



**UNIVERSIDAD NACIONAL
AUTÓNOMA DE MÉXICO**

**FACULTAD DE ESTUDIOS SUPERIORES
ACATLÁN**

**"FUNDAMENTOS MATEMÁTICOS DEL ANÁLISIS DE
CORRESPONDENCIA SIMPLE PARA LA INTERPRETACIÓN GRÁFICA DE
DATOS CATEGÓRICOS"**



**Tesis que para obtener el título de:
LICENCIADA EN MATEMÁTICAS APLICADAS Y COMPUTACIÓN**

Presenta:

GABRIELA ACOSTA ESPINOSA

Asesora:

M. en E. MARÍA DEL CARMEN GONZÁLEZ VIDEGARAY



MARZO 2004



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

A mis Padres,
las personas más importantes
para mí desde siempre.

A Ernesto y a todos los de OPHIS Alta Percepción,
por el tiempo, el espacio y el apoyo
incondicional que siempre me han brindado.
Mil Gracias.

A mis amigos y maestros de la Universidad,
por ser parte de una de las etapas
más importantes de mi vida.

ÍNDICE

INTRODUCCIÓN

CAPÍTULO I	ESCALAS DE MEDICIÓN Y ESTADÍSTICO Chi-Cuadrada	1
	Objetos	
	Variables	
	Clasificación	
	Medición y Escala	4
	Escala Nominal	
	Escala Ordinal	
	Escala Dicotómica	
	Escala de Intervalo	
	Escala Racional	
	Análisis de Correspondencia como Método Interdependiente	11
	Tablas de Contingencia	14
	Estadístico Chi-Cuadrada (χ^2)	19
CAPÍTULO II	FUNDAMENTOS MATEMÁTICOS DEL ANÁLISIS MULTIVARIADO	23
	Espacio Euclidiano y Bases del Espacio	24
	Dimensión del Espacio Vectorial	
	Base Ortogonal	
	Cambio de la Base del Espacio	25
	Transformaciones Ortogonales	28
	Rotación de los Vectores de la Base	
	<i>Rotación Impropia</i>	
	Rotación de Puntos	
	Matrices de Transformación sobre la Nueva Base	36
	Efectos Geométricos de Transformaciones	38
	Translación	
	Reflexión	
	Ajuste	
	Rotación	
	Transformaciones Compuestas	
	Matrices de Datos Estadísticos	45
	Eigenestructura de Matrices Simétricas	47
	Diagonalización de la Matriz de Transformación	52
	Reducción de la Dimensión	53
	Rango de una Matriz	
	Eigenestructura y Rango	
	Descomposición en Valores Singulares	56
CAPÍTULO III	ANÁLISIS DE CORRESPONDENCIA EN DATOS CATEGÓRICOS	63
	Representaciones Gráficas	63
	Perfiles de Columna y de Renglón	66
	Masas	71
	Distancia χ^2	72
	Distancias χ^2 en el Espacio Euclidiano	

Inercia	77
Análisis de Inercia	78
Descomposición de la Inercia	
<i>Contribuciones por Celda</i>	
<i>Inercias de Renglón y Columna</i>	
<i>Inercias Principales</i>	
Contribuciones Absolutas	
Contribuciones Relativas	
Calidad de la Gráfica en un Subespacio	
Anidamiento de los Ejes Principales	
Análisis por Renglón y por Columna	84
Factor de Escalamiento	
Álgebra y Geometría en el Análisis de Correspondencia	87
Matriz de Dispersión y Matriz de Correspondencia	
Descomposición en Valores Singulares	
Mapas Simétricos y Asimétricos	98
CAPÍTULO IV EJEMPLO DE APLICACIÓN	101
CONCLUSIONES	
REFERENCIAS BIBLIOGRÁFICAS	

INTRODUCCIÓN

El interés que se ha despertado en la búsqueda de nuevos procedimientos que permitan analizar y comprender los distintos fenómenos que ocurren en nuestro entorno ha ocasionado el desarrollo de técnicas y métodos con los cuales ha sido posible, hasta cierto punto, tener un mayor control sobre las situaciones que nos afectan. El gradual crecimiento de las aplicaciones de la ciencia en las distintas áreas del conocimiento también ha permitido el perfeccionamiento de métodos que teóricamente tuvieron sus orígenes hace varios años.

El Análisis de Correspondencia es un método estadístico que se ha visto beneficiado por los avances de la tecnología y actualmente es aplicado en campos tan variados como Sociología, Psicología, Medicina, Biología, Economía, Educación, Mercadotecnia, Publicidad, entre otros. El método es particularmente útil al trabajar con datos categóricos cuyas características especiales imponen ciertas restricciones en su manejo y, además, permite una forma sencilla de interpretar y comprender los resultados.

El objetivo del presente trabajo es explicar los fundamentos matemáticos del Análisis de Correspondencia Simple para dar a conocer los procedimientos con los que opera; para ello, el trabajo consta de cuatro capítulos. El primero de ellos hace referencia a conceptos básicos sobre las características y el manejo de conjuntos de datos que proporcionan determinada información. El Capítulo II se enfoca a explicar los aspectos relacionados con el Álgebra y la Geometría que conforman la estructura básica del análisis multivariado. En el Capítulo III se especifica sobre el Análisis de Correspondencia como un caso especial del análisis multivariado, mencionando definiciones de particular relevancia en la aplicación del método. Finalmente, el Capítulo IV se centra en emplear el Análisis de Correspondencia en un ejemplo práctico de un análisis de datos categóricos.

Cuando se aplica algún método o técnica previamente establecidos con el fin de descubrir comportamientos o situaciones que no se perciben fácilmente y que, al conocerlos, ayudan a tomar decisiones relevantes, pocas veces se toma conciencia de los mecanismos que operan detrás de tales métodos, sobre todo si se hace uso de programas de cómputo que agilizan las tareas y operaciones. Este trabajo pretende dar a conocer un poco más sobre tales procesos y, además, contribuir a enriquecer y promover el conocimiento y uso de herramientas matemáticas que están a nuestro alcance.

CAPÍTULO I

ESCALAS DE MEDICIÓN Y ESTADÍSTICO Chi-Cuadrada

Todo lo que se encuentra a nuestro alrededor ha sido, para la ciencia, objeto de estudio. Cualquier fenómeno o evento que se suscita, genera una serie de cuestionamientos que han querido ser resueltos. En el afán de querer comprender tales fenómenos, se han ido desarrollando diferentes ramas de investigación. Una de ellas es el Análisis Estadístico. Todas las observaciones que pueden hacerse sobre el medio que nos rodea constituyen una fuente potencial de datos, los cuales son la materia central del Análisis Estadístico.

El Análisis de Correspondencia, como método estadístico, permite realizar análisis de datos en situaciones que presentan ciertas condiciones específicas. Para saber con claridad cuándo es posible aplicar este método, es necesario conocer los elementos que deben considerarse y comprender algunos conceptos básicos relativos a ellos.

Para facilitar el manejo de los datos y evitar ambigüedades por describirlos con palabras, éstos pueden convertirse en números, sin embargo, esta transición debe ser definida estableciendo los conceptos bajo los cuales la relación entre las observaciones y los números es válida, haciendo posible también invertir el proceso e interpretar fácilmente lo que los números representan. Esto se conoce como **medición** de variables, lo que hace más fácil observar el comportamiento de los datos al aplicar métodos estadísticos.

Objetos

En primer lugar, debe establecerse una forma de referirse a las fuentes de datos. La expresión que ha sido más aceptada en el Análisis Estadístico para nombrar a estas fuentes es la de **objetos**, aunque también se hace referencia a ellas como “cosas” o “entidades” en algunos casos. Así, en un sentido muy general, los objetos se refieren no sólo a las fuentes de datos tangibles como individuos, animales, productos para el consumo, plantas o cualquier unidad física o biológica, sino también a las intangibles como estímulos, percepciones sensoriales, periodos de tiempo, pensamientos, eventos, etc., es decir, cualquier aspecto sobre el cual puedan hacerse observaciones.

Para fines de análisis estadísticos, los datos provienen, generalmente, de una muestra aleatoria de la población total (cualquier clase bien definida de objetos). Para que

esta muestra sea representativa, cada miembro de la población debe tener la misma probabilidad de pertenecer a ella.

Variables

Una vez definidos los objetos debe notarse que éstos, cualquiera que sea su naturaleza, son resultado de un conjunto de características o atributos. De hecho, un objeto no puede mantenerse independiente de sus características. Dado que éstas pueden diferir de un objeto a otro y tomar valores diferentes, es posible medir a los objetos tomando en cuenta las características que los identifican.

Cuando el número de valores o categorías diferentes que puede asumir un atributo o característica de un objeto es dos o más, entonces estos atributos se consideran **variables**. La fuerte relación que existe entre los objetos y las variables se explica al no poder concebir un objeto sin referirse a las variables que lo definen, del mismo modo, las variables no tienen significado si no se refieren a un objeto que posea tales atributos.

Al observar cualquier objeto se pueden percibir sus atributos (composición, color, forma, peso, tamaño, etc.); estas variables podrían tomar distintos valores. Así mismo, cualquier disciplina tiene variables de interés, por ejemplo, en Economía un producto (objeto) puede estudiarse bajo ciertos atributos como calidad, precio, costo, ventas, etc. (variables). En Psicología, la actitud (objeto) de un individuo se analiza por sus componentes cognoscitivos, afectivos y de comportamiento (variables).

Clasificación

Las variables poseen ciertas propiedades especiales que permiten hacer distinciones entre ellas y clasificarlas en diferentes categorías.

Una primera distinción se hace entre variables **cualitativas** y **cuantitativas**. Las variables cualitativas indican la presencia o ausencia de atributos, características o propiedades que describen a los objetos sin reflejar cantidades o distancias entre ellos. Cuando los datos son cualitativos, éstos se encuentran clasificados en categorías asumiendo valores no numéricos, es decir, son **datos no métricos** y se conocen también como **variables categóricas**. Por otro lado, las variables cuantitativas son **datos métricos** que

tienen la propiedad de que al ser medidos pueden observarse diferencias en cantidad o grado estableciendo distancias relativas entre ellos.

Otra forma de clasificación es la que distingue entre variables **discretas y continuas**. Las variables discretas son aquellas que toman un número finito y generalmente pequeño de valores. En la transición de un valor al siguiente valor no existe un punto intermedio que pueda considerarse como otro valor de la variable. Las variables medidas a través de frecuencias, donde se contabiliza el número de ocurrencias de un determinado evento, son ejemplos de variables discretas. Otros ejemplos son las categorías de afiliación religiosa, tipo de sangre, continentes, etc. En ocasiones el término discreto es sustituido por **cualitativo, nominal o categórico** (de cualidad, nombre y categoría).

Mientras que generalmente las variables discretas toman valores enteros (aunque no es regla general y también es posible utilizar términos fraccionarios), las variables continuas pueden tomar cualquier valor numérico dentro del rango de escala indicando la cantidad de la variable. Dados dos valores de la variable, siempre existe otro valor intermedio entre ellos. Estas variables también se conocen como cuantitativas o de intervalo. Por ejemplo, las variables edad, tiempo, medidas de peso, altura, tasas de interés, temperatura, distancia, promedios, etc., son continuas.

A veces la distinción entre una variable discreta y una continua no es muy clara y es posible que una variable que es continua en un sentido teórico sea tratada como discreta, sobre todo si existen muchas categorías que representan atributos cuantitativos. Por ejemplo, para medir el tiempo (variable continua) puede utilizarse un reloj digital para medirlo en términos de minutos o de horas (categorías discretas). O bien, para la variable edad pueden definirse categorías discretas que representen rangos de edad (valor 1 para edades de 0 a 5 años, valor 2 para edades de 6 a 10 años, etc.).

Finalmente, las variables pueden clasificarse de acuerdo con el número de valores que pueden tomar. Aquellas que sólo asumen dos valores son llamadas **variables dicotómicas**. Algunos ejemplos son el sexo (masculino o femenino), categorías de uso (usuario, no usuario), al agrupar al reino animal (vertebrados o invertebrados), actitudes (acuerdo o desacuerdo), creencias o convicciones (verdadero o falso), etc.

Como en el caso de las variables discretas, algunas variables pueden considerarse como continuas en un sentido estricto pero pueden ser tratadas como dicotómicas tomando solamente dos posibles valores de la variable, dependiendo del objetivo que sea de interés para el investigador; tal es el caso de actitudes u opiniones, por ejemplo, es posible que únicamente se requiera saber si un cliente está o no está satisfecho con algún servicio que se le brinda, sin importar el grado de satisfacción (muy satisfecho, poco satisfecho, etc.), y sólo considerar los valores satisfecho e insatisfecho.

En general, si en cualquier variable que teóricamente es continua se especifican intervalos en los valores que pueda asumir entonces es posible considerar a esta variable como discreta o dicotómica dependiendo del número de intervalos que se consideren.

Cualquiera que sea la clasificación, las categorías de las variables deben cumplir dos condiciones: ser exhaustivas y mutuamente exclusivas. La primera condición se refiere a tener el número suficiente de categorías para que todos los elementos de la población puedan pertenecer al menos a una de ellas. Las categorías serán mutuamente exclusivas si cada miembro de la población pertenece a una y sólo una categoría.

Para que las observaciones (datos) puedan representarse numéricamente debe establecerse alguna escala, la cual ya es propiamente un conjunto de números, donde lo importante no son los números en sí, sino la interpretación de éstos, ya que sus propiedades difieren en cada escala dependiendo del tipo de variables que representen.

De aquí la importancia de que al comenzar un análisis se identifique el tipo de datos que se han recolectado ya que éstos deben tener las características necesarias para ajustarse al tipo de análisis estadístico al que estarán sujetos. De este modo, la calidad y confiabilidad de los resultados permitirán una interpretación correcta de éstos.

MEDICIÓN Y ESCALA

Cuando se establece una escala se está determinando una forma de medir los datos. De esta manera, al hablar de medición se hace referencia a los métodos que se siguen para establecer las dimensiones o cantidades de ciertos fenómenos y que, en forma general, consideran 4 puntos o niveles principales que se ejemplifican en la figura (1.1):

- 1) La *variable*, atributo, característica o propiedad que se quiere medir.
- 2) La *categoría* o grado (modalidad) en que se manifiesta la variable medida.
- 3) El *valor* o forma de expresar de forma numérica el atributo.
- 4) La *relación* aceptable entre los distintos valores de la variable.

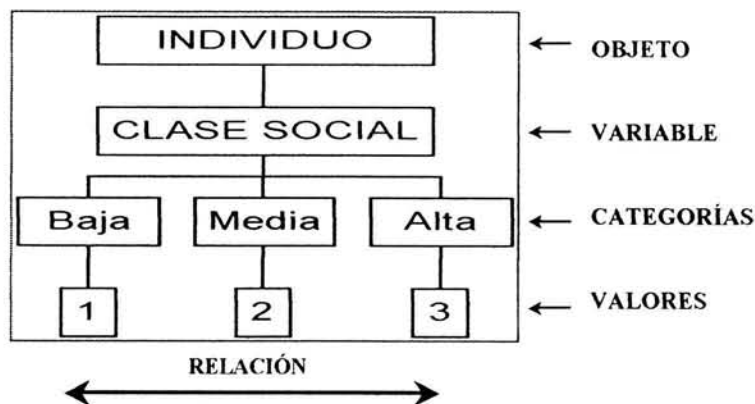


Fig. (1.1)

En resumen, medición puede definirse como la asignación de símbolos (números) a las características (atributos) de los objetos de acuerdo a ciertas reglas pre-especificadas. En este contexto, no se mide al objeto en sí, sino algunas de sus características, como actitudes, percepciones, preferencias, motivaciones, valores, necesidades, etc., donde debe existir una correspondencia uno a uno entre los números asignados y las categorías de los objetos que están siendo medidos, cada número debe ser asignado a una sola categoría y cada categoría debe identificarse con un único número.

Las reglas para asignar números a las características de los objetos deben estandarizarse y aplicarse uniformemente en un mismo análisis, es decir, no deben variar de un objeto a otro ni modificarse con el tiempo, de tal forma que las transformaciones que pueden aplicarse a la estructura numérica no afecten el significado original establecido. Estas reglas determinan la forma en que los números son asignados y cada conjunto de reglas determina un tipo general de escala. A continuación se definen con más detalle los aspectos referentes a los tipos de escala.

Las **escalas** reflejan la necesidad de "transformar" los datos a una forma cuantitativa para compararlos. Son números asignados a las características de los objetos en el proceso de medición y representan el nivel de medida que poseen los datos de acuerdo a su

naturaleza y al propósito de su análisis clasificándolos en base a ciertas características de interés. Los números, utilizados como “etiquetas”, poseen varias propiedades que generan diferentes niveles de medición, y diferentes niveles pueden proporcionar diferentes propiedades a los números.

Así, los datos también pueden diferenciarse de acuerdo al tipo de escala que se les asigne y que será definida según el tipo de transformaciones bajo las cuales dicha escala no varíe y conserve sus propiedades y significado iniciales.

Para definir las escalas se consideran algunas operaciones básicas que permiten diferenciarlas. Existen 4 escalas principales: **nominal**, **ordinal**, **de intervalo** y **racional**. En la escala nominal se establece una relación de igualdad que permite identificar, diferenciar y clasificar a los objetos. En la escala ordinal, se puede observar una relación de mayor o menor entre los objetos, ordenándolos y estableciendo diferencias entre ellos. Si estas diferencias pueden medirse y cuantificarse, entonces se tiene una escala de intervalo. Finalmente, en la escala racional es posible determinar un valor cero que indique la ausencia o falta del atributo medido.

Bajo estas condiciones, para medir los datos no métricos (cualitativos) se utiliza la escala nominal o la escala ordinal. Los datos métricos (cuantitativos) son medidos con la escala de intervalo o la racional. Sus características principales se muestran en la tabla (1.1).

Escala Nominal

La escala nominal representa el nivel más básico de medición en cuanto a precisión, ya que simplemente asigna números que sólo sirven como “etiquetas” para nombrar, identificar y clasificar a los objetos. Los valores de la variable que está siendo medida difieren únicamente por su tipo. Es llamada también escala categórica, discreta o simbólica. Cada valor o etiqueta representa una categoría específica y cada categoría es diferente de las demás, éstas no pueden ser relacionadas numéricamente ya que no indican la cantidad de las características que el objeto posee.

La escala nominal puede aplicarse en dos casos de acuerdo a su propósito:


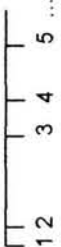
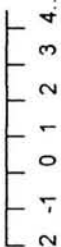
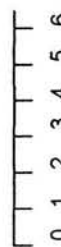
ESCALA	CARACTERÍSTICAS	OPERACIONES EMPÍRICAS BÁSICAS	EJEMPLOS	ESTADÍSTICOS DESCRIPTIVOS Y DE INFERENCIA
NOMINAL	Números (etiquetas) que identifican y clasifican objetos, indicando sólo diferencias en "tipo".	Relación de igualdades (Relación de Identidad)	Números de identificación personal, números asignados a jugadores en un equipo, marcas de automóviles, etc.	<p>  1 2 3 4 </p>
ORDINAL	Números que indican las posiciones relativas de los objetos pero no la magnitud de las diferencias entre ellos.	Relación de mayor o menor (Relación de Orden)	Orden de los participantes en una competencia, clases sociales	<p>  1 2 3 4 5 ... </p>
DICOTOMICA	Cuando sólo existen dos categorías, ya sean nominales u ordinales.	Relación de Identidad u Orden	Clasificación por sexo	<p> Media Varianza Hombre-Mujer </p>
INTERVALO	Las distancias entre los objetos pueden ser comparadas. El punto cero es arbitrario. Las diferencias iguales entre valores representan cantidades iguales.	Determinación de la igualdad de la diferencia de los intervalos (Distancia)	Escalas de temperatura, el tiempo medido en diferentes tipos de calendarios.	<p> Rango Media Desviación estándar  -2 -1 0 1 2 3 4... </p>
RACIONAL	Pueden calcularse proporciones y razones entre los valores de escala. El punto cero es fijo.	Existencia de un cero real (o absoluto)	Longitud, peso, edad, costos, tiempo de reacción.	<p> Media geométrica Coeficiente de variación  0 1 2 3 4 5 6 ... </p>

Tabla (1.1) Escalas de Medición

1) Para identificar a los objetos: en este caso existe una correspondencia uno a uno entre las etiquetas y los objetos, donde a cada etiqueta le corresponde uno y sólo un objeto y a cada objeto le corresponde una y sólo una etiqueta; por ejemplo, al asignar un número de identificación personal a un grupo de personas.

2) Para clasificar objetos: las etiquetas sirven para diferenciar clases o categorías, dichas clases deben ser mutuamente exclusivas y colectivamente exhaustivas. Los objetos que pertenecen a una misma clase se consideran equivalentes con respecto a la característica que representa la etiqueta de esa clase; por esto, dos clases no pueden tener la misma etiqueta. Un ejemplo es cuando se clasifica a las personas por su estado civil: soltero, casado, viudo, divorciado, o para determinar la afiliación política: demócratas, independientes, republicanos.

Las etiquetas asignadas no indican un orden explícito porque no puede establecerse una relación de mayor o menor entre sus valores y los intervalos que existen entre ellos no pueden compararse cuantitativamente. Los datos permiten únicamente contabilizar el número de ocurrencias en cada clase o categoría, por lo que la única operación estadística permitida es la frecuencia y, a partir de esto, pueden calcularse estadísticos como porcentajes, moda, estadístico chi-cuadrada (χ^2) y pruebas binomiales.

Escala Ordinal

En esta escala los números o etiquetas siguen una secuencia con un orden que no varía, permitiendo un nivel de mayor precisión que la escala nominal. Los objetos son ordenados de acuerdo a las etiquetas que indican la magnitud o cantidad relativa del atributo que poseen, de tal forma que pueden compararse e identificarse aquellos que tienen mayor o menor cantidad del atributo en cuestión. Es decir, existe una jerarquía entre los valores del atributo, establecida por la propiedad de transitividad (si $A > B$ y $B > C$, entonces $A > C$) la cual no posee la escala nominal.

En la escala ordinal las magnitudes de las distancias entre los objetos ordenados no son significativas, es decir, no puede determinarse la magnitud de las diferencias que existen entre los valores de la variable, ya que sólo se indica su posición relativa. Se puede establecer que $A > B$ y $B > C$ pero no necesariamente la distancia entre A y B es la misma que

entre B y C. Un ejemplo de esto se da al conocer el lugar obtenido por los ganadores en una competencia: primero, segundo y tercer lugar, etc.

Al igual que en la escala nominal, los objetos equivalentes en atributos tienen la misma posición y, mientras se preserve el orden, cualquier transformación de la escala es válida, por lo que además de las frecuencias, pueden calcularse estadísticos como percentiles, cuartiles y mediana.

Escala Dicotómica

Existe una escala que, debido a que los puntos de vista de algunos autores difieren, se ha optado por tratarla como un caso aparte de las cuatro escalas principales. Esta escala hace referencia a las variables dicotómicas. En algunos casos estas variables pueden tratarse con una escala nominal, como es el caso de la clasificación de las personas por su sexo: masculino o femenino y en otros casos es posible identificar un orden en las categorías, por ejemplo si se desea calificar el rendimiento de un alumno como alto o bajo. Por esto su clasificación no siempre resulta muy precisa y se considera un caso especial, con sus propias características y propiedades. Una escala dicotómica puede aplicarse, por lo tanto, cuando sólo existen dos categorías que satisfacen las condiciones de una escala nominal u ordinal.

Una característica de estas variables es que pueden aplicarse operaciones y calcularse estadísticos como la media y la varianza. Por ejemplo, teniendo una variable dicotómica a la que se le asignan los valores de 0 y 1 a sus respectivas categorías y se obtiene una media con valor de 0.62, indicaría que el 62% de los objetos medidos pertenecen a la categoría con etiqueta "1", lo que no sería significativo al aplicarse en casos donde se tienen más de dos variables

Escala de Intervalo

La escala de intervalo posee todas las características de la escala ordinal, además de otras propiedades. En ésta, la distancia o diferencia entre dos valores de escala adyacentes cualesquiera es igual. Estas distancias, que son numéricamente iguales, representan diferencias iguales en las características que están siendo medidas, es decir, los valores de

escala están separados por intervalos constantes. De esta forma, por ejemplo, la diferencia entre los valores 22 y 25 es la misma que entre 51 y 54.

El punto cero en este caso no es absoluto, porque no significa que el objeto no posea el atributo medido, ni tampoco es fijo, sino arbitrario, por lo que las razones entre los valores de esta escala no tienen un significado real; como puede verse en las escalas de temperatura; primero, un valor de 0° no indica que no hay temperatura, ya que incluso pueden observarse temperaturas bajo cero, y segundo, un valor de 0° medido en una determinada escala, por ejemplo Centígrados, no representa la misma temperatura medida en grados Fahrenheit, así, 100°F no debe interpretarse como el doble de 50°F , ya que si se convierten estas temperaturas a grados Centígrados, es decir 37.7°C y 10°C , la razón de éstas es 3.77 y no es lo mismo que la razón de 100°F y 50°F que es 2. Estas conclusiones son incorrectas debido a que ambas escalas posicionan el punto cero en un punto arbitrario y no en el "verdadero" punto cero que representaría cero temperatura.

Cualquier transformación lineal positiva de la forma $y=a+bx$ es válida y conserva las propiedades de la escala, donde x es el valor de escala original, y el valor transformado, a cualquier constante y b cualquier constante positiva.

Las operaciones permitidas son las mismas que en la escala nominal y la ordinal, además de las operaciones aritméticas básicas de suma y resta y las operaciones estadísticas como la media aritmética y la desviación estándar.

Escala Racional

Este es el mayor nivel de precisión en la medición. La escala racional posee todas las propiedades de la escala nominal, ordinal y de intervalo pero se caracteriza, además, por tener un punto cero absoluto o verdadero que permite referirse a la ausencia o carencia de alguna característica. Con esta escala es posible identificar o clasificar objetos, ordenarlos, comparar diferencias o intervalos y también calcular proporciones y razones de los valores de escala, donde razones iguales tienen igual significado debido a la presencia del cero absoluto. Por ejemplo, cuando se habla de longitudes, éstas pueden medirse con una escala en pies, metros, pulgadas, etc.; la variable peso se mide con una escala en libras,

kilogramos, etc.; el dinero con una escala en pesos, dólares, marcos, etc; el tiempo se mide en segundos, horas, años, etc.

