplo, suponiendo que las observaciones se distribuyen sobre un mapa regular como el representado en la Figura 1, hay varias alternativas para establecer el conjunto de vecinos para la celda a .

Una posibilidad es considerar vecinos a aquellas celdas que poseen un borde común, como se muestra en la Figura 1 (superior izquierdo). Otra posibilidad es considerar como vecinos, aquellas celdas que poseen un vértice común, Figura 1 (superior derecho). Las elecciones de estos conjuntos de vecinos son denominadas, respectivamente, criterio tipo torre y criterio tipo alfil, en analogía al movimiento de las piezas de ajedrez.

De igual forma, puede elegirse vecinos mediante una combinación de ambos criterios dando lugar al criterio tipo reina. (véase Figura 1 (inferior izquierda))

Las áreas o polígonos de un mapa pueden transformarse en puntos, y viceversa. La elección de un punto representativo de un polígono es típicamente resuelta por criterios geométricos mediante el punto central o centroide del polígono. Dado el centroide de cada celda, como se ilustra en la Figura 1 Contigüidad sobre Mapa Regular (inferior izquierda), es posible definir una red de puntos o nodos. En este caso, algún criterio de distancia permite definir a los vecinos de a tal que cada centroide que se encuentre dentro de una distancia máxima al centroide de a sería considerado vecino.

Un problema con la elección de vecinos por medio de la distancia (y contigüidad) es la existencia de puntos aislados que pueden no contener vecino alguno para un radio determinado. Esto sucede habitualmente cuando la densidad de los puntos sobre el plano no es regular o cuando se encuentran algunos nodos distribuidos por agrupamientos (clúster) y

otros aislados. Para salvar este problema se suele determinar un radio d de amplitud tal que asegure que cada observación tiene al menos un vecino.

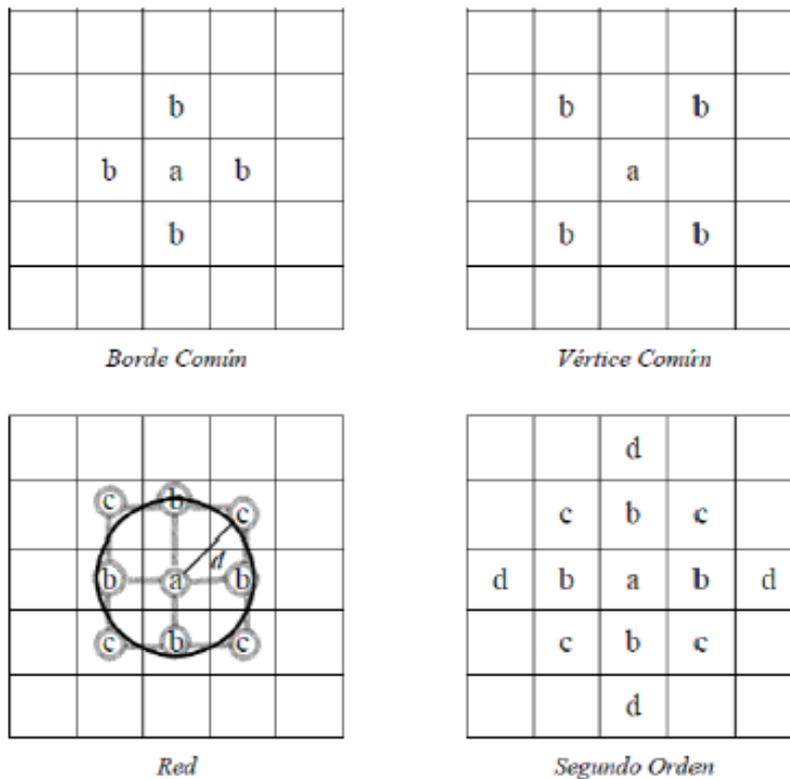
Un criterio alternativo es el de k -vecinos más cercanos. La ventaja de este criterio es que todas las unidades poseerán la misma cantidad de vecinos evitando el problema de unidades aisladas. Bajo mapas regulares, como los representados en la Figura 1, si k es igual a 4, los vecinos más cercanos a la localidad a serán los mismos que los elegidos por el criterio tipo torre y, si k es igual a 8, los vecinos más cercanos a la localidad a serán los mismos que los elegidos por el criterio tipo reina.

Por otro lado, es posible considerar varios órdenes de contigüidad o vecindad. En la Figura 1 (inferior derecha) muestra el grupo de vecinos de segundo orden para la celda a . Este conjunto de vecinos, identificado por las letras c y d , incluye a los vecinos de los vecinos de a definidos por el criterio tipo torre.

Como alternativa a los pesos binarios, pueden considerarse aquellas funciones de distancia entre unidades espaciales que combinan la distancia, el perímetro y otras características geográficas. Algunas especificaciones utilizadas en la literatura aplicada son mostradas en la Tabla 1. Diversas especificaciones de la matriz de pesos

Determinadas especificaciones plantean posibles problemas de endogeneidad que deberán ser considerados en el momento de la estimación del modelo econométrico. Este tipo de problema puede surgir con propuestas que utilizan variables socio-económicas, como el nivel de empleo o el producto bruto per cápita, para la elección de los pesos. Exceptuando estos casos, los procedimientos presentados en la Tabla 1. Diversas especificaciones de la matriz de pesos pueden encasillarse como procedimientos exógenos, es decir, aquellos que determinan la estructura de la matriz en función, únicamente, del arreglo espacial de los datos.

Figura 1 Contigüidad sobre Mapa Regular



Fuente: Anselin (1988).

Desde una perspectiva diferente, podrían seleccionarse los pesos espaciales mediante una descripción estadística de los datos. Las propuestas de este tipo suelen denominarse procedimientos endógenos. En este caso, los vecinos cercanos o lejanos son establecidos mediante la información brindada por los propios datos mediante un estadístico que sirve de guía, como en el modelo estadístico local, SLM, de (Getis & Aldstadt, 2004), el procedimiento AMOEBA de (Aldstadt & Getis, 2006): ambos referidos en (Herrera, et al., 2012), el de modelo de ecuaciones estructurales de (Folmer & Oud, 2008), la técnica de filtrado espacial de (Tiefelsdorf & Griffith, 2007), el procedimiento de máxima entropía general de (Fernández et al., 2008) o el procedimiento del coeficiente de correlación completo, CCC, de (Mur & Paelinck, 2011).

Una vez elegidos los pesos espaciales, es raro que se utilicen en su forma natural. Lo habitual es trabajar con alguna transformación de la matriz en el modelo espacial ya que mejora las propiedades estadísticas de los estimadores y sus estadísticos. La transformación más utilizada es la normalización por fila, en donde los nuevos pesos son obtenidos como $w_{ij}^* = w_{ij} / (\sum_j w_{ij})$ de tal forma que la suma de cada fila de la matriz sea igual a la unidad $\sum_j w_{ij}^* = 1$ (representamos como W^* a la matriz de contactos normalizadas)

Tabla 1. Diversas especificaciones de la matriz de pesos

Referencia	Modelo	Descripción
Dacey (1968)	$w_{ij} = d_{ij} \cdot \alpha_i \cdot \beta_{i(j)}$	<ul style="list-style-type: none"> . d_{ij}: distancia entre los puntos o regiones (i, j) . α_i: proporción de i sobre el área total de regiones. . $\beta_{i(j)}$: proporción del perímetro de i en contacto con j
Cliff y Ord (1973)	$w_{ij} = d_{ij}^{-a} [\beta_{i(j)}]^b$	<ul style="list-style-type: none"> . a, b: parámetros positivos
Bodson y Peeters (1975)	$w_{ij} = \sum_{n=1}^N K_n \left\{ \frac{a}{1 + b \cdot e^{-c_j d_{ij}}} \right\}$	<ul style="list-style-type: none"> . K_n: importancia del medio de comunicación n . N: total de medios de comunicación considerados . a, b, c: parámetros a estimar.
Anselin (1980)	$w_{ij} = d_{ij}^{-2}$	
Cliff y Ord (1981)	$w_{ij} = (c + d_{ij})^{-a}$	<ul style="list-style-type: none"> . c: término constante positivo
Case <i>et al.</i> (1993)	$w_{ij} = \frac{1}{ x_i - x_j }$	<ul style="list-style-type: none"> . x: variable socioeconómica (ej., PIB per cápita).
Molho (1995)	$w_{ij} = \frac{E_j^{-ad_{ij}}}{\sum_{k \neq i} E_k^{-ad_{ik}}}$; $\forall i \neq j$	<ul style="list-style-type: none"> . E: volumen de empleo
Ma <i>et al.</i> (1997)	$w_{ij} = e^{-d_{ij}^a}$ $w_{ij} = (l_{ij}/l_i)^a$ $w_{ij} = \frac{(l_{ij}/l_i)^a}{d_{ij}^{-b}}$	<ul style="list-style-type: none"> . l_{ij}: longitud de frontera entre las regiones (i,j) . l_i: perímetro de la región i
Toral (2000A,B)	$w_{ij} = \delta_{ij} \frac{k_i k_j P_i P_j}{d_{ij}^a}$	<ul style="list-style-type: none"> . $\delta_{ij} = 1$, si las unidades espaciales i, j tienen una frontera en común y cero, si no la tienen. . p: población; k: longitud (km) de carreteras . d_{ij}: distancia por carretera entre las capitales de i, j . a: parámetro positivo, con valores 0, 1 ó 2.
Van der Kruk (2001)	$W = \sum_{d=1}^D W_d$	<ul style="list-style-type: none"> . d: orden de vecindad . D: número máximo de órdenes de vecindad existentes

Fuente: (Yrigoyen, 2003)

1.3.2. Detección Global de Dependencia Espacial Univariante

La cuestión, ¿presenta la variable de estudio autocorrelación espacial? entonces, pasa por conocer cómo detectar la presencia de autocorrelación espacial. Desde una perspectiva descriptiva, una primera aproximación cualitativa puede realizarse mediante el análisis exploratorio de datos espaciales. (Haining, 2003).

El análisis exploratorio de datos espaciales (AEDE) permite la visualización mediante diferentes gráficos del comportamiento de la variable bajo estudio, siendo el mapa un elemento central. Existe una amplia variedad de mapas y formas para describir datos continuos sobre polígonos irregulares ("lattice data"), una de las más usuales es el mapa de

coropletas. En este tipo de mapa se representa la distribución espacial de una variable o atributo mediante diferentes tonalidades.

El análisis exploratorio espacial brindará los primeros indicios de dependencia, aunque se necesita técnicas inferenciales para poder obtener una cuantificación de la misma. La formulación de estas varía a través de diferentes estadísticos, aunque cada uno de ellos puede ser visto como un caso especial del test de Mantel referido en (Haining, 2003).

El test de Mantel estima la correlación existente entre dos matrices B y C donde dichas matrices capturan el grado de similitud en el espacio y en la distribución de la variable, respectivamente. Dicho estadístico se define como:

$$\Gamma = \sum_{i=1}^N \sum_{j=1}^N b_{ij}c_{ij} \quad \text{con } i \neq j \quad (8)$$

Donde b_{ij} es un elemento de la matriz B que captura la similitud espacial de las localizaciones i y j , y c_{ij} es un elemento de la matriz C que captura la similitud de los valores de una determinada variable entre las localizaciones i y j .

Reformulando del estadístico Γ , tal que los elementos b_{ij} sean los pesos espaciales de la matriz de contactos W y los elementos c_{ij} representen el cuadrado de la diferencia de valores de la variable de interés entre localizaciones, puede obtenerse el estadístico C de (1954): Geary

$$C = \frac{N-1 \sum_{i=1}^N \sum_{j=1}^N w_{ij} (y_i - y_j)^2}{2S_0 \sum_{i=1}^N (y_i - \bar{y})^2} \quad (9)$$

Donde $S_0 = \sum_{i=1}^N \sum_{j=1}^N w_{ij} = 1'W1$, siendo 1 un vector $(N \times 1)$ de unos y \bar{y} es la media muestral.

Los momentos del estadístico de Geary, bajo la hipótesis nula de aleatoriedad, son:

$$E[C] = 1 \quad (10)$$

$$V[C] = \frac{(2S_1 + S_2)(N-1) - 4S_0^2}{2(N+1)S_0^2} \quad (11)$$

Donde $S_1 = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (w_{ij} + w_{ji})^2$ y $S_2 = \sum_{j=1}^N (w_{ji} + w_{ij})^2$ siendo $w_j = \sum_{i=1}^N w_{ij}$

Su distribución asintótica es normal:

$$\sqrt{N}[C - 1] \sim_{as} N [0; V(C)] \quad (12)$$

Posiblemente el test más ampliamente utilizado, y que también puede derivarse del test de Mantel, es el estadístico I de Moran (1950):

$$I = \frac{N \sum_i \sum_j (y_i - \bar{y}) w_{ji} (y_j - \bar{y})}{S_0 \sum_{i=1}^N (y_i - \bar{y})} \quad (13)$$

Los momentos del estadístico de Moran, bajo la hipótesis nula de aleatoriedad, son:

$$E[I] = -\frac{1}{N-1} \quad (14)$$

$$V[I] = \frac{(3S_0^2 + S_1 R^2 - N S_2)}{S_0 (N+1)(N-1)} - \frac{1}{(N-1)^2} \quad (15)$$

Su distribución asintótica, nuevamente, es normal:

$$\sqrt{N}[I - E(I)] \sim_{as} N[0; V(I)] \quad (16)$$

Diferentes experimentos Monte Carlo (Anselin & Florax, 1995) revelan que la aproximación a la distribución normal funciona razonablemente bien para $N > 50$. El comportamiento del estadístico empeora sensiblemente cuando se reducimos el tamaño muestral o se añaden problemas nuevos como heteroscedasticidad, atípicos, etc., lo que significa que los resultados deben interpretarse con cautela.

Tanto el test C de Geary como el test I de Moran capturan la dependencia global, es decir, la autocorrelación existente en todo el espacio geográfico. Existen otros estadísticos que permiten capturar la dependencia local utilizando sub-muestras del espacio geográfico. Entre los más utilizados se destacan el análisis LISA de Anselin (1995) ,y los tests $G_i(d)$ de Getis y Ord, (Getis & Ord, 1992; Ord & Getis, 2001).

El rechazo de la hipótesis nula de aleatoriedad de alguno de estos tests de autocorrelación espacial implica un análisis más detallado sobre los determinantes de la variable bajo estudio.

1.4. Análisis exploratorio de datos espaciales

El AEDE para Anselin (1999), es “una colección de técnicas para describir y visualizar distribuciones espaciales, identificar localizaciones espaciales atípicas u outliers espaciales, descubrir patrones de asociación espacial, clúster o puntos calientes, y sugerir regímenes espaciales u otras formas heterogeneidad espacial”.

Según Yrigoyen (2003), “el AEDE tiene un carácter descriptivo (estadístico) más que confirmatorio (econométrico) pero, tras la detección de estructuras espaciales en las variables geográficas, hará posible la formulación de hipótesis previas para la modelización econométrica y, en su caso, posible predicción espacial de nuevos datos. Es decir, el AEDE debe constituir la fase previa a toda modelización econométrica espacial, sobre todo cuando no exista un marco formal o teoría previa acerca del fenómeno que se pretende explicar”.

A diferencia del análisis tradicional, en el análisis exploratorio de datos espaciales los mapas cobran especial importancia, ya que permiten visualizar en el mapa los casos atípicos observados en el histograma, los patrones espaciales, las asociaciones, etc. Permite responder preguntas tales como:

- ¿Dónde se encuentran en el mapa los casos atípicos observados en el histograma?
- ¿Dónde aparecen los valores de cierto atributo de una parte del mapa en el *scatterplot*?
- ¿Cuáles son los patrones espaciales y las asociaciones espaciales en este conjunto de datos?, etc.

Las distintas técnicas que se pueden emplear en el AEDE, son varias y pueden ser clasificadas según distintos criterios, pero la mayoría de autores consideran, por un lado, los métodos gráficos y por otro, los contrastes estadísticos, y ambos son complementarios. Se debe considerar si el análisis es univariante o multivariante y si queremos considerar la asociación espacial global o local. En las técnicas gráficas existen dos perspectivas que serían la geoestadística y la reticular o “lattice”, que se diferencian, según Anselin (1998), en la forma en que la proximidad espacial es formalizada. La geoestadística supone procesos espaciales continuos, que hacen usar la distancia métrica para organizar las observaciones y la lattice supone observaciones de carácter discreto, siendo clave el concepto de vecino espacial que conduce a la matriz de pesos espaciales y a las variables espaciales retardadas. Para ambos enfoques algunos autores (Anselin, 1999), (Moreno & Vayá, 2002) y (Yrigoyen, 2003) distinguen entre técnicas que se usan para representar la distribución espacial, de otras que reflejan la asociación espacial y un tercer bloque para la heterogeneidad espacial.

Recogen las siguientes técnicas gráficas:

- Función de distribución acumulativa. Refleja la distribución espacial, desde un enfoque geoestadístico para el caso univariante, lo mismo que el diagrama/mapa de puntos y el mapa de círculos.
- Diagrama/mapa de caja representa la distribución espacial, desde un enfoque lattice para el caso univariante.
- Diagrama dispersión-caja muestra la distribución espacial bajo la perspectiva lattice en el caso multivariante.
- Cartograma y gráfico de coordenadas paralelas muestran la distribución espacial bajo la perspectiva lattice y geoestadística en el caso multivariante.
- Diagrama de dispersión espacialmente retardado, nube del variograma y diagrama de caja del variograma recogen la asociación espacial global bajo la aproximación geoestadística.
- Puntos atípicos en el diagrama de caja del variograma y puntos atípicos en la nube del variograma recogen la asociación espacial local bajo la aproximación geoestadística.
- Diagrama/mapa de dispersión de Moran, mapa de contigüidades espaciales y gráfico del retardo espacial muestran la asociación espacial global bajo el enfoque lattice.
- Mapas LISA, diagrama de caja LISA y puntos atípicos en el diagrama de dispersión de Moran, muestran la asociación espacial local bajo el enfoque lattice.
- Nube del variograma multivariante representa la asociación espacial bajo el enfoque geoestadístico para el caso multivariante.
- Diagrama de dispersión multivariante de Moran representa la asociación espacial bajo enfoque lattice en el caso multivariante.
- Mapa del histograma de frecuencias estudia la heterogeneidad espacial con la perspectiva lattice lo mismo que el diagrama de dispersión.

Con los contrastes estadísticos se persigue detectar el grado de asociación espacial global o local. Los test de autocorrelación espacial global tratan de contrastar la independencia o no de la distribución de una variable en la totalidad de un área territorial, mientras que los de autocorrelación espacial local, se centran en la detección de relaciones espaciales en zonas o subzonas concretas del territorio, midiendo su aportación a las características de independencia de la variable en todo el espacio considerado.

Los test más conocidos serían, para la autocorrelación espacial global, el test I de Moran, del que se deriva el diagrama de dispersión de Moran, el test c de Geary y la G (d) de Getis y Ord. También estaría el Γ de Mantel. Para la autocorrelación espacial local estarían los estadísticos locales de Getis y Ord y los LISA de Anselin.

Además, están los diagramas espacio temporales como sería el coeficiente de autocorrelación espacio tiempo de Moran propuesto por Yrigoyen (2003), y los indicadores LISA para el caso

bivariante. También está el análisis de correlación espacial para tasas, que pretende corregir los problemas ocasionados al presentar las variables en ratios. Las técnicas serían la estandarización empírica bayesiana, el método directo de estandarización de tasas propuesto por (Saez & Saurina, 2007) y la adaptación del estadístico I de Moran propuesto por (Assuncao & Reis, 1999).

1.5. Estudios de población. Surgimiento y desarrollo

Siempre ha existido una preocupación por la dinámica demográfica en pos de una meta social considerada importante. Los principales planteamientos y teorías se comportan en consecuencia de los sistemas económicos sociales, naturales y sociales, de tal forma estos se han transitado desde el estímulo al poblamiento y crecimiento de la población hasta las restricciones y limitaciones para la multiplicación de los hombres.

Confucio (551-479 A.C.), en sus análisis de población relaciona factores demográficos con económicos al plantear que la celebración de matrimonios caros disminuye la nupcialidad y con ello el número de nacimientos para garantizar el aumento de a población.

Platón (428-347 A.C.), sugiere aumentar la población mediante la inmigración estableciendo un principio de optimalidad. El número ideal de habitantes para una ciudad-estado era de 5040 personas

Hay enfoques relevantes en el período de la ilustración, donde a la par del crecimiento de la población ocurren grandes cambios económicos, políticos y sociales y con ello una nueva concepción del mundo: se cree en el hombre y su capacidad de dominar la naturaleza, existe la convicción de poder llegar a una sociedad perfecta.

Ya para 1798 Malthus publica de forma anónima que el crecimiento de la población es en forma de progresión geométrica, mientras la producción de alimentos crece en progresión aritmética, a la hora de juzgar esos planteamientos se deben tener en cuenta su pensamiento social, la fuente de sus datos, tanto desde el punto de vista del alcance como veracidad y su sesgo intencionado o no, a la hora del análisis y las conclusiones.

Para la llamada Escuela Clásica (liderada por Ricardo y Smith), el tamaño absoluto de la población forma parte de la explicación para saber cómo se produce el excedente, saliendo la idea de rendimientos decrecientes es decir a mayor población menos rendimientos. La población es vista como fuerza de trabajo y no en su doble calidad de productora consumidora.

