UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

APLICACIÓN DEL ANÁLISIS DISCRIMINANTE A LA SEGMENTACIÓN DE MERCADOS EN PRUEBAS DE PRODUCTO

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIA

P R E S E N T A :

NANCY FLORES SORCIA

DIRECTOR DE TESIS:
DRA. MARÍA DEL PILAR ALONSO REYES

(2012)

1. Datos del alumno.
    Flores
    Sorcia
    Nancy
    0445529630818
    Universidad Nacional Autónoma de México
    Facultad de Ciencias
    Actuaría
    301223560

2. Datos del tutor.
    Dra.
    María del Pilar
    Alonso
    Reyes

3. Datos del sinodal 1.
    M en C.
    José Antonio
    Flores
    Díaz

4. Datos del sinodal 2.
    Dra.
    Ruth Selene
    Fuentes
    García

5. Datos del sinodal 3.
    Act.
    Francisco
    Sánchez
    Villarreal

6. Datos del sinodal 4.
    Act.
    Edna Gabriela
    López
    Estrada

7. Datos del trabajo escrito.
    APLICACIÓN DEL ANÁLISIS DISCRIMINANTE A LA SEGMENTACIÓN DE MERCADOS EN PRUEBAS DE PRODUCTO.
    87 p
    2012

ÍNDICE

Introducción

En la actualidad, la forma en la que se dan a conocer los nuevos productos o servicios, así como su comercialización dentro de un mercado ya sea competitivo o no, es de gran importancia para que éstos permanezcan en el gusto de los consumidores.

Las herramientas utilizadas para ello, son resultado de la interacción existente entre la mercadotecnia e investigación de mercados. La primera se encarga de tomar decisiones para crear políticas comerciales, mientras que la segunda provee la información necesaria para conformar estas políticas, ello, con el fin de obtener mejores resultados.

Debido a que la mayoría de los consumidores tienen diferentes deseos y necesidades, existen métodos de segmentación de mercados, los cuales derivan en un conocimiento más detallado de los individuos. Esto a su vez, ayuda a los encargados de la toma de decisiones, a establecer un mercado objetivo al cual se dirijan todos los esfuerzos, lo que representa un área de oportunidad en el caso de que no sea un mercado saturado de competidores.

En este trabajo se realizará una segmentación de mercados mediante la aplicación de uno de los métodos multivariados conocido como análisis discriminante. En el primer capítulo se dará una breve introducción a lo que se refiere mercadotecnia e investigación de mercados; en el segundo, se presentará un desarrollo teórico de la técnica utilizada; en el tercer capítulo, se tendrá la aplicación de esta técnica multivariada, con datos que pertenecen a estudios de mercado, tipo prueba de producto, tanto para bebidas de café como cosméticos; por último, se tiene la conclusión de los resultados obtenidos, así como, la bibliografía utilizada.

1. Mercadotecnia e investigación de mercados

Con frecuencia suelen considerarse por *investigación de mercados* y *mercadotecnia (marketing)* como conceptos similares, sin embargo, difieren al conceptualizarlos y aplicarlos, a pesar de que van de la mano uno con el otro, ya que la mayor parte de las decisiones de marketing involucran recursos limitados de la investigación de mercados.

A continuación se explican las diferencias entre ambas, así como su interacción involucrando a su vez otros conceptos.

1.1. Marketing o mercadotecnia

El *marketing* o *mercadotecnia* cuenta con una rica variedad de conceptos y herramientas que ayudan a la toma de decisiones, pero ¿cómo se define a esta disciplina?

> *La* ***mercadotecnia*** *es aquélla actividad humana cuyo objetivo principal está dirigido a satisfacer necesidades, carencias y deseos a través de procesos de intercambio; asimismo, la mercadotecnia es el proceso de planeación, ejecución y conceptualización de precios, premiación y distribución de ideas, mercancías y técnicas para crear intercambios que satisfagan objetivos individuales y organizacionales*[1]*.*

Se define como **consumidor** aquél ente que compra para consumir, el cual puede ser de tipo:

- **Independiente**. Las preferencias de otros consumidores no le afectan para tomar una decisión.
- **Ilimitado**. Su nivel de satisfacción se incrementa conforme compra.

[1] Definición según the American Marketing Association (AMA).

- **Consciente**. El consumidor conoce siempre y bajo cualquier circunstancia su nivel de preferencia hacia los productos o le son indiferentes.

Los factores que pueden influir en su conducta hacia un producto son de tipo psicológico, socioeconómico o por aprendizaje, retención o memorización. El primero involucra la personalidad del individuo y la percepción que tiene de sí mismo o auto concepto. El segundo trata de las características de los compradores como la edad, poder adquisitivo y estilos de vida. El tercero se refiere al proceso por el cual el individuo adquiere experiencia de compra y consumo que aplicará en su comportamiento futuro.

La percepción es el proceso mediante el cual el individuo selecciona, organiza e interpreta estímulos; con ello, toma decisiones de compra basadas en lo que percibe como necesario y bloquea estímulos desfavorables o ingratos.

Es así como la forma en que los productos se perciben es lo más importante para su éxito, más que las características reales que posea. Los productos que son percibidos favorablemente, como es obvio, tienen mejores posibilidades de ser comprados. De tal modo que, la importancia de la mercadotecnia reside en la contribución en forma directa a la venta de los productos de una empresa, además de crear oportunidades para realizar innovaciones en ellos.

Para producir una respuesta positiva de los consumidores se lleva a cabo una estrategia que engloba a cuatro elementos, conocidos como la mezcla de la mercadotecnia o las **4 P de la mercadotecnia**:

1. **Producto**. Aquello que es propenso a ser comprado, cambiado, traspasado, etcétera.

2. **Precio**. Indica la cantidad que un cliente paga por el producto, el cual es fijado aplicando estrategias que no siempre tienen que ver con el costo de producirlo, sino con el precio de la competencia.

3. **Plaza (Distribución)**. Es el lugar donde estará disponible el producto para su compra, ello, designa la forma en la que se distribuye, cómo se compra y cómo se vende.

4. **Promoción (Comunicación)**. Son las acciones que tienen por objeto cambiar el comportamiento del consumidor hacia el producto.

El medio por el cual interactúan los consumidores y los elementos antes señalados es el **mercado**, el cual dependiendo desde qué perspectiva se defina, puede referirse a un área geográfica, a un grupo de personas que realizan transacciones o a un ámbito dentro del cual las relaciones de oferta y demanda concurren para la fijación de un precio.

En el mismo se busca cubrir las necesidades, deseos y posibilidades de satisfacción de los consumidores; no obstante, lo antes mencionado es tan variado de persona a persona, que por ello la empresa debe analizar su mercado con el propósito de adaptar su oferta y estrategias de mercadotecnia a las necesidades de este. A este conjunto de personas se le conoce como **mercado meta**.

Por otro lado, de acuerdo a lo que el mercado ofrezca, éstos pueden ser:

- **De mercancías**. Se ofrecen bienes producidos específicamente para venderlos; ejemplo de ello, es el mercado de calzado, ropa, alimentos y bebidas, etcétera.

- **De servicios**. Se ofrecen bienes no tangibles, por ejemplo, el mercado de trabajo, turismo o servicios de transporte.

1.2. Investigación de mercados

Hoy en día la competencia entre diversas empresas por mantener o cubrir una parte de la demanda de productos del mercado ha llevado a la necesidad de conocer aún más el comportamiento y gustos de los consumidores, ello, se traduce en la recopilación y análisis de información de los mismos, lo cual es posible gracias a la investigación de mercados, con base a esto se puede ofrecer una definición de la misma a continuación:

> *La **investigación de mercados** es el enfoque sistemático y objetivo para el desarrollo del suministro de información para el proceso de toma de decisiones por la gerencia de marketing*[2].

Es así como la investigación de mercados es la función que enlaza al consumidor y los responsables de la toma de decisiones para llevar a cabo las políticas comerciales de una empresa o institución. En otras palabras, esta función especifica la información requerida para identificar las oportunidades, evaluar o perfeccionar acciones de marketing, monitorear el desempeño de un producto o servicio.

La forma de hacerlo es mediante el diseño para recolectar la información, en el que se dirige e implementa el proceso de recolección de datos, analiza los resultados, comunica los hallazgos y sus implicaciones. De este modo, se suministra de información oportuna a la gerencia de marketing para la toma de decisiones.

La esencia de la política comercial de las empresas, según Philip Kotler, se basa en:

- **La segmentación de mercados**. Se divide un mercado en distintos grupos de compradores que podrían requerir productos con diferentes características.

- **Fijación de metas**. Es el acto de desarrollar mediciones del atractivo comercial de cada segmento y seleccionar uno o más segmentos a los cuales entrar.

- **El posicionamiento del producto**. Se logra al identificar los conceptos posibles o variables que caracterizan a cada segmento. Consiste en planear la oferta y la imagen de la empresa de una manera tal que el mercado meta comprenda y valore su ubicación en relación con la competencia.

[2] Thomas C. Kinnear. Investigación de mercados, un enfoque aplicado.

1.2.1. Segmentación de mercados

Para que la fijación de metas y posicionamiento del producto sea exitoso, se requiere antes de la **segmentación de mercados**, el cual como ya se ha considerado busca dividir el mercado en grupos más pequeños o segmentos homogéneos que pueden ser elegidos como mercado meta de las empresas. Es así que se mejora la precisión del marketing de una empresa, ya que agrupa a individuos con necesidades semejantes.

El resultado de una segmentación adecuada será aquélla en la que se cumpla lo siguiente:

- Dentro de cada uno de los segmentos los individuos son similares entre sí, de esta manera sus respuestas a las variables que conforman la mezcla de mercadotecnia son similares.

- Son diferenciables entre los segmentos obtenidos, es decir, heterogéneos.

- Para poder garantizar la rentabilidad de cada segmento, deben ser bastante grandes.

- Operacionales, para identificar a los clientes y poder tomar decisiones referentes a la plaza y promoción.

Existen grandes beneficios derivados de la obtención de una segmentación de mercados, a continuación, se apuntan algunos:

- Permite identificar las necesidades de los clientes dentro de un segmento de mercado y el diseño más eficaz de la mezcla de mercadotecnia para satisfacerlas.

- Las empresas de tamaño mediano pueden crecer más rápido si obtienen una posición sólida en los segmentos especializados del mercado.

- La empresa enfrenta a menos competidores en un segmento en específico, con esto, se generan nuevas oportunidades de crecimiento y obtiene una ventaja competitiva.

La forma de obtener una segmentación de mercados es mediante la realización en primer lugar de un **estudio de mercado**, donde se examinan las necesidades específicas satisfechas por las ofertas actuales, las que no lo son y las que podrían ser reconocidas.

Se llevan a cabo entrevistas de exploración y organiza sesiones de grupos para entender mejor las motivaciones, actitudes y conductas de los consumidores. Se obtiene información sobre los atributos y la importancia que se les da, conciencia de la marca y calificaciones a la misma, patrones de uso y actitudes hacia la categoría de los productos.

De forma paralela, se obtiene información demográfica, psicológica y de tipo socioeconómico de los individuos. Características del tipo de cliente a tratar como: clientes consumidores o de un mercado en general; por ejemplo, dentro del universo de usuarias de cremas faciales existen las consumidoras de una crema en especial que produce la empresa o alguna de la competencia.

Asimismo, puede ser información referida al comportamiento observado del consumidor hacia el producto o servicio; tales como, la lealtad a la marca o beneficios adquiridos por el uso del mismo.

Las fuentes de datos pueden ser de tipo **primario** o **secundario**, la forma en la que se toman estos datos y su disponibilidad son las que hacen posible esta clasificación. Ejemplo del tipo secundario, son las estadísticas gubernamentales, reportes de estudios de mercado anteriores, estadísticas de la industria de su interés, bases de datos de compras u opinión de expertos en la materia.

Sin embargo, con el primer tipo en la mayoría de las ocasiones no es posible obtener suficiente información, así que se recurre a las fuentes de datos primarios que son aquéllas que el investigador recolecta por sí mismo; en un estudio de mercado, suele auxiliarse de encuestadores, vendedores, intermediarios o clientes de la empresa para obtener la información de interés. La ventaja de este es que se consigue la información de la fuente original.

El método de obtención de los datos, es mediante:

- **Encuesta**. Consiste en recolectar datos de entrevistas realizadas a un número limitados de personas, es decir, de una muestra seleccionada de un grupo más grande.

- **Observación**. Los datos son tomados observando alguna acción llevada a cabo por un comprador. Los consumidores no se dan cuenta de que están siendo observados, por lo que actuarán de manera natural, ya que el encargado de recopilar la información actuará como un comprador más en el establecimiento donde se lleva a cabo dicho ejercicio. Comúnmente, en las empresas de investigación de mercado esto se conoce como estudios de *mistery shopper.*

- **Experimental**. Para conseguir los datos primarios se requiere establecer un experimento controlado que simule la situación real de mercado tanto como sea posible. La aplicación más importante del método experimental son las **pruebas de producto**, en donde se fabrican pocas unidades de un producto y se dan a probar a algunos consumidores.

En segundo término se lleva a cabo la **recopilación de la información y análisis**, en este paso se interpretan los datos de una forma superficial, por lo que se recurre al cálculo de estadísticas descriptivas comunes como lo son la media, la moda y la varianza de algunas respuestas recopiladas en el estudio.

Posteriormente, se aplican métodos estadísticos más sofisticados, que en capítulos posteriores se estarán tratando, para eliminar las variables y agrupar o construir el segmento con los consumidores que comparten un aspecto en particular, que los distingue de los demás subconjuntos del mercado con necesidades diferentes.

El último paso es la obtención de **perfiles de consumidores**, aquí se prepara un perfil de cada segmento obtenido en términos de actitudes distintivas, conductas y otras características que se hayan coleccionado en el estudio.

Este procedimiento debe repetirse periódicamente, porque los segmentos cambian. En esta etapa, también se investiga la **jerarquía de atributos** que los consumidores consideran al escoger o mostrar una declinación de preferencia hacia una marca en especial.

2. Técnicas de análisis multivariado

El análisis multivariado es el conjunto de métodos estadísticos cuya finalidad es analizar datos multivariantes, es decir, hay más de una variable medida para cada individuo. Su razón radica en un mejor entendimiento del objeto de estudio, obteniendo información que los métodos estadísticos univariantes son incapaces de conseguir.

El análisis multivariado se compone de:

- Análisis de la dependencia. Técnicas aplicables cuando una o varias variables dependientes van a ser explicadas por un conjunto de variables independientes, que actúan como predictivas.

- Análisis de la interdependencia. Técnica que otorga la misma consideración a todas las variables objeto de estudio, sin distinguir entre dependientes e independientes, y que tienen como fin descubrir las interrelaciones y en definitiva la estructura subyacente en ellas.

- Otras técnicas. Se refiere a aquéllas más novedosas que surgen con el fin de superar el enfoque monocriterio de éstas o para permitir un tratamiento más eficaz y eficiente de las enormes cantidades de datos que se manejan en la actualidad.

De esta clasificación se desprenden varias técnicas, a continuación se presenta una breve explicación de las más conocidas en cada una.

Técnicas de análisis de la dependencia

- Análisis discriminante (AD). Técnica de clasificación que permite agrupar a los elementos de una muestra en dos o más categorías diferentes, predefinidas en una variable dependiente no métrica, en función de una serie de variables independientes métricas combinadas linealmente.

- Análisis multivariante de la varianza (MANOVA). Extensión de la ANOVA (método para contrastar si diversas muestras proceden de poblaciones con igual media), ésta se aplica a una combinación de variables dependientes relacionadas entre sí.

- Análisis multivariante de la covarianza. Proceso también conocido como MANCOVA, que comienza por emplear la regresión para eliminar la variación experimentada por las variables dependientes, producida por una variable independiente no controlada (covariable).

- Regresión lineal. Técnica que pretende determinar la combinación lineal de variables independientes cuyo comportamiento es el mejor predictor de los cambios experimentados por la variable dependiente; todas las variables que intervienen en la regresión son métricas, aunque admite la posibilidad de trabajar con no métricas, si se emplean variables ficticias para su transformación.

- Modelo logit. Método de elección discreta en el que la función de densidad de probabilidad de la variable perturbación es la función logística. En el caso de que la variable dependiente sea politómica[3] en lugar de dicotómica, se habla de un *Modelo logit multinomial.*

- Modelo probit. Método de elección discreta en el que la función de densidad de probabilidad de la variable perturbación es de tipo normal estándar.

- Modelos tobit. Modelo de regresión con datos censurados variante del probit, caracterizado por la existencia de un gran número de observaciones nulas o cuyo valor es una cota.

- Análisis conjunto. Se emplea para entender cómo conforman los individuos sus preferencias hacia los objetos, normalmente marcas o productos.

[3] Este tipo de variable toma tres o más valores posibles; por ejemplo, si se habla de la lealtad de compra hacia diferentes marcas de autos, en tal caso se podría tener la marca A, B o C. Por otro lado, si es de tipo dicotómica, los valores posibles son dos; ejemplo de ello, es vivo o muerto, le gusta un producto o no, etcétera.

- Segmentación jerárquica. Tiene por objeto distinguir conjuntos de elementos homogéneos en una población, a través de un proceso iterativo descendente de partición de la muestra total en sucesivos grupos, en virtud del valor adoptado por la variable dependiente.

- Análisis con ecuaciones estructurales. Mejor conocido como *análisis de estructuras de covarianzas*, permite analizar varias relaciones de dependencia que se presentan simultáneamente.

Técnicas de análisis de Interdependencia

- Análisis factorial. Técnica de análisis de la interdependencia presentada por un cierto número de variables susceptible de ser sintetizada en un conjunto de factores comunes que subyacen tras ella; estos pueden ser comunes (captan la variabilidad compartida por todas las variables), o específicos (captan la variabilidad propia de cada variable, sin relación con la demás).

- Análisis de componentes principales. Técnica de análisis de la interdependencia presentada por un cierto número de variables susceptible de ser sintetizada en un conjunto de factores comunes que subyacen tras ella. Dichos componentes buscan explicar la mayor proporción posible de la variabilidad total, lo que quiere decir que, a diferencia de lo que ocurre en el análisis factorial, no existen componentes específicos.

- Análisis de correspondencias simples. Técnica basada en el estudio de la asociación entre las categorías de variables no métricas, que persigue la elaboración de un mapa que ponga de manifiesto dicha asociación en modo gráfico. Cabe señalar que, en el ámbito de investigación de mercados, esta técnica es utilizada para reconocer los atributos más relacionados con diversas marcas o individuos estudiados.

- Análisis de conglomerados. También conocido como análisis cluster, es la técnica cuyo fin es clasificar sujetos u objetos en función de ciertas características de modo que los elementos de cada grupo sean muy similares entre sí.

- Escalamiento multidimensional. Técnica cuyo fin es elaborar una representación gráfica que permita describir el comportamiento de un grupo de individuos.

- Análisis con clases latentes. Busca distinguir en una muestra grupos de elementos homogéneos en función de los valores que adopta una variable latente no métrica. Tales valores son las categorías de esa variable, las cuales reciben el nombre de clases latentes.

Otras técnicas

- Elección multicriterio discreta. Conjunto de métodos de ayuda en la resolución de problemas de decisión en los que se han de tener en cuenta diferentes puntos de vista o criterios y en los que se mezcla un número finito de alternativas.

- Análisis con redes neuronales. Técnica cuya forma de proceder pretende replicar el funcionamiento del cerebro humano, intentando aprender de los errores cometidos para obtener el mejor resultado posible.

- Data mining. Llamada esta técnica *minería de datos*, consiste en la exploración y análisis de un gran volumen de información con el fin de descubrir relaciones, reglas o patrones de comportamiento en ellos, que sean de utilidad para el usuario en la toma de decisiones.

2.1. El análisis discriminante

Las primeras nociones del análisis discriminante (AD) datan de 1920, cuando el estadístico inglés Karl Pearson propuso la existencia del *coeficiente de similitud racial*[4], un tipo de distancia entre grupos. Por esa época, surgieron alrededor del mundo otras propuestas, uno de los ejemplos más claros fue en la India, donde diez años más tarde se obtendría una formalización de la misma en manos de P. C. Mahalanobis. Fue así, que finalmente la idea de una distancia intergrupal fue interpretada por R. A. Fisher en las años treinta como una combinación lineal de variables para discriminar entre grupos. Esta mezcla de términos, tales como la distancia y combinación lineal de variables, fue presentada por Fisher en su trabajo titulado *The use of multiple measurements in taxonomic problems*[5].

A partir de ese momento, la aplicación de este método inició en problemas esbozados en ciencias médicas y biológicas, en la actualidad el interés de su uso se ha incrementado en estudios referentes al marketing, educación, negocios, ingeniería y psicología.

