

---

**ANÁLISIS DE COMPONENTES PRINCIPALES CON DATOS  
GEOREFERENCIADOS.  
UNA APLICACIÓN EN LA INDUSTRIA TURÍSTICA.**

Diciembre de 2017

---

**Tesis para optar al grado académico  
de Magister en Estadística Aplicada**

Candidata: Lic. Laura Isabel Luna

Directora: PhD. Mónica Balzarini



Análisis de componentes principales con datos georeferenciados.  
Una aplicación en la industria turística. by Luna, Laura Isabel is  
licensed under a [Creative Commons Reconocimiento-  
NoComercial 4.0 Internacional License](https://creativecommons.org/licenses/by-nc/4.0/).

## AGRADECIMIENTOS

*Esta tesis es producto del esfuerzo realizado a lo largo de los últimos años, compaginado con trabajo y responsabilidades, en la cual han me han acompañado un grupo grande de compañeros y amigos.*

*Quiero en primer lugar agradecer a mi directora, la Dra. Mónica Balzarini, por compartir su inmenso conocimiento, por su generosidad y su energía positiva, como así también haberme guiado en este proceso.*

*También le agradezco a los docentes del Instituto de Estadística y Matemática de la Facultad de Ciencias Económicas, por compartir su sabiduría y sus valiosos consejos, en especial a la Dra. Norma Patricia Caro, la Mgter. Mariana Gonzalez, quienes me alentaron permantemente y al Dr. Fernando García por sus recomendaciones.*

*Agradezco a mis compañeros de la Dirección General de Estadística y Censos, en especial a todos los de la Dirección de Estadísticas Económicas por su apoyo permanente, por hacerme reír y por el esfuerzo diario puesto en la construcción de estadísticas oficiales.*

*Agradezco a la Mgter. Irene Cuello por enseñarme sobre turismo y muchas cosas más, así como sus valiosos aportes y contribuciones. También agradezco a los miembros de la Facultad de Turismo y Ambiente de la provincia.*

*Agradezco a cada uno de mis profesores de la maestría en estadística y a mis compañeros de los cuales aprendí muchísimo. Y a Gustavo Casanova siempre ayudando en las gestiones administrativas.*

*Agradezco a los Miembros del Tribunal de Tesis, Dra. María Inés Stimolo, Dra. Cecilia Bruno y Mgter Regina Cavallin por aceptar gentilmente formar parte del tribunal examinador y por dedicar su valioso tiempo a la revisión de este trabajo.*

*A todos los que forman parte de mi vida, muchas gracias.*

*A Marcos, Jano y Bastian*

*mis amores, mis motores*

*A Manuel, Susana, Gustavo y Liliana (en el cielo)*

*mis padres, por el apoyo permanente*

## RESUMEN

El análisis espacial de las actividades características del turismo permite generar información relativa a la estructura de la industria turística, información necesaria para la toma de decisiones. En este trabajo se propuso realizar un mapeo de las actividades características del turismo en departamentos de Córdoba, Argentina. La innovación metodológica consiste en la generación de estadísticas para datos espaciales multidimensionales. Se estudian y comparan en su desempeño métodos multivariados con y sin restricción espacial en el contexto de la aplicación. De esta comparación se pudo concluir que el grado de estructuración espacial de las componentes que sintetizan las actividades características del rubro fue mayor cuando se aplicó el análisis de componentes principales espaciales (MULTISPATI-PCA) que con el análisis de componentes principales clásico (PCA) sobre indicadores de la actividad. Los mapas de variables índices mostraron mayor estructura espacial con MULTISPATI-PCA. Se realizaron agrupamientos de departamentos en función de la participación de las actividades de interés en el valor agregado turístico, para ello se utilizaron componentes principales espaciales como input del análisis de conglomerados *fuzzy k-means*. Finalmente se realizó un mapeo en función de la participación del valor agregado turístico en el producto bruto regional de los departamentos y se analizó la variación en la participación de las distintas actividades que componen el agregado para el período 2001-2014.

Palabras clave: actividades características del turismo, análisis multivariado espacial, cluster fuzzy k-means

## **ABSTRACT**

The spatial analysis of the tourism characteristic activities allows us to generate information about the structure of tourism industry, which is necessary for decision making. In this work, tourism characteristic activities in the departments of Córdoba were mapped. The methodological innovation lies in the generation of statistics for multidimensional spatial data. Multivariate methods with and without spatial restrictions were studied and compared in their performance in the application context. The comparison showed that the spatial principal components analysis (MULTISPATI-PCA) yielded a higher degree of spatial structuring of the components that summarize tourism activities than principal components analysis (PCA). The maps of the summarized variables showed a higher level of structure with MULTISPATI-PCA. Departments were classified according to the participation of tourism activities in the value added of tourism using the spatial principal components obtained as input of the cluster fuzzy k-means analysis. Finally, a mapping was performed based on the participation of the tourism value added in the gross regional product of the different departments and the variations in the participation of the different activities that make up the aggregate was analyzed for the period 2001-2014.

**Key words:** activities characteristic of tourism, spatial multivariate analysis, cluster fuzzy k-means

## TABLA DE CONTENIDOS

CAPÍTULO 1. INTRODUCCIÓN GENERAL .....	1
ANTECEDENTES .....	1
OBJETIVO GENERAL.....	5
OBJETIVOS ESPECÍFICOS .....	5
CAPÍTULO 2. VARIABILIDAD ESPACIAL .....	6
INTRODUCCIÓN .....	6
MATERIALES Y MÉTODOS .....	8
Datos .....	8
Índices de Autocorrelación Espacial.....	9
RESULTADOS Y DISCUSIÓN .....	13
Cálculo de Índices de Moran y Geary .....	13
CAPÍTULO 3. ANALISIS MULTIVARIADO .....	18
INTRODUCCIÓN .....	18
MATERIAL Y MÉTODOS.....	20
Análisis de componentes principales clásico y con restricción espacial .....	20
RESULTADOS Y DISCUSIÓN .....	23
CAPITULO 4. CLASIFICACIÓN NO SUPERVISADA ESPACIAL MULTIVARIADA .....	31
INTRODUCCION .....	31
MATERIAL Y MÉTODOS.....	32
Análisis de clúster fuzzy k-means .....	32

Clasificación de áreas .....	37
RESULTADOS Y DISCUSIÓN .....	39
Clasificación de departamentos vía clúster fuzzy k-means .....	39
Participación del valor agregado turístico en el producto bruto regional .....	42
Clasificación de departamentos vía clúster fuzzy k-means en función de la variabilidad de las variables de entrada .....	45
CONCLUSIONES .....	47
CONCLUSIONES FINALES .....	49
REFERENCIAS BIBLIOGRÁFICAS .....	51
ANEXO 1 .....	54
ANEXO 2 .....	55
ANEXO 3 .....	58
ANEXO 4 .....	61
ANEXO 5 .....	63
ANEXO 6 .....	64

## LISTA DE TABLAS

Tabla 1: Índices de autocorrelación espacial actividades características del turismo (participación VARCT).....	14
Tabla 2: Descriptiva univariada para cada actividad. Participación en el valor agregado turístico (VARCT).....	15
Tabla 3: Autovalores, varianza espacial e índices de Moran de las componentes principales generadas a partir de MULTISPATI-PCA (sobre la participación promedio de las actividades).....	26
Tabla 4: Autovalores e índices de Moran de las componentes principales generadas a partir del PCA (sobre la participación promedio de las actividades).....	27
Tabla 5: Selección del número de clases de la partición de departamentos .....	40
Tabla 6: Autovalores, varianza espacial e índices de Moran de las componentes principales generadas a partir de MULTISPATI-PCA .....	45

## LISTA DE FIGURAS

Figura 1. (a) Triangulación de Delaunay; (b) Gráfico de Gabriel ; (c) Gráfico de Vecindad Relativa; (d) Gráfico Esfera de Influencia.....	11
Figura 2. Red de Conexión.....	14
Figura 3. Gráfico Biplot, Análisis de Componentes Principales obtenido a partir de la participación de las actividades (H, R, ST, AV, SC, SD, AV) en los departamentos de la provincia de Córdoba.....	26
Figura 4. Representación gráfica de los dos primeros ejes del PCA (izquierda) y MULTISPATI-PCA (derecha) que muestra la correlación entre las variables y entre éstas y las componentes principales. ....	28
Figura 5. Variabilidad espacial multivariada en contribución de las ramas características del turismo al VARCT. Los valores de variables sintéticas son logrados con la CP1 del PCA (izquierda) y CP1 del MULTISPATI-PCA (derecha).....	30
Figura 6. Mapas de variabilidad espacial multivariada logrados con la CP2 del PCA (izquierda) y CP2 del MULTISPATI-PCA (derecha).....	30
Figura 7. Mapa con clases delimitadas: a) dos (izquierda), b) tres (centro) y c) cuatro (derecha) clases.....	40
Figura 8. Mapa participación promedio del VARCT en el producto bruto regional.....	45

## LISTA DE SÍMBOLOS Y ABREVIATURAS

CP: Componente principal

CV: Coeficiente de variación

IM: Índice de Moran

IG: Índice de Geary

MULTISPATI- PCA: Análisis de componentes principales restringido espacialmente

PCA: Análisis de componentes principales

PC: Componentes principales

sPC: Componentes principales espaciales del MULTISPATI-PCA

AP: Agricultura de Precisión

KM-sPC: Análisis de cluster fuzzy k-means sobre componentes principales espaciales

PGB: Producto geográfico bruto

PBR: Producto bruto regional

VARCT: Valor agregado de las ramas características del turismo

H: Hoteles y Similares

SV: Segundas viviendas en propiedad

R: Restaurantes y similares

ST: Servicios de transporte de pasajeros por carretera

AV: Agencias de viajes y similares

SC: Servicios culturales

SD: Servicios de actividades deportivas y otras actividades de recreo

# CAPÍTULO 1. INTRODUCCIÓN GENERAL

## ANTECEDENTES

En ocasiones, cuando se trabaja con datos económicos es necesario incorporar al estudio referencias geográficas, dado que una cuestión importante se refiere a la identificación y explicación de la variabilidad espacial de las estructuras económicas. La identificación y medición de las estructuras espaciales ha sido un tema importante en la estadística aplicada. La importancia y relevancia de los métodos que analizan los efectos espaciales en los modelos económicos se ha ido incrementando de forma notable en estos últimos años, esto en parte, debido al renovado interés por el papel del espacio y la interacción espacial en la teoría económica, que han desarrollado algunos trabajos como los de la nueva geografía económica (Krugman, 1991), por otro lado la disponibilidad de nuevas fuentes estadísticas de información microterritorial (censos, directorios, etc.), así como mejores herramientas informáticas (desarrollo de software específicos para el tratamiento de datos espaciales) han dado lugar a una demanda creciente de este tipo de análisis espacial.

Los métodos de análisis multivariado se utilizan para resumir conjuntos de datos en variables nuevas construidas como transformaciones de las originales, con mínima pérdida de información. Dentro de estos métodos se destaca el análisis de componentes principales (PCA, Pearson, 1901), esta técnica permite identificar las variables que explican la mayor parte de la variabilidad total contenida en los datos, explorar las correlaciones entre variables y reducir la dimensión del análisis al combinar las variables en nuevos índices (variables

sintéticas); cada una de estas nuevas variables sintéticas es denominada componente principal (CP).

Las técnicas multivariantes estándar pueden ser utilizadas con éxito en conjuntos de datos para los cuales se conoce la posición espacial de cada unidad de análisis (datos georreferenciados). En estos casos, usualmente primero se obtienen variables sintéticas y luego se describe la variación espacial de éstas para obtener mapas de contorno en relación a valores de las nuevas variables índice. Sin embargo, en este tipo de análisis no se toman en cuenta las relaciones espaciales directamente en el cómputo de las componentes; las técnicas de ordenamiento, como PCA, no fueron específicamente diseñadas para identificar estructuras espaciales. La incorporación de la información geográfica se realiza usualmente a posteriori del PCA mediante la asignación de los valores de las componentes a cada uno de los sitios georreferenciados o ajustando semivariogramas (Schabenberger et al., 2002) y usando otras herramientas de la geoestadística clásica (Kriging) para lograr mapas de variabilidad espacial por interpolación.

Es posible detectar la presencia de autocorrelación espacial de las CPs utilizando estadísticos de autocorrelación univariados como el índice de Moran (Moran, 1948) o el de Geary (Geary, 1954). Estos análisis se abordan en forma univariada dificultando la interpretación de la variabilidad conjunta (Córdoba et al. 2012). Dray (2008), propone un método de análisis multivariado que incorpora la información espacial previa al análisis multivariado, el método es conocido como MULTISPATI-PCA. Éste se basa en el PCA pero incorpora la restricción dada por los datos espaciales mediante el cálculo del índice de Moran para medir la dependencia o correlación espacial entre las observaciones en un sitio y el promedio de las observaciones en el vecindario de ese sitio. Para delimitar los vecindarios utiliza una matriz

de pesos espaciales determinando cuáles y cuántas observaciones cercanas a cada sitio conforman el vecindario (Córdoba, 2014).

Estos nuevos métodos multivariados podrían servir para mapear el espacio geográfico según la estructura espacial de indicadores económicos como son los datos de actividades relacionadas al turismo que conforman el Producto Geográfico Bruto. El análisis de las actividades características del turismo<sup>1</sup>, permite generar información relativa a la estructura de la industria turística, esta información mejora la toma de decisiones, dado que ofrece elementos necesarios para la apropiada definición de estrategias de desarrollo local.

El turismo es un fenómeno que desempeña un papel muy importante en la economía, ya que genera por sí solo un efecto multiplicador sobre la producción, el empleo y la renta en una región. En la provincia de Córdoba el sector turístico es señalado como uno de los sectores estratégicos de desarrollo económico. Además existe interés por parte de los decisores políticos en identificar áreas y regiones de desarrollo sobre las cuales ajustar los planes de desarrollo de infraestructura, equipamiento turístico y promoción.

En este trabajo se realiza un mapeo de las actividades características del turismo en los departamentos de la provincia de Córdoba. La finalidad es innovar en cuanto a las estadísticas

---

<sup>1</sup>Según la Organización Mundial de Turismo las industrias turísticas (también conocidas como actividades turísticas) son aquellas que generan principalmente productos característicos del turismo. Los productos característicos del turismo son aquellos que cumplen uno o ambos de los siguientes criterios: (a) El gasto turístico en el producto debería representar una parte importante del gasto total turístico (condición de la proporción que corresponde al gasto/demanda). (b) El gasto turístico en el producto deberían representar una parte importante de la oferta del producto en la economía (condición de la proporción que corresponde a la oferta). Este criterio supone que la oferta de un producto característico del turismo se reduciría considerablemente si no hubiera visitantes". Para mayor información consultar Anexo 6 de este documento.

para datos espaciales multidimensionales y suministrar elementos para la toma de decisiones regionales.

En el Capítulo II se describen, analizan e ilustran procedimientos estadísticos que permiten describir la variabilidad espacialmente estructurada en las actividades características del turismo, las herramientas utilizadas son de naturaleza univariada. En el Capítulo III se exponen dos técnicas multivariadas (análisis de componentes principales y análisis de componentes principales restringido espacialmente). Se realiza un análisis comparativo de los resultados obtenidos con la implementación de un PCA clásico y de la versión restringida espacialmente (MULTISPATI-PCA).