La Escuela Marxista es la principal opositora del pensamiento malthusiano. Planteando lo siguiente el comportamiento de la población depende de la base económica y que en el proceso de acumulación se produce una superpoblación relativa (por la relación capital-

trabajo). Este grupo es independiente del tamaño de la población, para Marx toda formación económica social tiene su propia Ley de Población. Para el capitalismo es la ley de la superpoblación relativa; La Ley en el Socialismo es la satisfacción de las necesidades crecientes de la población

Los modelos demo-económicos argumentan la necesidad de reducir el ritmo de crecimiento de la población. Evalúan el impacto en la población del crecimiento económico. Algunos ejemplos son:

- Coale-Hoover: Ejercicio de impacto del crecimiento con hipótesis de alta y baja fecundidad.
- Harrod-Domar: La población activa permanece constante. Son modelos a corto plazo

El club de Roa sostiene que no solo es importante que el crecimiento poblacional se reduzca, sino que, si no se hace, los resultados serán catastróficos. En realidad, son neomalthusianos.

Se les critica que los límites del crecimiento no son físicos, sino socioeconómicos y políticos. También que el crecimiento de la población es exponencial, pero la ciencia y el progreso científico tienen una dinámica propia superior.

Todos estos estudios de población tienen una orientación clasista y con un pensamiento de derecha imperante en el mundo, no es de esperar planteamientos en aras de reducción de la pobreza, y con esto un equilibrio natural del crecimiento poblacional.

A partir de los años 70 hay un cambio en los estudios de población focalizándolos hacia la relación población desarrollo, en 1974 surge el (PAMP) Plan de Acción Mundial sobre Población, con ello se trata de eliminar las barreras ideológicas y se asumiera el problema demográfico como parte del problema global del futuro de la humanidad, tomándose en cuenta índices deficitarios del nivel de bienestar de la población como alimentos, educación, empleo, vivienda y servicios de salud.

En el marco de la Segunda Conferencia Mundial de Población en México 1984, se confirma como el principal objetivo del desarrollo social, económico y humano el mejorar el nivel y localidad de vida de las personas. Estableciendo como la base para la solución definitiva de los problemas de la población las transformaciones económicas y sociales; y que las políticas de población deben ser un elemento integrante de las políticas de desarrollo. Por tanto, el crecimiento económico es un medio para elevar el bienestar y facilitar el logro del desarrollo social y personal.

Para caracterizar las poblaciones de una manera más balanceada surge el Índice de Desarrollo Humano (IDH) este se construye de 4 indicadores los que a su vez expresan 3

aspectos del desarrollo humano: Ingresos (Producto Interno Bruto Per cápita), longevidad (esperanza de vida al nacer) y conocimientos (tasa de alfabetismo y tasa de matrícula combinada, es decir tasa de matrícula de las enseñanzas primaria, secundaria y terciaria). En 1996 el profesor cubano Cándido López crea el Índice de Desarrollo Humano Modificado (IDHM) que introduce en el cálculo una medida de la transferencia del ingreso al desarrollo humano o bienestar.

Las recomendaciones del Plan de Acción Mundial sobre Población respetan las condiciones económicas y sociales de cada país, además de sus valores culturales y éticos, acervo cultural y convicciones filosóficas, Esto es fundamental en temas tal álgidos como el aborto.

1.5.1 Evolución de los estudios sociodemográficos.

El auge de los estudios sobre la población tiene lugar precisamente cuando la población se comienza a ver como un problema. Mientras su crecimiento es lento, por una alta natalidad y mortalidad y una vez que la mortalidad se reduce los países con excedentes demográficos se pueden expandir a otros territorios, mientras el medio ambiente no parece perturbarse por la presencia del hombre, no hay razones para preocuparse.

Pero estos equilibrios comienzan a romperse a mediados del presente siglo, en la posguerra, cuando no hay nuevos territorios que colonizar y la mortalidad de los países pobres se reduce por la introducción de la medicina moderna y medidas higiénicas y la población tercermundista, que mantiene una fecundidad elevada, comienza a crecer a ritmos elevados. A su vez, comienzan a agotarse las fuentes de recursos no renovables y el medio ambiente demuestra su fragilidad.

En la postguerra también hay interés por la planificación del desarrollo económico y una mayor intervención del Estado. A ello se suma el desarrollo creciente de la mecanización en el proceso de información estadística y posteriormente la computación.

Los estudios demográficos han respondido fundamentalmente a la evolución de la transición demográfica, es decir, el paso de altos a bajos niveles de fecundidad y mortalidad.

1.5.2 Estadísticas y su vínculo con los estudios de población

Los estudios de población sin lugar a dudas deben ser abordados con técnicas estadísticas, la fuente de mayor aporte para un estudio de población es el censo de población actividad realizada cada 10 años en casi todos los países. Los censos de población recogen los datos de interés según sus objetivos, estos datos casi siempre son más de 30 por planilla y se les realizan a millones de personas, sin duda estamos en presencia de una gran base de datos;

y la mejor herramienta para obtener la información deseada a partir de esos datos es la estadística, dada su forma de trabajo. Además, en ellos también está presente la aleatoriedad.

Un levantamiento inicial digital agiliza los resultados globales pero las técnicas estadísticas son las que permiten un estudio profundo, pues no todos los datos tienen las mismas características a pesar de ser el mismo objeto; ejemplo las viviendas no son todas iguales atendiendo a su cantidad de habitaciones, este dato es importante en un censo de población y viviendas como el nuestro; las técnicas estadísticas permiten estudiar las viviendas a partir de la cantidad de habitaciones.

En Cuba la Oficina Nacional de Estadísticas e Información (ONEI) perteneciente al Ministerio de Economía y Planificación es la encargada de procesar las estadísticas cubanas. Dentro de ella el Centro de Estudios de Población y Desarrollo (CEPDE), atiende las estadísticas demográficas, cuyo soporte los constituye el Sistema de Información Estadístico Nacional de Demografía (SIEN-D).

Tabla 2. Fuentes de información del Sistema de Información Estadístico Nacional de Demografía

Estadísticas	Centro informante	Organismo rector	Documento	Observaciones
Nacimientos	Hospitales 99.8% Registro civil	MINSAP/ MINJUS	Cert. Nacimiento	Queda inscrito en el CIRP
Defunciones	Hospitales Registro civil	MINSAP/ MINJUS	Cert Defunción General y Neonatal	Talón desprendible
Migraciones Internas	Oficinas del CIRP	MININT	Talón de baja	
Migraciones externas	Dir. Inmigración y extranjería	MININT	Soporte magnético	Captación nacional
Matrimonios	Palacio de los matrimonios y Notarías	MINJUS	Acta matrimonial	
Divorcios	Bufetes colectivos /Tribunales y Notarías	MINJUS	Modelo estadístico	
Población por sexo y edades	Registro Cosumd. Y resto de subsistemas CPV/81	MINCIN	Soporte magnético	Captación nacional

Fuente: González E., CEPDE, 1999

Los estudios de población tienen una nomenclatura usual de términos demográficos y algunas técnicas propias de esa disciplina, como son la ecuación compensadora, el Diagrama de Lexis, las pirámides de población, coeficiente de vejez, razones, proporciones y tasas, estas 3 últimas son conocidas de la estadística general.

1.5.3 Distribución territorial de la población.

La forma en que se distribuye la población en el territorio, es decir, por provincias, municipios, zonas urbana o rural, etc., constituye uno de los elementos que determinan la distribución territorial de la sociedad (de la economía, los servicios, etc.). De ahí la importancia de su estudio. La definición más simple que puede darse es aquella que expresa que “la distribución de la población es la forma en que la población ocupa y usa el espacio físico”.

Otra definición plantea que la distribución de la población se encarga de la “localización espacial de las actividades económicas y de la población”. Esta definición da cuenta de la expansión o crecimiento de la población y de las fuerzas productivas en un territorio. De hecho, el doble carácter de la población como productora o elemento más activo de las fuerzas productivas y como consumidora, está condicionando que su asentamiento en un territorio implique la localización o emplazamiento de la producción material y los servicios.

De tal manera, puede plantearse la presencia de una relación recíproca entre el comportamiento demográfico de una población y el desarrollo socioeconómico del territorio donde está asentada la misma.

La distribución de la población en el territorio puede entenderse como el resultado de dos tipos de procesos poblacionales: el comportamiento diferenciado por territorios de la fecundidad y la mortalidad, y por las migraciones.

Las migraciones son las principales responsables de los cambios rápidos y más significativos de la distribución territorial de la población, y es por esto la variable que más se relaciona con ella.

Unidades de análisis de la distribución espacial de la población:

- Divisiones político administrativas (provincias, municipios)
- Áreas urbanas y rurales.
- Agrupaciones de población, según tamaño.
- Áreas metropolitanas
- Regiones, especialmente regiones económicas.
- Barrios u otro tipo de sector de la ciudad.

Los propósitos de la clasificación de la población por territorios son:

- Disponer de datos demográficos de divisiones administrativas como elementos útiles para programar, planificar y desarrollar los servicios sociales de dicho territorio (salud, educación, vivienda, servicios comunales).
- Disponer de información territorial sobre las variables demográficas y prever su comportamiento.
- Obtener información que sirva para construir marcos muestrales.
- Tener un mejor conocimiento de la migración interna (flujos, patrones de movilidad, selectividad).

La diversidad de las características de la población, según se distribuya espacialmente es notoria. Algunos aspectos que tienen diferencias espaciales son:

- La fecundidad, la mortalidad y las migraciones.
- La estructura por edades y sexos.
- La ocupación según ramas de actividad y los niveles de ingreso.
- El nivel de escolaridad y el grado de instrucción.
- La composición étnica
- El grado de participación en la sociedad.
- La movilidad profesional.
- La densidad geográfica.
- Los tipos de familia
- Las condiciones de habitabilidad.
- Las actitudes culturales.

Existen tres tipos de factores asociados a estas diferencias espaciales: los de tipo geográfico, los de tipo económico y los de tipo demográfico.

* De tipo geográfico: Dentro de éste tenemos el medio natural (clima, topografía, fertilidad del suelo, acceso a recursos energéticos y a las fuentes de materia prima) y la situación geográfica de las fuentes de producción agraria e industrial.

* De tipo socioeconómico: La influencia de las actividades económicas, el desarrollo científico-técnico, las políticas y la organización social y aspectos culturales.

* De tipo demográfico: El crecimiento natural, y fundamentalmente las diferencias en la fecundidad, señalan una tendencia a la agudización de las diferencias territoriales, y por otro lado la influencia de las migraciones, por lo general desde regiones de alto crecimiento migratorio a regiones de bajo crecimiento.

El tipo de economía y el grado de desarrollo de cada país o región implica una distribución espacial de la población diferente. Sin embargo, no es posible hoy un desarrollo equilibrado entre la distribución de la población y la estructura económica, pues ha ido más rápido el proceso de urbanización que el de industrialización.

Medidas de la distribución espacial de la población:

Las mismas permiten relacionar a la población con su espacio físico y conocer las diferencias entre los territorios.

1. Densidad de población: Relaciona la cantidad de personas por unidad de espacio físico. Es el indicador más elemental para medir la concentración de la población.

Curva de Lorentz: Relaciona las proporciones acumuladas de los asentamientos urbanos según tamaño con la población con la proporción de población allí residiendo.

En el eje de las ordenadas (y) se encuentra el porcentaje acumulado de las localidades ordenadas según tamaño. En el eje de las abscisas (x) se localiza el porcentaje acumulado de la población que reside en las localidades, según categorías de tamaño. Si la curva tiene un comportamiento muy parecido a la diagonal, la población urbana tiene una distribución uniforme, o sea, a incrementos relativos de la población urbana, se observan iguales incrementos de las localidades urbanas. En la medida que la curva se aleja de la diagonal, la concentración urbana es mayor.

Índice de concentración de Gini: Resume en una cifra la curva de Lorentz, por lo que permite realizar comparaciones con otros países y en el tiempo. Consiste en el cálculo de la distancia máxima entre la curva y la diagonal, a través de procedimientos matemáticos.

Índice de primacía: Expresa las veces que la población de una ciudad es mayor que la ciudad o las ciudades que le suceden en tamaño.

1.5.4 El tiempo en Demografía. El Diagrama de Lexis.

En Demografía el tiempo, y también la edad, son variables de gran importancia, presentes en todos los análisis demográficos. La variable tiempo puede medirse de dos formas diferentes: como reflejo de un instante dado (censos) y de forma continua (registros), como reflejo de un intervalo o período. De acuerdo al tipo de información la edad será recogida de diferente forma.

La noción de instante y de intervalo permite el análisis de la población de acuerdo a los conceptos de edad exacta y edad cumplida. Así, si preguntamos a una persona su edad, seguramente nos dirá los años que cumplió en su último cumpleaños, lo que significa que nos está diciendo su edad cumplida.

El **Diagrama de Lexis** (introducido en 1875 por el estadístico alemán Wilhem Lexis) es un método gráfico utilizado para representar hechos demográficos en el tiempo y facilitar la interpretación de tasas e indicadores demográficos. En el eje horizontal se representa el tiempo "t" y en el eje vertical la edad "x", desde el momento del nacimiento o el hecho demográfico en común (matrimonio, etc.). Ambos ejes deben tener la misma escala, por lo que la población será representada por una línea recta con inclinación de 45° con relación a cada eje, partiendo de la edad $x = 0$, o sea, el nacimiento, e interrumpiéndose al momento del fallecimiento (divorcio, en el caso del matrimonio).

Definición de Nivel, Patrón y Tendencias.

En el análisis demográfico comúnmente se hace referencia a diferentes conceptos para expresar las características de las variables en estudio. Por ejemplo, se expresan criterios sobre el **nivel** de la mortalidad, los **patrones** de la migración o la **tendencia** del crecimiento de la población.

Nivel: En lenguaje demográfico se entiende por nivel a la situación que existe en determinado momento en uno de los componentes del crecimiento demográfico (fecundidad, mortalidad, migraciones), medido a través de alguna herramienta demográfica. En dependencia del refinamiento de la medida que se utilice, se podrá determinar con mayor o menor exactitud el nivel de dicha variable.

Las medidas más adecuadas para medir el nivel de:

- Fecundidad: tasa neta de reproducción, que es el número promedio de hijas tenidas por las mujeres al final de su vida fértil, si las mismas siguen los patrones de fecundidad y mortalidad observados para un momento dado.

- Mortalidad: esperanza de vida al nacimiento, que es el número promedio de años que vivirán los nacidos en el año de referencia, si se mantienen las condiciones de mortalidad de ese año.

- Migraciones: tasa neta de migración interna. Número promedio anual de migrantes netos por cada 1000 habitantes. La consideración de que el nivel es alto o bajo estará en dependencia de los marcos históricos o biológicos en los que la medida fluctúa. Por ejemplo, considerando que la tasa neta de reproducción fluctúa entre las 0.6 y 3.5 hijas por mujer, la cubana se puede considerar baja, pues en 1997 fue de 0.77 (ONE-CEPDE., 2012), (ONEI, 2012) .

Patrón: La forma de participación de la población de un país en determinada característica o variable demográfica está dada por el aporte de los distintos segmentos de la población vinculados al hecho demográfico. Esta forma de distribución por los diferentes segmentos constituye el patrón de esta población con respecto a la variable estudiada.

El estudio de la fecundidad, por ejemplo, se apoya en el aporte que realizan las mujeres comprendidas en el llamado “período fértil”, de 15 a 49 años de edad. Dentro de este rango, la mayor fecundidad ocurre entre los 20-24 y 25-29 años. En dependencia del peso de cada uno de los grupos quinquenales dentro del total de la distribución, se dice que la población en estudio tiene un patrón de fecundidad temprana (si se concentra entre los 20-24 años), tardía (si se concentra entre los 25-29 años) o dilatada (si presenta valores poco diferenciados entre los 20-29 años).

Tendencia: El conocimiento de una secuencia de información sobre una variable demográfica durante un período de tiempo permite determinar la tendencia que va experimentando el comportamiento de dicha variable.

1.5.5 Razones, proporciones y tasas.

El estudio de la población no se limita al conocimiento de la población en términos absolutos, sino a obtener medidas relativas, en aras de realizar comparaciones con otros países o territorios y también comparaciones de la misma unidad geográfica en el tiempo.

Para ello son usadas tasas, razones y proporciones, algunas de ellas derivadas de la ecuación compensadora Razón (Relación): Cuando el dividendo y el divisor pertenecen a subconjuntos o categorías diferentes dentro del conjunto de información. Por ejemplo, la razón de sexo o relación de masculinidad:

RM = Hombres / Mujeres, generalmente se analiza en hombres por cien mujeres.

Proporción: Cuando el dividendo es una parte del divisor, o sea, el dividendo es un subconjunto del divisor. Las proporciones se conocen como porcentajes cuando están multiplicadas por cien, son un tanto por ciento del total. Por ejemplo:

Proporción de población económicamente activa (PEA) = PEA / Población total o tasa de actividad

La población total está constituida por activos y no activos.

Es importante destacar que la proporción es diferente a la probabilidad, la cual expresa la frecuencia relativa de ocurrencia de un acontecimiento o suceso no renovable. Se consideran los individuos expuestos al riesgo, con relación al tamaño de la cohorte o grupo de individuos.

Tasa: Expresa la frecuencia relativa de cierto fenómeno dentro de la población en un determinado período de tiempo. Por ejemplo, la tasa de natalidad:

Tipos de tasas: Las tasas pueden clasificarse de acuerdo a diferentes criterios.

1. De acuerdo al universo estudiado. Se clasifican como:

a) Tasas globales: Cuando el fenómeno se analiza para la población total en estudio o para el total de cada sexo, sin distinguir grupos de edades. Como resultado se obtiene un solo indicador. Dentro de este grupo se encuentran las tasas brutas y las tasas brutas tipificadas, estas últimas son las tasas brutas, depuradas. Las tasas brutas se denominan así, precisamente porque no tienen en cuenta el efecto de la diferente distribución por edades de la población, es decir, relacionan el hecho demográfico con el total de la población por lo que en ocasiones involucran a segmentos de la misma que no participan del hecho demográfico analizado. Entre ellos se encuentra, por ejemplo, la tasa bruta de mortalidad:

$$d^t = \frac{D^t}{N^t} \text{ siendo } N^t = (N \text{ inicial} + N \text{ final}) / 2$$

Su principal ventaja es que constituyen un indicador único y de fácil cálculo y su desventaja es que está afectada por la estructura por edades de la población, por lo que para hacer comparaciones con otros territorios o en el tiempo es necesario estandarizarla o tipificarla.

A modo de conclusiones, se puede resumir que los estudios de población son importantes para las políticas públicas, y pueden ser realizados con mayor profundidad si se aprovecha la información georeferenciadas que existe. Este tipo de datos generalmente puede ser tratado desde un enfoque de econometría espacial, el cual se distingue por la detección de los efectos espaciales, tales como, la heterogeneidad y la asociación espacial. Una metodología que válida es la de AEDE.

Capítulo 2 Procedimiento para la identificación de patrones de asociación espacial

Al trabajar con datos geográficos no es suficiente con visualizar dichos datos en un mapa para realizar un primer análisis exploratorio, por esta razón debemos auxiliarnos de la estadística descriptiva y estadígrafos específicos donde se considera el componente espacial

En el presente capítulo se presenta la estructura del procedimiento que se centra en variables e indicadores que describen la dinámica población y sus características.

Los especialistas de la ONEI refieren de la carencia de estudios de población donde se trabajen indicadores y variables de dinámica poblacional reflejados en un mapa, visualización muy útil para mostrarles a los decisores. Ellos, la ONEI, no son un órgano del gobierno, pero si le ofrecen información para su trabajo. El conocer patrones de asociación espacial con una base estadística para diferentes indicadores y variables es un estudio totalmente novedoso para esta oficina, con esta herramienta pueden sustentar o refutar tendencias supuestas por simple observación de algunos comportamientos.

El procedimiento que se propone combina el análisis estadístico clásico con el análisis exploratorio de datos desde la perspectiva de la econométrica espacial, trabajando variables e indicadores asociados a la dinámica poblacional y sus características. Se enfatiza en los efectos espaciales de la heterogeneidad y la autocorrelación espacial. La base informativa es del Censo de Población y Viviendas 2012. La propuesta puede ser aplicada a otros Consejos Populares de cualquier provincia del país

2.1 Antecedentes metodológicos

Existen varias metodologías para la identificación de patrones de asociación espacial mediante el empleo del Índice de Autocorrelación Espacial de Moran Global, así como, los indicadores locales de asociación espacial (LISA) y su representación cartográfica. Entre ellas se encuentra la propuesta de Luc Anselin para la obtención de indicadores locales de asociación espacial (Anselin, 1995).

Con el avance de los sistemas informáticos, el surgimiento y perfeccionamiento de los SIG que facilitan la identificación y representación de los patrones de asociación espacial, comienzan a aparecer también metodologías ajustadas a cada uno de estos sistemas, entre ellas, autocorrelación espacial e indicadores locales de asociación espacial. Importancia, estructura y aplicación (Celemín, 2009), que presenta un proceder estadístico paso a paso para el cálculo del Índice de Moran de forma manual, así como, su validación estadística y representación cartográfica mediante el software GeoDa.

