



TRANSITARE

Revista de Turismo, Economía y Negocios

transitare.anahuacoaxaca.edu.mx

Artículo de investigación

Clasificación de turistas asistentes a reuniones utilizando análisis multivariado con variables mixtas

Classification of Tourists Attending Meetings by Using Multivariate Analysis of Mixed Variables

Victoria Dowbley^{1*}, Gastón Miguel de la Cruz²

¹Facultad de Ciencias Económicas, Universidad Nacional de La Plata

²Facultad de Ciencias Económicas, Universidad de Buenos Aires

Argentina

Historia del artículo. *Recibido: 5 de julio de 2016; aceptado: 4 de septiembre de 2016.*

*Correo electrónico: v_dowbley@yahoo.com.ar

Esta obra está bajo [Licencia de Creative Commons Reconocimiento-NoComercial 4.0 Internacional](https://creativecommons.org/licenses/by-nc/4.0/).

Universidad Anáhuac de Oaxaca, Oaxaca, México; 2017.

Abstract. The objective of this work is to propose a multivariate methodology that considers a set of variables simultaneously to classify tourists attending meetings from a database of great dimension with variables of mixed nature. The database corresponds to the demand survey carried out by the Economic Observatory of Meetings Tourism of Argentina for the year 2015. This paper proposes a three-stage analysis strategy which takes into account the characteristics of the database. These stages are: preparation of the database (linked to the objective of the classification and its theoretical framework), application of the Cluster analysis with the non-hierarchical classification algorithm k-medias, characterization and validation of the obtained groups. Since the algorithm k-medias is efficient for large databases but only works with a few numerical variables, dimensions are first reduced by applying principal components analysis for mixed variables and then the algorithm is applied on a few numerical components. This paper proposes criteria to select a solution among a set of possible solutions according to the objective of the classification. The software used is: first and third stage SPSS version 20, second stage R version 3.1. As a result, a statistically validated solution of seven groups was obtained. This work is expected to contribute to the understanding of the diversity of tourists attending meetings and encourage the use of multivariate statistical methods in this area of research.

Keywords: Multivariate Analysis, Cluster Analysis, Meetings Tourism, Mixed Variables.

Resumen. El objetivo de este trabajo es proponer una metodología multivariada que considere un conjunto de variables simultáneamente para clasificar asistentes turistas a reuniones a partir de una base de datos de gran dimensión con variables de naturaleza mixta. La base de datos corresponde a la Encuesta de Demanda llevada a cabo por el Observatorio Económico de Turismo de Reuniones de Argentina correspondiente al año 2015. Se propone una estrategia de análisis en tres etapas que tiene en cuenta las características de la base de datos. Estas etapas son: preparación de la base de datos (relacionada con el objetivo de la clasificación y su marco teórico), aplicación del Análisis de Cluster con el algoritmo de clasificación no jerárquico k-medias, caracterización y validación de los grupos obtenidos. Dado que el algoritmo k-medias es eficiente para grandes bases de datos pero solo funciona con unas pocas variables numéricas, primero se reducen dimensiones aplicando Análisis de Componentes Principales para variables mixtas y luego se aplica el algoritmo sobre unas pocas componentes numéricas. Se plantean criterios para seleccionar una solución entre un conjunto de soluciones posibles acorde al objetivo de la clasificación. El software utilizado es: primera y tercera etapa SPSS versión 20, segunda etapa R versión 3.1. Como resultado se obtuvo una solución de siete grupos validada estadísticamente. Con este trabajo se espera contribuir a la comprensión de la diversidad de turistas asistentes a reuniones e incentivar el uso de los métodos estadísticos multivariados en esta área de investigación.

Palabras clave: Análisis multivariado, Análisis de cluster, Turistas asistentes a reuniones, Variables mixtas.

1. Introducción

El Turismo es una de las actividades de mayor relevancia económica a nivel mundial y de mayor crecimiento sostenido. En los últimos años el crecimiento de arribos internacionales creció por encima de las proyecciones esperadas y hoy en día la actividad representa el 7% de las exportaciones de bienes y servicios mundiales (UNWTO, 2016). Argentina acompaña este fenómeno de manera similar, al mostrar un crecimiento robusto de sus arribos internacionales a partir del año 2002.

A fin de asegurar el desarrollo turístico a largo plazo es necesario contar con un consenso entre los actores que intervienen en la actividad a fin de aunar los esfuerzos detrás de un objetivo común. Argentina entendió este desafío y por ello ha elaborado planes y programas tanto estratégicos como específicos. Tal es el caso del Turismo de Reuniones, modalidad de turismo que ha emprendido planes sectoriales propios como el Plan de Marketing de Turismo de Reuniones 2008-2011 y su actualización del año 2013-2016.

La necesidad de los actores del sector de turismo de reuniones de conocer en profundidad al heterogéneo grupo de turistas asistentes ha fomentado el interés por la clasificación y la caracterización de los mismos. A partir del año 2011 el Observatorio Económico del Turismo de Reuniones (OETR) recolecta, analiza y publica determinadas características propias del asistente turístico como la estadía promedio, el lugar de origen de los turistas, el gasto total, gasto promedio, entre otras características. No obstante, hay autores que reconocen que no sólo es necesario conocer las características agregadas como el gasto, el cual es abordado por la mayoría de los estudios, sino también conocer las características propias de los distintos grupos que componen al turismo de eventos (Tang & Turco, 2001).

Por este motivo, es necesario indagar sobre metodologías que logren una clasificación de los turistas asistentes efectiva en cuanto a los objetivos de incrementar el impacto económico de los eventos realizados. A su vez, estas metodologías deben permitir un balance adecuado entre el esfuerzo dedicado a las mismas y las limitaciones de tiempo y otros recursos del investigador.

En este trabajo se consideran los asistentes a reuniones que son turistas, ya sea que residan en Argentina o en el extranjero. El objetivo de este trabajo es proponer una metodología multivariada que considere un conjunto de variables simultáneamente para clasificar a dichos asistentes turistas a partir de una base de datos de gran dimensión con variables de naturaleza mixta. Para ello se utilizan los

resultados de la encuesta de demanda del Observatorio Económico de Turismo de Reuniones de la República Argentina del año 2015.

Se espera que este trabajo empírico contribuya al campo de los estudios en turismo, específicamente al campo de la investigación del mercado turístico y el estudio del comportamiento del turista, en donde se reconoce un faltante de evidencia empírica en investigación aplicada que puedan poner a prueba diversos modelos teóricos. (Horner & Swarbrooke, 2016).

El trabajo se estructura de la siguiente manera: en la segunda sección se presentan el marco teórico que sustentan la metodología empleada, la cual se detalla en la tercera sección. En la cuarta sección se muestran los resultados alcanzados para luego dar lugar a las discusiones de los resultados alcanzados, presentes en la quinta sección.

2. Marco conceptual

El *management* -y comercialización de los eventos- ha sido reconocido como un área de estudio de importancia tanto para los investigadores como para los profesionales del sector (Getz, 2005, 2008). Desde el lado académico, el turismo de eventos ha sido estudiado en profundidad dando cuenta de la contribución de las ciencias sociales a esta área de estudio interdisciplinar (Getz 2012, Getz en Page & Connell, 2012). En ese sentido, Getz (2007) hace una diferenciación entre el turismo y los eventos, como campos de estudios propios y que la conjunción de ambos da como resultado un nuevo campo de estudio denominado *Turismo de eventos*. Bajo este enfoque el autor diseña un modelo de gestión del turismo de eventos, en función del tipo de evento determinado entre otras cosas por las características de sus asistentes. Es aquí donde otras disciplinas pueden realizar aportes para la clasificación de los turistas asistentes.

