



**UNIVERSIDAD NACIONAL  
AUTÓNOMA DE MÉXICO**

**FACULTAD DE ESTUDIOS SUPERIORES  
ACATLÁN**



IMAGEN Y POSICIONAMIENTO: APLICACIÓN DEL ANÁLISIS DE  
CORRESPONDENCIAS A LA INVESTIGACIÓN DE MERCADOS

TESINA

QUE PARA OBTENER EL TÍTULO DE:

Lic. en Matemáticas Aplicadas y Computación

PRESENTA

Erika Capula Sánchez

Asesor: Ing. Elvira Beatriz Clavel Díaz

Abril, 2004



Universidad Nacional  
Autónoma de México



**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.

**A Dios:**

**Por su infinita bondad y eterno amor.**

**A mis padres:**

**Héctor y Socorro que siempre me brindan su apoyo  
incondicional.**

**Para ustedes con todo mi amor.**

**A mis hermanas:**

**Por su compañía, juegos y peleas.**

**A mis amigos:**

**Que han marcado mi vida con su presencia.**

**A mi asesora:**

**La profesora Beatriz Clavel, con toda mi admiración y  
cariño por su invaluable comprensión y ayuda.**

**Las matemáticas dan oportunidad, como ningún otro campo, de practicar el enérgico deporte de la precisión mental. Ser matemático es llevar una de las formas más deliciosas de la existencia: vivir en perpetuo asombro.**

**Alberto Barajas Celis.**

# ÍNDICE

<b>Introducción</b> .....	<b>1</b>
<b>Capítulo 1. Investigación de Mercados</b> .....	<b>4</b>
1.1 Definición .....	4
1.2 Etapas del proceso de investigación de mercados .....	5
1.2.1 Definición del problema y los objetivos de la investigación de mercado .....	6
1.2.2 Identificación de fuentes para la recopilación de datos .....	7
1.2.3 Implantación del plan de investigación .....	10
1.2.4 Interpretación e informe de los resultados de la investigación .....	10
1.3 Investigaciones de mercados basadas en encuestas .....	11
1.4 Tipos de estudios de investigaciones de mercados basadas en encuestas .....	26
1.5 Técnicas estadísticas útiles en las investigaciones de mercados basadas en encuestas .....	27
<b>Capítulo 2. Tablas de Contingencia</b> .....	<b>35</b>
2.1 Introducción .....	35
2.2 Prueba de independencia Ji-cuadrada de Pearson .....	38
2.3 Razón de verosimilitud Ji-cuadrada .....	41
2.4 Ji-cuadrada de Mantel-Haenszel ... ..	41
2.5 Prueba de bondad de ajuste Ji-cuadrada .....	42
2.6 Medidas de asociación .....	45
2.6.1 Coeficiente phi ( $\phi$ ) .....	45
2.6.2 Coeficiente de contingencia $C$ .....	46
2.6.3 $V$ de Cramer .....	46
2.7 Aplicación .....	47
2.8 Análisis de tablas de contingencia en SAS .....	49

<b>Capítulo 3. Análisis de Correspondencias .....</b>	<b>52</b>
3.1 Introducción .....	52
3.2 Objetivos del análisis de correspondencias .....	53
3.3 Matriz de correspondencias y masas de renglón y columna .....	53
3.4 Matriz de perfiles .....	55
3.5 Importancia de los perfiles .....	61
3.6 Espacio de los perfiles .....	61
3.6.1 Espacio de los perfiles sobre J .....	61
3.6.2 Espacio de los perfiles sobre I .....	63
3.7 Nube de puntos .....	64
3.8 Distancia Ji-cuadrada .....	65
3.9 Inercia .....	70
3.10 Coordenadas factoriales .....	71
3.11 Relaciones de transición .....	74
3.12 Mapa perceptual .....	75
3.13 Descomposición aditiva de la Ji-cuadrada .....	78
3.14 Interpretación de los planos de proyección .....	80
3.14.1 Contribuciones absolutas .....	80
3.14.2 Cosenos cuadrados .....	83
3.15 Análisis de correspondencias en SAS .....	85
<b>Capítulo 4. Estudio de Caso: Imagen y Posicionamiento de Golosinas .....</b>	<b>89</b>
4.1 Imagen y Posicionamiento .....	89
4.2 Objetivos de la investigación de mercado .....	90
4.3 Metodología de la investigación de mercado .....	91
4.4 Imagen de golosinas .....	92
4.5 Imagen de marcas de chocolates .....	96
4.6 Análisis de la frecuencia de consumo de chocolates .....	101
4.7 Hábitos de consumo y compra de chocolates .....	104
4.8 Análisis de resultados .....	111

<b>Conclusiones</b> .....	<b>112</b>
<b>Referencias Bibliográficas</b> .....	<b>115</b>
<b>Referencias Electrónicas</b> .....	<b>119</b>
<b>Anexo A</b> .....	<b>121</b>
<b>Anexo B</b> .....	<b>122</b>
<b>Anexo C</b> .....	<b>123</b>

## INTRODUCCIÓN

Se entiende por *análisis multivariante* a la rama de la estadística y del análisis de datos, que estudia, interpreta y elabora el material estadístico sobre la base de un conjunto de  $n > 1$  variables, que pueden ser de tipo cuantitativo, cualitativo o una mezcla de ambos. La información en análisis multivariante es, por lo tanto, de carácter multidimensional.

El análisis multivariante es una metodología estadística más complicada, pero también más potente, que utiliza extensamente los métodos del álgebra lineal, cálculo numérico y geometría. Las técnicas de análisis multivariante suelen clasificarse en técnicas de *dependencia* y técnicas de *interdependencia*. Una técnica de dependencia es aquella en la cual dentro de un conjunto de variables se pueden identificar tanto las variables dependientes como las variables independientes. El análisis de regresión múltiple, el análisis conjunto, el análisis discriminante múltiple y el análisis de varianza multivariante son ejemplos de técnicas de dependencia. Por otro lado, una técnica de interdependencia es aquella en la cual las variables no son divididas en dependientes o independientes, sino que todas las variables son analizadas simultáneamente como un solo conjunto. Ejemplos de este tipo de técnica son: el análisis factorial, el análisis de conglomerados y el *análisis de correspondencias*.

El análisis de correspondencias es una de las técnicas más recientes del análisis multivariante. Sus orígenes se remontan al año de 1962 con los trabajos del franco libanés Jean-Paul Benzécri, pero fue hasta 1965 cuando al dirigir la tesis doctoral de B. Escofier en la Facultad de Ciencias de Rennes, se mostraron todas sus propiedades algebraicas y geométricas. Concebido básicamente como una técnica gráfica para el estudio de tablas de contingencia, se ha revelado eficaz para el análisis de cualquier matriz de números no negativos. La principal virtud con la que cuenta el análisis de correspondencias, frente a las otras técnicas de interdependencia, es que es apropiado para variables tanto *cualitativas* como *cuantitativas*. Además, esta técnica estadística multivariante es ampliamente utilizada en diferentes campos de aplicación entre los que destacan: sociología, psicología,

investigación clínica, ecología, lingüística, genética e *investigación de mercados*, entre otros. En el campo de la investigación de mercados su objetivo consiste fundamentalmente en analizar las relaciones entre dos conjuntos de variables, normalmente objetos (personas, productos, marcas, servicios, empresas, etc.) y atributos o características, aunque es posible analizar cualquier otro tipo de variables que hayan sido incluidas en la investigación. La representación gráfica de las relaciones entre las variables, llamada comúnmente *mapa perceptual*, sirve para facilitar la interpretación del análisis de correspondencias. Por ejemplo, a través de ellos se puede observar fácilmente en una investigación de mercado de imagen y posicionamiento de qué forma son percibidas las marcas o productos que compiten en un mismo mercado.

El presente trabajo tiene como principal objetivo el mostrar el uso del análisis de correspondencias en una investigación de mercado de imagen y posicionamiento de golosinas para valorar la percepción de los consumidores con respecto a un producto (chocolates) analizando las tablas de contingencia obtenidas.

El trabajo se encuentra integrado de la siguiente forma:

En el primer capítulo se explican y se ejemplifican las etapas que conforman una investigación de mercados, enfocándose principalmente en aquellas que se basan en la realización de encuestas.

En el segundo capítulo se aborda el tema de tablas de contingencia. Se considera importante dedicar todo un capítulo al análisis de las tablas de contingencia, ya que prácticamente siempre en las investigaciones de mercado basadas en encuestas se requiere del desarrollo de una tabulación, es decir de un ordenamiento y resumen de la información obtenida a partir de un conteo que permita determinar las frecuencias de las respuestas. Además, gracias a la tabulación o creación de tablas de contingencia es posible realizar, en primer instancia, diferentes tipos de gráficos con los más relevantes resultados obtenidos, así como realizar pruebas de independencia entre variables, para finalmente aplicar el

análisis de correspondencias. Es por ello que en este capítulo se presentan los conceptos básicos necesarios para el análisis de tablas de contingencia de dos dimensiones del tipo  $n \times p$ , las cuáles se definen como un arreglo de frecuencias en  $n$  renglones y  $p$  columnas.

En el capítulo III se presentan los fundamentos matemáticos en los que se basa el análisis de correspondencias. Dicho análisis se desarrolla sobre una matriz de probabilidades (matriz de correspondencias) que se obtiene de una matriz o tabla de contingencia. De igual manera, se ejemplifica cada uno de los conceptos tratados y se utiliza el paquete estadístico SAS (Statistical Analysis System) versión 8.2 para la generación de resultados.

Finalmente, en el cuarto capítulo se aplica la teoría presentada en los capítulos anteriores en una investigación de mercado de imagen y posicionamiento de golosinas. Con el fin de enriquecer el análisis descriptivo de las variables de la investigación de mercado se presentan también una serie de gráficas con algunos de los más importantes resultados obtenidos.

## CAPÍTULO 1 – INVESTIGACIÓN DE MERCADOS

### 1.1 DEFINICIÓN

La investigación de mercados es el proceso de diseñar, recopilar, registrar, analizar, interpretar y comunicar los datos y resultados relacionados con situaciones específicas que permitan un mejor entendimiento del mercado. La investigación de mercados se ha convertido en una valiosa herramienta capaz de establecer una comunicación entre los consumidores (clientes) y los productores (empresas).

En la actualidad, cada vez son más las organizaciones que invierten en la investigación de mercados como un medio para obtener la información que les permita responder preguntas concretas para planear y ejecutar acciones de mercadeo minimizando riesgos. Por ejemplo, si una empresa pretende lanzar un nuevo producto al mercado, ésta necesitará de una investigación de mercados; es decir, necesitará obtener información relacionada con varios aspectos del mercado al que pretende incorporar su producto como quién será su eventual competencia, qué características debe de tener el producto, el precio adecuado, los puntos de venta, el perfil del posible consumidor, etc., de tal forma que la empresa minimice el riesgo de un fracaso con la introducción del nuevo producto al mercado.

*¿Quiénes realizan investigaciones de mercados?*

Las empresas, con sus propios recursos humanos, pueden desarrollar investigaciones de mercado, pero también pueden contratar compañías externas dedicadas a ofrecer este tipo de servicio. En México, algunas de las más importantes agencias de investigaciones de mercados son: ACNielsen, Gallup, y Berumen<sup>1</sup>.

Con más de treinta años de experiencia, ACNielsen es la empresa líder mundial en investigación, información y análisis de mercados para productos de consumo masivo y

---

<sup>1</sup> Las páginas en internet de estas compañías son: [www.acnielsen.com.mx](http://www.acnielsen.com.mx), [www.gallupmexico.com](http://www.gallupmexico.com) y [www.berumen.com.mx](http://www.berumen.com.mx) respectivamente.

servicios industriales. Los servicios principales de investigación de mercado que actualmente ofrece son: medición del comercio detallista (monitoreo continuo de las compras del consumidor a través de tecnología scanning y de auditorías directas en tiendas), investigación en el panel de hogares (servicio que proporciona información detallada de las compras realizadas por los miembros del hogar) e investigación ad hoc, la cual incluye estudios cualitativos y cuantitativos a la medida de las necesidades del cliente.

Fundada en 1970, Gallup es otra de las empresas que se ha consolidado como una prestigiosa empresa de investigación de mercados dedicada al desarrollo de estudios cualitativos y cuantitativos. Así mismo, es ampliamente reconocida por su participación en encuestas de opinión y conteos rápidos en contiendas electorales. En 1979, Gallup sacó al mercado “Omnibus”, el primer servicio de estudios de mercado continuos en México.

Berumen, otra de las compañías de más prestigio en nuestro país en la investigación de mercados, se fundó en abril de 1992 y desde entonces se ha dedicado de manera exitosa a estudios ad hoc de investigación cualitativa y cuantitativa ofreciendo sus servicios a empresas del sector privado, público y social.

## **1.2 ETAPAS DEL PROCESO DE INVESTIGACIÓN DE MERCADOS**

En el proceso de investigación de mercados se pueden distinguir cuatro etapas:

- Definición del problema y los objetivos de la investigación de mercado
- Identificación de fuentes para la recopilación de datos
- Implantación del plan de investigación
- Interpretación e informe de los resultados de la investigación

Estas etapas deben ser desarrolladas cabalmente y con estricto apego al orden arriba planteado.

### **1.2.1 DEFINICIÓN DEL PROBLEMA Y LOS OBJETIVOS DE LA INVESTIGACIÓN DE MERCADO**

El primer paso que se debe dar para planear una investigación de mercado es definir el problema de manera precisa. Esto último puede ser una tarea difícil pero sustantiva de la que se parte para planear el rumbo que habrá de seguir la investigación. Suponga el siguiente ejemplo hipotético: una empresa fabricante de productos en polvo para preparar agua de sabor para beber, percibe una disminución en sus ventas. Lo primero que se hace es determinar las causas específicas que originan esta disminución en ventas como puede ser la publicidad, las características del producto, su precio, etc. Después de realizar un sondeo, se descubre que el problema no surge de una mala campaña publicitaria, ni del precio, ya que éste se encuentra dentro del rango manejado por la competencia. Lo que originó la caída en las ventas fue un cambio en las preferencias del consumidor con respecto a las características del producto. La aparición de productos sin azúcar llamados “light” los cuales conservan un agradable sabor, acaparó fuertemente la atención del público generando un cambio en el mercado.

Una vez que se ha delimitado el problema, se procede a establecer el o los objetivos de la investigación de mercado. Conocer con precisión el fin de la investigación, garantizará que ésta sea más rentable y proporcione información útil para la empresa. Para el ejemplo expuesto anteriormente, algunos de nuestros objetivos podrían ser traducidos en necesidades específicas de información tales como conocer la opinión del público con respecto a nuestro producto y a los nuevos productos, la frecuencia de consumo, los sabores preferidos, el lugar de compra de este tipo de productos, etc.

En cuanto a su naturaleza, la investigación de mercados puede clasificarse en **investigación de mercado cualitativa** e **investigación de mercado cuantitativa**.

Si los datos que se necesitan para cumplir los objetivos son más de tipo psicológico que numérico, la **investigación de mercado cualitativa** suele ser la más apropiada. Las

entrevistas a profundidad y las sesiones de grupo son los métodos más utilizados en este tipo de investigaciones. Las *entrevistas a profundidad* generalmente son realizadas por expertos en el área de psicología y aplicadas de manera individual a cada entrevistado. Durante la entrevista se abordan libremente una serie de tópicos y no se requiere elaborar un cuestionario formal. Estas entrevistas son, por lo general, extensas y conllevan un costo mayor, es por esta razón que se utilizan pequeños tamaños de muestra. La mayoría de las veces, la entrevista a profundidad es grabada para la posterior realización del análisis de la información. Las *sesiones de grupo* son parecidas a las entrevistas a profundidad, sólo que en este caso se entrevista a varias personas a la vez, lo cual reduce su costo. Para la grabación de las sesiones de grupo, se suelen utilizar salas con cámaras de video conocidas como “cámaras de Gesell” las cuales se encuentran ocultas.

Por el contrario, si lo que se busca son datos numéricos, lo que necesitaremos es desarrollar una **investigación de mercado cuantitativa**. Estas investigaciones tienden a ser altamente estructuradas (a diferencia de las investigaciones de mercado cualitativas que son más flexibles) y además, permiten medir, evaluar y probar hipótesis del mercado.

Cabe resaltar que es posible combinar las investigaciones cualitativas y las investigaciones cuantitativas en el desarrollo de un proyecto de investigación de mercado. En general, se realiza primero la fase cualitativa, la cual proveerá de información importante para la planeación de la fase cuantitativa, especialmente cuando la investigación se realice por medio de encuestas y se necesite diseñar un cuestionario.

### **1.2.2 IDENTIFICACIÓN DE FUENTES PARA LA RECOPIACIÓN DE DATOS**

Durante el proceso de búsqueda y obtención de información se pueden distinguir dos tipos de datos: primarios y secundarios.

Se le llaman **datos primarios** a aquellos datos que se crean y obtienen por medio de una investigación específica, la cual se caracteriza por ser más costosa y que exige mayor tiempo y trabajo. A la investigación de mercados que se desarrolla sobre la base de datos primarios se le conoce como estudio prospectivo<sup>2</sup>. La recopilación de datos primarios se da principalmente sobre la base de *la observación, la experimentación y la encuesta*.

La *investigación observacional* consiste en recoger datos basándose en la observación de personas, acciones y situaciones. Por ejemplo, se puede hacer una visita a los puntos de venta de la competencia para realizar la observación y obtener información acerca de la variedad de productos, los precios de cada uno de ellos, las actividades promocionales, la distribución física, etc. La investigación observacional puede clasificarse en *natural y mecánica*. La investigación observacional natural es aquella que se realiza de manera personal por parte del investigador (observador). En cambio, la investigación observacional mecánica consiste en realizar la observación por medio de alguna clase de tecnología como puede ser circuitos cerrados de televisión, cámaras de video, ojos electrónicos ocultos, etc. En la mayoría de los casos, la información que se obtiene a través de este tipo de investigaciones no resulta ser la suficiente para la toma de decisiones de marketing.

La *investigación experimental* nos permite evaluar por medio de un experimento la relación causa-efecto entre una variable dependiente y otra variable denominada independiente<sup>3</sup>. Por ejemplo, si se deseara investigar la existencia de una factible relación causal entre los mensajes publicitarios (variable independiente) y las ventas de un producto (variable dependiente), se podría experimentar en dos mercados (dos ciudades o estados): en uno se colocaría el producto sin ninguna clase de publicidad y en otro con publicidad en todos los medios posibles. Un estudio en el cual se realiza alguna manipulación de acontecimientos y en donde se carece del control experimental total de un auténtico experimento se le conoce

---

<sup>2</sup> El estudio prospectivo implica la recolección de datos de primera mano después de que se ha definido el problema y los objetivos de la investigación para conseguir unos resultados propuestos como probables.

<sup>3</sup> Variable independiente o manipulada (X): Es la causa en la relación causa-efecto. Es la variable que el investigador manipula y sobre la que se basa la predicción.

Variable dependiente o de criterio (Y): Es el efecto en la relación causa-efecto. Es la variable que es predecible y medible según se manipula la variable independiente.

como *cuasiexperimento*. Generalmente el cuasiexperimento se aplica cuando no es posible realizar la selección aleatoria de los sujetos participantes de una investigación.

La *investigación por encuestas* es la más adecuada para la obtención de datos que produzcan información descriptiva<sup>4</sup>, la cual se genera preguntando directamente a los consumidores. El éxito de la investigación por encuestas dependerá básicamente del uso de un cuestionario adecuadamente estructurado.

Por otra parte, los **datos secundarios** son aquellos existentes y disponibles que fueron generados para otros fines. Estos datos se pueden obtener por medio de *fuentes internas* (bases de datos generadas por los diferentes departamentos de la empresa) o *externas* (publicaciones gubernamentales, publicaciones periódicas, libros, datos comerciales, etc.). A la investigación de mercado que gira en torno a la recolección de datos secundarios se le conoce como estudio retrospectivo<sup>5</sup>.

El acceso a datos secundarios se da con mayor facilidad, rapidez y a menor costo que los primarios, pero con la desventaja de que generalmente no proveen toda la información necesaria para el cumplimiento de los objetivos, además de que no se puede confiar plenamente en su calidad. La calidad de la información<sup>6</sup> se evalúa con base en cuatro factores que son: pertinencia (que se ajuste o adapte a las necesidades de la investigación de mercados), exactitud ( la confiabilidad de la información), actualidad e imparcialidad (se contempla la objetividad con la que se recabó). Por tal motivo se debe procurar la obtención de datos primarios.

---

<sup>4</sup> Se le llama información descriptiva a aquella información que sirve única y exclusivamente para describir la distribución existente de las variables, sin profundizar en sus causas.

<sup>5</sup> El estudio retrospectivo involucra la recolección de datos que ya existían antes de que el problema y los objetivos de la investigación fueran definidos, con el fin de identificar retrospectivamente los antecedentes o causas de algún fenómeno observado.

<sup>6</sup> LÓPEZ, Carlos (Junio,2003). La Investigación de Mercados II.

<http://www.gestiopolis.com/canales/demarketing/articulos/no%2010/investigacionmercadoII.htm>

### **1.2.3 IMPLANTACIÓN DEL PLAN DE INVESTIGACIÓN**

La implantación del plan de investigación de mercados consiste fundamentalmente en llevar a cabo la recopilación, el procesamiento y el análisis de la información. Esta es la etapa de la investigación que requiere de una mayor cantidad de tiempo.

En principio, se debe brindar una capacitación al personal encargado de recoger los datos en donde además se expongan claramente los objetivos de la investigación. Si la investigación que se va a realizar es de tipo cualitativo, se deben concertar las fechas en que los entrevistados se reunirán con el entrevistador o experto. Del mismo modo, si la investigación es de tipo cuantitativo será importante programar las fechas del levantamiento de información para realizar el trabajo de manera oportuna. Posteriormente, cuando se ha cumplido con la recopilación de datos, se procede a realizar el procesamiento de los mismos para, finalmente, elaborar el análisis de la información obtenida.

### **1.2.4 INTERPRETACIÓN E INFORME DE LOS RESULTADOS DE LA INVESTIGACIÓN**

Una vez aplicado el plan, la etapa final de la investigación de mercados consiste en interpretar e informar los resultados. El informe deberá ir acompañado de una serie de conclusiones y recomendaciones las cuales buscarán dar solución a los objetivos planteados al principio. Al final de la investigación se deberá haber logrado ampliar los conocimientos sobre el mercado y los consumidores para llevar a cabo estrategias comerciales encaminadas al éxito de la empresa y su producto.

### 1.3 INVESTIGACIONES DE MERCADOS BASADAS EN ENCUESTAS

La investigación de mercado cuantitativa más utilizada para la obtención de datos primarios es la que se realiza apoyada por una encuesta. Los diferentes tipos de encuestas que se aplican en la actualidad son:

**Encuestas personales.** Son aquéllas en las que el entrevistador obtiene la información directamente del entrevistado por medio de una serie de preguntas. Las respuestas son registradas por el entrevistador en los espacios previamente asignados en el cuestionario estructurado. Estas encuestas pueden ser realizadas en puntos de afluencia (centro comercial, parque, vía pública, etc.) o en la casa del entrevistado. Debido a los niveles de inseguridad que se viven en México, son más frecuentes las entrevistas en puntos de afluencia que en casa ya que la gente se encuentra temerosa de abrir la puerta a extraños y además a proporcionar cualquier tipo de información. Este es el método más utilizado, ya que permite obtener mejores resultados.

**Encuestas por teléfono.** Son similares a las personales sólo que la comunicación se lleva a cabo a través de una línea telefónica. Este método suele ser el más rápido y flexible para la obtención de información. Algunas empresas dedicadas a realizar entrevistas telefónicas cuentan con una importante herramienta conocida como CATI (Computer Assisted Telephone Interview) por medio de la cual es posible capturar las respuestas en una computadora y que éstas se integren a la base de datos al mismo tiempo en que se está realizando la entrevista.

**Encuestas por correo.** Los cuestionarios son enviados y regresados por correo postal o electrónico. Junto con el cuestionario se debe anexar una carta en donde se explique el objetivo del cuestionario y, en algunas ocasiones, se suele ofrecer algún tipo de premio para quién conteste y regrese el cuestionario. Este tipo de encuestas es muy poco confiable porque difícilmente se puede detectar quién es la persona que contesta el cuestionario, además de que generalmente la tasa de respuesta es muy baja debido a que los

entrevistados, por falta de tiempo u otras razones, no se muestran interesados en contestar el cuestionario y mucho menos a enviarlo de regreso. Es por ello que se debe enviar el mayor número posible de cuestionarios para lograr obtener la información suficiente para realizar el estudio. La principal ventaja de las encuestas por correo es que tienen un bajo costo.

**Encuestas a través de internet.** En este tipo de encuestas, el cuestionario se colocan en un servidor y cuando los entrevistados visitan ese URL<sup>7</sup> llenan el cuestionario en línea y los datos regresan al servidor en donde se compilan al instante. Para realizar este tipo de encuestas se requiere de una plataforma de investigación denominada CAWI (Computer Assisted Web Interview).

**Encuestas administradas.** Los cuestionarios se encuentran en lugares visibles y al alcance de la gente, los cuales pueden ser contestados en cualquier momento y por cualquier persona interesada en hacerlo. Por ejemplo, algunos hoteles, restaurantes y tiendas de ropa recurren a este tipo de encuestas para conocer de qué forma sus clientes evalúan la calidad del servicio y/o producto recibido.

Una vez que la empresa ha decidido implantar este tipo de investigación como instrumento para cubrir sus necesidades específicas de información, se procede a realizar cuidadosamente cada una de las siguientes acciones:

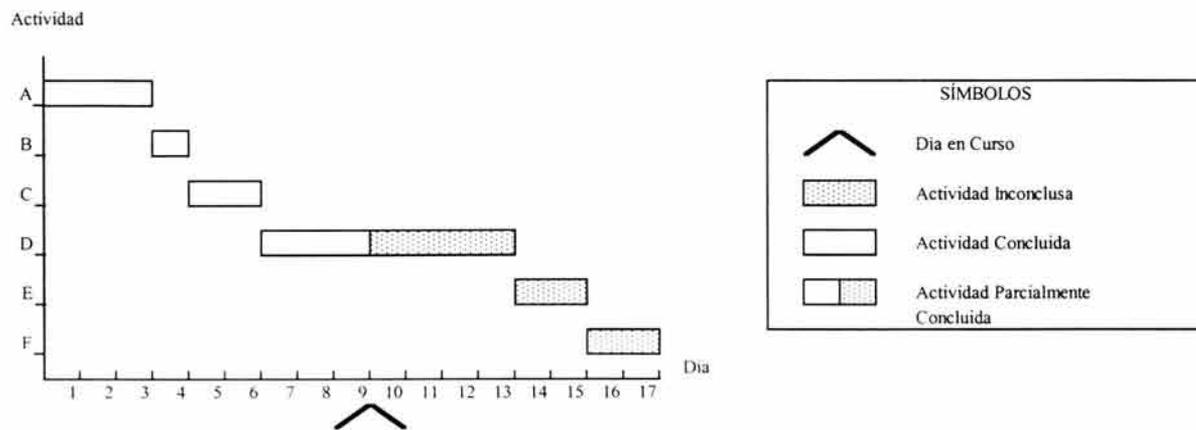
1 - Calendarización de actividades del área de campo y producción.

Es de suma importancia estimar la cantidad de tiempo que se requerirá para desarrollar las diferentes tareas que la investigación involucra, de tal forma que éstas se concluyan oportunamente. Así mismo, es importante darle seguimiento a la investigación de mercado de manera permanente comparando los tiempos programados con los realizados hasta el

---

<sup>7</sup> URL (Uniform Resource Locator): Localizador uniforme de recursos, dirección que identifica un documento o recurso en internet.

momento y en caso de ser necesario, acelerar o reprogramar actividades para que el estudio se pueda concluir a tiempo. Los diagramas de Gantt, gráficas de PERT o simples cronogramas (cuadros de doble entrada) son herramientas de gran utilidad para la programación de actividades. A continuación se muestra un diagrama de Gantt a manera de ejemplo en el cual se presenta la programación de seis diferentes tipos de actividades (A-F).



GRÁFICA 1.1  
Diagrama de Gantt

## 2 – Plan de Muestreo.

Es en esta etapa de la investigación donde se debe tomar decisiones en cuanto a tres factores: *la unidad de muestreo, el tamaño de la muestra y el procedimiento de muestreo.*

En principio, es importante identificar la población objetivo (población diana) y analizar si todos los elementos que la integran son factibles para obtener la muestra. Si esto no es así, será necesario eliminar de la población objetivo las partes de la población no factibles de muestrear y trabajar sólo en base de lo que será la población investigada (población muestreada). Por ejemplo, si se pretende realizar una investigación de mercado en torno a toallas sanitarias dirigida a mujeres en edad reproductiva en cierto estado del país (población objetivo), sería conveniente excluir a las zonas de extrema pobreza en las cuales

el consumo de este tipo de productos es prácticamente nulo y, de esta forma, redefinir a la población que estará sujeta a la realización de la investigación (población investigada).

La *unidad de muestreo* consiste en decidir a qué elementos de la población se va a entrevistar. En muchas ocasiones resulta evidente cuál es la unidad de muestreo. Continuando con el ejemplo expuesto anteriormente, si la investigación es acerca de toallas sanitarias, la unidad de muestreo serán mujeres en edad reproductiva. Algunas veces es necesario llevar a cabo un *submuestreo*, por ejemplo en el *muestreo por conglomerados*, debido a que una unidad de la población (*unidad primaria*) se puede dividir en cierto número de unidades más pequeñas o subunidades (*unidad secundaria*). Por ejemplo, una manzana territorial podría ser la unidad primaria y las viviendas que se localizan en dicha manzana serían las unidades secundarias.

El *tamaño de la muestra* establece el número total de personas que deben ser entrevistadas y que es adecuado para cometer un determinado error de muestreo prefijado (margen de error). Este error de muestreo dado en las encuestas resulta al considerar sólo una muestra y no al examinar a toda la población. Por otro lado, el sesgo en la selección y la imprecisión de las respuestas son ejemplos de los errores que no son de muestreo, los cuales no se pueden atribuir a la variabilidad entre las respuestas. Es importante señalar que un tamaño de muestra grande no siempre garantiza que la muestra sea representativa<sup>8</sup> y por ello resulta más conveniente utilizar una muestra pequeña pero que realmente sea representativa de la población. La teoría del muestreo proporciona un marco dentro del cual es posible obtener un tamaño de muestra óptimo para la investigación. Sin embargo, en la práctica muchas empresas utilizan para sus encuestas de opinión métodos a la medida por ejemplo, estableciendo un tamaño de muestra mínimo de 100 para cada grupo o segmento de la población que es de interés.