También se puede citar Análisis de Regresión Espacial en R (Bivand & Portnov, 2004; Edsall et al., 2008) que propone una metodología ajustada al software R, este mismo autor propone una metodología para la obtención de los patrones mediante la aplicación informática GeoDa de su propia autoría (Anselin, 2005; Anselin, et al., 2022)

Existen metodologías para el trabajo con aplicaciones como ArcGIS, QGIS y MapInfo (Österman, 2014). Las antes citadas son ampliamente utilizadas en la actualidad, con un notable empleo luego de la aparición de la epidemia causada por la COVID19. A continuación se listan algunas de las investigaciones que emplean alguna de las metodologías anteriores:

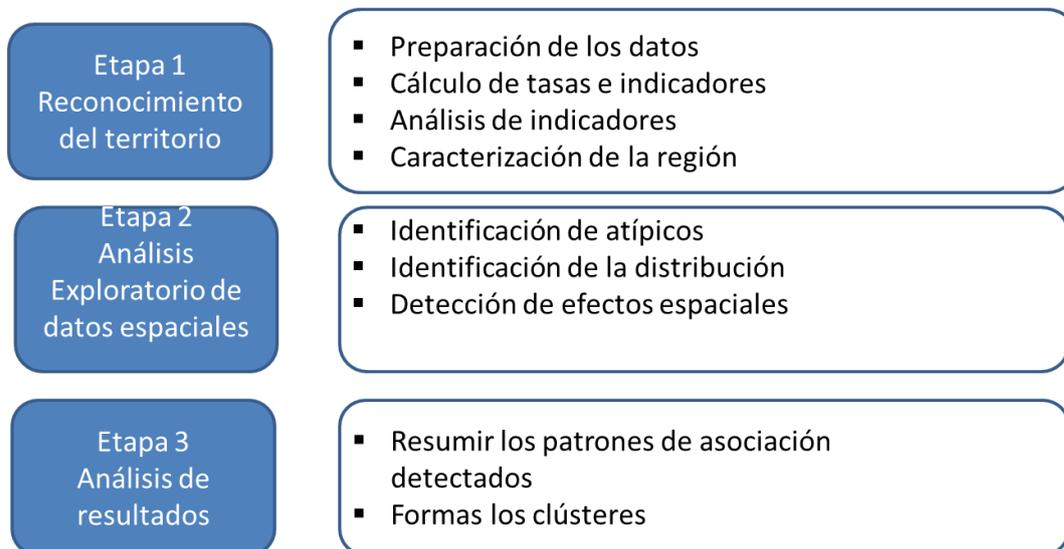
- Spatial dependency of arsenic, antimony, boron and other trace elements in the shallow groundwater systems of the Lower Katari Basin, Bolivian Altiplano (Lima et al., 2020).
- Investigating Commuting Time in a Metropolitan Statistical Area Using Spatial Autocorrelation Analysis (Miri & Miri, 2020).
- A Spatio-temporal Analysis of COVID-19 Outbreak in Italy (Ghosh & Cartone, 2020).
- Spatial Analysis of COVID-19 cases and intensive care beds in the State of Ceará, Brazil (Pedrosa & Albuquerque, 2020).
- La autocorrelación espacial y el desarrollo de la geografía cuantitativa (Siabato & Guzmán-Manrique, 2019).
- La calidad de vida de la población en la determinación de la calidad territorial. Un estudio de autocorrelación espacial aplicado a la ciudad de Mar del Plata, Argentina (Lucero & Celemín, 2014).
- An ecological study of immigrant segregation in established and new destination metros using Moran's I and LISA cluster designation (Zuchowski, 2017).
- El índice de moran en el desempleo juvenil por entidad federativa en México, 2005-2015 II (Aviles, 2017)

La gran mayoría de los trabajos emplean QGIS, ArcGis, y GeoDa, esto pudiera deberse en primer lugar a que presentan interfaz gráfica que abstrae al investigador de estudiar las interioridades de algún lenguaje de programación, y, en segundo lugar, porque son software especializados en el análisis de mapas y cuentan con muchas funcionalidades incorporadas.

El procedimiento que se expone a continuación, propone el uso de GeoDa para el análisis de la componente matemática del análisis exploratorio de datos espaciales (AEDE) por las posibilidades que brinda de crear y personalizar las funciones y resultados, para ello se utilizará como guía la metodología propuesta por Luc Anselin (Anselin, 2005) con algunas modificaciones que serán detalladas en próximos epígrafes.

2.1.1 Procedimiento para la identificación de patrones de asociación espacial.

El procedimiento estadístico se sustenta esencialmente en 2 aplicaciones informáticas que se complementan para lograr la detección de los posibles patrones de asociación y está compuesto por 3 etapas y 9 pasos.



Para una mejor comprensión del procedimiento se explicará de forma detallada cada etapa y sus pasos.

La primera etapa está relacionada con la confección de la matriz de datos y su análisis inicial, haciendo uso de Microsoft Excel o LibreOffice (este último, software libre). La segunda etapa recoge todo el proceso de análisis exploratorio de datos espaciales. La tercera etapa se emplea para la obtención de los patrones de asociación espacial en caso de existir y su tratamiento para posterior análisis, la visualización e interpretación de los patrones identificados.

2.2 Etapa 1: Reconocimiento del Territorio.

Como indican Anselin (1999), y Haining (1995), muchos de los análisis de tipo espacial que se realizan actualmente gracias a la creciente disponibilidad de datos geográficos, tienen que contrastarse con la literatura científica existente, dado que la gran dimensión de las bases de datos y la multitud de formas de obtenerlas dan lugar a un número de cuestiones que no siempre se resuelven satisfactoriamente y requieren de un mayor esfuerzo de investigación. Además, como ya se ha indicado, en la mayoría de los casos, el tratamiento de las bases de datos espaciales (sobre todo, las grandes bases microterritoriales) no debe llevarse a cabo directamente, tal como es proporcionada por los organismos o fuentes originales, sino que exige un tratamiento que haga posible la obtención de una información lo más precisa, depurada y operativa posible.

Esta etapa tiene como propósito delimitar las acciones de organización a desarrollar con respecto al estudio y análisis de la dinámica poblacional. Se obtendrá información disponible relacionada con elementos del territorio: base físico espacial, así como, una caracterización general.

Paso 1. Preparación de los datos.

Para la selección de las variables se tiene en cuenta la disponibilidad de información recopilada, esta información se obtiene de la Oficina Nacional de Estadística e Información (ONEI) y el nivel de particularización de dicha información, esto es, a nivel de provincia, municipio, consejo popular, asentamiento, etc.

En Cuba existen mecanismos para el recopilado de información periódica de múltiples dimensiones; sin embargo, la información para niveles como consejos populares o asentamiento, se limita exclusivamente a los censos de población y vivienda (CPV) que se realizan en el país cada diez años.

Por tanto, un elemento a definir en este paso, dada la información disponible es, la unidad de estudio a la que se asocia un dato espacial.

Un dato espacial se define como la observación de una variable asociada a una localización en el espacio geográfico, esta localización puede expresarse mediante tres tipos de objetos espaciales, los puntos, líneas y polígonos.

- Los puntos se encuentran determinados por las coordenadas terrestres de latitud y longitud, pudiendo corresponderse con personas, empresas, parques, etc.
- Las líneas son objetos que cubren una distancia dada y unen varios puntos o nodos, se determina por las coordenadas de los puntos o nodos que comunican, este sería el caso de las líneas de transmisión, infraestructuras viales, calles de una ciudad, etc.
- Los polígonos son figuras planas conectadas por distintas líneas u objetos cerrados que cubren un área determinada, se determinan mediante las coordenadas de todos los vértices del polígono, a este tipo de objeto espacial le pertenecen los países, provincias, municipios, consejos populares, etc.

A este tipo de dato asociado a un espacio geográfico se le denomina “observación georeferenciada”. Cuando se tienen observaciones georeferenciadas, se deben utilizar herramientas que permitan detectar ciertas características dentro de los datos, como son tendencia, valores atípicos, esquemas de asociación y dependencia espacial, concentración espacial o puntos calientes/fríos, entre otros.

En la actualidad se tiene gran cantidad de información georeferenciada, pero estos datos suelen ser tratados con herramientas del análisis de series temporales (o de corte transversal, no espacial), sin usar técnicas adecuadas para el análisis estadístico espacial.

Paso 2. Cálculo de tasas e indicadores socio-demográficos

A partir de las variables que conforman lo que se podría llamar una base de datos primaria, obtenidos de la ONEI, se puede calcular índices y tasas característicos dentro del análisis demográfico:

La forma más general de expresar la estructura de la población es a través de tres grandes grupos de edades, que representan el grupo de niños y jóvenes o pre-laboral, las edades laborales y los ancianos o edades post-laborales.

Es muy común que la población se estudie combinando los atributos sexo y edad con otros atributos o entre sí. Las dos formas más frecuentes de analizar la población por sexos y edad son la pirámide de población y la relación de masculinidad.

Para la selección de las variables se estudian las propuestas por los especialistas de la ONEI Provincial Cienfuegos que estudian la dinámica de población; la información se corresponde con los datos del Censo de Población y Viviendas del 2012.

Las variables socio-demográficas más comúnmente estudiadas se listan a continuación:

- Coeficiente de Dependencia
- Coeficiente de Vejez
- Tasa de Actividad
- Tasa de Ocupación
- Índice de Masculinidad
- Casados y Unidos
- Divorciados y Separados
- Viudos
- Solteros
- Índice de Educación Superior
- Índice de Técnico Medio
- Índice de Preuniversitario
- Índice de Secundaria
- Índice de Primaria
- Color de la Piel
- Índice de Masculinidad

Excepto el índice de masculinidad todas las demás variables se pueden agrupar de acuerdo al sexo y globalmente.

Generalmente los estudios realizados se enfocan en el análisis de una sola variable o fenómeno. Tal es el caso de los estudios realizados sobre envejecimiento en Cuba: (Barros & Aroca, 2014; Conde, 2020; Rodríguez, 2020)

A continuación, se esboza el significado de las tasas e indicadores más importantes.

Coefficiente de Dependencia

Coefficiente de dependencia (también conocido como Índice de Carga): Es la relación entre la Población No Económicamente Activa (PNEA) y la activa mide el grado de dependencia o carga que en promedio tiene que soportar cada persona disponible para la producción de bienes.

Su análisis debe tener en cuenta que no expresa plenamente el grado real de dependencia, ya que como se conoce la PEA incluye tanto a ocupados, como a no ocupados. Es por ello que los especialistas, siempre que se pueda, prefieren utilizar la relación de Población No Ocupada sobre la Ocupada.

$$CD = \frac{N_i}{N_a} \quad (17)$$

Donde: Ni: Población inactiva

Na: Población Activa

Se supone, este indicador está afectado por el grado de envejecimiento de la población, para este estudio se consideró población activa entre 15 y 64 años, a propuesta de la ONEI.

Coefficiente de Vejez

La causa fundamental del envejecimiento de la estructura por edades de una población lo constituye la reducción de la fecundidad. El indicador más conocido es el Coeficiente de vejez demográfica, que mide la relación que existe entre la población anciana y el total de la población (proporción de ancianos).

$$V = \frac{N_{60 \text{ y más}}}{N_{\text{total}}} \quad (18)$$

Tasa de Actividad

Es la proporción entre la Población Económicamente Activa(PEA), dividida por la Población Total

$$TA = \frac{N_{PEA}}{N_{total}} \quad (19)$$

Tasa de Ocupación

Es la que de manera más fiel refleja la participación de la población en la actividad económica, ya que tiene como punto base a la población ocupada.

$$\varphi = \frac{N_o}{N} \quad (20)$$

Donde: φ : Tasa de ocupación de la población total

N: Población total

N_o : Población ocupada

Razón de Masculinidad

Al analizar la población según sexos, se puede calcular la proporción de uno u otro dentro del total de la población; sin embargo, es más común que las diferencias entre los sexos se analicen a través de la relación o índice de masculinidad. Se obtiene a través de la división por cociente entre los hombres y las mujeres. Se expresa como la cantidad de hombres por cada 100 o 1000 mujeres y puede calcularse para la población total y por grupos de edades

$$RM = \left(\frac{N_{hombres}}{N_{mujers}} \right) 100 \quad (21)$$

En este paso, se reúnen, si está disponible, o calculan todas las tasas e indicadores de esta índole al nivel de unidad geográfica seleccionada. Esta sería una base de datos secundaria, que contiene toda la información sintetizada es indicadores socio-demográficos y que usualmente los especialistas de esta rama del saber, usan para explicar el fenómeno de la dinámica de la población. Será punto de partida para este estudio.

Paso 3. Análisis de indicadores socio-demográficos

La matriz base para el procesamiento se elabora en MS Excel o en la aplicación LibreOffice, donde son introducidos los datos y además calculadas las tasas, índices y coeficientes demográficos con la única condición de que sea guardada como formato dbf, acrónimo de data base format. La primera columna debe contener el identificador de la unidad geográfica seleccionada en el mismo orden en que aparecen en el archivo dbf del mapa asociado a utilizar. La matriz de datos puede organizarse ubicando en las columnas a las variables calculadas (X_j) y en las filas (Y_i), corresponde cada unidad geográfica de estudio (U.G.E), según la tabla 3.

Tabla 3. Matriz base para el procesamiento.

U.G.E	X_1	X_2	X_3	X_4	...	X_5
Y_1						
Y_2						
⋮						
Y_n						

Fuente: Elaboración propia

En los resúmenes demográficos es usual encontrar como técnica ilustrativa las pirámides de población. En la pirámide de población se relaciona la edad, de forma simple o por grupos quinquenales, con el sexo. Es una representación en forma de histograma de la distribución por edad según el sexo de una población. En el eje de las x se relacionan los valores absolutos o relativos de la población según edades y en el eje de las y, las edades. De forma convencional, la parte izquierda de la pirámide corresponde al sexo masculino y la derecha al sexo femenino.

La pirámide muestra de una forma fácil el tipo de estructura de la población, que es consecuencia de la fecundidad, mortalidad y migraciones del territorio estudiado. En la base de la pirámide se encuentran los grupos de edades más jóvenes, la cima representa a los más viejos. Cada barra representa el peso, absoluto o relativo, de una edad o un grupo de edades dentro de la población total.

Para comparar pirámides hay que tener en cuenta tengan la misma escala en la base. Generalmente se usan en valores porcentuales, para facilitar la comparabilidad, igualando las escalas porcentuales. (Erviti & Segura, 2000)

En este paso se realiza, además, a cada indicador el análisis descriptivo según la metodología AED con el objetivo de conseguir un entendimiento básico estos.

Paso 4. Caracterización de la región.

En esta etapa se obtendrá información disponible sobre el caso de estudio relacionadas con elementos del territorio: base físico espacial e infraestructuras construidas, así como una caracterización general considerando las dimensiones poblacional, económica, social y ambiental.

El reconocimiento de la dinámica demográfica básica, asumiendo para ello el estudio realizado para el Fondo de Población de Naciones Unidas (FPNU,2016), debe realizarse a través de indicadores demográficos básicos que permiten establecer la intensidad y tendencias en el comportamiento de las variables demográficas (fecundidad, mortalidad y migración) o mediante los cambios observados en los siguientes componentes: tamaño y crecimiento, estructura y composición, distribución espacial y movilidad de la población, los cuales se describen a continuación.

1. Tamaño y crecimiento

Mientras que el tamaño de población describe la magnitud de la población residente en la unidad de estudio en un momento determinado, el crecimiento informa sobre la velocidad con la que cambia el tamaño de la población y con la que cambiará a futuro debido al crecimiento natural o vegetativo y al crecimiento migratorio. El crecimiento natural o vegetativo de la población resulta del balance entre los nacimientos y las defunciones que se observan en un territorio particular. La población puede también aumentar por inmigración o disminuir por emigración (saldo migratorio). El crecimiento total de la población resulta del balance entre el saldo vegetativo y el saldo migratorio. De estos factores y su constante cambio dependerá el tamaño de la población que reside en el territorio.

2. Estructura y composición

Los individuos que componen una población son diversos en sus características, es por ello que se hace necesario considerar su estructura por edad y sexo. Usualmente la información sobre la población se agrupa en quinquenios de edad para construir las pirámides o perfiles de población, que permiten identificar equilibrios y desequilibrios entre sexos y generaciones.

La estructura poblacional, en la medida que ofrece informaciones de grupos etarios permite determinar demandas específicas de bienes y servicios en temas como educación, salud, oportunidades económicas, atención a grupos en situación de vulnerabilidad, protección de la familia y la niñez, entre otros.

Existen otras características sociales y económicas como nivel educativo, estado civil, actividad económica y condición socio-económica que son relevantes para la comprensión de la composición de la población de cada municipio.

3. Distribución espacial

Así como los distintos grupos de población tienen características diversas, su distribución en el territorio tampoco es homogénea. El análisis de distribución espacial indica la localización de la población dentro del territorio. Esta distribución espacial se concreta en el número de asentamientos poblacionales que existen en la unidad de estudio.

El análisis del asentamiento de la población en el territorio permite determinar demandas de recursos físicos y ambientales, como agua, suelo, alcantarillado, energía eléctrica, espacio público, infraestructura de transporte, infraestructura social como escuelas y centros de salud, y equipamientos comunitarios y económicos (empresas, servicios financieros, servicios relacionados con la producción, entre otros).

4. Movilidad

La movilidad hace referencia al desplazamiento de carácter temporal, estacional o intermitente de población residente y no residente en el municipio que hace uso del territorio en función de sus actividades. Es necesario precisar no solo la magnitud, sino las razones de movilidad (salud, educación, trabajo, etc.) de la población a fin de dimensionar adecuadamente la oferta de áreas y servicios según las necesidades.

La información sobre movilidad de la población permite hacer una buena definición de suelo urbano, rural y de expansión y de los usos asociados. El conocimiento de la movilidad de la población dentro de la unidad de estudio, y desde o hacia otros territorios vecinos, permite definir no solo los equipamientos a ese nivel de particularidad, sino también la necesidad de mejorar la conectividad entre áreas urbanas y rurales.

2.3 Etapa 2. Análisis Exploratorio de Datos Espaciales

El análisis exploratorio de datos espaciales - AEDE, constituye una disciplina reciente que ha adquirido una especial importancia debido principalmente al avance de la tecnología en las comunicaciones y la globalización de la economía.

Antes de comenzar describiendo cualquier método estadístico de análisis de datos espaciales es necesario definir qué se debe entender como datos espaciales. Un dato espacial puede ser definido como la observación de una variable asociada a una localización del espacio geográfico.

Cuando se tienen observaciones georeferenciadas, se deben utilizar herramientas que permitan detectar ciertas características dentro de los datos, como son tendencia, valores atípicos, esquemas de asociación y dependencia espacial, concentración espacial o puntos calientes/fríos, entre otros.

Aquí las herramientas gráficas juegan un papel muy importante en la visualización de los datos en el AEDE y los mapas se convierten en un elemento esencial en la visualización. La interacción entre las gráficas, los mapas, las tablas y los parámetros de las gráficas son tan importantes que se requiere obtener rápidamente los mapas apropiados, así como, una modificación de los parámetros de manera interactiva para poder proveer diferentes tipos de visualizaciones, lo cual se logra con enlaces entre la cartografía y las presentaciones estadísticas.

Los mapas y los gráficos estadísticos deben ser analizados conjuntamente ya que se complementan unos a otros en el análisis exploratorio de datos espaciales. La presentación de los datos hace referencia a lo que se muestra en un gráfico y qué tipo de gráfico se construye. Los gráficos para el caso univariado incluyen entre otros los histogramas, los boxplots, los gráficos cuantil-cuantil y los gráficos de probabilidad.

Los gráficos para el caso multivariado incluyen el scatterplot (donde los casos son representados como puntos), los gráficos de ejes coordenados paralelos, los gráficos en tres dimensiones (caso de tres variables) y los glifos (donde los casos son representados como símbolos complejos, como el caso de los rostros de Chernoff).

Teniendo en cuenta que las diferentes presentaciones de los datos sirven para identificar características dentro de estos, son presentados algunos de los diferentes tipos de gráficos que se pueden construir, de acuerdo con las características que se pueden obtener de ellos.

Paso 1. Identificación de atípicos

Para la detección de valores atípicos se cuenta con varias técnicas del análisis de datos (AED) y que se complementan con la visualización en mapas. Algunas de estas se comentan a continuación.

Diagrama de Caja - Mapa

El diagrama de caja, caja de bigotes o boxplot es un gráfico que corresponde a la representación gráfica de cinco índices numéricos de las observaciones realizadas, que son en orden ascendente, valor mínimo observado, cuartil inferior o primer cuartil, mediana, cuartil superior o tercer cuartil y valor máximo observado. Los bigotes se encuentran en los valores máximo y mínimo de las observaciones y la caja se encuentra trazada entre los dos cuartiles,

con una línea que corresponde a la localización de la mediana. Una representación más refinada de los diagramas de caja, localizan los bigotes en las siguientes ubicaciones:

$$\text{Bigote inferior} = \min(x_{\min}, Q_1 - 1,5(Q_3 - Q_1)) = \min(x_{\min}, Q_1 - 1,5RIQ)$$

$$\text{Bigote superior} = \min(x_{\max}, Q_3 + 1,5(Q_3 - Q_1)) = \max(x_{\min}, Q_3 - 1,5RIQ)$$

Donde:

Q_1 : Es el primer cuartil

Q_2 : Es la mediana

Q_3 : Es el tercer cuartil

$Q_3 - Q_1 = RIQ$: Es el rango intercuartílico

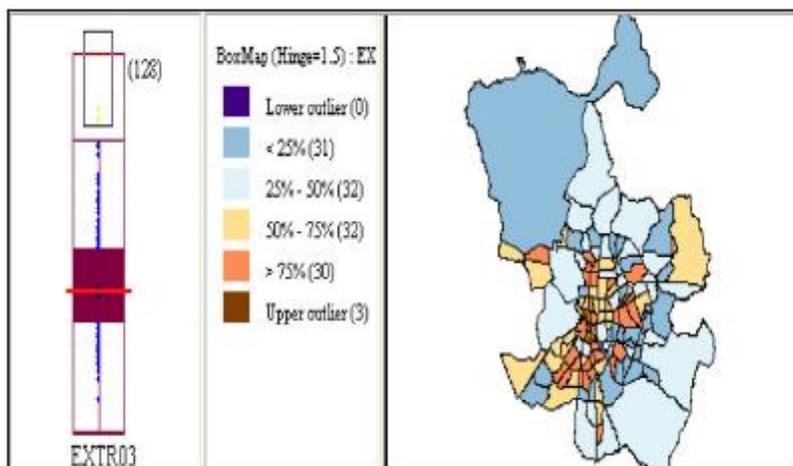
x_{\min} y x_{\max} : Son los valores mínimo y máximo observados

De esta forma se consideran valores atípicos o outliers las observaciones que se encuentran por encima del bigote superior o por debajo del bigote inferior. El rango de 1,5 veces el recorrido intercuartílico puede ser cambiado por 3, el cual constituye un criterio más estricto para definir una observación como valor atípico. Al trabajar con un paquete estadístico espacial y seleccionar en el diagrama de caja las observaciones atípicas, se podrá ver en el mapa a que unidades espaciales corresponden dichas observaciones.