En el Capítulo IV se utilizan las componentes principales espaciales obtenidas con MULTISPATI-PCA como input del análisis de clúster fuzzy k-means para la identificación de grupos homogéneos de departamentos, la hipótesis que subyace a la metodología propuesta es que la incorporación de la autocorrelación espacial a través del PCA espacial aplicado sobre las actividades características del turismo obtenidas a escala departamental, producirá agrupamientos que contienen menos “ruido”, i.e. conjunto de departamentos más homogéneos y con mayor coherencia espacial. Además en este capítulo posterior a la regionalización se pretende realizar una estratificación según sea la participación de la actividad en el producto bruto regional (PBR)<sup>2</sup> y otra estratificación en función del cambio en la participación a lo largo del período analizado.

---

<sup>2</sup>El Producto Geográfico Bruto se define como el valor total de los bienes y servicios generados en una provincia durante un período de tiempo, que generalmente es un año, libre de duplicaciones. Cuando se toma como referencia un territorio económico menor a una provincia, éste se denomina Producto Bruto Regional. Es decir que, el Producto Bruto Regional (PBR) valúa el producto bruto total de cada una de las regiones en que se subdivide la provincia, siendo su objetivo cuantificar el valor agregado generado por los establecimientos residentes agrupados por clase de actividad económica, en cada una de las regiones.

## **OBJETIVO GENERAL**

Evaluar la aplicación de técnicas de análisis multivariado para datos espaciales en el contexto de indicadores de actividad turística a escala regional.

## **OBJETIVOS ESPECÍFICOS**

- 1) Modelizar la variabilidad espacial de las variables indicadoras de la actividad turística relevadas a escala departamental en la Provincia de Córdoba, Argentina.
- 2) Construir índices que permitan un mapeo multidimensional de la estructura del turismo en la Provincia. Reflejar como las actividades características del turismo contribuyen al valor agregado
- 3) Comparar metodologías para el tratamiento de datos espaciales en el contexto multivariado, PCA clásico y PCA con restricción espacial, en el desempeño estadístico con datos a escala regional.
- 4) Realizar agrupamientos departamentales en función a la participación de las actividades de interés en el valor agregado turístico.

- 5) Mapear el territorio de la Provincia de Córdoba en función de la participación del valor agregado turístico en el Producto Geográfico Bruto de los diferentes departamentos.

## **CAPÍTULO 2. VARIABILIDAD ESPACIAL**

### **INTRODUCCIÓN**

En los últimos años, la economía regional ha experimentado un fuerte desarrollo metodológico basado en la necesidad de trabajar los denominados efectos espaciales: la heterogeneidad y la dependencia espacial. La heterogeneidad espacial consiste en la variación de las relaciones en el espacio. La heterogeneidad espacial, se refiere a la falta de estabilidad en el espacio del comportamiento de la variable bajo estudio, esto implica, por ejemplo, que la forma funcional y los parámetros de una regresión puedan variar según la localización, siendo por tanto, no homogéneos en toda la muestra. El tratamiento de la heterogeneidad espacial puede realizarse por medio de las técnicas econométricas estándar. El fenómeno de la dependencia o autocorrelación espacial surge cuando el valor de una variable en un lugar del espacio, está relacionado con su valor en otro u otros lugares del espacio. No obstante, y a diferencia de lo ocurrido con la heterogeneidad espacial, la dependencia espacial no puede ser tratada por la econometría estándar. Ello es debido a la multidireccionalidad que domina las relaciones de interdependencia entre unidades espaciales (Moreno Serrano et al., 2002). Este trabajo se enfoca en el segundo fenómeno.

La identificación y medición de las estructuras espaciales es un tema importante en la estadística aplicada. La dependencia o autocorrelación espacial ha sido ampliamente estudiada, desde los primeros trabajos estadísticos de Cliff y Ord (1973) y Cressie N. (1985). 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 un proceso. Además, este efecto espacial ha sido objeto de aplicaciones en numerosos campos, como el de los recursos naturales y medio ambiente (Anselin, 2001), delincuencia (Baller et al., 2001; Ceccato et al., 2001) entre otros.

En el terreno específicamente económico, se ha demostrado la existencia de autocorrelación espacial en fenómenos como la producción, desocupación, renta disponible, procesos de convergencia regional, demanda de transporte, impuestos, y precio de la vivienda.

El análisis económico convencional ha dado, usualmente, mayor importancia al papel del tiempo como dimensión clave de estudio, sin valorar adecuadamente el factor espacial (Chasco, 2003). Considerar la dimensión espacial de los fenómenos bajo estudio implica preguntarse sobre la ubicación de éstos y la forma en que dicha distribución geográfica condiciona o influye dichos fenómenos (Sánchez Peña, 2012).

El fenómeno de autocorrelación espacial obliga a tener siempre en cuenta la configuración espacial de la muestra de valores. Esta variabilidad espacial genera correlaciones entre las observaciones de una misma variable registrada repetidamente en el espacio y por tanto los datos no pueden tratarse estadísticamente como datos independientes. Por otro lado, la investigación de los mercados en su distribución territorial es una importante necesidad para las instituciones públicas y privadas, que necesitan abordar acciones como la planificación territorial, programación de la actividad comercial e industrial, urbanismo y medio ambiente

(Chasco, 2003). Cuando una variable se distribuye de forma sistemática en un espacio geográfico se dice que el fenómeno en cuestión presenta 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. Los primeros índices formales para detectar la presencia de autocorrelación espacial se deben a Moran (Moran, 1948) y Geary (Geary, 1954).

En el presente Capítulo se describen, analizan e ilustran los principales procedimientos estadísticos que permiten describir la variabilidad espacialmente estructurada en las variables de interés, específicamente en las actividades características del turismo. Las herramientas usadas en este Capítulo son de naturaleza univariada, es decir aplicables variable a variable. Se provee de rutinas del software R (R Core Team, 2017) (Anexo 1) desde las cuales se pueden implementar cada una de las aproximaciones descriptas.

## **MATERIALES Y MÉTODOS**

### **Datos**

Se utilizarán datos procesados por la Dirección de Estadísticas Económicas (DEE) de la provincia de Córdoba para el periodo 2001 - 2014, correspondientes a la provincia de Córdoba y a sus 26 departamentos. Se trabajará con la serie de Producto Geográfico Bruto a valores constantes, debido a que presenta mayor estabilidad que a valores corrientes y nos permite analizar la estructura económica.

Para calcular el aporte de la industria turística al valor agregado provincial, se trabaja con las ramas de actividades características del turismo correspondiente a las categorías de: a) Hoteles y similares b) Segundas viviendas en propiedad c) Restaurantes y similares d) Servicios de transporte de pasajeros por ferrocarril e) Servicios de transporte de pasajeros por

carretera f) Servicios de transporte aéreo de pasajeros g) Servicios anexos al transporte de pasajeros h) Alquiler de bienes de equipo de transporte de pasajeros i) Agencias de viajes y similares j) Servicios culturales k) Servicios de actividades deportivas y otras actividades de recreo.

Previo al análisis, los datos fueron sometidos a un procedimiento de depuración. En primer lugar se seleccionaron aquellas ramas de actividad en las cuales se tuviera información en todos los departamentos, debido a que algunas de ellas, solo estaban presentes en algunos departamentos (por ejemplo servicios de transporte aéreo de pasajeros). Se consideraron las siguientes actividades: 1) Hoteles y similares (H), 2) Segundas viviendas en propiedad (SV), 3) Restaurantes y similares (R), 4) Servicios de transporte de pasajeros por carretera (ST), 5) Agencias de viajes y similares (AV), 6) Servicios culturales (SC) y 7) Servicios de actividades deportivas y otras actividades de recreo (SD). Los datos se normalizaron a través del cálculo de participaciones relativas, definidas como el cociente entre el valor de la rama específica y el valor agregado total (VARCT<sup>3</sup>) de cada departamento, esto permite trabajar con la composición de la actividad y evita las diferencias que surgen en valores absolutos (monetarios) a causa de los diferentes tamaños en la economía de los departamentos.

Luego de realizar un análisis de componentes principales (ACP) para cada uno de los años, se decide trabajar con el valor promedio del periodo (2001-2014), dado que no se observaron diferencias significativas en los resultados para cada uno de los años.

## **Índices de Autocorrelación Espacial**

---

<sup>3</sup> VARCT: Valor agregado de las ramas características del turismo. Surge como suma del valor agregado de las ramas características. (Anexo 6)

Dos de los índices más usados para medir autocorrelación espacial son el Índice de Moran (Moran, 1950) y el Índice de Geary (Geary, 1954). Estos estadísticos permiten medir el grado de dependencia entre observaciones de una misma variable en un sitio y en su vecindario.

El cálculo del índice o coeficiente de Moran de autocorrelación espacial en un espacio continuo requiere la definición de una matriz de ponderación espacial y para obtener estos ponderadores se pueden seguir distintos procedimientos, entre los que se destacan el uso de redes de conexión (Dray, 2011). Las redes de conexión o gráficos de vecindario se generan conectando individuos vecinos en un mapa (Legendre y Legendre, 1998).

Existen diferentes opciones o alternativas metodológicas para definir los vecindarios que dependen de los diferentes tipos de muestreo presente en los datos (grilla regular, irregular o transectas). En esta investigación, se trabajó con muestreos irregulares, los métodos usados para construir la red de conexión se basan en la triangulación de Delaunay (Lee y Schachter, 1980), el gráfico de Gabriel (Gabriel y Sokal, 1969), el concepto de vecinos más cercanos y de distancia euclídea entre observaciones. Cuando se trabaja con áreas irregulares es usual elegir un punto para representar al polígono, a menudo referido como el centroide del polígono.

En la figura 1 se muestran los diferentes métodos. La triangulación de Delaunay es un método recomendado para construir gráficos de vecindario cuando las entidades se encuentran distribuidas en forma homogénea en el espacio. Sin embargo, puede conectar a entidades periféricas que no deberían estar relacionadas. El gráfico de Gabriel es un subconjunto del gráfico de Delaunay que no incluye las conexiones periféricas. Las redes de conexión pueden ser adaptadas manualmente pudiéndose excluir contactos entre puntos cercanos o incluir relaciones entre puntos lejanos, siguiendo criterios aceptables para el propósito del estudio.

a) Triangulación de Delaunay



b) Gráfico de Gabriel



c) Gráfico de Vecindad Relativa



d) Gráfico Esfera de Influencia



**Figura 1. (a) Triangulación de Delaunay; (b) Gráfico de Gabriel ; (c) Gráfico de Vecindad Relativa; (d) Gráfico Esfera de Influencia.**

Para calcular el índice de Moran se mide la variable  $X$  en un sitio, llamémoslo  $i$ -ésimo sitio y se compara su valor con el valor promedio de la variable en las localizaciones de su vecindario. La expresión del índice es:

$$I = \frac{N \sum_i \sum_j W_{i,j} (X_i - \bar{X})(X_j - \bar{X})}{(\sum_i \sum_j W_{i,j})(X_i - \bar{X})^2} \quad (1.1)$$

donde  $N$  es el número total de observaciones,  $X_i$  es el valor de la variable en una localización particular (posición  $i$ ) y  $X_j$  es el valor de la variable en otra localización (posición  $j$ ). El elemento  $W_{i,j}$  de la matriz de ponderaciones  $W$ , es el peso aplicado a la comparación de las

observaciones en la posición  $i$  y la posición  $j$  que puede ser entendido como un coeficiente de continuidad. Cuando se utilizan redes de conexión, la matriz  $W$  está compuesta por ceros y unos. Cuando la posición  $j$  es adyacente a la posición  $i$ , el término  $ij$  recibe un peso de 1 y si no, de 0. Otra posibilidad para construir la matriz  $W$  es relacionar los elementos con la distancia  $d_{ij}$  entre las posiciones de manera inversamente proporcional, es decir:  $W_{ij} = 1/d_{ij}$ .

El índice de Moran varía entre  $-1$  y  $1$  ya que es estandarizado por la variabilidad de  $X$ . Cuando la autocorrelación es alta, el coeficiente será alto. Un valor cercano a  $1$  indica una alta correlación positiva, mientras que valores cercanos a  $-1$  indican alta autocorrelación negativa.

El Índice  $C$  de Geary, es similar al índice de Moran, pero en su numerador no mide la interacción a través del producto cruzado de las desviaciones con respecto a la media, sino que expresa la magnitud de las desviaciones entre observaciones en las diferentes localizaciones. La expresión del índice de Geary es:

$$G = \frac{(N-1) \sum_i \sum_j W_{i,j} (X_i - X_j)^2}{2(\sum_i \sum_j W_{i,j})(X_i - \bar{X})^2} \quad (1.2)$$

El valor índice de Geary se encuentra en el intervalo  $[0,2]$ . Si no hay autocorrelación espacial, el valor esperado de  $C$  es  $1$ . Valores del índice entre uno y dos indican autocorrelación espacial negativa. Este índice se relaciona inversamente con el índice de Moran.

La significancia de los valores del índice de Moran observado puede ser probada a través del procedimiento de simulación de Monte-Carlo mediante el cual las ubicaciones son

permutadas para obtener una distribución del coeficiente de Moran bajo la hipótesis nula de que hay una distribución aleatoria es decir, no existe una correlación espacial entre las observaciones. Si el valor observado del coeficiente de Moran es mayor que el esperado bajo simulación, entonces existe un agrupamiento o una estructura de los datos mediada por la distancia a la cual se encuentran valores más cercanos entre sí son más parecidos (autocorrelación espacial positiva). Por el contrario, un valor significativo y negativo del coeficiente de Moran indica que valores cercanos espacialmente son más diferentes entre sí que los esperados por azar (autocorrelación espacial negativa).

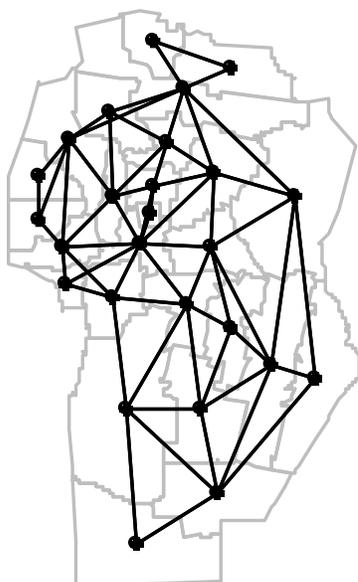
## **RESULTADOS Y DISCUSIÓN**

### **Cálculo de Índices de Moran y Geary**

El cálculo del índice o coeficiente de Moran de autocorrelación espacial en un espacio continuo requiere la definición de una matriz de ponderación espacial. La matriz de ponderación espacial puede ser representada en forma gráfica (como gráficos de vecindarios) donde los nodos corresponden a los sitios de los departamentos y los bordes a pesos espaciales no nulos. Para definir la matriz de ponderación espacial en esta tesis se consideró una matriz binaria. La especificación más simple de un vecindario es una matriz de conectividad  $C$ , en la que  $c_{ij} = 1$  si las unidades espaciales  $i$  y  $j$  son vecinos y  $c_{ij} = 0$ , dos departamentos se consideran vecinos si tienen un borde en común. La función que se utilizó para crear la red de conexión fue `poly2nb` de la librería “`spdep`” (Bivand, et. al., 2017). Se utilizó la función `nb2listw` del software R para determinar los sitios vecinos de cada sitio. La

matriz de pesos espaciales fue obtenida después de estandarizar por fila ( $style=W$ ). La red de conexión se presenta en la Figura 2.

