Revista Universitaria de Geografía
ISSN: 0326-8373
ISSN: 1852-4265
ceditorialdgyt@uns.edu.ar
Universidad Nacional del Sur
Argentina

Buzai, Gustavo D.
Matriz geográfica. Contexto y potencialidaden el análisis espacial cuantitativo°
Revista Universitaria de Geografía, vol. 33, núm. 2, 2024, Julio-Diciembre, pp. 11-43
Universidad Nacional del Sur
Bahía Blanca, Argentina

Disponible en: https://www.redalyc.org/articulo.oa?id=383281797001

Cómo citar el artículo
Número completo
Más información del artículo
Página de la revista en redalyc.org

Sistema de Información Científica Redalyc
Red de revistas científicas de Acceso Abierto diamante
Infraestructura abierta no comercial propiedad de la academia

Matriz geográfica. Contexto y potencialidad en el análisis espacial cuantitativo°

Gustavo D. Buzai*

Resumen

Hace seis décadas el geógrafo Brian J.L. Berry propuso el uso de la matriz geográfica como estructura de alcance conceptual y como herramienta para el análisis espacial cuantitativo. La primera pone en evidencia las formas básicas del abordaje geográfico y la segunda la posibilidad de avanzar sobre el trabajo tradicional y aplicar procedimientos de análisis multivariado con focalización en las unidades espaciales. La Geografía como ciencia espacial desarrolló una importante tradición que hace posible llegar a procedimientos de regionalización en base multivariada para delinear el modelo espacial y descubrir los aspectos subyacentes que intervienen en su organización. Considerando la importancia actual del contexto digital producido por el campo de la Geoinformática, el artículo propone un modelado de secuencias del tratamiento matricial en sus principales caminos metodológicos, el aporte fundamental de la matriz geográfica en la tarea de regionalización y las posibilidades en el análisis espacial cuantitativo actual.

Palabras clave: Matriz geográfica, Geografía Cuantitativa, Análisis Espacial, Estadística Espacial, Regionalización.

Geographic Matrix. Context and Potential in Quantitative Spatial Analysis

Abstract

Six decades ago, geographer Brian J.L. Berry proposed the use of the geographic matrix as both a conceptual framework and a tool for quantitative spatial analysis. The former highlights the fundamental approaches of geographical inquiry, while the latter enables advancements beyond traditional methods by applying multivariate analysis techniques focused on spatial units. As a spatial science, Geography has developed a strong tradition that facilitates multivariate-based regionalization processes, allowing for the delineation of spatial models and identification of underlying factors influencing spatial organization. Given the growing significance of the digital context shaped by the field of Geoinformatics, this article proposes

° https://doi.org/10.52292/j.rug.2024.33.2.0072
* Universidad Nacional de Luján, Instituto de Investigaciones Geográficas / CONICET. www.inigeo.unlu.edu.ar, gdbuzai@conicet.gov.ar

a modeling of sequential matrix-based approaches, outlining key methodological pathways, the critical role of the geographic matrix in regionalization efforts, and its potential applications in contemporary quantitative spatial analysis.

Keywords: Geographic matrix, Quantitative Geography, Spatial Analysis, Spatial Statistics, Regionalization

Introducción

A seis décadas de distancia de que fuera propuesta la *matriz geográfica* (Berry, 1964) el análisis espacial cuantitativo experimentó cambios notables al ingresar la mayoría de sus procedimientos al ámbito digital. Actualmente se tienen a disposición *toolbox* generales y de gran potencia en el interior del *software* específico para la resolución de cuestiones técnicas, aunque los procedimientos completos requieren del uso combinado de sistemas computacionales que se presentan eficientes cuando la secuencia de procedimientos y los alcances son conocidos previamente con claridad.

La Geografía Cuantitativa, al ser el paradigma de la Geografía que tiene mayor afinidad con los procedimientos técnicos de automatización, desarrolló diversos caminos metodológicos de análisis espacial y, de ellos, consideramos que el uso de matrices numéricas es uno de los principales desarrollos que requiere una síntesis actualizada.

El trabajo matricial en Geografía tiene una larga trayectoria como instrumento para la organización y presentación de datos geográficos, asimismo debemos destacar que la matriz geográfica surge el mismo año en que aparece el primer Sistema de Información Geográfica (SIG) (Buzai y Robinson, 2010) y a partir de allí tienen un recorrido conjunto de ampliación de confluencia en el interior del campo de la Geoinformática.

La Geografía Cuantitativa, en su posición paradigmática, es una forma de pensamiento y la matriz geográfica la apoya en perspectivas teóricas, metodológicas y aplicativas. Como forma de pensamiento queda demarcada en el campo de la lógica representacional y en la búsqueda de generalizaciones que brindan soluciones de conjunto a las problemáticas surgidas en el marco de las distribuciones espaciales. En su capacidad metodológica permite delinear una secuencia de procedimientos hacia la construcción de conocimientos y en la tarea aplicativa genera herramientas para la toma de decisiones en materia de ordenamiento territorial.

El presente trabajo introduce a la Geografía Cuantitativa como forma de pensamiento en cuanto a su lógica, como criterio demarcatorio y sus fundamentos como modo de acceder a la realidad. Se analiza el trabajo matricial y la matriz geográfica en cuanto a su formulación como herramienta estructuradora y las formas de abordaje desde un punto de vista sistémico hacia una síntesis conceptual que vincula la dicotomía general-regional. Finalmente, el objetivo es aportar una construcción estructurada como diagrama de secuencias del trabajo matricial en Geografía con la finalidad de estructurar una síntesis actual.

Geografía Cuantitativa

Lógica cuantitativa

Las aplicaciones matemáticas en Geografía atravesaron un extenso recorrido histórico desde la formulación de la disciplina realizada por Eratóstenes en el siglo III a.C., aunque se pueden encontrar raíces más lejanas en el mismo surgimiento de la ciencia (Buzai, 2021), las cuales llegan hasta la actualidad. La Geografía surge como una ciencia cuantitativa con el objetivo de calcular para conocer mejor el funcionamiento del universo y de nuestro planeta.

Es a finales del siglo XIX cuando se define la Geografía como ciencia humana. Se la considera orientada al estudio de la relación entre la sociedad y su medio, un abordaje que comienza a realizarse desde diferentes perspectivas conceptuales. Desde este comienzo se verifica una secuencia de cambios paradigmáticos (Kuhn, 1994) y a mediados del siglo XX surge la Geografía Cuantitativa, apoyada conceptualmente en la construcción de una Geografía científica (Schaefer, 1953; Buzai, 2023b) que intenta explicar, a través del uso de modelos y leyes del comportamiento espacial, una realidad generalizable.

Al presentarse la matemática como instrumento transversal a varios paradigmas a través de diferentes formas de uso, la lógica proposicional posibilita un claro criterio demarcatorio. El objetivo es poder determinar cuando el uso de métodos cuantitativos se realiza en el campo de la Geografía Cuantitativa, siendo (A) Geografía Cuantitativa, y (B) aplicación de técnicas cuantitativas en Geografía. La formulación es la siguiente:

[1] $A \longrightarrow B, A \therefore B$

[2] $A \longrightarrow B \neg B \therefore \neg A$

[3] $A \longrightarrow B, B \therefore A$

El modo *Ponens* [1] nos indica que si hay Geografía Cuantitativa se puede afirmar que habrá aplicación de técnicas cuantitativas, el modo *Tollens* [2] nos indica que si no hay aplicación de técnicas cuantitativas podemos afirmar que no hay Geografía Cuantitativa, y el modo de abducción [3] infiere que si hay aplicación de técnicas cuantitativas hay Geografía Cuantitativa, pero no es posible afirmarlo de forma lógica ya que representa la llamada falacia de afirmación del consecuente. Esto indica que la aplicación de técnicas cuantitativas en Geografía no garantiza hacer Geografía Cuantitativa, en este sentido, queda en evidencia la condición necesaria que corresponde a la cuantificación como forma de pensar la realidad.

La Geografía Cuantitativa en el interior de la perspectiva humana se desarrolla a mediados del siglo XX ante la preocupación por descubrir las leyes que explican las pautas de distribución espacial, apoyándose en la matemática como lenguaje científico en un avance que va desde las aplicaciones geométricas hacia las aplicaciones aritméticas, permitiendo el uso de procedimientos de análisis espacial con mayor flexibilidad.

En esta instancia, el trabajo matricial y la matriz geográfica adquieren una posición central en la investigación en Geografía Cuantitativa y estas estructuras de datos, más allá de constituir herramientas técnicas apoyan una forma de pensar, tanto para la construcción de conocimientos como para comprender la realidad.

Fundamentos

La construcción regional se sustenta en el método de superposición cartográfica sostenida conceptualmente por el paradigma de la Geografía Racionalista como actualización constructivista de la Geografía Clásica propuesta por Hartshorne (1939) (Rey Balmaceda, 1973). Este método combina aspectos físicos y humanos con el objetivo de obtener áreas homogéneas y descubrir relaciones de causalidad. Corresponde al método fundamental de la Geografía que hasta hoy se mantiene con el procedimiento algebra de mapas propuesto por Tomlin (1990) y que basa gran parte del análisis espacial con SIG.