---

<sup>8</sup> Una muestra representativa es aquella en la que todos los segmentos de la población están incluidos en la muestra en sus proporciones correctas de la población. RITCHEY, Ferris (2002). Estadística para las Ciencias Sociales. El Potencial de la Imagen Estadística. México: McGraw-Hill. p. 40

El *procedimiento de muestreo* consiste en determinar de qué forma habrá de seleccionarse a las personas que se entrevistará, lo que garantizará que la muestra sea representativa. Se pueden distinguir dos tipos de muestreo: probabilístico y no probabilístico.

En el **muestreo probabilístico** se requiere de una lista completa de las unidades de muestreo (*marco*) y de un procedimiento para seleccionar a los entrevistados conforme a una probabilidad conocida. Entre los tipos de muestreo probabilísticos más comunes se encuentran *el muestreo aleatorio simple, el muestreo aleatorio estratificado, el muestreo por conglomerados y el muestreo sistemático.*

En el *muestreo aleatorio simple*<sup>9</sup> o *muestreo irrestricto aleatorio* cada unidad tiene la misma oportunidad de selección que cualquier otra unidad en cada etapa sucesiva del proceso de selección.

En el *muestreo aleatorio estratificado* la población se divide en subpoblaciones (estratos) las cuales no se traslapan. De cada uno de los estratos establecidos se toma una muestra aleatoria simple. Si el muestreo aleatorio estratificado se realiza correctamente, casi siempre es posible obtener una varianza más pequeña que la dada aplicando un muestreo aleatorio simple. Y esto es posible ya que la varianza de las observaciones dentro de cada estrato es generalmente menor que la varianza en toda la población.

En el *muestreo por conglomerados* la población se divide en grupos o conglomerados y se seleccionan sólo algunos de ellos para conformar la muestra. A menudo se recomienda usar este tipo de muestreo probabilístico cuando no se disponga de una lista confiable de los elementos de la población (*marco*) y además, la construcción de dicha lista resulte demasiado costosa. Por ello, es común utilizar como marco divisiones geográficas ya establecidas como pueden ser municipios, delegaciones, ciudades, etc. Así mismo, se

---

<sup>9</sup> Una muestra aleatoria es aquella en la cual cada persona (u objeto) en la población tiene la misma oportunidad de ser seleccionada (o) para la muestra. Ibidem, RITCHEY, Ferris. p. 41

prefiere el muestreo por conglomerados debido al bajo costo que éste involucra ya que se reducen considerablemente los desplazamientos de parte del personal de campo.

En el *muestreo sistemático o pseudo muestreo aleatorio simple* las unidades de la población se enumeran bajo cierto orden y se van eligiendo de manera secuencial a intervalos iguales aquéllas que formarán parte de la muestra. A diferencia del muestreo aleatorio simple, el muestreo sistemático garantiza barrer toda la población ya que la muestra se reparte uniformemente sobre ésta. Además, en el muestreo sistemático es más fácil y rápido sacar la muestra y hacerlo sin cometer errores.

Por el contrario, en el **muestreo no probabilístico** no es necesario desarrollar un marco de muestreo y por lo tanto los costos y los problemas que éste representa son eliminados. Al seleccionar este tipo de muestreo se debe ser extremadamente cuidadoso porque la precisión en los datos obtenidos se puede ver afectada por una mala selección de la muestra, ya que no existe un soporte probabilístico, por ejemplo en el caso de la existencia de sesgos ocultos. El *muestreo de juicio*, el *muestreo por cuotas* y el *muestreo por conveniencia* son algunos de los tipos de muestreo no probabilístico más utilizados en la actualidad.

En el *muestreo de juicio*, los entrevistados son seleccionados simple y sencillamente a criterio del entrevistador o encuestador sin tomar en cuenta ninguna restricción. En el *muestreo por cuotas*, los entrevistados son seleccionados a criterio del entrevistador con la restricción de que se deben cumplir una serie de cuotas preestablecidas que se supone caracterizan a la población (por ejemplo sexo, edad, nivel socioeconómico, escolaridad, etc.). En el *muestreo por conveniencia* como su nombre lo indica, se hace una selección de entrevistados que resulten convenientes para realizar la investigación. Por ejemplo un grupo de suscriptores de cierta revista, un grupo de investigadores científicos, mujeres que asisten a una tienda departamental determinado día, etc.

### 3 - Diseño del cuestionario.

El diseño del cuestionario para un estudio basado en encuestas debe ser llevado a cabo con extremo cuidado con el fin de obtener por medio de éste respuestas significativas que faciliten la toma de decisiones. Al diseñar el cuestionario lo que se estará haciendo es traducir los objetivos de la investigación en preguntas específicas para la recolección de datos.

#### ➤ *Redacción y Formato*

La redacción de las preguntas y el formato del cuestionario son aspectos fundamentales que influyen en el éxito o fracaso del cuestionario. Al redactar las preguntas se debe utilizar un lenguaje natural y familiar que sea claro y preciso, y que permita tanto al entrevistador como al entrevistado su fácil comprensión. Así mismo, se deben evitar las preguntas tendenciosas y las preguntas dobles.

Algunas consideraciones importantes para el diseño del formato del cuestionario son:

- Disponer de suficiente espacio en blanco para anotar las respuestas y entre pregunta y pregunta.
- Mantener un estilo consistente.
- Agrupar preguntas del mismo tema.
- Elegir un tipo de letra que facilite su lectura.
- Numerar las preguntas.
- Evitar cortar preguntas por el hecho de que el espacio en la hoja se terminó y continuar la redacción en la siguiente página.
- Señalar claramente los saltos de página y de pregunta.
- Especificar si la pregunta es de respuesta múltiple (**RM**) o de respuesta sencilla (**RS**).

➤ *Tipos de Preguntas*

En el proceso del diseño del cuestionario se pueden distinguir tres tipos básicos de preguntas: *abiertas*, *cerradas* y *semiabiertas*.

Las *preguntas abiertas* son aquéllas en las que el entrevistado tiene la libertad de expresar la respuesta con sus propias palabras. Al incluir este tipo de preguntas en el cuestionario, es importante contemplar un espacio destinado a la posterior codificación de las respuestas emitidas. Un ejemplo de pregunta abierta sería:

¿Qué es lo que más le agradó del producto que acaba de probar?

_____	<b>RM</b>			
_____	<table border="1"><tr><td> </td><td> </td><td> </td></tr></table>			
_____	<table border="1"><tr><td> </td><td> </td><td> </td></tr></table>			
_____	<table border="1"><tr><td> </td><td> </td><td> </td></tr></table>			

En el ejemplo anterior, se esperaría obtener como máximo tres opciones de respuestas diferentes, las cuales se deberán señalar en las líneas correspondientes y posteriormente asignar a cada respuesta un código de tres dígitos que se especificará en los cuadritos que aparecen al lado derecho de las líneas.

En cambio, las *preguntas cerradas* se encuentran limitadas por el número de opciones de respuesta. Las preguntas cerradas que constan de sólo dos posibles alternativas de respuesta (por ejemplo si y no) se les conoce como preguntas dicotómicas. Por ejemplo, una pregunta cerrada de un cuestionario podría decir:

¿Cuántos televisores que funcionen tiene usted en su hogar?

	<b>RS</b>
0. Ninguno	0
1. Uno	1
2. Dos	2
3. Tres	3
4. Cuatro	4
5. Cinco	5
6. Más de Cinco	6

Las *preguntas semiabiertas o precodificadas* son aquellas en donde se cuenta con una codificación previa de algunas posibles respuestas pero además el entrevistado tiene la posibilidad de emitir una respuesta diferente a las opciones ya establecidas. La siguiente pregunta es un ejemplo al respecto:

¿A qué banco acude con mayor frecuencia?

	<b>RS</b>
0. Ninguno	0
1. Banamex/Citybank	1
2. BBVA/Bancomer	2
3. Banco Azteca	3
4. Bancrecer	4
5. Banorte	5
6. Santander/Serfin	6
7. Scotiabank/Inverlat	7
8. Otro ( <b>Especificar</b> )	8

\_\_\_\_\_

De acuerdo al número de respuestas que el entrevistado puede dar a una pregunta, las preguntas se pueden clasificar en *preguntas de respuesta múltiple* y *preguntas de respuesta sencilla*.

Las *preguntas de respuesta múltiple* permiten a los entrevistados dar una o más respuestas para una pregunta en específico. Por ejemplo al preguntar:

¿Qué tiendas departamentales ha visitado en los últimos tres meses?

	<b><i>RM</i></b>			
0. Ninguna	0			
1. Liverpool	1			
2. Sanborns	2			
3. Sears	3			
4. Palacio de Hierro	4			
5. Otras ( <i>Especificar</i> )	5			
_____	<table border="1" style="display: inline-table; border-collapse: collapse;"> <tr> <td style="width: 20px; height: 20px;"></td> <td style="width: 20px; height: 20px;"></td> <td style="width: 20px; height: 20px;"></td> </tr> </table>			
_____	<table border="1" style="display: inline-table; border-collapse: collapse;"> <tr> <td style="width: 20px; height: 20px;"></td> <td style="width: 20px; height: 20px;"></td> <td style="width: 20px; height: 20px;"></td> </tr> </table>			
_____	<table border="1" style="display: inline-table; border-collapse: collapse;"> <tr> <td style="width: 20px; height: 20px;"></td> <td style="width: 20px; height: 20px;"></td> <td style="width: 20px; height: 20px;"></td> </tr> </table>			

Nótese que al principio de las opciones de respuesta de esta pregunta se señala con las letras ***RM*** que la respuesta puede ser múltiple.

En las *preguntas de respuesta sencilla o única*, el entrevistado solamente puede emitir una respuesta única a una pregunta en cuestión. A continuación se ejemplifica este tipo de preguntas:

¿Cuál es su estado civil?

	<b><i>RS</i></b>
1. Soltero (a)	1
2. Casado (a)	2
3. Divorciado (a)	3
4. Viudo (a)	4
5. Unión Libre	5

Las letras ***RS*** especifican que la respuesta debe ser sencilla o única.

➤ *Escalas de Medida*

De manera frecuente, al momento de redactar las posibles opciones de respuesta para preguntas cerradas y semiabiertas se hace uso de escalas de medida, especialmente para la medición de actitudes. Dentro de la investigación de mercados, la actitud se define como la predisposición de los clientes o consumidores para responder de acuerdo a sus creencias o sentimientos emocionales, ante un producto o servicio.

Comúnmente se habla de cuatro niveles de medida los cuales son: escala nominal, escala ordinal, escala de intervalo y escala de razón, y por lo tanto, de cuatro tipos de variables: variable nominal, ordinal, de intervalo y de razón.

**Escala Nominal.** Los números funcionan como simples etiquetas para representar diferentes categorías, clases o palabras y por lo tanto las propiedades aritméticas carecen de sentido. La siguiente pregunta hace uso de una escala nominal.

¿Cuál es el medio de comunicación que más utiliza?

	<b>RS</b>
0.Ninguno	0
1.Radio	1
2.Televisión	2
3.Prensa	3
4.Revistas	4
5.Internet	5

**Escala Ordinal.** En este tipo de escala, las clases o categorías no sólo son diferentes entre sí, sino que además tienen una cierta relación entre ellas porque una es superior o inferior a la otra ya que guardan un orden. Por ejemplo, al pedir que se evalúe la siguiente afirmación, se hace uso de una escala ordinal:

La servilletas para la cocina de la marca “Dream” son las más absorbentes:

	<b>RS</b>
1. Totalmente de acuerdo	1
2. Casi totalmente de acuerdo	2
3. De acuerdo	3
4. Indeciso	4
5. En desacuerdo	5
6. Casi totalmente en desacuerdo	6
7. Totalmente en desacuerdo	7

Se debe tener especial cuidado en que las categorías de las variables nominales y ordinales sean exhaustivas, mutuamente excluyentes y basadas en un único principio clasificatorio.

**Escala de intervalo.** Escala de medición que permite calcular e interpretar diferencias (además de asignar nombre y orden) entre los datos. En este tipo de escala el cero se determina de manera arbitraria (cero arbitrario) y no significa la ausencia del atributo que se mide. Un ejemplo típico de una variable que se mide en esta escala es la temperatura, en donde se sabe que la diferencia entre 10°C y 15°C es la misma que entre 25°C y 30°C y si se dice que un líquido se encuentra a 0°C, no significa que no tiene temperatura.

**Escala de razón.** Las escalas de razón son similares a las de intervalo con respecto a que los intervalos entre los números son iguales, pero en este caso el cero (cero real o absoluto) corresponde con la ausencia del atributo que se mide. Existen muchas variables en la investigación de mercados que se miden en una escala de razón como por ejemplo la cantidad de dinero que acostumbra gastar en el supermercado, el número de autos que posee y el número de hijos que tiene.

A la escala que es de tipo nominal u ordinal se le suele llamar **escala no métrica** ya que la variable resultante es de tipo *cualitativo*. Por el contrario, a la escalas de intervalo y a las escalas de razón se les conoce como **escalas métricas** ya que las variables resultantes son de tipo *cuantitativo*.

Las variables cuantitativas pueden ser *discretas* o *continuas*. Las variables discretas sólo pueden asumir un número limitado de valores dentro de un intervalo y no se pueden subdividir en partes más pequeñas. En cambio, las variables continuas pueden asumir un número infinito de valores en un intervalo y se pueden subdividir en partes más pequeñas.

Con el fin de mejorar la medición de actitudes, se han desarrollado algunos tipos especiales de escalas (principalmente ordinales) llamadas **escalas indirectas**, entre las cuales podemos mencionar: la escala de Likert, el diferencial semántico, el escalograma de Guttman y la escala de Thurstone. La escala de Likert, el diferencial semántico y la escala de Thurstone son métodos centrados en los sujetos, mientras que el escalograma de Guttman es un método centrado en las respuestas.

**Escala de Likert.** Método desarrollado por Rensis Likert en 1932, en el cual se evalúa una frase o afirmación por medio de una escala ordinal que refleje el grado de acuerdo o desacuerdo con la misma. A cada grado de acuerdo o desacuerdo se le asigna un valor numérico. La puntuación de una escala de Likert se obtiene sumando los valores obtenidos en cada afirmación, es por ello que se le llama *escala aditiva*. Veamos el siguiente ejemplo en dónde se emplea una escala de Likert:

Los productos del departamento de frutas y verduras de esta tienda de autoservicio siempre están frescos:

	<b><i>RS</i></b>
1. Totalmente de acuerdo	1
2. De acuerdo	2
3. Ni de acuerdo ni en desacuerdo	3
4. En desacuerdo	4
5. Totalmente en desacuerdo	5

**Diferencial Semántico.** Este método para medir actitudes fue desarrollado por Osgood, Suci y Tannebaund en 1957 el cual requiere de un conjunto de adjetivos bipolares (opuestos o extremos) que permitan clasificar el concepto de interés. El procedimiento consiste en pedir al entrevistado que señale el punto exacto que indique su opinión con respecto al tema



- Estudiar la distribución de frecuencias de las clasificaciones para cada aseveración y eliminar aquéllas a las que los diferentes jueces hayan dado clasificaciones muy dispersas, es decir que se encuentren en diferentes renglones.
- Determinar el valor, en la escala, de cada una de las afirmaciones restantes, esto es, el número del renglón en el que cae la mediana de la distribución.
- Seleccionar una o dos afirmaciones de cada uno de los once renglones para la escala final. Aquéllas que presenten la oscilación más estrecha de clasificaciones son preferidas por ser las más factibles. Se enumeran entonces en un orden al azar para formar la escala final.

Las escalas de Thurstone no son muy empleadas en la investigación de mercados, probablemente porque la tarea de prepararlas requiere mucho tiempo.

#### 4 - Prueba piloto del cuestionario.

Para poder garantizar la elaboración de un buen cuestionario, es indispensable realizar, cuando menos, una prueba previa del mismo con una pequeña muestra bajo las mismas condiciones para las que está programado el estudio.

Por medio de la prueba piloto se puede examinar el documento en cuanto a claridad y extensión y, en caso de ser necesario, realizar los cambios pertinentes en el cuestionario con el fin de asegurar que la información que se obtenga sea la esperada.

#### 5 - Recopilación, codificación de preguntas abiertas, captura, tabulación, análisis y presentación de la información.

Una vez que se tiene el cuestionario final para la realización de la investigación, es necesario brindar capacitación al personal de campo para comenzar el trabajo de levantamiento de datos. Conforme van llegando los cuestionarios contestados, se procede a realizar una guía de codificación para las preguntas abiertas (asignación de un número a

cada categoría de respuestas) y a capturar todas las respuestas obtenidas para construir la base de datos.

Al cumplir con el tamaño de muestra establecido y con la base de datos completa, se genera una tabulación de la información, la cual consiste en hacer tablas de frecuencias para cada una de las preguntas del cuestionario. Finalmente se realiza el análisis de la información (apoyados principalmente en técnicas estadísticas) y la presentación de resultados.

#### **1.4 TIPOS DE ESTUDIOS DE INVESTIGACIONES DE MERCADOS BASADAS EN ENCUESTAS**

Dentro de las investigaciones de mercados basadas en encuestas se puede distinguir una amplia variedad de estudios entre los que destacan:

**Imagen y posicionamiento.** Este tipo de investigaciones tienen como objetivo descubrir de qué forma el consumidor percibe y registra en su mente (posiciona) cierto conjunto de productos, servicios o marcas en algunos conceptos clave. La imagen corresponde al resultado final de esa percepción.

**Usos, hábitos y actitudes.** El propósito de estas investigaciones es conocer aspectos específicos del comportamiento del consumidor tales como hábitos de compra y consumo de cierto producto o servicio, así como los usos que le da a este, etc.

**Recordación publicitaria.** Estos estudios intentan evaluar el grado de recordación de una campaña publicitaria en el consumidor, para determinar si ésta fue exitosa o no.

**Prueba de concepto/producto.** Por medio de estas investigaciones, los fabricantes pueden poner a prueba su producto antes de sacarlo al mercado y saber si éste responde a los gustos, necesidades y deseos del público consumidor.

**Perfil de cliente.** Estos estudios analizan las características de los consumidores regulares de un producto o servicio. Se investiga qué factores motivan su comportamiento de compra, así como conocer sus características sociodemográficas por medio de información relacionada con su edad, sexo, estado civil, escolaridad, ingreso familiar, etc. De igual forma, se indaga acerca de actividades específicas que realiza de manera frecuente y de sus gustos y preferencias.

**Segmentación de mercados.** Son investigaciones por medio de las cuales se trata de identificar los subconjuntos o segmentos de consumidores que comparten necesidades, características o comportamientos similares. Además, estos estudios permiten obtener la información necesaria para la elaboración y la ejecución de la estrategia de mercadotecnia de una empresa y garantizar la correcta asignación de recursos entre mercados y productos. Por ejemplo, la información que la segmentación de mercados proporciona puede ser utilizada por la compañía para dirigir de una mejor manera sus campañas publicitarias.

Algunos de los factores que se consideran para sentar las bases sobre las cuales se va a dividir el mercado son: factores geográficos (ciudad, país, etc.), factores demográficos (sexo, edad, estado civil, etc.), factores conductuales (relacionados con la preferencia y uso de productos), factores socioculturales (clase social, valores esenciales), características psicológicas y estilo de vida.

Estos tipos de estudios se pueden realizar de manera independiente o bien, se pueden combinar dos o más de ellos en una investigación de mercado.

## **1.5 TÉCNICAS ESTADÍSTICAS ÚTILES EN LAS INVESTIGACIONES DE MERCADOS BASADAS EN ENCUESTAS**

La mayoría de las investigaciones de mercado implican la recolección de una gran cantidad de datos de muchas variables (multivariantes). Es por ello, que resulta necesario el dominio de algunas técnicas estadísticas que permitan la identificación de interdependencias entre

una serie de variables, en especial de aquéllas que se usan con gran frecuencia en diversos tipos de estudios de mercado. Debido al enorme crecimiento de disponibilidad de hardware y software estadístico en los últimos años, se ha incrementado el uso de técnicas estadísticas multivariadas. Algunas de estas técnicas estadísticas que se suelen utilizar para llevar a cabo un análisis de mercado son:

- *Análisis de conglomerados*

Este tipo de análisis se realiza principalmente en los estudios de segmentación de mercados con el fin de identificar y describir los diferentes sectores que existen en determinado mercado, así como conocer el tamaño de cada uno.

El análisis de conglomerados busca dentro de los datos obtenidos en el cuestionario para identificar a los entrevistados que dan respuestas idénticas o parecidas a alguna determinada combinación de preguntas. Los entrevistados de cada grupo o conglomerado son diferentes de los entrevistados de los demás conglomerados.

El análisis de conglomerados puede aplicarse a datos de variables continuas, de escalas y categóricas (es decir escalas métricas y no métricas). Por medio de un diagrama de árbol conocido como *dendograma* puede mostrarse el procedimiento para formar los conglomerados. El primer paso para desarrollar un análisis de conglomerados consiste en identificar la similitud o proximidad entre dos objetos (entrevistados) por medio de lo que se conoce como la *medida de similitud*. Aunque no existe una forma ideal para medir dicha similitud, la distancia euclidiana es la más comúnmente usada. La fórmula a través de la cual se obtiene la distancia del punto *i* al punto *j* es:

$$d_{ij} = \left[ \sum_{v=1}^p (x_{iv} - x_{jv})^2 \right]^{1/2}$$

donde:  $d_{ij}$  = la distancia entre objetos (personas) *i* y *j*

$x_{iv}$  = el valor de la variable  $v$  para el objeto  $i$

$x_{jv}$  = el valor de la variable  $v$  para el objeto  $j$

$p$  = el número de variables

Por ejemplo<sup>11</sup>, suponga que se obtienen las respuestas de doce entrevistados para dos preguntas diferentes, en las cuales los entrevistados utilizaron una escala de 10 puntos para contestar. A continuación se muestran las respuestas de los entrevistados para las dos preguntas:

		Pregunta	
		1	2
Entrevistado	1	1	2
	2	5	2
	3	6	2
	4	5	3
	5	8	5
	6	2	8
	7	2	9
	8	3	9
	9	9	9
	10	4	10
	11	8	10
	12	10	10

**TABLA 1.1**  
**Matriz de Datos**

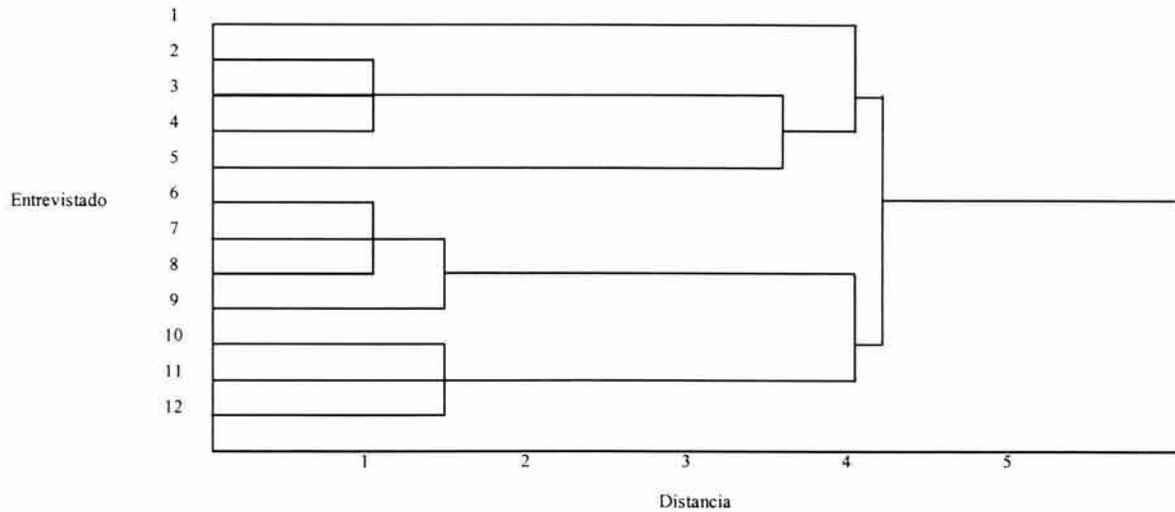
La tabla 1.2 muestra las distancias calculadas para las respuestas de los doce entrevistados:

		Entrevistado											
		1	2	3	4	5	6	7	8	9	10	11	12
Entrevistado	1	0	4.0	5.0	4.1	7.6	6.1	7.1	7.3	10.6	8.5	10.6	12.0
	2		0	1.0	1.0	4.2	6.6	7.6	7.3	8.1	8.1	8.5	9.4
	3			0	1.4	3.6	7.2	8.1	7.6	7.6	8.2	8.2	8.9
	4				0	3.6	5.8	6.7	6.3	7.2	7.1	7.6	8.6
	5					0	6.7	7.2	6.4	4.1	6.4	5.0	5.4
	6						0	1.0	1.4	7.1	2.8	6.3	8.2
	7							0	1.0	7.0	2.2	6.1	8.1
	8								0	6.0	1.4	5.1	7.1
	9									0	5.1	1.4	1.4
	10										0	4.0	6.0
	11											0	2.0
	12												0

**TABLA 1.2**  
**Matriz de Distancias**

<sup>11</sup> Ibidem, BOYD, Harper. pp. 615-620

Para esta matriz de distancias, el correspondiente dendograma quedaría de la siguiente forma:



**GRÁFICA 1.2**  
**Dendograma**

Los doce entrevistados están enumerados a lo largo del eje vertical, mientras que el eje horizontal representa las distancias entre ellos. Las líneas horizontales, que se originan en cada entrevistado y se extienden hacia la derecha, indican que si la distancia mínima (MD) se establece en algún número por debajo de 1.0, los dos entrevistados no están separados por la distancia mínima, de modo que no pueden formarse grupos. Si MD se establece en 1.0, se puede observar en la matriz de distancias que hay cuatro distancias entre entrevistados de 1.0 (entrevistados 2 y 3, 2 y 4, 6 y 7, 7 y 8). Por lo tanto, cuando MD es igual a 1.0, los entrevistados 3 y 4 se unen al 2 para formar un grupo, y los entrevistados 6 y 8 se unen al sujeto 7 para formar un segundo grupo. El dendograma muestra esta información de grupos al unir las tres líneas horizontales relacionadas con los entrevistados 2, 3 y 4 en una sola línea horizontal. El agrupamiento de los entrevistados 6, 7 y 8 también se identifica de la misma manera. Por ende, si el investigador había establecido la MD en 1.0, este análisis de grupos daría por resultado la formación de dos grupos, pero la mitad de los entrevistados (6 de 12) no están incluidos en ninguno de los grupos.

- *Análisis factorial*

El análisis factorial es utilizado principalmente en los datos de variables continuas o de escalas de intervalo (escalas métricas) para preguntas relacionadas a cierto producto o servicio.

Al contar con una escala de intervalo, el entrevistado podrá proporcionar su nivel de acuerdo o desacuerdo con una serie de afirmaciones las cuales son previamente redactadas en el diseño del cuestionario. De este modo, se podrá conocer las características o factores principales que los entrevistados consideran importantes. Al identificar las características principales que los consumidores consideran significativas en el producto, la compañía podría mejorar el diseño de éste y resaltar tales características en la publicidad.

El análisis factorial aplica un avanzado análisis de correlación a las respuestas dadas en la serie de afirmaciones. Si las respuestas de un grupo de tres o más afirmaciones están muy correlacionadas, entonces se dice que tales afirmaciones (variables observables) miden algún factor (variable no observable o latente) que es común a ellas y de esta forma reducir el número de variables a un conjunto más manejable<sup>12</sup>.

Veamos, por ejemplo, el siguiente estudio piloto hipotético<sup>13</sup> realizado por un banco a 15 personas para determinar si se deben desarrollar programas especiales de mercadotecnia para varios segmentos clave. Uno de los objetivos del estudio de investigación se refiere a las actitudes hacia la banca. A los entrevistados se les preguntó su opinión en una escala de cero a nueve, acuerdo-desacuerdo, sobre las siguientes cuestiones:

- 1) Los bancos pequeños cobran una menor tasa de interés que los bancos grandes.
- 2) Los bancos grandes tienen más probabilidades de cometer errores que los bancos pequeños.

---

<sup>12</sup> Las variables observables son aquellas variables que son medibles directamente.

Las variables no observables o latentes se definen en términos conceptuales que no pueden ser directamente medidos.

<sup>13</sup> AAKER, David y DAY (1990). Investigación de Mercados. Colombia: McGraw-Hill. Tercera Edición. pp. 469-471

- 3) Los cajeros no necesitan ser extremadamente corteses y amistosos, es suficiente con ser atentos.
- 4) Quiero ser personalmente conocido en mi banco y ser tratado con cortesía especial.
- 5) Después de ser tratado en una forma impersonal o descuidada por una institución financiera, nunca utilizaría esa organización otra vez.

Los datos del estudio piloto se muestran a continuación:

		<i>Datos de Insumo</i>				
		<i>Variable</i>				
		1	2	3	4	5
<i>Entrevistado</i>	1	9	6	9	2	2
	2	4	6	2	6	7
	3	0	0	5	0	0
	4	2	2	0	9	9
	5	6	9	8	3	3
	6	3	8	5	4	7
	7	4	5	6	3	6
	8	8	6	8	2	2
	9	4	4	0	8	8
	10	2	8	4	5	7
	11	1	2	6	0	0
	12	6	9	7	3	5
	13	6	7	1	7	8
	14	2	1	7	1	1
	15	9	7	9	2	1

**TABLA 1.3**  
**Matriz de Datos**

Y los resultados obtenidos por medio del análisis factorial son los que se indican en la siguiente tabla:

		<i>Cargas Factoriales</i>	
		<i>Factor</i>	
		1	2
<i>Variable</i>	1	-0.24	0.72
	2	0.06	0.87
	3	-0.94	0.33
	4	0.94	0.21
	5	0.93	0.26
<i>Porcentaje de variación explicado</i>		55%	36%

**TABLA 1.4**  
**Cargas Factoriales**

Las cargas factoriales proporcionan una indicación de qué variables originales están correlacionadas con cada factor y el grado de correlación. A mayor carga, mayor representatividad del factor. Las cargas factoriales cercanas a +1 ó -1 indican una clara asociación entre la variable y el factor. Por otro lado, cargas cercanas a 0 indican ausencia de asociación. Es decir,

Cargas  $> \pm 0.30$  (nivel mínimo)

Cargas  $> \pm 0.40$  (importantes)

Cargas  $> \pm 0.50$  (significativas)

El objetivo principal es identificar la carga significativa para cada variable. La situación ideal es que cada variable tenga una sola carga significativa sobre un factor, aunque en la realidad puede haber variables con cargas significativas sobre varios factores o bien variables que no cargan sobre ningún factor.

