

El análisis cluster y el marketing: una relación trascendental

Cluster analysis and marketing: a transcendental relationship

*Ms C. Jensy Tanda Díaz, jensy.tanda@umcc.cu, Lic. Yanlis Rodríguez Veiguela, Dr C. Manuel Marrero Marrero. Universidad de Matanzas "Camilo Cienfuegos".
Dr. C. Roberto C. Pons García. Universidad Central "Marta Abreu" de las Villas.*

Resumen

Los investigadores necesitan conocer más acerca de las bondades que brindan las técnicas estadísticas como herramientas para la consecución de los objetivos de sus investigaciones. El objetivo principal de este estudio es desarrollar la técnica del análisis cluster desde un abordaje teórico y práctico vinculando su aplicación al marketing, a la investigación y segmentación de mercados. Por lo trascendental y engorroso que resultan tales estudios en ciudades, los autores han seleccionado la aplicación del cluster para determinar los públicos objetivos de la ciudad de Matanzas, Cuba a partir de la evaluación de los atributos de imagen urbana.

Palabras claves: Análisis Cluster, Marketing, Segmentación de Mercado, Públicos Objetivos, Imagen Urbana.

Abstract

The investigators need to know more about the kindness that offer the statistical techniques as tools for the attainment of the objectives of their investigations. The main objective of this study is to develop the technique of the cluster analysis from a theoretical and practical boarding linking its application to the marketing, to the investigation and segmentation of market. For the important and annoying that its are such studies in cities, the authors have selected the application of the cluster to determine the objective publics of the city of Matanzas, Cuba starting from the evaluation of the attributes of urban image.

Keywords: Cluster Analysis, Marketing, Segmentation of Market, Objective Publics, Urban Image.

Desde hace algunos años los investigadores han reconocido la necesaria comprensión de la estadística como una herramienta para el desarrollo de la ciencia, muestra de ello son los métodos multivariados como los casos de la regresión y la correlación múltiple, el análisis discriminante múltiple, el de los componentes principales y el análisis factorial común, el análisis multivariante de varianza y covarianza, la correlación canonical, el análisis multidimensional y el análisis cluster, el cual será objeto de análisis en este artículo.

Esta técnica estadística es la denominación de un grupo de otras multivariantes cuyo principal propósito es la comparación y agrupación de objetos basándose en el valor y se ha denominado como análisis Q, construcción de tipología, análisis de clasificación y taxonomía numérica. Es una de las más utilizadas para

la formación de "tipos", tiene como propósito el análisis de conglomerados que no es más que agrupar las observaciones y clasificar objetos de forma que sean muy homogéneos dentro de los grupos (mínima varianza) y que sean lo más heterogéneo posible entre ellos (máxima varianza), con la que se puede comprender a los mismos y a la población de la que proceden.

Existe una coincidencia marcada entre diversos autores como Hair, J. F.; Anderson, R.E.; et al. (1999); Johnson, D. E. (1998); Moreno, N.L (2006) en los diferentes pasos a tener en cuenta para desarrollar el análisis cluster, a partir de los cuales los autores del presente artículo realizan la siguiente propuesta del análisis cluster combinado:

1. Selección de las variables del análisis cluster: los objetivos del análisis cluster no pueden separarse de la selección de las variables utilizadas para

caracterizar los objetos a agrupar, sin importar que el objetivo sea exploratorio o confirmatorio. La selección de las variables a incluir en el valor teórico del análisis cluster debe hacerse con relación a consideraciones teóricas, conceptuales y prácticas, y se basará en la lógica por la importancia de incluir solo aquellas variables que caracterizan realmente los objetos que se están agrupando, y se refieren específicamente a los objetivos del análisis. En esta investigación las variables a utilizar como criterio de selección son los atributos de imagen.

2. Selección del algoritmo cluster:

En primer lugar se debe establecer la medida de similitud, ya que la similitud es una medida de correspondencia, o parecido, entre los objetos que van a ser agrupados. El método que se propone utilizar es el de las medidas de distancia, ya que estas representan la similitud como la proximidad de las observaciones respecto a las otras, ya que en realidad estas son medidas de diferencia, donde los valores elevados indican una menor similitud. En la investigación se utilizará la distancia euclídea cuadrada, o absoluta, que es la suma de las diferencias al cuadrado sin tomar la raíz cuadrada. Es significativo destacar que esta última es la medida de distancia recomendada para los métodos de análisis cluster del centroide y de ward.