A mediados del siglo XX, el contexto histórico resulta tener un papel preponderante como historia externa de la disciplina que propicia un nuevo cambio paradigmático.

La Segunda Guerra Mundial y la posguerra impulsaron el desarrollo científico-tecnológico de ciertas líneas académicas ligadas a la física-matemática, las cuales tuvieron un notable impacto dentro de las ciencias en general y en la Geografía en particular (Johnston, 1987).

La primera crítica sistemática realizada a la Geografía Racionalista proviene del trabajo de Schaefer (1953) (Buzai, 2023b) y toma un papel central al generar la base conceptual que llevaría a un cambio paradigmático en nuestra ciencia a partir de la Geografía Cuantitativa. Propone un abordaje opuesto orientado a la búsqueda de generalidades, corresponde a la utilización de métodos cuantitativos y a formular una Geografía que conduce al descubrimiento de las leyes que rigen las pautas de distribución espacial.

Aparece la llamada Revolución Cuantitativa (Burton, 1963) entendida principalmente como una tendencia conceptual por su alto nivel de abstracción. En este sentido, si bien fue el término *Geografía Cuantitativa* el que fue utilizado principalmente para definir esta línea (Garrison y Marble, 1967; Cole y King, 1968) y que se mantiene hasta la actualidad (Fotheringham *et al*., 2000; Sarkar, 2013; Harris, 2016; Feulliet et al., 2019), también fueron destacadas sus características teórico-metodológicas al definirla como *Geografía Teórica* (Bunge, 1962, 1966), su aspecto revolucionario al denominarla Nueva Geografía (Manley, 1966), su alto contenido estadístico al llamarla Geografía Estadística (Berry y Marble, 1968) y su perspectiva de Macro Geografía en los estudios de la Física Social (Stewart y Warntz, 1958).

A partir del predominio del paradigma cuantitativo se rescatan los trabajos clásicos que habían utilizado principalmente a la Geometría como lenguaje espacial (Thünen, 1826; Weber, 1909; Burgess, 1925; Christaller, 1933; Lösch, 1939; Hoyt, 1939; Harris y Ullman, 1945). Estos estudios fueron ampliados posteriormente desde un punto de vista analítico utilizando la matriz geográfica propuesta por Berry (1964) como sistema organizador de los datos de naturaleza espacial con posibilidades de trabajarse en perspectivas regionales, generales y temporales.

La ampliación de las bases geométricas a partir de un fuerte componente matemático y la modalidad matricial, en la cual era posible aplicar las diversas metodologías estadísticas en el análisis del espacio geográfico, permitieron generar abordajes modelísticos de las distribuciones espaciales. Ante la gran variedad de aplicaciones surgen una serie de obras que sistematizaron inicialmente la perspectiva: Bunge (1962), Haggett (1965), Garrison y Marble (1967), Cole y King (1968), Harvey (1969) y Chorley (1973). A partir de ellas puede verse que la etapa inicial generó básicamente siete principios que caracterizan los estudios de análisis geográfico a partir de la cuantificación y que son esenciales:

— El abordaje geográfico es principalmente espacial: desde el inicio de la Geografía como ciencia humana existen dos formas principales de encarar la investigación geográfica, como el estudio de la relación hombre-medio (definición ecológica) o como el estudio de la diferenciación areal (definición corológica), las cuales se complementan. Indudablemente cualquiera de estas perspectivas tienen su sustento en el espacio geográfico y para la Geografía Cuantitativa no existe posibilidad de realizar estudios geográficos sin base espacial, esta base de asociaciones y correlaciones espaciales es la que brindaría la identidad que singulariza a la Geografía en el contexto de las ciencias.

— La región se construye: tomando el avance teórico generado por la perspectiva racionalista los estudios de Geografía Cuantitativa consideran que la región, como delimitación de espacio sobre la superficie terrestre, se construye. No existe una región como realidad objetiva previa al investigador, sino que éste le pondrá límites de acuerdo a los objetivos de la investigación a partir de las metodologías aplicadas.

— La metodología de construcción regional es cuantitativa: a diferencia de la perspectiva racionalista que construye sus espacios geográficos a partir de la superposición cartográfica en un nivel cualitativo, la Geografía Cuantitativa utiliza métodos matemáticos para cumplir el objetivo. Corresponde en su instancia final a los procedimientos de construcción regional con base en la matriz geográfica y las correlaciones incluidas en ella.

— Se busca la construcción de modelos: es una de las finalidades principales de la Geografía como ciencia nomotética, ya que atiende a las características espaciales generalizables. Los modelos espaciales son construcciones simplificadas de la realidad espacial que permiten comprenderla en sus rasgos fundamentales, también pueden actuar como una guía para la planificación territorial (Buzai, 2023a) y son base para la transmisión del conocimiento.

— Se descarta el excepcionalismo: la Geografía no es una ciencia excepcional y por lo tanto tiende a realizar estudios de aspectos generalizables que culminan en la construcción de modelos. Como cualquier otra ciencia, la Geografía puede utilizar métodos cuantitativos para estudiar la realidad, en este caso desde un punto de vista espacial a partir del cual el investigador pondrá a prueba sus hipótesis y generará otras.

— Se producen capacidades interdisciplinarias: la utilización de métodos cuantitativos brindan la posibilidad de compartir un lenguaje común a otras ciencias. De esta manera la Geografía adquiere una mayor capacidad interdisciplinaria en el intercambio basado en el trabajo de aplicación concreto, en este sentido, las demandas de un saber al otro se sustentan en una base compartida.

— Se logra una mayor objetividad: no se considera que la investigación científica sea objetiva ya que durante su desarrollo se toman muchas decisiones subjetivas en el interior de un proceso constructivista (Johnston, 1968). Sin embargo, se considera que se avanza hacia un mayor nivel de objetividad ante la posibilidad de replicar procedimientos y llegar a los mismos resultados en la construcción de conocimientos, aspecto que se encuentra al alcance de todo investigador.

Matriz geográfica

Formulación

El concepto de matriz tomó un lugar central en el ámbito de la investigación cuantitativa en Geografía a partir de la década de 1960 y, se transformó, a partir de la década de 1980, en base a las posibilidades brindadas en el campo de la Geoinformática al combinar planillas de cálculo, software de análisis estadístico y SIG.

En el campo de la Geografía, fue Berry (1964) quien introdujo, a través de esta organización, importantes posibilidades operativas para el trabajo aplicativo. Al tratamiento de variables se suma la posibilidad de introducir el tratamiento de unidades espaciales con la capacidad de aplicar el análisis multivariado en procesos de regionalización. La propuesta del autor se presenta como una síntesis conceptual y organizativa, a partir de la cual es posible identificar claramente los principales enfoques del análisis regional.

La matriz tradicional es una estructura que organiza los datos cuantitativos en una tabla de doble entrada en donde las filas representan las unidades espaciales y las columnas las diferentes variables medidas en ellas. Toda propuesta de aplicación estadística en esta organización estará centrada en el tratamiento de variables y el objetivo metodológico es la obtención de macrovariables formadas por un conjunto de variables de similares comportamientos. El resultado final se obtiene como procedimiento clasificatorio que, representadas en el mapa, cada macrovariable brinda una síntesis de la distribución espacial de sus componentes.

La matriz geográfica también almacena información sobre la observación de un fenómeno en un lugar determinado (hecho geográfico, geographical fact), sin embargo, su organización se produce ante una transposición de la anterior, en este caso las variables se encuentran ubicadas en el sentido de las filas y las unidades espaciales en las columnas. La propuesta de aplicación estadística difiere del abordaje tradicional al utilizar el análisis multivariado en el sentido de las columnas y, de esta manera, su objetivo es unir unidades espaciales. El resultado final corresponde a una regionalización vinculada a una matriz de datos con valores agrupados para las nuevas unidades espaciales.

En este contexto, la construcción matricial es una propuesta conceptual global que nos permite mostrar los diferentes tipos de abordajes realizados en el análisis regional y apoya la construcción cuantitativa. Definiendo a la Geografía a partir de su objeto formal, apoya la idea de que el geógrafo se caracteriza por su abordaje integrador para lo cual se utiliza un marco conceptual como sistema. El abordaje geográfico es espacial y los conceptos utilizados son de naturaleza espacial e integradora.

Líneas de abordaje

Al analizar la matriz geográfica en el sentido de las filas observamos de qué manera cada tema se distribuye espacialmente en el área de estudio (Geografía General) y si lo hacemos en el sentido de las unidades espaciales corresponde a la forma que, el conjunto de hechos geográficos, se combinan de manera específica en cada una de ellas (Geografía Regional). La dimensión temporal se logra por superposición de matrices de diferentes cortes temporales (Geografía Histórica).