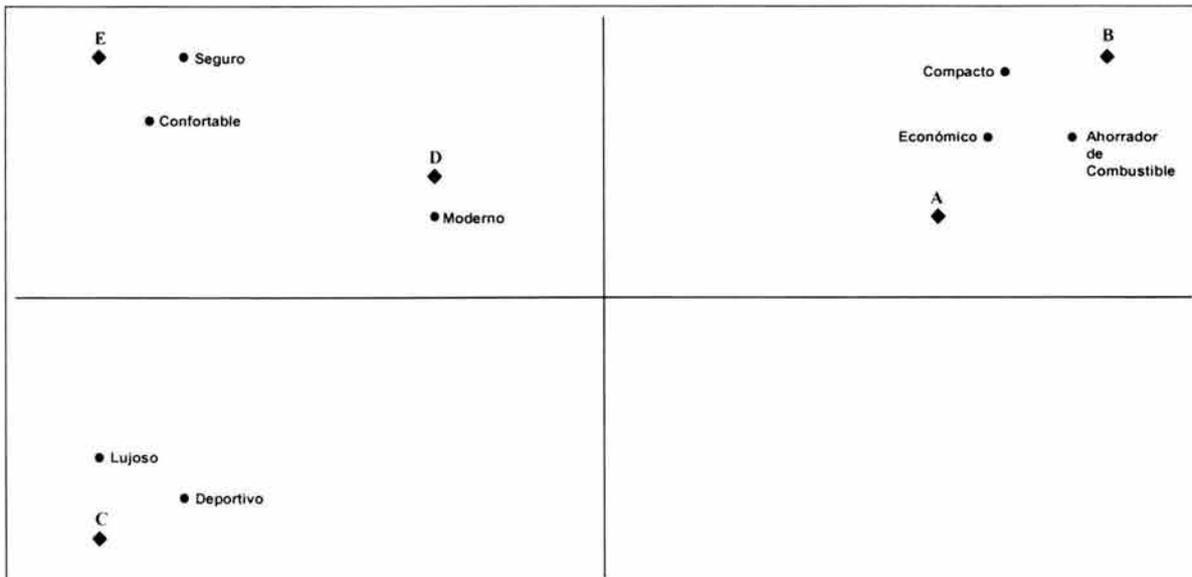
Entonces, volviendo al ejemplo, las variables 3, 4 y 5 se combinan para definir al primer factor, el cual podría ser denominado como un factor “personal”. El segundo factor está correlacionado en una forma más alta con las variables 1 y 2 y por tal motivo podría denominarse como un factor de “banco pequeño”.

- *Análisis de correspondencias*

Esta técnica se utiliza especialmente en las investigaciones de mercado de imagen y posicionamiento. Dicho análisis se desarrolla partiendo de una matriz de probabilidades (matriz de correspondencias) que se obtiene de una tabla de dos sentidos o de doble entrada de variables nominales llamada tabla de contingencia.

En términos generales, se puede decir que la principal ventaja del análisis de correspondencias es que permite visualizar claramente en un **mapa perceptual**, la asociación existente entre una serie de atributos, características o afirmaciones y ciertas marcas, productos o servicios. En la siguiente página se muestra un ejemplo de mapa perceptual.

La proximidad entre los atributos y los modelos de autos (A-E) indica que están asociados en la percepción del público. Por ejemplo, se puede observar que el auto del modelo C es percibido por los consumidores como lujoso y deportivo. Así mismo, se puede mencionar que el modelo de auto D es percibido principalmente como moderno y el modelo E como seguro y confortable. Finalmente, los modelos A y B compiten, ya que son percibidos de manera similar, pues comparten los siguientes atributos: ahorrador de combustible, económico y compacto.



**GRÁFICA 1.3**  
**Mapa Perceptual**

La teoría matemática del análisis de correspondencias y la construcción de mapas perceptuales será ampliamente explicada en el capítulo 3 de este trabajo.

## CAPÍTULO 2 – TABLAS DE CONTINGENCIA

### 2.1 INTRODUCCIÓN

Frecuentemente, los datos de una investigación de mercados incluye variables que son categóricas. Las **variables categóricas** son aquéllas con ciertas características de naturaleza cualitativa que no son cuantificables y que se pueden medir únicamente en cuanto a clases o categorías, las cuales son mutuamente excluyentes. Por ejemplo, el estado civil de una persona sólo se puede medir en cinco clases o categorías (soltero, casado, divorciado, viudo, separado). Por esto, se puede decir, que las variables categóricas son equivalentes a las escalas nominales.

Las tablas de datos cruzados de variables categóricas se conocen como **tablas de contingencia**. Una tabla de contingencia se construye generalmente con el propósito de analizar la relación entre clasificaciones.

Cuando solamente se cuenta con una variable de acuerdo con la cual los datos son clasificados se dice que se tiene una **tabla de contingencia de una dimensión**. A una tabla de contingencia que consta de dos variables de clasificación se le conoce como **tabla de contingencia de dos dimensiones ( $n \times p$ )**. La forma más sencilla de este tipo de tablas de contingencia es aquella que involucra dos variables dicotómicas (es decir, una tabla de contingencia con un total de cuatro celdas de frecuencias absolutas), comúnmente llamada **tabla de contingencia  $2 \times 2$** . Por otro lado, a la tabla de contingencia que implica un mayor nivel de complejidad por referirse a tres ( $n \times p \times q$ ) o más variables categóricas se la llama **tabla de contingencia multidimensional**. A lo largo de este capítulo se mostrará la forma de realizar el análisis de tablas de contingencia de dos dimensiones ( $n \times p$ ).

El análisis de tablas de contingencia es una técnica estadística ampliamente utilizada en la investigación social, médica y del comportamiento. Esta técnica intenta descubrir la

relación existente entre dos o más variables de tipo cualitativo (sexo, nivel socioeconómico, actitud, etc.), que son el tipo de variables que más se emplean en la ciencias sociales y en la investigación de mercados.

La forma general de una tabla de contingencia  $n \times p$  ( $n$  renglones y  $p$  columnas) que muestra la relación entre dos variables categóricas es:

		Columna						Total Renglón
		1	2	...	j	...	p	
I \ J	1	$k_{11}$	$k_{12}$	...	$k_{1j}$	...	$k_{1p}$	$k_{10}$
	2	$k_{21}$	$k_{22}$	...	$k_{2j}$	...	$k_{2p}$	$k_{20}$
	...	.	.		.		.	.
	...	.	.		.		.	.
Renglón	i	$k_{i1}$	$k_{i2}$	...	$k_{ij}$	...	$k_{ip}$	$k_{i0}$
	...	.	.		.		.	.
	... <td>.</td> <td>.</td> <td></td> <td>.</td> <td></td> <td>.</td> <td>.</td>	.	.		.		.	.
	n	$k_{n1}$	$k_{n2}$	...	$k_{nj}$	...	$k_{np}$	$k_{n0}$
Total Columna		$k_{01}$	$k_{02}$	...	$k_{0j}$	...	$k_{0p}$	$k_{00}$

**TABLA 2.1**  
**Tabla de Contingencia K**

En donde:

$k_{ij}$  son las frecuencias observadas

$k_{i0}$  son los totales por renglón

$k_{0j}$  son los totales por columna

$k_{00}$  es el total general de la tabla de contingencia

Si  $k_{ij}$  es la frecuencia observada en el renglón  $i$  y la columna  $j$  de la tabla de contingencia, sea  $p_{ij}$  la probabilidad (desconocida) de que una observación caiga en esta misma celda. De manera que las probabilidades de las celdas de una tabla de contingencia  $n \times p$  son:

I \ J	Columna						Total Renglón	
	1	2	...	j	...	p		
Renglón	1	$p_{11}$	$p_{12}$	...	$p_{1j}$	...	$p_{1p}$	$P_{10}$
	2	$p_{21}$	$p_{22}$	...	$p_{2j}$	...	$p_{2p}$	$P_{20}$
	...	...	...	...	...	...	...	...
	...	...	...	...	...	...	...	...
	i	$p_{i1}$	$p_{i2}$	...	$p_{ij}$	...	$p_{ip}$	$P_{i0}$
	...	...	...	...	...	...	...	...
n	$p_{n1}$	$p_{n2}$	...	$p_{nj}$	...	$p_{np}$	$P_{n0}$	
Total Columna	$p_{01}$	$p_{02}$	...	$p_{0j}$	...	$p_{0p}$	$p_{00}=1$	

**TABLA 2.2**  
**Matriz de Probabilidades P**

Obsérvese que la suma de las probabilidades totales por renglón es igual a 1:

$$p_{10} + p_{20} + \dots + p_{n0} = 1$$

y lo mismo sucede para las probabilidades totales por columna:

$$p_{01} + p_{02} + \dots + p_{0p} = 1$$

Si las clasificaciones por renglón y por columna son independientes entre sí, la probabilidad de una celda deberá ser igual al producto de sus respectivas probabilidades de renglón y columna:

$$p_{ij} = p_{i0} p_{0j}$$

## 2.2 PRUEBA DE INDEPENDENCIA JI-CUADRADA DE PEARSON<sup>14</sup>

Al pretender realizar un análisis de tablas de contingencia lo que realmente nos interesa saber es si la clasificación por renglón es independiente de la clasificación por columna. Y para ello, el primer paso que se debe dar es realizar el cálculo de las **frecuencias esperadas** (o teóricas) para cada una de las celdas de la tabla de contingencia. La frecuencia esperada para cualquier celda de una tabla de contingencia  $n \times p$  se obtiene multiplicando el total del renglón y el total de la columna a los que pertenece y dividiendo este producto entre el total de la tabla:

$$\text{frecuencia esperada} = \frac{\text{total renglón} \times \text{total columna}}{\text{total tabla}}$$

Estas frecuencias esperadas junto con las frecuencias observadas de la tabla de contingencia original nos permitirán obtener el valor del **estadístico de la prueba ji-cuadrada  $X^2$** . Este estadístico fue propuesto por Karl Pearson en 1900 y se puede interpretar como una medida de distancia entre las frecuencias observadas y las frecuencias esperadas. Como toda distancia, su valor siempre será cero (cuando las frecuencias observadas y esperadas sean idénticas) o positivo.

La fórmula matemática del estadístico es:

$$X^2 = \sum \frac{(\text{frecuencia observada} - \text{frecuencia esperada})^2}{\text{frecuencia esperada}}$$

La suma es de todas las celdas de la tabla.

<sup>14</sup> La prueba ji-cuadrada de Pearson es el procedimiento de inferencia estadística más antiguo que aún se utiliza.

Una vez que se obtiene el estadístico de la prueba  $X^2$ , éste se compara con el valor crítico de  $\chi^2$  con  $(n-1)(p-1)$  grados de libertad y un nivel de significancia establecido ( $\alpha$ ) el cual se interpreta como la probabilidad de rechazar la hipótesis nula de independencia ( $H_0$ ) siendo cierta.

El término “grados de libertad” se refiere al número de celdas cuyas frecuencias pueden variar libremente. Así, si  $t$  es el número total de celdas en una tabla de contingencia ( $t=n \times p$ ) se le debe restar un grado de libertad porque la suma de todas las frecuencias observadas debe ser igual a  $k_{00}$ ; es decir,  $k_1 + k_2 + \dots + k_t = k_{00}$ . Además se debe restar  $(n-1)$  porque es necesario estimar  $(n-1)$  probabilidades de renglón (no es necesario estimar todas las probabilidades porque la última probabilidad de renglón se determina automáticamente ya que la suma de todas las probabilidades de renglón es igual a uno). De igual forma, se resta  $(p-1)$  porque se deben estimar  $(p-1)$  probabilidades de columna.

Por todo lo anterior, los grados de libertad asociados con el estadístico de la prueba son:

gl =  $t$  – un grado de libertad por cada restricción lineal independiente impuesta sobre las frecuencias observadas de las celdas

$$\begin{aligned} \text{gl} &= np - 1 - (n-1) - (p-1) \\ &= np - 1 - n + 1 - p + 1 \\ &= np - n - p + 1 \\ &= (n-1)(p-1) \end{aligned}$$

Finalmente, es importante señalar que la prueba de independencia ji-cuadrada de Pearson se puede utilizar de forma segura con los valores críticos de la distribución ji-cuadrada cuando no más del 20% de las frecuencias esperadas sean menores que 5 y además todas éstas sean mayores o iguales que 1, esto es debido a que frecuencias esperadas mayores dan más precisión a la prueba. Además se sugiere que el tamaño de la muestra sea grande ( $k_{00} \geq 50$ ).

Resumiendo la prueba de independencia ji-cuadrada de Pearson se tiene:

*Supuestos:*

- Las frecuencias observables provienen de una muestra aleatoria de la población de interés.
- Cada observación es clasificada en sólo una categoría por renglón y por otra en columna.

*Hipótesis:*

$H_0$  : las dos clasificaciones son independientes:  $p_{ij}=p_{i0}p_{0j} \quad \forall i,j$

$H_a$  : las dos clasificaciones son dependientes:  $p_{ij}\neq p_{i0}p_{0j} \quad \forall i,j$

*Estadístico de la prueba:*

$$X^2 = \sum_{i=1}^n \sum_{j=1}^p \frac{(k_{ij}-E_{ij})^2}{E_{ij}}$$

donde:

$k_{ij}$  = la frecuencia observada en el renglón  $i$ , columna  $j$

$E_{ij}$  = la frecuencia esperada en el renglón  $i$ , columna  $j = \frac{k_{i0}k_{0j}}{k_{00}}$

*Región de Rechazo:*

$X^2 > \chi^2$ , donde  $\chi^2$  tiene  $(n-1)(p-1)$  grados de libertad. Es decir, la región crítica queda definida por todos los valores mayores a  $\chi^2$ .

### 2.3 RAZÓN DE VEROSIMILITUD JI-CUADRADA

La razón de verosimilitud ji-cuadrada es una alternativa más para realizar la prueba de independencia entre los renglones y columnas de una tabla de contingencia.

*Estadístico de la prueba:*

$$G^2 = 2 \sum_{i=1}^n \sum_{j=1}^p k_{ij} \ln \left[ \frac{k_{ij}}{E_{ij}} \right]$$

donde:

$k_{ij}$  = la frecuencia observada en el renglón  $i$ , columna  $j$

$E_{ij}$  = la frecuencia esperada en el renglón  $i$ , columna  $j = \frac{k_{i0}k_{0j}}{k_{00}}$

*Región de Rechazo:*

$G^2 > \chi^2$ , donde  $\chi^2$  tiene  $(n-1)(p-1)$  grados de libertad. Es decir, la región crítica queda definida por todos los valores mayores a  $\chi^2$ .

### 2.4 JI-CUADRADA DE MANTEL-HAENSZEL

La prueba ji-cuadrada de Mantel-Haenszel se aplica de manera especial en tablas de contingencia de dimensión  $2 \times 2$ , manejando solamente un grado de libertad.

*Estadístico de la prueba:*

$$Q = \frac{(k_{11} - E_{11})^2}{V_{11}}$$

donde:

$k_{11}$  = la frecuencia observada en el renglón  $l$ , columna  $l$

$E_{11}$  = la frecuencia esperada en el renglón  $l$ , columna  $l = \frac{k_{10}k_{01}}{k_{00}}$

$$V_{11} = \frac{k_{10} k_{20} k_{01} k_{02}}{(k_{00})^2 (k_{00}-1)}$$

*Región de Rechazo:*

$Q > \chi^2$ , donde  $\chi^2$  tiene  $l$  grado de libertad. Es decir, la región crítica queda definida por todos los valores mayores a  $\chi^2$ .

En relación a  $X^2$  y  $Q$  es sencillo comprobar que:

$$X^2 = \left[ \frac{n}{(n-1)} \right] Q$$

## **2.5 PRUEBA DE BONDAD DE AJUSTE JI-CUADRADA**

La prueba ji-cuadrada puede ser usada para determinar qué tan bien se ajustan las distribuciones teóricas (normal, binominal, etc.) a distribuciones empíricas (las cuales se obtienen de datos muestrales).

Para poder aplicar la prueba de bondad de ajuste ji-cuadrada, las frecuencias observadas deben aparecer en una tabla de contingencia de una dimensión ( $l \times p$ ) como la siguiente:

Celdas							Total
1	2	3	. . .	j	. . .	p	
$k_1$	$k_2$	$k_3$	. . .	$k_j$	. . .	$k_p$	$k_{00}$

**TABLA 2.3**  
**Tabla de Contingencia 1xp**

*Supuestos:*

- La muestra de  $k_{00}$  observaciones es aleatoria.

*Hipótesis:*

$H_0$  : la probabilidad de que una observación

caiga en la celda  $j$  es igual a  $p^*_j$  :  $p_j = p^*_j \quad \forall j$

$H_a$  : al menos una  $p_j$  es diferente de  $p^*_j$  :  $p_j \neq p^*_j \quad \forall j$

*Estadístico de la prueba:*

$$X^2 = \sum_{i=1}^k \frac{(k_j - k_{00}p^*_j)^2}{k_{00}p^*_j}$$

donde:

$k_j$  = la frecuencia observada en el renglón  $j$

*Región de Rechazo:*

$X^2 > \chi^2$ , donde  $\chi^2$  tiene  $(p-1)$  grados de libertad ( $p$  es el número de celdas). Es decir, la región crítica queda definida por todos los valores mayores a  $\chi^2$ .

Para ejemplificar, suponga que se entrevista a 200 personas y entre otras preguntas se les hace la siguiente: ¿Qué día de la semana es el que acude con mayor frecuencia a realizar sus compras de despensa en una tienda de autoservicio? Obteniéndose la siguiente información:

Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo	Total
8	5	5	12	40	50	80	200

**TABLA 2.4**  
**Tabla de Contingencia 1x7**

La distribución teórica que el investigador esperaba que siguieran los datos es:

Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo	Total
10	10	10	10	30	60	70	200

**TABLA 2.5**  
**Distribución Teórica**

Por lo tanto nuestras hipótesis son:

$$H_0 : \quad p_1= 0.05 \quad p_2= 0.05 \quad p_3= 0.05 \quad p_4= 0.05$$

$$p_5= 0.15 \quad p_6= 0.30 \quad p_7= 0.35$$

$$H_a : \quad \text{al menos una } p_j \text{ es diferente de } p^*_j$$

Calculando el estadístico de prueba:

$$X^2 = \frac{(8-10)^2}{10} + \frac{(5-10)^2}{10} + \frac{(5-10)^2}{10} + \frac{(12-10)^2}{10} + \frac{(40-30)^2}{30} + \frac{(50-60)^2}{60} + \frac{(80-70)^2}{70}$$

$$= 0.4 + 2.5 + 2.5 + 0.4 + 3.3 + 1.7 + 1.4 = 12.2$$

Se compara  $X^2$  con el valor crítico de una  $\chi^2$  con  $(p-1) = (7-1) = 6$  grados de libertad y utilizando  $\alpha=0.05$ , rechazaremos la hipótesis nula si  $X^2 > 12.592$ . Dado que el valor del estadístico de la prueba,  $X^2 = 12.2$ , no es mayor que el valor crítico de  $\chi^2$ , no se puede rechazar la hipótesis nula, llegando a la conclusión de que el ajuste entre la distribución teórica y la distribución empírica es aceptable.

## 2.6 MEDIDAS DE ASOCIACIÓN

Las medidas de asociación se aplican de manera alternativa y permiten cuantificar la fuerza de la asociación entre las dos variables.

En las siguientes secciones se presentarán las fórmulas para calcular algunas de estas medidas de asociación, las cuales se generan también en el paquete estadístico SAS. Vale la pena mencionar que una dificultad común a todas las medidas de asociación propuestas hasta el día de hoy es la dificultad para interpretarlas.

### 2.6.1 COEFICIENTE PHI ( $\varphi$ )

El coeficiente  $\varphi$  se aplica únicamente en tablas de contingencia de  $2 \times 2$ . Cuando existe independencia entre los renglones y las columnas de la tabla de contingencia,  $\varphi$  es cercano a cero.

$$\varphi = \left[ \frac{X^2}{k_{00}} \right]^{1/2}$$

Desafortunadamente, el valor de  $\varphi$  puede ser mayor que uno. Es por ello que Pearson, en 1904, propuso una modificación y estableció lo que ahora se conoce como coeficiente de contingencia  $C$ .

### 2.6.2 COEFICIENTE DE CONTINGENCIA C

Esta medida de asociación toma valores entre 0 y  $C_{m\acute{a}x}$ . El valor maximo de  $C$ ,  $C_{m\acute{a}x}$ , depende del tamano de la tabla de contingencia y ocurre cuando las dos clasificaciones son perfectamente dependientes o asociadas.

$$C = \left[ \frac{X^2}{X^2 + k_{00}} \right]^{1/2}$$

Para una tabla de contingencia con el mismo nmero de renglones y columnas ( $n \times n$ ) el valor mximo de  $C$  es:

$$C_{m\acute{a}x} = \left[ \frac{n-1}{n} \right]^{1/2}$$

Por ejemplo, para una tabla de contingencia ( $3 \times 3$ ),  $C_{m\acute{a}x} = 0.82$ .

Cuando en una tabla de contingencia el nmero de renglones y columnas es diferente ( $n \times p$ ), el valor mximo de  $C$  se calcula con la misma frmula anterior pero utilizando el mnimo entre  $n$  y  $p$ . De este modo, para una tabla de contingencia ( $4 \times 5$ ),  $C_{m\acute{a}x} = 0.87$ .

Esta medida de asociacin tiene la limitante de que  $C < 1$ , lo cual fue resuelto por Cramer en 1946 cuando propuso  $V$ .

### 2.6.3 V DE CRAMER

Es un valor del nivel de asociacin existente entre variables nominales, que es correcto para tablas de contingencia de cualquier tamano. La  $V$  de Cramer puede tomar cualquier valor entre 0 y 1. Un valor de  $V$  cercano a 1 indica un nivel de asociacin elevado. En otras

palabras, si  $V=1$  entonces existe una correlación perfecta y si  $V=0$ , las variables son independientes.

$$V = \left[ \frac{X^2}{k_{00}[\min(n-1, p-1)]} \right]^{1/2}$$

## 2.7 APLICACIÓN

Para ejemplificar todo lo expuesto en este capítulo, suponga que se tiene la siguiente tabla de contingencia<sup>15</sup> que se obtuvo después de realizar una encuesta a 403 personas y en donde se les pidió a los encuestados asociar cada una de las marcas de productos con una característica en especial. Los renglones corresponden a cuatro características del producto y las columnas representan a cinco diferentes marcas de productos.

		Marca					Total
		A	B	C	D	E	
Característica	1	8	11	5	7	12	43
	2	5	15	3	1	6	30
	3	105	32	11	23	37	208
	4	32	57	6	2	25	122
Total		150	115	25	33	80	403

**TABLA 2.7**  
**Tabla de Contingencia**

Si lo que se quiere es probar la hipótesis nula de que las dos clasificaciones son independientes, se iniciará calculando las frecuencias esperadas. De modo que, para las dos primeras celdas de la tabla de contingencia se tendría:

para la celda  $k_{11}$  se tiene:  $\frac{43 \times 150}{403} = 16.00$

<sup>15</sup> Los datos de la tabla de contingencia 4x5 se extrajeron de JOBSON, J. (1992). Applied Multivariate Data Analysis. Volume II. USA: Springer. p. 434

para la celda  $k_{12}$  se tiene:  $\frac{43 \times 115}{403} = 12.27$

Realizando estas operaciones para todas las celdas de la tabla de contingencia, la tabla de frecuencias esperadas sería la siguiente:

		Marca					Total
		A	B	C	D	E	
Característica	1	16.00	12.27	2.67	3.52	8.54	43
	2	11.17	8.56	1.86	2.46	5.96	30
	3	77.42	59.35	12.90	17.03	41.29	208
	4	45.41	34.81	7.57	9.99	24.22	122
Total		150	115	25	33	80	403

**TABLA 2.8**  
**Frecuencias Esperadas**

Utilizando las frecuencias esperadas y observadas de las celdas de la tabla de contingencia, procedemos a calcular el valor del estadístico de la prueba ( $X^2$ ).

$$\begin{aligned}
 X^2 = & \frac{(8-16)^2}{16} + \frac{(11-12.27)^2}{12.27} + \frac{(5-2.67)^2}{2.67} + \frac{(7-3.52)^2}{3.52} + \frac{(12-8.54)^2}{8.54} + \\
 & \frac{(5-11.17)^2}{11.17} + \frac{(15-8.56)^2}{8.56} + \frac{(3-1.86)^2}{1.86} + \frac{(1-2.46)^2}{2.46} + \frac{(6-5.96)^2}{5.96} + \\
 & \frac{(105-77.42)^2}{77.42} + \frac{(32-59.35)^2}{59.35} + \frac{(11-12.90)^2}{12.90} + \frac{(23-17.03)^2}{17.03} + \frac{(37-41.29)^2}{41.29} + \\
 & \frac{(32-45.41)^2}{45.41} + \frac{(57-34.81)^2}{34.81} + \frac{(6-7.57)^2}{7.57} + \frac{(2-9.99)^2}{9.99} + \frac{(25-24.22)^2}{24.22} \\
 = & 70.917
 \end{aligned}$$

Ahora, se compara  $X^2$  con el valor crítico de una  $\chi^2$  con  $(n-1)(p-1) = (4-1)(5-1) = (3)(4) = 12$  grados de libertad. Si utilizamos  $\alpha=0.05$ , rechazaremos la hipótesis nula de que las clasificaciones son independientes si  $X^2 > 21.0261$ . Dado que el valor del estadístico de la prueba,  $X^2 = 70.917$ , es mayor que el valor crítico de  $\chi^2$ , rechazamos la hipótesis nula de independencia entre las dos clasificaciones a un nivel de significación de  $\alpha=0.05$ .

Aplicando de manera alternativa el estadístico  $G^2$  de la razón de verosimilitud ji-cuadrada obtenemos:  $G^2 = 73.342$ . Al igual que con el estadístico  $X^2$ , con  $\alpha=0.05$  y 12 grados de libertad, se rechaza nuevamente la hipótesis nula de independencia dado que  $G^2 > \chi^2$ .

Finalmente, para medir la fuerza de asociación (relación o dependencia) entre las dos variables, calculamos el coeficiente de contingencia  $C$  y la  $V$  de Cramer.

$$C = \left[ \frac{70.917}{70.917+403} \right]^{1/2} = 0.387$$

$$V = \left[ \frac{70.917}{403(3)} \right]^{1/2} = 0.242$$

De acuerdo a los valores obtenidos para  $C$  y  $V$ , se puede concluir que existe una considerable fuerza de asociación entre las dos variables de clasificación.

## **2.8 ANÁLISIS DE TABLAS DE CONTINGENCIA EN SAS**

El paquete estadístico SAS cuenta con procedimientos que permiten desarrollar el análisis de tablas de contingencia. El programa en dicho paquete para analizar la tabla de contingencia presentada como ejemplo en el subcapítulo anterior se encuentra en el Anexo A y la salida de resultados se muestra a continuación:

The FREQ Procedure

Table of car by tienda

car	tienda					Total
	A	B	C	D	E	
CAR1	8	11	5	7	12	43
	16.005	12.27	2.6675	3.5211	8.536	
	1.99	2.73	1.24	1.74	2.98	10.67
	18.60	25.58	11.63	16.28	27.91	
	5.33	9.57	20.00	21.21	15.00	
CAR2	5	15	3	1	6	30
	11.166	8.5608	1.861	2.4566	5.9553	
	1.24	3.72	0.74	0.25	1.49	7.44
	16.67	50.00	10.00	3.33	20.00	
	3.33	13.04	12.00	3.03	7.50	
CAR3	105	32	11	23	37	208
	77.419	59.355	12.903	17.032	41.29	
	26.05	7.94	2.73	5.71	9.18	51.61
	50.48	15.38	5.29	11.06	17.79	
	70.00	27.83	44.00	69.70	46.25	
CAR4	32	57	6	2	25	122
	45.409	34.814	7.5682	9.9901	24.218	
	7.94	14.14	1.49	0.50	6.20	30.27
	26.23	46.72	4.92	1.64	20.49	
	21.33	49.57	24.00	6.06	31.25	
Total	150	115	25	33	80	403
	37.22	28.54	6.20	8.19	19.85	100.00

The FREQ Procedure

Statistics for Table of car by tienda

Statistic	DF	Value	Prob
Chi-Square	12	70.9167	<.0001
Likelihood Ratio Chi-Square	12	73.3422	<.0001
Mantel-Haenszel Chi-Square	1	4.9951	0.0254
Phi Coefficient		0.4195	
Contingency Coefficient		0.3868	
Cramer's V		0.2422	

Sample Size = 403

En esta salida de resultados en SAS, la primera parte muestra la tabla de contingencia con los datos crudos (frecuencias observadas), frecuencias esperadas y porcentajes (totales, por renglón y por columna). En la segunda parte, aparecen los valores de los estadísticos (value):  $X^2$  (Chi-Square),  $G^2$  (Likelihood Ratio Chi-Square) y  $Q$  (Mantel-Haenszel Chi-Square), señalando los grados de libertad (DF) y los valores del nivel de significancia de la prueba (Prob). De igual forma, se muestran los valores de las medidas de asociación:  $\phi$  (Phi Coefficient),  $C$  (Contingency Coefficient) y  $V$  (Cramer's V), así como el tamaño de la muestra  $k_{00}$  (Sample Size).

Comparando los resultados de la corrida en SAS con los calculados en el subcapítulo anterior, se puede observar que los valores obtenidos son similares y por lo tanto se les da la misma interpretación. Aunque obviamente, a través del uso de SAS los resultados se obtienen de una manera más rápida y sencilla pues solamente se necesita realizar un pequeño código de programa.

## CAPÍTULO 3 – ANÁLISIS DE CORRESPONDENCIAS

### 3.1 INTRODUCCIÓN

El análisis de correspondencias (AC) es una técnica exploratoria parecida al análisis factorial y al de componentes principales que permite ver de una manera clara y sencilla, a través de un *mapa perceptual*, la asociación existente entre dos variables categóricas de una tabla de contingencia. Dentro de la investigación de mercados, específicamente en los estudios de imagen y posicionamiento, esta técnica se utiliza para visualizar a través de un mapa perceptual de qué forma se encuentran asociadas una serie de atributos o características con ciertas marcas, productos o servicios.