En cualquiera de estos casos, cuando una variable es medida con diferentes escalas, lo importante es considerar las unidades en que éstas están siendo medidas, ya que una escala alternativa es sólo una transformación de otra y debe elegirse aquella que resulte más práctica y conveniente.

Las transformaciones permitidas son proporcionales de la forma $y=bx$, donde x es el valor de escala original, y el valor transformado y b cualquier constante positiva.

En este caso pueden calcularse las operaciones de multiplicación y división y todas las operaciones estadísticas, como la media geométrica y el coeficiente de variación.

ANÁLISIS DE CORRESPONDENCIA COMO MÉTODO INTERDEPENDIENTE

Como se mencionó anteriormente, la importancia de reconocer el tipo de datos que se presentan en una investigación se verá reflejada al obtener los resultados del análisis e interpretarlo para tomar las decisiones pertinentes. La obtención de resultados confiables también dependerá de elegir el método estadístico adecuado a los datos.

Una clasificación general de los métodos estadísticos es la que los divide en métodos univariados y métodos multivariados.

Los métodos univariados pueden aplicarse cuando se va a analizar un único atributo o característica de los elementos de la muestra, o bien, cuando se tienen varias variables de cada elemento pero donde cada una se analiza por separado.

A su vez, el análisis univariado se subdivide de acuerdo al tipo de datos, ya sean métricos (donde se utilizan las escalas de intervalo y la racional), o no métricos (medidos con la escala nominal y ordinal). También se considera el número de muestras independientes ya que se puede trabajar con una sola muestra o tener dos o más de ellas.

Esta clasificación genera distintos métodos en los que se pueden aplicar diferentes pruebas estadísticas, como se muestra en la figura (1.2):

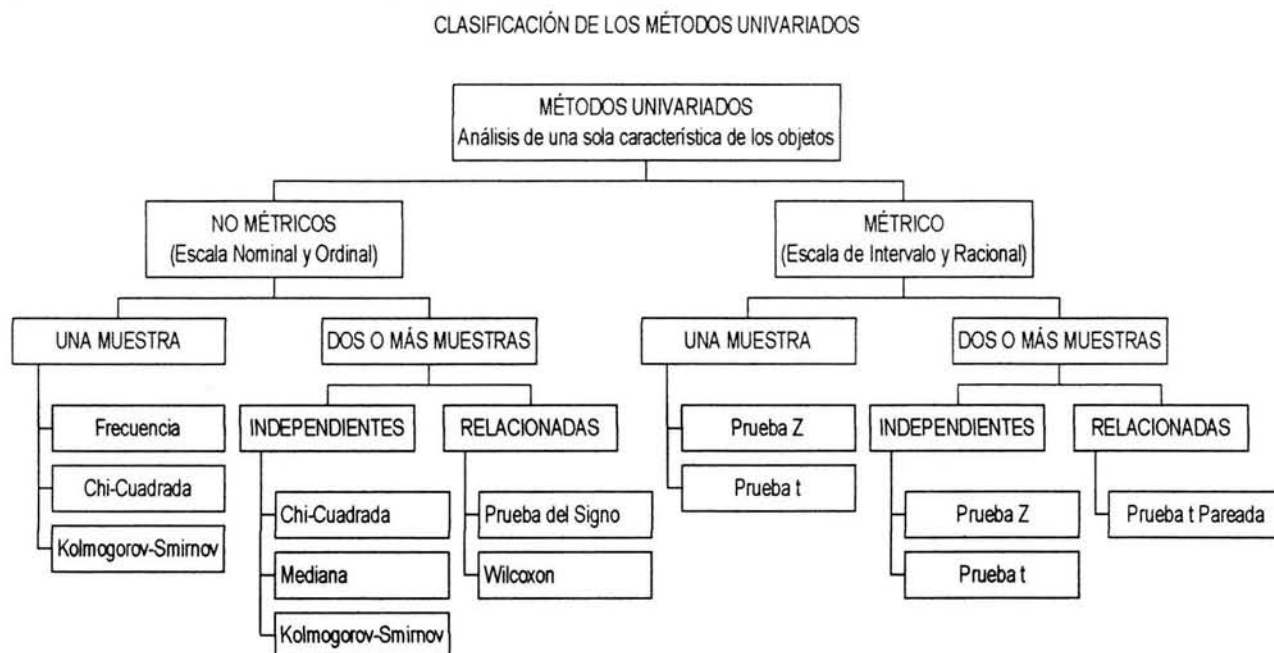


Fig. (1.2) Métodos Univariados

En el caso del análisis multivariado se analizan varias variables de cada objeto simultáneamente, es decir, se refiere a estudiar las relaciones existentes entre dos o más variables. En este caso todas las variables deben ser aleatorias y no sólo es el número de variables lo que importa sino las combinaciones entre ellas, medidas a través de correlaciones y covarianzas para medir, explicar y predecir el grado de relación existente.

Al igual que en el análisis univariado, existe una clasificación para los métodos multivariados. Éstos pueden clasificarse en técnicas dependientes e interdependientes. En las primeras, se presentan una o más variables dependientes y el resto son variables independientes; las interdependientes se enfocan a agrupar los datos o los objetos basándose en las similitudes o relaciones entre ellas, en este caso no hay variables que sean dependientes o independientes. La figura (1.3) muestra la clasificación del análisis multivariado y los métodos que pueden aplicarse en cada caso.

CLASIFICACION DE LOS MÉTODOS MULTIVARIADOS

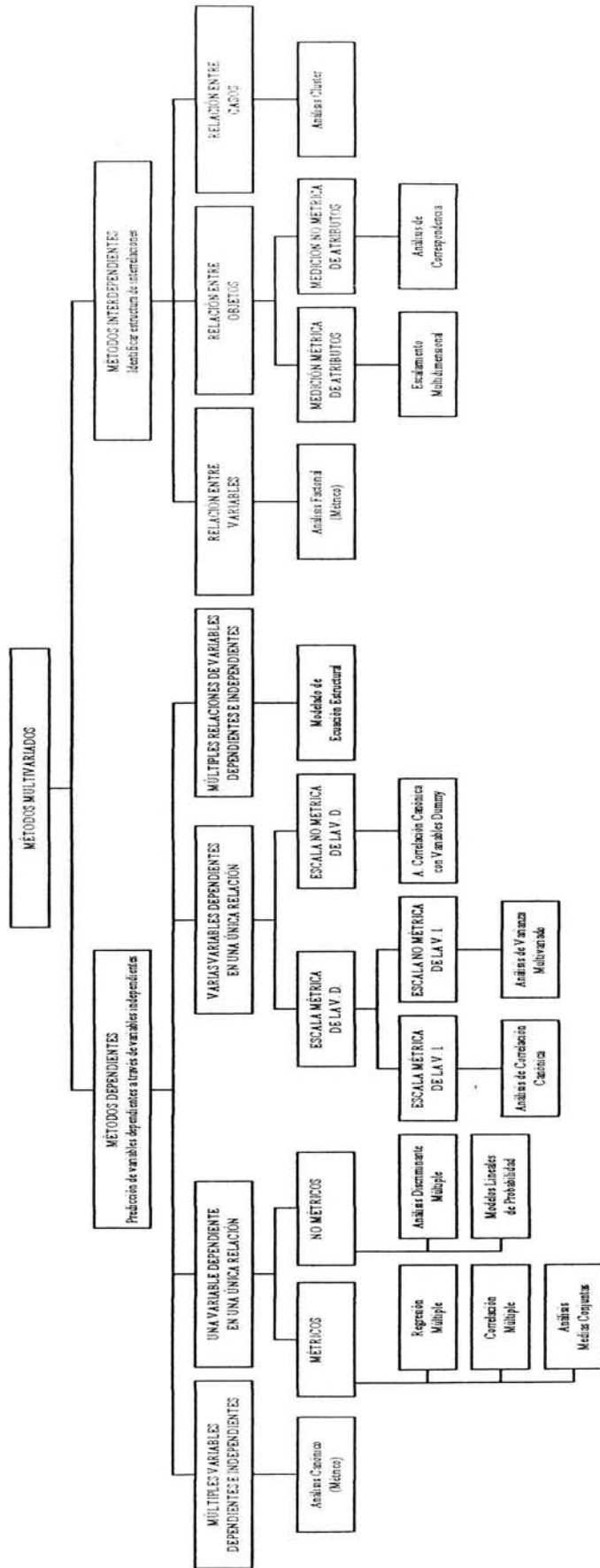


Fig. (1.3) Métodos Multivariados

Esta clasificación proporciona una idea más completa de dónde se ubica el Análisis de Correspondencia, las similitudes y diferencias con otros métodos y los casos en los que puede o no puede ser aplicado de acuerdo al tipo de datos que se tengan y las escalas que se manejen. En este contexto, para propósitos del presente trabajo, se observa que el Análisis de Correspondencia es un método estadístico multivariado e interdependiente que se enfoca a descubrir relaciones entre los objetos estudiados con base en ciertas características o atributos, con datos cualitativos que describen propiedades de los objetos y donde las variables de interés son categóricas y se miden utilizando una escala nominal.

Cuando se tiene este tipo de datos y escalas existen algunas herramientas que facilitan su manejo e interpretación, una de ellas es la tabla de contingencia, que será parte fundamental en la aplicación del Análisis de Correspondencia.

TABLAS DE CONTINGENCIA

De acuerdo a las propiedades de la escala nominal, la única operación estadística que puede calcularse con datos categóricos es la frecuencia.

Algunas veces es necesario obtener información a partir del estudio de una variable que describa características de los objetos, en este caso, se puede examinar la distribución de frecuencia a través de tablas que muestren los valores del conteo, los porcentajes y las frecuencias acumuladas que se representan gráficamente por medio de histogramas. Esto también puede hacerse cuando se tienen varias variables pero analizando cada una de ellas por separado.

Sin embargo, en ocasiones resulta más interesante y más útil establecer si existen relaciones entre varias variables. El Análisis de Correspondencia Simple es la forma más básica donde se analizan relaciones entre dos variables categóricas haciendo uso de las **tablas de contingencia** o **tablas cruzadas**. Sus fundamentos y procedimientos pueden extenderse al caso de más de dos variables constituyendo el Análisis de Correspondencia Múltiple.

Las tablas de contingencia reflejan la distribución conjunta de dos o más variables con un número limitado de categorías, asumiendo que cada variable tiene una escala nominal u ordinal. Las celdas en una tabla de contingencia contienen las frecuencias observadas de cada combinación entre las categorías de las variables. Si lo que se busca es identificar asociaciones entre k variables cualitativas, entonces se hace referencia a tablas de contingencia de k dimensiones. El caso más sencillo es en el que se analizan dos variables. La forma general de una tabla de contingencia de dos dimensiones se muestra en la tabla (1.2).

		Columnas (Variable 2)					Total	
		1	2	.	.	c		
Renglones (Variable 1)	1	n_{11}	n_{12}	.	.	.	n_{1c}	n_{1+}
	2	n_{21}	n_{22}	.	.	.	n_{2c}	n_{2+}
	r	n_{r1}	n_{r2}	.	.	.	n_{rc}	n_{r+}
Total		n_{+1}	n_{+2}				n_{+c}	$n_{++} = n$

Tabla (1.2) Forma general de una tabla de contingencia bidimensional

En ella se clasifican las observaciones de una muestra de tamaño N respecto a dos variables cualitativas, una de ellas con r categorías que se representan en los renglones y la otra con c categorías representadas en las columnas.

En cada celda se muestran las **frecuencias observadas** para la i -ésima categoría de la variable renglón (variable 1) y la j -ésima categoría de la variable columna (variable 2); estas frecuencias se denotan por n_{ij} para los valores de $i=1\dots r$ y $j=1\dots c$.

Al sumar los valores sobre un renglón o columna completos se obtienen las frecuencias observadas para cada una de las categorías de las variables, estas frecuencias son los *totales marginales* que se denotan por n_{i+} en el caso de las categorías de los renglones y por n_{+j} para las de las columnas.

Siguiendo la misma notación, el número total de observaciones (N) se representa por n_{++} o simplemente n .

De lo anterior se tiene que:

$$n_{i+} = n_{i1} + n_{i2} + \dots + n_{ic} = \sum_{j=1}^c n_{ij}$$

$$n_{+j} = n_{1j} + n_{2j} + \dots + n_{rj} = \sum_{i=1}^r n_{ij}$$

$$n_{++} = \sum_{i=1}^r \sum_{j=1}^c n_{ij} = \sum_{i=1}^r n_{i+} = \sum_{j=1}^c n_{+j}$$

La distribución de frecuencia de cada variable representada en la tabla es subdividida respecto a las categorías de la otra variable. Los renglones y las columnas de la tabla de contingencia no tienen un significado predefinido, es decir, las categorías de una variable pueden ser representadas ya sea en los renglones o en las columnas sin afectar su interpretación en el análisis. En ocasiones, puede resultar más conveniente representar una variable con valores en los renglones y en otras ocasiones en las columnas; esto depende únicamente del enfoque que se desee en el análisis, ya que los resultados, de una u otra manera, serán los mismos, como se verá más adelante.

Con los datos de la tabla de contingencia, es posible determinar si las variables cualitativas involucradas son contingentes (dependientes) una de la otra. La independencia de las variables se da en el sentido de que los resultados obtenidos para una de las variables no están influenciados o relacionados con los valores obtenidos de la otra variable, es decir, que la ocurrencia de una de ellas no determina estrictamente la ocurrencia de la otra, y aún si se descubriera que ambas variables son dependientes esto no debe interpretarse como que una de ellas es causa de la otra.

Para determinar la relación entre las variables deben considerarse las diferencias existentes entre los valores observados y los valores esperados. Esto puede visualizarse fácilmente en tablas de contingencia pequeñas, como el caso especial de una tabla de 2x2. Por ejemplo, supóngase que se pregunta a 100 hombres y a 100 mujeres si están de acuerdo (A) o en desacuerdo (D) en relación a una determinada situación. Los datos obtenidos se muestran en la tabla (1.3).

	Hombres	Mujeres	Total
A	20	60	80
D	80	40	120
Total	100	100	200

Tabla (1.3)

Con estos datos se pueden evaluar las desviaciones entre las frecuencias observadas y las **frecuencias esperadas**, estas últimas pueden establecerse a partir de los totales marginales ya que no se conocen *a priori* dichos valores.

Asumiendo en principio que no existe relación entre las variables de género (hombre-mujer) y de actitud (acuerdo-desacuerdo), se esperaría que los resultados se reflejaran en las mismas proporciones dadas por los totales marginales, es decir, si en total 80 personas estuvieron de acuerdo, representando el 40% de toda la muestra, se esperaría que el 40% de los hombres y el 40% de las mujeres haya optado por la actitud A. Las frecuencias observadas, sin embargo, difieren de las esperadas ya que 20 hombres y 60 mujeres están de acuerdo con la situación que se les planteó y estas diferencias deben cuantificarse para determinar el grado de la relación entre las variables.

Algo similar ocurre en tablas de contingencia más grandes de rc (donde r y c representan el número de renglones y columnas respectivamente y ambos son mayores a 2). Supóngase que se tiene la siguiente tabla de contingencia (Tabla (1.4)) que muestra los temas de mayor interés que tiene un grupo de 626 personas clasificadas por su nivel educativo, donde el grado de estudios del grupo I se refiere a personas cuyo máximo nivel obtenido es Secundaria terminada, el grupo II tiene como máximo nivel la Preparatoria concluida y el grupo III tiene al menos una carrera universitaria.

INTERESES	GRUPOS			Total
	I	II	III	
Religión	10 (0.333)	14 (0.467)	6 (0.200)	30
Ciencia	36 (0.214)	92 (0.548)	40 (0.238)	168
Música	38 (0.218)	58 (0.333)	78 (0.448)	174
Deportes	24 (0.119)	80 (0.396)	98 (0.485)	202
Política	6 (0.115)	14 (0.269)	32 (0.615)	52
Total	114 (0.182)	258 (0.412)	254 (0.406)	626

Tabla (1.4)

Para determinar si existe alguna relación entre ambas variables se analizan las frecuencias observadas y las esperadas. En la tabla se visualizan los valores observados para las diferentes categorías de las variables cualitativas involucradas y los totales marginales. Los valores entre paréntesis representan la proporción de cada categoría de las

columnas con respecto a la categoría por renglón (por ejemplo, para n_{11} , $0.333 = 10/30$). Así mismo, dividiendo los totales marginales de columna entre el número total de observaciones n , se obtienen las proporciones para cada categoría de la variable columna (por ejemplo, para la columna 1 se tiene $n_{+1}=0.182=114/626$). Estos tres valores (0.182, 0.412 y 0.406) son los valores esperados para las categorías de la variable columna.

A partir de los totales marginales para cada categoría n_{i+} , se obtienen las frecuencias esperadas denotadas por F_{ij} que representarían los valores esperados en cada una de las celdas de la tabla; es decir $F_{ij}=E(n_{ij})$. Así, si hay 52 personas que prefieren los temas de Política, se esperaría que 18.2% de ellas pertenecieran al grupo I, 41.2% al grupo II y 40.6% al grupo III, es decir, 9.47 personas en el grupo I, 21.42 en el grupo II y 21.11 en el grupo III. Estas son las frecuencias esperadas F_{ij} . Es importante mantener las fracciones en el cálculo aunque no sea posible aplicarlas cuando se trata de objetos indivisibles.

Si no hubiera diferencia entre las variables analizadas se esperaría que los valores observados en cada celda difirieran de los valores esperados sólo por razones atribuibles al muestreo o a la aleatoriedad. Al asumir que no hay diferencia, se establece una "hipótesis de independencia" o "hipótesis de homogeneidad"; así, la hipótesis nula H_0 es que no existe asociación entre las variables de la tabla de contingencia, es decir, son independientes.

El cálculo de las frecuencias esperadas para cada celda se basa en el principio de las proporciones. Supóngase que la probabilidad de que un elemento de la población pertenezca a la i -ésima categoría de la variable renglón y a la j -ésima categoría de la variable columna es p_{ij} . La frecuencia esperada para la ij -ésima celda a partir de una muestra de tamaño n (n observaciones independientes) es:

$$F_{ij} = E(n_{ij}) = np_{ij} \quad (1.1)$$

Si la probabilidad de que un elemento pertenezca a la i -ésima categoría de la variable renglón es p_{i+} , la probabilidad de que pertenezca a la j -ésima categoría de la variable columna es p_{+j} y se asume que las variables son independientes, entonces:

$$p_{ij} = p_{i+} p_{+j} \quad (1.2)$$

Por lo tanto, de (1.1):

$$F_{ij} = n p_{i+} p_{+j} \quad (1.3)$$

Las probabilidades p_{i+} y p_{+j} pueden ser estimadas a partir de los valores observados considerando los totales marginales, así, los mejores estimadores para p_{i+} y p_{+j} son:

$$\hat{p}_{i+} = \frac{n_{i+}}{n} \quad \text{y} \quad \hat{p}_{+j} = \frac{n_{+j}}{n} \quad (1.4)$$

donde n_{i+} es el total marginal del renglón i , n_{+j} es el total marginal de la columna j y n es el total de observaciones. Lo anterior permite estimar las frecuencias esperadas en la ij -ésima celda:

$$E_{ij} = n \hat{p}_{i+} \hat{p}_{+j} = n \left(\frac{n_{i+}}{n} \right) \left(\frac{n_{+j}}{n} \right) = \frac{n_{i+} n_{+j}}{n} \quad (1.5)$$

donde E_{ij} es el estimador de las frecuencias esperadas F_{ij} . Esto comprueba del ejemplo anterior que:

$$E_{51} = \frac{(52)(114)}{626} = 9.47, \quad E_{52} = \frac{(52)(258)}{626} = 21.42 \quad \text{y} \quad E_{53} = \frac{(52)(254)}{626} = 21.1$$

Una vez obtenidas las frecuencias esperadas, éstas deben compararse con los valores observados y después establecer si existe independencia o no entre las variables dependiendo de los resultados obtenidos.

ESTADÍSTICO Chi-Cuadrada (χ^2)

Para conocer el grado de asociación entre las variables cualitativas de una tabla de contingencia puede calcularse el Estadístico Chi-cuadrada (χ^2). Éste es utilizado en el Análisis de Correspondencia no sólo por medir la significancia estadística y las discrepancias en una tabla de contingencia sino porque es la base para el desarrollo de uno de los aspectos más relevantes y complejos de este método estadístico desde el punto de vista geométrico: la distancia chi-cuadrada, la cual se analizará más adelante.

El estadístico χ^2 es considerado como una prueba de independencia entre variables, con la cual se determina si las frecuencias observadas que muestran la ocurrencia de valores categóricos de una variable cualitativa permiten rechazar la hipótesis de independencia sobre las frecuencias esperadas o teóricas.

El estadístico χ^2 está dado por:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - E_{ij})^2}{E_{ij}} \quad (1.6)$$

donde n_{ij} es la frecuencia observada de la ij -ésima celda y E_{ij} es la frecuencia esperada de la misma celda. Aquí se observa que su magnitud depende de los valores de las diferencias $(n_{ij}-E_{ij})$ y que puede interpretarse como una distancia entre dichos valores, por lo tanto este estadístico sólo tomará valores positivos o el valor de cero cuando los valores observados sean iguales a los valores esperados.

Entre más grande sea el estadístico, mayor discrepancia hay entre los valores y es una evidencia en contra de la hipótesis de independencia H_0 . De esta forma, para aceptar H_0 se esperarían valores de χ^2 más pequeños que los que se obtendrían si la hipótesis fuera falsa.

La aceptación o rechazo de la hipótesis se determina a través de la prueba chi-cuadrada, en donde se determina la probabilidad de obtener el valor del estadístico χ^2 estimado asumiendo que H_0 es verdadera.

Aunque el estadístico χ^2 asume valores discretos puede aproximarse a una distribución continua (distribución chi-cuadrada). En este caso el estadístico puede compararse con los valores tabulados de dicha distribución (valores críticos). La forma de la distribución depende del número de variables independientes involucradas, es decir, de los **grados de libertad**, denotados por v que son el número de términos independientes en (1.6). En una tabla de contingencia bidimensional, los grados de libertad están dados por el número de categorías de cada variable. No todas las frecuencias en las celdas de la tabla pueden variar libremente. Teniendo una tabla de $r \times c$ (donde r es el número de renglones y c el número de columnas) se tendrían rc términos (frecuencias), sin embargo, de éstos, algunos ya están determinados por los totales marginales que son fijos y se conocen. Así, si se tienen las frecuencias de las primeras celdas de cada renglón de la tabla, el valor de la última celda de cada renglón se determina automáticamente. Al conocer el valor de r , los términos independientes se reducen a $rc-r$, es decir, se elimina un término por cada renglón. Lo mismo sucede con las columnas, pero en este caso, ya se ha eliminado una de ellas a

través de los r términos determinados en los renglones, por lo que en realidad se fijan $c-1$ términos. Por lo tanto, el número de términos independientes es:

$$\begin{aligned} &= rc-r-(c-1) \\ &= r(c-1)-(c-1) \\ &= (c-1)(r-1) \end{aligned}$$

Así, los grados de libertad de una tabla de contingencia de rc es el número de renglones menos uno multiplicado por el número de columnas menos uno:

$$\text{g.l.} = v = (r-1)(c-1)$$

La figura (1.4) muestra algunas formas que adquiere la distribución dependiendo de los grados de libertad.

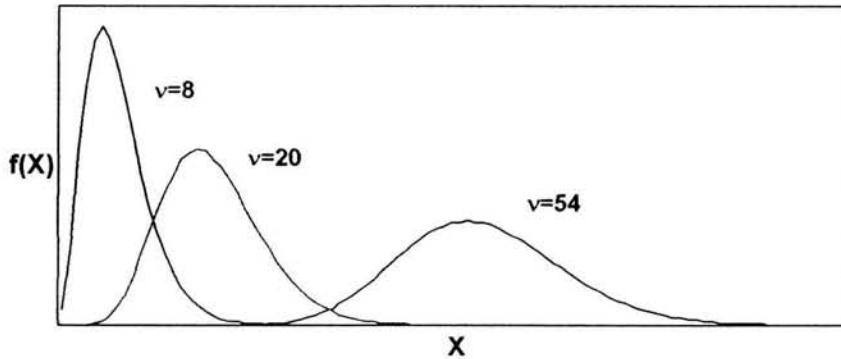


Fig. (1.4) Distribución χ^2

El valor calculado de χ^2 se compara con el valor crítico dado por las tablas de la distribución (denotado por $\chi^2_{v,\alpha}$) para v grados de libertad y un nivel de significancia α dado (generalmente 0.01 ó 0.05) que indica la probabilidad de que la relación entre las variables ocurra sólo por azar y que, en realidad, tal relación no exista en la población. Por lo tanto, entre mayor sea el valor α , la confiabilidad de que dicha relación exista disminuye.

La hipótesis nula será rechazada si el valor calculado χ^2 es mayor que el valor crítico $\chi^2_{v,\alpha}$, ya que esto indica que es muy poco probable (menos de $100\alpha\%$) que el valor calculado de χ^2 ocurra sólo por azar y, por lo tanto, las variables no se consideren independientes.

Por ejemplo, para la tabla (1.3), el valor χ^2 es:

$$\chi^2 = \frac{(20-40)^2}{40} + \frac{(60-40)^2}{40} + \frac{(80-60)^2}{60} + \frac{(40-60)^2}{60} = 33.33$$

El valor crítico para un nivel de significancia $\alpha=.05$ y un grado de libertad es $\chi^2_{1}=3.841$, por lo que $\chi^2 > \chi^2_v$, concluyendo que la actitud de acuerdo o desacuerdo es dependiente o contingente del género.

En el caso de la tabla (1.4) el valor $\chi^2=49.034$ excede el valor crítico $\chi^2_{8}=15.507$ para 8 grados de libertad y un nivel de significancia $\alpha=.05$, por lo que la hipótesis nula se rechaza y se concluye que existe dependencia entre las variables, o bien, que existe algún tipo de relación entre el nivel educativo de una persona y el tipo de temas que le son de mayor interés.

La aproximación de la distribución χ^2 es muy satisfactoria cuando las frecuencias esperadas son mayores o iguales a 5, aunque aún con frecuencias más pequeñas también se obtienen resultados aceptables. Un aspecto esencial en el uso del estadístico χ^2 es que se está tratando con frecuencias y no con porcentajes por lo que los resultados en forma de porcentajes o proporciones deben ser convertidas a números absolutos para realizar el análisis. Además las frecuencias de las celdas deben ser independientes en el sentido de que una observación no debe aparecer en más de una celda; éstas deben ser mutuamente exclusivas y exhaustivas con respecto a los posibles valores de las variables.

