



UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

FACULTAD DE CIENCIAS

Posicionamiento en el mercado mediante técnicas
de análisis multivariado

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

Actuaria

PRESENTA:

Palacios Ramírez Karla Vianey

TUTORA

Ruth Selene Fuentes García



México, D.F.

2015



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

AGRADECIMIENTOS

A todas las personas que hicieron posible este proyecto,

A mi asesora de tesis, la Dra. Ruth Selene Fuentes García por la ayuda, paciencia y orientación que me brindo al realizar este proyecto, por su apoyo y amistad que me permitieron aprender más sobre temas relacionados con estadística, por ser una excelente profesora.

A mi papá que siempre me ha entendido y cuidado, por su apoyo incondicional y su cariño, por todos sus sacrificios.

A mi mamá por ser mi guía para poder llegar a este punto de mi carrera, que con su ejemplo y palabras siempre me apoyo en todas las decisiones importantes, por ser la mejor mamá, te amo.

A mis hermanas por siempre estar ahí apoyándome y alentándome a seguir adelante, gracias por todos los buenos momentos.

A mis abuelos que siempre fueron un apoyo y un pilar fundamental en mi trayectoria, por sus palabras y amor.

A mi abuelita Santa que fue una gran consejera y un gran apoyo en los momentos difíciles.

A mis amigos, que han estado conmigo en malos y buenos momentos, me han ayudado a seguir adelante y a nunca dejar mis sueños.

Índice general

1. INTRODUCCIÓN	3
2. FUNDAMENTO DEL PROBLEMA	6
3. MARCO TEÓRICO	11
3.1. CONCEPTOS GENERALES DE ÁLGEBRA MATRICIAL	11
3.2. ANÁLISIS DE COMPONENTES PRINCIPALES	16
3.3. ANÁLISIS DE FACTORES	20
3.3.1. ESTIMACIÓN EN EL MODELO DE FACTORES	22
3.3.2. ROTACIÓN EN EL MODELO DE FACTORES	23
3.4. CONSTRUCCIÓN DEL MAPA DE POSICIONAMIENTO	24
3.5. CONSIDERACIONES PARA LA CONSTRUCCIÓN DE UNA ESCALA, ALPHA DE CRONBACH	24
4. DESARROLLO DEL PROBLEMA	28
4.1. ELABORACIÓN DE LAS DIMENSIONES O EJES	28
4.2. ANÁLISIS DE FACTORES PARA LA CONSTRUCCIÓN DE ÍNDICES.	28
4.3. MODIFICACIÓN DE LA ESCALA ORIGINAL	31
4.4. ALPHA DE CRONBACH CALCULADA A LOS ÍNDICES	32
4.5. ANÁLISIS DE COMPONENTES PRINCIPALES APLICADO A LOS ÍN- DICES	35
4.6. REPRESENTACIÓN ESPACIAL DE LOS PRODUCTOS	37
4.7. CONSTRUCCIÓN DEL MAPA DE POSICIONAMIENTO	46
5. RESULTADOS	48
6. CONCLUSIONES	53
7. ANEXOS	55
8. BIBLIOGRAFÍA	58

Capítulo 1

INTRODUCCIÓN

Una empresa es la unidad productiva en la que se lleva a cabo la producción de bienes y servicios que sirven para satisfacer las necesidades de los individuos que componen una sociedad. La economía está formada por miles de empresas que interactúan dentro de los mercados competitivos. La exigencia de los consumidores de comprar productos de mayor calidad y precio competitivo es una de las partes determinantes sobre las decisiones de producción de una empresa, lo que las lleva a investigar su posicionamiento actual en el mercado y su proyección futura para llegar a obtener cierto poder de mercado. Las empresas normalmente reformulan su estrategia de venta varias veces durante la vida de un producto a medida que sus condiciones económicas y las del mercado cambian, la competencia lanza nuevos productos o los consumidores tienen nuevas exigencias; el impacto que pueda tener el posicionamiento de una marca en las tendencias del consumidor es parte del estudio fundamental que debe realizar cualquier empresa que pretenda que su producto se mantenga relevante dentro del mercado.

El análisis del posicionamiento de una marca en el mercado suele tomar dos referencias esenciales:

- Posicionamiento en segmentos de mercado: mediante el conocimiento del tamaño de los segmentos que constituyen la demanda en el mercado de un bien y la competencia en la parte productora de cada uno, la empresa debe decidir si va a especializarse en un solo segmento del mercado o se va a dirigir a un grupo específico de ellos.

- Posicionamiento frente a las marcas competidoras: una vez seleccionado el segmento en el cual se enfocará la empresa, es necesario conocer la percepción de los consumidores respecto a la oferta de la empresa y frente a la oferta de los competidores.

Los resultados del análisis del posicionamiento de una empresa se plasman de forma relacionada en un mapa de posicionamiento; un mapa de posicionamiento es una herramienta muy útil en el estudio de la mercadotecnia porque proporciona una representación visual de las percepciones de los consumidores sobre la marca y la competencia, utilizando atributos (dimensiones) relevantes para los consumidores.

Se tienen dos formas básicas de analizar el posicionamiento;

- Comparación de productos sin referirse a atributos concretos, utilizando la técnica de escalamiento multidimensional no métrico.
- Tener en cuenta las evaluaciones que reciben los atributos en cada producto, mediante el análisis de factores, el análisis cluster u otras técnicas multivariadas.

El objetivo de este trabajo es abordar el interés que tiene una empresa de posicionarse en el mercado; se pretenderá utilizar técnicas de análisis multivariado como lo es el análisis de componentes principales y el análisis de factores, técnicas que resultan adaptables a los propósitos de los investigadores de ésta área aunque con frecuencia no son usadas de manera adecuada.

El posicionamiento estadístico de productos requiere de dos partes: una serie de ejes o dimensiones que representarán los grupos principales de rasgos considerados por la población encuestada, y la ubicación en el espacio de cada uno de los productos. Se buscará crear un mapa de posicionamiento mediante el análisis de los factores o los componentes principales, donde los ejes del mapa serán extraídos utilizando el análisis estadístico pertinente y cada marca tendrá un factor de calificación que refleje el grado en que posee una característica específica.

En el presente proyecto se abordará uno de los mayores problemas que tienen las empresas al utilizar éstas técnicas, pues muchas veces los datos a trabajar son producto de una encuesta en donde las variables utilizadas no son continuas,

sino que en realidad se trata de variables ordinales o categóricas en su mayoría, y la teoría empleada en estas técnicas se basa en un contexto continuo que difícilmente se obtiene realizando dicha dinámica. Se propondrá corregir este problema mediante la creación de índices que contengan las variables originales, de modo que la construcción de la escala no afecte al contenido inicial y de igual manera permita utilizar los métodos multivariados de manera correcta para llegar a resultados consistentes.

El texto presente es una breve introducción a ciertos conceptos generales de utilidad, en el Capítulo II se mencionará el problema a resolver y las bases necesarias para lograrlo.

En el capítulo III y IV se tratan y definen los conceptos importantes para este proyecto, se muestran aspectos como el tipo de investigación, las técnicas y procedimientos que fueron utilizados para llevar a cabo dicha investigación.

En el capítulo V se presentarán los resultados obtenidos del análisis de datos y de la metodología propuesta, así como una estrategia de mercado a seguir.

Por último, en el capítulo VII se incluye un Anexo que consta del código realizado en R para obtener el análisis propuesto.

Capítulo 2

FUNDAMENTO DEL PROBLEMA

La motivación de este proyecto fue el interés por conocer cómo se comporta el mercado en un área específica, tratar de resolver la pregunta que tienen muchas empresas sobre los cambios que podrían realizar a sus productos para aumentar las ventas y estar entre las marcas competitivas. El impacto que pueda tener el análisis sobre el comportamiento o las características de una marca específica en una empresa puede traer grandes beneficios a futuro, como la reducción de costos y una mejor inversión en la renovación del producto para poder aumentar las ventas y por tanto las ganancias de la empresa.

Vamos a suponer que tenemos una empresa de botanas saladas y lanzamos al mercado nuestra marca llamada “PRODUCTO 1”, se buscará conocer el posicionamiento de nuestra marca respecto a los productos de competencia, en el estudio se comparará con 4 productos más y un producto “IDEAL” mediante un análisis estadístico basándonos en las respuestas de individuos acerca de factores específicos sobre el producto. Entenderemos por producto “IDEAL” al producto que la población consumidora considera que cumple con las características más importantes en una botana salada.

Se realizó una encuesta a 402 personas donde se preguntaban datos respecto a la botana ideal, además de ciertas características sobre cinco productos de botanas saladas que actualmente están a la vanguardia en el mercado, incluyendo el “PRODUCTO 1”.

Para tener información acerca de lo que los consumidores consideran su botana “IDEAL” se realizaron 39 preguntas referentes a la importancia de ciertos atri-

butos que pudieran tener los productos, las cuales contestaron con una escala ordinal de acuerdo al grado de importancia, donde el valor más pequeño corresponde a un grado de importancia menor y el valor más grande corresponde a un grado importancia mayor. La escala de respuesta es la siguiente:

1. NADA IMPORTANTE
2. POCO IMPORTANTE
3. NI MUCHO NI POCO IMPORTANTE
4. IMPORTANTE
5. MUY IMPORTANTE

La evaluación que hizo cada individuo fue su perspectiva respecto a qué tan importante consideraban las siguientes características:

1. Antojo del producto
2. Es crujiente
3. Su sabor es rico
4. Sabor a recién hechas
5. Puedes comerlas sin hartarte
6. Satisface tu antojo
7. Tiene una variedad de sabores
8. Es ligera, no llena.
9. Tiene empaque metalizado
10. Es para toda la familia
11. Es una marca de tradición
12. Tiene buena apariencia
13. Es un marca moderna y actual
14. Es para jóvenes

15. Es para gente como tú
16. Es para adultos
17. Es divertida
18. Es una marca confiable
19. Es una marca para niños
20. Es una marca de prestigio
21. Tiene un empaque atractivo
22. Lanza nuevos sabores constantemente
23. Tiene buena publicidad
24. Tiene buenas promociones
25. Tiene la cantidad suficiente de producto en la bolsa
26. Está elaborada con ingredientes naturales
27. Calma tu hambre
28. Vale lo que cuesta
29. Es un producto de calidad
30. La bolsa viene llena, no puro aire
31. Tiene un precio justo
32. Tiene las presentaciones que necesitas
33. Puedes consumirla en cualquier lugar
34. Es ideal para fiestas o reuniones
35. Es para compartir
36. Combina con otros ingredientes
37. Se puede encontrar en cualquier parte
38. Se puede consumir haciendo cualquier cosa

39. Puede acompañar otros alimentos

Con la información obtenida en las 39 preguntas realizadas podemos darnos una idea de la perspectiva que tienen los consumidores respecto a una botana (botana ideal), qué consideran como características básicas en el producto y cuáles son irrelevantes, para así poder construir las dimensiones o ejes del mapa de posicionamiento.

Realizaremos las mismas 39 preguntas a los 402 encuestados, pero ahora irán dirigidas a cada producto en particular, es decir, para cada uno de los cinco productos se preguntará qué tan de acuerdo está el encuestado en que el producto cumple con las propiedades, con el fin de poder tener una representación espacial de los productos en relación con los dos ejes del mapa de posicionamiento.

La respuesta sigue siendo una variable ordinal y su escala es la siguiente:

1. COMPLETAMENTE EN DESACUERDO
2. EN DESACUERDO
3. NI EN ACUERDO NI EL DESACUERDO
4. DE ACUERDO
5. COMPLETAMENTE DE ACUERDO

Se tiene una relación entre ésta escala y la anterior (Nada importante,..., Muy importante) ya que cuando los consumidores (encuestados) consideran que una característica es muy importante en una botana salada, le darán un valor de 5 y los productos que tengan esta característica estarán cumpliendo con las expectativas del consumidor. De igual modo cuando tome el valor de 1 significa que para los consumidores no es relevante el hecho de que un producto tenga esa característica.

En cuanto a la segunda escala cuando se le asigne el valor de 5= COMPLETAMENTE DE ACUERDO será equivalente a decir que el producto cumple con darle mucha importancia a esa característica y cuando se le asigne el valor de 1= COMPLETAMENTE EN DESACUERDO será equivalente a decir que el producto no le da importancia a dicha característica.

Nos gustaría que nuestra marca le diera importancia a las características que el consumidor considera relevantes y que ahorrara costos en las características que los consumidores consideren irrelevantes, o bien, que nuestro PRODUCTO 1 debe ser lo más parecido que sea posible a la botana “IDEAL”.

Capítulo 3

MARCO TEÓRICO

3.1. Conceptos generales de álgebra matricial

Notación de matrices y vectores:

X_{ij} = j-ésima variable del i-ésimo individuo

x_{ij} = valor observado de la j-ésima variable del i-ésimo individuo

$i = 1, \dots, n$ y $j = 1, \dots, p$

Esperanzas y varianzas de vectores aleatorios

Sea X un vector aleatorio de dimensión p .