En segundo lugar se establecerá el método cluster a realizar, se propone su empleo combinado para obtener los beneficios de los métodos jerárquico y no jerárquico. En un primer momento, la técnica jerárquica puede establecer el número de conglomerados, los perfiles de los centros de conglomerados y la identificación de cualquier atípico obvio. Una vez que se han eliminado estos, las observaciones restantes pueden ser agrupadas mediante el método no jerárquico, estableciendo como dato de entrada el número de clusters resultantes del método jerárquico.

3. Reespecificación de los cluster obtenidos: el análisis cluster es sensible a los atípicos, los cuales representarán tanto observaciones verdaderamente "aberrantes", que no son representativas de la población en general; o una muestra reducida del grupo (s) de la muestra; en ambos casos, estos distorsionan la verdadera estructura de la población. Por lo que una vez identificados los atípicos independientemente del método utilizado, estas observaciones deben analizarse y evaluarse a efectos de su representatividad en la población y eliminarlos del análisis si se consideran no representativos.

4. Interpretación de los cluster obtenidos: este paso implica el examen de cada conglomerado en términos de su valor teórico y la asignación de una

etiqueta precisa que describa su naturaleza para lo que se realiza la interpretación de los grupos, mediante el etiquetado de los cluster, con la ayuda de la tabla de los centroides finales que se obtiene mediante el uso del programa estadístico SPSS 11.5, cuando se ejecuta el no jerárquico K-medias. Estos centroides permiten conocer cuáles fueron las mejores valoraciones promedio realizadas a los diferentes indicadores analizados, lo que consecuentemente hace posible identificarlos a través de una etiqueta.

5. Validar los resultados de la aplicación del análisis cluster: es importante asegurarse de su relevancia práctica y como validación una de las vías es utilizar otro método para agrupar datos como la correlación canónica, la regresión múltiple, la chi-cuadrado y el análisis discriminante que es el más utilizado, utilizar el propio análisis cluster para muestras distintas es muy costoso. A partir de lo anterior se selecciona el análisis discriminante, su objetivo y forma de aplicación será brevemente descrito a continuación:

- El análisis discriminante: persigue una correcta clasificación de las observaciones en grupos predeterminados con antelación, ya que se debe contar con un conjunto de variables que describan a cada una de las observaciones o individuos, además de una variable dependiente de las anteriores que determine la correcta clasificación excluyente de todas y cada una de las variables en grupos o conjuntos distintos. El objetivo principal es predecir la probabilidad de pertenencia de una observación a un grupo determinado, al igual que distinguir cuáles son las variables que presentan la mayor diferenciación de los individuos entre grupos, mediante la creación de una función discriminante, que realiza la predicción de pertenencia de nuevos individuos a los grupos establecidos.

El criterio de selección de las variables que se utilizará es el de wilks lambda, que plantea la selección de las variables en las funciones discriminantes, y se lleva a cabo mediante la maximización del cociente entre el determinante de la varianza intergrupos y el de las varianzas intragrupos (F), que es una medida recíproca a la lambda de wilks. Las funciones discriminantes obtenidas o factores, serán ortogonales entre sí.

6. Descripción de los perfiles de los cluster obtenidos: describir los perfiles de los cluster resultantes se logran a partir de la descripción de las características de cada conglomerado, lo que implica el uso del análisis discriminante, utilizando para ello datos no incluidos previamente en el análisis cluster, estos normalmente son variables de clasificación (sociodemográficas, socioculturales, etc.). Para describirlos, se toma a los conglomerados o factores (clusters resultantes

del análisis de K-medias) como variables dependientes y se comparan cada uno con las independientes, resaltándose en cada segmento aquellas características que lo distinguen de los otros.

La investigación de mercado y el análisis cluster

La investigación de mercados ayuda a identificar diferentes problemas y oportunidades, así como a evaluar y desarrollar alternativas de acción, teniendo en cuenta las necesidades y deseos del mercado. En el caso de las ciudades, contribuye a una mejor comprensión del entorno, para lo cual se auxiliará de una correcta segmentación de mercado, ya que a decir de Santesmases, M. (2004) esta va a permitir establecer prioridades en el proceso de planificación.