Hablando en un sentido más estricto, la principal propiedad del AD se refiere a la clasificación de uno o más individuos dentro de algún grupo con elementos similares a él, visto éste como un subconjunto de la población. Ello, a partir de la construcción de una función que tenga como variables independientes, características mesurables y como variable dependiente, una de tipo nominal. El resultado es una función, combinación lineal de variables independientes, esta relación se puede visualizar de la siguiente forma:

$$y = f(x_1, x_2, x_3, ..., x_n) \qquad \text{.................... } \mathbf{(1)}$$

Donde *y* es la variable dependiente, mientras que $x_1, x_2, x_3, ..., x_n$ son las diferentes variables de la población, que tienen alguna métrica asociada para su medición.

[4] También conocido como, CRL se refiere a las siglas en inglés de *coefficient of racial likeness.*
[5] Documento publicado en 1936 en *Annals of Eugenics.*

Es por ello, que se busca conformar grupos que muestren simultáneamente una varianza intragrupos *mínima* y otra intergrupos *máxima*, así se obtendrá una función o regla discriminante capaz de asignar a los elementos de forma óptima. De este modo, es posible enumerar los objetivos del AD, los cuales se enlistan a continuación:

1. Determinar si los datos son aptos para la aplicación del AD

2. Determinar cuáles variables se tomarán en cuenta para explicar las diferencia entre los grupos;

3. Determinar las combinaciones lineales de variables predictivas para separar a los grupos, maximizando la variación entre los mismos y minimizando la variación dentro de ellos;

4. Asignar nuevos objetos, empresas o individuos, cuyos perfiles son conocidos, pero no su identidad de grupo, ello con base a la función discriminante.

Algunas aplicaciones del AD

Como fue reportado el AD es uno de los métodos del análisis multivariado y tal vez el más utilizado, debido a la gran gama de aplicaciones que posee, por ejemplo:

- El éxito o fracaso de un producto que ya ha sido lanzado al mercado o está en proceso de aprobación. Se realizaría con base al análisis de comportamientos hacia el producto, por parte de los consumidores.

- Fracaso empresarial. Si un investigador está interesado en pronosticar quiebras de negocios, tal vez pueda agrupar a las empresas que quebraron o no con el paso del tiempo con base a las variables independientes como ubicación, razones financieras o cambios en la administración, construir una función que le permita conocer el éxito de una empresa.

- La esperanza de vida de una persona, ello se podría derivar de su nivel de vida, la cual estará deducida por factores como nivel socioeconómico, ingresos, edad, estado civil, etcétera. Ya que el nivel de vida que tenga la persona, podría influir de manera directa en su esperanza.

- En un ámbito académico, se puede estar interesado en estimar el rendimiento de un alumno de nuevo ingreso, en especial de aquéllos que podrían dejar la escuela. Los elementos que serían de ayuda son datos recopilados de otros grupos ya estudiados, obteniendo así calificaciones, interés por el estudio u horas dedicadas a esta actividad, etcétera.

- Si es de carácter contable podría consistir en identificar que individuos de los que declaran sus impuestos podrían estar subestimando o exagerando en sus utilidades o impuestos.

- Un ejecutivo de préstamos de una compañía hipotecaria debe decidir si aprueba un préstamo hipotecario a un solicitante, esta decisión se toma determinando si las características se apegan más a las de personas que en el pasado pagaron oportunamente sus préstamos, que las de aquéllas que no cumplieron con sus pagos. La información acerca de estos grupos podría incluir factores como edad, ingresos, estado civil, deudas pendientes y posesión de bienes duraderos.

En cada uno de estos casos existe una o más variables predictivas o explicativas, en algunos de ellos la variable de criterio es continua y en su mayoría medible haciendo uso de una escala ordinal, en cuyo caso un análisis de regresión sería llevado a cabo. En otros casos, las variables de criterio son medidas con escala nominal, dicotómica o politómica, de ser así, el AD es de utilidad.

A su vez, existen otros supuestos que son necesarios de satisfacer para que el AD sea aplicable, éstos se abordarán posteriormente, antes de ello se revisarán algunos temas que serán de importancia en el desarrollo de esta técnica.

2.2. Revisión de tópicos a utilizar en análisis discriminante

Con frecuencia se recurre al manejo de diversas herramientas algebraicas en las técnicas de análisis multivariado; por lo que, el análisis discriminante no es la excepción. Como se apreciará en los apartados siguientes, el AD obtiene información a partir del ordenamiento y exploración de los datos, éstos últimos son manipulados hasta llegar a una función como resultado, capaz de discriminar a nuevos elementos en el grupo correspondiente. Es por ello que, a continuación se hace un breve compendio de los insumos necesarios para llegar al producto deseado.

Escalas de medición

Cuando en estadística se habla de la medición de un fenómeno, se refiere generalmente a la asignación de valores a observaciones, de modo que los números sean susceptibles de análisis por medio de operaciones de acuerdo con ciertas reglas; en otras palabras, la relación entre los objetos se está observando y los números, es tan directa que es posible obtener nueva información de estos individuos.

La teoría de la medición está formada por un conjunto de conjeturas, cada una referida a un nivel diferente de medición. Las operaciones permitidas con un conjunto dado, dependen del nivel de medida que se logre, ya sea nominal, ordinal o de intervalo. De estas últimas, se presentará su descripción y pruebas aplicables, ya que son de uso común para el análisis discriminante.

- **Escala nominal o clasificatoria**.- La medición se da a un nivel elemental cuando los números u otros símbolos se emplean para la clasificación de objetos, personas o características. Cuando se usan con el fin de distinguir entre sí los grupos al que pertenecen varios sujetos, los números o símbolos constituyen una escala nominal o clasificatoria.

Por ejemplo, cuando un médico psiquiátrico diagnostica a una persona como esquizofrénica, paranoica, maniaco-depresiva o psiconeurótica, se vale de un símbolo para representar la clase a la que pertenece cada individuo, por tanto, emplea el escalamiento nominal.

- **Escala ordinal o escala de rango**.- Puede suceder que los objetos de una categoría de escala no sean precisamente diferentes a los de otra, sino que están relacionadas entre sí. Ejemplo de ello, son las que comparan altura, preferencia, dificultad, perturbación, madurez, etcétera. Tales relaciones pueden formularse con el signo de “menor que” o “mayor que”; de igual modo, se pueden designar como “es preferible a”, “es más alto que”, “es difícil que”, etcétera. Su significado específico depende de la naturaleza de la relación que define la escala.

  Un ejemplo claro de este tipo de escala, se visualiza en la definición de status socioeconómico; es decir, en prestigio o aceptación social, todos los miembros de la clase media superior, son mayores que todos los de la clase media inferior. A su vez, los de la clase media inferior son superiores a los de baja inferior; asimismo, la relación de igual que, se mantiene en los miembros de la misma clase.

- **Escala intervalar**.- Tiene todas las características de una escala ordinal y cuando, se conoce la distancia entre dos números cualesquiera se tiene una medida considerablemente más fuerte que la ordinal. En tal caso, la medición se ha ejecutado en el sentido de una escala intervalar, esto es, si la asignación de números a varias clases de objetos es tan precisa que se sabe la magnitud de los intervalos entre todos los objetos de la escala.

  Una escala intervalar está caracterizada por una unidad de medida común y constante que asigna un número real a todos los pares de objetos de un conjunto ordenado, el punto cero y la unidad de medida son arbitrarios. Un ejemplo claro de ello, se percibe en la medición de la temperatura, ya que el cero de temperatura no representa ausencia de esa característica, simplemente es un punto de partida.

- **Escala de razón**.- Tiene todas las características de una escala intervalar y en su origen tiene un punto cero real, es llamada escala de razón o de proporción. En ella, la razón de un punto a otro cualquiera de la escala es independiente de la unidad de medida.

Un ejemplo claro de este tipo, se da al medir la masa o peso de algún objeto, en este caso sí se tiene un verdadero punto cero, ello en ausencia de algún cuerpo. Asimismo, se presenta este tipo de escala cuando se habla de medir temperatura absoluta, altura, tiempo de reacción, distancia, etcétera.

2.3. Desarrollo analítico del análisis discriminante

Como ya se había mencionado, la idea esencial del AD, se basa en la obtención de una regla, que permita hacer la integración de un elemento a un grupo ya definido, esto se obtiene mediante la construcción de una función que relaciona las variables incluidas en el estudio y la variable dependiente de tipo nominal. De esta forma, se debe asegurar que los datos satisfacen ciertas condiciones, para que esta técnica sea aplicable, ya que de no ser así el análisis realizado no será válido.

2.3.1. Condiciones para la aplicación del análisis discriminante

La aplicación del AD demanda un conjunto de variables discriminantes y de resultado, las primeras harán referencia a mediciones de características de los individuos y las segundas al grupo al que pertenece cada individuo.

Asimismo, el conjunto de mediciones deberá corresponder a dos o más grupos mutuamente excluyentes. Cuando sea desconocida la adscripción de ciertos casos a los conjuntos formados, es conveniente excluirlos antes de comenzar el análisis, posteriormente, cuando se haya definido la función discriminante con los casos donde sí era conocida la pertenencia, se retomarán y podrán ser clasificados con la misma.

Las variables discriminantes deberán estar en escala ordinal, de intervalo o razón, con ello se garantiza la posibilidad de medir varianzas y demás operaciones necesarias en el desarrollo del mismo. Mientras que, las variables que definen los grupos deberán de ser de tipo nominal.

Respecto al tamaño de los grupos, no hay inconveniente en que sean de diferentes dimensiones.

Los supuestos para aplicar el AD son:

- **Número mínimo de grupos y observaciones**. Una de las condiciones esenciales para la aplicación de este método, es la necesidad de que haya al menos dos grupos, ya que de lo contrario no habría subconjuntos hacia los cuales se discriminarán los elementos. También, cada grupo deberá contar con al menos dos o más casos, esto para realizar la medición de algunas características de los mismos.

- **Número máximo de variables discriminantes.** Se podrá utilizar cualquier número de variables discriminantes, siempre y cuando sea inferior al número de casos menos dos.

- **Número máximo de funciones discriminantes**. El número máximo de funciones discriminantes que se pueden calcular podrá ser igual al número de variables discriminantes, siempre y cuando su número no sea mayor que el número de grupos menos uno.

$$Número\ de\ \ funciones = mín\ \{n, k-1\}$$

Con *n* el número de variables discriminantes y *k* el número de grupos. Así, cuando el número de variables es menor que el número de grupos, el máximo de funciones discriminantes coincide con *n,* en tal situación la función discriminante no opera trasladando las coordenadas de los puntos de un espacio *n-dimensional* a un espacio menor, sino que sólo se realiza un cambio de ejes dentro de esa misma dimensión.

Cabe recordar que, si se aplican las funciones discriminantes a los valores observados en las variables de cada individuo, se tendrá como resultado lo que se llama *puntuaciones discriminantes*[6] para ese individuo.

- **Normalidad multivariante.** Se refiere a que cada grupo de mediciones representa una muestra aleatoria extraída de una población con distribución normal

[6] En ocasiones se refiere a estos como *scores.*

multivariable sobre las variables discriminantes. Es necesario su cumplimiento para la obtención de probabilidades de pertenencia a un grupo, así como pruebas de significancia.

En caso de que se trate de una población con distribución normal multivariable y con un tamaño de muestra suficientemente grande, entonces las distancias se distribuyen como Ji-cuadrada con *n* grados de libertad, donde *n* es el número de variables discriminantes.

- **Homogeneidad de matrices de varianzas y covarianzas**. Este supuesto encierra el cumplimiento de igualdad entre matrices de varianzas y covarianza, tanto de las poblaciones como de los subconjuntos o muestras obtenidas de las correspondientes.

- **Linealidad.** Este punto engloba la existencia de *relaciones lineales* dentro de cada grupo. Este supuesto implica que el modelo a contrastar predice los valores de las variables dependientes, siempre que se produzca una modificación en las independientes. En el AD, la medición de correlaciones será de utilidad en ocasiones, por tanto, es condición necesaria la existencia de asociaciones lineales entre las mismas.

- **Ausencia de multicolinealidad y singularidad.** Se habla de multicolinealidad cuando dos o más variables de la matriz de correlaciones presentan una correlación perfecta o casi perfecta con respecto a otras, lo cual implica que se comportan del mismo modo y la información aportada por alguna de ellas es reiterativa. Por otro lado, la singularidad se representa mediante la correlación perfecta, debido a que puede darse el caso en el que una variable sea aproximadamente la combinación de otras consideradas en el análisis.

Cuando se da alguna de las situaciones antes reportadas, el rango de la matriz de varianzas y covarianzas desciende, con ello el determinante de la misma es cero, lo que implicaría la imposibilidad de obtener su inversa. En caso de que la multicolinealidad y singularidad no se manifiesten perfectamente, el determinante será cercano a cero, admitiendo así que la inversa contenga valores muy inestables, sensibles al mínimo cambio en las correlaciones presentes en la matriz original.

2.3.2. Función discriminante de Fisher

Si bien se tiene la alternativa de asignar nuevos elementos mediante el cálculo de distancias, otra forma de llevarla a cabo, es obteniendo una regla o función discriminante, mediante el *método de Fisher,* éste se basa en el cumplimiento de algunos supuestos por parte de las poblaciones en estudio. Por lo cual, una vez que ya se haya realizado la comprobación del cumplimiento de todos los factores que hacen posible su aplicación, es permisible proceder a la construcción de la *función discriminante lineal de Fisher*, la que se procederá a desarrollar a continuación.

Se visualizan las observaciones para un grupo *i*, en forma de un arreglo matricial, su forma será como la que a continuación se presenta:

$$X_i = \begin{bmatrix} X_{i11} & X_{i21} & \cdots & X_{in1} \\ X_{i12} & X_{i22} & \cdots & X_{in2} \\ \vdots & \vdots & \vdots & \vdots \\ X_{i1m_i} & X_{i2m_i} & \cdots & X_{inm_i} \end{bmatrix}$$

Entonces, cada entrada de dicha matriz se denotará por X_{iwj}, que se entenderá por la *j-ésima* observación del *i-ésimo* grupo, en la variable *w*.

Por tanto, cada uno de los subíndices tendrá las siguientes características:

- $i = 1, 2, 3, \ldots, k$, con k números de grupos.
- $w = 1, 2, 3, \ldots, n$ hace referencia a la variable en cuestión y n es el total de variables discriminantes utilizadas en el proceso.
- $j = 1, 2, 3, \ldots, m_i$, con m_i el número de observaciones del grupo i

También, la función discriminante Z_i , para el i-ésimo grupo estará dada en forma matricial:

$$Z_i = X_i U$$

Donde U representa el vector de parámetros, es decir, $U' = (U_1, U_2, \ldots U_n)$ desarrollando, se obtiene lo siguiente:

$$\begin{bmatrix} Z_{i1} \\ Z_{i2} \\ \vdots \\ Z_{im_i} \end{bmatrix} = \begin{bmatrix} X_{i11} & X_{i21} & \cdots & X_{in1} \\ X_{i12} & X_{i22} & \cdots & X_{in2} \\ \vdots & \vdots & \vdots & \vdots \\ X_{i1m_i} & X_{i2m_i} & \cdots & X_{inm_i} \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{bmatrix}$$

De este modo, la función discriminante para cualquier observación *j* en el grupo *i*, quedará expresada por:

$$Z_{ij} = U_1 X_{i1j} + U_2 X_{i2j} + \ldots + U_n X_{inj}$$

Y la media de Z en el *i-ésimo* grupo será:

$$\overline{Z_i} = U_1 \overline{X}_{i1} + U_2 \overline{X}_{i2} + \ldots + U_n \overline{X}_{in}$$

Ahora bien, la desviación de cualquier Z_{ij} con respecto a $\overline{Z_i}$:

$$Z_{ij} - \overline{Z_i} = (U_1 X_{ij1} + U_2 X_{ij2} + \ldots + U_n X_{ijn}) - (U_1 \overline{X}_{i2} + U_2 \overline{X}_{i2} + \ldots + U_n \overline{X}_{in})$$

$$= U_1 (X_{ij1} - \overline{X}_{i1}) + U_2 (X_{ij2} - \overline{X}_{i2}) + \ldots + U_n (X_{ijn} - \overline{X}_{in})$$

$$= \sum_{w=1}^{n} U_w (X_{iwj} - \overline{X}_{iw})$$

Sea $x_{iwj} = (X_{iwj} - \overline{X}_{iw})$, entonces, $Z_{ij} - \overline{Z_i} = (U_1, U_2, \ldots U_n)\begin{pmatrix} x_{i1j} \\ x_{i2j} \\ \vdots \\ x_{inj} \end{pmatrix}$ la cual corresponde a la expresión matricial.

Tomando el cuadrado de estas diferencias, se tiene:

$$(Z_{ij} - \overline{Z_i})^2 = \left[(U_1, U_2, \ldots U_n)\begin{pmatrix} x_{i1j} \\ x_{i2j} \\ \vdots \\ x_{inj} \end{pmatrix} \right] \left[(U_1, U_2, \ldots U_n)\begin{pmatrix} x_{i1j} \\ x_{i2j} \\ \vdots \\ x_{inj} \end{pmatrix} \right]'$$

$$= \left[(U_1, U_2, \ldots U_n)\begin{pmatrix} x_{i1j} \\ x_{i2j} \\ \vdots \\ x_{inj} \end{pmatrix} \right] \left[\begin{pmatrix} x_{i1j} \\ x_{i2j} \\ \vdots \\ x_{inj} \end{pmatrix}' (U_1, U_2, \ldots U_n)' \right]$$

$$= (U_1, U_2, \ldots U_n)\begin{pmatrix} x_{i1j} \\ x_{i2j} \\ \vdots \\ x_{inj} \end{pmatrix} (x_{i1j}, x_{i2j}, \ldots, x_{inj}) \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{pmatrix}$$

$$= U' \begin{pmatrix} x_{i1j} \\ x_{i2j} \\ \vdots \\ x_{inj} \end{pmatrix} (x_{i1j}, x_{i2j}, \ldots, x_{inj}) U$$

$$= U' \begin{pmatrix} x_{i1j}x_{i1j} & x_{i1j}x_{i2j} & \cdots & x_{i1j}x_{inj} \\ x_{i2j}x_{i1j} & x_{i2j}x_{i2j} & \cdots & x_{i2j}x_{inj} \\ \vdots & \vdots & \ddots & \vdots \\ x_{inj}x_{i1j} & x_{inj}x_{i2j} & \cdots & x_{inj}x_{inj} \end{pmatrix} U$$

Se toman estas diferencias para todos los elementos del *i-ésimo grupo*:

$$\sum_{j=1}^{m_i}(Z_{ij} - \overline{Z_i})^2 = U'\begin{pmatrix} \sum_{j=1}^{m_i} x_{i1j}x_{i1j} & \sum_{j=1}^{m_i} x_{i1j}x_{i2j} & \cdots & \sum_{j=1}^{m_i} x_{i1j}x_{inj} \\ \sum_{j=1}^{m_i} x_{i2j}x_{i1j} & \sum_{j=1}^{m_i} x_{i2j}x_{i2j} & \cdots & \sum_{j=1}^{m_i} x_{i2j}x_{inj} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^{m_i} x_{inj}x_{i1j} & \sum_{j=1}^{m_i} x_{inj}x_{i2j} & \cdots & \sum_{j=1}^{m_i} x_{inj}x_{inj} \end{pmatrix}U$$

Dentro de dicha expresión, $\begin{pmatrix} \sum_{j=1}^{m_i} x_{i1j}x_{i1j} & \sum_{j=1}^{m_i} x_{i1j}x_{i2j} & \cdots & \sum_{j=1}^{m_i} x_{i1j}x_{inj} \\ \sum_{j=1}^{m_i} x_{i2j}x_{i1j} & \sum_{j=1}^{m_i} x_{i2j}x_{i2j} & \cdots & \sum_{j=1}^{m_i} x_{i2j}x_{inj} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^{m_i} x_{inj}x_{i1j} & \sum_{j=1}^{m_i} x_{inj}x_{i2j} & \cdots & \sum_{j=1}^{m_i} x_{inj}x_{inj} \end{pmatrix}$ es la matriz de sumas de cuadrados y productos cruzados dentro del grupo *i*, la cual se denotará por $m_i W_i$. Entonces,

$$\sum_{j=1}^{m_i}(Z_{ij} - \overline{Z_i})^2 = U'\, m_i W_i\, U$$

Generalizando para los *k* grupos, se tiene:

$$\sum_{i=1}^{k}\sum_{j=1}^{m_i}(Z_{ij} - \overline{Z_i})^2 = \sum_{i=1}^{k} U'\, m_i W_i\, U = U'\left[\sum_{i=1}^{k} m_i W_i\right]U$$

Sea $W = \sum_{i=1}^{k} m_i S_i$, a su vez, *W* se conoce como la suma de cuadrados dentro de los grupos.