Las tablas de contingencia pueden aplicarse a análisis donde intervengan más de dos variables (Análisis de Correspondencia Múltiple). Cuando esto sucede, la introducción de una nueva variable a una tabla de contingencia (bivariada) puede resultar en varias posibilidades: confirmar y reforzar la asociación observada entre las variables originales, indicar que dicha asociación no es verdadera, que la asociación inicial no cambia aún con la nueva variable involucrada o revelar información que no fue detectada en un principio.

CAPÍTULO II

FUNDAMENTOS MATEMÁTICOS DEL ANÁLISIS MULTIVARIADO

El Análisis de Correspondencia es uno de los métodos que conforman el análisis multivariado. Estos métodos estadísticos pueden ser usados para muchos propósitos de distintas disciplinas a través del análisis de datos, encontrando relaciones entre múltiples variables con el objetivo de descubrir asociaciones entre ellas y medir la naturaleza y el grado de estas asociaciones. Medir la variación entre un conjunto de variables es lo que, en gran medida, permite comprender el comportamiento de los datos, por lo que es muy frecuente el cálculo de medidas que dan información sobre éstos tales como medias, varianzas o covarianzas. El análisis multivariado ha sido aplicado en diversas disciplinas, sobre todo en aquellas relacionadas con las ciencias exactas, aunque sus aplicaciones en las ciencias administrativas y del comportamiento han ido adquiriendo cada vez mayor importancia; en ellas el interés es encontrar la interdependencia entre las variables e interpretar la información que el análisis provee.

Los métodos estadísticos multivariados que pueden trabajar con variables que tienen una escala ordinal o nominal son de más reciente aplicación pero, en general, todos los métodos para el análisis multivariado involucran aspectos matemáticos similares. Particularmente, se fundamentan en aspectos de Álgebra Lineal y Geometría, además de recurrir a conceptos de Estadística. Los métodos multivariados están estructurados por un conjunto de procedimientos organizados que pueden aplicarse a problemas específicos con ciertas características. Estudiar estos métodos, significa estudiar transformaciones lineales y realizar operaciones relacionadas con vectores, matrices, eigenestructuras y algunas formas de descomposición de matrices, siendo éstos los aspectos más relevantes que se describen en este capítulo. Los fundamentos del Análisis de Correspondencia, y de otros métodos multivariados, se desarrollan a partir de definiciones básicas hasta llegar a procedimientos más complejos que, finalmente, permiten obtener una representación numérica y gráfica que facilitan la interpretación de los resultados.

Aunque los principios básicos son los mismos, el Análisis de Correspondencia se distingue de los demás métodos por el tipo de escala de las variables involucradas donde se considera el manejo de datos categóricos.

ESPACIO EUCLIDIANO Y BASES DEL ESPACIO

Uno de los aspectos importantes del análisis multivariado es que las operaciones que se realizan en términos numéricos (algebraicos) pueden representarse geoméricamente, lo que permite entender el significado de tales operaciones. El “lugar” sobre el cual se representan diferentes conceptos como puntos, distancias, vectores, ángulos, entre otros, es el **espacio Euclidiano**. Un espacio Euclidiano de dimensión finita n es la colección de todos los vectores de n componentes para los cuales las operaciones de suma y multiplicación de vectores están definidas y, además, para dos vectores cualesquiera en el espacio existe un número no negativo llamado **distancia Euclidiana** entre vectores.

Dimensión del Espacio Vectorial

La **independencia lineal** está relacionada con el número mínimo de vectores con los cuales cualquier vector del espacio puede expresarse como combinación lineal de ellos. Un conjunto de p vectores $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_p$ ¹ es linealmente independiente si en la ecuación lineal:

$$k_1\mathbf{a}_1' + k_2\mathbf{a}_2' + \dots + k_p\mathbf{a}_p' = \mathbf{0}$$

la única solución para los escalares k_1, k_2, \dots, k_p es la solución trivial, es decir, $k_1 = k_2 = \dots = k_p = 0$.

Un conjunto de vectores linealmente independientes determina el “volumen” (dimensión) del espacio expandido por ellos. Si los vectores son linealmente dependientes entonces existe redundancia ya que al menos uno de ellos puede escribirse como combinación lineal de los otros. En general, si se tiene un conjunto de p vectores de n componentes y $p > n$, entonces el conjunto siempre es linealmente dependiente, si $p < n$ entonces los vectores no son suficientes para definir un espacio de n dimensiones. La **dimensión** de un espacio vectorial es igual al número máximo de vectores linealmente independientes en ese espacio.

Base Ortogonal

Cualquier conjunto de n vectores linealmente independientes en un espacio de dimensión n donde cualquier otro vector α de ese espacio puede expresarse como combinación lineal de ellos constituye una **base** de ese espacio, es decir:

¹ Se asume que un vector \mathbf{a} es un vector columna $\mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$; \mathbf{a}' es el vector transpuesto.

$$\alpha = k_1 \mathbf{a}_1 + k_2 \mathbf{a}_2 + \dots + k_n \mathbf{a}_n$$

donde el conjunto de vectores $\{\mathbf{a}_i\}$ es una base del espacio n -dimensional.

La **base estándar** o natural de un espacio de dimensión n es el conjunto de n vectores que se "alinean" sobre las coordenadas:

$$\mathbf{e}_1 = (1, 0, \dots, 0)$$

$$\mathbf{e}_2 = (0, 1, \dots, 0)$$

$$\vdots$$

$$\mathbf{e}_n = (0, 0, \dots, 1)$$

Los vectores de una base estándar $\{\mathbf{e}_i\}$ son mutuamente ortogonales (es decir, el producto interno de cualquier par de ellos es igual a cero) y son unitarios (la longitud de cada vector es igual a 1)². La base estándar también se denota como E .

Si cualquier par de vectores renglón y vectores columna de una base son mutuamente ortogonales y además los vectores son unitarios, la base es **ortogonal** (algunos autores la llaman base **ortonormal**).

CAMBIO DE LA BASE DEL ESPACIO

Los conceptos que se han mencionado son la base para describir los principios y procedimientos que se llevan a cabo en el análisis multivariado. Como se verá posteriormente, el resultado de estos procedimientos podrá ser interpretado, bajo las condiciones que el Análisis de Correspondencia establece, de una forma relativamente sencilla y de acuerdo a las características del problema que haya sido planteado. En los temas siguientes, además del fundamento algebraico, se pondrá énfasis en la descripción geométrica, sintetizando visualmente la información generada por el método.

² La longitud de un vector \mathbf{a} se define como $\|\mathbf{a}\| = \left[\sum_{i=1}^n a_i^2 \right]^{1/2}$

El producto interno (o escalar) de dos vectores \mathbf{a} y \mathbf{b} se define como $\mathbf{a} \cdot \mathbf{b} = \sum_{k=1}^n (a_k b_k) = \|\mathbf{a}\| \|\mathbf{b}\| \cos \theta_{ab}$

Usualmente, cuando se hace referencia a un espacio vectorial de n dimensiones se concibe un espacio cuyos ejes son perpendiculares entre sí y la base con la cual se expresan las coordenadas o los vectores es la base estándar $\{\mathbf{e}_i\}$. Por ejemplo, en un espacio de dos dimensiones se considera un plano cartesiano donde los ejes de las abscisas y de las ordenadas son perpendiculares (ortogonales) y los vectores pueden escribirse como combinación lineal de la base estándar E ; es decir, un vector α puede representarse como:

$$\alpha = a_1 \mathbf{e}_1 + a_2 \mathbf{e}_2$$

$$\alpha = a_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + a_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

donde los escalares a_1 y a_2 son los componentes (coordenadas) del vector α relativos a la base estándar E .

Los vectores referidos a una base dada pueden también ser referidos a cualquier otra base del espacio. Por ejemplo, supóngase que se tiene un vector \mathbf{a} representado en el espacio vectorial definido por la base estándar $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ y $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, con las coordenadas $\begin{bmatrix} .9 \\ .5 \end{bmatrix}$, es decir:

$$\mathbf{a} = a_1 \mathbf{e}_1 + a_2 \mathbf{e}_2$$

$$= 0.9 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 0.5 \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

El mismo vector \mathbf{a} puede también ser expresado en términos de otra base F , donde los vectores \mathbf{f}_1 y \mathbf{f}_2 son combinaciones lineales de la base E :

$$\mathbf{f}_1 = c_1 \mathbf{e}_1 + c_2 \mathbf{e}_2$$

$$\mathbf{f}_2 = d_1 \mathbf{e}_1 + d_2 \mathbf{e}_2$$

donde c_1 , c_2 , d_1 y d_2 son escalares. Dando valores a $c_1=0.707$, $c_2=0.707$, $d_1=0.940$ y $d_2=0.342$, entonces:

$$\mathbf{f}_1 = 0.707 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 0.707 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} .707 \\ .707 \end{bmatrix}$$

$$\mathbf{f}_2 = 0.940 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 0.342 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} .940 \\ .342 \end{bmatrix}$$

Los vectores f_1 y f_2 son linealmente independientes y son unitarios, pero no son ortogonales ya que el producto interno $f_1 \cdot f_2$ no es cero, por lo que el ángulo que forman ambos vectores no es de 90° , como sucede en la base E; de este modo, f_1 y f_2 constituyen ejes cartesianos oblicuos. La figura (2.1) muestra el vector a referido a ambas bases.

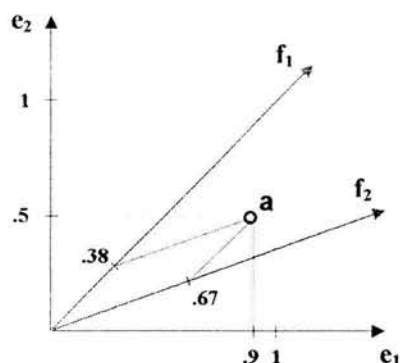


Fig. (2.1) Vector referido a dos bases distintas del espacio bidimensional

Si el vector a es referido a la nueva base F, entonces la expresión para el vector es:

$$a = a_1^* f_1 + a_2^* f_2$$

donde los escalares (coordenadas) a_1^* y a_2^* son relativas a los vectores f_1 y f_2 respectivamente. En este caso, los valores de a_1^* y a_2^* se hallarían resolviendo:

$$\begin{aligned} a &= a_1^* f_1 + a_2^* f_2 = a_1 e_1 + a_2 e_2 \\ &= a_1^* \begin{bmatrix} .707 \\ .707 \end{bmatrix} + a_2^* \begin{bmatrix} .940 \\ .342 \end{bmatrix} = \begin{bmatrix} .9 \\ .5 \end{bmatrix} \end{aligned}$$

y se tiene que $a_1^* = 0.38$ y $a_2^* = 0.67$, lo cual se observa en la figura extendiendo líneas paralelas a los ejes f_1 y f_2 .

De la misma forma, si a estuviera referido a la base F y se deseara encontrar las coordenadas en términos de la base original E entonces, como f_1 y f_2 están definidos en términos de e_1 y e_2 :

$$\begin{aligned} a &= a_1^* f_1 + a_2^* f_2 \\ &= a_1^* (c_1 e_1 + c_2 e_2) + a_2^* (d_1 e_1 + d_2 e_2) \\ &= (a_1^* c_1 + a_2^* d_1) e_1 + (a_1^* c_2 + a_2^* d_2) e_2 \end{aligned}$$

Si en términos de E las coordenadas son a_1 y a_2 , entonces, sustituyendo $(a_1^* c_1 + a_2^* d_1)$ por a_1 y $(a_1^* c_2 + a_2^* d_2)$ por a_2 , se tiene:

$$a = a_1 e_1 + a_2 e_2$$

donde:

$$a_1 = 0.38(0.707) + 0.67(0.940) = 0.9$$

$$a_2 = 0.38(0.707) + 0.67(0.342) = 0.5$$

representan las coordenadas originales del vector \mathbf{a} relativas a \mathbf{e}_1 y \mathbf{e}_2 y que son las proyecciones de \mathbf{a} sobre esos ejes.

Así, un vector descrito en términos de una base también puede describirse en términos de otra base; lo importante es conocer la forma en que los vectores de ambas bases están conectados. En este ejemplo la relación de los vectores de F con los de E está dada por medio de:

$$\mathbf{f}_1 = 0.707\mathbf{e}_1 + 0.707\mathbf{e}_2$$

$$\mathbf{f}_2 = 0.940\mathbf{e}_1 + 0.342\mathbf{e}_2$$

donde \mathbf{f}_1 y \mathbf{f}_2 son oblicuos.

TRANSFORMACIONES ORTOGONALES

La relación establecida entre los vectores de dos bases distintas E y F se determina mediante las combinaciones lineales:

$$\mathbf{f}_1 = c_1\mathbf{e}_1 + c_2\mathbf{e}_2$$

$$\mathbf{f}_2 = d_1\mathbf{e}_1 + d_2\mathbf{e}_2$$

donde la transformación de los vectores de la base E depende de los valores c_1 , c_2 , d_1 y d_2 , es decir, la matriz T:

$$T = \begin{bmatrix} c_1 & c_2 \\ d_1 & d_2 \end{bmatrix}$$

es la **matriz de transformación** del conjunto de vectores \mathbf{e}_1 y \mathbf{e}_2 . Los vectores resultantes \mathbf{f}_1 y \mathbf{f}_2 están dados por:

$$\mathbf{f}_1 = c_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$$

$$\mathbf{f}_2 = d_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + d_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$$

Arreglando en forma matricial lo anterior, se tiene:

$$F = ET = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} c_1 & d_1 \\ c_2 & d_2 \end{bmatrix} = \begin{bmatrix} \mathbf{f}_1 & \mathbf{f}_2 \\ c_1 & d_1 \\ c_2 & d_2 \end{bmatrix}$$

donde las columnas de F son los vectores de la nueva base en términos de E , obtenidos bajo la matriz de transformación T y T' es la matriz transpuesta de T ; de lo anterior se tiene que $F=T'$, por lo que T' representa la nueva base³. En el ejemplo anterior, los vectores de F son unitarios, pero no ortogonales, es decir, los nuevos ejes son oblicuos.

ROTACIÓN DE LOS VECTORES DE LA BASE

Existe un caso especial donde al obtener una nueva base se conservan los ángulos, las magnitudes y las distancias entre los puntos o vectores. Las transformaciones lineales que permiten mantener estas relaciones son las **transformaciones ortogonales**, o bien, **rotaciones**, en donde la matriz de transformación A es ortogonal y tiene la propiedad:

$$A'A=AA'=I$$

donde A' es la matriz transpuesta de A e I es la matriz identidad.

El efecto geométrico de una matriz de transformación A es una rotación de los ejes de la base original, ya sea en el sentido de las manecillas del reloj o en el sentido contrario, con un ángulo θ determinado.

Para que un vector (eje) pueda ser definido en un espacio de dos dimensiones, debe cumplirse que:

$$\cos^2\alpha + \cos^2\beta = 1$$

donde α y β son los ángulos directores y la suma de los cuadrados de los cosenos directores debe ser igual a 1. Esto aplica para cualquier espacio de dimensión n .

Supóngase que se tiene un vector $a'=(1 \ 2)$ sobre los ejes x y y como se ve en la figura (2.2). El vector a' genera un ángulo α con el eje x y otro ángulo β con el eje y .

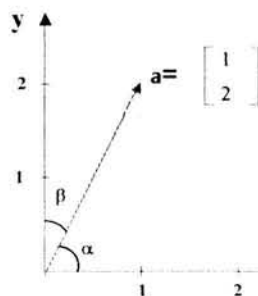


Fig. (2.2) Ángulos directores de un vector

³ En general, se representa por M' a la matriz transpuesta de M

La magnitud de \mathbf{a} es $\|\mathbf{a}\|=\sqrt{5}$, por lo que los cosenos directores son:

$$\cos \alpha = 1/\sqrt{5}, \quad \cos \beta = 2/\sqrt{5}$$

correspondientes a los ángulos $\alpha=63.43^\circ$ y $\beta=26.56^\circ$ que determinan la posición del vector \mathbf{a} .

En el caso de las rotaciones, el ángulo que forman los ejes originales (denotados por i) con los ejes nuevos (denotados por j) es el ángulo θ_{ij} . Por ejemplo, si se quieren rotar los ejes \mathbf{e}_1 y \mathbf{e}_2 con un ángulo de 45° en el sentido contrario a las manecillas del reloj, representado por el símbolo (\uparrow), entonces:

θ_{11} involucra una rotación de 45° en sentido (\uparrow), donde $\cos 45^\circ=0.707$

θ_{12} involucra una rotación de 135° en sentido (\uparrow), donde $\cos 135^\circ=-0.707$

θ_{21} involucra una rotación de 45° en sentido (\downarrow), donde $\cos 45^\circ=0.707$

θ_{22} involucra una rotación de 45° en sentido (\uparrow), donde $\cos 45^\circ=0.707$

Lo anterior se muestra en la figura (2.3) con las rotaciones descritas.

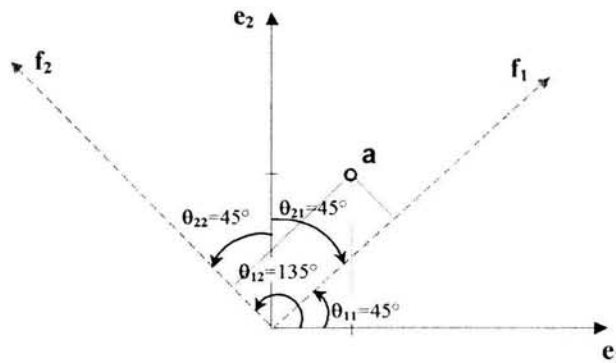


Fig. (2.3) Rotación de ejes

Los nuevos ejes \mathbf{f}_1 y \mathbf{f}_2 también son una combinación lineal de \mathbf{e}_1 y \mathbf{e}_2 , dada por los cosenos de los ángulos que forman entre ellos, es decir:

$$\mathbf{f}_1 = \cos \theta_{11} \mathbf{e}_1 + \cos \theta_{21} \mathbf{e}_2 = 0.707 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 0.707 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} .707 \\ .707 \end{bmatrix}$$

$$\mathbf{f}_2 = \cos \theta_{12} \mathbf{e}_1 + \cos \theta_{22} \mathbf{e}_2 = -0.707 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 0.707 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -.707 \\ .707 \end{bmatrix}$$

Los ejes f_1 y f_2 están expresados en términos de e_1 y e_2 , la base original, y son, además, las coordenadas que los representan. Esta es la rotación obtenida con la matriz de transformación A, análoga a la matriz T mencionada anteriormente:

$$A = \begin{bmatrix} \cos\theta_{11} & \cos\theta_{21} \\ \cos\theta_{12} & \cos\theta_{22} \end{bmatrix} = \begin{bmatrix} .707 & .707 \\ -.707 & .707 \end{bmatrix}$$

por lo que:

$$F = EA' = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} .707 & -.707 \\ .707 & .707 \end{bmatrix}$$

Ahora supóngase que se tiene el vector $a'=(1 \ 2)$ referido a la base estándar e_1 y e_2 :

$$a' = 1(1 \ 0) + 2(0 \ 1) = (1 \ 2)$$

Las nuevas coordenadas de este vector, referidas a la base F y que se denotan por a_1° y a_2° , están dadas por la transformación:

$$\begin{bmatrix} a_1^\circ \\ a_2^\circ \end{bmatrix} = \begin{bmatrix} \cos\theta_{11} & \cos\theta_{21} \\ \cos\theta_{12} & \cos\theta_{22} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} .707 & .707 \\ -.707 & .707 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2.12 \\ .707 \end{bmatrix}$$

obteniendo el vector a° con coordenadas $a_1^\circ=2.121$ y $a_2^\circ=0.707$, referidas a los ejes f_1 y f_2 respectivamente, y se obtuvieron premultiplicando la matriz de transformación A por el vector a, es decir:

$$a^\circ = Aa$$

En general, si se obtiene una nueva base y las coordenadas de un punto fijo x relativas a la base original son transformadas, el punto transformado con coordenadas relativas a la nueva base se denotará como x° .

Dado que la matriz de transformación A genera una rotación, entonces A debe ser ortogonal y cumplir con:

$$A'A = \begin{bmatrix} .707 & -.707 \\ .707 & .707 \end{bmatrix} \begin{bmatrix} .707 & .707 \\ -.707 & .707 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$$

$$AA' = \begin{bmatrix} .707 & .707 \\ -.707 & .707 \end{bmatrix} \begin{bmatrix} .707 & -.707 \\ .707 & .707 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$$

y también cumple con los cosenos directores:

$$\cos^2 \theta_{11} + \cos^2 \theta_{21} = (0.707)^2 + (0.707)^2 = 1$$

$$\cos^2 \theta_{12} + \cos^2 \theta_{22} = (-0.707)^2 + (0.707)^2 = 1$$

Características de la Transformación de la Base

En particular, las matrices de transformación que generan una rotación de los vectores de la base poseen algunas características especiales que ya han sido descritas. Por ejemplo, si en el caso de la base estándar para un espacio de dos dimensiones, se aplica una transformación representada por la composición lineal:

$$\mathbf{f}_1 = t_{11}\mathbf{e}_1 + t_{21}\mathbf{e}_2$$

el vector obtenido \mathbf{f}_1 es la resultante de sumar el vector $\mathbf{e}_1' = (1 \ 0)$ multiplicado por el escalar t_{11} más el vector $\mathbf{e}_2' = (0 \ 1)$ multiplicado por el escalar t_{21} , (Figura (2.4)).

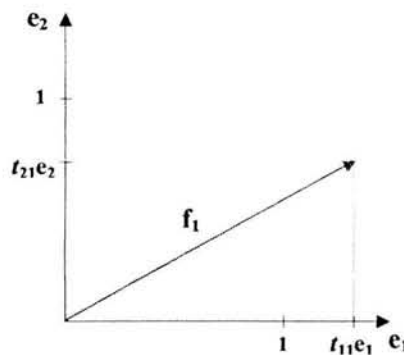


Fig. (2.4) Transformación de un eje

Como los vectores $\{\mathbf{e}_i\}$ son una base ortogonal, su rotación requiere hallar un segundo vector \mathbf{f}_2 que sea ortogonal a \mathbf{f}_1 para garantizar que las magnitudes, distancias y ángulos originales entre vectores se conserven aún después de la transformación. Así, para el vector $\mathbf{f}_2 = t_{12}\mathbf{e}_1 + t_{22}\mathbf{e}_2$, el vector $\mathbf{t}_2' = (t_{12} \ t_{22})$ debe ser ortogonal al vector $\mathbf{t}_1' = (t_{11} \ t_{21})$ para generar la nueva base ortogonal F a partir de T (los valores de t_{ij} son cosenos directores).

$$F = ET' = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{bmatrix}$$

donde T' contiene los vectores \mathbf{t}_1 y \mathbf{t}_2 .

Rotación Impropia

El tipo de rotación que genera una matriz A se conoce como **rotación propia** o **rígida**, sin embargo, no todas las matrices ortogonales originan este tipo de rotación. Si se modificara la matriz A y se multiplicara uno de sus renglones por -1 , entonces se tendría la siguiente matriz B:

$$\begin{bmatrix} \cos(\theta)_{11} & \cos(\theta)_{21} \\ -\cos(\theta)_{12} & -\cos(\theta)_{22} \end{bmatrix}$$

Esta matriz también es ortogonal y genera una rotación, la diferencia es que mientras el nuevo eje f_1 rota de la misma manera, el eje f_2 es reflejado después de rotar, como se muestra en la figura (2.5):

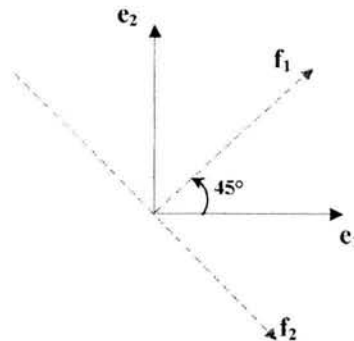


Fig. (2.5) Rotación impropia

Cuando un número impar de renglones de la matriz de transformación es multiplicado por -1 la rotación obtenida es una **rotación impropia**.

El tipo de rotación que produce una matriz de transformación A puede conocerse a través de su determinante ya que éste siempre será igual a 1 ó -1 . Si $|A|=1$ la rotación es propia, si $|A|=-1$ se genera una rotación impropia.

ROTACIÓN DE PUNTOS

El tipo de transformación lineal que se ha explicado involucra una rotación de ejes a partir de una matriz de transformación A , donde A es ortogonal. Sin embargo, existe un procedimiento alternativo que genera los mismos resultados: que la base original no se transforme en una nueva para referir los vectores o puntos a esa nueva base, sino que sean los puntos los que se roten manteniendo fija la base original. En el ejemplo de la matriz A , se rotaron los ejes e_1 y e_2 un ángulo de 45° en el sentido contrario de las manecillas del reloj y posteriormente se encontraron las coordenadas del vector a referidas a la nueva base F . Otra forma de obtener esto es rotar el vector a con un ángulo de 45° en el sentido de las manecillas del reloj. La figura (2.6) muestra gráficamente la rotación del punto.

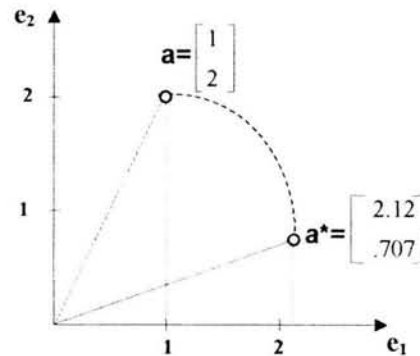


Fig. (2.6) Rotación de un punto sobre la misma base

donde $a^*=Aa$ está referido a la base original E . En este caso, dado que A es ortogonal, la rotación de puntos se obtiene de la misma forma que la rotación de ejes. En general, un punto x transformado por medio de una matriz ya sea ortogonal o no, es representado por x^* .

Generalización de las Transformaciones

Para resumir lo anterior, las transformaciones de los vectores de la base se refieren a transformar la base original $\{e_i\}$ en un nuevo conjunto de vectores $\{f_i\}$, el cual es expresado en términos de $\{e_i\}$, mientras que las coordenadas de un punto x son ahora expresadas como x° en términos de la nueva base $\{f_i\}$. En la figura (2.7) se muestran las generalizaciones de estas transformaciones cuando se trata de matrices de transformación ortogonales.

