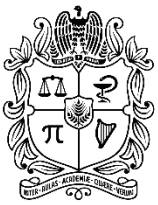
PROPUESTA DE TRABAJO DE GRADO



UNIVERSIDAD
NACIONAL
DE COLOMBIA

TITULO: Análisis De Clúster Automático
TUTOR: Juan David Velasquez Henao
DIRECTOR PROPUESTO: Juan David Velasquez Henao

AUTOR:

Marisol Correa Henao	CC. 1017230592
mcorreah@unal.edu.co	TEL. 3196683034

PROGRAMA: Maestría En Ingeniería Analítica **PERFIL:** Profundización

Universidad Nacional de Colombia - Sede Medellín
Facultad de Minas
Medellín
2019

TABLA DE CONTENIDO

RESUMEN EJECUTIVO

En este documento se desarrollará el proceso de software de análisis de clúster automático, aunque hoy en día existen varias librerías que permiten realizar análisis de clúster se busca automatizar el proceso y lograr diferentes opciones de clúster centralizadas en un mismo paquete, facilitando el análisis y la parametrización de los modelos, para ello se usarán las librerías ya existentes en python, tomando como base lo que se tiene para el software SAS, de manera que se puedan usar tanto por una persona con conocimientos básicos como por una persona con conocimientos profundos que quiera parametrizar sus análisis.

Palabras Clave: Análisis, Cluster, Software, Python, Librería, aprendizaje de máquinas automático

ABSTRACT

In this document, the automatic cluster analysis software process will be developed, although today there are several libraries that allow you to perform cluster analysis, you want to automate the process and achieve different centralized cluster options in the same package, facilitating analysis and parameterization of the models, so that the libraries could already be used and python, based on what it has for SAS software, so that it can be used both by a person with basic knowledge and by a person with deep knowledge who wants Parameterize your analysis.

Keywords: Analysis, Cluster, Software, Python, Packages, automatic machine learning

1. INTRODUCCIÓN

En este documento se encontrará una breve descripción de las diferentes técnicas de análisis de agrupamiento o análisis de clúster, además de las metodologías a utilizar para el aprovechamiento de los datos, incluyendo técnicas de clúster para grandes volúmenes de datos, de manera que se logre la automatización de los procesos luego se hace una descripción de los paquetes existentes en diferentes software y lenguajes de programación y ventajas, limitaciones o dificultades que se encuentran en cada uno y los elementos que podrían facilitar el proceso de detección de grupos, por último se seleccionará la metodología de interés y se presenta las técnicas a utilizar que permitirán un proceso automatizado cuyo producto final sea una librería que se pueda implementar en cualquier proyecto de manera intuitiva, útil y sencilla para cualquier tipo de usuario.

2. JUSTIFICACIÓN

La estadística ha estado y está dedicada al análisis de grandes volúmenes de datos. Sangüesa (2000) habla del reto que ha supuesto la prospección de datos para la estadística, debido a la necesidad de crear instrumentos de trabajo que den respuestas de fácil comprensión a usuarios no siempre expertos en su funcionamiento. De hecho buena parte de los métodos de prospección de datos proceden de la estadística, por ejemplo, los métodos de clasificación y agregación de datos, los modelos de predicción, las redes bayesianas y los métodos heurísticos (métodos de investigación que inventan nuevos procesos, nuevas formas de organización, a partir de los que se están llevando a cabo con el objetivo de mejorarlos).

Las ciencias y la economía sostienen una dependencia notoria con la estadística. Para poder analizar cualquier realidad científica, social o económica de interés es imprescindible la utilización de diferentes métodos estadísticos que permitan la observación del fenómeno y la recolección de datos. Dado esto, se busca lograr una mayor compresión de la realidad analizada que permita la toma de decisiones (Carrasco, 2005).

Lo anterior aunado con la nueva tendencia de automatización del trabajo está empezando a permear en las empresas gracias, en gran medida, a las tecnologías y técnicas de Machine Learning, pero también intervienen otros factores como la incorporación de avances en interacción, información e integración (Ashenden, Ward-Dutton y Wentworth, 2016).