El departamento que posee una mayor cantidad de vecinos es Santa María (9 links), mientras que entre los departamentos con un menor número de vecinos figuran, Capital (2 links), General Roca (2 links), Minas (2 links), Río Seco (2 links) y Sobremonste (2 links). En promedio fueron cinco los sitios vecinos de cada punto.



**Figura 2. Red de Conexión.**

Tabla 1: Índices de autocorrelación espacial actividades características del turismo (participación VARCT).

Variable	Índice de Moran		Índice de Geary	
	MI	p-valor	GI	p-valor
H	0,23	0,019	0,77	0,070
SV	0,27	0,013	0,68	0,017
R	0,07	0,173	0,85	0,142
ST	0,18	0,035	0,69	0,018
AV	0,14	0,088	0,79	0,055
SC	0,27	0,011	0,52	0,001
SD	0,61	0,001	0,42	0,001

H:Hoteles y similares, SV: Segundas viviendas, R: Restaurantes, ST: servicios de Transporte, SAT: Servicios Anexos de transporte, AV: Agencias de viaje, SC: Servicios culturales, SD: Servicios deportivos

La mayoría de las variables analizadas mostraron autocorrelación espacial positiva (Tabla 1).

Al menos con el índice de Moran global (MI) o con el de Geary (GI), a excepción de restaurantes (R) y agencias de viaje (AV), las autocorrelaciones espaciales fueron estadísticamente significativas ( $p < 0,05$ ). Para ambos índices, la variable con mayor autocorrelación global positiva fue la relacionada a la disponibilidad de servicios deportivos (SD) y servicios culturales (SC). En la tabla 2 se presentan los resultados del análisis descriptivo. Se muestran media, desvío estándar, coeficiente de variación (CV%) para cada una de las variables. La actividad que presenta menor variabilidad entre departamentos es R (CV=16%), el resto de las actividades presenta valores altos de CV. Las agencias de viaje, AV, constituyen la actividad característica mas variable entre departamentos en la Provincia (CV=114%).

Tabla 2: Descriptiva univariada para cada actividad. Participación en el valor agregado turístico (VARCT).

Variable	Media	DE	CV	Mín	Máx
H	0,09	0,08	90,80	0,00	0,33
SV	0,18	0,14	75,54	0,01	0,51
R	0,42	0,07	16,00	0,24	0,51
ST	0,16	0,07	42,12	0,03	0,34
AV	0,02	0,02	121,66	0,00	0,08
SC	0,03	0,03	114,62	0,00	0,12
SD	0,08	0,06	73,11	0,00	0,21

H:Hoteles y similares, SV: Segundas viviendas, R: Restaurantes, ST: servicios de Transporte, SAT: Servicios Anexos de transporte, AV: Agencias de viaje, SC: Servicios culturales, SD: Servicios deportivos

La participación de la actividad hotelera (H) en el total del agregado turístico se destaca en los departamentos de San Alberto (0,33), Punilla (0,28) y Calamuchita (0,27), esto se debe a que la oferta de alojamiento se encuentra concentrada en estos departamentos. En la provincia de Córdoba los valles de Punilla, Calamuchita y Traslasierra generan una gran afluencia de turistas, lo que ha propiciado el desarrollo de la actividad hotelera en la región, debemos mencionar que en el primero se destacan ciudades como Villa Carlos Paz, La Falda y Cosquín, en el segundo la ciudad de Villa General Belgrano, Santa Rosa de Calamuchita, y en el tercero la ciudad de Mina Clavero y Nono, por mencionar algunas de ellas, pues todos estos departamentos poseen puntos turísticos de reconocida afluencia. El departamento Capital también posee una importante oferta hotelera, sin embargo el peso de la actividad sobre las otras es proporcionalmente menor. En este departamento actividades vinculadas al servicio de transporte (ST) y a las actividades culturales (SC) cobran mayor peso. En este punto es importante aclarar que el peso de la actividad de segundas viviendas (SV) debido a

la manera en se construye posee mucho peso en valor absoluto<sup>4</sup>, a los fines analíticos en este trabajo muestra aquellos departamentos con menor producto bruto, por ello los departamentos de Pocho y Minas presentan mayor peso.

Se observan diferencias entre departamentos en la participación de la mayoría de las actividades, Capital posee una estructura distinta al resto, en este departamento se destacan las actividades vinculadas al transporte, los servicios culturales y deportivos. Cuando se analiza la participación de las actividades a lo largo del período se aprecia una mayor estabilidad en las series.

---

<sup>4</sup> Esto se debe a la manera en que se estima la actividad en las cuentas provinciales (ya que el principal componente es el valor locativo de la vivienda, ya sea alquilada u ocupada por sus propios dueños)

## **CAPÍTULO 3. ANALISIS MULTIVARIADO**

### **INTRODUCCIÓN**

Existen interacciones subyacentes entre las diferentes actividades económicas. Por ello el análisis de las covariaciones o correlaciones entre las ramas características del turismo es un aspecto que debe ser considerado en los estudios económicos. No obstante, es importante remarcar que la estructura de covariación reflejada por un análisis multivariado clásico puede verse afectada por los patrones espaciales subyacentes en los datos. En el análisis clásico no se toman en cuenta las relaciones espaciales en el cómputo de los indicadores o variables sintéticas que describen la estructura de distancias multivariadas entre los casos en estudio; las técnicas de ordenamiento, como PCA, no fueron específicamente diseñadas para identificar estructuras espaciales. Los índices de autocorrelación espacial, como el de Moran (Moran, 1948) y el de Geary (Geary, 1954), fueron los primeros usados para medir y analizar el grado de dependencia entre observaciones de una misma variable ubicadas en diferentes posiciones de un contexto geográfico, pero al ser univariados, dificultan la interpretación de la variabilidad espacial conjunta entre las variables en estudio.

Las componentes principales (PC) son apropiadas para resumir variabilidad multidimensional, pero no están diseñadas para revelar patrones espaciales. El objetivo de este trabajo es utilizar una metodología que resuma la variabilidad y revele estructuras espaciales al mismo tiempo; existen hoy métodos que abarcan estos dos objetivos. Dray et al. (2008), proponen un método de análisis multivariado que incorpora la información espacial previo al análisis multivariado, el método es conocido como MULTISPATI-PCA. Éste se basa en el PCA pero incorpora la restricción dada por los datos espaciales mediante el cálculo del índice de Moran antes de obtener las PC. El objetivo es encontrar variables sintéticas independientes que optimicen el producto de la varianza total y el coeficiente de Moran como medida de autocorrelación espacial.

Para delimitar los vecindarios, MULTISPATI-PCA utiliza una matriz de pesos espaciales determinando cuáles y cuántas observaciones cercanas a cada área económica deben ser consideradas para el cálculo del índice de Moran. Este análisis permite estudiar las relaciones entre las variables considerando su estructura espacial. La técnica ha mostrado ser eficiente en estudios de ecología (Dray et al., 2008), en agricultura (Córdoba, 2014), y en este trabajo se prueba para variables económicas.

El objetivo del presente capítulo es comparar metodologías para el tratamiento de datos espaciales en el contexto multivariado PCA clásico y PCA con restricción espacial en su desempeño para el tratamiento de datos de las ramas características de la actividad turística en una región.

## **MATERIAL Y MÉTODOS**

### **Análisis de componentes principales clásico y con restricción espacial**

El análisis de componentes principales clásico (PCA), tiene como objetivo principal describir una matriz de datos, reducir el número de variables que explican las principales variaciones e identificar correlaciones entre las variables medidas para generar nuevas variables no correlacionadas (i.e de fácil interpretación), llamadas componentes principales. El PCA consiste en encontrar transformaciones ortogonales de las variables originales para conseguir un nuevo conjunto de variables que no estén correlacionadas. El análisis de componentes principales explica la variación total en los datos y para obtener una reducción efectiva de su dimensión es necesario que las variables estén correlacionadas. Las componentes son combinaciones lineales de las variables originales y se espera que solo unas pocas (las primeras) recojan la mayor parte de la variabilidad de los datos, obteniéndose una reducción de la dimensión en los mismos. En resumen, el análisis de componentes principales encuentra los pesos o ponderaciones para cada variable con el fin de construir combinaciones lineales de variables capaces de maximizar la varianza entre las áreas de muestreo. Las combinaciones lineales obtenidas (CPs) son ortogonales (independientes) y en conjunto explican toda la variabilidad de los datos originales. La primera componente (CP1) explica la mayor parte de la variación total en el conjunto de datos y la segunda (CP2), la mayor parte de la variabilidad remanente o no explicada por la CP1. Luego, el propósito fundamental de la técnica consiste en la reducción de la dimensión de los datos con el fin de simplificar el problema en estudio.

Los resultados del PCA se pueden visualizar en un gráfico denominado Biplot (Gabriel, 1971) el cual permite representar en un plano óptimo para el estudio de variabilidad, las diferencias entre sitios, la correlación entre variables y las variables que mejor explican las principales variaciones. La incorporación de la información geográfica o la característica espacial de los datos puede realizarse a posteriori del PCA mediante la asignación de los valores de las componentes a cada uno de los sitios georreferenciados o bien ajustando semivariogramas (Schabenberger y Pierce, 2002) a las CP. Una ventaja de la utilización de variables sintéticas para mapeo es que se colapsa la caracterización multidimensional de las observaciones, permitiendo la construcción de mapas sintéticos de variabilidad espacial. Esta técnica permite visualizar el patrón de la variabilidad espacial y explorar gráficamente la estructura espacial de las variables analizadas. También se puede estudiar la presencia de autocorrelación espacial en las PC utilizando estadísticos de autocorrelación univariados como el índice de Moran (Moran, 1948) o el de Geary (Geary, 1954). Estos índices son utilizados para medir y analizar el grado de dependencia entre observaciones en un contexto geográfico (Cliff y Ord, 1973). Los datos multivariados son generalmente registrados en una matriz  $X$  con  $n$  filas (observaciones) y  $p$  columnas (variables). El diagrama de dualidad provee un marco teórico que define la estructura de numerosos métodos de análisis multivariado usando tres matrices  $(X, Q, D)$ . La teoría del diagrama de dualidad incluye métodos estándar como el análisis de componentes principales (PCA) y la extensión de éste a datos espaciales. Se considera la matriz de datos  $X$   $n \times p$  (originales o transformados) como parte del triplete  $(X, Q, D)$ , donde  $Q_{p \times p}$  y  $D_{n \times n}$  usualmente matrices simétricas definidas positivas usadas como métricas de relaciones o distancias.

Para la realización del PCA restringido espacialmente, denominado MULTISAPTI-PCA, es necesario primero definir cómo la información espacial es introducida en el análisis. En

MULTISAPTI - PCA, la detección de la estructura espacial se realiza a través del índice de Moran (MI). Esta aproximación entonces requiere que los sitios vecinos sean definidos. Esto en general se consigue por la construcción de una red de conexión (también llamada gráficos de vecinos) la cual usa un criterio objetivo para definir qué entidades son vecinas y cuáles no. Existen diferentes opciones o alternativas metodológicas para definir los vecindarios que dependen de los diferentes tipos de muestreo presente en los datos (grilla regular, irregular o transectas) (Bivand, 2008). Para muestreos irregulares los métodos se basan en el gráfico de Gabriel (Gabriel y Sokal, 1969), la triangulación de Delaunay (Lee y Schachter, 1980), los vecinos más cercanos (Cover y Hart, 1967) y la distancia Euclidea entre otros.

Una vez que la red de conexión es definida, la información espacial es almacenada en una matriz de conexión binaria (en la cual si las unidades espaciales son vecinas asume el valor 1 o 0 en caso contrario), la cual es simétrica y sus filas y columnas corresponden a la misma entidad (como una matriz de distancias). Esta matriz de conectividad en general es escalada para obtener la matriz de pesos espaciales  $W$ . La matriz es una representación matemática de la disposición geográfica de los puntos en la región (Bivand, 2008). Los pesos espaciales reflejan a priori la ausencia presencia o intensidad de la relación espacial entre las ubicaciones de interés. Una vez que los pesos espaciales han sido definidos, el índice de autocorrelación MI es computado. El método MULTISPATI-PCA introduce una matriz de pesos espaciales estandarizada ( $W$ ) por fila mediante un el análisis del triplete estadístico  $(X, Q, D)$ . La matriz  $\tilde{X} = WX$  está compuesta por los promedios ponderados de los valores de los vecinos de acuerdo a la matriz de conexión espacial, es llamada matriz lagged. Las dos tablas  $X$  y  $\tilde{X}$  tienen las mismas columnas (variables) y las mismas filas (observaciones). El análisis MULTISPATI-PCA consiste en el análisis de este par de tablas  $(X$  y  $\tilde{X})$  mediante un análisis de coinerencia (Dray et al., 2003). Para establecer la significación estadística de la

estructura espacial de la tabla, puede usarse un procedimiento basado en permutación. El estadístico usado es igual a la traza ( $X^T DWXQ$ ). El p-valor es computado por comparación del valor observado a aquellos obtenidos por permutación de las filas de  $X$  (Córdoba et al., 2012). MULTISPATI-PCA maximiza el producto escalar entre una combinación lineal de las variables originales y una combinación lineal de variables lagged (Saby et al., 2009). La ventaja de MULTISPATI-PCA respecto al PCA es que las componentes principales espaciales del MULTISPATI-PCA (sPC) maximizan la autocorrelación espacial entre los sitios. Por lo tanto, las sPC del MULTISPATI-PCA muestran fuertes estructuras espaciales sobre los primeros pocos ejes (Arrouays et al., 2011). MULTISPATI-PCA puede producir valores negativos de los CPs (debido a la posible autocorrelación negativa en los datos).

## **RESULTADOS Y DISCUSIÓN**

La combinación lineal de cada autovector con el valor observado para dicha variable conforma la componente principal. Se denomina CP1a la componente principal con mayor autovalor, es decir, la CP que explique la mayor variabilidad de los datos, luego, se denominará CP2 a la siguiente componente que explique alta variabilidad, pero ésta explicará menos que la CP1 y más que la CP3. Las dos primeras CP (CP1+CP2) explican el 68% de la variabilidad total de los datos (Tabla 4).