Mapa de Caja o Boxmap

Un mapa de caja es una versión del mapa de cuartiles, en el cual aparecen resaltados los valores atípicos o outliers que se encuentran por encima o por debajo de 1.5 (o 3.0) veces el rango intercuartil. La clasificación es la misma que la realizada en el diagrama de caja o Boxplot.

Figura 2. Gráfico y Mapa de cajas.



Fuente: (Yrigoyen, 2006)

Diagrama de Dispersión - Mapa

En el caso multivariado para dos variables es útil realizar el gráfico de dispersión o scatterplot donde cada una de las variables es representada en cada uno de los ejes y cada punto correspondería a una unidad espacial. Los puntos que se encuentran alejados de la nube de puntos serán los atípicos y utilizando los gráficos interactivos, las unidades espaciales correspondientes a dichos puntos, podrían ser identificados fácilmente en el mapa.

Paso 2. Identificación de la distribución

Histograma – Mapa

El histograma es la gráfica que representa las frecuencias de los valores observados y es una aproximación discreta de la función de densidad de probabilidad de una variable aleatoria.

Al construir el histograma se puede observar la distribución de la variable observada en el área de estudio, al trabajar con diagramas interactivos se puede observar en el mapa las unidades espaciales que pertenecen a cada uno de los intervalos. Adicionalmente si se quiere ver la distribución de la variable en una sub-región perteneciente al área de estudio, al seleccionar la subregión se puede observar en el histograma resaltada la distribución de la variable para un subconjunto de unidades espaciales, es decir el histograma regional. Si al comparar el histograma del área de estudio con el histograma regional resaltado estos difieren entre sí, este hecho sugeriría la posibilidad de la existencia del fenómeno espacial denominado heterogeneidad espacial.

Mapa de cuantiles

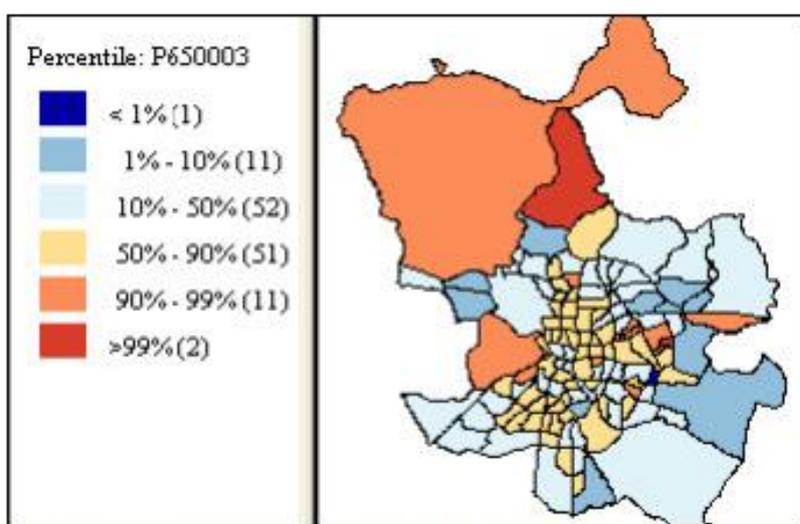
Es un mapa donde aparecen la distribución de las unidades espaciales de acuerdo con los cuantiles, es decir, que las unidades que pertenecen al mismo cuartil tienen el mismo color. Puede ser elaborado con cuartiles, quintiles o percentiles.

Mapa de percentiles

El mapa de percentiles es un caso particular del mapa de cuantiles donde los datos se dividen y agrupan en una serie de categorías con igual número de observaciones. Por ejemplo, un mapa de percentiles será aquel cuya distribución se divide en 100 grupos con igual número de unidades territoriales.

Este tipo de mapa no es útil en casos en los que la variable de interés contenga un gran número de observaciones con valores parecidos.

Figura 3. Mapa de percentiles



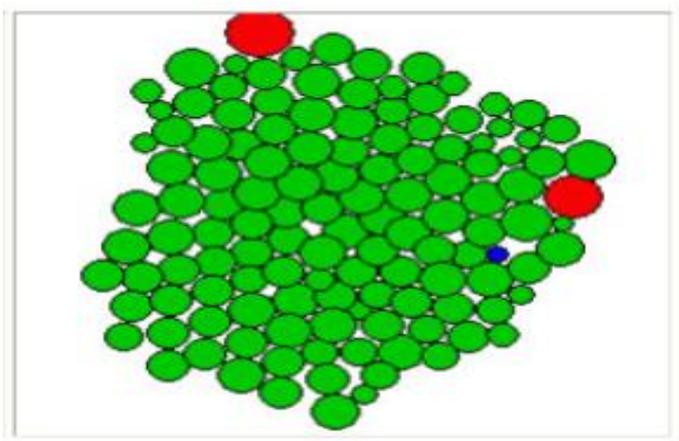
Fuente: (Yrigoyen, 2006)

Paso 3. Detección de efectos espaciales

Cartograma

El cartograma es una representación gráfica del mapa donde cada unidad espacial es reemplazada por un círculo cuya área es proporcional al valor de la variable en dicha unidad espacial. El cartograma permite identificar valores atípicos y patrones espaciales en las unidades espaciales del área en estudio, ya que las unidades espaciales con los valores atípicos se resaltan con colores diferentes del resto de las unidades

Figura 4. Cartograma



Fuente: (Yrigoyen, 2006)

Diagrama de dispersión de Moran

El scatterplot de Moran es un diagrama de dispersión donde se representa la variable en estudio y el retardo espacial de dicha variable. El diagrama se obtiene representando en un plano cartesiano n pares de valores.

$$\left\{ \sum_{j \in N(i)} w^*(i, j) z(j), z(i) \right\}_{i=1, \dots, n}$$

Donde:

$w^*(i, j)$: denota la (i, j) -ésima entrada de la matriz estandarizada de ponderaciones

$N(i)$: denota el conjunto de vecinos de la unidad i ($i \neq N(i)$)

El eje vertical es usado para los valores de los vecinos espacialmente promediados y el eje horizontal para la unidad espacial que se encuentra en el centro del promedio espacial. Anselin ha denominado este como el scatterplot de Moran (Anselin 1998). El estadístico I de Moran de la autocorrelación espacial es el estimador de la pendiente de la regresión por mínimos cuadrados ordinarios de $\left\{ \sum_{j \in N(i)} w^*(i, j) z(j), z(i) \right\}_{i=1, \dots, n}$ en $\{z(i)\}_{i=1, \dots, n}$. Cuanto mayor sea la pendiente obtenida mayor será el grado de autocorrelación espacial.

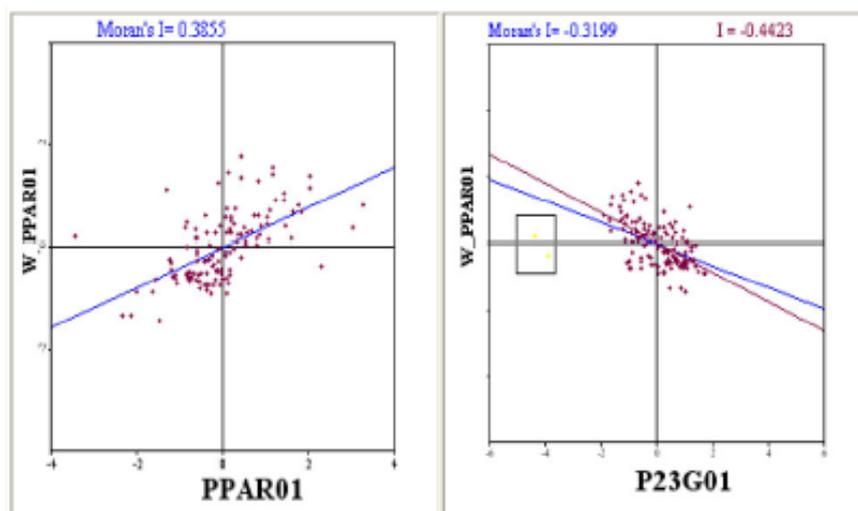
Este diagrama de dispersión divide el tipo de asociación espacial en cuatro categorías:

- Dos para autocorrelación espacial positiva, localizadas en los cuadrantes I, donde unidades espaciales con valores altos de la variable analizada están rodeadas de unidades espaciales con valores también altos en la característica analizada; y III, donde

unidades con valores bajos de la variable están rodeadas de unidades con valores bajos de la variable.

- Dos para autocorrelación espacial negativa, localizadas en los cuadrantes II, donde unidades espaciales con valores bajos de la variable en estudio están rodeadas de unidades espaciales con valores altos de la variable: y IV en el cual unidades espaciales con valores altos en la variable se encuentran rodeadas de unidades espaciales cuyos valores de la variable son bajos.

Figura 5. Mapa de dispersión de Morán.

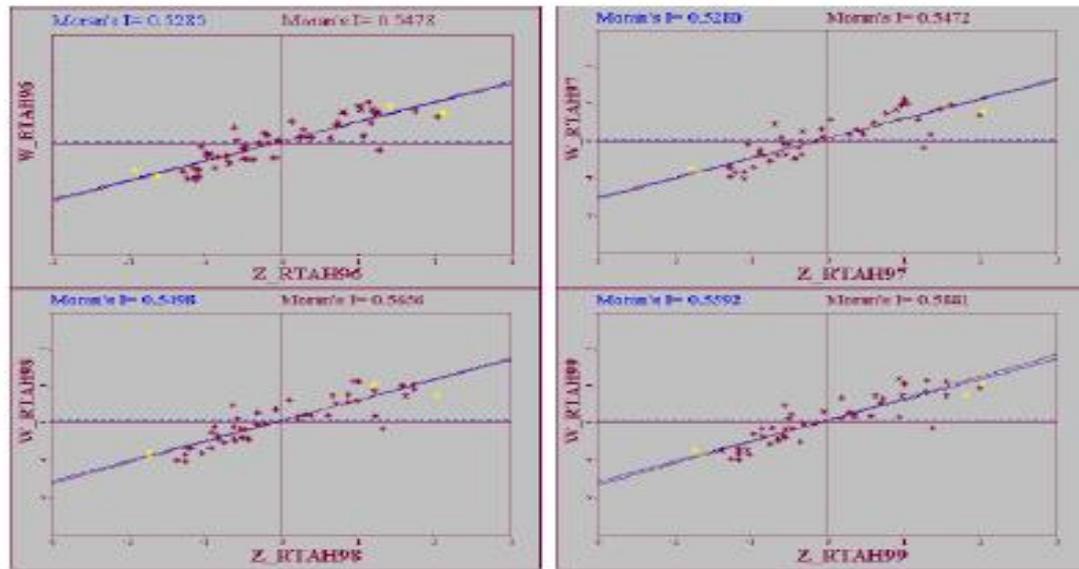


Fuente: (Yrigoyen, 2006)

Diagrama de dispersión de Moran Multivariante

Este diagrama multivariante es implantado en un entorno dinámico, permite comparar el comportamiento del fenómeno de asociación espacial en varios indicadores, así como, analizar el comportamiento de determinadas unidades geográficas que, por ejemplo, se presentan con valores atípicos o extremos.

Figura 6. Mapa de dispersión de Morán Multivariante.



Fuente: (Yrigoyen, 2006)

Boxplots Rezagados

Los Boxplots calculan separadamente los subconjuntos espaciales y luego los superponen, su comparación puede indicar diferencias entre las áreas. Las unidades espaciales que componen cada sub-área pueden no ser equivalentes, por ejemplo, pueden variar en tamaño, por lo tanto, la ponderación podría ser considerada de la siguiente manera.

Sea $z(1), \dots, z(n(R)), (n(r))$: denota los valores para la región r y $w(1), \dots, w(n(R))$ las correspondientes ponderaciones para cada una de las unidades espaciales, con $r = 1; \dots; R$.

El propósito es hacer comparaciones sobre R subregiones con $R \geq 2$.

Para encontrar la mediana ponderada se ordenan las observaciones de tal forma que $z([1]) \leq z([2]) \dots \leq z([R])$. Las ponderaciones se reordenan de tal forma que correspondan con el orden en $\{z(i)\}$.

La suma acumulada de las ponderaciones reordenadas es:

$$S(k) = \sum_{j=1}^k w([j]) \quad k = 1, 2, \dots, n(r) \quad (22)$$

El índice t de la mediana ponderada es el valor más pequeño de k , tal que $S(k)$ es al menos la mitad de la suma total de las ponderaciones:

$$t = \min \{k | S(k) \geq S(n(r))/2\}$$

Si $S(t)$ es estrictamente mayor que $S(n(r))/2$ la mediana ponderada es $z(t)$. Si $S(t)$ es igual a $S(n(r))/2$ la mediana ponderada es el promedio entre $z(t)$ y $z(t + 1)$. Los cuartiles superior e inferior pueden ser calculados usando el mismo principio y el boxplot construido.

Las medianas ponderadas son utilizadas en varios contextos que incluyen la suavización de los datos espaciales. Las ponderaciones se pueden realizar basándose en el área, en la distancia hasta un punto dado o el tamaño de la población entre otras.

Los boxplots espacialmente rezagados son una serie ordenada de boxplots. El orden es con respecto a una área o sitio determinado. Cada boxplot usa el subconjunto de los datos que están a una distancia dada del área escogida.

Identificación de clúster

Un clúster puede ser definido como la agrupación de unidades espaciales localizadas en un área específica, que concentran valores más elevados o bajos de lo que cabría esperar en caso de encontrarnos ante una distribución homogénea, dominando la aleatoriedad en el resto del territorio estudiado.

Una de las dificultades con las que se han encontrado los analistas de fenómenos espaciales ha sido la identificación de clúster espaciales en los mapas. Existen diversos métodos para realizar la identificación de estos clústeres, en el presente trabajo se recogen las ideas centrales del desarrollado realizado por Jared Aldstadt y Arthur Getis, quienes propusieron un algoritmo denominado AMOEBA (A Multidirectional Optimum Ecotope-Based Algorithm), para la construcción de clústeres espaciales teniendo como referencia las colas de la distribución. Este método es multidireccional, ya que se comienza en una unidad espacial que se considera "semilla" y la búsqueda se realiza en todas las direcciones haciendo comparaciones entre el conjunto total de unidades espaciales. Este algoritmo comparte algunas características con los métodos de segmentación de imágenes, cuyo objetivo es partir un gran número de píxeles en regiones homogéneas a partir de un número pequeño de píxeles semilla.

El algoritmo utiliza el estadístico local de Getis-Ord, el cual para una unidad espacial i está definida como:

$$G_i^* = \frac{\sum_{j=1}^N w_{ij} x_j - \bar{x} \sum_{j=1}^N w_{ij}}{S \sqrt{\frac{\sum_{j=1}^N w_{ij}^2 - (\sum_{j=1}^N w_{ij})^2}{N-1}}} \sim N(0,1) \quad (23)$$

$$S = \sqrt{\frac{\sum_{j=1}^N x_j^2}{N-1} - (\bar{x})^2}$$

(24)

Donde:

N Es el número total de unidades espaciales

x_j Es el valor del fenómeno en estudio en la unidad j

\bar{x} Es la media de todos los valores

$$w_{ij} = \begin{cases} 1 & j \text{ es vecino de } i \\ 0 & \text{en otro caso} \end{cases}$$

Los elementos utilizados para una prueba de hipótesis son:

H₀: No existe autocorrelación

H₁: Existe autocorrelación

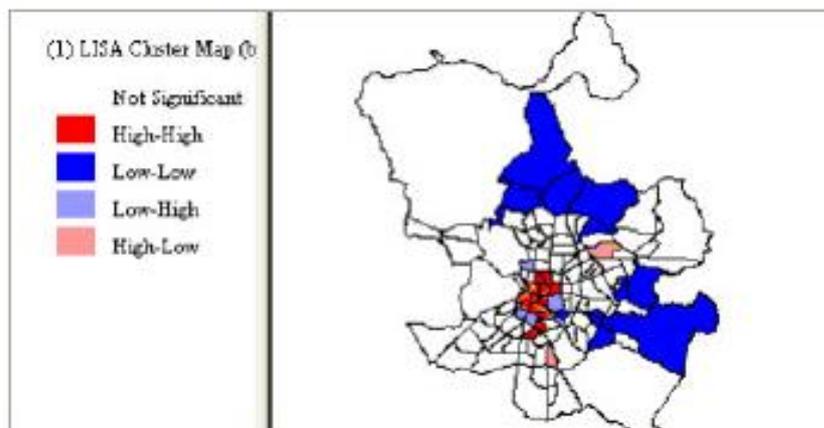
Mapa de LISA

Las letras LISA significan “Local Indicator of Spatial Association” (Indicador local de asociación espacial). En el mapa LISA se representan las localizaciones con valores significativos en indicadores estadísticos de asociación espacial local permitiendo observar la presencia de puntos calientes (“hot spots”) o atípicos espaciales, cuya mayor o menor intensidad dependerá de las medidas de asociación espacial global.

Goodchild (1987) define la autocorrelación espacial (AE) como la concentración o dispersión de los valores de una variable en un mapa; es decir refleja el grado en que objetos en una unidad geográfica son similares a otros objetos en unidades geográficas próximas.

El análisis de autocorrelación espacial global realiza un examen conjunto de todas las unidades que componen la muestra para determinar si las unidades espaciales se encuentran distribuidas aleatoriamente o si, por el contrario, lo hacen conforma a un patrón determinado.

Figura 7. Mapa de LISA.



Fuente: (Yrigoyen, 2006)

Según lo comentado por Moreno y Vallá (2002) “entre dichos estadísticos globales, los cuales resumen en un indicador único el esquema general de dependencia, se encuentran la I de Moran (Moran, 1948) y la G(d) (Getis y Ord, 1992)”, se presentan algunas pruebas de autocorrelación global.

Tabla 4. Estadísticos de Asociación global

I de Moran	G(d) de Getis y Ord
$I = \frac{N}{S_0} * \frac{\sum_{i \neq j}^N w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{S_{i=1}^N (x_i - \bar{x})^2}$ <p> x_i: valor de la variable x en la región i \bar{x}: media muestral de la variable x w_{ij}: pesos de la matriz W N: tamaño muestral $S_0 = \sum_i \sum_j w_{ij}$ </p>	$G(d) = \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij}(d) x_i x_j}{\sum_{i=1}^N \sum_{j=1}^N x_i x_j} \quad i \neq j$ <p>Donde los pares de regiones i y j son vecinas si se encuentran dentro de una distancia determinada</p>
<p>Interpretación tras su estandarización:</p> <ul style="list-style-type: none"> • Z(I)>0 y signif.: autocorrelación (+) • Z(I)<0 y signif.: autocorrelación (-) 	<p>Interpretación tras su estandarización:</p> <ul style="list-style-type: none"> • Z(G)>0 y signif.: concentración de valores elevados • Z(G)<0 y signif.: concentración de valores bajos

Fuente: (Moreno & Vayá, 2002)

Los estadísticos presentados en el anterior permiten contrastar la hipótesis nula de no autocorrelación espacial, es decir, la hipótesis nula de la existencia de una distribución aleatoria de la variable a lo largo del territorio.

Sobre los estadísticos de autocorrelación espacial local Moreno y Vallá (2002) mencionan lo siguiente:

Las pruebas de autocorrelación espacial global analizan todas las observaciones de la muestra de forma conjunta, no resultan sensibles a situaciones donde predomine una importante inestabilidad en la distribución espacial de la variable objeto de estudio (procesos no estacionarios espacialmente), es decir, no contemplan la posibilidad de que el esquema de dependencia detectado a nivel global (por ejemplo, ausencia de autocorrelación espacial) pueda no mantenerse en todas las unidades del espacio analizado. Dicha limitación es fácilmente superable por medio del cálculo de los denominados contrastes locales de asociación espacial entre los que se encuentran el estadístico local de Moran I_i , (Anselin, 1995) y los test New-Gi(d) y New-Gi*(d) (Ord y Getis, 1996).

