

8
25

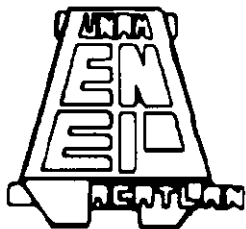


UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

ESCUELA DE ESTUDIOS PROFESIONALES
"ACATLÁN"

ANÁLISIS DISCRIMINANTE Y LA REGRESIÓN LOGÍSTICA EN EL ESTUDIO DEL PERFIL DE USUARIOS DE SERVICIOS BANCARIOS

T E S I S
QUE PARA OBTENER EL TÍTULO DE:
A C T U A R I O
P R E S E N T A :
PATRICIA TORRES SANTILLÁN



ASESOR DE TESIS: JAVIER GONZALEZ ROSAS



270174
FEBRERO 1999.

TESIS CON
ALLA DE ORIGEN



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

ÍNDICE

Introducción

I.	Perfil de usuarios	Pág
I.1	¿Qué es el mercado?	5
I.2	Segmentación del mercado	13
I.3	Características de los usuarios de servicios bancarios	28
II.	Modelos estadísticos y perfil de usuarios	
2.1	Modelos explicativos	29
2.2	Análisis discriminante	34
2.3	Regresión logística simple	55
2.4	Comparación de modelos	65
III.	Aplicación de los modelos	
3.1	Descripción de datos	66
3.2	Estimación de parámetros en análisis discriminante	73
3.3	Estimación de parámetros en regresión logística	80
3.4	Resultados	83
	Conclusiones	97
	Anexo I	101
	Anexo II	106
	Anexo III	110
	Bibliografía	117

PROLOGO

El mundo en que vivimos es muy complejo, y tratar de explicar lo que ocurre a nuestro alrededor siempre ha sido preocupación constante de todos los seres humanos. Así, el modelo religioso ha ofrecido una explicación a todo lo que nos rodea postulando que todo sucede por obra y gracia de un creador de todo lo visible e invisible.

Otro enfoque es el mundo de las probabilidades. Según este enfoque, todo en la vida está gobernado por leyes de probabilidad; de tal manera que, existen eventos cuya probabilidad de ocurrencia es muy alta y por eso suceden casi a diario, mientras que otros tienen una probabilidad muy pequeña de ocurrir y debido a esto casi no los vemos ocurrir a nuestro alrededor.

Los eventos que tienen poca probabilidad de ocurrir han sido interpretados a través de la historia de la humanidad de diferentes maneras. Algunos los asocian con milagros, pero en realidad se debe a que tienen una probabilidad de ocurrencia muy pequeña. Otros los han asociado con la suerte y cuando ocurren esos eventos con probabilidad pequeña, lo atribuyen a la buena o mala suerte.

Todos los seres humanos tenemos que fallecer algún día, tarde o temprano este evento tendrá que ocurrir, ya que hasta ahora no se sabe de alguien que viva eternamente. Sin embargo, esa probabilidad de un momento a otro, puede incrementar o decrementar; por ejemplo, la gente que vive a la orilla de un río, tiene una probabilidad más alta de morir que los que viven alejados del río, ya que, aunque la probabilidad de que se desborde el río puede ser pequeña, cuando ocurra el desbordamiento es más probable que mueran los que están más cerca del río que los que están lejos y esto no tiene nada que ver ni con castigos divinos, ni con mala suerte ni con nada por el estilo.

En todo esto, quizá el enfoque probabilista puede ser más real o menos contradictorio, ya que, de acuerdo a las probabilidades es factible que mueran al desbordarse el río, niños inocentes, pero el enfoque religioso se contradice, ya que la muerte de niños inocentes se debe al desbordamiento del río, pero el desbordamiento es obra y gracia del creador, por lo que entonces, la muerte de esos niños ¿será obra y gracia de él?

La investigación que presenta Paty en esta tesis, es la aplicación de modelos derivados de la teoría de probabilidades y descansan en el hecho de que todos tenemos una probabilidad de ser usuarios de algún banco, sin embargo, esa probabilidad es más grande o más pequeña dependiendo de ciertas características socio-demográficas que los modelos permiten identificar, pero, después de todo ¿será decisión de algún ser superior el tipo de banco que usamos?, yo creo que no, ¿y usted?

Xavier González Rosas
Director de tesis

Hoy en día, ninguna empresa puede operar en todos los mercados ni satisfacer todas las necesidades de los diferentes usuarios. Tampoco puede realizar una buena labor dentro de un mercado extenso. En este sentido, las empresas funcionan mejor cuando definen cuidadosamente sus mercados objetivo y, realizan adecuadamente su labor al preparar un programa de mercadotecnia a la medida de cada mercado meta, definiendo las necesidades del cliente, desde el punto de vista de éste, no desde la perspectiva de la empresa.

Todo producto implica alternativas mediadas, y la administración no puede saber cuáles, sin hablar ni investigar con los clientes. Así, al comprador de un auto le gustaría uno de gran rendimiento, que nunca falle, que sea seguro, de línea atractiva y de bajo costo. Puesto que todas estas características no pueden combinarse en un solo automóvil, los diseñadores deben hacer elecciones rigurosas, no con base en lo que a ellos les gusta, sino en lo que el cliente prefiere o espera. Esto es importante, debido a que las ventas de una empresa provienen de dos grupos: los nuevos clientes y los clientes asiduos.

Siempre cuesta más atraer a nuevos compradores que retener a los clientes asiduos. Por tanto, la conservación del cliente es más importante que la atracción de éste, ya que un cliente satisfecho:

- compra una y otra vez;
- habla favorablemente de la compañía a otros;
- presta menos atención a las marcas y publicidad de la competencia; y
- adquiere productos adicionales de la misma empresa.

Por ello, es conveniente que la empresa mida regularmente la satisfacción de éstos. No puede confiar sólo en las quejas voluntarias de los clientes cuando están insatisfechos, ya que la mayoría de éstos nunca lo manifiestan a la empresa, por lo tanto, es importante establecer sistemas de sugerencias para maximizar la oportunidad del cliente para quejarse. Pero escuchar no es suficiente, la empresa debe responder constructivamente a las quejas.

El propósito de la mercadotecnia es ayudar a las organizaciones a alcanzar sus metas. En las empresas privadas, la meta principal son las utilidades; en el caso de las

organizaciones públicas y no lucrativas, es sobrevivir y recabar fondos suficientes para desempeñar sus funciones. Ahora bien, la clave no es lograr utilidades como tales, sino conseguirlas como consecuencia de haber realizado un buen trabajo.

La investigación de mercados relaciona al consumidor, al cliente y al público con el especialista de mercadeo a través de la información, la cual utiliza para:

- identificar y definir las oportunidades y los problemas de mercadeo;
- generar, perfeccionar y evaluar las acciones de mercadeo;
- monitorear el desempeño del mercadeo; y
- mejorar la comprensión del mercado como un proceso.

Para esto, es necesario conocer el mercado al cual se atiende; es decir, conocer las necesidades y deseos del consumidor, así como también sus características sociales, económicas, culturales, etcétera. Esta tesis se enfoca en algunas metodologías que se pueden aplicar para identificar las características socioeconómicas de los clientes; en particular, de los usuarios de bancos, para así saber por ejemplo, quiénes pueden ser los clientes potenciales de algún servicio o producto, y a través de qué medios informativos se puede llegar a ellos.

El trabajo consta de 3 capítulos; en el capítulo I, se definen conceptos que permiten entender el objetivo y la función de la mercadotecnia, tales como *necesidades, deseos, demandas, productos, intercambio, transacciones, relaciones, etcétera*. Asimismo, se explica en que consiste la segmentación del mercado y su importancia en la mercadotecnia, se describen las formas fundamentales de segmentación, además de tratar sobre la utilidad de mapas perceptuales para conocer las características del mercado objetivo. En el segundo se aborda las bases teóricas de los modelos estadísticos utilizados en este trabajo, a saber, análisis discriminante y regresión logística y, el último capítulo trata sobre la aplicación de éstos en la determinación de las características de los usuarios bancarios, el análisis de los resultados obtenidos, el procedimiento de validación de los modelos; es decir, la manera de medir el grado de error que se comete al usar los resultados del modelo en la explicación de la realidad y, por último las conclusiones que se obtienen.

CAPÍTULO I

PERFIL DE USUARIOS

1.1 ¿ QUÉ ES EL MERCADO ?

Necesidades, deseos y demandas

El punto de partida de la mercadotecnia, radica en las necesidades y deseos humanos. La gente necesita para sobrevivir alimento, aire, agua, vestido y albergue. Además tiene un fuerte deseo de diversión, educación y otros servicios, así como marcadas preferencias de bienes y servicios básicos.

Es conveniente hacer una distinción entre *las necesidades, los deseos y las demandas*. La *necesidad* humana es el estado en el que se siente la carencia de algunos satisfactores básicos. Es decir, la gente requiere de alimento, vestido, abrigo, seguridad, sentido de pertenencia, estimación y algunas otras cosas para sobrevivir. Estas necesidades no han sido creadas por la sociedad o por los comerciantes, sino que forman parte integral del proceso biológico y de la condición del ser humano.

Los *deseos* son los satisfactores específicos para estas necesidades profundas. Es decir, una persona necesita alimento y se le antoja una hamburguesa; necesita un vestido y quiere un traje Pierre Cardin; necesita, estima y compra un Cadillac. En otras sociedades estas necesidades se satisfacen de forma diferente; por ejemplo, los balineses¹ satisfacen su hambre con mangos; sus necesidades de vestido, con un taparrabo y sus necesidades de estima, con un collar de conchas.

En tanto que las necesidades son pocas, los deseos son muchos y éstos continuamente están siendo modelados y remodelados por fuerzas comerciales e instituciones tales como iglesias, escuelas, familias y corporaciones comerciales.

¹ Los balineses son originarios de una isla de Indonesia, entre Java y Lombok, 5,561 Km²; 2,120 hab. Cap Singaraja.

Las **demandas** son los requerimientos de productos específicos que están respaldados por la capacidad y voluntad de adquirirlos. Los deseos se tornan en exigencias cuando están respaldados por el poder adquisitivo. Mucha gente quiere un Cadillac, pero sólo unas cuantas quieren y pueden comprarlo. Por lo tanto, las empresas no sólo deben medir cuántas personas quieren sus productos, sino también, lo que es aún más importante, cuántas realmente desearían y podrían comprarlo.

Estas distinciones aclaran las frecuentes críticas que se le hace a la mercadotecnia en el sentido de que "los mercadólogos crean necesidades" o de que "los mercadólogos hacen que la gente compre cosas que no quiere". Éstos no crean necesidades; las necesidades son anteriores a ellos. Los mercadólogos, conjuntamente con otras influencias en la sociedad, influyen solamente en los deseos. Ellos son los que sugieren al consumidor que un Cadillac podría satisfacer las necesidades de *status* social de una persona. No crean el *status*, solamente tratan de señalar que un producto específico satisfecería esa necesidad. Los mercadólogos tratan de influir en la demanda, haciendo al producto atractivo, accesible y de fácil disposición.

Productos

La gente satisface sus necesidades y deseos, con productos; se pueden definir entonces a los productos en general, **como todo aquello que puede ofrecerse a alguien para satisfacer una necesidad o deseo**. Normalmente la palabra producto trae a la mente un objeto físico, como un automóvil, un televisor o una bebida gaseosa y, por lo general, empleamos la expresión productos y servicios, para distinguir entre los objetos físicos y los intangibles. Al pensar en los productos físicos, su importancia descansa, no tanto en su posesión, sino en su empleo para satisfacer nuestros deseos: es decir, no compramos un auto sólo para mirarlo, sino para poder transportarnos; no compramos un horno de microondas para admirarlo, sino para poder cocinar en él. Así, los productos físicos son realmente vehículos que nos proporcionan servicios.

Por lo demás, los servicios también están suministrados por otros medios, tales como personas, lugares, actividades, organizaciones e ideas. Si estamos aburridos, podemos ir a un centro nocturno a ver a un cómico (persona); viajar a algún centro vacacional como Cancún (lugar); practicar algún ejercicio físico (actividad); afiliarnos a un club de

corazones solitarios (organización); o adoptar otra filosofía sobre la vida (idea). En otras palabras, los servicios pueden realizarse mediante objetos físicos y otros medios. Por lo tanto, utilizaremos el término producto; para referirnos a los productos físicos, a los servicios y otros vehículos capaces de satisfacer un deseo o una necesidad. Ocasionalmente utilizaremos otros términos para el producto, como ofertas, satisfactores o recursos.

Los fabricantes se crean muchas dificultades al prestar mayor atención a los productos físicos que a los servicios que éstos otorgan. Los comerciantes aman sus productos, pero olvidan que el cliente los adquiere porque satisfacen una necesidad. La gente no compra los productos físicos por ellos mismos. Un lápiz labial se adquiere para que proporcione un servicio: ayudar a que la persona luzca mejor. Un taladro se adquiere para que proporcione un servicio: hacer el orificio que se necesita. Un objeto físico es el medio en el que se aloja un servicio. **La función del experto en mercadotecnia es: más que la descripción de sus características, la venta de los beneficios o servicios contenidos en los productos físicos.** Los vendedores que se concentran en el producto, en vez de las necesidades del cliente, se dice que sufren de "...opía de mercadotecnia".

Intercambio, transacciones y relaciones

El hecho de que la gente tenga necesidades y deseos, y pueda dar un valor a los productos, no define totalmente la mercadotecnia. Ésta surge cuando la gente decide satisfacer sus necesidades y deseos a través del intercambio. Existen cuatro maneras mediante las cuales la gente puede obtener los productos que desea.

La primera forma es la **autoproducción**. La gente puede aliviar su hambre, cazando, pescando y recolectando fruta. No requiere interacción con nadie más. En este caso no hay mercado ni mercadotecnia.

La segunda forma es la **coacción**. La gente hambrienta puede arrebatar o robar alimentos a otros. No se ofrece ningún beneficio a los demás, excepto el de no ser dañados.

La tercera forma es la **mendicidad**. La gente hambrienta puede acercarse a otros y suplicarles que les den alimentos. No tienen nada tangible que ofrecer, excepto gratitud.

La cuarta manera es el **intercambio**. La gente hambrienta puede acercarse a otros y ofrecerles algún recurso a cambio, como dinero, un producto o algún servicio.

La mercadotecnia surge de esta última forma de adquisición de productos y, se entiende al intercambio como el acto de obtener de alguien un producto deseado, ofreciendo algo a cambio, es el concepto determinante en el que se sustenta la mercadotecnia. Para que tenga lugar el intercambio, deben darse cinco condiciones:

1. Que existan al menos dos partes.
2. Que cada parte posea algo que pueda tener valor para la otra parte.
3. Que cada parte sea capaz de comunicarse y hacer entrega.
4. Que cada parte tenga libertad para aceptar o rechazar la oferta.
5. Que cada parte considere que es apropiado o deseable negociar con la otra parte.

Si se dan estas condiciones, hay posibilidades potenciales para el intercambio. Para que éste realmente tenga lugar, ambas partes deben convenir en los términos del intercambio, lo que les permitiría quedar en mejores condiciones (o al menos no peores) que las que tenían antes del intercambio. Éste es el sentido por el cual el intercambio se describe como un proceso de creación de valores; es decir, normalmente hace que ambas partes queden en mejor situación que la que tenían antes de realizarlo.

El intercambio debe verse como un proceso más que como un acontecimiento. Se dice que dos partes están comprometidas en el intercambio si hacen negociaciones y si se encaminan a un acuerdo. Si se llega a un acuerdo, decimos que tiene lugar una transacción, la cual se define como la unidad básica del intercambio y consiste en el comercio de valores entre dos partes. Podemos decir entonces que A dio X a B, y recibió Y a cambio; Juan Acevedo dio \$400 a Raúl Herrera y obtuvo un televisor. Ésta es la clásica transacción monetaria. Sin embargo, las transacciones no siempre requieren dinero como uno de los valores comercializados. Una transacción de trueque consistiría en que Juan diera un refrigerador a Raúl a cambio de un televisor; también podría consistir en la comercialización de servicios en lugar de mercancías, como cuando el notario Acevedo hace un testamento para el Dr. Herrera a cambio de un examen médico.

Una transacción comprende varias dimensiones: al menos dos objetos de valor, acuerdo sobre las condiciones, un tiempo convenido y un lugar establecido. Por lo común surge un sistema legal para apoyar y dar cumplimiento a estas negociaciones. Las transacciones pueden fácilmente dar lugar a conflictos basados en la mala interpretación o la mala fe. Sin una "legislación por contratos", la gente realizaría las transacciones con cierta desconfianza y todos saldrían perdiendo.

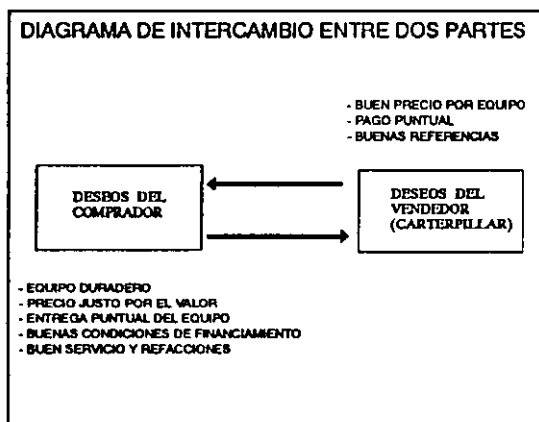
Una transacción difiere de una transferencia. En una transferencia A da a B un obsequio, un subsidio o una contribución de beneficencia, a lo cual llamamos transferencia y no transacción. Sin embargo, el comportamiento de la transferencia puede también entenderse a través del concepto de intercambio. Generalmente la transferencia contiene ciertas expectativas al otorgarse un obsequio, tales como obtener gratitud o buscar el buen comportamiento del receptor. Los recaudadores profesionales de fondos están conscientes de los motivos de "reciprocidad" que subyacen en la conducta del donador, y tratan de proporcionarle beneficios, como notas de agradecimiento, revistas e invitaciones especiales a eventos. Los mercadólogos han ampliado recientemente el concepto de mercadotecnia, incluyendo el estudio de la conducta de transferencia y de la conducta de transacción.

En el sentido más genérico, el mercadólogo busca provocar una respuesta conductual de la otra parte. Una empresa comercial quiere una respuesta llamada compra, un candidato político quiere una respuesta llamada voto, una iglesia quiere una respuesta llamada unión, un grupo de acción social quiere una respuesta llamada adopción de la idea. ***La mercadotecnia consiste en tratar de provocar las respuestas deseadas de un público meta, hacia algún objeto.***

Para efectuar intercambios de éxito, el mercadólogo analiza lo que cada parte espera dar y recibir. El intercambio simple puede ilustrarse mostrando a los dos actores, los deseos y ofrecimientos que fluyen entre ellos. Supongamos por ejemplo que Caterpillar, el fabricante más grande del mundo de maquinaria pesada, investiga los beneficios que una compañía constructora desea obtener al adquirir un trascabo, los cuales aparecen en el diagrama de intercambio en la figura 1. La constructora quiere un equipo de alta calidad, precio justo, entrega puntual, buenas condiciones de financiamiento y un buen servicio. No todos los deseos tienen la misma importancia y pueden variar de un comprador a otro. Una de las tareas de Caterpillar es descubrir la importancia de los diversos deseos

del comprador. Al mismo tiempo, éste tiene una lista de deseos. Caterpillar quiere un buen precio por el equipo, pago puntual y buena publicidad verbal. Si hay la concordancia suficiente en las listas de deseos, existen las bases para una transacción. La labor de Caterpillar es formular una oferta y motivar a la compañía constructora para que compre su equipo. La constructora podría a su vez hacer una contraoferta. El proceso de tratar de llegar a términos convenientes para ambas partes recibe el nombre de negociación. La **negociación** conduce, ya sea hacia términos aceptables para ambas partes, o a la decisión de no hacer la transacción.

FIGURA 1



Fuente: Dirección de la Mercadotecnia, Análisis, Planificación y Control, pág. 9.

Hasta aquí se ha descrito la naturaleza de la transacción en mercadotecnia. Sin embargo cabe resaltar que ésta, es parte de una idea mayor, la de las relaciones de mercadotecnia. Los mercadólogos hábiles tratan de establecer relaciones a largo plazo, confiables y seguras con clientes, distribuidores, comerciantes y proveedores. Esto se logra prometiendo y entregando a la otra parte alta calidad, buen servicio y precios justos todo el tiempo. Se consigue fortaleciendo los lazos económicos, técnicos y sociales entre los miembros de las dos organizaciones. Ambas partes desarrollan más confianza, un mejor conocimiento y un mayor interés en ayudarse entre sí. Las relaciones de mercadotecnia reducen los costos y el tiempo de la transacción; en el mejor de los casos, la transacción pasa, de ser negociada cada vez, a una simple rutina.

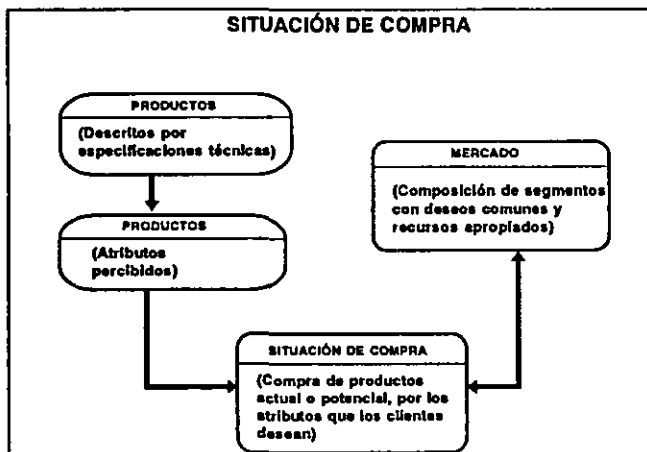
El resultado final de las relaciones de mercadotecnia es la estructuración de un bien único de la empresa, llamado **red de mercadotecnia**. Una red de mercadotecnia está formada por la empresa y compañías con las cuales se han establecido relaciones comerciales sólidas y confiables. Cada vez más la mercadotecnia está tratando de sustituir la maximización de utilidades en cada transacción individual, por la maximización de las relaciones provechosas con las otras partes. El principio de operación es: **establece buenas relaciones y las transacciones redituables enseguida se darán.**

Mercados

El intercambio conduce necesariamente al concepto de mercado, el cual se define como **todos los clientes potenciales que comparten una necesidad o deseo específico y que podrían estar dispuestos o tienen la capacidad de realizar un intercambio para satisfacer esa necesidad o deseo.** Así, el tamaño del mercado depende del número de personas que muestren la necesidad, tengan los recursos que interesen a otros y estén dispuestos a ofrecer dichos recursos en intercambio para satisfacer esa necesidad o deseo.

El grado de interés por un producto, depende de la proporción en que satisface las necesidades del consumidor. Como se muestra en la figura 2, si el producto satisface ampliamente alguna necesidad entonces habrá como resultado una compra.

FIGURA 2



Fuente: *Ibidem*, pág. 10

Mercadotecnia

El concepto de mercado nos remite directamente al concepto de mercadotecnia. La mercadotecnia es una actividad humana que está relacionada con los mercados, implica trabajar con éstos para actualizar intercambios potenciales, con el objeto de satisfacer necesidades y deseos humanos y, ***se define como un proceso social y administrativo mediante el cual, grupos e individuos obtienen lo que necesitan y desean, a través de la creación y el intercambio de productos; tratando de actualizar los mercados potenciales.***

El interés por la mercadotecnia se intensifica en la medida en que más organizaciones del sector negocios, del sector no lucrativo y, del sector internacional reconocen la forma en que la mercadotecnia contribuye a un mejor desempeño en el mercado. Debido a que en la actualidad las empresas no pueden sobrevivir por el simple hecho de realizar un buen trabajo; para tener éxito en los mercados locales y extranjeros, que se caracterizan por un crecimiento lento y una fuerte competencia, por lo cual deben realizar una excelente labor. Los consumidores y compradores de las empresas tienen ante sí muchos proveedores para elegir los satisfactores de sus necesidades y, por consiguiente, buscan excelencia en la calidad, el valor o el costo cuando eligen a sus proveedores. Estudios recientes han demostrado que la clave para una operación rentable de la empresa es el conocimiento y satisfacción de los clientes con ofertas competitivas superiores, en donde la función de la mercadotecnia es, definir los clientes objetivo y la mejor forma de satisfacer sus necesidades y deseos de una manera competitiva y redituable.