Si se analiza la Figura 3, donde el gráfico Biplot permite visualizar simultáneamente las observaciones y las variables del archivo de datos, se puede observar que departamentos que poseen una proyección hacia la derecha de la componente principal 1 (CP1), tienen mayor peso de las actividades servicios de transporte de pasajeros (ST), de los servicios culturales

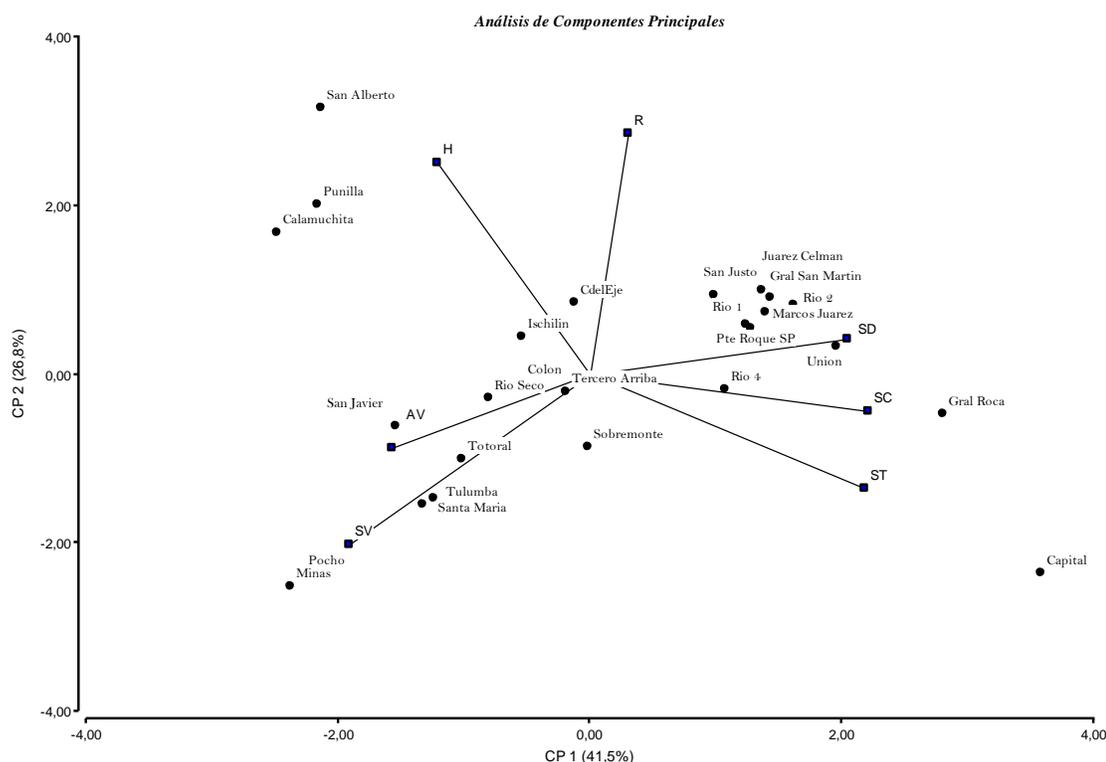
(SC) y de las actividades deportivas (SD), mientras que los departamentos que poseen su proyección hacia la izquierda de la CP1, reciben su aporte principal de la actividad segundas viviendas (SV). La variación en la CP2 fue fuertemente impulsada por Hoteles (H).

Se puede observar que departamentos como General Roca, Río Cuarto, Marcos Juárez, San Justo y Unión entre otros se encuentran hacia la derecha del gráfico, mientras que los departamentos San Javier, Río Seco, Totoral y Tulumba se encuentran hacia la izquierda del gráfico. A partir del análisis de las observaciones, podemos inferir que los departamentos que se encuentran hacia la derecha poseen una estructura diferente a aquellos departamentos que se encuentran hacia la izquierda. En el sentido ortogonal, sobre la CP2, se destacan la proyección San Alberto, Calamuchita y Punilla. El departamento Capital figura en el extremo derecho separándose del resto. Otra situación especial se visualiza con los departamentos de Pocho y Minas.

Los autovectores, coeficientes asociados a las variables son ponderadores o pesos, mientras más grande sea este coeficiente mayor inercia o “peso” tendrá esa variable para explicar la variabilidad de las observaciones. El signo del autovalor indica el sentido, si es negativo la proyección de la variable será hacia la izquierda de la CP1, mientras que si es positivo su proyección será hacia la derecha de la CP1. Se observa que las variables servicios culturales, servicios de transporte y actividades deportivas, poseen los coeficientes más altos (0,48, 0,47 y 0,44 respectivamente). Los vectores de esta variable, se encuentran hacia la derecha del gráfico Biplot (Figura 3), la proyección de estas variables sobre la CP1 se encuentran en el extremo derecho del eje y están cerca a las proyecciones de los departamentos como General Roca, Río Cuarto, Marcos Juárez, San Justo y Unión. De allí que podemos inferir que estos departamentos reciben su principal aporte de los servicios culturales, de los servicios de

transporte y de las actividades deportivas. Esto coincide con la distinción entre departamentos con mayor y menor valor de producto geográfico bruto regional (Anexo 5). Esto muestra que el peso de la actividad de transporte (ST) es alto en términos relativos con respecto al resto de las actividades características del turismo, dado que son departamentos con mayor actividad económica en general.

En el sentido Norte-Sur, sobre la CP2, se destacan la proyección de los departamentos San Alberto (Mina Clavero, Villa Cura Brochero, Nono, etc.), Calamuchita (Villa General Belgrano, Santa Rosa de Calamuchita, etc.) y Punilla (Villa Carlos Paz, La Falda y Cosquín, etc.), esto se debe a que en la CP2 la actividad hoteles y similares (H) posee uno de los coeficientes más altos (0,54). Mientras que los que tienen una proyección hacia arriba tienen mayor peso de la actividad segunda vivienda (SV) donde se destacan los departamentos de Pocho y Minas.



**Figura 3. Gráfico Biplot, Análisis de Componentes Principales obtenido a partir de la participación de las actividades (H, R, ST, AV, SC, SD, AV) en los departamentos de la provincia de Córdoba**

En las tablas 3 y 4 se presentan las varianzas y los coeficientes de autocorrelación de cada una de las CPs generadas a partir de MULTISPATI-PCA y PCA, respectivamente. MULTISPATI-PCA maximiza el producto entre la varianza espacial y la autocorrelación mientras que PCA maximiza la varianza total. Los resultados muestran que con MULTISPATI-PCA se explica una menor proporción de la varianza acumulada en los dos primeros ejes o CPs, en esta aplicación, respecto de PCA (2,643 vs. 2,903 para el eje 1 y 1,258 vs. 1,876 para el eje 2).

No obstante, los valores del índice de Moran calculados para las dos primeras CP sugieren que la estimación de autocorrelación aumentó cuando se usó MULTISPATI-PCA respecto de la contenida en las CPs del PCA (0,543 vs. 0,436 para el eje 1; 0,429 vs. 0,087 para el eje 2). Este resultado sugiere que existirá una visualización mejor de la variabilidad espacial si se usa el análisis restringido espacialmente. Por el contrario, a nivel de las CPs 3 este comportamiento fue inverso. Luego, se puede decir que no siempre la varianza disminuye a causa de la estimación de autocorrelación y que la maximización de la variabilidad espacial depende de las características de la autocorrelación presente.

Tabla 3: Autovalores, varianza espacial e índices de Moran de las componentes principales generadas a partir de MULTISPATI-PCA (sobre la participación promedio de las actividades)

Eje	Autovalores	Varianza Espacial	Proporción (varianza espacial)	Proporción Acumulada	Índice de Moran
-----	-------------	-------------------	--------------------------------	----------------------	-----------------

1	1,436	2,643	0,378	0,378	0,543
2	0,539	1,258	0,180	0,557	0,429
3	0,105	1,282	0,183	0,740	0,082
4	0,005	0,094	0,013	0,754	0,055
5	-0,013	0,252	0,036	0,790	-0,051
6	-0,106	0,452	0,065	0,854	-0,236
7	-0,186	1,019	0,146	1,000	-0,183

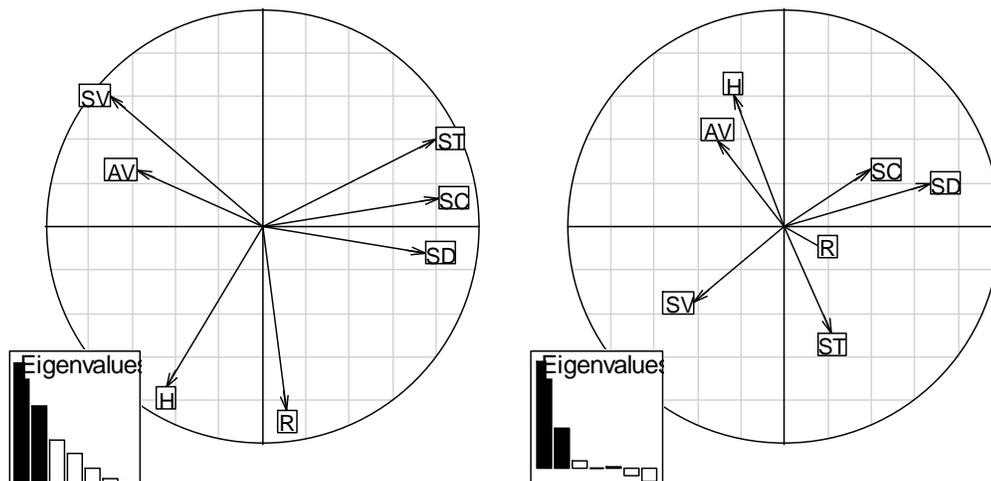
Tabla 4: Autovalores e índices de Moran de las componentes principales generadas a partir del PCA (sobre la participación promedio de las actividades)

Eje	Autovalores	Proporción	Proporción Acumulada	Indice de Moran
1	2,903	0,415	0,415	0,436
2	1,876	0,268	0,683	0,087
3	1,025	0,146	0,829	0,384
4	0,719	0,103	0,932	0,013
5	0,343	0,049	0,981	-0,161
6	0,118	0,017	0,998	0,046

7	0,016	0,002	1,000	-0,041
---	-------	-------	-------	--------

En la figura 4 se muestra la representación gráfica de las dos primeras componentes del PCA (izquierda) y también de las dos primeras componentes del MULTISPATI-PCA (derecha) así como los autovalores asociados a cada uno de esos ejes (gráfico de barras).

Las barras de color negro corresponden a la cantidad de ejes seleccionados que fueron utilizados para la representación gráfica e interpretación de la variabilidad subyacente, en este caso las dos primeras CPs. Las barras negras junto con las blancas indican la cantidad de ejes que se obtuvieron en el análisis. La altura de cada barra representa la proporción de la variabilidad total reflejada por cada CP. Así, para MULTISPATI-PCA analizar las dos primeras CPs aporta suficiente información para el análisis, ya que las siguientes CP no representan una contribución importante (barras blancas).



**Figura 4. Representación gráfica de los dos primeros ejes del PCA (izquierda) y MULTISPATI-PCA (derecha) que muestra la correlación entre las variables y entre éstas y las componentes principales.**

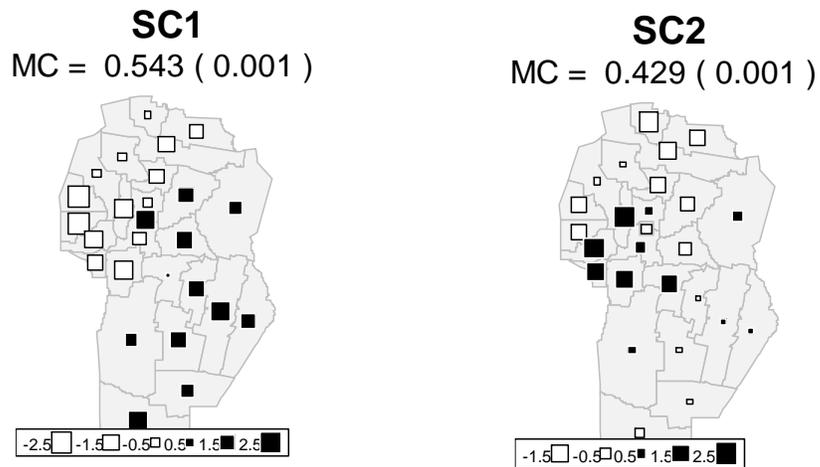
La variación en el eje 1, para ambos métodos, fue fuertemente impulsada por servicios culturales (SC) y servicios deportivos (SD) (variable de mayor proyección sobre el eje de las abscisas). No obstante, se visualizaron diferencias entre ambos análisis a nivel de la estructura de co-variación entre las dos mediciones de servicios de transporte (ST).

Las relaciones entre servicios culturales (SC) y servicios deportivos (SD) no cambiaron de manera importante como para impactar las conclusiones como sí sucedió con las correlaciones entre hoteles (H) y agencias de viaje (AV). El reposicionamiento de servicios de restaurantes (R) sobre el primer eje de MULTISPATI-PCA produjo un cambio de ponderación de las variables sobre la CP2 (eje de ordenadas) haciendo que quede más correlacionada agencia de viajes (AV) con hoteles (H).

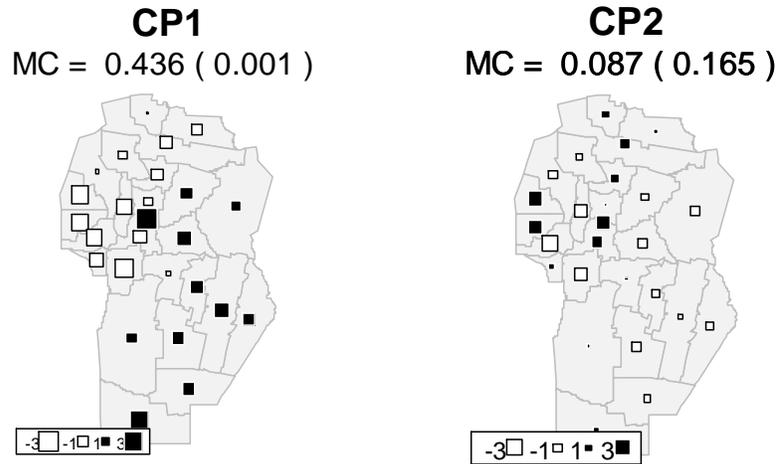
La correlación, entre agencia de viajes (AV) y hoteles (H), fue rescatada en menor medida por el PCA según muestra el ángulo entre los dos vectores de la variable y se hizo más notoria con MULTISPATI-PCA. Las relaciones entre SC y SD no cambiaron de manera importante como para impactar las conclusiones.

El reposicionamiento de R sobre el primer eje de MULTISPATI-PCA para indicar la presencia de correlación negativa entre H y ST, produjo un cambio de ponderación de las variables sobre la CP2 (eje de ordenadas) haciendo que ésta quede más correlacionada con hoteles (H), agencias de viaje (AV), permitiendo así analizar la variabilidad del desde otra dimensión distinta.

Las figuras 5 y 6 muestran los mapas logrados mediante la asignación cada uno de los sitios georreferenciados de los valores de la CP1 y CP2 del MULTISPATI-PCA y PCA, respectivamente.



**Figura 5. Variabilidad espacial multivariada en contribución de las ramas características del turismo al VARCT. Los valores de variables sintéticas son logrados con la CP1 y CP2 del MULTISPATI-PCA.**



**Figura 6. Mapas de variabilidad espacial multivariada logrados con la CP1 y CP2 de PCA.**

Los mapas de variabilidad espacial multivariada logrados con la CP1 de ambos métodos resultaron similares. Este resultado se justifica en el hecho de que en ambos análisis la CP1 se encuentra altamente correlacionada positivamente con servicios culturales (SC) y servicios deportivos (SD) y negativamente con segundas viviendas (SV).

Por otra parte, en los mapas construidos a partir de la CP2 del PCA se representa principalmente la variación de la hoteles (H) mientras que en el mapa logrado con la CP2 del MULTISPATI-PCA la variabilidad si bien es debida principalmente a la hoteles (Figura 6) lo hace en mayor medida junto con otras actividades como agencia de viajes (AV) y contrasta con aquellas cuya variación está vinculada a servicios de transporte (ST); los cuadrados negros representan aquí valores de mayor hoteles (H) que los cuadrados blancos según pone de manifiesto la estructura de correlación de las variables (Figura 6).