Los orígenes del análisis de correspondencias se remontan a la década de los sesenta con los trabajos del franco libanés Jean-Paul Benzécri. El análisis de correspondencias se desarrolla como una técnica estadística aplicada de naturaleza euclidiana que parte del análisis de tablas de contingencia, cuyo objeto es obtener una representación geométrica de los datos que ayude a visualizar y localizar las relaciones entre las variables de grandes matrices de datos. Este nuevo enfoque del estudio de tablas de contingencia ha derivado posteriormente en nuevas formulaciones estadísticas para el análisis de datos multivariados. De este modo es plausible utilizarlo en cualquier estructura (arreglo tabular o matricial) que contenga valores positivos<sup>16</sup>.

El AC es útil en trabajos exploratorios donde no existen hipótesis previas sobre el comportamiento de la población, tanto en las vertientes correlacionales como en las experimentales. Permite extraer relaciones de dependencia a partir de variables categóricas expresadas en tablas de contingencia. Por otra parte, es posible analizar la estructura de esta asociación, en forma de proximidades que nos ayudan a identificar las "causas" de la

---

<sup>16</sup> HERRERO, Francisco y CUESTA (Junio,2003). Introducción al Análisis de Correspondencias: Uso en Análisis Multidimensionales. [http://www.uniovi.es/~Psi/Dpto\\_Psicologia/metodos/hardcopy/98\\_1.html](http://www.uniovi.es/~Psi/Dpto_Psicologia/metodos/hardcopy/98_1.html)

asociación medida, acentuando la representación gráfica como método óptimo de obtención de conclusiones. Es decir, como una herramienta básica de información de tipo cualitativo.

### 3.2 OBJETIVOS DEL ANÁLISIS DE CORRESPONDENCIAS

El objetivo principal del análisis de correspondencias es básicamente descriptivo y pretende poner de manifiesto la existencia estructurada de relaciones de semejanza o diferencia entre las variables de una tabla de contingencia a través de una mapa perceptual, permitiendo establecer escenarios prospectivos.

Los objetivos del análisis de correspondencias responden a la necesidad de profundizar en las relaciones de dependencia que se establecen entre dos o más variables cualitativas observadas en una misma población y en particular insistiendo en explicar cómo las distintas categorías o modalidades de ambas variables se relacionan unas con otras. Es decir, representar geométricamente la estructura de relación entre los datos. Puede considerarse como una radiografía que revela estructuras presentes en la tabla de datos y que están ocultas a nuestros sentidos<sup>17</sup>.

### 3.3 MATRIZ DE CORRESPONDENCIAS Y MASAS DE RENGLÓN Y COLUMNA

Como ya se dijo, el análisis de correspondencias debe asumirse como una técnica estadística que puede ser usada para analizar la asociación entre dos variables categóricas de una tabla de contingencia. En el capítulo anterior se presentó la notación para una tabla de contingencia de  $n$  renglones y  $p$  columnas a través de la tabla 2.1, la cual ahora definimos como  $K$  en donde  $k_{ij}$  representa a las celdas de frecuencias absolutas para  $i=1,2,\dots,n$  y  $j=1,2,\dots,p$ . Ahora, para realizar el análisis de correspondencias, primeramente se debe convertir la tabla de contingencia  $K$  en una tabla de frecuencias relativas  $F$  llamada

---

<sup>17</sup> CORNEJO, José (1988). Técnicas de Investigación Social: El Análisis de Correspondencias. Barcelona: PPU. p. 97

matriz de probabilidades o matriz de correspondencias, la cual también es una tabla de doble entrada compuesta de  $n$  renglones y  $p$  columnas y en donde el valor de cada celda  $f_{ij}$  se obtiene de la división entre los elementos  $k_{ij}$  y  $k_{00}$  de la tabla de contingencia  $\mathbf{K}$ , como se muestra en la tabla 3.1.

		Columna						Masa de Renglón
		1	2	...	j	...	p	
I	J							
	Renglón	1	$f_{11}$	$f_{12}$	...	$f_{1j}$	...	$f_{1p}$
2		$f_{21}$	$f_{22}$	...	$f_{2j}$	...	$f_{2p}$	$f_{20}$
...		...	...		...		...	...
...		...	...		...		...	...
...		...	...		...		...	...
i		$f_{i1}$	$f_{i2}$	...	$f_{ij}$	...	$f_{ip}$	$f_{i0}$
...		...	...		...		...	...
n	$f_{n1}$	$f_{n2}$	...	$f_{nj}$	...	$f_{np}$	$f_{n0}$	
Masa de Columna		$f_{01}$	$f_{02}$	...	$f_{0j}$	...	$f_{0p}$	$f_{00}=1$

**TABLA 3.1**  
**Matriz de Correspondencias F**

En donde las celdas son probabilidades dadas por:

$$f_{ij} = \frac{k_{ij}}{k_{00}}$$

$$f_{i0} = \frac{k_{i0}}{k_{00}}$$

$$f_{0j} = \frac{k_{0j}}{k_{00}}$$

Siendo:

$$\sum_{i=1}^n f_{i0} = \sum_{j=1}^p f_{0j} = \sum_{i=1}^n \sum_{j=1}^p f_{ij} = 1$$

El vector  $(n \times 1)$  de los totales marginales de renglón  $f_{i0}$ , para  $i=1,2,\dots,n$ , es denotado por  $\mathbf{n}$  y de manera similar, el vector  $(p \times 1)$  de los totales marginales de columna  $f_{0j}$ , para  $j=1,2,\dots,p$ , es denotado por  $\mathbf{p}$ . Estos dos vectores son conocidos como *masas de renglón  $\mathbf{n}$*  y *masas de columna  $\mathbf{p}$*  respectivamente.

La masa de renglón y columna debe entenderse como la importancia relativa de los renglones y columnas dentro de la tabla de datos, es decir, sirve para atenuar la preponderancia que podría tener algún renglón o columna en el análisis.

A lo largo de este capítulo se utilizará la tabla de contingencia presentada en el capítulo anterior (Tabla 2.7) para ejemplificar el análisis de correspondencias. La matriz de correspondencias  $F$  para esta tabla sería:

		Marca					Masa de Renglón
		A	B	C	D	E	
Característica	1	0.020	0.027	0.012	0.017	0.030	0.107
	2	0.012	0.037	0.007	0.002	0.015	0.074
	3	0.261	0.079	0.027	0.057	0.092	0.516
	4	0.079	0.141	0.015	0.005	0.062	0.303
Masa de Columna		0.372	0.285	0.062	0.082	0.199	1

**TABLA 3.3**  
**Matriz de Correspondencias F**

En esta tabla podemos ver que para las masas de renglón, la que corresponde a la tercer característica (renglón) del producto es la que se considera de mayor importancia dado su valor (0.516), seguida en orden descendente por las masas de las características (renglones) 4,1,2. De manera similar se interpretan las masas de columna (marca), es decir, la marca (columna) A es la que se considera de mayor importancia seguida en orden descendente por las marcas B, E, D y C.

### 3.4 MATRIZ DE PERFILES

Partiendo de la tabla de frecuencias absolutas  $K$  (Tabla 2.1), es posible construir la *matriz de perfiles renglón*  $N$  dividiendo cada celda de la tabla entre el total correspondiente al renglón en el que se encuentra, como se puede ver a continuación:

I \ J	Columna						Masa de Renglón
	1	2	. . .	j	. . .	p	
1	$k_{11}/k_{10}$	$k_{12}/k_{10}$	. . .	$k_{1j}/k_{10}$	. . .	$k_{1p}/k_{10}$	$k_{10}/k_{10}=1$
2	$k_{21}/k_{20}$	$k_{22}/k_{20}$	. . .	$k_{2j}/k_{20}$	. . .	$k_{2p}/k_{20}$	$k_{20}/k_{20}=1$
⋮	⋮	⋮		⋮		⋮	⋮
⋮	⋮	⋮		⋮		⋮	⋮
Renglón i	$k_{i1}/k_{i0}$	$k_{i2}/k_{i0}$	. . .	$k_{ij}/k_{i0}$	. . .	$k_{ip}/k_{i0}$	$k_{i0}/k_{i0}=1$
⋮	⋮	⋮		⋮		⋮	⋮
⋮	⋮	⋮		⋮		⋮	⋮
n	$k_{n1}/k_{n0}$	$k_{n2}/k_{n0}$	. . .	$k_{nj}/k_{n0}$	. . .	$k_{np}/k_{n0}$	$k_{n0}/k_{n0}=1$
Masa de Columna	$k_{01}/k_{00}$	$k_{02}/k_{00}$	. . .	$k_{0j}/k_{00}$	. . .	$k_{0p}/k_{00}$	$k_{00}/k_{00}=1$

**TABLA 3.4**  
**Matriz de Perfiles Renglón N**

De manera que para definir un perfil de renglón se establece la siguiente notación:

$$f^i_j = \{ f^i_j, j \in J \} \quad \text{con} \quad f^i_j = \begin{cases} \frac{f_{ij}}{f_{i0}} & \text{si } f_{i0} \neq 0, \\ \frac{k_{ij}}{k_{i0}} & \text{si } k_{i0} \neq 0. \end{cases}$$

Si se construye una matriz diagonal  $D_n$  en la cual los elementos de la diagonal son los elementos del vector  $n$  (masa de renglón), en nomenclatura matricial la matriz de perfiles renglón se obtiene a través de  $N = D^{-1} F$ .

De manera similar, para construir la *matriz de perfiles columna*  $P$  es necesario dividir cada celda de la tabla entre el total correspondiente a la columna en la que se encuentra, como se puede ver a continuación:

I \ J		Columna						Masa de Renglón
		1	2	...	j	...	p	
Renglón	1	$k_{11}/k_{01}$	$k_{12}/k_{02}$	...	$k_{1j}/k_{0j}$	...	$k_{1p}/k_{0p}$	$k_{10}/k_{00}$
	2	$k_{21}/k_{01}$	$k_{22}/k_{02}$	...	$k_{2j}/k_{0j}$	...	$k_{2p}/k_{0p}$	$k_{20}/k_{00}$
	...	...	...	...	...	...	...	...
	i	$k_{i1}/k_{01}$	$k_{i2}/k_{02}$	...	$k_{ij}/k_{0j}$	...	$k_{ip}/k_{0p}$	$k_{i0}/k_{00}$
	...	...	...	...	...	...	...	...
	n	$k_{n1}/k_{01}$	$k_{n2}/k_{02}$	...	$k_{nj}/k_{0j}$	...	$k_{np}/k_{0p}$	$k_{n0}/k_{00}$
Masa de Columna		$k_{01}/k_{01}$	$k_{02}/k_{02}$	...	$k_{0j}/k_{0j}$	...	$k_{0p}/k_{0p}$	$k_{00}/k_{00}=1$

**TABLA 3.5**  
**Matriz de Perfiles Columna P**

El perfil de columna se denota como:

$$f^j_i = \{ f^j_i, i \in I \} \quad \text{con} \quad f^j_i = \begin{cases} \frac{f_{ij}}{f_{0j}} & \text{si } f_{0j} \neq 0, \\ \frac{k_{ij}}{k_{0j}} & \text{si } k_{0j} \neq 0. \end{cases}$$

Si de igual forma, se construye una matriz diagonal  $D_p$  en la cual los elementos de la diagonal son los elementos del vector  $p$  (masa de columna), en nomenclatura matricial la matriz de perfiles columna se obtiene a través de  $P=(D^{-1}F)'$ . Además, puede observarse que para las matrices  $D_n$  y  $D_p$ , la suma de sus elementos de la diagonal es igual a 1.

Estadísticamente, cada perfil es una distribución condicionada, de manera que  $f^i_j$  es la distribución de  $J$  condicionada a  $i$ , así como  $f^j_i$  es la distribución de  $I$  condicionada a  $j$ . De tal forma que:

$$\sum f^i_j = 1$$

$$\sum f^j_i = 1$$

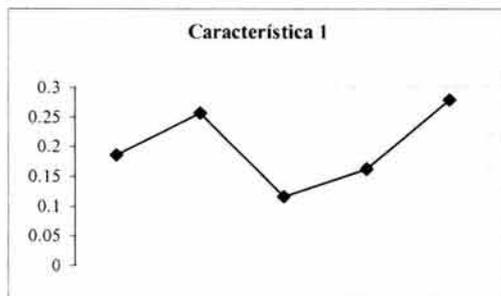
El perfil renglón o perfil columna identifica cada modalidad en cuanto a su importancia relativa.

Por ejemplo, considerando nuevamente la tabla de contingencia 2.7, la matriz de perfiles renglón  $N$  sería:

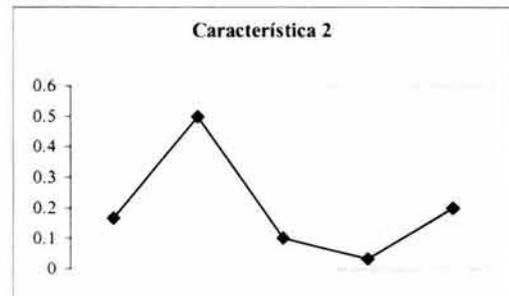
		Marca					Total
		A	B	C	D	E	
Característica	1	0.186	0.256	0.116	0.163	0.279	1
	2	0.167	0.500	0.100	0.033	0.200	1
	3	0.505	0.154	0.053	0.111	0.178	1
	4	0.262	0.467	0.049	0.016	0.205	1
Masa de Columna		0.372	0.285	0.062	0.082	0.199	1

**TABLA 3.6**  
**Matriz de Perfiles Renglón N**

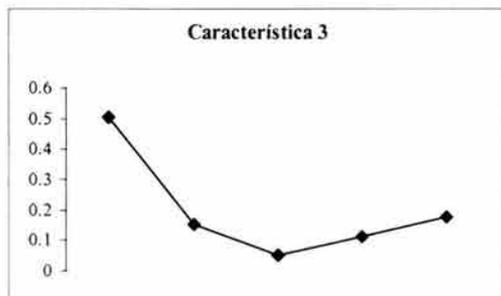
Con el fin de facilitar la visualización de los perfiles de cada renglón se presentan las siguientes gráficas:



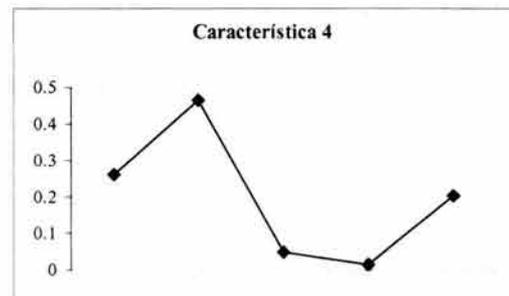
**GRÁFICA 3.1**  
**Perfiles de Renglón 1**



**GRÁFICA 3.2**  
**Perfiles de Renglón 2**



**GRÁFICA 3.3**  
**Perfiles de Renglón 3**



**GRÁFICA 3.4**  
**Perfiles de Renglón 4**

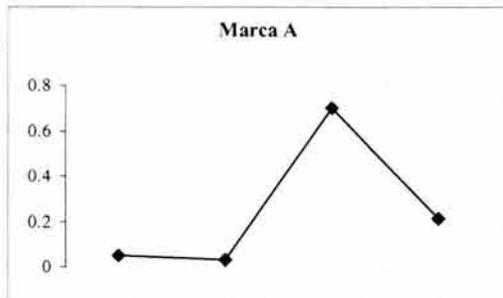
Analizando la matriz de perfiles renglón  $N$  y las gráficas anteriores, se puede observar que las características (renglones) 2 y 4 son bastante similares a diferencia de las características (renglones) 1 y 3 que son muy diferentes entre sí.

La matriz de perfiles columna  $P$  queda como sigue:

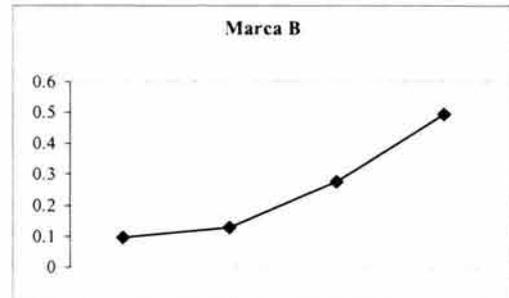
		Marca					Masa de Renglón
		A	B	C	D	E	
Característica	1	0.053	0.096	0.200	0.212	0.150	0.107
	2	0.033	0.130	0.120	0.030	0.075	0.074
	3	0.700	0.278	0.440	0.697	0.463	0.516
	4	0.213	0.496	0.240	0.061	0.313	0.303
Total		1	1	1	1	1	1

**TABLA 3.7**  
**Matriz de Perfiles Columna P**

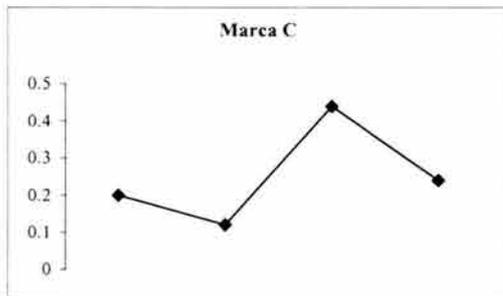
Las gráficas de los perfiles columna quedarían de la siguiente forma:



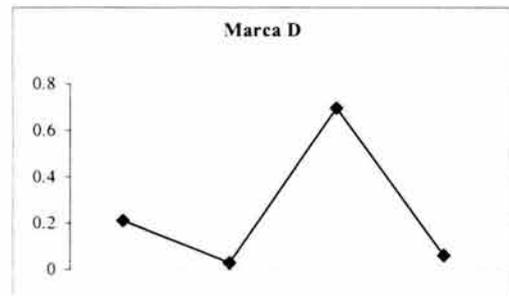
**GRÁFICA 3.5**  
**Perfiles de Columna 1**



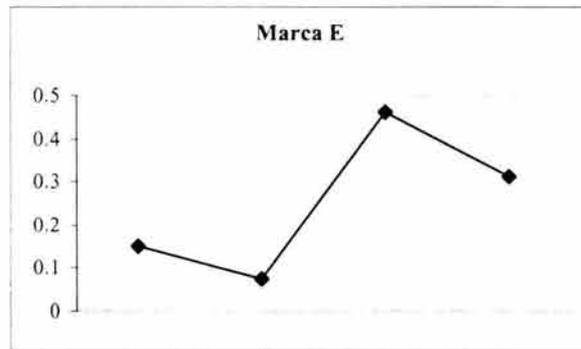
**GRÁFICA 3.6**  
**Perfiles de Columna 2**



**GRÁFICA 3.7**  
**Perfiles de Columna 3**



**GRÁFICA 3.8**  
**Perfiles de Columna 4**



**GRÁFICA 3.9**  
**Perfiles de Columna 5**

Analizando la matriz de perfiles columna  $P$  y sus respectivas gráficas, se puede observar que existe una gran similitud entre los perfiles de las marcas (columnas) A y D, así como entre los perfiles de las marcas (columnas) C y E. Por el contrario, los perfiles de la marca (columna) B son muy diferentes a los perfiles de las otras cuatro marcas (columnas).

Las matrices diagonales  $D_n$  y  $D_p$  quedarían de la siguiente forma:

$$D_n = \begin{bmatrix} 0.107 & 0 & 0 & 0 \\ 0 & 0.074 & 0 & 0 \\ 0 & 0 & 0.516 & 0 \\ 0 & 0 & 0 & 0.303 \end{bmatrix}$$

$$D_p = \begin{bmatrix} 0.372 & 0 & 0 & 0 & 0 \\ 0 & 0.285 & 0 & 0 & 0 \\ 0 & 0 & 0.062 & 0 & 0 \\ 0 & 0 & 0 & 0.082 & 0 \\ 0 & 0 & 0 & 0 & 0.199 \end{bmatrix}$$

### 3.5 IMPORTANCIA DE LOS PERFILES

Para la matriz de correspondencias  $F$ , cualquier elemento  $i$  de  $I$  está representado geoméricamente por un vector en  $R^p$ . El  $j$ -ésimo componente de este vector tiene los valores  $f^i_j$ , los cuales son los valores del perfil de  $i$  sobre  $J$ . A la inversa, cualquier elemento  $j$  de  $J$  está representado geoméricamente por un vector en  $R^n$ . El  $i$ -ésimo componente de este vector tiene los valores  $f^j_i$ , los valores del perfil de  $j$  en  $I$ . El modelo básico del análisis de correspondencias involucra cálculos basados sobre los perfiles de  $I$  y  $J$ . De este modo, los dos elementos  $i$  y  $i'$  de  $I$  son considerados semejantes si y sólo si sus perfiles asociados  $f^i_J$  y  $f^{i'}_J$  son semejantes, es decir, si el  $i$ -ésimo renglón y el  $i'$ -ésimo renglón de la tabla de contingencia son proporcionales. De manera análoga, los dos elementos  $j$  y  $j'$  de  $J$  son considerados semejantes si y sólo si sus perfiles asociados  $f^j_I$  y  $f^{j'}_I$  son semejantes, es decir, si la  $j$ -ésima columna y la  $j'$ -ésima columna de la tabla de contingencia son proporcionales<sup>18</sup>.

### 3.6 ESPACIO DE LOS PERFILES

El conjunto  $I$  está representado en el espacio de perfiles sobre  $J$  y el conjunto  $J$  está representado en el espacio de perfiles sobre  $I$  esto es,

$\{f^i_J, i \in I\}$  = conjunto de perfiles de  $i \in I$ .

$\{f^j_I, j \in J\}$  = conjunto de perfiles de  $j \in J$ .

#### 3.6.1 ESPACIO DE LOS PERFILES SOBRE J

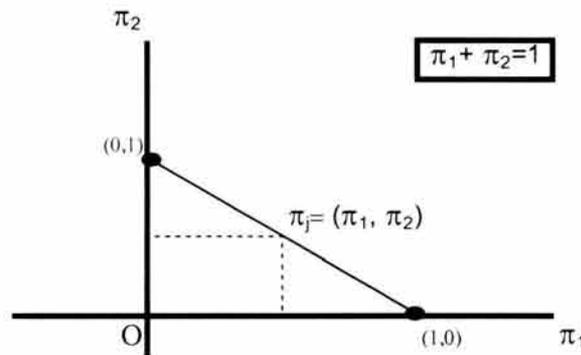
Un punto de este espacio es un perfil sobre  $J$ , es decir un conjunto de ceros o números positivos suscritos en  $J$  y de los cuales el total es 1. Es decir,

<sup>18</sup> JAMBU, Michel (1991). Exploratory and Multivariate Data Analysis. USA: Academic Press. pp. 175-176

$\pi_j = \{\pi_j \mid j \in J\}$  en donde  $\pi_j$  es cero o un número real positivo ( $\pi_j \geq 0$ ) y  
 $\sum \{\pi_j \mid j \in J\} = 1$

Se dice que el espacio de perfiles sobre  $J$  es un espacio de  $\text{card}(J)-1$  dimensiones (número de elementos de  $J$  menos uno). Por ejemplo, si  $J$  es el conjunto de 10 categorías o modalidades, el espacio de perfiles sobre  $J$  es de dimensión 9.

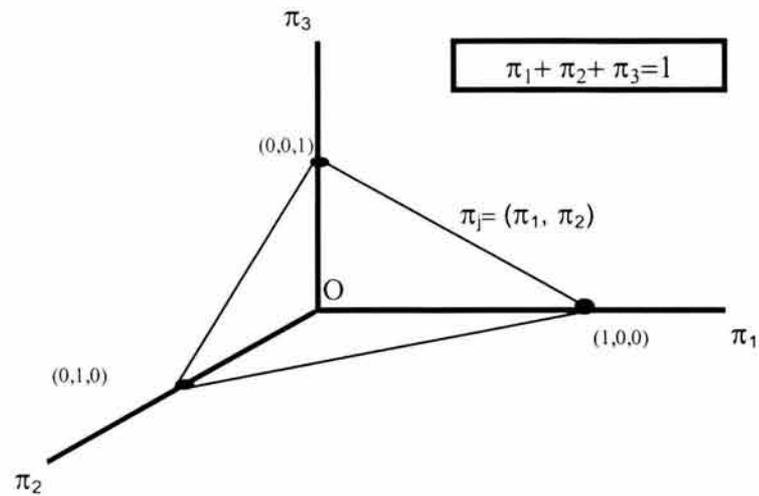
Cuando la dimensión de  $J=2$ :  $J=\{1,2\}$  ( $J$  está constituido de dos elementos llamados 1 y 2). En el espacio de dimensión 2 (el cual corresponde a un plano) se trazan dos ejes llamados  $O\pi_1$  y  $O\pi_2$ :



GRÁFICA 3.10

Resulta evidente que las condiciones  $\pi_1 \geq 0$  y  $\pi_2 \geq 0$  significan que el punto  $\pi_j$  se encontrará en el primer cuadrante. Además la condición de  $\pi_1 + \pi_2 = 1$  implica que el punto  $\pi_j$  estará localizado dentro del área que se forma a partir de la recta que une los puntos  $(1,0)$  y  $(0,1)$ .

Para el caso donde  $J=3$ :  $J=\{1,2,3\}$  ( $J$  está constituido de tres elementos llamados 1, 2 y 3) la gráfica sería:



GRÁFICA 3.11

### 3.6.2 ESPACIO DE LOS PERFILES SOBRE I

Un perfil sobre  $I$  es un conjunto de ceros o números reales positivos suscritos en él:

$$\pi_i = \{\pi_i \mid i \in I\} \quad \text{en donde } \pi_i \text{ es cero o un número real positivo } (\pi_i \geq 0) \text{ y}$$

$$\sum \{\pi_i \mid i \in I\} = 1$$

El espacio de perfiles sobre  $I$  es de dimensión  $\text{card}(I) - 1$ , es decir, el número de elementos de  $I$  menos uno. Por ejemplo, si  $I$  es un conjunto de 20 categorías o modalidades, el espacio de perfiles sobre  $I$  es de dimensión 19.

### 3.7 NUBE DE PUNTOS

La nube de puntos o diagrama de dispersión como también se le conoce, es una representación gráfica de un conjunto de puntos en el plano. Según la forma de la nube de puntos se puede obtener la siguiente información<sup>19</sup>:

- Conocer si existe una relación directa o inversa entre las variables.
- Saber si esa relación es fuerte o débil.
- Determinar si la relación se ajusta a un modelo lineal o bien a otro modelo matemático (por ejemplo a un modelo curvilíneo).

En el análisis de correspondencias, la construcción del mapa perceptual se logra a través de la representación gráfica de las *nubes de puntos*  $N(I)$  y  $N(J)$ .

La *nube de puntos*  $N(I)$  es el conjunto de elementos de  $i \in I$ , cuyas coordenadas son los componentes del perfil  $f^i_j = \{f^i_j, j \in J\}$  y cuya masa es  $f_{i0}$ .

La *nube de puntos*  $N(J)$  es el conjunto de elementos de  $j \in J$ , cuyas coordenadas son los componentes del perfil  $f^j_i = \{f^j_i, i \in I\}$  y cuya masa es  $f_{0j}$ .

Por lo tanto,

$$N(I) = \{(f^i_j, f_{i0}) \mid i \in I\} \subset \mathbb{R}^p$$

$$N(J) = \{(f^j_i, f_{0j}) \mid j \in J\} \subset \mathbb{R}^n$$

<sup>19</sup> HERRERO, Francisco y CUESTA (Junio,2003). Relación Bivariada.  
<http://www.uniovi.es/UniOvi/Apartados/Departamento/Psicologia/metodos/tutor.6>

### 3.8 DISTANCIA JI-CUADRADA

Para calcular la distancia entre dos perfiles, Benzécri (1973) propuso utilizar la distancia ji-cuadrada en lugar de la ya conocida distancia euclídea. El nombre de distancia ji-cuadrada se debe a la relación que guarda con la prueba de independencia ji-cuadrada para tablas de contingencia. Esta distancia difiere de la métrica euclídea usual en la ponderación de cada cuadrado por los inversos de las frecuencias correspondientes a cada término.

Así, la distancia ji-cuadrada ponderada entre dos perfiles renglón  $(i, i')$  cualesquiera en el espacio  $R^p$  está dada por:

$$\begin{aligned} d^2(i, i') &= \sum_{j=1}^p \frac{(f_{ij}^i - f_{ij}^{i'})^2}{f_{0j}} = \sum_{j=1}^p \frac{1}{f_{0j}} \left[ \frac{f_{ij}}{f_{i0}} - \frac{f_{i'j}}{f_{i'0}} \right]^2 \\ &= \sum_{j=1}^p \left[ \frac{f_{ij}}{f_{i0}\sqrt{f_{0j}}} - \frac{f_{i'j}}{f_{i'0}\sqrt{f_{0j}}} \right]^2 \end{aligned}$$

De igual manera, la distancia ji-cuadrada ponderada entre dos perfiles de columna  $(j, j')$  cualesquiera en el espacio  $R^n$  está dada por:

$$\begin{aligned} d^2(j, j') &= \sum_{i=1}^n \frac{(f_{ij}^j - f_{ij}^{j'})^2}{f_{i0}} = \sum_{i=1}^n \frac{1}{f_{i0}} \left[ \frac{f_{ij}}{f_{0j}} - \frac{f_{ij'}}{f_{0j'}} \right]^2 \\ &= \sum_{i=1}^n \left[ \frac{f_{ij}}{f_{0j}\sqrt{f_{i0}}} - \frac{f_{ij'}}{f_{0j'}\sqrt{f_{i0}}} \right]^2 \end{aligned}$$

Esta distancia ji-cuadrada, a diferencia de la distancia euclídea habitual, posee la *propiedad de equivalencia distribucional*. Esto implica que si por ejemplo, dos renglones  $i$  y  $i'$  tienen idénticos perfiles (distribucionalmente equivalentes), estos renglones pueden ser

reemplazados en la tabla de datos crudos por un sólo renglón que se obtenga de sumar ambos renglones. Esta unión de los dos renglones no afectará la geometría (distancia) de los perfiles columna. De manera similar, un renglón puede ser subdividido en dos o más renglones en la tabla de datos crudos sin afectar la geometría (distancia) de las columnas. Si dos perfiles de renglón son idénticos, éstos ocuparán posiciones idénticas en el espacio de renglón. Esta propiedad garantiza una cierta *invarianza*<sup>20</sup> en los resultados con respecto al número de categorías en que se subdividen las variables. De igual forma se garantiza la *robustez*, pues no se pierde información agrupando categorías homogéneas ni se gana descomponiendo una categoría en subcategorías homogéneas.