Tabla 5. Estadísticos de Asociación Espacial Local

Local de Moran I_i	G(d) de Getis y Ord
$I = \frac{Z_i}{\sum_i Z_i^2 / N} \sum_{j \in J_i} w_{ij} Z_j$ <p>z_i: valor correspondiente a la región i de la variable normalizada J_i: conjunto de regiones vecinas a i</p>	$New - G_i^* = \frac{\sum_{j=1}^N w_{ij} x_j - W_i^* \bar{x}}{S \{ [N S_{ii}^* - W_i^{*2}] / (N - 1) \}^{1/2}}$ $W^* = W_i + w_{ir}$ $S_{ii}^* = \sum_j w_{ij}^2$ $S^2 = \frac{1}{N - 1} \sum_j (x_j - \bar{x})^2$
<p>Interpretación tras su estandarización:</p> <ul style="list-style-type: none"> • $Z(I_i) > 0$ y signif.: cluster de valores similares alrededor de i • $Z(I_i) < 0$ y signif.: cluster de valores disímiles alrededor de i 	<p>Interpretación tras su estandarización:</p> <ul style="list-style-type: none"> • $New - G_i^* > 0$ y signif.: cluster de valores similares elevados alrededor de i • $New - G_i^* < 0$ y signif.: cluster de valores similares bajos alrededor de i

Fuente: (Moreno & Vayá, 2002)

Los estadísticos presentados en el cuadro anterior obtienen un valor para cada observación de la muestra, con el objetivo de analizar la situación de cada unidad espacial por separado.

Cuando no existe un marco formal o teoría previa acerca del fenómeno que se analiza, deben de utilizarse las técnicas del AEDE. Esta situación se plantea cuando se analizan grandes

bases de datos geográficos cuya distribución no se conoce a priori. El AEDE combina el análisis estadístico con el gráfico, dando lugar a lo que podría denominarse una “visualización científica” (Haining, 2003) que, a los contrastes estadísticos sobre los efectos espaciales de dependencia y heterogeneidad, une un amplio marco de gráficos o “vistas” múltiples y dinámicas sobre la información geográfica .

Tabla 6. Resumen de algunas técnicas de visualización de AEDE.

	Perspectiva de Econometría Espacial
Visualización de distribuciones espaciales	Mapa de Caja, Histograma, análisis de la varianza exploratorio espacial.
Visualización de asociación espacial global	Gráficos del retardo espacial, diagrama y mapa de dispersión de Moran.
Visualización de asociación espacial local	Mapas LISA, atípicos en diagrama de dispersión de Moran.
Asociación espacial multivariante	Diagrama de dispersión multivariante de Moran
Heterogeneidad espacial	Mapa , histograma, diagrama de dispersión.

Fuente: (Acevedo & Velázquez, 2008)

2.4 Etapa 3. Análisis de Resultados

Articulando todos los resultados obtenidos en cada una de las etapas y pasos se puede elaborar una descripción detallada de cada uno de los indicadores analizados, identificando localizaciones atípicas, descubriendo formas de asociación (autocorrelación espacial) que, a su vez, pueden ser de carácter global o local, y sugiriendo estructuras en el espacio geográfico (heterogeneidad espacial).

Paso 1: Resumir los patrones de asociación detectados

Los patrones de asociación deben ser explicados atendiendo a dos aristas fundamentales.

- 1- Explicar la presencia o no de heterogeneidad.
- 2- Explicar la presencia de autocorrelación local y global.

Los patrones de asociación detectados, pueden ser visualizados en una tabla como se muestra:

Tabla 7. Tabla resumen de resultados

Indicadores	Alto-Alto	Bajo-Bajo	Bajo-Alto	Alto-Bajo

Paso 2: Formar los clústeres

Se deben utilizar ciertas técnicas especialmente diseñadas para encontrar algún “sentido” en los datos, es decir, para la detección en ellos de estructuras o conglomerados.

De este modo, es posible obtener una estructura explicativa de los datos a través de técnicas que combinan herramientas de la estadística básica (descriptivos, correlaciones, tablas de frecuencias o de correlación cruzada) con análisis multivariante avanzado, especialmente diseñado para identificar formas en grandes bases de datos (análisis clúster, escalas multidimensionales, análisis logit, correlación canónica, análisis de correspondencias, árboles de clasificación, etc.).

En resumen, del capítulo, se concluye que el procedimiento diseñado debe permitir procesar los datos asociados a cualquier unidad geográfica, vinculado a un sistema de información geográfica, ofreciendo resultados apoyado en la capacidad de visualización a través de mapas. El procedimiento permite a través de la metodología AEDE detectar valores atípicos, así como, patrones de asociación espacial y de heterogeneidad. Los resultados pueden ser presentados a través de mapas y gráficas, que ofrecen información de los grupos resultantes de este análisis.

Capítulo 3. Aplicación del procedimiento para la identificación de patrones de asociación espacial.

3.1 Introducción

En el presente capítulo se aplica el procedimiento propuesto para el caso de los consejos populares y expone los resultados para cada paso del procedimiento.

3.2 Etapa 1 Reconocimiento del territorio

La ONEI está interesada en estudiar los Consejos Populares pues estos constituyen la unidad de gobierno más cercana a la población, esta decisión favorece la investigación pues con una población de 404 228 habitantes, trabajar en 275 asentamientos pueden aparecer asimetrías, asentamientos muy poblados y otros, poco poblados. En el asentamiento de Cienfuegos está el 27% de la población de la provincia. En el estudio preliminar se observaron indicadores nulos en los asentamientos¹, esto no ocurre a nivel de consejos populares, tributando esto a la coherencia de los datos.

Para los estudios de población la ONEI trabaja las variables a nivel de municipio, provincia y país, pero no a nivel de Consejo Popular, siendo esta la unidad elemental a nivel de gobierno.

Paso: 1 Preparación de los datos.

La ONEI facilita la información en una base de datos en formato Excel con la siguiente estructura:

Tabla 8. Base de Datos primaria

	P4_SEXO		P17		P4_SEXO		P17		P4_SEXO		P17		P4_SEXO		P17	
	Total	Masculino	Femenino	Total	Masculino	Femenino	Total	Masculino	Femenino	Total	Masculino	Femenino	Total	Masculino	Femenino	
Elpidio Gómez	3118	1602	1516	1252	880	372	18	12	6	23	18	5	6	6		
Arriete-Ciego Montero	4886	2536	2350	1888	1358	530	43	18	25	31	22	9	14	12		
Espartaco	3892	1985	1907	1334	930	404	51	20	31	43	35	8	7	6		
Palmira Sur	6092	3057	3035	2435	1521	914	67	29	38	61	50	11	22	15		
Palmira Norte	8626	4315	4311	3655	2267	1388	121	30	91	139	69	70	29	10		
San Fdo de Camarones	6325	3322	3003	1986	1332	654	57	28	29	108	99	9	29	23		
Paradero de Camarones	1623	830	793	559	383	176	14	3	11	15	13	2	1	0		
Piragua	1068	562	506	424	322	102	9	4	5	5	4	1	3	3		
Ajuria	2108	1152	956	860	636	224	14	4	10	29	22	7	6	5		

¹ (Conde, 2020)

Esta base de datos primaria es un libro que contiene 5 hojas, que refieren a: actividad, estado civil, edad, color de la piel y nivel educacional, todas ellas tienen los datos totalizados y desglosados por sexo.

La hoja Actividad tiene la cantidad de personas contempladas en las categorías siguientes: Trabajo, Tenía trabajo pero no trabajó, Buscaba porque había perdido, Buscaba trabajo 1ra vez, Jubilado o Pensionado por edad, Otros Pensionados, Rentista o recibe ayuda económica, Quehaceres del hogar, Estudiante, Incapacitado para el trabajo, No realiza ninguna actividad, En hospital, asilo, etc., Otra Situación, No aplica (menos de 12 años)

La hoja Estado civil tiene la cantidad de personas contempladas en las categorías siguientes: Casado, Unido, Divorciado, Separado, Viudo, Soltero, No aplica (por edad). Para el estudio se unen casados con unidos y divorciado y separado y se trabaja como un solo indicador.

La hoja Edad tiene la cantidad de personas contempladas en las categorías siguientes: está compuesta por grupos atareos de 5 años comenzando de 0 a 4 años, y terminando de 80 a 84 años y más de 84 años.

La hoja Color de la piel tiene la cantidad de personas contempladas en las categorías siguientes por: Negros, Mestizos y Blancos.

La hoja Nivel Educacional tiene la cantidad de personas contempladas en las categorías siguientes: Ninguno (quienes no tiene vencido el nivel primario), Primaria, Secundaria básica, Obrero calificado, Preuniversitario, Técnico Medio, Pedagogía Nivel Medio, Superior o Universitario, No aplica (menores de 12 años).

Paso 2: Cálculo de tasas e indicadores socio-demográficos

Con esta base de datos primaria se procede a la confección de una matriz base, como se aprecia en la Tabla 10, que contiene como variables a los indicadores, tasas y coeficientes, socio – demográficos (véase Tabla 9), que fueron calculados a nivel de consejo popular, que será procesada estadísticamente.

Tabla 9. Resumen de la descripción de las variables

Variable	Indicador	Descripción del indicador
Dependencia Económica	CoeDepT	Coeficiente de dependencia global
	CoeDepM	Coeficiente de dependencia masculino
	CoeDepF	Coeficiente de dependencia femenino
Envejecimiento	CoeVejT	Coeficiente de vejez global
	CoeVejM	Coeficiente vejez masculino
	CoeVejF	Coeficiente vejez femenino
Ocupación	TasOcuT	Tasa de ocupación global

Actividad	TasOcuM	Tasa de ocupación masculino
	TasOcuF	Tasa de ocupación femenino
	TasAtcT	Tasa de actividad global
	TasAtcM	Tasa de actividad masculino
	TasAtcF	Tasa de actividad femenino
Masculinidad	IndMasc	Hombre/Mujeres

Fuente: Elaboración propia

Otras variables bajo estudio son:

- Estado Civil
 - Casados y unidos
 - Divorciados y separados
 - Solteros
 - Viudos
- Educación
 - Universidad
 - Técnico Medio
 - Preuniversitario
 - Secundaria
 - Primaria
- Color de la Piel
 - Negro
 - Mestizo
 - Blanco

Las variables se tratan como coeficientes y tasas según corresponde, para el total y por sexo, en las tasas se considera $k=1$. El resto de los indicadores se calculan como proporciones con respecto a la población, por sexo y total. Al usar una proporción con respecto a la cantidad de habitantes por Consejo Popular se amortizan las asimetrías. Excepto el índice de masculinidad todas están desglosadas por sexo, reflejando en total 55 variables.

Tabla 10. Matriz base de indicadores calculados para Consejo Popular

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	CP	CoeDepT	CoeDepM	CoeDepF	CoeVejT	CoeVejM	CoeVejF	TasOcuT	TasOcuM	TasOcuF	TasAtcT	TasAtcM	TasAtcF	CasUniT	CasUnim
2	Elpidio Gómez	0,31	0,30	0,32	0,20	0,20	0,19	0,41	0,56	0,25	0,69	0,70	0,62	0,56	0,55
3	Arriete-Ciego Montero	0,32	0,32	0,32	0,19	0,18	0,19	0,40	0,54	0,24	0,68	0,68	0,62	0,53	0,52
4	Espartaco	0,32	0,31	0,32	0,21	0,21	0,21	0,36	0,48	0,23	0,68	0,69	0,62	0,55	0,54
5	Palмира Sur	0,31	0,32	0,30	0,19	0,19	0,20	0,41	0,51	0,31	0,69	0,68	0,64	0,54	0,53
6	Palмира Norte	0,30	0,29	0,31	0,16	0,15	0,17	0,44	0,53	0,34	0,70	0,71	0,65	0,50	0,50
7	San Fdo de Camarones	0,33	0,33	0,33	0,23	0,22	0,24	0,32	0,41	0,23	0,67	0,67	0,60	0,53	0,52
8	Potrerillo	0,35	0,34	0,37	0,25	0,24	0,26	0,37	0,53	0,20	0,65	0,66	0,58	0,55	0,53
9	Isla de Pinos	0,31	0,30	0,33	0,20	0,17	0,22	0,46	0,56	0,36	0,69	0,70	0,62	0,49	0,48
10	Paradero de Camarones	0,32	0,30	0,33	0,19	0,17	0,22	0,35	0,47	0,24	0,68	0,70	0,61	0,56	0,55
11	Mal Tiempo	0,33	0,32	0,34	0,21	0,21	0,20	0,38	0,48	0,27	0,67	0,68	0,62	0,57	0,54
12	Las Nubes	0,33	0,31	0,34	0,21	0,19	0,23	0,44	0,55	0,34	0,67	0,69	0,60	0,50	0,51
13	San José	0,33	0,32	0,34	0,21	0,20	0,23	0,44	0,56	0,32	0,67	0,68	0,60	0,51	0,51
14	Marta Abreú	0,30	0,28	0,32	0,17	0,17	0,17	0,37	0,46	0,27	0,70	0,72	0,63	0,52	0,50
15	Chicharones-El No	0,34	0,34	0,35	0,22	0,22	0,22	0,33	0,46	0,17	0,66	0,66	0,60	0,62	0,58
16	Lajas Norte	0,32	0,32	0,32	0,20	0,19	0,20	0,50	0,58	0,41	0,68	0,68	0,64	0,51	0,50
17	Lajas Sur	0,33	0,33	0,32	0,22	0,22	0,23	0,48	0,58	0,38	0,67	0,67	0,61	0,53	0,52
18	Piragua	0,32	0,31	0,34	0,21	0,20	0,21	0,41	0,58	0,21	0,68	0,69	0,60	0,58	0,55
19	Ajuria	0,32	0,33	0,31	0,18	0,18	0,17	0,41	0,56	0,24	0,68	0,67	0,64	0,60	0,56
20	Caracas	0,32	0,32	0,33	0,23	0,23	0,22	0,40	0,53	0,25	0,68	0,68	0,61	0,61	0,61
21	Ramón Balboa	0,32	0,31	0,33	0,19	0,19	0,20	0,35	0,50	0,20	0,68	0,69	0,62	0,55	0,53
22	Salto- Modelo	0,33	0,33	0,34	0,22	0,23	0,21	0,32	0,44	0,18	0,67	0,67	0,61	0,60	0,58
23	5 de Septiembre	0,30	0,31	0,30	0,19	0,19	0,19	0,39	0,54	0,21	0,70	0,69	0,64	0,52	0,48
24	Santiago de Cartagena	0,31	0,31	0,31	0,18	0,18	0,18	0,34	0,51	0,13	0,69	0,69	0,66	0,57	0,53
25	Ariza	0,31	0,31	0,30	0,18	0,18	0,18	0,44	0,58	0,30	0,69	0,69	0,65	0,53	0,51
26	Cartagena	0,32	0,32	0,32	0,22	0,22	0,22	0,40	0,53	0,27	0,68	0,68	0,61	0,54	0,53
27	14 de Julio	0,30	0,30	0,30	0,19	0,19	0,19	0,45	0,57	0,31	0,70	0,70	0,64	0,51	0,49

Fuente: Elaboración propia

Paso 3. Análisis de indicadores socio-demográficos

Mediante el paquete estadístico SPSS V.23 se obtienen los estadísticos descriptivos de todas las variables como se puede apreciar en el Anexo 2, entre los calculados está la media, trimedia, máximo, mínimo, varianza, coeficiente de variación, etc. Una muestra de estos aparece en la Tabla 11. Resumen de estadísticos descriptivos a seguir.

Tabla 11. Resumen de estadísticos descriptivos

	Media	Tri-media 5%	test Grubbs	p-value	Mediana	Mínimo	Máximo	Varianza	CV	Asimetría	Curtosis
CoeDept	0.308	0.309	3.765	0.011	0.31	0.23	0.35	0.000	6.718	-1.297	3.251
CoeDepM	0.305	0.306	3.656	0.019	0.31	0.23	0.34	0.000	6.704	-1.149	2.543
CoeDepF	0.311	0.312	3.826	0.008	0.31	0.22	0.37	0.001	7.651	-0.878	2.432
CoeVejT	0.185	0.187	2.931	0.389	0.19	0.09	0.25	0.001	17.538	-0.761	0.773
CoeVejM	0.182	0.183	3.103	0.205	0.19	0.09	0.24	0.001	16.267	-0.664	0.990
CoeVejF	0.188	0.189	2.542	1.000	0.19	0.09	0.28	0.001	20.508	-0.498	0.372
TasOcuT	0.414	0.415	2.87	0.482	0.41	0.26	0.54	0.003	12.959	-0.272	0.026
TasOcuM	0.529	0.532	4.224	0.001	0.54	0.32	0.62	0.002	9.340	-1.308	3.231
TasOcuF	0.291	0.29	2.302	1.000	0.27	0.13	0.47	0.006	26.70	0.190	-0.751
TasAtcT	0.692	0.691	3.765	0.011	0.69	0.65	0.77	0.000	2.988	1.297	3.251
TasAtcM	0.695	0.694	3.656	0.019	0.69	0.66	0.77	0.000	2.937	1.149	2.543
TasAtcF	0.638	0.636	3.42	0.056	0.64	0.58	0.74	0.001	4.660	0.923	1.396
IndMasc	0.516	0.516	2.729	0.784	0.52	0.46	0.57	0.000	3.969	-0.118	0.119

Fuente: Elaboración propia

Se puede apreciar a través del valor de la media (0.308), lo que representa aproximadamente un 31%, que la dependencia económica es considerable, dado que es un índice demográfico que expresa la proporción de personas dependientes sobre la población activa, que se encuentra en edad de trabajar. En este caso se puede decir que un tercio de la población. Es notable como la dependencia por sexo se muestra ligeramente mayor en el caso femenino. Si se compara con los datos de Cuba para este rubro respecto al año 2012 cuyo valor era 20.8², este se muestra más elevado.

Por su parte el envejecimiento, expresa la relación entre la cantidad de personas adultas mayores y la cantidad de niños y jóvenes. Se adoptó para Cuba y sus territorios una tipología similar a la elaborada por el Centro Latinoamericano y Caribeño de Demografía (CELADE) - División de Población de CEPAL, para clasificar a los países de la región. Los grupos se

² Coeficiente de dependencia de personas de la tercera edad (coeficiente de la población mayor de 65 años por ciento de la población entre 20-64 años). <https://knoema.es/atlas/Cuba/topics/Datos-demogr%C3%A1ficos/Coeficientes-de-dependencia/Coeficiente-de-dependencia-de-personas-de-la-tercera-edad-mayores-de-65-a%C3%B1os-por-la-poblaci%C3%B3n-entre-20-64-a%C3%B1os?mode=amp>

establecen de acuerdo a intervalos que indican el grado de envejecimiento, como se explica a continuación³:

- Grupo de Envejecimiento I (GEI): menos del 10.0 por ciento de la población de 60 años y más respecto al total.
- Grupo de Envejecimiento II (GEII): entre el 10.0 y el 15.0 por ciento de la población de 60 años y más respecto al total.
- Grupo de Envejecimiento III (GEIII): mayor del 15.0 por ciento de la población de 60 años y más respecto al total.

El coeficiente de envejecimiento, dado su valor 0.182 (en promedio) clasifica en el tercer grupo.

La tasa de ocupación cuyo valor promedio, según los datos obtenidos es del 0.54, prácticamente la mitad de la población. La tasa de actividad, por su parte es de 0.77, superior a la tasa de ocupación. Esto se debe a que la segunda considera no solo a los que están ocupados, sino al total de personas en edad laboral, ejerzan o no.

El test de Grubbs para detectar presencia de valores atípicos en la variable. Definido por las hipótesis: H_0 : *no hay presencia de atípicos*, en contraposición de la hipótesis alternativa, H_i : *hay al menos un atípico*. Se puede tomar la decisión sobre la base del estadístico de contraste o sobre el *p-value*, si este es menor que $\alpha = 0.05$, valor prefijado para esta investigación, se rechaza la hipótesis de nulidad. En la tabla se resaltan en color rojo en la columna del *p-value* aquellas variables en los que se rechaza la hipótesis nula.

Paso 4. Caracterización de la región.

Cienfuegos es una de las provincias más pequeñas de Cuba, ocupando el decimotercer lugar en extensión con 4 188.61 Km². Representando el 3.81% de la superficie total del país y su población representa el 3.6% del total de Cuba.

La provincia de Cienfuegos, ubicada en la zona centro-sur de Cuba, tenía en el censo del año 2012 una población total de 404 228 personas, de ella, el 49.3% eran mujeres y el 50.7% hombres, en valores absolutos se ha incrementado en 8 443 residentes su variación poblacional como promedio anual en los últimos 10 años es de 2.3 %, crecimiento que esta dado fundamentalmente por el municipio cabecera 6.3%. Decreciendo los municipios Cumanayagua -4.5 %, Cruces son -3.2 % y Lajas con -2.7%.

³ Fuente: ONEI (2022) Sistema de Información Estadístico Nacional (SIEN), de Demografía 2021.

La provincia de Cienfuegos en el año 2012 mostraba una estructura de población típica de país desarrollado, estrecha en la base, resultado de las bajas tasas de natalidad que tiene el país desde el año 1978 y del crecimiento de la esperanza de vida al nacer que tiene el territorio de 78.28 años, siendo para las mujeres de 80.17 años y para los hombres de 76.56 años. El 54.6 % de la población de la provincia se encuentra en los municipios Cienfuegos y Cumanayagua, cuyas cabeceras son las únicas que clasifican como ciudad en la provincia. La ubicación geográfica, unida a factores socioeconómicos de la cabecera provincial, genera que el principal centro de atracción sea la ciudad de Cienfuegos.

La relación de masculinidad es de 1 028 hombres por cada mil mujeres y se comporta con el nivel inferior en el municipio cabecera, en el cual las mujeres representan el 50.5% del total de la población. El 82.0 % de la población reside en asentamientos urbanos, con una densidad de 101.6 hab./Km², pero en el municipio cabecera es de 483.5 hab./Km². El municipio de Cumanayagua presenta los valores más bajos en estos indicadores con una densidad de 45.0 hab./Km² y solo el 66.0 % de su población residiendo en asentamientos urbanos.