Por lo tanto, la ***variación dentro de grupos o intragrupos***, estará dada por:

$$\sum_{i=1}^{k}\sum_{j=1}^{m_i}(Z_{ij} - \overline{Z_i})^2 = U'\, W\, U \qquad \textbf{.................... (2)}$$

Por otro lado, se quiere conocer la ***variación entre grupos***, para ello se recurre a $\overline{Z}$, la cual es el estimador de la media general de la variable *Z* ($\overline{Z}$). De este modo, se obtiene la diferencia $\overline{Z_i} - \overline{Z}$ para $i = 1, 2, 3, \dots, k$, es decir, para los k grupos; en donde $\overline{Z} = U_1 \overline{X}_1 + U_2 \overline{X}_2 + \dots + U_n \overline{X}_n$, con $\overline{X}_w$ la media general de la *w-ésima* variable.

Entonces,

$$\overline{Z_i} - \overline{Z} = (U_1 \overline{X}_{i1} + U_2 \overline{X}_{i2} + \dots + U_n \overline{X}_{in}) - (U_1 \overline{X}_1 + U_2 \overline{X}_2 + \dots + U_n \overline{X}_n)$$

$$= U_1(\overline{X}_{i1} - \overline{X}_1) + U_2(\overline{X}_{i2} - \overline{X}_2) + \dots + U_n(\overline{X}_{in} - \overline{X}_n)$$

$$= \sum_{w=1}^{n} U_w(\overline{X}_{iw} - \overline{X}_w)$$

En términos matriciales:

$$\overline{Z_i} - \overline{Z} = (U_1, U_2, \dots, U_n)\begin{pmatrix} \overline{X}_{i1} - \overline{X}_1 \\ \overline{X}_{i2} - \overline{X}_2 \\ \vdots \\ \overline{X}_{in} - \overline{X}_n \end{pmatrix} = (U_1, U_2, \dots, U_n)\begin{pmatrix} \overline{x}_{i1} \\ \overline{x}_{i2} \\ \vdots \\ \overline{x}_{in} \end{pmatrix} \text{ con } \overline{x}_{iw} = \overline{X}_{iw} - \overline{X}_w$$

Elevando al cuadrado, se tiene:

$$(\overline{Z_i} - \overline{Z})^2 = \left[(U_1, U_2, \dots, U_n)\begin{pmatrix} \overline{x}_{i1} \\ \overline{x}_{i2} \\ \vdots \\ \overline{x}_{in} \end{pmatrix} \right] \left[(U_1, U_2, \dots, U_n)\begin{pmatrix} \overline{x}_{i1} \\ \overline{x}_{i2} \\ \vdots \\ \overline{x}_{in} \end{pmatrix} \right]'$$

$$= \left[(U_1, U_2, \ldots, U_n) \begin{pmatrix} \bar{x}_{i1} \\ \bar{x}_{i2} \\ \vdots \\ \bar{x}_{in} \end{pmatrix} \right] \left[\begin{pmatrix} \bar{x}_{i1} \\ \bar{x}_{i2} \\ \vdots \\ \bar{x}_{in} \end{pmatrix}' (U_1, U_2, \ldots, U_n)' \right]$$

$$= (U_1, U_2, \ldots, U_n) \begin{pmatrix} \bar{x}_{i1} \\ \bar{x}_{i2} \\ \vdots \\ \bar{x}_{in} \end{pmatrix} (\bar{x}_{i1}, \bar{x}_{i2}, \ldots, \bar{x}_{in}) \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{pmatrix}$$

$$= U' \begin{pmatrix} \bar{x}_{i1} \\ \bar{x}_{i2} \\ \vdots \\ \bar{x}_{in} \end{pmatrix} (\bar{x}_{i1}, \bar{x}_{i2}, \ldots, \bar{x}_{in}) U$$

$$= U' \begin{bmatrix} \bar{x}_{i1}\bar{x}_{i1} & \bar{x}_{i1}\bar{x}_{i2} & \ldots & \bar{x}_{i1}\bar{x}_{in} \\ \bar{x}_{i2}\bar{x}_{i1} & \bar{x}_{i2}\bar{x}_{i2} & \ldots & \bar{x}_{i2}\bar{x}_{in} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{x}_{in}\bar{x}_{i1} & \bar{x}_{in}\bar{x}_{i2} & \ldots & \bar{x}_{in}\bar{x}_{in} \end{bmatrix} U$$

En donde:

$$\begin{bmatrix} \bar{x}_{i1}\bar{x}_{i1} & \bar{x}_{i1}\bar{x}_{i2} & \ldots & \bar{x}_{i1}\bar{x}_{in} \\ \bar{x}_{i2}\bar{x}_{i1} & \bar{x}_{i2}\bar{x}_{i2} & \ldots & \bar{x}_{i2}\bar{x}_{in} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{x}_{in}\bar{x}_{i1} & \bar{x}_{in}\bar{x}_{i2} & \ldots & \bar{x}_{in}\bar{x}_{in} \end{bmatrix}$$

corresponde a la matriz de productos cruzados de diferencias de los centroides del *i-ésimo grupo* con respecto al centroide general, esta matriz se denotará como $m_i B_i$.

Ahora para todos los grupos se tendría lo siguiente:

$$\sum_{i=1}^{k}(\overline{Z_i}-\overline{Z})^2 = U'\begin{bmatrix} \sum_{i=1}^{k}\overline{x}_{i1}\overline{x}_{i1} & \sum_{i=1}^{k}\overline{x}_{i1}\overline{x}_{i2} & \cdots & \sum_{i=1}^{k}\overline{x}_{i1}\overline{x}_{in} \\ \sum_{i=1}^{k}\overline{x}_{i2}\overline{x}_{i1} & \sum_{i=1}^{k}\overline{x}_{i2}\overline{x}_{i2} & \cdots & \sum_{i=1}^{k}\overline{x}_{i2}\overline{x}_{in} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{k}\overline{x}_{in}\overline{x}_{i1} & \sum_{i=1}^{k}\overline{x}_{in}\overline{x}_{i2} & \cdots & \sum_{i=1}^{k}\overline{x}_{in}\overline{x}_{in} \end{bmatrix}U$$

$$= U'\left[\sum_{i=1}^{k} m_i B_i\right]U$$

Sea $B = \sum_{i=1}^{k} m_i B_i$ la suma de cuadrados de las diferencias de los centroides de cada grupo con respecto al centroide general.

Por lo tanto, la ***variación entre grupos*** estará dada por la siguiente expresión:

$$\sum_{i=1}^{k}(\overline{Z_i}-\overline{Z})^2 = U'BU \qquad \text{………………..} \ \mathbf{(3)}$$

Posteriormente, para encontrar los componentes de la función discriminante, se procede a:

1. Disminuir las variaciones dentro de cada grupo, como en la expresión **(2)***,* con el fin de evitar cruces entre conjuntos.

2. Maximizar diferencias entre los centros de los grupos, es decir, las variaciones entre grupos deberán ser mayores, correspondientes a la variable Z, dados por la expresión **(3)**.

Ahora bien, el cociente constituido con las características anteriores y que se tendrá que maximizar es el siguiente:

$$\lambda = \frac{Variación\ entre\ grupos}{Variación\ dentro\ de\ grupos}$$

Es así como, el cociente λ a maximizar estará dado por:

$$\lambda = \frac{U'B\,U}{U'W\,U} \qquad \text{………………….. (4)}$$

Para ello, se toma la derivada parcial de λ con respecto de U .

$$\frac{\partial\lambda}{\partial U} = \frac{\partial}{\partial U}\,\frac{U'B\,U}{U'W\,U}$$

$$\frac{\partial\lambda}{\partial U} = \frac{U'W\,U\,\dfrac{\partial\,U'B\,U}{\partial U} - U'B\,U\,\dfrac{\partial\,U'W\,U}{\partial U}}{(U'W\,U)^2}$$

$$\frac{\partial\lambda}{\partial U} = \frac{U'W\,U\,(2BU) - U'B\,U\,(2WU)}{(U'W\,U)^2}$$

$$\frac{\partial\lambda}{\partial U} = \frac{2\,(U'W\,UBU - U'B\,UWU)}{(U'W\,U)^2}$$

$$\frac{\partial\lambda}{\partial U} = \frac{2\,(U'W\,UBU - U'B\,UWU)}{(U'W\,U)^2}\;\frac{\dfrac{1}{U'WU}}{\dfrac{1}{U'WU}}$$

$$\frac{\partial\lambda}{\partial U} = \frac{2\,(U'W\,UBU\,\dfrac{1}{U'WU} - U'B\,UWU\,\dfrac{1}{U'WU})}{U'W\,U}$$

$$\frac{\partial\lambda}{\partial U} = \frac{2\,(U'W\,UBU\,\dfrac{1}{U'WU} - U'B\,UWU\,\dfrac{1}{U'WU})}{U'W\,U}$$

$$\frac{\partial \lambda}{\partial U} = \frac{2\,(BU - U'B\,UWU\,\frac{1}{U'WU})}{U'W\,U}$$

Ya que $U'W\,U$ y $U'B\,U$ son escalares, se tiene lo siguiente:

$$\frac{\partial \lambda}{\partial U} = \frac{2\,(BU - WU\,\frac{U'B\,U}{U'WU})}{U'W\,U}$$

$$\frac{\partial \lambda}{\partial U} = \frac{2\,(BU - WU\lambda)}{U'W\,U}$$

Posteriormente, $\frac{\partial \lambda}{\partial U}$ se iguala al vector cero:

$$\frac{2\,(BU - WU\lambda)}{U'W\,U} = 0$$

Lo cual sucede, si y sólo si, $2\,(BU - WU\lambda) = 0$, es decir, si $BU - WU\lambda = 0$, factorizando U :

$$(B - W\lambda)\,U = 0$$

En seguida, se multiplica por W^{-1}, siempre y cuando W sea no singular, de lo contrario no será posible.

$$W^{-1}(B - W\lambda)\,U = 0$$

$$(W^{-1}B - W^{-1}W\lambda)\,U = 0$$

$$(W^{-1}B - W^{-1}W\lambda)\,U = 0$$

$$(W^{-1}B - \lambda I)\,U = 0 \qquad \text{……………….. } \mathbf{(5)}$$

Por lo tanto, el problema se sintetiza en encontrar los valores y vectores propios de la ecuación anterior.

Es importante destacar que para el caso en el que se tiene dos grupos, es decir, *k=2* la variación entre grupos será:

$$\sum_{i=1}^{2}(\overline{Z_i}-\overline{Z})^2=\left[\overline{Z_1}-\frac{\overline{Z_1}+\overline{Z_2}}{2}\right]^2+\left[\overline{Z_2}-\frac{\overline{Z_1}+\overline{Z_2}}{2}\right]^2$$

$$=\left[\frac{\overline{Z_1}-\overline{Z_2}}{2}\right]^2+\left[\frac{\overline{Z_2}-\overline{Z_1}}{2}\right]^2=\frac{\left[\overline{Z_1}-\overline{Z_2}\right]^2}{2}$$

Sea $\underline{X_i}'=(\overline{X}_{i1},\overline{X}_{i2},\ldots,\overline{X}_{in})$ el vector de medias para el i-ésimo grupo, con lo que:

$$(\overline{Z_1}-\overline{Z_2})=U'\,\underline{\overline{X}_1}-U'\,\underline{\overline{X}_2}$$

Elevando al cuadrado, de ambos lados, con las correspondientes operaciones permisibles en ambos lados, se tiene:

$$(\overline{Z_1}-\overline{Z_2})^2=(U'\,\underline{\overline{X}_1}-U'\,\underline{\overline{X}_2})\;(U'\,\underline{\overline{X}_1}-U'\,\underline{\overline{X}_2})'$$

$$(\overline{Z_1}-\overline{Z_2})^2=(U'\,\underline{\overline{X}_1}-U'\,\underline{\overline{X}_2})\,(\underline{\overline{X}_1}'\,U''-\underline{\overline{X}_2}'\,U'')$$

$$(\overline{Z_1}-\overline{Z_2})^2=(U'\,\underline{\overline{X}_1}-U'\,\underline{\overline{X}_2})\,(\underline{\overline{X}_1}'\,U-\underline{\overline{X}_2}'\,U)$$

$$(\overline{Z_1}-\overline{Z_2})^2=U'\,\underline{\overline{X}_1}\,\underline{\overline{X}_1}'U-U'\,\underline{\overline{X}_1}\,\underline{\overline{X}_2}'U-U'\,\underline{\overline{X}_2}\,\underline{\overline{X}_1}'U+U'\,\underline{\overline{X}_2}\,\underline{\overline{X}_2}'U$$

$$(\overline{Z_1}-\overline{Z_2})^2=U'\,(\underline{\overline{X}_1}\,\underline{\overline{X}_1}'U-\underline{\overline{X}_1}\,\underline{\overline{X}_2}'U-\underline{\overline{X}_2}\,\underline{\overline{X}_1}'U+\underline{\overline{X}_2}\,\underline{\overline{X}_2}'U)$$

$$(\overline{Z_1}-\overline{Z_2})^2=U'\,(\underline{\overline{X}_1}\,\underline{\overline{X}_1}'-\underline{\overline{X}_1}\,\underline{\overline{X}_2}'-\underline{\overline{X}_2}\,\underline{\overline{X}_1}'+\underline{\overline{X}_2}\,\underline{\overline{X}_2}')\,U$$

$$(\overline{Z_1} - \overline{Z_2})^2 = U' (\underline{X_1} (\underline{X_1}' - \underline{X_2}') - \underline{X_2}\, \underline{X_1}' + \underline{X_2}\, \underline{X_2}') U$$

$$(\overline{Z_1} - \overline{Z_2})^2 = U' (\underline{X_1}(\underline{X_1}' - \underline{X_2}') - \underline{X_2}(\underline{X_1}' - \underline{X_2}')) U$$

$$(\overline{Z_1} - \overline{Z_2})^2 = U' (\underline{X_1} - \underline{X_2})(\underline{X_1}' - \underline{X_2}') U$$

$$(\overline{Z_1} - \overline{Z_2})^2 = U' (\underline{X_1} - \underline{X_2})(\underline{X_1} - \underline{X_2})' U$$

Por lo tanto, $B = \dfrac{(\underline{X_1} - \underline{X_2})(\underline{X_1} - \underline{X_2})'}{2}$

Entonces cuando *k=2,* la solución de la ecuación **(5)**, será la siguiente:

$$(W^{-1}B - \lambda I)\, U = (W^{-1} \frac{(\underline{X_1} - \underline{X_2})(\underline{X_1} - \underline{X_2})'}{2} - \lambda I)\, U = 0$$

$$\lambda I U = W^{-1} \frac{(\underline{X_1} - \underline{X_2})(\underline{X_1} - \underline{X_2})'}{2} U$$

$$\lambda U = W^{-1} \frac{(\underline{X_1} - \underline{X_2})(\underline{X_1} - \underline{X_2})'}{2} U$$

Por tanto, el valor de los escalares de la función discriminante, con *k=2* será:

$$U = W^{-1}(\underline{X_1} - \underline{X_2})\, C$$

Con C un escalar, resultado de una multiplicación de matrices, en la práctica este escalar se suele igualar a uno, ya que no influye de forma definitiva en la discriminación de los elementos.

$$C_{1\times 1} = \frac{\dfrac{(\underline{\overline{X}}_1 - \underline{\overline{X}}_2)'_{1\times n}}{2} U_{n\times 1}}{\lambda_{1\times 1}}$$

Entonces, el punto de corte para asignar los elementos, será el promedio de las funciones discriminantes de cada grupo y se asignará el elemento $\underline{x}$ a la población uno, en el caso en el que suceda lo siguiente:

$$Z_0 > \frac{\overline{Z_1} + \overline{Z_2}}{2}$$

Sustituyendo el valor de Z para cada uno de los grupos, se tiene:

$$U'\underline{x} > \frac{(W^{-1}(\underline{\overline{X}}_1 - \underline{\overline{X}}_2))'\,\underline{\overline{X}}_1 + (W^{-1}(\underline{\overline{X}}_1 - \underline{\overline{X}}_2))'\,\underline{\overline{X}}_2}{2}$$

Aplicando propiedades de la transpuesta de un producto,

$$U'\underline{x} > \frac{(\underline{\overline{X}}_1 - \underline{\overline{X}}_2)'W^{-1}\underline{\overline{X}}_1 + (\underline{\overline{X}}_1 - \underline{\overline{X}}_2)'W^{-1}\,\underline{\overline{X}}_2}{2}$$

Factorizando,

$$U'\underline{x} - \frac{(\underline{\overline{X}}_1 - \underline{\overline{X}}_2)'W^{-1}(\underline{\overline{X}}_1 + \underline{\overline{X}}_2)}{2} > 0$$

$$U'\underline{x} - \frac{(W^{-1}(\underline{\overline{X}}_1 - \underline{\overline{X}}_2))'(\underline{\overline{X}}_1 + \underline{\overline{X}}_2)}{2} > 0$$

$$U'\underline{x} - \frac{U'(\underline{\overline{X}}_1 + \underline{\overline{X}}_2)}{2} > 0$$

$$U'\left[\underline{x} - \frac{(\underline{\overline{X}}_1 + \underline{\overline{X}}_2)}{2}\right] > 0$$

Por lo tanto, se asigna el elemento $\underline{x}$ de la siguiente forma:

$$\begin{cases} Si \quad U'\left[\underline{x} - \dfrac{(\underline{\bar{X}}_1 + \underline{\bar{X}}_2)}{2}\right] > 0 \quad \text{Asignar a } \boldsymbol{P_1} \\ \\ Si \quad U'\left[\underline{x} - \dfrac{(\underline{\bar{X}}_1 + \underline{\bar{X}}_2)}{2}\right] \leq 0 \quad \text{Asignar a } \boldsymbol{P_2} \end{cases}$$

En cualquier caso se procede a asignar la observación $\underline{x}$ a la *i-ésima* población según su valor en la puntuación discriminante y sus diferencias con respecto a las obtenidas por las medias muestrales, ello se expresa mediante la siguiente ***regla de asignación de Fisher***:

$$\left| U'(\underline{x} - \underline{\bar{X}}_i) \right| < \left| U'(\underline{x} - \underline{\bar{X}}_j) \right| \qquad \forall \;\; i \neq j \qquad \textbf{.................. (6)}$$

2.3.3. Regla de máxima verosimilitud

Una forma alternativa de clasificar algún individuo a la población más cercana es por medio de la *regla de máxima verosimilitud* o por la *regla de Bayes*, ello partiendo de la hipótesis de que las variables observadas en el *i-ésimo* grupo tienen una distribución de probabilidad conocida de parámetros ϕ_i en P_i, la probabilidad o verosimilitud de cualquier observación $\underline{x} = (x_1, x_2, \ldots, x_n)$ en P_i, estará dada por $f(x_1, x_2, \ldots, x_n; \phi_i)$[7].

$$L_i(\underline{x}) = f(x_1, x_2, \ldots, x_n; \phi_i) \text{ ; con } \; i = 1,2,3,\ldots,k$$

Es decir,

$$L_i(\underline{x}) = f(x_1, x_2, \ldots, x_n; \phi_i) = \prod_{j=1}^{n} f(x_j; \phi_i) \qquad \textbf{.................. (7)}$$

[7] Esta función también puede tomarse como una probabilidad condicional, denotada por $f(\underline{x} \mid i)$.

Por ello, una regla de clasificación aceptable será asignar la observación a la población P_i para la cual la verosimilitud de la observación es mayor, es decir, se asignará el elemento x a P_j si:

$$L_j(\underline{x}) = máx\{L_1(x), L_2(x), \ldots, L_k(x)\} \qquad \textbf{.................. (8)}$$

Por lo tanto, en el caso *k=2,*

$$Ln\ L_i(\underline{x}) - Ln\ L_j(\underline{x}) > 0$$

La función discriminante que obedece a este criterio, estará dada por:

$$V(\underline{x}) = Ln\ L_i(\underline{x}) - Ln\ L_j(\underline{x})$$

Por lo tanto, la ***regla de decisión por máxima verosimilitud*** queda de la siguiente manera, asignar a:

$$\begin{array}{lll} P_i & si & V(\underline{x}) > 0 \\ P_j & si & V(\underline{x}) \leq 0 \end{array} \qquad \textbf{.................. (9)}$$

De este modo, en poblaciones normales**,** esta regla se expresará de la forma en la que a continuación se desarrollará.