De acuerdo con la distancia ji-cuadrada, las modalidades perfil renglón están representadas por una configuración de  $n$  puntos en un espacio euclídeo  $R^p$  de coordenadas:

$$P_i = \left[ \frac{f_{i1}}{f_{i0}\sqrt{f_{01}}}, \dots, \frac{f_{ij}}{f_{i0}\sqrt{f_{0j}}}, \dots, \frac{f_{ip}}{f_{i0}\sqrt{f_{0p}}} \right]$$

separados por la distancia euclídea ordinaria.

Para aclarar, suponga que se pretende realizar un análisis de correspondencias con la siguiente tabla de contingencia  $K$ :

		Variable 2			Total
		A	B	C	
Variable 1	1	40	70	35	145
	2	20	30	40	90
	3	80	120	160	360
Total		140	220	235	595

**TABLA 3.8**  
**Tabla de Contingencia K**

<sup>20</sup> Invarianza: Una propiedad de un conjunto de variables o un estadístico que permanece sin cambios después de realizar alguna transformación u operación. Por ejemplo, la varianza de un conjunto de observaciones es invariante bajo transformaciones lineales de los datos. EVERITT, B. (1999). The Cambridge Dictionary of Statistics. United Kingdom: Cambridge University Press. p. 168

La matriz de correspondencias  $F$  sería la siguiente:

		Variable 2			Masa de Renglón
		A	B	C	
Variable 1	1	0.067	0.118	0.059	0.244
	2	0.034	0.050	0.067	0.151
	3	0.134	0.202	0.269	0.605
Masa de Columna		0.235	0.370	0.395	1

**TABLA 3.9**  
**Matriz de Correspondencias F**

Las matrices de perfiles renglón  $N$  y columna  $P$  quedarían de la siguiente forma:

		Variable 2			Total
		A	B	C	
Variable 1	1	0.276	0.483	0.241	1
	2	0.222	0.333	0.444	1
	3	0.222	0.333	0.444	1
Masa de Columna		0.235	0.370	0.395	1

**TABLA 3.10**  
**Matriz de Perfiles Renglón N**

		Variable 2			Masa de Renglón
		A	B	C	
Variable 1	1	0.286	0.318	0.149	0.244
	2	0.143	0.136	0.170	0.151
	3	0.571	0.545	0.681	0.605
Total		1	1	1	1

**TABLA 3.11**  
**Matriz de Perfiles Columna P**

Las distancias ji-cuadrada entre columnas están dadas por:

$$d^2(A,B) = \left[ \frac{0.067}{0.235\sqrt{0.244}} - \frac{0.118}{0.370\sqrt{0.244}} \right]^2 + \left[ \frac{0.034}{0.235\sqrt{0.151}} - \frac{0.050}{0.370\sqrt{0.151}} \right]^2 + \left[ \frac{0.134}{0.235\sqrt{0.605}} - \frac{0.202}{0.370\sqrt{0.605}} \right]^2 = 0.006$$

$$d^2(A,C) = \left[ \frac{0.067}{0.235\sqrt{0.244}} - \frac{0.059}{0.395\sqrt{0.244}} \right]^2 + \left[ \frac{0.034}{0.235\sqrt{0.151}} - \frac{0.067}{0.395\sqrt{0.151}} \right]^2 + \left[ \frac{0.134}{0.235\sqrt{0.605}} - \frac{0.269}{0.395\sqrt{0.605}} \right]^2 = 0.100$$

$$d^2(B,C) = \left[ \frac{0.118}{0.370\sqrt{0.244}} - \frac{0.059}{0.395\sqrt{0.244}} \right]^2 + \left[ \frac{0.050}{0.370\sqrt{0.151}} - \frac{0.067}{0.395\sqrt{0.151}} \right]^2 + \left[ \frac{0.202}{0.370\sqrt{0.605}} - \frac{0.269}{0.395\sqrt{0.605}} \right]^2 = 0.156$$

Claramente se puede observar que en la matriz de perfiles renglón  $N$ , los renglones 2 y 3 de la variable 1 son idénticos. Por lo tanto se procede a realizar la unión de ambos renglones para construir la nueva tabla de contingencia  $K$ :

		Variable 2			Total
		A	B	C	
Variable 1	1	40	70	35	145
	2	100	150	200	450
Total		140	220	235	595

**TABLA 3.12**  
**Tabla de Contingencia K**

La tabla de correspondencias  $F$  y las matrices de perfiles  $N$  y  $P$  serán:

		Variable 2			Masa de Renglón
		A	B	C	
Variable 1	1	0.067	0.118	0.059	0.244
	2	0.168	0.252	0.336	0.756
Masa de Columna		0.235	0.370	0.395	1

**TABLA 3.13**  
**Matriz de Correspondencias F**

		Variable 2			Total
		A	B	C	
Variable 1	1	0.276	0.483	0.241	1
	2	0.222	0.333	0.444	1
Masa de Columna		0.235	0.370	0.395	1

**TABLA 3.14**  
**Matriz de Perfiles Renglón N**

		Variable 2			Masa de Renglón
		A	B	C	
Variable 1	1	0.286	0.318	0.149	0.244
	2	0.714	0.682	0.851	0.756
Total		1	1	1	1

**TABLA 3.15**  
**Matriz de Perfiles Columna P**

Finalmente, se puede verificar que con la unión de los renglones 2 y 3 no se modificaron las distancias ji-cuadrada entre las columnas, cumpliéndose así la propiedad de equivalencia distribucional. Es decir,

$$d^2(A,B) = \left[ \frac{0.067}{0.235\sqrt{0.244}} - \frac{0.118}{0.370\sqrt{0.244}} \right]^2 + \left[ \frac{0.168}{0.235\sqrt{0.756}} - \frac{0.252}{0.370\sqrt{0.756}} \right]^2 = 0.006$$

$$d^2(A,C) = \left[ \frac{0.067}{0.235\sqrt{0.244}} - \frac{0.059}{0.395\sqrt{0.244}} \right]^2 + \left[ \frac{0.168}{0.235\sqrt{0.756}} - \frac{0.336}{0.395\sqrt{0.756}} \right]^2 = 0.100$$

$$d^2(B,C) = \left[ \frac{0.118}{0.370\sqrt{0.244}} - \frac{0.059}{0.395\sqrt{0.244}} \right]^2 + \left[ \frac{0.252}{0.370\sqrt{0.756}} - \frac{0.336}{0.395\sqrt{0.756}} \right]^2 = 0.156$$

### 3.9 INERCIA

La inercia se define como una medida de dispersión total de la nube de puntos respecto a su baricentro o centro de gravedad y el cual se interpreta como un perfil medio. De esta forma, las inercias o varianzas de las nubes de puntos  $N(I)$  y  $N(J)$  son:

$$\begin{aligned} \text{Inercia } [N(I)] &= \sum \{ f_{i0} \cdot d^2(i, G_i) \quad , i \in I \} \\ &= \sum \{ (f_{ij} - f_{i0}f_{0j})^2 \frac{1}{f_{i0}f_{0j}} , i \in I, j \in J \} = \frac{X^2}{k_{00}} = \phi^2 \end{aligned}$$

$$\begin{aligned} \text{Inercia } [N(J)] &= \sum \{ f_{0j} \cdot d^2(j, G_j) \quad , j \in J \} \\ &= \sum \{ (f_{ij} - f_{i0}f_{0j})^2 \frac{1}{f_{i0}f_{0j}} , i \in I, j \in J \} = \frac{X^2}{k_{00}} = \phi^2 \end{aligned}$$

De lo anterior se puede observar que la inercia o varianza ( $\phi^2$ ) es la misma para ambas nubes de puntos. El valor más bajo que puede tomar este estadístico es cero el cual indica no asociación entre los renglones y columnas de la tabla de contingencia, y el valor más

alto sería  $card(I)-1$ , es decir el número de renglones de la tabla menos uno, el cual indicaría una perfecta asociación de cada renglón con cada columna. Lleva este nombre por la fórmula de física que dice que  $masa * distancia^2 = inercia$ , en donde la *masa* está representada por las  $f_{i0}$  y  $f_{0j}$ , y  $distancia^2$  por  $d^2(i, G_i)$  y  $d^2(j, G_j)$ .

Antes de continuar, ejemplifiquemos lo anterior considerando la tabla de contingencia de 4x5 que cruza los datos de 4 características de producto con 5 marcas diferentes (Tabla 2.7). Dado que  $X^2 = 70.917$ , entonces:

$$\phi^2 = 70.917 / 403 = 0.17597$$

Este valor nos indica que existe una débil asociación entre los renglones y columnas de la tabla de contingencia.

### 3.10 COORDENADAS FACTORIALES

A continuación se presenta a manera de resumen la notación matricial que se va a utilizar para encontrar los ejes principales de inercia, es decir, aquéllas direcciones que absorben el máximo de inercia de cada nube de puntos<sup>21</sup>.

$F = [f_{ij}]$  es la matriz de correspondencias de dimensión  $n \times p$ .

$D_n = [f_{i0}]$  es la matriz diagonal de las masas de renglón de dimensión  $n \times n$ .

$D_p = [f_{0j}]$  es la matriz diagonal de las masas de columna de dimensión  $p \times p$ .

$D_n^{-1} F = [f^i_j]$  es la matriz diagonal de los perfiles de renglón de dimensión  $n \times p$ .

$D_p^{-1} F' = [f^j_i]$  es la matriz diagonal de los perfiles de columna de dimensión  $p \times n$ .

<sup>21</sup> JOARISTI, L. y LIZASOAIN (2000). Análisis de Correspondencias. Madrid: La Muralla, S.A. pp. 41-44

En el espacio  $R^p$  en que se sitúan los renglones:

$D^{-1}_n F$  son, como se acaba de indicar, las coordenadas  $f^i_j$  de los renglones y  $D^{-1}_p$  es la métrica bajo la cual:

$$d^2(i,i') = \sum_{j=1}^p \frac{(f^i_j - f^{i'}_j)^2}{f_{0j}}$$

El criterio a optimizar en  $R^p$ , establecido con relación al origen  $O$ , es:

$$\begin{aligned} \text{Max } \{ \sum f_{i0} \cdot d^2(i,O) \quad , i \in I \} &= \text{Max } \{ D_n [(D^{-1}_n F) D^{-1}_p u]^2 \} = \\ \text{Max } \{ u' D^{-1}_p F' D^{-1}_n F D^{-1}_p u \} \end{aligned}$$

con la restricción de normalización  $u' D^{-1}_p u = 1$ .

Así,  $u$  es un vector propio ortonormal<sup>22</sup> de la matriz  $F' D^{-1}_n F D^{-1}_p$ , de valor propio  $\lambda$ :

$$[F' D^{-1}_n F D^{-1}_p] u = \lambda u$$

Los puntos de renglón tienen en cada eje factorial, las siguientes coordenadas:

$$\begin{aligned} \psi &= (D^{-1}_n F) D^{-1}_p u \\ \psi(i) &= \sum \{ (f^i_j / f_{0j}) u_j \quad , j \in J \} = \sum \{ (f_{ij} / f_{i0} f_{0j}) u_j \quad , j \in J \} \end{aligned}$$

$$\begin{aligned} \text{con} \quad E(\psi) &= \sum \{ f_{i0} \psi(i) \quad , i \in I \} = 0 \\ \text{Var}(\psi) &= \sum \{ f_{i0} \psi^2(i) \quad , i \in I \} = \lambda \end{aligned}$$

<sup>22</sup> Vectores ortonormales: Dos o más vectores mutuamente ortogonales, cada uno de longitud 1. Una serie de pares de vectores ortonormales se llama conjunto ortonormal.

Vectores ortogonales: Se dice que dos vectores  $u$  y  $v$  de un espacio vectorial son ortogonales si su producto interno es cero; es decir,  $\langle u, v \rangle = 0$ . MORRIS, Christopher (1996). Diccionario Enciclopédico de Ciencia y Tecnología. Tomo III. México: Prentice-Hall Hispanoamericana. pp. 1709-1710

En el espacio  $R^n$  en que se sitúan las columnas:

$D_p^{-1} F'$  son las coordenadas  $f_i^j$  de las columnas y  $D_n^{-1}$  es la métrica bajo la cual:

$$d^2(j, j') = \sum_{i=1}^n \frac{(f_i^j - f_i^{j'})^2}{f_{i0}}$$

El criterio a optimizar en  $R^n$ , establecido con relación al origen  $O$ , es:

$$\begin{aligned} \text{Max } \{ \sum f_{0j} \cdot d^2(j, O) \quad , j \in J \} &= \text{Max } \{ D_p [(D_p^{-1} F') D_n^{-1} v]^2 \} = \\ \text{Max } \{ v' D_n^{-1} F D_p^{-1} F' D_n^{-1} v \} \end{aligned}$$

con la restricción de normalización  $v' D_n^{-1} v = 1$ .

Cada  $v$  es un vector propio ortonormal de la matriz  $F D_p^{-1} F' D_n^{-1}$ , con el mismo valor propio asociado  $\lambda$ :

$$[F D_p^{-1} F' D_n^{-1}] v = \lambda v$$

Además,  $u$  y  $v$  son los vectores unitarios de los distintos ejes factoriales.

Los puntos de columna tienen en cada eje factorial, las siguientes coordenadas:

$$\begin{aligned} \varphi &= (D_p^{-1} F') D_n^{-1} v \\ \varphi(j) &= \sum \{ (f_i^j / f_{i0}) v_i \quad , i \in I \} = \sum \{ (f_{ij} / f_{i0} f_{0j}) v_i \quad , i \in I \} \end{aligned}$$

$$\begin{aligned} \text{con} \quad E(\varphi) &= \sum \{ f_{0j} \varphi(j) \quad , j \in J \} = 0 \\ \text{Var}(\varphi) &= \sum \{ f_{0j} \varphi^2(j) \quad , j \in J \} = \lambda \end{aligned}$$

En extrapolación de la física a la estadística, la varianza se interpreta como la inercia de los distintos puntos respecto del origen de coordenadas, que no es más que una suma ponderada de las distancias de los puntos a otro, que en este caso es el origen. El que  $\psi$  y  $\varphi$  tengan la misma varianza implica que están medidas en la misma escala.

Es sencillo comprobar que:

$$\sum \{ \lambda_{\alpha}, \alpha=1, 2, \dots, h \} = \mathbf{X}^2 / k_{00}$$

donde  $h$  es el mínimo entre  $n-1$  y  $p-1$ , prescindiendo del primer valor propio, cuyo valor es 1 y representa la solución trivial. Los valores propios son inferiores a 1.

### 3.11 RELACIONES DE TRANSICIÓN

Partiendo de las expresiones de las coordenadas factoriales:

$$\begin{aligned} \psi &= (D_n^{-1} F) D_p^{-1} u \\ \varphi &= (D_p^{-1} F') D_n^{-1} v \end{aligned}$$

se puede llegar, para cada factor, a estas dos expresiones, llamadas relaciones de transición:

$$\begin{aligned} \psi(i) &= 1/\sqrt{\lambda} \{ f_j^i \cdot \varphi(j) \quad , j \in J \} , \forall i \in I \\ \varphi(j) &= 1/\sqrt{\lambda} \{ f_i^j \cdot \psi(i) \quad , i \in I \} , \forall j \in J \end{aligned}$$

Es decir, tales fórmulas permiten representar los puntos de un espacio en función de las coordenadas de los puntos del otro, siendo la consecuencia inmediata que se puede proceder a una representación simultánea de las dos nubes de puntos<sup>23</sup>.

<sup>23</sup> Ibidem, JOARISTI, L. y LIZASOAIN. pp. 44-45

### 3.12 MAPA PERCEPTUAL

A partir de la obtención de las coordenadas factoriales, es posible realizar una representación gráfica de los puntos en un mapa de 2 ó 3 dimensiones (siempre y cuando recojan información suficiente) llamada comúnmente dentro del campo de la investigación de mercados como *mapa perceptual*. En éste, la cercanía entre puntos indica la semejanza entre los objetos que representan, mientras que su alejamiento señala que uno y otro tienen poco o nada que ver entre sí.

El mapa perceptual resulta ser una herramienta poderosa y valiosa para la toma de decisiones en cuestión de posicionamiento porque en él se puede<sup>24</sup>:

- Identificar las fuerzas y debilidades de los productos bajo estudio para que posteriormente se encuentren aquellos puntos que se puedan reforzar, desarrollar o aprovechar al máximo para su comercialización.
- Entender la estructura competitiva del mercado desde la perspectiva del consumidor, lo que permite saber los límites de la categoría a la que pertenece la marca de interés, y ayuda a detectar oportunidades para atender nuevos mercados o necesidades que no han sido aprovechadas por las marcas existentes.
- Evaluar la efectividad del posicionamiento de una marca a través de determinar cuáles ventajas y desventajas percibe el consumidor, lo que da pie a una revisión del programa de mercadotecnia para posicionarla o reposicionarla.
- Determinar la posibilidad de éxito o fracaso para extender la línea del producto y estimar si la posición de una nueva marca se puede lograr.

---

<sup>24</sup> (Junio,2003) Una Imagen vale más que mil palabras.  
<http://148.205.64.181/microseg/Administrador/Uploader/material/Una%20Imagen%20Vale%20Mas%20que%20Mil%20Palabras.PDF>

- Identificar las diferencias entre segmentos de mercado al comparar los mapas perceptuales de distintos grupos de consumidores.

Retomando los datos de la tabla de contingencia (4x5) de características vs. marcas con la que se ha venido trabajando (Tabla 2.7), se procede a realizar los cálculos pertinentes para la obtención de las coordenadas factoriales. Los valores propios quedan como sigue:

	Valor	Porcentaje
$\lambda_1$	0.14191	80.65%
$\lambda_2$	0.03286	18.67%
$\lambda_3$	0.00120	0.68%
$\phi^2$	0.17597	100.00%

**TABLA 3.16**  
**Valores Propios y**  
**Procentaje de Inercia Explicado**

De este modo, se puede decir que la información de la tabla de contingencia se explica por la primera dimensión el 80.65%, mientras que en la segunda la explicación es del 18.67% y en la tercera del 0.68%.

Las coordenadas por renglón necesarias para construir el mapa perceptual para las dos primeras dimensiones son:

Característica		Dimensión	
		1	2
	1	0.04	0.50
	2	0.55	0.09
	3	-0.35	-0.05
	4	0.44	-0.11

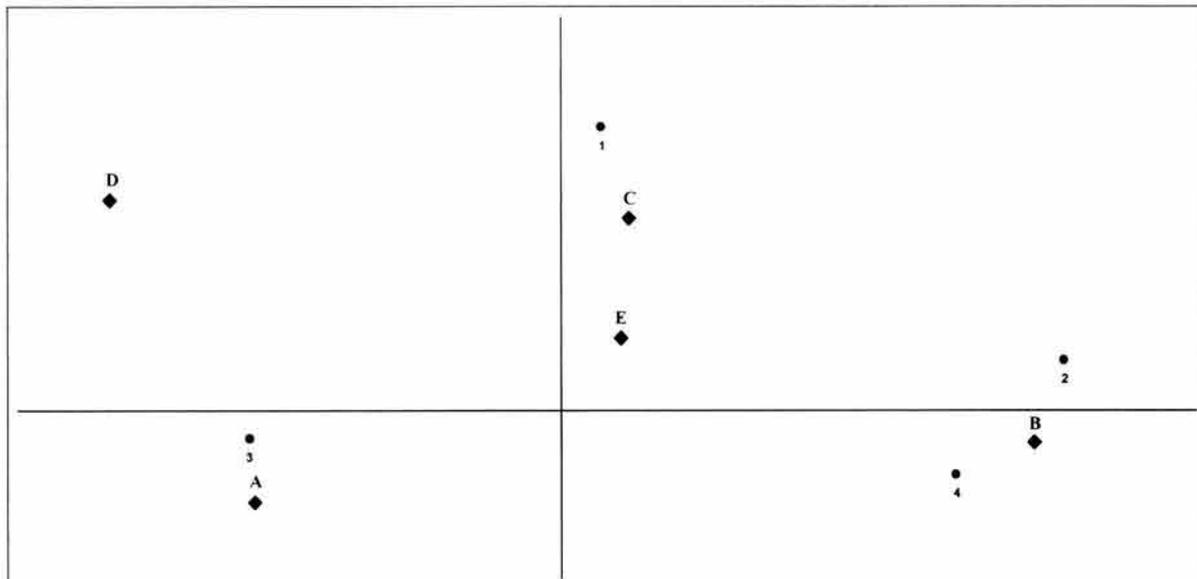
**TABLA 3.17**  
**Coordenadas por Renglón**

Y las coordenadas por columna quedarían como:

		Dimensión	
		1	2
Marca	A	-0.34	-0.16
	B	0.52	-0.06
	C	0.08	0.34
	D	-0.50	0.37
	E	0.07	0.13

**TABLA 3.18**  
**Coordenadas por Columna**

Una vez obtenidas las coordenadas, se construye la siguiente gráfica en Power Point con la representación simultánea:



**GRÁFICA 3.12**  
**Mapa Perceptual**

A través del mapa perceptual es posible visualizar la asociación entre los renglones y columnas de la tabla de contingencia. Los cuatro renglones o características del producto se especifican con los números 1, 2, 3 y 4. Así mismo, las cinco columnas o marcas de producto están simbolizadas con las letras A, B, C, D y E. De esta forma se puede observar que la marca B se asocia con las características 2 y 4, la marca A se asocia con la característica 3 y la marca C se asocia con la característica 1. Por el contrario, las marcas D y E no se muestran asociadas a ninguna característica.

### 3.13 DESCOMPOSICIÓN ADITIVA DE LA JI-CUADRADA

Para encontrar la dimensión significativa, se puede utilizar el método sugerido por Williams (1952), que consiste en descomponer la ji-cuadrada total en suma de ji-cuadradas correspondientes a los valores propios.

La descomposición es la siguiente:

Valor Propio	Ji-Cuadrada	Grados de Libertad
$\lambda_1$	$X^2_2 = k_{00}\lambda_1$	$n + p - 3$
$\lambda_2$	$X^2_3 = k_{00}\lambda_2$	$n + p - 5$
...	...	...
$\lambda_h$	$X^2_h = k_{00}\lambda_h$	$n + p - (2p - 1) = n - p + 1$
	$X^2$	$(n-1)(p-1)$

**TABLA 3.19**  
**Descomposición Aditiva de la Ji-Cuadrada**

El eje factorial correspondiente a  $\lambda_h$  no es significativo si  $X^2_h = k_{00}\lambda_h$  con  $(n - p + 1)$  grados de libertad no es significativa, es decir si  $X^2_h < \chi^2$ .

Aunque no hay normas precisas para decidir cuántos ejes escoger en el análisis de correspondencias, una regla empírica simple revela que, una vez ordenados los valores propios, en forma descendente, se toman los primeros valores propios, antes de cambiar bruscamente hacia abajo, el porcentaje de variación explicada. Otros investigadores escogen todos los valores propios mayores a la relación:  $[1 / (\text{inf}(n,p) - 1)]$ , valor que conllevaría cada uno de los ejes si la repartición de la inercia fuera igual para todos los ejes. En este punto, es importante mencionar que uno de los objetivos del análisis de correspondencias es la simplificación. Por lo general, en la práctica se escogen los dos primeros valores propios (máximo, los tres primeros), aunque el porcentaje de variación sea pequeño, pues lo que interesa es simplificar la información<sup>25</sup>.

Si utilizamos los valores propios obtenidos anteriormente para realizar la descomposición aditiva ji-cuadrada del ejemplo que se ha venido desarrollando (Tabla 2.7) se tendría:

Valor Propio	Ji-Cuadrada	Grados de Libertad	Ji-Cuadrada Crítico	
$\lambda_1 = 0.14191$	$X^2_1 = k_{00}\lambda_1$ = (403) (0.14191) = 57.18973	$n+p-3 = 4+5-3 = 6$	$X^2_{0.05} = 12.5916$	$X^2_1 > X^2_{0.05}$ (significativa)
$\lambda_2 = 0.03286$	$X^2_2 = k_{00}\lambda_2$ = (403) (0.03286) = 13.24258	$n+p-5 = 4+5-5 = 4$	$X^2_{0.05} = 9.48773$	$X^2_2 > X^2_{0.05}$ (significativa)
$\lambda_3 = 0.00120$	$X^2_3 = k_{00}\lambda_3$ = (403) (0.00120) = 0.4836	$n+p-7 = 4+ 5-7 = 2$	$X^2_{0.05} = 5.99147$	$X^2_3 < X^2_{0.05}$ (no significativa)
	$X^2 = 70.91591$	$(n-1) (p-1) = (3) (4)$ = 12	$X^2_{0.05} = 21.0261$	$X^2 > X^2_{0.05}$ (significativa)

**TABLA 3.20**  
**Descomposición Aditiva de la Ji-Cuadrada**

<sup>25</sup>GARNICA, Elsy (Enero,2003). Análisis Multivariante sobre la vivienda. <http://iies.faces.ula.ve/revista/articulos/revi10garnicao.pdf>

De los resultados presentados en la tabla anterior, se puede observar que las dos primeras dimensiones son significativas y con las cuales se tendría explicada el 99.32% de la información inicial, por lo tanto se hace plausible reducir la dimensionalidad a dos únicos componentes.

### 3.14 INTERPRETACIÓN DE LOS PLANOS DE PROYECCIÓN

Los valores que toman las contribuciones absolutas y los cosenos cuadrados son decisivos al definir o interpretar los ejes o planos de proyección.

#### 3.14.1 CONTRIBUCIONES ABSOLUTAS

Para una interpretación correcta de los planos de proyección hay que tener en cuenta su contribución absoluta, esto es su peso en la determinación de cada eje.

La contribución absoluta se obtiene multiplicando la coordenada del eje al cuadrado por la masa. Para el caso de la contribución absoluta del renglón al eje quedaría:

$$CTA(i) = (f_{i0} \psi^2) / \lambda \quad (\sum CTA (i) = 1)$$

De manera similar, la contribución absoluta de la columna al eje se obtiene por medio de la siguiente formula:

$$CTA(j) = (f_{j0} \varphi^2) / \lambda \quad (\sum CTA (j) = 1)$$

Los puntos  $i \in I$  para los cuales los valores de las contribuciones son superiores al promedio son considerados como puntos explicativos. El conjunto de puntos explicativos de  $I$  puede entonces ser seleccionado como se expone a continuación:

$$\{ i \in I ; (CTA(i) = (f_{i0} \psi^2) / \lambda \geq 1/n) \}$$

donde  $n$  es el número de elementos de  $I$  (número de renglones).

Lo mismo se puede expresar para  $J$ . El conjunto de puntos explicativos de  $J$  puede entonces ser seleccionado de acuerdo a:

$$\{ j \in J ; (CTA(j) = (f_{0j} \varphi^2) / \lambda \geq 1/p) \}$$

donde  $p$  es el número de elementos de  $J$  (número de columnas).

De aquí se puede concluir que las categorías que tienen una fuerte contribución absoluta en el eje son las que determinarán su posición y por lo tanto, sirven para interpretar su significado.

Los cálculos para obtener las contribuciones absolutas por renglón para la primer dimensión de nuestro ejemplo quedan como sigue:

$$CTA(i=1) = \left[ \frac{(0.107)(0.04)^2}{0.14191} \right] = 0.00$$

$$CTA(i=2) = \left[ \frac{(0.074)(0.55)^2}{0.14191} \right] = 0.16$$

$$CTA(i=3) = \left[ \frac{(0.516)(0.35)^2}{0.14191} \right] = 0.43$$

$$CTA(i=4) = \left[ \frac{(0.303)(0.44)^2}{0.14191} \right] = 0.41$$

Las contribuciones absolutas por renglón para la primera y segunda dimensión se presentan a continuación:

		Dimensión	
		1	2
Característica	1	0.00	<b>0.82</b>
	2	0.16	0.02
	3	<b>0.43</b>	0.04
	4	<b>0.41</b>	0.12
		1.00	1.00

**TABLA 3.21**  
**Contribuciones Absolutas**  
**por Renglón**

Analizando la tabla anterior y localizando los puntos que son mayores o iguales a  $1/4=0.25$ , se puede observar que las características (renglones) 3 y 4 contribuyen básicamente a la formación de la primera dimensión, mientras que la segunda dimensión está determinada por la característica (renglón) 1.

Las contribuciones absolutas por columna son:

		Dimensión	
		1	2
Marca	A	<b>0.30</b>	0.30
	B	<b>0.55</b>	0.03
	C	0.00	<b>0.22</b>
	D	0.14	<b>0.35</b>
	E	0.01	0.10
		1.00	1.00

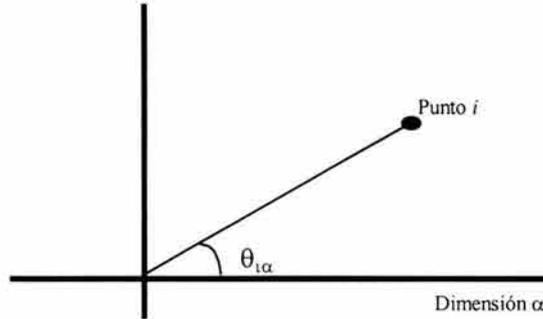
**TABLA 3.22**  
**Contribuciones Absolutas**  
**por Columna**

Localizando en esta tabla los puntos que son mayores a  $1/5=0.20$ , se puede observar que las marcas (columnas) A y B contribuyen básicamente a la formación de la primera dimensión, mientras que la segunda dimensión está determinada por las marcas (columnas) C y D.

### 3.14.2 COSENOS CUADRADOS

Una herramienta que sirve como complemento a la hora de realizar la interpretación de los planos de proyección son los cosenos cuadrados.

Cada punto renglón unido con el centro de gravedad (origen), forma un ángulo con su proyección en el espacio de los primeros ejes, cuyo coseno cuadrado coincide con la calidad de representación del i-ésimo elemento en el plano (espacio) sobre el que se está proyectando.



GRÁFICA 3.13

Si se denota por  $\theta_{i\alpha}$  al ángulo entre las dos líneas se tiene:

$$\cos^2 \theta_{i\alpha} = \psi^2_{i\alpha} / \sum_{\alpha=1}^h \psi^2_{i\alpha}$$

Los cosenos cuadrados permiten describir la calidad de la representación de cada una de las categorías sobre el eje en consideración. Cuanto más grande sea el valor de este indicador ( $\approx 1$ ), mejor será la calidad de la representación del punto (categoría) respectivo.