Considerando esta perspectiva pueden realizarse diferentes abordajes de análisis: (1) Un tema en la totalidad de unidades espaciales (distribución espacial, un mapa temático), (2) Una unidad espacial en la totalidad de contenido (combinación de diferentes hechos geográficos en un sitio), (3) varios temas en la totalidad de unidades espaciales (asociación espacial de variables), (4) varias unidades espaciales en la totalidad de contenidos (diferenciación areal), (5) Todos estos estudios pueden realizarse en un subconjunto como recorte espacial y temático en el interior de la matriz y (6) en niveles de profundidad incluyendo la dimensión temporal.

Los procedimientos corresponden a la metodología de un proceso de generalización ya que se parte de la mayor diferenciación interna para ir construyendo agrupamientos más amplios y con mayor superficie de homogeneidad. Es lo que se denomina regionalización por agregación de sitios (Harvey, 1969), en este caso son unidades espaciales. Es un procedimiento que se dirige desde lo individual a lo general.

En síntesis, la matriz geográfica desde su conceptualización a mediados de la década de 1960 se convirtió en una de las herramientas de mayor operatividad para la investigación en Geografía cuando el objetivo de la tarea demanda un eficiente almacenamiento y tratamiento de la información cuantitativa utilizada. Resulta también importante porque a través de su estructura brinda una integración satisfactoria en los criterios demarcatorios de diferentes perspectivas de abordaje geográfico y avanza hacia la aplicación de procesos de regionalización.

En la actualidad la matriz de datos tradicional corresponde a la tabla de atributos del SIG vectorial y la matriz geográfica es utilizada con software de análisis estadístico en el campo de la Geoinfomática (Buzai y Baxendale, 2012). El análisis conceptual realizado por Berry (1964) se hace operativo con gran amplitud, en donde las mayores posibilidades metodológicas le brindan una notable vigencia.

Modelado de secuencias

Se propone un diseño de procedimientos presentado en la figura 1 y seguidamente se analizan cada uno de sus componentes:

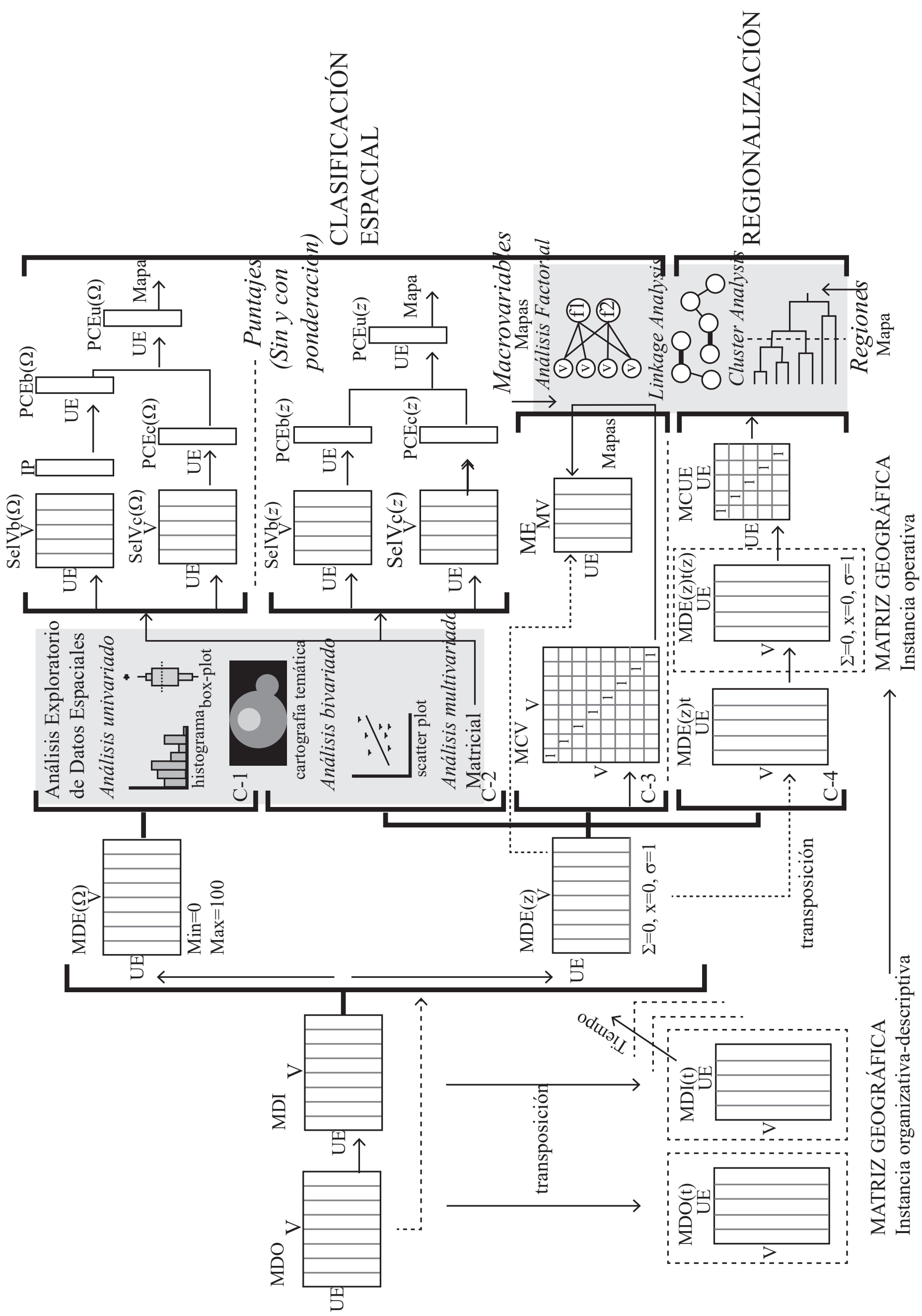

Figura 1. Modelado de secuencias del trabajo matricial en Geografía. Fuente: elaboración propia.

Matrices iniciales: MDO, MDI

El modelo inicia con la realización de la matriz de datos originales (MDO), la cual contiene los datos básicos de la investigación, organizados en una tabla de doble entrada donde las unidades espaciales se encuentran en las filas y las variables en las columnas.

En general, cuando la aplicación se realiza en el ámbito de la Geografía Humana, esta matriz se realiza a partir de la utilización de datos que surgen de atributos poblacionales, por lo tanto, los valores que la componen son numéricos y las variables representan cada uno de los temas considerados. Estos datos se encuentran en diferentes unidades de medida, pudiendo incorporarse en valores absolutos, en porcentaje o en por miles, encontrándose incluidos en unidades de conteo de población, hogares y viviendas, porcentajes de temas específicos o por miles en tasas demográficas.

Con la finalidad de superar el inconveniente producido ante la existencia de unidades espaciales de diferentes tamaños poblacionales, en donde aquellas que cuentan con mayor población generalmente tendrán mayores valores en gran parte de las variables, se procede a realizar una primera transformación a través del cálculo de porcentajes para la totalidad de variables con el objetivo de analizar el peso relativo del dato en cada unidad espacial. Estos cálculos generan la matriz de datos índice (MDI).

Los datos de cada variable se hacen relativos respecto de la variable base a partir de la cual tienen origen. Las de población se relacionan con el valor del total de población, las de hogares con el total de hogares y las de vivienda con el total de vivienda. Las variables bases se calculan respecto de su propio total que, en la columna resultante, pasará a tener valor . La transposición de estas matrices lleva las unidades espaciales a las columnas y las variables a las filas. De esta manera, cada una se convierte en matriz geográfica en instancias de organización y descripción de los datos. A partir de aquí se aplican las líneas de abordaje presentadas.

Matrices estandarizadas: MDE(Ω), MDE(z)

Se procede a la estandarización de la MDI mediante el uso de dos diferentes cálculos de puntajes:

La *matriz de datos estandarizada* por puntaje omega, MDE(Ω), se obtiene aplicando:

[4] $\Omega_i = \frac{x_i - m}{M - m} * 100$

Donde, Ω_i es el puntaje omega para la unidad espacial i, x_i es cada uno de los datos individuales, *m* y *M* son el dato menor y el mayor de la serie de datos respectivamente. La comprobación para cada columna estandarizadas se produciría en cálculos que arrojen los siguientes valores: Min = 0 y Max = 100.

La *matriz de datos estandarizada* por puntaje *z*, MDE(*z*), se obtiene aplicando:

[5] $z_i = \frac{x_i - \bar{x}}{\sigma}$

Siendo,

[6] $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

[7] $\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})}{n}}$

Donde, z_i es el puntaje estándar para cada unidad espacial *i*, x_i es cada uno de los datos individuales, $\bar{x}$ es la media y σ es el desvío estándar de la serie de datos. La comprobación para cada columna estandarizada se produciría en cálculos que arrojen los siguientes resultados: $\Sigma = 0$, $\bar{x} = 0$ y $\sigma = 1$. El puntaje *z* se presenta en valores positivos y negativos, como unidades de desvío, respecto de $\bar{x} = 0$.

Habiendo construido la MDO, MDI, MDE(Ω) y MDE(*z*) se puede aplicar la primera serie de procedimientos de análisis espacial entre los que se incluyen análisis univariados (cartografía temática, histograma y *box-plot*) y bivariado (gráfico de dispersión).