Kotler, P. (2000) resalta que la segmentación junto con el posicionamiento van a ser el corazón del moderno marketing estratégico, mientras que para Vázquez, C. R.; Trespalacios, J. A. (1994) es uno de los conceptos más trascendentes del marketing moderno. Ellos consideran que para que la segmentación de mercados sea más efectiva desde el punto de vista estratégico tendrá como primera etapa la identificación de los segmentos de mercados, en un segundo momento la selección del mercado meta, dejando en estas etapas las condiciones preparadas para la proyección estratégica de marketing.

En la actualidad numerosas investigaciones de mercado se centran en la búsqueda de las necesidades y deseos de las personas a nivel de sus diversos públicos, en el caso particular de la ciudad, esto llegan a alcanzar connotaciones altas, por el tipo amplio y diverso de necesidades y deseos que habría que interpretar para luego satisfacer y delimitar la imagen de la ciudad, que no es más que la suma de creencias, ideas e impresiones que una persona o un grupo tiene de la urbe (Friedmann, R. 2005); constituiría uno de los puntos de partida para la planificación del futuro, es por ello que el análisis cluster se vuelve una herramienta eficaz para el logro de estos fines, ya que es usado precisamente para la segmentación de los mercados; comprensión del comportamiento del comprador (identificación de grupos de compradores homogéneos, para analizar el comportamiento de cada grupo por separado); identificar oportunidades para nuevos productos, seleccionar mercados de prueba, reducción de datos con el fin de facilitar el manejo conveniente de la información, entre otros aspectos.

Igualmente la selección adecuada de los grupos objetivos que se pretendan alcanzar o lograr, se hace sin perder de vista su situación actual y sus criterios acerca del atractivo de una ciudad respecto a las

otras, es determinante. Sanz de la Tajada, L. A. (1994) afirmaría que son, por lo tanto, individuos que se conocen en función de su perfil, al que se llega a partir de características de diversos tipos, entre las que caben destacar las siguientes: sociodemográficas –el sexo, la edad, la región en que habitan, etc.–, socioculturales –la clase social, el nivel de estudios, la profesión u ocupación, etc.–, socioeconómicas –el nivel de ingresos, la situación patrimonial, la estructura del gasto, etc.–, sociopolíticas –la afiliación o simpatía por uno u otro partido o central sindical, su peso como autoridad pública, etc.–; y psicológicas –las actitudes, los factores de la personalidad, los estilos de vida, etc.

Por lo que el análisis cluster teniendo en cuenta todo lo abordado en el presente artículo y las bondades que permite en la investigación de mercados, su uso se ha desarrollado ya no solo en términos empresariales, sino que se ha diversificado su aplicación a otros sectores como la ciudad; lo que será ejemplificado a través de su aplicación en la ciudad de Matanzas, Cuba, en el caso del análisis cluster combinado. (Ver Diagrama 1).

Definición de los públicos objetivos de la ciudad de Matanzas a partir de la evaluación de los atributos de imagen urbana

El instrumento aplicado fue una encuesta a través de cuestionarios, la ficha técnica de la misma se encuentra en la Tabla 1 de los anexos. Las variables de selección fueron: arquitectura y urbanismo, patrimonio histórico, oferta cultural, actitud de la ciudad, aspecto físico, prestigio de la ciudad, servicios, transporte e infraestructura de comunicación y medio ambiente; el comportamiento de las mismas durante el procedimiento se encuentran en las Tablas 2,2.1 y 2.2 de los anexos.

La validación de los conglomerados se realiza mediante el análisis discriminante, teniendo en cuenta las variables no utilizadas a la hora de formarlos; pero que se conoce que cambian.

Se utilizarán las variables sociodemográficas y socioculturales como variables independientes, mientras que como variable categórica dependiente se introduce el conjunto de conglomerados resultante del cluster. La utilización del discriminante con fines de reclasificación arrojó que la clasificación realizada por el análisis cluster combinado fue correcta con 94.2% de los casos originalmente agrupados correctamente, validando así el método.

Conglomerado 1: Críticos optimistas moderados

Está conformado por el 20.55% de los encuestados, se caracteriza por ser de evaluaciones moderadas,

que peor evalúan la arquitectura y urbanismo y la oferta cultural de la ciudad.

El conglomerado está compuesto mayoritariamente por encuestados adultos (50) entre 50-64 años, para un 43,10% del mismo. Igualmente, predominan los residentes del Consejo Popular Naranjal (32) para un 27,59%. Referente al nivel educativo, destacan los bachilleres (67), con un 57,76% del conglomerado total. Mientras que respecto a los años de residencia, está compuesto por un 81,90% de encuestados con más de 20 años de residencia en la ciudad, por lo que resulta el segundo conglomerado que mayor vivencia tiene en la urbe.