Los cosenos cuadrados por renglón para la tabla de contingencia (Tabla 2.7) con la que se ha trabajado son:

		Dimensión	
		1	2
Característica	1	0.01	<b>0.99</b>
	2	<b>0.94</b>	0.02
	3	<b>0.98</b>	0.02
	4	<b>0.93</b>	0.06

**TABLA 3.23**  
**Cosenos Cuadrados**  
**por Renglón**

En esta tabla se puede ver que los valores de los cosenos cuadrados que más se destacan en la formación del primer eje son los de las características (renglones) 2, 3 y 4; mientras que la característica (renglón) 1 destaca en la formación del segundo eje.

Los cosenos cuadrados por columna son:

		Dimensión	
		1	2
Marca	A	<b>0.81</b>	0.19
	B	<b>0.99</b>	0.01
	C	0.04	<b>0.87</b>
	D	<b>0.64</b>	0.36
	E	0.19	<b>0.71</b>

**TABLA 3.24**  
**Cosenos Cuadrados**  
**por Columna**

Como se puede apreciar, los valores de los cosenos cuadrados que más se destacan en la formación del primer eje son los de las marcas (columnas) A, B y D; mientras que las marcas (columnas) C y E destacan en la formación del segundo eje.

### 3.15 ANÁLISIS DE CORRESPONDENCIAS EN SAS

El código del programa que se diseñó en este trabajo para realizar el análisis de correspondencias para la tabla de contingencia presentada como ejemplo a lo largo de este capítulo (Tabla 2.7) se puede encontrar en el Anexo B.

La salida con los resultados obtenidos sería la siguiente:

```

The CORRESP Procedure

      Inertia and Chi-Square Decomposition

Singular   Principal   Chi-      Cumulative
Value      Inertia    Square   Percent   Percent
-----+-----+-----+-----+-----+
0.37672    0.14191    57.1917  80.65     80.65
0.18127    0.03286    13.2414  18.67     99.32
0.03464    0.00120    0.4835   0.68     100.00

      Total      0.17597    70.9167  100.00

Degrees of Freedom = 12

      Row Coordinates

                Dim1      Dim2
C1              0.0443     0.5035
C2              0.5541     0.0892
C3             -0.3447    -0.0503
C4              0.4359    -0.1137

      Summary Statistics for the Row Points

                Quality      Mass      Inertia
C1              0.9970     0.1067     0.1554
C2              0.9632     0.0744     0.1383
C3              0.9998     0.5161     0.3561
C4              0.9966     0.3027     0.3503
    
```

Partial Contributions to Inertia for the Row Points

	Dim1	Dim2
C1	0.0015	0.8232
C2	0.1610	0.0180
C3	0.4322	0.0397
C4	0.4053	0.1191

Indices of the Coordinates that Contribute Most to Inertia for the Row Points

	Dim1	Dim2	Best
C1	0	2	2
C2	0	0	1

The CORRESP Procedure

Indices of the Coordinates that Contribute Most to Inertia for the Row Points

	Dim1	Dim2	Best
C3	1	0	1
C4	1	0	1

Squared Cosines for the Row Points

	Dim1	Dim2
C1	0.0077	0.9893
C2	0.9389	0.0243
C3	0.9790	0.0208
C4	0.9331	0.0635

Column Coordinates

	Dim1	Dim2
A	-0.3384	-0.1634
B	0.5219	-0.0582
C	0.0751	0.3420
D	-0.4982	0.3728
E	0.0663	0.1293

Summary Statistics for the Column Points

	Quality	Mass	Inertia
A	0.9995	0.3722	0.2989
B	0.9999	0.2854	0.4473
C	0.9171	0.0620	0.0471
D	0.9995	0.0819	0.1802
E	0.8997	0.1985	0.0265

Partial Contributions to Inertia for the Column Points

	Dim1	Dim2
A	0.3004	0.3024
B	0.5478	0.0294
C	0.0025	0.2209
D	0.1432	0.3464
E	0.0061	0.1010

The CORRESP Procedure

Indices of the Coordinates that Contribute Most to Inertia for the Column Points

	Dim1	Dim2	Best
A	2	2	2
B	1	0	1
C	0	2	2
D	0	2	2
E	0	0	2

Squared Cosines for the Column Points

	Dim1	Dim2
A	0.8106	0.1889
B	0.9877	0.0123
C	0.0421	0.8750
D	0.6407	0.3588
E	0.1873	0.7124

Como se puede apreciar, los resultados obtenidos en esta salida de SAS son prácticamente idénticos a los mostrados a lo largo de este capítulo. Cabe resaltar que gracias a la existencia de paquetes estadísticos como el SAS, es posible realizar el análisis de correspondencias de una manera más rápida y sencilla ya que de realizarse de manera manual nos encontraríamos ante serios problemas en el desarrollo de cálculos tediosos y engorrosos. Una desventaja que a menudo se encuentra tanto en SAS como en otros paquetes estadísticos, es la construcción del mapa perceptual ya que en la mayoría de las ocasiones se cuenta con un gran número de categorías o modalidades que tienen una puntuación muy cercana y resulta imposible asignar a cada punto su “etiqueta o nombre” completo.

## CAPÍTULO 4 – ESTUDIO DE CASO: IMAGEN Y POSICIONAMIENTO DE GOLOSINAS

### 4.1 IMAGEN Y POSICIONAMIENTO

Dentro del entorno tan fuertemente competitivo de las empresas, cobra especial importancia el concepto de *posicionamiento*, ya sea de marca, productos, servicios, empresas, países, organismos, etc. Pero el posicionamiento no se refiere al producto en sí, sino al lugar que ocupa en la mente del consumidor de acuerdo a sus percepciones y dentro de un marco de referencia delimitado por los productos de la competencia. En la actualidad, el posicionamiento se ha convertido en la pieza clave para que una empresa logre permanecer en el mercado.

Así pues, el posicionamiento se refiere al modo cómo un conjunto de consumidores, usuarios, suscriptores o compradores, perciben y registran en su mente determinado producto o marca, dentro de un marco de referencia delimitado por los productos competidores. De esta manera, también se establecen las preferencias de dicho grupo de consumidores entre las diversas marcas que conforman una categoría de productos.

La *percepción* que se forma la gente acerca de las marcas comienza por la vinculación al recuerdo de éstas a través de sus múltiples experiencias, creencias, impresiones, sensaciones, actitudes y de la información recibida, ya sea directamente de las propias marcas o de parte de terceras personas. A través de estos vínculos es como se forma la *imagen* de la marca. De este modo, una marca que sea fuertemente vinculada con atributos o características determinantes y relevantes para los consumidores, tendrá de una buena imagen la cual le permitirá gozar de importantes ventajas competitivas, aumentando las posibilidades de compra y la lealtad de sus clientes<sup>26</sup>.

---

<sup>26</sup> HERNÁNDEZ, José y SAIZ (1996). Marketing y Tecnología. España: Ediciones Pirámide. pp. 81-86

Dado que la imagen de una marca no es algo permanente, las empresas se ven en la necesidad de recurrir de manera periódica a las investigaciones de mercado de imagen y posicionamiento con el fin de poder establecer acciones oportunas cuando se observen variaciones poco favorables de la imagen correspondiente.

#### **4.2 OBJETIVOS DE LA INVESTIGACIÓN DE MERCADO**

El mercado de golosinas es muy variado en lo que a productos de consumo masivo se refiere integrándolo desde gomas de mascar, chocolates, dulces, pastillas refrescantes, botanas, etc.

Adicionalmente, considerando que México es un país donde casi la mitad de la población tiene menos de 20 años y que las golosinas son productos que se consumen desde la infancia, es decir, que tiene su principal nicho de mercado en los niños y jóvenes, se confirma la importancia de medición y monitoreo del canasto.

El objetivo principal para la realización de esta investigación de mercado es determinar los hábitos de compra y de consumo, así como conocer la imagen y posicionamiento de golosinas enfocándose en el mercado de chocolates.

### 4.3 METODOLOGÍA DE LA INVESTIGACIÓN DE MERCADO

La recopilación de la información se llevó a cabo a través de un cuestionario estructurado (Anexo C) aplicado en lugares de afluencia del Distrito Federal tales como parques, escuelas y centros comerciales durante la última semana de enero de 2003. La forma de entrevista fue personal y se consideraron los siguientes requisitos:

- Residencia en el Distrito Federal mínima de 6 meses.
- No trabajar en agencias de mercadotecnia o publicidad, ni en compañías que se dediquen a la venta o distribución de golosinas.
- Haber consumido alguna clase de golosina en la última semana.

Se realizó un muestreo no probabilístico por cuotas, con un tamaño de la muestra de 300 entrevistas segmentadas por sexo y edad de la siguiente manera:

Sexo		
Masculino	50%	150
Femenino	50%	150

**TABLA 4.1**  
**Cuotas por Sexo**

Edad		
9-13	20%	60
14-18	20%	60
19-25	20%	60
26-35	20%	60
36-45	20%	60

**TABLA 4.2**  
**Cuotas por Edad**

#### 4.4 IMAGEN DE GOLOSINAS

El número de entrevistas efectivas que se lograron levantar para esta investigación de mercado fue de 309 (hubo una sobrecuota de 9 entrevistas). A continuación se presenta la tabla de contingencia correspondiente a la pregunta número seis en donde se evalúa la imagen de nueve diferentes tipos de golosinas (*señalados como P01-P09*) por medio de nueve atributos o características previamente codificados (*señalados como A01-A09*):

Estudio de Golosinas  
Sección I. Consumo de golosinas entre comidas  
Tabla 1. Imagen de Golosinas  
P6. ¿Cuál de los productos que aparecen en esta tarjeta cree usted que ...?

Base: Total Entrevistados

Atributo	Producto									Total
	P01-Botanas	P02-Chocolates	P03-Galletas	P04-Dulces	P05-Pastillas	P06-Chicles	P07-Pastelitos	P08-Hielados	P09-Untables	
A01-'Da energía'	28	167	10	9	5	5	8	12	45	289
A02-'Es el más rico'	49	81	16	19	4	8	19	76	26	298
A03-'Es para toda la familia'	96	49	56	15	2	4	20	30	20	292
A04-'Nos pone alegres y contentos'	67	67	12	44	5	10	17	52	12	286
A05-'Es para consumirlo en un momento de descanso'	91	32	34	20	13	14	21	46	23	294
A06-'Es para rufos'	19	50	17	128	4	13	16	20	25	292
A07-'Es para acompañar un café'	5	6	215	6	0	3	57	3	9	304
A08-'Es para tener siempre con usted'	33	31	31	32	78	73	5	5	8	296
A09-'Quita el mal aliento'	8	5	1	4	198	84	2	0	2	304
Total	396	488	392	277	309	214	165	244	170	2655

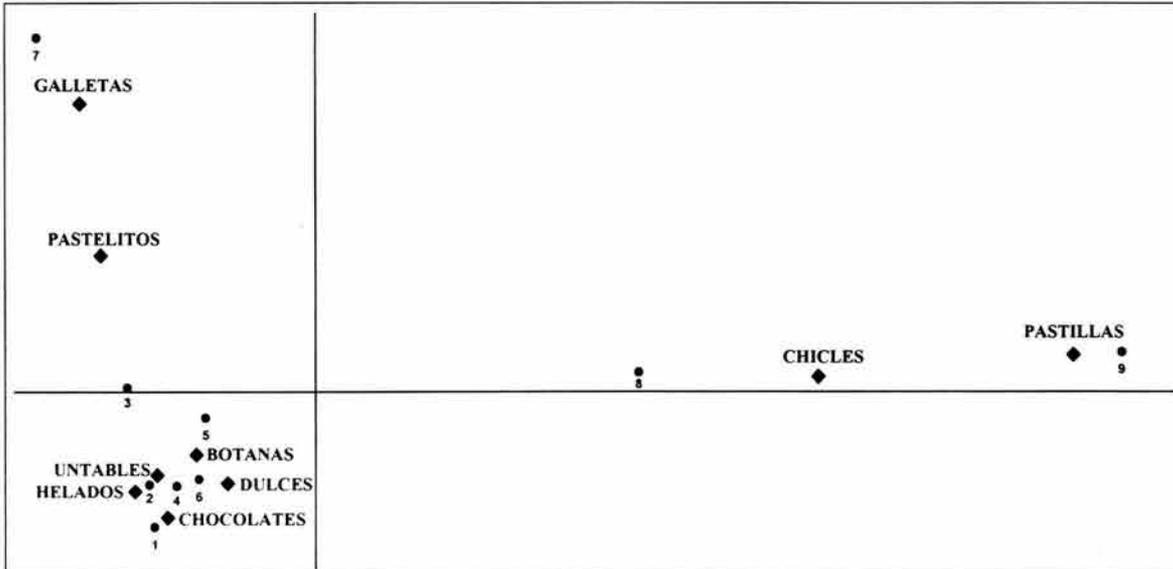
Datos Muestrales

TABLA 4.3  
Tabla de Contingencia

Como se puede apreciar los marginales de los atributos (renglones) no suman 309, esto es porque se eliminaron aquéllas respuestas no específicas como es el caso de 'Ninguna/No contestó'.

Para realizar la prueba ji-cuadrada de independencia se encuentra que  $X^2=3450.543$ . Comparando el valor de este estadístico  $X^2$  con el valor crítico de una  $\chi^2$  con  $(n-1)(p-1) = (9-1)(9-1) = (8)(8) = 64$  grados de libertad y utilizando  $\alpha=0.001$ , se rechaza la hipótesis nula de que las clasificaciones son independientes dado que  $X^2 > \chi^2$ . Además, calculando el coeficiente de contingencia  $C$  y la  $V$  de Cramer se tiene que:  $C=0.752$  y  $V=0.403$ , por lo que se concluye que existe un fuerte grado de dependencia entre las variables.

Aplicando el análisis de correspondencias, se obtiene el siguiente mapa perceptual en dos dimensiones:



- 1=Da energía
- 2=Es el más rico
- 3=Es para toda la familia
- 4=Nos pone alegres y contentos
- 5=Es para consumirlo en un momento de descanso

- 6=Es para niños
- 7=Es para acompañar un café
- 8=Es para tener siempre con usted
- 9=Quita el mal aliento

**GRÁFICA 4.1**  
**Mapa Perceptual**

La calidad de la representación viene dada por la inercia o cantidad de varianza explicada, es decir, la cantidad de información recogida en el mapa. En este caso, el total de información explicada es del 75.83%, 45.83% del eje 1 y 30.00% del eje 2.

Las contribuciones absolutas y los cosenos cuadrados para los renglones (atributos) son:

		Dimensión	
		1	2
Atributo	A01	0.025607	<b>0.102665</b>
	A02	0.028005	0.050303
	A03	0.035605	0.000097
	A04	0.018924	0.049038
	A05	0.012092	0.003805
	A06	0.013590	0.042998
	A07	0.080559	<b>0.739386</b>
	A08	<b>0.106701</b>	0.002165
	A09	<b>0.678918</b>	0.009544
		1.00	1.00

**TABLA 4.4**  
**Contribución Absoluta por Atributo**

		Dimensión	
		1	2
Atributo	A01	0.106708	0.280075
	A02	0.273057	0.321087
	A03	0.460608	0.000823
	A04	0.284975	0.483443
	A05	0.196666	0.040515
	A06	0.056932	0.117926
	A07	0.140761	<b>0.845766</b>
	A08	<b>0.814745</b>	0.010821
	A09	<b>0.980620</b>	0.009025

**TABLA 4.5**  
**Cosenos Cuadrados por Atributo**

Así mismo, las contribuciones absolutas y los cosenos cuadrados para las columnas (productos) quedarían de la siguiente manera:

		Dimensión	
		1	2
Producto	P01	0.019301	0.030316
	P02	0.036618	<b>0.151162</b>
	P03	0.074240	<b>0.633845</b>
	P04	0.007268	0.044988
	P05	<b>0.608053</b>	0.008543
	P06	<b>0.186699</b>	0.000947
	P07	0.026078	0.059553
	P08	0.027230	0.047414
	P09	0.014514	0.023232
		1.00	1.00

**TABLA 4.6**  
**Contribución Absoluta por Producto**

		Dimensión	
		1	2
Producto	P01	0.140779	0.144756
	P02	0.155986	0.421545
	P03	0.150990	<b>0.843929</b>
	P04	0.029081	0.117846
	P05	<b>0.978025</b>	0.008995
	P06	<b>0.917177</b>	0.003045
	P07	0.381332	0.570085
	P08	0.211911	0.241554
	P09	0.312726	0.327715

**TABLA 4.7**  
**Cosenos Cuadrados por Producto**

Como se mencionó anteriormente, el primer factor explica el 45.83% de la variabilidad de la nube. Dos atributos en el campo positivo explican casi la totalidad del factor: *Es para tener siempre con usted (A08)* y *Quita el mal aliento (A09)*; ya que éstos tienen contribuciones absolutas altas. En el espacio de productos, también son dos los que determinan el factor: *Pastillas refrescantes (P05)* y *Chicles (P06)*; ambos en el campo positivo. Además todas estas modalidades tienen una buena calidad de representación sobre el eje si se observan los valores de los cosenos cuadrados.

El segundo factor explica el 30.00% de la varianza. Debido a sus altas contribuciones absolutas, dos atributos explican casi la totalidad del factor: *Da energía (A01)* y *Es para acompañar un café (A07)*, el primero en el campo negativo y el segundo en el campo positivo. En el espacio de productos: *Chocolates (P02)* y *Galletas (P03)* son los que determinan el factor, el primero en el campo negativo y el segundo en el campo positivo. Únicamente, *Es para acompañar un café (A07)* y *Galletas (P03)* tienen una buena calidad de representación sobre el eje de acuerdo a los valores de los cosenos cuadrados. Finalmente se puede concluir que *Chocolates (P02)* está fuertemente asociado al atributo *Da energía (A01)* y *Galletas (P03)* a *Es para acompañar un café (A07)*.

Analizando directamente el mapa perceptual se pueden resumir las siguientes relaciones:

<b>Producto:</b>	<b>Se asocia con atributo:</b>
P01=Botanas	A05='Es para consumirlo en un momento de descanso' A06='Es para niños'
P02=Chocolates	A01='Da energía' A02='Es el más rico' A04='Nos pone alegres y contentos'
P03=Galletas	A07='Es para acompañar un café'
P04=Dulces	A06='Es para niños'
P05=Pastillas	A09='Quita el mal aliento'
P08=Helados	A02='Es el más rico'
P09=Untables	A02='Es el más rico' A04='Nos pone alegres y contentos'

**TABLA 4.8**  
**Tabla Resumen de Imagen de Productos**

### 4.5 IMAGEN DE MARCAS DE CHOCOLATES

Ahora toca el turno de analizar la tabla de contingencia correspondiente a la pregunta número siete en donde se evalúa la imagen de quince diferentes marcas de chocolates (A-O) por medio de veintidós atributos o características previamente codificados (A01-A22).

Estudio de Golosinas  
 Sección II. Imagen de Marcas de Chocolates  
 Tabla 2. Imagen de Marcas de Chocolates  
 P7. ¿Cuál de las compañías que aparecen en esta tarjeta cree usted que ...?

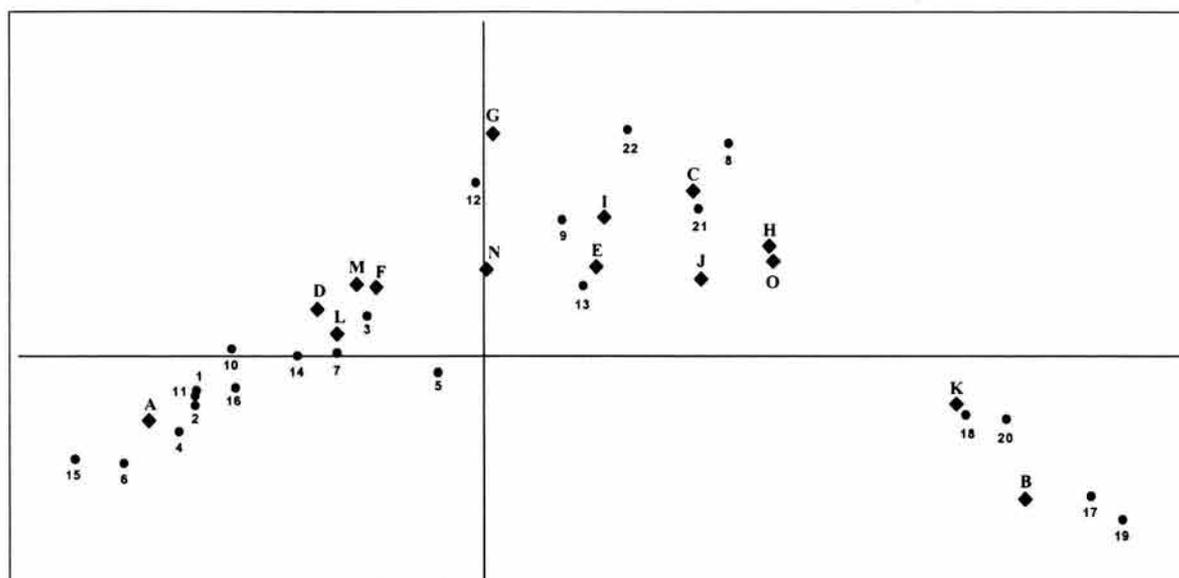
Base: Entrevistados que consumen chocolates

Atributo \ Marca	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	Total
A01='Me hace quedar bien'	166	7	8	35	8	13	6	5	1	5	2	3	5	1	1	266
A02='Es de buena calidad'	172	10	7	17	16	17	1	3	2	1	1	7	5	5	0	264
A03='Es el más rico'	105	15	30	29	27	19	7	5	7	3	3	4	8	0	1	263
A04='Tiene mejor presentación'	188	15	6	24	5	10	1	2	2	3	1	5	3	3	0	268
A05='Tiene mejores anuncios'	100	48	14	37	9	12	18	3	5	5	1	3	1	2	0	258
A06='Es cara'	220	13	3	5	9	3	3	0	0	0	1	8	1	0	0	266
A07='Es el mejor chocolate'	127	19	28	13	18	25	4	8	4	1	2	8	4	0	0	261
A08='Es fácil de encontrar en las tiendas'	17	18	130	9	18	1	3	35	4	16	1	4	5	1	1	263
A09='Es para toda la familia'	42	15	64	28	22	13	5	19	5	17	2	5	4	10	2	253
A10='Es para adultos'	141	3	17	18	19	11	8	4	4	2	0	9	7	7	0	250
A11='Ideal para regalar'	159	12	10	60	3	15	0	3	2	2	0	2	2	2	0	272
A12='Es para jóvenes'	25	7	24	60	22	36	38	9	23	6	6	5	1	2	2	266
A13='Tiene variedad de presentaciones/empaques'	33	36	22	33	55	22	4	2	4	20	5	8	7	4	3	258
A14='Es importado'	130	16	2	9	12	23	16	1	23	2	3	6	1	0	3	247
A15='Es refinado/elegante'	235	4	4	9	3	5	1	1	0	0	1	4	4	1	0	272
A16='No puede faltar en una fiesta importante'	149	10	12	32	9	12	3	4	0	11	2	5	2	0	1	252
A17='Es para niños'	5	181	18	6	3	1	0	28	5	16	9	0	0	0	0	272
A18='Tiene mucha leche'	7	106	12	1	50	6	0	1	12	4	45	4	1	3	10	262
A19='Los niños lo piden con más frecuencia'	2	198	24	3	2	2	3	11	5	8	11	1	0	2	0	272
A20='Los padres lo dan con más confianza a sus hijos'	5	140	48	3	21	7	0	10	5	6	12	1	1	1	2	262
A21='Da energía'	14	25	41	1	36	11	17	7	7	1	6	3	3	1	5	178
A22='Quita el hambre'	15	10	29	3	15	8	43	15	23	7	6	1	3	4	2	184
Total	2057	908	553	435	382	272	181	176	143	136	120	96	68	49	33	5609

Datos Muestrales

TABLA 4.9  
 Tabla de Contingencia

Para esta tabla de contingencia no es seguro realizar la prueba de independencia ji-cuadrada dado que el 28% de las celdas presentan valores esperados menores a 5. Sin embargo, es posible realizar el análisis de correspondencias, por lo que el mapa perceptual en dos dimensiones correspondiente sería:



- 1=Me hace quedar bien
- 2=Es de buena calidad
- 3=Es el más rico
- 4=Tiene mejor presentación
- 5=Tiene mejores anuncios
- 6=Es cara
- 7=Es el mejor chocolate
- 8=Es fácil de encontrar en las tiendas

- 9=Es para toda la familia
- 10=Es para adultos
- 11=Ideal para regalar
- 12=Es para jóvenes
- 13=Tiene variedad de presentaciones/empaques
- 14=Es importado
- 15=Es refinado/elegante
- 16=Nopuede faltar en una fiesta importante

- 17=Es para niños
- 18=Tiene mucha leche
- 19=Los niños lo piden con más frecuencia
- 20=Los padres lo dan con más confianza a sus hijos
- 21=Da energía
- 22=Quita el hambre

**GRÁFICA 4.2**  
**Mapa Perceptual**

En este caso, el total de información explicada es del 69.05%, 47.27% del eje 1 y 21.78% del eje 2.

Las contribuciones absolutas y los cosenos cuadrados para los renglones (atributos) y las columnas (marcas de chocolates) se presentan en las tablas 4.10, 4.11, 4.12 y 4.13.

Los atributos que más contribuyen a la formación del eje 1 son: *Es cara (A06)*, *Es refinado/elegante (A15)*, *Es para niños (A17)*, *Tiene mucha leche (A18)*, *Los niños lo piden con más frecuencia (A19)* y *Los padres lo dan con más confianza a sus hijos (A20)*. De los cuales *Es cara (A06)* y *Es refinado/elegante (A15)* están en el campo negativo. A excepción

de *Es cara (A06)* y *Tiene mucha leche (A18)*, todos los demás atributos tienen una buena calidad de representación sobre el eje. En el espacio de las marcas de productos, las que más contribuyen en la formación de este eje son: *A* y *B*, *A* en el campo negativo y *B* en el campo positivo. Ambas marcas tienen una buena calidad de representación.

Para el segundo eje, los atributos que más contribuyen a su formación son: *Es fácil de encontrar en las tiendas (A08)*, *Es para toda la familia (A09)*, *Es para jóvenes (A12)*, *Da energía (A21)* y *Quita el hambre (A22)*, mientras que en el espacio de las marcas las que más contribuyen son: *C* y *G*. Tanto los atributos como las marcas de productos se encuentran en el campo positivo. Ninguna de estas coordenadas factoriales tiene una buena calidad de representación sobre el eje.

Atributo		Dimensión	
		1	2
A01		0.037051	0.004774
A02		0.037123	0.010244
A03		0.006055	0.006624
A04		0.041971	0.023945
A05		0.000959	0.001054
A06		<b>0.057757</b>	0.047876
A07		0.009552	0.000036
A08		0.026602	<b>0.189305</b>
A09		0.002600	<b>0.074254</b>
A10		0.026760	0.000166
A11		0.038090	0.006745
A12		0.000028	<b>0.126531</b>
A13		0.004234	0.019975
A14		0.014395	0.000001
A15		<b>0.076085</b>	0.045836
A16		0.026088	0.003935
A17		<b>0.167555</b>	0.083109
A18		<b>0.101672</b>	0.014446
A19		<b>0.185266</b>	0.114430
A20		<b>0.119938</b>	0.016430
A21		0.013773	<b>0.060800</b>
A22		0.006446	<b>0.149485</b>
		1.00	1.00

**TABLA 4.10**  
Contribución Absoluta por Atributo

Atributo		Dimensión	
		1	2
A01		<b>0.858227</b>	0.050935
A02		0.795874	0.101164
A03		0.370764	0.186842
A04		0.771229	0.202687
A05		0.055867	0.028276
A06		0.619815	0.236672
A07		0.479148	0.000833
A08		0.116886	0.383162
A09		0.040480	0.532552
A10		0.756228	0.002155
A11		0.573831	0.046812
A12		0.000227	0.467383
A13		0.062851	0.136597
A14		0.289618	0.000006
A15		<b>0.702691</b>	0.195001
A16		0.741319	0.051502
A17		<b>0.726222</b>	0.165930
A18		0.458496	0.030009
A19		<b>0.736618</b>	0.209581
A20		<b>0.893304</b>	0.056369
A21		0.216941	0.441140
A22		0.044775	0.478346

**TABLA 4.11**  
Cosenos Cuadrados por Atributo

		Dimensión	
		1	2
Producto	A	0.387467	0.136425
	B	0.444534	0.291608
	C	0.040808	0.236167
	D	0.020409	0.015318
	E	0.008113	0.047953
	F	0.005343	0.020007
	G	0.000026	0.141986
	H	0.024164	0.033628
	I	0.003451	0.043201
	J	0.010772	0.012598
	K	0.044893	0.004232
	L	0.003517	0.000783
	M	0.001869	0.005593
	N	0.000000	0.005791
	O	0.004634	0.004771
		1.00	1.00

**TABLA 4.12**  
Contribución Absoluta por Producto

		Dimensión	
Producto	A	0.837250	0.135793
	B	0.755937	0.228426
	C	0.166252	0.443211
	D	0.181466	0.062739
	E	0.080866	0.220177
	F	0.115378	0.199019
	G	0.000161	0.400488
	H	0.307734	0.197273
	I	0.047073	0.271455
	J	0.244096	0.131506
	K	0.365034	0.015853
	L	0.318554	0.032662
	M	0.135913	0.187361
	N	0.000011	0.115437
	O	0.155485	0.072804

**TABLA 4.13**  
Cosenos Cuadrados por Producto

Analizando directamente el mapa perceptual se puede observar que por ejemplo, las marcas de chocolate *B* y *K* aparecen de una forma apartada en el cuarto cuadrante, ya que se ubican como productos dirigidos especialmente *para niños* (*A17*, *A19* y *A20*). Por otro lado, las marcas de productos *J*, *H* y *O* no cuentan con una imagen bien definida, ya que no se mostraron asociadas a ningún atributo. Algo similar sucede con el atributo *Tiene mejores anuncios* (*A05*), ya que éste aparentemente no se encuentra asociado a alguna marca de chocolates. En el caso de la marca *A*, que es la marca que solicitó la investigación de mercado, se puede observar que es asociado con varios atributos a la vez, entre los que se encuentran: *Me hace quedar bien* (*A01*), *Es de buena calidad* (*A02*), *Tiene mejor presentación* (*A04*), *Es caro* (*A06*), *Ideal para regalar* (*A11*), *Es refinado/elegante* (*A15*) y *No puede faltar en una fiesta importante* (*A16*).