Cartografía temática

La cartografía temática es un campo de estudio que tiene por objeto la realización de mapas de cualquier tipo de tema que exceda la representación de los rasgos del terreno al presentar distribuciones espaciales que no pueden ser vistas de manera empírica en el área de estudio. Sus resultados brindan apoyo a diversos campos científicos en actividades aplicadas y en la comunicación sintética de temas dirigidos a una gran amplitud de usuarios.

Teniendo como soporte la base cartográfica del área de estudio, nos centramos en las formas de construir intervalos de clase para realizar mapas coropléticos, en el cual se asocia la intensidad de un color con la intensidad del valor que presenta la variable en cada unidad espacial. Esta es la principal modalidad para la representación de datos censales.

Los principales métodos básicos para la determinación de intervalos de clase son el de cortes naturales a partir de los mayores saltos encontrados en la curva de datos, de intervalos iguales al dividir en n intervalos la amplitud de los valores extremos, de cuantiles al incluir la misma cantidad de unidades espaciales en cada intervalo y de desvíos estándar cuando se realizan cortes sobre los datos estandarizados en *z* (Buzai *et al*., 2016).

El mapa de cortes naturales representa más fielmente la estructura de los datos, el mapa de cuantiles permite comparaciones perfectas considerando únicamente al orden de las unidades espaciales y el método de intervalos iguales permite realizar comparaciones cuando la amplitud de los datos de las variables es la misma, ya sea en % o en *z*.

La selección de la cantidad de intervalos de clase se ubicará en el interior de un continuo formado entre la individualidad y la generalidad. Si no se cuenta con una decisión sobre el nivel de detalle es posible utilizan el cálculo de *k* de Sturges:

[8] $k = 1 + 3{,}3 \, log \, n$

Donde *n* representa la cantidad de datos y *log* es el logaritmo de base 10.

Todas las variables pueden ser cartografiadas por separado, con diferentes métodos y cantidad de intervalos de clase por lo cual la cantidad de mapas a obtener es infinita, proponiéndose en esta instancia un actual concepto de Atlas (Buzai *et al*., 2022; Principi *et al*, 2023), en el que se privilegia el formato digital, flexible y centrado en el usuario del sistema.

Representaciones gráficas exploratorias univariadas

El análisis exploratorio de datos (EDA, *Exploratory Data Analysis*) incluye un conjunto de técnicas orientadas a describir gráficamente la estructura de los datos, considerándose procedimientos indispensables al momento de realizar las primeras aproximaciones aplicativas. Actualmente estas técnicas se potencian al incorporar la dimensión espacial vinculando las bases de datos alfanuméricas con las bases cartográficas y, de esta manera, toma el nombre de análisis exploratorio de datos espaciales (ESDA, *Exploratory Spatial Data Analysis*).

Con posterioridad a la realización de cartografía temática, la utilización de estas técnicas permite avanzar en el análisis espacial de los datos en cada variable (análisis univariado) y de la relación en el comportamiento de dos variables (análisis bivariado).

Las posibilidades tecnológicas actuales en el campo del ESDA se ven potenciadas con el uso de los SIG (Grecousis, 2020) vinculados a la geovisualización basada en la interactividad digital (Bosque Sendra y Zamora Ludovic, 2002) y apoyada en una perspectiva racionalista en la construcción de conocimiento a partir de diferentes representaciones gráficas que serán detalladas a continuación.

El histograma utiliza barras verticales para representar la frecuencia en que aparecen un conjunto de datos y de esta manera poder ver su distribución en una dimensión. El eje *x* contiene los intervalos seleccionados (*k*) y el eje *y* los valores. La amplitud de los intervalos se calcula mediante:

[9] $w = \frac{r}{k}$

Donde *w* es la extensión del intervalo, *r* es la amplitud de los datos (mayor – menor) y *k* es la cantidad de intervalos o cantidad de barras que tendrá el histograma. En el ESDA, la selección de cada barra permitirá ver a que unidades espaciales corresponde en el mapa y, con ello, su distribución espacial.

El *box-plot* (gráfico de caja) es un método que muestra la distribución de los datos en cuartiles y tiene por objetivo resaltar los valores atípicos (*va*) (*outliers*). De esta manera, a las cuatro categorías estándares representadas por el 25 % de las unidades espaciales se destacan dos clases al separar los valores extremos. Los componentes en orden creciente son los siguientes.

6° valores atípicos superiores
5° intervalo: > 75 % (Cuarto cuartil: *Q4*)
4° intervalo: 50% al 75 % (tercer cuartil: *Q3*)
3° intervalo: 25% al 50 % (segundo cuartil: *Q2*)
2° intervalo: < 25 % (primer cuartil: *Q1*)
1° valores atípicos inferiores

El 1° y 2° nivel contiene el 25 % de las unidades espaciales, el 3° el 50 %, el 4° el 75 % y el 5° y 6° llegan al 100 %. El punto medio de la distribución está representado por la mediana:

[10] $Mdna(i) = \frac{n+1}{2}$

[11] $Mdna(p) = \frac{n}{2}\ , \frac{n+2}{2}$

Siendo *Mdna*(*i*) y *Mdna*(*p*) la mediana para un número impar y par de observaciones. Un valor se considera atípico cuando supera el punto de corte o *hinge* (*h*):

[12] $h < Q1 - a * RIQ$

[13] $h > Q3 + a * RIQ$

Donde *h* es el punto de corte, *Q1* y *Q3* son los valores de la posición 25 % y 75 % respectivamente (extremos de la caja), *a* es un valor de amplitud que logra el extremo, en general 1,5 o 3,0, al sumarse a la posición *Q3* para los extremos superiores y restarse a la posición *Q1* para los extremos inferiores y *RIQ* es el rango intercuartil, *Q3 – Q1*. De acuerdo a la consideración de Tukey (1977) los valores que se encuentran a la distancia de 1.5 y 3 son valores atípicos leves y los que superan 3 son valores atípicos extremos. Bastará seleccionar los puntos de cada cuartil y los extremos para representar los resultados en un mapa que se denomina *box-map*.

Representación gráfica exploratoria bivariada

La posibilidad de un ESDA bivariado se asocia con las propiedades del gráfico de dispersión (*scatter diagram*). Su aplicación brinda como resultado una configuración en la cual cada variable queda representada por ejes ortogonales (*x*-*y*) y cada unidad espacial aparece como un punto localizado en el espacio de relaciones a partir de los valores que tiene en cada variable.

Cuando los datos están en puntajes estándar *z*, cada uno de estos ejes se ubica en el centro del gráfico por x=0, y=0 que corresponden a la media y quedan definidos cuatro cuadrantes. El cuadrante inferior izquierdo y el superior derecho contienen unidades espaciales que contienen valores que se encuentran por debajo y por arriba de la media en ambas variables respectivamente. El cuadrante superior izquierdo corresponde a valores por debajo de la media en *x* y por arriba de la media en *y*, y el cuadrante inferior derecho presenta la situación inversa.

El análisis del sentido de la relación entre los resultados producidos en ambas variables se calcula por la recta de regresión formada en base a la nube de puntos que corresponde a las posiciones de cada unidad espacial en el sistema de coordenadas. Si el sentido de la recta es desde el cuadrante - - al cuadrante + + la relación muestra una tendencia positiva y si va desde el cuadrante - + al cuadrante + - es negativa. Cuando la nube de puntos es redondeada y no se puede definir una recta de regresión la relación es aleatoria. La ecuación de la recta es:

[14] $y = a \pm bx$

Siendo

[15] $a = \frac{\sum y - b \sum x}{n}$

[16] $b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$

Donde *a* es la ordenada al origen, es decir el punto por donde la recta cruza el eje y cuando *x*=0 y *b* representa la pendiente de la recta, es decir, el incremento de *y* ante los cambios de valores en *x*. Este cálculo se utiliza en el análisis de regresión cuando se intentan predecir los valores de una variable dependiente (*y*) respecto de los valores conocidos de la variable independiente (*x*).

La interpretación de estas relaciones será diferente según sea el tipo de variables relacionadas. Las *variables de beneficio* (vb) son aquellas en las que sus máximos puntajes indican situaciones favorables (ej. Máximo nivel educativo alcanzado: terciario-universitario completo) y las *variables de costo* (vc) se comportan de manera inversa al presentar en sus máximos puntajes situaciones desfavorables (ej. Necesidades Básicas Insatisfechas).

Aunque no se encuentra vinculado a la matriz geográfica, cabe mencionar aquí que, desde inicios de la década de 1990, junto al desarrollo de *software* considerados Sistemas de Ayuda a la Decisión Espacial (SADE), acompañando las representaciones gráficas exploratorias univariadas y bivariada el ESDA incorpora la posibilidad de visualizar distribuciones espaciales que representan patrones de autocorrelación espacial local (Anselin, 1996). En este caso corresponde a la determinación de cúmulos espaciales que verifican si las asociaciones espaciales tienen mayor fuerza entre unidades espaciales vecinas y estos valores disminuyen con la distancia.