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}$$

■ Esperanza:

$$\mu = E(X) = \begin{pmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix}_{p \times 1}$$

■ Matriz de Varianzas-Covarianzas:

$$\Sigma = Var(X) = Cov(X, X) = E\{(X - \mu)(X - \mu)'\} =$$

$$E \begin{pmatrix} (X_1 - \mu_1)^2 & (X_1 - \mu_1)(X_2 - \mu_2) & \cdots & (X_1 - \mu_1)(X_p - \mu_p) \\ (X_2 - \mu_2)(X_1 - \mu_1) & (X_2 - \mu_2)^2 & \cdots & (X_2 - \mu_2)(X_p - \mu_p) \\ \vdots & \vdots & \ddots & \vdots \\ (X_p - \mu_p)(X_1 - \mu_1) & (X_p - \mu_p)(X_2 - \mu_2) & \cdots & (X_p - \mu_p)^2 \end{pmatrix}$$

Los elementos de Σ se denotan como:

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix}_{(p \times p)}$$

donde $\sigma_{ij} = Cov(X_i, X_j)$

■ Correlaciones

$$P = Corr(X) = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{pmatrix}$$

donde $\rho_{ij} = Corr(X_i, X_j) = \frac{\sigma_{ij}}{\sqrt{\sigma_{jj}\sigma_{ii}}}$ para $i \neq j$.

Las estadísticas descriptivas (multivariadas), describen el comportamiento de la realización de una muestra aleatoria X_1, X_2, \dots, X_n de una distribución multivariada, es decir, para $i = 1, \dots, n$.

$$X_i = \begin{pmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{ip} \end{pmatrix}$$

cada X_i es una variable aleatoria multivariada de dimensión p .

Por lo tanto, un conjunto de datos está formado por n realizaciones de p variables aleatorias

$$X = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{pmatrix}_{(n \times p)}$$

- Varianza muestral:

$$\hat{\Sigma} = \frac{1}{n-1} \left\{ \sum_{i=1}^n (X - \mu)(X - \mu)' \right\}$$

sus elementos se denotan por:

$$\hat{\Sigma} = \begin{pmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{12} & \cdots & \hat{\sigma}_{1p} \\ \hat{\sigma}_{21} & \hat{\sigma}_{22} & \cdots & \hat{\sigma}_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_{p1} & \hat{\sigma}_{p2} & \cdots & \hat{\sigma}_{pp} \end{pmatrix}_{(p \times p)}$$

$$\text{donde } \hat{\sigma}_{ij} = \frac{1}{n-1} \left\{ \sum_{i=1}^n (X_{ij} - \mu_i)(X_{ij} - \mu_j)' \right\}$$

Definición 1 Determinante

El determinante de una matriz cuadrada $A = \{a_{ij}\}$ de $k \times k$ es:

$$\det(A) = |A| = \sum_{j=1}^k a_{ij} A_{ij}$$

donde $A_{ij} = (-1)^{1+j} |A^{1j}|$ y $|A^{1j}|$ es la matriz obtenida a partir de A al eliminar su primer renglón y su j-ésima columna.

Definición 2

Una matriz cuadrada es no singular si $\underset{(k \times k)}{A} \underset{(k \times 1)}{x} = \underset{(k \times 1)}{0}$ implica que $\underset{(k \times 1)}{x} = \underset{(k \times 1)}{0}$. Si una matriz no cumple eso se llamará matriz singular.

Definición 3

Sea $A = \{a_{ij}\}$ una matriz cuadrada de $k \times k$. La traza de la matriz A, denotada $\text{tr}(A)$, es la suma de los elementos de la diagonal, es decir,

$$\text{tr}(A) = \sum_{i=1}^k a_{ii}$$

Definición 4 Eigenvalores y eigenvectores

Los eigenvalores y eigenvectores de una matriz cuadrada $A = \{a_{ij}\}$ de $k \times k$ son aquellos que satisfacen

$$Ae = \lambda e$$

donde λ es un eigenvalor y e es un eigenvector.

Los eigenvectores se obtienen como solución a la ecuación:

$$|A - \lambda I| = 0$$

con I la matriz identidad.

La ecuación anterior la podemos expresar de la siguiente forma, como una ecuación polinomial en λ de grado k .

$$b_1\lambda^k + b_2\lambda^{k-1} + \dots + b_k\lambda + b_{k+1} = 0$$

cuyas raíces son los eigenvalores de A .

Si A es una matriz simétrica, sus eigenvalores son reales y pueden escribirse de forma descendente, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k$.

De modo que para cada eigenvalor λ_i existe un eigenvector e_i que satisface la ecuación

$$Ae = \lambda e$$

Propiedades

$$tr(A) = \sum_{i=1}^k \lambda_i$$

$$|A| = \prod_{i=1}^k \lambda_i$$

Definición 5

Una matriz es definida positiva si todos sus eigenvalores son positivos.

Definición 6

Una matriz es semi-definida positiva si todos sus eigenvalores son no negativos.

Resultado 1

Sean A y B matrices de $k \times k$ y c un escalar, se cumplen las siguientes propiedades:

1. $tr(cA) = ctr(A)$
2. $tr(A \pm B) = tr(A) \pm tr(B)$
3. $tr(AB) = tr(BA)$
4. $tr(B^{-1}AB) = tr(A)$
5. $tr(AA') = \sum_{i=1}^k \sum_{j=1}^k a_{ij}^2$

Lema 1 Descomposición Espectral

La descomposición espectral nos permite expresar la inversa de una matriz cuadrada en términos de sus eigenvectores y eigenvalores.

Sea $A_{k \times k}$ una matriz positiva definida con descomposición espectral.

$$A = \sum_{i=1}^k \lambda_i e_i e_i'$$

Sean los eigenvectores normalizados las columnas de otra matriz $P = [e_1, e_2, \dots, e_k]$

Entonces,

$$A_{k \times k} = \sum_{i=1}^k \lambda_i \begin{matrix} e_i & e_i' \\ (k \times 1) & (1 \times k) \end{matrix} = \begin{matrix} P & \Lambda & P' \\ (k \times k) & (k \times k) & (k \times k) \end{matrix}$$

donde $PP' = P'P = I$ y Λ es la matriz diagonal

$$\Lambda_{(k \times k)} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_k \end{bmatrix} \quad \text{con } \lambda_i > 0$$

Además

$$A^{-1} = P\Lambda^{-1}P = \sum_{i=1}^k \frac{1}{\lambda_i} e_i e_i'$$

Lema 2

Maximización de formas cuadráticas para puntos en el espacio unitario.

Sea $B_{p \times p}$ una matriz positiva definida con eigenvalores $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_p \geq 0$ y eigenvectores e_1, e_2, \dots, e_p normalizados.

Entonces,

$$\begin{aligned} \max_{x \neq 0} \frac{x' B x}{x' x} &= \lambda_1 \text{ sujeto a que } x = e_1 \\ \max_{x \neq 0} \frac{x' B x}{x' x} &= \lambda_p \text{ sujeto a que } x = e_p \end{aligned} \quad (2,1)$$

En general

$$\max_{x \perp e_1, e_2, \dots, e_k} \frac{x' B x}{x' x} = \lambda_{k+1} \text{ sujeto a que } x = e_{k+1} \text{ con } k = 1, 2, \dots, p-1. \quad (2,2)$$

3.2. Análisis de componentes principales

Uno de los problemas centrales del análisis de datos multivariados es la reducción de la dimensión, tratar de explicar p variables por un pequeño subconjunto $k < p$ de ellas a costa de una pequeña pérdida de información. El análisis de componentes principales busca explicar la estructura de la varianza y covarianza de un grupo de variables a través de combinaciones lineales de éstas. Su objetivo general es reducir la dimensión de los datos y poder interpretarlos de una manera más sencilla.

La principal utilidad de éste método es:

- Representar óptimamente en un espacio de dimensión pequeña observaciones de un espacio general p -dimensional.
- Transformar variables originales, en general correlacionadas, en un nuevo grupo de variables no correlacionadas, facilitando el manejo de datos para la verificación de agrupaciones.

Se necesitan p componentes para reproducir la variabilidad total del sistema, muchas veces la variabilidad puede ser explicada por un número k de componentes principales con $k < p$ en ese caso hay tanta información en los primeros k componentes como la hay en el total de los p componentes originales. Los k componentes principales que cumplan eso pueden reemplazar a las p variables iniciales, de modo que en lugar de tener n mediciones de p variables, se reducirá el conjunto de datos a n mediciones de k componentes principales.

Desde el punto de vista algebraico, las componentes principales son combinaciones lineales de las p variables observadas X_1, X_2, \dots, X_p .

Geométricamente, esas combinaciones lineales representan la selección de un nuevo sistema de coordenadas obtenido al rotar el sistema original con X_1, \dots, X_p como los ejes coordenados. Los nuevos ejes representan las direcciones de máxima variabilidad y proporcionan una descripción más simple de la estructura de la covarianza.

Los componentes principales dependen únicamente de la matriz de covarianza Σ (o bien la matriz de correlaciones ρ) de las X_1, X_2, \dots, X_p variables.

No se requiere asumir el comportamiento de una normal multivariada, sin embargo cuando los componentes principales se derivan de una población con el

comportamiento normal multivariado se tienen interpretaciones muy útiles en términos de la densidad constante de los elipsoides. Adicionalmente, se puede hacer inferencia sobre los componentes de la muestra cuando la población sigue la distribución normal multivariada.

Sea $X' = \{X_1, X_2, \dots, X_p\}$ el vector aleatorio cuya matriz de covarianza es Σ con valores propios (eigenvalores) $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_p \geq 0$.

Consideremos las siguientes combinaciones lineales

$$Y_1 = a'_1 X = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p$$

$$Y_2 = a'_2 X = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p$$

$$Y_3 = a'_3 X = a_{31}X_1 + a_{32}X_2 + \dots + a_{3p}X_p$$

...

$$Y_p = a'_p X = a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p$$

Calculando la respectiva varianza y covarianza tenemos lo siguiente:

$$Var(Y_i) = a'_i \Sigma a_i \text{ con } i = 1, 2, 3, \dots, p$$

$$Cov(Y_i, Y_k) = a'_i \Sigma a_k \text{ con } i, k = 1, 2, 3, \dots, p$$

Las componentes principales serán esas combinaciones lineales no correlacionadas $Y_1, Y_2, Y_3, \dots, Y_p$ cuyas varianzas resultan lo más grande posible.

La primer componente principal es la combinación lineal con máxima varianza, es decir, que maximiza $Var(Y_i) = a'_i \Sigma a_i$. Donde es claro que la varianza $Var(Y_i) = a'_i \Sigma a_i$ puede ser incrementada al multiplicar cualquier a_i por una constante muy grande, así que para eliminar el problema se agrega la restricción de que los coeficientes asociados a los vectores tengan norma uno.

Tomando en cuenta lo expuesto anteriormente definimos las componentes principales de la siguiente manera:

La primer componente principal es la combinación lineal $a'_1 X$ que maximiza $Var(a'_1 X)$ sujeto a que $a'_1 a_1 = 1$.

La segunda componente principal es la combinación lineal $a'_2 X$ que maximiza $Var(a'_2 X)$ sujeto a que $a'_2 a_2 = 1$ y $Cov(a'_1 X, a'_2 X) = 0$.

La tercer componente principal es la combinación lineal a'_3X que maximiza $Var(a'_3X)$ sujeto a que $a'_3a_3 = 1$ y $Cov(a'_3X, a'_2X) = Cov(a'_3X, a'_1X) = 0$.

Al i -ésimo paso, la i -ésima componente principal es la combinación lineal a'_iX que maximiza $Var(a'_iX)$ sujeto a que $a'_ia_i = 1$ y $Cov(a'_iX, a'_kX) = 0$ para $k < i$.

RESULTADO I

Sea Σ la matriz de covarianza asociada al vector aleatorio $X' = \{X_1, X_2, \dots, X_p\}$ y suponga que Σ tiene las siguientes parejas ordenadas de eigenvalores - eigenvectores $(\lambda_1, e_1), (\lambda_2, e_2), (\lambda_3, e_3), \dots, (\lambda_p, e_p)$ donde $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_p \geq 0$. Entonces la i -ésima componente principal está dada por

$$Y_i = e'_iX = e_{i1}X_1 + e_{i2}X_2 + \dots + e_{ip}X_p \quad i = 1, 2, 3, \dots, p$$

con

$$\begin{aligned} Var(Y_i) &= e'_i\Sigma e_i = \lambda_i & i = 1, 2, 3, \dots, p \\ Cov(Y_i, Y_k) &= e'_i\Sigma e_k = 0 & i \neq k \end{aligned}$$

Si alguna λ_i es igual, implicaría que los correspondientes vectores e_i son iguales y por tanto los componentes principales Y_i no son únicos.

Demostración

Usando 2.1 del *Lema 2*, con $B = \Sigma$ se tiene que:

$$\max_{a \neq 0} \frac{a'\Sigma a}{a'a} = \lambda_1 \quad \text{cuando } a = e_1$$

Pero $e'_1e_1 = 1$ pues los eigenvectores están normalizados, entonces

$$\max_{a \neq 0} \frac{a'\Sigma a}{a'a} = \lambda_1 = \frac{e'_1\Sigma e_1}{e'_1e_1} = e'_1\Sigma e_1 = Var(Y_1)$$

De manera similar usando 2.2 del *Lema 2*, obtenemos lo siguiente

$$\max_{a \perp e_1, e_2, \dots, e_k} \frac{a'\Sigma a}{a'a} = \lambda_{k+1} \quad k = 1, 2, \dots, p-1$$

Como $a = e_{k+1}$, con $e'_{k+1}e_i = 0$ se tiene que $Var(Y_{k+1}) = \lambda_{k+1}$.