2.1. *Conceptos del marketing*

En primer lugar es válido citar algunos conceptos de Marketing y sus disciplinas derivadas como la Investigación de Mercados y el Comportamiento del consumidor. En párrafos anteriores se mencionó el modelo de gestión Getz (2007) en donde uno de los determinantes de éxito es el tipo de asistentes; en otras palabras, el éxito del turismo de eventos depende de la cantidad y la cualidad del

turista que asiste al evento. Es por ello que muchas organizaciones y responsables de la gestión de un evento se esfuerzan en mejorar sus estrategias de comercialización en un marco de eficiencia y efectividad. Y es por estas razones que los organizadores deciden segmentar sus mercados.

Segmentar representa la acción de agrupar a determinados elementos (en este caso personas que consumen bienes, servicios) de modo que ese grupo tenga homogeneidad entre dichos elementos y a la vez este grupo sea heterogéneo frente a otros grupos. En otras palabras, la segmentación consiste en la visualización de un mercado heterogéneo como un número de mercados más pequeños y más homogéneos (Kotler, García de Madariaga, Flores Zamora, Bowen & Makens, 2013).

Por tanto, a la hora de realizar una segmentación es necesario determinar aquellas características que determinan la homogeneidad y/o la heterogeneidad del grupo, y esto es una de los objetos de estudio del turismo de eventos. Como ejemplo se cita el trabajo de Shaw, Agarwal & Bull (2000), quienes identifican tres áreas de estudio: la naturaleza de la demanda turística, la expresión de esto en términos de patrones de viaje, y el impacto de los turistas en la comunidad local. Por otro lado, se elaboraron trabajos de revisión sobre las segmentaciones en la demanda turística, a fin de categorizar los estudios de segmentación de la demanda turística en España (Talaya & Lara, 1996; González & Molina, 2007). No obstante, estos enfoques son parciales en cuanto nuestros intereses de hallar una metodología que permita clasificar a los turistas de eventos. El trabajo de Tkaczynski & Rundle-Thiele (2011) resuelve esta cuestión al analizar 120 segmentaciones de la demanda de eventos, a fin de categorizar:

- Las variables más recurrentes en las investigaciones
- Los instrumentos de recolección de datos más recurrentes en las investigaciones
- La técnica estadística del análisis de los datos más recurrente
- Los instrumentos de procesamiento de datos más recurrentes

Además, categoriza a las segmentaciones en base a los cuatro fundamentos de segmentación de Kotler: demográfico, geográfico, psicográfico y comportamental. Se destaca que, de los 120 estudios analizados, sólo 16 utilizan Análisis de Cluster como método de segmentación.

2.1. Conceptos de la estadística

Tal lo señalado anteriormente la segmentación se realiza dentro de lo que se conoce como la investigación de mercados, actividad que corresponde a la aplicación de distintas técnicas tendientes a identificar y/o cuantificar determinados hechos de un mercado. En función del objetivo de una investigación, los recursos con que se cuentan y el tipo de fenómeno a estudiar, la investigación puede valerse de técnicas cualitativas o cuantitativas. En este último caso corresponde utilizar técnicas estadísticas que, en función de la profundidad de la investigación serán descriptivas, explicativas o predictivas.

Respecto del término “segmentación” cabe hacer una aclaración. En el área de la estadística existen muchos métodos multivariados orientados a la clasificación de elementos: uno de ellos es denominado Análisis de Segmentación y otro Análisis de Cluster. En particular, el Análisis de Segmentación se usa para aproximar la probabilidad de que un elemento pertenezca a un grupo determinado por una variable objetivo en función de las otras variables. Esto es, básicamente, un modelo de predicción, donde hay una variable explicada y múltiples variables explicativas; es decir, existe un conocimiento a priori sobre el tipo de relaciones entre las variables. En el Análisis de Cluster, el conocimiento a priori no existe y se trabaja directamente el conjunto de variables de la base de datos con la intención de encontrar patrones o agrupamientos.

Esta investigación, como se mencionó anteriormente, tiene el objetivo de clasificar a los asistentes turistas a eventos del tipo Congresos en 2015 utilizando un método estadístico multivariado. En particular, se aplica el Análisis de Cluster (AC) definido como un conjunto de técnicas multivariadas exploratorias cuyo objetivo es agrupar los elementos de una población o de una muestra según las características que poseen. Los grupos resultantes contendrán un alto grado de homogeneidad entre los elementos de cada grupo y un alto grado de heterogeneidad entre grupos. Este método parte de la información recolectada en censos o en encuestas y considera, en forma simultánea, un conjunto de variables para clasificar. Para ello se utilizan medidas de disimilitud o de distancia para formar grupos de tal forma que la variabilidad dentro de los grupos sea mínima y entre grupos sea máxima (Hair, Anderson & Tatham, 1999; Everitt, Landau & Leese, 2001; Peña, 2002). Además, permite obtener un exhaustivo abanico de grupos y se caracterizan por no establecer a priori los valores de corte de las variables seleccionadas ni responder a criterios de informantes calificados o expertos.

Existen varios tipos de algoritmos de clasificación dentro del Análisis de Cluster. En este trabajo se optó por aplicar el algoritmo de clasificación de las k-medias (Mac Queen, 1967). Este es un método de reasignación que permite que un elemento asignado a un grupo en un determinado paso del proceso sea reasignado a otro grupo en un paso posterior si se optimiza un criterio de selección. El proceso termina cuando no quedan elementos cuya reasignación permita optimizar el resultado que se ha conseguido. Este algoritmo es eficiente para grandes bases de datos, pero solo funciona con unas pocas variables numéricas (Mac Queen, 1967; Anderberg, 1973). Por lo tanto, dada la cantidad de observaciones y de variables categóricas y numéricas que tiene la encuesta de demanda del OETR, se aplicó primero una estrategia de análisis en tres etapas que tiene en cuenta las características de la base de datos. Se explica en detalle cada etapa en la siguiente sección.

3. Metodología

La estrategia de análisis de la base de datos consta de tres etapas. En la Etapa 1 se preparó la base de datos: se resumió la información eliminando algunas variables originales siguiendo distintos criterios teóricos y con la realización de un Análisis de Componentes Principales (ACP) para variables mixtas. En la Etapa 2 se aplicó el algoritmo de clasificación no jerárquico k-medias. Finalmente, en la Etapa 3 se validaron los grupos obtenidos.

Etapa 1. Preparación de la base de datos

En esta etapa se eliminó la información redundante o irrelevante para los objetivos de la clasificación, de forma tal que se simplificaran los cálculos y se facilitara la identificación de los grupos (Berdegué y Escobar, 1990; Everitt et al., 2001; Chian, Long Chen & Wei Su, 2007; Ahmad y Dey, 2011; Chatzis, 2011). Para lograr esto, se realizaron las siguientes tareas:

- a) Se definió el objetivo de la clasificación: La finalidad de la clasificación es identificar segmentos de mercado para los actores del sector y la definición de políticas orientadas a su desarrollo.

b) Se consideró un marco teórico en función del objetivo de la clasificación. Se tuvieron en cuenta aquellos aspectos que han sido considerados relevantes por la comunidad científica para estudiar los tipos de asistentes turistas en función del objetivo planteado (Tkaczynski & Rundle-Thiele, 2011).

c) Se seleccionaron las variables de la base de datos de la Encuesta de Demanda que permiten hacer operativo el marco teórico:

- Lugar de residencia: Argentina – Extranjero.
- Tamaño del grupo: cantidad de personas que comparten un mismo presupuesto de viaje
- Cantidad de noches en Argentina que dura el viaje
- Tipo de alojamiento (Hotel 5 estrellas, Hotel 4 estrellas, Hotel 3 estrellas, Hotel 1-2 estrellas, Apart Hotel, Hotel Boutique, Hostería – posada, Hospedaje – Residencial, Cabaña – Bungalow, Albergue – Hostel B&B, Establecimiento rural, Vivienda propia, de familiares o amigos, Vivienda alquilada, Ninguno – No se hospedó).
- Forma de organización del viaje: contratación directa de alojamiento y/o transporte, por cuenta de terceros, a través de un paquete armado por el organizador del evento, a través de un paquete armado por una agencia de viajes.
- Medio de transporte utilizado para arribar a la ciudad destino: avión, automóvil, ómnibus, ferrocarril, barco, otro.
- Artículos comprados durante el viaje: ropa, libros, souvenir, música, artículos de cuero, productos regionales, otros.
- Actividades realizadas durante el viaje: visita a museos, espectáculos de tango / folklore, Teatro/Cine, City Tour, Excursión contratada, Experiencia en establecimientos rurales, Golf, otra.
- Monto gastado durante todo el viaje a la Argentina (se refiere a gastos del grupo de viaje en conjunto)

d) Se redefinieron, las variables originales seleccionadas a través de variables de síntesis.