La función de verosimilitud para la *i-ésima población* será.

$$L_i(\underline{x}) = \frac{1}{(2\pi)^{n/2}\left|\Sigma\right|^{1/2}} \exp\left\{-\frac{1}{2}(\underline{x}-\underline{\mu}_i)'\Sigma^{-1}(\underline{x}-\underline{\mu}_i)\right\} \qquad \textbf{..................(10)}$$

Entonces la expresión **(7)**, quedará como sigue**:**

$$L_i(\underline{x}) - L_j(\underline{x}) > 0$$

$$\frac{1}{(2\pi)^{n/2}\left|\Sigma\right|^{1/2}}\exp\left\{-\frac{1}{2}(\underline{x}-\underline{\mu}_i)'\Sigma^{-1}(\underline{x}-\underline{\mu}_i)\right\}-\frac{1}{(2\pi)^{n/2}\left|\Sigma\right|^{1/2}}\exp\left\{-\frac{1}{2}(\underline{x}-\underline{\mu}_j)'\Sigma^{-1}(\underline{x}-\underline{\mu}_j)\right\}>0$$

$$\frac{1}{(2\pi)^{n/2}\left|\Sigma\right|^{1/2}}\left[\exp\left\{-\frac{1}{2}(\underline{x}-\underline{\mu}_i)'\Sigma^{-1}(\underline{x}-\underline{\mu}_i)\right\}-\exp\left\{-\frac{1}{2}(\underline{x}-\underline{\mu}_j)'\Sigma^{-1}(\underline{x}-\underline{\mu}_j)\right\}\right]>0$$

$$\exp\left\{-\frac{1}{2}(\underline{x}-\underline{\mu}_i)'\Sigma^{-1}(\underline{x}-\underline{\mu}_i)\right\}>\exp\left\{-\frac{1}{2}(\underline{x}-\underline{\mu}_j)'\Sigma^{-1}(\underline{x}-\underline{\mu}_j)\right\}$$

$$Ln\left[\exp\left\{-\frac{1}{2}(\underline{x}-\underline{\mu}_i)'\Sigma^{-1}(\underline{x}-\underline{\mu}_i)\right\}\right]>Ln\left[\exp\left\{-\frac{1}{2}(\underline{x}-\underline{\mu}_j)'\Sigma^{-1}(\underline{x}-\underline{\mu}_j)\right\}\right]$$

Es decir:

$$\left[-\frac{1}{2}(\underline{x}-\underline{\mu}_i)'\Sigma^{-1}(\underline{x}-\underline{\mu}_i)\right]-\left[-\frac{1}{2}(\underline{x}-\underline{\mu}_j)'\Sigma^{-1}(\underline{x}-\underline{\mu}_j)\right]>0$$

Con ello, a lo que se llega es:

$$V(\underline{x})=U'(\underline{x}-\underline{\mu})$$

Donde $U'=\left[\Sigma^{-1}(\underline{\mu}_i-\underline{\mu}_j)\right]'$ y $\underline{\mu}=\frac{1}{2}(\underline{\mu}_i+\underline{\mu}_j)$, con lo que se tendría que asignar a la *población i* si $V(\underline{x})>0$, en otro caso, a la *población j*, lo cual coincide con lo mostrado para el desarrollo con el método de Fisher, que a diferencia del de verosimilitudes, muestra un enfoque no paramétrico para hallar la función discriminante.

Ahora bien, cuando ya se tiene la función de verosimilitud y suponiendo que se conoce la probabilidad con la que un elemento pertenece al *i-ésimo* grupo[8], denotado como q_i, es viable establecer una regla discriminante que incluya ambos supuestos, es decir:

[8] Este tipo de probabilidad es conocida comúnmente como *probabilidad a priori,* ya que ésta se conoce incluso antes de realizar algún experimento o en este caso una asignación de un individuo a una población.

$$\underline{x} \text{ es de } P_i \text{ si } q_i L_i(\underline{x}) = máx\left\{ q_1 L_1(\underline{x}), q_2 L_2(\underline{x}), \ldots, q_k L_k(\underline{x}) \right\}$$

Aplicando el Teorema de Bayes[9], se tiene que:

$$P(i \cap \underline{x}) = P(\underline{x})\ P(i \mid \underline{x}) = q_i P(\underline{x} \mid i)$$

Por tanto, $P(i \mid \underline{x}) = \dfrac{q_g P(\underline{x} \mid i)}{P(\underline{x})}$

Pero $P(\underline{x}) = \sum_{i'=1}^{k} q_i P(\underline{x} \mid i')$, por lo que $P(i \mid \underline{x})$ quedaría expresada como:

$$P(i \mid \underline{x}) = \frac{q_i P(\underline{x} \mid i)}{\sum_{i'=1}^{k} q_{i'} P(\underline{x} \mid i')}$$

Formulado en términos de funciones de máxima verosimilitud, la probabilidad de que el elemento pertenezca a alguno de los grupos[10] dado $\underline{x}$, queda expresada como:

$$P(i \mid \underline{x}) = \frac{q_i L_i(\underline{x})}{\sum_{i'=1}^{k} q_{i'} L_{i'}(\underline{x})} \quad \text{para algún} \quad i = 1, 2, \ldots, k$$

Entonces la regla de decisión será, asignar la unidad *x* a la *i-ésima* población si:

$$P(i \mid \underline{x}) > P(i' \mid \underline{x}) \quad \text{con } i \neq i'$$

[9] Si los eventos $B_1, B_2, B_3, .., B_k$ constituyen una partición del espacio muestral *S,* donde $P(B_i) \neq 0$ para $i = 1,2,3,..., k$. Entonces para cualquier evento *A* en *S,* tal que $P(A) \neq 0$,

$$P(B_r \mid A) = \frac{P(B_r \cap A)}{\sum_{i=1}^{k} P(B_i \cap A)} = \frac{P(B_r)P(A \mid B_r)}{\sum_{i=1}^{k} P(B_i)P(A \mid B_i)} \text{ para } r = 1,2,3,..., k.$$

[10] Este tipo de probabilidad es conocida como *probabilidad a posteriori,* ya que se obtiene a partir de una *probabilidad a priori* y una probabilidad deducida por algún experimento, en este caso, obtenida mediante el comportamiento de las poblaciones.

La función discriminante asociada *k=2,* es la siguiente:

$$\frac{q_i L_i(\underline{x})}{\sum_{i'=1}^{k} q_{i'} L_{i'}(\underline{x})} > \frac{q_j L_j(\underline{x})}{\sum_{i'=1}^{k} q_{i'} L_{i'}(\underline{x})}$$

Debido a que el denominador en ambos lados es el mismo, esta desigualdad se cumple si $q_i L_i(\underline{x}) > q_j L_j(\underline{x})$. Además, ya que en ambos lados se tienen cantidades mayores a cero, es posible aplicar el logaritmo:

$$Ln\{\ q_i L_i(\underline{x})\ \} > Ln\{\ q_j L_j(\underline{x})\ \}$$

$$Ln\ q_i + Ln(L_i(\underline{x})) > Ln\ q_j + Ln\ (L_j(\underline{x}))$$

$$Ln\ q_i - Ln\ q_j + Ln(L_i(\underline{x})) - Ln(L_j(\underline{x})) > 0$$

$$Ln\ (L_i(\underline{x})) - Ln\ (L_j(\underline{x})) + Ln\ \frac{q_i}{q_j} > 0$$

Con esto, la ***regla discriminante de Bayes*** será asignar el nuevo elemento a la *i-ésima* población si $B(\underline{x}) > 0$, de no ser así se asignan a *j,* con $B(\underline{x})$ dado por:

$$B(x) = Ln\ (L_i(\underline{x})) - Ln\ (L_j(\underline{x})) + Ln\ \frac{q_i}{q_j} \qquad \textbf{……………….(11)}$$

Es importante destacar que, usando la regla de Bayes, el número total de clasificaciones erróneas será minimizado, ello se debe al uso de la función de máxima verosimilitud y por el uso de probabilidades a priori.

Se observa que cuando $q_i = q_j = \frac{1}{2}$ se tiene que $B(\underline{x}) = V(\underline{x})$.

2.3.4. Criterio de clasificación por medio de distancias

Una forma alternativa de clasificar a un individuo dentro de un grupo es partiendo del supuesto de que éste pertenecerá a un conjunto si sus características son similares, es decir, la distancia de ese vector de observaciones con respecto al comportamiento promedio mostrado por el grupo, es menor a la que se obtiene con respecto a otros.

Existen diversos modelos de distancias estadísticas que podrían ser utilizados, cuando no se toma en cuenta la distribución de los datos objeto de estudio y éstos son de tipo cuantitativo. Algunos de ellos se enuncian a continuación.

Sea $\underline{x} = (x_1, x_2, \ldots, x_n)$ una observación en $\Re^n$, $\underline{y} = (\overline{y_1}, \overline{y_2}, \ldots, \overline{y_n})$ el vector de medias de la población *y,* entonces la distancia de Minkowsky se define como:

$$d_q(\underline{x}, \underline{y}) = \left[\sum_{i=1}^{n} | x_i - \overline{y_i} |^q \right]^{1/q}$$

Casos particulares de la distancia d_q son:

- Distancia Ciudad:

$$d_1(\underline{x}, \underline{y}) = \sum_{i=1}^{n} | x_i - \overline{y_i} |$$

- Distancia Euclídea

$$d_2(\underline{x}, \underline{y}) = \sqrt{\sum_{i=1}^{n} (x_i - \overline{y_i})^2}$$

- Distancia Dominante

$$d_\infty(\underline{x}, \underline{y}) = \max_{1 \le i \le n} \left\{ | x_i - \overline{y_i} | \right\}$$

Tienen también interés en las aplicaciones, la distancia normalizada por el rango R_i de la variable *i.*

$$d_G(\underline{x},\underline{y}) = \frac{1}{n}\sum_{i=1}^{n}\frac{|x_i - \overline{y_i}|}{R_i}$$

y, cuando los valores de las variables son positivos, la métrica de Canberra:

$$d_C(\underline{x},\underline{y}) = \frac{1}{n}\sum_{i=1}^{n}\frac{|x_i - \overline{y_i}|}{x_i + \overline{y_i}}$$

Es importante mencionar que, $d_G(\underline{x},\underline{y})$ y $d_C(\underline{x},\underline{y})$ son invariantes por cambios de escala.

Por otro lado, si se toma en cuenta la correlación entre variables, entonces se define la distancia de Mahalanobis entre los centroides de dos poblaciones de la siguiente manera:

$$D_{ij}^2 = (\overline{\underline{X}_i} - \overline{\underline{X}_j})' W^{-1} (\overline{\underline{X}_i} - \overline{\underline{X}_j}) \quad \textbf{..................(12)}$$

Donde $\overline{\underline{X}_i}$ es el vector de medias del grupo i, $\overline{\underline{X}_j}$ el vector de medias del grupo *j* y W^{-1} es la matriz inversa de varianzas y covarianzas dentro de los grupos.

La relación existente entre el método de Fisher y la distancia de Mahalanobis, se puede obtener si se sustituye el valor de U, encontrado mediante el método de Fisher en el cociente $\lambda = \frac{U'BU}{U'WU}$.

Se tiene lo siguiente:

$$\lambda = \frac{U'BU}{U'WU} = \frac{U'(\overline{\underline{X}_i} - \overline{\underline{X}_j})(\overline{\underline{X}_i} - \overline{\underline{X}_j})'U}{U'WU}$$

$$\lambda = \frac{(\overline{\underline{X}_i} - \overline{\underline{X}_j})'W^{-1}(\overline{\underline{X}_i} - \overline{\underline{X}_j})(\overline{\underline{X}_i} - \overline{\underline{X}_j})'W^{-1}(\overline{\underline{X}_i} - \overline{\underline{X}_j})}{(\overline{\underline{X}_i} - \overline{\underline{X}_j})'W^{-1}W(W^{-1}(\overline{\underline{X}_i} - \overline{\underline{X}_j}))}$$

$$\lambda = \frac{(\underline{\overline{X}}_i - \underline{\overline{X}}_j)'W^{-1}(\underline{\overline{X}}_i - \underline{\overline{X}}_j)\,(\underline{\overline{X}}_i - \underline{\overline{X}}_j)'W^{-1}(\underline{\overline{X}}_i - \underline{\overline{X}}_j)}{(\underline{\overline{X}}_i - \underline{\overline{X}}_j)'\,I\,(W^{-1}(\underline{\overline{X}}_i - \underline{\overline{X}}_j))}$$

$$\lambda = \frac{(\underline{\overline{X}}_i - \underline{\overline{X}}_j)'W^{-1}(\underline{\overline{X}}_i - \underline{\overline{X}}_j)\,(\underline{\overline{X}}_i - \underline{\overline{X}}_i)'W^{-1}(\underline{\overline{X}}_i - \underline{\overline{X}}_j)}{(\underline{\overline{X}}_i - \underline{\overline{X}}_j)'\,W^{-1}(\underline{\overline{X}}_i - \underline{\overline{X}}_j)}$$

$$\lambda = \frac{D_{i,j}^2 D_{i,j}^2}{D_{i,j}^2}$$

Debido a que $D_{i,j}^2 = (\underline{\overline{X}}_i - \underline{\overline{X}}_j)'\,W^{-1}(\underline{\overline{X}}_i - \underline{\overline{X}}_j)$, por lo tanto $\lambda = D_{i,j}^2$.

De esta forma, la asignación de un elemento se realiza comparando sus coordenadas con las del centro de cada grupo. La ***regla de asignación por distancia de Mahalanobis*** queda como sigue:

Asignar x a P_i $\quad si \quad D_{x,i}^2 < D_{x,j}^2$ **……………….. (13)**

Para cualquier j.

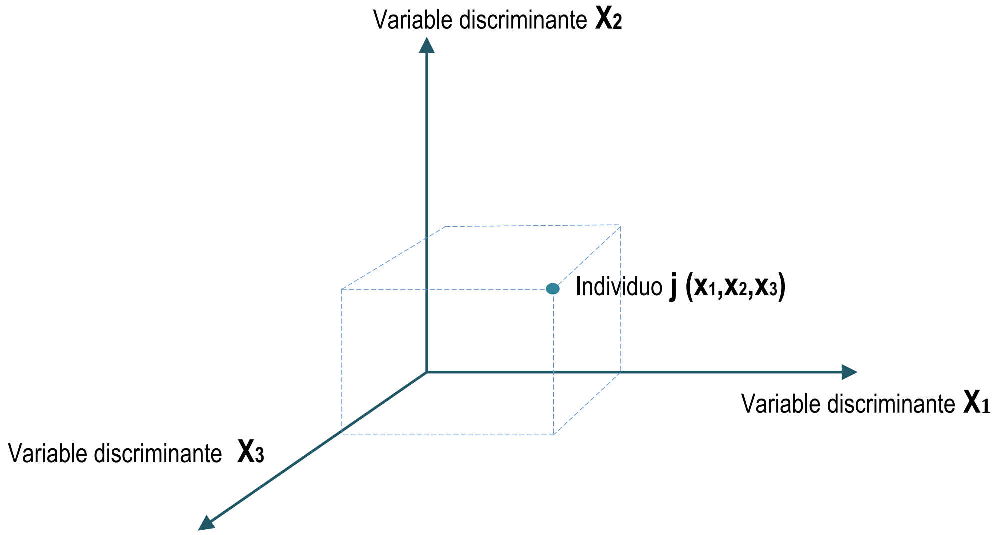
Cabe mencionar que, cualquiera que sea la distancia a elegir, la regla de clasificación basada en distancias será siempre la misma, es decir, se asignará al individuo $\underline{x} = (x_1, x_2, \ldots, x_n)$ a la población más próxima.

2.3.5. Interpretación geométrica de la función discriminante

La interpretación geométrica del AD, dará una noción más tangible de lo que ocurrirá al obtener las *funciones discriminantes,* para ello se considera a cada uno de los individuos que conforman la población a estudiar, como un punto en el espacio *n-dimensional* definido, ello al tomar a las variables discriminantes como ejes de tal espacio.

Supóngase que, *n=3*, entonces los valores alcanzados por un punto *j* en las tres variables constituirán las coordenadas de ese punto en el espacio tridimensional, como se muestra a continuación:

Ubicación de un punto j en el espacio definido por las variables discriminantes



De esta forma, no es difícil imaginar que si los individuos de un mismo grupo se comportan de manera similar respecto a las variables discriminantes, sus coordenadas serán análogas y con ello se distinguirán como un conjunto de puntos excluyentes unos de otros para cada grupo.

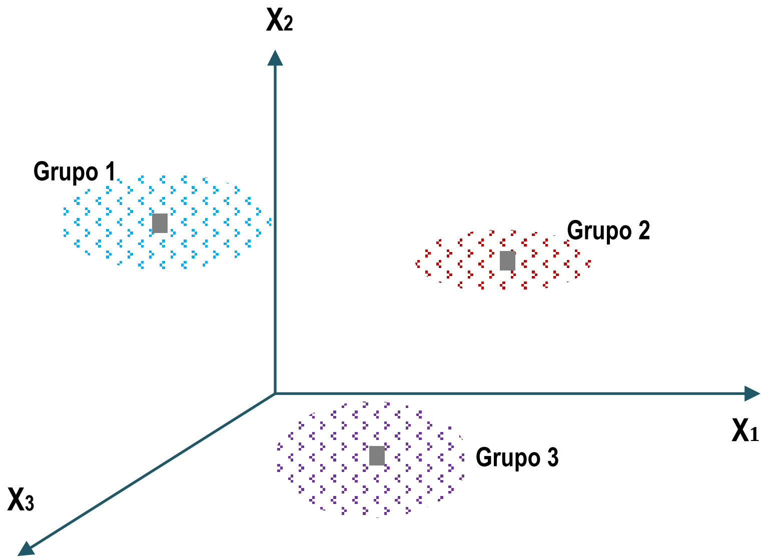
De este modo, la posición de cada grupo se verá caracterizada por su *centroide*, éste quedará definido como el punto que se obtiene al considerar los valores medios que el grupo de individuos presenta para cada una de las variables discriminantes.

Por lo tanto, para el *i-ésimo* grupo en cada una de las variables, el centroide quedará denotado como:

$$g_i\ (\overline{X}_{i1}, \overline{X}_{i2}, \overline{X}_{i3})$$

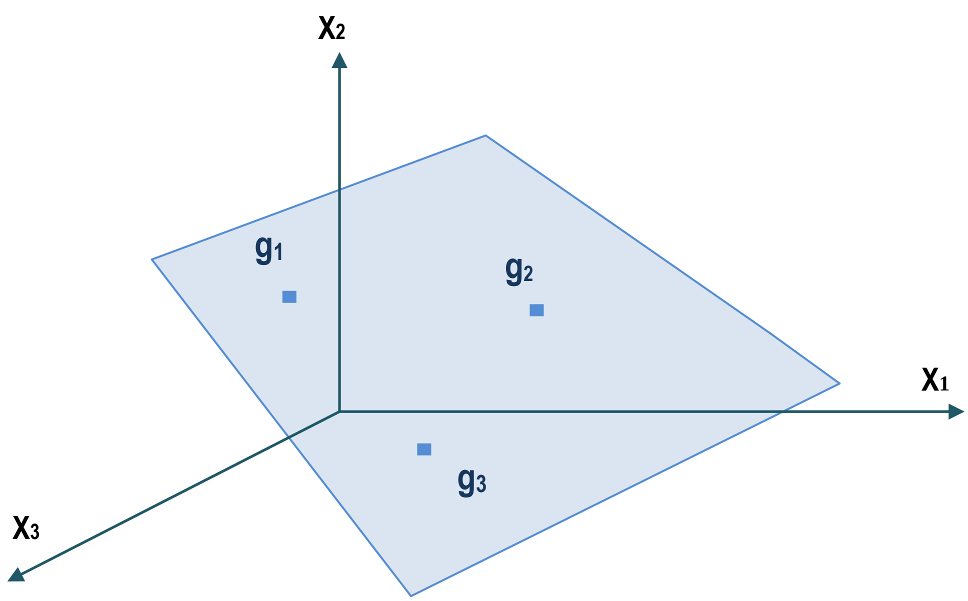
Entonces, si el problema consiste en describir diferencias entre los grupos en función de las variables discriminantes, a lo que se recurrirá es a examinar la posición de los centroides para determinar si éstos quedan suficientemente diferenciados.

Ubicación de los centroides de cada grupo y los puntos cercanos a los mismos



Para representar las posiciones relativas de los centroides correspondientes a cada uno de los grupos, no siempre es necesario permanecer en un espacio *n-dimensional*; por ejemplo, una recta será adecuada para ubicar los centroides si $n=2$, si $n=3$ se tratará de un plano, etcétera. Generalizando esta cuestión, se tiene que n centroides definen un espacio de dimensión $n-1$.