Queda por demostrar que e_i es ortogonal a e_k (es decir, $e'_ie_k = 0$, $i \neq k$) resultando $Cov(Y_i, Y_k) = 0$. Los eigenvectores de Σ son ortogonales si todos los $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_p$ son distintos. Si los eigenvalores no son todos distintos, el eigenvector correspondiente a los eigenvalores comunes puede elegirse de forma que resulte ortogonal. Por lo tanto, para cualquiera dos eigenvectores e_i, e_k , se tiene $e'_ie_k = 0$ si $i \neq k$. Ya que $\Sigma e_k = \lambda_k e_k$, si multiplicamos por e'_i del lado

izquierdo, resulta:

$$Cov(Y_i, Y_k) = e_i' \Sigma e_k = e_i' \lambda_k e_k = \lambda_k e_i' e_k = 0$$

para toda $i \neq k$.

RESULTADO II

Sea Σ la matriz de covarianza asociada al vector aleatorio $X' = \{X_1, X_2, \dots, X_p\}$ y suponga que Σ tiene las siguientes parejas ordenadas de eigenvalores - eigenvectores $(\lambda_1, e_1), (\lambda_2, e_2), (\lambda_3, e_3), \dots, (\lambda_p, e_p)$ donde $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_p \geq 0$.

Sean

$$Y_1 = e_1' X, Y_2 = e_2' X, \dots, Y_p = e_p' X$$

los componentes principales, entonces

$$\sum_{i=1}^p Var(X_i) = \lambda_1 + \lambda_2 + \dots + \lambda_p = \sum_{i=1}^p Var(Y_i)$$

Demostración

Por la *Definición 3*, tenemos que $\sigma_{11} + \sigma_{22} + \dots + \sigma_{nn} = tr(\Sigma)$. Usando *Lema 1* con $A = \Sigma$ podemos escribir $\Sigma = P \Lambda P'$ donde Λ es la matriz diagonal de eigenvalores en sus entradas y $P = [e_1, e_2, \dots, e_p]$ tal que $PP' = P'P = I$, usando propiedades de la traza de una matriz tenemos lo siguiente:

$$tr(\Sigma) = tr(P \Lambda P') = tr(\Lambda P' P) = tr(\Lambda) = \lambda_1 + \lambda_2 + \dots + \lambda_p$$

Por lo tanto

$$\sum_{i=1}^p Var(X_i) = tr(\Sigma) = tr(\Lambda) = \sum_{i=1}^p Var(Y_i)$$

El *RESULTADO II* implica que la proporción total de varianza explicada por el i -ésimo componente es

$$\frac{\lambda_i}{\lambda_1 + \lambda_2 + \dots + \lambda_p} \quad i = 1, 2, \dots, p.$$

Si la mayoría de la varianza total (digamos el 80 % o 90 %) puede ser explicada por el primer, segundo o tercer componente, entonces estos componentes pueden reemplazar a las p variables originales sin la pérdida de mucha información.

3.3. Análisis de factores

El objetivo principal del análisis de factores es describir la relación de covarianza entre variables, mediante cantidades subyacentes pero no observadas llamadas factores. Supone que las variables pueden ser agrupadas conforme a su correlación, es decir, que las variables dentro de un grupo particular están altamente correlacionadas entre ellas y tienen una correlación baja con las variables en otros grupos, así cada grupo de variables es lo que constituye un factor.

Consideremos el vector aleatorio X , con p variables, media μ y matriz de covarianzas Σ , entonces el modelo de factores sostiene que X es linealmente independiente dado F_1, F_2, \dots, F_m factores aleatorios y p fuentes adicionales de variación llamadas errores $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$ que dependen únicamente de la variable respuesta X_i .

De tal modo que tenemos el siguiente modelo de factores:

$$X_1 - \mu_1 = l_{11}F_1 + l_{12}F_2 + \dots + l_{1m}F_m + \varepsilon_1$$

$$X_2 - \mu_2 = l_{21}F_1 + l_{22}F_2 + \dots + l_{2m}F_m + \varepsilon_2$$

$$X_3 - \mu_3 = l_{31}F_1 + l_{32}F_2 + \dots + l_{3m}F_m + \varepsilon_3$$

...

$$X_p - \mu_p = l_{p1}F_1 + l_{p2}F_2 + \dots + l_{pm}F_m + \varepsilon_p$$

O bien, en su forma matricial tenemos:

$$X_{(p \times 1)} - \mu_{(p \times 1)} = L_{(p \times m)}F_{(m \times 1)} + \varepsilon_{(p \times 1)} \dots (1)$$

Donde la matriz L llamada matriz de cargas, contiene las cargas l_{ij} de la i -ésima variable en el j -ésimo factor.

De modo que podemos ver a $X_1 - \mu_1, \dots, X_p - \mu_p$ en términos de $p + m$ variables aleatorias no observables.

Haremos algunos supuestos acerca de los vectores aleatorios F y ε en el modelo (1) para poder establecer algunas relaciones entre las covariables.

Supondremos que:

$$E[F] = 0_{(m \times 1)}$$

$$Cov[F] = E[FF'] = I_{(m \times m)}$$

$$E[\varepsilon] = 0_{(p \times 1)}$$

$$Cov[\varepsilon] = E[\varepsilon\varepsilon'] = \Psi_{(p \times p)} = \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \psi_p \end{bmatrix}$$

Además que $F \perp \varepsilon$ (F independiente de ε) entonces:

$$Cov[\varepsilon, F] = E[\varepsilon F'] = 0_{(p \times m)}$$

El modelo anterior es conocido como modelo ortogonal de factores e implica una estructura para la covarianza de X de la siguiente forma:

$$\begin{aligned} \Sigma &= E[(X - \mu)(X - \mu)'] = E[(LF + \varepsilon)(LF + \varepsilon)'] \\ &= E[(LF + \varepsilon)((LF)' + \varepsilon')] \\ &= E[(LF)(LF)' + \varepsilon(LF)' + LF\varepsilon' + \varepsilon\varepsilon'] \\ &= LE(FF')L' + E(\varepsilon F')L + LE(F\varepsilon') + E(\varepsilon\varepsilon') \\ &= LL' + \Psi \end{aligned}$$

Además

$$Cov(X, F) = E(X - \mu)F' = LE(FF') + E(\varepsilon F') = L$$

El modelo $X - \mu = LF + \varepsilon$ es lineal respecto a los factores F_i . Ésta es una condición para que la matriz de covarianzas de X tenga la forma expresada anteriormente, pero por construcción del modelo se cumple.

La estructura de la covarianza en el modelo ortogonal de factores se ve de la siguiente forma:

1) $Cov(X) = LL' + \Psi$ es decir

- $Var(X_i) = l_{i1}^2 + \dots + l_{im}^2 + \psi_i$
- $Cov(X_i, X_k) = l_{i1}l_{k1} + \dots + l_{im}l_{km}$

2) $Cov(X, F) = L$ es decir

- $Cov(X_i, F_j) = l_{ij}$

Veamos que la $Var(X_i) = l_{i1}^2 + \dots + l_{im}^2 + \psi_i$ se debe a una parte consistente en la proporción contribuida de varianza de la i -ésima variable por los m factores

comunes, llamada comunalidad y otra parte debida a la proporción de varianza específica de la variable X_i .

De modo que la expresión anterior la podemos ver como:

$$\text{Var}(X_i) = h_i^2 + \psi_i$$

con $h_i^2 = l_{i1}^2 + \dots + l_{im}^2$ (comunalidad).

No siempre se puede factorizar la matriz de covarianzas Σ de las variables observadas de la forma $LL' + \Psi$ cuando $m \ll p$, por lo que existen métodos para aproximarla.

3.3.1. Estimación en el modelo de factores

En la práctica debemos encontrar una estimación \hat{L} de las cargas L y una estimación $\hat{\Psi}$ de la varianza específica Ψ , tales que $S = \hat{L}\hat{L}' + \hat{\Psi}$ donde S es la matriz de covarianza empírica de X .

Los métodos más comunes para encontrar la estimación son los siguientes:

El método de máxima verosimilitud, está basado en asumir normalidad de los factores y los errores. Las ecuaciones resultantes de maximizar la log-verosimilitud asumiendo $S = LL' + \Psi$, son complicadas y se resuelven por medio de algoritmos numéricos.

El método de factores principales con una estimación preliminar de \hat{h}_i^2 y $\hat{\psi}_i = 1 - \hat{h}_i^2$. En el siguiente paso, la matriz de cargas es estimada por la descomposición espectral de la matriz de covarianzas reducida $S - \hat{\Psi}$, el procedimiento se repite hasta alcanzar la convergencia.

El método de componentes principales, se obtienen cargas estimadas \hat{L} mediante la descomposición de la matriz S .

La matriz estimada de cargas está dada por:

$$\hat{L} = [(\hat{\lambda}_1 \hat{e}_1)^{1/2}; (\hat{\lambda}_2 \hat{e}_2)^{1/2}; \dots; (\hat{\lambda}_m \hat{e}_m)^{1/2}]$$

con $(\hat{\lambda}_i \hat{e}_i)$ el i -ésimo par de eigenvalor-eigenvector de la matriz S .

La varianza específica será estimada por los elementos de la diagonal de la matriz $S - \hat{L}\hat{L}'$, entonces

$$\hat{\Psi}_{(p \times p)} = \begin{bmatrix} \hat{\psi}_1 & 0 & \cdots & 0 \\ 0 & \hat{\psi}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \hat{\psi}_p \end{bmatrix}$$

con $\Psi_i = s_{ii} - \sum_{j=1}^m \hat{l}_{ij}^2$

y la comunalidad es estimada como:

$$\hat{h}_i^2 = \hat{l}_{i1}^2 + \dots + \hat{l}_{im}^2$$

3.3.2. Rotación en el modelo de factores

Si $\hat{L}_{p \times m}$ es la matriz estimada de las cargas de los factores, obtenida por cualquiera de los métodos anteriores entonces,

$$\hat{L}^* = \hat{L}T$$

donde $TT' = T'T = I$ es una matriz de rotación de las cargas.

Sin embargo, la matriz de covarianzas no cambia, pues:

$$\hat{L}\hat{L}' + \hat{\Psi} = \hat{L}TT'\hat{L}' + \hat{\Psi} = \hat{L}^*\hat{L}^{*'} + \hat{\Psi}$$

y la matriz de residuos

$$S - \hat{L}\hat{L}' - \hat{\Psi} = S - \hat{L}^*\hat{L}^{*'} - \hat{\Psi}$$

de modo que \hat{h}_i^2 y $\hat{\psi}_i$ no se alteran.

En el caso de $m = 2$, podemos graficar el par de cargas de factores $(\hat{l}_{i1}, \hat{l}_{i2})$ para las p variables, en ese caso la rotación se puede hacer hacia algún ángulo específico, usando

$$\hat{L}^* = \hat{L}T$$

donde

$$T = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix}$$

es la rotación en el sentido de las manecillas del reloj, y

$$T = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}$$

es la rotación en el sentido contrario de las manecillas del reloj.

Al aplicar una rotación a los factores se puede facilitar la interpretación.

3.4. Construcción del mapa de posicionamiento

El análisis de factores agrupará rasgos que los encuestados consideren como similares, identifica a los grupos cuyo tamaño depende del número de declaraciones que son consideradas como similares, de modo que los ejes del mapa de posicionamiento de factores pueden ser determinados antes del análisis.

El primer eje del mapa de posicionamiento será el correspondiente al primer factor (el grupo más grande), el segundo eje del mapa será el correspondiente al segundo factor (el segundo grupo más grande) y así sucesivamente.

Es importante considerar las variables que podrían estar influyendo en la construcción de estos ejes, como el entorno donde fue realizada la encuesta, las bases que consideran las personas encuestadas para determinar si un producto es similar o es distinto, también lo que los encuestados buscan de un producto.

3.5. Consideraciones para la construcción de una escala, Alpha de Cronbach

Las escalas son instrumentos de aplicación para la evaluación de un atributo específico. La construcción de escalas puede definirse como la asignación de un número a objetos o a eventos mediante una regla específica, donde los números asignados son conocidos como “puntajes de la escala” y el método o regla que se utilice para asignar los puntajes es conocido como “escala”.

Existen diferentes formas de construir una escala, ya sea de manera nominal, ordinal o por intervalos, el problema surge cuando se trata de asignar puntajes a variables que no son tangibles como habilidad, necesidad, valor, actitud o preferencia dado que no pueden ser observadas de manera directa y dependerán de ciertos factores psicológicos. Es preferible que los ítem estén formulados de manera que no susciten respuestas socialmente deseables.

Utilizaremos un método para determinar si nuestra escala es confiable y así poder confiar en las respuestas proporcionadas, lo haremos mediante la alpha de Cronbach.

La principal ventaja de ese método es que lo único que requiere es una administración de la prueba; además, es un coeficiente sencillo de calcular. Mide la fiabilidad de la prueba en función de la longitud del test y la proporción

de varianza total de la prueba debida a la covarianza entre sus partes, analiza concretamente la consistencia interna de la escala como una dimensión de su fiabilidad mediante el cálculo de la correlación entre los ítem de la escala.

