



Análisis de Correspondencias Múltiples aplicado a una distribuidora de bebidas

Mauricio Cuozzo
UNT-Facultad de Ciencias Económicas
Mauriciocuozzo@gmail.com

RESUMEN

Las organizaciones históricamente se enfrentaron a problemas complejos, afectados por múltiples variables, hoy en día tienen la posibilidad gracias a los avances tecnológicos de brindar un mayor valor a partir de la aplicación de técnicas de análisis multivariado que permita a las empresas mejorar su gestión de los grandes volúmenes de datos que ellas mismas generan.

Se presenta como objeto de estudio el sector de “Equipos De Frío” (EDF) perteneciente a una embotelladora dedicada a la producción, comercialización y distribución de bebidas ubicada en territorio tucumano. En el mismo, se identifican como recursos escasos los EDF, los cuales deben ser asignados y/o reasignados a los clientes

Se utiliza el Análisis de correspondencias múltiple como una técnica descriptiva o exploratoria cuyo objetivo es resumir una gran cantidad de datos en un número reducido de dimensiones, con la menor pérdida de información posible.

Este trabajo responde a un diseño cuantitativo no experimental de corte longitudinal con un alcance exploratorio y descriptivo. En una investigación no experimental se observan fenómenos tal como se dan en su contexto natural, para posteriormente analizarlos.

Palabras Clave: Análisis multivariado – Correspondencia – Analítica de negocios

INTRODUCCIÓN

Nos encontramos actualmente en la era del conocimiento donde la información es uno de los pilares fundamentales en la sociedad y se ha convertido en el recurso más valioso de cada empresa, organización, fundación, etc. Los avances tecnológicos permiten almacenar y procesar grandes volúmenes de información en las instituciones con o sin fines de lucro, para lograr soporte de las decisiones financieras, comerciales, administrativas y económicas.

La tendencia actual es analizar la información histórica para descubrir conocimiento que ayude a optimizar el proceso de toma de decisiones. Si un gerente está bien informado de lo que sucede en la empresa con datos reales, puede tomar decisiones mucho más acertadas y lograr que su empresa sea competitiva en el mercado y más eficiente internamente.

Hair (1999) relata que a partir de la década del 80 se fueron desarrollando los programas estadísticos que facilitaron el análisis de gran cantidad de datos cuyo origen estaba en encuestas o en bases de datos que provenían de fuentes secundarias de información. Los fundamentos teóricos o estadísticos de las técnicas multivariantes que permitían el análisis de estos datos habían sido desarrollados con anterioridad, pero sólo cuando los ordenadores tuvieron la capacidad de cálculo y memoria necesarios para llevar a cabo el análisis multivariante.



La organización bajo estudio es una distribuidora de bebidas ubicada en el territorio tucumano, que además de comercializar bebidas a otras organizaciones, les ofrece equipos de frío (EDF) en comodato que les permite a sus clientes conservar sus productos en el mejor estado posible hasta que la bebida llegue a las manos del consumidor final. Ese contrato de comodato establece entre otras situaciones que el cliente deberá cumplir con una cierta cantidad de bebidas dependiendo de la marca del EDF y de la cantidad de puertas con la que cuenta el equipo.

El propósito de este trabajo es analizar la base de datos facilitada por la organización con el fin de generar información valiosa para el gerente del sector a partir de la aplicación de un análisis de correspondencias múltiples.

Planteamiento del Problema

La administración de los EDF en la empresa por varios años no tuvo una atención “especializada”, pero dado el incremento tanto en clientes como en equipos, la empresa decidió definir recientemente un sector que se encargue específicamente de la gestión de estos equipos de frío.

Para el sector es clave la condición de los clientes en relación al cumplimiento o no del estándar establecido por la organización, que está representada por aquellas heladeras que en base a la cantidad de puertas que tienen, cumplen o no con el estándar de ventas determinados por puerta (20 unidades por puerta) por lo que es sumamente importante que este sector logre a partir de la reasignación de equipos, en coordinación con la fuerza de venta, poder reducir el ratio de equipos improductivos.

Teniendo en cuenta entonces lo mencionado anteriormente y sumado a la gran cantidad de datos que se genera en el sector, se considera oportuno emplear técnicas de análisis multivariado que permitan al encargado del sector analizar correctamente los problemas complejos que enfrenta estando al mando del sector.

Preguntas de investigación:

- ¿Cuáles son las variables que afectan a la condición de un cliente?
- ¿Qué tipo de clientes son los que no cumplen con el estándar?
- ¿Cuál es la situación actual de los clientes?

Objetivos

Objetivo General:

Definir un perfil de cliente que se encuentre asociado a las dos condiciones que puede tener según su cumplimiento con el estándar.

Objetivos Específicos:

- Definir y clasificar las variables que afectan a la condición de un cliente
- Determinar la situación actual de los clientes

MARCO TEORICO

Análisis multivariante



Según Hair (1999) el análisis multivariante no es fácil de definir. En un sentido amplio, se refiere a todos los métodos estadísticos que analizan simultáneamente medidas múltiples de cada individuo u objeto sometido a investigación. Cualquier análisis simultáneo de más de dos variables puede ser considerado aproximadamente como un análisis multivariante. En sentido estricto, muchas técnicas multivariantes son extensiones del análisis univariante (análisis de distribuciones de una sola variable) y del análisis bivariante (clasificaciones cruzadas, correlación, análisis de la varianza y regresiones simples utilizadas para analizar dos variables). Por ejemplo, una regresión simple (con una variable predictor) se extiende al caso multivariante para incluir varias variables predictor. De la misma forma, la variable dependiente que se encuentra en el análisis de la varianza se extiende para incluir múltiples variables dependientes en el análisis multivariante de la varianza. Como veremos más adelante, en muchas ocasiones las técnicas multivariantes son un medio de representar en un análisis simple aunque lo que requirió varios análisis utilizando técnicas univariantes. Otras técnicas multivariantes, sin embargo, están diseñadas exclusivamente para tratar con problemas multivariantes, tales como el análisis factorial que sirve para identificar la estructura subyacente de un conjunto de variables o el análisis discriminante que sirve para diferenciar entre grupos basados en un conjunto de variables. El análisis multivariante incluirá tanto técnicas multivariantes como técnicas multivariantes, debido a que los autores creen que el conocimiento de las técnicas multivariantes es un primer paso esencial en la comprensión del análisis multivariante.