- Artículos comprados: cada una de las opciones se representa con una variable dicotómica en la base de datos. Se creó una variable numérica: cantidad de tipos de artículos comprados.

- Actividades realizadas: al igual que con los artículos comprados, cada una de las opciones de actividades se representa con una variable dicotómica en la base de datos. Se resumieron estas variables en una sola: cantidad de actividades realizadas.

e) Se calcularon nuevas variables e indicadores:

- Con la cantidad de noches en Argentina se calculó la estadía (medida en cantidad de días).

- Se calculó el Gasto diario per cápita: (Gasto total / Tamaño del Grupo) / Estadía

f) Para las variables cuantitativas Gasto Total, Gasto diario per cápita, Estadía y Tamaño de grupo, se detectaron y eliminaron los casos atípicos mediante la determinación del percentil 98 (Hair et al., 1999; Peña, 2002.) Para la variable cualitativa Medio de transporte se eliminaron los casos con modalidades de baja frecuencia: barco y ferrocarril (juntas representaban un total de 78 casos).

g) Dado que las variables cuantitativas originales poseen una distribución marcadamente asimétrica y esto dificulta el buen desempeño del ACP (Hair et al., 1999; Peña, 2002) se optó por transformar las variables sumando 1 y aplicando logaritmo antes de estandarizarlas:

$$\text{nueva variable} = \ln(\text{variable original} + 1)$$

h) Se analizó la variabilidad de las variables numéricas, la bibliografía indica que se deben descartar aquellas variables que presentan bajo poder discriminatorio en la construcción de los grupos. Para ello se calculó el coeficiente de variación de cada una de las variables numéricas. En este caso todas las variables numéricas consideradas tenían coeficientes de variación superiores al 80% y ninguna fue eliminada.

i) Se analizó la correlación entre las variables numéricas mediante el cálculo de la matriz de correlación total. La bibliografía indica que se deben identificar los pares de variables altamente correlacionadas y descartar una de las variables de cada par. La elección de la variable a descartar depende del criterio del analista. En este caso no fue necesario eliminar variables ya que ningún par de correlaciones fueron superiores a 0,90 (Berdegué y Escobar, 1990; Köbrich, Rehman & Khan, 2003).

j) Se estandarizaron las variables numéricas para eliminar los efectos de escala y de unidad de medida (Berdegué y Escobar, 1990; Köbrich et al., 2003). Por lo tanto, su valor se interpreta como la cantidad de desvíos estándar que se encuentran entre un valor cualquiera de una variable numérica y su media.

k) Se unieron modalidades¹ en las variables categóricas para reducir la complejidad de la clasificación sin pérdida de información. Hay dos formas de hacerlo teniendo en cuenta si la variable es nominal u ordinal. En el caso de las variables nominales, cada categoría puede ser unida con cualquier otra categoría; este procedimiento se denomina *libre*. En el caso de las variables ordinales, una modalidad sólo puede ser unida a otra si es contigua. Este procedimiento se conoce como *monótono* (Escobar, 1998). Así, las modalidades con baja frecuencia que pueden distorsionar el ACP mixto de la misma manera que lo harían en un Análisis de Correspondencias Múltiples (ACM), se unen con las modalidades contiguas para dar mejores resultados.

- En función del objetivo de clasificación se optó por unir las 14 modalidades de tipo de alojamiento en tres: *Alojamiento con cocina*, *Alojamiento sin cocina* y *No se hospedó*.

- Las modalidades de la variable Organización del viaje se unieron en dos: *Contratación directa* y *Contratación por medio de un tercero*.

Así, al finalizar la Etapa 1 de análisis, la base quedó constituida por 2.996 casos y 11 variables, tal como se muestra en la Tabla 1.

¹ Se entiende por modalidad: cada una de las categorías de la variable categórica.

Tabla 1. *Variables consideradas al final de la Etapa 1 de análisis*

| Número de variable | Nombre de la variable | Tipo de variable |
|--------------------|--|---|
| 1 | Gasto total | Numérica |
| 2 | Gasto Diario per cápita | Numérica |
| 3 | Cantidad de actividades realizadas | Numérica |
| 4 | Cantidad de artículos comprados | Numérica |
| 5 | Tamaño del grupo | Numérica |
| 6 | Estadía | Numérica |
| 7 | Lugar de residencia | Catórica: Argentina - Extranjero |
| 8 | Tipo de alojamiento | Catórica: Con cocina - Sin cocina - No se hospedó |
| 9 | Medio de transporte | Catórica: Avión - Automóvil - Ómnibus |
| 10 | Forma de organización del viaje | Catórica: Contratación directa - Contratación por medio de terceros |
| 11 | Compra de artículos elaborados en origen | Catórica: SI - NO |

Fuente: Elaboración propia sobre la base de información provista por el OETR.

Etapa 2. Aplicación del algoritmo de clasificación no jerárquico para la obtención de los grupos

Siguiendo el criterio sugerido por Lebart, Morineau & Piron (1995), se realizó en primer lugar un ACP para variables mixtas y luego se aplicó el algoritmo *k-medias*. Cabe recordar que los componentes son ortogonales y que un número reducido de ellos suelen explicar un alto porcentaje de la información contenida en la matriz de datos original. Teniendo en cuenta que cada componente principal es una variable sintética construida a partir de las variables originales, cada caso (asistente turista) puede ser identificado por sus coordenadas respecto de cada uno de los componentes. Por lo tanto, dichos componentes pueden ser utilizados como variables de clasificación en el AC. De esta forma, se asigna en la clasificación un peso relativo a las variables originales que surge naturalmente de los datos. A diferencia de la etapa anterior donde se reduce el número de variables originales, aquí se reduce las dimensiones del espacio conservando dichas variables. A su vez, los componentes constituyen variables numéricas por lo que permiten la aplicación del algoritmo *k-medias*.

Por otra parte, debe notarse que hay muchos algoritmos del tipo *k-medias* dependiendo de la medida de disimilitud utilizada, de la selección de los *k* centros iniciales y de las estrategias para calcular la media de los grupos. En particular, en este trabajo se utiliza el ACP para variables mixtas y luego se aplica el algoritmo *k-medias* utilizando la distancia euclídea con las componentes seleccionadas. Se eligen aleatoriamente los *k* centros iniciales y se calcula la media de los grupos en cada reasignación de elementos (turista asistente).

Esta obra está bajo Licencia de Creative Commons Reconocimiento-NoComercial 4.0 Internacional.

Universidad Anáhuac de Oaxaca, Oaxaca, México; 2017.

El ACP se realizó con el comando `dudi.mix` del paquete `ade4` (Dray & Dufuor, 2007) del *software* R sobre las 11 variables resultantes de la Etapa 1 (seis numéricas, dos dicotómicas y dos categóricas con tres modalidades) y expuestas en la Tabla 1. Los porcentajes de la varianza de los datos explicada por los componentes resultantes se presentan en la Tabla 2.