Por tanto, el estadístico alpha de Cronbach puede ser considerado como un coeficiente de correlación.

Nos indicará si los diferentes ítem de la escala están midiendo una realidad común, debemos tener una elevada correlación en las respuestas a estos ítem de otro modo algunas declaraciones de la escala no son medidas fiables.

Los valores de la alpha de Cronbach oscilan entre 0 y 1, cuando toma el valor de cero nos diría que los ítem no están correlacionados con los demás y viceversa.

La denotaremos como α_c y está dada por la siguiente expresión:

$$\alpha_c = \left(\frac{k}{k-1} \right) \left(1 - \frac{\sum_{i=1}^k S_i^2}{S_{sum}^2} \right)$$

donde k es el número de ítem en la prueba, S_i^2 es la varianza y S_{sum}^2 es la varianza de la prueba total.

No existe un regla concreta para interpretar en que rango puede considerarse aceptable el valor de α_c , para éste trabajo tomaremos como criterio general, George y Mallery (2003, p231) que sugieren las recomendaciones siguientes para evaluar los coeficientes del alpha de Cronbach (Tabla 3.1).

RANGO	CRITERIO
$\alpha_c < ,5$	inaceptable
$,5 \leq \alpha_c < ,6$	pobre
$,6 \leq \alpha_c < ,7$	cuestionable
$,7 \leq \alpha_c < ,8$	aceptable
$,8 \leq \alpha_c < ,9$	bueno
$,9 \leq \alpha_c$	excelente

Cuadro 3.1: Evaluación del coeficiente.

Encontramos diversas limitaciones con el alpha de Cronbach ya que es una cota inferior a la fiabilidad y en muchos casos suele ser una subestimación. Existen mejores alternativas para el alpha pero apenas se conocen y mucho menos se utilizan para evaluar la fiabilidad, otro problema es que es incorrectamente to-

mado como una medida de la estructura interna de la prueba y por lo tanto como argumento de que los elementos de la prueba "miden lo mismo". Sin embargo, esta estadística no proporciona al investigador este tipo de información, el resultado de esta mala interpretación es que un valor alto implica un rasgo de validez que a menudo se da por sentado cuando, de hecho, no ha sido investigado en absoluto. Esto produjo varios resultados interesantes que fueron recogidos por muchos psicólogos y llevaron a la interpretación de alfa como una medida de la consistencia interna de una prueba.

Es seguro que la interpretación que se refiere al alpha de Cronbach como medida de consistencia interna ha ganado más presencia que la interpretación de cota inferior. La consistencia interna se refiere a la interrelación de un conjunto de elementos. En la construcción de la prueba práctica, el uso de alpha a menudo va mano a mano con el análisis de componentes principales. Una interpretación generalizada aunque informal de consistencia interna de la prueba es que el primer valor propio de la matriz de correlación íter ítem está altamente relacionado con el segundo valor propio pero no es claro qué tan alta es la relación.

Por lo tanto, todo lo que alpha puede revelar acerca de la interrelación de los elementos es su grado medio de interrelación, teniendo en cuenta que el alfa también depende del número de elementos en la prueba, pero esto en realidad dice muy poco o nada de la consistencia interna, sin importar la forma en que se defina. El pensar en un solo número alpha que exprese tanto la fiabilidad y consistencia interna de la prueba, que sugiere que los ítems "midan lo mismo", es una medida para la evaluación de la calidad de la prueba. Mientras tanto, como se mencionó anteriormente el alpha de Cronbach es sólo una cota inferior de la fiabilidad y ni siquiera una realista.

Se podría argumentar que no se pierde nada al usar una cota inferior como lo es el alpha de Cronbach en la práctica, porque las estimaciones de baja fiabilidad innecesarias pueden tener un efecto positivo hacia el investigador al hacer todo lo posible para construir una prueba de alta calidad, así se puede concluir que el alpha dice muy poco o nada sobre estructuras de elementos de múltiples factores. Por otra parte, depende del número de elementos y se muestra que el alfa crece a medida que crece el número de elementos.

Es difícil defender convincentemente usando uno de los límites inferiores más pequeños (alpha), dada la disponibilidad de muchas cotas inferiores más grandes, por ejemplo la cota "GLB" (Greatest Lower Bound) que al tomar el extremo inferior representa la fiabilidad más pequeña posible de acuerdo a la matriz de covarianza observada bajo la restricción de que la suma de los errores de las

varianzas se maximiza por errores que tienen correlación cero con otras variables. Para los fines del análisis propuesto, el alpha de Cronbach resulta ser una estadística de gran ayuda, pues buscamos un acercamiento a la fiabilidad de la prueba, es decir, si la escala propuesta es correcta o bien, no captura la información general sobre los datos originales. Teniendo en cuenta las limitaciones que pudiéramos encontrar en el alpha de Cronbach, con reserva puede ser útil para analizar la interrelación de las variables propuestas en la escala.

Capítulo 4

DESARROLLO DEL PROBLEMA

4.1. Elaboración de las dimensiones o ejes

Se analizaron las respuestas dadas por el encuestado respecto a la importancia de ciertas características en una botana salada, se supuso que las personas encuestadas realmente sabían la importancia que tiene cada característica y también suponiendo que están diciendo lo que realmente perciben, en resumen estas calificaciones nos reflejan lo que las personas creen que influye más al momento de elegir una botana salada en el mercado. En el caso de la encuesta estudiada, se tiene una escala de 5 puntos para cada una de las 39 características preguntadas, respecto a la importancia de éstas, dicha escala como se mencionó anteriormente es ordinal, pues el valor 5 se refiere que la presencia de cierta característica es muy importante y de ahí van decreciendo en orden de importancia hasta llegar al valor de 1.

4.2. Análisis de factores para la construcción de índices.

El análisis de factores de las calificaciones anteriores realizado a las 39 variables (características) produjo dos grupos básicos de rasgos, las cargas de los factores se observan en el Cuadro 4.1. El análisis realizado se hizo tomando en cuenta dos factores, pues éstos son los que se proponen como ejes del mapa de posicionamiento; sin embargo, como se observa en el Cuadro 4.1 el porcentaje que se está

explicando con sólo dos factores es del 26 % lo cual es un resultado muy pobre y de poca ayuda para los propósitos del análisis. Nos arroja cierta incertidumbre del buen ajuste del modelo a los datos.

Uno de los objetivos principales de este proyecto fue el hacer un uso correcto de las herramientas de análisis estadístico para realizar una propuesta o una mejora en el producto analizado; generalmente se tiene poco cuidado en analizar el cumplimiento de ciertos supuestos o en la potencia de la prueba utilizada, por ejemplo en éste análisis observamos que el p-valor para la prueba χ^2 (ji-cuadrada) es muy pequeño por lo que se rechaza la hipótesis nula (H_0) donde dos factores son suficientes, de modo que los resultados arrojados por este análisis son poco confiables y es necesaria otra rutina estadística para poder dar una perspectiva correcta del problema.

Se realizó el análisis para 3 factores (Cuadro 4.2), tratando de corregir el porcentaje de varianza acumulado, se observó que el porcentaje de varianza acumulada fue de 28 % la cual sigue siendo muy pobre para el fin del análisis, usar más factores complica la idea de realizar un mapa de posicionamiento.

Como se mencionó anteriormente el problema con estas técnicas es la falta de atención a la cantidad de información verdadera arrojada por los métodos, bien podríamos tener algún resultado coherente, sin embargo no siempre sería real, en algunos casos podríamos darle alguna idea a la empresa para mejorar su posicionamiento en el mercado usando las cargas obtenidas en el análisis pero al tener un 26 % o 28 % de información capturada, es probable que las propuestas no estén bien dirigidas, o bien, no se refleje una mejora significativa en la empresa. Una alternativa para corregir este problema, es el realizar índices alternativos que permitan reducir un poco el número de variables (39) de modo que la escala discreta no afecte tanto y se pueda realizar otro tipo de análisis estadístico, como un análisis de componentes principales.

	<i>Factor1</i>	<i>Factor2</i>
<i>Empaque metalizado</i>	0,165	0,184
<i>Cantidad suficiente en la bolsa</i>		0,567
<i>Se le antoja</i>	0,142	0,491
<i>Ingredientes naturales</i>	0,214	0,424
<i>Presentaciones necesarias</i>	0,426	0,226
<i>Para la familia</i>	0,450	0,169
<i>Crujiente</i>	0,293	0,320
<i>Marca de tradición</i>	0,543	
<i>Buena apariencia</i>	0,482	0,305
<i>Consumir en cualquier lugar</i>	0,480	0,243
<i>Moderna y actual</i>	0,494	0,173
<i>Nuevos sabores</i>	0,345	0,317
<i>Calma el hambre</i>	0,252	0,318
<i>Sabor rico</i>	0,147	0,495
<i>Para jóvenes</i>	0,550	
<i>Para compartir</i>	0,464	0,110
<i>Sabor a recién hechas</i>	0,215	0,479
<i>Ideal para fiestas</i>	0,382	0,290
<i>Para gente como tú</i>	0,488	0,277
<i>Vale lo que cuestan</i>		0,455
<i>Producto de calidad</i>	0,215	0,396
<i>Combina con otros ingredientes</i>	0,398	0,277
<i>Es para adultos</i>	0,590	
<i>Puedo comerlas son hartarme</i>	0,392	0,221
<i>Satisface el antojo</i>	0,271	0,365
<i>Bolsa llena/sin aire</i>		0,577
<i>Encuentra en cualquier parte</i>	0,165	0,478
<i>Divertida</i>	0,540	
<i>Confiable</i>	0,440	0,289
<i>Buena publicidad</i>	0,532	0,199
<i>Se puede consumir haciendo cualquier cosa</i>	0,453	0,239
<i>Variedad de sabores</i>	0,315	0,374
<i>Ligera/no llenadora</i>	0,374	0,427
<i>Precio justo</i>	0,104	0,489
<i>Buenas promociones</i>	0,376	0,378
<i>Puede acompañar otros alimentos</i>	0,385	0,298
<i>Marca para niños</i>	0,569	
<i>Marca de prestigio</i>	0,519	0,152
<i>Empaque atractivo</i>	0,474	0,289
<i>SS loadings</i>	5,861	4,271
<i>Proporción de varianza</i>	0,150	0,110
<i>Varianza acumulada</i>	0,150	0,260

Test of the hypothesis that 2 factors are sufficient. The chi square statistic is 1789.2 on 664 degrees of freedom.

The p-value is 4.99e-104

Cuadro 4.1: Cargas de los 2 factores de 39 rasgos.

<i>SS loadings</i>	4,642	3,753	2,849
<i>Proporción de varianza</i>	0,119	0,096	0,073
<i>Varianza acumulada</i>	0,119	0,215	0,288

Test of the hypothesis that 3 factors are sufficient. The chi square statistic is 1561.62 on 627 degrees of freedom.
The p-value is 2.95e-81

Cuadro 4.2: Análisis de 3 factores aplicados a 39 rasgos.

4.3. Modificación de la escala original

Se construyeron 4 índices de evaluación, relacionados con las preguntadas realizadas, mediante una escala aditiva de las respuestas dadas. El objetivo de este cambio es movernos a un contexto continuo para poder utilizar de manera adecuada las técnicas multivariadas. La construcción de los índices se basa en agrupar las preguntas que se relacionen.

ÍNDICES CONSTRUIDOS

- I_1 CARACTERÍSTICAS SENSORIALES DEL PRODUCTO (SABOR, GUSTO) (*preguntas 1-8*)
- I_2 IMAGEN (*preguntas 09-21*)
- I_3 PUBLICIDAD/ PRECIO/VALOR/CALIDAD (*preguntas 22-31*)
- I_4 ACCESIBILIDAD DEL PRODUCTO (*preguntas 32-39*)

Los índices construidos tuvieron un nuevo rango de medición, siguen siendo variables discretas aunque se conserva la propiedad de que los valores grandes se refieran a mucha importancia y valores pequeños se refieran a poco importancia para cada índice.

Sea P_i la respuesta a la pregunta número i , con $P_i = 1, 2, 3, 4, 5$ los índices serán la suma de las P_i 's, de modo que el rango donde toman valores los índices, de acuerdo al número de preguntas que contienen es el siguiente:

$$I_1 = 8, 9, 10, \dots, 40.$$

$$I_2 = 13, 14, \dots, 65.$$

$$I_3 = 10, 11, \dots, 50.$$

$$I_4 = 8, 9, \dots, 40.$$

Se referirá en lo consecutivo a los índices como:

1. ÍNDICE 1 = Características Sensoriales
2. ÍNDICE 2 =Imagen
3. ÍNDICE 3 =Publicidad
4. ÍNDICE 4 =Accesibilidad

4.4. Alpha de Cronbach calculada a los índices

Utilizamos la función `cronbach (items)` que se encuentra en la librería `multilevel` del software estadístico R. La función recibe como objeto una matriz o un `data frame` donde cada columna representa un ítem del índice analizado, arroja como resultado el valor estimado del Alpha de Cronbach calculado con la fórmula propuesta anteriormente, así como el número de observaciones sobre las cuales se calculó.