Escalas de medida

Según Anderson el análisis de los datos implica la separación, identificación y medida de la variación en un conjunto de variables, tanto entre ellas mismas como entre una variable dependiente y una o más variables independientes. El término clave aquí es medida, dado que el investigador no puede separar o identificar una variación a menos que pueda ser medible. La medida es importante para representar con precisión el concepto de nuestro interés y es crucial en la selección del método de análisis multivariante apropiado.

Existen dos tipos básicos de datos: no métricos (cualitativos) y métricos (cuantitativos). Los datos no métricos son atributos, características o propiedades categóricas que identifican o describen a un sujeto. Describen diferencias en tipo o clase indicando la presencia o ausencia de una característica o propiedad. Muchas propiedades son discretas porque tienen una característica peculiar que excluye todas las demás características. Por ejemplo, si uno es hombre, no puede ser mujer. No hay cantidad de «género», sólo la condición de ser hombre o mujer. Por el contrario, las medidas de datos métricos están constituidas de tal forma que los sujetos pueden ser identificados por diferencias entre grado o cantidad. Las variables medidas métricamente reflejan cantidades relativas o grado. Las medidas métricas son las más apropiadas para casos que involucran cantidad o magnitud, tales como el nivel de satisfacción o la demanda de trabajo.

Las medidas no métricas pueden tener escalas nominales u ordinales. La medida con una escala nominal asigna números que se usan para etiquetar o identificar sujetos u objetos. Las escalas nominales, también conocidas como escalas de categoría, proporcionan el número de ocurrencias en cada clase o categoría de la variable que se está estudiando. Por tanto, los números o símbolos asignados a los objetos no tienen más significado cuantitativo que indicar la presencia o ausencia del atributo o característica bajo investigación. Los ejemplos de datos con escala nominal incluyen el



sexo, la religión o el partido político de una persona. Para trabajar con estos datos, el analista puede asignar números a cada categoría, por ejemplo, 2 para mujeres y 1 para hombres. Estos números sólo representan categorías o clases y no implican cantidades de un atributo o característica. Para el presente trabajo esta es la escala que poseen las variables.

Análisis de correspondencias múltiples

El análisis de correspondencias es una técnica de interdependencia recientemente desarrollada que facilita tanto la reducción dimensional de una clasificación de objetos (por ejemplo, productos, personas, etc.) sobre un conjunto de atributos y el mapa perceptual de objetos relativos a estos atributos. Los investigadores se enfrentan constantemente a la necesidad de «cuantificar datos cualitativos» que encuentran en variables nominales. El análisis de correspondencias difiere de otras técnicas de interdependencia discutidas antes en su capacidad para acomodar tanto datos no métricos como relaciones no lineales. En su forma más básica, el análisis de correspondencias emplea una tabla de contingencia, que es la tabulación cruzada de dos variables categóricas. A continuación transforma los datos no métricos en un nivel métrico y realiza una reducción dimensional (similar al análisis factorial) y un mapa perceptual (similar al análisis multidimensional). A modo de ejemplo, las preferencias por una marca de los encuestados pueden ser tabuladas de forma cruzada con variables demográficas (por ejemplo, género, categorías de renta, ocupación) indicando cuánta gente que prefiere cada una de las marcas entra dentro de cada categoría de las variables demográficas. A través del análisis de correspondencias, la asociación o «correspondencia» de marcas y las características distintivas de aquellos que prefieren cada marca se muestran en un mapa bi o tridimensional, tanto de marcas como características de los encuestados. Las marcas percibidas como similares están localizadas en una cercana proximidad unas de otras. De la misma forma, las características más distintivas de los encuestados que prefieren cada marca están determinadas también por la proximidad de las categorías de las variables demográficas respecto de la posición de la marca. El análisis de las correspondencias proporciona una representación multivariante de la interdependencia de datos no métricos que no es posible realizar con otros métodos.

Además Uriel comenta que el objetivo del análisis de correspondencias es muy parecido al del escalamiento multidimensional pero en el caso del escalamiento multidimensional el mapa perceptual que se obtiene muestra las distancias percibidas entre diversos sujetos u objetos, atendiendo a un conjunto de variables que los caracterizaban. Mientras que en el análisis de correspondencias, el mapa mostrara las distancias entre los distintos niveles de variables no métricas, por lo que se suele decir que el análisis de correspondencias sirve para visualizar tablas de contingencia.

Tablas de contingencias

Bustamante explica que si el objetivo es analizar la interrelación (dependencia) entre los valores de variables nominales. En este caso la herramienta de análisis para dos variables es la denominada tabla de contingencia. En esta tabla se calcula la distribución de los casos (las frecuencias de aparición) para las distintas combinaciones de valores de las dos variables, como se observa en la figura siguiente.



		Variable 2				totales 1
		valor 1	valor 2	...	valor p2	
Variable 1	valor 1	n_{11}	n_{12}	...	n_{1p2}	$t_{1.}$
	valor 2	n_{21}	n_{22}	...	n_{2p2}	$t_{2.}$

	valor p1	n_{p11}	n_{p12}	...	n_{p1p2}	$t_{p1.}$
	totales 2	$t'_{.1}$	$t'_{.2}$...	$t'_{.p2}$	t

A partir de la tabla de contingencia podemos calcular las probabilidades marginales de los valores de la variable 1 como $P_i = t_i/t$, que representa la probabilidad de que la variable 1 tome el valor i . Del mismo modo podemos calcular las probabilidades para la variable 2 como $P_j = t'_j/t$. A partir de las probabilidades marginales podemos calcular los casos “esperados”, bajo la hipótesis a cuestionar de independencia entre variables. Para calcular el valor esperado se multiplica el número total de casos por la probabilidad de que la variable 1 tome el valor i y la variable 2 tome el valor j , es decir, $E_{ij} = t (t_i/t)(t'_j/t) = t i t'_j / t$. Obsérvese que únicamente bajo la hipótesis de independencia podemos calcular la probabilidad conjunta como un producto de probabilidades.

Medidas de similitud

Según Black La similitud entre objetos es una medida de correspondencia, o parecido, entre objetos que van a ser agrupados. La similitud entre objetos puede medirse de varias formas, pero tres métodos dominan las aplicaciones del análisis multivariante. Medidas de correlación, medidas de distancia y medidas de asociación. Cada uno de los métodos representa una perspectiva particular de similitud, dependiendo tanto de sus objetivos como del tipo de datos. Tanto las medidas de distancia como la correlación exigen datos métricos, mientras que las medidas de asociación son para datos no métricos.