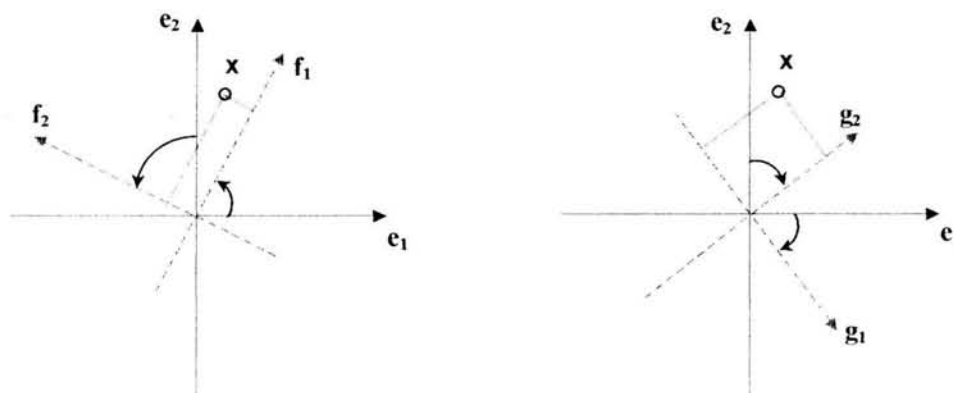


Fig. (2.7) Transformaciones de la base (rotaciones en ambos sentidos)

En el caso de las matrices no ortogonales (T) las posibles transformaciones están dadas por:

- *Transformación de Puntos* (donde \mathbf{x} es el vector original y \mathbf{x}^* es el vector transformado)

$$\mathbf{x}^* = T\mathbf{x} \quad \text{con los vectores } \mathbf{x}^* \text{ y } \mathbf{x} \text{ en términos de } E$$

$$\mathbf{x} = T^{-1}\mathbf{x}^* \quad \text{con los vectores } \mathbf{x} \text{ y } \mathbf{x}^* \text{ en términos de } E^4$$

- *Transformación de Vectores de la Base* (donde \mathbf{x} es el vector de la base original, \mathbf{x}° es el vector transformado y $F=ET'$)

$$\mathbf{x}^\circ = (T')^{-1}\mathbf{x} \quad \text{con el vector } \mathbf{x}^\circ \text{ en términos de } F \text{ y el vector } \mathbf{x} \text{ en términos de } E$$

$$\mathbf{x} = T'\mathbf{x}^\circ \quad \text{con el vector } \mathbf{x} \text{ en términos de } E \text{ y el vector } \mathbf{x}^\circ \text{ en términos de } F$$

En el caso especial de matrices ortogonales A, se tiene:

- *Transformación de Puntos*

$$\mathbf{x}^* = A\mathbf{x} \quad \text{con los vectores } \mathbf{x}^* \text{ y } \mathbf{x} \text{ en términos de } E$$

$$\mathbf{x} = A'\mathbf{x}^* \quad \text{con los vectores } \mathbf{x} \text{ y } \mathbf{x}^* \text{ en términos de } E$$

- *Transformación de Vectores de la Base* (donde $F=EA'$)

$$\mathbf{x}^\circ = A\mathbf{x} \quad \text{con el vector } \mathbf{x}^\circ \text{ en términos de } F \text{ y el vector } \mathbf{x} \text{ en términos de } E$$

$$\mathbf{x} = A'\mathbf{x}^\circ \quad \text{con el vector } \mathbf{x} \text{ en términos de } E \text{ y el vector } \mathbf{x}^\circ \text{ en términos de } F$$

En los ejemplos vistos se ha trabajado con espacios de dos dimensiones. Las matrices de transformación T y A, han sido, por lo tanto, matrices de 2x2, lo que simplifica las operaciones realizadas. Sin embargo, todo lo que se ha descrito puede extenderse al caso de más de dos dimensiones. Las rotaciones pueden realizarse usando la matriz de cosenos directores apropiada donde, por ejemplo, en el caso de 3 dimensiones, las coordenadas de un punto a en la base original pueden expresarse como \mathbf{a}° en la nueva base a través de:

$$\begin{bmatrix} a_1^\circ \\ a_2^\circ \\ a_3^\circ \end{bmatrix} = \begin{bmatrix} \cos\theta_{11} & \cos\theta_{21} & \cos\theta_{31} \\ \cos\theta_{12} & \cos\theta_{22} & \cos\theta_{32} \\ \cos\theta_{13} & \cos\theta_{23} & \cos\theta_{33} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

Las matrices de transformación ortogonales tienen un papel central en el análisis multivariado por sus propiedades especiales. En el Análisis de Correspondencia se hace referencia a estas matrices para encontrar un resultado óptimo en el problema planteado.

⁴ Se denota por T^{-1} a la matriz inversa de T

MATRICES DE TRANSFORMACIÓN SOBRE LA NUEVA BASE

Así como se mostró que una matriz de transformación T mapea un vector \mathbf{x} en un vector \mathbf{x}^* , ambos en términos de E , se mostrará cómo puede hallarse una nueva matriz de transformación T° sobre una nueva base F que mapea un vector \mathbf{x}° en otro vector $\mathbf{x}^{*\circ}$, ambos en términos de F . Se ejemplificará con un caso en el que la base F es oblicua y, posteriormente, se mencionará el caso especial para bases ortogonales.

Supóngase que se tiene una matriz T dada por:

$$T = \begin{bmatrix} .87 & .50 \\ .34 & .94 \end{bmatrix}$$

y una matriz L que transformna a los vectores de la base E en una nueva base oblicua F , ya que L contiene cosenos directores pero el producto interno de los renglones no es cero:

$$L = \begin{bmatrix} .95 & .33 \\ .50 & .87 \end{bmatrix}$$

La base F está dada por:

$$F = EL' = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} .95 & .50 \\ .33 & .87 \end{bmatrix} = \begin{bmatrix} .95 & .50 \\ .33 & .87 \end{bmatrix}$$

La transformación de un punto \mathbf{x} a través de T , relativa a la base E es:

$$\mathbf{x}^* = T\mathbf{x}$$

Si $\mathbf{x}' = (1 \ 2)$, entonces el punto transformado en términos de E es:

$$\mathbf{x}^* = \begin{bmatrix} .87 & .50 \\ .34 & .94 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1.87 \\ 2.22 \end{bmatrix}$$

En el panel I de la figura (2.8) se muestra la transformación. Ahora, se desea encontrar una matriz T° que transforme al punto $\mathbf{x}^{\circ} = (1 \ 2)$ referido a la base F en el punto $\mathbf{x}^{*\circ}$ referido también a F , es decir:

$$\mathbf{x}^{*\circ} = T^\circ \mathbf{x}^\circ$$

como se muestra en el panel II de la figura (2.8).

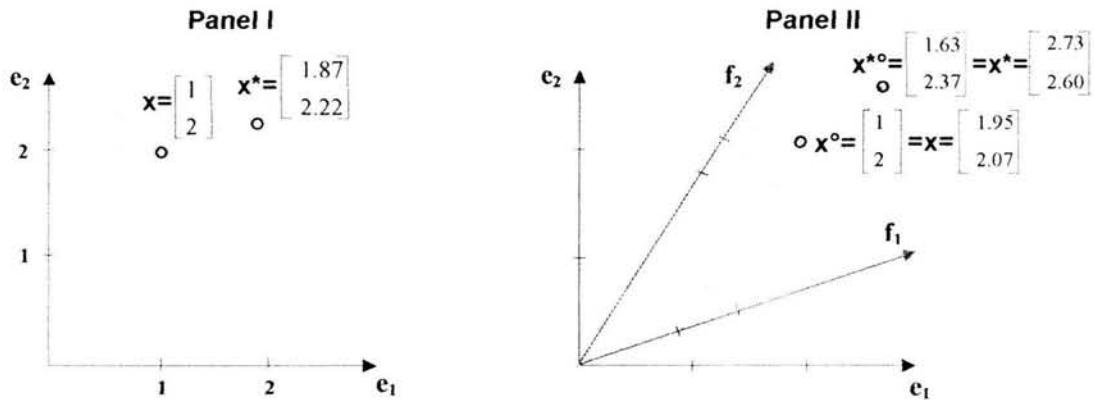


Fig. (2.8) Transformación de puntos en bases distintas

Para encontrar T° es necesario:

1. Mapear x° en x por medio de L' (donde x es relativo a la base E):

$$L' \quad x^\circ \quad x \\ \begin{bmatrix} .95 & .50 \\ .33 & .87 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1.95 \\ 2.07 \end{bmatrix}$$

2. Mapear x en x^* por medio de T (donde x^* es relativo a la base E):

$$T \quad x \quad x^* \\ \begin{bmatrix} .87 & .50 \\ .34 & .94 \end{bmatrix} \begin{bmatrix} 1.95 \\ 2.07 \end{bmatrix} = \begin{bmatrix} 2.73 \\ 2.60 \end{bmatrix}$$

3. Mapear x^* en $x^{*\circ}$ por medio de $(L')^{-1}$ (donde $x^{*\circ}$ es relativo a la base F):

$$(L')^{-1} \quad x^* \quad x^{*\circ} \\ \begin{bmatrix} 1.32 & -.76 \\ -.50 & 1.44 \end{bmatrix} \begin{bmatrix} 2.73 \\ 2.60 \end{bmatrix} = \begin{bmatrix} 1.63 \\ 2.37 \end{bmatrix}$$

Las transformaciones aplicadas al punto x° para obtener su imagen $x^{*\circ}$, ambos referidos a F, conforman el producto que constituye la matriz de transformación T° sobre la nueva base, por lo tanto, la matriz T° es:

$$T^\circ = (L')^{-1}TL'$$

es decir:

$$T^\circ = (L')^{-1} \quad T \quad L' \\ \begin{bmatrix} 1.32 & -.76 \\ -.50 & 1.44 \end{bmatrix} \begin{bmatrix} .87 & .50 \\ .34 & .94 \end{bmatrix} \begin{bmatrix} .95 & .50 \\ .33 & .87 \end{bmatrix} = \begin{bmatrix} .83 & .40 \\ .41 & .98 \end{bmatrix}$$

por lo que el procedimiento anterior para hallar el punto transformado se reduce a:

$$x^{*\circ} = T^\circ x^\circ = \begin{bmatrix} .83 & .40 \\ .41 & .98 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1.63 \\ 2.37 \end{bmatrix}$$

El punto \mathbf{x}° referido a F es equivalente al punto \mathbf{x} relativo a E, si a éste se le aplica la transformación T, entonces \mathbf{x}^* equivale al resultado de aplicar T° a \mathbf{x}° , es decir, \mathbf{x}^* relativo a F. Si T transforma a \mathbf{x} en \mathbf{x}^* en la base E, T° es la matriz que transforma a \mathbf{x}° en \mathbf{x}^* relativos a F, con resultados equivalentes.

La razón de encontrar T° es obtener una matriz de transformación con respecto a una nueva base F que realice las transformaciones equivalentes a las que realiza T sobre la base E. Si L fuera ortogonal, entonces se tiene el caso especial para la inversa de L':

$$(L')^{-1}=L$$

por lo tanto:

$$T^\circ=LTL'$$

EFFECTOS GEOMÉTRICOS DE TRANSFORMACIONES

Las operaciones matriciales, como las transformaciones vistas, pueden ser representadas geoméricamente cuando se tienen hasta tres dimensiones; pero los principios se conservan en los espacios de cualquier dimensión. Las operaciones a las que se hará referencia a continuación permitirán comprender lo que sucede geoméricamente al realizar la descomposición de matrices, un aspecto que es determinante para obtener los resultados tanto numéricos como geoméricos en el Análisis de Correspondencia. Estas operaciones ya no se refieren a la transformación de un sólo vector \mathbf{x} , sino a la transformación de un conjunto de ellos arreglados en forma matricial (matriz X).

Se ha establecido que un vector dado \mathbf{x} es un vector columna, sin embargo, en la matriz X, cada renglón representa un vector, por lo que para obtener el mismo resultado, se transpone la matriz de transformación A y se postmultiplica por la matriz X. Considérese una matriz original X con vectores en dos dimensiones, denotados por letras de la a a la i. A esta matriz se le aplicarán las transformaciones de translación, reflexión, ajuste (o reescalamiento) y rotación, representadas por una matriz A que presenta formas especiales que a menudo se encuentran en el análisis multivariado. Las transformaciones, en forma general, están dadas por $X^*=XA$.

En la figura (2.9) se muestra la matriz original X y la representación gráfica de los puntos.

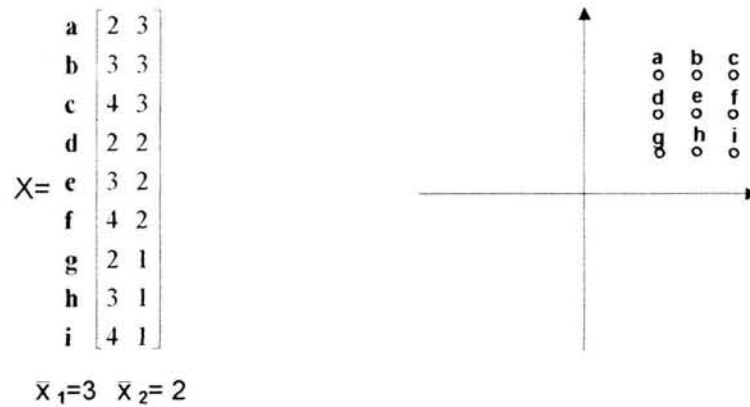


Fig. (2.9) Matriz original y representación gráfica de los puntos

Translación

La primera transformación que se aplicará a la matriz X es muy frecuente en el análisis multivariado: la translación. En términos generales, la translación se refiere al desplazamiento de los puntos hacia un nuevo origen de interés, expresándose como desviaciones respecto a ese nuevo origen; si este origen es el centroide (o media aritmética) de cada conjunto de coordenadas, entonces, en términos estadísticos, la media para cada dimensión de la matriz X es:

$$\bar{x}_1 = 3 \qquad \bar{x}_2 = 2$$

Sustrayendo la media de cada variable a los puntos (observaciones) originales se obtienen las desviaciones o medias corregidas denotadas por:

$$x_{dij} = x_{ij} - \bar{x}_j$$

donde x_{ij} son los valores de la matriz original X y \bar{x}_j es la media de la variable j .

La matriz de desviaciones o medias corregidas X_d se obtiene con la sustracción:

$$X_d = X - \bar{X}$$

y, en este caso, se obtienen las dos columnas de los valores de las desviaciones. En la figura (2.10) se muestra la transformación matricial y su gráfica respectiva:

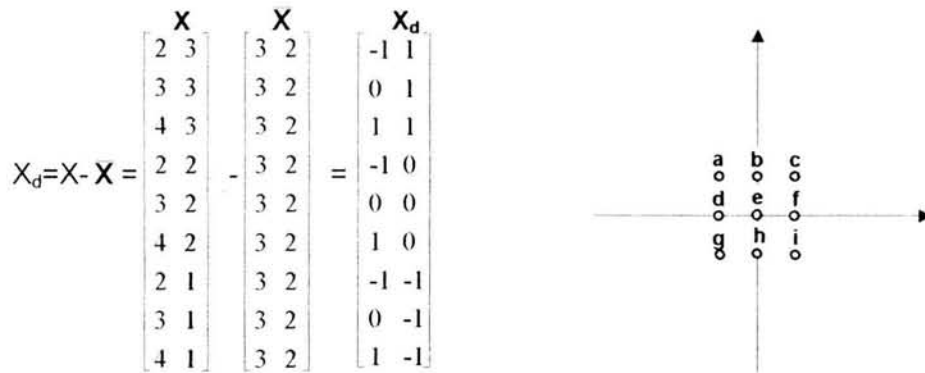


Fig. (2.10) Obtención de la matriz de desviaciones de las medias y su representación gráfica (Traslación de la matriz X hacia un nuevo origen –centroide-.)

Esta transformación no involucra una multiplicación de matrices, sino que se obtiene mediante una sustracción. Lo que resulta es un desplazamiento de los puntos de acuerdo a la desviación que presentan con respecto al centroide, es decir, gráficamente la nueva posición de los puntos indica la diferencia que tienen respecto a la media de cada dimensión.

Reflexión

Anteriormente se mencionó un caso especial de las rotaciones que involucraban la multiplicación de un número impar de renglones de una matriz de transformación por -1 y que originaba la reflexión de los puntos. Este es el caso de una rotación impropia donde a la rotación de los puntos le sigue una reflexión.

En la figura (2.11) se muestra la reflexión del eje 2 de los puntos de la matriz X_d , en donde las coordenadas de este eje se multiplican por -1 ; esto está dado por la matriz:

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

por lo que:

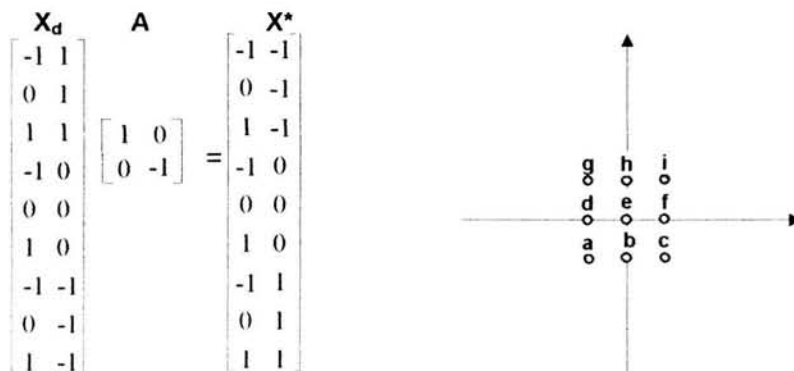


Fig. (2.11) Reflexión de los puntos a través del eje 1

Si se deseara obtener la reflexión del eje 1, o bien, realizar una reflexión a través del origen (es decir, una rotación de 180°), las matrices de transformación serían, respectivamente:

$$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{y} \quad \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$$

De lo anterior, se concluye que una reflexión de un número par de dimensiones es una rotación propia, mientras que si el número de dimensiones reflejadas es impar, se obtiene una rotación propia seguida de una reflexión (rotación impropia).

Ajuste

Cuando la matriz de transformación es una matriz diagonal cuyos valores λ_{ii} son distintos, el efecto que se produce es una "expansión" o "contracción" de las coordenadas de los puntos relativos a los ejes originales. Por ejemplo, en la figura (2.12), la matriz X_d es transformada por la matriz diagonal:

$$\begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$$

donde $\lambda_{11}=3$ expande las coordenadas del eje 1 tres veces más que las coordenadas originales.

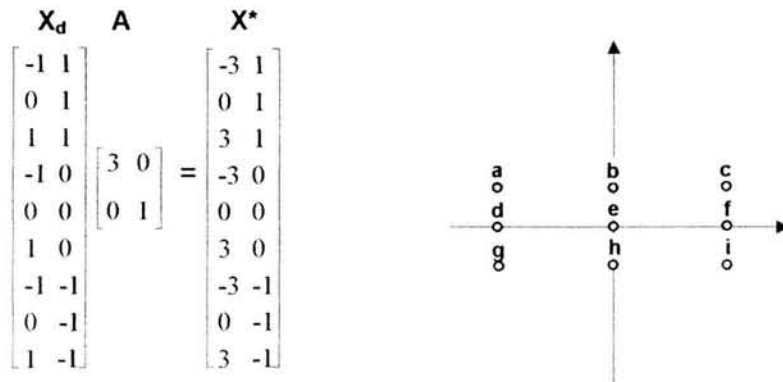


Fig. (2.12) Expansión de los puntos sobre el eje 1

Los casos para este tipo de transformación son:

Si $\lambda_{ii} > 1$, el ajuste es una expansión de las coordenadas.

Si $\lambda_{ii} = 1$, las coordenadas permanecen sin cambio.

Si $|\lambda_{ii}| < 1$, se obtiene una contracción de las coordenadas.

Aunque el término ajuste se restringe a valores positivos de λ_{ii} , si alguno de éstos es negativo se obtiene una expansión o contracción seguida de una reflexión.

En el ajuste se asume que las transformaciones son ajustes no uniformes; si todos los valores de la diagonal son iguales entonces el tipo de transformación se denomina dilatación central.

Rotación

En los temas anteriores se estableció que si la matriz de transformación es ortogonal, el efecto que se produce es una rotación. Por ejemplo, en dos dimensiones, la matriz ortogonal:

$$\begin{bmatrix} \cos 45^\circ & \text{sen } 45^\circ \\ -\text{sen } 45^\circ & \cos 45^\circ \end{bmatrix}$$

genera una rotación de 45° en sentido contrario a las manecillas del reloj. En la matriz se muestran los valores que se obtienen con las relaciones trigonométricas de ángulos complementarios:

$$\begin{aligned} \cos (90^\circ + \psi) &= -\text{sen } \psi \\ \cos(90^\circ - \psi) &= \text{sen } \psi \end{aligned}$$

donde $\psi = 45^\circ$ (esto sólo se puede aplicar en dos dimensiones). La figura (2.13) muestra la rotación de los puntos de la matriz X_d gráficamente:

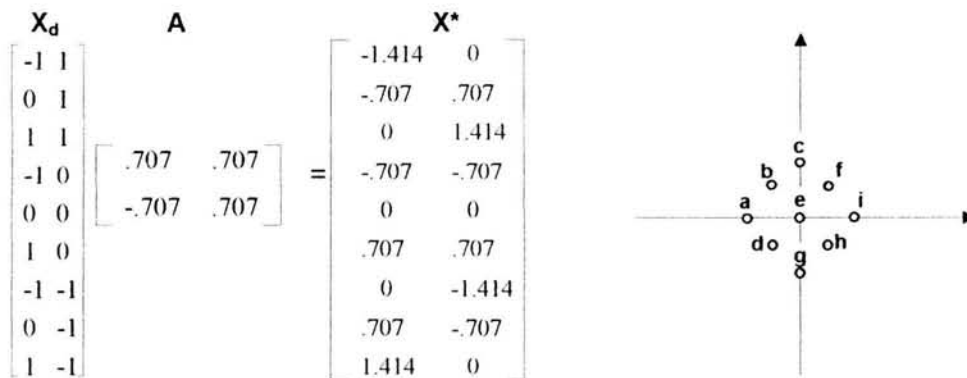


Fig. (2.13) Rotación de los puntos 45° en el sentido contrario de las manecillas del reloj

Este resultado también puede obtenerse rotando los ejes de la base un ángulo de 45° en el sentido de las manecillas del reloj.

TRANSFORMACIONES COMPUESTAS

Las transformaciones de translación, reflexión, ajuste y rotación han sido definidas por separado, pero el efecto que resulta de mayor interés es el que se origina cuando a una matriz no sólo se le aplica una de estas transformaciones sino una sucesión de ellas. Por ejemplo, si se tiene una matriz de transformación S que transforma a una matriz que a su vez ya ha sido transformada por otra matriz de transformación T, entonces las matrices S y T forman una transformación compuesta a través del producto TS. Si X_d es la matriz original, entonces las matrices obtenidas mediante las transformación compuesta por T y S son:

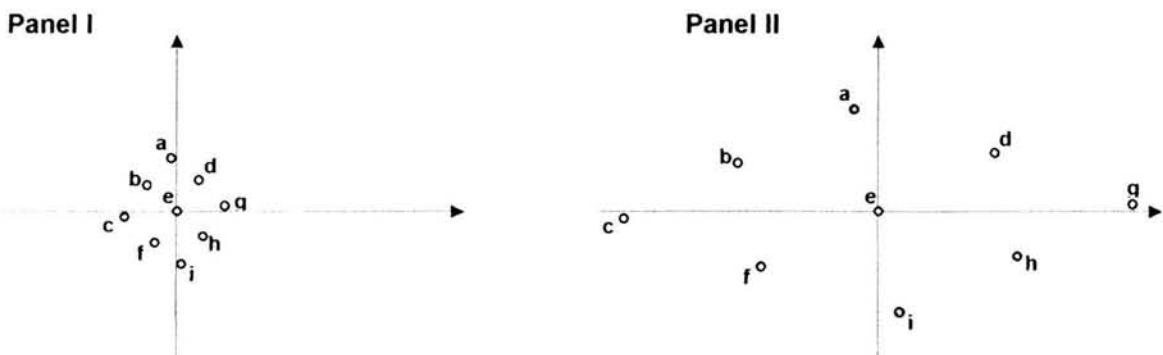
$$X_d \rightarrow X_d T \rightarrow X_d T S$$

Los resultados de aplicar la transformación TS a la matriz X_d no son los mismos que si se aplicara la transformación ST, por lo que el orden es importante en este tipo de transformaciones. Además, esto sólo se puede aplicar si todas las matrices son relativas a una misma base.

El tipo de transformación compuesta que es de mayor interés para el Análisis de Correspondencia es aquella en la que primero se realiza una rotación (posiblemente una rotación impropia), posteriormente se obtiene un ajuste y finalmente se realiza otra rotación. Esta transformación será aplicada a la misma matriz X_d de los ejemplos anteriores. Primero, supóngase que la rotación impropia está dada por la matriz P:

$$P = \begin{bmatrix} -.64 & -.77 \\ -.77 & .64 \end{bmatrix}$$

P es ortogonal, rota a los puntos 50° en el sentido de las manecillas del reloj y produce una reflexión del eje 1. El efecto gráfico se muestra en el panel I de la figura (2.14).



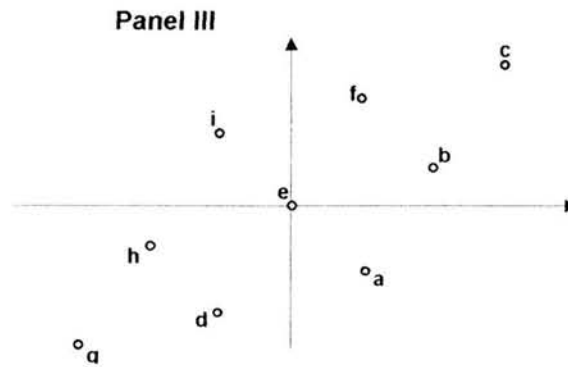


Fig. (2.14) Transformación compuesta

A los puntos obtenidos con X_dP se les aplica un ajuste con la matriz diagonal Δ :

$$\Delta = \begin{bmatrix} 5.12 & 0 \\ 0 & 1.95 \end{bmatrix}$$

Esta matriz expande las coordenadas del primer eje de la matriz X_dP a razón de 5.12 veces por unidad y las del segundo eje 1.95 veces por unidad. El panel II de la figura (2.14) muestra esta transformación.

Por último, se aplica una rotación más, dada por la matriz ortogonal Q' :

$$Q' = \begin{bmatrix} -.851 & -.526 \\ .526 & -.851 \end{bmatrix}$$

Los puntos del producto de matrices $X_dP\Delta$ son rotados 148° en el sentido de las manecillas del reloj y el resultado final se observa en el panel III de la figura (2.14).