Tabla 2. *Porcentaje de la varianza explicada por los componentes principales del ACP*

| Componente | Autovalor | Porcentaje de la varianza explicada | Porcentaje de acumulado de la varianza acumulada |
|------------|-----------|-------------------------------------|--|
| 1 | 3,32 | 25,52 | 25,52 |
| 2 | 1,89 | 14,52 | 40,04 |
| 3 | 1,41 | 10,88 | 50,92 |
| 4 | 1,34 | 10,29 | 61,21 |
| 5 | 1,05 | 8,09 | 69,30 |
| 6 | 0,85 | 6,53 | 75,84 |
| 7 | 0,78 | 5,97 | 81,80 |
| 8 | 0,71 | 5,45 | 87,25 |
| 9 | 0,52 | 3,97 | 91,23 |
| 10 | 0,42 | 3,21 | 94,43 |
| 11 | 0,40 | 3,09 | 97,52 |
| 12 | 0,31 | 2,41 | 99,93 |
| 13 | 0,01 | 0,07 | 100,00 |

Fuente: Elaboración propia sobre la base de información provista por el OETR.

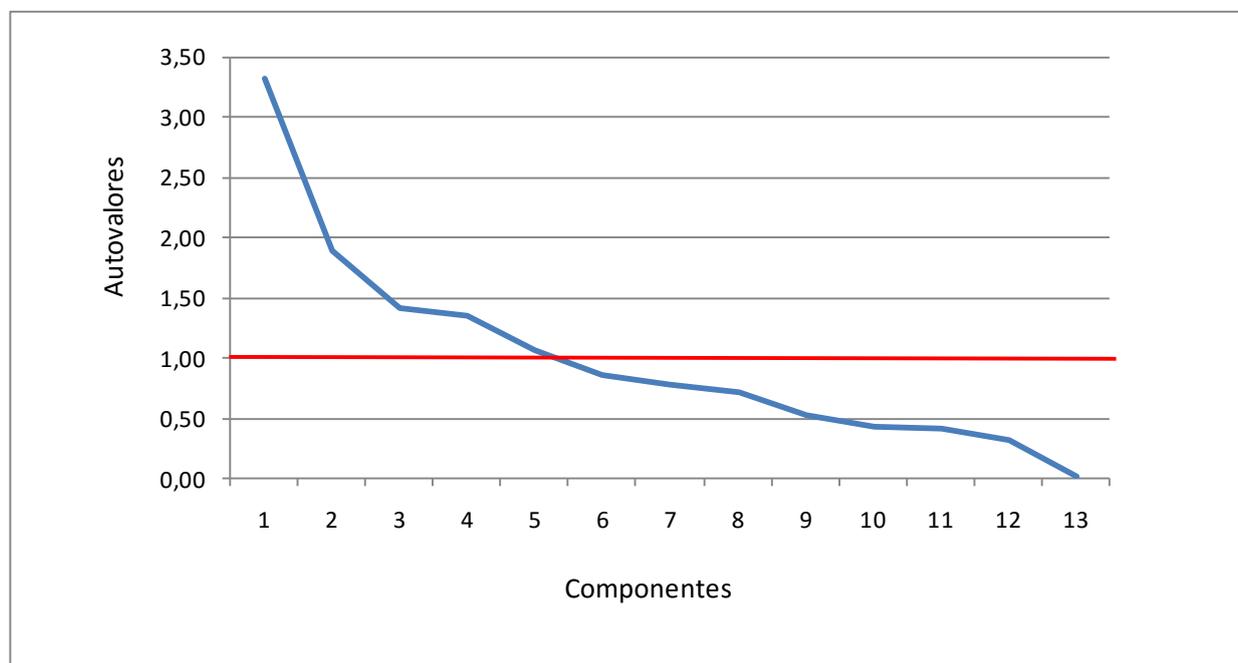
La determinación de la cantidad de componentes a retener (sobre los cuales se aplicará el algoritmo *k-medias*) puede basarse en diferentes criterios. Aquí se comentan tres de ellos, expuestos por Hair et al. (1999) y Peña (2002).

1) Establecer un porcentaje *acceptable* de la varianza explicada acumulada. En la literatura, hay consenso en establecer ese porcentaje por encima del 50% para las Ciencias Sociales. Según este criterio podrían retenerse los cuatro primeros componentes explicando el 61,21% de la varianza de los datos.

2) Aplicar el criterio de cambio abrupto. En el gráfico de sedimentación (Figura 1) se observa una pendiente abrupta entre la primera componente y la segunda. Luego la pendiente se hace más suave. Siguiendo este criterio se podrían seleccionar los dos primeros componentes y conservar el 40,04% de la varianza explicada.

3) Conservar aquellos componentes que posean autovalores iguales o mayores a 1. Siguiendo este criterio, habría que retener los primeros cinco componentes y conservar el 69,3% de la varianza explicada. Sin embargo, esta regla aplica solo para variables numéricas estandarizadas y no está claro en la literatura si debería aplicarse en el caso de variables mixtas.

Figura 1. *Gráfico de sedimentación correspondiente a los autovalores del ACP.*



Fuente: Elaboración propia sobre la base de información provista por el OETR.

Es importante destacar que, a medida que se agregan dimensiones al AC, los grupos resultantes poseen mayor varianza intragrupos; es decir, presentan mayor heterogeneidad. Por este motivo, la decisión de retener un componente más para introducirlo al AC debe tomarse teniendo en cuenta el

aporte que haga dicho componente a la conformación y a la interpretación de los grupos resultantes en contraposición a la mayor varianza intragrupos que genere.

Por lo expuesto, se optó por conservar los primeros cuatro componentes. En la Tabla 3, se presentan las cargas factoriales y a continuación se describen las componentes. El comando `dudi.mix` de R crea una variable para cada modalidad de las variables categóricas, tal como haría en un ACM.

Tabla 3. *Cargas factoriales de las primeras cuatro componentes ortogonales.*

| Variable | Componentes | | | |
|---|-------------|--------|--------|--------|
| | 1 | 2 | 3 | 4 |
| Gasto total | -0,884 | 0,023 | -0,147 | 0,245 |
| Gasto Diario per cápita | -0,614 | -0,557 | 0,343 | 0,054 |
| Cantidad de actividades realizadas | -0,343 | 0,549 | 0,066 | 0,099 |
| Cantidad de artículos comprados | -0,386 | 0,489 | 0,590 | -0,240 |
| Tamaño del grupo | 0,035 | 0,613 | -0,297 | 0,482 |
| Estadía | -0,616 | 0,275 | -0,501 | -0,187 |
| Lugar de residencia: Argentina | 0,344 | 0,163 | 0,069 | 0,098 |
| Lugar de residencia: Extranjero | -1,218 | -0,577 | -0,243 | -0,346 |
| Tipo de alojamiento: Sin cocina | -0,412 | -0,065 | 0,014 | 0,252 |
| Tipo de Alojamiento: Con cocina | 0,681 | 0,446 | -0,490 | -0,898 |
| Tipo de Alojamiento: No se hospedó | 1,763 | -0,995 | 1,688 | 0,732 |
| Medio de transporte: Automóvil | 0,394 | 0,321 | 0,241 | 0,899 |
| Medio de transporte: Avión | -0,921 | -0,393 | -0,099 | -0,249 |
| Medio de transporte: Ómnibus | 0,657 | 0,142 | -0,109 | -0,540 |
| Forma de organización del viaje: Contratación por medio de terceros | -0,226 | -0,186 | -0,017 | 0,787 |
| Forma de organización del viaje: Contratación directa | 0,084 | 0,069 | 0,006 | -0,293 |
| Compra de artículos elaborados en origen: NO | 0,170 | -0,239 | -0,250 | 0,087 |
| Compra de artículos elaborados en origen: SI | -0,801 | 1,120 | 1,176 | -0,410 |

Fuente: Elaboración propia sobre la base de información provista por el OETR.