Para nuestra marca de interés *A*, no es posible definir a través del mapa perceptual cuál marca es una fuerte competencia para ella. Aunque, es importante considerar que las marcas *F*, *L* y *M* están asociadas con atributos que se consideran importantes como lo son: *Es el más rico* (*A03*) y *Es el mejor chocolate* (*A07*). También hay que observar que en el

subcapítulo anterior se descubrió que el *chocolate* es percibido como un producto que *da energía* y ahora en este nuevo mapa perceptual, la marca *C* (líder en ventas) está asociada con este atributo (*A21*).

Por todo lo explicado anteriormente, en términos generales se puede concluir que la marca *A* se encuentra bien posicionada en el mercado, aunque no hay que perder de vista que tres atributos de gran importancia como lo son *Es el más rico* (*A03*), *Es el mejor chocolate* (*A07*) y *Da energía* (*A21*) están siendo asociados con la competencia. Tal vez valdría la pena esforzarse por realizar una campaña publicitaria con el fin de buscar cambiar la percepción del mercado en cuanto a que el producto es un producto caro y además haciendo énfasis en los tres atributos antes mencionados.

Finalmente, se pueden resumir las siguientes relaciones:

Producto:	Se asocia con atributo:
A	A01='Me hace quedar bien' A02='Es de buena calidad' A04='Tiene mejor presentación' A06='Es caro' A11='Ideal para regalar' A15='Es refinado/elegante' A16='No puede faltar en una fiesta importante'
B	A17='Es para niños' A19='Los niños lo piden con más frecuencia'
C	A08='Es fácil de encontrar en las tiendas' A21='Da energía'
F y M	A03='Es el más rico'
E	A09='Es para toda la familia' A13='Tiene variedad de presentaciones/empaques'
I	A09='Es para toda la familia'
K	A18='Tiene mucha leche' A20='Los padres lo dan con más confianza a sus hijos'
L	A03='Es el más rico' A07='Es el mejor chocolate'

**TABLA 4.14**  
**Tabla Resumen de Imagen de Marcas de Chocolates**

## 4.6 ANÁLISIS DE LA FRECUENCIA DE CONSUMO DE CHOCOLATES

Ahora toca el turno de analizar la tabla de contingencia correspondiente a la pregunta número ocho, la cual se cruza con la edad del entrevistado:

Estudio de Golosinas  
Sección III. Hábitos de Consumo y Compra de Chocolates  
Tabla 3. Frecuencia de Consumo de Chocolates  
P8. ¿Con qué frecuencia acostumbra consumir chocolates?

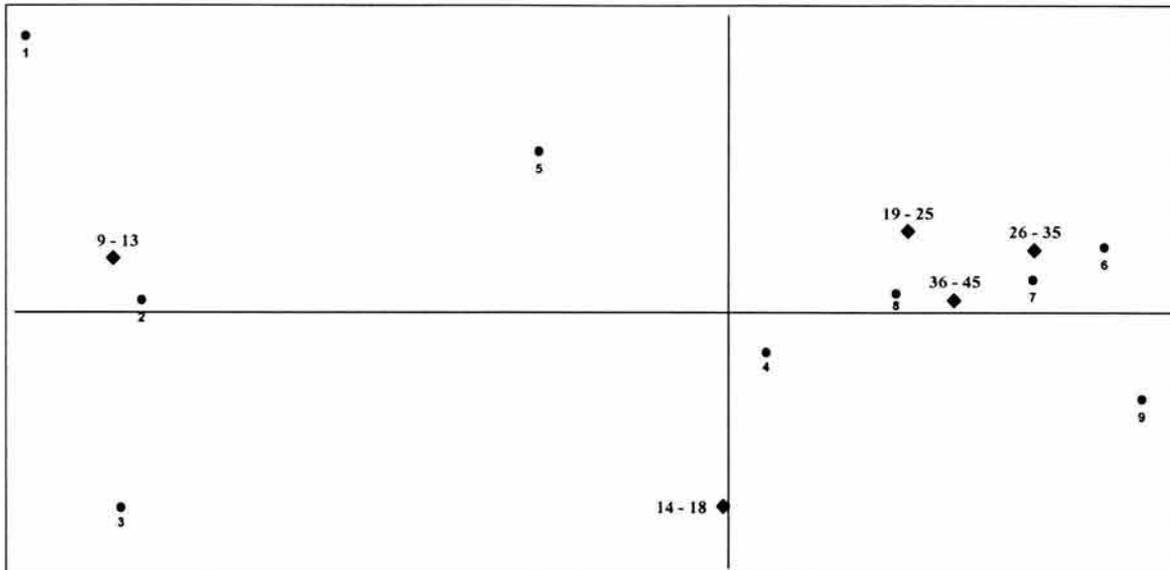
Base: Entrevistados que consumen chocolates

Frecuencia	Edad					Total
	1= '09 a 13'	2= '14 a 18'	3= '19 a 25'	4= '26 a 35'	5= '36 a 45'	
F01='Varias veces al día'	1	0	1	0	0	2
F02='Diario'	9	4	3	1	4	21
F03='4 ó 6 veces a la semana'	9	10	3	1	1	24
F04='2 ó 3 veces a la semana'	16	20	11	19	14	80
F05='1 vez a la semana'	17	3	10	12	9	51
F06='Cada 15 días'	2	5	10	7	1	25
F07='1 vez al mes'	3	4	5	5	8	25
F08='1 vez cada 2 ó 3 meses'	1	1	1	1	2	6
F09='Rara vez'	2	12	8	9	11	42
Total	60	59	52	55	50	276

Datos Muestrales

**TABLA 4.15**  
**Tabla de Contingencia**

Para esta tabla de contingencia tampoco es seguro realizar la prueba de independencia ji-cuadrada dado que el 53% de las celdas presentan valores esperados menores a 5. Sin embargo, es posible realizar el análisis de correspondencias y el mapa perceptual en dos dimensiones correspondiente sería:



1=Varias veces al día  
 2=Diario  
 3=4 ó 6 veces a la semana  
 4=2 ó 3 veces a la semana  
 5=1 vez a la semana

6=Cada 15 días  
 7=1 vez al mes  
 8=1 vez cada 2 ó 3 meses  
 9=Rara vez

**GRÁFICA 4.3**  
**Mapa Perceptual**

En este caso, el total de información explicada es del 71.78%, 44.41% del eje 1 y 27.37% del eje 2.

Las contribuciones absolutas y los cosenos cuadrados para los renglones (frecuencias) quedan como sigue:

		Dimensión	
		1	2
Frecuencia	F01	0.030906	0.051869
	F02	<b>0.225784</b>	0.000964
	F03	<b>0.277058</b>	0.307790
	F04	0.003521	0.043622
	F05	0.058095	<b>0.444476</b>
	F06	0.108050	0.035859
	F07	0.071117	0.008584
	F08	0.005216	0.000647
	F09	<b>0.220254</b>	0.106189
		1.00	1.00

**TABLA 4.16**  
**Contribución Absoluta por Frecuencia**

		Dimensión	
		1	2
Frecuencia	F01	0.278205	0.287701
	F02	<b>0.854401</b>	0.002248
	F03	0.547875	0.375043
	F04	0.040591	0.309899
	F05	0.167257	<b>0.788512</b>
	F06	0.245526	0.050209
	F07	0.460421	0.034243
	F08	0.146406	0.011190
	F09	<b>0.718965</b>	0.213590

**TABLA 4.17**  
**Cosenos Cuadrados por Frecuencia**

De igual manera, las contribuciones absolutas y los cosenos cuadrados para las columnas (edades) toman los siguientes valores:

Edad	Dimensión	
	1	2
EDAD1	0.710349	0.058736
EDAD2	0.000052	0.751019
EDAD3	0.052056	0.115779
EDAD4	0.159049	0.072260
EDAD5	0.078493	0.002205
	1.00	1.00

**TABLA 4.18**  
Contribución Absoluta por Edad

Edad	Dimensión	
	1	2
EDAD1	0.946683	0.048234
EDAD2	0.000109	0.964218
EDAD3	0.156134	0.213977
EDAD4	0.554280	0.155172
EDAD5	0.195754	0.003389

**TABLA 4.19**  
Cosenos Cuadrados por Edad

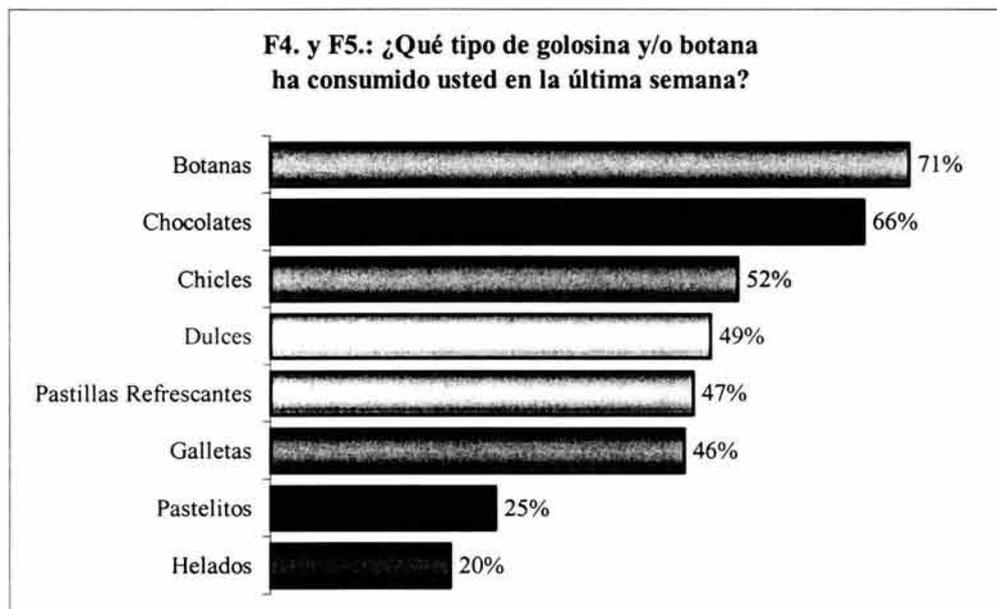
Para la formación del primer eje, las frecuencias que más contribuyen son: *Diario (F02)*, *4 ó 6 veces a la semana (F03)* y *Rara vez (F09)*, las dos primeras en el campo negativo y la tercera en el campo positivo. A excepción de *4 ó 6 veces a la semana (F03)*, todas las demás frecuencias poseen una buena calidad de representación. En el espacio de las edades, el rango de edad que más contribuye a la formación del eje 1 es: *de 9 a 13 años (EDAD1)* en el campo negativo, el cual además tiene una buena calidad en la representación.

En lo que se refiere al segundo eje, la frecuencia que más contribuye a su formación es: *1 vez a la semana (F05)*, en el campo positivo. Para el espacio de las edades, el rango de edad que más contribuye es: *de 14 a 18 años (EDAD2)*, en el campo negativo. Ambas modalidades cuentan con una buena calidad en la representación.

Finalmente, analizando directamente el mapa perceptual anterior, se puede concluir que sí existe una fuerte asociación entre la edad y la frecuencia de consumo ya que a menor edad mayor es la frecuencia de consumo de chocolates y por el contrario, las personas adultas realizan un consumo más moderado del producto.

#### 4.7 HÁBITOS DE CONSUMO Y COMPRA DE CHOCOLATES

En lo que sigue se hará una representación gráfica de los principales resultados obtenidos en la investigación de mercado de imagen y posicionamiento de golosinas. Primeramente, se presentan los resultados del consumo total de golosinas correspondientes a las preguntas F4. y F5. de la Sección 0. Filtros:

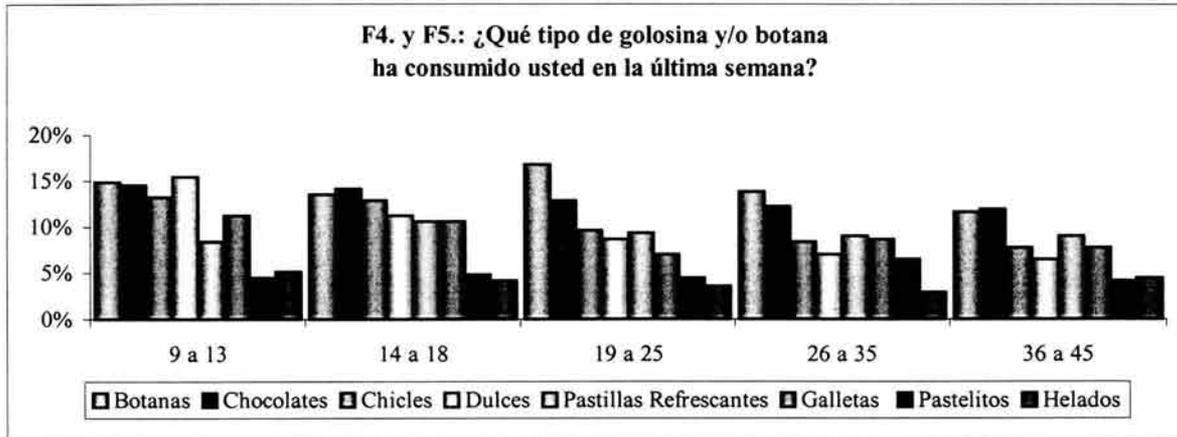


**GRÁFICA 4.4**

Como se puede apreciar en este gráfico, las botanas y los chocolates son los productos que más se consumen fuera de las horas de comida. En el caso contrario se encuentran los helados, los cuales sólo fueron mencionados por el 20% de los entrevistados. El consumo de galletas, pastillas refrescantes, dulces y chicles fue mencionado por casi el 50% de los entrevistados. Se utilizó este tipo de gráfico de barras<sup>27</sup> ya que es una pregunta de respuesta múltiple y la suma de porcentajes no da 100%.

<sup>27</sup> Gráfico de Barras: Serie de barras verticales u horizontales, donde la longitud de la barra representa la frecuencia porcentual de una categoría de una variable nominal u ordinal. Los gráficos de barras son especialmente eficaces para ilustrar la competencia entre categorías. Ibidem, RITCHEY, Ferris. p. 79

Ahora, esta misma información se analiza en un gráfico similar pero mostrando los porcentajes obtenidos en cada grupo de edad:

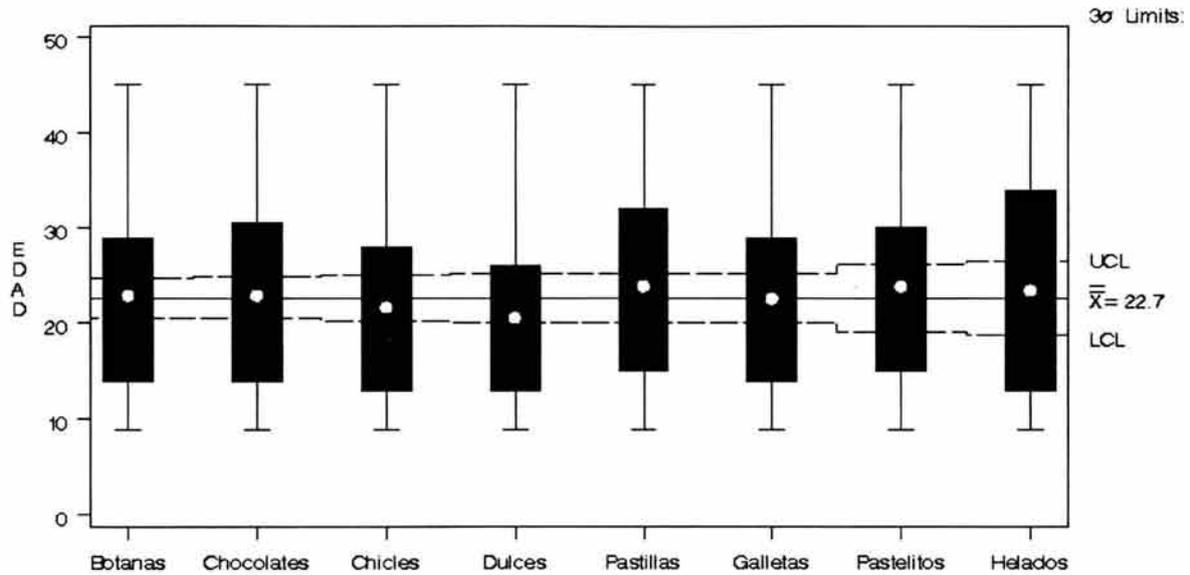


GRÁFICA 4.5

De acuerdo a lo mostrado en esta gráfica, se puede afirmar que en especial el consumo de chocolates se mantiene de manera constante en los cinco grupos de edad. Vale la pena señalar que en el caso de las botanas, para el rango de edad de 19 a 25, se registró un mayor consumo de éstas. De igual manera, el consumo de dulces presenta un “pico” en las respuestas de los entrevistados de 9 a 13 años. Por el contrario, la golosina que menor porcentaje de consumo registró fueron los helados, especialmente entre los entrevistados de 26 a 35 años.

Ahora bien, presentemos estos mismo resultados utilizando un diagrama de caja<sup>28</sup>:

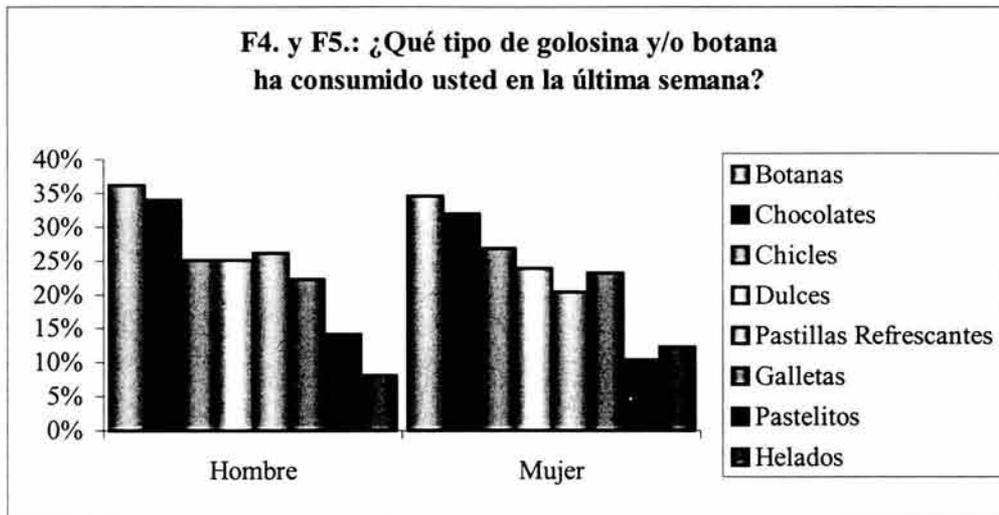
<sup>28</sup> Diagrama de caja: Es una técnica gráfica de gran utilidad para detectar la existencia de valores atípicos (observaciones o datos que están a más de 1.5 veces el rango intercuartílico del correspondiente cuartil), así como las puntuaciones mínimas y máximas en una distribución. Dicho gráfico consta de una caja determinada por el primer y el tercer cuartil, dentro de la caja se representa mediante una línea la mediana (segundo cuartil) y mediante un pequeño círculo la media.



**GRÁFICA 4.6**  
**Diagrama de Caja**

En nuestro caso, los diagramas de caja no se usan con la finalidad de detectar la existencia de valores atípicos ya que la muestra fue tomada de acuerdo a cuotas establecidas y todos los entrevistados caen dentro del rango de edad de 9 a 45 años. Sin embargo, dichos diagramas nos permiten obtener nueva información. Por ejemplo, enfocándonos en la mediana de chocolates, chicles, dulces y helados, se puede observar que el consumo tales golosinas se agrupa hacia la parte inferior de la distribución, es decir, hacia los entrevistados de 9 a 20 años de edad. Para el caso de los helados, se puede ver que el consumo se dispersa ampliamente entre un mayor rango de edades. Lo contrario sucede con el consumo de dulces, el cual se dispersa en un menor rango de edades.

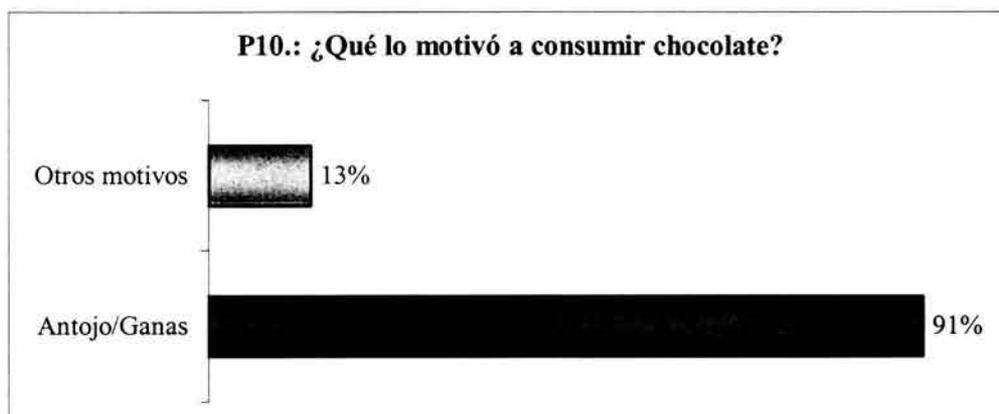
Analicemos ahora los datos obtenidos en estas mismas preguntas, pero segmentando la información por sexo:



**GRÁFICA 4.7**

En la gráfica anterior se puede observar que el consumo de botanas y chocolates se comporta de manera similar en ambos sexos. También se puede comentar que los hombres muestran una ligera inclinación por el consumo de pastillas refrescantes, mientras que las mujeres por el consumo de chicles.

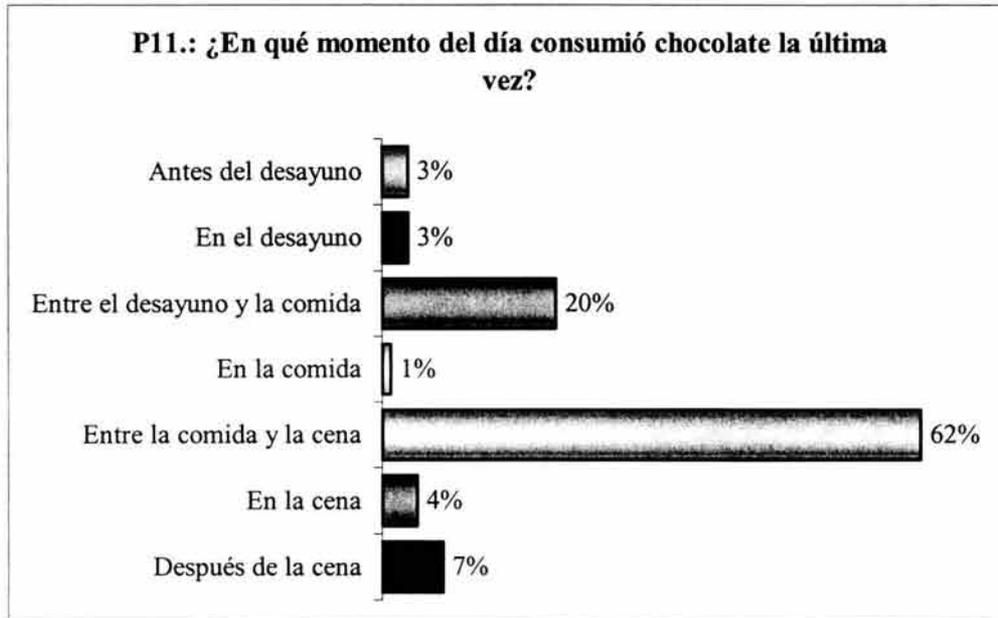
Para la pregunta número diez se obtienen los siguientes resultados:



**GRÁFICA 4.8**

Sin lugar a dudas, el antojo/ganas se distingue como el principal motivo que orilla a los entrevistados a consumir chocolate.

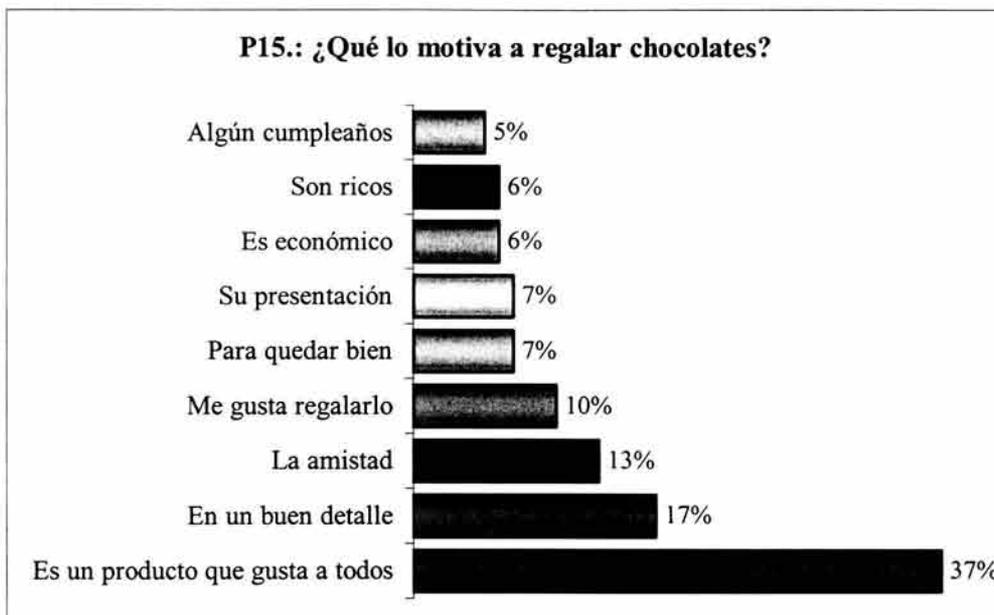
De manera similar, se obtuvo una respuesta contundente de parte de los entrevistados para la pregunta once:



**GRÁFICA 4.9**

El tiempo que se da entre la comida y la cena resulta ser el momento en el que la mayoría de los entrevistados señalaron que realizan el consumo de chocolate.

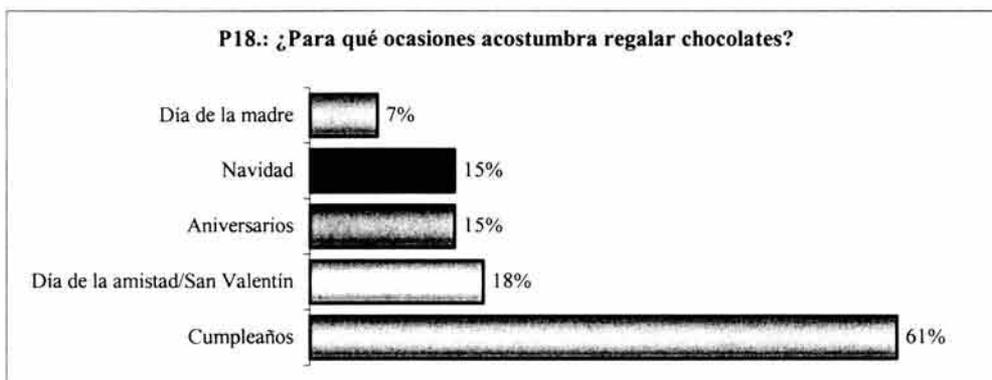
Un 50% de los entrevistados que consumen chocolate mencionó que también acostumbra regalar chocolates a otras personas. Con esta nueva base de datos se muestran los resultados de la pregunta quince:



GRÁFICA 4.10

Un 37% de los entrevistados mencionan que lo que los motiva a regalar chocolates es que éstos son un producto del gusto de todos. Así mismo, un 17% considera que los chocolates son un buen detalle para regalar.

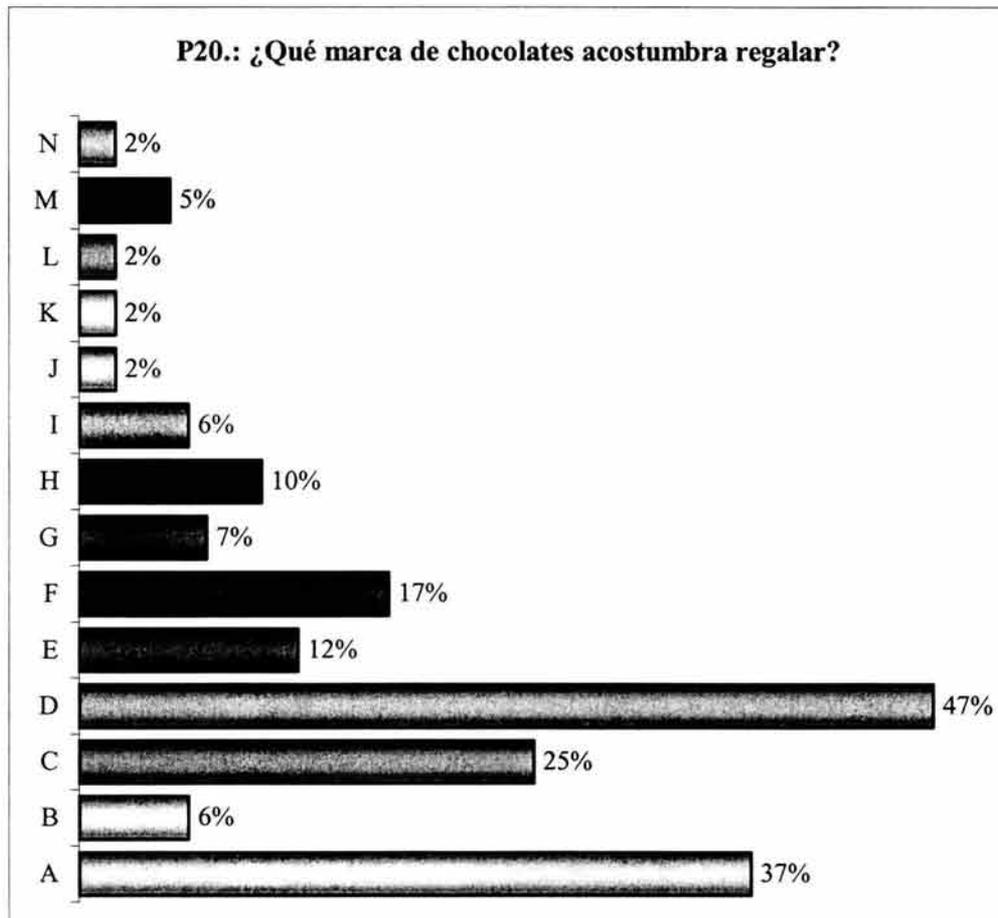
En el gráfico correspondiente a la pregunta dieciocho aparecen las principales ocasiones para las cuales los entrevistados acostumbran regalar chocolates:



GRÁFICA 4.11

La principal ocasión que los entrevistados mencionan con un 61%, para regalar chocolates son los cumpleaños.