Conglomerado 3: Críticos moderados

Está conformado por el 24,83% de los encuestados, y se caracteriza por ser el más moderado y equilibrado en sus evaluaciones, lo que se evidencia en que peor evalúan solamente el patrimonio histórico, y que solo también son los segundos más críticos en lo que respecta al componente emocional de la imagen (actitud y prestigio). Mientras, que en lo positivo, solo son los mejores al evaluar el aspecto físico de la ciudad que el resto de los conglomerados y segundos en la evaluación del transporte e infraestructura.

Está compuesto mayoritariamente por encuestados comprendidos en edades entre 35-49 años (64) para un 44,14% y entre 20-34 años (50) para un 34,48%, que representan el 78,62% del conglomerado total. Igualmente, predominan los residentes en Matanzas Oeste (39) y Versalles (38), para el 26,90% y el 26,20% respectivamente. Referente al nivel educativo, destacan los bachilleres (70), para un 48,28% y los universitarios (62) para un 42,76%. Mientras que respecto a los años de residencia en la ciudad, priman los comprendidos entre 13-19 años (71), para un 48,97%, y en segundo lugar, los mayores de 20 años (65), para un 44,83%.

Conglomerado 4: Críticos pesimistas moderados

Lo conforma el 17,47% de los encuestados, y se caracteriza por ser de evaluaciones negativas, aunque estos se consideren moderados en comparación con las evaluaciones negativas del conglomerado 2 (críticos pesimistas), por lo que el 4 es el segundo que peor evalúa la imagen de la ciudad. Son los más críticos en evaluar la arquitectura y urbanismo, la oferta cultural, el prestigio y el transporte e infraestructura de comunicación de la ciudad. También son los segundos peores evaluadores del patrimonio histórico, de los servicios y del medio ambiente.

El conglomerado está compuesto mayoritariamente por encuestados mayores de 65 años (39), para un 38,24%, y por los comprendidos entre 50-64 años

(34) para un 33,33%, representando un 71,57% de encuestados mayores de 50 años, por lo que resulta el conglomerado de edades más avanzadas. Igualmente, predominan los residentes del Consejo Popular Peñas Altas (29) para un 28,48%. Referente al nivel educativo, destacan los de nivel secundario (49) para un 48,04% del total. Mientras que respecto a los años de residencia, está compuesto por un 91,18% de encuestados con más de 20 años de residencia en la ciudad, por lo que resulta el que mayor vivencia tiene de la misma.

Conglomerado 5: Críticos optimistas

Está conformado por el 17,29% de los encuestados, y se caracteriza por ser el de mejores evaluaciones. Son los menos críticos al evaluar aspectos concernientes al diseño urbano de la ciudad como su arquitectura y urbanismo, su patrimonio histórico, así como los relacionados con la oferta cultural y de servicios, además de su transporte e infraestructura y su medio ambiente. También son los segundos que mejor evalúan los componentes emocionales de la imagen (actitud y prestigio de la ciudad) y su aspecto físico.

Se destaca por ser el que agrupa más jóvenes, por lo que existe predominio de encuestados entre 20-34 años (58), para un 57,43%. Donde son mayoría los residentes en los consejos Playa (24), para un 23,76% y Matanzas Este (22), para un 21,78%. Referente al nivel educativo, destaca por ser el de mayor instrucción, con un 54,45% de universitarios y un 39,60% de bachilleres. Mientras que respecto a los años de residencia en la ciudad, priman los de menos de 6 años viviendo en la misma (46), para un 45,54% siendo el conglomerado de menor vivencia de la ciudad.

Conclusiones

Las técnicas estadísticas, y en particular el análisis cluster, constituyen una herramienta vital para alcanzar los resultados propuestos por el marketing. En el caso particular de este estudio permitió determinar los públicos objetivos de la ciudad de Matanzas, Cuba a partir de la evaluación de los atributos de imagen urbana.

Bibliografía

- Friedmann, R. 2005. Marketing estratégico y participativo de ciudades. [En línea]. [Consulta: Febrero, 2006]. Disponible en: <http://www.redinternacional.org.mx>
- Hair, J. F.; Anderson, R.E... et al. 1999. Análisis multivariante. Madrid: Prentice-Hall Iberia. p. 302.
- Johnson, D. E.1998. Métodos multivariados aplicados al análisis de datos. México: Editorial International Thomson Editores. p. 351.