Análisis multivariado: Puntajes de clasificación espacial W

Un procedimiento inicial del análisis multivariado avanza en la construcción de puntajes de clasificación espacial (PCE) a partir de la combinación de variables estandarizadas en puntajes Ω. Surgen dos posibilidades, trabajar con variables de beneficio *Vb*(Ω) o variables de costo *Vc*(Ω). Cabe aclarar que cuanto más alto sean los datos de la variable las primeras presentarán situacione más favorables y las segundas más desfavorables.

Las variables fueron estandarizadas con el puntaje Ω [4], las cuales, en su orientación como *Vb*, pueden ser consideradas indicadores de planificación (IP) (Buzai y Baxendale, 2008) si se las combina con un puntaje de objetivo (PO), el cual mide el grado de apartamiento a un valor deseable previamente definido:

[17] $PO = 100 - |x_i - vo| * UA$

Donde x_i es el valor de la variable en cada unidad espacial, *vo* es el valor objetivo y *UA* es una unidad de ajuste

Siendo:

[18] $UA = \frac{100}{|vo - vmd|}$

Donde *vmd* es el valor más distante al valor objetivo. El puntaje 100 es el valor objetivo y cada x_i muestra la distancia que la aparta del objetivo. Entonces, combinado con variables de beneficio siempre el valor 100 y 0 representan las mejores y peores situaciones respectivamente.

Cuando no se incluyen PO y las variables se combinan entre ellas y se obtienen directamente puntajes de clasificación espacial de beneficio, PCEb(Ω) o de costo, PCEc(Ω). Estos puntajes surgen al promediar los valores estandarizados en las columnas (temas) seleccionadas y representan una síntesis que brinda la estructura espacial del conjunto. Los mapas mostrarán con colores más intensos los mayores valores y en ese sentido estos representarán las mejores o peores áreas respectivamente.

Si quisieran combinarse variables de beneficio y de costo en una resolución que genere un puntaje de clasificación espacial unificado, PCEu(Ω), normalmente las variables de costo se invierten *VC*(Ω)*i* al sentido del beneficio mediante:

[19] $VC(\Omega)i = 100 - VC\Omega$

Aunque también se pueden estandarizar directamente de forma inversa utilizando:

[20] $\Omega Ci = \frac{M - x_i}{M - m} * 100$

Donde, es el puntaje omega de costo invertido para la unidad espacial *i*, es cada uno de los datos individuales, *m* y *M* son el dato menor y el mayor respectivamente de la serie de datos. La comprobación para cada columna estandarizadas se produciría en cálculos que arrojen los siguientes valores: Min = 0 y Max = 100.

Mediante estas combinaciones se obtiene el PCEu(Ω) que permite incluir en una misma resolución variables de beneficio y costo, siendo que el mapa se interpreta en el sentido del beneficio.

Análisis multivariado: Puntajes de clasificación espacial z

Un segundo procedimiento avanza en la construcción de puntajes de clasificación espacial (PCE) a partir de la combinación de variables estandarizadas en puntajes

z. Al igual que en el caso anterior, surgen dos posibilidades, trabajar con variables de beneficio, Vb(z) o variables de costo Vc(z).

Las variables fueron estandarizadas de acuerdo a [5] y con ellas pueden obtenerse puntajes de clasificación espacial de beneficio, PCEb(z) o de costo, PCEc(z). Estos puntajes se obtienen al promediar las columnas seleccionadas y representan una síntesis de la estructura espacial del conjunto. Los mapas mostrarán colores más intensos en los mayores valores y en ese sentido estos representarán las mejores o peores áreas respectivamente. En Buzai (2014) se presenta cartografía temática comparable con amplitudes de considerando: Muy bajo (<-1), Bajo (-1 a -0,5), Medio (-0,5 a 0,5), Alto (0,5 a 1) y Muy alto (>1,5). De esta manera se crea un intervalo intermedio, dos superiores y dos inferiores.

Si quisieran combinarse variables de beneficio y de costo en una resolución que genere un puntaje de clasificación espacial unificado, PCEu(z), generalmente las variables de costo se invierten al sentido del beneficio mediante:

[21] $VC(z)_i = z_i * -1$

Donde el multiplicador -1 cambia el signo del puntaje.

Mediante estas combinaciones se obtiene el PCEu(z) que combina variables de beneficio y costo y cuyo mapa se interpreta en el sentido del beneficio.

Ponderaciones

La construcción de PCE de acuerdo a los dos puntos previos, se apoya en el supuesto de que todas las variables tienen similar importancia, sin embargo, es posible que esto no sea así y que teóricamente tengamos conocimiento para saber que algunas variables son más importantes que otras. La forma para determinar la importancia relativa de las variables puede ser simple como la consideración de un valor de ponderación por conocimiento temático o a partir de la aplicación de un procedimiento de cálculo específico.

La definición aplicativa considera que la importancia del tema es de un 100 % y que cada variable representa un porcentaje de importancia del total, siendo que matemáticamente se resuelve a partir de proporciones con sumatoria 1.

Existen diferentes métodos para el cálculo de ponderaciones (w) (Buzai y Baxendale, 2012), de ellos seleccionamos el método por ranking recíproco propuesto por Malczewski (1999) el cual se basa en el siguiente cálculo:

[22] $w_j = \frac{1/r_j}{\sum 1/r_j}$

Donde w_j es el valor de ponderación otorgado a cada variable y r_j el número de orden (ranking) que se le brinda a la variable de acuerdo a su importancia.

Los valores de ponderación cumplen las siguientes condiciones:

[23] $0 \leq w_j < 1$

y

[24] $\sum_{i=1}^{n} w_j = 1$

El método de resolución cambia y corresponde a lo que se denomina combinación lineal ponderada a partir del cual se obtiene el resultado final:

[25] $I_j = \sum_{i=1}^{n} V_j w_j$

Donde I_j es el valor índice para la unidad espacial *i*, es la sumatoria de los resultados brindados por la totalidad de variables, V_j es cada variable y w_j es su valor de ponderación. De esta manera cada variable aportará una proporción al resultado final de acuerdo a su grado de importancia en la temática total.

Correlaciones

El coeficiente de correlación tiene su fundamento en la variabilidad conjunta de dos variables a partir del cálculo de la covarianza:

[26] $Cov_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n}$

Al incluir el desvío estándar se obtiene la fórmula del coeficiente *r* de Pearson:

[27] $r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n\sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})}{n}}\sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})}{n}}}$

Utilizando la nomenclatura del desvío estándar puede simplificarse como:

[28] $r_{xy} \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n\sigma_x\sigma_y}$

Finalmente, si consideramos que a partir de [5] los puntajes estándar *z* de ambas variables es posible avanzar en su máxima simplificación:

[29] $r_{xy} = \frac{\sum_{i=1}^{n} z_x z_y}{n}$

De esta manera, se multiplican las columnas *z* y el cálculo del promedio corresponde al coeficiente de correlación *r* de Pearson.

La significatividad de *r* depende del contexto de trabajo, las situaciones controladas de laboratorio pueden arrojar resultados mucho más elevados que las aplicaciones a situaciones empíricas. También el valor puede ser interpretado de diferente manera según sea el tamaño de las poblaciones.

Con la finalidad de determinar la relevancia del valor obtenido se utiliza el test *t* de Student al comparar el valor *t* obtenido con el valor esperado para el límite de confianza $p < 0{,}05$ o $p < 0{,}1$ lo cual indica que se producirá, con dos grados de libertad, en menos del 5 % y 10 % de los casos respectivamente.

[30] $t_{xy} = \frac{r_{xy}\sqrt{n-2}}{\sqrt{1-r_{xy}^2}}$

Finalmente puede calcularse el coeficiente de determinación (*cdet*) que representa el porcentaje de explicación que una variable tiene sobre otra, en general cuanto es explicada la variable dependiente *y* respecto de la variable independiente o explicativa *x* de acuerdo a la ecuación [7] como resultado de la recta de regresión.

[31] $cdet = [\frac{\sum_{i=1}^{n} z_x z_y}{n}]^2 = r^2$

De todas formas, se considera que las relaciones de causalidad no surgen de la técnica, sino que se sustenta en aspectos conceptuales que atañen al conocimiento temático.

Matrices de correlaciones

Es una matriz cuadrada que contiene los valores *r* de Pearson que indican el sentido e intensidad de las relaciones medidas. La cantidad de cálculos (c) para su realización está dada por:

[32] $c = \frac{(n*n)-n}{2}$

Donde *n* es la cantidad de unidades, la resta elimina la diagonal de $r = 1$ y la división por 2 corresponde a un triángulo interno de la matriz, que atañen al conocimiento temático.