La provincia posee un envejecimiento poblacional del 18.3 % (porcentaje de personas con 60 años y más con respecto al total de población). La provincia es la quinta más envejecida del país, mostrando valores altos, los municipios de Cruces y Lajas con 21.0 % y 20.7 % respectivamente. Asociado a este proceso, se manifiestan bajas tasas de fecundidad al igual que viene ocurriendo en el país.

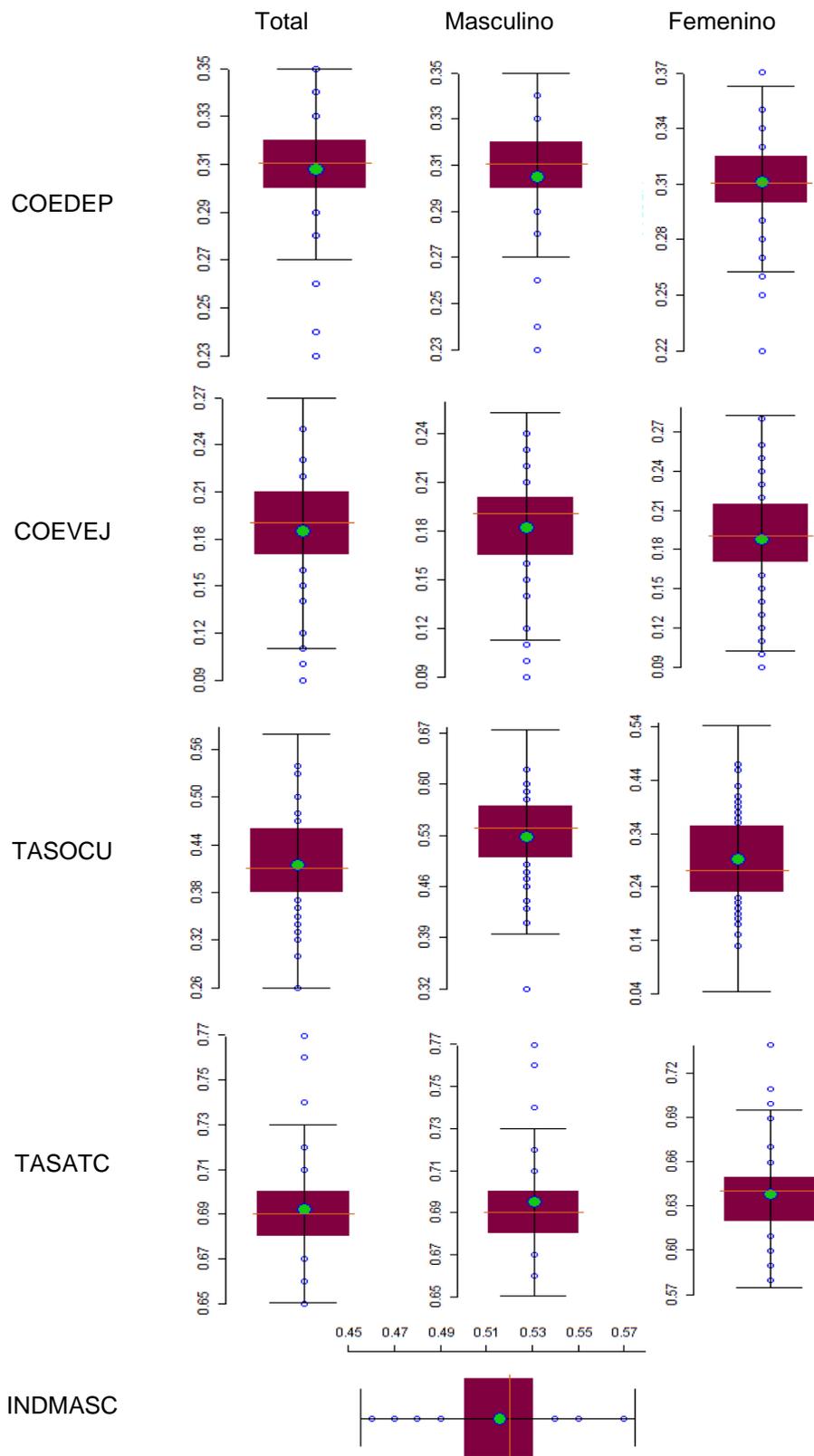
En cuanto al Color de la Piel, el 75.8 % de la población se declaró Blanco, superando la media del país que tiene un 64.1 % y el 24.2 % No blanco, cifra está que no varía mucho respecto al 2002 donde el 75.9 % se declaró Blanco, la que crece ligeramente es los mestizos o mulatos que varían del 15.8 % en el 2002 al 16.8 % en el año 2012. El 44.1 % de la población mayor de 12 años en la provincia no formaliza oficialmente su estado civil o conyugal, declarándose, unido, incrementándose con relación al 2002 que representaba el 40.2 % superando esta proporción en los municipios Abreus con el 56.5 % y Cumanayagua con el 50.4 %.

3.3 Etapa 2. Análisis Exploratorio de Datos Espaciales

A partir de una base cartográfica obtenida de la ONEI se agrega la matriz base a una tabla con extensión *.dbf (data base format), que permite el uso de un Sistema de Información Geográfico (SIG) como el software GeoDa.

Se utiliza un archivo de tipo shp (shapefile) que es típicamente el que describe las características de forma que se visualizan en un mapa, tales como, puntos, polígonos etc. En

Figura 9. Diagramas de cajas y bigotes

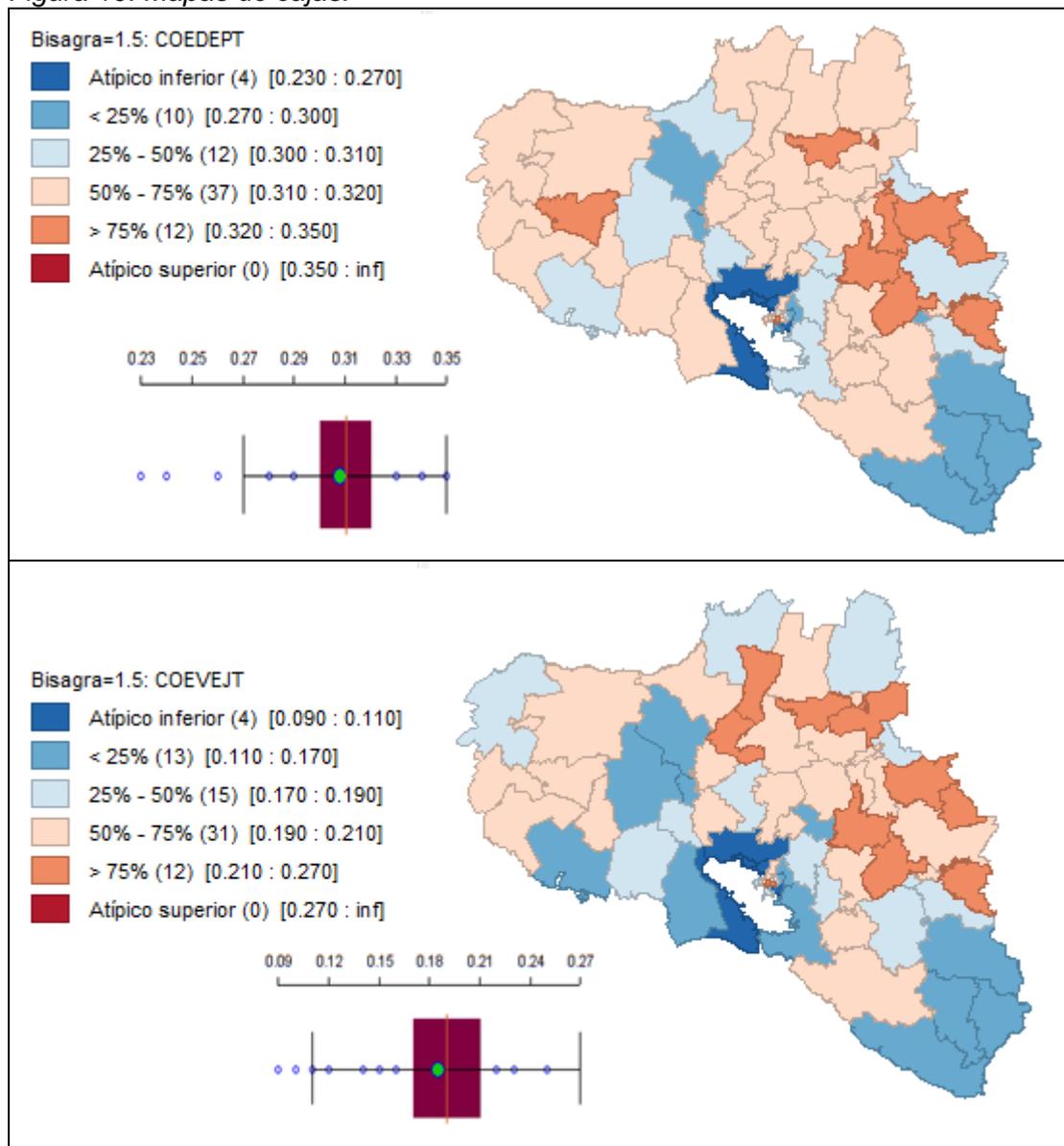


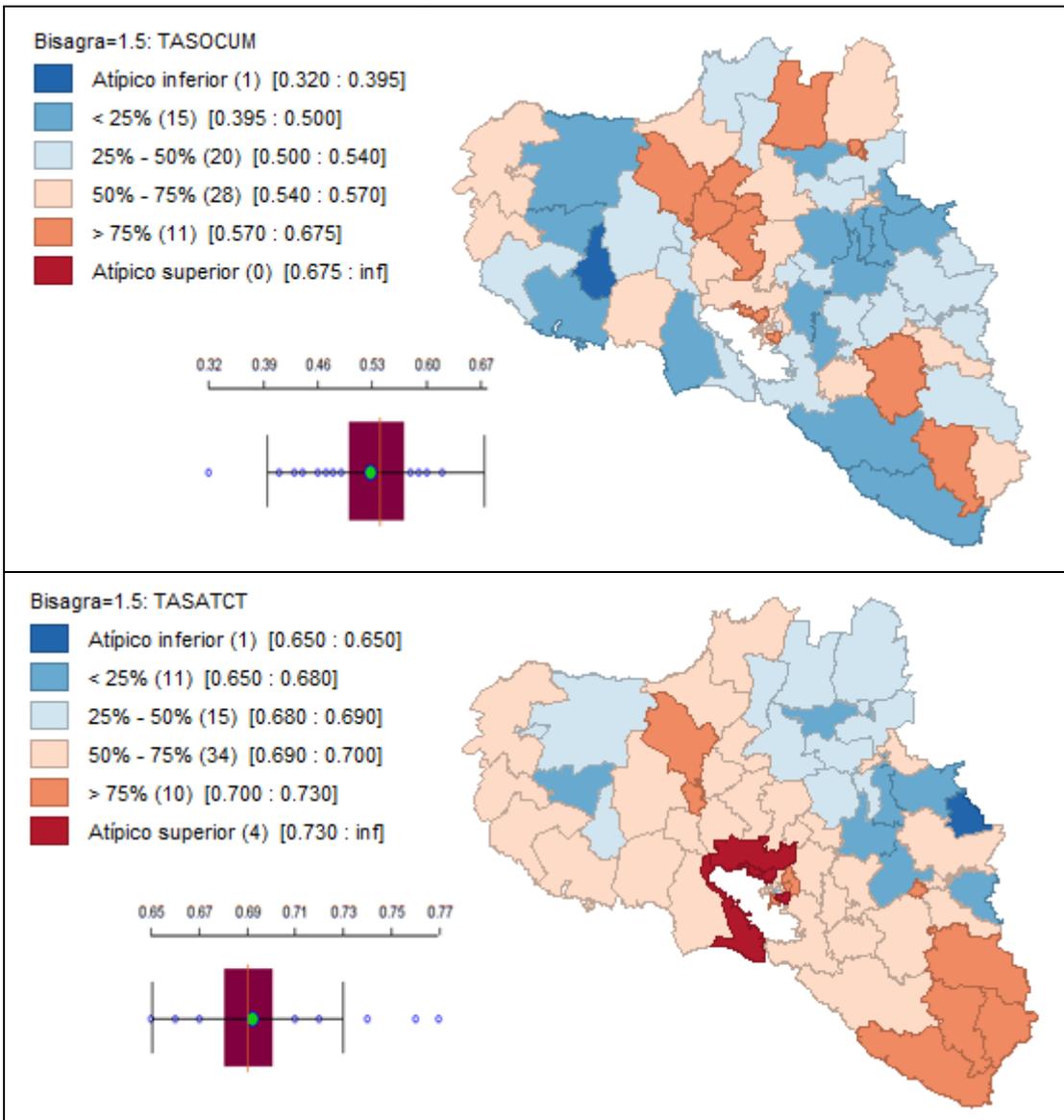
En la tabla 11 se resaltó la presencia de valores atípicos en las variables referidas a la Dependencia económica, así como, en la tasa de ocupación masculina, y las tasas de actividad total y masculina. Sin embargo, al complementar esta prueba con el diagrama de

cajas, se hace notar que el coeficiente de vejez, también aparece con valores que están por debajo del bigote inferior.

El mapa de caja para cada caso permite ubicar el consejo o los consejos populares que presentan la discontinuidad en cuanto a la distribución de la variable, tal como, se muestra en la Figura 10.

Figura 10. Mapas de cajas.





En cada caso es notable la representación de los estadísticos de posición que caracterizan el diagrama de caja, esta vez, en forma de una tabla de frecuencia que sirve como leyenda de colores, que, a su vez, es completada con el mapa a su derecha.

A modo de resumen se presenta en la Tabla 12, la cantidad de atípicos que aparecen por cada variable en la que se detectó la presencia de estos.

Tabla 12. Resumen de valores atípicos por variables

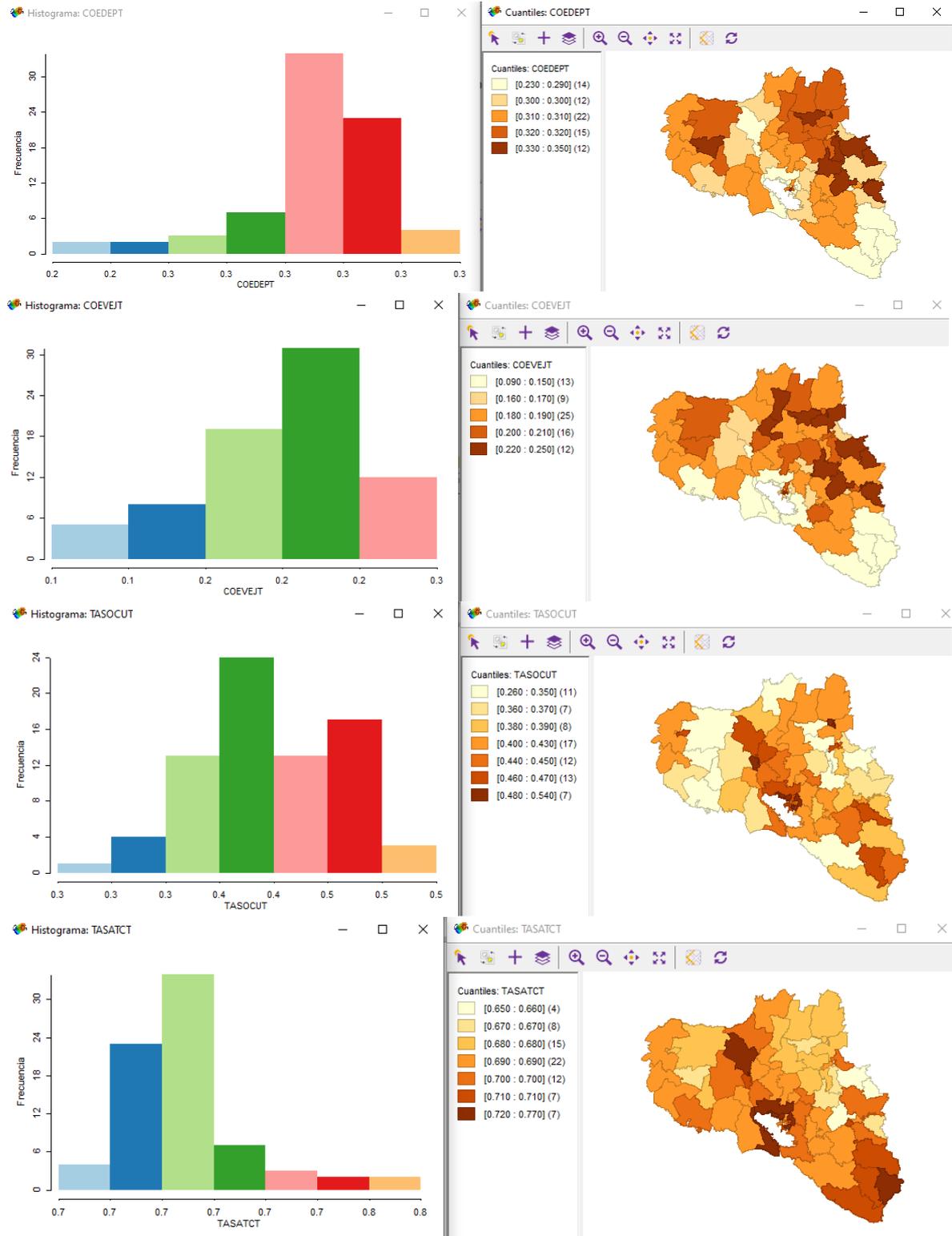
Variable	Atípico Inferior	Atípico Superior
Coefficiente de Dependencia	4	0
Coefficiente de Envejecimiento	4	0
Tasa de Ocupación	1	0
Tasa de Actividad	1	4
Índice de Masculinidad	0	0

Fuente: Elaboración propia

Paso 2. Identificación de la distribución

A partir del histograma, en la parte derecha de la Figura 11, se puede apreciar la forma de la distribución de cada variable con cierta similitud a la distribución normal, lo que puede ser confirmado con alguna prueba de normalidad como Shapiro-Wilk y de utilidad posterior si se desea hacer inferencias.

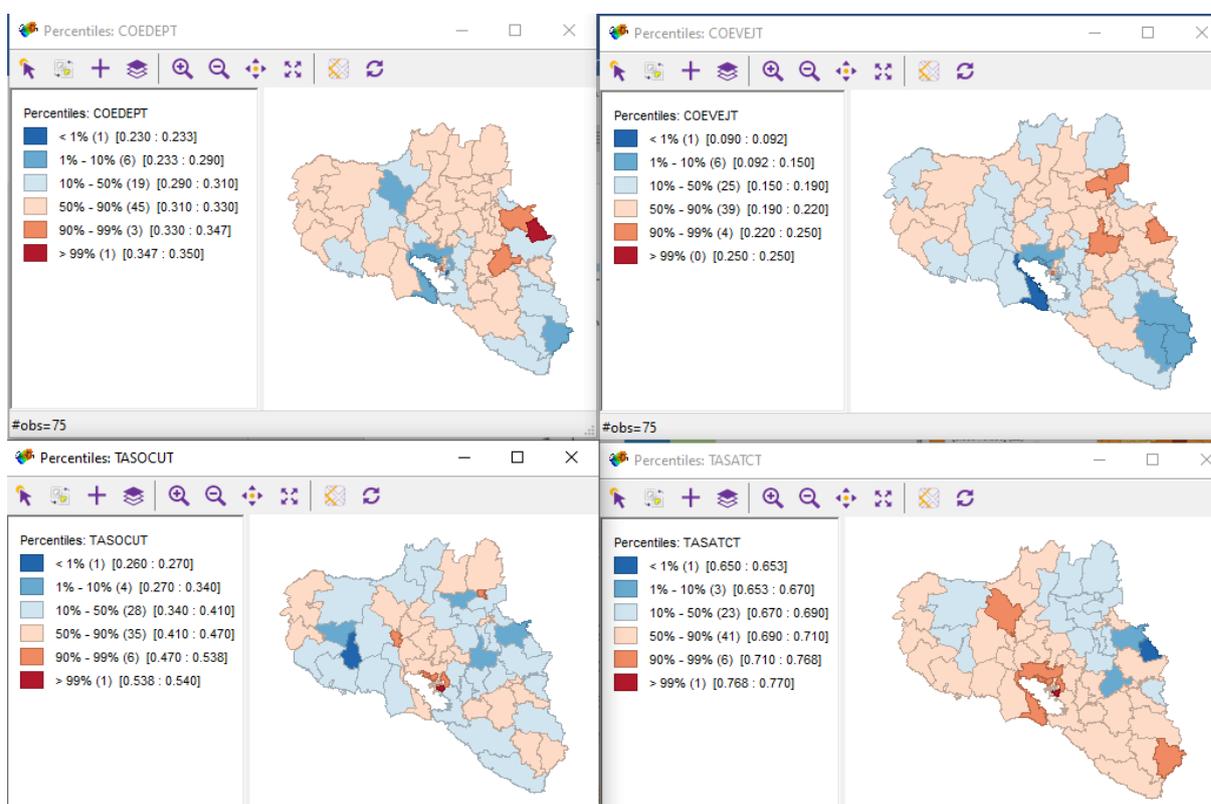
Figura 11. Resumen de histograma y mapa de cuantiles.