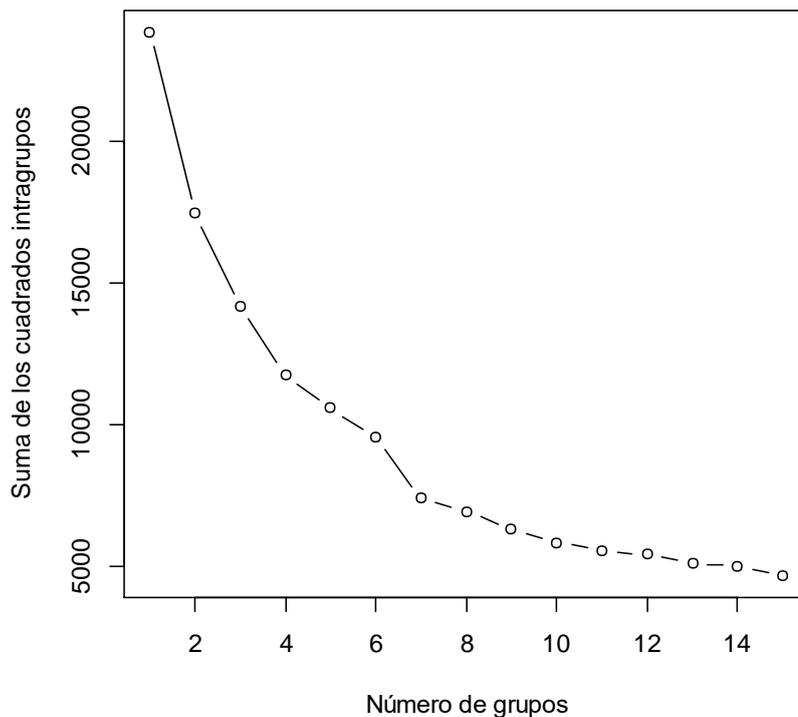
Las cargas factoriales presentadas en la Tabla 3 correspondientes a las variables cuantitativas, son los coeficientes de correlación lineal entre cada una de ellas y los componentes, mientras que las correspondientes a cada modalidad de las variables categóricas son el promedio de las cargas en cada componente de los individuos que presentan esa modalidad.

Respecto de las variables numéricas, la primera componente está muy correlacionada negativamente con las variables *Gasto total*, *Gasto diario per cápita* y *Estadía*. A su vez, este componente se vincula con la modalidad *Extranjero* de la variable Lugar de residencia, con la

modalidad *Avión* de la variable Medio de transporte y, débilmente, con la modalidad *Alojamiento sin cocina* de la variable Tipo de Alojamiento. En esta misma componente se asocia la modalidad *Argentina* de la variable Lugar de residencia, las modalidades *Ómnibus* de la variable Medio de transporte y las modalidades *Alojamiento con cocina* (esta modalidad se relaciona aún más con la cuarta componente) y *No se hospedó* de la variable Tipo de Alojamiento.

La segunda componente se correlaciona positivamente con la variable *Cantidad de actividades realizadas* y *Tamaño de grupo*. La tercera componente se correlaciona positivamente con *Cantidad de artículos comprados* y se asocia con la modalidad *SI* de la variable *Compra de artículos elaborados en origen*. A su vez esta componente se relaciona (aunque en forma débil) con la modalidad *NO* de la variable *Compra de artículos elaborados en origen*.

Figura 2. *Suma de los cuadrados intragrupos según cantidad de grupos aplicando el algoritmo k-medias a las 4 primeras componentes del ACP*



Fuente: Elaboración propia sobre la base de información provista por el OETR.

La cuarta componente se relaciona con la modalidad *Alojamiento sin cocina* de la variable *Tipo de Alojamiento*, con la modalidad *Automóvil* de la variable *Medio de transporte* y con las modalidades de la variable *Forma de organización del viaje*.

Para realizar el AC sobre los cuatro componentes retenidos, se debió determinar antes el número de grupos deseados, ya que es un requisito del algoritmo *k-medias*. Con este fin, se aplicó el algoritmo para distintas cantidades de grupos y se calculó en cada caso la suma de los cuadrados intragrupos para los componentes retenidos. En la Figura 2 se representan los resultados obtenidos para el caso de retener las cuatro primeras componentes.

Se observa que la suma de cuadrados intragrupos disminuye drásticamente al comienzo, luego disminuye marcadamente al pasar de 6 a 7 grupos y finalmente disminuye suavemente. De acuerdo con esta figura resulta razonable considerar una solución que tuviera siete grupos, ya que la suma de los cuadrados intragrupos disminuye muy poco más allá de siete grupos y se incrementarían los esfuerzos de interpretación de los mismos a la vez que disminuiría la robustez de la solución.

Etapa 3. Validación estadística de los grupos resultantes

La validación estadística implica hacer pruebas de consistencia de la lógica y de los resultados. Es decir, se debe repetir el ejercicio de clasificación de la Etapa 2 utilizando distintos centros iniciales y comparando los resultados finales. El supuesto es que la clasificación debe ser estable en el sentido que los resultados (los grupos formados) deben ser semejantes independientemente de los centros iniciales elegidos.

Por este motivo se repitió el procedimiento cuatro veces comenzando con centros aleatorios diferentes cada vez. Luego, se verificó si los turistas asistentes quedaban agrupados aproximadamente de la misma manera y se obtuvo una tasa de error (definida como el porcentaje de turistas asistentes que quedaban en diferentes grupos respecto del total de turistas asistentes) para todas las comparaciones posibles entre los cuatro agrupamientos obtenidos. Dichas tasas de error estuvieron entre 0,1% y 16,9%. Estos porcentajes se consideran se consideran bajos y por lo tanto la solución con 7 grupos es estable. En el Tabla 4 se presenta la composición de los grupos para esta solución.

Tabla 4. *Composición de los grupos resultantes de la clasificación de los turistas asistentes*

| Grupo | Cantidad de asistentes turistas | En % |
|-------|---------------------------------|--------|
| 1 | 221 | 7,4% |
| 2 | 523 | 17,5% |
| 3 | 404 | 13,5% |
| 4 | 549 | 18,3% |
| 5 | 414 | 13,8% |
| 6 | 192 | 6,4% |
| 7 | 693 | 23,1% |
| Total | 2.996 | 100,0% |

Fuente: Elaboración propia sobre la base de información provista por el OETR.

4. Resultados

A continuación, se caracterizan los siete grupos resultantes de la solución planteada. Cabe destacar que, en general, los grupos obtenidos mediante AC no reciben un nombre, sino un número. Cualquier esfuerzo por nombrar a los grupos necesariamente reduciría su caracterización a una o a unas pocas variables clasificatorias.

- **Grupo 1 (221 casos)**

Este grupo se caracteriza por estar conformado casi exclusivamente por residentes argentinos (99,5%) y por hospedarse por lo menos una noche, principalmente en un tipo de alojamiento con cocina (58,4%). La mayoría de los integrantes de este grupo contratan directamente los servicios de alojamiento y transporte (91%) y arriban a destino mediante transporte terrestre (Ómnibus 52%, Automóvil 44,3%). Sólo el 10% compra artículos elaborados en origen (el segundo porcentaje más alto entre todos los grupos) y la mayoría no realizan actividades turísticas (98,2%).

A su vez este grupo se caracteriza por tener los menores valores de gasto diario per cápita (la mediana es \$349, la más baja de todos los grupos), no realizar compras (75%) y viajar en grupos grandes (el 73% de los integrantes del grupo viaja en grupos de 3 o más).

- Grupo 2 (523 casos)

Este grupo posee el segundo porcentaje más alto de participación de extranjeros (31,9%). El 84,7% se hospedó en alojamientos sin cocina mientras que el 15,1 % optó por un tipo de alojamiento con cocina. La mayoría de los integrantes de este grupo contratan directamente los servicios de alojamiento y transporte (80,1%) y arriban a destino principalmente por avión (54,1%). El 86,2% de los integrantes de este grupo compra artículos elaborados en origen y el 19,3% realiza actividades turísticas (los porcentajes más altos de los 7 grupos).

En este grupo el gasto diario per cápita mediano es \$1.722, la tercera mayor en relación a todos los grupos. Todos los integrantes realizan compras de al menos un artículo (el 64% compra 2 ó más tipos de artículos) y la mayoría realiza por lo menos una actividad recreativa (68%). El 79,5% de los integrantes viaja solo o acompañado.