Si se multiplican las matrices $P\Delta Q'$ anteriores se obtiene la matriz de transformación V :

$$V = P\Delta Q' = \begin{bmatrix} -.64 & -.77 \\ -.77 & .64 \end{bmatrix} \begin{bmatrix} 5.12 & 0 \\ 0 & 1.95 \end{bmatrix} \begin{bmatrix} -.851 & -.526 \\ .526 & -.851 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 4 & 1 \end{bmatrix}$$

Si se aplicara esta transformación directamente a la matriz X_d , es decir, X_dV , se obtendría el mismo resultado del panel III, donde se realiza $X_dP\Delta Q'$. Este resultado es de suma importancia para el análisis multivariado, donde la descomposición de una matriz arbitraria (V) en matrices más sencillas que describen transformaciones básicas, permite comprender el sentido geométrico de las transformaciones.

Las condiciones que una matriz debe cumplir para que sea posible realizar una descomposición única como la descrita son que sea no singular y que sus entradas sean valores reales. Una matriz T es **no singular** si su determinante es distinto de cero ($|T| \neq 0$). Si $|T|=0$, la matriz es singular y su inversa no está definida.

MATRICES DE DATOS ESTADÍSTICOS

La aplicación de transformaciones lineales permite describir muchas operaciones estadísticas dado que el Álgebra de matrices es muy concisa en sus procedimientos y permite una notación "compacta" de la información referente a los datos que están siendo analizados. Las matrices que se mencionan a continuación juegan un papel importante en el análisis multivariado por lo que resulta necesario saber el tipo de información que éstas proporcionan. Supóngase una matriz de datos originales con valores relativos a 3 variables X_1, X_2 y X_3 (Tabla 2.1):

	X_1	X_2	X_3		X_1X_2	X_1X_3	X_2X_3			
A=	1	2	1		2	1	2			
	2	1	1		2	2	1			
	0	1	2		0	0	2			
	1	3	2		3	2	6			
	4	10	4		40	16	40			
	10	11	6		110	60	66			
	7	12	5		84	35	60			
	4	3	4		12	16	12			
	5	3	8		15	40	24			
	8	2	7		16	56	14			
	$\Sigma X^2 =$	276	402	216	$\Sigma X_1X_2 =$	284	$\Sigma X_1X_3 =$	228	$\Sigma X_2X_3 =$	227

Tabla (2.1) Matriz de datos originales y productos cruzados

Denotando por A a la matriz anterior, se puede obtener la matriz B llamada **momento de producto mínimo** dada por:

$$B = A'A$$

$$B = \begin{bmatrix} 276 & 284 & 228 \\ 284 & 402 & 227 \\ 228 & 227 & 216 \end{bmatrix}$$

Los valores de B son los productos cruzados (productos escalares) para cada par de variables originales; naturalmente, los valores de la diagonal son la suma de los cuadrados de los elementos de cada variable X_1, X_2 y X_3 ; por ejemplo: $\Sigma X_1^2 = 276, \Sigma X_2X_3 = 227$.

Por otro lado, como ya se mencionó, la translación de los puntos hacia un nuevo origen (centroide o punto medio) genera la **matriz de medias corregidas**. La matriz de medias corregidas para la matriz A es:

$$A_d = A - \bar{A}$$

	$A_d = A - \bar{A}$	
$A_d = A - \bar{A} =$	$\begin{bmatrix} 1 & 2 & 1 \\ 2 & 1 & 1 \\ 0 & 1 & 2 \\ 1 & 3 & 2 \\ 4 & 10 & 4 \\ 10 & 11 & 6 \\ 7 & 12 & 5 \\ 4 & 3 & 4 \\ 5 & 3 & 8 \\ 8 & 2 & 7 \end{bmatrix} - \begin{bmatrix} 4.2 & 4.8 & 4 \\ 4.2 & 4.8 & 4 \\ 4.2 & 4.8 & 4 \\ 4.2 & 4.8 & 4 \\ 4.2 & 4.8 & 4 \\ 4.2 & 4.8 & 4 \\ 4.2 & 4.8 & 4 \\ 4.2 & 4.8 & 4 \\ 4.2 & 4.8 & 4 \\ 4.2 & 4.8 & 4 \end{bmatrix} = \begin{bmatrix} -3.2 & -2.8 & -3 \\ -2.2 & -3.8 & -3 \\ -4.2 & -3.8 & -2 \\ -3.2 & -1.8 & -2 \\ -2 & 5.2 & 0 \\ 5.8 & 6.2 & 2 \\ 2.8 & 7.2 & 1 \\ -2 & -1.8 & 0 \\ .8 & -1.8 & 4 \\ 3.8 & -2.8 & 3 \end{bmatrix}$	

donde la matriz \bar{A} contiene las medias de cada variable y en la matriz A_d se representan las desviaciones de los puntos respecto al punto medio. Con el momento de producto mínimo de esta matriz se obtiene la matriz de suma de cuadrados y productos cruzados corregidos dada por:

$$S = A_d' A_d$$

$$S = \begin{bmatrix} 99.6 & 82.4 & 60 \\ 82.4 & 171.6 & 35 \\ 60 & 35 & 56 \end{bmatrix}$$

Si se divide a la matriz S entre el número de observaciones (m), se obtiene la **matriz de covarianza** que es el promedio de los productos cruzados de las medias corregidas (S):

$$C = (1/m)S$$

$$C = \begin{bmatrix} 9.96 & 8.24 & 6 \\ 8.24 & 17.16 & 3.5 \\ 6 & 3.5 & 5.6 \end{bmatrix}$$

Los valores de la diagonal son las varianzas $Var(X) = \sum(X - \bar{X})^2 / m$ y los elementos restantes son las covarianzas definidas por $Cov(X_1 X_2) = \sum x_1 x_2 / m$.

Todas las matrices descritas permiten expresar de forma concisa algunas de las operaciones estadísticas que son de uso frecuente en el análisis multivariado. En particular, será de interés la matriz de covarianza ya que con ésta es posible conocer las relaciones establecidas entre cada par de variables y su comportamiento.

EIGENESTRUCTURA DE MATRICES SIMÉTRICAS

Uno de los principales objetivos del análisis multivariado es explicar las variaciones o relaciones entre las variables que se analizan. Un medio para lograr esto es a través de transformaciones que permitan obtener espacios multidimensionales que expliquen la mayor varianza posible y para ello se recurre al análisis de matrices de dispersión como la matriz de covarianza ya que son éstas las que brindan información sobre las variaciones existentes en los datos. En el Análisis de Correspondencia, sin embargo, es necesario hacer algunas consideraciones por el tipo de escala de las variables que se manejan; las matrices de dispersión que se definen en este método se construyen de una forma específica para los fines que se pretenden alcanzar, tal como se mostrará en el siguiente capítulo.

En términos generales, cuando se tiene un conjunto de datos X , puede analizarse su comportamiento por medio de matrices como las mencionadas en el apartado anterior. Específicamente, para los fines que aquí se persiguen, se hará referencia a la matriz de medias corregidas X_d y a la matriz de covarianza $C(X)$. La matriz X_d muestra las desviaciones de los puntos con respecto a la media de cada dimensión, es decir, proporciona información inicial con respecto a la variación de los datos. El objetivo es encontrar una nueva base dada por una matriz ortogonal de transformación T tal que los nuevos ejes desplieguen la mayor varianza posible.

De aquí que, teniendo las variables $X_{d1}, X_{d2}, \dots, X_{dn}$, deben aplicarse transformaciones lineales z_i de la forma:

$$z_i = t_1 X_{d1} + t_2 X_{d2} + \dots + t_n X_{dn}$$

donde los valores t_1, t_2, \dots, t_n conforman un vector de transformación \mathbf{t} para $i=1 \dots m$, donde m es el número de observaciones, de tal manera que se maximice la variación dada por:

$$\text{Var}(z_i) = \sum_{i=1}^m (z_i - \bar{z})^2 / m$$

donde \bar{z} es la media de la variable transformada y el vector $\mathbf{t}' = (t_1 \ t_2 \ \dots \ t_n)$ está restringido a que su magnitud sea igual a 1 para evitar valores indefinidamente grandes de la varianza. Este vector debe maximizar la varianza y sobre él se proyectan los puntos originales X_d ; dichas proyecciones están dadas por los valores de z_i . Esto es equivalente a minimizar la suma de las distancias cuadradas entre los puntos originales y vector \mathbf{t} .

Dado que se está realizando una rotación de la base, la matriz ortogonal de transformación T contendrá el conjunto de vectores \mathbf{t} que generan la nueva base. Estos vectores son llamados **ejes principales** y sus componentes son cosenos directores.

Para obtener lo anterior se recurre a la matriz de covarianza de los datos originales ya que a partir de ésta se realiza un procedimiento que permite obtener la matriz T que genera la rotación óptima para hallar la base requerida. Tal procedimiento se refiere a encontrar la eigenestructura (valores y vectores propios) de la matriz de covarianza. Los eigenvectores obtenidos representan a los vectores de la matriz T y los eigenvalores indican los valores de la dispersión sobre cada nuevo vector.

Cuando se tiene una nueva base conformada por eigenvectores la transformación adquiere, geoméricamente, una forma particularmente simple. El problema de la eigenestructura puede explicarse de la siguiente manera:

Supóngase que se tiene una matriz cuadrada A . La ecuación:

$$A\mathbf{x}=\lambda\mathbf{x}$$

donde \mathbf{x} es un vector y λ un escalar, conduce a hallar aquellos vectores \mathbf{x} que al aplicarles la transformación A se mapean en ellos mismos o en múltiplos de ellos mismos (es decir, que mantengan su dirección) de tal forma que esas imágenes están dadas por $\mathbf{x}^*=\lambda\mathbf{x}$. Los vectores \mathbf{x} se llaman vectores invariantes bajo la transformación A y el problema para hallarlos se reduce a resolver la ecuación:

$$(A-\lambda I)\mathbf{x}=0$$

donde I es la matriz identidad y la solución deseada es aquella para valores de $\mathbf{x}\neq\mathbf{0}$, lo que condiciona a que el determinante de $(A-\lambda I)$ sea igual a cero.

Si además la matriz A es simétrica, las propiedades de su eigenestructura se pueden describir en forma más sencilla que las de matrices no simétricas.

La matriz de covarianza $C(X)$ es cuadrada y simétrica, por lo que constituye un caso especial para la eigenestructura. Esta matriz se puede obtener a partir del momento de producto mínimo de la matriz de medias corregidas:

$$C(X)=1/m(X_d'X_d)$$

donde m es el número de observaciones y X_d la matriz de medias corregidas de X . En esta matriz los valores son reales y no negativos y, por lo tanto, todos sus eigenvalores también serán reales y no negativos.

Para ilustrar lo que sucede al obtener la eigenestructura de la matriz simétrica $C(X)$, considérese primero la matriz de medias corregidas X_d que resulta de una matriz de datos originales X (Tabla (2.2))

$$X_d = X - \bar{X} = \begin{array}{c} X \\ \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 0 & 1 \\ 1 & 3 \\ 4 & 10 \\ 10 & 11 \\ 7 & 12 \\ 4 & 3 \\ 5 & 3 \\ 8 & 2 \end{bmatrix} \\ \bar{X} \\ \begin{bmatrix} 4.2 & 4.8 \\ 4.2 & 4.8 \\ 4.2 & 4.8 \\ 4.2 & 4.8 \\ 4.2 & 4.8 \\ 4.2 & 4.8 \\ 4.2 & 4.8 \\ 4.2 & 4.8 \\ 4.2 & 4.8 \\ 4.2 & 4.8 \end{bmatrix} \\ X_d \\ \begin{bmatrix} -3.2 & -2.8 \\ -2.2 & -3.8 \\ -4.2 & -3.8 \\ -3.2 & -1.8 \\ -0.2 & 5.2 \\ 5.8 & 6.2 \\ 2.8 & 7.2 \\ -0.2 & -1.8 \\ 0.8 & -1.8 \\ 3.8 & -2.8 \end{bmatrix} \end{array}$$

Tabla (2.2) Matriz de datos originales y de medias corregidas

La matriz de covarianza $C(X)$ es:

$$C(X) = (1/m)X_d'X_d$$

$$C(X) = \begin{bmatrix} 9.96 & 8.24 \\ 8.24 & 17.16 \end{bmatrix}$$

Los elementos de la diagonal representan las varianzas para las variables x_1 y x_2 respectivamente, mientras que los elementos restantes son las covarianzas de x_1 y x_2 .

La ecuación característica de $C(X)$ es:

$$|C(X) - \lambda I| = 0$$

Así, para obtener la eigenestructura de la matriz $C(X)$ anterior se tiene que:

$$|C(X) - \lambda I| = \begin{vmatrix} 9.96 - \lambda & 8.24 \\ 8.24 & 17.16 - \lambda \end{vmatrix} = 0$$

el polinomio característico es:

$$p(\lambda) = \lambda^2 - 27.12\lambda + 103.01$$

igualando a cero y resolviendo el polinomio se obtienen los eigenvalores ordenados de mayor a menor $\lambda_1 = 22.55$ y $\lambda_2 = 4.57$. La interpretación de los resultados es que λ_1 representa

el valor máximo de la varianza descrita a lo largo de una dimensión representada por el eigenvector asociado a λ_1 ; este valor es alcanzado a través de una composición lineal de los vectores de la base original. Por otro lado, λ_2 describe el máximo valor de la varianza restante a lo largo de otra dimensión, dada por un eigenvector ortogonal al primero. Los vectores correspondientes a los valores λ_i son:

$$\mathbf{t}_1 = \begin{bmatrix} .654k \\ k \end{bmatrix} \quad \mathbf{t}_2 = \begin{bmatrix} -1.528k \\ k \end{bmatrix}$$

Ambos vectores son ortogonales ($\mathbf{t}_1 \cdot \mathbf{t}_2 = 0$), pero no son unitarios, y esa es una restricción que debe satisfacerse, por lo tanto, para que los vectores sean unitarios se divide cada componente entre la magnitud del vector. Los vectores unitarios son:

$$\mathbf{t}_1(\text{norm}) = \frac{\mathbf{t}_1}{|\mathbf{t}_1|} = \begin{bmatrix} .548 \\ .837 \end{bmatrix} \quad \mathbf{t}_2(\text{norm}) = \frac{\mathbf{t}_2}{|\mathbf{t}_2|} = \begin{bmatrix} -.837 \\ .548 \end{bmatrix}$$

Por lo tanto, la matriz de transformación T que genera la rotación de la base y que describe la máxima varianza para cada dimensión es:

$$T = \begin{bmatrix} .548 & -.837 \\ .837 & .548 \end{bmatrix}$$

que contiene a los vectores \mathbf{t}_1 y \mathbf{t}_2 (ejes principales).

Los aspectos geométricos de estos resultados se muestran en la figura (2.15) donde puede notarse que los eigenvalores λ_i indican la magnitud de los ajustes y los eigenvectores \mathbf{t}_i indican las nuevas direcciones sobre los cuales se efectúan dichos ajustes; es decir, cualquier punto que caiga sobre el vector que define \mathbf{t}_1 será expandido sobre ese mismo vector en la razón 22.55:1 y del mismo modo, los puntos sobre \mathbf{t}_2 serán expandidos a razón 4.57:1.

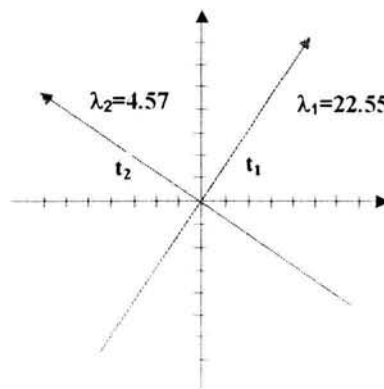


Fig. (2.15) Vectores invariantes bajo la matriz de transformación T

Aplicando esta transformación a la matriz X_d se obtiene la matriz Z que contiene los vectores z_1 y z_2 dados por las transformaciones:

$$z_1 = 0.548 X_{d1} + 0.837 X_{d2}$$

$$z_2 = -0.837 X_{d1} + 0.548 X_{d2}$$

y que representan las coordenadas de los puntos referentes a la nueva base obtenida mediante T (figura (2.16)):

$$Z = X_d T = \begin{bmatrix} -3.2 & -2.8 \\ -2.2 & -3.8 \\ -4.2 & -3.8 \\ -3.2 & -1.8 \\ -0.2 & 5.2 \\ 5.8 & 6.2 \\ 2.8 & 7.2 \\ -0.2 & -1.8 \\ 0.8 & -1.8 \\ 3.8 & -2.8 \end{bmatrix} \begin{bmatrix} .548 & -.837 \\ .837 & .548 \end{bmatrix} = \begin{bmatrix} -4.097 & 1.144 \\ -4.386 & -0.241 \\ -5.482 & 1.433 \\ -3.26 & 1.692 \\ 4.243 & 3.017 \\ 8.368 & -1.457 \\ 7.561 & 1.602 \\ -1.616 & -0.819 \\ -1.068 & -1.656 \\ -0.261 & -4.715 \end{bmatrix}$$

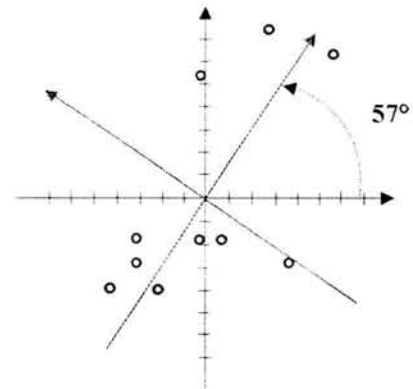


Fig. (2.16) Matriz Z de los puntos referidos a la nueva base

Esta transformación corresponde a una rotación de la base de 57° en el sentido contrario a las manecillas del reloj.

Para calcular la matriz de covarianza de la nueva matriz Z se tiene:

$$C(Z) = (1/m) Z'Z$$

m es el número de observaciones y Z es la matriz de los puntos transformados. Como $Z = X_d T$, entonces:

$$C(Z) = (1/m) [(X_d T)' (X_d T)] = (1/m) [T' X_d' X_d T] = T' C(X) T$$

Por lo anterior, $C(Z)$ se puede obtener directamente a través de T y $C(X)$:

$$C(Z) = \begin{bmatrix} T' & C(X) & T \end{bmatrix} = \begin{bmatrix} .548 & .837 \\ -.837 & .548 \end{bmatrix} \begin{bmatrix} 9.96 & 8.24 \\ 8.24 & 17.16 \end{bmatrix} \begin{bmatrix} .548 & -.837 \\ .837 & .548 \end{bmatrix} = \begin{bmatrix} 22.57 & 0 \\ 0 & 4.57 \end{bmatrix}$$

En esta matriz los valores de la diagonal son los eigenvalores de $C(X)$ y representan las varianzas para las variables Z_1 y Z_2 respectivamente, además, los elementos fuera de la diagonal son cero, lo que indica que después de la transformación las variables no están correlacionadas, ya que la covarianza entre cada par de ellas es cero, es decir, los productos cruzados en $C(Z)$ desaparecen. Esta matriz no sólo representa la mejor explicación de la variación de los datos sobre la nueva base, sino que también es de relevancia al considerar otro aspecto de las matrices de transformación: la diagonalización.

DIAGONALIZACIÓN DE LA MATRIZ DE TRANSFORMACIÓN

Anteriormente se mostró que una matriz de transformación T° mapea vectores \mathbf{x}° en otros vectores $\mathbf{x}^{*\circ}$, ambos referidos a la nueva base F , donde es necesario conocer las matrices T (la matriz de transformación relativa a la base E), y L (la matriz que conecta a las bases E y F) para obtener T° .

La matriz L se escogió arbitrariamente y los vectores de la nueva matriz F estaban dados por L' . Si L' asumiera ciertas características especiales y sus columnas estuvieran dadas por los eigenvectores de T , entonces la matriz T° sería una matriz diagonal cuyos valores serían los eigenvalores de T .

Recordando que $T^\circ = (L')^{-1}TL'$ y simplificando la notación, llámese D a la matriz T° , U a la matriz L' y A a la matriz T . Dada una matriz cuadrada A , se dice que A es **diagonalizable** si:

$$D = U^{-1}AU$$

siempre que U y D se puedan obtener. La matriz D es diagonal y sus entradas son los eigenvalores de A , U contiene los eigenvectores de A y debe ser no singular para poder hallar U^{-1} .

La matriz D es, entonces, una transformación que se aplica sobre una nueva base de eigenvectores y que mapea puntos mediante ajustes dados por los eigenvalores de la diagonal; esta transformación simplemente expande los puntos sobre la nueva base por lo que representa una forma geométrica más simple.

Así, la matriz A representa una transformación sobre la base original y la matriz D representa transformaciones sobre la nueva base dada por U .

La matriz A se puede escribir como:

$$A = UDU^{-1}$$

El hecho de trabajar con matrices simétricas garantiza que éstas pueden diagonalizarse y si este es el caso entonces $U^{-1} = U'$, por lo que $D = U'AU$ y $A = UDU'$.

Por lo anterior, considerando que $C(X)$ es simétrica, otra forma de obtener $C(Z)$ es mediante la diagonalización de $C(X)$ ya que análogamente a la generalización de la matriz A se tiene:

$$C(Z)=T'C(X)T$$

de donde:

$$C(X)=TC(Z)T'$$

En resumen, al aplicar la rotación de la base original a través de T se ha obtenido un conjunto de vectores de una nueva base donde las proyecciones de los puntos originales sobre los nuevos ejes despliegan la máxima varianza. La matriz de transformación $D=C(Z)$ relativa a la nueva base transforma los puntos referidos a ella a través de un ajuste expandiéndolos a razón de $\lambda_k:1$ con respecto a cada nuevo vector k .

En general, al calcular la eigenestructura de una matriz cuadrada A , ya sea simétrica o no, se cumple que:

- La suma de sus eigenvalores es igual a la suma de los elementos de su diagonal:

$$\sum_{i=1}^m \lambda_i = \sum_{i=1}^n a_{ii} \quad y,$$

- El producto de sus eigenvalores es igual a su determinante:

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

por lo que si al menos un valor $\lambda_i=0$, entonces la matriz A es singular.

Otro aspecto que debe considerarse es referente a valores repetidos o nulos de λ_i , ya que estos determinan la dimensión del espacio sobre el cual se representan los valores.

REDUCCIÓN DE LA DIMENSIÓN

Las transformaciones que se han realizado han mapeado vectores que se encuentran en un espacio de dimensión n en otros espacios de la misma dimensión. Por ejemplo, si el vector $\mathbf{x}'=(1 \ 2)$ es transformado en el vector \mathbf{x}^* a través de la matriz $T=\begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}$, entonces:

$$\mathbf{x}^* = T\mathbf{x}' = \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 9 \\ 8 \end{bmatrix}$$

Ambos vectores \mathbf{x} y \mathbf{x}^* se encuentran en un espacio de dos dimensiones. Sin embargo, no siempre las matrices de transformación mapean vectores en espacios de la misma dimensión. La dimensión del espacio imagen está determinada por el **rango** de la matriz de transformación.

Rango de una Matriz

Existen varias formas de definir y de calcular el rango de una matriz A (cuadrada o no) denotado por $r(A)$; en ellas se hace referencia a la independencia lineal. Se dice que el rango de una matriz A es el máximo número de renglones o columnas de A que sean linealmente independientes. De aquí que, si A es rectangular de $m \times n$, entonces su rango no puede exceder el mínimo de m ó n , es decir:

$$r(A) = k \leq \min(m, n)$$

En otras palabras, pueden tenerse m puntos en un espacio de n dimensiones, o bien, n puntos en un espacio de dimensión m , donde el número menor (m ó n) representa el espacio que contiene a los puntos.

Geoméricamente, el rango de una matriz determina las características de la dimensión del espacio original que se conservan bajo una matriz de transformación. Por ejemplo, si el punto $\mathbf{x}' = (1 \ 3)$ es transformado bajo la matriz de transformación $T = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, entonces:

$$\mathbf{x}^* = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

el vector imagen \mathbf{x}^* aparentemente se proyecta sobre un espacio de dos dimensiones (figura (2.17)), sin embargo, al aplicar la misma transformación a otros vectores $\mathbf{x}_1' = (1 \ 1)$, $\mathbf{x}_2' = (2 \ 3)$ y $\mathbf{x}_3' = (3 \ 1)$ se tiene que $\mathbf{x}_1^* = (1 \ 0)$, $\mathbf{x}_2^* = (2 \ 0)$ y $\mathbf{x}_3^* = (3 \ 0)$; todos ellos son proyectados sobre una recta, por lo que sólo una dimensión es suficiente para representar las transformaciones.

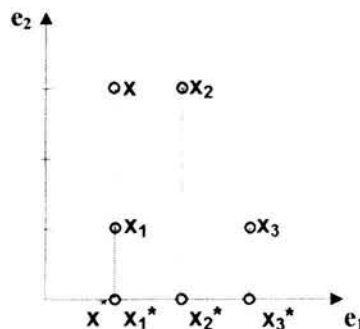


Fig. (2.17) Reducción de la dimensión

Las transformaciones hechas por T se mapean sobre e_1 y es de notarse que cualquier vector al que se le aplique T conservará su primer componente con el mismo valor, mientras que el segundo, independientemente de su valor, siempre se hará cero y por lo tanto se pierde información sobre los vectores originales. Estos resultados se dan debido a que el rango de la matriz T es $r(T)=1$, por lo que el espacio imagen es una recta, es decir, de una dimensión.

En particular, para las matrices de transformación A cuadradas de $n \times n$ para la cual existe A^{-1} se cumple que:

- $|A| \neq 0$, por lo que es no singular.
- $r(A)=n$, A es de rango completo.
- Los renglones y columnas de A son linealmente independientes.
- El espacio que se obtiene con la transformación A conserva las características del espacio original.
- Existe una correspondencia uno a uno (biunívoca) entre los puntos preimagen y los puntos imagen.
- Se puede obtener la preimagen de los vectores a partir de las imágenes aplicando la transformación A^{-1} .

Eigenestructura y Rango

Finalmente, la eigenestructura de una matriz también determina su rango, aún si ésta es rectangular. Si la matriz cuadrada A de $n \times n$ es simétrica, el rango será el número de eigenvalores distintos de cero, donde $r(A)=k \leq n$.

Si A no es simétrica o es rectangular, su rango se obtiene calculando la matriz de momento de producto mínimo ($A'A$) o mayor (AA'), ya que:

$$r(A)=r(A'A)=r(AA')$$

Si $A_{m \times n}$ (m renglones y n columnas), entonces $A'A$ es de orden n y AA' es de orden m ; dependiendo de los valores de m y n será más conveniente obtener aquella de menor orden para facilitar el cálculo, aunque con cualquiera se obtiene el mismo resultado.