A partir de la MDE(*z*) se realiza la matriz de correlaciones de variables, MCV, la cual es cuadrada al presentar la misma cantidad de variables en las filas y en las columnas. Su forma de realización implica el cálculo del coeficiente *r* para cada par de entidades y completarla con los resultados obtenidos. Sin embargo, cuando partimos de los datos estandarizados *z* puede resolverse con una multiplicación de matrices utilizando sus transposiciones *t* (Baxendale, 1992):

[33] $MCV = MDE_{(z)t} * MDE_{(z)}$

[34] $MCUE = MDE_{(z)t(z)t} {}^{*} MDE_{(z)t(z)}$

La multiplicación de matrices puede realizarse si la cantidad de columnas de la primera es igual a la cantidad de filas de la segunda, por ese motivo primero debe considerarse la transpuesta. La obtención de las matrices de correlaciones amplía el camino metodológico que lleva a la aplicación de procedimientos estadísticos para su reducción en macrovariables.

Cuando la MDE(*z*) se transpone y nuevamente se estandariza en el sentido de las unidades espaciales surge la matriz geográfica en instancia operativa como MDE(*z*)t(*z*). Ambas constituyen los insumos para la realización de las matrices de correlaciones y la posterior aplicación de las metodologías de análisis multivariado con el objetivo de unir variables en macrovariables o unidades espaciales en regiones respectivamente.

Las técnicas principales aplicadas son el *factor analysis* (análisis factorial), *linkage analysis* (análisis de enlaces) y *cluster anlysis* (análisis de conglomerados) que serán analizadas seguidamente.

Factor Analysis (Análisis factorial)

Procedimiento estadístico multivariado que parte de la MCV con el objetivo de explicar la estructura de las covariaciones entre variables (*V*) mediante la definición de una cierta cantidad de factores (*F*), siendo $F < V$. De esta manera, las variables observadas se pueden expresar como combinaciónes lineales de los factores. Su aplicación se realiza principalmente para descubrir relaciones que lleven a determinar "dimensiones latentes" como causas profundas que no son directamente observables y que actúan como responsables de las manifestaciones visibles permitiendo una interpretación estructural.

De lo anterior surge que al análisis de factores podría ser utilizado con doble propósito: como método que permite reducir la dimensionalidad del sistema y como método para descubrir su estructura subyacente a partir de los factores.

El método de los componentes principales (PCA *Principal Components Analysis*) se utiliza para el logro del primer objetivo a partir de definir componentes principales que explican consecutivamente la mayor proporción de la varianza total. El método del análisis factorial (FA *Factorial Anlaysis*) define factores como dimensiones latentes que explican la mayor proporción de la varianza común (covarianzas). Los componentes se consideran reales y los factores hipotéticos (Kline, 1994).

El análisis factorial permite determinar cuánto influye cada factor en el comportamiento de la variable, lo cual queda expresado en la primera ecuación factorial:

[35] $$V_j = a_{j1}F_1 + a_{j2}F_2 + a_{j3}F_3 + \cdots + a_{jm}F_m + b_jS_j + c_j\varepsilon_j$$

Por lo tanto, la variable V_j es dependiente de los factores comunes ($F_1 \ldots F_m$), de los factores específicos (S_j) y de los factores de error (ε_j) de cada variable individual como factores de error. Los coeficientes de cada variable en cada factor se denominan pesos factoriales o saturaciones. A partir de aquí, la segunda ecuación factorial presenta la varianza de V_j explicada por la totalidad de los factores.

[36] $$\sigma_j^2 = 1 = a_{j1}^2 + a_{j2}^2 + a_{j3}^2 + \cdots + a_{jn}^2 + b_j^2 + c_j^2$$

La suma de los pesos factoriales al cuadrado es igual a 1, correspondiendo al total de la explicación, siendo a^2 es la varianza común, b^2 la varianza específica y c^2 la varianza de error.

El procedimiento obtiene tantos factores como variables y esto cubrirá el 100 % de la variabilidad [33], aunque en general se eligen los factores que tienen el mayor poder explicativo [34], principalmente aquellos que explican más que una variable de manera individual.

[37] $$F + S + \varepsilon = 1\,,\, \sigma_j^2 = 1$$

[38] $$F + S + \varepsilon < 1\,,\, \sigma_j^2 < 1$$

A continuación, se desarrolla el procedimiento matemático para la obtención de los factores, denominado método del centroide, el cual ubica cada eje factorial en el centroide de vectores en el espacio *n*-dimensional. Fue desarrollado en el campo de la psicología por Thurstone (1947), ejemplificado por Cortada de Kohan (1980, 1994) y puesto en práctica por Buzai (2003) para el análisis socioespacial urbano. Contempla los siguientes pasos a partir de la MCV:

— Cambiar la diagonal principal de la matriz de correlaciones, reemplazando los valores (como se trata de la matriz de correlaciones, la diagonal debería tener valores 1) por la máxima correlación encontrada en cada columna, en valores absolutos.

— Sumar a modo de comprobación: los valores de cada columna, y verificar que den igual que la suma de los valores de cada fila (en ambas sumas no incluir los valores de la diagonal). En símbolos: $\sum r_{j\cdot} = \sum r_{i\cdot}$

— Se calcula un nuevo vector fila, con la suma de los valores de cada columna ($\sum r_{j\cdot}$)y los valores de las nuevas comunalidades. Llamaremos este nuevo vector fila $t_{\bullet 1}$

— Se suman los valores del vector fila $t_{\bullet 1}$. Al resultado lo llamaremos *T1,* luego se calcula $\sqrt{T1}$ obteniéndose los valores de saturación para el primer factor:

[39] $aj_1 = \frac{tj_1}{\sqrt{T1}}$

Siendo que la comprobación está dada por:

[40] $\sum aj_1 = \sqrt{T1}$

Los valores obtenidos como saturaciones factoriales deben consignarse en la matriz factorial como puntajes de cada variable en el Factor I.

Para obtener el Factor II

— Realizar la matriz residual parcial (MRP) como matriz cuadrada que contiene las saturaciones factoriales en filas y columnas siempre con valor positivo. Se completa multiplicando los valores y las comunalidades parciales corresponden al puntaje factorial al cuadrado.

— Realizar la primera matriz residual a partir de restar el valor de correlación de la MCV original por los valores obtenidos en el MRP y consignarlo en la celda correspondiente a cada uno:

[41] $\rho_{jk} = r_{jk} - (a_{j1} * a_{k1})$

Es posible realizar una nueva comprobación de la exactitud del procedimiento cuando se obtiene un resultado cero (0) al sumar los valores de cada columna sin su comunalidad.

— Colocar el valor de la comunalidad estimada sobre la comunalidad calculada. Esta estimación es el mayor valor de correlación de la columna, independientemente del signo de esta relación se consigna siempre en positivo.

— Obtener desestimando las comunalidades de la suma y si existe un valor negativo como máximo *r* en la nueva fila, se procede a la reflexión de la variable con la finalidad de simplificar los cálculos en un espacio positivo.

En el caso de que más de un valor de la nueva fila sea negativo, el proceso de reflexión se inicia con la variable cuyo valor negativo sea el más alto (en valor absoluto). Cuando aparecen correlaciones negativas, los vectores que representan las variables producen relaciones angulares con $\theta > 90^{\circ}$, lo que significa que sus direcciones al graficarlos están en sentidos opuestos (hasta llegar a 180º, lo que representaría la máxima correlación negativa r=-1). El procedimiento de Análisis Factorial por método de centroide basa su procedimiento en la ubicación de los ejes factoriales en un espacio de signo positivo, por lo tanto, aquellos vectores que son negativos y se encuentran debajo de un plano imaginario, deberán ser reflejados con la finalidad de atribuirles puntajes positivos que favorecen los cálculos sin alterar, en una instancia posterior, el resultado final.

— Bajar el valor seleccionado a una nueva fila pasándola con valor positivo, se determina la fila-columna a ser reflejada y se genera una nueva fila para la variable sumándole el puntaje de la fila anterior, en este caso .

— Realizar la sumatoria de los valores obtenidos para la fila de la variable y verificar la existencia de valores negativos, si los hubiera se procede a una siguiente reflexión si se encontrara un mayor valor negativo.

Desde un punto de vista geográfico, cabe destacar que, además de la necesidad de realizar una correcta interpretación de los factores seleccionados en la MF resulta esencial poder ver la forma en que cada uno de los factores se distribuye espacialmente. Esta última etapa se logra a partir de la generación de la matriz de calificaciones factoriales (MCF) a partir de la siguiente multiplicación de matrices y como último procedimiento técnico del análisis factorial aplicado al análisis espacial:

[42] $\mathrm{MCF} = \mathrm{MDE}(z) * \mathrm{MF}$

En el resultado, cada fila de la MCF representa una unidad espacial y cada columna un factor. Por consiguiente, cada calificación factorial expresa la intensidad en que cada factor se encuentra presente en cada una de las unidades espaciales.

Finalmente se nombra cada factor en base al análisis del peso relativo en el que predominan las diferentes variables y se realiza el mapeo de cada uno pudiéndose captar la estructura socioespacial subyacente en el área de estudio. En este sentido cabe mencionar que esta línea de aplicaciones generó las bases para la definición de la Ecología Factorial como campo aplicativo que evoluciona desde los estudios de Sociología Urbana de la Escuela de Chicago (Berry y Kasarda, 1977).