De acuerdo a la marca de chocolates que los entrevistados acostumbran regalar se encuentra lo siguiente:



GRÁFICA 4.12

En este gráfico resulta evidente que la marca D se muestra como líder en cuanto al consumo de chocolates para regalo, seguida por nuestra marca de interés A y por la marca líder en ventas C.

## **4.8 ANÁLISIS DE RESULTADOS**

A través del mapa perceptual de imagen y posicionamiento de golosinas se pudo determinar que el chocolate es percibido por los consumidores como un producto que da energía, es rico y hace sentir a uno alegre y contento. También fue interesante observar que curiosamente la marca de chocolates líder en ventas (llamada marca C) es percibida como aquella que da energía y es fácil de encontrar en las tiendas, atributos que por sí mismos posicionan de manera apropiada a la marca. Por otra parte, la empresa de la marca de chocolate que solicitó la investigación de mercado (A) se encontró ante el problema de ser percibida como una marca cara.

En esta misma investigación se pudo observar que:

- Las botanas son la golosina que registró un mayor consumo de parte de los entrevistados. Los chocolates ocuparon la segunda posición.
- El principal motivo por el que los entrevistados consumen chocolates es el antojo/ganas.
- La hora entre la comida y la cena es el momento en el que más frecuentemente se realiza el consumo de chocolates.
- El principal motivo por el que los entrevistados regalan chocolates es porque consideran que es un producto del gusto de todos, además de que la ocasión en que más acostumbran regalarlos es en los cumpleaños.

## CONCLUSIONES

A lo largo de este trabajo se ha pretendido exponer a través de una aplicación, la eficacia del análisis de correspondencias en el estudio de grandes masas de información, ya que permite analizar todo tipo de variables –nominales, ordinales, de intervalo o de razón– y mediante una reducción de la dimensionalidad se manifiesta como una herramienta de síntesis notable que ofrece una representación gráfica de la información.

Se ha visto que el análisis de correspondencias es una técnica factorial que consiste básicamente en un doble análisis de componentes principales, uno realizado sobre los renglones y otro realizado sobre las columnas de una tabla de datos, centrado en los individuos, a diferencia del análisis factorial que se enfoca en las variables. Así mismo, se han señalado las principales propiedades del análisis de correspondencias: el uso de la distancia especial de origen euclidiano llamada ji-cuadrada y la representación simultánea de todas las categorías de las variables por medio de un mapa perceptual. En general, se puede decir que al interpretar los resultados del análisis de correspondencias a través de un mapa perceptual lo que se hace es examinar la posición de los puntos renglón y de los puntos columna con el fin de descubrir qué categorías de la variable renglón son similares, qué categorías de la variable columna son similares y qué categorías de renglón y columna son similares entre sí.

De igual forma, se ha podido aprender que una posible limitación que se puede encontrar al aplicar el análisis de correspondencias es que al utilizar menos dimensiones que el número máximo de dimensiones posibles, la calidad de la solución no sea tan buena y esto puede ocasionar que se llegue a conclusiones equivocadas al interpretar los resultados.

Vale la pena considerar que el hecho de que esta técnica sea originaria de Francia, y dada la poca importancia que los norteamericanos prestan a los trabajos de otros países, es posiblemente uno de los motivos por el que en un principio no haya sido tan ampliamente difundida, como sucedió por ejemplo con el análisis factorial.

Sin embargo, actualmente se puede afirmar que el análisis de correspondencias es más conocido y empleado por investigadores de todas las ramas científicas. Tal es el caso de la mercadotecnia con la investigación de mercados, área en la cual encuentra su mayor campo de aplicación en las investigaciones que buscan evaluar la forma en que las diferentes marcas (productos o servicios) son percibidas por los consumidores: investigaciones de mercado de imagen y posicionamiento.

Por otro lado, el análisis de tablas de contingencia se ha manifestado como una poderosa herramienta de investigación que permite realizar pruebas de independencia a través de las cuales se puede saber si dos principios de clasificación están o no relacionados significativamente. Por medio del cálculo de medidas de asociación tales como la V de Cramer, el coeficiente de contingencia C y la misma inercia  $\phi^2$  es posible determinar el grado de esa relación. Es importante tomar en cuenta que la prueba de independencia ji-cuadrada es un método aproximado que se hace cada vez más exacto a medida que el número de las observaciones de las celdas de la tabla de contingencia se incrementan.

Como se pudo observar en la investigación de mercado de imagen y posicionamiento de marcas de chocolates presentada en el capítulo 4, el análisis de correspondencias se aplicó a las tablas de contingencia que se obtuvieron al realizar una encuesta entre los consumidores a través de un cuestionario estructurado. En dicha investigación se pudo conocer un poco más acerca del mercado de este tipo de golosina, la cual es percibida por los consumidores fundamentalmente como un producto que da energía. De acuerdo a los resultados obtenidos en la investigación, se puede concluir que todo el trabajo que ha realizado la empresa con el fin de posicionar la marca ha servido de mucho, aunque está presente la imagen de ser un chocolate caro. Es por ello que se recomienda seguir trabajando en cuanto al posicionamiento de la marca con publicidad a través de los diferentes medios de comunicación. Dado que el antojo/ganas es el principal motivo por el que los entrevistados mencionan que se sienten motivados a consumir chocolate y que además el momento de mayor consumo se da entre la hora de la comida y la cena; se recomienda que se realicen

campañas publicitarias a través de radio y televisión dentro de un horario de tres a siete de la tarde.

Es importante mencionar que las investigaciones de mercado de imagen y posicionamiento no se aplican únicamente en el estudio de las marcas y productos, sino ahora también en ámbitos como las preferencias de los electores, esto como resultado del reciente surgimiento de la mercadotecnia política.

Gracias a la disponibilidad de software estadístico como el SAS (Statistical Analysis System) y el SPSS (Statistical Package for the Social Sciences), la elaboración de los cálculos del análisis de tablas de contingencia, del análisis de correspondencias y la creación de los mapas perceptuales se ha ido facilitando lo que a su vez ha contribuido a su mayor uso, ya que de otro modo se necesitaría de una mayor cantidad de tiempo y esfuerzo para el desarrollo de los cálculos requeridos.

Finalmente, a través de este trabajo fue posible mostrar que para cualquier empresa que ofrece algún producto o servicio, el análisis de correspondencias es una técnica estadística práctica y efectiva que permite conocer de qué manera es percibida tanto ésta como la competencia por los consumidores y por lo tanto resulta imprescindible el uso de esta técnica para plantear estrategias de mercadotecnia que construyan o moldeen la identidad de la empresa o sus productos. Además, el desarrollo de este trabajo me permitió obtener un nuevo panorama del análisis de datos, integrando los conocimientos adquiridos durante la carrera y la práctica profesional en diferentes temas como lo son estadística, álgebra lineal, cálculo numérico y geometría.

## REFERENCIAS BIBLIOGRÁFICAS

AAKER, David y DAY (1990). Investigación de Mercados. Colombia: McGraw-Hill. Tercera Edición.

AZORIN, Francisco (1972). Curso de Muestreo y Aplicaciones. España: Aguilar.

BENZÉCRI, Jean (1992). Correspondence Analysis Handbook. New York: Marcel Dekker, Inc.

BERENSON, Mark y LEVINE (1987). Estadística para Administración y Economía. Conceptos y Aplicaciones. México: Nueva Editorial Interamericana.

BISQUERRA, Rafael (1989). Introducción Conceptual al Análisis Multivariante. Un Enfoque Informático con los Paquetes SPSS-X, BMDP, LISREL y SPAD. Volumen II. Barcelona: PPU.

BLASIUS, Jorg y GREENACRE (1998). Visualization of Categorical Data. USA: Academic Press.

BOYD, Harper (1996). Investigación de Mercados. Texto y Casos. México: Limusa. Quinta Edición.

CAMPBELL, Donald y STANLEY (1993). Diseños Experimentales y Cuasiexperimentales en la Investigación Social. Argentina: Amorrortu Editores. Sexta Edición.

CECIL, Frederick (1969). Métodos Estadísticos. España: Aguilar.

CHAO, Lincoln (1993). Introducción a la Estadística. México: Compañía Editorial Continental. Séptima Edición.

COCHRAN, William (1998). Técnicas de Muestreo. México: Compañía Editorial Continental. Decimocuarta Edición.

CORNEJO, José (1988). Técnicas de Investigación Social: El Análisis de Correspondencias. Barcelona: PPU.

CRAMER, Duncan (1998). Fundamental Statistics for Social Research. Step-by-step calculations and computer techniques using SPSS for Windows. London: Routledge.

CUADRAS, C. (1981). Métodos de Análisis Multivariante. Barcelona: EUNIBAR.

DILLON, William (1996). La Investigación de Mercados en un entorno de Marketing. España: Irwin. Tercera Edición.

EVERITT, B. y DER (1996). A Handbook of Statistical Analyses Using SAS. Great Britain: Chapman & Hall.

EVERITT, B. (1977). The Analysis of Contingency Tables. USA: Chapman & Hall.

EVERITT, B. (1999). The Cambridge Dictionary of Statistics. United Kingdom: Cambridge University Press.

GREENACRE, Michael (1984). Theory and Applications of Correspondence Analysis. USA: Academic Press.

HAIR, Joseph y ANDERSON (1987). Multivariate Data Analysis with Readings. New York: Macmillan Publishing Company.

HERNÁNDEZ, José y SAIZ (1996). Marketing y Tecnología. España: Ediciones Pirámide.

HOEL, Paul (1976). Estadística Elemental. México: Compañía Editorial Continental. Tercera edición.

JAMBU, Michel (1991). Exploratory and Multivariate Data Analysis. USA: Academic Press.

JOARISTI, L. y LIZASOAIN (2000). Análisis de Correspondencias. Madrid: La Muralla, S.A.

JOBSON, J. (1992). Applied Multivariate Data Analysis. Volume II. USA: Springer.

KENDALL, Kenneth y KENDALL (1991). Análisis y diseño de sistemas. México: Prentice Hall Hispanoamericana.

KLOTTER, Philip (1989). Mercadotecnia. México: Prentice Hall Hispanoamericana.

LEBART, Ludovic (1984). Multivariate Descriptive Statistical Analysis. USA: John Wiley & Sons.

LOHR, Sharon (1999). Muestreo: Diseño y Análisis. México: International Thomson Editores.

MENDENHALL, William (1994). Estadística Matemática con Aplicaciones. México: Grupo Editorial Iberoamérica. Segunda Edición.

MORRIS, Christopher (1996). Diccionario Enciclopédico de Ciencia y Tecnología. Tomo III. México: Prentice-Hall Hispanoamericana.

PEÑA, Daniel y ROMO (1997). Introducción a la Estadística para las Ciencias Sociales. España: McGraw-Hill.

PÉREZ, César (2000). Técnicas de Muestreo Estadístico. Teoría, Práctica y Aplicaciones Informáticas. México: Alfaomega.

RITCHEY, Ferris (2002) Estadística para las Ciencias Sociales. El Potencial de la Imaginación Estadística. México: McGraw-Hill.

SÁNCHEZ, Juan (1992). Análisis de Tablas de Contingencias. España: Siglo XXI. Segunda Edición.

STOKES, Maura (2000). Categorical data analysis using the SAS system. USA: SAS Institute.

## REFERENCIAS ELECTRÓNICAS

(Junio,2003) Una Imagen vale más que mil palabras.

<http://148.205.64.181/microseg/Administrador/Uploader/material/Una%20Imagen%20Vale%20Mas%20que%20Mil%20Palabras.PDF>

CARRASCO, Salvador (Junio,2003). Análisis de Correspondencias Simples.

<http://www.uv.es/~carrascs/PDF/ACS.PDF>

GARNICA, Elsy (Junio,2003). Análisis Multivariante sobre la vivienda.

<http://iies.faces.ula.ve/revista/articulos/revi10garnicao.pdf>

HERRERO, Francisco y CUESTA (Junio,2003). Introducción al Análisis de Correspondencias: Uso en Análisis Multidimensionales.

[http://www.uniovi.es/~Psi/Dpto\\_Psicologia/metodos/hardcopy/98\\_1.html](http://www.uniovi.es/~Psi/Dpto_Psicologia/metodos/hardcopy/98_1.html)

HERRERO, Francisco y CUESTA (Junio,2003). Relación Bivariada.

<http://www.uniovi.es/UniOvi/Apartados/Departamento/Psicologia/metodos/tutor.6>

LÓPEZ, Carlos (Junio,2003). La Investigación de Mercados I.

<http://www.gestiopolis.com/canales/demarketing/articulos/no%209/invmDOS.htm>

LÓPEZ, Carlos (Junio,2003). La Investigación de Mercados II.

<http://www.gestiopolis.com/canales/demarketing/articulos/no%2010/investigacionmercadoII.htm>

MICHELOUD, Francois (Junio,2003). Correspondence Analysis.

<http://www.micheloud.com/FXM/COR/E/index.htm>

MOLINA, Cuauhtémoc (Junio,2003). La Investigación de Mercados.

<http://cmg-uv.tripod.com/merca/b02.htm>

SALINAS, Oscar (Junio,2003). La Investigación de Mercados III.

<http://www.gestiopolis.com/canales/demarketing/articulos/no11/invmdos3.htm>

---

**ANEXO A**

```
data ejemplo1;
input car $ tienda $ number;
cards;
CAR1 A 8
CAR1 B 11
CAR1 C 5
CAR1 D 7
CAR1 E 12
CAR2 A 5
CAR2 B 15
CAR2 C 3
CAR2 D 1
CAR2 E 6
CAR3 A 105
CAR3 B 32
CAR3 C 11
CAR3 D 23
CAR3 E 37
CAR4 A 32
CAR4 B 57
CAR4 C 6
CAR4 D 2
CAR4 E 25
proc freq;
tables car*tienda / expected chisq;
weight number;
run;
```

**ANEXO B**

```
libname TESINA 'C:\TESINA\SAS';
```

```
*-----+  
| DATOS DE JOBSON, J. (1992). Applied Multivariate Data Analysis. |  
| Volume II. USA: Springer |  
+-----+;
```

```
data TESINA.MAP_1;  
input prod $ A B C D E;  
CARDS;  
C1 8 11 5 7 12  
C2 5 15 3 1 6  
C3 105 32 11 23 37  
C4 32 57 6 2 25  
;
```

```
RUN;
```

```
proc corresp data=TESINA.MAP_1 out=TESINA.MAP_2;  
var A B C D E;  
id prod;  
run;
```

## ANEXO C

**Jóvenes y Adultos**

Entrevista

Día   Mes

Hora Inicio   :

---

*Buenos días/tardes, mi nombre es \_\_\_\_\_ y trabajo para una agencia de investigación de mercados. Estamos llevando a cabo una encuesta sobre "productos que se consumen entre comidas" y nos gustaría contar con su valiosa colaboración. La información que nos proporcione será manejada en una forma estrictamente confidencial y utilizada sólo para fines estadísticos.*

---

**Carátula - Información del Entrevistado**

Sexo		Edad Años Cumplidos	
1. Hombre	<input style="width: 20px; height: 20px; border: 1px solid black;" type="text"/>	1. 9 a 13	<input style="width: 20px; height: 20px; border: 1px solid black;" type="text"/>
2. Mujer	<input style="width: 20px; height: 20px; border: 1px solid black;" type="text"/>	2. 14 a 18	<input style="width: 20px; height: 20px; border: 1px solid black;" type="text"/>
		3. 19 a 25	<input style="width: 20px; height: 20px; border: 1px solid black;" type="text"/>
		4. 26 a 35	<input style="width: 20px; height: 20px; border: 1px solid black;" type="text"/>
		5. 36 a 45	<input style="width: 20px; height: 20px; border: 1px solid black;" type="text"/>

---

**Sección 0. Filtros**

**F1. ¿Vive usted en esta ciudad?**

0. No  → **Terminar Entrevista**

1. Sí

**F2. ¿Cuánto tiempo lleva viviendo en esta ciudad?**

1. Menos de 6 meses  → **Terminar Entrevista**

2. 6 meses o más

**F3. ¿Trabaja usted o alguien de su familia para alguna de las siguientes compañías?**

<b>RM</b>	
<input style="width: 20px; height: 20px; border: 1px solid black;" type="text"/>	
<input style="width: 20px; height: 20px; border: 1px solid black;" type="text"/>	→ <b>Terminar Entrevista</b>
<input style="width: 20px; height: 20px; border: 1px solid black;" type="text"/>	
<input style="width: 20px; height: 20px; border: 1px solid black;" type="text"/>	
<input style="width: 20px; height: 20px; border: 1px solid black;" type="text"/>	

0. No trabaja en ninguna

1. Agencia de publicidad

2. Agencia de promociones

3. Agencia de investigación de mercados

4. Compañía de venta o distribución de golosinas y/o botanas tales como chocolates, dulces, pastillas refrescantes, papas fritas, etc.

*Vamos a pensar en productos que se consuman entre comidas, para calmar el hambre o para darse un gusto.*

F4. Pensando en la última semana, podría decirme, ¿qué productos tipo golosina y/o botana ha consumido usted fuera de las comidas principales? ¿Algún otro? ¿Qué otro?  
*(Atn. Investigador: Anotar primera respuesta en columna de 1a. Mención y otras respuestas en columna de Otras Menciones.)*

F5. Y, ¿en la última semana consumió ...?  
*(Atn. Investigador: Preguntar por productos No mencionados en pregunta anterior.)*

	F4.		F5.
	1a. Mención	Otras Menciones	Ayudado
		RM	RM
99. Ninguno/Ningún otro	99	99	99
98. No recuerda	98	98	
01. Botanas (papas fritas, cacahuates, palomitas, etc.)	01	01	01
02. Chocolates	02	02	02
03. Galletas	03	03	03
04. Dulces	04	04	04
05. Pastillas refrescantes	05	05	05
06. Chicles	06	06	06
07. Pastelitos envueltos en celofán	07	07	07
08. Helados	08	08	08
09. Untables	09	09	09
10. Otros. <i>(Especificar)</i>	10	10	10
_____	<input type="text"/>		
_____		<input type="text"/>	
_____		<input type="text"/>	

*(Atn. Investigador: Aplicar siguiente pregunta sólo si mencionó Ninguno en las dos preguntas anteriores.)*

F6. ¿Por qué razones no consumió ningún producto fuera de comidas?

	RM
1. Estoy a dieta	1
2. Problemas de salud	2
3. No como entre comidas	3
4. Otras. <i>(Especificar)</i>	4
_____	<input type="text"/>
_____	<input type="text"/>
_____	<input type="text"/>

*(Atn. Investigador: Si aplicó pregunta anterior Terminar Entrevista.)*

**Sección I. Consumo de golosinas entre comidas**

(*Atn. Investigador: Aplicar esta sección si mencionó Dulces en F4. ó F5., si no mencionó Dulces pasar a P6.*)

**DULCES**

P1. ¿Qué tipo de dulces consumió?

_____	<input type="text"/>	<input type="text"/>	<input type="text"/>
_____	<input type="text"/>	<input type="text"/>	<input type="text"/>
_____	<input type="text"/>	<input type="text"/>	<input type="text"/>
_____	<input type="text"/>	<input type="text"/>	<input type="text"/>
_____	<input type="text"/>	<input type="text"/>	<input type="text"/>

P2. ¿Con qué frecuencia acostumbra consumir dulces?

- |                                   |                      |                      |                      |
|-----------------------------------|----------------------|----------------------|----------------------|
| 01. Varias veces al día           | <input type="text"/> | <input type="text"/> | <input type="text"/> |
| 02. Diario                        | <input type="text"/> | <input type="text"/> | <input type="text"/> |
| 03. 4 ó 6 veces a la semana       | <input type="text"/> | <input type="text"/> | <input type="text"/> |
| 04. 2 ó 3 veces a la semana       | <input type="text"/> | <input type="text"/> | <input type="text"/> |
| 05. 1 vez a la semana             | <input type="text"/> | <input type="text"/> | <input type="text"/> |
| 06. Cada 15 días                  | <input type="text"/> | <input type="text"/> | <input type="text"/> |
| 07. 1 vez al mes                  | <input type="text"/> | <input type="text"/> | <input type="text"/> |
| 08. 1 vez cada 2 ó 3 meses        | <input type="text"/> | <input type="text"/> | <input type="text"/> |
| 09. Rara vez                      | <input type="text"/> | <input type="text"/> | <input type="text"/> |
| 10. Otros. ( <i>Especificar</i> ) | <input type="text"/> | <input type="text"/> | <input type="text"/> |

\_\_\_\_\_

P3. Hablando de la última vez en que consumió dulces, ¿qué lo motivó a consumirlos?

- |                                      |                      |
|--------------------------------------|----------------------|
|                                      | <b>RM</b>            |
| 1. Antojo/Ganas                      | <input type="text"/> |
| 2. Consentirme/Mimarme               | <input type="text"/> |
| 3. Tenía un vacío en el estómago     | <input type="text"/> |
| 4. Quitarme el aburrimiento          | <input type="text"/> |
| 5. Necesitaba un momento de descanso | <input type="text"/> |
| 6. Premiarme                         | <input type="text"/> |
| 7. Otros. ( <i>Especificar</i> )     | <input type="text"/> |

\_\_\_\_\_

P4. ¿En qué momento del día consumió dulces la última vez?

- |                                  |   |
|----------------------------------|---|
| 1. Antes del desayuno            | 1 |
| 2. En el desayuno                | 2 |
| 3. Entre el desayuno y la comida | 3 |
| 4. En la comida                  | 4 |
| 5. Entre la comida y la cena     | 5 |
| 6. En la cena                    | 6 |
| 7. Después de la cena            | 7 |

P5. ¿Qué marca de dulces consumió la última vez?

\_\_\_\_\_ 

--	--	--

- |                  |     |
|------------------|-----|
| 999. No recuerda | 999 |
| 998. Sin marca   | 998 |
| 997. A granel    | 997 |

*A continuación le voy a leer una serie de frases sobre algunas razones del porqué se consumen algunos productos.*

P6. ¿Cuál de los productos que aparecen en esta tarjeta cree usted que ...?  
 (Mostrar Tarjeta)                      (Rotar Conceptos)                      (Anotar Sólo un Producto)

- |   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|---|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
| 1. Da energía                                   | <table border="1" style="width: 100%; height: 100%;"><tr><td style="width: 20px; height: 15px;"></td><td style="width: 20px; height: 15px;"></td></tr><tr><td style="width: 20px; height: 15px;"></td><td style="width: 20px; height: 15px;"></td></tr></table> |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2. Es el más rico                               |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 3. Es para toda la familia                      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 4. Nos pone alegres y contentos                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 5. Es para consumirlo en un momento de descanso |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 6. Es para niños                                |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 7. Es para acompañar un café                    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 8. Es para tener siempre con usted              |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 9. Quita el mal aliento                         |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

**Lista de Códigos de Productos**

- 01. Botanas
- 02. Chocolates
- 03. Galletas
- 04. Dulces
- 05. Pastillas refrescantes
- 06. Chicles
- 07. Pastelitos envueltos en celofán
- 08. Helados
- 09. Untables
- 99. Ninguna/No contestó



**Sección III. Hábitos de consumo y compra de chocolates**

**P8.** ¿Con qué frecuencia acostumbra consumir chocolates?

- |                                   |    |
|-----------------------------------|----|
| 01. Varias veces al día           | 01 |
| 02. Diario                        | 02 |
| 03. 4 ó 6 veces a la semana       | 03 |
| 04. 2 ó 3 veces a la semana       | 04 |
| 05. 1 vez a la semana             | 05 |
| 06. Cada 15 días                  | 06 |
| 07. 1 vez al mes                  | 07 |
| 08. 1 vez cada 2 ó 3 meses        | 08 |
| 09. Rara vez                      | 09 |
| 10. Otros. ( <i>Especificar</i> ) | 10 |

\_\_\_\_\_ 

--	--	--

**P9.** Regularmente, ¿quién compra el chocolate que usted consume?

- |                                   |           |
|-----------------------------------|-----------|
|                                   | <b>RM</b> |
| 01. Yo mismo (a)                  | 01        |
| 02. Esposo (a)                    | 02        |
| 03. Hijos (as)                    | 03        |
| 04. Padre/Madre                   | 04        |
| 05. Hermanos (as)                 | 05        |
| 06. Novio (a)                     | 06        |
| 07. Amigos (as)                   | 07        |
| 08. Compañeros (as) de oficina    | 08        |
| 09. Compañeros (as) de escuela    | 09        |
| 10. Otros. ( <i>Especificar</i> ) | 10        |

\_\_\_\_\_ 

--	--	--

\_\_\_\_\_ 

--	--	--

**P10.** Hablando de la última vez en que consumió chocolates, ¿qué lo motivó a consumirlos?

- |                                      |           |
|--------------------------------------|-----------|
|                                      | <b>RM</b> |
| 1. Antojo/Ganas                      | 1         |
| 2. Consentirme/Mimarme               | 2         |
| 3. Tenía un vacío en el estómago     | 3         |
| 4. Quitarme el aburrimiento          | 4         |
| 5. Necesitaba un momento de descanso | 5         |
| 6. Premiarme                         | 6         |
| 7. Otros. ( <i>Especificar</i> )     | 7         |

\_\_\_\_\_ 

--	--	--

\_\_\_\_\_ 

--	--	--

P11. ¿En qué momento del día consumió chocolate la última vez?

- |                                  |                          |
|----------------------------------|--------------------------|
| 1. Antes del desayuno            | <input type="checkbox"/> |
| 2. En el desayuno                | <input type="checkbox"/> |
| 3. Entre el desayuno y la comida | <input type="checkbox"/> |
| 4. En la comida                  | <input type="checkbox"/> |
| 5. Entre la comida y la cena     | <input type="checkbox"/> |
| 6. En la cena                    | <input type="checkbox"/> |
| 7. Después de la cena            | <input type="checkbox"/> |

P12. ¿En compañía de quién consumió chocolate la última vez?

- |                                   |                          |
|-----------------------------------|--------------------------|
|                                   | <b>RM</b>                |
| 01. Sólo                          | <input type="checkbox"/> |
| 02. Esposo (a)                    | <input type="checkbox"/> |
| 03. Hijos (as)                    | <input type="checkbox"/> |
| 04. Padre/Madre                   | <input type="checkbox"/> |
| 05. Hermanos (as)                 | <input type="checkbox"/> |
| 06. Novio (a)                     | <input type="checkbox"/> |
| 07. Amigos (as)                   | <input type="checkbox"/> |
| 08. Compañeros (as) de oficina    | <input type="checkbox"/> |
| 09. Compañeros (as) de escuela    | <input type="checkbox"/> |
| 10. Otros. ( <i>Especificar</i> ) | <input type="checkbox"/> |

\_\_\_\_\_

\_\_\_\_\_

P13. ¿Qué marca de chocolate consumió la última vez?

\_\_\_\_\_

- |                  |                          |
|------------------|--------------------------|
| 999. No recuerda | <input type="checkbox"/> |
| 998. Sin marca   | <input type="checkbox"/> |
| 997. A granel    | <input type="checkbox"/> |

P14. ¿Acostumbra regalar chocolates?

- |       |                          |                       |
|-------|--------------------------|-----------------------|
| 0. No | <input type="checkbox"/> | → Terminar Entrevista |
| 1. Sí | <input type="checkbox"/> |                       |

P15. ¿Qué lo motiva a regalar chocolates?

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

P16. ¿Con qué frecuencia acostumbra comprar chocolates para regalarlos?

- |                               |   |
|-------------------------------|---|
| 1. Más de una vez a la semana | 1 |
| 2. Una vez a la semana        | 2 |
| 3. Cada 15 días               | 3 |
| 4. Cada mes                   | 4 |
| 5. Cada 2 ó 3 meses           | 5 |
| 6. Cada 4 ó 6 meses           | 6 |
| 7. Cada 6 a 12 meses          | 7 |
| 8. Cada 12 meses o más        | 8 |

P17. Generalmente, ¿a quién le regala chocolates?

- |                                   |           |
|-----------------------------------|-----------|
|                                   | <b>RM</b> |
| 01. Esposo (a)                    | 0 1       |
| 02. Hijos (as)                    | 0 2       |
| 03. Padre/Madre                   | 0 3       |
| 04. Hermanos (as)                 | 0 4       |
| 05. Novio (a)                     | 0 5       |
| 06. Amigos (as)                   | 0 6       |
| 07. Compañeros (as) de oficina    | 0 7       |
| 08. Compañeros (as) de escuela    | 0 8       |
| 09. Maestros (as)                 | 0 9       |
| 10. Otros. ( <i>Especificar</i> ) | 1 0       |

\_\_\_\_\_

\_\_\_\_\_

P18. ¿Para qué ocasiones acostumbra regalar chocolates?

- |                                   |           |
|-----------------------------------|-----------|
|                                   | <b>RM</b> |
| 0. Ninguna en especial            | 0         |
| 1. Cumpleaños                     | 1         |
| 2. Día de la madre                | 2         |
| 3. Día del padre                  | 3         |
| 4. Día del maestro (a)            | 4         |
| 5. Día de la amistad/San Valentin | 5         |
| 6. Navidad                        | 6         |
| 7. Aniversarios                   | 7         |
| 8. Otras. ( <i>Especificar</i> )  | 8         |

\_\_\_\_\_

\_\_\_\_\_

P19. Aproximadamente, ¿cuánto gasta cada vez que regala chocolates?

\$

P20. ¿Qué marca de chocolates acostumbra regalar?

\_\_\_\_\_

- |                  |     |
|------------------|-----|
| 999. No recuerda | 999 |
| 998. Sin marca   | 998 |
| 997. A granel    | 997 |

<b>Información del Entrevistado</b>	
Nombre	_____
	_____
<b><i>Dirección:</i></b>	
Calle	_____
Entre	_____ y _____
Colonia	_____ Delegación/Municipio _____
C.P.	_____ Teléfono _____
	Hora Final <input type="text"/> <input type="text"/> : <input type="text"/> <input type="text"/>
<b>Con esto terminamos la entrevista, muchas gracias por su cooperación y su tiempo, que tenga buen día.</b>	
Yo _____ declaro haber llenado este cuestionario con la información que el entrevistado, cuyo nombre aparece arriba, me proporcionó.	
Plaza	_____ a _____ Día _____ de _____ Mes _____ de 2002.
Firma	_____
Supervisor (a)	_____ Firma _____
Estuvo presente por el cliente:	
Nombre	_____ Firma _____