- Grupo 3 (404 casos)

Este grupo está integrado casi totalmente por residentes argentinos (97,5%) que se hospedan en alojamientos sin cocina (96%). La mayoría de los integrantes de este grupo contratan por medio de terceros los servicios de alojamiento y transporte (61,6%) y arriban a destino principalmente por automóvil (77%). Este grupo se caracteriza por tener turistas asistentes que no compra artículos elaborados en origen (98%) ni realiza actividades turísticas (99,3%).

En este grupo el gasto diario per cápita mediano es \$1.167. El 75% de los integrantes del grupo no realiza compras y la mayoría realiza por lo menos una actividad recreativa (71,5%). El 89% de los integrantes viaja acompañado.

- Grupo 4 (549 casos)

Este grupo posee el porcentaje más alto de participación de extranjeros (83,8%). El 94% de los integrantes del grupo se hospedó en alojamientos sin cocina y el 6% en alojamientos con cocina. El 29,9 % contrató por medio de terceros los servicios de alojamiento y transporte (el segundo mayor porcentaje). Casi la totalidad de los integrantes arribaron a destino principalmente por avión (96,2%). El 98,2% de los integrantes de este grupo no compra artículos elaborados en origen y el 6,2% realiza actividades turísticas (el segundo porcentaje más alto de los siete grupos).

Este grupo posee el mayor gasto diario per cápita mediano de todos los grupos (\$2.850). La mayoría de los integrantes no realiza compras (73,6%) ni actividades recreativas (65%). El 75,3% de los integrantes viaja solo.

- Grupo 5 (414 casos)

Este grupo se caracteriza por estar conformado casi exclusivamente por residentes argentinos (99,3%) y por hospedarse por lo menos una noche, principalmente en un tipo de alojamiento con cocina (94,4%). La mayoría de los integrantes de este grupo contratan directamente los servicios de alojamiento y transporte (88,6%) y arriban a destino mediante ómnibus (80%). Sólo el 4,3% compra artículos elaborados en origen y la mayoría no realizan actividades turísticas (99,3%).

Este grupo tiene un gasto diario per cápita mediano de \$668. Se caracteriza por no realizar compras (72,2%) ni actividades recreativas (75,8%) y por viajar sin acompañante (86%).

- Grupo 6 (192 casos)

Este grupo está integrado casi totalmente por residentes argentinos (97,9%) que no se hospedan (98,4%). La mayoría de los integrantes de este grupo contrata directamente los servicios de alojamiento y transporte (77,1%) y arriban a destino principalmente por automóvil (77%) u ómnibus (37%). Este grupo se caracteriza por tener turistas asistentes que no compra artículos elaborados en origen (97,4%).

Además, ningún integrante de este grupo realiza actividades turísticas (constituye el caso más extremo entre los siete grupos). Este grupo posee un gasto diario per cápita mediano es \$950. El 82% de los integrantes del grupo no realiza compras y la mayoría no realiza ninguna actividad recreativa (92%). El 81,7% de los integrantes viaja solo.

- Grupo 7 (693 casos)

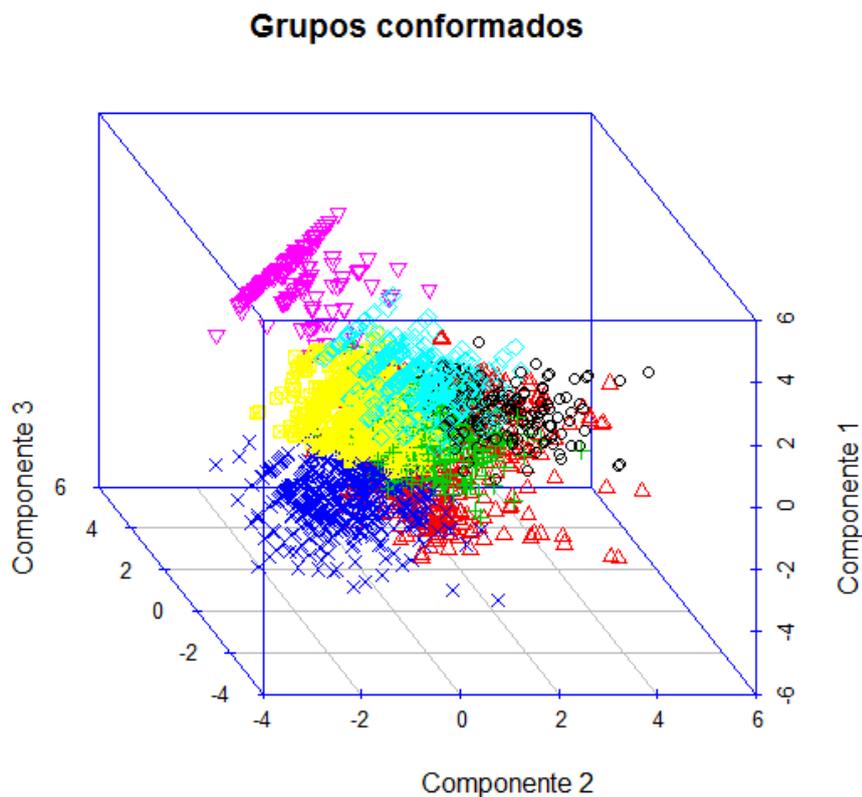
Este grupo está integrado casi totalmente por residentes argentinos (97,8%) que se hospedan en alojamientos sin cocina (98,4%). La mayoría de los integrantes de este grupo contrata directamente los servicios de alojamiento y transporte (73,4%) y arriban a destino principalmente por ómnibus (45,2%) o avión (29,3%). Este grupo se caracteriza por tener turistas asistentes que no compra artículos elaborados en origen (98,6%) ni realizan actividades turísticas (99,3%).

Este grupo posee un gasto diario per cápita mediano es \$1.767. El 44,6% de los integrantes del grupo realiza entre una y tres tipos de compras y el 33% realiza entre una y tres actividades recreativas. El 22% de los integrantes viaja acompañado.

Representación gráfica de los grupos

El diagrama de dispersión, tanto en su forma matricial como en su forma tridimensional (Figura 3), no permite visualizar apropiadamente todos los grupos cuando se poseen cientos o miles de casos y, además, se agrupa considerando más de tres variables. Esto sucede porque los puntos (que representan los casos) se superponen.

Por este motivo, se optó por utilizar un método de estimación de densidad para representar gráficamente los grupos resultantes en el espacio de las componentes. Se utilizó al paquete ks del software R que permite realizar estimaciones de densidad con el método no paramétrico de Kernel para tres dimensiones.

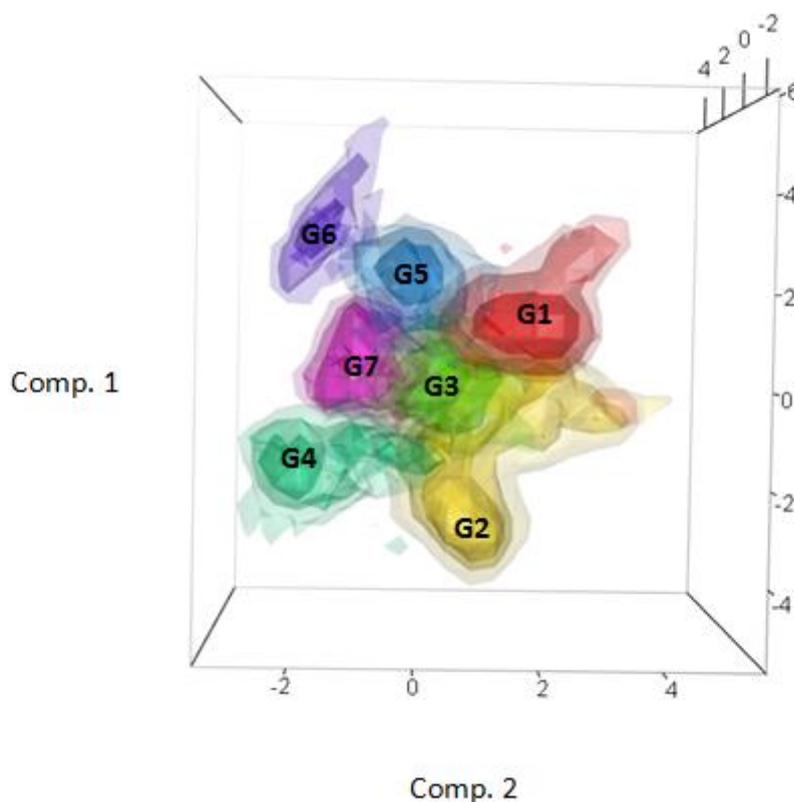
Figura 3. *Diagrama de dispersión de los turistas asistentes en las tres primeras componentes.*

Nota: Se indica la pertenencia a cada grupo con un color y una forma diferentes.