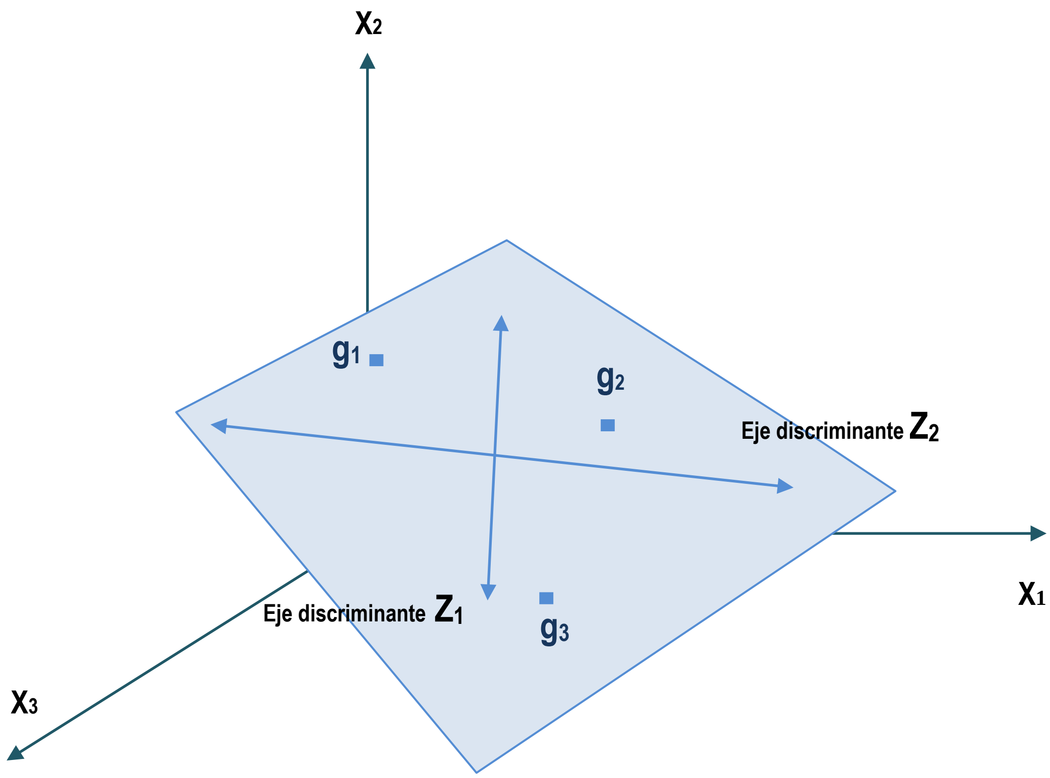
Posición relativa de los centroides cuando el número de grupos es igual a tres



Una vez que se tengan ubicados los centroides de cada grupo, se procede a encontrar el origen del espacio definido por los mismos, éste corresponde al *centroide general* o vector de medias todos los casos.

Posteriormente, se localizan los ejes y sus direcciones, el primer eje se localiza en la misma dirección en la que se produce la máxima dispersión; el segundo eje es del mismo modo con la condición extra de que este debe ser perpendicular al primero; lo mismo sucederá para los siguientes ejes. Para el caso en el que *n=3,* se tendría algo semejante a la figura que a continuación se muestra.

Localización de los ejes discriminantes Z_1 y Z_2



Finalmente, los ***ejes Z_1 y Z_2*** son las ***funciones discriminantes****,* ya que son interpretadas como ejes que maximizan las diferencias entre los grupos y resultan ortogonales entre sí. La obtención de estas funciones facilitará la transformación de las coordenadas en el espacio *n-dimensional* definido por las variables al nuevo sistema de ejes en el espacio definido por los centroides.

3. Ejemplos de segmentación de mercados

3.1. Prueba de producto para café

En este ejemplo se tomará una base de la empresa X, la cual es el resultado de una prueba de producto realizada a dos bebidas de café, que se llamará producto del cliente y producto del competidor. El objetivo es conocer cuáles son las características que más impactan en la preferencia de alguno de los productos y con base a las evaluaciones que se obtengan, predecir para los futuros encuestados cuál es el que más les agradará.

Se trata de una prueba comparativa, es decir, a cada individuo se le dio primero un producto y posteriormente a ello, el de la competencia. Para asegurar que los datos no estuvieran sesgados, se fue rotando aleatoriamente; por ejemplo, si al individuo 1 se le da a probar primero la bebida del cliente y luego el de la competencia, habrá otro que probará un orden inverso a este primero. De este modo, se tendrán 75 personas que habrán probado primero un producto y el resto el de la competencia.

La base consta de 150 entrevistas para la Ciudad de México, cuyas edades de los entrevistados oscilan entre 30 y 45 años, para el nivel socioeconómico (C+ / C)[11], divididas a su vez por

- ✓ Sexo
- ✓ Perfil de consumo (medium y heavy)
- ✓ Tipo de taza en la que se probó el producto (taza blanca o negra)

Asimismo, se evaluaron diferentes características, como lo son:

- ✓ Overall liking, que se refiere a el gusto en términos generales mostrado hacia los productos probados (escala 1-7)
- ✓ Apariencia en general del polvo (escala 1-7)
- ✓ Color del polvo (escala1-5)
- ✓ Apariencia en general de la bebida (escala 1-7)
- ✓ Aroma en general (escala 1-7)

[11] Se refieren a los niveles medio alto y medio, respectivamente.

- ✓ Intensidad del aroma en general (escala 1-5)
- ✓ Sabor en general (escala 1-7)
- ✓ Intensidad del sabor a café (escala 1-5)
- ✓ Amargo (escala 1-5)
- ✓ Sabor suave (escala 1-5)
- ✓ Suavidad (escala 1-5)
- ✓ Sabor que deja en su boca (escala 1-7)
- ✓ Intensidad del sabor que deja en su boca (escala 1-5)

Algunas preguntas tienen una escala de respuesta de tipo ordinal, que oscilan entre 1 y 7. Un ejemplo de ellas es la siguiente:

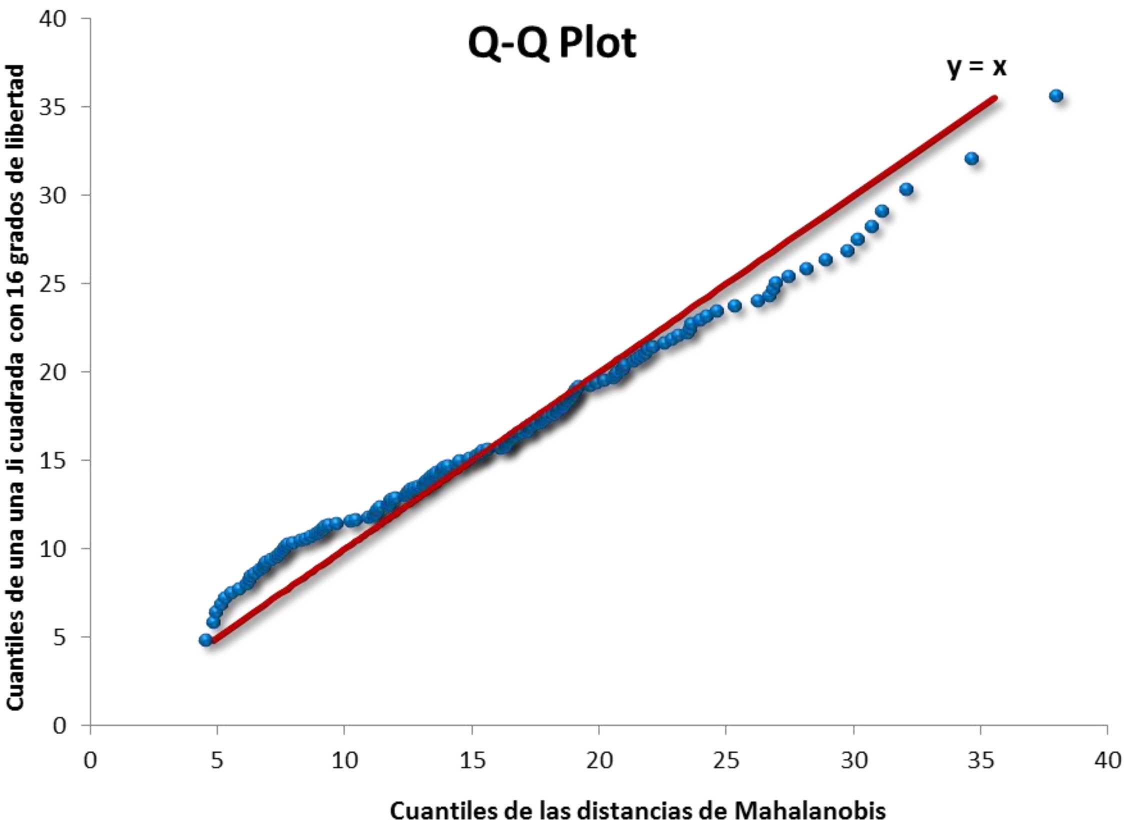
- ¿En términos generales, cómo evaluaría el sabor en general del café?
  Las opciones son:
  7-Excelente, 6-Muy bueno, 5-Algo bueno, 4-Ni bueno ni malo, 3-Un poco desagradable, 2-Muy desagradable y 1-Extremadamente desagradable.

Otras preguntas con escala de respuesta también ordinal oscilan de 1 a 5. Un ejemplo de éstas es:

- ¿Cómo evaluaría el color del polvo?
  Las respuestas son:
  1-El polvo es demasiado oscuro, 2-El polvo es muy oscuro, 3-Justo como a mí me gusta, 4-El polvo es muy claro, 5-El polvo es demasiado claro.

El estudio realizado muestra la diferencia de calificaciones entre los productos del cliente y el del competidor, es decir, que si en la columna de *overall liking* se tiene una puntuación negativa significa que el producto del competidor ganó en ese rubro para el individuo encuestado. En el caso de puntuaciones positivas se trata de puntos fuertes a favor del producto del cliente.

Es importante recordar que, uno de los supuestos que deberían cumplirse es que cada grupo de mediciones representa una muestra aleatoria extraída de una población con distribución normal multivariada. La prueba que se llevó a cabo con los datos, fue gráficamente con un Q-Q Plot; los cuantiles de la función de distribución empírica de las distancias de Mahalanobis contra los cuantiles de la función de distribución Ji cuadrada (teóricos)[12]. El resultado es el que a continuación se muestra[13]:



Con lo cual se concluye que no siguen exactamente una distribución normal multivariada, sin embargo, el análisis discriminante es una técnica robusta, es decir, aunque los datos no cumplan las condiciones, el resultado de la clasificación es próximo al mejor resultado posible.

Una vez que se tiene claro que tipo de variables están presentes, se prosigue a aplicar el análisis discriminante con el software estadístico *PASW Statistics 18*.

[12] La distancia de Mahalanobis sigue una distribución Ji cuadrada con grados de libertad igual al número de variables incluidas en el estudio.
[13] El Q-Q Plot se construyó en R y Excel.

Al importar la base y seleccionar este método en el menú de opciones, se incluyen los criterios de solicitados, los cuales se realizan a continuación.

La variable dependiente o de agrupación será la de **preferencia,** con valores de 1 a 2, ya que el 3 es indefinido, una vez que se tenga el modelo se puede definir a que grupo pertenecen estas observaciones. Por su parte, las variables independientes serán de tipo ordinal y serán las puntuaciones obtenidas, tales como *el overall liking, Apariencia, Aroma, etc*.

Asimismo se incluirán variables como el sexo, el perfil de consumo y el tipo de taza en que tomó el café, con el fin de conocer si son relevantes en la preferencia. Los resultados obtenidos se analizarán a continuación.

Para el procesamiento de datos, es importante hacer notar que hay un caso que no fue incluido ya que se refiere a una preferencia indefinida, es decir, tiene código 3. Al finalizar, se estará en posibilidad de identificar a qué grupo pertenece esta observación con ayuda de la función discriminante.

Resumen del procesamiento para el análisis de casos

Casos no ponderados		N	Porcentaje
Válidos		149	99.3
Excluidos	Códigos de grupo para perdidos o fuera de rango	1	.7
	Perdida al menos una variable discriminante	0	.0
	Perdidos o fuera de rango ambos, el código de grupo y al menos una de las variables discriminantes.	0	.0
	Total excluidos	1	.7
Casos Totales		150	100.0

El número de casos cuya preferencia es hacia el producto del cliente son 95, en cambio, para el competidor, son 54 individuos.

En la tabla de estadísticos se puede identificar algunas variables cuyas medias se diferencian del resto en los dos grupos con gran variabilidad, por tanto, son fuertes candidatas a quedarse en el análisis como variables altamente discriminativas, esto es de forma superficial, ya que habrá otros estadísticos que indicarán si se incluyen o no.

Por el momento, se podría decir que las variables que marcan una gran diferencia entre los grupos son *overall liking, apariencia del polvo, apariencia general, aroma en general, sabor en general, sabor que deja en la boca, sexo, perfil de consumo y la taza en la que bebieron el producto al probarlo.*

Los estadísticos de grupo para los que tuvieron preferencia por el producto del cliente son los siguientes:

Estadísticos de Preferencia 1

Característica	Media	Desv. típ.
Ov_ Liking	1.65	1.192
Apariencia_polvo	1.20	1.470
Color_polvo	.73	.856
Apariencia_gral	1.12	1.080
Aroma_gral	1.17	1.136
Intensidad_aroma	.63	.946
Sabor_gral	1.68	1.307
Int_sab_café	.56	1.173
Amargo	.15	1.263
Ev_sabor_suave	-.41	1.685
Suavidad	-.04	1.557
Sabor_boca	1.63	1.384
Int_sabor_boca	.47	1.184
Sexo	1.58	.496
Perfil_de_Consumo	1.52	.502
Taza	1.29	.458

Los estadísticos de grupo para los que mostraron preferencia por el producto de la competencia son los siguientes:

Estadísticos de Preferencia 2

Característica	Media	Desv. típ.
Ov_ Liking	-1.24	1.331
Apariencia_polvo	.24	1.601
Color_polvo	.48	.818
Apariencia_gral	-.20	1.337
Aroma_gral	-.89	1.562
Intensidad_aroma	-.31	1.195
Sabor_gral	-1.30	1.238
Int_sab_café	.07	1.257
Amargo	.67	1.099
Ev_sabor_suave	-.54	1.777
Suavidad	-.28	1.235
Sabor_boca	-1.50	1.255
Int_sabor_boca	.11	1.269
Sexo	1.37	.487
Perfil_de_Consumo	1.59	.496
Taza	1.24	.432

Por otro lado, observando la matriz de correlaciones, se encuentra que existe una correlación inferior a -0.3 entre las variables de *evaluación del sabor suave* y *suavidad, con intensidad del sabor a café*, así como, *amargo.*

Asimismo, las variables *intensidad y sabor que deja en la boca*, poseen una alta correlación con las variables referentes a *suavidad, intensidad del sabor a café y amargo.*

A continuación se presenta la matriz de correlaciones:

Matrices intra-grupo combinadas[a]

Correlación	Ov_ Liking	Apariencia_polvo	Color_polvo	Apariencia_gral	Aroma_gral	Intensidad_aroma	Sabor_gral	Int_sab_café	Amargo	Ev_sabor_suave	Suavidad	Sabor_boca	Int_sabor_boca	Sexo	Perfil_de_Consumo	Taza
Ov_ Liking	1.000	.193	-.115	.349	.440	.125	.784	.088	-.002	.103	.045	.659	.131	.021	-.047	.010
Apariencia_polvo	.193	1.000	.048	.328	.221	-.004	.223	.124	.254	-.188	-.076	.142	.079	-.144	-.067	-.117
Color_polvo	-.115	.048	1.000	.091	.040	.160	-.142	.118	.117	-.229	-.122	-.166	.043	.089	.032	-.119
Apariencia_gral	.349	.328	.091	1.000	.518	.136	.343	.123	.176	-.042	-.082	.222	.133	-.109	-.025	.005
Aroma_gral	.440	.221	.040	.518	1.000	.444	.393	.202	.171	-.116	-.164	.214	.170	.027	-.123	.010
Intensidad_aroma	.125	-.004	.160	.136	.444	1.000	.045	.378	.295	-.237	-.190	-.017	.271	-.032	-.090	-.110
Sabor_gral	.784	.223	-.142	.343	.393	.045	1.000	.141	-.008	.040	.078	.649	.192	.025	.000	.103
Int_sab_café	.088	.124	.118	.123	.202	.378	.141	1.000	.667	-.500	-.435	-.044	.737	.055	.150	.018
Amargo	-.002	.254	.117	.176	.171	.295	-.008	.667	1.000	-.475	-.344	-.193	.549	-.005	-.018	-.035
Ev_sabor_suave	.103	-.188	-.229	-.042	-.116	-.237	.040	-.500	-.475	1.000	.506	.237	-.402	-.005	.042	-.049
Suavidad	.045	-.076	-.122	-.082	-.164	-.190	.078	-.435	-.344	.506	1.000	.210	-.404	-.115	.046	-.044
Sabor_boca	.659	.142	-.166	.222	.214	-.017	.649	-.044	-.193	.237	.210	1.000	.036	-.090	.072	.111
Int_sabor_boca	.131	.079	.043	.133	.170	.271	.192	.737	.549	-.402	-.404	.036	1.000	.020	.036	-.059
Sexo	.021	-.144	.089	-.109	.027	-.032	.025	.055	-.005	-.005	-.115	-.090	.020	1.000	.049	.030
Perfil_de_Consum	-.047	-.067	.032	-.025	-.123	-.090	.000	.150	-.018	.042	.046	.072	.036	.049	1.000	-.035
Taza	.010	-.117	-.119	.005	.010	-.110	.103	.018	-.035	-.049	-.044	.111	-.059	.030	-.035	1.000

a. La matriz de covarianzas tiene 147 grados de libertad

Lo anterior sugiere la existencia de colinealidad entre las variables antes referidas, por lo que podrían no incluirse de forma simultánea en el análisis, ya que se trata de información que tal vez una sola podría expresar. Es importante recordar que, la existencia de colinealidad podría terminar en un problema de inestabilidad en la matriz de varianzas y covarianzas para obtener su inversa.

Ahora se procede a conocer los resultados de la prueba de igualdad de medias, ello con ayuda del estadístico lambda de Wilks, que es otra medida del potencial discriminativo de las variables. Valores pequeños de ésta indican que la variable es buena para diferenciar mejor a los grupos.

La lambda de Wilks para un conjunto de n variables mide las desviaciones dentro de cada grupo respecto a las deviaciones globales, sin diferenciar conjuntos, en el espacio n-dimensional construido a partir de sus valores.

El valor de lambda se calcula a partir de las diferencias entre los grupos y la homogeneidad dentro de los mismos. Para n variables, si se denomina B y W respectivamente a las matrices de nxn de sumas de cuadrados y productos cruzados intergrupos e intragrupos, el estadístico lambda de Wilks estará dado por el cociente:

$$\Lambda = \frac{|W|}{|B+W|}$$

Donde $|W|$ corresponde a las desviaciones intergrupos y $|B+W|$ se refiere a las desviaciones globales. Por lo tanto, si la varianza total en gran parte está conformada por la varianza entre grupos, el valor de lambda es muy pequeño. Por el contrario, si el valor tiende a 1, se dice que no hay diferencia entre las medias de las variables estudiadas.

Cabe señalar que, el valor de Λ puede transformarse en una estadística multivariante general F, que permite contrastar la existencia de diferencias significativas entre los grupos. Este valor es el mismo que se calcula en el análisis de media simple, es decir, es el cociente entre la media cuadrática dentro de grupos y la media cuadrática total.

En el siguiente cuadro se presenta las pruebas de igualdad de las medias de grupos:

Pruebas de igualdad de las medias de los grupos

Características	Lambda de Wilks	F	gl1	gl2	Sig.
Ov_ Liking	.441	186.317	1	147	.000
Apariencia_polvo	.915	13.735	1	147	.000
Color_polvo	.981	2.907	1	147	.090
Apariencia_gral	.773	43.092	1	147	.000
Aroma_gral	.632	85.467	1	147	.000
Intensidad_aroma	.838	28.375	1	147	.000
Sabor_gral	.442	185.951	1	147	.000
Int_sab_café	.964	5.559	1	147	.020
Amargo	.958	6.380	1	147	.013
Ev_sabor_suave	.999	.186	1	147	.667
Suavidad	.994	.911	1	147	.341
Sabor_boca	.438	188.287	1	147	.000
Int_sabor_boca	.980	3.066	1	147	.082
Sexo	.960	6.159	1	147	.014
Perfil_de_Consumo	.995	.812	1	147	.369
Taza	.997	.498	1	147	.481

Al parecer en todas las variables se presenta una diferencia entre los grupos, excepto para las características *evaluación del sabor suave, suavidad, perfil de consumo y taza.* Por lo tanto, no se incluirán en el análisis las características antes mencionadas, ya que tienen una significancia mayor a 0.10.