Como $A'A$ y AA' son cuadradas y simétricas, todos sus eigenvalores serán reales y no negativos. El número de eigenvalores positivos será el rango de A , aún cuando existan valores λ_i repetidos; si este es el caso, pueden hallarse vectores ortonormales para estos valores mediante el proceso de ortonormalización de Gram-Schmidt.

- Si $r(A)=n$, para $A_{n \times n}$, todos sus renglones y columnas son linealmente independientes.
- Si $r(A)=n < m$, para $A_{m \times n}$, entonces los renglones de A son linealmente dependientes.
- Si $r(A)=m < n$, para $A_{m \times n}$, entonces las columnas de A son linealmente dependientes.
- Si $r(A)=k < n \leq m$, entonces $r(A)=k$ es el mayor número de vectores linealmente independientes de A , donde k es el número de eigenvalores positivos de A , $A'A$ ó AA' .

Las matrices de covarianza pertenecen al caso donde se calcula $A'A$ ó AA' . De esta manera, las propiedades antes descritas se aplican también a estas matrices, por lo que todos los eigenvalores de $C(X)$ son no negativos y su rango está determinado por los valores λ_i que son positivos.

DESCOMPOSICIÓN EN VALORES SINGULARES

Los temas que se han tratado hasta el momento constituyen la base para explicar uno de los aspectos fundamentales del Análisis de Correspondencia que conforma la parte numérica o abstracta sobre la que pocas veces se profundiza y se refiere a la descomposición de matrices en valores singulares, el cual será aplicado en el siguiente capítulo.

Cuando se mencionaron los efectos geométricos de ciertas transformaciones especiales se observó que una matriz de transformación arbitraria puede descomponerse en matrices más sencillas que explican los efectos obtenidos al aplicar dicha matriz (donde $V=P\Delta Q'$). Este tipo de descomposición es la más general y puede aplicarse a cualquier matriz, cuadrada o rectangular, simétrica o asimétrica, singular o no singular; encontrando la estructura básica que simplifica su naturaleza geométrica.

Suponiendo el caso más general, donde A es rectangular de $m \times n$, su descomposición en el triple producto de matrices que producen una rotación (seguida o no por una reflexión), un ajuste y otra rotación, es llamada **Descomposición en Valores Singulares (DVS)**.

Si $m > n$ y el rango $r(A) = k \leq n < m$, la DVS está dada por:

$$A_{m \times n} = P_{m \times k} \Delta_{k \times k} Q'_{k \times n}$$

Las matrices P , Δ y Q se definen por medio de la eigenestructura de las matrices de producto mínimo y producto mayor de A . La matriz Δ es llamada **matriz de valores singulares** y los valores de la diagonal corresponden a las raíces cuadradas de los k eigenvalores de la matriz $A'A$ ó AA' ordenados de mayor a menor ($\alpha_k = \sqrt{\lambda_k}$). La matriz Q está conformada por los eigenvectores principales de $A'A$, y la matriz P corresponde a los eigenvectores principales de AA' .

Las relaciones entre $A = P\Delta Q'$ y las matrices de producto mínimo y producto mayor se establecen por medio del análisis de éstas últimas; es decir, si $A = P\Delta Q'$, entonces los productos $A'A$ y AA' corresponden a:

$$\begin{aligned} A'A &= (P\Delta Q')'(P\Delta Q') & AA' &= (P\Delta Q')(P\Delta Q')' \\ &= Q\Delta P'P\Delta Q' & &= P\Delta Q'Q\Delta P' \\ &= Q\Delta^2 Q' & &= P\Delta^2 P' \end{aligned}$$

Si se denota por D a la matriz diagonal Δ^2 , entonces:

$$A'A = QDQ' \quad AA' = PDP'$$

Aunque la matriz A no sea cuadrada, los productos $A'A$ y AA' sí lo son y, además, son matrices simétricas por lo que ambas tienen la forma de descomposición $A = UDU'$, es decir, son el triple producto de una matriz ortogonal (U), una matriz diagonal (D) y la matriz ortogonal transpuesta (U').

Como $r(A) = r(A'A) = r(AA')$ entonces $A'A$ y AA' tienen el mismo rango definido por sus k eigenvalores distintos de cero; estos eigenvalores, además, son los mismos para ambos productos, por lo que la matriz D asociada a $A'A$ y AA' es la misma. Los valores de D son reales, no negativos y están ordenados de mayor a menor. La matriz Δ puede entonces definirse como:

$$\Delta_{k \times k} = D^{1/2}_{k \times k}$$

Para que P y Q sean ortogonales y propiamente constituyan una rotación es necesario que $k=m=n$, es decir, A debe ser cuadrada y de rango completo, así, A puede escribirse como:

$$A=UDU^{-1}$$

donde U^{-1} es la matriz inversa de U . Si además A es simétrica, la DVS es equivalente al caso especial de diagonalización en donde:

$$A=UDU'$$

En ambos casos se cumple que $P'P=PP'=I$ y $Q'Q=QQ'=I$. Sin embargo, en el caso de matrices no simétricas o rectangulares, la matriz $P_{m \times k}$ son las primeras k columnas de la matriz ortogonal $P_{m \times m}$ formada por los eigenvectores de AA' y $Q_{n \times k}$ son las primeras k columnas de la matriz ortogonal $Q_{n \times n}$ que contiene los eigenvectores de $A'A$ correspondientes a los valores ordenados de D y, por lo tanto, de Δ . En este caso P y Q son ortogonales por columna, por lo que se cumple que $P'P=Q'Q=I$, donde $k < n < m$. Las matrices se muestran en el panel I de la figura (2.18).

Para que P y Q puedan considerarse ortogonales propiamente, la matriz Δ puede estar contenida en el bloque superior izquierdo de una matriz rectangular más grande de $m \times n$ con $m-k$ renglones y $n-k$ columnas de ceros, en este sentido, P y Q constituyen matrices de rotación de orden m y n respectivamente. Este caso se conoce como DVS completa (panel II de la figura (2.18)), aunque su importancia no va más allá de considerar que P y Q pueden ser ortogonales si se agregan los renglones y columnas necesarios ya que, de cualquier manera, las dimensiones que se le añaden desaparecen debido a los $m-k$ renglones y $n-k$ columnas de ceros de la matriz que contiene a Δ .

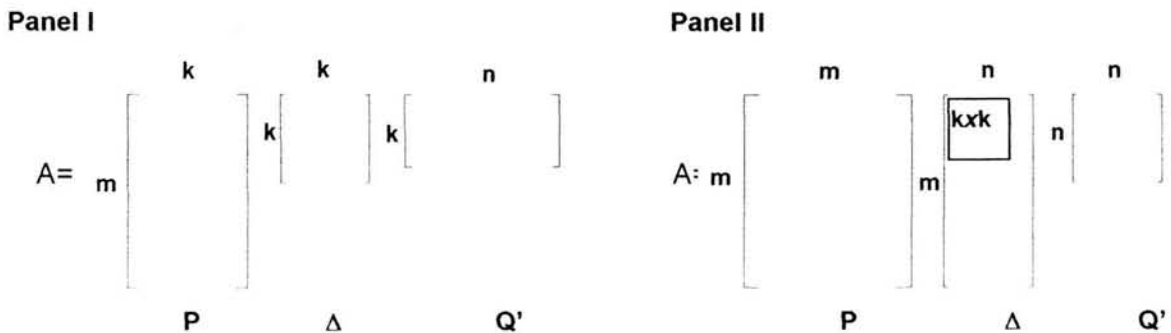


Fig. (2.18) Descomposición en Valores Singulares y DVS Completa

Para cada matriz A existe sólo una matriz Δ , por lo que los valores singulares siempre son únicos. Si todos los valores singulares de Δ son distintos, entonces P y Q' son únicas. Si algunos de éstos están repetidos sólo las porciones de P y Q' que están relacionadas a valores singulares distintos son únicas; las porciones de P y Q' correspondientes a valores repetidos no son únicas, en este caso, pueden hallarse vectores linealmente independientes por el proceso de Gram-Schmidt en el bloque repetido de r eigenvalores (que serán únicos en el subespacio r -dimensional) y encontrar vectores mutuamente ortogonales.

Para obtener la DVS de una matriz $A_{m \times n}$, donde $n \leq m$, el procedimiento es el siguiente:

- Se calcula el momento de producto mínimo $A'A$; por ejemplo, si $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 2 \end{bmatrix}$ entonces:

$$A'A = \begin{bmatrix} 2 & 3 \\ 3 & 6 \end{bmatrix}$$

El resultado siempre será una matriz cuadrada simétrica de $n \times n$.

- Se encuentra la eigenestructura de $A'A$:

$$|(A'A) - \lambda I| = \begin{vmatrix} 2-\lambda & 3 \\ 3 & 6-\lambda \end{vmatrix} = (2-\lambda)(6-\lambda) - 9 = 0$$

$$\lambda^2 - 8\lambda + 3 = 0$$

Eigenvalores

$$\lambda_1 = 7.61$$

$$\lambda_2 = 0.39$$

Eigenvectores

$$Q = \begin{bmatrix} -.472 & .882 \\ -.882 & -.472 \end{bmatrix}$$

Los eigenvalores conforman la matriz Δ^2 de rango k ($k \leq n$)

$$\Delta^2 = \begin{bmatrix} 7.61 & 0 \\ 0 & .39 \end{bmatrix}$$

La matriz Q es la matriz de los eigenvectores principales asociados con los eigenvalores positivos de $A'A$ y es llamada matriz principal de **vectores singulares de la derecha** de A , cuyas columnas son ortogonales. Los eigenvalores de $A'A$ serán los mismos que los eigenvalores de AA' .

- Se calcula la raíz cuadrada de los valores de la diagonal de Δ^2 .

$$\Delta = \begin{bmatrix} 2.76 & 0 \\ 0 & .62 \end{bmatrix}$$

Estos son los **valores singulares** de A (raíces cuadradas de los eigenvalores de A'A ó AA').

- Se obtiene Δ^{-1} (la matriz inversa de Δ), que son los recíprocos de los valores singulares.

$$\Delta^{-1} = \begin{bmatrix} 1/2.76 & 0 \\ 0 & 1/.62 \end{bmatrix}$$

- Se calcula P por medio de la eigenestructura de AA', o bien, de forma más sencilla a través de:

$$P=AQ\Delta^{-1} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} -.472 & .882 \\ -.882 & -.472 \end{bmatrix} \begin{bmatrix} 1/2.76 & 0 \\ 0 & 1/.62 \end{bmatrix} = \begin{bmatrix} -.49 & .66 \\ -.32 & -.76 \\ -.81 & -.10 \end{bmatrix}$$

P es la matriz cuyas columnas son los eigenvectores principales de AA' asociados con los k eigenvalores positivos ordenados de $D=\Delta^2$. Las columnas de P son llamadas **vectores singulares de la izquierda** de A.

La DVS de A, por lo tanto, es:

$$A=P\Delta Q' = \begin{bmatrix} -.49 & .66 \\ -.32 & -.76 \\ -.81 & -.10 \end{bmatrix} \begin{bmatrix} 2.76 & 0 \\ 0 & .62 \end{bmatrix} \begin{bmatrix} -.472 & -.882 \\ .882 & -.472 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 2 \end{bmatrix}$$

La descomposición realizada se muestra en la figura (2.19), donde se aprecian los puntos de P, su ajuste a través de Δ (expansión sobre el eje 1 y contracción sobre el eje 2) y la rotación de 118° dada por Q', dando como resultado los puntos originales de la matriz A:

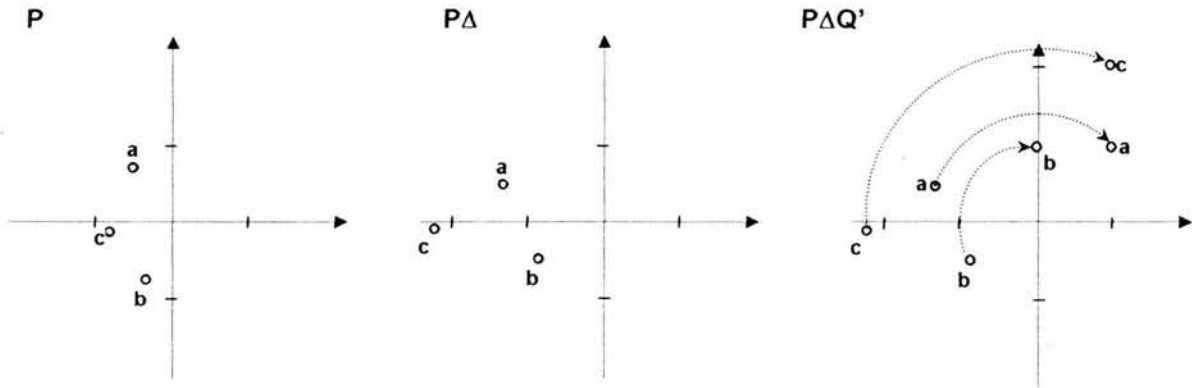


Fig. (2.19) Transformaciones a través de la DVS

Si en la matriz $A_{m \times n}$, n es mayor que m , entonces se aplicaría el mismo procedimiento pero sobre la matriz A' , es decir:

$$A' = P_1 \Delta_1 Q_1'$$

donde P_1 contiene los eigenvectores principales de $A'A$, Δ_1 es la matriz de valores singulares y Q_1 contiene los eigenvectores principales de AA' , y al volver a transponer:

$$A = (P_1 \Delta_1 Q_1')' = Q_1 \Delta_1 P_1'$$

por lo tanto:

$$P \equiv Q_1, \quad \Delta \equiv \Delta_1, \quad Q \equiv P_1$$

Si en lugar de calcular la eigenestructura de $A'A$ se calculara la de AA' , entonces:

$$AA' = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 1 & 2 \\ 3 & 2 & 5 \end{bmatrix}$$

cuya ecuación característica es:

$$-\lambda^3 + 8\lambda^2 - 3\lambda = 0$$

los eigenvalores son:

$$\lambda_1 = 7.61$$

$$\lambda_2 = 0.39$$

$$\lambda_3 = 0$$

Si se considerara el valor $\lambda_3 = 0$ en la matriz Δ^2 entonces los vectores multiplicados por él se desvanecen; de este modo, la dimensión está dada por $r(A'A) = r(AA') = 2$ considerando

únicamente los valores positivos de λ_{ii} y obteniendo el resultado anterior con la eigenestructura de $A'A$.

En cuanto a las gráficas que se pueden obtener, debe considerarse que en la mayoría de los casos será necesario reducir la dimensión del espacio original determinado por k , con lo cual se pierde información. Sin embargo, la DVS permite que la suma de desviaciones cuadradas de A y $P\Delta Q'$ sea mínima; uno de los aspectos que contribuyen a esto es que los valores de Δ están ordenados de mayor a menor, de ese modo, si no se consideran todos los valores de la diagonal de Δ , primero se tomarán en cuenta los valores λ_i más grandes que contribuyen a explicar en mayor medida la variación encontrada en A .

CAPÍTULO III

ANÁLISIS DE CORRESPONDENCIA EN DATOS CATEGÓRICOS

El Análisis de Correspondencia es un método estadístico que se aplica en estudios donde los datos recolectados son categóricos y que analiza las relaciones o “correspondencias” entre las categorías de variables medidas en una escala nominal, permitiendo realizar análisis de datos no métricos. Las frecuencias, en su forma más básica, se despliegan en una tabla de contingencia y los resultados se interpretan gráficamente a través de mapas de posicionamiento cuya dimensión puede ser reducida dependiendo de las condiciones del problema. En estos mapas se representan conjuntamente las categorías de los renglones y de las columnas de la tabla de contingencia.

El Análisis de Correspondencia “cuantifica” los datos que proporcionan información sobre los atributos o características que identifican a los objetos bajo estudio (datos cualitativos) con el fin de descubrir y/o comprender comportamientos y relaciones que no pueden percibirse inmediatamente y que al ser comparados y evaluados explican las razones o motivos que originan un fenómeno determinado. Dichos objetos pueden ser tangibles (cualquier objeto físico) o intangibles (estímulos, acciones, percepciones sensoriales o pensamientos).

Como se mencionó en el Capítulo I, cuando se tienen únicamente dos variables categóricas el método es llamado Análisis de Correspondencia Simple. Si el número de variables es mayor el método se conoce como Análisis de Correspondencia Múltiple y su procedimiento es una generalización del Análisis de Correspondencia Simple, del cual a continuación se mencionan los aspectos generales.

REPRESENTACIONES GRÁFICAS

Los medios comúnmente utilizados para desplegar y representar cierto tipo de información son las gráficas, éstas permiten visualizar los datos de una forma clara y concreta y dan una idea más general de su comportamiento.

Un conjunto de datos puede ser representado a través de gráficas de dispersión en donde las variables involucradas pueden estar definidas por distintos tipos de escala. Por

ejemplo, supóngase que en un país se van a realizar votaciones para elegir al nuevo presidente; en el Estado "X" de ese país el partido político en el poder ha determinado el tiempo que destinará a cada medio de comunicación para hacer campaña antes de que se realicen las elecciones. El partido ha decidido que durante ese tiempo se destinarán 12 días para publicidad a través de la televisión, 8 días a través de la radio y 10 días en periódicos; en total, la última etapa de campaña involucra 30 días de publicidad.

Estos datos pueden representarse gráficamente como se muestra en la figura (3.1), donde el eje vertical de la izquierda representa el número de ocasiones en las que se utilizó cada medio de comunicación y el de la derecha representa el porcentaje para cada uno de ellos; ambos son datos numéricos. El eje horizontal difiere en que representa a los diferentes medios de comunicación y son datos no métricos.

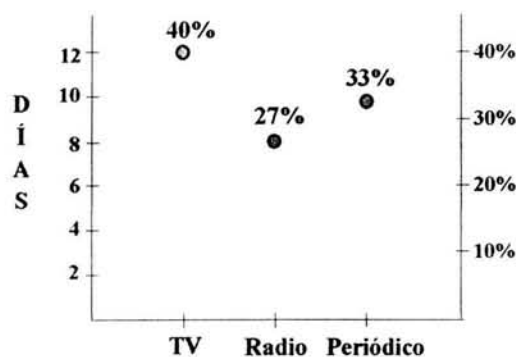


Fig. (3.1) Representación gráfica de los datos (frecuencias y porcentajes)

Dado que el eje horizontal representa una variable categórica, no tiene sentido establecer distancias entre las categorías, el hecho de que estén separadas por intervalos iguales es arbitrario; por lo tanto, no pueden interpretarse distancias entre los puntos desplegados en la gráfica. Únicamente cuando las variables que definen los ejes son de la misma naturaleza numérica y sus escalas son comparables es posible establecer distancias en cualquier dirección de la gráfica. En estos casos, las gráficas se consideran como mapas, donde pueden medirse las diferencias entre los puntos desplegados.

En la gráfica de la figura (3.1) se puede obtener la información completa y exacta de los datos analizados, es decir, no hay pérdida de información ya que la gráfica es una descripción precisa de los datos, pero no representa un medio para realizar inferencias estadísticas o servir como hipótesis sobre el comportamiento de los datos. En otras palabras,

la gráfica es un instrumento que transforma la información y que expresa los datos de una forma diferente, sin modelarlos ni resumirlos.

Ahora supóngase que en el Estado “Y” la planeación de la campaña del mismo partido ha sido distinta y se han contemplado 40 días de publicidad, distribuidos en 16 días en televisión, 10 días en radio y 14 días en periódicos. Estos datos y lo anteriores se muestran en la tabla (3.1):

Medios de Comunicación				
	Televisión	Radio	Periódico	Total
Estado X	12	8	10	30
Estado Y	16	10	14	40

Tabla (3.1) Frecuencias de los datos

Al graficar los datos se obtienen las frecuencias de cada Estado para los distintos medios de comunicación y sus porcentajes; las gráficas de la figura (3.2) se muestran por separado debido a que las bases para los porcentajes son diferentes (30 y 40 respectivamente).

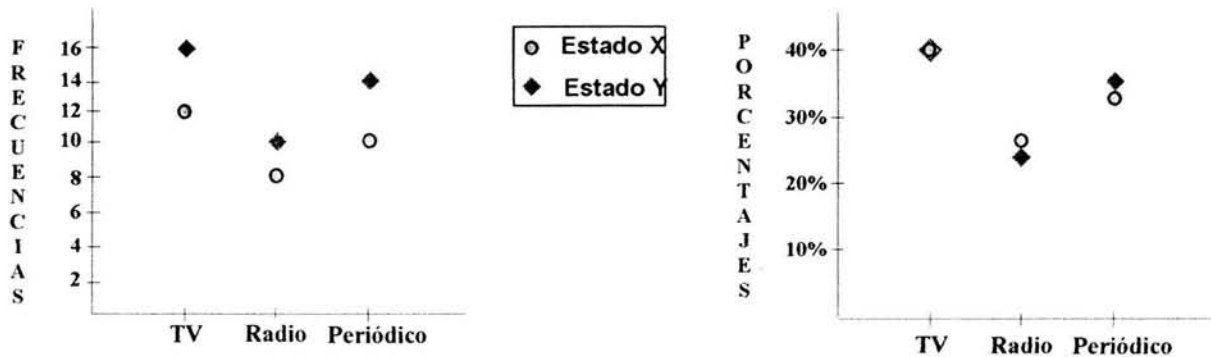


Fig. (3.2) Frecuencias y porcentajes para el Estado X y el Estado Y

Como el Estado “Y” destinó más días de publicidad las frecuencias de los medios utilizados son mayores, estos son valores absolutos. En cambio, en la gráfica de porcentajes no resulta tan inmediato descubrir cómo se comportaron las preferencias por los medios de comunicación en ambos estados, estos son los números relativos de las ocasiones en que se utilizó cada medio de comunicación; puede deducirse que en el Estado “Y” hubo mayor tendencia a utilizar el periódico, se prefirió menos el uso de la radio y el uso de la televisión tuvo la misma tendencia en ambos estados.

Las gráficas de la tabla (3.1), con 2 renglones y 3 columnas para las categorías de las variables, se visualizan de manera clara. Sin embargo, en muchas ocasiones los conjuntos de datos son más grandes; por ejemplo, si se hubieran querido analizar 8 Estados en lugar de dos y se agregaran más opciones de medios de comunicación, se tendría una tabla con un mayor número de renglones y columnas y su representación gráfica sería confusa por el número de puntos y símbolos que tendrían que ser desplegados. Debido a esto, es necesario contar con otros métodos para la representación gráfica de conjuntos grandes de datos, un aspecto fundamental en el Análisis de Correspondencia.

PERFILES DE COLUMNA Y DE RENGLÓN

Supóngase que se hace una encuesta en la que se les pregunta a 300 personas cuál compañía de telefonía celular, de las cinco opciones que se les presentan, es la de su preferencia. Las personas están clasificadas en tres categorías de acuerdo a la edad: Categoría I (de 18 a 25 años), Categoría II (de 26 a 45 años) y Categoría III (más de 45 años); así mismo, las compañías se identifican con las letras A, B, C, D y E. Los resultados obtenidos se muestran en la tabla de contingencia (3.2).

Compañía	Categorías			Total
	I	II	III	
A	20	60	0	80
B	0	0	50	50
C	0	35	35	70
D	20	35	15	70
E	10	10	10	30
Total	50	140	110	300

Tabla (3.2) Tabla de contingencia

Si se considera a cada renglón como un conjunto de frecuencias y éstas se dividen entre su total correspondiente se obtienen los **perfiles de renglón**, donde la suma de los elementos de cada perfil es 1. Es decir, sea

$$a_{ij} = \frac{n_{ij}}{n_{i+}}$$

el j -ésimo elemento del perfil del renglón i donde n_{ij} es la frecuencia observada de la ij -ésima celda y n_{i+} es el total marginal del renglón i , entonces:

$$\sum_j a_{ij} = 1 \quad \forall i$$

En la tabla (3.3a) se muestran los perfiles de renglón relacionados con la tabla anterior. Por ejemplo, para la compañía D el perfil correspondiente es (0.29 0.50 0.21), que significa que del total de las personas que prefieren la compañía D (70), el 29% tiene entre 18 y 25 años, el 50% tiene entre 26 y 45 años y el 21% tiene más de 45 años. Lo mismo sucede para cada compañía. Los perfiles de renglón son los perfiles de las compañías contra las categorías de edad. El último renglón de la tabla (3.2) es el total para cada una de las categorías que fueron consideradas y, al igual que para cada compañía, es posible obtener su perfil dividiendo las frecuencias entre el total general, obteniéndose los valores (0.17 0.46 0.37). Este perfil representa las proporciones para cada categoría independientemente de las compañías y es llamado **perfil promedio de renglón**.

De la misma forma, las columnas pueden considerarse como conjuntos de frecuencias y obtener los **perfiles de columna**; sea

$$b_{ij} = \frac{n_{ij}}{n_{+j}}$$

el *i*-ésimo elemento del perfil de la columna *j* donde n_{+j} es el total marginal de la columna *j*, entonces:

$$\sum_i b_{ij} = 1 \quad \forall j$$

En la tabla (3.3b) se muestran los perfiles de columna o los perfiles de las categorías de edad contra las compañías, donde para la categoría I (50 en total) se obtiene el perfil (0.40 0 0 0.40 0.20) y el **perfil promedio de columna** es (0.27 0.17 0.23 0.23 0.10).

Compañía	Categorías		
	I	II	III
A	0.25	0.75	0
B	0	0	1
C	0	0.50	0.50
D	0.29	0.50	0.21
E	0.33	0.33	0.33
Promedio	0.17	0.46	0.37

Tabla (3.3a) Perfiles de renglón

Compañía	Categorías			Promedio
	I	II	III	
A	0.40	0.43	0	0.27
B	0	0	0.45	0.17
C	0	0.25	0.32	0.23
D	0.40	0.25	0.14	0.23
E	0.20	0.07	0.09	0.10

Tabla (3.3b) Perfiles de columna

La elección de trabajar con perfiles de renglón o de columna dependerá de lo que se necesite analizar. Si lo que se considera más importante son las compañías, deben elegirse los perfiles por renglón, pero si el comportamiento de las categorías de edad con respecto a sus preferencias es de mayor interés lo mejor es trabajar con los perfiles de columna. De cualquier forma, la interpretación del análisis por renglón o por columna es equivalente y las conclusiones son similares.