El concepto de mercadotecnia sostiene que la clave para alcanzar las metas organizacionales, consiste en determinar las necesidades y deseos de los mercados objetivo, entregar los satisfactores deseados de forma más eficaz y eficiente que los competidores, para lograr esto se requiere de tres etapas:

- 1) La segmentación del mercado.
- 2) Establecimiento de metas de mercado, y
- 3) El posicionamiento del producto.

I.2 SEGMENTACIÓN DEL MERCADO

Qué es la segmentación

El mercado de casi todos los productos no es una masa amorfa de personas o empresas, sino una serie de subgrupos con diferentes necesidades y deseos. La **segmentación del mercado** es el proceso de dividir un mercado heterogéneo en segmentos más o menos homogéneos e identificables con el fin de crear una mezcla de comercialización² que satisfaga las necesidades de los clientes en los segmentos interesantes para la compañía.

Hoy en día, los fabricantes de automóviles reconocen la existencia de infinidad de segmentos que, en su conjunto, constituye el mercado automotriz. Hay varios segmentos: el de los amantes de la velocidad, las madres con niños chicos, los matrimonios jóvenes, el de aquellos que buscan prestigio social e incluso los estados de la Unión Americana (por ejemplo, California tiene estrictas normas sobre la contaminación ambiental que exigen ciertas modificaciones en el motor).

La segmentación les ayuda a escoger los objetivos mercadológicos, porque entraña el examen de varias oportunidades de beneficios y de participación en el mercado. La segmentación representa, pues, el primer paso en el proceso de planeación que, en definitiva, decide la forma y dónde se asignarán los recursos mercadológicos.

La segmentación permite definir con precisión a sus consumidores y concentrarse en el estudio y conocimiento de las necesidades de un grupo específico. A la larga les sirve para identificar mejores oportunidades de conseguir beneficios y buenos resultados con la asignación eficiente de los recursos.

² Es decir, los productos que se ofrecerá a cada segmento, dependerá de las características de éstos.

Mercadotecnia indiferenciada

Hay empresas que no segmentan sus mercados; las que optan por esta estrategia realizan una **mercadotecnia indiferenciada**. Utilizan un producto y una mezcla para todos los segmentos potenciales. No se hace intento alguno por cultivar varios estratos del público consumidor. Hay muchos negocios pequeños y grandes que todavía siguen tal estrategia. Por ejemplo, algunas grandes panificadoras adoptan un método indiferenciado en la venta del pan de caja. Con ello pretenden satisfacer el mayor número posible de necesidades semejantes de los consumidores. Una de las ventajas del mercado indiferenciado es el ahorro en los costos de producción y de mercadotecnia. Como no se elabora más que un artículo, la empresa hace ahorros en la producción masiva. Por otra parte, también los costos de mercadotecnia suelen ser menores, ya que no hay más que un producto que promover y un solo canal de distribución.

Segmentación para la identificación de características

Características socioeconómicas: Existen 4 características estrechamente interrelacionadas en esta categoría: ingreso, ocupación, educación y clase social. Cada una tiene una influencia significativa sobre los patrones de gasto y uso de los usuarios de algún servicio, así como sobre otras características. La educación está relacionada estrechamente con ingreso y ocupación. La ocupación, a la vez, es un componente importante de la clase social, junto con variables como ingreso, área residencial y tipo de vivienda. Una clase social es un grupo con valores, intereses y patrones de comportamiento similares.

Características organizacionales: Así como hay características distintivas que influyen en el comportamiento como compradores y usuarios, así existen organizaciones. Las empresas y algunas instituciones gubernamentales, pueden ser clasificadas de acuerdo al tipo, actividad y tamaño. Asimismo, en términos de ventas, por valor agregado o número de empleados. Otras variables posibles son beneficio y crecimiento, entonces bajo-crecimiento y sin ganancias tienen respuestas muy diferentes a nuevos productos. La fabricación de un producto eléctrico puede clasificar organizaciones, por sus características prácticas de compra.

Los productos o servicios actualmente son adquiridos debido a su utilidad; sin embargo, - esto frecuentemente implica segmentar directamente, de acuerdo a la situación. Por ejemplo, en segmentación de mercado para servicios telefónicos de larga distancia, el tipo de organización puede disfrazar usos diferentes de teléfono.

Métodos para la segmentación de mercados

Si una empresa decide seguir la estrategia de segmentación de mercados puede hacerlo en dos formas. En primer lugar, analiza un objetivo general de mercadeo, en busca de segmentos que ofrezcan oportunidades de beneficio. Hace algunos años General Foods realizó un estudio muy amplio de segmentación sobre el mercado de los alimentos Cycle para perros. También Xerox e IBM llevaron a cabo este tipo de investigación antes de desarrollar productos para los pequeños negocios.

Un segundo método de la segmentación se aplica cuando una compañía ya cuenta con un producto y, sin embargo, examina a sus clientes y a los posibles consumidores a fin de averiguar sus motivos, índice de uso y datos afines. Estos datos sirven después para idear su estrategia de promoción, distribución y fijación de precios, a fin de cumplir con los objetivos de la empresa. Abbot Laboratories logro vender durante años Selsun, un medicamento cuya venta requería receta médica. Selsun contiene 2.5% de solución de sulfuro de selenio, para el control de la caspa. Abbot se dio cuenta de que este artículo era demasiado fuerte para el mercado de consumidores. Previo estudio de segmentación, identificó un grupo que estaba buscando un control más eficaz de la caspa y que era atendido por el líder en el ramo: el champú Head and Shoulders. Basándose en los datos aportados por la investigación, se creó un producto de sulfuro de selenio al 1% y se le puso el nombre Selsun Blue para venta libre. En la actualidad es uno de los champús anticaspa de más venta. Otro ejemplo común de la necesidad de segmentar lo encontramos en la industria de las líneas aéreas. En los estudios de segmentación se basan las nuevas rutas, las tarifas, las frases promocionales, la decoración interior de los aviones, y el aspecto exterior, por citar algunas de las características más conocidas. Braniff, por ejemplo, lleva muchos años considerándose la línea aérea que más utilizan los hombres de negocios. En cambio, Air Jamaica se centra en un segmento totalmente distinto.

La segmentación no siempre significa dividir o fraccionar un mercado total. Si una compañía estudia a sus clientes y los divide por edad o ingresos, obtendrá poca o nula información. La segmentación suele ser un proceso multidimensional e incluye muchas variables y más importante aún: la segmentación consiste a menudo en un proceso de agregar más que de fraccionar. Lo que se pretende es reunir grupos de personas o de compañías que respondan más a determinado aspecto del producto. El proceso de agregación prosigue y se interrumpe cuando han sido identificados todos los compradores potenciales de un artículo. A continuación se analiza al grupo para precisar sus características comunes y saber si hay suficientes compradores para alcanzar los objetivos de utilidades de la empresa.

Las formas fundamentales de segmentación

Hay cuatro formas básicas de segmentación:

1. demográfica (por edad, sexo e ingresos);
2. geográfica (por región: urbana o rural);
3. conductual (clasificación de uso: usuario ligero, medio e intensivo) y,
4. psicográfica (personalidad: compulsiva, sociable, autoritaria, ambiciosa).

1) Demográfica: Son las variables estadísticas que describen las curvas nacionales o estatales de nacimientos, muertes, formación de familias y edad en la sociedad; así como las afiliaciones religiosas, raciales y los orígenes nacionales de la población. Por lo general, para una buena cantidad de trabajo de mercadotecnia, ha sido útil considerar conjuntamente el estado civil, así como la presencia y edades de niños en una variable del ciclo de vida familiar. Estilos de vida generales y las tasas de compra del consumidor en categorías como seguros, aparatos electrodomésticos y casas, se ven significativamente influenciadas por dicha variable.

2) Geográfica: Esta es una variable ampliamente usada para la segmentación y es el apoyo principal de organización para censos estadísticos. Es una consideración crítica en el sector público, donde las jurisdicciones gubernamentales influyen en los recursos, las leyes y otros determinantes del éxito del programa. La demanda agregada del producto, los gustos y preferencias son significativamente diferentes entre regiones y países. En todos los mercados, es necesario reconocer como varían los costos y el mercado potencial, de acuerdo a la localización del mismo.

La demográfica y la geográfica, quizá sean los tipos más frecuentes de segmentación, en el cuadro I, se muestran las formas más comunes de estas formas de segmentación.

CUADRO I. TIPOS DE SEGMENTACIÓN DEMOGRÁFICA Y GEOGRÁFICA

	DEMOGRÁFICA	GEOGRÁFICA
CONSUMIDOR	Edad, tamaño de familia, educación, ingresos familiares, nacionalidad, ocupación, raza, religión, residencia, sexo, clase social	Región del país, tamaño de la ciudad, densidad del mercado, clima
INDUSTRIA	Número de empleados, cantidad de ventas, monto de las utilidades, tipo de la línea de productos	Región del país
GOBIERNO	Tipo de organismo, monto del presupuesto, grado de autonomía	Federal, estatal y municipal

Fuente: Elaboración propia

3) Conductual: Hasta ahora nos hemos limitado a ilustrar la segmentación de mercado realizada con base en los factores demográficos y geográficos, con los cuales obtenemos las características para la identificación o denominación de la persona u organización. Estudiaremos a continuación otro criterio básico de esta estrategia: la **segmentación conductual**, que mide el comportamiento o actitudes dentro de la categoría del producto que está siendo analizada. Se desarrolla el segmento en razón de las cualidades que aporta el producto y no a las características del consumidor. Uno de los primeros campos en que se aplicó este criterio fue el mercado de los dentífricos. Con base en los resultados de la investigación las empresas lo dividieron así: 1) personas que deseaban un sabor agradable, 2) personas que querían evitar la caries dental, 3) personas activas que deseaban tener dientes más brillantes, 4) personas deseosas de economizar.

4) Psicográficas: Hay un espacio continuo desde las características muy generales de la personalidad, las cuales describen la orientación y la respuesta del usuario de algún servicio, un individuo a su medio ambiente, (por ejemplo, agresividad y necesidad de logro y dominación) hasta las mediciones específicas del estilo de vida y de la actividad, de los intereses y de las opiniones. Algunas veces estas últimas medidas se agrupan dentro de la categoría de características psicográficas. Un estilo de vida es un modo distinto de orientación, de usuarios de algún servicio o segmento, hacia el consumo, el trabajo y el juego. (Por ejemplo, volubles, jugadores, etcétera). Las actividades, intereses y opiniones son, en contraste, bastante específicos para la compra y el uso de la categoría del producto o del servicio. Sin embargo, son suficientemente generales y orientados hacia las personas, y pueden ser usados en un número de categorías de producto.

En la siguiente cuadro, se muestran los cuatro tipos de segmentación explicados anteriormente, así como sus descriptores, es decir las unidades de análisis del tipo de segmentación.

CUADRO II. PRINCIPALES VARIABLES DE SEGMENTACIÓN PARA LOS MERCADOS DE CONSUMO

TIPO DE SEGMENTACIÓN	VARIABLE	CLASIFICACIÓN GENERAL
GEOGRÁFICA	Región	Pacífico, Montañas, Noroeste, Suroeste, Noreste y Sureste Central, Atlántico del Sur y Medio, Nueva Inglaterra.
	Tamaño del condado	A, B, C, D
	Tamaño de la ciudad	Menos de 5, 5-20, 20-50, 50-100, 100-250, 250-500, 500-1000, 1000-4000, 4000 o más (Cifras en miles).
	Densidad	Urbana, suburbana, rural.
	Clima	Nórdico, meridional.
DEMOGRÁFICA	Edad	Menos de 6, 6-11, 12-19, 20-34, 35-49, 50-64+
	Sexo	Masculino, femenino.
	Tamaño de la familia	1-2, 3-4, 5+
	Ciclo de vida de la familia	Joven, soltero; joven, casado, sin hijos; joven, casado, el hijo más pequeño menor de 6 años; joven, casado, el hijo más pequeño de 6 años o mayor; mayor, casado, con hijos; mayor, casado, sin hijos menores de 18 años; mayor, soltero.
	Otros ingresos	Menores de \$10, \$10-15, \$15-20, \$20-25, \$25-30, \$30-50, \$50 y más (Cifras en miles).
	Ocupación	Profesional y técnico; directivos, funcionarios y propietarios; eclesiásticos, ventas; artesanos, operadores; agricultores; jubilados, estudiantes; amas de casa; desempleados.
	Educación	Escuela primaria o menos, algo de secundaria, secundaria terminada, universitaria, título profesional.
	Religión	Católica, protestante, judía, otra raza blanca, negra, oriental.
	Nacionalidad	Estadounidense, británica, francesa, alemana, escandinava, italiana, latinoamericana, del medio oriente, japonesa.
PSICOGRÁFICA	Clase social	Baja-baja, baja-alta, clase trabajadora, media, media-alta, alta-baja, alta-alta.
	Estilo de vida	Conservadora, mundanos, intelectuales.
	Personalidad	Compulsiva, sociable, autoritaria, ambiciosa.
CONDUCTUAL	Ocasiones	Regularmente, ocasiones especiales.
	Beneficios	Calidad, servicio, economía.
	Condición del usuario	No usuario, ex-usuario, usuario potencial, primerizo y regular.
	Clasificación del uso	Usuario ligero, medio e intensivo.
	Estado de lealtad	Ninguna, media, sólida, absoluta.
	Etapas de disposición	No consciente, consciente, informado, interesado, deseos, pretende comprar.
	Actitud hacia el producto	Entusiasta, positiva, indiferente, negativa, hostil.

Fuente: Dirección de la Mercadotecnia, Análisis, Planificación y Control, pág. 301.

Si un gerente de mercadotecnia conoce los beneficios que desea el consumidor puede modificar la mezcla para que corresponda a esas preferencias. Si no está en condiciones de hacerlo, quizá le convenga desarrollar un nuevo producto. Le será más fácil elaborar un programa de promoción una vez efectuada la segmentación, puesto que ella le permite descubrir los aspectos del producto que quiere el cliente.

El gerente de mercadotecnia tiene además la obligación de investigar cuáles artículos satisfacen mejor las exigencias del público; por tanto, se entera de las debilidades y ventajas de la competencia. La información suministrada por la demografía le permite analizar las características de los que buscan determinado tipo de beneficios. Luego, la compara con los perfiles conseguidos de los medios de comunicación, a fin de hacer concordar el público que los ve y los oye con el mercado en cuestión. Las personas que van al Burger King lo hacen porque buscan un restaurante con buena imagen en la comunidad, que sirva comida sabrosa, y cuyos empleados sean atentos y serviciales. Muchos de los que comen en los Mc'Donalds quieren un lugar que goce de popularidad entre los niños y además acuden a ellos por razones de comodidad; conceden poca importancia a la diversidad de menús. Se debe idear una mezcla de mercadotecnia para ambas cadenas de alimentos rápidos si se desea favorecer más la satisfacción de esas necesidades.

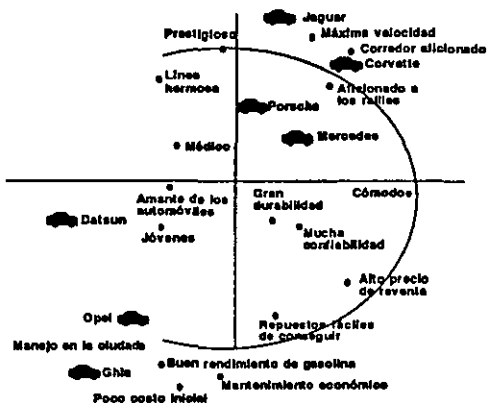
Los *mapas perceptuales* se utilizan a menudo para destacar las ventajas de un producto. Los atributos de varias marcas y productos se trazan en "mapas", con métodos estadísticos muy modernos. Cuanto más cercanas estén las marcas en el mapa, mayor será su semejanza. Por otra parte, cuanto más próximo esté el atributo a determinada marca, mayores probabilidades habrá de que los consumidores piensan que poseen este atributo.

Si la explicación anterior le parece al lector un poco confusa, examine con detenimiento la figura 3. Este mapa perceptual muestra las relaciones de atributos de productos, marcas de automóviles deportivos y el típico propietario de un auto. Por ejemplo, "gran durabilidad" y "mucho confiabilidad" se hallan en estrecha proximidad con "alto valor de reventa", lo cual indica que para el consumidor ambas dimensiones son determinantes importantes de un gran valor de reventa. Observe también que ninguno de los automóviles deportivos se aproxima a estos beneficios, circunstancia que posiblemente indica una oportunidad de mercadotecnia. En forma análoga, el Opel se consideraba

básicamente un automóvil para la ciudad, mientras que el Corvette se le percibía en función de la velocidad y de las carreras, pero sin el prestigio del Jaguar.

FIGURA 3

MAPA PERCEPTUAL DE MARCAS DE AUTOMÓVILES DEPORTIVOS,
ATRIBUTOS DEL PRODUCTO Y DUEÑOS CAPRICHOSOS DE AUTOMÓVILES



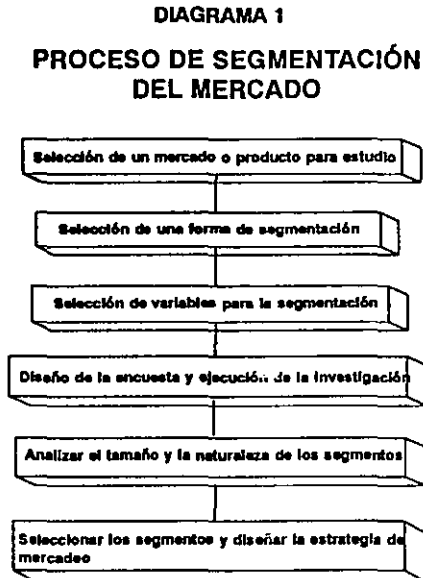
Fuente: Curso de mercadotecnia, pág 78

Los mapas perceptuales ayudan al mercadólogo a distinguir entre las marcas competidoras y a encontrar resquicios vacíos en el mercado. Con el examen de una marca en el mapa perceptual, el gerente puede evaluar los puntos fuertes y débiles de la marca. Por ejemplo, con referencia de nuevo a la figura 3, se considera que el Ghia es económico y de mantenimiento barato, pero no bello, veloz o prestigioso.

Otro ejemplo de este criterio de segmentación es por volumen de consumo. Hay usuarios *intensos, moderados y ligeros y no usuarios* casi para cualquier producto. A menudo los "usuarios intensos" representan una gran parte del consumo de un producto. Por ello, la compañía quizás quiera enfocar su mezcla para comercializar al segmento de usuarios intensos. Miller lo hizo cuando dejó de anunciar su cerveza como "la champaña de las cervezas embotelladas" y la enfocó hacia los trabajadores manuales; éstos son los "usuarios intensos" de cerveza y no se suelen identificar ellos mismos con la imagen de la "champaña". Ahora, en los anuncios actuales aparecen los "interesantes" miembros de las profesiones no especializadas, como los capitanes de remolcadores y los perforadores de pozos petroleros, todos ellos bebiendo una cerveza Miller.

Proceso más común en estudios de segmentación de mercados

En el diagrama 1, se muestra la forma más común de estudios de segmentación de mercado.



Fuente: *Ibidem*, pág. 30.

1. Se selecciona una área de mercados o de productos.
2. Se escoge una forma de segmentar el mercado (de las que hemos explicado o una combinación de ellas).
3. Se seleccionan descriptores de la segmentación en la forma escogida como factores psicográficos (personalidad: compulsiva, sociable, autoritaria, ambiciosa); factores demográficos (edad, ingresos, raza); factores geográficos (región, estado, municipios, etcétera).
4. Se diseña la encuesta y se realiza la investigación.
5. Se analizan el tamaño y la naturaleza de los segmentos descubiertos en la investigación.
6. Se seleccionan los segmentos y se establece la estrategia de mercadeo (desarrollo de producto, frases promocionales, estrategias de fijación de precios y canales de distribución) para llegar a los consumidores.

Selección de un mercado o producto para estudio. El área de mercado o de productos dónde se realizará la investigación de la segmentación depende de los objetivos de la empresa: *dónde quiere estar presente, a quiénes desea servir y recursos de los que dispone*. A menudo la investigación se aplica a subsegmentos ya servidos. Por ejemplo, el estudio efectuado por Anheuser-Busch sobre el mercado de las cervezas ligeras constituye un segmento del mercado total. A veces esta clase de trabajos se llevan a cabo en campos afines, como cuando Frito-Lay examinó los mercados de los productos para untar y quesos derretidos.

En otros casos, la investigación se hace cuando ha habido un notable adelanto tecnológico. Por ejemplo, uno de los principales fabricantes de plaguicidas inventó una sustancia que impide la reinfestación por pulgas. Los productos actuales matan a las pulgas pero no a los huevecillos; por tanto, cuando incuban los huevecillos reaparece el problema. El fabricante llevó a cabo un estudio exhaustivo de segmentación a fin de analizar los diversos mercados y los sistemas de aplicación. Se entrevistó a los propietarios de tiendas de animales domésticos, a los dueños de criaderos de perros, y a los poseedores de gatos y de perros. También se examinó en cada segmento la duración de la eficacia (un beneficio). Entre las opciones de aplicación figuraban atomizadores, pulverizadores de bomba, aerosoles, polvos y champúes.

Selección de una forma de segmentación. Una vez seleccionada una área de investigación, los mercadólogos escogen una o varias formas de segmentación. Pueden hacerlo con base en factores demográficos, psicográficos, geográficos, el beneficio o el volumen. La base escogida puede ser resultado de su opinión y de sus experiencias. Un fabricante de tintes para el cabello posiblemente quiera analizar los segmentos de mercado con un criterio demográfico y a partir del estilo de vida. Al segmentar su mercado de telefonemas de larga distancia, quizá Bell Telephone prefiera el volumen de esas llamadas como criterio de segmentación.

Selección de variables para la segmentación. Luego de escoger una forma, el gerente de mercadotecnia ha de seleccionar los *descriptores de la segmentación*. Estos son las unidades de análisis del tipo de segmentación. Por ejemplo, si una compañía decide utilizar la demografía como base de su segmentación, posiblemente use la edad, ocupación e ingresos como variables. La que aplica la segmentación geográfica quizá seleccione como variables el contraste entre urbano y rural. En el ejemplo del control de

plagas de pulgas intervienen diversas variedades de segmentación (ver cuadro III):

CUADRO III. DESCRIPTORES SEGÚN TIPO DE SEGMENTACIÓN

Tipo	Descriptor
Conductual (beneficio)	Duración del control de plagas de pulgas.
Psicográficos (estilo de vida)	Cazadores (con perro) o no cazadores.
Factores demográficos	Dueños de perreras, veterinarios, tipo de animal doméstico, tamaño del animal doméstico.
Factores geográficos	Urbano, rural, fincas.

Fuente: elaboración propia

Diseño de la encuesta y ejecución de la investigación. Es obvio que el investigador debe entrevistar a los consumidores pertinentes. Perdería el tiempo si aplicara un estudio sobre productos contra las pulgas entre personas que no tengan un animal doméstico. Además se deben examinar muestras representativas de los consumidores a quienes desea analizar la compañía.

Analizar el tamaño y la naturaleza de los segmentos. En esta etapa los segmentos ya han sido integrados con la clasificación de los entrevistados en varias categorías. Se examinan los perfiles y los beneficios deseados; se estima el tamaño de los segmentos. Encontramos, por ejemplo, un segmento de 1.5 millones de dueños, en los suburbios donde hay perros que pesan más de 25 kg.; esas personas afirman que "a toda costa probarán el producto". Este grupo quería comprar la presentación en aerosol y con eficacia durante cinco meses. Están dispuestos a pagar \$ 5.95 por él. Un análisis más detallado del segmento reveló cuáles son los medios de comunicación que escuchaban, que veían y leían, así como los productos de control de pulgas que utilizaban. Se analizó información similar en otros segmentos que satisficieran las metas de utilidades establecidas por el fabricante de plaguicidas.