Por otro lado, el valor de la M de Box[14], evidencia si existe homogeneidad de varianzas y covarianzas. Innegablemente, en la práctica este supuesto es difícil de cumplir y corroborar; no obstante, es permisible la comprobación del mismo mediante esta prueba su estadístico está dado por la siguiente expresión:

$$M = (m-k)\ In\left|S\right| - \sum_{i=1}^{k}(m_i - 1)\ In\left|S_i\right|$$

[14] La prueba M de Box es una generalización del *test de Barlett* para la comprobación de la homogeneidad de varianzas univariadas, se basa en los determinantes de las matrices de varianzas y covarianzas para cada grupo.

Con m el total de individuos de la muestra, k el número de grupos, m_i los sujetos en el grupo i, S_i la matriz de varianzas y covarianzas para el i-ésimo grupo y S la matriz global de varianzas y covarianzas intragrupos.

Por otro lado, si se tiene el estadístico C, definido como:

$$C = \frac{2n^2 + 3n - 1}{6\,(n+1)\,(k-1)} \left(\sum_{i=1}^{k} \frac{1}{m_i - 1} - \frac{1}{m-k} \right)$$

Entonces, el estadístico $M \cdot (1 - C)$ se distribuirá aproximadamente como χ^2 con $\frac{n\,(n+1)\,(k-1)}{2}$ grados de libertad, con *n* variables discriminantes y *k* número de grupos.

Cabe considerar que, esta prueba resulta muy sensible a la violación de supuestos de normalidad multivariante, es decir, las matrices podrían reflejar diferencias significativas, a pesar de que en realidad no sea ese el caso; de tal modo que, este supuesto raramente se satisface mediante estas pruebas.

Los resultados de esta prueba se muestran en el siguiente cuadro.

Resultados de la prueba

M de Box		14.39
F	Aprox.	1.39
	gl1	10.00
	gl2	57,218.18
	Sig.	0.177

Contrasta la hipótesis nula de que las matrices de covarianzas poblacionales son iguales.

Se concluye, que de acuerdo al valor de significancia no se rechaza la hipótesis nula, lo que sugiere que las matrices de covarianzas son iguales.

Por otro lado, tomando en cuenta los valores de los logaritmos de los determinantes, se tiene otra medida de variabilidad de los grupos. Estas cantidades son pequeñas, con lo cual se puede concluir que ambas poblaciones tienen matriz de varianzas y covarianzas aproximadamente iguales, ya que si fueran más grandes ello correspondería a mayor variabilidad dentro de los grupos, o mejor dicho, indicaría que éstos tienen una matriz de varianzas y covarianzas diferente.

Prueba de Box sobre la igualdad de las matrices de covarianza

Logaritmo de los determinantes

Preferencia	Rango	Logaritmo del determinante
1	4	.314
2	4	.595
Intra-grupos combinada	4	.513

Los rangos y logaritmos naturales de los determinantes impresos son los de las matrices de covarianzas de los grupos.

En seguida, con el objetivo de introducir sólo las variables que discriminan más, se procede a seleccionarlas paso a paso.

Esencialmente, el procedimiento coincide con el algoritmo utilizado en la regresión múltiple. Se establece un método dinámico articulado en una serie de pasos de inclusión de variables de acuerdo a su capacidad discriminante. El procedimiento llega a su fin cuando no queda en el exterior del modelo alguna variable que cumpla con el criterio de entrada y, simultáneamente, todas las que están incluidas observen los criterios de permanencia.

Las variables deberán ir entrando y saliendo en función del menor valor con que este estadístico se vaya presentando. El número de pasos posibles es el doble de variables independientes contabilizadas. Finalmente, las variables introducidas serán *sabor que deja en su boca, aroma en general, sabor en general y color del polvo.*

Estadísticos por pasos

Variables introducidas / excluidas[a,b,c,d]

Paso	Variables Introducidas	Lambda de Wilks							
		Estadístico	gl1	gl2	gl3	F exacta			
						Estadístico	gl1	gl2	Sig.
1	Sabor_boca	.438	1	1	147	188.287	1	147	.000
2	Aroma_gral	.390	2	1	147	114.249	2	146	.000
3	Sabor_gral	.372	3	1	147	81.601	3	145	.000
4	Color_polvo	.358	4	1	147	64.474	4	144	.000

En cada paso se introduce la variable que minimiza la lambda de Wilks global.
a. El número máximo de pasos es 32.
b. La F parcial mínima para entrar es 3.84.
c. La F parcial máxima para salir es 2.71
d. El nivel de F, la tolerancia o el VIN son insuficientes para continuar los cálculos.

Por su parte, un modo de valorar la importancia discriminante de cada una de las funciones consiste en compararlas entre sí, de forma que se conocerá cuáles destacan en relación a las demás. Bastaría sumar todos los autovalores y dividir por esa cantidad a cada uno de ellos. Este cálculo conduciría a los porcentajes relativos, los cuales indican el porcentaje que una función posee sobre el poder discriminante total acumulado por el conjunto de funciones.

Otro modo de juzgar la importancia de las funciones se basa en el cálculo del coeficiente de correlación canónica que, al igual que el autovalor, mide las desviaciones de las puntuaciones discriminantes entre los grupos respecto a las desviaciones dentro de los grupos. El coeficiente de correlación canónica está relacionado con el autovalor mediante la expresión, referida a una función discriminante i.

$$r_i^* = \sqrt{\frac{\lambda_i}{1+\lambda_i}}$$

Este coeficiente proviene del análisis de correlaciones canónicas, éste estudia el grado de asociación entre dos variables medidas en escala de intervalo (Levine, 1977). Se desarrolla creando q pares de combinaciones lineales, siendo q el número de variables en el conjunto más pequeño. Las combinaciones en cada par se generan maximizando la correlación entre ambas. Para el primer par se tendrá el mayor grado de asociación, para el segundo, se determinan de modo que presenten el mayor grado de asociación entre sí, pero con la condición de que no esté correlacionada con las del primer par, y así sucesivamente. El coeficiente de correlación canónica es una medida idéntica al coeficiente de correlación de Pearson entre las dos combinaciones lineales de un par.

Un valor alto para este coeficiente indicaría que existe una relación entre el grupo de pertenencia y los valores de la función discriminante. Es decir, la función adopta diferentes valores de los considerados y responde satisfactoriamente al propósito de discriminar entre los grupos.

Mientras el porcentaje relativo indica cuál es la función más potente, la correlación canónica indica en que grado éste resulta relevante. En una situación en la que los grupos no sean suficientemente diferentes respecto a las variables analizadas, se podrá determinar una función discriminante que representa al mayor porcentaje relativo. Sin embargo, con ayuda del coeficiente de correlación canónica se rechazará la función en caso de que este valor correspondiente sea bajo.

El porcentaje relativo o varianza vinculada a la función discriminante es de 100%, debido a que se trata de una sola. El valor del coeficiente de correlación canónica de .801, expresa que hay una buena discriminación.

Resumen de las funciones canónicas discriminantes

Autovalores

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	1.791[a]	100.0	100.0	.801

a. Se han empleado las 1 primeras funciones discriminantes canónicas en el análisis.

Es importante considerar que, el estadístico lamba de Wilks (Λ) también constituye una medida de las diferencias entre los grupos debidas a varias funciones discriminantes. Esta cantidad expresa cual es la proporción de varianza total en las puntuaciones discriminantes que no está explicada por las diferencias entre grupos.

Cuando el valor de lambda es cercano a 1, la dispersión será debida a las diferencias dentro de los grupos, y al representarlos en el espacio discriminante estarán poco separados. Por el contrario, valores pequeños de lambda indican un mayor grado de discriminación por parte de la función.

A esta medida se le asocia un contraste Ji cuadrada, donde la hipótesis nula hace referencia a la no existencia de diferencia en las puntuaciones otorgadas a las diferentes variables, independientes entre las dos categorías de referencia, en este caso es la preferencia por el producto del cliente o la competencia. Debido a que la significancia es de 0.000 se rechaza la hipótesis nula. Por lo tanto, se exalta la existencia de diferencias entre los grupos.

Lambda de Wilks

Contraste de las funciones	Lambda de Wilks	Ji-cuadrado	gl	Sig.
1	.358	148.825	4	.000

La matriz de estructura expresa la correlación entre los valores de la función discriminante y las variables independientes.

Matriz de estructura

Características	Función
	1
Sabor_boca	.846
Sabor_gral	.840
Ov_ Liking[a]	.780
Aroma_gral	.570
Apariencia_gral[a]	.431
Apariencia_polvo[a]	.243
Intensidad_aroma[a]	.173
Int_sabor_boca[a]	.155
Int_sab_café[a]	.116
Color_polvo	.105
Taza[a]	.078
Suavidad[a]	.074
Ev_sabor_suave[a]	.062
Amargo[a]	-.036
Perfil_de_Consumo[a]	.014
Sexo[a]	-.012

Correlaciones intra-grupo combinadas entre las variables discriminantes y las funciones discriminantes canónicas tipificadas
Variables ordenadas por el tamaño de la correlación con la función.
a. Esta variable no se emplea en el análisis.

Es una forma de expresar en qué medida cada variable contribuye a la discriminación de los grupos. En este ejemplo, la variable que tiene una relación lineal mayor es el *sabor que deja en la boca.*

La matriz de confusión o mejor conocida como los resultados de la clasificación, contiene el número de casos que han sido clasificados de forma correcta e incorrecta sobre el total de la muestra. En el caso, se tiene que 93 individuos fueron asignados de forma correcta los que tienen preferencia por el producto del cliente y sólo dos han sido asignados de forma incorrecta.

Por su parte, los casos asignados a la preferencia por el producto de la competencia, se tiene que el 88.9% de los casos se clasificó correctamente y sólo el 11.1% no fue correcto.

Resultados de la clasificación[a]

		Preferencia	Grupo de pertenencia pronosticado		Total
			1	2	
Original	Recuento	1	93	2	95
		2	6	48	54
		Casos	0	1	1
	%	1	97.9	2.1	100.0
		2	11.1	88.9	100.0
		Casos	.0	100.0	100.0

a. Clasificados correctamente el 94.6% de los casos agrupados originales.

El procesamiento de los datos se llevó con éxito en todos los casos, esto se resume en el cuadro del proceso de clasificación.

Estadísticos de clasificación

Resumen del proceso de clasificación

Procesados		150
Excluidos	Código de grupo perdido o fuera de rango	0
	Perdida al menos una variable discriminante	0
Usados en los resultados		150

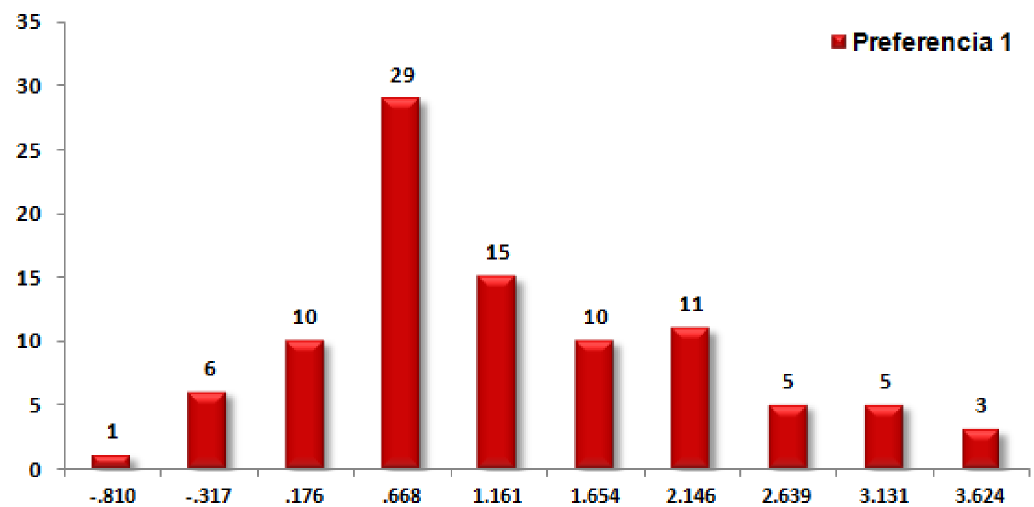
En los resultados obtenidos, también se puede identificar a las puntuaciones de los centroides de las funciones, para el producto del cliente se tiene 1.002 y para la competencia es de -1.763.

Funciones en los centroides de los grupos

Preference	Función
	1
1	1.002
2	-1.763

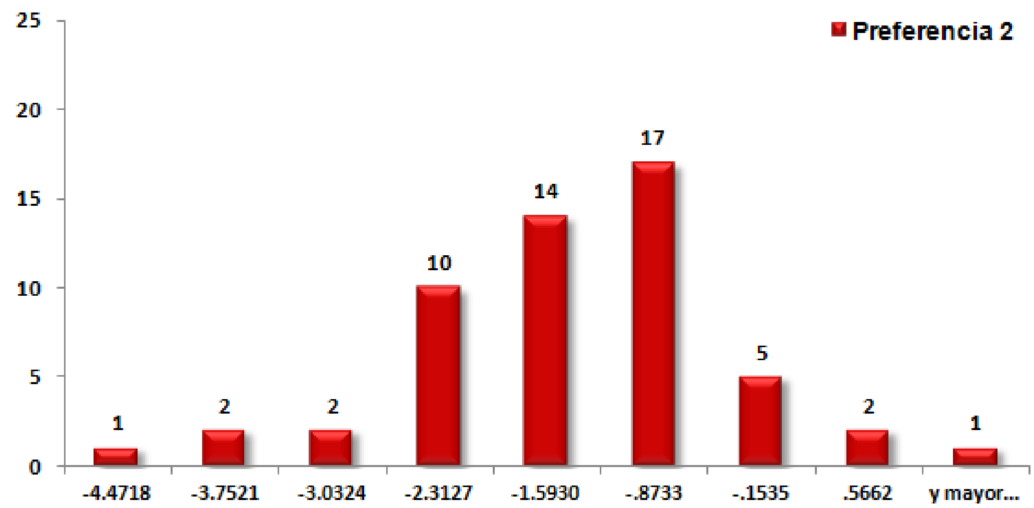
Funciones discriminantes canónicas no tipificadas evaluadas en las medias de los grupos

Si se grafican las puntuaciones discriminantes obtenidas con las funciones canónicas, separadas por grupos, se tiene lo siguiente:



Estos histogramas permiten formarse una idea aproximada tanto de la representación de la distribución, como del grado de dispersión de los individuos dentro de su propio grupo.

Se observa que la dispersión dentro de los grupos no es tan grande, lo cual es bueno para poder diferenciarlos.



Finalmente, la función discriminante efectiva, cuyos coeficientes son los siguientes, incluye a aquéllas variables que resultaron más significativas o que discriminan mejor, las cuales son de mayor importancia al establecer una preferencia por alguno de los productos.

Coeficientes de las funciones canónicas discriminantes

Características	Función 1
Color_polvo	.290
Aroma_gral	.217
Sabor_gral	.306
Sabor_boca	.426
(Constante)	-.673

Coeficientes no tipificados

.Por lo tanto, la combinación lineal de las variables independientes que permitirá el cálculo de la puntuación discriminante para cualquier individuo será la siguiente:

Z_1**=** -0.673 + 0.290 **Color del Polvo** + 0.217 **Aroma en general** + 0.306 **Sabor en General** + 0.426 **Sabor que deja en la boca**

Con ayuda de los *coeficientes de las funciones canónicas discriminantes,* se obtiene la puntuación para cada observación, así como los centroides de los grupos. Posteriormente, se toma la diferencia al cuadrado y se ve cuál es la mínima cantidad; por ejemplo, si la distancia calculada entre las puntuaciones discriminantes de un individuo al del centroide del grupo uno es de 0.427 contra 4.461 para el segundo, se deduce que, ese individuo se asigna al primer grupo.

Derivados de los coeficientes anteriores, es posible obtener los factores estandarizados con ayuda de las puntuaciones discriminantes. Éstos tienen la misma interpretación que los presentados en un análisis de regresión lineal.

A diferencia de estos últimos, los obtenidos por este método, evitan el efecto de escala y son medidas, de la dirección de la relación positiva o negativa, así como de su intensidad, que sería el valor absoluto del coeficiente. Por ejemplo, se puede decir que la característica más influyente en la preferencia por algún producto, está relacionada en un grado fuerte con la calificación observada en *sabor que deja en la boca.*

Coeficientes estandarizados de las funciones discriminantes canónicas

Características	Función
	1
Color_polvo	.244
Aroma_gral	.284
Sabor_gral	.393
Sabor_boca	.571

Por último, se tiene la función lineal de Fisher, con ayuda de ésta también se puede clasificar a un individuo, es decir, se asigna a la población en la que sea máxima su puntuación para la función dada.

Coeficientes de la función de clasificación

Característica	Preferencia	
	1	2
Color_polvo	1.299	.497
Aroma_gral	.278	-.322
Sabor_gral	.629	-.218
Sabor_boca	.597	-.581
(Constante)	-2.101	-1.856

Funciones discriminantes lineales de Fisher

Por ejemplo si un individuo muestra las siguientes puntuaciones:

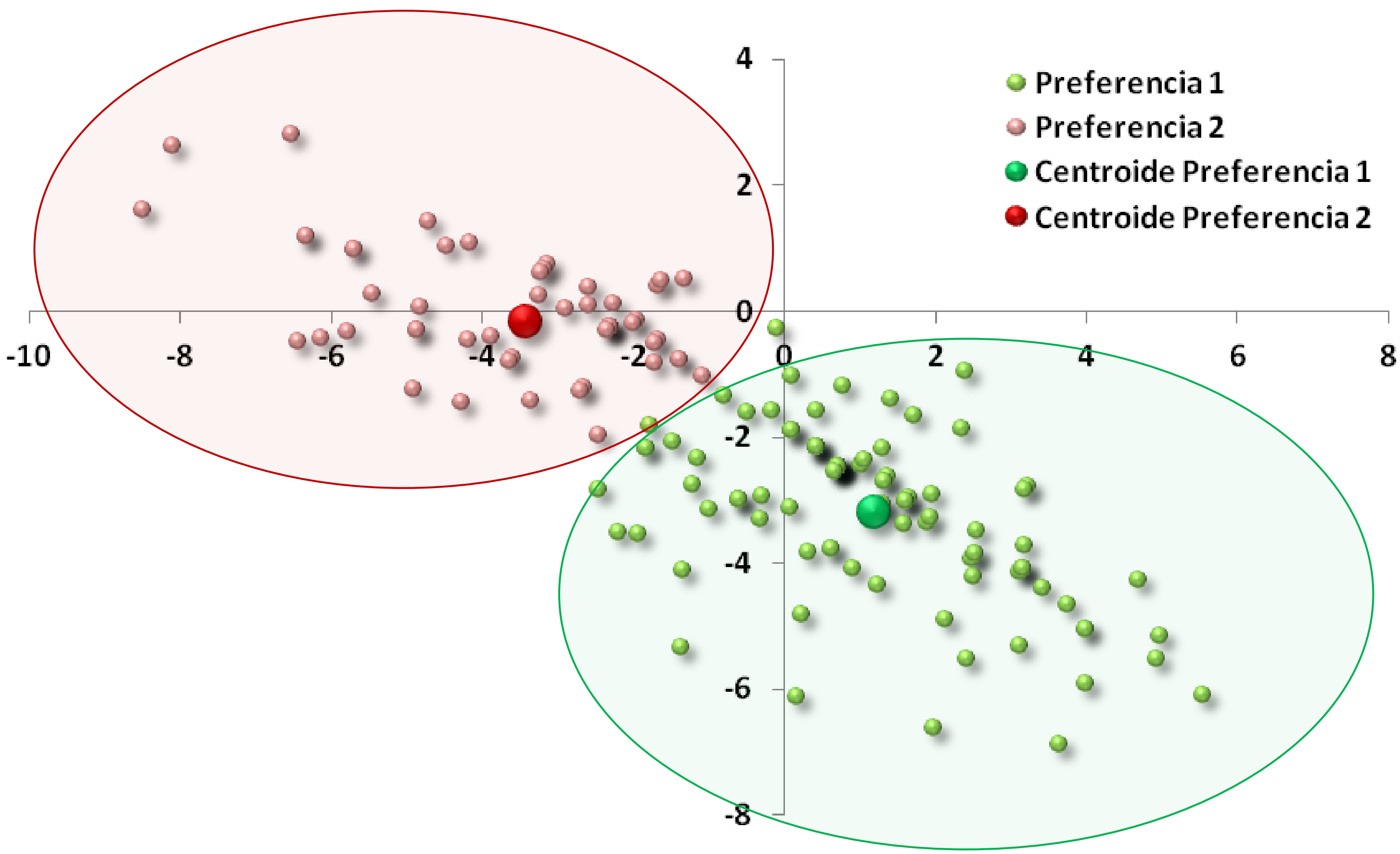
- Color del polvo: 1
- Aroma en general: 0
- Sabor en general: 1
- Sabor que deja en la boca: 1

La evaluación en la función de clasificación para la preferencia 1 es de 0.424, en cambio para la segunda es de -2.158, lo cual indica que su preferencia se declina por el producto del cliente, es decir, el primero.