Para el presente trabajo se hará hincapié en las medidas de asociación de similitud que se utilizan para comparar objetos cuyas características se miden sólo en términos no métricos (medida nominal y ordinal). Una medida de asociación podría evaluar el grado de acuerdo o de acercamiento entre cada par de encuestados o de sujetos pertenecientes a cualquier base de datos. La forma más simple de medida de asociación sería el porcentaje de veces que existió acuerdo (ambos dicen sí o ambos dicen no) para el mismo conjunto de cuestiones. Se han desarrollado extensiones de este simple coeficiente de ajuste para acomodar variables nominales de varias categorías o incluso medidas ordinales. Muchos programas informáticos, sin embargo, dan un apoyo limitado a las medidas de asociación, y el investigador está en muchas ocasiones forzado a calcular en primer lugar las medidas de similitud y a continuación introducir la matriz de similitud en los diversos programas.

MARCO METODOLOGICO

Este trabajo responde a un diseño cuantitativo no experimental de corte transversal con un alcance exploratorio y descriptivo. En una investigación no experimental se observan fenómenos tal como se dan en su contexto natural, para posteriormente analizarlos.



La población bajo estudio es el registro de los EDF del área durante el período 2021 y 2022. Como técnicas de recolección de datos se recurre a entrevistas, observación directa y acceso a los datos en los diferentes sistemas que cuenta la empresa. Como técnicas de análisis se utiliza el análisis de correspondencias múltiples y para el procesamiento y visualización de datos se utiliza el software RStudio. Para el tratamiento de los datos se procedió en primer lugar a realizar actividades de data cleaning que consistieron en limpiar y depurar los datos detectando, corrigiendo y eliminando registros corruptos o imprecisos de aquellas entradas que contenían información incorrecta, incompleta, mal formateada o duplicada.

DESARROLLO

Base de datos

Para el presente trabajo se seleccionaron 5 variables para poder ser analizadas mediante técnicas de análisis multivariantes, estas variables son las siguientes:

- **Tamaño:** Referido a una clasificación que se le da al cliente según la cantidad de equipos con los que cuenta, su nivel de ventas y las dimensiones de su local físico. Se encuentra dividido en 4 categorías que son "Pequeño", "Mediano", "Grande" y "Extra Grande".
- **Clasificación:** Es una denominación que realiza la organización bajo estudio para identificar a su clientela según ciertas características que los diferencia a unos de otros por sus actividades empresariales que ejerce.
- **Condición:** Para esta variable se enfrenta al nivel de ventas real de cada cliente en el último mes contra el nivel de ventas mensual definido por el estándar, donde los clientes que igualan o superan el estándar son clasificados como "Cumple" y quienes se encuentran por debajo del nivel que marca el estándar son clasificados como "NoCumple".

Las variables presentadas anteriormente pueden clasificarse como *variables cualitativas de escala nominal* ya que cuentan con categorías que son mutuamente excluyentes y exhaustivas. A excepción del caso de la variable tamaño que se clasifica como *variable cualitativa de escala ordinal*.

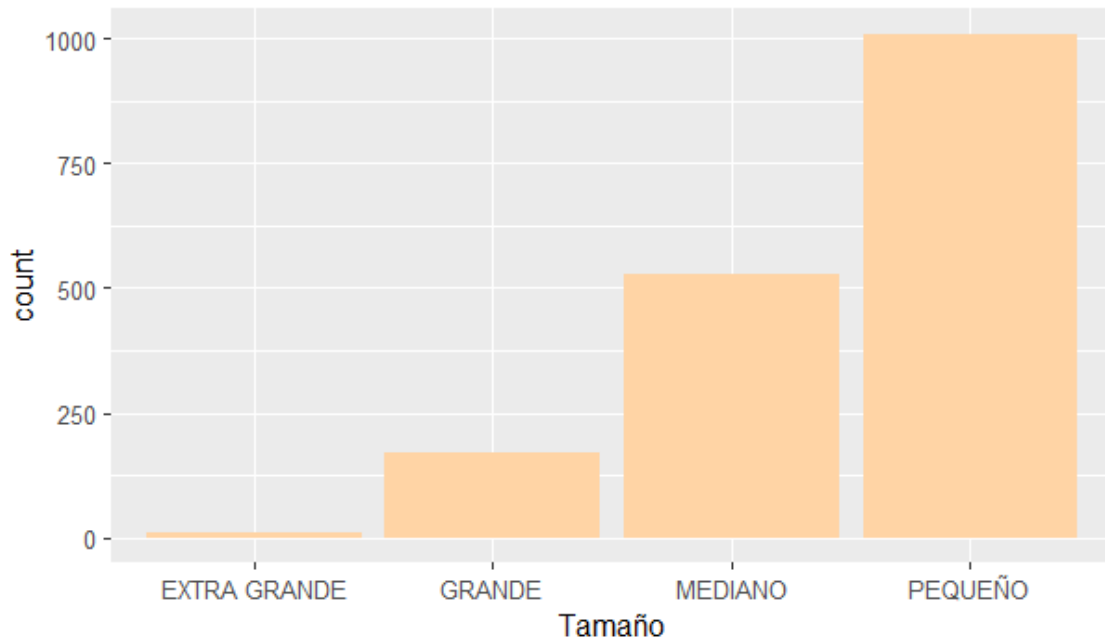
Análisis de correspondencias múltiples

Teniendo en cuenta la problemática presentada y que en la base de datos se encuentran variables nominales categóricas, resulta oportuno aplicar un análisis de correspondencias múltiples que permita representar en un espacio multidimensional reducido la relación existente entre las categorías con el fin de que tras descubrir las relaciones entre las variables se puedan realizar sugerencias para la organización bajo estudio.



Una vez definida la técnica a utilizar se procede a visualizar en gráficos de barras las 5 variables con la intención de describir las variables bajo estudio para obtener una mejor comprensión de la base de datos.