Considerando que los perfiles promedio son los “valores esperados” para cada renglón o columna (según sea el caso), se puede notar si existen diferencias significativas entre éstos y los perfiles de cada categoría. Por ejemplo, para el perfil de la compañía A (renglón) se tiene un valor en la categoría II muy por arriba del promedio de renglón, es decir, hay una proporción mayor de personas entre 26 y 45 años (75%) que prefieren la compañía A que el promedio (46%). En los perfiles de columna también se observa esta diferencia, ya que mientras que en el perfil promedio de columna la preferencia para la compañía A es 27%, en el perfil de la categoría II el 43% prefiere la compañía A.

Si se deseara representar gráficamente los datos de la tabla (3.2) de la misma forma en que se hizo en el ejemplo de los Estados “X” y “Y” (considerando los porcentajes únicamente) se obtendría la siguiente gráfica:

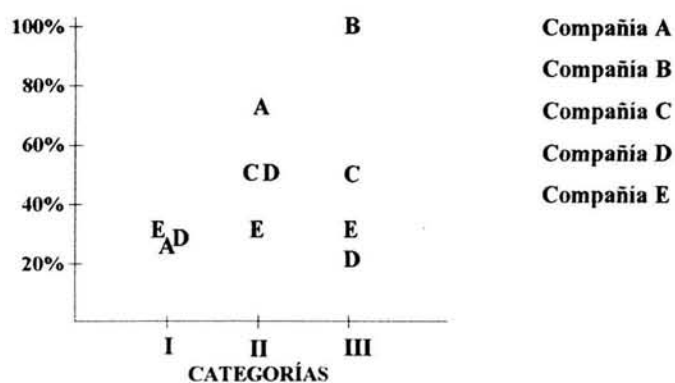


Fig. (3.3) Porcentajes de preferencia de las compañías

La interpretación de esta gráfica resulta confusa, aún cuando son sólo 5 los renglones considerados; además, esta representación no permite realizar un análisis más profundo del comportamiento de las categorías, lo único que se muestra son los porcentajes que tiene cada compañía en cada categoría de edad.

Una manera diferente de graficar los datos es utilizando los perfiles de renglón. Los ejes que definen el espacio donde se posicionarán las compañías son los que representan a las categorías de edad, en este caso son 3, por lo que se obtendrá un espacio tridimensional. Sobre cada eje de este espacio descansan vectores unitarios o puntos unitarios que al unirse forman un triángulo equilátero. Cada perfil de renglón consta de 3 elementos y éstos serán las coordenadas que definirán la posición de los puntos que representan a cada compañía en el mapa tridimensional. La figura (3.4) muestra los perfiles en el espacio definido por los 3 ejes de las categorías I, II y III.

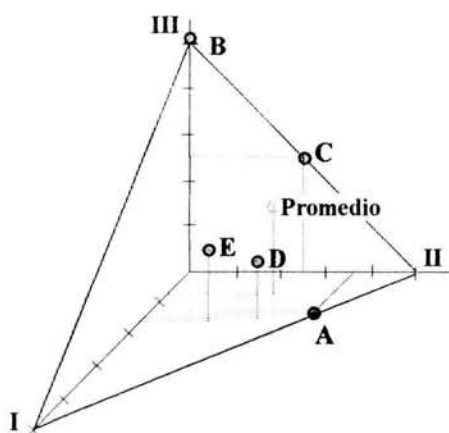


Fig. (3.4) Gráfica de los perfiles de renglón y perfil promedio de renglón

Por ejemplo, el perfil de la compañía C es $(0 \ 0.50 \ 0.50)$, indicando que está posicionado a cero unidades sobre el eje I, a media unidad a partir del eje II y a media unidad a partir del eje III. De igual forma, el perfil promedio de renglón se posiciona con las coordenadas $(0.17 \ 0.46 \ 0.37)$. El perfil de la compañía B cae exactamente en el vértice del eje III, ya que la totalidad de las personas para este perfil se concentra en esa categoría de edad; éste es un **perfil extremo**.

Es de notarse que los 6 puntos caen en el plano definido por los puntos extremos $(1 \ 0 \ 0)$, $(0 \ 1 \ 0)$ y $(0 \ 0 \ 1)$ que son los vértices. Extrayendo esta región se obtiene un sistema de coordenadas triangulares representado en la figura (3.5)

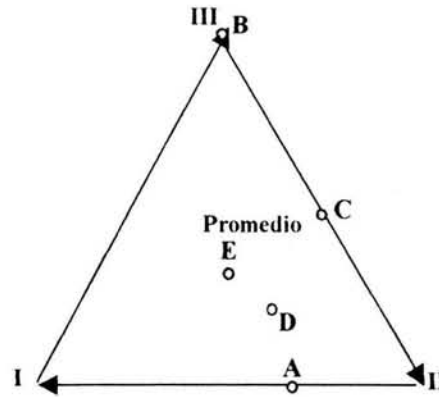


Fig. (3.5) Sistema de coordenadas triangulares con los perfiles de renglón

Los perfiles pueden posicionarse directamente en el triángulo sin ser necesario desplegarlos primero en el espacio tridimensional. Dado que el triángulo es equilátero, sobre cada uno de sus lados se recorre una distancia igual al valor del elemento del perfil correspondiente y después se trazan líneas paralelas a los respectivos lados del triángulo, estas líneas se intersectan en un punto que representa el perfil de la compañía. Por ejemplo, para posicionar el perfil de la compañía D (0.29 0.50 0.21), se recorre una distancia de 0.29 sobre el eje I, 0.50 sobre el eje II y 0.21 sobre el eje III, las líneas paralelas a partir de estos puntos se intersectan en el punto D como se muestra en la figura (3.6).

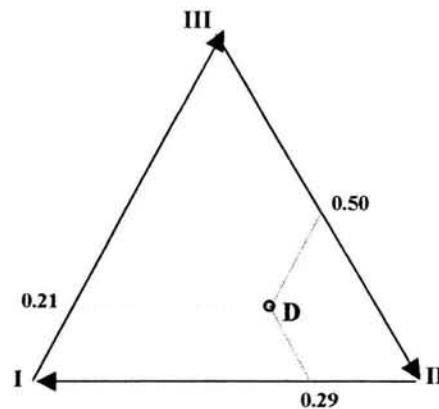


Fig. (3.6) Posición del punto de perfil de renglón

Este procedimiento puede hacerse, sin embargo, sólo cuando los perfiles tienen 3 elementos. Si el perfil consta de m elementos, los puntos se graficarán en un espacio de dimensión $m-1$. Este espacio es una región restringida por la unión de todos los pares de m vectores unitarios que están sobre los m ejes perpendiculares, donde cada punto unitario representa un vértice del espacio. De esto se deduce que si un perfil tiene 4 elementos, aún podrá representarse en un espacio tridimensional, pero la representación gráfica no será posible para perfiles con un número mayor de elementos.

MASAS

Las posiciones que ocupan los puntos de perfil en la figura (3.5) en relación con los perfiles respectivos de la tabla (3.3a) se interpretan de una forma sencilla. Cada perfil está compuesto por 3 valores, cada uno de los cuales se refiere a una categoría de edad representadas en los vértices. La posición del punto de perfil dependerá de estos valores tendiendo a estar más cerca de los vértices para los cuales sus valores correspondientes son mayores. Es decir, cada valor en el perfil le da un peso a cada vértice, y el punto de perfil será "atraído" en mayor medida hacia el vértice que tenga mayor peso. Por ejemplo, el punto A se encuentra más cerca del vértice II que del vértice I y a la mayor distancia del vértice III. Para el punto E, los 3 vértices tienen el mismo peso y por lo tanto contribuyen de la misma manera en el perfil, colocándolo exactamente en el centro del triángulo. Un caso extremo es el punto B que cae exactamente en el vértice III pues todo su peso recae en esa categoría de edad.

Las posiciones de los perfiles son los **centroides** o **baricentros** (centros de gravedad) de los vértices, es decir, son promedios ponderados de los vértices, donde los pesos están dados por cada uno de los elementos (valores) del perfil.

El punto de perfil promedio no sólo es un centroide de los vértices, sino que también es un centroide de los perfiles. La posición del perfil promedio representa el promedio de todos los elementos de la tabla (300) concentrados en diferentes locaciones representadas por los perfiles (5 en este caso). Por lo tanto, para considerar al perfil promedio de renglón como un centroide de los perfiles de renglón debe tenerse en cuenta que cada uno de ellos tiene una masa asociada que lo pondera. Cada perfil tiene asociado un valor (total por renglón) que contribuye al total general, por ejemplo, la compañía A contribuye con 80 personas, la B con 50 y así sucesivamente. Los pesos asignados para cada perfil son, por lo tanto, $(0.27 \ 0.17 \ 0.23 \ 0.23 \ 0.10)$, donde $0.27=80/300$, $0.17=50/300$, etc. Estos pesos reciben un nombre especial: **masas**. Los pesos asignados a cada perfil de renglón son las masas por renglón y, como puede observarse, éstas son iguales a los valores del perfil promedio de columna.

La misma lógica se aplica en el caso del análisis de columna, donde las masas por columna serían (0.17 0.46 0.37), que son iguales al perfil promedio de renglón. Por lo tanto, denotando por r_i a la masa del i -ésimo renglón, se tiene que:

$$r_i = \frac{n_{i+}}{n}$$

donde n_{i+} es el total marginal del renglón i y n es el total de observaciones, además r_i es el i -ésimo elemento del perfil promedio de columna, por ejemplo, $r_1=0.27$, $r_2=0.17$. Análogamente, si c_j es la masa de la j -ésima columna entonces:

$$c_j = \frac{n_{+j}}{n}$$

donde n_{+j} es el total marginal de la columna j y n es el total de observaciones, además c_j es el j -ésimo elemento del perfil promedio de renglón, por ejemplo, $c_1=0.17$, $c_2=0.46$.

DISTANCIA χ^2

El mapa de la figura (3.5) se originó a partir de un espacio tridimensional en donde se posicionaron los puntos de los perfiles de renglón. La posición de estos puntos da una idea de cómo se comportan los datos y el punto de perfil promedio sirve como referencia para interpretarlos. Sin embargo, las relaciones o asociaciones no sólo se dan entre los puntos de perfil y el perfil promedio, sino también entre éstos y los vértices. En un mapa construido como el de las figuras (3.4) y (3.5) se está considerando que las categorías que conforman los vértices tienen el mismo peso, es decir, no se establecen diferencias entre ellas en cuanto a las frecuencias que cada una pueda tener. Resulta claro que en la mayoría de los casos las frecuencias de las categorías que definen los vértices del mapa son distintos. Esto también debe reflejarse en los mapas para que sea posible hacer una interpretación más adecuada y obtener información más precisa.

Las distancias Euclidianas no son lo suficientemente flexibles para permitir un tratamiento geométrico ni estadístico que permita reflejar estos aspectos, sin embargo, bajo ciertas transformaciones, es posible generar un método que tenga la robustez necesaria para alcanzar estos objetivos satisfactoriamente.

En el Capítulo I se mencionaron aspectos relativos a las tablas de contingencia y las operaciones que pueden calcularse sobre los datos. Entre ellos se mencionó el estadístico χ^2 que mide la discrepancia entre las frecuencias observadas y las esperadas bajo una hipótesis de independencia. Bajo ese mismo orden de ideas, pero en el contexto del Análisis de Correspondencia, se señala la habilidad del estadístico χ^2 para medir la heterogeneidad en términos de perfiles y orientarlo en un sentido geométrico.

Para calcular la distancia Euclidiana entre dos puntos de perfil que se encuentran en un espacio multidimensional (como el de la figura (3.4)) se aplica la fórmula:

$$d = \sqrt{\sum_j (x_j - y_j)^2} \quad (1)$$

donde x_j son los elementos del perfil del primer punto y y_j son los elementos del perfil correspondiente al segundo punto.

Recordando el cálculo del estadístico χ^2 se observa que:

$$\chi^2 = \sum_i \sum_j \frac{(n_{ij} - E_{ij})^2}{E_{ij}} \quad \text{para } i=1\dots r \quad j=1\dots c \quad (2)$$

Desarrollando la fórmula (2) se obtiene una suma de rc términos (r renglones y c columnas) que pueden agruparse en r grupos de c términos cada uno (un grupo por cada renglón de la tabla), donde cada grupo tendría la forma:

$$\dots + \left[\frac{(n_{i1} - E_{i1})^2}{E_{i1}} + \frac{(n_{i2} - E_{i2})^2}{E_{i2}} + \dots + \frac{(n_{ic} - E_{ic})^2}{E_{ic}} \right] + \dots \quad (3)$$

Cada uno de los r grupos de términos se asemeja al cálculo del cuadrado de una distancia Euclidiana entre dos puntos, excepto que cada término está dividido por un factor, lo que le da una connotación diferente. En general, cada término es de la forma:

$$\frac{(n_{ij} - E_{ij})^2}{E_{ij}} \quad (4)$$

Dividiendo el numerador y el denominador entre el cuadrado del total por renglón, es decir, el total marginal n_{i+} , entonces:

$$\frac{(n_{ij}/n_{i+} - E_{ij}/n_{i+})^2}{E_{ij}/n_{i+}^2} \quad (5)$$

Pero n_{ij}/n_{i+} es el j -ésimo elemento del perfil del renglón i , denotado por a_{ij} . Además, se sabe que $E_{ij} = \frac{(n_{i+} \cdot n_{+j})}{n}$, entonces $\frac{E_{ij}}{n_{i+}} = \frac{n_{+j}}{n}$ que es la masa de la j -ésima columna c_j . Por lo tanto, de (5) se obtiene:

$$n_{i+} \left[\frac{(a_{ij} - c_j)^2}{c_j} \right] \quad (6)$$

Entonces, cada grupo de términos en (3) queda como:

$$n_{i+} \left[\sum_j \frac{(a_{ij} - c_j)^2}{c_j} \right] \quad (7)$$

Si en la expresión entre corchetes no aparecieran los denominadores c_j , se estaría calculando el cuadrado de la distancia Euclidiana entre el perfil del renglón i y el perfil promedio de renglón. El efecto que producen los denominadores en cada término es ponderar cada diferencia cuadrada. Cuando los factores de ponderación son los elementos del perfil esperado en el denominador, la distancia Euclidiana ponderada se conoce como **distancia chi-cuadrada**.

Entonces si \mathbf{a}_i es el perfil del renglón i y \mathbf{c} el perfil promedio por renglón, la distancia χ^2 entre ambos es:

$$\|\mathbf{a}_i - \mathbf{c}\|_c = \sqrt{\sum_j \frac{(a_{ij} - c_j)^2}{c_j}} \quad (8)$$

La distancia χ^2 entre dos perfiles de renglón \mathbf{a}_i y \mathbf{a}_r está dada por:

$$\|\mathbf{a}_i - \mathbf{a}_r\|_c = \sqrt{\sum_j \frac{(a_{ij} - a_{rj})^2}{c_j}} \quad (9)$$

De forma equivalente, en un análisis por columna, si \mathbf{b}_j es el perfil de la columna j y \mathbf{r} el perfil promedio de columna, la distancia χ^2 es:

$$\|\mathbf{b}_j - \mathbf{r}\|_r = \sqrt{\sum_i \frac{(b_{ij} - r_i)^2}{r_i}} \quad (10)$$

La distancia χ^2 entre dos perfiles de columna \mathbf{b}_j y \mathbf{b}_j' es:

$$\|\mathbf{b}_j - \mathbf{b}_j'\|_r = \sqrt{\sum_i \frac{(b_{ij} - b_{ij}')^2}{r_i}} \quad (11)$$

Distancias χ^2 en el Espacio Euclidiano

La figura (3.4) muestra los puntos de perfil desplegados en un espacio Euclidiano, donde las distancias entre ellos son distancias Euclidianas calculadas con la fórmula (1). Sobre este mismo espacio Euclidiano se pueden graficar los puntos considerando las distancias χ^2 , para ello se toma en cuenta que las coordenadas originales (elementos de los perfiles) están ponderados por los elementos del perfil promedio. Así, los elementos $(x_j - y_j)^2$

son transformados en la forma $\frac{(x_j - y_j)^2}{c_j}$, o bien $\left(\frac{x_j}{\sqrt{c_j}} - \frac{y_j}{\sqrt{c_j}} \right)^2$, es decir, los elementos de

los perfiles serán divididos entre la raíz cuadrada del elemento correspondiente del perfil promedio.

Dado que los valores de c_j son menores a 1, los valores de las coordenadas se incrementarán dependiendo del valor de c_j . Aquellos términos que contengan un valor c_j relativamente menor a los demás incrementarán los valores de sus coordenadas en una cantidad relativamente mayor, y, por el contrario, si el valor de c_j es mayor a los demás, el valor de las coordenadas se incrementará en menor medida. En otras palabras, las categorías que tengan una menor frecuencia aumentarán más el valor de sus coordenadas que aquellas que observen una frecuencia mayor.

Gráficamente, la transformación para visualizar la distancia χ^2 se hace ajustando los ejes sobre los que se va a graficar. En la figura (3.4), los vértices se encuentran a una unidad de distancia del origen generando un triángulo equilátero (puntos extremos). Al hacer la transformación de las coordenadas, los vértices ocuparán las posiciones definidas por los valores de c_j .

Las coordenadas originales de los vértices son vectores unitarios, estos vectores ahora serán ponderados por las masas de las columnas. Por ejemplo, los valores del perfil promedio son (0.17 0.46 0.37); el primer vértice, con coordenadas (1 0 0) se posicionará en el punto (2.43 0 0) ya que $1/\sqrt{c_1} = 1/\sqrt{0.17} = 2.43$, de igual forma, el segundo vértice tendrá coordenadas (0 1.47 0) y el tercero (0 0 1.64). Con esta transformación, las distancias del origen a cada vértice no son iguales, lo que generará que el triángulo que definen tenga diferentes longitudes en sus lados y sobre éste se posicionarán los puntos de

perfil utilizando las coordenadas originales pero considerando las nuevas escalas de los ejes o, lo que es lo mismo, dividiendo cada elemento del perfil de renglón entre la raíz cuadrada del elemento del perfil promedio correspondiente ($\sqrt{c_i}$), como se muestra en la figura (3.7).

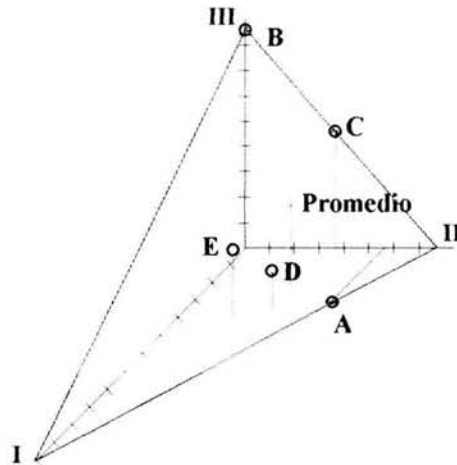


Fig. (3.7) Perfiles de renglón en el sistema de coordenadas tridimensional, con los ejes ajustados a las distancias χ^2

El plano triangular en el que se despliegan los perfiles de renglón ya no es equilátero; al considerar las distancias χ^2 los ejes se ajustan de acuerdo a la ponderación dada por los elementos c_i ; el plano ajustado se muestra en la figura (3.8).

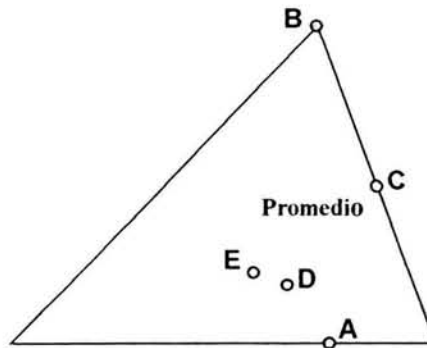


Fig. (3.8) Plano triangular con distancias χ^2

Las escalas sobre los ejes “ajustados” no son idénticas, por lo que un intervalo fijo sobre un eje puede ser mayor o menor que el mismo intervalo en los otros dos ejes. De esta manera, las distancias Euclidianas están ponderadas por los valores de las masas, constituyendo distancias χ^2 .

Al considerar las distancias χ^2 se pretende estandarizar los datos debido a la existencia de discrepancias inherentes en las varianzas de conjuntos de frecuencias. Una forma de estabilizar las varianzas es dividiendo las frecuencias entre la raíz cuadrada de las frecuencias esperadas.

La preferencia por el uso de la distancia χ^2 también se debe a que ésta es la única que posee la propiedad de garantizar estabilidad en las distancias entre renglones cuando las columnas de la tabla son particionadas, o bien, fusionadas; este es el "Principio de Equivalencia Distribucional". Si, por ejemplo, una columna o renglón es particionado, la distancia χ^2 no será afectada por tal subdivisión y así las columnas o renglones que tengan perfiles parecidos y que tienen casi la misma posición en el espacio, pueden agruparse sin afectar la geometría de los renglones o columnas.

INERCIA

Los ejemplos anteriores han mostrado aspectos referentes a la posición de los puntos de perfil en los espacios multidimensionales. Ahora resulta necesario establecer un medio por el cual se pueda conocer el grado de dispersión de los puntos de perfil y esto se hace, precisamente, a través del estadístico χ^2 .

Retomando la expresión (7), la fórmula del estadístico χ^2 sería:

$$\chi^2 = \sum_i n_{i+} \sum_j \left[\frac{(a_{ij} - c_j)^2}{c_j} \right] \quad (12)$$

El término n_{i+} (totales marginales por renglón) multiplica al cuadrado de cada distancia χ^2 involucrada en la fórmula, si éste es dividido entre el total de la tabla (n), entonces se obtiene la masa del i -ésimo renglón, es decir $r_i = n_{i+}/n$.

La fórmula (13) representa, entonces, un promedio ponderado del cuadrado de las distancias χ^2 entre los perfiles de renglón y su perfil promedio, expresado como:

$$\begin{aligned} \frac{\chi^2}{n} &= \sum_i r_i \sum_j \left[\frac{(a_{ij} - c_j)^2}{c_j} \right] \\ &= \sum_i r_i \| \mathbf{a}_i - \mathbf{c} \|_c^2 \end{aligned} \quad (13)$$

Este valor recibe el nombre de **inercia** o **inercia total**, y mide el grado de dispersión entre los puntos de perfil. De este modo la inercia también se obtiene al dividir el valor del estadístico χ^2 entre el total general de la tabla de contingencia (χ^2/n).

Para el análisis por columna se obtiene un resultado equivalente:

$$\begin{aligned} \frac{\chi^2}{n} &= \sum_j c_j \sum_i \left[\frac{(b_{ij} - r_i)^2}{r_i} \right] \\ &= \sum_j c_j \|\mathbf{b}_j - \mathbf{r}\|_r^2 \end{aligned} \quad (14)$$

La inercia de los puntos está en el rango $[0, n-1]$ donde $n-1$ es la dimensión del espacio sobre el que se posicionan los puntos de perfil en una matriz de $m \times n$ con $n < m$. Entre mayor sea el valor de la inercia, mayor es la dispersión de los puntos y éstos caen más cerca de los vértices, indicando mayor asociación o correlación entre renglones y columnas. El mínimo valor que puede obtenerse es cero y resulta cuando los perfiles son iguales y caen en el mismo punto (perfil promedio). Si todos los puntos de perfil caen exactamente en los vértices se alcanza el máximo valor de la inercia.

De esta forma, se muestra la relación entre el estadístico χ^2 y la inercia; si en lugar de asociar a cada distancia cuadrada con su masa respectiva se asocia con su total marginal que es n veces la masa por renglón ($n_{i+} = nr_i$), entonces la suma ponderada es igual al estadístico χ^2 , tal como se muestra en (12). Así, el estadístico χ^2 y la inercia pueden interpretarse geoméricamente como el grado de dispersión de un conjunto de puntos de perfil respecto a su promedio. En el caso de la tabla (3.2) la inercia de los puntos de perfil es $\chi^2/n = 0.549$, representada en la figura (3.8). El máximo valor que podría alcanzarse es 2.

ANÁLISIS DE INERCIA

La inercia es un punto fundamental en el Análisis de Correspondencia ya que de ella se desprende todo un análisis que permite obtener información de los mapas que se generan como resultado de la aplicación del método.

Descomposición de la Inercia

Supóngase que, en términos generales, la distancia χ^2 entre un punto de perfil a_i y el

perfil promedio c es $d_i = \sqrt{\sum_j \frac{(a_{ij} - c_j)^2}{c_j}}$ representada en la figura (3.9).

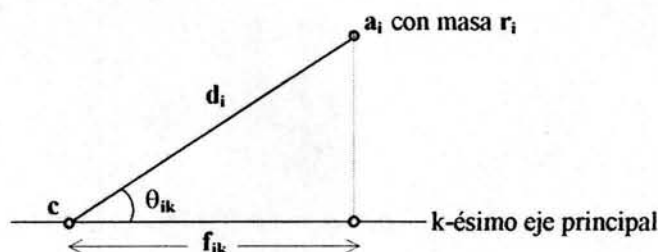


Fig. (3.9) Distancia χ^2 de un punto de perfil al perfil promedio y proyección sobre un eje principal

Entonces la inercia total es:

$$\frac{\chi^2}{n} = \sum_i r_i \sum_j \frac{(a_{ij} - c_j)^2}{c_j} = \sum r_i d_i^2$$

La inercia total puede descomponerse de varias maneras y obtener distintos elementos que contribuyen con una cierta cantidad al total de la variación. Así, se tienen las contribuciones por celda, inercias por renglón o columna y las inercias principales. Estas cantidades pueden expresarse como porcentajes de inercia, explicando una parte del total de ésta.

Contribuciones por Celda

Las contribuciones a la inercia total que hace cada una de las celdas de una tabla de contingencia es la cantidad que resulta del cuadrado de la diferencia ponderada entre un elemento del perfil y el elemento correspondiente del perfil promedio, multiplicada por la masa de ese perfil (r_i), es decir:

$$\text{Contr. de la Celda } ij = r_i \frac{(a_{ij} - c_j)^2}{c_j}$$

que representa a cada uno de los elementos para calcular la distancia entre un perfil y el perfil promedio. Esta contribución se divide entre la inercia total (χ^2/n) para obtener el porcentaje de inercia con el cual contribuye la celda. Las celdas cuya diferencia $(a_{ij} - c_j)$ sea

mayor, contribuirán en mayor medida a la inercia, ya que, en el sentido de la hipótesis de independencia, éstos valores son los más improbables.

Inercias de Renglón y Columna

Otra descomposición de la inercia es la que se hace con respecto a los renglones y columnas. La inercia de renglón (INR) es la distancia cuadrada entre éste y el perfil promedio multiplicada por la masa, es decir:

$$INR = r_i \sum_j \frac{(a_{ij} - c_j)^2}{c_j} = r_i d_i^2$$

donde d_i es la distancia representada en la figura (3.9).