Los valores calculados del estadístico de Cronbach para los cuatro índices del producto IDEAL se pueden apreciar en el Cuadro 4.3. Se observó que el valor del estadístico se encuentra entre .70 y .80 en el ÍNDICE 1, ÍNDICE 3 e ÍNDICE 4, comparando con los valores propuestos por George y Mallery podemos decir que éstos índices son aceptables. El alpha de Cronbach del ÍNDICE 2 tiene un valor de .8220, basándonos de nuevo en George y Mallery, el ÍNDICE 2 es bueno.

Tomando en consideración este resultado usaremos los nuevos índices construidos para el análisis, pues de acuerdo a este estadístico al unir las preguntas en estos cuatro índices se conserva la variabilidad y se sigue capturando la información original.

ÍNDICE	ALPHA DE CRONBACH
1 Características Sensoriales	0,7020
2 Imagen	0,8220
3 Publicidad	0.7209
4 Accesibilidad	0.7025

Cuadro 4.3: Alpha de Cronbach para cada índice PRODUCTO IDEAL.

Realizamos el mismo análisis para cada producto estudiado con el objetivo de examinar la veracidad de los índices creados. El cálculo de la estadística Alpha de Cronbach se realizó de manera análoga al procedimiento anteriormente expuesto.

En el Cuadro 4.4 aparecen los valores para el PRODUCTO 1, los índices referentes a las Características Sensoriales del producto y a la publicidad no tiene mayor problema, pues ambas exceden el valor de .70 ubicándose en un rango aceptable mientras que los índices restantes tales como la Imagen del producto y Accesibilidad toman valores alrededor de 0.68. En realidad no se ven muy alejados del valor aceptable y tomando en consideración que una diferencia menor de 0.02 no es muy alta para éste tipo de problema, concluimos que la estadística calculada indica una buena agrupación de las variables, basándonos en que la variabilidad original se conserva.

ÍNDICE	ALPHA DE CRONBACH
1 Características Sensoriales	0,6844
2 Imagen	0,7906
3 Publicidad	0.6808
4 Accesibilidad	0.7148

Cuadro 4.4: Alpha de Cronbach para cada índice PRODUCTO 1.

Siguiendo con el análisis, los valores obtenidos de la estadística para el PRODUCTO 2 se encuentran en el Cuadro 4.5 donde percibimos que los ÍNDICES 1 y 4, es decir, Características Sensoriales y Accesibilidad son los únicos en los que habría de tener cuidado pues el valor del Alpha de Cronbach está por debajo del 0.70, no a una distancia mayor de 0.50 por lo que se conserva la idea de que esta forma de agrupar es buena, al menos para los índices ya analizados.

ÍNDICE	ALPHA DE CRONBACH
1 Características Sensoriales	0,6666
2 Imagen	0,7968
3 Publicidad	0.7037
4 Accesibilidad	0.6744

Cuadro 4.5: Alpha de Cronbach para cada índice PRODUCTO 2.

Lo que respecta al PRODUCTO 3 y al PRODUCTO 4, la estadística calculada arroja resultados similares para ambos, Cuadro 4.6 y Cuadro 4.7 respectivamente, siendo consistente en un valor alrededor de 0.80 para el índice de Imagen (INDICE 2) y valores entre 0.68 para los índices restantes, el valor más chico es de 0.6673 en el índice Características Sensoriales (INDICE 1) para el PRODUCTO 3.

ÍNDICE	ALPHA DE CRONBACH
1 Características Sensoriales	0,6673
2 Imagen	0,7811
3 Publicidad	0.6840
4 Accesibilidad	0.6628

Cuadro 4.6: Alpha de Cronbach para cada índice PRODUCTO 3.

INDICE	ALPHA DE CRONBACH
1 Características Sensoriales	0,6843
2 Imagen	0,8041
3 Publicidad	0.6959
4 Accesibilidad	0.6853

Cuadro 4.7: Alpha de Cronbach para cada índice PRODUCTO 4.

Por último en el Cuadro 4.8 se ubican los valores calculados para el PRODUCTO 5, siendo todos valores superiores al 0.76 decimos que para el producto en cuestión la reducción de variables en una escala aditiva, es decir, en los índices construidos es consistente con la información original.

INDICE	ALPHA DE CRONBACH
1 Características Sensoriales	0,7715
2 Imagen	0,8370
3 Publicidad	0.8256
4 Accesibilidad	0.7625

Cuadro 4.8: Alpha de Cronbach para cada índice PRODUCTO 5.

En general para los productos de interés incluyendo el IDEAL, después de analizar los valores obtenidos de esta estadística, se concluye que agruparlos en los 4 índices es una buena medida de reducción del número de variables. El siguiente análisis se realizó a partir de estas nuevas variables (índices construidos), pero siempre recordando el origen de ellos, para así poder retomar la información en algún momento del estudio que se busque una interpretación más profunda o bien una estrategia más específica para la empresa.

4.5. Análisis de componentes principales aplicado a los índices

Se realizó un análisis de componentes principales a los cuatro índices creados anteriormente, observamos en el Cuadro 4.9 las cargas de los componentes, así como la proporción de varianza acumulada que representan; si nos quedamos con dos componentes la varianza explicada es del 89.99 % la cual para el tipo de variable analizada es bastante alta, si estuviéramos hablando de algún caso clínico o algún tema delicado sería bueno quedarnos con tres componentes puesto que tendríamos un 95.64 % de varianza acumulada, sin embargo por tratarse de un análisis de mercadotecnia un 89.99 % ya nos explica gran parte del problema. Observamos que a comparación del análisis de factores realizado anteriormente (Cuadro 4.2) aquí tenemos mucha más certeza en los resultados del análisis pues la varianza explicada en el primero era tan sólo del 26 % mientras que con la corrección de la escala y el nuevo análisis realizado tenemos casi un 90 % (Cuadro 4.9).

Nos quedamos con los primeros dos componentes y para fines prácticos del análisis al graficar e interpretar realizamos una reflexión sobre los componentes al multiplicarlos por -1 (Cuadro 4.10). Notemos que la reflexión mencionada no afecta los resultados pues la distancia (euclidiana) se conserva.

El primer componente principal se ve explicado prácticamente por la Imagen del producto con una carga de 0.768 mientras que las Características Sensoriales, la Publicidad y la Accesibilidad del Producto influyen casi en la misma proporción y con una carga alrededor de 0.30 y 0.40 (Cuadro 4.10). El segundo componente se ve explicado por la Publicidad del producto (0.652) y las características sensoriales del producto (0.439) de manera positiva y con la imagen (-0.598) de manera negativa, es decir, de manera inversa con los dos índices anteriores.

De modo que la componente uno está relacionada con la Imagen del producto, desde la forma de su empaque, a quién va dirigida (jóvenes, adultos, gente como tú, etc.), el tipo de marca ya sea divertida, confiable. La componente dos está relacionada con el sabor y calidad del producto, que sea crujiente y tenga sabor a recién hecho, se antoje, que valga lo que cueste o bien un precio justo.

Los componentes analizados anteriormente los utilizaremos para la representación espacial de nuestro PRODUCTO 1 respecto a los productos de competencia en el mercado. Notemos que el análisis se realiza sobre la importancia declarada que las personas consideran tuvo cada ítem en el índice, pues es lo que la población encuestada considera que una botana en el mercado deba cumplir o su ideal y como una empresa de botanas nos interesa estar lo más próximo al ideal que sea posible.

	<i>Componente 1</i>	<i>Componente 2</i>
<i>Características sensoriales del producto</i>	-0,342	-0,439
<i>Imagen</i>	-0,768	0,598
<i>Publicidad</i>	-0,382	-0,652
<i>Accesibilidad del producto</i>	-0,382	-0,155
<i>Desviación estándar</i>	11,197259	4,8958328
<i>Proporción de varianza</i>	0,7555335	0,1444386
<i>Varianza acumulada</i>	0,7555335	0,8999721
	<i>Componente 3</i>	<i>Componente 4</i>
<i>Características sensoriales del producto</i>	0,339	0,758
<i>Imagen</i>	-0,209	
<i>Publicidad</i>	-0,587	-0,288
<i>Accesibilidad del producto</i>	0,705	-0,577
<i>Desviación estándar</i>	3,06210307	2,68753955
<i>Proporción de varianza</i>	0,05650278	0,04352512
<i>Varianza acumulada</i>	0,95647488	1,00000000

Cuadro 4.9: Cargas de los componentes por Índice.

	<i>Componente 1</i>	<i>Componente 2</i>
<i>Características sensoriales del producto</i>	0,342	0,439
<i>Imagen</i>	0,768	-0,598
<i>Publicidad</i>	0,382	0,652
<i>Accesibilidad del producto</i>	0,382	0,155

Cuadro 4.10: Reflexión de los primeros dos componentes principales.

4.6. Representación espacial de los productos

Calculamos los cuatro índices para los cinco productos a analizar, incluyendo el “PRODUCTO 1”. Para poder calcular los componentes principales por cada producto debemos tener una medida resumen de cada índice, es decir, poder darle un número que englobe la información de cada índice en los productos, de manera que al calcular su respectivo componente se reduzca a la combinación lineal de los índices con coeficientes iguales a las cargas obtenidas en el análisis de componentes principales realizado a la importancia declarada (producto IDEAL) por los encuestados.

Por cada índice calcularemos lo siguiente:

$$\text{Componente 1} = 0,342 * \text{Caract.Sensoriales} + 0,768 * \text{Imagen} + 0,382 * \text{Publicidad} + 0,382 * \text{Accesibilidad}$$

$$\text{Componente 2} = 0,439 * \text{Caract.Sensoriales} - 0,598 * \text{Imagen} + 0,652 * \text{Publicidad} + 0,155 * \text{Accesibilidad}$$

Donde *Caract.Sensoriales*, *Imagen*, *Publicidad* y *Accesibilidad* se refieren al valor que tomemos como medida resumen del índice, ya sea la media o mediana según convenga.

Para elegir una medida resumen del índice, se realizó un análisis descriptivo de cada uno para así tener un buen resumen de su comportamiento. Un histograma es una representación gráfica de estadísticas de una variable en forma de barras, donde las frecuencias son representadas en el eje vertical mientras que en el horizontal se representan los valores de cada una de las variables.

En la Figura 4.1 se observan los histogramas de cada uno de los índices, notamos que los cuatro están cargados hacia el lado derecho, es decir, tienen una cola ligera del lado izquierdo, que puede provocar que la media se recorra hacia el lado contrario, también se observa que no siguen un comportamiento

que ajuste a una distribución de probabilidad como una normal que para los fines de análisis reduciría bastante información, dado que no podemos ajustarles una distribución seguiremos con el análisis descriptivo.

Las gráficas de caja o “boxplot” se encuentran en la Figura 4.2 para cada índice, la mediana se observa toma un valor alto, debida a la mayor densidad de datos en esa escala, también se observan ciertos datos atípicos o fuera de la caja, no obstante dado que la cantidad es mínima no es de preocuparse.

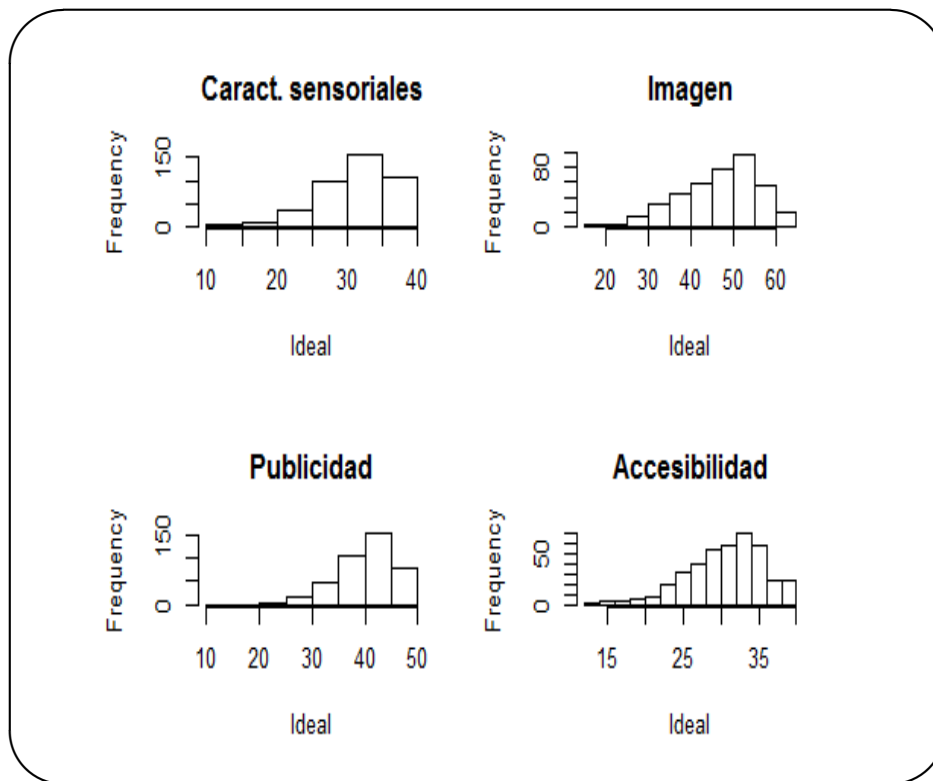


Figura 4.1: Histograma de los índices.