Seleccionar los segmentos y diseñar la estrategia de mercadeo. Tras examinar los resultados de la investigación del estudio, la empresa identificó siete segmentos que le permitirían lograr sus objetivos de utilidades. Desafortunadamente la disponibilidad de recursos y la capacidad de producción a corto plazo hará que ésta no rebase las 900,000 unidades. En ese momento el fabricante clasificó sus segmentos a partir de la

rentabilidad y escogió segmentos (comenzando por los de mayor rentabilidad) hasta que le fue imposible satisfacer las necesidades de la producción. Al final se quedó con cuatro segmentos: veterinarios especializados en el cuidado de los animales que viven en las zonas de clima muy cálido; las perreras situadas en esas regiones, los dueños de perros que pesan más de 25 kg. y los de perros cuyo peso fluctúa entre 10 y 25 kg.

A continuación se preparó una mezcla de mercadotecnia que llegase a cada segmento. Cada uno necesitaba un producto y un método especial de distribución. Los productos se vendían por medio de diferentes canales. En el caso de dueños de perros, se hizo una promoción diferente para cada producto. Al mercado de veterinarios se llegó a través de publicidad en revistas especializadas, se utilizó el correo directo en el caso de perreras y tiendas de animales domésticos, para dar a conocer el producto a los dueños de perros, se utilizaron las revistas *Better Homes and Gardens*, *Outdoor Life* y *Good Housekeeping*, etcétera. Se aplicaron diferentes estrategias de precios a cada segmento del mercado.

Problemas, falacias y contrasegmentación. La segmentación no es una panacea y tampoco satisface las exigencias de todas las empresas. En realidad, hay costos que se elevan al aplicarla. Por ejemplo, la diferenciación de productos significa menor economía de escala³ y costos mayores de producción e inventarios: una mayor variedad a menudo implica mayor complejidad del producto y más problemas técnicos. Las diferentes estrategias promocionales para diversos segmentos, pueden significar mayores gastos de promoción. Si se requieren diferentes canales de distribución, estos costos también serán más elevados.

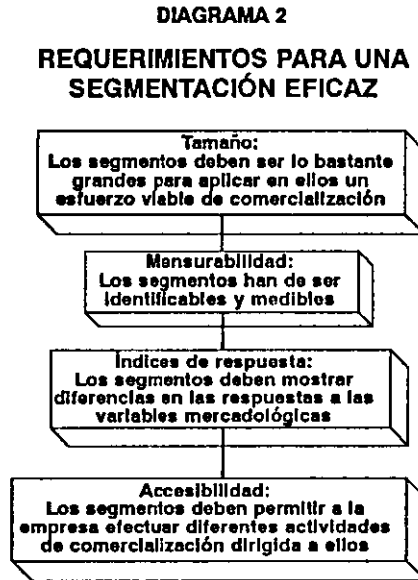
Además de las consideraciones con los costos, algunos mercados pueden ser demasiado pequeños para segmentarlos de un modo económico. Quizá un mercado no tenga suficientes características distintivas para segmentarlo. O bien, el mercadólogo puede descubrir una base para la segmentación pero quizá no pueda identificar y llegar a los consumidores objetivo. En algunos casos, los usuarios intensos constituyen una proporción tan grande del volumen de ventas, que son el único objetivo pertinente. Si un grupo reducido produce la mayoría de las ventas, los esfuerzos de comercialización se deben enfocar a ese grupo.

³

Es decir, las ganancias en la producción y/o en los costos resultantes del aumento del tamaño de la planta, empresa o industria.

Requerimientos para una segmentación eficaz

Para que sean de utilidad máxima, los segmentos del mercado deben mostrar las siguientes características (diagrama 2):



Fuente: Dirección de la Mercadotecnia, Análisis, Planificación y Control, pág. 312.

1. **Tamaño.** Este primer criterio indica simplemente que el segmento o segmentos definidos han de ser bastante amplios para que la empresa alcance sus objetivos de utilidades. Si se lleva a extremos ridículos, podríamos afirmar que cada miembro de un mercado constituye un segmento. Es decir, todos somos diferentes en una forma u otra. Es evidente que resulta absurda esta actitud tan exagerada. Si un gerente de mercadotecnia empieza con un mercado que tenga potencial limitado de ventas, quizá la segmentación no sea una opción viable. El mercado de aviones 747 equipados al gusto de cada cliente es tan reducido, que Boeing cometería un error imperdonable si tratara de segmentarlo. En cambio, el mercado total del 747 se segmenta entre usuarios militares o comerciales, las líneas aéreas y transporte de carga o de pasajeros.

2. **Mensurabilidad.** Es el segundo criterio, el gerente de mercadotecnia debe ser capaz de medir y cuantificar una posible segmentación a fin de evaluar su variabilidad. La Oficina de Turismo del gobierno canadiense quería efectuar un estudio sobre los estadounidenses que posiblemente decidan pasar sus vacaciones en Canadá. Como quería ampliar el número de vacacionistas que arriban a Canadá, no bastaba con examinar el 5% de la población estadounidense que ya había visitado ese país. Tampoco todos los que toman vacaciones se deben considerar parte del mercado.

En este caso el estudio se basaba en los que habían tomado una semana de vacaciones el año anterior y habían recorrido una distancia equivalente a un viaje a Canadá (hubieran o no ido a ese país).

Supongamos que una fábrica de cosméticos quisiera efectuar un estudio de segmentación para averiguar el tamaño del mercado y las características demográficas de mujeres casadas que desean vehementemente tener un amante aunque sea platónico. La identificación y medición de este mercado representaría una ardua tarea en el mejor de los casos. Los investigadores han descubierto que es difícil medir el éxito de la operación de lucha contra el crimen, porque los estudios de segmentación de delitos señalan que algunos no quieren denunciar las agresiones que les cometen a ellos o a sus familias. Además, algunos delitos (entre ellos los hurtos en tiendas) rara vez se comunican a la autoridad.

3. **Índices de respuestas.** El tercer criterio de segmentación significa que, si varios segmentos emiten una reacción similar ante la mezcla de mercadotecnia, no se necesita idear una especial para cada uno. Por ejemplo, si todos los segmentos respondieron en forma idéntica ante los cambios de precios, no hay necesidad de fijar precios individuales para cada segmento.
4. Finalmente, la **accesibilidad** exige detectar y, luego, medir las características de la segmentación. Y lo más importante: el mercadólogo ha de tener acceso a los mercados segmentados. Supongamos que, con la investigación de mercado, una compañía de licores descubre que quienes más consumen una nueva marca de whisky escocés son las mujeres divorciadas y de disipada vida social. La segmentación será inútil si no se logra averiguar los medios de comunicación que

este grupo ve, lee o escucha, y si no se investiga dónde vive o hace sus compras. Desde luego, quizá no se cuenta con un medio que llegue a esas personas.

1.3 CARACTERÍSTICAS DE LOS USUARIOS DE SERVICIOS BANCARIOS

Una vez definidos los términos esenciales para entender la función y la importancia de la investigación de mercados, surgen algunas preguntas que pueden ser de interés para el investigador de mercados: ¿Cuál es el perfil de mis clientes? ¿Qué características diferencian a mis clientes de los de otros bancos?

Utilizando la lista de posibles bases del cuadro II (pág. 19), que aunque no es exhaustiva, presenta las cualidades de las variables importantes, que pueden ser usadas para definir y describir las características de una población objetivo, la cual en esta investigación será la población de usuarios de diferentes instituciones bancarias, en particular interesará conocer las siguientes características: edad, sexo, ingresos, nivel educativo, nivel social, estado civil, condición de actividad (trabaja o no trabaja) y ocupación.

El conocer el perfil de los clientes de cada banco, puede servir al mercadólogo a distinguir entre los bancos competidores, así como para elaborar promocionales con el objetivo de atraer nuevos clientes de acuerdo a estas características (mediante un análisis más detallado del segmento que revele cuáles son los medios de comunicación que escuchaban, que ven o leen).

Algunos investigadores consideran las características del consumidor o usuarios de servicios bancarios, independientemente del interés específico del producto. Generalmente utilizan características geográficas, demográficas y psicográficas. Después ven si estos segmentos de clientes o usuarios muestran diferentes respuestas hacia el producto o servicio. Otros investigadores consideran las **respuestas del consumidor** al producto, tales como beneficios buscados, ocasiones de uso, marcas y lealtad. Una vez que están formados los segmentos, el investigador analiza si las diferentes características del cliente o usuario de servicios están relacionadas con cada segmento. Por ejemplo, se podría investigar si las personas que quieren "calidad en la atención" vs. "tasas bajas de interés" al adquirir una tarjeta de crédito, difieren en su formación geográfica, demográfica y psicográfica.

CAPÍTULO 2

MODELOS ESTADÍSTICOS Y EL PERFIL DE USUARIOS

2.1 MODELOS EXPLICATIVOS

El Modelo de Regresión Lineal

Una manera de introducir los problemas relacionados con el *Análisis Discriminante* (AD) y la *Regresión Logística* (RL), es tratar primero el tema de Análisis de Regresión (AR). Abordar el tema en este orden se debe a dos razones; en primer lugar, porque AR proporciona los antecedentes teóricos de modelos que pretenden explicar desde el punto de vista probabilístico el comportamiento de una variable aleatoria Y_i , en términos de variables explicativas (X_1, X_2, \dots, X_N), y en segundo lugar, porque aunque están dirigidos a resolver situaciones diferentes, tienen muchas cosas en común.

En Análisis de Regresión, el investigador supone que tiene una variable real⁴ X , que toma n valores distintos, es decir; x_1, x_2, \dots, x_n . Además el investigador observa que asociada a cada uno de los valores x_i existe una variable aleatoria (v.a.) Y_i continua, es decir:

x_1	Y_1
x_2	Y_2
.	.
x_n	Y_n

Algunas preguntas de interés son: ¿Existirá alguna relación entre la variable x y la variable y ?, en caso de existir alguna relación entre las variables ¿se podrá predecir de manera exacta el valor que asumirá ésta en cualquier situación de interés. Si este es el caso, solamente se podrá hablar de las probabilidades de observar valores de interés de la v.a. Y_i asociados con el valor x_i de la variable real X .

⁴ Los valores de una variable real no tienen asociadas probabilidades de ocurrencia.

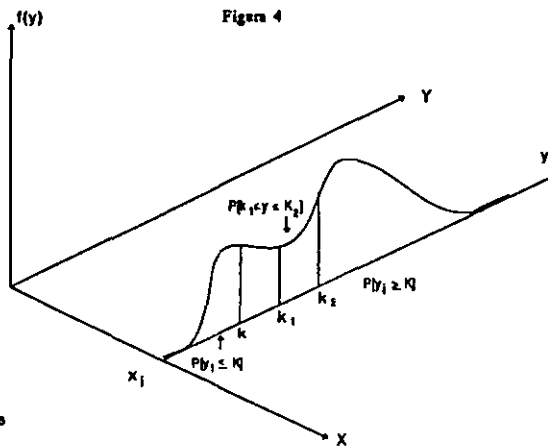
Como Y_i es continua, entonces para el valor x_i , algunas probabilidades de interés serán las siguientes:

$$P[Y_i \geq k]$$

$$P[Y_i \leq k]$$

$$P[k_1 \leq Y_i \leq k_2]$$

ya que la probabilidad de que $Y_i = k$ es cero, por ser Y_i variable aleatoria continua. (Ver figura 4).



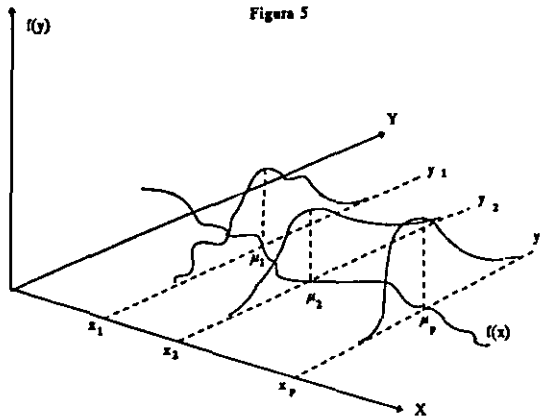
Para poder calcular las probabilidades anteriores es necesario conocer completamente la ley de distribución de probabilidades de cada variable aleatoria Y_i .

Como cada Y_i puede ser diferente para cada x_i , tratar de estudiar la situación así planteada sería muy complicado, ya que la diferencia en las variables aleatorias Y_i se puede dar en cualquiera de los siguientes casos:

- a) Si sus valores son diferentes.
- b) Si la media es diferente.
- c) Si la varianza es diferente.
- d) Si la función de densidad es diferente.

⁵ Todas las figuras de este capítulo han sido elaboradas por la autora.

Es decir, las variables Y_i son diferentes si al menos una de las características anteriores se cumple (ver figura 5). Por lo que, el Modelo de Regresión Lineal simplifica la situación anterior al establecer los siguientes supuestos:



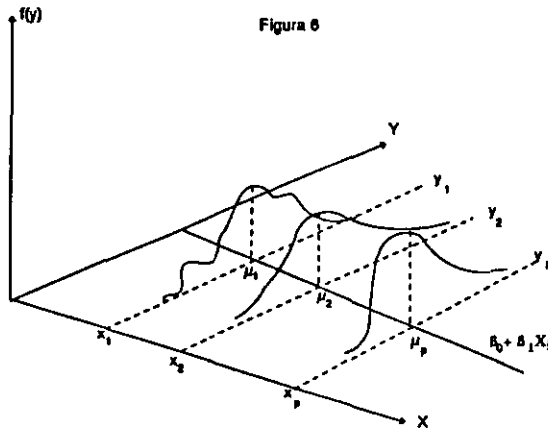
Fuente: (ver nota pág. 30)

Supuesto 1: Las medias de las variables Y_i , están dadas por la siguiente relación:

$$E(Y_i) = \beta_0 + \beta_1 X_i \quad (1)$$

donde: $E(Y_i)$ denota la media de la variable aleatoria Y_i ; y
 X_i es el valor de X asociado con la v.a. Y_i

Con este supuesto, la línea que une las medias ya no es arbitraria como en la figura 2.2, ya que éstas se calcularán de manera exacta mediante la ecuación (1), dado cualquier valor x_i de X . Gráficamente implica que las medias estarán ubicadas sobre una línea recta (figura 6).

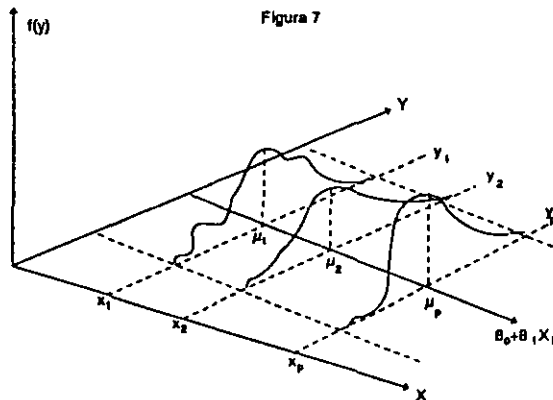


Fuente: (ver nota pág. 30)

Supuesto 2: Las varianzas de las variables aleatorias Y_i son iguales, es decir:

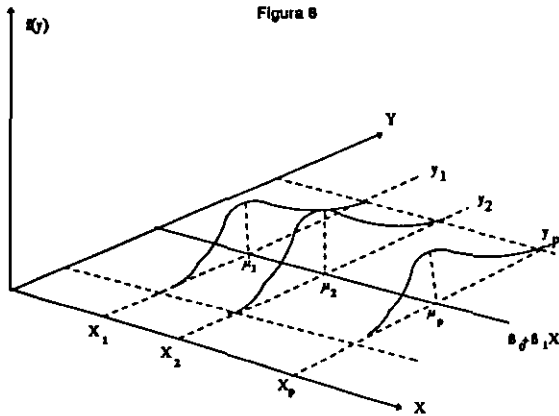
$$\text{VAR}(Y_i) = \sigma^2 \quad \text{para toda } i$$

Este supuesto restringe a las variables Y_i , a que el grado de dispersión a los lados de la media sea el mismo. Es decir, las variables aleatorias pueden tomar diferentes valores, pero el grado de dispersión debe ser el mismo para cada x_i . En la figura 7, se observa que, la variación de cualquier Y_i dado x_i , está restringido a tomar valores entre las líneas punteadas.



Supuesto 3: Las variables aleatorias Y_i tienen una ley de distribución de probabilidades normal.

Gráficamente este supuesto se traduce, en que la forma de la función de densidad de las variables Y_i sea de campana (figura 8).



Fuente: (ver nota pág. 30)

Supuesto 4: Las variables Y_i son no correlacionadas, es decir:

$$\text{COV}(Y_i, Y_j) = 0 \quad \text{para } i \neq j$$

Este último supuesto restringe a las variables Y_i y Y_j a que sus valores y sus probabilidades sean independientes para cualquier valor x_i y x_j respectivamente. Este supuesto implica que lo que sucede con Y_i no se puede utilizar para predecir lo que sucederá con Y_j .

De esta manera, el problema inicial se ha simplificado ya que la única diferencia entre las variables aleatorias Y_i es la media. Pero la media de la variable aleatoria Y_i se predice de manera exacta y está dada por $E(Y_i) = \beta_0 + \beta_1 x_i$ para un determinado valor de X .

Es importante dejar claro, que el Modelo de Regresión Lineal no podrá predecir valores individuales de la variable aleatoria Y_i para una x_i dada, sólo tratará de predecir de manera exacta la media de la v.a. Y_i , es decir; para cualquier valor de x_i el modelo determinará sólo de manera exacta la media de la v.a. Y_i . Sin embargo, con dicha media y los supuestos anteriores, se puede conocer cualquier probabilidad relacionada con los valores de cada v.a. Y_i .

2.2 Análisis Discriminante

Supongamos ahora, que se tienen Y, X_1, X_2, \dots, X_p variables aleatorias, de tal manera que se piensa que la ley de distribución de probabilidad (comportamiento) de Y , depende de los valores que asumen las otras variables. Si Y es continua, el modelo de regresión lineal múltiple se puede usar para intentar explicar la relación entre estas variables.

Recordemos que en regresión Y, X_1, X_2, \dots, X_p son mediciones de los elementos de una población. Ahora supóngase el caso en que se tengan éstas mismas mediciones para las x 's características y Y no represente una medición, o variable continua como antes, sino más bien una característica presente o ausente en los elementos, es decir una variable categórica y si el interés se centra en estudiar algún tipo de relación entre las variables, entonces esta situación no se puede modelar con Análisis de Regresión, para este caso existen dos modelos alternativos que se pueden usar, el **Análisis Discriminante y la Regresión Logística**.

Ahora bien, si definimos a $Y=1$ cuando el elemento de la población tenga el atributo de interés y a $Y=0$ cuando no lo tenga⁶, y si nos interesa analizar como cambia $P(Y=1)$ y la $P(Y=0)$ en función de X_1, X_2, \dots, X_p , entonces la teoría adecuada es la **Regresión Logística**. Si por el contrario, nos interesa determinar qué características están asociadas con los elementos que tienen el atributo de interés y cuáles están asociadas con los elementos que no la tienen, entonces un modelo adecuado podría ser **Análisis Discriminante**.

⁶ Se le conoce a Y como variable dicotómica o "dummy"

La geometría de las combinaciones lineales

Una manera de visualizar cómo este modelo determina las características de interés, es analizando la geometría de las combinaciones lineales cuando se involucran solamente dos variables X_1 y X_2 . Por ejemplo, supongamos la siguiente combinación lineal:

$$Z = 4 X_1 - 3 X_2 \quad (1)$$

donde:

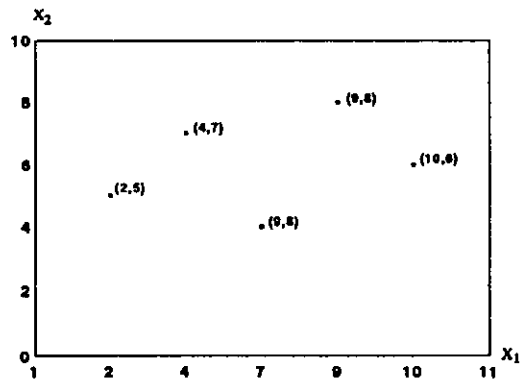
Z es una nueva variable que se forma a partir de X_1 y X_2 .

y, supongamos que se tienen 5 mediciones de X_1 y X_2 , cuyos valores se pueden observar en la cuadro IV y los cuales están representados en la figura 9.

Cuadro IV

MEDICIÓN	X_1	X_2
1	2	5
2	4	7
3	7	4
4	9	8
5	10	6

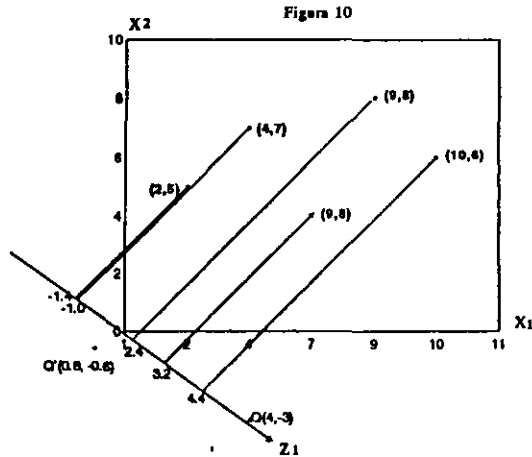
Figura 9



Fuente: (ver nota pág. 30)

Como se mencionó anteriormente, la combinación lineal de X_1 y X_2 da origen, al sustituir los valores de las mediciones en la ecuación a una nueva variable Z , cuyos valores son: -7, -5, 16, 12 y 22.

El problema ahora es, encontrar la manera de representar en la misma figura los valores de Z . Un método es, dibujando una línea que una el origen y el punto definido con los coeficientes 4 y -3 de la combinación lineal, el cual se denota en la figura 10, con la letra Q.



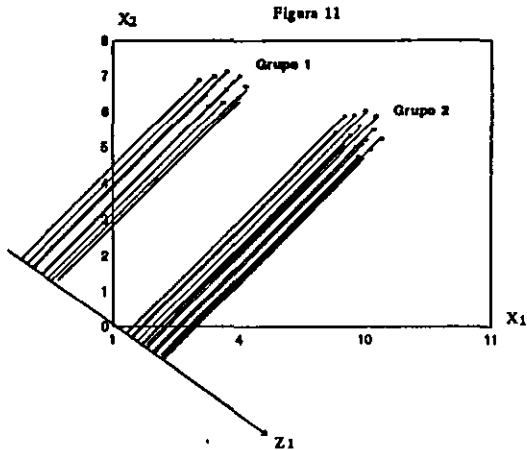
Fuente: (ver nota pág. 30)

Obsérvese que sobre la recta existen otros puntos con los cuales se puede definir la misma recta y con cuyas coordenadas se puede formar otra combinación lineal. Como el punto Q'(0.8, -0.6), cuyas coordenadas definen la siguiente combinación lineal:

$$Z_1 = 0.8 X_1 - 0.6 X_2 \quad (2)$$

De tal manera que al sustituir los valores de X_1 y X_2 en la ecuación (2) que representa esta nueva combinación lineal, se obtienen los siguientes valores: -1.4, -1.0, 2.4, 3.2, y 4.4 que gráficamente representan las intersecciones de las líneas trazadas de manera perpendicular de los puntos P a la recta Z_1 . Obsérvese que el origen del plano definido por X_1 y X_2 también es el origen de la recta Z_1 , y que por lo tanto, los valores de Z_1 se distribuyen a la izquierda y a la derecha del origen, según si los valores son negativos o positivos.

La importancia del hecho anterior radica en que, si en el plano definido por X_1 y X_2 existen dos grupos de puntos, entonces las proyecciones también quedan agrupadas en el eje Z_1 . (Ver figura 11).



Fuente: (ver nota pág. 30)

Pero el hecho todavía más importante es, que no es necesario representarlo gráficamente ya que los valores de las intersecciones están dados por la combinación lineal (2). Es decir, si al graficar los valores de Z_1 se observan dos grupos de puntos, quiere decir que en el espacio de dos dimensiones definidos por X_1 y X_2 , también existen dos grupos de puntos.