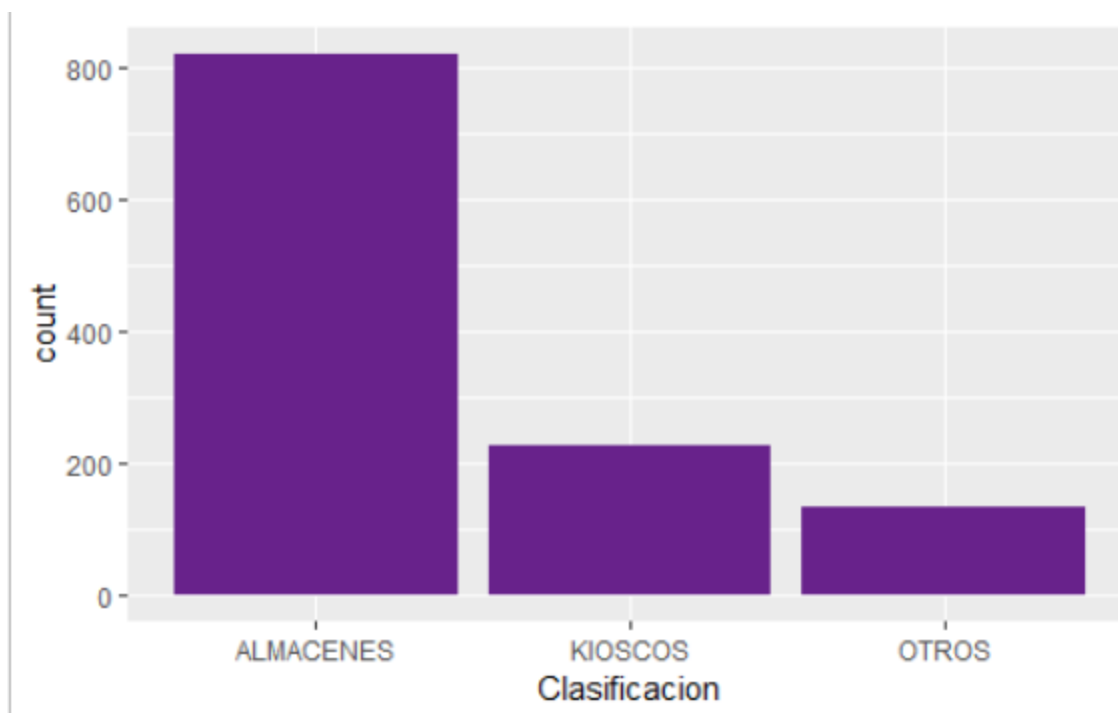
Grafico 1: Frecuencias de Tamaño en grafico de barras.



Fuente: Elaboración propia en RStudio.

En el grafico 1 se puede visualizar la cantidad de veces que aparece cada categoría de Tamaño, se destaca la baja frecuencia con la que aparece la categoría “Extra Grande” por lo que con el objetivo de mejorar la calidad del análisis serán eliminados aquellos sujetos que pertenezcan a esta categoría.

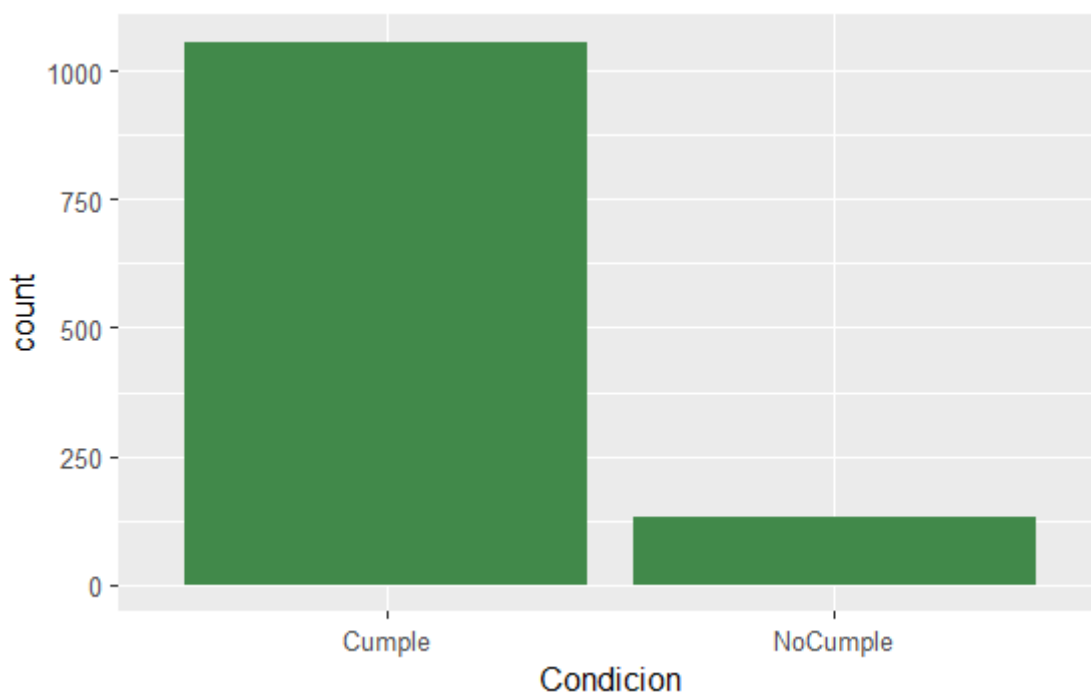
Grafico 2: Frecuencias de Clasificación en grafico de barras.



Fuente: Elaboración propia en RStudio.

En el caso del gráfico 3 se advierte que los tipos de negocios están divididos en 3 grupos en donde los denominados como Almacenes cuentan con alrededor de 800 observaciones mientras que los Kioscos y Otros cuentan con más o menos 250 y 150 observaciones respectivamente.

Gráfico 3: Frecuencias de Condición en gráfico de barras.