Elaboración propia sobre la base de información provista por el OETR.

En la Figura 4 se presenta la visualización de los siete grupos en las tres primeras componentes resultantes del ACP. Cada grupo está representado por un color. La figura muestra la probabilidad de encontrar en ese espacio un elemento del grupo (un turista asistente en este caso) y se favorece, principalmente, la visualización de las componentes 1 y 2. La intensidad del color refleja la densidad estimada: el tono más oscuro indica la menor superficie en que puede encontrarse el 25% de la densidad, el segundo tono indica el 50% y el tono más claro, el 75%.

Figura 4. *Representación de los siete grupos en el espacio de las tres primeras componentes del ACP en la que se favorece la visualización de las componentes 1 y 2.*



Nota: Representación según método no paramétrico de estimación de la densidad de Kernel.

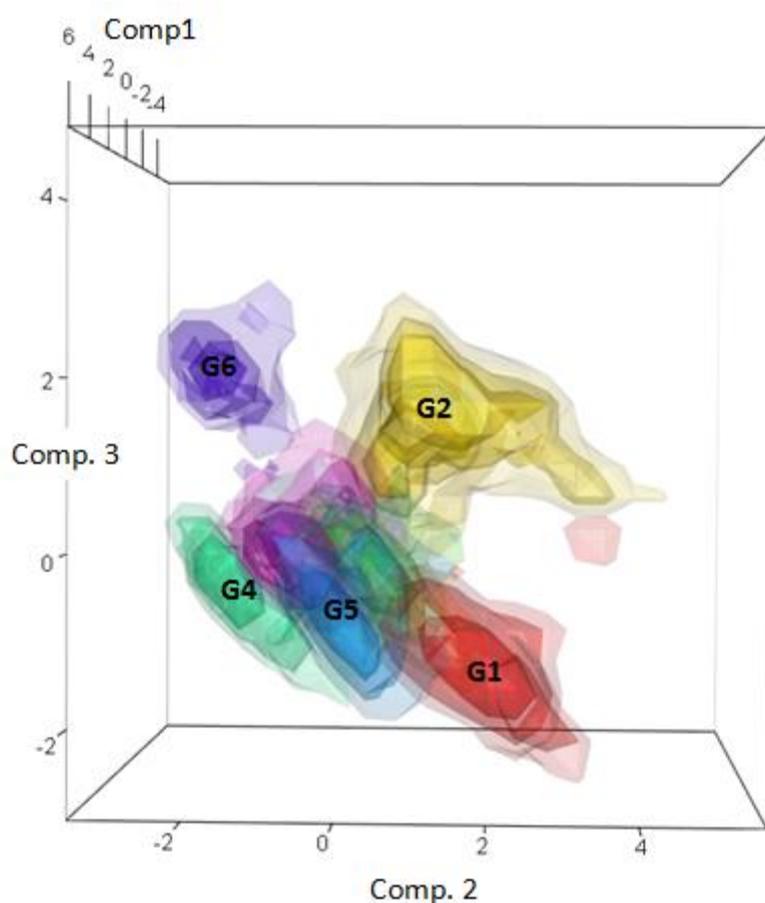
Elaboración propia sobre la base de información provista por el OETR.

Los grupos ubicados hacia los valores negativos del componente 1 tienen mayores gastos totales, mayores gastos diarios per cápita, mayores estadías, más integrantes que viajan en avión y se hospedan en alojamientos sin cocina. Los grupos ubicados hacia los valores positivos del componente tienen más integrantes residentes en Argentina, no se hospedan o lo hacen en alojamientos con cocina y utilizan medios de transporte terrestre (automóvil u ómnibus). Esto se corresponde con la caracterización realizada previamente. Los grupos G1, G5 y G6 poseen los menores valores de gasto mientras que los grupos G2 y G4 presentan mayores valores de gasto. A su vez los grupos G2 y G4

tienen estadías más largas y los grupos G1, G5 y G6 tienen los mayores porcentajes de integrantes residentes en Argentina.

Los grupos ubicados hacia los valores positivos de la componente 2 están integrados por turistas que realizan más cantidad de actividades recreativas y viajan en grupos de mayor tamaño. Así en los grupos G1, G2 y G3 los turistas asistentes suelen viajar con más acompañantes que en los grupos G4 y G6. Asimismo, los grupos G1 y G2 realizan más actividades recreativas que el resto de los grupos.

Figura 5. *Representación de los siete grupos en el espacio de las tres primeras componentes del ACP en la que se favorece la visualización de las componentes 2 y 3.*



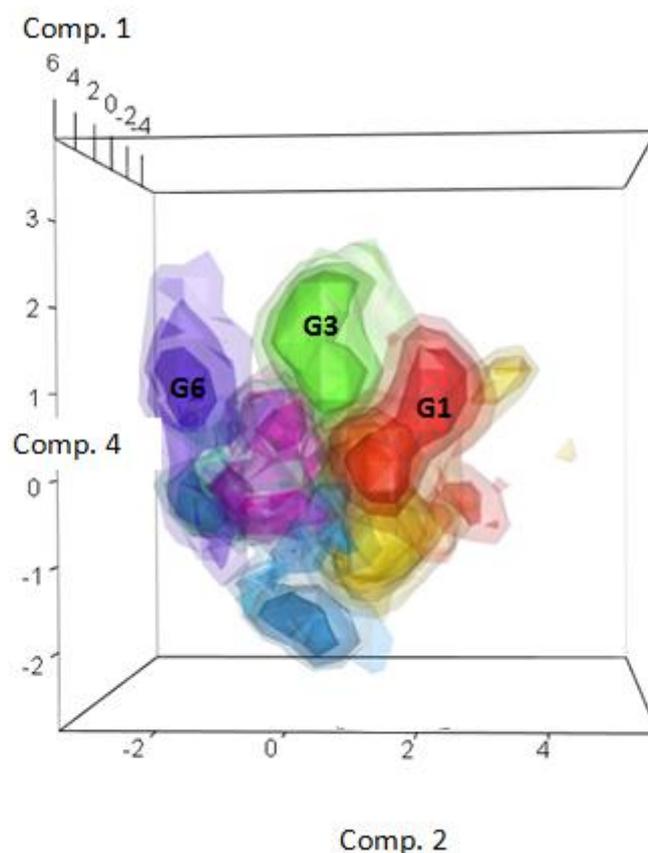
Nota: Representación según método no paramétrico de estimación de la densidad de Kernel.

Elaboración propia sobre la base de información provista por el OETR.

En cuanto a la componente 3 (ver Figura 5), los valores positivos reflejan la mayor cantidad de tipos de artículos comprados y se asocian a los compradores de artículos elaborados en origen. Así los grupos G1 y G2 se caracterizan porque sus integrantes compran este tipo de artículos (10,9% y 86,2% respectivamente).

En la Figura 6, los grupos ubicados en los valores positivos de la componente 4 poseen mayores porcentajes de asistentes turistas que viajan en automóvil, que contratan los servicios de alojamiento y transporte por medio de terceros y que se hospedan en alojamientos sin cocina. Así los grupos G1, G3 y G6 se ubican en la parte superior de la Figura 4. En particular se observa que en la dimensión de los componentes 1, 2 y 4, se distingue más el G3.