## **CAPITULO 4. CLASIFICACIÓN NO SUPERVISADA ESPACIAL MULTIVARIADA**

### **INTRODUCCION**

En este capítulo se utilizan las componentes principales espaciales obtenidas con MULTISPATI-PCA como input del método de conglomerados denominado fuzzy k-means para la identificación de áreas o regiones provinciales mediante el agrupamiento de departamentos. El método, denominado KM-sPC (Córdoba et al., 2012), se diferencia de los propuestos por Oliver y Webster (1989) y Bourgault et al. (1992), ya que la autocorrelación espacial es tomada en cuenta antes de la obtención de la matriz de distancias para aplicar el

algoritmo fuzzy k-means. La hipótesis que subyace a la metodología usada es que la incorporación de la autocorrelación espacial a través del PCA espacial aplicado sobre las actividades características del turismo, producirá áreas que contienen menos “ruido”, i.e. conjunto de departamentos más homogéneos. Si bien este procedimiento ha sido usado a escala fina con datos continuos, su aplicación a datos que representan departamentos a una escala regional como la que operan los indicadores de turismo, es novedosa.

En el presente capítulo se ilustra y evalúa la metodología usada como herramienta para la identificación de áreas en la provincia con características similares respecto a la contribución de los rubros de la actividad turística. Esto resulta útil dado que existe interés por parte de los decisores de política pública en identificar regiones, con el fin de establecer estrategias de desarrollo regional, concretamente en el ámbito del turismo. Esto se ha puesto de manifiesto en la ley provincial 10312 de regionalización turística.

Con el fin de complementar los resultados obtenidos, se realiza un mapeo del territorio provincial en función de la participación total del valor agregado de las ramas características (VARCT) sobre el producto bruto regional (PBR). Por último se analiza la dispersión en la participación promedio de las actividades a lo largo del período analizado en los departamentos, a través de una estratificación.

## **MATERIAL Y MÉTODOS**

### **Análisis de clúster fuzzy k-means**

Existen tres matrices primarias que participan en el análisis fuzzy k-means. La primera de ellas es la matriz de datos a clasificar ( $X$ ). La matriz  $X$  se forma a partir de  $n$  observaciones multivariadas, cada una con  $p$  variables. La segunda matriz ( $V$ ), consta de los  $k$  centroides

correspondientes a cada clúster localizado en el espacio de los atributos definido por las  $p$  variables. La tercera, es la matriz de pertenencia difusa ( $U$ ), que contiene los valores o asignaciones parciales de cada una de las  $n$  observaciones en cada uno de los  $k$  clúster o conglomerados, limitada por la restricción que se muestra en (4.1), debiéndose cumplir ésta para cualquier  $i = 1, \dots, n$

$$\mu_{i,j} \in [0,1] \quad \forall_{i,j} \quad \text{y} \quad \sum_{j=1}^k \mu_{i,j} = 1, \forall_j \quad (4.1)$$

La partición difusa óptima de los datos es la que minimiza la función objetivo  $J_m$  igual a la suma ponderada de las distancias cuadráticas entre las observaciones y los centroides de cada clúster:

$$J_m(U, v) = \sum_{i=1}^n \sum_{j=1}^k (\mu_{ij})^m (d_{ij})^2 \quad (4.2)$$

donde  $m$  es el coeficiente de ponderación difuso ( $1 \leq m < \infty$ ) cuya función es controlar el grado de solapamiento que se establece entre los clusters y  $(d_{ij})^2$  es el cuadrado de la distancia en el espacio de los atributos entre el punto  $i$  y la clase centroide  $j$ , que se puede calcular de la siguiente manera:

$$(d_{ij})^2 = \|x_i - v_j\|^2 = (x_i - v_j)^T \mathbf{A} (x_i - v_j) \quad (4.3)$$

donde  $x_i$  es la observación  $i$ -ésima de la matriz de datos  $\mathbf{X}$ ,  $v_j$  el centroide del clúster  $j$ , y  $\mathbf{A}$  es la matriz de pesos definida positiva ( $p \times p$ ) que induce norma por el producto interno. La matriz de ponderación  $\mathbf{A}$  define un procedimiento de normalización de la distancia. El resultado representa la distancia entre dos puntos o vectores en un espacio vectorial lineal.

Fridgen et al. (2004) aconsejan tomar  $\mathbf{A} = \mathbf{I}$  (matriz identidad), únicamente cuando las variables sean estadísticamente independientes y presenten la misma varianza. La métrica obtenida es, por tanto, la distancia Euclídea entre la observación  $i$ -ésima y el centroide. En el caso de que las varianzas de las variables sean distintas, es recomendable estandarizar las variables mediante la utilización de una matriz diagonal ( $\mathbf{A} = \mathbf{D}$ ) cuyos términos sean precisamente las varianzas de las variables en estudio o bien trabajar con las variables previamente estandarizadas. Finalmente, la tercera posibilidad es tomar  $\mathbf{A} = \mathbf{S}$  (matriz de varianzas y covarianzas de  $\mathbf{X}$ ), con lo que la métrica resultante es la distancia de Mahalanobis. Se utiliza esta distancia cuando las variables de clasificación no solo muestran varianzas distintas sino que están correlacionadas entre sí. Mientras que el algoritmo iterativo fuzzy k-means siempre converge a un mínimo local de  $J_m$  a partir de un determinado  $\mathbf{U}$  inicial, una aleatorización diferente de  $\mathbf{U}$  podría dar lugar a un mínimo local diferente (Xie y Beni; 1991; Bezdek, 1981).

El algoritmo difuso fuzzy k-means utiliza un proceso iterativo para la obtención del par ( $\mathbf{U}$ ,  $\mathbf{V}$ ) que hace óptima la partición difusa de los datos  $\mathbf{X}$ . La estructura del algoritmo (Bezdek, 1981) se muestra a continuación.

1. Se elige el número de grupos o clusters  $k$ , con  $2 \leq k \leq n$ .
2. Se fija el valor del exponente difuso  $m$ , con  $1 < m < \infty$ .
3. Se selecciona una medida apropiada de similaridad o distancia  $d_{ij}$ .
4. Se selecciona el valor del criterio de convergencia (finalización) del algoritmo.
5. Se selecciona el número máximo de iteraciones,  $l_{\max}$ .

6. Se inicializa la matriz  $U^0$  con valores aleatorios y según la restricción especificada en (4.1).

7. En las sucesivas iteraciones  $l=1,2,3,\dots$ , se recalculaba  $V^l$  (matriz de centroides) a partir de  $U^{(l-1)}$ , utilizando la siguiente expresión:

$$v_{i,j} = \frac{\sum_{j=1}^n (u_{ij})^m x_j}{\sum_{j=1}^n (u_{ij})^m} \quad (4.4)$$

8. La minimización de (4.2) mediante el método iterativo de Picard hace posible el cálculo (actualización) de  $U^l$  a partir de la matriz actualizada  $V^l$ , según:

$$u_{ij} = \left[ \sum_{j=1}^k \left( \frac{d_{ij}}{d_{lj}} \right)^{2/m-1} \right]^{-1}, \quad i = 1, \dots, n \quad j = 1, \dots, k \quad (4.5)$$

9. Se interrumpe el algoritmo cuando se alcanzaba el número máximo de iteraciones ( $l_{\max}$ ), o cuando  $\|U^l - U^{(l-1)}\| \leq \epsilon$ ; en otro caso, se volvía al paso 7.

10. Se computaban finalmente los índices para validar los clúster.

Para evaluar la clasificación conseguida con un determinado número de grupos, existen diferentes índices como el coeficiente de partición (o fuzziness performance index-FPI, Bezdek, 1981), el índice de entropía de la clasificación (o normalized classification entropy-NCE, Bezdek, 1981), el índice de Xie-Beni (Xie y Beni, 1991) y el de Fukuyama-Sugeno (Fukuyama y Sugeno, 1989), entre otros.

El coeficiente de partición (CP) mide el grado de separación o solapamiento (grado de fuzziness) entre los grupos formados. Se considera que mientras menos difusa es la partición, mejor es la clasificación. Su cálculo resulta de la siguiente expresión:

$$CP(U) = \frac{\sum_{i=1}^n \sum_{j=1}^k u_{ij}^2}{n} \quad (4.6)$$

En este caso el óptimo se da al maximizar CP y equivale a una clasificación en la que cada observación pertenece a un único clúster. El mínimo se da cuando cada observación pertenece, con el mismo grado, a cada clúster (mayor incertidumbre).

La entropía de la partición (EP) estima la cantidad de desorganización creada por la partición difusa de la matriz de datos X con un número específico de clusters. Para este índice los valores de EP próximos a 0 son indicativos de una mejor clasificación, es decir, con mayor grado de organización.

$$EP(U) = \frac{\sum_{i=1}^n \sum_{j=1}^k u_{ij} \log(u_{ij})}{n} \quad (4.7)$$

En el índice de Xie-Beni (XB) se incorpora a v y X. Este índice prefiere particiones cuya distancia intra-clúster sea mínima y la distancia inter-clúster máxima:

$$XB(U, v, X) = \frac{\sum_{i=1}^n \sum_{j=1}^k \|x_i - v_j\|^2}{n \left( \min_{j \neq j'} \|v_j - v_{j'}\|^2 \right)} \quad (4.8)$$

El índice XB es considerado como una medida de compactación dentro del clúster. Un valor pequeño de XB, representa un agrupamiento donde los grupos son compactos y bien separados. Por consiguiente, la mejor partición se obtiene mediante la minimización de XB.

El índice Fukuyama-Sugeno (FS) está formado por la diferencia entre la medida de compactación y la medida de separación entre los centroides de los grupos y la media de todos los centroides. El mínimo de FS corresponde a una partición difusa con clases compactas y bien separadas.

$$FS(U, v, X) = \sum_{i=1}^n \sum_{j=1}^k u_{ij}^m \left( \|x_i - v_j\|^2 - \|v_j - \bar{v}\|^2 \right) \quad (4.9)$$

Es importante considerar que para un conjunto de datos, los índices no son necesariamente consistentes entre sí e incluso pueden contradecirse, es decir, sugerir diferentes números de clúster como partición óptima. Una solución es obtener un único índice que resuma los anteriores (Galarza et al., 2013). En los índices mencionados excepto para el coeficiente de partición (CP), un valor menor del índice implica una mejor clasificación. Por ello, se recalcula CP como  $CP^* = 1/CP$  para que el valor mínimo en todos los índices represente la mejor elección. Adicionalmente, se normalizan los valores de los índices entre 0 y 1 dividiendo cada valor por el máximo alcanzado por el índice en las diferentes clasificaciones. Luego, se calcula la distancia Euclídea para cada clasificación utilizando los valores de los índices normalizados y se selecciona la clasificación con menor valor de este nuevo índice (Córdoba, 2014).

Los índices aportan información sobre cual podrá ser la clasificación óptima. La selección final de la cantidad de clúster debe seguir una relación de compromiso entre lo sugerido por los índices y el criterio económico.

## **Clasificación de áreas**

Siguiendo a Córdoba (2012) la propuesta para la delimitación de áreas se basa en los análisis MULTISPATI-PCA y clúster fuzzy k-means (Dray et al., 2008; Fridgen et al., 2004; Bezdek, 1981). Además de las variables registradas, la base de datos debe incluir las coordenadas espaciales de cada punto de datos. Las coordenadas geográficas son generalmente convertidas a coordenadas cartesianas. Esto permite que las distancias se muestren como

absolutas (metros) en lugar de distancias relativas (grados). La etapa de pre-procesamiento se puede realizar utilizando cualquier Sistemas de Información Geográfica (SIG). El próximo paso del algoritmo es aplicar MULTISPATI-PCA a las actividades características del turismo (H, ST, SC, AV, SD, R y SV) y obtener las componentes principales espaciales (sPC). El análisis puede ser realizado con los paquetes "ade4" (función multispati, Chessel et al., 2004) y "spdep" (Bivand et al., 2013) del software R (R CoreTeam, 2017). En este caso se definió una matriz binaria donde dos departamentos son considerados vecinos si tienen un borde en común.

La función que se utilizó para crear la red de conexión fue poly2nb del paquete "spdep". Usando el paquete "ade4" (Chessel et al., 2004), el software R devuelve un objeto de clase multispati, que contiene varios elementos, entre ellos las sPC. Un conjunto reducido de estas variables sintéticas resultantes, que explican una gran cantidad de la variación total ( $\geq 70\%$ ), son posteriormente usadas como input del análisis de cluster fuzzy k-means.

Finalmente se realiza la aplicación del análisis de clúster fuzzy k-means usando las componentes principales espaciales como variables en las que se basa la clasificación. Así la matriz de datos utilizada en el análisis fuzzy k-means incluye las n observaciones cada una con  $a < p$  componentes principales espaciales. La distancia Euclídea se utiliza como medida de similitud en la función de optimización del fuzzy k-means, ya que las componentes principales son independientes y se estandarizan cuando se realiza el análisis MULTISPATI-PCA, por lo tanto sus varianzas no difieren. El exponente difuso se fija en el valor convencional de 1,30 (Odeh et al., 1992). Alternativamente, el algoritmo fuzzy k-means puede ser implementado desde otros software como FuzMe (Minasny y McBratney, 2002) que además de trabajar con la distancia Euclídea permiten utilizar las distancias de

Mahalanobis o Diagonal que son apropiadas cuando las variables no son estadísticamente independientes y/o presentan varianzas diferentes. El coeficiente de partición (conocido también como fuzziness performance index, FPI) y la entropía de clasificación normalizada (normalized classification entropy, NCE) (Odeh et al., 1992) se pueden utilizar para determinar el número óptimo de clusters. Este se obtiene cuando ambos índices se reducen al mínimo, lo que representa el menor solape entre los grupos (FPI) o el mayor grado de organización (NCE) como consecuencia del proceso de agrupación de los datos (Fridgen et al., 2004). Para ejecutar el nuevo algoritmo denominado KM-sPC, se utilizaron los scripts desarrollados por Córdoba et al. (2012) en el software R con adaptaciones necesarias para incorporar la actividades a escala departamental (R CoreTeam, 2017).

## **RESULTADOS Y DISCUSIÓN**

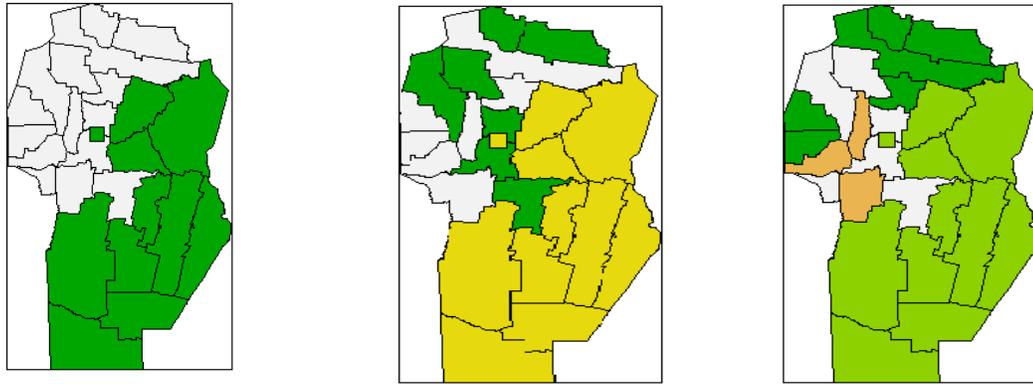
### **Clasificación de departamentos vía clúster fuzzy k-means**

En el mapa de la Figura 7 se presentan los mapas resultantes de la clasificación de departamentos realizada a partir de la participación de las actividades características del turismo (H, ST, SC, AV, AD, R y SV) y mediante el análisis de cluster fuzzy k-means. Se observa que la clasificación con tres y cuatro clusters producen mayor fragmentación de las clases delimitadas.