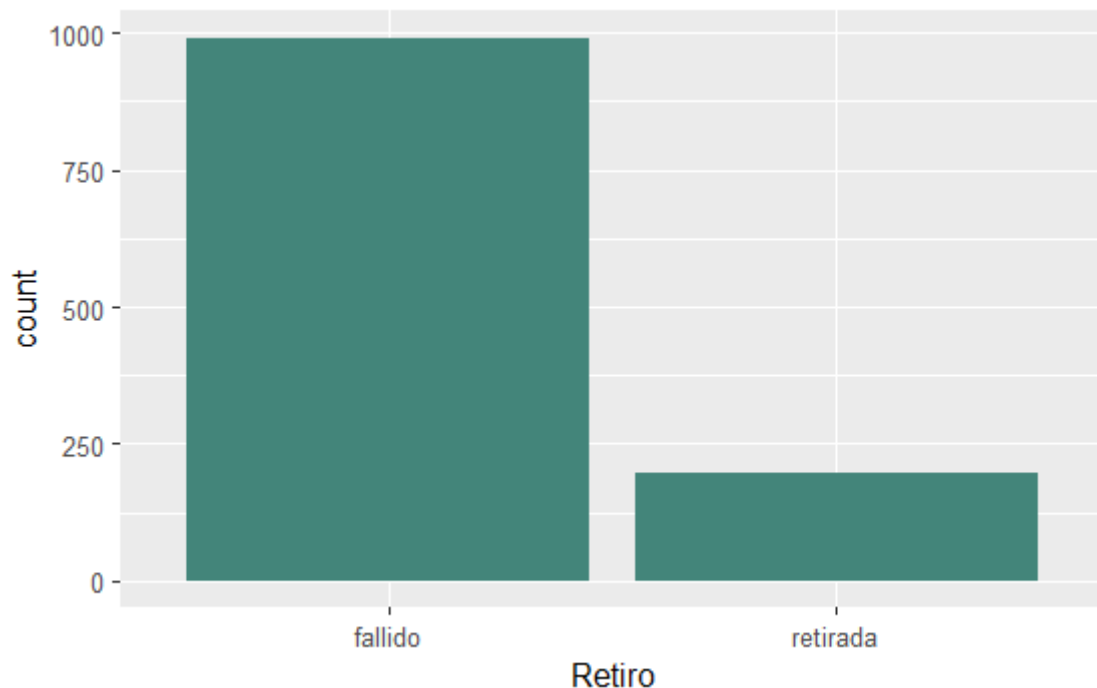




Fuente: Elaboración propia en RStudio.

Para el caso del grafico n° 4 se puede atisbar que existe un numero bajo de clientes que no cumplen con el estándar para esta base pero uno de los objetivos de este estudio es poder relacionar a aquellos clientes que tienen esta condición con las demás variables con la finalidad de poder definir un perfil que ayude a la gestión de la organización.

Grafico 4: Frecuencias de Retiro en grafico de barras.



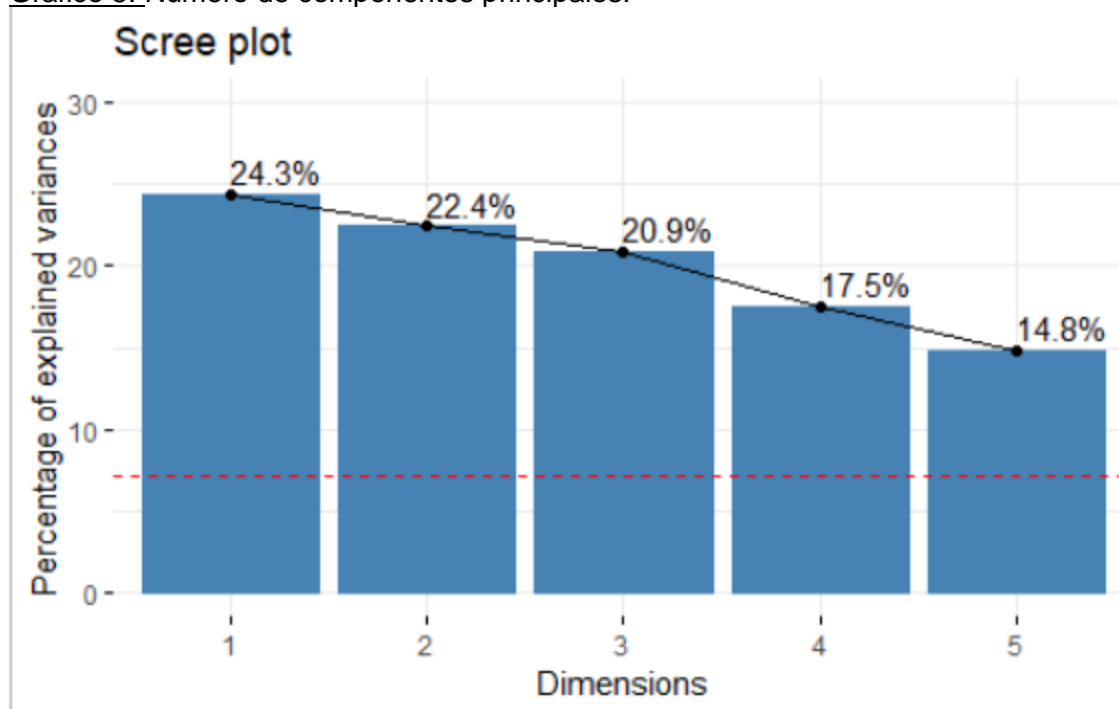
Fuente: Elaboración propia en RStudio.

A diferencia del grafico anterior, en el grafico n° 5 se puede observar que la condición de retiro fallido (la cual es la problemática) tiene un mayor número de observaciones que los clientes que no cumplen con el estándar y esto representa una gran cantidad de dinero perdido por esfuerzos y por capital inmovilizado para realizar un retiro que termina siendo fallido.

Tras haber realizado un análisis descriptivo de las variables, se prosigue con el análisis de correspondencias múltiples recordando que se eliminó la categoría "Extra Grande" de la variable "Tamaño".



Grafico 5: Numero de componentes principales.