En la Figura 11, parte derecha, se muestra el mapa de cuantiles para algunas variables. Es apreciable como espacialmente, cada fenómeno analizado se distribuye con patrón Norte – Sur, tal es el caso del envejecimiento y la dependencia económica; o Centro – Periferia, como se observa en la ocupación y la actividad.

En la Figura 12 se muestra el conjunto de mapas de percentiles para la Dependencia, Envejecimiento, así como, la Actividad y Ocupación respectivamente. Se hace visible el predominio de la tonalidad rosa en el coeficiente de dependencia económica a través de la mayoría de los consejos populares, de ellos solo tres clasifican por encima del percentil 90%, dicho de otra forma, los consejos populares que presentan mayor valor de dicha variable. En el caso del coeficiente de envejecimiento sucede similar. Es notable que en ambos casos, los consejos populares del municipio cabecera aparecen con coloración azul, denotando menor valor de dependencia y envejecimiento.

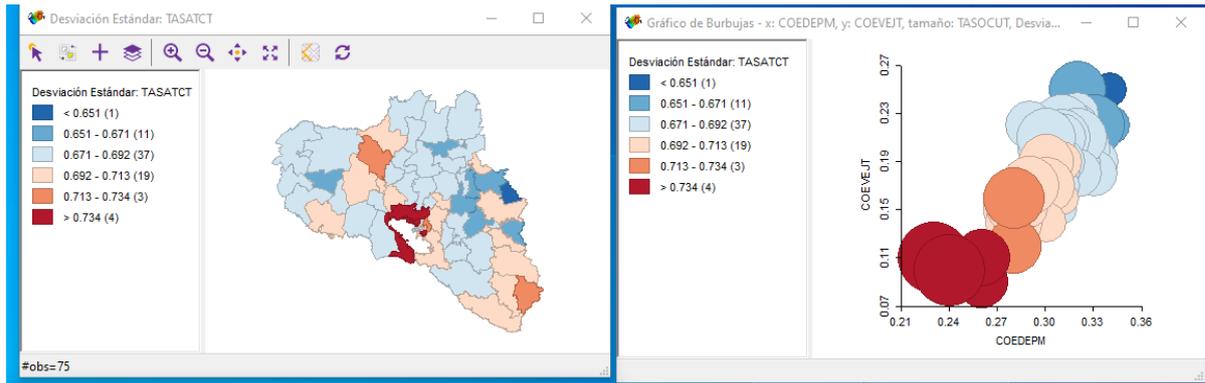
Figura 12. Mapas de percentiles



Por su parte, la tasa de ocupación es mayor en los CP de la cabecera municipal. Análogamente, se comporta la tasa de actividad, pero esta última clasifica desde el percentil 50% al 90%, a 41 consejos populares del total.

Los mayores valores de la desviación estándar de la tasa de ocupación unido a valores bajos de la dependencia económica y el envejecimiento son apreciables en los consejos populares del municipio Cienfuegos, como puede apreciarse en la Figura 13.

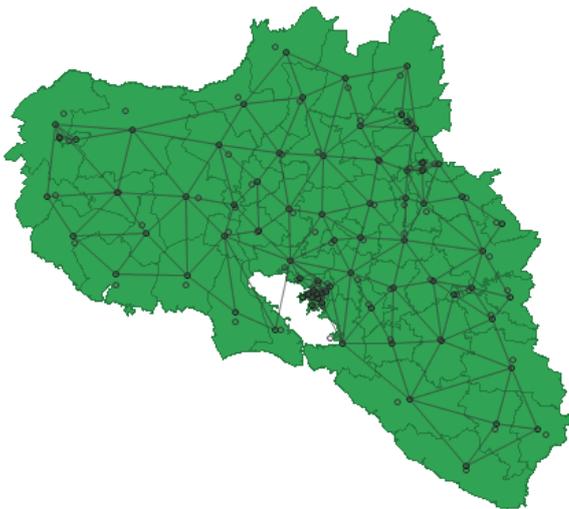
Figura 13. Mapa de desviación estándar



Paso 3. Detección de efectos espaciales

Para la detección de efectos espaciales, tanto locales como globales, se construye la matriz de vecindad y pesos. Se utiliza contigüidad tipo reina, con un mínimo de 1 vecino y como máximo, 10. A continuación se muestra en la figura 14, el gráfico de conectividad.

Figura 14. Gráfico de conectividad

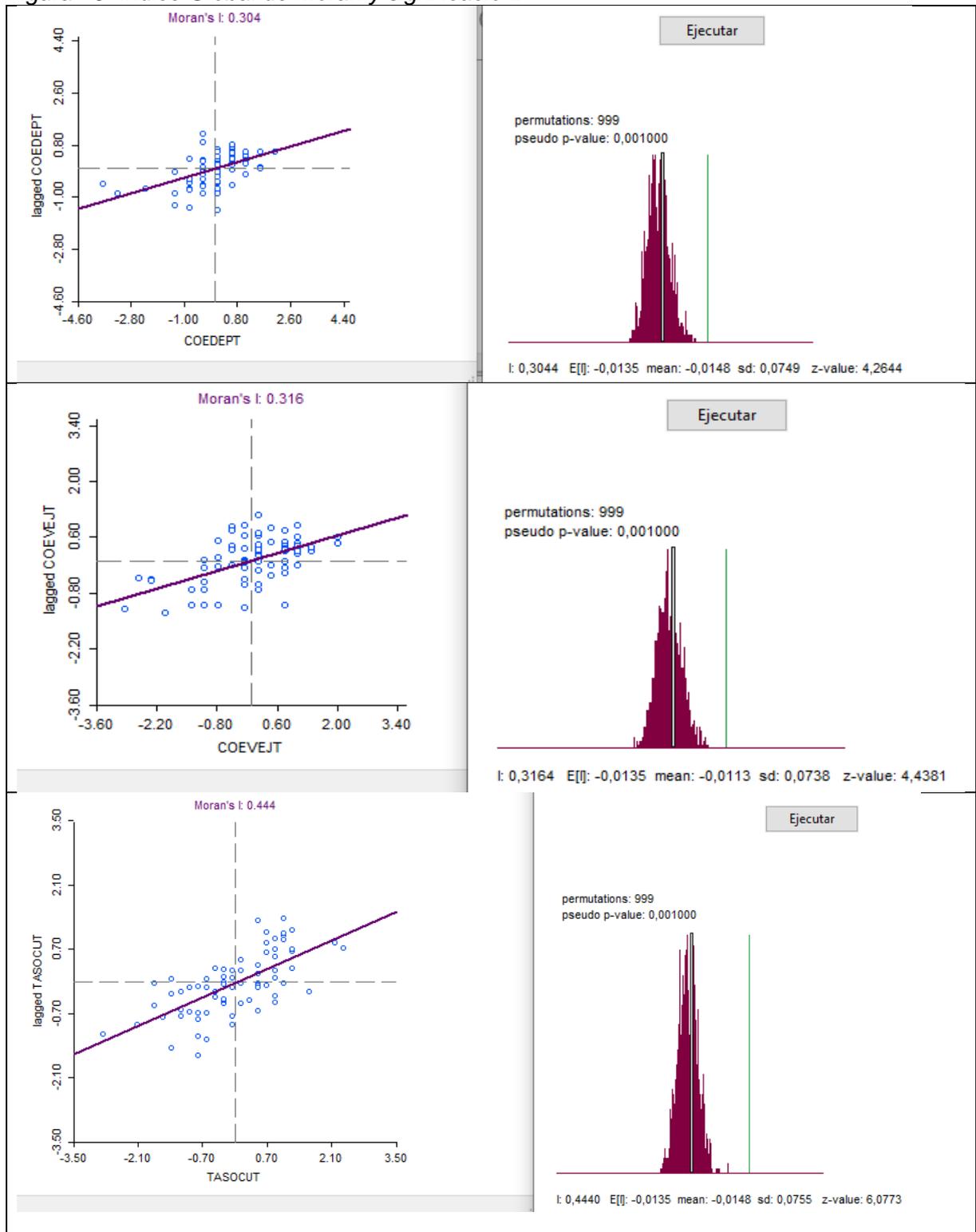


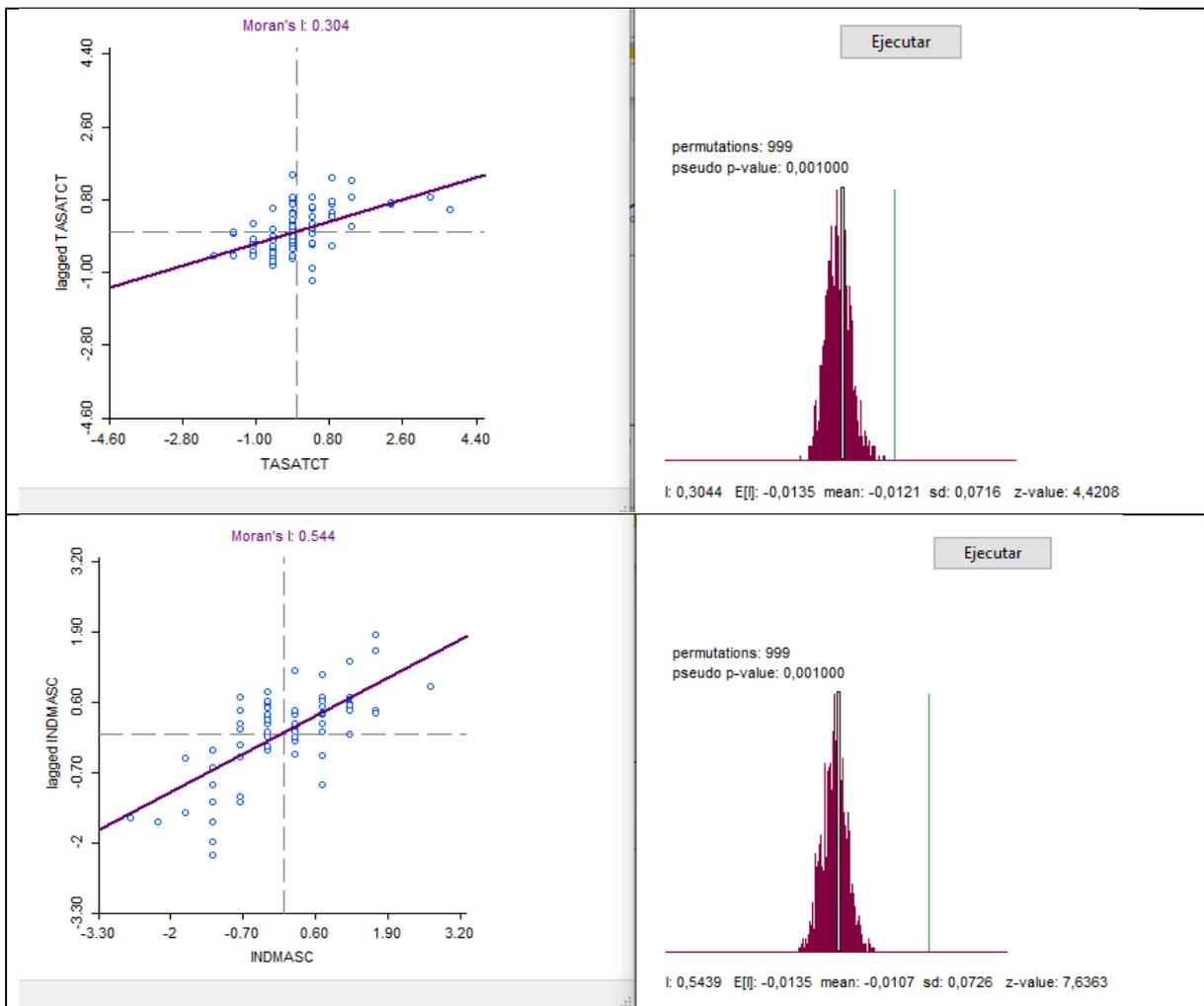
Se calcula el índice de Moran, su significación y los consejos populares se clasifican según su posición en el diagrama de dispersión de Moran.

El Índice Global de Moran consiste en la medición de la presencia o ausencia de autocorrelación espacial de una variable sobre la base de la dócima, H_0 : no existe correlación espacial frente a la alternativa H_i : existe correlación espacial. Los patrones de asociación espacial de las variables bajo estudio son estadísticamente significativos y tienden a estar agrupados en el espacio, de acuerdo con la evidencia estadística se puede rechazar la hipótesis nula de no autocorrelación espacial en favor de la H_i . Este resultado se puede apreciar en la Figura 15, en la gráfica de la derecha, donde los datos se agrupan alrededor de la línea de tendencia que recorre los cuadrantes I y III.

Esto indica que la distribución espacial de los valores altos y los valores bajos en los datos está más agrupada espacialmente de lo que se esperaría si los procesos espaciales subyacentes fueran aleatorios.

Figura 15. Índice Global de Moran y significación.





Un resumen del cálculo del índice de Moran, así como, el cálculo de la significación para el contraste de la hipótesis, calculada mediante método Monte Carlo se presentan en la tabla 13.

Tabla 13. Resumen del Índice global de Moran

Variables	I de Moran	MCs	Z-value	p-value
CoeDept	0.304	0.3044	4.4208	0.001
CoeDepM	0.265	0.2546	3.7837	0.001
CoeDepF	0.322	0.3223	4.1741	0.001
CoeVejT	0.316	0.3164	4.4381	0.001
CoeVejM	0.286	0.2862	4.0294	0.001
CoeVejF	0.339	0.3386	4.7275	0.001
TasOcuT	0.444	0.444	6.0773	0.001
TasOcuM	0.207	0.2073	3.1636	0.002
TasOcuF	0.545	0.5452	7.1649	0.001
TasAtcT	0.304	0.3044	4.4208	0.001
TasAtcM	0.265	0.2646	3.7837	0.001
TasAtcF	0.29	0.2898	4.1982	0.001
IndMasc	0.544	0.5439	7.6363	0.001

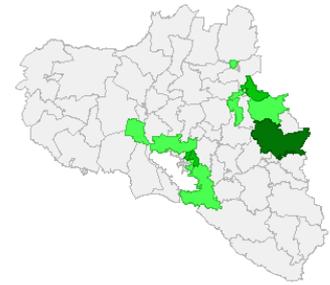
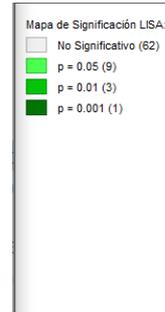
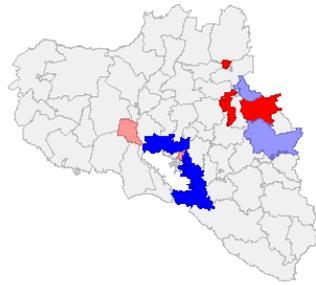
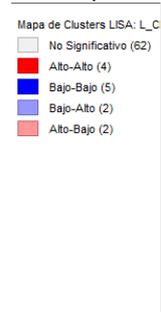
Fuente: Elaboración propia

Se puede apreciar el índice de correlación espacial positivo, lo que indica que los valores tienden a agruparse; además, la correlación es significativa.

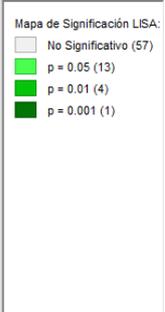
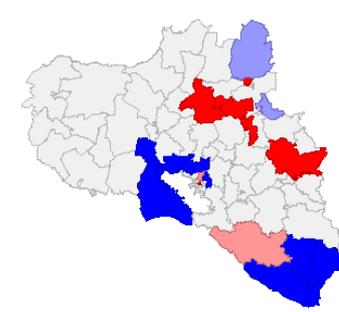
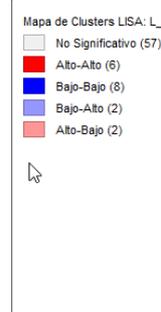
El Índice Local de Asociación Espacial (LISA) permite la identificación de patrones locales de asociación espacial, descomponiendo el Índice Moran para evaluar la influencia de ubicaciones individuales en la estadística global, como se puede apreciar en la Figura 16.

Figura 16. Resumen de mapas LISA

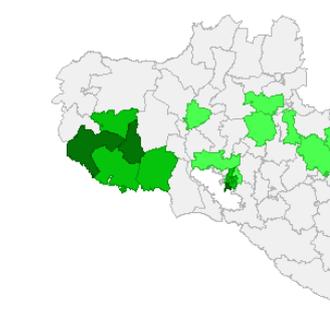
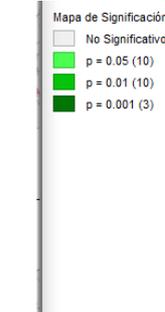
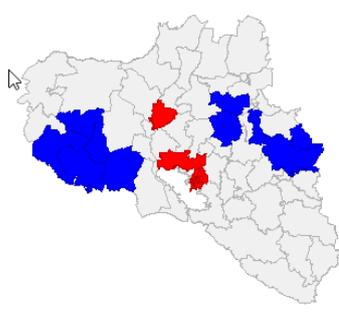
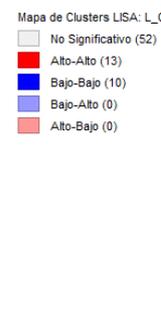
CoeDept



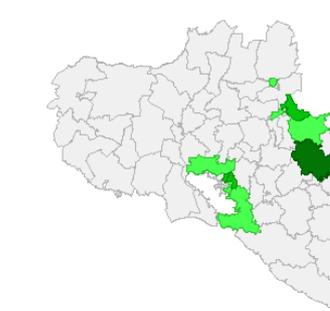
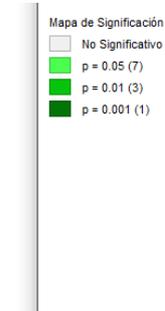
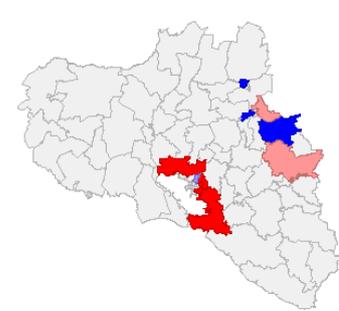
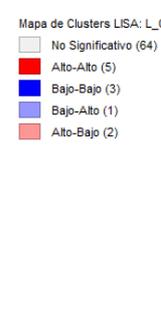
CoeVejT



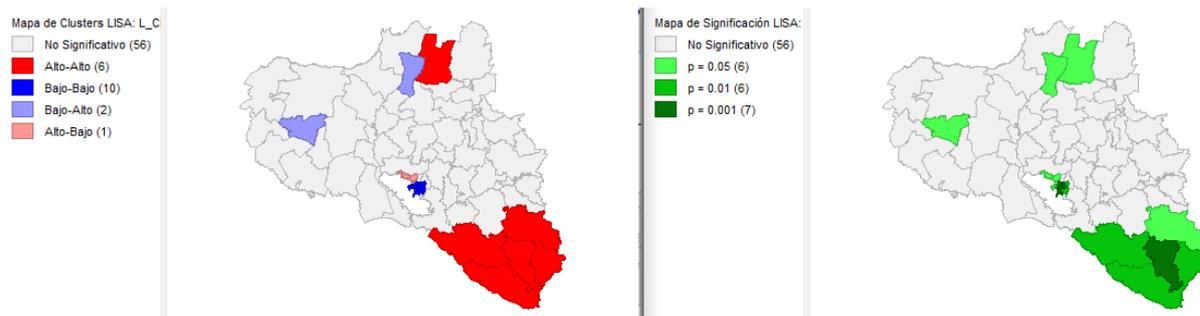
TasOcuT



TasAtcT



IndMasc



Si bien los mapas LISA, informan en forma de clúster que agrupan los consejos populares en cinco categorías: No significativos, Alto-Alto, Bajo-Bajo, Bajo-Alto, Alto-Bajo. Estas categorías ofrecen una explicación sobre cómo se distribuye localmente el Índice de Moran respecto al valor de sus vecinos, o sea, describe el valor de ese consejo popular y sobre el valor de los consejos vecinos. Adjunto se puede ver la significación estadística para ese comportamiento a través del p-value para cada categoría. En la tabla se puede apreciar

Tabla 14. Resumen de la cantidad de consejos populares por categorías

	No Significativo	Alto-Alto	Bajo-Bajo	Bajo-Alto	Alto-Bajo
COEDEP	62	4	5	2	2
COEVEJ	57	6	8	2	2
TASOCU	52	13	10	0	0
TASATC	64	5	3	1	2
IMASCU	56	6	10	2	1

Fuente: Elaboración propia

3.4 Etapa 3. Análisis de Resultados

La correlación que se dan entre CP de diferentes municipios, en zonas de desarrollo o crecimiento, incorpora la mirada del territorio y del espacio, sin necesariamente mirar la DPA, que es la tradicional. Por otro lado, en la zona norte de Cumanayagua se aprecia un comportamiento similar al del municipio Cruces.

Llama la atención el no encontrar en los CP de las cabeceras municipales los mayores niveles de ocupación.

Dado el impacto que tendrá a la vuelta de unos pocos años los altos niveles de envejecimiento poblacional que tiene el país, una mirada a las reservas en cada uno de los territorios para suplir el déficit de fuerza laboral, los municipios del noroeste de la provincia son los más envejecidos, en estos se encuentran los CP Martha Abreu y Ajuria, tienen bajo coeficiente de vejez y están rodeados de CP con alto coeficiente de vejez.

Llama poderosamente el similar comportamiento que tienen los CP del municipio Cienfuegos, muestra de los diferenciales que tienen con el resto de la provincia.

A modo de conclusiones del capítulo se tiene que; aplicando el procedimiento se obtienen tablas, diferentes mapas y gráficos donde se muestran los valores atípicos, los patrones de asociación, así como el resultado de análisis bivalente. La forma de representación gráfica facilita la interpretación de los datos.

