



ENCUENTRO REGIONAL ADENAG
ZONAS CENTRO Y NOROESTE
19 y 20 mar 2020

ANÁLISIS DE CORRESPONDENCIA MÚLTIPLE CON SOFTWARE R COMO HERRAMIENTA PARA EL ANÁLISIS EXPLORATORIO DE DATOS

Miguel Arnone – Javier Antonio García

Facultad de Ciencias Económicas, Universidad Nacional de Tucumán

arnonemiquel1@gmail.com – jagarcia@face.unt.edu.ar

RESUMEN

El presente trabajo tiene como objetivo explorar e identificar relaciones entre determinadas características de los consumidores con ciertos atributos del calzado, las cuales se obtendrán con una encuesta. Lo dicho se realiza para un negocio de calzado femenino, de modo que, con la información obtenida, se pueda obtener un punto de partida para mejorar los procesos de compra y comercialización a través del consecuente análisis de datos. Para ello, se emplea como metodología el análisis de correspondencia múltiple, con el cual es posible analizar las variables bajo estudio de forma simultánea. Por último, se obtienen las conclusiones pertinentes al estudio.

PALABRAS CLAVE: Consumidores - Variables Cualitativas - Comercialización - TTDD

1. INTRODUCCIÓN

La transformación vertiginosa del mercado exige decisiones rápidas, las cuales deben estar fundamentadas y no deben surgir únicamente de la mera intuición o buen juicio de los administradores. Dicha fundamentación se basa en la recopilación y análisis de información pertinente y oportuna.

En este trabajo, se propone el Análisis de Correspondencia Múltiple como una herramienta complementaria para la interpretación de la información. A través del mismo, se realiza un estudio simultáneo de múltiples variables, lo cual nos permite obtener un entendimiento más amplio y profundo de los datos, en contraste al análisis estadístico convencional univariado.

Dicha herramienta será utilizada para mejorar las compras, precisamente la mezcla de productos, y la comercialización en un negocio dedicado a la venta de calzado femenino de la Provincia de Tucumán. El mismo, funciona desde fines del año 2017, en principio, con formato de showroom, para luego en 2019, expandirse y concretar la apertura de una sucursal en una de las ubicaciones más concurridas de San Miguel de Tucumán. El comercio cuenta actualmente con un proveedor principal y se encuentra en búsqueda de nuevos proveedores secundarios para abarcar un mayor segmento de consumidores. Actualmente, su mercado objetivo está determinado directamente por el producto que se compra al proveedor principal, el mismo, abarca una edad promedio de consumidoras de entre 16 a 20 años.

Al estar expuesto al consumo masivo, el comercio se enfrenta a una amplia gama de clientes, los cuales es necesario conocer, comprender e identificar. Al estar ligado a un solo proveedor, el negocio no explota sus ventajas de ubicación, así como las de su llegada por redes sociales, por lo que se encuentra acotado al segmento que dicta el producto que vende, sin posibilidades de expandir su mercado.

El objetivo de esta investigación es, a través de una encuesta, realizar un análisis exploratorio de los segmentos de clientes potenciales a los cuales el comercio puede abarcar. Asimismo, en base a la información obtenida, esbozar la posible oferta de nuevos productos y de su posterior comercialización.



ENCUENTRO REGIONAL ADENAG
ZONAS CENTRO Y NOROESTE
19 y 20 mar 2020

2. ENFOQUE METODOLÓGICO

El análisis de correspondencia es 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. El AC difiere de otras técnicas de interdependencia en su capacidad para utilizar tanto datos no métricos como relaciones no lineales (Hair, 1999)

Se utiliza para analizar, desde un punto de vista gráfico, las relaciones de dependencia e independencia de un conjunto de variables categóricas. Para ello asocia a cada una de las modalidades de la tabla, un punto en el espacio R^n (generalmente $n=2$) de forma que las relaciones de cercanía/lejanía entre los puntos calculados reflejen las relaciones de dependencia y semejanza existentes entre ellas. (Greenacre, 2008)

El objetivo del presente estudio es, a través del análisis de correspondencia múltiple, explorar la base de datos derivada de una encuesta realizada a los consumidores.