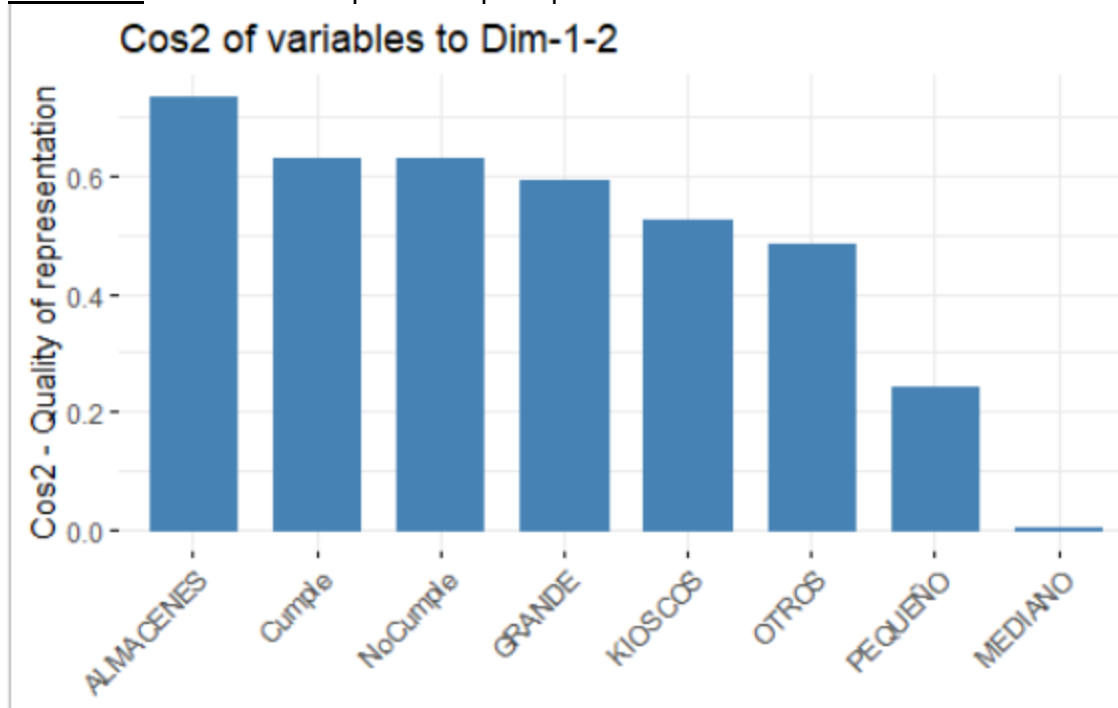


Fuente: Elaboración propia en RStudio.

Con el grafico 6 se busca determinar el número de componentes principales, se puede visualizar mediante el Scree Plot presentado, que es un modelado de los eigenvalues ordenados de mayor a menor. El número de componentes es determinado en el punto más allá del cual los eigenvalues restantes son todos relativamente pequeños y de tamaño comparable que en este caso serán dos dimensiones ya que al elegir esta cantidad de dimensiones se puede visualizar las distancias entre las variables en un eje de coordenadas y permitirá explicar el 46,7% de la variabilidad.



Grafico 6: Numero de componentes principales.



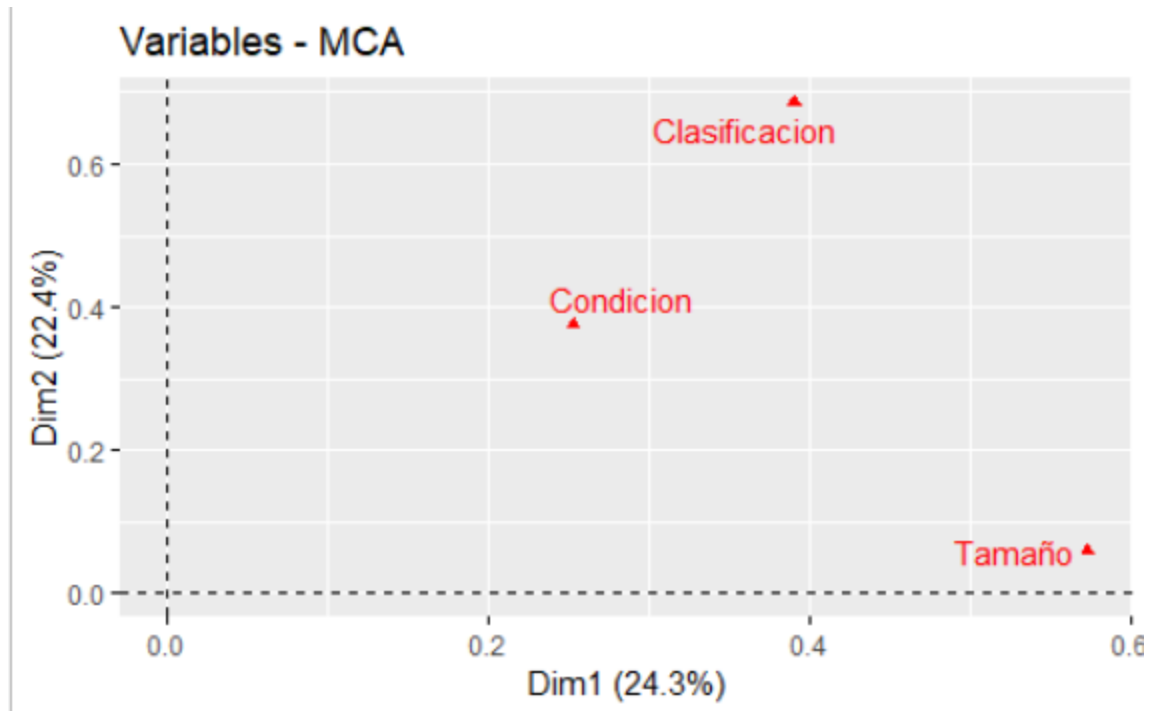
Fuente: Elaboración propia en RStudio.

La calidad de la representación se llama el coseno cuadrado (Cos2), el cual mide el grado de asociación entre las categorías de las variables y un eje en particular. Si la categoría de una variable está bien representada por dos dimensiones, la suma del cos2 es cercana a uno. Para algunos ítems de las filas, más de dos dimensiones son requeridas para representar perfectamente los datos.

Las categorías peor representadas por las dimensiones 1 y 2 parecen ser MEDIANO y PEQUEÑO. Esto significa que la posición de sus correspondientes puntos en el diagrama de dispersión debe ser interpretada con precaución.



Grafico 7: Relación entre variables y dimensiones.



Fuente: Elaboración propia en RStudio.

Después de definir la cantidad de dimensiones se busca establecer cuáles son las variables que más preponderancia tienen sobre cada dimensión y como se puede ver en el grafico 7 la variable Tamaño es la variable que se encuentra más a la derecha por lo que es la variable que mejor explica a la dimensión 1 mientras que para la dimensión 2 hay que fijarse en la variable que se encuentre más arriba en la gráfica que en este caso será la Clasificación la variable que mejor explique la variabilidad de la segunda dimensión.