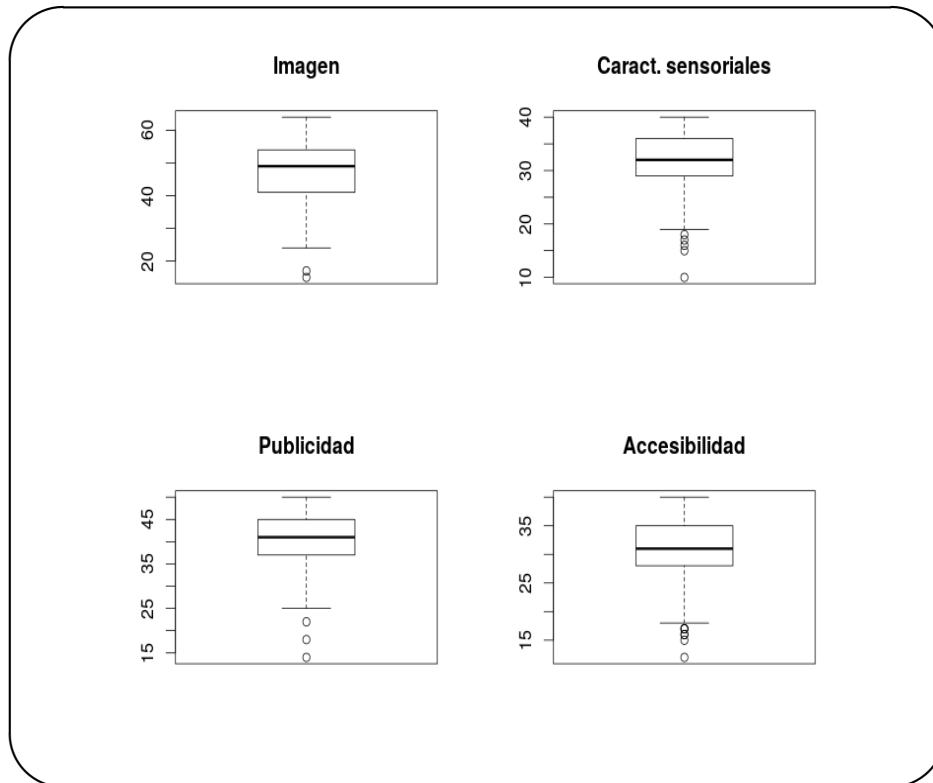


Figura 4.2: Boxplot de los índices.

Analizamos los valores máximo y mínimo que toman los índices, así como los valores de la media y mediana. Los valores del índice de Características Sensoriales se aprecian en el Cuadro 4.11, donde observamos que la media y la mediana difieren 0.19 unidades, el valor menor que alcanza es de 10 y el máximo de 40. El resumen descriptivo del índice de Imagen se observa en el Cuadro 4.12, tomando valores entre 15 y 64, con una media de 47.46 no muy lejana a la mediana, este índice tiene un mayor rango de movimiento. En el Cuadro 4.13 encontramos los valores que toma el índice de Publicidad, con una media de 40 y tomando valores en un intervalo que va de 14 a 50 unidades. Por último el índice de Accesibilidad tiene una media y una mediana muy parecidas, alrededor de 30, como se observa en el Cuadro 4.14.

Valor mínimo que alcanza	10.00
Primer cuartil	20.00
Mediana	32.00
Media	31.81
Tercer cuartil	36.00
Valor máximo que alcanza	40.00

Cuadro 4.11: Resumen descriptivo del índice de Características Sensoriales.

Valor mínimo que alcanza	15.00
Primer cuartil	41.00
Mediana	49.00
Media	47.46
Tercer cuartil	54.00
Valor máximo que alcanza	64.00

Cuadro 4.12: Resumen descriptivo del índice de Imagen.

Valor mínimo que alcanza	14.00
Primer cuartil	37.00
Mediana	41.00
Media	40.68
Tercer cuartil	45.00
Valor máximo que alcanza	50.00

Cuadro 4.13: Resumen descriptivo del índice de Publicidad.

Valor mínimo que alcanza	12.00
Primer cuartil	28.00
Mediana	31.00
Media	31.07
Tercer cuartil	35.00
Valor máximo que alcanza	40.00

Cuadro 4.14: Resumen descriptivo del índice de Accesibilidad.

Se observó que en general para los cuatro índices el valor de la media es cercano a la mediana, no obstante debemos tener cuidado en el índice Imagen pues es donde la media y la mediana difieren en más de una unidad, pero en realidad sigue siendo poca pues recordemos que el rango de movimiento era grande, de 15 a 64 unidades. Teniendo en consideración el análisis anterior, se tomó como medida resumen el valor de la mediana por el hecho de que las distribuciones están un poco sesgadas hacia el extremo derecho, como se apreció en los histogramas (Figura 4.1).

Las medidas resumen por índice son las agregadas en el Cuadro 4.15, recordemos que este es valor de la medida resumen pero para el IDEAL, es decir, el que la población consumidora desearía tener. Se procedió a calcular el valor de los componentes principales, simplemente evaluando estos valores en las combinaciones lineales anteriormente obtenidas.

Índice	Medida Resumen
Características Sensoriales	32.00
Imagen	49.00
Publicidad	41.00
Accesibilidad	31.00

Cuadro 4.15: Medida Resumen

La solución a la combinación lineal respecto a cada componente es la siguiente:
Componente 1 = $0,342 * 32,00 + 0,768 * 49,00 + 0,382 * 41,00 + 0,382 * 31,00 = 76,1$
Componente 2 = $0,439 * 32,00 - 0,598 * 49,00 + 0,652 * 41,00 + 0,155 * 31,00 = 16,3$
 Antes de comenzar con la construcción del mapa de posicionamiento, comparemos los descriptivos del producto IDEAL contra los descriptivos del PRODUCTO 1 como un primer acercamiento de la diferencia que existe entre el producto de interés en el estudio y el producto que la población considera bueno o con las características importantes para comprarlo.

En la Figura 4.3 se aprecian los histogramas por cada índice respecto a los productos contrastados, si nos enfocamos en la Imagen del producto notamos que el PRODUCTO 1 toma valores un poco más grandes que el producto IDEAL.

Aunque ambas gráficas tienen un comportamiento parecido en general, en cuanto al índice Características Sensoriales pasa lo contrario, el producto IDEAL

toma valores más grandes con mayor frecuencia que el PRODUCTO 1. Los índices referidos a la Publicidad y Accesibilidad del producto a simple vista tienen una distribución similar en ambos productos analizados, es importante recordar que éste sólo es un primer acercamiento del comportamiento de los índices por cada producto, para tener una estrategia sólida es necesario hacer un análisis más profundo de éstos.

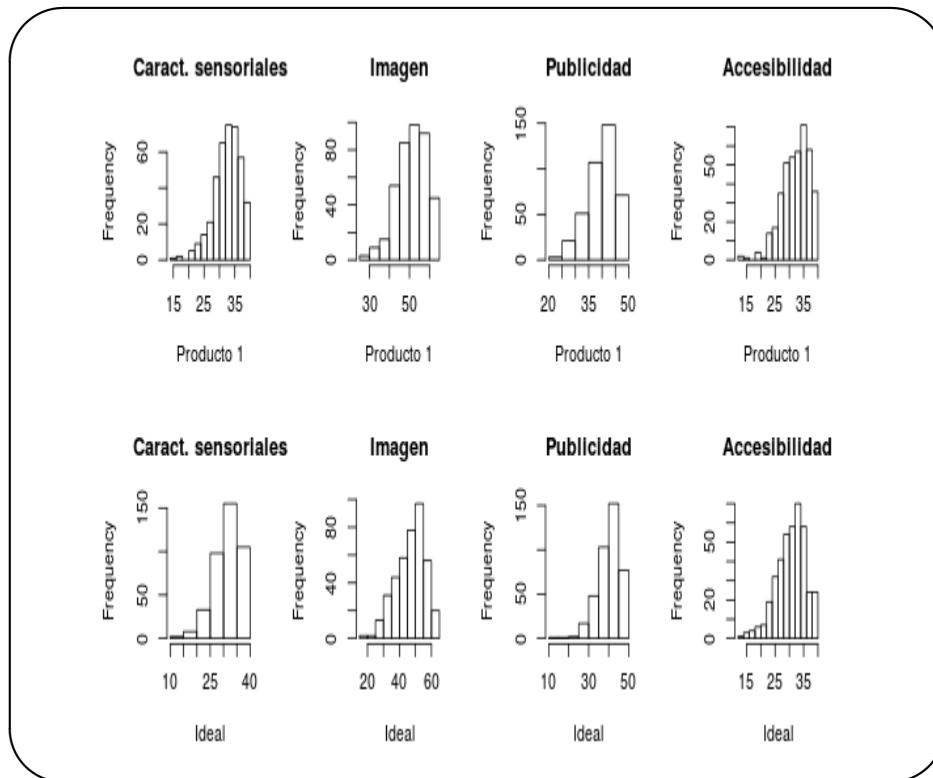


Figura 4.3: Descriptivos del IDEAL vs PRODUCTO 1.

Dentro de la Figura 4.4, 4.5, 4.6 y 4.7 se puede examinar el comportamiento de los productos comparados con el producto IDEAL, a modo de percibir diferencias significativas entre ellos. En la Figura 4.4 se contrastó el PRODUCTO 2 con el IDEAL, resultando histogramas muy parecidos en la mayoría de los índices contrastados, en el índice Publicidad se encontró quizá un comportamiento menos estable que en el obtenido en el producto IDEAL, conservando el valor de concentración más alto (43-45); el PRODUCTO 2 resultó ser el más parecido al producto IDEAL en cuanto a los histogramas analizados.

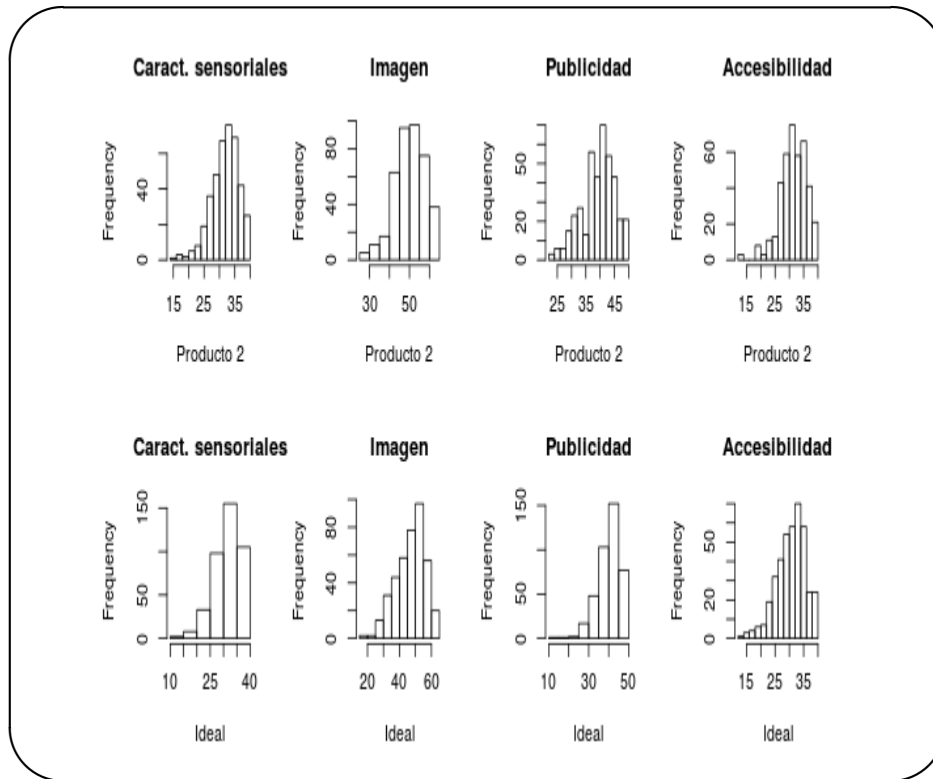


Figura 4.4: Descriptivos del IDEAL vs PRODUCTO 2

En cuanto al PRODUCTO 2 y PRODUCTO 3 los histogramas generados para el índice de publicidad (Figura 4.5 y 4.6) , están un poco sesgados a la derecha en comparación con el producto IDEAL, de igual forma en el índice Características Sensoriales e Imagen referentes al PRODUCTO 3.