Las preguntas para la encuesta se eligieron en base a la segmentación demográfica de los consumidores. La demografía es el estudio de las poblaciones humanas en términos de magnitud, densidad, ubicación, edad, género, raza, ocupación y otros datos estadísticos (Kotler, 2007) principalmente en cuanto a la edad, establecida en segmentos según generación (Kotler, 2007), y en cuanto a la dedicación la cual fue definida según la principal actividad del encuestado, sin ahondar en las características de la misma para obtener mayor solidez en la recolección de datos.

En cuanto a las variables relacionadas a los atributos del calzado, se definieron ocho variables a analizar, dos de ellas relacionadas a los consumidores y las restantes relacionadas al calzado propiamente dicho. La categorización de las mismas fue realizada mediante el juicio de los investigadores en base al análisis del rubro al cual se dedica el negocio bajo estudio

En la actualidad las empresas reconocen que no pueden atraer a todos los compradores del mercado, o al menos no en la misma forma. Los compradores son demasiado numerosos, están ampliamente distribuidos, y tienen necesidades y prácticas de compra muy distintas. Además, las propias empresas varían mucho en su capacidad para atender a distintos segmentos del mercado. Por eso, una empresa debe identificar las partes del mercado a las que podría servir mejor, y de las cuales obtener mayores utilidades; necesita diseñar estrategias para desarrollar las relaciones adecuadas, con los clientes correctos (Kotler, 2007). En concordancia con lo que pregona Kotler, se pretende explorar los segmentos del mercado para poder comprender y satisfacer a los consumidores de forma más directa y estructurada.

El software para realizar el análisis es The R Project, el cual es un programa de fuente libre y código abierto para computación estadística y gráficos.

3. DESARROLLO

Inicialmente, se recopilaron datos sobre las variables bajo estudio a través de un cuestionario auto administrado, bajo el uso de la plataforma Google Forms, y distribuido principalmente mediante las redes sociales del negocio bajo estudio. Se obtuvieron 455 respuestas, por lo que la base de datos consiste en una matriz de 455 filas (individuos) y de 8 columnas (variables bajo estudio), para un total de 3640 observaciones. No obstante, ante las primeras pruebas de análisis, se decidió descartar algunas variables debido a que los resultados no fueron concluyentes. De este modo, la base de datos fue reducida (una parte de la misma se muestra en la Tabla 1), dejando cuatro variables, dos demográficas y dos de calzado para poder generar información relevante y útil. De esa forma, la base de datos a



ENCUENTRO REGIONAL ADENAG
ZONAS CENTRO Y NOROESTE
19 y 20 mar 2020

trabajar quedó conformada 455 respuestas sobre cuatro variables, brindando un total de 1820 observaciones.

Para realizar el análisis, se procedió a categorizar cada una de las variables bajo estudio. Una vez que se obtuvieron las respuestas al cuestionario, se procedió a depurar la base de datos a través del uso de preguntas filtro, las cuales fueron seleccionadas específicamente con ese el fin de apartar aquellos registros que no son relevantes para el estudio. Aquellas categorías eliminadas, se consensuaron con el comercio en cuestión, en base a los objetivos del mismo en congruencia con los de la investigación.

Una vez aplicado el filtro, las variables y categorías resultaron de la siguiente manera:

Edad	Dedicacion	Tipo Calzado	Altura
Menor de 18	Estudiante	Borcegos	Bajo
Entre 19 y 25	Trabajadora	Bota Caña alta	Medio
Entre 26 y 40		Bota caña baja	Alto
		Zapatillas	

Tabla 2: Variables y categorías filtradas.

De esta forma, la base de datos se conformó por un total final de 388 registros, de los cuales se obtuvieron 1552 observaciones de 4 variables.

A continuación, se procede a utilizar el software R para aplicar el análisis de correspondencia múltiple.