El problema ahora es, determinar los coeficientes de la combinación lineal que tiene esa propiedad. La solución se obtiene al dividir cada uno de los coeficientes de la combinación lineal (1) entre 5, es decir:

$$\sqrt{4^2 + (-3)^2} = \sqrt{25} = 5$$

$$Z_1 = \frac{4}{5} X_1 - \frac{3}{5} X_2 = 0.8 X_1 - 0.6 X_2$$

En general, si la combinación lineal está dada por: $Z = a X_1 + b X_2$ y, si Z , se define como:

$$Z_1 = \frac{a}{W} X_1 + \frac{b}{W} X_2$$

donde: $W = \sqrt{a^2 + b^2}$

Entonces, Z, tiene la propiedad de agrupar en una dimensión los posibles grupos que existan en dos dimensiones. En general, si se tiene un espacio p-dimensional, entonces la combinación lineal buscada, formada por p variables. Está dada por:

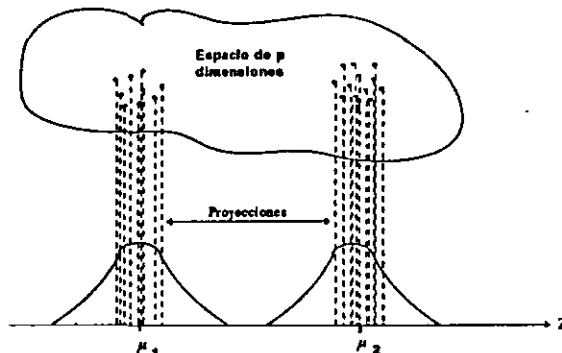
$$Z = \frac{a_1}{W} X_1 + \frac{a_2}{W} X_2 + \dots + \frac{a_p}{W} X_p \quad (3)$$

donde: Las X's representan variables
Las a's son constantes desconocidas y,

$$W = \sqrt{\sum_{j=1}^p a_j^2}$$

Aunque en este caso no podemos visualizar un espacio en p dimensiones, lo importante es, que si en dicho espacio existen 2 grupos de puntos, entonces los valores de Z que se obtienen con la combinación lineal (3), se proyectan en el espacio unidimensional, definido por Z, también en 2 grupos de puntos (figura 12).

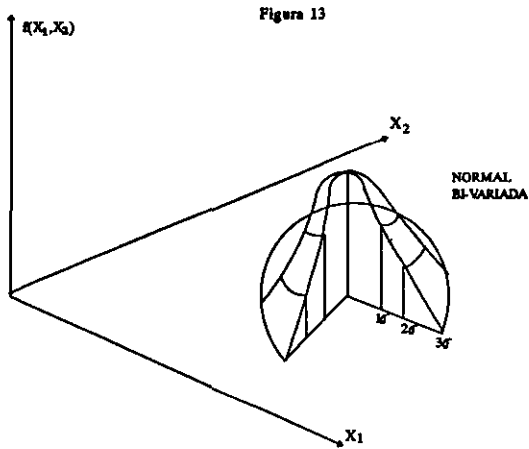
Figura 12



Fuente: (ver nota pág. 30)

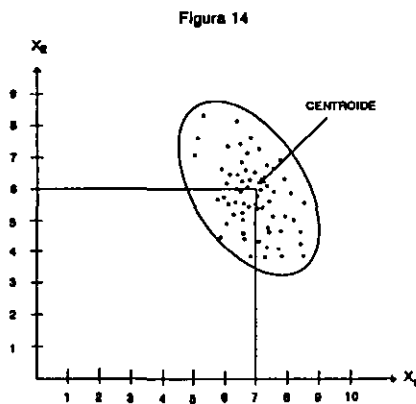
La distribución de una combinación lineal

Supongamos que X_1 y X_2 tienen una ley de distribución de probabilidades normal bi-variada, entonces su función de densidad tiene la forma descrita en la figura 13.



Fuente: (ver nota pág. 30)

Si no consideramos la función de densidad $f(X_1, X_2)$, entonces los puntos se dispersan entre una elipse (figura 14). La gráfica muestra la dispersión de los puntos cuyas coordenadas X_1 y X_2 siguen una distribución normal bi-variada con centroide en el punto $(7,6)$, que son las medias de X_1 y X_2 respectivamente.



Fuente: (ver nota pág. 30)

Si ahora proyectamos los puntos en el eje Z_1 , definido como:

$$Z_1 = 0.8 X_1 - 0.6 X_2$$

entonces, el punto (7, 6) se proyectará en $Z_1 = 2$. De esta manera, una pregunta interesante que surge en este momento es saber qué pasará con las proyecciones de los demás puntos. ¿Tendrán un comportamiento predecible?

Se puede demostrar formalmente que, si X_1 y X_2 siguen una ley de distribución normal bi-variada, entonces la combinación lineal $Z_1 = a X_1 + b X_2$, tiene una ley de distribución normal univariada con media y varianza dadas por:

$$\mu = a E(X_1) + b E(X_2)$$

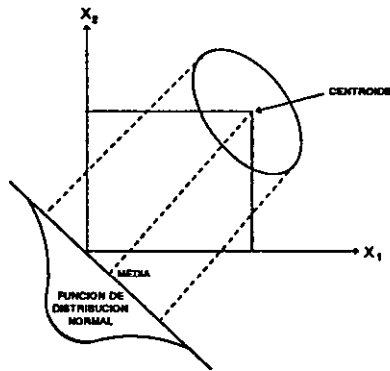
$$\sigma^2 = a^2 \sigma_1^2 + b^2 \sigma_2^2 + 2 a b r_{12} \sigma_1 \sigma_2$$

donde: σ_1 y σ_2 son las desviaciones estándar de X_1 y X_2 , y r_{12} es su coeficiente de correlación.

Con base en el resultado anterior, se puede predecir algo con respecto a la variable Z , por ejemplo que:

1. La mayoría de los puntos en el plano que estén cercanos al centroide se proyectarán en el eje Z cercanos a su media,
2. La mayoría de los puntos que estén alejados del centroide se proyectarán en Z alejados de su media y,
3. La forma de la función de densidad de la variable aleatoria Z será en forma de campana (figura 15).

Figura 15

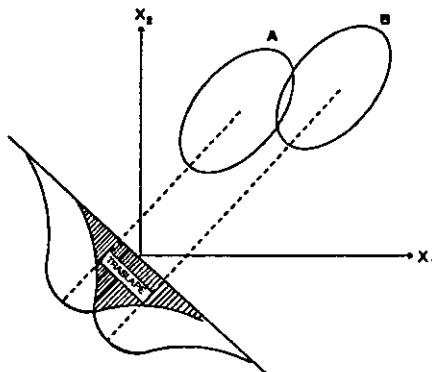


Fuente: (ver nota pág. 30)

Grados de traslape en distribuciones

Sean A y B dos grupos de puntos en el plano definido por X_1 y X_2 , distribuidos según una ley de distribución normal bi-variada cada uno, y supongamos que proyectamos los puntos del plano en un eje definido por una combinación lineal, como se ilustra en la figura 16.

Figura 16



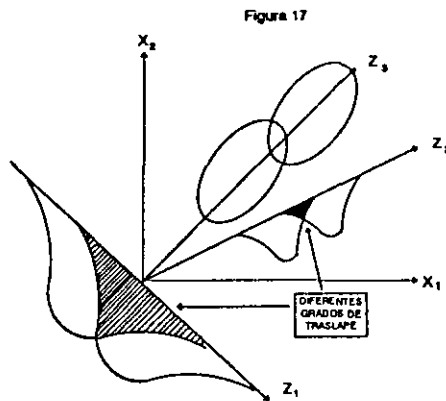
Fuente: (ver nota pág. 30)

En esta figura, observamos que al hacer la proyección, existe un cierto grado de traslape de las distribuciones. Una medida del grado de traslape está dado por la siguiente relación:

$$\frac{(\mu_1 - \mu_2)^2}{\sigma^2} = f^2$$

donde: μ_1 es la media de X_1 ,
 μ_2 es la media de X_2 , y
 σ^2 es la mayor de las varianzas de X_1 y X_2

f^2 se interpreta de la siguiente manera: el grado de traslape es inversamente proporcional a su valor; es decir, entre más pequeño es el valor de f^2 el grado de traslape es mayor, de tal manera que, entre más alejadas estén las medias de las distribuciones, el grado de traslape es menor. Pero además la separación entre las medias se hace más pequeña (se pondera) al dividir entre el grado de dispersión mayor de las variables aleatorias. Obsérvese también que si las medias son iguales, el grado de separación entre las distribuciones es cero y por lo tanto se obtiene el traslape máximo. En la figura 17 se observa que el grado de traslape o de separación es diferente, dependiendo de la combinación lineal.



Fuente: (ver nota pág. 30)

Ahora bien, supongamos que tenemos las siguientes combinaciones lineales de X_1 y X_2 . De tal manera que existen dos grupos de puntos en el plano, definido por X_1 y X_2 . Cada una de las combinaciones separa en mayor o menor medida a los grupos:

$$Z_1 = 0.80 X_1 - 0.60 X_2$$

$$Z_2 = 0.97 X_1 - 0.24 X_2$$

$$Z_3 = 0.98 X_1 - 0.18 X_2$$

$$Z_4 = 0.71 X_1 - 0.71 X_2$$

Al hacer los cálculos de la medida de separación de las distribuciones, se tiene que:

$$f^2(Z_1) = 3.46$$

$$f^2(Z_2) = 9.44$$

$$f^2(Z_3) = 10.65$$

$$f^2(Z_4) = 9.70$$

De esta manera observamos que la tercera combinación lineal separa más a los dos grupos. Así, la pregunta de interés en este momento sería: ¿Cómo encontrar la combinación lineal que tuviera una separación máxima?

Para esto es necesario expresar la medida de separación f^2 en términos de los coeficientes de la combinación lineal de X_1 y X_2 y entonces derivar parcialmente con respecto de a y b , para encontrar los coeficientes que maximizan a $f^2(a,b)$. En general, si se tienen p variables entonces se busca maximizar a $f^2(a_1, a_2, \dots, a_p)$. Ésta es precisamente la combinación lineal que se obtiene al aplicar el modelo de Análisis Discriminante, es decir, se determina la combinación lineal que separa lo más posible a los grupos.

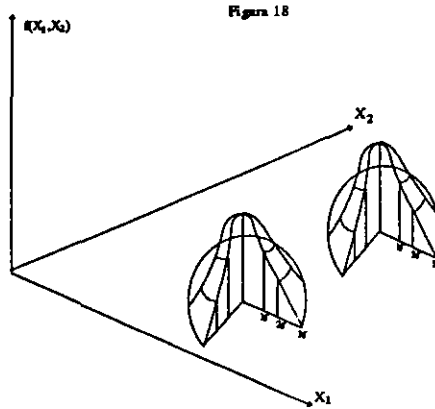
Supuestos del Análisis Discriminante

Para encontrar la combinación lineal que minimiza el grado de traslape de los grupos, el modelo de Análisis Discriminante establece los siguientes supuestos:

- 1) Cada grupo de variables X_1, X_2, \dots, X_p , tiene una ley de distribución normal p-variada.
- 2) Las matrices de varianza-covarianza de los grupos son iguales.

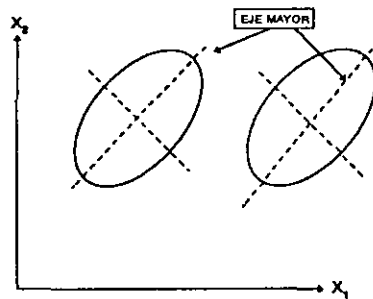
Cualquier violación a los supuestos anteriores, se refleja en la combinación lineal calculada, la cual no produce la separación máxima de las distribuciones proyectadas en el eje Z.

En la figura 18, se observa la interpretación geométrica del supuesto 1, en el caso en que $P=2$. Es decir, en este caso existen dos distribuciones en forma de campana y solamente difieren por la localización del centroide.



En la figura 19, se observa el plano formado por las variables X_1 y X_2 . Las elipses se forman al hacer cortes de las campanas con planos paralelos al plano formado por X_1 y X_2 . El supuesto 2, implica tener dos elipses exactamente del mismo tamaño, de tal manera que los ejes mayores son paralelos.

Figura 19



Distribuciones normales bivariadas con matrices de varianzas-covarianzas iguales, implican elipses iguales con ejes mayores paralelos

Gráficamente, la violación a los supuestos del modelo discriminante se observa en las figuras 20, 21 y 22.

Figura 20

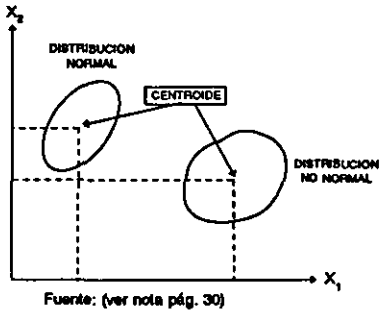
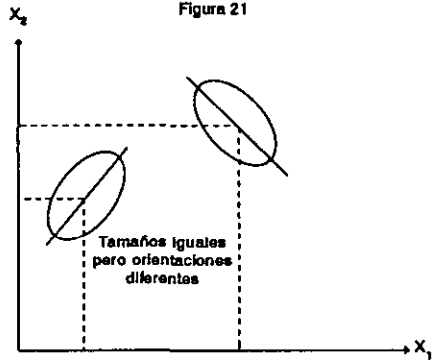
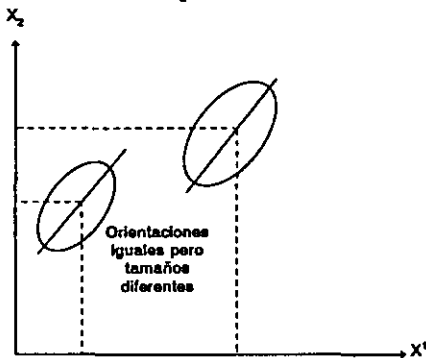


Figura 21



DISTRIBUCIONES NORMALES PERO MATRICES DE VARIANZAS-COVARIANZAS DIFERENTES

Figura 22



DISTRIBUCIONES NORMALES PERO MATRICES DE VARIANZAS-COVARIANZAS DIFERENTES

Es importante subrayar que, la interdependencia (correlación) entre variables afecta a cualquier combinación lineal. Dicho efecto se traduce en:

- 1) Puede cambiar el signo del coeficiente de una variable que el investigador espera de acuerdo a su experiencia que sea negativo o positivo; y
- 2) Se pueden estimar coeficientes muy grandes en algunas variables de la combinación lineal.

Por lo tanto, es necesario analizar las correlaciones entre pares de variables antes de aplicar la teoría de A.D. a un conjunto de datos. Si dos variables están altamente correlacionadas⁷ entonces, es necesario eliminar alguna de las dos, ya que no aporta mucha información adicional a Y. El criterio para la eliminación es, dejar la variable que se identifica más con los objetivos de la investigación.

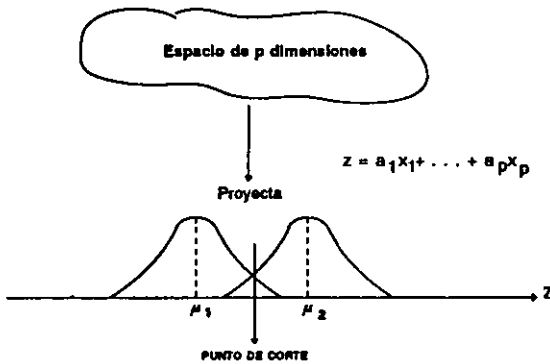
Regla para clasificar individuos

Sean A y B, dos grupos de individuos o elementos de una población en estudio y X_1, X_2, \dots, X_p variables continuas que representan mediciones de los individuos. Como se sabe el centroide de cada distribución se proyecta en la media μ_a y μ_b respectivamente.

Al punto medio entre μ_a y μ_b se le llama *punto de corte*, de tal manera que, los individuos se clasifican en el grupo A, si su puntaje (el valor de Z) es menor que el punto de corte; en caso contrario, se clasifica en el grupo B (figura 23).

⁷ En la práctica correlaciones mayores o iguales a 0.70 se consideran altas.

Figura 23



Fuente: (ver nota pág. 30)

Ejemplo:

Supóngase que se tiene la siguiente función discriminante para los grupos A y B:

$$Z = 0.030 X_1 + 0.204 X_2 + 0.010 X_3 + 0.443 X_4$$

y, que el vector de medias de los grupos está dado por:

$$\mu_A = (12.57, 9.57, 11.49, 7.97)$$

$$\mu_B = (8.75, 5.33, 8.5, 4.75)$$

Sustituyendo los valores de las medias de los grupos A y B en la función discriminante tenemos que:

$$Z_{\mu_A} = 5.97$$

$$Z_{\mu_B} = 3.54$$

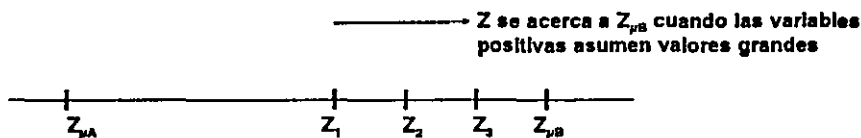
De tal forma que, el punto de corte (f) se calcula de la siguiente manera:

$$f = \frac{Z_{\mu_A} + Z_{\mu_B}}{2} = \frac{5.97 + 3.54}{2} = 4.76$$

Interpretación de los coeficientes

Cuando se tienen dos grupos, el enfoque tradicional que se usa para la interpretación de los coeficientes es, examinar el signo y la magnitud de los coeficientes estandarizados de cada variable en la combinación lineal. En general, las variables con coeficientes negativos estarán más asociadas con uno de los grupos mientras que, variables con coeficientes positivos con el otro grupo. Para determinar con cuál grupo están asociadas las variables se calcula la proyección de los centroides de estos grupos; de tal manera que, las variables con signos positivos se asocian con el grupo cuyo centroide sea el mayor de los dos. Ya que, los coeficientes positivos contribuyen más a que el valor de Z se localice al lado derecho del punto de corte (ver figura 24).

Figura 24



Fuente: (ver nota pág. 30)

Análisis Discriminante Múltiple

Considérense ahora G_1, G_2, \dots, G_k grupos de elementos de una población tal que:

- La unión de todos los grupos es igual a la población total.
- Un elemento de la población no puede pertenecer a más de un grupo.

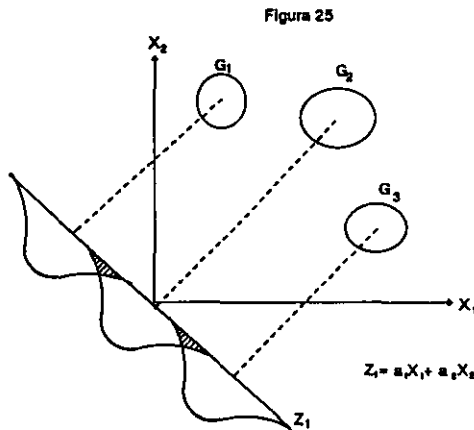
Ahora bien, si a cada uno de los individuos de la población se le asigna una variable Y_i que asume los siguientes valores:

- 1 Si el i -ésimo individuo pertenece al grupo 1
- 2 Si el i -ésimo individuo pertenece al grupo 2
- .
- .
- K Si el i -ésimo individuo pertenece al grupo K

Es importante dejar claro que, los valores anteriores son solamente etiquetas o categorías, implicando una medición de distancia u orden.

Por otro lado, si el interés es modelar las probabilidades $P(Y_i=1)$, $P(Y_i=2)$, \dots , $P(Y_i=K)$, el modelo adecuado es la **regresión politómica**, que es una extensión de la regresión logística. Pero, si el interés se centra en determinar las características que hacen diferentes a los individuos de los grupos G_1 , G_2 , \dots , G_k , entonces, la teoría adecuada es el Análisis Discriminante Múltiple (ADM).

Para aclarar ideas, supóngase que se tienen tres grupos de individuos y dos variables X_1 y X_2 medidas en cada uno de los grupos (figura 25).



Fuente: (ver nota pág. 30)

Al igual que en AD simple, el objetivo es buscar una combinación lineal que separe los puntos proyectados lo más posible. El primer problema en el caso múltiple es, que la medida de separación definida para dos grupos no se puede utilizar debido a que se tienen más de dos medias y por lo tanto no se puede definir de manera única. Por lo que la medida de separación para más de dos grupos se define en términos de la homogeneidad y de la heterogeneidad de los grupos debido a las siguientes razones:

- 1) La homogeneidad: es una medida de la similitud entre los individuos de un mismo grupo.
- 2) La heterogeneidad: es una medida de qué tan diferentes son los individuos entre grupos.

Como la homogeneidad se mide en función de la suma de cuadrados dentro de los grupos y la heterogeneidad se mide en función de la suma de cuadrados entre grupos, entonces la medida de separación para más de dos grupos se define como:

$$\lambda = \frac{\text{Suma de cuadrados entre grupos}}{\text{Suma de cuadrados dentro de los grupos}}$$

A λ se le conoce como el criterio discriminante y representa, como ya se mencionó una medida de separación entre más de dos grupos. Otra de las propiedades de λ es que también se puede aplicar cuando sólo existen dos grupos. Cabe resaltar que, la suma de cuadrados dentro de los grupos debe ser lo más pequeña posible, porque entre más pequeña sea, más parecidos serán los elementos de los grupos, mientras que la suma de cuadrados entre grupos debe ser lo más grande posible, ya que entre más grande sea, menos parecidos serán los elementos entre los mismos grupos. Por lo tanto, entre más grande sea λ , más heterogéneos y homogéneos simultáneamente serán los grupos. Es por esto que ADM busca expresar el criterio discriminante λ como una función de los coeficientes de la combinación lineal:

$$Z = a_1 X_1 + a_2 X_2 + \dots + a_p X_p$$

Y después, por métodos de cálculo diferencial encuentra los coeficientes a_i que hace que el criterio discriminante sea máximo.

En general, si se tienen K grupos, la teoría encuentra $k-1$ combinaciones lineales Z_1, Z_2, \dots, Z_{k-1} , también conocidas como funciones discriminantes, tales que:

$$Z_1 = a_{11} X_1 + a_{12} X_2 + \dots + a_{1p} X_p$$

Tiene el criterio discriminante λ más grande posible entre todas las combinaciones lineales de las p variables explicativas.

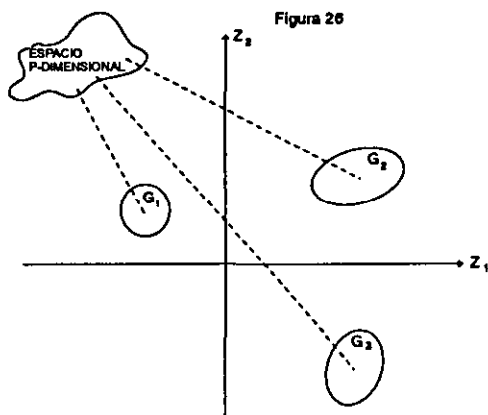
$$Z_2 = a_{21} X_1 + a_{22} X_2 + \dots + a_{2p} X_p$$

Tiene la siguiente propiedad: entre todas las combinaciones lineales de X_1, X_2, \dots, X_p , que no están correlacionadas con Z_1 , tiene el criterio discriminante λ más grande posible.

$$Z_k = a_{k1} X_1 + a_{k2} X_2 + \dots + a_{kp} X_p$$

Tiene la siguiente propiedad: entre todas las combinaciones lineales de X_1, X_2, \dots, X_p , que no están correlacionadas con Z_1, Z_2, \dots, Z_{k-1} , tiene el criterio discriminante λ más grande posible.

Ahora bien, supongamos que se tienen tres grupos G_1, G_2 y G_3 , en un espacio de p dimensiones; para obtener la máxima separación de los grupos, la proyección se hace en un plano definido por Z_1 y Z_2 (ver figura 26).



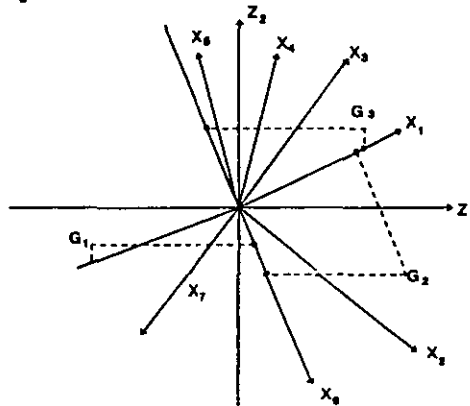
Fuente: (ver nota pág. 30)

Para clasificar a los individuos en el Análisis Discriminante Múltiple, se calcula la proyección de todos los puntos del espacio p -dimensional en el eje Z_1 y Z_2 , después se calcula la distancia entre el punto y los centroides proyectados en Z_1 y Z_2 , y se asigna al grupo cuya distancia es más pequeña. La clasificación se hace con base en Z_1 y Z_2 , porque proporcionan la máxima separación entre los grupos.