Conclusiones

1. Se estudian las características y evolución de la econometría espacial y las técnicas y métodos para el análisis de datos espaciales y su aplicación en los estudios de dinámica de población.
2. Se identifica como aplicar los métodos de Análisis Exploratorio de Datos Espaciales para el estudio de patrones de asociación espacial de los Consejos Populares en la provincia Cienfuegos.
3. Se elabora un procedimiento para la identificación de patrones de asociación espacial de los Consejos Populares en Cienfuegos
4. Se identifican patrones de asociación espacial de los indicadores demográficos a nivel de Consejos Populares de la provincia de Cienfuegos y se visualizan a través de las herramientas visuales del GeoDa.
5. El procedimiento puede aplicarse a otros datos de panel y a diferentes estructuras espaciales

Recomendaciones

Extender esta aplicación a otros datos del censo.

Referencias Bibliográficas

- Acevedo, I., & Velázquez, E. (2008). Algunos conceptos de la econometría espacial y el análisis exploratorio de datos espaciales. *Ecós de Economía*, (27), 9-34.
- Aja Díaz, A. (2015). La experiencia de Cuba en la realización de los proyectos de capacitación e investigación en población y Desarrollo Local (CEDEM/UNFPA). *Revista Novedades en Población*, 11(22).
- Aldstadt, J., & Getis, A. (2006). Using AMOEBA to Create a Spatial Weights Matrix and Identify Spatial Clusters. *Geographical Analysis*, 38(4), 327-343. <https://doi.org/10.1111/j.1538-4632.2006.00689.x>
- Alonso, Ó. L. (2016). *La Econometría Espacial. Un Enfoque para el Estudio a Nivel de Datos Microterritoriales de la Políticas de Desarrollo Rural de la Unión Europea. El Caso de Asturias*. Servicio de Publicaciones y Divulgación Científica.
- Anselin, L. (1988a). Lagrange Multiplier Test Diagnostics for Spatial Dependence and Spatial Heterogeneity. *Geographical Analysis*, 20(1).
- Anselin, L. (1988b). *Spatial econometrics: Methods and models 4*. Springer Science & Business Media.
- Anselin, L. (1995, abril). Local Indicators of Spatial Association. *Geographical Analysis*, 27(2).
- Anselin, L. (1999). The Future of Spatial Analysis in the Social Sciences. *Geographic Information Sciences*, 5(2), 67-76. <https://doi.org/10.1080/10824009909480516>
- Anselin, L. (2000). The Alchemy of Statistics, or Creating Data Where No Data Exist. *Annals of the Association of American Geographers*, 90(3), 586-592. <https://doi.org/10.1111/0004-5608.00212>
- Anselin, L. (2001). Rao's score test in spatial econometrics. *Journal of Statistical Planning and Inference*, 97(1), 113-139. [https://doi.org/10.1016/S0378-3758\(00\)00349-9](https://doi.org/10.1016/S0378-3758(00)00349-9)
- Anselin, L. (2005). Exploring spatial data with GeoDaTM: a workbook. *Center for spatially integrated social science*, 1963, 157.
- Anselin, L. (2010). Thirty years of spatial econometrics. *Papers in Regional Science*, 89(1), 3-25. <https://doi.org/10.1111/j.1435-5957.2010.00279.x>
- Anselin, L., & Florax, R. J. G. M. (1995). *New directions in spatial econometrics*. Springer Verlag.
- Anselin, L., Gallo, J. L., & Jayet, H. (2008). Spatial Panel Econometrics. En L. Mátyás & P. Sevestre (Eds.), *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice* (pp. 625-660). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-75892-1_19
- Anselin, L., Li, X., & Koschinsky, J. (2022). GeoDa, from the desktop to an ecosystem for exploring spatial data. *Geographical Analysis*, 54(3), 439-466.
- Araujo González, R. (2016). Desarrollo local y trabajo comunitario: La experiencia del CEDEM-UH en el marco de PADIT. *Revista Novedades en Población*, 12(24), 84-93.

- Araujo, R. (2017). Demografía y desarrollo local: Un algoritmo de trabajo. *Revista Novedades en Población*, 13(26), 145-156.
- Assuncao, R. M., & Reis, E. A. (1999). A new proposal to adjust Moran's I for population density. *Statistics in medicine*, 18(16), 2147-2162.
- Aviles, G. V. (2017). *El índice de moran en el desempleo juvenil por entidad federativa en México, 2005-2015 ii.*
- Baller, R. D., Anselin, Luc., Messner, S. F., Deane, G., & Hawkins, D. F. (2006). Structural covariates of U.S. County homicide rates: Incorporating spatial effects. *Criminology*, 39(3), 561-588.
- Barros, O., & Aroca, P. (2014). Econometría espacial y el análisis sociodemográfico.: Aplicación en la formación de agrupaciones espaciales de envejecimiento en Cuba, período 2003-2009. *Revista novedades en población*, 10(20), 1-10.
- Becerra, F. (2004). *Evolución del desarrollo socio-económico a escala territorial: El caso de la provincia de Cienfuegos* [PhD Thesis].(Tesis de Doctoral).Universidad de Cienfuegos.
- Bivand, R. S., & Portnov, B. A. (2004). Exploring spatial data analysis techniques using R: The case of observations with no neighbors. En *Advances in Spatial Econometrics* 121-142.
- Brock, W. A., & Durlauf, S. N. (2001). Discrete Choice with Social Interactions. *The Review of Economic Studies*, 68(2), 235-260.
- Cabrera, E. (2016). *Un enfoque prospectivo para el desarrollo sostenible en ecosistemas de montaña. Caso Guamuhaya.* (Tesis Doctoral). Universidad de La Habana.
- Celemín, J. P. (2009). Autocorrelación espacial e indicadores locales de asociación espacial: Importancia, estructura y aplicación. *Revista Universitaria de Geografía*, 18(1), 11-31.
- Conde, Y. (2020). *Procedimiento estadístico para la identificación de patrones de asociación espacial. Caso: Montañas de Guamuhaya, Cienfuegos.* Universidad de Cienfuegos.
- de Orueta, E. L. B. (2003). La población en la planificación del territorio. *La dinámica geodemográfica protagonista del territorio: actas del VIII Congreso de la Población Española, Santiago de Compostela, 13-15 de junio de 2003*, 19-62.
- Díaz Gispert, L. (2011). *Evaluación del desarrollo sostenible para ecosistemas de montaña.* (Tesis Doctoral). Universidad de la Habana.
- Edsall, R., Andrienko, G., Andrienko, N., & Buttenfield, B. (2008). Interactive maps for exploring spatial data. *Manual of geographic information systems*, 837-858.
- Erviti, B., & Segura, T. (2000). Estudios de población. *Centro de estudios demográficos, Universidad de La Habana.*
- Fernández, E., Mayor, M., & Rodríguez, J. (2008). Estimating Spatial Autoregressive Models by GME-GCE Techniques. *International Regional Science Review*, 32(2), 148-172.
<https://doi.org/10.1177/0160017608326600>

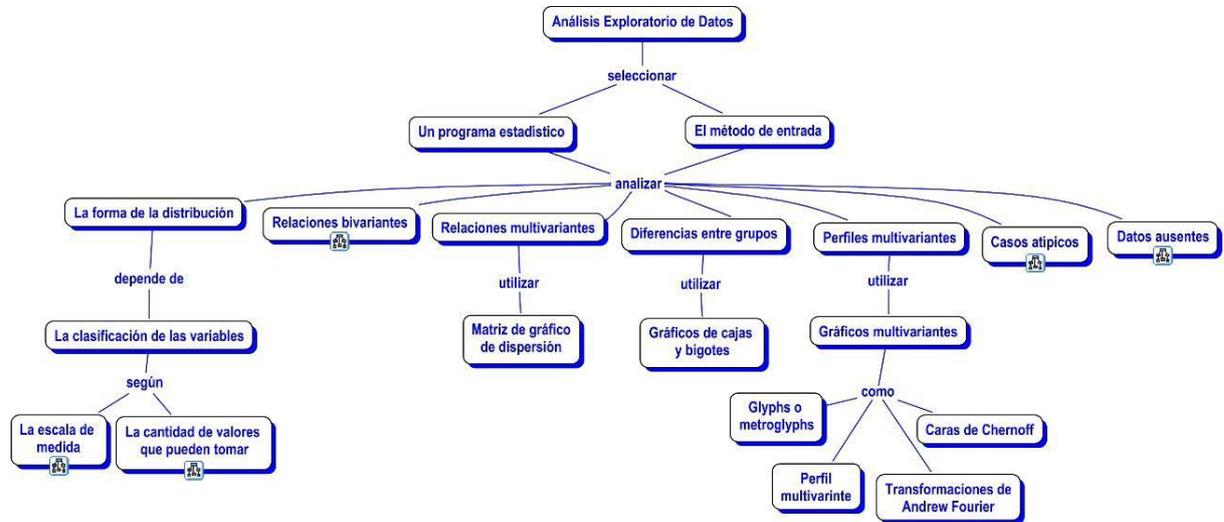
- Folmer, H., & Oud, J. (2008). How to Get Rid of W: A Latent Variables Approach to Modelling Spatially Lagged Variables. *Environment and Planning A: Economy and Space*, 40(10), 2526-2538.
<https://doi.org/10.1068/a4078>
- Fujita, M., & Krugman, P. (2004). La nueva geografía económica: Pasado, presente y futuro. *Investigaciones Regionales-Journal of Regional Research*, 4, 177-206.
- Getis, A., & Aldstadt, J. (2004). Constructing the spatial weights matrix using a local statistic. *Geographical analysis*, 36(2), 90-104.
- Getis, A., & Ord, J. K. (1992). The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*, 24(3), 189-206. <https://doi.org/10.1111/j.1538-4632.1992.tb00261.x>
- Ghosh, P., & Cartone, A. (2020). A Spatio-temporal analysis of COVID-19 outbreak in Italy. *Regional Science Policy & Practice*, 12(6), 1047-1062.
- Goodchild, M. F. (1987). A spatial analytical perspective on geographical information systems. *International Journal of Geographical Information Systems*, 1(4), 327-334.
<https://doi.org/10.1080/02693798708927820>
- Haining, R. (1995). *Spatial data analysis in the social and environmental sciences*. Cambridge university press.
- Haining, R. (2003). *Spatial Data Analysis: Theory and Practice*. Cambridge University Press.
<https://doi.org/10.1017/CBO9780511754944>
- Herrera, M., Paz, J., & Cid, J. (2012). Introducción a la Econometría Espacial. Una aplicación al estudio de la fecundidad en la Argentina usando R. *MPRA Paper*, 41138.
- Iglesias, G., Freire, J., & Martínez, M. (2022). *Espacio geográfico, procesos participativos y desarrollo local*.
- Kubrin, C. E. (2003). Structural covariates of homicide rates: Does type of homicide matter? *Journal of Research in Crime and Delinquency*, 40(2), 139-170.
- Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of Econometrics*, 95(2), 391-413. [https://doi.org/10.1016/S0304-4076\(99\)00044-5](https://doi.org/10.1016/S0304-4076(99)00044-5)
- LeSage, J., & Pace, R. K. (2009). *Introduction to spatial econometrics*. Chapman and Hall/CRC.
- Lima, I. Q.; Ramos Ramos, O.; Ormachea Muñoz, M.; Quintanilla Aguirre, J.; Duwig, C.; Maity, J. P.; Sracek, O.; & Bhattacharya, P.; (2020). Spatial dependency of arsenic, antimony, boron and other trace elements in the shallow groundwater systems of the Lower Katari Basin, Bolivian Altiplano. *Science of The Total Environment*, 719, 137505.
<https://doi.org/10.1016/j.scitotenv.2020.137505>
- Lucero, P. I., & Celemín, J. P. (2014). La calidad de vida de la población en la determinación de la calidad territorial. Un estudio de autocorrelación espacial aplicado a la ciudad de Mar del Plata, Argentina. *GeoFocus. International Review of Geographical Information Science and Technology*, 0(8), 94-114.

- Mazaira, Z. (2011). *Estudio del bienestar de las familias: El caso de la provincia Cienfuegos*. (Tesis de Doctoral). Universidad de Cienfuegos.
- Miri, S. H., & Miri, S. B. (2020). Investigating Commuting Time in a Metropolitan Statistical Area Using Spatial Autocorrelation Analysis. *American Journal of Engineering and Applied Sciences*, 13(1).
- Moreno, R., & Vayá, E. V. (2002). Econometría espacial: Nuevas técnicas para el análisis regional. Una aplicación a las regiones europeas. *Investigaciones Regionales-Journal of Regional Research*, 1, 83-106.
- Mur, J., & Paelinck, J. H. P. (2011). Deriving the W-matrix via p-median complete correlation analysis of residuals. *The Annals of Regional Science*, 47(2), 253-267. <https://doi.org/10.1007/s00168-010-0379-3>
- ONE-CEPDE. (2012). Anuario Demográfico de Cuba. *Plaza de la Revolución, Cuba: Oficina Nacional de Estadísticas*.
- ONEI. (2012). Informe final Censo de población y viviendas 2012. Resumen adelantado. *Plaza de la Revolución, Cuba: Oficina Nacional de Estadísticas*.
- Ord, J. K., & Getis, A. (2001). Testing for Local Spatial Autocorrelation in the Presence of Global Autocorrelation. *Journal of Regional Science*, 41(3), 411-432. <https://doi.org/10.1111/0022-4146.00224>
- Österman, A. (2014). *Map visualization in ArcGIS, QGIS and MapInfo* (2014). KTH, Geodesy and Geoinformatics.
- Paelinck, J. (1978). Spatial econometrics. *Economics Letters*, 1(1), 59-63.
- Pedrosa, N., & Albuquerque, N. (2020). Spatial Analysis of COVID-19 cases and intensive care beds in the State of Ceará, Brazil. *SciELO Preprints*. <https://doi.org/10.1590/SciELOPreprints.181>
- Phélan, M. (2007). La red observatorios locales de Barcelona, España. Un estudio de casos para diseñar una propuesta nacional. *Fermentum. Revista Venezolana de Sociología y Antropología*, 17(48), 96-122.
- Rabelo, M. G. (2006). *Las teorías acerca del subdesarrollo y el desarrollo: Una visión crítica*. Félix Varela. <https://books.google.com.cu/books?id=XNwstwAACAAJ>
- Rodríguez, Y. (2020). *Procedimiento estadístico para la identificación de patrones de asociación espacial en la población urbana del municipio Cienfuegos*. Universidad de Cienfuegos.
- Saez, M., & Saurina, C. (2007). Estadística y epidemiología espacial. *Girona: Edicions a Petició*.
- Siabato, W., & Guzmán-Manrique, J. (2019). La autocorrelación espacial y el desarrollo de la geografía cuantitativa. *Cuadernos de Geografía: Revista Colombiana de Geografía*, 28(1), 1-22.

- Tiefelsdorf, M., & Griffith, D. A. (2007). Semiparametric Filtering of Spatial Autocorrelation: The Eigenvector Approach. *Environment and Planning A: Economy and Space*, 39(5), 1193-1221. <https://doi.org/10.1068/a37378>
- Tobler, W. R. (1979). Cellular Geography. En S. Gale & G. Olsson (Eds.), *Philosophy in Geography* (pp. 379-386). Springer Netherlands. https://doi.org/10.1007/978-94-009-9394-5_18
- Torres, H. da G. (2006). Información demográfica y políticas públicas a escala regional y local. En: *Reunión de expertos sobre insumos sociodemográficos para la gestión y el desarrollo local-LC/L. 2581-P-2006*, 33-53.
- Varga, M. M. F. A. (2000). Geographic Knowledge Spillovers and University Research: Some Evidence from Austria. *Discussion Papers of the Institute for Economic Geography and GIScience*, 74(00), Article 00.
- Yrigoyen, C. C. (2003). *Econometría Espacial Aplicada a la Predicción-Extrapolación de Datos Microterritoriales*.
- Yrigoyen, C. C. (2006). Análisis estadístico de datos geográficos en geomarketing: El programa GeoDa. *Distribución y consumo*, 178, 34.
- Zuchowski, R. J. (2017). *An ecological study of immigrant segregation in established and new destination metros using Moran's and LISA cluster designation*.

Anexos

Anexos 1 Mapa conceptual sobre el análisis exploratorio de datos.



Anexo 2. Resumen de Estadísticos descriptivos de AED.

	Mean	Trimmed mean	Test Grubbs	p-value	Median	Min	Max	Var	C.V.	Skewness	Kurtosis
COEDEPT	0.308	0.309	3.765	0.011	0.310	0.230	0.350	0.000	6.72	-1.30	3.251
COEDEPM	0.305	0.306	3.656	0.019	0.310	0.230	0.340	0.000	6.70	-1.15	2.543
COEDEPF	0.311	0.312	3.826	0.008	0.310	0.220	0.370	0.001	7.65	-0.88	2.432
COEVEJT	0.185	0.187	2.931	0.389	0.190	0.090	0.250	0.001	17.54	-0.76	0.773
COEVEJM	0.182	0.183	3.103	0.205	0.190	0.090	0.240	0.001	16.27	-0.66	0.990
COEVEJF	0.188	0.189	2.542	1.000	0.190	0.090	0.280	0.001	20.51	-0.50	0.372
TASOCUT	0.414	0.415	2.870	0.482	0.410	0.260	0.540	0.003	12.96	-0.27	0.026
TASOCUM	0.529	0.532	4.224	0.001	0.540	0.320	0.620	0.002	9.34	-1.31	3.231
TASOCUF	0.291	0.290	2.302	1.000	0.270	0.130	0.470	0.006	26.70	0.19	-0.751
TASATCT	0.692	0.691	3.765	0.011	0.690	0.650	0.770	0.000	2.99	1.30	3.251
TASATCM	0.695	0.694	3.656	0.019	0.690	0.660	0.770	0.000	2.94	1.15	2.543
TASATCF	0.638	0.636	3.420	0.056	0.640	0.580	0.740	0.001	4.66	0.92	1.396
CASUNIT	0.530	0.530	2.546	1.000	0.530	0.430	0.620	0.002	7.41	0.15	0.238
CASUNIM	0.515	0.514	2.793	0.631	0.520	0.430	0.610	0.001	6.64	0.13	0.450
CASUNIF	0.548	0.549	2.743	0.748	0.550	0.410	0.660	0.003	9.20	-0.12	0.345
DIVSEPT	0.056	0.055	2.953	0.359	0.050	0.010	0.090	0.000	27.78	0.17	0.526
DIVSEPM	0.046	0.046	4.422	0.000	0.050	0.010	0.100	0.000	26.40	0.84	4.924
DIVSEPF	0.064	0.063	2.525	1.000	0.060	0.020	0.120	0.000	34.50	0.63	0.215
VIUDOST	0.041	0.041	2.566	1.000	0.040	0.020	0.070	0.000	28.11	0.09	-0.276
VIUDOSM	0.021	0.021	1.906	1.000	0.020	0.010	0.030	0.000	27.40	0.01	0.080
VIUDOSF	0.060	0.060	2.485	1.000	0.060	0.020	0.100	0.000	26.98	-0.12	0.156
SOLTERT	0.239	0.239	2.201	1.000	0.240	0.170	0.310	0.001	13.48	0.01	-0.104
SOLTERM	0.284	0.284	2.622	1.000	0.280	0.210	0.370	0.001	11.48	0.30	0.128
SOLTERF	0.191	0.192	2.469	1.000	0.190	0.100	0.280	0.001	19.33	-0.25	-0.162
EDUSUPT	0.079	0.075	3.214	0.132	0.060	0.020	0.230	0.002	59.55	1.06	0.829
EDUSUPM	0.063	0.058	3.626	0.022	0.050	0.020	0.220	0.002	69.04	1.51	2.374
EDUSUPF	0.097	0.094	2.991	0.312	0.080	0.010	0.250	0.003	52.79	0.71	0.062
TECMEDT	0.108	0.107	2.154	1.000	0.100	0.040	0.180	0.001	31.15	0.15	-0.764
TECMEDM	0.117	0.117	2.070	1.000	0.110	0.050	0.180	0.001	27.66	-0.03	-0.603
TECMEDF	0.098	0.097	2.209	1.000	0.100	0.040	0.180	0.001	38.22	0.20	-0.938
INDMASC	0.516	0.516	2.729	0.784	0.520	0.460	0.570	0.000	3.97	-0.12	0.119
BLANCOT	0.783	0.789	3.336	0.080	0.790	0.460	0.950	0.009	12.38	-0.72	0.535
BLANCOM	0.406	0.407	2.625	1.000	0.410	0.250	0.510	0.004	14.62	-0.24	-0.552
BLANCOF	0.378	0.380	3.821	0.009	0.390	0.210	0.460	0.002	11.63	-1.02	1.855
TOTALN	0.067	0.066	2.638	1.000	0.060	0.010	0.150	0.001	47.22	0.41	-0.407
NEGROM	0.035	0.035	2.705	0.849	0.040	0.010	0.080	0.000	46.42	0.37	-0.248
NEGROF	0.031	0.030	2.531	1.000	0.030	0.010	0.070	0.000	49.90	0.43	-0.325
TOTALME	0.150	0.145	3.939	0.005	0.150	0.030	0.450	0.006	50.77	1.08	2.141
MESTIZOM	0.457	0.457	3.003	0.299	0.460	0.340	0.580	0.002	9.01	-0.23	1.160
MESTIZOF	0.075	0.073	3.541	0.033	0.070	0.020	0.210	0.001	50.77	0.84	0.959