Primero, se copia la base de datos trabajada en Excel al software R. Para ello utilizamos el siguiente script:

```
Datos<-read.delim("clipboard", dec=",")
```

Con el mismo, ingresamos la base de datos de las respuestas de los encuestados al software. Seguido, se deben cargar los paquetes necesarios para correr el ACM con los siguientes scripts:

```
library("ggplot2")  
library("factoextra")  
library("FactoMineR")
```

Una vez listos los paquetes, se procede a ejecutar los comandos necesarios para realizar el análisis de correspondencia múltiple propiamente dicho:

```
res.mca <- MCA(Datos, graph = FALSE)  
print(res.mca)
```

A través del último script, se obtienen los resultados del análisis, los cuales se pueden apreciar:

name	description
1 "\$eig"	"eigenvalues"
2 "\$var"	"results for the variables"
3 "\$var\$coord"	"coord. of the categories"
4 "\$var\$cos2"	"cos2 for the categories"
5 "\$var\$contrib"	"contributions of the categories"



ENCUENTRO REGIONAL ADENAG
ZONAS CENTRO Y NOROESTE
19 y 20 mar 2020

6 "\$var\$v.test"	"v-test for the categories"
7 "\$ind"	"results for the individuals"
8 "\$ind\$coord"	"coord. for the individuals"
9 "\$ind\$cos2"	"cos2 for the individuals"
10 "\$ind\$contrib"	"contributions of the individuals"
11 "\$call"	"intermediate results"
12 "\$call\$marge.col"	"weights of columns"
13 "\$call\$marge.li"	"weights of rows"

Una vez ejecutado el análisis, se realiza el gráfico correspondiente mediante el siguiente script:

```
fviz_mca_var(res.mca, col.var = "blue", addEllipses = FALSE, repel = TRUE) +  
theme_minimal()
```

El gráfico obtenido es el siguiente:

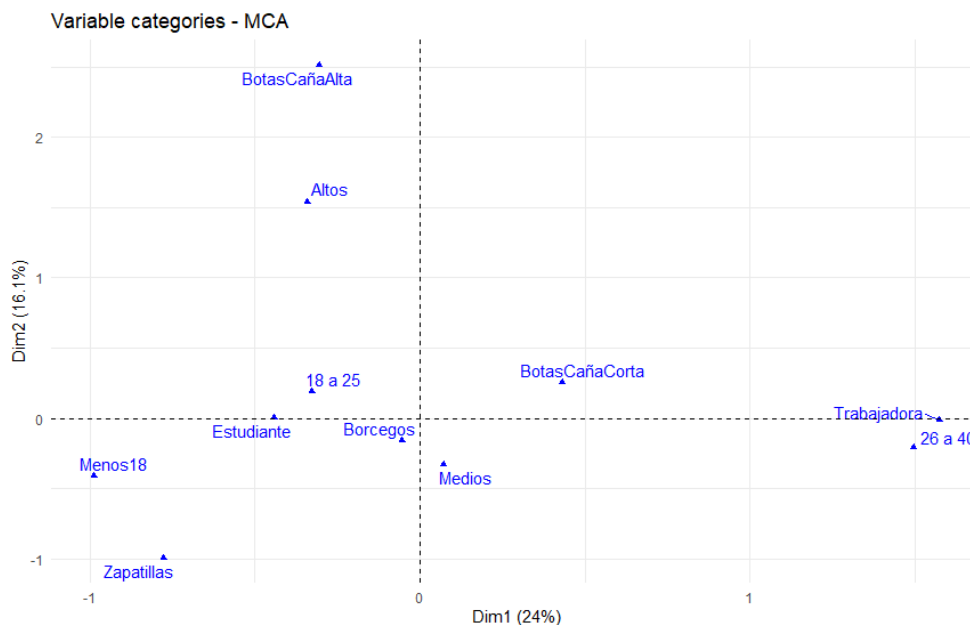


Gráfico 1: Análisis de correspondencia múltiples

En primera instancia, se puede apreciar una relación entre aquellas mujeres que tienen una edad mayor a 25 años y trabajan con un tipo de calzado preferentemente de bota con caña corta. Por otro lado, se percibe una relación entre aquellas jóvenes que tienen una edad menor a 25 años y estudian con el tipo de calzado borcego y zapatillas para aquellas más adolescentes.