Interpretación de las variables

En la práctica para identificar qué variables caracterizan a los grupos involucrados, usualmente se grafican los centroides en las primeras 2 funciones discriminantes, **ya que dichas funciones tienen la propiedad de separar a los elementos (individuos, objetos, etcétera) lo más posible**. En la figura 27, se representa el caso de tres grupos y ocho variables.

Figura 27



Fuente: (ver nota pág. 30)

La figura muestra que la primera función discriminante es la principal fuente de diferencias entre los grupos dos y tres con respecto al grupo uno, porque los centroides aparecen muy separados con relación a Z_1 . La segunda función discriminante produce diferencias significativas entre los grupos uno y dos con respecto al grupo tres, porque en relación a Z_2 están muy separados.

La longitud y la dirección de las flechas de la figura 27 están dadas por los coeficientes de las variables que forman las combinaciones lineales de Z_1 y Z_2 y se utilizan para determinar en primer lugar qué variables están más asociadas con las funciones Z_1 y Z_2 y en segundo lugar, para determinar qué grupos tienen en mayor o en menor *medida* la característica asociada con la variable de interés.

Ejemplo:

Si X_1 midiera la escolaridad, entonces al trazar perpendiculares de los centroides de los tres grupos al vector definido por los coeficientes de la variable X_1 , en las combinaciones lineales de Z_1 y Z_2 , se concluye que los grupos G_1 y G_2 son más escolarizados que el grupo G_3 , porque las intersecciones están más cerca de la punta de la flecha de X_1 , y entre ellos la diferencia en escolaridad es pequeña, porque la separación de las perpendiculares asociadas con G_2 y G_3 es menor. Si X_2 midiera el número de hijos nacidos vivos, al trazar perpendiculares a ese vector, se concluye que, los grupos G_1 y G_2 tienen más hijos que el grupo G_3 , ya que las intersecciones están más cerca de la punta de la flecha de X_2 ; porque, entre más grandes son los valores de X_2 , más se contribuye a que el punto se localize cerca del grupo G_1 o del grupo G_2 .

Validación del modelo

El Análisis Discriminante calcula los coeficientes de la combinación lineal que maximiza la separación de los grupos bajo dos grandes supuestos, el primero de ellos se refiere a la normalidad en cada grupo y el segundo a la igualdad de matrices de varianza-covarianza entre éstos.

Una manera de validar el supuesto de normalidad es analizando cada una de las variables individualmente. Si alguna de ellas tiene una distribución no normal univariada entonces, la distribución del vector (X_1, X_2, \dots, X_p) no puede ser *normal p-variada*. De entrada, si alguna de las variables X_i hasta X_p son no continuas (dicotómicas, ordinales o numéricas) entonces, éstas no pueden tener una distribución normal univariada y, por lo tanto, el supuesto de *normal p-variada* tampoco se cumple.

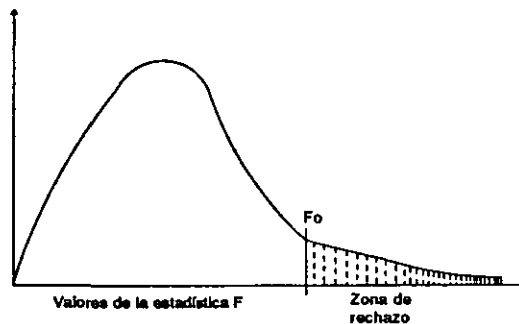
Existen varias pruebas para verificar la hipótesis de igualdad de matrices de dispersión. En este caso, hablaremos de la prueba de **BOX**. La hipótesis que se plantea es la siguiente:

H_0 : *Las matrices de varianzas-covarianzas entre grupos, son iguales.*

La estadística de prueba se construye bajo el supuesto de que H_0 es cierta, y se demuestra que tiene una ley de distribución F . La estadística se construye de tal manera que, si F toma valores cercanos a cero, entonces es un indicio de que H_0 no se rechaza; por el contrario, si la estadística toma valores grandes, indica que H_0 se rechaza.

En la figura 28, se muestra la zona de rechazo de la hipótesis anterior, de tal manera que, si la estadística toma valores mayores que F_0 , entonces H_0 se rechaza, es decir, se concluye que las matrices no son iguales. En caso contrario, H_0 no se rechaza y se concluye que las matrices son iguales.

Figura 28



Fuente: (ver nota pág. 30)

Es importante tener presente que, si no se cumplen los supuestos anteriores, entonces la combinación lineal obtenida para el caso de dos grupos, o las combinaciones lineales obtenidas para el caso de tres o más grupos, no separan a éstos de manera óptima. Sin embargo, la combinación lineal obtenida, separa en alguna medida a los grupos, pero las conclusiones que se obtengan con esta combinación lineal, deben tomarse con las reservas del caso.

El uso o no de los resultados obtenidos con una combinación que no es la óptima, depende finalmente del tipo de decisiones que tome el investigador. Por ejemplo, si la decisión está relacionada con la vida o la muerte de personas de la población, puede ser muy riesgoso usar los resultados. Si la decisión está asociada con asignar recursos de más o de menos a los elementos de la población, puede ser que no sea muy riesgoso usar los resultados.

2.3 Regresión Logística Simple

Supongamos que se tienen n individuos de una población, en los que se mide una variable continua X_i (edad, peso, talla, etcétera) y nos interesa investigar si existe alguna relación entre esta variable y la presencia o ausencia de un atributo de interés en los individuos. Si definimos para cada individuo una variable Y_i de la siguiente manera:

$$Y_i = \begin{cases} 1 & \text{si el } i\text{-ésimo individuo tiene el atributo de Interés.} \\ 0 & \text{si el } i\text{-ésimo individuo no tiene el atributo de Interés.} \end{cases}$$

Entonces el problema se traduce en encontrar la relación entre las variables X_i y Y_i . Como Y_i es aleatoria, la relación no puede estar dada en términos de tratar de predecir el valor que asumirá Y_i dado un valor específico de X_i , más bien, trataremos de predecir la probabilidad de los valores de Y_i .

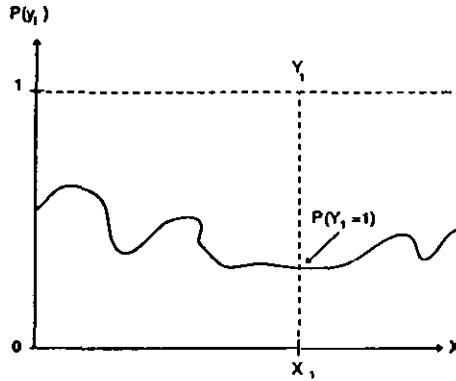
Como Y_i asume solamente dos posibles valores, entonces, si se determina la probabilidad de ocurrencia de alguno de ellos, por complemento se conocerá la probabilidad del otro. Es decir, si:

$$P(Y=1)=K \quad \text{entonces: } P(Y=0)=1-K; \quad \text{para } 0 \leq K \leq 1$$

Por lo anterior podemos suponer, sin perder generalidad que, el interés se centra en determinar si existe alguna manera de predecir el valor de $P(Y_i=1)$ dado que X_i asume algún valor específico.

Las probabilidades $P(Y_i=1)$, pueden tener un comportamiento arbitrario en términos de los valores de X_i , o, no tener ninguna relación. En la figura 29, la curva representa el comportamiento (arbitrario) de las probabilidades para las variables X_i ,

Figura 29



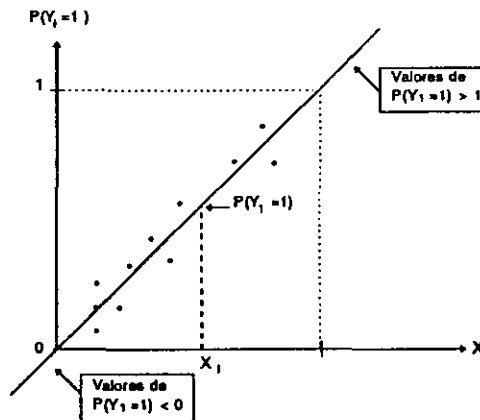
Fuente: (ver nota pág. 30)

Obsérvese que la curva se encuentra entre los valores 0 y 1, ya que se están modelando probabilidades. Un primer intento de modelar la situación anterior es suponer que:

$$P(Y_i=1) = \alpha + \beta X_i$$

Sin embargo, esta función no es completamente adecuada, ya que para algunas X_i la $P(Y_i=1)$, puede ser mayor que uno o menor que cero (ver figura 30), lo cual contradice la definición de probabilidad.

Figura 30



Fuente: (ver nota pág. 30)

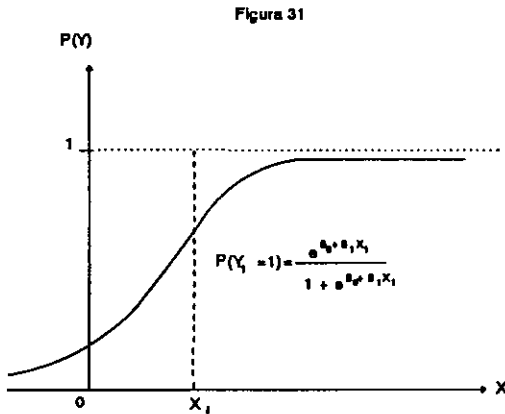
Es por esto que, para modelar probabilidades es necesario suponer que el comportamiento de $P(Y_i=1)$ en términos de las X_i está gobernado por una función matemática que solamente puede tomar valores entre 0 y 1 inclusive. Dicha función es conocida como la **función logística** y su ecuación es la siguiente:

$$P(Y_i=1) = \frac{e^{\beta_0 + \beta_1 X_i}}{1 + e^{\beta_0 + \beta_1 X_i}} \quad (1)$$

ya que parte de:

$$\ln \left[\frac{P(Y_{i-1})}{1 - P(Y_{i-1})} \right] = \beta_0 + \beta_1 X$$

La gráfica de la función se muestra en la figura 31.



Fuente: (ver nota pág. 30)

Otro supuesto adicional para completar el modelo es, que las variables aleatorias Y_i y Y_j para toda $i \neq j$, son estocásticamente independientes.

Es importante resaltar sin embargo que, en este caso no es necesario establecer ningún supuesto en relación a la **Ley de Distribución de Probabilidades** de Y_i , como en el modelo de regresión, ya que por la manera en que fue definida Y_i , ésta tiene una ley de distribución de probabilidades **Bernoulli**. El supuesto de varianza constante que se hace en regresión lineal, no se cumple en este modelo ya que:

$$\text{Var}(Y_i) = P(Y_i=1) P(Y_i=0)$$

y, como $P(Y_i=1)$ depende de X_i , entonces la varianza de Y_i también depende de X_i y por lo tanto es no constante.

Por otro lado, si en lugar de tener solamente una variable explicativa, se tienen X_1, X_2, \dots, X_p variables explicativas del comportamiento probabilístico de la variable aleatoria Y_i con *Ley de Distribución Bernoulli*, el modelo queda de la siguiente manera:

$$1. \quad E(Y_i) = P(Y_i=1) = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}} \quad (2)$$

2. Y_i y Y_j son independientes para $i \neq j$

Al igual que la regresión lineal, la fuerza del modelo (2), radica en que permite modelar el efecto de muchas variables, algunas de las cuales pueden estar en escalas de medida diferente⁶. Es importante resaltar que, si alguna de las variables independientes X_i son tales como sexo, trabajo, estado civil, etcétera, es inapropiado incluirlas en el modelo como si fueran continuas. Esto es básicamente porque los números usados para representar los diferentes niveles son meramente identificadores y no tienen ningún significado numérico, ya que los códigos de la variable respectiva sólo identifican y no cuantifican.

Por ejemplo, supóngase que, en una encuesta se le asigna el 1 a la categoría hombre y 2 a la categoría mujer, esto no quiere decir que 2 sea mayor que 1. Ante esta situación el método a usar es una colección de variables dicotómicas con códigos cero y uno.

Si la variable cualitativa tiene dos categorías es necesario construir una variable D_1 cuyos códigos son cero y uno. Si la variable cualitativa tiene tres categorías es necesario crear dos variables D_1 y D_2 cuyos códigos son cero y uno. En general si la variable cualitativa tiene k categorías se tendrán que generar $k-1$ variables dicotómicas con códigos cero y uno. En la siguiente tabla se muestran los códigos que asumen las variables D_i , según el código de X_i .

⁶ Las variables pueden ser continuas, categóricas, ordinales o cualitativas.

**CÓDIGO DE LA
VARIABLE CUALITATIVA**

VARIABLES DICOTÓMICAS

	D_1	D_2	...	D_{k-1}
1	1	0	...	0
2	0	1	...	0
3	0	0	...	0
.
k-1	0	0	...	1
k	0	0	...	0

Interpretación de los coeficientes

En la ecuación (1), si β_1 es positivo, entonces, a medida que la variable X_1 se incrementa es más probable que el individuo tenga el atributo de interés, mientras que si es negativo cuando X_1 se incrementa es menos probable que se presente el atributo.

En la ecuación (2), los coeficientes β 's positivos hacen que la probabilidad $P(Y_i=1)$ tienda a uno a medida que los valores de las variables respectivas crecen, mientras que los coeficientes negativos hacen que $P(Y_i=1)$ se acerque a cero a medida que los valores de las variables respectivas se incrementen.

Una ventaja de este modelo es, que existe otra manera de interpretar los coeficientes que permite comparar grupos de la población estudiada, la cual está dada en términos del cociente de momios. Este cociente es una medida de asociación entre dos variables que indica:

Qué tanto más probable es que y_i sea igual a 1, cuando otra variable (x_i) asume ciertos valores

Ejemplo:

Si Y_i denota la presencia (código 1) o ausencia (código 0) de cáncer y X_i denota a la persona que fuma (código 1) o no (código 0); entonces, un cociente de momios entre estas variables igual a 2 indica que el cáncer ocurre dos veces más entre fumadores que entre no fumadores.

Para interpretar el cociente de momios en el Modelo de Regresión Logística, consideremos los siguientes casos:

CASO I. Supongamos que X_i es dicotómica y que sus valores son 0 y 1, si:

$$P(Y_i=1) = \frac{e^{\beta_0 + \beta_1 X_i + \dots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_i + \dots + \beta_p X_p}}$$

El cociente de momios entre Y_i y X_i se calcula identificando el coeficiente de la variable X_i y se aplica la exponencial, es decir:

$$\psi = e^{\beta_i}$$

donde:

- ψ Representa el cociente de momios,
- β_i Es el coeficiente de la variable X_i en el modelo logístico y,
- e Representa la función exponencial.

El cociente se interpreta de la siguiente manera:

Qué tanto más probable es que $y=1$,
cuando $x=1$ con respecto a $x=0$

CASO II. Supongamos X_i con tres categorías 1, 2 y 3; en este caso se tienen que construir dos variables diseño

$D_1 = 1$ Si $X=1$
 $D_1 = 0$ en otro caso

$D_2 = 1$ Si $X=2$
 $D_2 = 0$ en otro caso

Si
$$P(Y_i=1) = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_{D1} X_{D1} + \beta_{D2} X_{D2} + \dots + \beta_m X_m + \dots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_{D1} X_{D1} + \beta_{D2} X_{D2} + \dots + \beta_m X_m + \dots + \beta_p X_p}}$$

el cociente de momios de Y_i y D_1 y Y_i y D_2 es respectivamente:

$\psi = e^{\beta_{D1}}$ y $\psi = e^{\beta_{D2}}$

los cuales se interpretan respectivamente, de la siguiente manera:

Qué tanto más probable es que $y=1$, cuando $x_i=1$ con respecto a $x_i=3$

Qué tanto más probable es que $y=1$, cuando $x_i=2$ con respecto a $x_i=3$

CASO III. Supongamos X_i continua. En este caso el cociente de momios $\Psi=e^{\beta}$ se interpreta de la siguiente manera:

En cuanto se incrementa la probabilidad o el riesgo de que $y=1$ por cada unidad de medida en que se incrementa x_i

Validación del modelo

Una manera de medir el grado de ajuste del modelo es a través de una tabla de clasificación, en la cual se contabilizan las observaciones que, de acuerdo al modelo se clasifican correctamente, tanto para $Y_i=1$ como para $Y_i=0$ y además con respecto al total. Para la clasificación, si $P(Y_i=1)$ calculada con el modelo es mayor o igual a 0.5, entonces a Y_i se le asigna el valor de 1, en caso contrario se asigna el 0. Ejemplo:

Observación codificada con	Asignado a		% correcto de clasificación
	1	2	
1	28	5	84.85
2	7	13	65.00
Total			77.36

En la tabla anterior se muestra que, de 33 observaciones codificadas con 1, el 84.85% fue clasificada correctamente y de 20 observaciones codificadas con 0, el 65% fue clasificado correctamente. También se reporta que el total de clasificación correcta es del 77.36%.

Otra manera de medir el ajuste es comparando las observaciones con las predicciones. Es decir, si se tienen:

Observaciones	Predicciones
Y_1	Y_{1p}
Y_2	Y_{2p}
.	.
.	.
.	.
Y_n	Y_{np}

Por lo que una medida individual de la discrepancia entre las predicciones y las observaciones podría ser:

$$(Y_i - Y_{ip})^2$$

De tal manera que, la discrepancia general estaría dada por:

$$D = \sum (Y_i - Y_{ip})^2$$

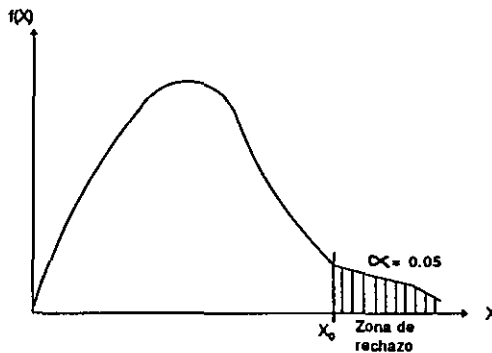
En términos generales, se busca que D sea pequeña e idealmente que sea igual a cero. Así, si el modelo ajusta bien a los datos, D debe ser pequeña, en caso contrario debe tomar valores grandes. Para probar si D es grande o pequeña se plantea la siguiente hipótesis:

H_0 : El modelo ajusta bien; la cual es equivalente a:

H_0 : D es pequeña.

La distribución de D , bajo la hipótesis de que el modelo es correcto, se distribuye como χ^2 con $n-(p+1)$ grados de libertad. Para probar H_0 se define una zona de rechazo que depende del nivel de significancia (ver figura 32).

Figura 32



Fuente: (ver nota pág. 30)

En la figura, X_0 denota el valor de una χ^2 con $n-(p+1)$ grados de libertad, tal que, a su derecha queda una probabilidad de 0.05. Si el valor de D al sustituir la información del modelo y las observaciones es mayor que X_0 , entonces se rechaza la hipótesis H_0 , es decir, se concluye que el modelo no ajusta bien porque las diferencias entre lo observado y lo predicho son grandes.

⁹ p representa el número de variables en el modelo logístico.

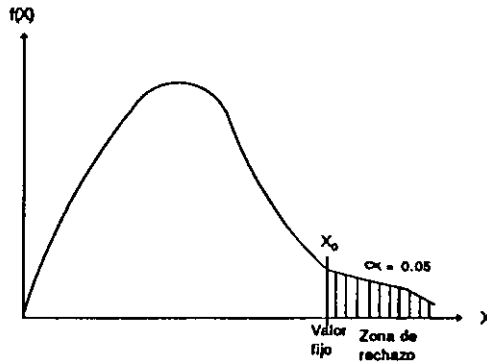
Otra manera de medir el grado de ajuste es probando la siguiente hipótesis:

$$H_0: \beta_1 = 0, \beta_2 = 0, \dots, \beta_p = 0$$

y que se refiere a que los coeficientes del modelo logístico son simultáneamente igual a cero. Esto significa también que $P(Y_i=1)$ no depende de las variables X_i .

Bajo el supuesto de que todos los coeficientes son cero, la distribución de la estadística para probar H_0 es una **Ji-cuadrada** con p grados de libertad. Los valores de la estadística que reflejarían que la hipótesis no se rechaza, son valores pequeños cercanos a cero. Por el contrario, si la estadística toma valores grandes, estaría indicando que la hipótesis H_0 es falsa; lo anterior se traduce en rechazar o no rechazar la hipótesis de acuerdo a un valor fijo que se determina según el nivel de significancia de la prueba (ver figura 33).

Figura 33



Fuente: (ver nota pág. 30)

Si el valor de la estadística es mayor que el valor fijo χ_0 , entonces se rechaza la hipótesis H_0 y se estaría concluyendo que al menos una de las λ 's es diferente de cero en el modelo de regresión logística. Si el valor es menor o igual que χ_0 , entonces no se rechaza H_0 y se estaría concluyendo implícitamente que no existe relación alguna entre la $P(Y_i=1)$ y las variables X_1, X_2, \dots, X_p .

2.4 Comparación de modelos

En el siguiente cuadro, se hace una comparación de los supuestos de los modelos de análisis discriminante y regresión logística.

CUADRO V. SUPUESTOS DE LOS MODELOS: ANÁLISIS DISCRIMINANTE Y REGRESIÓN LOGÍSTICA

ANÁLISIS DISCRIMINANTE		REGRESIÓN LOGÍSTICA	
1.	Variable dependiente dicotómica	1.	Variable dependiente dicotómica
2.	$z = a_1 X_1 + a_2 X_2 + \dots + a_p X_p$	2.	$P(Y_i=1) = \frac{e^{\beta_0 + \beta_1 X_1}}{1 + e^{\beta_0 + \beta_1 X_1}}$
3.	Cada grupo de variables X_1, X_2, \dots, X_p tienen una distribución normal p-variada	3.	Y_i y Y_j son independientes para $i \neq j$
4.	Las matrices de varianza-covarianza de los grupos son iguales		

Fuente: Elaboración propia

Con base en el cuadro comparativo, se puede concluir que el modelo de análisis discriminante tiene dos supuestos muy fuertes, el de **distribución normal p-variada** y el de **igualdad de matrices de varianza-covarianza**. El término fuerte, es en el sentido de que son conceptos muy técnicos que no son manejables por personas que solamente les interesa aplicar el modelo, los conceptos son generalmente manejados por expertos en estadística y probabilidad.

El modelo de regresión logística, no tiene supuestos fuertes, por lo que se pudiera pensar que esta es una ventaja del modelo, con respecto al de Análisis Discriminante. Sin embargo, el modelo discriminante es más fácil de interpretar para cuando la variable dependiente es politómica. En el modelo logístico la interpretación se complica aunado al hecho de que no existe en la actualidad *software* estadístico para un análisis completo del modelo.

CAPÍTULO 3

APLICACION DE LOS MODELOS

3.1 DESCRIPCIÓN DE DATOS

La aplicación de los modelos se hizo con base en dos muestras distintas. Las inferencias se obtuvieron mediante muestreos por cuotas, en el que los porcentajes de las diferentes clases sociales de la población de estudio (Distrito Federal) y de la muestra fueron los mismos. El tamaño de la primera muestra fue de 212 casos y el de la segunda muestra de 200 casos. En cada muestra se captaron características sociales, económicas y demográficas de usuarios de servicios bancarios.

Descripción de la primera muestra

El instrumento que se utilizó para captar la información fue un cuestionario y el método de recolección se realizó mediante una entrevista. De la lista de variables, se seleccionaron aquellas que permiten analizar el perfil¹⁰ económico, social y demográfico de la población de estudio; dichas variables son:

- nivel social
- edad agrupada
- sexo
- estado civil
- trabajo
- ocupación

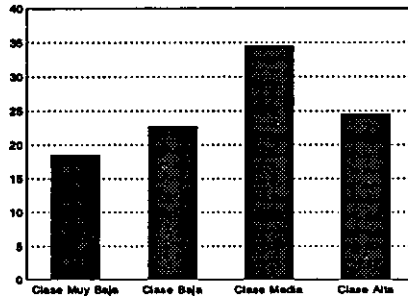
En las siguientes figuras se muestra la distribución de frecuencias para cada una de las variables antes mencionadas

¹⁰ Se entiende por perfil, al conjunto de características que describen a una persona.