**a)**

**b)**

**c)**



**Figura 7. Mapa con clases delimitadas: a) dos (izquierda), b) tres (centro) y c) cuatro (derecha) clases.**

Los índices utilizados para la selección número óptimo de clases no fueron coincidentes en la indicación del número de cluster a retener (Tabla 5). El Coeficiente de Partición que representa el menor solape entre los grupos y el índice de Fucuyama-Sugeno sugieren que la partición óptima se encuentra con cuatro clases, mientras que para el índice de Entropía que presenta el mayor grado de organización como consecuencia del proceso de agrupación de los datos de la clasificación y el de índice de Xie-Beni el óptimo sugerido, es de tres clases. Un índice resumen que contiene información de cada uno de los índices previamente calculados, indicó que la partición óptima es de tres clases.

Tabla 5: Selección del número de clases de la partición de departamentos

Índice	2 clases	3 clases	4 clases
Coeficiente de Partición	1,053	1,039	<b>1,053</b>
Entropía de Partición	0,090	<b>0,058</b>	0,106
Xie - Beni	0,013	<b>0,004</b>	0,012
Fukuyama - Sugeno	-50,531	-63,157	<b>-95,585</b>
Indice Resumen	1,929	<b>1,710</b>	2,526



Para cada índice se indica el número óptimo de clases sugerido colocando en negrita el mejor valor del índice

Teniendo en cuenta las estructuras económicas subyacentes en la industria turística, cuatro clases (mapa de la derecha, c) reflejan las particularidades y similitudes entre los departamentos, y aporta un elemento adicional para el análisis de desarrollo regional.

Por un lado se tiene un área (clase 1, color marrón claro) conformada por los departamentos de Punilla, Calamuchita y San Alberto. Luego otra área (clase 2, verde claro) conformada por los departamentos del sureste de la provincia y el departamento Capital, una tercera área (clase 3, color blanco) compuesta por los departamentos de San Javier, Ischilín, Cruz del Eje, Colón y Santa María y por último una cuarta clase (verde oscuro) donde se encuentran los departamentos del norte (Sobremonte, Río Seco, Tulumba y Totoral) y del oeste (Minas y Pocho) de la provincia.

Los departamentos que conforman la clase 1 presentan mayor desarrollo turístico esto se observa tanto por el lado de la demanda, la región recibió la mayor afluencia de turistas, 69,8% para el período 2014/2015 (según lo informado por la agencia Córdoba Turismo); como por el lado de la oferta, son los departamentos con mayor participación del VARCT en su producto bruto regional (entre el 16,6% y el 29,6%, Figura 8). En estos departamentos en los últimos tiempos, se ha producido un crecimiento en la oferta de alojamiento (sobre todo en la tipología cabañas y otras no convencionales); al igual que en el sector gastronómico, lo cual implica un impacto en la generación de empleo en los sectores específicos y en los complementarios. La generación de empleo impacta directamente en las economías locales.

La clase 2, está conformada por los departamentos del sureste de la provincia y el departamento Capital, departamentos con estructuras orientadas al sector agropecuario e industrial, importante desarrollo de las actividades del transporte, culturales y deportivas.

Luego, la clase 3, con departamentos con un desarrollo importante del sector servicios y por último la clase 4 compuesta por los departamentos de Pocho, Minas, Río Seco, Sobremonte, Totoral y Tulumba departamentos más pobres, con menor producto bruto regional, donde es necesario un fortalecimiento de las economías regionales, y su posible inserción en la actividad turística por compatibilidad y complementación. El desarrollo del turismo rural, aparece como una alternativa que posibilita el desarrollo de la actividad.

### **Participación del valor agregado turístico en el producto bruto regional**

En la introducción a este capítulo se mencionó el interés de complementar los resultados obtenidos con otros datos disponibles relacionados a la regionalización. En este apartado se calcula la participación de las ramas características del turismo total (VARCT) en el producto bruto regional. Hasta ahora se había trabajado con la participación de cada actividad (H, SC, SD, AV, R, etc.) en el total de agregado turístico, esto permitió realizar agrupamientos entre los diferentes departamentos de acuerdo a su composición interna.

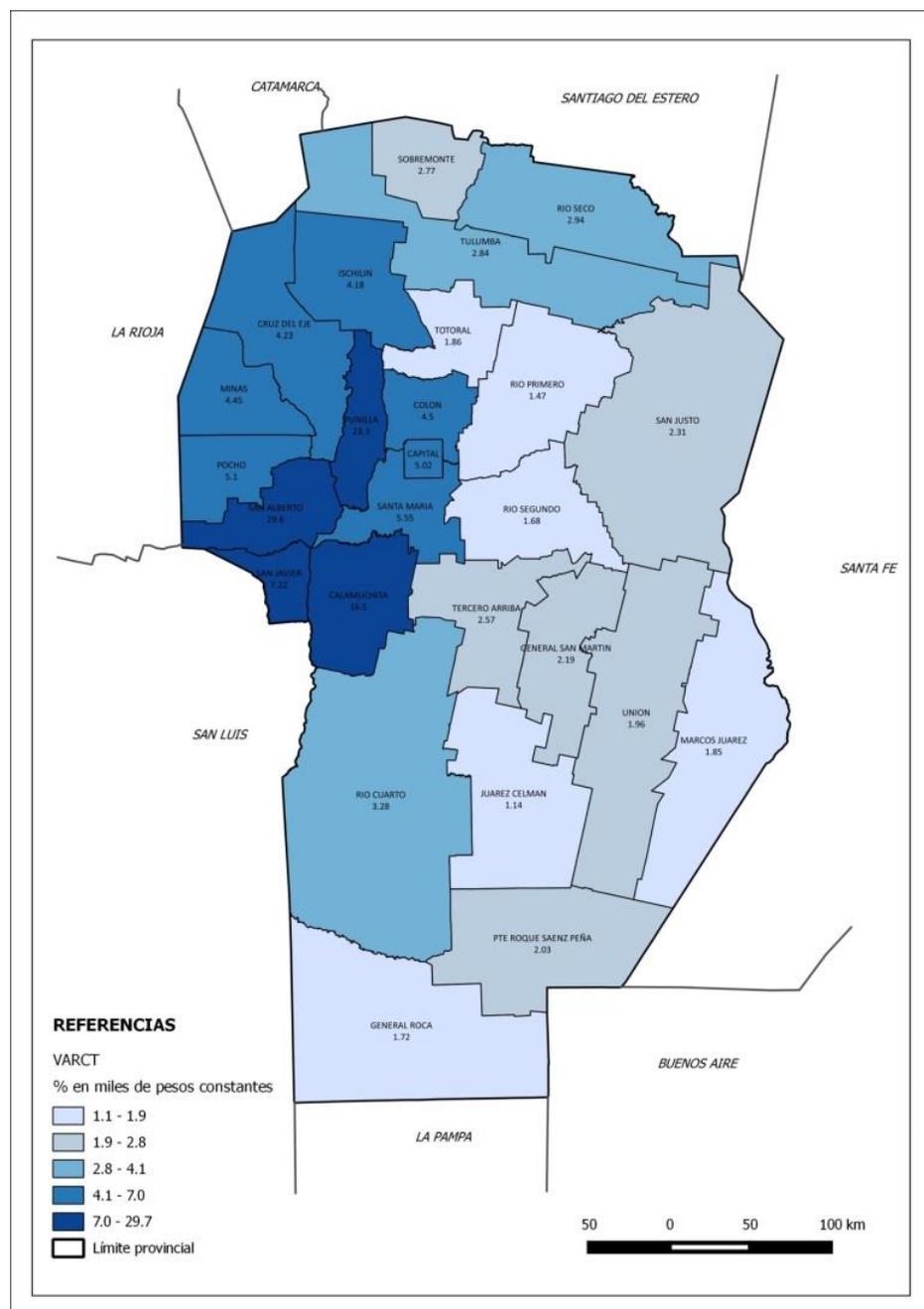
En la figura 8 se muestra la participación del VARCT (%) en el total del producto bruto en los diferentes departamentos. Para el cálculo se promedió la participación a lo largo de todo el período (2001-2014) y se construyeron cinco intervalos en función de la participación promedio. En el intervalo con mayor participación (de 7,0% a 29,7%), se encuentran los departamentos de San Alberto (29,6%), Punilla (23,3%), Calamuchita (16,6%) y San Javier

(7,2%). En estos departamentos, como se mencionó en párrafos anteriores se encuentran las principales puntos turísticos de la provincia de Córdoba (Villa Carlos Paz, Villa General Belgrano, Mina Clavero, entre otras) lo cual explica un mayor desarrollo de las actividades vinculadas al turismo, la actividad de hoteles (H) se destaca en estos departamentos, de acuerdo a datos sobre hoteles registrados de la Agencia Córdoba Turismo (ACT) en el año 2014 el departamento de Punilla concentró el 42,0% de las plazas disponibles, Calamuchita el 14,0% y San Alberto el 12,0%. El segundo intervalo (4,1% a 7,0%) se encuentran los departamentos de Santa María (5,6%), Pocho (5,1%), Capital (5,0%), Colón (4,5%), Minas (4,5%), Cruz del Eje (4,3%) e Ischillín (4,2%). En el tercer intervalo (2,8% a 4,1%) se encuentran Río Cuarto (3,3%), Río Seco (2,9%) y Tulumba (2,8%). En el cuarto intervalo (1,9% a 2,8%) se encuentran Sobremonte (2,8%), Tercero Arriba (2,6%), San Justo (2,3%), General San Martín (2,2%), Presidente Roque Saenz Peña (2,0%) y Unión (1,9%). Por último en el intervalo de menor participación (1,3% a 1,9%) se destacan los departamentos Totoral (1,9%), Marcos Juárez (1,8%), General Roca (1,7%), Río Segundo (1,7%), Río Primero (1,4%) y Juárez Celman (1,1%).

Los departamentos que tienen una menor participación del VARCT, coinciden en su mayoría con los departamentos provinciales que tienen una estructura productiva orientada hacia los sectores productores de bienes (agricultura y ganadería, industria, construcción, electricidad, gas y agua y minería) con una menor participación de los sectores de servicios. Mientras lo contrario ocurre con aquellos departamentos que cuentan con una mayor participación del sector servicios.

Según el informe de la Dirección de estadísticas económicas (DEE) para el año 2014, la agricultura, ganadería, caza y silvicultura fue la actividad con mayor peso relativo en el

Producto Bruto Regional de doce departamentos provinciales. Le siguió la industria manufacturera, actividad más importante en seis departamentos. Por detrás estuvo hoteles y restaurantes con mayor participación en el PBR de tres departamentos, al igual que el transporte, almacenamiento y comunicaciones.



**Figura 8. Mapa participación promedio del VARCT en el producto bruto regional**

### **Clasificación de departamentos vía clúster fuzzy k-means en función de la variabilidad de las variables de entrada**

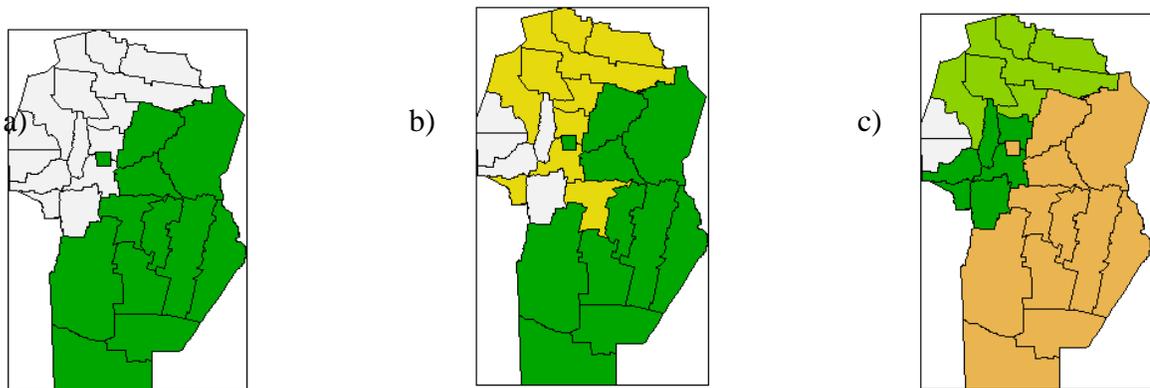
A los fines de completar el análisis realizado en los párrafos anteriores, donde se trabajó con la participación promedio de las actividades, en este apartado se trabaja con la dispersión en la participación, interesa conocer la estratificación que surge al analizar como ha cambiado la participación promedio de las actividades a lo largo del período analizado en los departamentos. Se realizó un PCA espacial, reteniéndose las tres primeras componentes principales espaciales (tabla 6), que explican una gran cantidad de la variación total (71,3%), para ello se utilizó la librería “ade4” (Chessel et al., 2004). Finalmente se realiza la aplicación del análisis de clúster fuzzy k-means usando las componentes principales espaciales como variables en las que se basa la clasificación.

Tabla 6: Autovalores, varianza espacial e índices de Moran de las componentes principales generadas a partir de MULTISPATI-PCA

Eje	Autovalores	Varianza Espacial	Proporción (varianza espacial)	Proporción Acumulada	Indice de Moran
1	1,870	3,168	0,453	0,453	0,590
2	0,676	1,541	0,220	0,673	0,438
3	0,034	0,285	0,041	0,713	0,121
4	0,019	0,262	0,037	0,751	0,072

5	-0,056	0,539	0,077	0,828	-0,104
6	-0,070	0,550	0,079	0,906	-0,128
7	-0,173	0,655	0,094	1,000	-0,264

En la Fig. 9 se presentan los mapas resultantes de la nueva clasificación de departamentos realizada a partir del desvío en la participación de las actividades (H, ST, SC, AV, AD, R y SV) y mediante el análisis de cluster fuzzy k-means. Se observa que la clasificación con tres y cuatro clusters producen mayor fragmentación de las clases delimitadas.



**Figura 9. Mapa con áreas delimitadas: a) dos (izquierda), b) tres (centro) y c) cuatro (derecha) clases.**

Al igual que antes los índices utilizados para la selección número óptimo de clases no fueron coincidentes en la indicación del número de cluster a retener (Tabla 7). El Coeficiente de Partición sugiere que la partición óptima se encuentra con dos clases mientras que para el índice de Entropía de la clasificación y el de Xie-Beni el óptimo sugerido, es de tres clases y el y el índice de Fukuyama-Sugeno indica 4 clases. El índice resumen que contiene

información de cada uno de los índices previamente calculados, indicó que la partición óptima es de tres clases.