Se determina cuáles son las variables que más contribuyen en cada eje:

```
fviz_contrib(res.mca, choice = "var", axes = 1, top = 4)
```



ENCUENTRO REGIONAL ADENAG ZONAS CENTRO Y NOROESTE 19 y 20 mar 2020

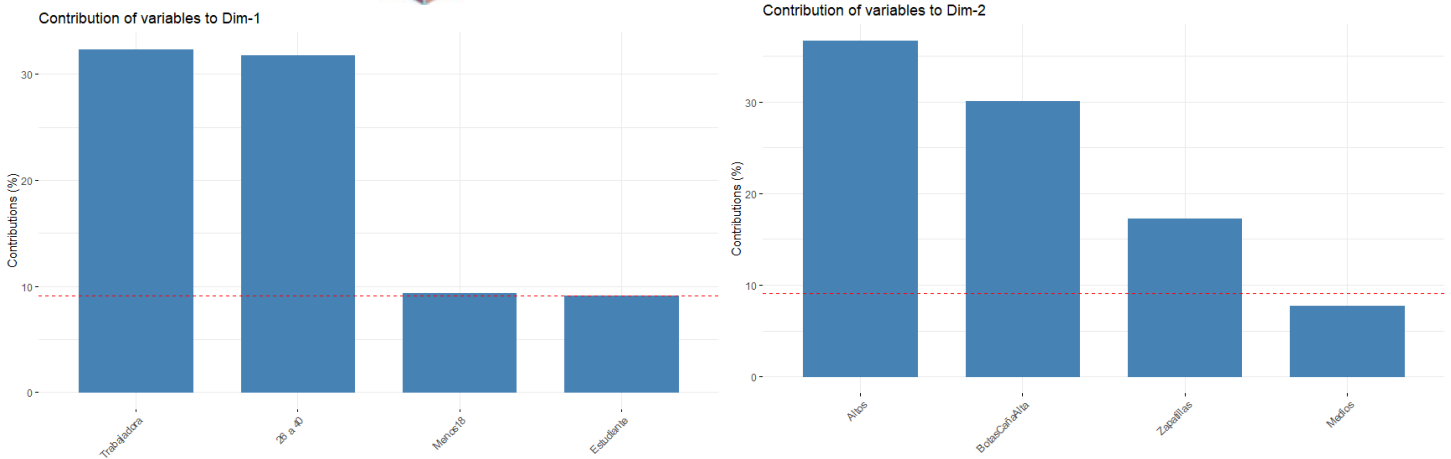


Gráfico 2: contribución de las vbles a los ejes X e Y

Se aprecia como en el eje X, la variable que mayor explicación tiene para dicho eje es la categoría “trabajadora”. Por otro lado, para el eje Y, sería la categoría “altos”, perteneciente a la variable Altura.

Igualmente, se logra distinguir que los ejes no son lo suficientemente significativos. Por ende, no se puede ser concluyente en cuanto a si las cuatro variables presentan una relación de manera simultánea. Por ello, se decidió ahondar en el análisis de las variables una a una, para lograr obtener información significativa.

De todos los análisis de correlación llevados a cabo, se resaltaron los siguientes:

- Análisis de correspondencia múltiple entre las variables: Edad, dedicación y altura.