De forma similar se lleva a cabo el cálculo de puntuaciones para todas las observaciones. Con ayuda de estos se obtiene un *diagrama de dispersión para las funciones discriminantes*, en el eje de las abscisas la primera función y la segunda en el eje de las ordenadas.

En caso de que no sea posible distinguir a que grupo pertenece alguna observación, se podrá recurrir al cálculo de distancias de Mahalanobis hacia el centroide de cada uno de los grupos, por lo tanto, se asignará la observación dada a aquélla población donde la distancia sea mínima.

Diagrama de Dispersión para las puntuaciones en las funciones discriminantes de Fisher



Función 1

Función 2

3.2. Prueba de producto para cosméticos

A continuación se presentará este método aplicado a una base de la empresa Z, obtenida de un estudio en el cual se encuestó a usuarias de tres diferentes marcas de cosméticos. Las preguntas están enfocadas a la evaluación de campañas que presentaron estas marcas, en las cuales se envía un mensaje en específico sobre la misma.

Se desea conocer cuáles son las características percibidas en las campañas por una usuaria, de modo que los pondera con mayor importancia para ser fiel a esa marca

La base consta de 435 entrevistas para la Ciudad de México, cuyas edades de las entrevistadas oscilan entre 18 y 40 años.

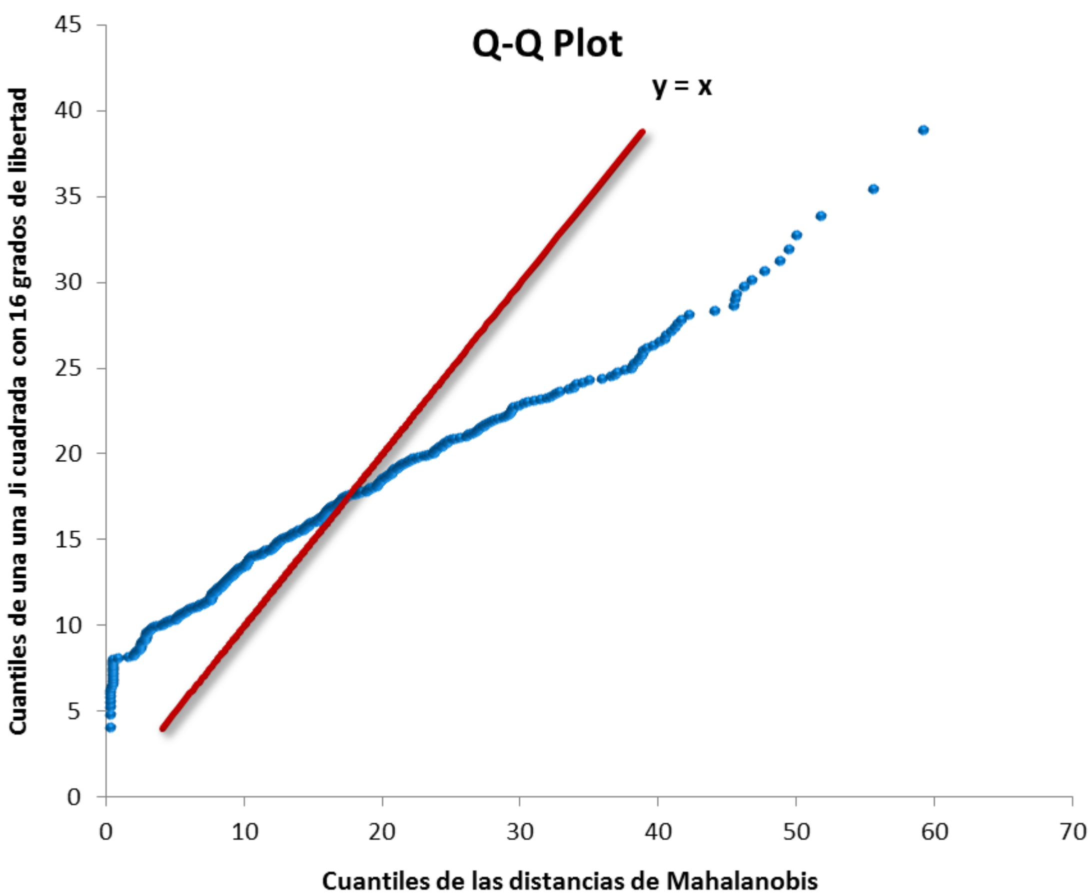
Por su parte, todas las mediciones están en una escala de Likert, que indican el grado de acuerdo o desacuerdo con cada atributo, es decir, *1=Totalmente en desacuerdo* y *5=Totalmente de acuerdo*.

Los atributos evaluados son los siguientes:

- ✓ Confianza
- ✓ Calidad
- ✓ Eficacia
- ✓ Vanguardia
- ✓ Innovación
- ✓ Prestigio
- ✓ Elegante
- ✓ Natural
- ✓ Cercana
- ✓ Sinceridad
- ✓ Buen servicio
- ✓ Sexy

- ✓ Atrevida
- ✓ Urbana
- ✓ Precio justo
- ✓ Moderna

Es importante destacar que uno de los supuestos que serán de mayor importancia para el AD, es que los datos tengan un comportamiento normal multivariado. Ello se podrá saber si realizamos una prueba gráfica, es decir, construir un Q-Q Plot de las distancias de Mahalanobis contra los cuantiles de la Ji cuadrada correspondientes. El gráfico se presenta a continuación:



No es difícil observar que las distancias no siguen exactamente una distribución Ji cuadrada con 16 grados de libertad; con lo cual se dice que la muestra no proviene de una población normal multivariada; no obstante, habrá que recordar que el análisis discriminante es una técnica robusta.

Una vez que se tiene claro que tipo de variables están presentes, se prosigue a aplicar el análisis discriminante con el software estadístico *PASW Statistics 18*.

La variable que se utilizará como dependiente será *usuaria*, cuyos valores son de 1 a 3, donde el primero pertenece a la marca del cliente y los otros dos son de los competidores de mercado evaluados. Las variables independientes, serán las evaluaciones de cada atributo.

Al ejecutar el proceso con estas especificaciones, los primeros resultados son los relacionados al procesamiento de los datos, que de acuerdo al cuadro arrojado, no hubo valores perdidos al procesar, es decir, el número de casos válidos es el mismo que el de tamaño de la base.

Resumen del procesamiento para el análisis de casos

Casos no ponderados		N	Porcentaje
Válidos		435	100.0
Excluidos	Códigos de grupo para perdidos o fuera de rango	0	.0
	Perdida al menos una variable discriminante	0	.0
	Perdidos o fuera de rango ambos, el código de grupo y al menos una de las variables	0	.0
	Total excluidos	0	.0
Casos Totales		435	100.0

Se obtiene los cuadros de estadísticos de grupo, cada uno contiene 145 usuarias.

Estadísticos de Usuaria 1

Características	Media	Desv. típ.
Confianza	4.12	.949
Calidad	3.99	.917
Eficacia	3.94	.919
Vanguardia	4.01	.972
Innovación	4.02	.982
Prestigio	3.99	.954
Elegante	4.08	.958
Natural	4.07	.962
Cercana	3.88	1.079
Sinceridad	3.78	.975
Buen_servicio	3.95	.967
Sexy	3.92	1.055
Atrevida	3.96	.971
Urbana	3.99	.968
Precio_Justo	3.65	1.134
Moderna	3.97	1.013

Estadísticos de Usuaria 2

Características	Media	Desv. típ.
Confianza	4.26	.888
Calidad	4.21	.865
Eficacia	4.17	.890
Vanguardia	4.17	.819
Innovación	4.21	.772
Prestigio	4.22	.820
Elegante	4.18	.839
Natural	4.19	.833
Cercana	4.11	.875
Sinceridad	4.07	.805
Buen_servicio	4.15	.836
Sexy	3.96	.985
Atrevida	3.98	.989
Urbana	4.06	.888
Precio_Justo	3.88	1.024
Moderna	4.13	.868

Estadísticos de Usuaria 3

Características	Media	Desv. típ.
Confianza	2.59	1.010
Calidad	2.61	.994
Eficacia	2.72	1.091
Vanguardia	2.68	1.104
Innovación	2.69	.997
Prestigio	2.66	1.083
Elegante	2.70	1.137
Natural	2.73	1.168
Cercana	2.72	1.141
Sinceridad	2.60	.996
Buen_servicio	2.72	1.083
Sexy	2.68	1.116
Atrevida	2.76	1.101
Urbana	2.89	1.197
Precio_Justo	2.72	1.077
Moderna	2.88	1.115

Es importante recordar, que esta información puede ofrecer un punto de partida para conocer cuales son las variables que podrían ser de mayor importancia para discriminar entre grupos, sin embargo, para este caso las características evaluadas no muestran diferencias tan evidentes. Aunque es posible que el atributo de *confianza* sea de utilidad debido a que es la que presenta mayor variabilidad entre los tres grupos.

En seguida, se obtiene la matriz de correlaciones, de este modo se conocerá si existe alguna variable correlacionada con otra, ello, con el fin de evitar que la matriz sea singular, ya que alguna variable pueda ser expresada como combinación lineal de otra.

Como se observa en dicha matriz, existen muchas correlaciones mayores a 0.5, por ejemplo, entre *confianza* y *calidad, eficacia, vanguardia, innovación, prestigio, elegante, natural, cercana* y *buen servicio.*

Un comportamiento similar tiene el resto de las variables, por lo que se sugiere hacer una revisión más exhaustiva de cada variable, con el fin de evitar redundancia en la información dada.

En este cuadro se muestra la matriz de correlaciones:

Matrices intra-grupo combinadas[a]

Correlación	Confianza	Calidad	Eficacia	Vanguardia	Innovación	Prestigio	Elegante	Natural	Cercana	Sinceridad	Buen_servicio	Sexy	Atrevida	Urbana	Precio_Justo	Moderna
Confianza	1.000	.719	.628	.593	.582	.578	.565	.540	.550	.453	.534	.470	.477	.493	.412	.497
Calidad	.719	1.000	.700	.645	.618	.594	.593	.584	.584	.506	.570	.456	.484	.509	.467	.489
Eficacia	.628	.700	1.000	.655	.565	.609	.581	.545	.549	.493	.542	.433	.505	.502	.499	.529
Vanguardia	.593	.645	.655	1.000	.667	.613	.608	.604	.563	.552	.538	.447	.581	.503	.419	.532
Innovación	.582	.618	.565	.667	1.000	.646	.611	.614	.535	.490	.587	.465	.471	.458	.435	.450
Prestigio	.578	.594	.609	.613	.646	1.000	.680	.626	.506	.508	.595	.512	.500	.491	.404	.483
Elegante	.565	.593	.581	.608	.611	.680	1.000	.701	.562	.509	.592	.527	.470	.485	.407	.533
Natural	.540	.584	.545	.604	.614	.626	.701	1.000	.653	.584	.630	.490	.573	.539	.443	.518
Cercana	.550	.584	.549	.563	.535	.506	.562	.653	1.000	.592	.547	.464	.510	.481	.484	.470
Sinceridad	.453	.506	.493	.552	.490	.508	.509	.584	.592	1.000	.484	.437	.402	.417	.459	.447
Buen_servicio	.534	.570	.542	.538	.587	.595	.592	.630	.547	.484	1.000	.429	.502	.518	.439	.515
Sexy	.470	.456	.433	.447	.465	.512	.527	.490	.464	.437	.429	1.000	.568	.469	.439	.471
Atrevida	.477	.484	.505	.581	.471	.500	.470	.573	.510	.402	.502	.568	1.000	.601	.464	.517
Urbana	.493	.509	.502	.503	.458	.491	.485	.539	.481	.417	.518	.469	.601	1.000	.391	.483
Precio_Justo	.412	.467	.499	.419	.435	.404	.407	.443	.484	.459	.439	.439	.464	.391	1.000	.592
Moderna	.497	.489	.529	.532	.450	.483	.533	.518	.470	.447	.515	.471	.517	.483	.592	1.000

a. La matriz de covarianzas tiene 432 grados de libertad

En seguida se realiza la prueba de igualdad de medias, para conocer el potencial discriminativo entre las variables mediante el estadístico lambda de Wilks.

Como ya se había mostrado en el ejemplo anterior, valores pequeños de ésta indican si la variable es buena o no para discriminar entre los grupos.

Pruebas de igualdad de las medias de los grupos

Características	Lambda de Wilks	F	gl1	gl2	Sig.
Confianza	.612	137.046	2	432	.000
Calidad	.631	126.188	2	432	.000
Eficacia	.697	93.710	2	432	.000
Vanguardia	.678	102.742	2	432	.000
Innovación	.649	116.626	2	432	.000
Prestigio	.657	112.974	2	432	.000
Elegante	.681	101.288	2	432	.000
Natural	.694	95.252	2	432	.000
Cercana	.743	74.881	2	432	.000
Sinceridad	.680	101.659	2	432	.000
Buen_servicio	.700	92.575	2	432	.000
Sexy	.758	69.045	2	432	.000
Atrevida	.761	67.801	2	432	.000
Urbana	.785	59.223	2	432	.000
Precio_Justo	.822	46.726	2	432	.000
Moderna	.765	66.435	2	432	.000

Se observa que los valores más pequeños, respecto al resto, son para *confianza, calidad, eficacia, vanguardia, innovación, prestigio, elegante, natural* y *sinceridad.* Con lo cual se reduce el número de variables que podrán estar en el análisis.

Posteriormente, se realiza la prueba M de Box, con lo que se pretende contrastar en qué medida las matrices de varianzas y covarianzas para cada uno de los grupos proceden o no de la misma población.

Resultados de la prueba

M de Box		47.31
F	Aprox.	3.90
	gl1	12.00
	gl2	904,408.62
	Sig.	0.000

Contrasta la hipótesis nula de que las matrices de covarianzas poblacionales son iguales.

Se tiene que el nivel de significancia es de 0.000, con ello se rechaza la hipótesis nula y se puede afirmar la existencia de diferencias entre grupos.

Asimismo, tomando en cuenta los valores de los logaritmos de los determinantes, se tiene otra medida de variabilidad de los grupos; cantidades mínimas indican que las matrices de varianzas y covarianzas para estos conjuntos son aproximadamente iguales, lo cual no sucede en esta ocasión, de acuerdo a lo que se ve en el cuadro siguiente:

Prueba de Box sobre la igualdad de las matrices de covarianza

Logaritmo de los determinantes

Usuaria	Rango	Logaritmo del determinante
1	3	-1.148
2	3	-1.379
3	3	-1.269
Intra-grupos	3	-1.156

Los rangos y logaritmos naturales de los determinantes impresos son los de las matrices de covarianzas de los grupos.

Llevando a cabo la inclusión de variables en el análisis paso a paso, haciendo llamado a la parsimonia en el modelo, solo se tomarán las características de *confianza, sinceridad e innovación.* Los estadísticos por pasos son los siguientes:

Estadísticos por pasos

Variables introducidas / excluidas[a,b,c,d]

Paso	Variables Introducidas	Lambda de Wilks							
		Estadístico	gl1	gl2	gl3	F exacta			
						Estadístico	gl1	gl2	Sig.
1	Confianza	.612	1	2	432.000	137.046	2	432.000	.000
2	Sinceridad	.563	2	2	432.000	71.758	4	862.000	.000
3	Innovación	.546	3	2	432.000	50.616	6	860.000	.000

En cada paso se introduce la variable que minimiza la lambda de Wilks global.
a. El número máximo de pasos es 32.
b. La F parcial mínima para entrar es 3.84.
c. La F parcial máxima para salir es 2.71
d. El nivel de F, la tolerancia o el VIN son insuficientes para continuar los cálculos.

Por su parte, se procede a calcular el coeficiente de correlación canónica. Un valor alto indicará que existe una relación entre el grupo de pertenencia y los valores de la función discriminante, con ello responde al propósito de discriminar entre los grupos.

Resumen de las funciones canónicas discriminantes

Autovalores

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	.819[a]	99.2	99.2	.671
2	.007[a]	.8	100.0	.081

a. Se han empleado las 2 primeras funciones discriminantes canónicas en el análisis.

Se observa que el porcentaje de varianza vinculada a la primera función discriminante es de 99.2% y para la segunda es de solo 0.8%. Los resultados apuntan a que la primera representa una dimensión mucho más relevante de cara a la discriminación entre los tres tipos de usuarias de cosméticos. Por otro lado, se tiene que los valores de lambda de Wilks son:

Lambda de Wilks

Contraste de las funciones	Lambda de Wilks	Ji-cuadrado	gl	Sig.
1 a la 2	.546	260.689	6	.000
2	.993	2.838	2	.242

Una vez extraída la primera función, la discriminación residual en el sistema es escasa, ya que el valor de la significancia para la segunda función, no permite rechazar la hipótesis nula de igualdad de medias entre los grupos para un nivel de significancia de 0.05. Ello no sucede en el primer caso, porque la significancia es de 0.000, con lo cual se sugiere una diferencia entre grupos.

Por lo tanto, se tendrá que considerar significativa solo la primera de las dos funciones discriminantes obtenidas en el análisis.

Es importante recordar que, la matriz de estructura proporciona la correlación entre los valores de la función discriminante y las variables independientes, es una forma de expresar en qué medida cada cualidad contribuye a la función discriminante.

Se puede observar de los resultados que las variables más correlacionadas con la primera función son *confianza* e *innovación*, en cambio, *sinceridad* aporta en mayor cantidad a la segunda función.

Matriz de estructura

Características	Función	
	1	2
Confianza	.879*	-.417
Innovación	.812*	-.007
Calidad[a]	.763*	-.114
Sinceridad	.756*	.622
Vanguardia[a]	.729*	.044
Prestigio[a]	.699*	.012
Natural[a]	.695*	.123
Eficacia[a]	.692*	-.047
Elegante[a]	.680*	.025
Cercana[a]	.677*	.121
Buen_servicio[a]	.647*	.026
Moderna[a]	.569*	.020
Urbana[a]	.559*	-.006
Sexy[a]	.556*	.034
Atrevida[a]	.550*	-.008
Precio_Justo[a]	.524*	.108

Correlaciones intra-grupo combinadas entre las variables discriminantes y las funciones discriminantes canónicas tipificadas. Variables ordenadas por el tamaño de la correlación con la función.

*. Mayor correlación absoluta entre cada variable y cualquier función discriminante.

a. Esta variable no se emplea en el análisis.

Si se visualiza la matriz de confusión, se podrá percibir el porcentaje total de observaciones correctamente clasificadas, que es de 59.3%.

Es importante resaltar que el mayor número de casos mal clasificados proviene del grupo uno y dos, que tienen un porcentaje de clasificación correcta de 29% y 66.9% respectivamente.

Resultados de la clasificación[a]

		Usuaria	Grupo de pertenencia pronosticado			Total
			1	2	3	
Original	Recuento	1	42	78	25	145
		2	37	97	11	145
		3	15	11	119	145
	%	1	29.0	53.8	17.2	100.0
		2	25.5	66.9	7.6	100.0
		3	10.3	7.6	82.1	100.0

a. Clasificados correctamente el 59.3% de los casos agrupados originales.

El procesamiento de datos, se llevó de forma satisfactoria, ya que se procesaron los datos en su totalidad, con cero casos excluidos.