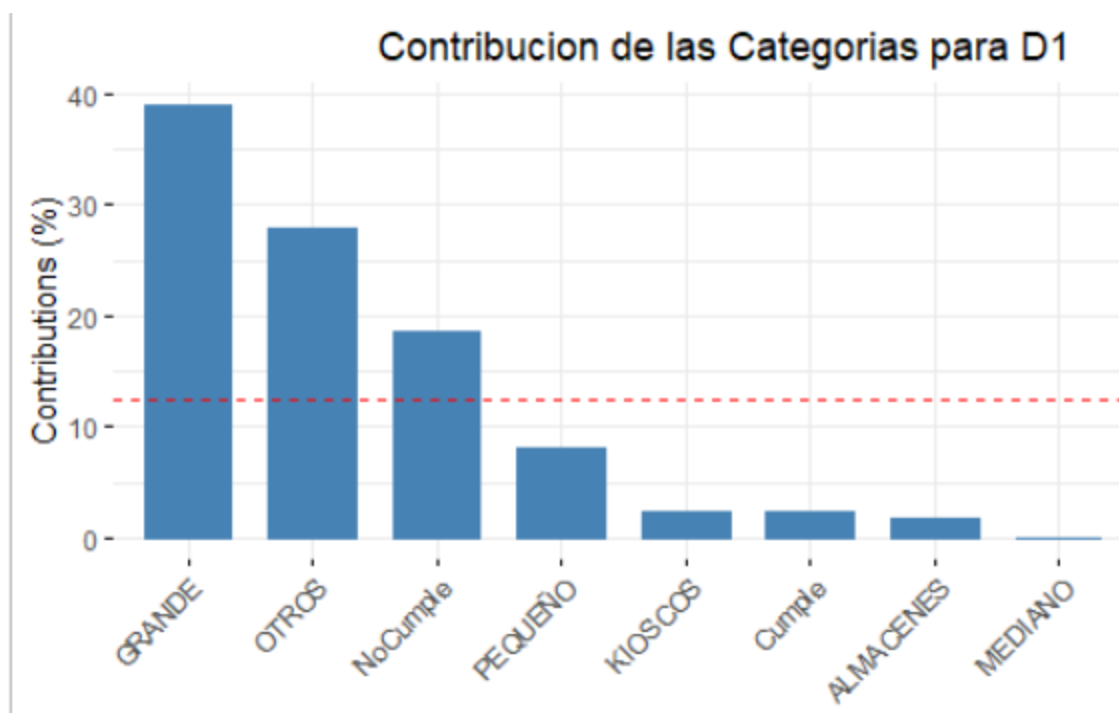
Los valores que toman las variables en el grafico 7 están representando el valor de R-cuadrado de cada variable que son los siguientes:

Variable	R2 para D1	R2 para D2
Tamaño	0.5725713	0.05881062
Clasificación	0.3902837	0.68622220
Condición	0.2532755	0.37571953

Como se puede observar la variable Tamaño tiene un R-cuadrado más alto por lo que se confirma lo visto en grafico que es la variable que más influencia tiene sobre la dimensión 1 , mientras que para el caso de la segunda dimensión se confirma que Clasificación es la variable de mayor influencia.



Grafico 8: Contribución de categorías para dimensión 1.

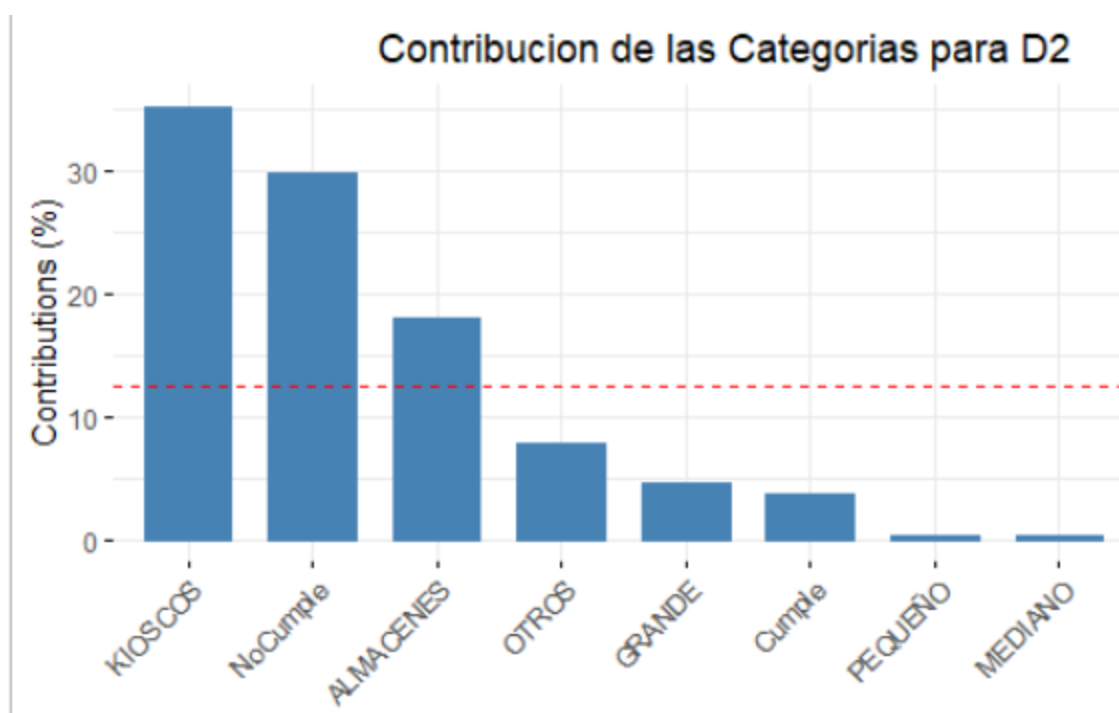


Fuente: Elaboración propia en RStudio.

En este grafico se puede observar la contribución que realiza cada categoría para la dimensión 1, la línea roja indica el valor promedio esperado si las contribuciones fueran uniformes. En este caso, las categorías GRANDE, OTROS y NoCumple son las más importantes en la definición de la primera dimensión.



Grafico 9: Contribución de categorías para dimensión 2.