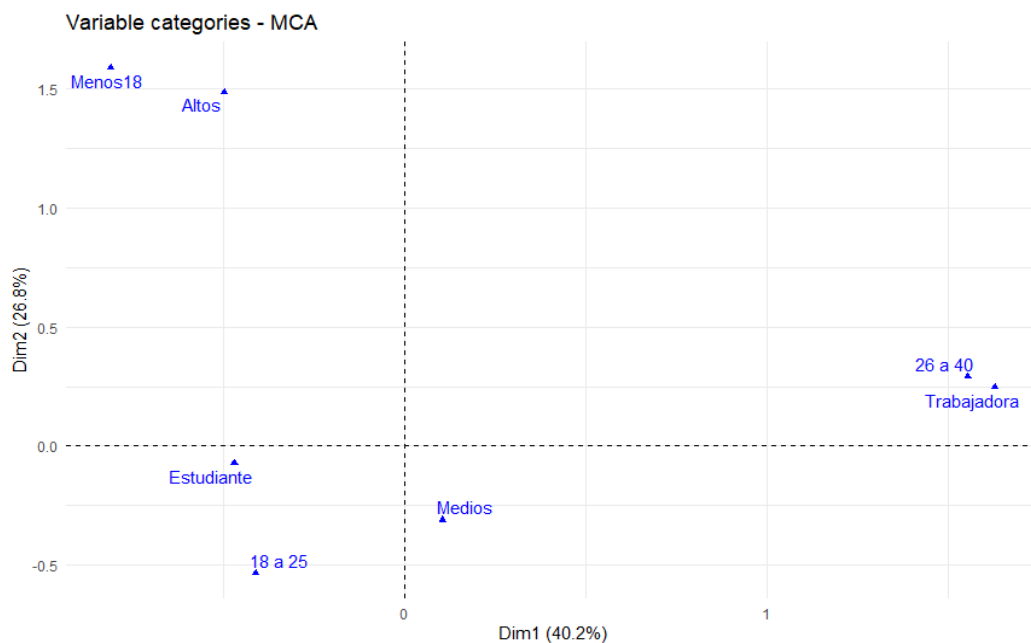


Gráfico 3: ACM de edad, dedicación y altura del calzado.



ENCUENTRO REGIONAL ADENAG
ZONAS CENTRO Y NOROESTE
19 y 20 mar 2020

Contribución de las variables a los ejes:

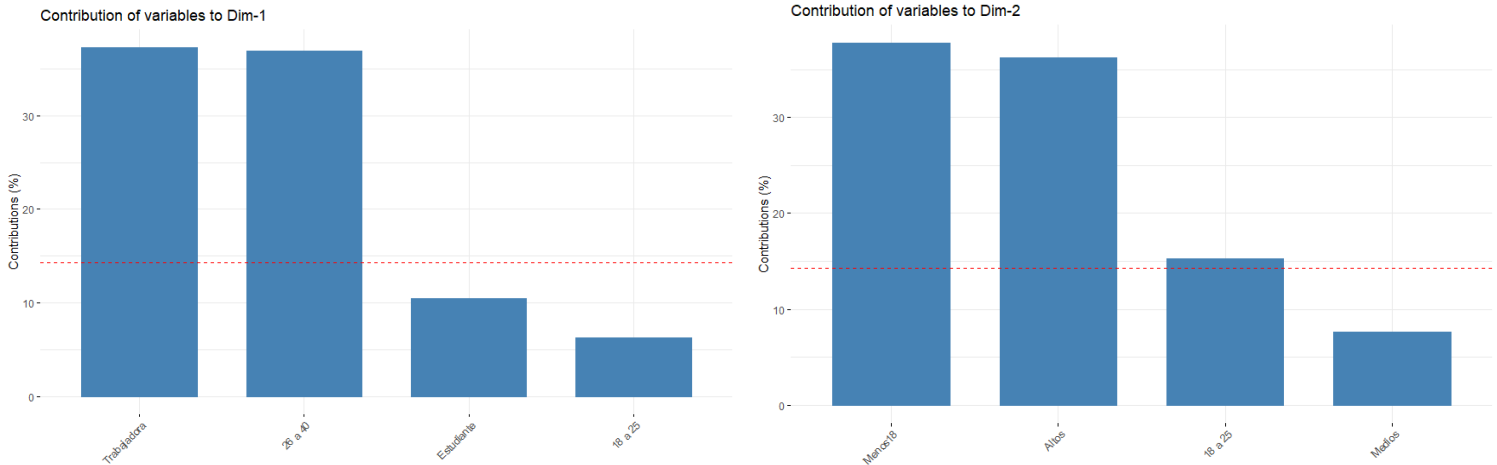


Gráfico 4: contribución de las variables a los ejes X e Y

En este caso, la variable que más contribuye a la explicación del eje X es “Trabajadora” y la que más contribuye al eje Y es “Menor de 18 años”.