Estadísticos de clasificación

Resumen del proceso de clasificación

Procesados		435
Excluidos	Código de grupo perdido o fuera de rango	0
	Perdida al menos una variable discriminante	0
Usados en los resultados		435

Los valores de las funciones evaluadas en las medias de los grupos son los siguientes:

Funciones en los centroides de los grupos

Usuaria	Función	
	1	2
1	.509	-.105
2	.759	.092
3	-1.267	.013

Funciones discriminantes canónicas no tipificadas evaluadas en las medias de los grupos

Por otro lado, los coeficientes de las funciones canónicas discriminantes son los siguientes:

Coeficientes de las funciones canónicas discriminantes

Características	Función	
	1	2
Confianza	.552	-.928
Innovación	.360	.009
Sinceridad	.383	1.095
(Constante)	-4.662	-.451

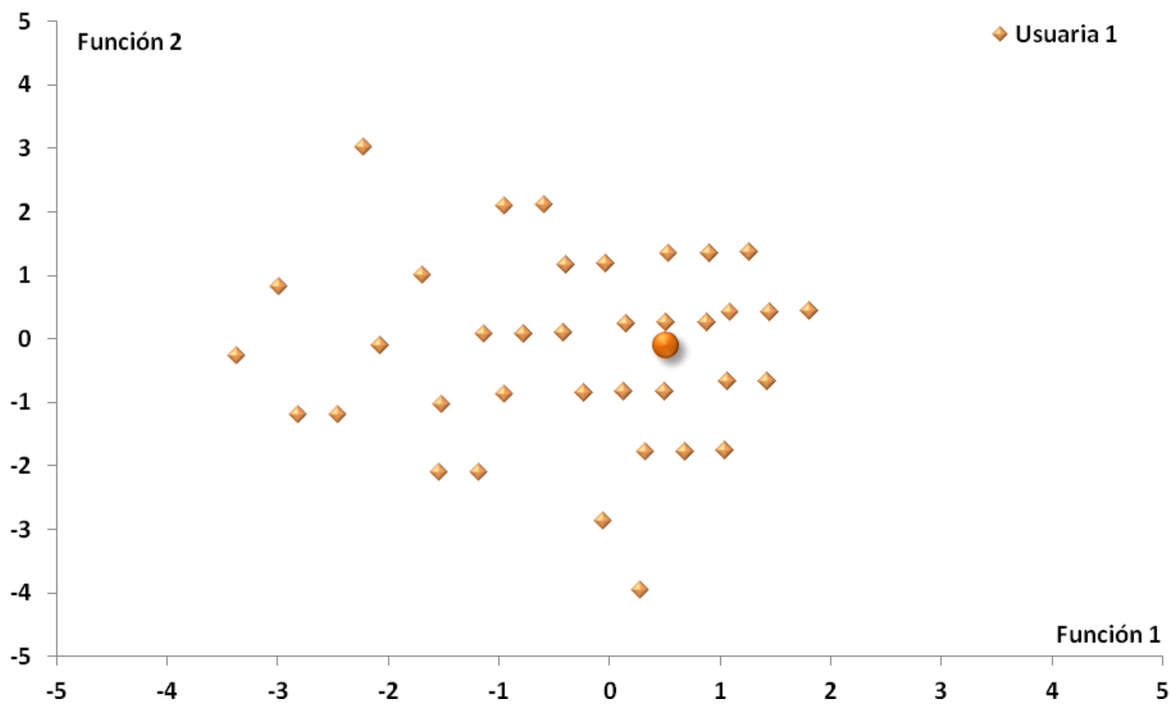
Coeficientes no tipificados

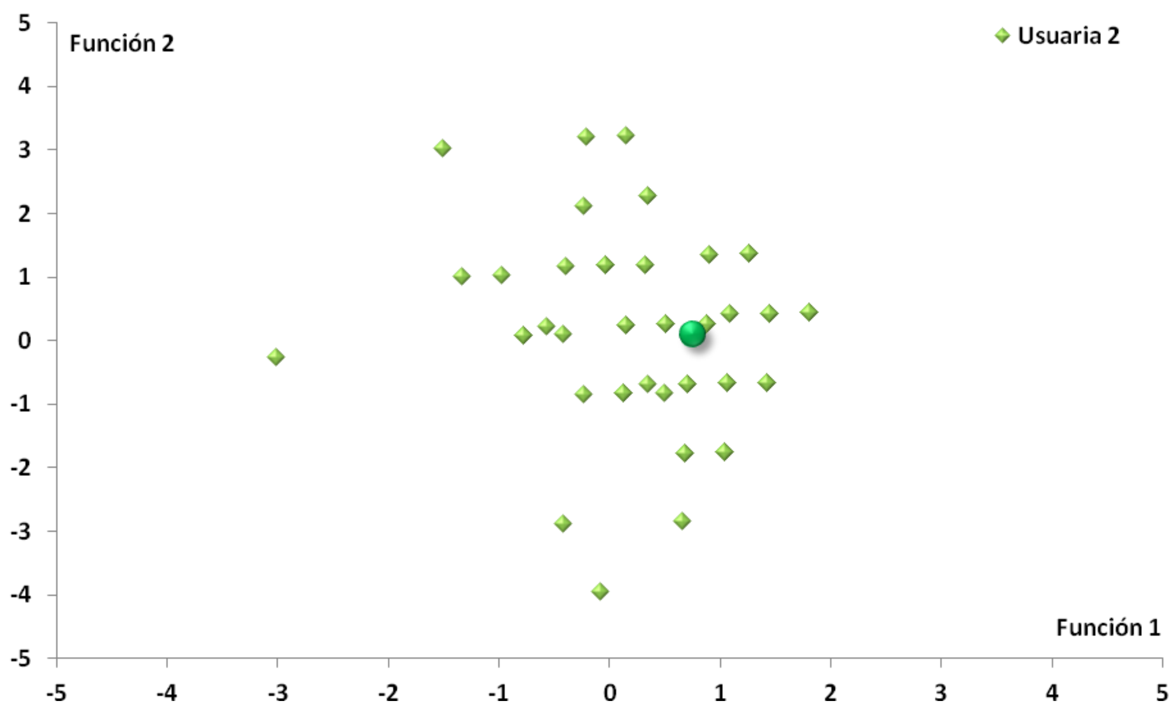
Y los coeficientes estandarizados, que como ya se había señalado, son similares en su interpretación a los que se obtienen por análisis de regresión, se observa que *confianza* aporta con mayor intensidad a la función uno; mientras que, *sinceridad* contribuye más a la segunda. Los valores dados se presentan a continuación:

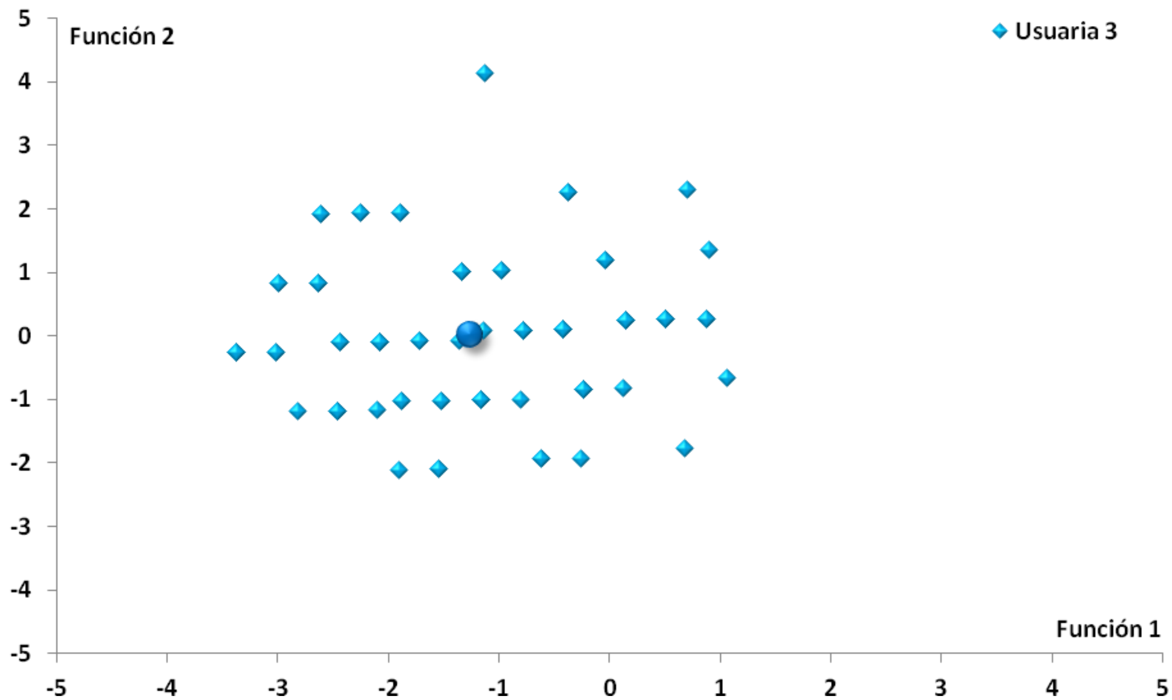
Coeficientes estandarizados de las funciones discriminantes canónicas

Características	Función	
	1	2
Confianza	.525	-.882
Innovación	.332	.009
Sinceridad	.356	1.017

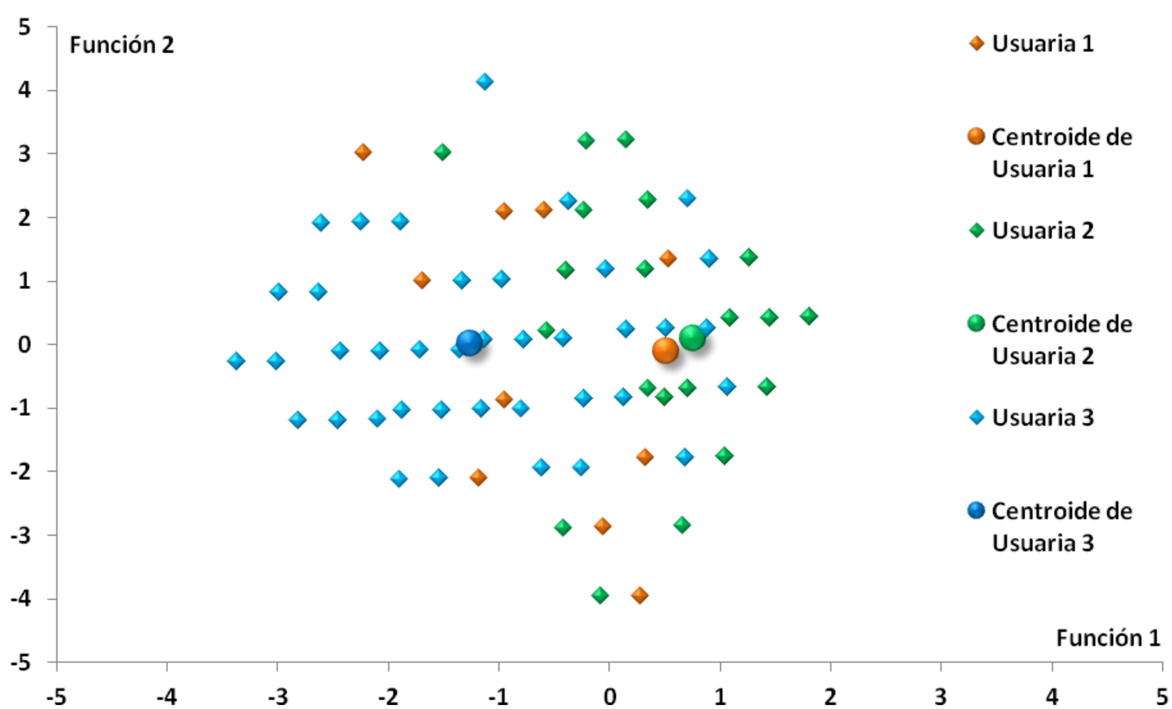
Las gráficas de las puntuaciones discriminantes por grupos separados, se muestran a continuación:



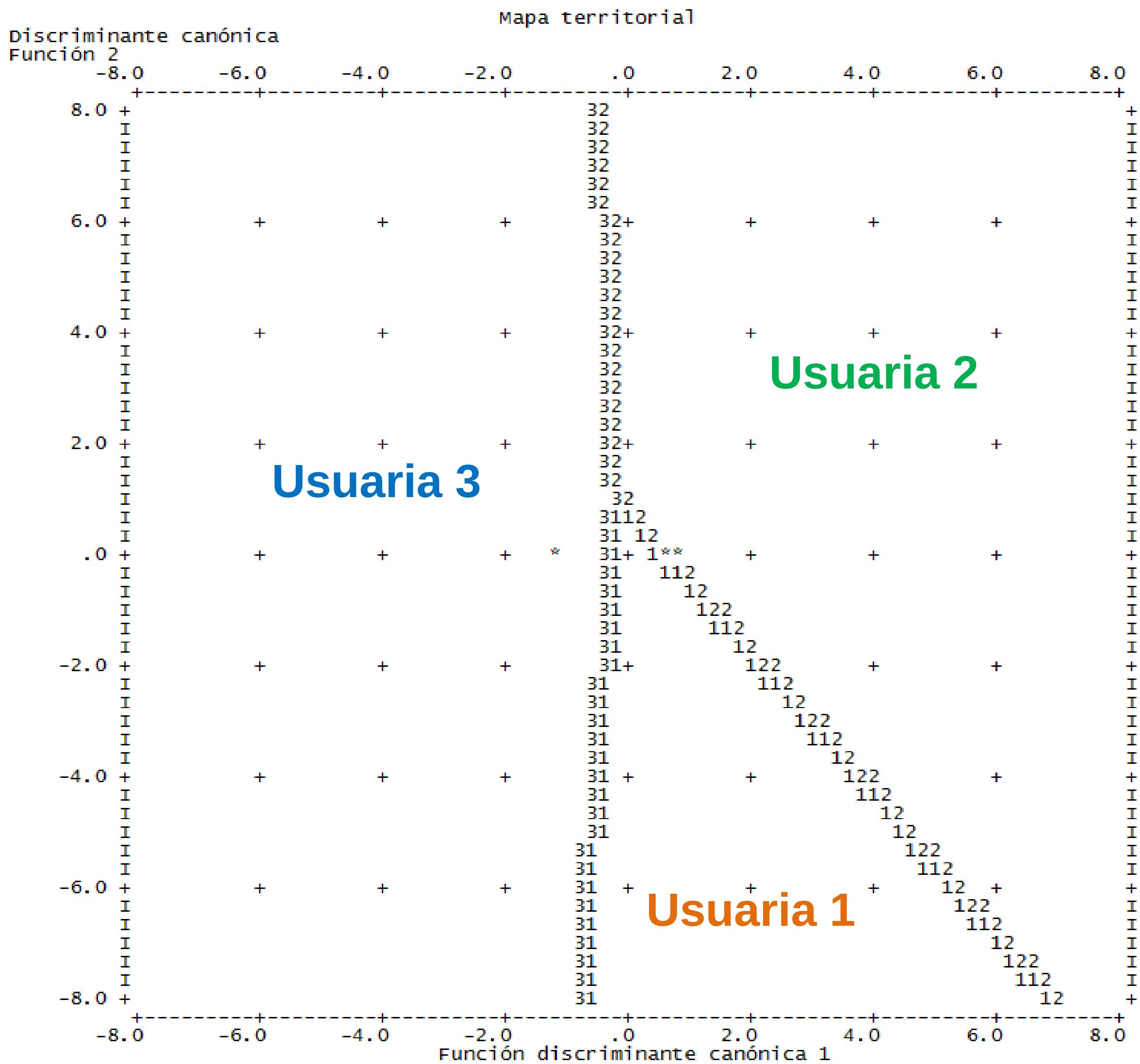




A continuación se muestra el diagrama de dispersión de todos los casos utilizados en el análisis, definido por las puntuaciones discriminantes. La mayor utilidad de este gráfico radica en la posibilidad de identificar casos atípicos difíciles de clasificar, lo cual no se percibe en este caso.



Por su parte, un mapa territorial representa el *territorio* que corresponde a cada uno de los grupos en el plano definido por las dos funciones discriminantes, la primera función en el eje de las abscisas y la segunda en el eje de las ordenadas. Las secuencias de los números que aparecen dividiendo al plano corresponden a las fronteras dadas por la regla de clasificación.



Una forma gráfica de conocer en donde se podría clasificar a un individuo, sería ubicándolo en dicho mapa, con ello se situaría en el grupo al que pertenece.

Si la clasificación se quiere llevar a cabo solo con las puntuaciones discriminantes, las combinaciones lineales de las variables independientes que permitirá el cálculo de éstas, para cualquier individuo serán las siguientes:

$$Z_1 = -4.662 + 0.552\ \textbf{Confianza} + 0.360\ \textbf{Innovación} + 0.383\ \textbf{Sinceridad}$$

$$Z_2 = -0.451 - 0.928\ \textbf{Confianza} + 0.009\ \textbf{Innovación} + 1.095\ \textbf{Sinceridad}$$

Con éstas se calcularía la distancia entre las puntuaciones discriminantes de una observación a clasificar y las evaluaciones de las funciones canónicas en los centroides; de esta forma, el individuo se asigna al grupo cuya distancia al centroide resulte mínima.

Finalmente, se obtienen los coeficientes de las funciones lineales de Fisher:

Coeficientes de la función de clasificación

Características	Usuaria		
	1	2	3
Confianza	2.319	2.274	1.229
Innovación	2.239	2.331	1.602
Sinceridad	2.212	2.524	1.662
(Constante)	-14.563	-15.975	-7.006

Funciones discriminantes lineales de Fisher

En caso de utilizar el criterio de discriminar a los grupos con ayuda de la funciones lineales de Fisher, lo que se debe conseguir es la evaluación del individuo en cada ecuación obtenida, entonces, se asigna éste a aquélla población que obtenga mayor puntuación.

Conclusiones.

La importancia de tener una segmentación de mercados, se ha vuelto hoy en día una necesidad en los campos de alta competencia para la venta de productos; ya que permite a la persona que posea esta información, tomar decisiones importantes que le ayuden a adelantarse a los hechos. De este modo, se reúne más información del mercado objetivo; lo cual deriva en algunos casos, en la planeación de campañas más agresivas para llegar a ellos, las cuales se enfocan en las características que los individuos tienden a tomar como de mayor relevancia antes de realizar una elección.

El análisis multivariado brinda un abanico de oportunidades para el estudio de datos, entre los que destacan el análisis discriminante para encontrar una fórmula de segmentación. Lo anterior, tiene como objetivo, el establecer una regla para clasificar o segmentar a futuras observaciones en los estudios similares que se lleven a cabo, ello, en base a lo analizado en un grupo de entrenamiento, que podría obtenerse en una prueba piloto o bien de estudios anteriores.

De acuerdo a los resultados en este trabajo, se concluye que es posible encontrar una función discriminante de Fisher o de coeficientes canónicos, para clasificar los datos. Se considera exitosa la obtención de esta regla, ya que el porcentaje de observaciones bien clasificadas rebasa la mitad de casos. En el estudio de café se obtuvo una tasa de clasificación correcta del 94.6%; por su parte, para el caso de cosméticos resultó más baja, con 59.3%. Asimismo, se obtuvieron las características que realmente se captan como importantes para decidir la preferencia en bebidas de café y cosméticos, así como su peso en específico sobre su fallo final.

Finalmente, es importante decir que para aquellas tasas de clasificación que resulten bajas en efectividad, podría considerarse utilizar otros métodos de segmentación como el de distancias, ya que para su aplicación basta la obtención de una métrica y no es necesario suponer alguna distribución de los datos.

Bibliografía

- AMÓN Uribe, Iván.
  "Guía metodológica para la selección de técnicas de depuración de datos"
  Universidad Nacional de Colombia
  Medellín, Colombia
  2010
  120 págs.

- CUADRAS, Carles M.
  "Nuevos Métodos de Análisis Multivariante"
  Ed. CMC Editions
  Barcelona, España
  2012
  297 págs.

- CHMELIK, Samantha.
  "Know! Market Research and Analysis - the Fundamentals"
  Ed. Knowledge inForm, Incorporated
  2005
  21 págs.

- GIL Flores, Javier, et.al.
  "Análisis Discriminante"
  Ed. La Muralla
  Madrid, España
  2001
  Págs.7-51
  125 págs.

- HIAM, Alexander.
  "Marketing For Dummies"
  Ed. John Wiley & Sons
  2009
  384 págs.

- HUBERTY, Carl J.
  "Applied discriminant analysis"
  Ed. J. Wiley
  1994
  466 págs.

- LÉVY Mangin, Jean Pierre, et.al.
  "Análisis Multivariable de las ciencias sociales"
  Ed. Pearson Education
  Madrid, España
  2003
  896 págs.

- LÓPEZ, Heriberto.
  "Los Niveles Socioeconómicos y la distribución del gasto"
  2009
  Comité de Niveles Socioeconómicos de la AMAI e Instituto de Investigaciones Sociales SC
  44 págs.

- MANLY, Bryan F.J.
  “Multivariate Statistical Methods: A primer”
  Ed. Chapman & Hall
  Segunda ed.
  1986
  75 págs.

- RENCHER, Alvin C.
  “Methods of Multivariate Analysis"
  Ed. Wiley-Interscience
  2002
  708 págs.

- SIDNEY, Siegel
  “Estadística no paramétrica aplicada a las ciencias de la conducta“
  Ed. Trillas
  Segunda reimpresión en español
  1970
  399 págs.

- WALPOLE, Ronald
  “Probabilidad y Estadística para ingenieros”
  Ed. Prentice-Hall Hispanoamericana
  Sexta ed.
  México, 1999
  752 págs.