Kotler, P... et al. 2000. Dirección de marketing. La edición del milenio. México: Prentice-Hall. p. 223.
 Moreno Delgado, N.L. 2006. Modelo para la definición y evaluación del concepto de productos turísticos culturales. Estudio en el contexto Varadero-Matanzas. Matanzas. Tesis en opción al grado científico de Doctor en Ciencias Económicas. p. 53.
 Santesmases, M. 2004. Marketing. Conceptos y estrategias. Madrid: Editorial Pirámide. p. 85.

Sanz de la Tajada, L. A. 1994. Integración de la identidad y la imagen de la empresa, Desarrollo conceptual y aplicación práctica. Madrid: ESIC Editorial. 189 p.
 Vázquez, C. R.; Trespalacios, J. A. 1994. Marketing: estrategias y aplicaciones sectoriales. Madrid: Editorial Civitas. p. 125.

Tabla 1: Ficha Técnica

Ficha Técnica	
Procedimiento Metodológico	Encuestas a través de cuestionarios con preguntas abiertas, cerradas, de contenido y control.
Universo	Personas de ambos sexos, mayores de 20 años de edad, que residen en la ciudad de Matanzas, la misma fue estimada en 95 920 ciudadanos.
Tamaño muestral	624 encuestas válidas realizadas en 7 puntos diseminados representativamente por toda la ciudad de Matanzas.
Error muestral	± 4% para un supuesto de máxima variabilidad (P=Q).
Nivel de confianza	95,45%, que sigue una distribución normal Z= 1,96
Diseño muestral	Estratificado, con asignación proporcional. La muestra definida se disemina proporcionalmente por los barrios seleccionados.

Fuente: Elaboración propia.

Tabla 2: Centros de los conglomerados finales

	Conglomerado				
	1	2	3	4	5
Arquitectura y urbanismo	2,80	2,39	2,53	1,72	3,04
Patrimonio Histórico	2,42	1,96	2,53	1,71	3,17
Oferta Cultural	2,91	2,54	2,89	1,62	3,23
Actitud de la ciudad	3,74	2,85	2,88	3,13	3,21
Aspecto físico	2,51	2,14	3,34	2,61	3,11
Prestigio de la ciudad	3,76	3,08	3,02	3,00	3,28
Servicios	2,36	1,54	2,12	1,60	3,07
Transporte e infraestructura	1,84	1,56	1,95	1,56	2,42
Medio ambiente	2,73	1,49	2,47	2,21	2,82

Fuente: Salida SPSS

ANOVA

	Conglomerado		Error		F	Sig.
	Media cuadrática	gl	Media cuadrática	gl		
Arquitectura y urbanismo	25,895	4	,204	579	126,544	,000
Patrimonio Histórico	33,183	4	,317	579	104,754	,000
Oferta Cultural	37,819	4	,261	579	144,961	,000
Actitud de la ciudad	15,900	4	,276	579	57,782	,000
Aspecto físico	26,125	4	,321	579	90,787	,000
Prestigio de la ciudad	12,101	4	,273	579	44,352	,000
Servicios	48,570	4	,312	579	131,188	,000
Transporte e infraestructura	13,004	4	,177	579	73,421	,000
Medio ambiente	32,365	4	,380	579	85,189	,000

Recibido: 13-1-10
 Aceptado: 11-9-10

Tabla 2.1: Distancias entre los centros de los conglomerados finales

Conglomerado	1	2	3	4	5
1		2,052	1,503	2,292	1,570
2	2,052		1,804	1,481	2,880
3	1,503	1,804		1,887	1,554
4	2,292	1,481	1,887		3,169
5	1,570	2,880	1,554	3,169	

Las pruebas F solo se deben utilizar con una finalidad descriptiva puesto que los conglomerados han sido elegidos para maximizar las diferencias entre los casos en diferentes conglomerados. Los niveles críticos no son corregidos, por lo que no pueden interpretarse como pruebas de la hipótesis de que los centros de los conglomerados son iguales.

Fuente: Salida SPSS.

Tabla 2.2: Número de casos en cada conglomerado

Conglomerado	
1	120,000
2	116,000
3	145,000
4	102,000
5	101,000
Válidos	584,000
Perdidos	,000

Fuente: Salida SPSS