Tabla 7: Selección del número de clases de la partición de departamentos a partir del análisis fuzzy k-means para la variabilidad

Índice	2 clases	3 clases	4 clases
Coeficiente de Partición	<b>1,071</b>	1,023	1,038
Entropía de Partición	0,116	<b>0,044</b>	0,079
Xie - Beni	0,010	<b>0,003</b>	0,008
Fukuyama - Sugeno	-56,211	-79,111	<b>-117,358</b>
Índice Resumen	2,000	<b>1,765</b>	2,515

‡ Para cada índice se indica el número óptimo de clases sugerido colocando en negrita el mejor valor del índice

Si se considera la dispersión en la participación de las actividades a lo largo del período, tres áreas o clases quedan definidas. Una clase compuesta (color blanco) por los departamentos de mayor desarrollo turístico, Punilla, San Alberto y Calamuchita y los departamentos más pobres Pocho y Minas, en los cuales la participación ha sido similar a lo largo de los años analizados. La clase 2 compuesta por los departamentos del sureste y Capital, esta coincide con la obtenida en la figura 7. La tercera clase está compuesta por el resto de los departamentos, esta coincide en parte con la clase 3 y resto de la clase 4 obtenidas en la figura 7.

## CONCLUSIONES

Teniendo en cuenta las estructuras económicas subyacentes en la industria turística, cuatro clases reflejan las particularidades y similitudes entre los departamentos, y esta

estratificación proporciona un elemento analítico adicional para la regionalización de las áreas.

Al analizar estos resultados, conjuntamente con el peso total de la actividad turística (figura 8) y la dispersión en la participación (figura 9), se puede observar que existen diferencias. Al observar el peso total, se observa que a veces se diluyen otros efectos en el agregado y quedan ocultas estructuras que están vinculadas a la composición de la actividad que aparecen cuando se considera la participación de las actividades en términos relativos (figura 7).

## CONCLUSIONES FINALES

Considerar la dimensión espacial en el análisis económico puede mejorar la interpretabilidad de los resultados. La variabilidad espacial genera correlaciones entre las observaciones de una misma variable registrada repetidamente en el espacio y por tanto los datos no pueden tratarse estadísticamente como independientes. Las técnicas multivariadas facilitan la interpretación de complejas relaciones entre variables, reducen la dimensión de la base de datos para mapear la variabilidad espacial, permiten detectar estructuras y revelan nuevas relaciones espaciales que pueden no ser evidentes cuando las variables económicas se analizan individualmente.

En esta investigación se analizó a través de técnicas multivariadas a las actividades características del turismo con el objetivo de poder identificar las diferencias en las estructuras departamental del territorio provincial.

La técnica multivariada MULTISPATI-PCA, diseñada para contemplar las relaciones entre las variables y su estructura espacial (autocorrelación), resultó apropiada para la visualización y exploración simultánea de datos de varias variables regionalizadas. En la comparación de MULTISPATI-PCA vs. PCA, utilizando el análisis restringido espacialmente la selección del número de componentes principales para la interpretación de la variabilidad fue no ambigua. El grado de estructuración espacial fue mayor con MULTISPATI-PCA que con PCA no restringido espacialmente. Esto se evidenció en los mapas de las variables sintéticas donde la estructura espacial fue más clara cuando se usaron las componentes principales espaciales de MULTISPATI-PCA.

Al aplicar el análisis de clúster fuzzy k-means incluyendo una dimensión espacial a través de la utilización de las componentes principales espaciales (MULTISPATI-PCA), se observó que la clasificación en cuatro clúster departamentales resulta apropiada a los índices utilizados para la selección número óptimo de clases no fueron coincidentes en la indicación del número de cluster a retener. Cuatro clases reflejan las particularidades y similitudes entre los departamentos teniendo en cuenta las estructuras económicas subyacentes en la industria turística. Esta estratificación proporciona un elemento analítico adicional para la regionalización de áreas turísticas.

La mayor participación promedio del VARCT (%) sobre el total del producto bruto regional la presentaron los departamentos de San Alberto (29,6%), Punilla (23,3%), Calamuchita (16,6%) y San Javier (7,2%). Sin embargo en el agregado, quedan ocultas estructuras relevantes para la comprensión del funcionamiento de la actividad económica que surgen del análisis de clúster realizado.

Se observa que la utilización de estas técnicas novedosas de mapeo multidimensional proporciona una herramienta que permite abordar nuevas dimensiones de análisis, que contribuyen a pensar en diferentes alternativas de desarrollo regional.

## REFERENCIAS BIBLIOGRÁFICAS

- Anselin, L. 2001. Spatial Effects in Econometric Practice in Environmental and Resource Economics. *American Journal of Agricultural Economics* 83, no. 3: 705–710.
- Arrouays D., Saby N.P.A., Thioulouse J., Jolivet C., Boulonne L. and Ratié C. 2011. Large trends in French topsoil characteristics are revealed by spatially constrained multivariate analysis. *Geoderma* 161, 107–114.
- Baller, R. D., Anselin L., Messner S. F., Glenn D., and Hawkins D. F.. 2001. Structural Covariates of US County Homicide Rates: Incorporating Spatial Effects. *Criminology* 39, no. 3: 561–588.
- Balzarini, M.; Teich, I.; Bruno, C.. 2011. Making Genetic Biodiversity Measurable: A Review of Statistical Multivariate Methods to Study Variability at Gene Level. *R*, n.d., 261–275.
- Bezdek J.C. 1981. *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York.
- Bivand R. 2008. Implementing representations of space in economic geography. *J. Reg. Sci.* 48(1): 1–27.
- Bivand, R. *Creating Neighbours*. 2014. <https://cran.r-project.org/web/packages/spdep>.
- Bivand R., Altman M., Anselin L., Assunção R., Berke O., Bernat A., Blanchet G., Blankmeyer E., Carvalho M., Christensen B., Chun Y., Dormann C., Dray S., Halbersma R., Krainski E., Legendre P., Lewin-Koh N., Li H., Ma J., Millo G., Mueller W., Ono H., Peres-Neto P., Piras G., Reder M., Tiefelsdorf M. and Yu D. 2017. *Spdep: Spatial Dependence: Weighting Schemes, Statistics and Models*. R Package Version 0.5-56. <https://cran.r-project.org/web/packages/spdep>.
- Ceccato V., Haining R., and Signoretta P.. 2002. Exploring Offence Statistics in Stockholm City Using Spatial Analysis Tools. *Annals of the Association of American Geographers* 92, no. 1: 29–51.
- CEPAL. *Indicadores sociales en América Latina y el Caribe*. 2005. <http://repositorio.cepal.org>.
- Chasco, M.. 2003. *Econometría espacial aplicada a la predicción-extrapolación de datos microterritoriales*. Consejería de Economía e Innovación Tecnológica.

- Chessel D, Dufour AB, Thioulouse J. 2004. The ade4 package-I- One-table methods. *R News* 4:5-10.
- Cliff, A. D., and Ord J. K.. 1973. *Spatial Autocorrelation*. London. Pion Limited.
- Cliff, A.D., and Ord, J. K.. 1982. *Spatial Processes: Models & Applications*. A. D. Neal L. Oden. *The Quarterly Review of Biology* 57:2, 236-236
- Córdoba, M., Balzarini, M., Bruno, C., Costa, J., 2012. Identificación de zonas de manejo sitio-específico a partir de la combinación de variables de suelo. *Corpoica. Ciencia y Tecnología Agropecuaria*.
- Córdoba, M. 2014. Herramientas estadísticas para el monitoreo y uso de la variabilidad espacial del rendimiento y propiedades de suelo intralote. <https://rdu.unc.edu.ar/handle/11086/1493>.
- Cressie N.A.C.1985. Fitting variogram models by weighted least squares. *Math. Geol.* 17:563-586.
- Dray, S. Chessel, and Thioulouse, J. 2003. Co-Inertia Analysis and the Linking of Ecological Data Tables. *Ecology* 84, no. 11: 3078–89. doi:10.1890/03-0178.
- Dray, S., Saïd S., and Débias F.. 2008. Spatial Ordination of Vegetation Data Using a Generalization of Wartenberg’s Multivariate Spatial Correlation. *Journal of Vegetation Science* 19, no. 1: 45–56. doi:10.3170/8-18312.
- Dray, S., and Jombart, T. 2011. Revisiting guerry’s data: introducing spatial constraints in multivariate analysis.” *The Annals of Applied Statistics* 5, no. 4. 2278–99.
- Fridgen J.J., Kitchen N.R., Sudduth K.A., Drummond S.T., Wiebold W.J. and Fraisse C.W. 2004. Management Zone Analyst (MZA): Software for Subfield Management Zone Delineation. *Agron. J.* 96: 100–108.
- Fukuyama Y. and Sugeno M. 1989. A new method of choosing the number of clusters for the fuzzy c-means method, *Proc. 5th Fuzzy Syst. Symp.*, p. 247-250.
- Gabriel, K. and Sokal R.. 1969. A New Statistical Approach to Geographic Variation Analysis. *Systematic Biology* 18, no. 3. : 259–278.
- Galarza R., Mastaglia N., Albornoz E.M. y Martínez C.E. 2013. Identificación automática de zonas de manejo en lotes productivos agrícolas. V Congreso Argentino de Agroinformática (CAI) - 42da. JAIIO, Córdoba.
- Geary, R. 1954. The Contiguity Ratio and Statistical Mapping. *The Incorporated Statistician* 5, no. 3: 115–146.
- Krugman, P. 1991. *Geography and Trade*. MIT Press, Cambridge MA.

- Der-Tsai L., Schachter B. 1980. Two Algorithms for Constructing a Delaunay Triangulation. *International Journal of Computer & Information Sciences* 9, no. 3: 219–242.
- Minasny B., McBratney A.B., 2002. FuzME version 3.0, Australian Centre for Precision Agriculture, The University of Sydney, Australia.
- Moran, P. 1948. The Interpretation of Statistical Maps. *Journal of the Royal Statistical Society. Series B (Methodological)* 10, no. 2: 243–251.
- Moreno Serrano, R., Valcarce E., 2002. *Econometría espacial: nuevas técnicas para el análisis regional. Una aplicación a las regiones europeas*. *Investigaciones Regionales - Journal of Regional Research*.
- Odeh I.O.A., Chittleborough D.J. and McBratney A.B. 1992. Soil Pattern Recognition with Fuzzy-c-means: Application to Classification and Soil-Landform Interrelationships. *Soil Sci. Soc. Am. J.* 56: 505.
- Pearson, K. 1901. LIII. On Lines and Planes of Closest Fit to Systems of Points in Space. *Philosophical Magazine Series 6* 2, no. 11 (November 1, 1901): 559–72.
- Peña, Daniel. 2002. *Análisis de datos multivariantes*. Editorial S.A. McGraw-Hill / Interamericana de España.
- R Core Team . 2017. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Saby N.P.A., Thioulouse J., Jolivet C.C., Ratié C., Boulonne L., Bispo A., Arrouays D. 2009. Multivariate analysis of the spatial patterns of 8 trace elements using the French soil monitoring network data. *Sci. Total Environ.* 407, 5644–5652.
- Sánchez-Peña, Landy L. 2012. Alcances y Límites de Los Métodos de Análisis Espacial Para El Estudio de La Pobreza Urbana. *Papeles de Población* 18, no. 72: 147–180.
- Schabenberger, O.; Pierce, F. J. , 2002. “Contemporary Statistical Models for the Plant and Soil Sciences.” *CRC Press*, February 24, 2017. <https://www.crcpress.com/Contemporary-Statistical-Models-for-the-Plant-and-Soil-Sciences/Schabenberger-Pierce/p/book>.
- Team, R. Core. *R: 2014. A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Xie L.X. and Beni G. 1991. Validity measure for fuzzy clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* 3(8): 841–847.

# ANEXO 1

Códigos R. Cálculo de los índices de autocorrelación espacial de Moran y de Geary y Análisis descriptivo.

```
### Capítulo 2 ###
```

```
# Cargar Librería
```

```
library(spdep)
```

```
# Lectura de Datos
```

```
load("F:/Tes/Bases/gVARCT.rda")
```

```
load("F:/Tes/Bases/df1.rda")
```

```
# Medidas Descriptivas
```

```
summary(df1)
```

```
# Definir vecinos
```

```
nb1 <- poly2nb(gVARCT) #Criterio queen
```

```
cards <- card(nb1) # cuenta el numero de vecinos en la lista de vecinos
```

```
summary(nb1)
```

```
# Opción 1 Definir matriz
```

```
lw1 <- nb2listw(nb1, style = "W") # Sin pesos estandarizo por fila
```

```
# Gráfico
```

```
plot(gVARCT, col = "grey95", border = "grey")
```

```
plot(lw1, coordinates(gVARCT), add = TRUE, pch = 20, lwd = 1.5, cex = 1)
```

```
## Coeficiente de Moran y Geary ##
```

```
i.moran <- lapply(df1, moran.mc, lw1, 999)
```

```
i.moran
```

```
i.geary <- lapply(df1, geary.mc, lw1, 999)
```

```
i.geary
```

## ANEXO 2

Códigos R. Estudios multivariados de datos georreferenciados

```
### Capítulo 3 ###
```

```
# Cargar Librería
```

```
library(ade4)  
library(spdep)  
library(geoR)  
library(gstat)
```

```
load("F:/Tes/Bases/gVARCT.rda")
```

```
## Abrir la base de datos
```

```
load("F:/Tes/Bases/df1.rda")
```

```
##Definir vecinos##
```

```
# poly2nb es una función que construye una lista de vecinos basada en regiones  
# que estan contiguas por compartir más de un punto en sus fronteras
```

```
nb1 <- poly2nb(gVARCT) #Criterio queen  
cards <- card(nb1) # cuenta el numero de vecinos en la lista de vecinos  
summary(nb1)
```

```
# Opción 1
```

```
lw1 <- nb2listw(nb1, style = "W") # Sin pesos estandarizo por fila
```

```
## Grafico ##
```

```
plot(gVARCT, col = "grey95", border = "grey")  
plot(lw1, coordinates(gVARCT), add = TRUE, pch = 20, lwd = 1.5, cex = 1)
```

```
# Análisis de componentes Principales
```

```
pca1 <- dudi.pca(df1, scannf = FALSE, nf = 7)
```

```
##Ver que objetos tiene
```

```
str(pca1)
```

```
#Para sacar los autovectores
```

```
pca1$e1
```

```
#Para sacar los autovalores
```

```

pca1$eig
# Que proporción de la varianza total explica
pca1$eig/sum(pca1$eig) * 100

# Análisis de componentes Principales espaciales (ACPe)

ms1 <- multispati(pca1, lw1, scannf = FALSE)
ms1

# Comparación entre ACP y ACPe
sum.ms1 <- summary(ms1)
sum.ms1

ms1$c1

sPC.lifgVARCT<-ms1$li
sPC.VARCT<-cbind(dep.names, ms1$li,xy)

# Representación de los resultados del ACPe

plot(ms1)
s.arrow(ms1$c1, clabel = 0.6)
s.corcircle(ms1$c1, clabel = 0.7)
add.scatter.eig(ms1$eig, xax = 1, yax = 2, posi = "topleft", ratio = 0.2)

png(file = "figs/fig-fig5d.png", width = 6, height = 3, units = "in",
     res = 72)
par(mar = rep(0.1, 4))
par(mfrow = c(1, 2))
plot(gVARCT, col = col.region[region.names], border = "transparent")
s.value(xy, bet$s[, 1], add.plot = TRUE)
plot(gVARCT, col = col.region[region.names], border = "transparent")
s.value(xy, bet$s[, 2], add.plot = TRUE)
dev.off

# Correlación con Multispati

mc.mpca<- lapply(ms1$li, moran.mc, lw1, 999)

par(mar = rep(0.1, 4))
par(mfrow = c(1, 2))

plot(gVARCT, col = "grey95", border = "grey",main="SC1")
s.value(xy, ms1$li[, 1], add.plot = TRUE)
mtext(paste("MC = ", round(mc.mpca[[1]]$statistic,3),paste("(", mc.mpca[[1]]$p.value,
  ")"), cex = ""))