En los diferentes lenguajes de programación y software estadísticos se encuentran paquetes para análisis de clúster o agrupamiento que permiten al usuario realizar sus agrupaciones pero que no son muy intuitivos al momento de parametrizarlos, lo que lleva a malas prácticas y a que los resultados no converjan de manera adecuada a lo que se desea (Eluri et al.,2016), entre los errores más comunes se encuentra el mal manejo de datos, ya que cada tipo de dato es diferente y según el algoritmo que se desee implementar necesita un tratamiento diferente, hay algoritmos que sólo reciben variables numéricas como entrada para sus procesos, por lo que en el caso de las variables categóricas es necesario que estas variables sean dicotomizadas o se realice algún procedimiento de transformación a variables numérica, otras particularidades de algunos algoritmos es que son sensibles a datos atípicos o a escalas diferentes entre variables, por lo que en el caso de las variables numéricas también es necesario realizar tratamiento de atípicos o estandarización de los datos, de manera que el modelo no se vea afectado o sesgado ante la escala o el peso de una variable particular y que este no sea

el resultado deseado, además de preparación de datos como imputación de datos faltantes o de irregularidades en los datos (Moran,2009).

Hoy en día hay algoritmos sofisticados que exigen una parametrización específica, la cual necesita un mayor entendimiento del problema y del algoritmo, por lo que personas inexpertas pueden usar el algoritmo bajo los parámetros por defecto que pueden no ser lo que en realidad necesita y lo que exige los datos en particular.
Dado que estas librería disponibles en los diferentes frameworks y lenguajes de programación no son lo suficientemente intuitivas se hace necesario un paquete o software que permita al usuario realizar una parametrización automática de acuerdo a sus datos o una parametrización manual con un muy buen concepto de cada parámetro, esta segunda solución se ha abarcado por parte de la compañía SAS, la cual en su asistente de clúster permite parametrizar el problema y brinda un muy buen manual de usuario, sin embargo, los valores por defecto y la parametrización que presenta no logra cubrir los problemas particulares de las bases de datos (SAS Institute, 2012).

2.1. HIPÓTESIS

Es posible diseñar algoritmos de detección de grupos que incorporen elementos del aprendizaje automático de máquinas tal que se facilite el proceso de modelación para el usuario final.

3. MARCO TEÓRICO

Las sociedades primitivas se organizaron para lograr trascender su ámbito local, lo que los llevó a enfrentar la necesidad de tomar decisiones que exigían un conocimiento numérico de los recursos disponibles. Esto dio lugar al desarrollo de las primeras técnicas estadísticas basado en el recuento y presentación de datos. A partir de allí, la humanidad ha mejorado las técnicas e implementado la estadística para resolver problemas en casi todos los ámbitos de estudio como las ciencias, la sociedad y la economía (Fernández, S. F.,et al., 2002)

Dentro de la evolución de las necesidades para el análisis de datos, surgió a principios del siglo XXI el término *ciencia de datos* atribuido al autor Cleveland (2001), quién lo define como una continuación de algunos campos de análisis de datos como la estadística, pone la nueva disciplina propuesta en el contexto de la informática y el trabajo contemporáneo en la minería de datos.

Así mismo, en el mismo año, los autores Breiman, et al. (2001), plantean “Si nuestro objetivo como campo es utilizar datos para resolver problemas,

entonces debemos alejarnos de la dependencia exclusiva de los modelos de datos y adoptar un conjunto de herramientas más diverso".

Según (Hazen et al., 2014) se define como campo emergente de la ciencia de datos que combina, ciencias de la computación, estadística, matemática, y la experiencia ciencia del comportamiento para la toma de decisiones. Otro elemento agregado posteriormente es su relación con la minería de datos (Provost and Fawcett, 2013), la minería de procesos (Van der Aalst, 2016) y el aprendizaje automático (Chojnacki et al., 2017).

Desde la *estadística*, se presenta dos ramas principales para el análisis de datos, la *estadística descriptiva* relacionada con la descripción y clasificación de los datos, y la *estadística inferencial* relacionada con la extracción de conclusiones a partir de los datos.

Existen diferentes métodos estadísticos para manipular un conjunto de datos como lo son el *análisis univariante*, el cual es definido por Cuadras (1996) como la descripción de una sola variable mediante características muestrales, o la realización de inferencias sobre la información de una muestra obtenida bajo una o varias condiciones experimentales; y e*l análisis multivariante* que cumple también con los objetivos descriptivos e inductivo, pero trabajando simultáneamente con varias variables en lugar de una sola, convirtiéndola en una metodología estadística más compleja, pero al mismo tiempo más potente.