Linkage Analysis (Análisis de enlaces)

El método fue propuesto por McQuitty (1957) para su aplicación en el estudio de test psicológicos. Desde un punto de vista geográfico fue presentado inicialmente por Racine y Reymond (1973) y se incluye en Haggett (1977). Su aplicación en el campo de la Geoinformática y los SIG fue realizado en Buzai y Baxendale (2006).

Los procedimientos metodológicos tradicionalmente fueron aplicados a la MCV(z) en la definición de macrovariables y en Geografía se amplió la posibilidad a la MCUE(z) en la construcción de regiones. Los diferentes pasos técnicos incluyen:

— Seleccionar los valores máximos de correlación positiva en cada columna de la matriz.

— Encontrar a qué par de unidades espaciales corresponden estas máximas correlaciones positivas y realizar un listado donde quede expresada la unión de unidades y el correspondiente coeficiente de correlación. Ejemplo de 6 entidades:

[43] $(A \longrightarrow D)(B \longrightarrow E)(C \longrightarrow B)(D \longrightarrow A)(E \longrightarrow B)(F \longrightarrow D)$

— Encontrar los pares recíprocos. Unidades en que se produce la máxima correlación de manera bidireccional. Realizar su listado. Continuación:

[44] $(A \Leftrightarrow D)(B \Leftrightarrow E)$

— Dibujar los pares recíprocos en orden uniendo las unidades en líneas gruesas.

— Volver al listado y determinar las unidades residuales. Son las que quedaron sin unir y serán vinculadas a alguna variable del par recíproco. Continuación:

[45] $(A \Leftrightarrow D \longleftarrow F)(C \longrightarrow B \Leftrightarrow E)$

— Realizar la matriz de especificidad de variables por grupos promediando los valores del par recíproco obtenidos en la MDE(z) o de unidades espaciales por grupos obtenidos en la MDE(z)t(z) (matriz geográfica en instancia operativa).

La técnica de *linkage analysis* presenta una solución única y clara en la que cada variable o unidad espacial se incluye en un grupo específico que tiende hacia el valor promedio como centroide del par recíproco.

Cluster Analysis (Análisis de conglomerados)

Es un procedimiento utilizado para la obtención de grupos homogéneos de entidades en base a la totalidad de variables medidas en ellas (consideradas n-dimensiones). El objetivo general es generar clases, dentro de las cuales los elementos componentes tendrán máxima similitud. El objetivo es maximizar la homogeneidad en el interior del grupo y la heterogeneidad entre los grupos.

Parte de la matriz de correlaciones en donde el coeficiente r se lo considera una medida de distancia entre cada par de entidades. En base a esto hay una serie de métodos que pueden ser utilizados para realizar los agrupamientos: vecino más próximo, vecino más alejado, promedio, centroide y método de *Ward*.

Utilizar los coeficientes r de Pearson para aplicar la técnica del vecino más próximo implica trabajar con los mayores valores de correlación. Los diferentes pasos técnicos incluyen:

— Encontrar el máximo valor de correlación (desestimando la diagonal $r = 1$) y a que unidades corresponde.

— Realizar una nueva matriz de correlaciones de n-1 x n-1 en la cual aparecen unidas las unidades seleccionadas. Para ello se genera una nueva columna a partir del promedio de las dos que se unen en MDE(z) y MDE(z)t(z) y se procede al cálculo de correlaciones. Se calculan los valores de correlación entre esta nueva entidad y todas las demás.

— Determinar el valor de correlación máximo en la nueva matriz y continuar con el procedimiento con una nueva matriz que desde la original sería de n-2 x n-2. Se realiza el mismo procedimiento, n-3 x n-3 hasta n-(n-1) x n-(n-1) en donde queda reducido en un único número.

El procedimiento brinda como resultado un dendrograma de correlaciones que se debe cortar para obtener la solución óptima. Existirá la posibilidad de realizar un agrupamiento dinámico de unión de variables formando macrovariables o de unidades espaciales en un proceso de regionalización. Es posible incluir una restricción de contigüidad cuando se eligen los mayores valores de correlación ante la comprobación de que sean unidades espaciales vecinas en base a un grado de contigüidad en regiones homogéneas.

El método *cluster analysis* brinda una serie de excelentes resultados clasificatorios para trabajos empíricos (Buzai y Baxendale, 2012), los presenta a través de una visión general a partir de la selección de diferentes puntos de corte y permite identificar claramente relaciones que difícilmente son accesibles a partir del análisis de unidades espaciales individuales.

Contexto computacional

El diagrama de secuencias presentado en la figura 1 y detallado en el presente texto, en la actualidad, se resuelven en el interior del campo de la Geoinformática, donde se combinan principalmente planillas de cálculo, programas de análisis estadístico y los SIG que permiten la combinación de resultados cuantitativos con las bases cartográficas digitales.

Muchas de las técnicas presentadas son operativas a partir del simple uso de los comandos de *software* específico. El paso por la MDO, MDI, MDE(Ω) y MDE(z) se realiza mediante el uso de planillas de cálculo de forma externa o como trabajo interno dentro del trabajo matricial en los SIG.

A partir de allí el primer bloque de procedimientos de análisis espacial se aplica a estas matrices para la realización de cartografía temática y exploración mediante el ESDA, cuyas posibilidades se encuentran disponibles a través de los denominados Sistemas de Ayuda a la Decisión Espacial (SADE) los cuales, de acuerdo a proceso detallado por Eastman (2007), permitieron que los SIG incorporan nuevas potencialidades por verticalización, un proceso de crecimiento continuo que a partir de las nuevas versiones de *software* incorpora nuevas capacidades o desarrolla aún más las existentes.

Los procedimientos para el trabajo con indicadores de planificación (IP) o la realización de puntajes de clasificación espacial (PCE) se resuelven en el interior del SIG mediante el trabajo en tablas y la selección de las variables estandarizadas correspondientes.

La MCV surge de la MDE(z) y en ella se aplican procedimientos para la reducción de la dimensionalidad estructural. La MCUE surge de la MDE(z)t(z). En ella se aplican procedimientos para la generación de regiones a partir de la unión de unidades espaciales que dividen el espacio geográfico en unidades de base. Estos procedimientos se realizan con programas de análisis estadístico y en el caso del trabajo geográfico eluden la transposición al poder explicar técnicas indistintamente en el nivel de las columnas o filas indistintamente.

Los procedimientos del *Factor Analysis*, *Linkage Analysis* y *Cluster Analysis* llegan a la clasificación espacial mediante variables y a la regionalización mediante unidades espaciales, en esta se trabaja con la matriz geográfica en una instancia operativa y resulta el nivel superior de la construcción regional como aporte de la Geografía Cuantitativa.

Conclusiones

El trabajo matricial con datos geográficos resulta ser fundamental en el análisis espacial y en el aporte que el paradigma de la Geografía Cuantitativa brinda para el estudio de las estructuras socioespaciales y su regionalización. En este sentido, se presenta como herramienta para la construcción de modelos de la organización espacial.

Durante seis décadas, esta línea de aplicación transitó un camino que actualmente confluye con los SIG para seguir avanzando en construcción de aspectos conceptuales y operativos. El campo de la Geoinformática proporcionó claras alternativas de resolución. Actualmente los caminos metodológicos permiten delinear un modelado de secuencia completo y esta construcción constituye el objetivo central de este trabajo.

El proceso parte de la MDO y, a partir de ella, avanza en sucesivas transformaciones con diversos alcances en la búsqueda de conocer la realidad geográfica como emergencia de diversos aspectos subyacentes. Asimismo, se pueden apreciar los momentos en que puede ser utilizada la matriz geográfica en las instancias de organización descripción y aplicación.

El trabajo con matrices tradicionales permite la realización de cartografía temática, la aplicación de las técnicas del ESDA, el cálculo de IP y PCE para finalizar con la obtención de macrovariables síntesis de las temáticas analizadas. El trabajo con la matriz geográfica lleva a la regionalización como resultado final de una síntesis espacial estructural, surgida de la MCUE en la que se combinan la totalidad de datos incorporados en el estudio y, como resultado final, la regionalización surge como máxima síntesis.

Teniendo en cuenta la importancia actual de los entornos digitales y la automatización computacional para el tratamiento de datos espaciales en la investigación en Geografía fue considerado que realizar una síntesis global del tratamiento matricial resultaba necesario ante el intento de demarcar, en base a las condiciones actuales, procedimientos centrales que la Geografía Cuantitativa. El diagrama de secuencias resume el camino recorrido que comenzó hace medio siglo y muestra de que manera apoya la esencia disciplinar con base en el análisis espacial.

Bibliografía

Anselin, L. (1996). *Interactive Techniques and Exploratory Spatial Data Analyisis*. Regional Research Institute Working Papers. Morgantown: University of West Virginia.

Baxendale, C.A. (1992). *Apuntes del curso Manejo de información cuantitativa atribuida espacialmente*, dictado por Darío C. Sánchez. Buenos Aires: Universidad del Salvador.

Berry, B.J.L. (1964). Approaches to Regional Analysis: A Synthesis. *Annals of the Association of American Geographers*, 54, 2-11.