Nivel social

En la figura 34, se puede observar la distribución porcentual de esta variable. La clase social *media*, es la que presenta el porcentaje más alto (34.5%); mientras que el más bajo corresponde a la clase social *muy baja*, con 18.4%.

FIGURA 34
DISTRIBUCION PORCENTUAL DEL NIVEL SOCIAL, 1992

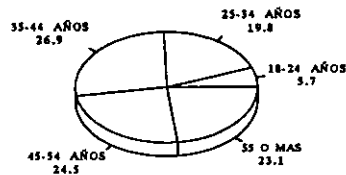


Fuente: Anexo I, listado A, pág 102.

Edad agrupada

En la figura 35, se muestra la distribución porcentual de la edad. Según los datos, la población se caracteriza por estar compuesta en su mayoría de personas adultas, ya que los porcentajes más altos se registraron en los grupos de mayor edad, es decir: 35-44 años, 26.9%; de 45-54 años, 24.5% y de 55 o más, 23.1%.

FIGURA 35
DISTRIBUCION PORCENTUAL DE LA EDAD, 1992

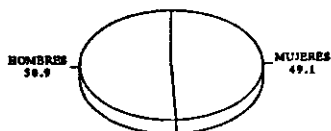


Fuente: Ibidem

Sexo

Con base en los datos se estima que en la población, la distribución de hombres y mujeres, es casi la misma (50.9%) y (49.1%) respectivamente (figura 36).

FIGURA 36
DISTRIBUCION PORCENTUAL DEL SEXO, 1992

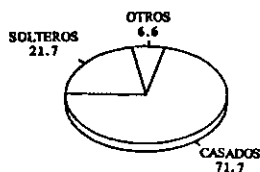


Fuente: Ibidem

Estado civil

La figura 37, muestra la distribución porcentual del estado civil, presentándose el mayor porcentaje en el grupo de personas casadas (71.7%). Los solteros y el código otros¹¹ representan el 21.7% y el 6.6% respectivamente.

FIGURA 37
DISTRIBUCION PORCENTUAL DEL ESTADO CIVIL, 1992



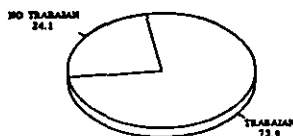
Fuente: Ibidem

¹¹ Para fines de este estudio, se agrupó en el rubro OTROS, a valores cuya frecuencia fue baja, como son: divorciados, unión libre y viudos.

Trabajo

En la figura 38, se observa que el porcentaje de personas que *trabajan* (75.9%), es mayor que el de aquellas que *no trabajan* (24.1%). Esto puede estar asociado al hecho de que la población de estudio se caracteriza por estar compuesta de personas adultas, ya que la probabilidad de trabajar crece a medida que aumenta la edad.

FIGURA 38
DISTRIBUCION PORCENTUAL DE PERSONAS QUE
TRABAJAN O NO, 1992

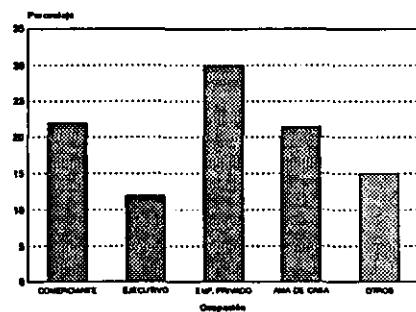


Fuente: *Ibidem*, pág. 103

Ocupación

La distribución porcentual de la ocupación (figura 39), se caracteriza por el hecho de que, el valor más alto corresponde al grupo de personas clasificadas como empleado privado (28.3%), y el más bajo para aquellas que se desempeñan como ejecutivos (11.3%)¹²

FIGURA 39
DISTRIBUCION PORCENTUAL DE LA OCUPACION, 1992



Fuente: *Ibidem*.

¹² En el código OTROS, de esta variable, se agruparon los siguientes valores: industrial, empleado de gobierno, persona física con empleo actual.

Descripción de la segunda muestra

En esta investigación, al igual que la anterior, el instrumento utilizado para captar la información fue un cuestionario y el método de recolección, una entrevista. Las variables que permiten analizar el perfil económico, social y demográfico de la población de estudio (usuarios bancarios), son:

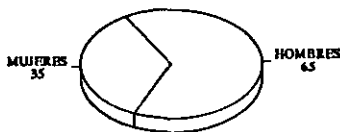
- sexo
- edad
- nivel social
- ocupación
- escolaridad del jefe de familia
- ingreso mensual familiar

En las siguientes figuras se muestra la distribución de frecuencias para cada una de las variables antes mencionadas.

Sexo

Según los datos de la muestra se estima que en la población, la distribución de hombres y mujeres es de 65% y 35% respectivamente, lo que representa un mayor porcentaje de hombres en la población.

FIGURA 40
DISTRIBUCION PORCENTUAL DE SEXO, 1992

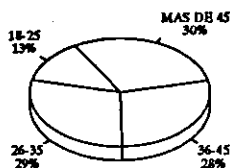


Fuente: Anexo I, listado B, pág. 104

Edad agrupada

En la figura 41, se muestra la distribución porcentual de la edad. Según los datos muestrales, la población se caracteriza por estar compuesta en su mayoría por personas mayores de 26 años, ya que los porcentajes más altos se registraron en los siguientes grupos: 26-35 (29%), 35-45 (28%) y de más de 45 (30%).

FIGURA 41
DISTRIBUCION PORCENTUAL DE EDAD, 1992

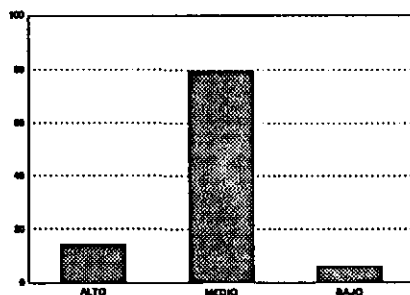


Fuente: Ibidem.

Nivel social

Según los datos de la muestra se estima que, la mayor parte de la población pertenece al nivel social medio (79.5%), siguiendo los de nivel alto (14.5%) y, por último, los de nivel bajo (6.0%) (figura 3).

FIGURA 3
DISTRIBUCION PORCENTUAL DE NIVEL SOCIAL, 1992



Fuente: Ibidem.

Ocupación

En la figura 43, observamos la distribución porcentual de la ocupación, según las estimaciones el valor más alto corresponde al grupo de personas clasificadas como profesionista libre (27%), mientras que el más bajo corresponde a aquellas cuya ocupación es empleado (12%)¹³.

FIGURA 43
DISTRIBUCION PORCENTUAL DE OCUPACION, 1992

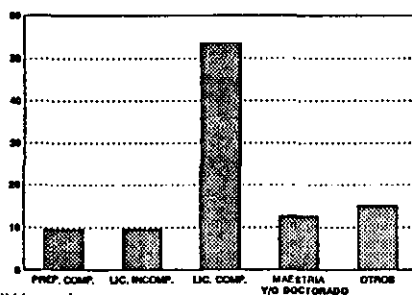


Fuente: *Ibidem*.

Escolaridad del jefe de familia

Según los datos de la muestra, se estima que la mayor parte de la población cuenta con licenciatura completa (53.3%) en tanto que, el nivel más bajo fue de preparatoria completa y licenciatura incompleta con un 9.5 % cada uno (figura 44).

FIGURA 44
DISTRIBUCION PORCENTUAL DE
LA ESCOLARIDAD DEL JEFE DE FAMILIA, 1992



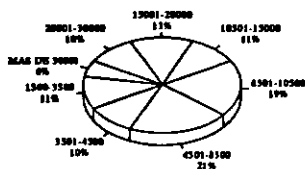
Fuente: *Ibidem*, pág. 105

¹³ En el código OTROS, de esta variable, se agruparon los siguientes valores: ejecutivo, técnico especializado, obrero calificado, obrero, mensajero, estudiante, ama de casa, jubilado/pensionado y desempleado.

Ingreso mensual familiar

En la figura 45, se muestra que la mayor parte de la población (21%) obtiene un ingreso mensual que se encuentra en el rango de \$ 4,501 a \$ 8,500, mientras que solamente un 6% de la población gana más de \$ 3,0001.

FIGURA 45
DISTRIBUCION PORCENTUAL DEL INGRESO
MENSUAL FAMILIAR



Fuente: *ibidem*.

3.2 ESTIMACIÓN DE PARÁMETROS EN ANÁLISIS DISCRIMINANTE

I. Perfil de usuarios de la primera muestra

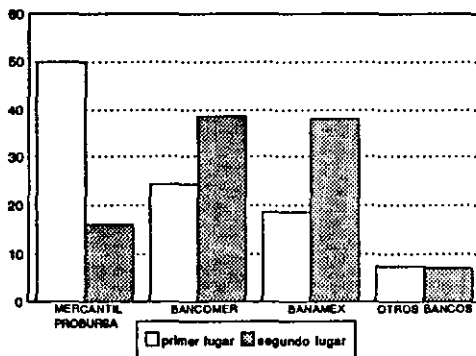
Ahora bien, como el objetivo de este estudio es analizar el perfil de los usuarios de bancos. En el cuestionario se planteó una pregunta que permitió detectar qué banco prefieren utilizar en primer o segundo lugar para sus servicios; es decir:

- Banco de preferencia en primer lugar y,
- Banco de preferencia en segundo lugar.

La distribución de la variable *preferencia* se muestra en la figura 46. Según los datos de la muestra, el banco que se prefiere en primer lugar es *Mercantil Probusca*, con un porcentaje de 48.1%. Mientras que, el que se prefiere después de éste es *Bancocmer* con

un porcentaje de 23.6%. De acuerdo a estos datos y con la preferencia en primer lugar se formaron los siguientes grupos: usuarios de Mercantil Probusa, usuarios de Bancomer, usuarios de Banamex y, por último, usuarios de Otros Bancos¹⁴, que incluye bancos cuya frecuencia fue baja.

FIGURA 46
DISTRIBUCION PORCENTUAL DE BANCOS
SEGUN PREFERENCIA, PRIMER Y SEGUNDO LUGAR, 1992



Fuente: Anexo I, listado A, pág 103.

Una vez que se determinan los grupos, las preguntas que surgen son: ¿Cuáles son las características de los usuarios de Mercantil Probusa? y ¿Cuáles las de los usuarios de Bancomer?. Es decir, en general cuáles son las características de las personas que utilizan el banco X.

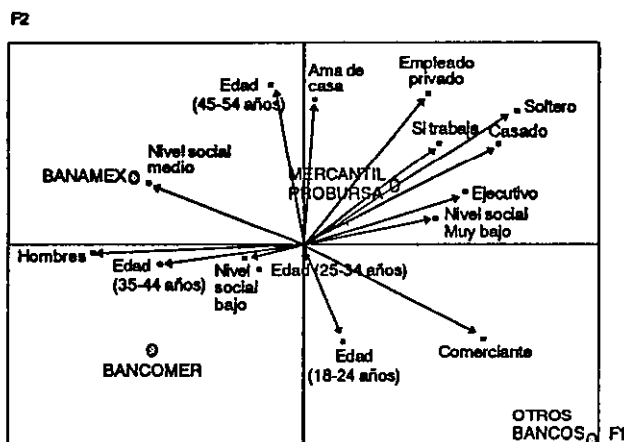
En este caso, el número de grupos que se formaron son 4 (Mercantil Probusa, Bancomer, Banamex y Otros). Las estimaciones se hicieron utilizando el modelo estadístico de Análisis Discriminante. Con los resultados de las primeras dos funciones discriminantes F_1 y F_2 (ya que éstas son las que separan más a los grupos), se definió una gráfica, donde el eje X toma los valores de F_1 y el eje Y los valores de F_2 . En primer lugar se grafican los centroides para definir la posición de los bancos.

Después se ubican las variables de interés (edad, nivel social, sexo, etcétera) utilizando los coeficientes que definen cada una de las combinaciones lineales F_1 y F_2 . Los

¹⁴ En el código OTROS, del banco de preferencia en primer lugar, quedaron agrupados los siguientes bancos: Serfin, Promex y Banorte.

resultados se pueden observar en la figura 47.

FIGURA 47
CARACTERISTICAS DE LOS USUARIOS BANCARIOS, 1992



Fuente: Anexo III, cuadros 1 y 2

Se puede analizar la figura, primero dividiendo en cuadrantes izquierdos y derechos, en los primeros se encuentran posicionados los bancos Banamex y Bancomer, asimismo, en los cuadrantes del lado derecho se encuentran posicionados Mercantil Probursa y Otros Bancos; se puede observar que las flechas que apuntan hacia los cuadrantes de el lado izquierdo corresponden a las siguientes variables: Edad (45-54 años), nivel social medio, hombres, edad (35-44 años), nivel social bajo y edad (25-34 años); por lo que, dichas variables son las que caracterizan a los bancos Banamex y Bancomer. De igual forma, las flechas que apuntan hacia el lado derecho corresponden a las siguientes variables: Ama de casa, empleado privado, trabaja, soltero, casado, ejecutivo, nivel social muy bajo, comerciante y edad (18-24 años); por lo que dichas variables son las que caracterizan a los bancos que se ubican en los cuadrantes del lado derecho (Mercantil Probursa y Otros Bancos).

Entonces, se puede decir, por ejemplo, que los usuarios de Mercantil Probursa y Otros Bancos tienen edades menores que los usuarios de Banamex y Bancomer; que el mayor porcentaje de personas que trabajan se ubican en Mercantil Probursa y Otros Bancos en relación con Banamex y Bancomer; además que la ocupación de los usuarios de estos

bancos es generalmente: empleado privado, ejecutivo, comerciante o ama de casa en relación con Banamex y Bancomer; etcétera.

El cuadro VI, muestra resumido el perfil de los usuarios bancarios, con base en los resultados antes mencionados.

CUADRO VI. PERFIL DE LOS USUARIOS BANCARIOS

BANCO	PERFIL
Banamex y Bancomer	Edad (45-54 años), nivel social medio, hombres, edad (35-44 años), nivel social bajo y edad (25-34 años)
Mercantil Probursa y Otros Bancos	Ama de casa, empleado privado, trabaja, soltero, casado, ejecutivo, nivel social muy bajo, comerciante y edad (18-24 años)

Fuente: figura 47, pág. 75

También al analizar la figura 47, es posible determinar las características de cada uno de los bancos. Es decir, podemos encontrar diferencias entre los usuarios de Banamex y Bancomer y, de manera completamente análoga se pueden encontrar diferencias entre los usuarios de Mercantil Probursa y de Otros Bancos. Los resultados se muestran en el cuadro VII.

CUADRO VII. PERFIL DE LOS USUARIOS BANCARIOS SEGÚN BANCO

BANCO	PERFIL
Banamex	Edad (45-54 años), nivel social medio
Bancomer	Hombres, edad (35-44 años), nivel social bajo y edad (25-34 años)
Mercantil Probursa	Ama de casa, empleado privado, trabaja, soltero, casado, ejecutivo, nivel social muy bajo
Otros Bancos	comerciante y edad (18-24 años)

Fuente: Ibidem.

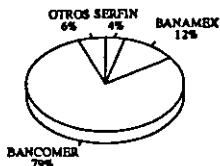
II. Perfil de usuarios de la segunda muestra

En este caso, también se incluyó una pregunta que permitió detectar el banco que prefieren utilizar las personas, para el manejo de sus cuentas. La pregunta fue:

- *Banco de preferencia en primer lugar*

La distribución de esta variable se muestra en la figura 48. Según los datos de la muestra, el banco que se prefiere en primer lugar es Bancomer, con un porcentaje de 79%; mientras que, Banamex ocupa el segundo lugar, con un porcentaje de 12%. De acuerdo a estos datos se formaron los siguientes grupos; usuarios de Bancomer, usuarios de Banamex, usuarios de Serfin y, por último, el grupo OTROS¹⁵, que incluye bancos cuya frecuencia fue baja.

FIGURA 48
DISTRIBUCION PORCENTUAL
DEL BANCO DE PREFERENCIA, 1992

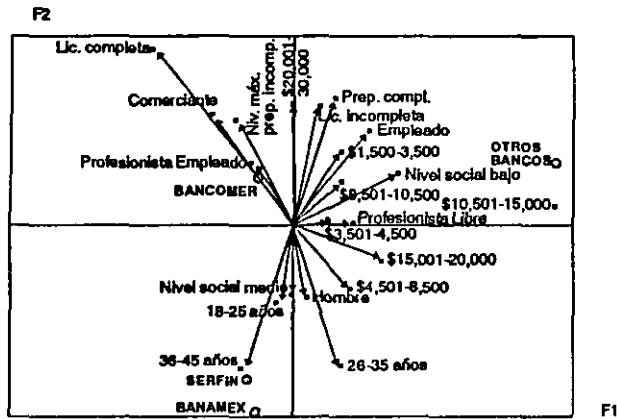


Fuente: Anexo I, Estado B, pág. 105

Siguiendo el procedimiento análogo al utilizado en el análisis anterior, se posicionaron tanto los bancos como las variables de interés (sexo, edad, nivel social, etcétera).

¹⁵ En el código OTROS, de esta variable, quedaron agrupados los siguientes bancos: Comermex y Atlántico.

FIGURA 49
CARACTERISTICAS DE LOS USUARIOS BANCARIOS



Fuente: anexo III, cuadro 3 y 4

De manera similar al ejemplo anterior, los resultados que a simple vista se pueden obtener con base en la figura 49 son: en los cuadrantes del lado izquierdo se encuentran posicionados los siguientes bancos: Bancomer, Serfin y Banamex, en este caso observamos que, en un cuadrante inferior izquierdo están posicionados dos bancos: Serfin y Banamex, lo que indica que dichos bancos son similares, (no necesariamente debe posicionarse un banco en cada cuadrante); asimismo, del lado derecho se ubica solamente Otros Bancos; las flechas que apuntan hacia el lado izquierdo, como ya se explicó anteriormente, corresponderán a las variables que caracterizan a los bancos ubicados de ese lado y son las siguientes: escolaridad del jefe de familia: hasta preparatoria incompleta o Licenciatura completa; comerciante, profesionista empleado; nivel social medio; edad: 18-25 años, 36-45 años. De manera similar, las flechas que apuntan hacia el lado derecho y que caracterizan a Otros Bancos corresponden a las siguientes variables: Ingreso mensual de: \$ 1,500-3,500, \$ 4,501-8,500, \$ 8,501-10,500, \$ 10,501-15,000, \$ 15,001-20,000, \$ 20,001-30,000, escolaridad del jefe de familia: preparatoria completa, licenciatura incompleta; empleado, profesionista libre; nivel social bajo, sexo masculino y edad de 26-35 años.

Entonces se puede decir; por ejemplo, que los usuarios de Otros Bancos, en general son de nivel social bajo, en comparación con Bancomer, Banamex y Serfin; que el mayor porcentaje de personas cuyo nivel máximo de estudios es de preparatoria completa se

encuentran en Otros Bancos, en relación con Bancomer, Banamex y Serfin; además, el mayor número de personas del sexo masculino se ubican en Otros Bancos, en relación con los bancos antes mencionados, etcétera.

El cuadro VIII, muestra resumido el perfil de los usuarios bancarios, con base en los resultados antes mencionados.

CUADRO VIII. PERFIL DE LOS USUARIOS BANCARIOS

BANCO	PERFIL
Bancomer, Serfin y Banamex	Ingresos de \$ 20,001-30,000, nivel máximo preparatoria incompleta, licenciatura completa, comerciante, profesionista empleado, nivel social medio, 18-25 años, 36-45 años.
Otros Bancos	Licenciatura incompleta, preparatoria completa, empleado, profesionista libre, ingresos: \$ 1,500-3,500, \$ 3,501-4,500, \$ 4,501-8,500 y \$ 8,500-10,500, \$ 10,501-15,000, sexo masculino, edad (26-35 años), nivel social bajo.

Fuente: figura 49, pág. 78

Analizando el cuadrante de la figura 49, en donde se posicionan Bancomer, Serfin y Banamex se pueden encontrar diferencias entre los usuarios de estos bancos. Los resultados se muestran en el cuadro IX.

CUADRO IX. PERFIL DE LOS USUARIOS BANCARIOS SEGÚN BANCO

BANCO	PERFIL
Bancomer	Ingresos de \$ 20,001-30,000, personas cuyo nivel máximo de estudios es de preparatoria incompleta, licenciatura terminada, comerciante, profesionista empleado.
Serfin y Banamex	Nivel social medio, edad (18-25, 26-35 y de 36-45 años), sexo masculino, ingresos de \$ 4,501-8,500 y de \$ 15,000-20,000.
Banamex	
Otros Bancos	Profesionista libre, empleado, ingresos de \$ 1,500-3,500, \$ 3,501-4,500, \$ 8,501-10,500 y de \$ 10,501-15,000, nivel social bajo, preparatoria completa, licenciatura incompleta.

Fuente: Ibidem.

3.3 ESTIMACIÓN DE PARÁMETROS EN REGRESIÓN LOGÍSTICA

I. Perfil de usuarios de la primera muestra

La Regresión Logística es otro método estadístico que permite determinar las características de los usuarios bancarios. El siguiente cuadro muestra los resultados obtenidos¹⁶ al aplicar este modelo.

CUADRO X

Variables	Banco			
	Banamex	Bancomer	Mercantil Probursa	Otros Bancos
Nivel social				
Muy Bajo			1.011	
Bajo		.314		
Medio		.602	.083	
Edad				
18-24 años		.699		1.190
25-34 años	.253	.281		.167
35-44 años	.861	.288		
45-54 años	.756		.238	
Sexo masculino	.293	.876		
Estado civil				
Soltero			1.376	
Casado			1.172	
Trabaja	.011		.656	.289
Ocupación				
Comerciante			.175	1.560
Ejecutivo			.862	.739
Empleado privado	.032		.818	
Ama de casa			.689	
Grado de ajuste ¹⁷	81.6	79.6	61.5	92.5

Fuente: Anexo III, cuadro 5.

¹⁶ Es importante aclarar que, el cuadro X sólo contiene las cifras positivas obtenidas en la aplicación del modelo, ya que son estas las que nos indican a que banco caracteriza cada variable y, en caso de ser más de uno, a cuál de ellos caracteriza mejor. En el anexo III, cuadro 5 se encuentran la totalidad de las cifras.

¹⁷ Este modelo estadístico proporciona el porcentaje de usuarios correctamente clasificadas por banco; a diferencia del Análisis Discriminante que proporciona esta información de manera general; es decir, del total de usuarios, el porcentaje que se clasificó correctamente dentro de cada banco.

Los valores en las celdas indican, el peso que tienen las variables en cada uno de los bancos. Si una variable caracteriza a más de un banco, entonces el valor más alto indicará a cuál de ellos caracteriza mejor; es así como que, con base en los datos del cuadro X, podemos concluir que la variable *nivel social muy bajo* caracteriza a Mercantil Probursa, *nivel social bajo* caracteriza a Bancomer y Mercantil Probursa, siendo el primer banco el que cuenta con un valor mayor, y por lo tanto, al que mejor caracteriza; *edad (18-24 años)* caracteriza a Bancomer y Otros Bancos, siendo este último el de mayor valor, entonces la variable caracteriza mejor a este banco, etcétera. Con base en lo anterior, se obtiene el siguiente perfil de usuarios (cuadro XI).

CUADRO XI. PERFIL DE USUARIOS BANCARIOS SEGÚN BANCO

BANCO	PERFIL
Banamex	Personas cuya edad está entre los 35 y 54 años
Bancomer	Nivel social bajo y medio, sexo masculino, 25-34 años
Mercantil Probursa	Nivel social muy bajo, personas casadas y solteras, trabajan, ejecutivo, empleado privado y ama de casa
Otros Bancos	Personas entre 18 y 24 años, comerciantes

Fuente: cuadro X, pág. 80

II. Perfil de usuarios de la segunda muestra

De igual forma se aplicó el modelo de Regresión Logística a la segunda muestra (ver cuadro XII¹⁸), los valores en las celdas indican, como en el caso anterior, el peso que tienen las variables en cada banco.

¹⁸ Al igual que en el caso anterior, este cuadro sólo contiene las cifras positivas obtenidas, para cotejar la totalidad de los datos ver anexo III, cuadro 6.