```

```
plot(gVARCT, col = "grey95", border = "grey",main="SC2")
s.value(xy, ms1$li[, 2], add.plot = TRUE)
mtext(paste("MC = ", round(mc.mpca[[2]]$statistic,3),paste("(", mc.mpca[[2]]$p.value,
")), cex = ""))
```

## ANEXO 3

Códigos R. Clasificación espacial multivariada fuzzy k-means

### Capítulo 4 ###

# Cargar Librería

```
library(rgdal)
```

```
library(e1071)
```

```
library(raster)
```

```
library(nlme)
```

### Abrir archivos

```
load("F:/Tes/Bases/basefuzzy.rda")
```

```
load("F:/Tes/Bases/gVARCT.rda")
```

```
load("F:/Tes/Bases/SpP.rda")
```

### Delimitación

```
pca2 <- dudi.pca(basefuzzy[,3:9], center=T,scannf = FALSE, nf = 7)
```

```
scatter(pca2, xax = 1, yax = 2,clab.r=0.4, clab.c=0.9)
```

```
cord <- coordinates(basefuzzy[,1:2])
```

```
nb1 <- poly2nb(gVARCT) #Criterio queen
```

```
lw1 <- nb2listw(nb1, style = "W")# Sin pesos estandarizo por fila
```

```
ms2 <- multispati(pca2, lw1, scannf = F, nfposi = 3)
```

```
sum.ms2 <- summary(ms2)
```

```
s.arrow(ms2$c1,xax = 1, yax = 2, clabel = 1)
```

```
add.scatter.eig(ms2$eig, xax = 1, yax = 2, posi = "bottomleft", ratio =  
0.2)
```

```
CS <- ms2$li[,1:3]
```

```
PredAM <- cbind(basefuzzy,CS) ;PredAM
```

### Delimitacion

```
CM2<-cmeans(PredAM[,10:12],2,100,method="cmeans",m=1.3)
```

```
CM3<-cmeans(PredAM[,10:10],3,100,method="cmeans",m=1.3)
```

```
CM4<-cmeans(PredAM[,10:12],4,100,method="cmeans",m=1.3)
```

```
CM22<-as.data.frame(CM2$cluster)
```

```
CM33<-as.data.frame(CM3$cluster)
```

```
CM44<-as.data.frame(CM4$cluster)
```

### Delimitacion de areas de manejo

```

I2CM <- fclustIndex(CM2,PredAM[,10:12], index=c("xie.beni",
      "fukuyama.sugeno",
      "partition.coefficient", "partition.entropy"))
I3CM <- fclustIndex(CM3,PredAM[,10:12], index=c("xie.beni",
      "fukuyama.sugeno",
      "partition.coefficient", "partition.entropy"))
I4CM <- fclustIndex(CM4,PredAM[,10:12], index=c("xie.beni",
      "fukuyama.sugeno",
      "partition.coefficient", "partition.entropy"))

Indices0 <- cbind(I2CM,I3CM,I4CM)
XieBeni <-Indices0[1,]
FukSug <-Indices0[2,]
CoefPart_1 <-Indices0[3,]
CoefPart <- 1/CoefPart_1
EntrPart <-Indices0[4,]
Indices <- as.data.frame(rbind(XieBeni,FukSug,CoefPart,EntrPart))
Indices

XieBeniMax<-max(Indices[1,])
FukSugMax<-max(Indices[2,])
CoefPartMax<-max(Indices[3,])
EntrPartMax<-max(Indices[4,])
XieBeniN<- XieBeni/XieBeniMax
FukSugN<- FukSug/FukSugMax
CoefPartN<- CoefPart/CoefPartMax
EntrPartN<-EntrPart/EntrPartMax
IndicesN <- as.data.frame(rbind(XieBeniN,FukSugN,CoefPartN,EntrPartN))
IndicesN2 <- (IndicesN)^2
Indice2CM <- sqrt(sum(IndicesN2[,1]))
Indice3CM <- sqrt(sum(IndicesN2[,2]))
Indice4CM <- sqrt(sum(IndicesN2[,3]))
Indice2CM
Indice3CM
Indice4CM

base00 <- cbind(CM22,CM33,CM44)

base000=SpatialPolygonsDataFrame(SpP, base00, match.ID = F)

plot(base000, col = 1:26, pbg="white")

#coordinates(base00) = ~x+y
#gridded(base00)=TRUE

par(mfrow = c(1, 2))
splot(base000["CM2$cluster"],col.regions=terrain.colors(100),colorkey= F)

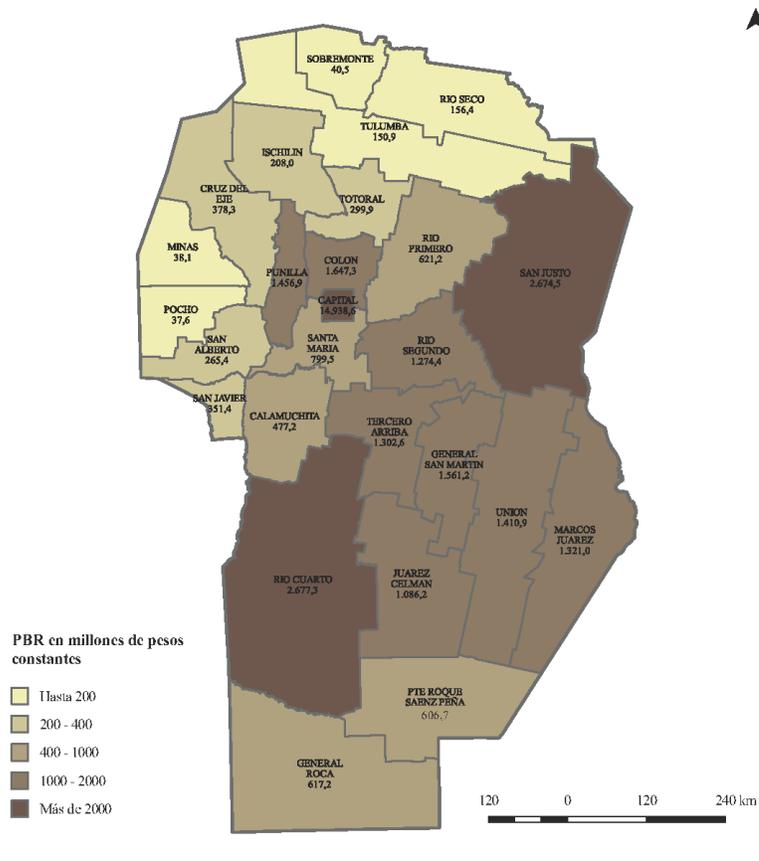
```

```
splot(base000["CM3$cluster"],col.regions=terrain.colors(100),colorkey= F)  
splot(base000["CM4$cluster"],col.regions=terrain.colors(100),colorkey= F)
```

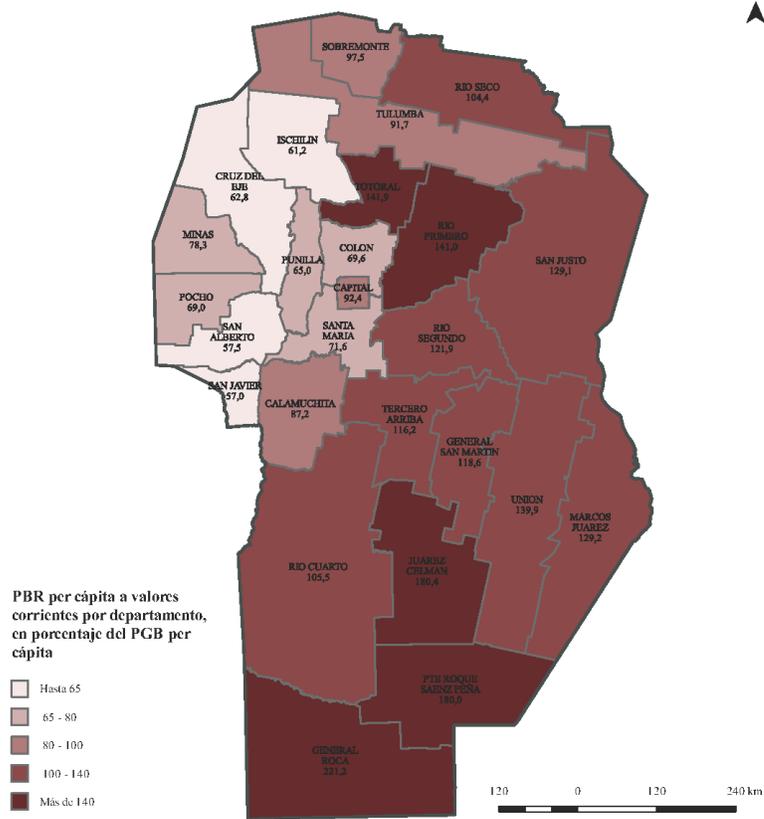
## ANEXO 4

Mapa Producto Bruto Regional a valores constantes y pér cápita. Año 2014.

Provincia de Córdoba.  
Producto Bruto Regional a Valores Constantes. Año 2014



## Producto Bruto Regional a Valores Corrientes como porcentaje del Producto Geográfico Bruto Per Capita. Año 2014



Fuente: Elaboración DEE - Subdirección de Cartografía y SIG en base a diversas fuentes

## ANEXO 5

Participación de VARCT sobre el producto bruto regional

Nom_dpto	VARCT Promedio
Calamuchita	16,6%
Capital	5,0%
Colón	4,5%
Cruz del Eje	4,2%
General Roca	1,7%
General San Martín	2,2%
Ischillín	4,2%
Juárez Celman	1,1%
Marcos Juárez	1,9%
Minas	4,4%
Pocho	5,1%
Presidente Roque Saenz Peña	2,0%
Punilla	23,3%
Río Cuarto	3,3%
Río Primero	1,5%
Río Seco	2,9%
Río Segundo	1,7%
San Alberto	29,6%
San Javier	7,2%
San Justo	2,3%
Santa María	5,5%
Sobremonte	2,8%
Tercero Arriba	2,6%
Totoral	1,9%
Tulumba	2,8%
Unión	2,0%

## ANEXO 6

### **Valor Agregado de las ramas características del turismo. Definiciones.**

Es difícil delimitar con precisión el ámbito del turismo, ya que en realidad abarca varios sectores de la realidad económica, constituyendo lo que se denomina un sector transversal de los restantes. Es por ello que para calcular el Valor Agregado de este sector, así como para estimar el empleo utilizado en el mismo se debe construir la Cuenta Satélite de Turismo según lo establecido por el Sistema de Cuentas Nacionales. Sin embargo, existen algunas limitaciones estadísticas en la elaboración de la Cuenta Satélite (CST) a nivel regional, especialmente en ausencia de un marco de compilación nacional para las estadísticas de turismo.

Es por este motivo que las recomendaciones de organismos internacionales especializados en temas de turismo sugieren realizar aproximaciones al valor agregado y al empleo en turismo mediante otras metodologías, siendo la más utilizada aquella referida a las ramas características de turismo.

Teniendo presente estas limitaciones, y dado que todavía no se cuenta con un cálculo de la CST a nivel nacional, se intenta realizar una aproximación al valor agregado generado por el sector turístico, utilizando la estructura de cálculo del Producto Geográfico Bruto (PGB) que se realiza a nivel provincial.

Las industrias turísticas (también conocidas como actividades turísticas o ramas características del turismo) son aquellas que generan principalmente productos característicos del turismo. Los productos característicos del turismo son aquellos que cumplen uno o ambos de los siguientes criterios:

- (a) El gasto turístico en el producto debería representar una parte importante del gasto total turístico (condición de la proporción que corresponde al gasto/demanda).
- (b) El gasto turístico en el producto deberían representar una parte importante de la oferta del producto en la economía (condición de la proporción que corresponde a la oferta). Este

criterio supone que la oferta de un producto característico del turismo se reduciría considerablemente si no hubiera visitantes<sup>5</sup>.

La Organización Mundial del Turismo (OMT) ha realizado un listado de los productos y actividades asociados a ellos que se consideran característicos del turismo.

Las diez primeras filas agrupan los considerados característicos del turismo a escala mundial y por lo tanto poseen comparabilidad internacional. Las filas 11 y 12 están reservadas para los productos característicos del turismo específicos de cada país o región.

Lista de categorías de productos de consumo característicos del turismo y de actividades características del turismo

Productos	Actividades
1. Servicios de alojamiento para visitantes	1. Alojamiento para visitantes
2. Servicios de provisión de alimentos y bebidas	2. Actividades de provisión de alimentos y bebidas
3. Servicios de transporte de pasajeros por ferrocarril	3. Transporte de pasajeros por ferrocarril
4. Servicios de transporte de pasajeros por carretera	4. Transporte de pasajeros por carretera
5. Servicios de transporte de pasajeros por agua	5. Transporte de pasajeros por agua
6. Servicios de transporte aéreo de pasajeros	6. Transporte aéreo de pasajeros
7. Servicios de alquiler de equipos de transporte	7. Alquiler de equipos de transporte
8. Agencias de viajes y otros servicios de reservas	8. Actividades de agencias de viajes otros servicios de reservas
9. Servicios culturales	9. Actividades culturales
10. Servicios deportivos y recreativos	10. Actividades deportivas y recreativas
11. Bienes característicos del turismo, específico de cada país	11. Comercio al por menor de bienes característicos del turismo, específicos
12. Servicios característicos del turismo, específicos de cada país	12. Otras actividades características del turismo, específicas de cada país

Fuente: Recomendaciones internacionales para Estadísticas del Turismo, Naciones Unidas, 2008.

Debido a que no existe consenso internacional acerca de qué actividades deben incluirse en

<sup>5</sup> Definición glosario básico de la Organización Mundial del Turismo, OMT, <http://www2.unwto.org/es>

el grupo correspondiente a los bienes y servicios característicos específicos de cada país (filas 11 y 12), y a los fines de poder realizar comparaciones interregionales, sólo se considerará el primer grupo de actividades (filas 1 a 10).

Esta metodología resulta la mejor opción desde el punto de vista técnico, pero puede producir sobreestimación o subestimación de la medición económica del turismo, por no considerar la actividad generada en ramas no características o sobrevalorarla en las ramas características (que también producen para no turistas). Por ejemplo, el comercio no ha sido incluido, sin embargo, el turista consume estos bienes, y de esta manera participa en la generación de empleo e ingreso, principalmente en las localidades con gran afluencia de turistas.

Para el cálculo del valor agregado se utiliza la serie de Producto Geográfico Bruto (PGB) a valores constantes.