Existen diversos objetivos que motivan la realización de análisis de datos multivariantes, entre ellos, Peña (2002) señala los siguientes:

- Simplificar el número de variables del que disponemos inicialmente, minimizando la pérdida de información, utilizando únicamente cantidad necesaria para explicar la realidad analizada, obteniendo sus propiedades.

- Agrupar los datos disponibles según similitudes entre las propiedades de los mismos.

- Clasificar nuevas observaciones en grupos que ya han sido determinados previamente.

- Vincular dos grupos de variables, examinando la relación que existe entre ambas.

Autores como Hair, Tatham y Black (1999), exponen la posibilidad de clasificación de las variables entre *dependientes o independientes*. Esto, permite catalogar los métodos en a*nálisis de dependencia*, en el que una o más variables dependientes se explicarán por un conjunto de variables

independientes conocidas; y en *análisis de interdependencia*, definido como aquel que no diferencia entre variables dependientes e independientes, tratándose todas como un conjunto.

Los autores Díaz, M., et al (2016) exponen las técnicas de dependencia que más se utilizan como:

- Análisis de varianza: Se caracteriza por presentar una variable o más métrica dependiente, y una o más no métricas independientes.
- Análisis de regresión: Se caracteriza por presentar una variable dependiente métrica y una o más variables independientes métricas. En este caso es común que la variable dependiente se expresa como una combinación lineal de las variables independientes.
- Análisis discriminante: Se caracteriza por presentar una variable dependiente no métrica y una o más variables independientes métricas. En este caso la variable dependiente se expresa como la combinación lineal de las variables independientes, este último bajo un supuesto de normalidad multivariada.
- Análisis de regresión logística: Se caracteriza por presentar una variable dependiente no métrica y una o más variables independientes métricas. En esta técnica no se requiere el supuesto de normalidad de las variables independientes como en el caso anterior.

Los autores Díaz, M., et al (2016) exponen las técnicas de interdependencia que más se utilizan como:

- Análisis de Componentes principales: Mediante el cual a partir de las variables originales, se buscan generar nuevas no correlacionadas que recogen la variabilidad de las originales.
- Análisis factorial: Se busca generar a partir de las variables métricas del problema una menor cantidad de variables, por lo general, no correlacionadas, que representen la misma información (o por lo menos en un alto porcentaje) que las variables métricas originales.
- Análisis de correspondencia: Busca describir dimensiones fundamentales de un fenómeno, pero trabajando con variables categóricas.
- Análisis clúster: Comprende diferentes técnicas, que al darse un conjunto de variables, se obtienen subconjuntos, teniendo en cuenta criterios de homogeneidad interna, pero guardando una diferencia significativa entre los conjuntos en sí.

Otras técnicas para analizar datos están orientadas al uso de inteligencia artificial como es el caso de la prospección de datos (data mining) se inscribe en un proceso más amplio como es el del descubrimiento de conocimiento dentro de grandes bases de datos (KDD o knowledge discovery in data bases).

Este proceso, no trivial, consiste en descubrir patrones válidos en un conjunto de datos, que deben ser potencialmente útiles en relación al objetivo propuesto en el proceso de prospección de datos y comprensibles para el usuario (Aluja, 2001).

Molina (2002) define *el data mining* partiendo, precisamente, de la distinción entre datos, información y conocimiento. El data mining trabaja en un nivel superior buscando patrones de conducta, agrupaciones, secuencias, tendencias o asociaciones de datos que puedan generar algún modelo que permita entender mejor el dominio con el objetivo de facilitar la toma de decisiones.