CUADRO XII

Variables	Banco			
	Bancomer	Banamex	Serfin	Otros Bancos
Sexo masculino		.753	1.415	.851
Edad				
18-25 años		1.333	.804	
26-35 años		1.524	1.555	
36-45 años		1.512	2.156	
Nivel social				
Bajo	.214			4.230
Medio		.793		
Ocupación				
Profesionista libre				1.131
Profesionista empleado	.888			
Comerciante	1.696			
Empleado	.916			1.801
Escolaridad del jefe de fam.				
Hasta preparatoria incompleta	1.091			
Preparatoria completa	.788		2.132	.063
Licenciatura incompleta	.902		1.218	.928
Licenciatura completa	1.547			
Ingresos				
\$ 1,500-3,500	.417		.827	7.640
\$ 3,501-4,500		.027	1.063	6.891
\$ 4,501-8,500		.301	9.591	7.295
\$ 8,501-10,500			11.310	7.523
\$ 10,501-15,000			.981	10.032
\$ 15,001-20,000			10.241	8.812
\$ 20,001-30,000	7.713		.598	
Grado de ajuste	80.5	88.5	96.5	95.5

Fuente: Anexo III, cuadro 8

Con base en los resultados, se concluye que, Banamex, Serfin y Otros Bancos se caracterizan por tener el mayor número de usuarios *hombres*, pero esta variable caracteriza mejor a Serfin, por ser el valor de este banco el mayor de los tres; la variable *edad (18-25 años)* caracteriza a los siguientes bancos: Banamex y Serfin, pero al que mejor caracteriza, es a Banamex por ser este el banco con el mayor valor; la variable *nivel social bajo* caracteriza mejor a Otros Bancos, la variable *comerciante* caracteriza a Bancomer, etcétera. El cuadro XIII, muestra en resumen las variables que caracterizan a cada banco, es decir, el perfil de usuarios bancarios.

CUADRO XIII. PERFIL DE USUARIOS BANCARIOS SEGÚN BANCO

BANCO	PERFIL
Bancomer	Profesionista empleado, comerciante, personas cuyo nivel máximo de estudios es preparatoria incompleta, licenciatura terminada, ingresos de \$20,001-30,000
Serfin	Sexo masculino, personas cuya edad está entre los 26-45 años, preparatoria completa, licenciatura incompleta, ingresos de \$ 4,500-8,500, \$ 8,501-10,500 y de \$ 15,001-20,000
Banamex	Edad (18-25 años), nivel social medio
Otros Bancos	Nivel social bajo, profesionista libre, empleado, ingresos de \$ 1,500-3,500, \$ 3,501-4,500 y de \$10,501-15,000

Fuente: cuadro XII, pág. 82

3.4 RESULTADOS

Los resultados que se obtuvieron, al aplicar cada uno de los modelos a una misma muestra, son similares; es decir, si comparamos las características de los usuarios bancarios de la primera muestra al aplicar el modelo de A.D. con las que se obtuvieron al aplicar el modelo de R.L, se observa que coinciden para cada uno de los bancos, siendo las diferencias mínimas (ver cuadro XIV).

**CUADRO XIV.
PRIMERA MUESTRA.**

PERFIL DE LOS USUARIOS BANCARIOS POR MODELO UTILIZADO SEGÚN BANCO

BANCO	PERFIL	
	ANÁLISIS DISCRIMINANTE	REGRESIÓN LOGÍSTICA
Banamex	Edad (45-54 años), <i>nivel social medio</i>	Edad (35-44 y de 45-54 años)
Bancomer	Hombres, edad (25-34 y de 35-44 años), nivel social bajo	Hombres, edad (25-34 años), nivel social bajo y <i>medio</i>
Mercantil Probursa	Ama de casa, empleado privado, trabaja, soltero, casado, ejecutivo, nivel social muy bajo	Ama de casa, empleado privado, trabaja, soltero, casado, ejecutivo, nivel social muy bajo
Otros Bancos	Comerciante, edad (18-24 años)	Comerciante, edad (18-24 años)

Fuente: cuadro VII, pág. 76

Fuente: cuadro XI, pág. 81

De igual forma, para la segunda muestra se obtuvieron los siguientes resultados para cada uno de los bancos, al aplicar los modelos de *AD* y *RL*.

CUADRO XV
SEGUNDA MUESTRA
PERFIL DE LOS USUARIOS BANCARIOS POR MODELO UTILIZADO SEGÚN BANCO

BANCO	PERFIL	
	ANÁLISIS DISCRIMINANTE	REGRESIÓN LOGÍSTICA
Bancomer	Ingresos de \$ 20,001-30,000, personas cuyo nivel máximo de estudios es de preparatoria incompleta, licenciatura terminada, comerciante, profesionista empleado	Profesionista empleado, comerciante, personas cuyo nivel máximo de estudios es preparatoria incompleta, licenciatura terminada, ingresos de \$ 20,001-30,000
Serfin	Nivel social medio, edad (18-25, 26-35 y de 36-45 años), sexo masculino, ingresos de \$ 4,501-8,500 y de \$ 15,000-20,000	Personas cuya edad está entre los 26-45 años, sexo masculino, preparatoria completa, licenciatura incompleta, ingresos de \$ 4,500-8,500, \$ 8,501-10,500 y de \$15,001-20,000
Banamex		Edad (18-25 años), nivel social medio
Otros Bancos	Profesionista libre, empleado, ingresos de \$ 1,500-3,500, \$ 3,501-4,500, \$ 8,501-10,500 y de \$ 10,501-15,000, nivel social bajo, preparatoria completa, licenciatura incompleta.	Profesionista libre, empleado, ingresos de \$1,500-3,500, \$ 3,501-4,500, y de \$ 10,501-15,000, nivel social bajo,

Fuente: cuadro IX, pág. 79

Fuente: cuadro XIII, pág. 83

Validación del modelo

Todo modelo es una simplificación de la realidad, pero la realidad no es tan simple, por lo que siempre tendrá asociado un grado de error. Validar un modelo significa evaluar el grado de error en que se incurre al usar los resultados del modelo en la explicación de la realidad. Por ejemplo, cada usuario pertenece a un banco en la realidad, el modelo

utiliza la información del usuario y lo clasifica en algún tipo de banco, al hacerlo puede clasificarlo correctamente o no; sin embargo, un buen modelo debe clasificar correctamente en la mayoría de los casos. Aún cuando por el tipo de información sea preferible identificar más exactamente a un grupo que al otro. Por ejemplo: en una prueba para la detección del SIDA es preferible que se cometan errores con personas que no están infectadas, al decir que si lo están, ya que se puede hacer una segunda prueba más detallada posteriormente. Sin embargo, un error más grave es diagnosticar que un individuo no tiene SIDA, cuando en realidad si tiene, debido entre otras cosas, a que no se someterá a una segunda prueba y, que además se convertirá en un potencial foco de infección.

Análisis Discriminante

Muestra I

Los datos para validar el modelo indican que en este estudio, dicho modelo explica un 37.95%, es decir, aproximadamente el 38% de la población total, se clasificó correctamente en los grupos de usuarios de banco definidos, en el siguiente cuadro se observa el porcentaje de correcta clasificación según lo observado y lo predicho por el modelo, como se puede ver la mayoría de los usuarios fueron clasificados correctamente.

CUADRO XVI. RELACIÓN DE CLASIFICACIONES

Observación codificada con:	No. de usuarios	Asignación de usuarios por el modelo			
		1	2	3	4
Grupo 1 (Mercantil Probursa)	97	30 30.9%	19 19.6%	24 24.7%	24 24.7%
Grupo 2 (Bancomer)	46	7 15.2%	16 34.8%	13 28.3%	10 21.7%
Grupo 3 (Banamex)	37	5 13.5%	8 21.6%	20 54.1%	4 10.8%
Grupo 4 (Otros bancos)	15	4 26.7%	2 13.3%	1 6.7%	8 53.3%

Fuente: anexo II, pág. 107

Porcentaje total de clasificación correcta: 37.95%.

Otra manera de validar el modelo es comparando lo que predice éste con los porcentajes reales, el siguiente cuadro muestra el perfil obtenido de los usuarios por banco según análisis discriminante y datos reales (anexo III, cuadro A).

CUADRO XVII. PERFIL DE USUARIOS POR BANCO SEGÚN MODELO

BANCO	Modelo Discriminante	Datos reales
Banamex	edad (45-54 años), nivel social medio	Edad (35-44 años), edad (45-54 años), empleado privado
Bancomer	hombres, edad (35-44 años), edad (25-34 años), nivel social bajo	Nivel social medio, hombres, casado, ama de casa
Mercantil Probursa	ama de casa, empleado privado, trabaja, soltero, casado, ejecutivo, nivel social muy bajo	ejecutivo, nivel social muy bajo
Otros Bancos	comerciante, edad (18-24 años)	Nivel social bajo, edad (18-24), edad (25-34 años), soltero, trabaja, comerciante

Fuente: Gráfica 47, pág. 75

Fuente: anexo III, cuadro A

Según este enfoque, el modelo discriminante predice que la proporción de individuos con edades de 45 a 54 años es mayor en Banamex. Al analizar los datos reales de este grupo de edad se observa que el 27.8% pertenece a Mercantil Probursa (MP), el 13% corresponde a Bancomer (Bc), el 32.4% a Banamex (Bx) y el 13% a Otros Bancos (Ob); siendo (Bx) el banco con el mayor porcentaje, por lo que en este caso ambos enfoques coinciden. Asimismo, según el modelo discriminante, la proporción de individuos con *nivel social medio* es mayor en Banamex, y los datos reales indican que el 30.9% pertenece a (MP), el 43.5% a (Bc), el 40.5% a (Bx) y, el 13.3 a (Ob), en este caso los mayores porcentajes corresponden a (Bc) y (Bx), con una diferencia de 3 puntos porcentuales, por lo que al ubicar esta variable en Banamex no se comete un error significativo; otra variable a analizar es la siguiente, según el modelo discriminante, el mayor porcentaje de *hombres* se encuentra en Bancomer, los datos reales indican que, el 47.4% pertenece a (MP), el 58.7% a (Bc), el 51.3% a (Bx) y, el 40.0% a (Ob), por lo que, ambos enfoques coinciden, como último ejemplo, la proporción de individuos con edades de 35 a 44 años es mayor en Bancomer según el modelo discriminante; los datos reales indican que, el banco que cuenta con el mayor porcentaje es (Bx) con 37.8%, sin embargo, Bancomer (32.6%) es el segundo que cuenta con un mayor porcentaje después de (Bx), con una diferencia de alrededor de 5 puntos porcentuales, lo que indica que al ser ubicada esta variable como característica de Bancomer no se comete error significativo, de manera similar se pueden analizar las variables restantes, y se observa que el perfil de usuarios obtenidos con ambos procedimientos pueden utilizarse indistintamente cometiendo un error que realmente no es significativo.

Regresión Logística

Los datos para validar el modelo indican que en este estudio, cada banco presentó los siguientes porcentajes de clasificación correcta: Mercantil Probursa (MP), 61.5%, Bancomer (Bc), 79.6%, Banamex (Bx), 81.6% y Otros Bancos (Ob), 92.5%, es decir, el porcentaje de población clasificada correctamente en los grupos de usuarios de banco definidos es alto. En los siguientes cuadros se observa el porcentaje de correcta clasificación según lo observado y la predicción del modelo en cada banco.

Mercantil Probursa

Observación codificada con:	Asignado a		Porcentaje correcto
	Usuarios	No usuarios	
Usuarios	64	34	65.3%
No usuarios	41	56	57.7%
		Total	61.5%

Bancomer

Observación codificada con:	Asignado a		Porcentaje correcto
	Usuarios	No usuarios	
Usuarios	153	2	98.7%
No usuarios	39	7	15.2%
		Total	79.6%

Banamex

Observación codificada con:	Asignado a		Porcentaje correcto
	Usuarios	No usuarios	
Usuarios	164	0	100.0%
No usuarios	37	0	0.0%
		Total	81.6%

Otros bancos

Observación codificada con:	Asignado a		Porcentaje correcto
	Usuarios	No usuarios	
Usuarios	186	0	100.0%
No usuarios	15	0	0.0%
		Total	92.5%

Fuente: anexo II, pág. 106

Al igual que en Análisis Discriminante, otro enfoque es comparar los resultados del modelo con los porcentajes reales, en el siguiente cuadro, se muestra el perfil obtenido de los usuarios por banco según el modelo de regresión logística y datos reales (anexo III, cuadro B).

CUADRO XVIII. PERFIL DE USUARIOS POR BANCO SEGÚN MODELO

BANCO	PERFIL DE USUARIOS POR BANCO SEGÚN MODELO	
	Regresión Logística	Datos reales
Banamex	Edad (35 a 44 años) Edad (45 a 54 años)	Edad (35-44 años) edad (45-54 años) empleado privado
Bancomer	Nivel social bajo Nivel social medio Hombres 25-34 años	Nivel social medio hombres casado ama de casa
Mercantil Probursa	Nivel social muy bajo casados solteros trabajan ejecutivo empleado privado ama de casa	ejecutivo nivel social muy bajo
Otros Bancos	Edad (18 a 24 años) comerciante	Nivel social bajo edad (18 a 24 años) edad (25 a 34 años) soltero trabaja comerciante

Fuente: cuadro XI, pág. 81

Fuente: anexo III, cuadro B

Según este enfoque, el modelo logístico predice que la proporción de individuos con edades de *35 a 44 años* es mayor en Banamex. Al analizar los datos reales de este grupo de edad se observa que el 23.7% pertenece a Mercantil Probursa (MP), el 32.6% corresponde a Bancomer (Bc), el 37.8% a Banamex (Bx) y el 20.0% a Otros Bancos (Ob); siendo (Bx) el banco con el mayor porcentaje, por lo que en este caso ambos enfoques coinciden. Asimismo, según el modelo logístico, la proporción de individuos con

edades de 45 a 54 años es mayor en Banamex, y los datos reales indican que el 27.8% pertenece a (MP), el 13.0% a (Bc), el 32.4% a (Bx) y, el 13.3 a (Ob), en este caso también coinciden los dos enfoques, ya que Banamex cuenta con el mayor porcentaje; otra variable a analizar es el *nivel social bajo*, según el modelo logístico, el mayor porcentaje de personas que pertenecen a este nivel se encuentra en Bancomer, mientras que los datos reales indican que, el 19.6% pertenece a (MP), el 21.7 a (Bc), el 24.3% a (Bx) y, el 26.7% a (Ob), aunque en este caso los bancos a los cuales les corresponde los mayores porcentajes son (Bx) y (Ob), Bancomer no difiere por mucho de estos porcentajes (entre dos y cinco puntos porcentuales) por lo que al ubicar esta variable como característica de dicho banco no se comete un error significativo, como último ejemplo, la proporción de individuos de *nivel social medio* es mayor en Bancomer según el modelo discriminante; los datos reales indican que, el banco que cuenta con el mayor porcentaje es (Bc) con 43.5%, por lo que en este caso también los enfoques coinciden.

Muestra II

Análisis Discriminante

De manera similar que con la muestra I, se explica la validación para la muestra II, los datos para validar el modelo para esta muestra indican que en este estudio, dicho modelo explica un 53.5%, es decir, aproximadamente el 54% de la población total, se clasificó correctamente en los grupos de usuarios de banco definidos, en el siguiente cuadro se observa el porcentaje de correcta clasificación según lo observado y lo predicho por el modelo; como se puede observar la mayoría de los usuarios fueron clasificados correctamente.

CUADRO XIX. RELACIÓN DE CLASIFICACIONES

Observación codificada con:	No. de usuarios	Asignado a			
		1	2	3	4
Grupo 1 (Bancomer)	158	78 49.4%	31 19.6%	30 19.0%	19 12.0%
Grupo 2 (Banamex)	23	3 13.0%	15 65.2%	2 8.7%	3 13.0%
Grupo 3 (Serfin)	8	1 12.5%	1 12.5%	6 75.0%	0 0.0%
Grupo 4 (Otros bancos)	11	2 18.2%	0 0.0%	1 9.1%	8 72.7%

Fuente: Anexo II, pág. 107

Porcentaje total de clasificación correcta: 53.5%.

Como ya se ha explicado existe otra forma de validar el modelo y es comparando lo que predice éste con los porcentajes reales, el siguiente cuadro muestra el perfil obtenido de los usuarios por banco según Análisis Discriminante y datos reales (anexo III, cuadro B).

CUADRO XX. PERFIL DE USUARIOS POR BANCO SEGÚN MODELO

BANCO	Análisis Discriminante	Datos reales
Bancomer	Ingresos: \$ 20,001-30,000	edad: 18-25 años
	preparatoria incompleta	preparatoria incompleta
	licenciatura terminada	ingresos: \$ 1,500-3,500
	comerciante	\$ 3,501-4,500
	profesionista empleado	\$ 20,001-30,000
Serfin	Nivel social medio	sexo masculino
	edad: 18-25 años	profesionista empleado
	26-35 años	comerciante
	36-45 años	ingresos: \$ 8,501-10,500
	sexo masculino	\$ 15,000-20,000
Banamex		edad: 36-45 años
		nivel social medio
Otros Bancos		ingresos: \$ 4,501-8,500 ¿?
		licenciatura terminada
	Profesionista libre	Profesionista libre
	empleado	empleado
	ingresos: \$ 1,500-3,500	edad: 26-35 años
\$ 3,501-4,500	ingresos: \$ 10,501-15,000	
\$ 8,501-10,500	nivel social bajo	
\$ 10,501-15,000	preparatoria completa	
nivel social bajo	licenciatura incompleta	
preparatoria completa		
licenciatura incompleta		

Fuente: cuadro IX, pág. 79

Fuente: anexo III, cuadro B

Según este enfoque, el modelo discriminante predice que la proporción de individuos con ingresos entre \$ 20,000 a 30,000 es mayor en Bancomer. Al analizar los datos reales de este grupo de ingresos se observa que el mayor porcentaje pertenece a Bancomer con el 12.6% por lo que ambos enfoques coinciden. Asimismo, según el modelo discriminante, la proporción de individuos con *preparatoria incompleta* es mayor en Bancomer, y los datos reales indican que el 16.5% pertenece a Bancomer (Bc), el 8.7% a Banamex (Bx), el 12.5% a Serfin (Sr) y, el 9.1 a Otros Bancos (Ob), por lo que también en este caso

ambos enfoques coinciden al presentar a Bancomer con el mayor porcentaje; otra variable a analizar es *licenciatura terminada*, según el modelo discriminante, el mayor porcentaje de personas con esta característica pertenecen a Bancomer, y los datos reales indican que, el 55.7% pertenece a (Bc), el 56.5% a (Bx), el 37.5% a (Sr) y, el 27.3% a (Ob), (Bx) cuenta con el mayor porcentaje, sin embargo, Bancomer es el segundo banco que cuenta con un mayor porcentaje después de (Bx), con una diferencia de cerca de 1 punto porcentual, lo que indica que al ser ubicada esta variable como característica de Bancomer no se comete error significativo, como último ejemplo, la proporción de individuos con actividad de *comerciante* es mayor también en Bancomer según el modelo discriminante; los datos reales indican que, el banco con el mayor porcentaje de personas que se dedican al comercio es Serfin (25.0%), sin embargo, Bancomer (17.7%) es el segundo banco que cuenta con un mayor porcentaje después de Serfin, con una diferencia de alrededor de 7 puntos porcentuales, lo indica, de igual manera que, al ser ubicada esta variable como característica de Bancomer no se comete error significativo, de manera similar se pueden analizar las variables restantes, y observar que el perfil de usuarios obtenido en ambos procedimientos pueden utilizarse indistintamente cometiendo un error que realmente no es significativo.

Regresión Logística

Los datos para validar el modelo indican que en este estudio, cada banco presentó los siguientes porcentajes de clasificación: Bancomer (Bc), 80.5%, Banamex (Bx), 88.5%, Serfin (Sr), 96.5% y Otros Bancos (Ob), 95.5%, es decir, el porcentaje de población clasificada correctamente en los grupos de usuarios de banco definidos es alto. En los siguientes cuadros se observa el porcentaje de correcta clasificación según lo observado y la predicción del modelo en cada banco.

Bancomer

Observación codificada con:	Asignado a:		Porcentaje correcto
	Usuarios	No usuarios	
Usuarios	8	34	19.0%
No usuarios	5	153	96.8%
		Total	80.5%

Banamex

Observación codificada con:	Asignado a:		Porcentaje correcto
	Usuarios	No usuarios	
Usuarios	175	2	98.9%
No usuarios	21	2	8.7%
		Total	88.5%

Serfin

Observación codificada con:	Asignado a:		Porcentaje correcto
	Usuarios	No usuarios	
Usuarios	192	0	100.0%
No usuarios	7	1	12.5%
		Total	96.5%

Otros Bancos

Observación codificada con:	Asignado a:		Porcentaje correcto
	Usuarios	No usuarios	
Usuarios	188	1	99.5%
No usuarios	8	3	27.3%
		Total	95.5%

Fuente: Anexo II, pág. 109

Al igual que en Análisis Discriminante, otro enfoque es comparar los resultados del modelo con los porcentajes reales, en el siguiente cuadro se muestra el perfil obtenido de los usuarios por banco según el modelo de regresión logística y datos reales.

CUADRO XXI. PERFIL DE USUARIOS POR BANCO SEGÚN MODELO

PERFIL DE USUARIOS POR BANCO SEGÚN MODELO		
BANCO	Regresión Logística	Datos reales
Bancomer	Profesionista empleado	edad: 18-25 años
	comerciante	preparatoria incompleta
	preparatoria incompleta	ingresos: \$ 1,500-3,500
	licenciatura terminada	\$ 3,501-4,500
	ingresos: \$ 20,001-30,000	\$ 20,001-30,000
Serfin	Sexo masculino	sexo masculino
	personas entre los 26-45 años,	profesionista empleado
	preparatoria completa	comerciante
	licenciatura incompleta	ingresos: \$ 8,501-10,500
	ingresos: \$ 4,500-8,500	\$ 15,000-20,000
	\$ 8,501-10,500	
	\$ 15,001-20,000	
Banamex	Edad (18 a 25 años)	edad: 36 a 45 años
	nivel social medio	nivel social medio
		ingresos: \$ 4,501-8,500
		licenciatura terminada
Otros Bancos	Nivel social bajo	Profesionista libre
	profesionista libre	empleado
	empleado	edad: 26-35 años
	ingresos: \$ 1,500-3,500,	ingresos: \$ 10,501-15,000
	\$ 3,501-4,500	nivel social bajo
\$ 10,501-15,000	preparatoria completa	
	licenciatura incompleta	

Fuente: cuadro XIII, pág. 83

Fuente: anexo III, cuadro B

Según el modelo logístico, la proporción de individuos cuya actividad es *profesionista empleado* es mayor en Bancomer. Al analizar los datos reales se observa que el 18.3% pertenece a Bancomer (Bc), el 21.7% corresponde a Banamex (Bx), el 25.0% a Serfin (Sr) y el 9.1% a Otros Bancos (Ob); siendo (Sr) y (Bx) los bancos con el mayor

porcentaje, y en tercer lugar se encuentra Bancomer, en este caso no coinciden los resultados, sin embargo, la diferencia entre Bancomer y estos dos bancos se encuentra entre los 4 y 7 puntos porcentuales. Asimismo, según el modelo logístico, la mayor proporción de personas del *sexo masculino* se encuentran en Serfin, y los datos reales indican que el 62.0% pertenece a (Bc), el 73.9% a (Bx), el 87.5% a (Sr) y, el 72.7% a (Ob), en este caso coinciden los dos enfoques, ya que Serfin cuenta con el mayor porcentaje de hombres; otra variable a analizar es la siguiente, según el modelo logístico, el mayor porcentaje de personas cuyas edades están entre los 18 y 25 años se ubican en Banamex, y los datos reales indican que, el 13.3% pertenece a (Bc), el 13.0 a (Bx), el 12.5% a (Sr) y, el 9.1% a (Ob), aunque en este caso el banco al cual le corresponde el mayor porcentaje es (Bc), (Bx) no difiere por mucho de este porcentaje, lo que indica que al ser ubicada esta variable como característica de (Bx) no se comete un error significativo, como último ejemplo, la proporción de individuos de nivel social bajo es mayor en Otros Bancos según el modelo logístico; los datos reales nos indican que, el banco que cuenta con el mayor porcentaje es (Ob) con 18.2%, por lo que en este caso también los enfoques coinciden.