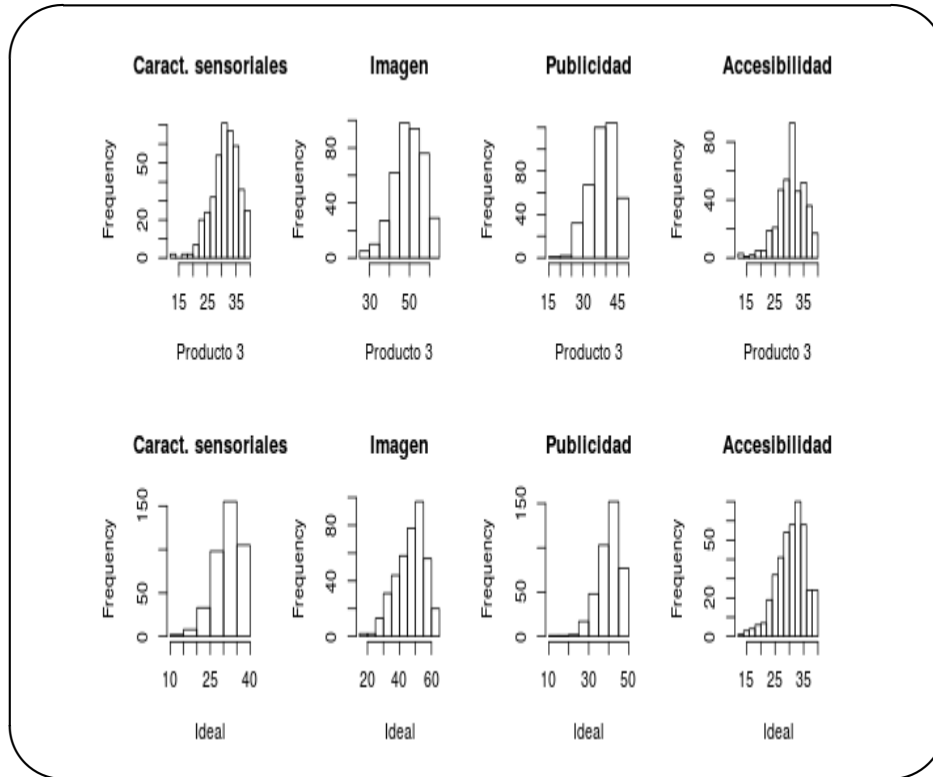


Figura 4.5: Descriptivos del IDEAL vs PRODUCTO 3

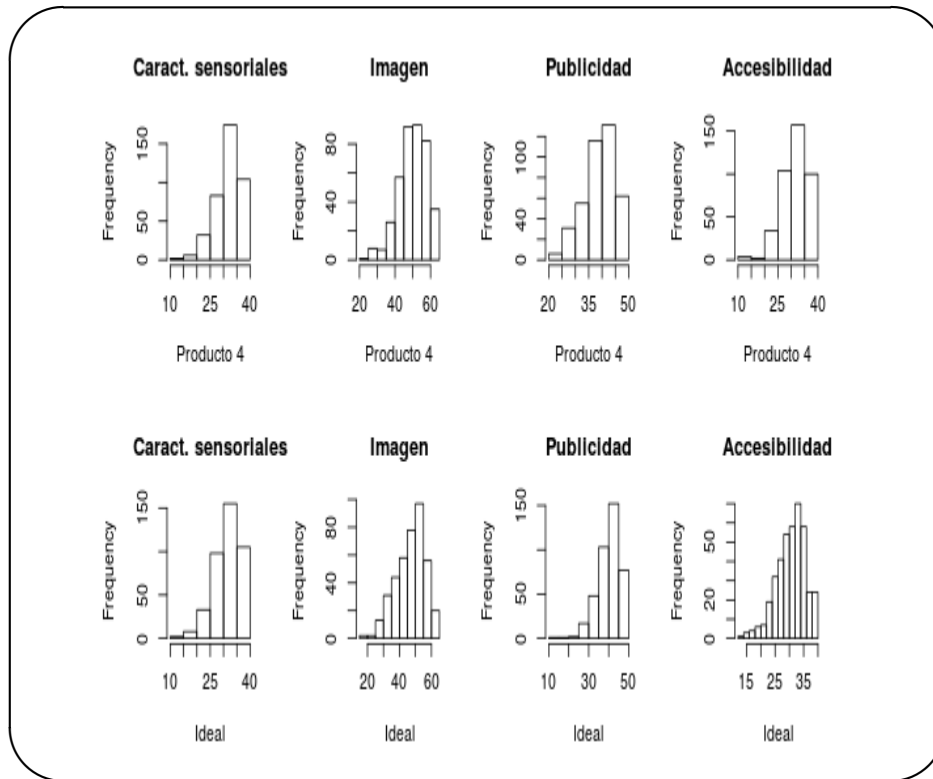


Figura 4.6: Descriptivos del IDEAL vs PRODUCTO 4

El cuadro comparativo de histogramas del PRODUCTO 5 contra el producto IDEAL, ubicado en la Figura 4.7, sugiere que el índice de Accesibilidad está alejado del respectivo índice del Ideal puesto que llega a tomar valores muy pequeños como 8, mientras el índice Accesibilidad dentro del producto IDEAL toma el 12 como valor mínimo, entonces la distribución de dicho índice esta recorrida un poco hacia la derecha en comparación con la distribución en el producto IDEAL.

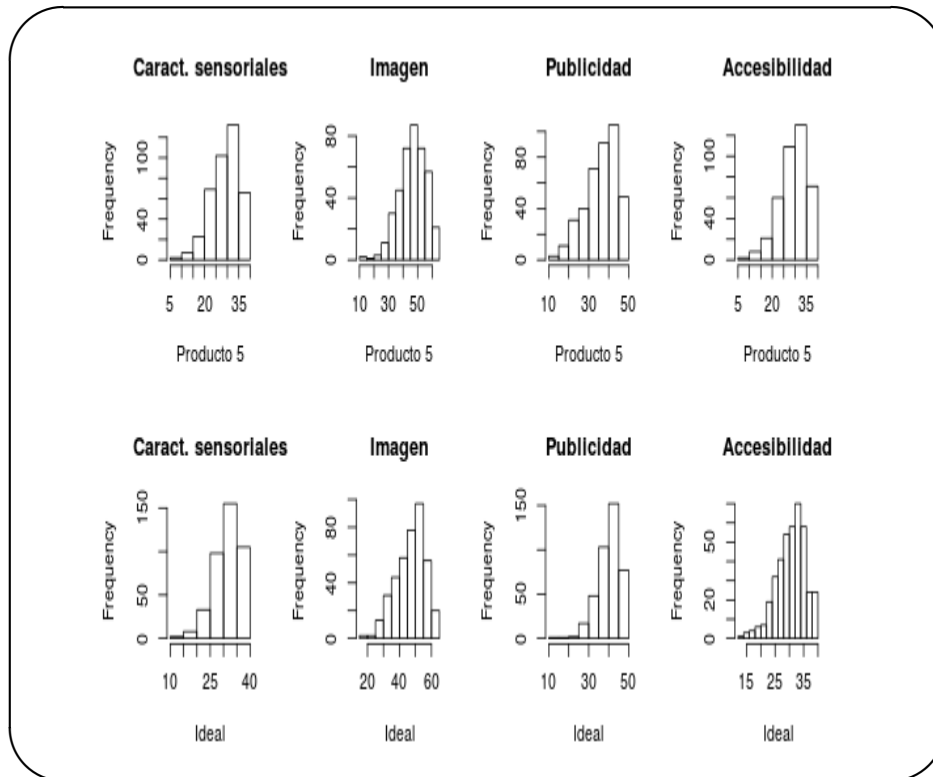


Figura 4.7: Descriptivos del IDEAL vs PRODUCTO 5

4.7. Construcción del mapa de posicionamiento

En el apartado anterior se calculó la medida resumen para el producto Ideal, el siguiente paso fue calcular la respectiva medida resumen de los cuatro índices para cada uno de los productos a analizar, PRODUCTO 1, PRODUCTO 2, PRODUCTO 3, PRODUCTO 4 y PRODUCTO 5. Los resultados aparecen en el Cuadro 4.16 y con esos valores se calcularon los dos Componentes Principales por cada producto, de manera análoga al razonamiento anterior, denotaremos

$$\text{Componente } ij = \text{Componente } i \text{ del índice } j \text{ con } i = 1, 2;$$

$$j = \begin{cases} 1 & \text{Características Sensoriales} \\ 2 & \text{Imagen} \\ 3 & \text{Publicidad} \\ 4 & \text{Accesibilidad} \end{cases}$$

Índice	Medida Resumen	Medida Resumen
	PRODUCTO 1	PRODUCTO 2
Características Sensoriales	34.00	33.00
Imagen	52.00	51.00
Publicidad	41.00	41.00
Accesibilidad	33.00	32.00
	PRODUCTO 3	PRODUCTO 4
Características Sensoriales	32.00	33.00
Imagen	50.00	51.00
Publicidad	39.00	40.00
Accesibilidad	32.00	32.00
	PRODUCTO 5	
Características Sensoriales	30.00	
Imagen	48.00	
Publicidad	38.00	
Accesibilidad	31.00	

Cuadro 4.16: Medida Resumen

$$\text{Componente 11} = 0,342 * 34,00 + 0,768 * 52,00 + 0,382 * 41,00 + 0,382 * 33,00 = 79,8$$

$$\text{Componente 21} = 0,439 * 34,00 - 0,598 * 52,00 + 0,652 * 41,00 + 0,155 * 33,00 = 15,7$$

$$\text{Componente 12} = 0,342 * 33,00 + 0,768 * 51,00 + 0,382 * 41,00 + 0,382 * 32,00 = 78,3$$

$$\text{Componente 22} = 0,439 * 33,00 - 0,598 * 51,00 + 0,652 * 41,00 + 0,155 * 32,00 = 15,7$$

$$\text{Componente 13} = 0,342 * 32,00 + 0,768 * 50,00 + 0,382 * 39,00 + 0,382 * 32,00 = 76,5$$

$$\text{Componente 23} = 0,439 * 32,00 - 0,598 * 50,00 + 0,652 * 39,00 + 0,155 * 32,00 = 14,5$$

$$\text{Componente 14} = 0,342 * 33,00 + 0,768 * 51,00 + 0,382 * 40,00 + 0,382 * 32,00 = 78$$

$$\text{Componente 24} = 0,439 * 33,00 - 0,598 * 51,00 + 0,652 * 40,00 + 0,155 * 32,00 = 15$$

$$\text{Componente 15} = 0,342 * 30,00 + 0,768 * 48,00 + 0,382 * 38,00 + 0,382 * 31,00 = 73,5$$

$$\text{Componente 25} = 0,439 * 30,00 - 0,598 * 48,00 + 0,652 * 38,00 + 0,155 * 31,00 = 14$$

Capítulo 5

RESULTADOS

Posteriormente se construyó el mapa de posicionamiento, Figura 5.1, simplemente como una gráfica en dos dimensiones (X,Y) tomando como eje X el valor del Componente 1 y como eje Y el valor del componente 2, resultando una gráfica bidimensional (Componente 1, Componente 2). En el Cuadro 5.1 se muestran los datos y en la Figura 5.1 su respectiva gráfica o como le hemos llamado “Mapa de Posicionamiento”.

PRODUCTO	Componente 1	Componente 2
IDEAL	76.1	16.3
PRODUCTO 1	79.8	15.7
PRODUCTO 2	78.3	15.7
PRODUCTO 3	76.5	14.5
PRODUCTO 4	78	15
PRODUCTO 5	73.5	14

Cuadro 5.1: Componentes Principales.



Figura 5.1: MAPA DE POSICIONAMIENTO.

El interés principal del análisis recordemos es conocer la posición en el mercado de un producto, en particular se supuso que la empresa está interesada en un producto que lanzó al mercado (PRODUCTO 1), la comparación se hará conforme a lo expuesto anteriormente y con base en el mapa de posicionamiento ya construido. El PRODUCTO 1 podrá compararse tanto con el Producto Ideal así como con los demás PRODUCTOS i ($i = 2, 3, 4, 5$) de competencia en el mercado.

La iniciativa del proyecto realizado es generar un mapa de posicionamiento de los productos analizados, incluyendo un producto IDEAL que es básicamente la idea que los consumidores tienen sobre la importancia de ciertos factores o características acerca de una botana. No sólo es comparar entre productos de competencia en el mercado sino también en poder ver qué tan cerca o lejos nuestro producto de interés se encuentra respecto al producto IDEAL y con base en esto poder tomar decisiones en cuanto a modificaciones que se le hagan al producto lanzado al mercado.

Cuando nuestro producto se encuentre en una posición alejada del producto IDEAL, como es el caso, se buscará moverlo en direcciones y magnitudes según convenga y sea posible, para quedar más cerca del producto IDEAL. Por construcción, el mapa de posicionamiento tiene como eje X la primera Componente Principal y como eje Y la segunda Componente, de modo que si buscamos movernos a la derecha es análogo a pensar en subir la Componente 1, si lo que se busca es moverse a la izquierda se tratará de bajar el valor de la Componente 1. El razonamiento es el mismo cuando queremos subir o bajar el producto en el mapa, pues simplemente subimos o bajamos la segunda Componente (eje Y) respectivamente.

Enfocándonos en lo obtenido en el análisis, la propuesta es mover el PRODUCTO 1 hacia la izquierda y un poco hacia arriba, pero como prioridad nos interesa el moverlo a la izquierda pues es donde más se aleja de la posición del producto IDEAL.