Al día de hoy los modelos más utilizados en el campo de la analítica son los modelos que provienen del aprendizaje de máquinas automático (machine learning) es la rama de la inteligencia artificial que estudia el modo en que los sistemas inteligentes son capaces de desarrollar conocimiento y habilidades nuevos a partir de su experiencia. Los métodos de aprendizaje automático buscan la extracción de nuevo conocimiento a partir de la observación de los datos de su entorno o del mismo comportamiento del sistema inteligente. Este campo ha aportado a la prospección de datos una gran parte de los métodos basados en la lógica, el aprendizaje basado en casos, las redes neuronales, reglas de clasificación, etc (Aluja, 2001; Mena, 1999), lo que ha permitido un avance significativo en estos modelos y diferentes enfoques, abriendo un campo grande al aprendizaje automático o automated machine learning, el cual busca, automatizar todos los procesos basados en técnicas de aprendizaje de máquinas con el fin de que personas no expertas hicieran uso de estas técnicas sin necesidad de convertirse un experto en este campo, además de ofrecer soluciones más simples y más rápidas que las soluciones convencionales que ofrecen estas técnicas (Feurer et al, 2015).

Adicional a todas estas técnicas de modelado y dado que la recolección de información cada vez más se vuelve más valiosa para el mercado y el aprovechamiento de la información permite que todas las ciencias, tengan avance continuo ha sido necesario incorporar en todas estas metodologías técnicas para análisis de grandes volúmenes de datos donde se utilizan técnicas de optimización que van desde la etapa de recolección de datos, organización y depuración, procesamiento hasta el análisis y la utilización de la información para inferencia (Aguilar, 2016).

4. ESTADO DEL ARTE

El análisis clúster, también conocido como análisis de conglomerados, se engloba dentro de los métodos conocidos como interdependientes. Se trata de una técnica creada para la clasificación de observaciones en grupos. Persigue

reunir las observaciones en grupos lo más homogéneos posible, de manera que los elementos integrantes de los conglomerados sean muy parecidos. Al mismo tiempo, se busca la máxima heterogeneidad entre clústeres (Uriel y Aldás, 2005).

Para el clúster de variables la técnica es la misma, maximizar la homogeneidad de los grupos de variables creados, de forma que las variables cuya relación sea más fuerte, y que por lo tanto proporcionan la misma información, constituyen un clúster (Chavent, et al., 2012, p. 3). Gráficamente, si la clasificación se realizó correctamente, la distancia de los elementos dentro de los conglomerados será reducida, mientras que los grupos estarán muy distantes los unos de los otros (Hair et al., 1999, p. 492).

A diferencia de lo que ocurre con otras técnicas multivariantes, como el análisis multivariante de la varianza, el análisis de conglomerados es una técnica descriptiva en lugar de explicativa. No utiliza la estadística para explicar la situación objeto de estudio. Se centra en describir la realidad que está analizando, mostrando las peculiaridades de los grupos alcanzados (Sánchez, 1990).

Aldenderfer y Blashfield (1984) dividen el análisis clúster en cinco etapas diferentes, las cuales serán definidas a continuación, indicando sus características:

- Selección de la muestra que se dividirá en grupos durante el análisis.
- Definición de las variables que van a ser incluidas en la muestra.
- Cómputo de las similitudes o distancias entre elementos
- Uso de los métodos de análisis clúster para la creación de grupos con entidades similares.

Dependiendo de la técnica empleada se distinguen dos clases de métodos, los *métodos jerárquicos* y los *métodos no jerárquicos*. La elección de uno u otro será tomada por el investigador, pudiendo combinar ambos métodos. Cabe mencionar que el método jerárquico es el comúnmente utilizado para realizar clusterización (Ferrán, 1996).

Los métodos jerárquicos presentan grupos que se van formando de forma progresiva, uniendo o separando grupos o individuos en función de su similitud. A medida que se avanza en el proceso de agrupamiento se va desarrollando una estructura en forma de árbol, a partir de la cual se decide el número de grupos a obtener, ya que es algo que no se conoce a priori. Los métodos jerárquicos pueden dividirse en dos tipos: *de aglomeración* y *decisivos.*

En los métodos no jerárquicos se conoce a priori el número de grupos que se

desea formar, por tanto, se persigue asignar cada observación al grupo que presente características más parecidas a las suyas, en donde la distancia de cada observación con respecto a un punto central, conocido como semilla sea reducida. El método no jerárquico habitual es el de k-medias. Como dice Vidal Díaz de Rada (2002), el objetivo es formar grupos uniformes, maximizando la varianza entre grupos y minimizándola dentro de cada grupo. Estos métodos son recursivos, puesto que un vez conseguido el resultado, se recalculan los centroides y se vuelve a realizar la asignación, otras técnicas se incorporan como complemento de k-medias como k-medianas y técnicas basadas en la mediana de los datos de manera que los datos atípicos no afecten o sesgen el resultado, adicional en respuesta al tratamiento de datos llegan metodologías que permiten realizar agrupamiento con variables categóricas como k-prototypes (Uriel y Aldás, 2005).