Berry, B.J.L.; Kasarda, J. (1977). *Contemporary Urban Ecology*. New York: McMillan.

Berry, B.J.L.; Marble, D. (1968). *Spatial Analysis: A Reader in Statistical Geography*. New York: Englewood Cleefs, Prentice Hall.

Bosque Sendra, J., Zamora Ludovic, (2002). Visualización geográfica y nuevas cartografías, *GeoFocus*, 2, 61-77.

Bunge, W. (1962). *Theoretical Geography*. Lund: The Royal University of Lund, Lund Studies in Geography 1.

Bunge, W. (1966) *Apendix to Theoretical Geography*. Lund: The Royal University of Lund, Lund Studies in Geography 6.

Burgess, E.W. (1925). The growth of the city: an introduction to a research project. En R.E. Park, E.W. Burgess [Eds.]. *The City*. Chicago: The Chicago University Press.

Burton, I. (1963). The Quantitative Revolution and Theoretical Geography, *The Canadian Geographer*, 72, 1, 151-162.

Buzai, G.D. (2003). *Mapas Sociales Urbanos*. Buenos Aires: Lugar Editorial. (1° ed.)

Buzai, G.D. (2014). *Mapas Sociales Urbanos*. Buenos Aires: Lugar Editorial. (2° ed.)

Buzai, G.D. (2021) The World Map by Anaximander (Miletus, 5th Century BC): Modeling Geographical Space at the Beginning of Science. *Anales de la Sociedad Científica Argentina*, 272(3), 5-18.

Buzai, G.D. (2023a). Geografía, mapas y modelización. Criterios desde el realismo y la cuantificación espacial. *Revista Universitaria de Geografía*, 32(1), 133-157.

Buzai, G.D. (2023b). Por una Geografía científica: Fred K. Schaefer_1953. *Pleamar*, 3, 85-100.

Buzai, G.D.; Baxendale, C.A. (2006). *Análisis socioespacial con Sistemas de Información Geográfica*. Buenos Aires: Lugar Editorial.

Buzai, G.D., Baxendale, C.A. (2008). Clasificación de unidades espaciales mediante el uso de indicadores de planificación. *Publicaciones del PROEG*, 6, 1-57.

Buzai, G.D., Baxendale, C.A. (2012). *Análisis Socioespacial con Sistemas de Información Geográfica*. (Tomo 2). Buenos Aires: Lugar Editorial.

Buzai, G.D., Baxendale, C.A., Humacata, L., Principi, N. (2016). *Sistemas de Información Geográfica. Cartografía temática y análisis espacial*. Buenos Aires: Lugar Editorial.

Buzai, G.D., Lanzelotti, S., Principi, N., Montes Galbán, E., Humacata, L., Acuña Suárez, G., Baxendale, C.A. (2022). Atlas regionales con Sistemas de Informacion Geografica en la generacion de resultados de transferencia. Atlas de Geografia Humana de la cuenca del rio Lujan, Provincia de Buenos Aires, Argentina. En: Gutiérrez Gallego, J.A., Schnabel, S., Lavado Contador, J.F., Castro Serrano, J. (Coord.). *La transferencia de los resultados de investigación para el desarrollo territorial sostenible*, Madrid, Dykinson, pp. 69-89.

Buzai, G.D.; Robinson, D.J. (2010). Geographical Information Systems (GIS) in Latin America, 1987-2010: A Preliminary Overview. *Journal of Latin American Geography*, 9(3), 9-31.

Christaller, W. (1933). *Die Zentralen Orte in Süddeuschland*, Jena. (Trad: Central Places in Southern Germany, New York, Englewood Cliffs – Prentice Hall, 1966).

Chorley, R. [Ed.] (1973). *Directions in Geography*. London: Methuen. (Trad: Nuevas tendencias en Geografía, IEAL, Madrid, 1973).

Cole, S.; King, L. (1968). *Quantitative Geography*. London: John Wiley & Sons.

Cortada de Kohan, N. (1980). El análisis factorial en la investigación geográfica. *Cuadernos de Geografía*, 8.

Cortada de Kohan, N. (1994). *Diseño Estadístico*. Buenos Aires: EUDEBA.

Eastman, J.R. (2007). La verticalización de los Sistemas de Información Geográfica: En: Buzai, G.D. (Ed.) *Memorias de la XI Conferencia Iberoamericana de Sistemas de Información Geográfica*, Buenos Aires: Universidad Nacional de Luján, pp. 183-195.

Feuillet, T.; Cossart, E.; Commenges, H. (2019). *Manuel de géographie quantitative. Conceps, outils, methodes*. Paris: Armand Colin.

Fotheringham, A.S.; Brunsdon, Ch.; Charlton, M. (2000). *Quantitative Geography. Perspectives on Spatial Analysis*. London: SAGE.

Garrison, W.; Marble, D. [Eds.] (1967). *Quantitative Geography*. Evansthon: Northwestern Studies in Geography.

Grecousis, G. (2020). *Spatial Analysis Methods and Practice*. Cambridge: Cambridge University Press.

Haggett, P. (1975). *Locational Analysis in Human Geography*. London: Edward Arnold.

Haggett, P. (1977). *El análisis locacional en la Geografía Humana*. Barcelona: Gustavo Gili.

Harris, Ch.; Ullaman, E. (1945). The nature of cities, *Annals of the American Academy of Polictical and Social Sciences*, 242, 7-17.

Harris, R. (2016). *Quantiative Geography. The Basis*. London: SAGE.

Hartshorne, R. (1939). The Nature of Geography: A Critical Survey of Current Thought in the Light of the Past, *Annals of the Association of American Geographers*, 29, 173-658.

Harvey, D. (1969). *Explanation in Geography*. London: Edward Arnold.

Hoyt, H. (1939). *The structure and growth of residential neighborhoods in American cities*. Washington: Federal Housing Administration.

Johnston, J.R. (1968). Choice in Classification: The Subjetivity of Objective Methods. *Annals of the Association of American Geographers*, 58(3), 575-589.

Johnston, J.R. (1987). *Geography and Geographers. Anglo-American Human Geography Since 1945*. London: Hodder Arnold.

Kline, P. (1994). *An easy guide of factor analysis*. London: Routledge.

Kuhn, T.S. (1994) *La estructura de las revoluciones científicas*. México: Fondo de Cultura Económica. Original en inglés: 1962.

Lösch, A. (1939). *Die räumliche Ordnung der Wirtschaft*. Jena. (Trad: The Economics of Location, Yale University, New Haven, 1954).

Malczewski, J. (1999). *GIS and multicriteria decision analysis*. New York: John Wiley & Sons.

Manley, G. (1966). A New Geography, *The Guardian*, March 17th.

McQuitty, L.L. (1957). Elementary Linkage Analysis for Isolating Orthogonal and Oblique Types and Typal Relevancies, *Educational and Psichological Measurement*, 17, 207-229.

Principi, N., Montes Galbán, E., Buzai, G.D. (2023). Hacia un atlas-web interactivo: Una propuesta de extensión universitaria para la enseñanza en Geografía. En: Nieto Masot, A., Cárdenas Alonso, G., Gutiérrez Gallego, J.A., Engelmo Moriche, A. (Eds.). *Actas de la XVIII Conferencia Iberoamericana de Sistemas de Información Geográfica*, Cáceres: Universidad de Extremadura, 469-474.

Racine, J.B.; Reymond, B. (1973). *L´Analyse quantitative en géographie*. Paris: Presses Universitaires de France.

Rey Balmaceda, R.C. (1973). *Geografía Regional. Teoría y Aplicación*. Buenos Aires: Estrada.

Sarkar, A. (2013). *Quantitative Geography: Techniques and Presentations*. Hyderabad: Orient BlackSwan.

Schaefer, F. (1953). Excepcionalism in Geography: A Methodological Examination, *Annals of the American Association of Geographers*, 43(3), 226-249.

Stewart, J.; Warntz, W. (1958). Macrogeography and social science, *Geographical Review*, 48, 167-184.

Thurstone, L.L. (1947). *Multiple Factor Analysis: A Development and Expansion of Vectors of the Mind*. Chicago: The University of Chicago Press.

Tomlin, C.D. (1990). *Geographic Information Systems and Cartographic Modeling*. New Jersey: Prentice Hall, Englewood Clift.

Tukey, J.W. (1977). *Exploratory Data Analysis*, Reading: Addison-Wesley.

Thünen, H. von (1826). *Der Isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie*. Rostok (Trad: The Isolated State, Oxford, Pergamo Press, 1966.

Weber, A. (1909). Über *den Standort der Industrien*. Tübingen. (Trad: Theory of Location of Industries, Chicago, 1929).

Fecha de recepción: 22 de diciembre de 2023
Fecha de aceptación: 10 de julio de 2024

© 2024 por los autores; licencia otorgada a la Revista Universitaria de Geografía. Este artículo es de acceso abierto y distribuido bajo los términos y condiciones de una licencia Atribución-NoComercial 4.0 Argentina de Creative Commons. Para ver una copia de esta licencia, visite https://creativecommons.org/licenses/by-nc/4.0/deed.es