➤ Análisis de correspondencia entre edad y tipo de calzado:

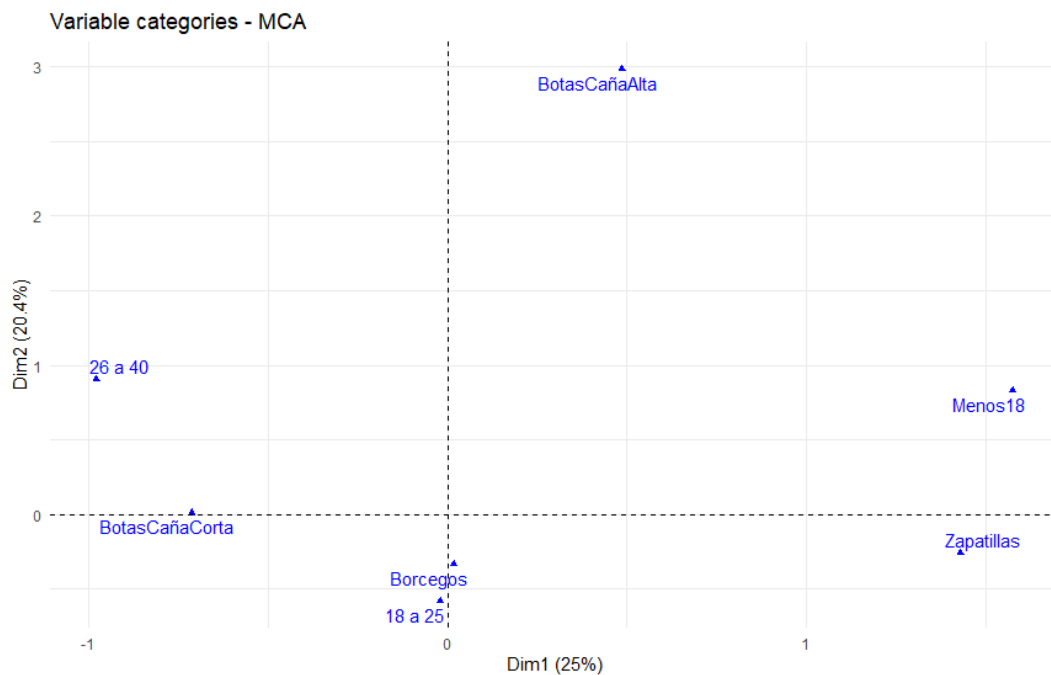


Gráfico 5: AC entre edad y tipo de calzado



ENCUENTRO REGIONAL ADENAG ZONAS CENTRO Y NOROESTE 19 y 20 mar 2020

Contribución de las variables a los ejes:

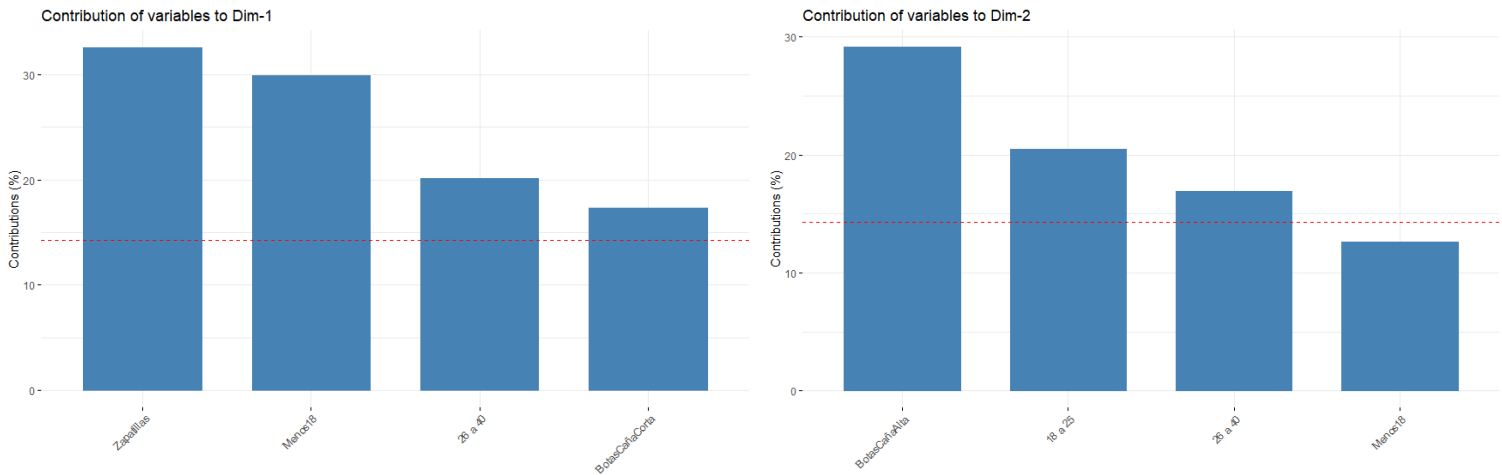


Gráfico 6: Contribución de las variables a los ejes X e Y

Es claro cómo, para el eje X, la variable que más contribuye es “Zapatillas”, categoría que pertenece a la variable Tipo de Calzado. Asimismo, para el eje Y, la variable que más contribuye es la categoría “Botas caña alta”.

4. ANALISIS DE RESULTADOS

Como se aprecia en el desarrollo, se obtuvieron pocos resultados significativos en relación al objetivo final del estudio. No obstante, existen variables que tienen un grado de relación aceptable, entre ellas;

- Gráfico 2: edad, dedicación y altura del calzado. Se puede entrever una relación entre aquellas consumidoras que tienen entre 18 y 25 años y que estudian con una altura de zapatos media. Asimismo, aquellas clientes menores de 18 años, prefieren los zapatos altos. Por último, se confirma que aquellas encuestadas mayores de 25 años tienden a trabajar como actividad principal.
- Gráfico 3: edad y tipo de calzado. Se puede apreciar que aquellas consumidoras de entre 18 a 25 años prefieren el calzado tipo borrego, mientras que aquellas menores de 18 años, tienen a elegir zapatillas. En cuanto al segmento más 25, el resultado no es concluyente.

5. CONCLUSIONES

Del estudio se puede concluir que los gustos y preferencias de los consumidores están determinados por otras (o más) variables fuera del alcance de este trabajo.

También, se recomienda la utilización y complementación del estudio con los datos de las ventas del comercio en cuestión. De este modo, se podrían determinar mezclas de producto en base a la concatenación de ambas bases de datos.

Por otro lado, se sugiere obtener los datos de los clientes reales a través de sistemas de información o de planillas organizadas, de modo que los mismos puedan ser analizados con herramientas de índole estadístico como el Análisis de Correspondencia Múltiple. Con



ENCUENTRO REGIONAL ADENAG
ZONAS CENTRO Y NOROESTE
19 y 20 mar 2020

esta metodología, se obtendrían datos sólidos con un alto grado de relevancia para determinar el mercado objetivo del negocio, así como para complementar la mezcla de producción y su posterior comercialización.

Se tiene en cuenta también que los canales de distribución del cuestionario fueron acotados, de modo que los resultados pueden estar condicionados en cierta forma por ello. Igualmente, un mayor número de encuestados se reflejaría en una mayor precisión a la hora de esbozar los resultados del estudio, ya que, a medida que la muestra crece, la misma se hace más representativa de la población bajo estudio.

6. TABLAS

n°	Edad	Dedicacion	Altura	Tipo
1	Menos18	Estudiante	Medios	Zapatillas
2	18 a 25	Estudiante	Medios	Zapatillas
3	26 a 40	Estudiante	Medios	Borcegos
4	26 a 40	Estudiante	Medios	BotasCañaCorta
5	Menos18	Estudiante	Altos	Borcegos
6	18 a 25	Estudiante	Altos	BotasCañaCorta
7	18 a 25	Estudiante	Medios	Borcegos
8	Menos18	Estudiante	Medios	Borcegos
9	18 a 25	Estudiante	Medios	Borcegos
10	18 a 25	Estudiante	Altos	Borcegos

Tabla 1: Primeras 10 observaciones de los encuestados

7. REFERENCIAS

HAIR (1999): Análisis Multivariante, quinta edición. Prentice Hall Iberia - Madrid, España.

GREENACRE (2008): La práctica del análisis de correspondencias. Fundación BBVA - España.

KOTLER (2007): Marketing, Versión para Latinoamérica, decimoprimer edición. PEARSON EDUCACION, México.