En los métodos no jerárquicos el número de grupos que deben formarse viene determinado, bien por experiencias o estudios previos, o bien a través de los resultados obtenidos al realizar primero un procedimiento jerárquico. Además, hay que señalar que un número reducido de grupos puede desencadenar un resultado muy pobre al estudiar grupos demasiado heterogéneos, mientras que trabajar con muchos grupos puede dificultar la interpretación (Luque, 2012).

Para aplicaciones en estos clúster existen software estadísticos que facilitan la aplicación en caso de R con paquetes como rattle o SAS que presentan un asistente de clustering el cuál pide una parametrización del tipo de clúster a implementar como cantidad de grupos, tratamiento a realizar en los datos y la salida esperada, siendo un entorno gráfico es más amigable para el usuario, sin embargo es necesario tener un conocimiento para realizar la parametrización del tipo de clúster a utilizar y del tratamiento de los datos.
En los diferentes lenguajes de programación se han implementado paquetes para realizar análisis de agrupamiento donde destacan R, Python, mahout, spark, al ser las más usadas, según la necesidad que se tenga.

Se ha realizado una sofisticación de las técnicas de agrupamiento con mayor robustez pero también mayor complejidad, ejemplo de esto son las técnicas de clúster implementadas en mahout de modelos de lógica difusa y modelos de aprendizaje profundo como agrupamiento espectral (Ng, 2002).

Herramientas como hadoop, especialmente spark y librerías como skitlearn han hecho una apuesta para trabajar con grandes volúmenes de datos permitiendo que más áreas de la ciencia se incorporen a realizar sus análisis y al aprovechamiento de datos y logrando una optimización de los modelos que se implementan.

Estas técnicas descritas necesitan un nivel de conocimiento en el tema o en los algoritmos a implementar y en el caso de herramientas como spark o librerías como skitlearn también exigen un nivel de programación o conocimiento de los lenguajes, como consecuencia todas estas técnicas se someten al conocimiento y por tanto a la subjetividad del modelados, con el fin de reducir esta subjetividad es necesario automatizar decisiones como cantidad de grupos óptimos, selección y tratamiento de variables e incluso selección del modelo óptimo según las condiciones y el volumen de los datos, facilitando la implementación y la detección de grupos.

5. OBJETIVOS

5.1. Objetivo General

Desarrollar prototipo de librería que permita análisis automático de clústers mediante metodologías del estado del arte y que facilite las decisiones de modelado por parte del usuario

5.2. Objetivo Específicos

1. Seleccionar, al menos, tres metodologías del estado del arte para la detección de grupos.
2. Diseñar o adaptar mecanismos de aprendizaje automático de máquinas para la automatización de la detección de grupos.
3. Implementar los algoritmos propuestos en un prototipo de librería que facilite su uso por parte del usuario.
4. Realizar pruebas de validación que permitan comprobar la librería.

6. METODOLOGÍA

Para la realización de un paquete de software que permita la automatización y facilite el proceso de detección de grupos se trabajará en cuatro fases.

1. Selección de metodologías
2. Diseño de mecanismos
3. Implementación
4. Pruebas

Donde se priorizan las siguientes actividades.

1. Selección de metodologías
    - 1.1. Análisis de bibliografía existente
    - 1.2. Revisión del estado del arte
    - 1.3. Selección de herramientas, lenguajes y plataformas
    - 1.4. Revisión y pruebas de las metodologías en las herramientas seleccionadas
    - 1.5. Selección de las metodologías a utilizar

1.6. Revisión y pruebas

2. Diseño de mecanismos
   2.1. Revisión requerimientos mínimos
   2.2. Revisión de librería necesarias
   2.3. Diseño de la plataforma
   2.4. Diseño de funcionalidades
   2.5. Diseño experiencia de usuario
3. Implementación
   3.1. Instalación de herramientas
   3.2. Instalación de librerías
   3.3. Desarrollo de funcionalidades de entrada de datos
   3.4. Desarrollo de funcionalidades de clúster
   3.5. Desarrollo de funcionalidades de salida de datos
   3.6. Empaquetado
   3.7. Documentación del paquete
4. Pruebas
   4.1. Testeo de errores
   4.2. Validación con diferentes conjuntos de datos
   4.3. Validación funcionamiento integral del paquete
   4.4. Pruebas de coherencia en la documentación

5. REFERENCIAS

Aguilar, L. J. (2016). Big Data, Análisis de grandes volúmenes de datos en organizaciones. Alfaomega Grupo Editor.