Lo moveremos a la izquierda disminuyendo el valor de la Componente 1, para esto recordemos como se definió la primer Componente Principal:

Componente 1 = 0,342 * *Caract.Sensoriales* + 0,768 * *Imagen* + 0,382 * *Publicidad* + 0,382 * *Accesibilidad*

La imagen del producto es el índice al que la Componente 1 le dio más peso, con un valor de 0.768, el PRODUCTO 1 tiene un valor muy alto en ésta componente lo que puede ser interpretado como que nuestra empresa le da mucha importancia a la imagen del producto, recordemos que el índice Imagen está compuesto de características como:

- Empaque metalizado
- Es para toda la familia
- Es una marca de tradición
- Buena apariencia
- Es un marca moderna y actual
- Es para jóvenes
- Es para gente como tú
- Es para adultos
- Divertida
- Es una marca confiable

- Es una marca para niños
- Es una marca de prestigio
- Empaque atractivo

La empresa interesada en el estudio, busca darle importancia a los aspectos que la gente considera importantes en una botana, el producto IDEAL no tiene un valor muy alto en la primer componente lo cual nos indica que la población consumidora considera importante la imagen del producto, sin embargo no lo ve como una prioridad, es decir, podemos no invertir tanto dinero en aspectos como el empaque metalizado, la apariencia y la percepción que queremos darle a la población de nuestro producto. Analizamos el segundo componente (eje Y) buscando mover nuestro producto un poco hacia arriba cerca del producto IDEAL.

Componente 2 = 0,439 * *Caract.Sensoriales* - 0,598 * *Imagen* + 0,652 * *Publicidad* + 0,155 * *Accesibilidad*

Dentro de los pesos que asigna esta componente a los índices, la publicidad juega un papel importante pues este índice tiene un peso de 0.652, por la estructura de la combinación lineal referente a la segunda componente, para lograr aumentar su valor hay diversas alternativas, una de ellas es aumentar el valor del índice Publicidad; conformado por preguntas tales como si la marca ofrece promociones, la buena calidad del producto, el precio y el valor de la botana. La siguiente alternativa sería disminuir el valor del índice Imagen, pues está inversamente relacionado con la componente al tener un peso de -0.598, mientras el valor del índice Imagen aumente la segunda componente disminuye y viceversa. La última alternativa es aumentar el valor del índice Características Sensoriales compuesto por el sabor del producto y el antojo que la población consumidora tiene hacia la botana.

El índice Imagen como vimos en los dos casos anteriores, es de suma importancia en el estudio pues para mover el PRODUCTO 1 hacia el IDEAL en ambas situaciones, tanto moverlo en el eje X o Y, llegamos a la conclusión de que la empresa le da más importancia de la necesaria a éste índice, descuidando los índices de Publicidad y Características Sensoriales que según el estudio las personas le dan más peso.

En cuanto a los productos de competencia, el PRODUCTO 2 se encuentra cercano al producto de interés lanzado por la empresa (PRODUCTO 1), en realidad en el Componente 2 tienen prácticamente el mismo valor, mientras que en el Componente 1 referido a la imagen del producto, el tipo de marca

ya sea confiable o divertida, etc. se tiene una distancia considerable entre el PRODUCTO 1 y PRODUCTO 2, podemos decir que en el mercado, estos dos productos son parecidos en cuanto a su sabor, calidad o precio, pero difieren en cualidades tales como la imagen o idea de los consumidores acerca del producto. Podría ser de interés el aproximar nuestro producto lanzado al mercado (PRODUCTO 1) con el PRODUCTO 2, por diversas razones, entre las principales podría preocuparnos el hecho de que el producto IDEAL o bien la importancia declarada por la población encuestada no sea del todo verídica ya sea por factores externos al estudio como el humor de las personas a la hora de la entrevista, o bien por factores internos como el elegir una zona específica de estudio donde tal vez no se refleje con la totalidad el concepto que pudiera tener toda la población consumidora acerca de la importancia de ciertos aspectos.

De igual forma podría pasar que por la experiencia de la empresa, sea cierto que el PRODUCTO 2 siempre ha estado a la vanguardia en el mercado siendo la competencia más fuerte hasta el momento, y podría interesarnos cambiar aspectos de nuestro producto para hacerlo lo más parecido posible al PRODUCTO 2, además de que el PRODUCTO 2 es el más cercano al producto IDEAL en cuestión.

La estrategia para acercarnos al PRODUCTO 2 será la siguiente, de nuevo basándonos en el mapa de posicionamiento, nos moveremos un poco a la izquierda sobre el eje X que recordemos es la Primer Componente.

$$\text{Componente 1} = 0,342 * \text{Caract.Sensoriales} + 0,768 * \text{Imagen} + 0,382 * \text{Publicidad} + 0,382 * \text{Accesibilidad}$$

Se reducirá el valor del índice Imagen, tratando de mantener un equilibrio en los índices restantes. La inversión destinada a la imagen del producto se repartirá de manera que se mejoren las características sensoriales del producto, su sabor, calidad, precio, accesibilidad y publicidad, pues logrando esto la combinación lineal que le da mucho peso a la Imagen se compensará con un menor peso a los demás aspectos que se refieren a los índices restantes.

Capítulo 6

CONCLUSIONES

Es importante siempre tener cuidado con el tipo de análisis a realizar y la información obtenida en los resultados, se obtuvo una alternativa para el estudio del posicionamiento en el mercado cuando las variables son categóricas, generando nuevas variables de tipo continuo (índices) y se observó que el análisis mejoró en cuanto a la proporción de varianza explicada.

En general, la idea de un análisis estadístico conlleva cierto cuidado en el cumplimiento de supuestos teóricos importantes. Como fue el caso, pues el análisis de factores no ajustó de manera correcta y no resumía la información de manera confiable debido tal vez a la estructura de la matriz de correlación de los datos y el incumplimiento de algún supuesto teórico referente a dicha matriz. Notamos que el análisis de componentes principales fue más acertado, debido a la corrección en cuanto a la escala de las variables originales, precisamente es en lo que debe tenerse cuidado, pues bien pudimos quedarnos con un análisis de factores que nos permitiera rescatar un 26 % de la información original. En cambio, optamos por analizar más a fondo las variables y tratar de corregir la exactitud de la prueba, de modo que al final nos permitiera rescatar el 90 % de la información.

El proyecto realizado es una alternativa para estudiar el posicionamiento en el mercado de un producto cuando el análisis de factores es realizado a variables categóricas y no genera resultados confiables, pues ya que generalmente en las publicaciones referentes a este ámbito se hace mención únicamente del análisis de factores como herramienta para realizar el estudio.

Tomando en cuenta las alternativas anteriores y el fin de nuestro análisis, la estrategia a seguir en la empresa es disminuir la inversión destinada a factores relacionados con la imagen, tal vez si el producto es de empaque metalizado

hacerlo de un material más barato o no darle tanta importancia al exterior de nuestro producto sino enfocarnos en el interior, como el sabor y la calidad, de manera conjunta invertir un poco más en publicidad, dar nuevas promociones e incluso elevar el número de comerciales o anuncios acerca de nuestra botana, así como cuidar que el precio de nuestro producto sea justo, que valga lo que cueste.

Capítulo 7

ANEXOS

Código en R

- Análisis de factores realizado al producto IDEAL, usando dos factores.

```
#####  
#####ANALISIS DE FACTORES#####  
#####  
library(stats)  
#IDEAL#  
Indices=data.frame(Ideal)  
#Ideal son los datos de la encuesta original,  
#las 39 preguntas a los 402 encuestados.  
analisisideal=factanal(Indices,factors=2)  
analisisideal
```

- Alpha de Cronbach calculado para cada producto.

```
#####  
#####ALFA DE CRONBACH#####  
#####  
library(multilevel)  
###Producto IDEAL (IMPORTANCIA)###  
###Producto 1###  
###Producto 2###  
###Producto 3###  
###Producto 4###  
###Producto 5###  
a1=cronbach(Indice1)  
a2=cronbach(Indice2)  
a3=cronbach(Indice3)  
a4=cronbach(Indice4)  
tabla=rbind(a1$Alpha, a2$Alpha, a3$Alpha, a4$Alpha)
```

```
rownames(tabla)=c("Indice1","Indice2","Indice3","Indice4")
colnames(tabla)="Alpha de Cronbach"
tabla
```

- **Análisis de Componentes Principales realizado al producto IDEAL.**

```
#####
###MAPA POSICIONAMIENTO###
#####
COMP=princomp(Indices)
summary(COMP)
COMP$loadings
```

- **Análisis descriptivo de los productos.**

Se realizará el siguiente código para cada producto.

```
#####
#####DESCRITIVOS DE LOS INDICES#####
#####PRODUCTO i#####
#####
par(mfrow = c(2, 4))
Pi=cbind(PiIndice1,PiIndice2,PiIndice3,PiIndice4,PiIndice5)
hist(PiIndice1,main="Caract. sensoriales",xlab="Producto 1")
hist(PiIndice2,main="Imagen",xlab="Producto 1")
hist(PiIndice3,main="Publicidad",xlab="Producto 1")
hist(PiIndice4,main="Accesibilidad",xlab="Producto 1")
#####
#####PRODUCTO IDEAL#####
#####
Ideal=cbind(Indice1,Indice2,Indice3,Indice4,Indice5)
hist(Indice1,main="Caract. sensoriales",xlab="Ideal")
hist(Indice2,main="Imagen",xlab="Ideal")
hist(Indice3,main="Publicidad",xlab="Ideal")
hist(Indice4,main="Accesibilidad",xlab="Ideal")
boxplot(Indice1,main="Caract. sensoriales")
boxplot(Indice2,main="Imagen")
boxplot(Indice3,main="Publicidad")
boxplot(Indice4,main="Accesibilidad")
summary(Ideal)
```

- **Medida resumen de cada producto por índice y su respectiva componente principal.**

```
#####
#####MEDIDA RESUMEN#####
#####COMPONENTE PRINCIPAL#####
#####
I=c(median(Indice1),median(Indice2),median(Indice3),median(Indice4))
P1=c(median(Puno1),median(Puno2),median(Puno3),median(Puno4))
P2=c(median(Pdos1),median(Pdos2),median(Pdos3),median(Pdos4))
```

```

P3=c(median(Ptres1),median(Ptres2),median(Ptres3),median(Ptres4))
P4=c(median(Pcuatro1),median(Pcuatro2),median(Pcuatro3),median(Pcuatro4))
P5=c(median(Pcinco1),median(Pcinco2),median(Pcinco3),median(Pcinco4))
comp1=c(0.342,0.768,0.382,0.382)
comp2=c(0.439,-0.598,0.652,0.155)
PRODUCTO1=c(sum(comp1*P1),sum(comp2*P1))
PRODUCTO2=c(sum(comp1*P2),sum(comp2*P2))
PRODUCTO3=c(sum(comp1*P3),sum(comp2*P3))
PRODUCTO4=c(sum(comp1*P4),sum(comp2*P4))
PRODUCTO5=c(sum(comp1*P5),sum(comp2*P5))
IDEAL=c(sum(comp1*I),sum(comp2*I))

```

■ Mapa de Posicionamiento

```

#####
###MAPA POSICIONAMIENTO###
#####
X=c(PRODUCTO1[1],PRODUCTO2[1],PRODUCTO3[1],PRODUCTO4[1],PRODUCTO5[1],IDEAL[1])
Y=c(PRODUCTO1[2],PRODUCTO2[2],PRODUCTO3[2],PRODUCTO4[2],PRODUCTO5[2],IDEAL[2])
plot(X,Y,pch = 1, cex = 1, col =c("red"),xlim = c(73.5,80),
ylim=c(13.5,16.5), xlab="Imagen/ Confiabilidad/ Tipo de marca",
ylab="Sabor/ Calidad/ Precio" ,xaxt="n",yaxt="n",
main="MAPA DE POSICIONAMIENTO") text(X[1],Y[1],
"PRODUCTO 1",pos=2,cex=.8) text(X[2],Y[2],"PRODUCTO 2",pos=2,cex=.8)
text(X[3],Y[3],"PRODUCTO 3",pos=2,cex=.8)
text(X[4],Y[4],"PRODUCTO 4",pos=2,cex=.8)
text(X[5],Y[5],"PRODUCTO 5",pos=4,cex=.8)
text(X[6],Y[6],"IDEAL",pos=4,cex=.8)
abline(h=15,v=76.75)

```

Capítulo 8

BIBLIOGRAFÍA

1. Johnson, R. A. & Wichern, D. W. (2002). *Applied multivariate statistical analysis*.
2. Hair, J. F., Anderson, R. E., Tatham, R. L. & Black, W. (1998). *Multivariate data analysis*. Prentice Hall College Division.
3. Gliem, J & Gliem, R. (2003). *Calculating, Interpreting, and Reporting Cronbach's Alpha Reliability Coefficient for Likert-Type Scales*. Conference in Adult, Continuing, and Community Education. Midwest Research to Practice.
4. Cronbach, L. J. (1951). *Coefficient alpha and the internal structure of tests*. *Psychometrika*, 16, 297-334.
5. Ten Berge, J. M. F., & Socan, G. (2004). *The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality*. *Psychometrika*, 69, 613-616.
6. Sijtsma, K. (2009). *On the use, the misuse and the very limited of the Cronbach's Alpha*. *Psychometrika*, 74(1) 107-120. DOI: 10.1007/S11336-008-9101-0.
7. Levy, J.P y Varela, J. (2003). *Análisis Multivariable para las Ciencias Sociales*. Prentice-Hall.
8. Lehmann, D.; Gupta, S. y Steckel, J.H. (1998):. *Marketing Research*. Addison Wesley.