Figura 6. *Representación de los siete grupos en el espacio de las tres primeras componentes del ACP en la que se favorece la visualización de las componentes 2 y 4.*



Nota: Representación según método no paramétrico de estimación de la densidad de Kernel.

Elaboración propia sobre la base de información provista por el OETR.

Esta obra está bajo Licencia de Creative Commons Reconocimiento-NoComercial 4.0 Internacional.

Universidad Anáhuac de Oaxaca, Oaxaca, México; 2017.

5. Discusión de los resultados

El trabajo contribuye en dos aspectos. Por un lado, a la comprensión de la diversidad de asistentes turistas a través de la aplicación de métodos estadísticos multivariados. Por otro lado, incentiva el uso de estos métodos, a fin de que se incremente el número de investigaciones aplicadas que puedan poner a prueba los diversos postulados teóricos del turismo en general y del turismo de reuniones en particular.

Además, surgen dos grandes diferencias respecto a las investigaciones relevadas por Tkaczynski & Rundle-Thiele (2011.). La primera es el alcance del presente trabajo, que es mayor en cuanto a alcance geográfico, alcance temporal y cantidad de casos analizados. La segunda es su correspondiente encuadre dentro del campo de estudio del turismo de eventos definido por Getz (2007), puesto que los estudios relevados por el citado trabajo no diferenciaban asistentes turistas de asistentes no turistas.

Cabe mencionar que este trabajo se concentró en el tipo de evento Congresos. Se deja para futuras investigaciones la consideración de otros tipos de eventos como: Simposios, Seminarios, Ferias, Exposiciones, Eventos deportivos, entre otras.

Se destaca que la metodología propuesta es satisfactoria en tanto se alcanza el objetivo planteado de lograr una buena clasificación de los asistentes turistas, en términos de la homogeneidad intragrupos y la heterogeneidad entre grupos.

En la bibliografía se señala al AC como un método objetivo de clasificación, pero en el desarrollo del trabajo se reconoce la presencia de la subjetividad del investigador al elegir las variables iniciales, las componentes a retener y la solución elegida dentro del conjunto de soluciones posibles (cantidad de grupos). En este trabajo se optó por una solución con siete grupos en vez de seis debido a la baja en la varianza intragrupos. Sin embargo, no hay reglas estrictas para definir la cantidad de grupos y se espera que una validación técnica a futuro ayude en ese sentido. En otras palabras, los grupos de asistentes turistas serán confiables en la medida que un actor clave o referente pueda reconocerlos y acepte que hay motivos para pensar que existen diferencias entre ellos.

6. Agradecimientos

Al equipo técnico del Observatorio de Turismo de Reuniones de la República Argentina, que cedieron los datos para la realización de este trabajo.

7. Referencias

- Ahmad, A. & Dey, L. (2011). A K-Means Type Algorithm for Subspace Clustering of Mixed Numeric and Categorical Datasets. *Pattern Recognition Letters*, 32, 1062-1069.
- Anderberg, M.R. (1973). *Cluster Analysis for Applications*. Academic Press.
- Berdegúe, J.; Escobar, G. (1990). *Tipificación de Sistemas de Producción Agrícola. RIMISP*. Chile, p. 13-44.
- Chatzis, S. P. (2011). A Fuzzy C-Means-Type Algorithm for Clustering of Data with Mixed Numeric and Categorical Attributes Employing a Probabilistic Dissimilarity Functional. *Expert Systems with Applications*, 38, 8684-8689.
- Chian Hsu, C.; Long Chen, C. & Wei Su, Y. (2007). Hierarchical Clustering of Mixed Data Base with Distance Hierarchy. *Information Sciences*, 177, 4474-4492.
- Dray, S.; Dufour, A. B. (2007). The Ade4 Package: Implementing The Duality Diagram for Ecologists. *Journal of Statistical Software*, 22 (4).
- Duong, T. (2007). Ks: Kernel Density Estimation and Kernel Discriminant Analysis for Multivariate Data in R. *Journal of Statistical Software*, 21 (7).
- Escobar, M. (1998). Las aplicaciones de análisis de segmentación: el procedimiento CHAID. *Revista Metodología de las Ciencias Sociales*, 1, 13 – 49.
- Everitt, B.; Landau, S. & Leese M. (2001). *Cluster Analysis*. Londres. Editorial Arnold, 2001.
- Getz, D. (2005). *Event management and event tourism* (2nd ed.). New York: Cognizant.
- Getz, D. (2008). Event Tourism: Definition, Evolution and Research. *Tourism Management*, 29(3), 403-428.
- Getz, D. (2012). *Event Studies: Theory, Research and Policy for Planned Events* (2nd ed.). Abingdon: Routledge.
- González, P. R., & Molina, Ó. M. (2007). La segmentación de la demanda turística española. *Metodología de encuestas*, 9(1), 57-92.
- Hair, J.F.; Anderson, R.E; Tatham, R.L. (1999). *Análisis Multivariante*. España: Prentice Hall.
- Hill, M.O.; Smith, A. J. E. (1976). Principal Component Analysis of Taxonomic Data with Multi-State Discrete Characters. *Taxon*, 25, 249-255.
- Horner, S., & Swarbrooke, J. (2016). *Consumer Behaviour In Tourism*. Routledge.
- Huang, Z. (1997). *Clustering Large Data Sets with Mixed Numeric and Categorical Values*. Australia: CSIRO Mathematical and Information Sciences.
- Köbrich, C. Rehman, T.; Khan M. (2003). Typification of Farming Systems for Constructing Representative Farm Models: Two Illustrations of the Application of Multivariate Analyses in Chile and Pakistan. *Agricultural Systems*, 76, 146-157.

- Kotler, P.; García De Madariaga M., J.; Flores Zamora, J.; Bowen, J. T. & Makens, J. C. (2013). *Marketing turístico* (5a ed.). Madrid: Pearson
- Lebart, J.L.; Morineau, A. & Piron, M. (1995). *Statistique exploratoire multidimensionnelle*. Dunod.
- Lebart, L.; Morineau, A & Warwick, K. (1984). *Multivariate Descriptive Statistical Analysis*. US: John Wiley & Sons, Inc.
- Page, S. & Connell, J. (Eds.). *Routledge Handbook of Events*. London: Routledge.
- Peña D. (2002). *Análisis de datos multivariantes*. España: Mac Graw-Hill.
- R Core Team (2012). *R: A Language and Environment for Statistical Computing*. Viena, Austria: R Foundation for Statistical Computing. Recuperado de: <http://www.R-project.org/>.
- Shaw, G., Agarwal, S. & Bull, P. (2000). Tourism Consumption and Tourist Behaviour: A British Perspective. *Tourism Geographies*, 2(3), 264-289.
- Talaya, Á. E. & Lara, E. M. R. (1996). La investigación de la demanda turística en España: recopilación y análisis. *Estudios turísticos*, (129), 81-104.
- Tang, Q., & Turco, D. M. (2001). Spending Behaviors of Event Tourists. *Journal of Convention & Exhibition Management*, 3(2), 33-40.
- Tkaczynski, A., & Rundle-Thiele, S. R. (2011). Event Segmentation: A Review and Research Agenda. *Tourism Management*, 32(2), 426-434.
- UNWTO (2016). *UNWTO Tourism Highlights*. Madrid: UNWTO.
- Young, F. W. (1981). Quantitative Analysis Of Qualitative Data. *Psychometrica*, 4, 357-388.
- Young, F.W.; Takane, Y. & Leeuw, J. (1978). The Principal Components of Mixed Measurement Level Multivariate Data: An Alternating Least Squares Method with Optimal Scaling Features. *Psychometrica*, 2, 279-281.

Para citar este artículo (estilo APA):

Dowbley, V. & De la Cruz, G. M. (2017). Clasificación de turistas asistentes a reuniones utilizando análisis multivariado con variables mixtas. *Transitare*, 3, 112-138