Aldenderfer, M. S. y Blashfield, R. K. (1984). Cluster analysis. California: Sage Publications

Aluja, T. (2001). La minería de datos, entre la estadística y la inteligencia artificial. Qüestiió: quaderns d'estadística i investigació operativa, 25(3), 479-498.

Ashenden,A., Ward-Dutton, N., y Wentworth, C., (2016). La nueva tendencia de automatización: Machine Learning y más. MWD Advisors. Disponible en: https://www.ibm.com/downloads/cas/M1PG1J23.

Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

Carrasco, E. C., (2005, September). Herramientas para la toma de decisiones empresariales: mapas tecnológicos. In IX Congreso de Ingeniería de Organización (p. 50).

Chavent, M., Kuentz-Simonet, V., Liquet, B. y Saracco, J. (2012). ClustOfVar: An R
package for the clustering of variables. Journal of Statistical Software, 50(13), 1-16.

Chojnacki, A., Dai, C., Farahi, A., Shi, G., Webb, J., Zhang, D.T., Abernethy, J., Schwartz, E., 2017. A Data Science Approach to Understanding Residential Water Contamination in Flint, in: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '17. ACM, New York, NY, USA, pp. 1407–1416.

Cuadras, C. M. (1996). Métodos de Análisis Multivariante (EUB). Barcelona: SL.

Díaz, M., León, Á., Alvin, H., & Díaz Mora, M. E. (2016). Introducción al análisis estadístico multivariado aplicado. Experiencia y casos en el Caribe colombiano. Universidad del Norte.

Eluri, V. R., Ramesh, M., Al-Jabri, A. S. M., & Jane, M. (2016, March). A comparative study of various clustering techniques on big data sets using Apache Mahout. In 2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC) (pp. 1-4). IEEE.

Fernández, S. F., Sánchez, J. M. C., Córdoba, A., & Largo, A. C. (2002). Estadística descriptiva. Esic Editorial.

Ferrán, M. (1996). SPSS para Windows: programción y análisis estadístico. Madrid:
McGraw-Hill.

Feurer, M., Klein, A., Eggensperger, K., Springenberg, J., Blum, M., & Hutter, F. (2015). Efficient and robust automated machine learning. In Advances in neural information processing systems (pp. 2962-2970).

Hair, J. F., Anderson, R. E., Tatham, R. L. y Black, W. C. (1999). Análisis multivariante (5ª ed.). Madrid: Prentice Hall Iberia

Hazen, B.T., Boone, C.A., Ezell, J.D., Jones-Farmer, L.A., 2014. Data quality for data science, predictive analytics, and big data
in supply chain management: An introduction to the problem and suggestions for research and applications. Int. J. Prod.
Econ. 154, 72–80. https://doi.org/10.1016/j.ijpe.2014.04.018

Luque, T. (Ed). (2012). Técnicas de análisis de datos en investigación de mercados (2ª ed.). Madrid: Ediciones Pirámide

Peña, D. (2002). Análisis de datos multivariantes. Madrid: McGraw-Hill.

Provost, F., Fawcett, T., 2013. Data Science for Business: What you need to know about data mining and dataanalytic
thinking. O'Reilly Media, Inc.

Sangüesa Solé, Ramón (coord.) (2000). Data mining: una introducción. Barcelona: Universitat Oberta de Catalunya.

SAS Institute. (2012). SAS/OR 9.3 User's Guide: Mathematical Programming Examples. SAS institute.

Uriel, E. y Aldás, J. (2005). Análisis Multivariante Aplicado. Madrid: Thomson

Van der Aalst, W.M., 2016. Process mining: data science in action. Springer.

Vidal Díaz de Rada Iguzquiza, V. D. (2002). Técnicas de análisis multivariante para investigación social y comercial: Ejemplos prácticos utilizando SPSS versión 11. Ra-Ma.