CONCLUSIONES

De tipo general

Con base en los resultados obtenidos, se concluye que se puede utilizar cualquiera de los modelos (Análisis Discriminante y Regresión Logística) para determinar las características de los usuarios bancarios y, en general de cualquier grupo de interés, siempre y cuando dicho grupo esté definido de forma similar al de este trabajo; por ejemplo, nunca usuarias de métodos anticonceptivos, usuarias y ex-usuarias; personas que compran en tiendas Liverpool, en Palacio de Hierro y en Sears.

Cabe mencionar, que los modelos deben utilizarse de manera complementaria, por lo que no se puede afirmar que un modelo sea mejor que otro y, en caso de que se tuviera que elegir entre los resultados de los dos modelos utilizados, se tendría que seleccionar el modelo que resulte con un mayor grado de ajuste.

De tipo particular

Con base en los resultados obtenidos, para la primer muestra, se tiene que, las características de los usuarios bancarios según banco son las siguientes: los usuarios de Banamex se caracterizan por ser personas adultas, es decir, con edad entre los 45 y 54 años. Los usuarios de Bancomer se caracterizan por ser del *sexo masculino*, con edades entre los 25-44 años y de *nivel social bajo*. Los usuarios de Mercantil Probusa se caracterizan por ser *solteros, casados, de nivel social muy bajo, personas que cuentan con empleo, amas de casa, empleados privados o ejecutivos* y, por último, las variables que caracterizan a los usuarios de Otros Bancos son: *comerciantes, así como personas jóvenes (18-24 años)*.

Para el caso de la segunda muestra se tiene que, en general, las características de los usuarios de Bancomer son: personas cuyo ingreso está entre \$ 20,001-30,000, personas cuyo nivel máximo de estudios es de *preparatoria incompleta, licenciatura terminada* y cuya ocupación es *comerciante o profesionista empleado*; los usuarios de Serfin son personas cuya edad se encuentra entre los 26 y los 45 años; cuya ocupación es

ejecutivo, profesionista empleado, profesionista libre, comerciante, técnico especializado y cuyo grado máximo de estudios alcanzado es de preparatoria incompleta y, por último las personas usuarias de Otros Bancos son en general hombres.

La información obtenida del perfil de usuarios, puede ser utilizada para atraer clientes (potenciales) o a un grupo específico de éstos, por ejemplo, si Mercantil Probusa decidiera atraer nuevos clientes, con base en la información obtenida en este trabajo, debería investigar lo siguiente: los programas, revistas, periódicos, etcétera, que acostumbran ver y/o leer sus clientes (que se caracterizan, como ya hemos visto anteriormente por ser solteros, casados, de nivel social muy bajo, personas que trabajan, ama de casa, empleado privado o ejecutivo), para promocionar en dichos medios de comunicación sus servicios y así llegar de una manera eficaz y obtener mejores resultados en su plan de atraer clientes potenciales. Otra manera de utilizar esta información, sería para atraer específicamente a un grupo de clientes, por ejemplo, en Bancomer, los datos muestran que este grupo de usuarios está formado en su mayoría por hombres, personas de nivel social bajo y personas entre 25 y 44 años, por lo que, al banco le puede interesar atraer ahora a clientes del sexo femenino, para esto necesita llevar a cabo otra investigación sobre cuáles son las características que llaman la atención a las mujeres para elegir un banco, y de esa forma hacer los cambios necesarios para atraer a este grupo de usuarios.

Análisis Multivariado: Análisis de más de dos variables simultáneamente.

Confiabilidad: Grado hasta el cual el proceso de medición está libre de errores aleatorios.

Cuestionario: Procedimiento formalizado para recolección de datos de los encuestados.

Desviación estándar: Medida de dispersión para datos de intervalos.

Grados de libertad: Cantidad de observaciones independientes sobre la variable de interés menos la cantidad de estadísticos calculados.

Hipótesis: Enunciado conjetural sobre el valor de una variable o la relación entre dos o más variables.

Hipótesis nula: Hipótesis que afirma que un parámetro de la población obtiene un valor o conjunto de valores determinado.

Hipótesis alternativa: Hipótesis que niega lo que se enuncia en la hipótesis nula.

Media: Medida de tendencia central para datos de intervalo como el valor promedio.

Medidas de dispersión: tipo de estadístico descriptivo que incluye medidas como desviación estándar, coeficiente de variación, rango intercuartil, y frecuencias relativas y absolutas.

Mezcla de mercadeo: Aquellas variables independientes, sobre las cuales una organización ejerce cierto grado de control (producto, precio, plaza y promoción).

Nivel de confianza: Probabilidad de no rechazar la hipótesis nula cuando es verdadera

LISTADO A
FRECUENCIAS DE LA PRIMERA MUESTRA

NIVEL SOCIAL

Value Label	Value	Frequency	Percent	Percent	Percent
Muy baja	1.00	39	18.4	18.4	18.4
Baja	2.00	48	22.6	22.6	41.0
Media	3.00	73	34.4	34.4	75.5
Alta	4.00	52	24.5	24.5	100.0
		-----	-----	-----	
	Total	212	100.0	100.0	
Valid cases	212	Missing cases	0		

EDAD AGRUPADA

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
18-24	1.00	12	5.7	5.7	5.7
25-34	2.00	42	19.8	19.8	25.5
35-44	3.00	57	26.9	26.9	52.4
45-54	4.00	52	24.5	24.5	76.9
55 o más	5.00	49	23.1	23.1	100.0
		-----	-----	-----	
	Total	212	100.0	100.0	
Valid cases	212	Missing cases	0		

SEXO

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Mujer	.00	104	49.1	49.1	49.1
Hombre	1.00	108	50.9	50.9	100.0
		-----	-----	-----	
	Total	212	100.0	100.0	
Valid cases	212	Missing cases	0		

EDO CIV

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Soltero	1.00	46	21.7	21.7	21.7
Casado	2.00	152	71.7	71.7	93.4
Otro	3.00	14	6.6	6.6	100.0
		-----	-----	-----	
	Total	212	100.0	100.0	
Valid cases	212	Missing cases	0		

TRABAJA

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
No	.00	51	24.1	24.1	24.1
Si	1.00	161	75.9	75.9	100.0
		-----	-----	-----	
	Total	212	100.0	100.0	
Valid cases	212	Missing cases	0		

OCUPACION

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Comerciante	1.00	44	20.8	21.9	21.9
Ejecutivo	3.00	24	11.3	11.9	33.8
Empleado privado	5.00	60	28.3	29.9	63.7
Ama de casa	6.00	43	20.3	21.4	85.1
Otros	8.00	30	14.2	14.9	100.0
	.	11	5.2	Missing	
		-----	-----	-----	
	Total	212	100.0	100.0	
Valid cases	201	Missing cases	11		

PRIMER BANCO DE PREFERENCIA

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
MercProb	1.00	102	48.1	49.8	49.8
Bancomer	2.00	50	23.6	24.4	74.1
Banamex	3.00	38	17.9	18.5	92.7
Otros	4.00	15	7.1	7.3	100.0
	.	7	3.3	Missing	
		-----	-----	-----	
	Total	212	100.0	100.0	
Valid cases	205	Missing cases	7		

SEGUNDO BANCO DE PREFERENCIA

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
MercProb	1.00	31	14.6	16.0	16.0
Bancomer	2.00	75	35.4	38.7	54.6
Banamex	3.00	74	34.9	38.1	92.8
Otros	4.00	14	6.6	7.2	100.0
	.	18	8.5	Missing	
		-----	-----	-----	
	Total	212	100.0	100.0	
Valid cases	194	Missing cases	18		

LISTADO B
FRECUENCIAS DE LA SEGUNDA MUESTRA

SEXO

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Mujer	.00	70	35.0	35.0	35.0
Hombre	1.00	130	65.0	65.0	100.0
		Total	200	100.0	100.0
Valid cases	200	Missing cases	0		

EDAD AGRUPADA

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
18-25	1.00	26	13.0	13.0	13.0
26-35	2.00	58	29.0	29.0	42.0
36-45	3.00	56	28.0	28.0	70.0
Más de 45	4.00	60	30.0	30.0	100.0
		Total	200	100.0	100.0
Valid cases	200	Missing cases	0		

NIVEL SOCIAL

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Baja	1.00	12	6.0	6.0	6.0
Media	2.00	159	79.5	79.5	85.5
Alta	3.00	29	14.5	14.5	100.0
		Total	200	100.0	100.0
Valid cases	200	Missing cases	0		

OCUPACION

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Profesionista libre	2.00	54	27.0	27.0	27.0
Profesionista empleado	3.00	37	18.5	18.5	45.5
Comerciante	4.00	31	15.5	15.5	61.0
Empleado	7.00	24	12.0	12.0	73.0
Otro	8.00	54	27.0	27.0	100.0
		Total	200	100.0	100.0
Valid cases	200	Missing cases	0		

ESCOLARIDAD DEL JEFE DE FAMILIA

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Prep. incompleta	1.00	30	15.0	15.0	15.0
Prep. completa	2.00	19	9.5	9.5	24.5
Lic. incompleta	3.00	19	9.5	9.5	34.0
Lic. completa	4.00	107	53.5	53.5	87.5
Maestría y doctorado	5.00	25	12.5	12.5	100.0
		-----	-----	-----	
	Total	200	100.0	100.0	
Valid cases	200	Missing cases	0		

INGRESO MENSUAL FAMILIAR

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
1,500-3,500	1.00	22	11.0	11.0	11.0
3,501-4,500	2.00	20	10.0	10.0	21.0
4,501-8,500	3.00	42	21.0	21.0	42.0
8,501-10,500	4.00	38	19.0	19.0	61.0
10,501-15,000	5.00	21	10.5	10.5	71.5
15,001-20,000	6.00	26	13.0	13.0	84.5
20,001-30,000	7.00	20	10.0	10.0	94.5
MAS DE 30,001	8.00	11	5.5	5.5	100.0
		-----	-----	-----	
	Total	200	100.0	100.0	
Valid cases	200	Missing cases	0		

PRIMER BANCO DE PREFERENCIA

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Bancomer	1.00	158	79.0	79.0	79.0
Banamex	2.00	23	11.5	11.5	90.5
Serfin	3.00	8	4.0	4.0	94.5
Otros	4.00	11	5.5	5.5	100.0
		-----	-----	-----	
	Total	200	100.0	100.0	
Valid cases	200	Missing cases	0		

ANEXO II

- - - - - D I S C R I M I N A N T A N A L Y S I S - - - - -

PRIMERA MUESTRA

Classification Results -

Actual Group	No. of Cases	Predicted Group Membership			
		1	2	3	4
Group MercProb	97	30 30.9%	19 19.6%	24 24.7%	24 24.7%
Group Bancomer	46	7 15.2%	16 34.8%	13 28.3%	10 21.7%
Group Banamex	37	5 13.5%	8 21.6%	20 54.1%	4 10.8%
Group Otros	15	4 26.7%	2 13.3%	1 6.7%	8 53.3%
Ungrouped Cases	6	0 .0%	1 16.7%	3 50.0%	2 33.3%

Percent of "grouped" cases correctly classified: 37.95%

SEGUNDA MUESTRA

Classification Results -

Actual Group	No. of Cases	Predicted Group Membership			
		1	2	3	4
Group Bancomer	158	78 49.4%	31 19.6%	30 19.0%	19 12.0%
Group Banamex	23	3 13.0%	15 65.2%	2 8.7%	3 13.0%
Group Serfin	8	1 12.5%	1 12.5%	6 75.0%	0 .0%
Group Otros	11	2 18.2%	0 .0%	1 9.1%	8 72.7%

Percent of "grouped" cases correctly classified: 53.50%

PRIMERA MUESTRA

----- LOGISTIC REGRESSION -----

Classification Table for VRL1

		Predicted		Percent Correct
		.00	1.00	
Observed	0	64	34	65.31%
	1	41	56	57.73%
Overall				61.54%

Classification Table for VRL2

		Predicted		Percent Correct
		.00	1.00	
Observed	0	153	2	98.71%
	1	39	7	15.22%
Overall				79.60%

Classification Table for VRL3

		Predicted		Percent Correct
		.00	1.00	
Observed	0	164	0	100.00%
	1	37	0	.00%
Overall				81.59%

Classification Table for VRL4

		Predicted		Percent Correct
		.00	1.00	
Observed	0	186	0	100.00%
	1	15	0	.00%
Overall				92.54%

SEGUNDA MUESTRA

- - - - - LOGISTIC REGRESSION - - - - -

Classification Table for VRL1

		Predicted		
		.00	1.00	
Observed	.00	0	1	19.05%
	0	8	34	
Observed	1.00	5	153	96.84%
	1			
			Overall	80.50%

Classification Table for VRL2

		Predicted		
		.00	1.00	
Observed	.00	0	1	98.87%
	0	175	2	
Observed	1.00	21	2	8.70%
	1			
			Overall	88.50%

Classification Table for VRL3

		Predicted		
		.00	1.00	
Observed	.00	0	1	100.00%
	0	192	0	
Observed	1.00	7	1	12.50%
	1			
			Overall	96.50%

Classification Table for VRL4

		Predicted		
		.00	1.00	
Observed	.00	0	1	99.47%
	0	188	1	
Observed	1.00	8	3	27.27%
	1			
			Overall	95.50%

ANEXO III

CUADRO 1
VALORES DE LOS CENTROIDES POR GRUPO

Grupo	Función Discriminante	
	1	2
Mercantil Probusa	.251	-.197
Bancomer	-.414	-.374
Banamex	-.457	.228
Otros Bancos	.778	.692

CUADRO 2
COEFICIENTES DE LAS VARIABLES UTILIZADAS
PARA DEFINIR EL PERFIL POR FUNCIÓN DISCRIMINANTE

Variables	Función Discriminante	
	1	2
Nivel social		
Muy bajo	.357	.092
Bajo	-.162	-.047
Medio	-.417	.214
Edad		
18-24 años	.110	-.342
25-34 años	-.119	-.090
35-44 años	-.391	-.073
45-54 años	-.090	.551
Sexo masculino	-.574	-.032
Estado civil		
Soltero	.582	.462
Casado	.534	.355
Trabaja	.375	.353
Ocupación		
Comerciante	.488	-.327
Ejecutivo	.436	.182
Empleado privado	.343	.520
Ama de casa	.035	.502

CUADRO A

Medias de los grupos

BANCO	Niv.soc.muy bajo	Niv.soc.bajo	Niv.soc.medio	Niv.soc.alto
Merc. Prob.	.25773	.19588	.30928	.23711
Bancomer	.15217	.21739	.43478	.19565
Banamex	.02703	.24324	.40541	.32432
Otros Banc.	.20000	.26667	.13333	.40000
Total	.18462	.21538	.34359	.25641

BANCO	18-24 años	25-34 años	35-44 años	45-54 años
Merc. Prob.	.05155	.21649	.23711	.27835
Bancomer	.06522	.19565	.32609	.13043
Banamex	.00000	.16216	.37838	.32432
Otros Banc.	.13333	.33333	.20000	.13333
Total	.05128	.21026	.28205	.24103

BANCO	55 o más	Hombres	Soltero	Casado
Merc. Prob.	.21649	.47423	.24742	.72165
Bancomer	.28261	.58696	.15217	.73913
Banamex	.13514	.51351	.18919	.70270
Otros Banc.	.20000	.40000	.33333	.60000
Total	.21538	.50256	.22051	.71282

BANCO	Otros edos.civ.	Si trabaja	Comerciante	Ejecutivo
Merc. Prob.	.03093	.80412	.18557	.14433
Bancomer	.10870	.71739	.26087	.08696
Banamex	.10811	.81081	.16216	.08108
Otros Banc.	.06667	.93333	.46667	.13333
Total	.06667	.79487	.22051	.11795

BANCO	Empleado privado	Ama de casa	Otras ocup.
Merc. Prob.	.34021	.21649	.11340
Bancomer	.19565	.26087	.19565
Banamex	.35135	.21622	.18919
Otros Banc.	.20000	.06667	.13333
Total	.29744	.21538	.14872

**CUADRO 3
VALORES DE LOS CENTROIDES POR GRUPO**

Grupo	Función Discriminante	
	1	2
Bancomer	-.113	.185
Banamex	-.089	-1.088
Serfin	-.155	-.558
Otros Bancos	1.921	.018

**CUADRO 4
COEFICIENTES DE LAS VARIABLES UTILIZADAS
PARA DEFINIR EL PERFIL POR FUNCIÓN DISCRIMINANTE**

Variables	Función Discriminante	
	1	2
Sexo masculino	.053	-.264
Edad		
18-25 años	-.052	-.279
26-35 años	.161	-.507
36-45 años	-.156	-.520
Nivel social		
Bajo	.344	.191
Medio	.004	-.246
Ocupación		
Profesionista libre	.202	.006
Profesionista empleado	-.133	.234
Comerciante	-.260	.409
Empleado	.246	.352
Escolaridad del jefe de familia		
hasta preparatoria	-.178	.390
incompleta	.141	.473
Preparatoria completa	.089	.439
Licenciatura incompleta	-.450	.648
Licenciatura completa		
Ingresos mensuales		
\$ 1,500-3,500	.156	.267
\$ 3,501-4,500	.117	.022
\$ 4,501-8,500	.193	-.233
\$ 8,501-10,500	.159	.160
\$ 10,501-15,000	.837	.068
\$ 15,001-20,000	.288	-.123
\$ 20,001-30,000	.001	.451

CUADRO B

Media de los grupos

BANCO	Hombres	18-25 años	26-35 años	36-45 años	Más de 45
Bancomer	.62025	.13291	.27215	.26582	.32911
Banamex	.73913	.13043	.30435	.43478	.13043
Serfin	.87500	.12500	.37500	.37500	.12500
Otros Bancos	.72727	.09091	.45455	.09091	.36364
Total	.65000	.13000	.29000	.28000	.30000

BANCO	Nivel social bajo	Nivel social medio	Nivel social alto	Profesionista libre	Profesionista empleado
Bancomer	.06329	.77848	.15823	.24051	.18354
Banamex	.00000	.91304	.08696	.39130	.21739
Serfin	.00000	.87500	.12500	.25000	.25000
Otros Bancos	.18182	.72727	.09091	.45455	.09091
Total	.06000	.79500	.14500	.27000	.18500

BANCO	Comerciante	Empleado	Otras ocupaciones	Hasta prep. incompleta	Prep. completa
Bancomer	.17722	.12658	.27215	.16456	.08861
Banamex	.04348	.04348	.30435	.08696	.00000
Serfin	.25000	.00000	.25000	.12500	.25000
Otros Bancos	.00000	.27273	.18182	.09091	.27273
Total	.15500	.12000	.27000	.15000	.09500

BANCO	Licenciatura incompleta	Licenciatura completa	Maestría y doctorado	\$ 1,500-3,500	\$ 3,501-4,500
Bancomer	.08861	.55696	.10127	.12025	.10759
Banamex	.08696	.56522	.26087	.08696	.08696
Serfin	.12500	.37500	.12500	.00000	.00000
Otros Bancos	.18182	.27273	.18182	.09091	.09091
Total	.09500	.53500	.12500	.11000	.10000

BANCO	\$ 4,501-8,500	\$ 8,501-10,500	\$ 10,501-15,000	\$ 15,001-20,000	\$ 20,001-30,000	Más de \$30,001
Bancomer	.18354	.19620	.08861	.12025	.12658	.05696
Banamex	.39130	.08696	.08696	.17391	.00000	.08696
Serfin	.25000	.50000	.00000	.25000	.00000	.00000
Otros Bancos	.18182	.09091	.45455	.09091	.00000	.00000
Total	.21000	.19000	.10500	.13000	.10000	.05500

CUADRO 5

Variables	Banco			
	Mercantil Probursa	Bancomer	Banamex	Otros Bancos
Nivel social				
Muy bajo	1.011	-.022	-2.293	-.207
Bajo	-.129	.314	-.134	-.541
Medio	.083	.602	-.075	-2.021
Edad				
18-24 años	-.288	.699	-6.212	1.190
25-34 años	-.291	.281	.253	.167
35-44 años	-.705	.288	.861	-.000
45-54 años	.238	-.861	.756	-.914
Sexo masculino	-.432	.876	.293	-1.090
Estado civil				
Soltero	1.376	-1.163	-.598	-.106
Casado	1.172	-.691	-.697	-.139
Trabaja	.656	-.986	.011	.289
Ocupación				
Comerciante	.175	-.327	-.676	1.560
Ejecutivo	.862	-1.082	-.639	.739
Empleado privado	.818	-.997	-.105	-.384
Ama de casa	.689	-.309	.032	-1.304

CUADRO 6

Variables	Banco			
	Bancomer	Banamex	Serfin	Otros Bancos
Sexo masculino	-.774	.753	1.415	.851
Edad				
18-25 años	-.741	1.333	.804	-.019
26-35 años	-1.330	1.524	1.555	-.267
36-45 años	-1.022	1.512	2.156	-2.433
Nivel social				
Bajo	.214	-7.670	-9.948	4.230
Medio	-.357	.793	-.583	-.014
Ocupación				
Profesionista libre	-.021	-.087	-1.868	1.131
Profesionista empleado	.888	-.711	-1.619	-1.586
Comerciante	1.696	-1.829	-1.585	-8.099
Empleado	.916	-1.739	-11.369	1.801
Escolaridad del jefe de familia				
Hasta preparatoria	1.091	-1.154	-.262	-1.035
Incompleta	.788	-9.230	2.132	.063
Preparatoria completa	.902	-1.491	1.218	.928
Licenciatura incompleta	1.547	-1.179	-.482	-1.815
Licenciatura completa				
Ingresos				
\$ 1,500-3,500	.417	-1.029	.827	7.640
\$ 3,501-4,500	-.227	.027	1.063	6.891
\$ 4,501-8,500	-.840	.301	9.591	7.295
\$ 8,501-10,500	-.280	-1.417	11.310	7.523
\$ 10,501-15,000	-1.240	-.822	.981	10.032
\$ 15,001-20,000	-1.018	-.101	10.241	8.812
\$ 20,001-30,000	7.713	-8.770	.598	-1.490

BIBLIOGRAFÍA

- Anderson S., 1980, *Statistical Methods for Comparative Studies*, Ed. John Wiley and Sons. New York, Chichester, Brisbane, Toronto and Singapore.
- Carl McDaniel Jr., 1986, *Curso de Mercadotecnia*, 2ª edición, por Harla, S.A. de C.V.
- Gordon A. D., 1981, *Clasificación*, Chapman and Hall, London, New York.
- Hair J., 1987, *Multivariate Data Analysis*, Ed. MacMillan Publishing Company, New York and London.
- Hosmer D. and Lemeshow S., 1989, *Applied Logistic Regression*, Ed. John and Sons. New York, Chichester, Brisbane, Toronto and Singapore.
- Mardia K. V., Kent J. T. y Bibby J. M., 1979, *Multivariate Analysis*, Academic Press, London, New York.
- Montgomery D. and Peck E., 1982, *Introduction to Linear Regression Analysis*, Ed. John Wiley and Sons. New York, Chichester, Brisbane, Toronto and Singapore.
- Mood A., Graybill F. and Boes D., 1974, *Introduction to the Theory of Statistics*, 3ª edición, McGraw Hill Book Company México, Printed in Singapore.
- Morrison D., 1976, *Multivariate Statistical Methods*, Ed. McGraw-Hill, Book Company, New York, San Francisco, Kuala Lumpur, London, México, Montreal, New Delhi, Panamá, París, Tokyo y Toronto.
- Philip Kotler, Dirección de la Mercadotecnia, *Análisis, Planificación y Control*, 7ª edición, Ed. Prentice-Hall Hispanoamérica, S.A.
- Tatsuoka M., 1970, *Discriminant Analysis*, Ed. Institute for Personality and Ability Testing. U.S.A.

Thomas C. Kinnear y James R. Taylor, *Investigación de Mercados. Un Enfoque Aplicado*,
4ª edición, Ed. McGraw-Hill.

William J. Stanton, Michael J. Etzel y Bruce J. Walker, *Fundamentos de Marketing*,
9ª edición, Ed. McGraw-Hill.