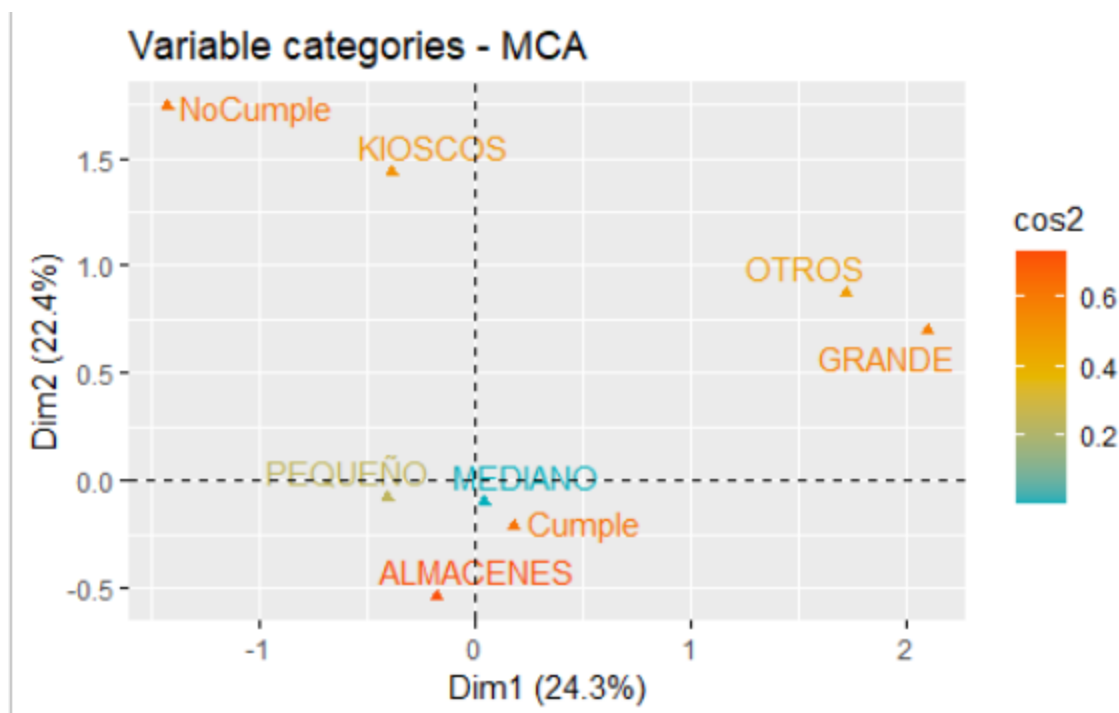


Fuente: Elaboración propia en RStudio.

En el grafico número 9 se puede visualizar que para el caso de la dimensión 2 serán KIOSCOS, NoCumple y ALMACENES las categorías que más contribuyan para explicar a la segunda dimensión.



Grafico 10: Relacion entre variables.



Aquí se observa la relación y asociación entre las categorías de las variables y se puede interpretar como:

- Las categorías de variables con un perfil similar están agrupadas juntas.
- Categorías de variables correlacionadas negativamente están posicionadas en lados opuestos del origen de la gráfica (cuadrantes opuestos).
- La distancia entre los puntos de cada categoría y el origen mide la calidad de la categoría de la variable en el mapa de factores. Los puntos de cada categoría que estén lejos del origen están bien representados en el mapa.
- Los colores de las variables están dados por el valor del Coseno-cuadrado que explica la calidad de representación.

Tras haber explicado lo que se observa genéricamente de este grafico, se procede a comentar las relaciones que se observan para este caso particular:

- Existe una fuerte asociación entre Almacenes, Cumple, Mediano y Pequeño. Pero teniendo en cuenta el bajo valor de Cos2 que tiene tanto la categoría de Mediano y Pequeño hay que tener cuidado con asociarlos a la condición de cumplimiento del estándar.



- Por otro lado se encuentra una relación entre NoCumple y Kioscos por lo que se puede establecer una correspondencia entre el hecho de que un cliente sea clasificado como Kiosco y este mismo no cumpla con el estándar.
- En contrapartida los Almacenes se encuentran a gran distancia de NoCumple como es el caso de los Kioscos con la condición de Cumple por lo que entre esas variables existe una correspondencia negativa.
- Por último se puede observar que es complicado relacionar a los locales clasificados como Otros y Grandes con la condición de que cumplan o no con el estándar.

RECOMENDACIONES

- En cuanto a la recolección de datos se recomienda para mejorar la calidad de análisis captar más datos que ayuden a crear un perfil de cliente como puede ser por ejemplo en que zona de la ciudad se encuentra ubicado el local del cliente.
- Referido a los resultados obtenidos la condición crítica para la organización es la de los clientes que no cumplen con el estándar que en este caso se dio que los que tenían más relación con esa condición son los denominados como kioscos, por lo que se recomienda complementar este análisis con un análisis de las posibles causas por las que se pueden encontrar relacionados los Kioscos con no cumplir con el estándar.

CONCLUSIONES

Teniendo en cuenta que los problemas que atraviesan las organizaciones se ven afectados por múltiples variables resulta oportuno destacar la importancia de las técnicas de análisis multivariante para poder analizar de manera correcta esas situaciones que afectan a las organizaciones con el fin de poder encontrar soluciones para esas problemáticas multivariadas.

Sera determinante para aplicar correctamente cualquier técnica de análisis multivariado el conocer y clasificar correctamente las variables que se están estudiando, así como también entender las escalas de medida de las mismas.

Se destaca al software RStudio como una importante herramienta para realizar diversos análisis estadísticos tanto por el funcionamiento del software como también por la gran comunidad científica que se encuentra por detrás de este programa.

En cuanto al análisis de correspondencias múltiples se la resalta como una técnica poderosa para analizar un conjunto de datos, sus individuos, variables y categorías con la finalidad de conocer las asociaciones entre las categorías de las variables para determinar perfiles de individuos que ayuden a las organizaciones a entender mejor a los individuos con los que se relaciona.



REFERENCIAS

Aldás, J., Uriel, E. (2017). *Análisis multivariante aplicado con R*. Paraninfo.

Anderson, R. Black, W.C. Hair, Jr. J.F. (1999). *Análisis Multivariante*. Madrid, España.

Editorial: Prentice Hall Iberia.

García, J., Molina, J., Berlanga, A., Patricio M., Bustamante, L., Padilla, W. (2018). *Ciencia de Datos. Técnicas analíticas y aprendizaje estadístico*. Publicaciones Altaria, S.L.

Hernández Sampieri, R., Mendoza, C. (2018). *Metodología de la investigación*.

Las rutas cuantitativa, cualitativa y mixta. McGraw-Hill.

Lind, D., Marchal, W., Wathen, S. (2012). *Estadística aplicada a los negocios y la economía*. McGraw-Hill.