La contribución relativa a la inercia total es el porcentaje $\frac{r_i d_i^2}{\chi^2/n}$. Equivalentemente,

esto se aplica en el caso de las columnas.

Inercias Principales

Las inercias principales son las de mayor importancia en el Análisis de Correspondencia. Cada eje que se calcula al reducir la dimensión del espacio original y que mejor se ajusta a los datos en términos de mínimos cuadrados es un **eje principal**; éstos son los vectores \mathbf{t} que se obtienen al calcular los eigenvectores de una matriz cuadrada simétrica.

La cantidad de inercia que aporta cada eje principal se llama **inercia principal**. Así, la k -ésima inercia principal es la suma ponderada de las distancias cuadradas de las proyecciones de los perfiles de renglón (o columna) sobre el k -ésimo eje principal con respecto al perfil promedio o centroide.

Gráficamente, como se muestra en la figura (3.9), la k -ésima inercia principal es:

$$\lambda_k = \sum_i r_i f_{ik}^2$$

donde f_{ik} es la coordenada principal del perfil a_i con respecto al k -ésimo eje principal (la proyección se encuentra a una distancia f_{ik} del centroide). Cada f_{ik}^2 representa a los k -ésimos componentes de d_i^2 . Los valores de λ_k corresponden a los eigenvalores de la matriz cuadrada simétrica que indican la magnitud del ajuste realizado a través de la transformación para obtener una nueva base de vectores ortogonales (vectores t) y corresponden a la inercia explicada por cada uno de ellos. Dependiendo del número de ejes se tendrá un "primer eje principal", un "segundo eje principal", etc., y todos tienen como origen el perfil promedio. Cada uno de ellos aporta una cantidad a la inercia total, siendo el primer eje el que contribuye con un mayor porcentaje, el segundo eje principal aporta la mayor cantidad de la inercia restante y así sucesivamente. Es por esto que los eigenvalores de la matriz de transformación están ordenados de mayor a menor.

De lo anterior se concluye que la suma de las inercias de renglón o de columna y la suma de las inercias principales es igual a la inercia total, es decir:

$$\frac{\chi^2}{n} = \sum_i r_i d_i^2 = \sum_k \lambda_k$$

Las contribuciones a la inercia (inercias principales) pueden expresarse como contribuciones absolutas o relativas.

Contribuciones Absolutas

Las contribuciones absolutas son las contribuciones de los perfiles a los ejes principales en donde se observa cuáles son los puntos que han determinado en mayor medida la orientación de los ejes principales y son representadas por cada término $r_i f_{ik}^2$ (contribución del i -ésimo renglón al k -ésimo eje principal) del cálculo de λ_k , es decir, se tienen i contribuciones absolutas para cada uno de los k ejes. La contribución de cada perfil indicará cuáles de ellos han influido más en determinar la posición del k -ésimo eje, lo que dependerá de la posición del perfil y de su masa asociada.

Contribuciones Relativas

Las contribuciones relativas son las contribuciones de los ejes principales a los perfiles; éstas muestran qué tan bien está representado un punto en el mapa dependiendo del porcentaje de la contribución para cada eje; si el porcentaje es relativamente alto, la

interpretación del perfil será confiable, mientras que si la contribución es baja los puntos no están explicados adecuadamente y esto debe ser considerado al interpretarlos.

Para calcular las contribuciones relativas se aplica el Teorema de Pitágoras para descomponer las distancias cuadradas en una suma de componentes (uno para cada eje), de la siguiente forma:

$$d_i^2 = \sum_k f_{ik}^2$$

Descomponiendo la inercia de renglón en una suma de términos para cada eje principal:

$$r_i d_i^2 = r_i \sum_k f_{ik}^2$$

cada término $r_i f_{ik}^2$ de la inercia del renglón es la contribución del k -ésimo eje a la inercia del i -ésimo renglón, por lo que se tendrán k contribuciones relativas para cada uno de los i puntos de perfil.

Estas contribuciones también pueden interpretarse geoméricamente como un coeficiente de correlación entre los puntos de perfil y los ejes principales; así, la relación f_{ik}/d_i es el coseno del ángulo θ_{ik} mostrado en la figura (3.9); el cuadrado del coseno es la contribución del k -ésimo eje principal a la inercia del i -ésimo renglón.

$$\cos^2 \theta_{ik} = \left(\frac{f_{ik}}{d_i} \right)^2$$

Calidad de la Gráfica en un Subespacio

Si un vector que va del centroide hacia un punto de perfil, con una distancia d_i se descompone en k componentes, entonces la suma de los cuadrados de los cosenos de los ángulos que forma d_i con cada eje principal debe ser igual a 1 para que se explique la inercia total del renglón (o columna).

$$\cos^2 \theta_{i1} + \cos^2 \theta_{i2} + \dots + \cos^2 \theta_{ik} = 1$$

Dado que los puntos de perfil no pueden ser representados gráficamente en el espacio multidimensional original de dimensión m (donde $m = \min(r,c) - 1$ en una tabla de $r \times c$), salvo en algunos casos, sino que son proyectados sobre un subespacio para poder visualizarlos, no se conocen sus posiciones reales. Al reducir la dimensión del espacio se

pierde información y esta pérdida se cuantifica como un porcentaje de la inercia total. Este porcentaje mide la calidad del subespacio para representar los puntos. Entre menor sea la pérdida de información, mayor será la calidad de la aproximación del subespacio.

La calidad total de la representación de un punto de perfil en una gráfica se obtiene sumando las contribuciones relativas para el número de ejes que la componen, dando el total del porcentaje de inercia explicado para esos puntos. Si algunos perfiles no están adecuadamente representados significa que éstos caen más sobre los ejes sucesivos; entre más cerca estén del subespacio, mejor serán representados por éste, si un punto está muy alejado, éste contribuirá en mayor medida a la pérdida de información y se tendrá un residuo o error más grande. Aunque geoméricamente sólo es posible visualizar los puntos hasta el eje $k=3$, numéricamente sí se puede conocer la contribución de todos los ejes y saber cuáles son los puntos que se encuentran más alejados del eje, plano o espacio tridimensional y que, por lo tanto, sus posiciones no son reflejadas adecuadamente. Conforme la dimensión del subespacio aumenta, mejor representados están los puntos de perfil. En una gráfica, las distancias observadas de los puntos proyectados siempre serán menores o iguales a las distancias reales, porque estas últimas son la suma de un conjunto de m componentes cuadrados, mientras que las distancias observadas son la suma de un número menor de estos componentes.

Anidamiento de los Ejes Principales

Aunque no siempre sea posible graficar todos los ejes del subespacio, un subespacio de cualquier dimensión que represente satisfactoriamente a los puntos (óptimo) contiene a otro subespacio óptimo de menor dimensión. Es decir, el primer eje principal que conforma al espacio unidimensional óptimo y que explica la mayor inercia estará contenido en el espacio bidimensional, donde el segundo eje principal explica la mayor inercia restante; los dos ejes que conforman a éste serán parte, a su vez, del espacio óptimo en tres dimensiones y así sucesivamente. Por esto, el espacio óptimo de mayor dimensión contendrá a todos los subespacios óptimos de menor dimensión.

ANÁLISIS POR RENGLÓN Y POR COLUMNA

Se ha mencionado que cuando se realiza un análisis por renglón también se analiza en forma simultánea el análisis por columna. El problema en el Análisis de Correspondencia consiste en encontrar el subespacio óptimo que represente mejor a los puntos de perfil. En el caso de la figura (3.7), por ejemplo, los ejes que mejor se ajustan a los puntos se muestran en la figura (3.10). Este es un análisis por renglón y deben obtenerse resultados equivalentes en un análisis por columna. La relación entre ambos análisis, aplicados a una misma tabla de contingencia, se observa en tres aspectos fundamentales: la dimensión del subespacio obtenido, la inercia total o inercias principales y las posiciones relativas.

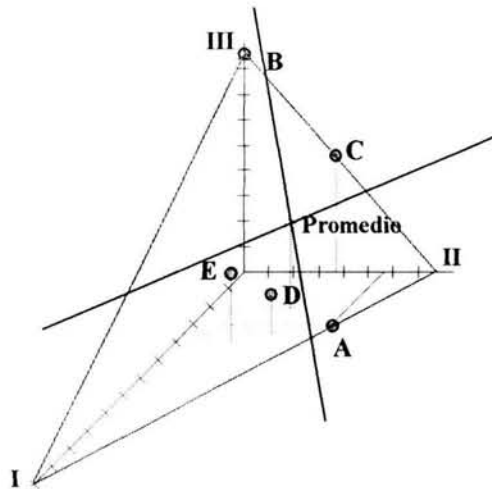


Fig. (3.10) Ejes que mejor se ajustan a los puntos del plano

Misma Dimensión

El mapa bidimensional que resulta al considerar los ejes de la figura (3.10) es el que se muestra en el panel I de la figura (3.11). Este es un análisis de renglón y así como los perfiles son proyectados en un subespacio, los vértices también pueden representarse en éste presentando una mayor dispersión dado que son los perfiles extremos y servirán como referencia para interpretar la dispersión de los puntos de perfil, cuyo valor de inercia es 0.549.

Si en lugar de analizar los perfiles de renglón se analizaran los perfiles de columna se tendrían 5 vértices (puntos de renglón) que formarían el espacio multidimensional sobre el cual se posicionan los 3 puntos de perfil correspondientes a las columnas. Sin embargo, estos puntos no "llenan" el espacio de cuatro dimensiones que forman los vértices sino que

son proyectados en un plano del mismo modo en que fueron proyectados los perfiles de renglón. Esto está relacionado con el rango de la matriz a partir de la cual se obtienen los ejes que describen las inercias principales λ_k . En el panel II de la figura (3.11) se muestra el plano que mejor se ajusta en un análisis por columna.

Los subespacios sobre los cuales se proyectan los perfiles de renglón y los perfiles de columna no son los mismos, pero tienen la misma dimensión.

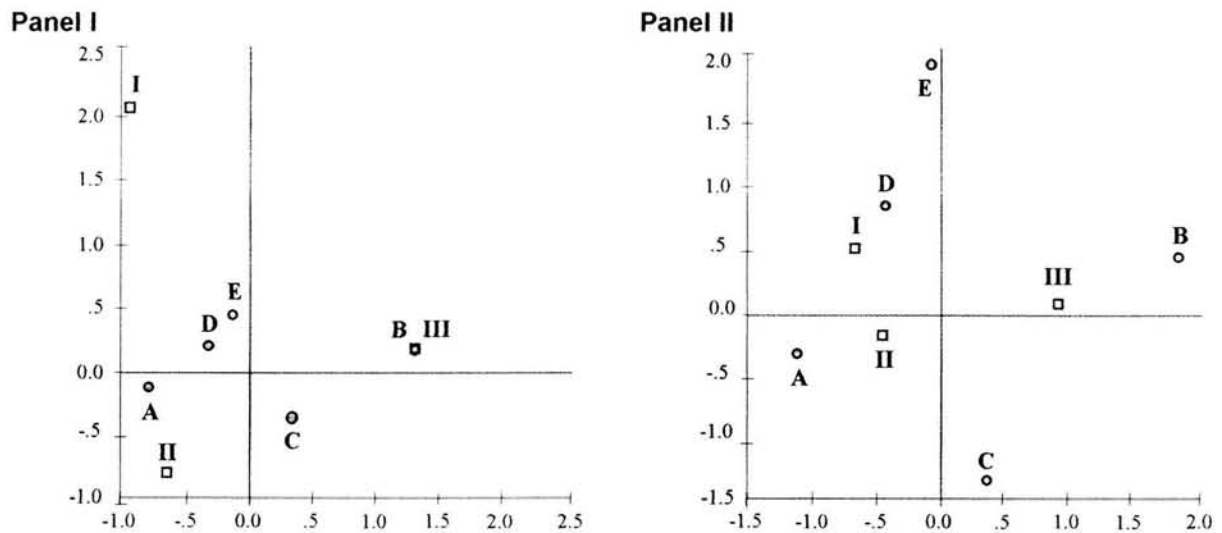


Fig. (3.11) Mapas de análisis por renglón y análisis por columna

Misma Inercia Total e Inercias Principales

Además de tener la misma dimensión, en los dos análisis se tiene la misma cantidad de inercia total. En cada uno de ellos se obtiene el mismo número de ejes que mejor se ajustan a los puntos para los cuales los valores de λ_k son los mismos. El k-ésimo eje principal en un análisis por renglón tiene la misma inercia principal que el k-ésimo eje principal en un análisis por columna, es decir, la inercia total se descompone en inercias principales de la misma forma para cualquier tipo de análisis, y por lo tanto, la calidad de la representación gráfica será la misma en ambos.

Mismas Posiciones Relativas

Observando de nuevo la figura (3.11) se percibe otro aspecto en el que se relacionan los análisis de renglón y de columna: en ambas gráficas los puntos tienen las mismas

posiciones relativas para cada eje principal, la diferencia radica en que en la primera, las coordenadas de los perfiles de renglón están en una escala más reducida que las coordenadas de los vértices en la gráfica del análisis de columna, y las posiciones de los vértices en el análisis de renglón son una expansión de las posiciones de los perfiles de columna en el análisis de columna.

Factor de Escalamiento

Con los puntos anteriormente descritos referentes a la conexión entre el análisis de renglón y de columna se puede explicar por qué al realizar cualquiera de ellos se realiza simultáneamente el otro tipo de análisis. Como los puntos de perfil y los vértices tienen las mismas posiciones relativas en ambos análisis, se deduce que existe un factor que modifica las escalas de los puntos. Este factor de escala que permite convertir las coordenadas de los perfiles a coordenadas de los vértices y las de los vértices a perfiles sobre el k -ésimo eje principal es la raíz cuadrada de la k -ésima inercia principal, es decir, son los valores singulares denotados por $\alpha_k = \sqrt{\lambda_k}$. En el caso de datos categóricos la máxima correlación entre variables de renglón y de columna es la correlación canónica que surge por la asignación de un valor numérico a las categorías de las variables en relación a los ejes principales; si estos valores son los valores singulares entonces la correlación no se afectará si los puntos cambian de escala a través de ellos. Por lo tanto, si las inercias principales de los mapas de la figura (3.11) son $\lambda_1 = 0.487$ y $\lambda_2 = 0.062$ para el primero y segundo ejes principales respectivamente, entonces los factores correspondientes son $\alpha_1 = 0.698$ y $\alpha_2 = 0.250$. Para pasar de los perfiles de renglón del panel I a los vértices de renglón del panel II se dividen las coordenadas del primer eje entre α_1 y las del segundo eje entre α_2 . Para obtener las coordenadas de los perfiles de columna a partir de las de los vértices por columna, éstas se multiplican por α_1 para el primer eje y por α_2 para el segundo eje.

En general, las coordenadas de los perfiles con respecto a los ejes principales se llaman **coordenadas principales** y las coordenadas de los vértices se llaman **coordenadas estándar**, por lo tanto, para obtener las coordenadas de un análisis a partir del otro, la transformación que se requiere hacer es:

$$k\text{-ésimas coordenadas principales} = (k\text{-ésimas coordenadas estándar}) * \alpha_k$$

De lo anterior, es posible utilizar los términos de k -ésimas coordenadas principales de renglón, k -ésimas coordenadas principales de columna, k -ésimas coordenadas estándar de renglón y k -ésimas coordenadas estándar de columna.

Así como la inercia total mide el grado de dispersión de un conjunto de perfiles respecto a los vértices, las inercias principales λ_k y los factores de escala α_k son también medidas de la dispersión de los perfiles comparados con los vértices sobre cada eje principal. El mayor valor que se puede obtener para α_k es 1, lo que indica que los perfiles siempre se encontrarán en una posición al "interior" a sus vértices correspondientes.

ÁLGEBRA Y GEOMETRÍA EN EL ANÁLISIS DE CORRESPONDENCIA

Las definiciones anteriores permiten explicar el Análisis de Correspondencia a nivel conceptual. Los cálculos que son necesarios para obtener valores numéricos se fundamentan en los aspectos descritos en el capítulo anterior referentes al Álgebra Lineal donde, específicamente, se recurre al cálculo de valores y vectores propios de una matriz para obtener la información relativa a la inercia total y sus distintas formas de descomposición, los valores que determinan las coordenadas de los vértices y de los perfiles, y las contribuciones absolutas y relativas, sin olvidar, por supuesto, el punto de vista geométrico que conduce a la obtención de los mapas.

Los procedimientos descritos para el análisis multivariado se aplican en el Análisis de Correspondencia, considerando algunos puntos importantes que deben tomarse en cuenta y que son específicos de este método estadístico. Primero, las variables involucradas en el Análisis de Correspondencia tratan con datos cualitativos que no tienen todas las características de los datos métricos y su manejo presenta algunas restricciones en cuanto a las operaciones que se les puede aplicar y que son válidas en términos estadísticos. Por ejemplo, cuando se calcula la correlación entre variables cuantitativas (a partir de las covarianzas), las variables involucradas son referentes a un mismo conjunto de datos métricos. En el Análisis de Correspondencia se tienen datos categóricos donde existe una correlación canónica dada por un factor de escalamiento que indica la máxima correlación encontrada entre las categorías de las variables de renglón y columna, por lo que las matrices que son descompuestas en valores singulares difieren de las matrices de covarianza a las que se hizo referencia en el análisis multivariado.

En cuanto a la aplicación del problema de mínimos cuadrados para hallar los ejes que mejor se ajustan a un conjunto de puntos, en el Análisis de Correspondencia se aplica el método de la suma ponderada de distancias cuadradas, refiriéndose a distancias χ^2 que permiten manejar los datos para los propósitos deseados.

Bajo estas circunstancias, sin embargo, los principios siguen siendo los mismos y el lugar donde se desarrollan los aspectos geométricos es el espacio Euclidiano, donde se conservan las características de la dimensión del espacio, la independencia lineal, los cambios de la base, las transformaciones ortogonales y sus efectos geométricos.

El Análisis de Correspondencia se realiza, naturalmente, a partir de una tabla de contingencia que contiene los datos originales en términos de frecuencias. En la mayoría de los casos el número de renglones y columnas de una tabla puede llegar a ser más grande que los ejemplos vistos y los puntos de un espacio multidimensional deben representarse en algún subespacio de menor dimensión que permita visualizar e interpretar las posiciones de estos puntos. Los cálculos pueden ser desarrollados por paquetes de computación que facilitan el manejo de cantidades grandes de datos. Con el fin de ejemplificar los pasos del método, se trabajará con una matriz de datos pequeña cuyo manejo es relativamente sencillo y donde los cálculos no resultan tan complejos.

Matriz de Dispersión y Matriz de Correspondencia

Inicialmente, se tiene una matriz de r renglones y c columnas con las frecuencias originales del problema y con un total general representado por n . Retomando la tabla de contingencia (1.4) del Capítulo I, donde las variables son categóricas, y realizando un análisis por renglón, las matrices de datos originales y de perfiles de renglón son:

INTERESES	GRUPOS			Total
	I	II	III	
Religión	10	14	6	30
Ciencia	36	92	40	168
Música	38	58	78	174
Deportes	24	80	98	202
Política	6	14	32	52
Total	114	258	254	626

Tabla (3.4) Tabla de contingencia con datos originales

INTERESES	GRUPOS		
	I	II	III
Religión	0.333	0.467	0.200
Ciencia	0.214	0.548	0.238
Música	0.218	0.333	0.448
Deportes	0.119	0.396	0.485
Política	0.115	0.269	0.615
MASA	0.182	0.412	0.406

Tabla (3.5) Perfiles de renglón

El número de vértices que forman el espacio sobre el cual se posicionan los puntos de perfil es 3, lo que indica que los puntos caerán en un subespacio de 2 dimensiones. Para graficar los vértices, sus coordenadas ya ponderadas son (2.34 0 0), (0 1.56 0) y (0 0 1.57), estos puntos determinan el triángulo sobre el que caen los puntos de perfil como se muestra en la figura (3.12).

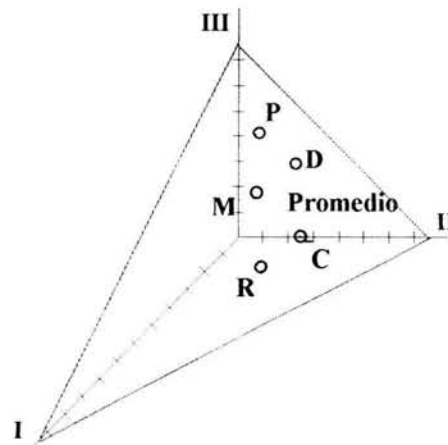


Fig. (3.12) Perfiles de renglón con distancias χ^2

“Extrayendo” el triángulo del espacio tridimensional, deben encontrarse los ejes que determinan el espacio bidimensional que mejor se ajusta a los puntos desplegados en el plano. Los ejes tendrán su origen en el punto de perfil promedio y los factores que determinarán su orientación serán las posiciones de los puntos y sus masas respectivas. Así, los ejes encontrados se muestran en la figura (3.13), primero sobre el espacio original y después considerando ya la reducción de la dimensión

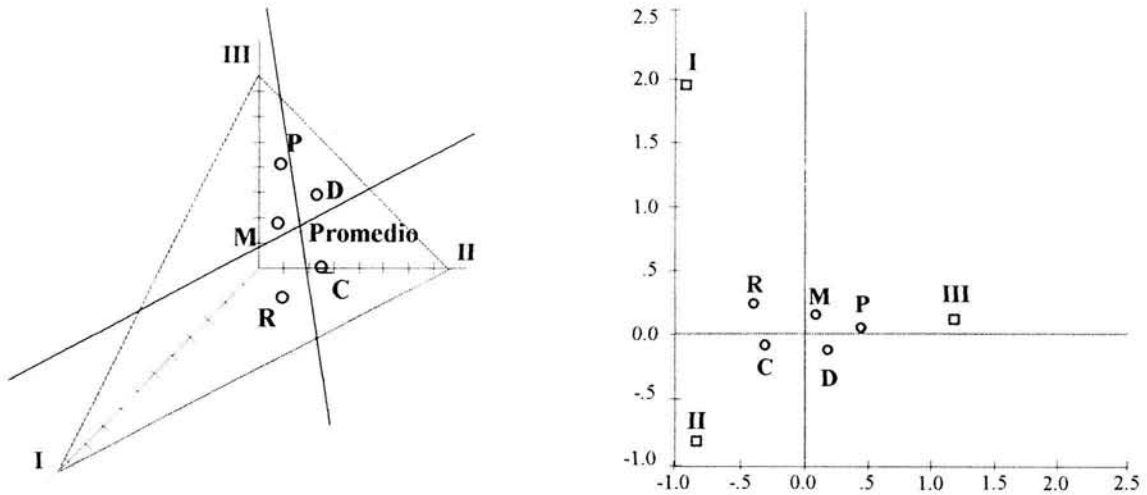


Fig. (3.13) Ejes del espacio óptimo

El procedimiento para obtener el subespacio es similar al problema de los mínimos cuadrados, ya que se desea encontrar los ejes para las cuales la suma de los cuadrados de las distancias de los puntos hacia éstos sea mínima.

Cuando se obtuvo la eigenestructura de la matriz de covarianza $C(X)$ lo que se hizo fue maximizar el producto de matrices $T'C(X)T=C(Z)$, donde T , la matriz de transformación, está conformada por los vectores ortogonales t que maximizan la varianza para cada dimensión (ejes principales) y que son los eigenvectores correspondientes a los eigenvalores de $C(X)$.

En el Análisis de Correspondencia debe considerarse que las distancias no son Euclidianas sino distancias χ^2 desplegadas en el espacio Euclidiano donde cada perfil tiene una masa asociada que lo pondera y lo sitúa en un lugar determinado. Por esto, el criterio utilizado es el de la suma ponderada del cuadrado de las distancias χ^2 , con el objetivo de encontrar el subespacio sobre el cual los puntos y los vértices se proyecten adecuadamente.

Si M representa al subespacio que mejor se ajusta al conjunto de perfiles y si la distancia χ^2 del i -ésimo perfil al subespacio M se denota como $d_i(M)$ entonces la "cercanía" de todos los perfiles a M es:

$$\text{Cercanía a } M = \sum r_i [d_i(M)]^2$$

donde r_i es la masa del i -ésimo perfil. Esto indica que el subespacio óptimo será aquel que minimice este criterio. Al minimizar la suma ponderada de las distancias cuadradas se busca explicar en la mayor medida posible la dispersión existente entre los puntos de perfil. En el Análisis de Correspondencia es preciso contar con datos que muestren información sobre la dispersión de los puntos; para ello se requiere analizar la inercia total, ya que ésta es una medida de las variaciones existentes. Si se desea minimizar:

$$\sum r_i [d_i(M)]^2$$

entonces M debe maximizar la variación dada por la inercia total:

$$\frac{\chi^2}{n} = \sum r_i d_i^2$$

Es necesario, por lo tanto, obtener una matriz que contenga datos sobre la variación de los puntos y que sea análoga a la matriz de desviaciones X_d/\sqrt{m} cuyo momento de producto mínimo es $C(X)$. Esta matriz es llamada **matriz de dispersión S** y contiene las raíces cuadradas de las contribuciones de cada celda de la tabla de contingencia a la inercia total. Cada una de estas celdas contribuye a la inercia total a través de:

$$r_i \frac{(a_{ij} - c_j)^2}{c_j}$$

Así, cada celda de S contiene los valores:

$$s_{ij} = \sqrt{r_i \frac{(a_{ij} - c_j)^2}{c_j}} = \frac{\sqrt{r_i}}{\sqrt{c_j}} (a_{ij} - c_j) = \frac{\sqrt{r_i}}{\sqrt{r_i}} \frac{r_i(a_{ij} - c_j)}{\sqrt{r_i c_j}} = \frac{(r_i a_{ij} - r_i c_j)}{\sqrt{r_i c_j}}$$

donde los términos $r_i a_{ij} = \left(\frac{n_{i+}}{n}\right) \left(\frac{n_{ij}}{n_{i+}}\right) = \frac{n_{ij}}{n}$ son las $r \times c$ frecuencias de cada celda de la tabla

de contingencia divididas entre el total general n . La matriz que contiene estos valores es la

matriz de correspondencia P donde el valor de cada celda es $p_{ij} = \frac{n_{ij}}{n}$. Por ejemplo, la matriz

P relacionada a la tabla (3.4) es:

				r_i
	.016	.022	.010	.048
	.058	.147	.064	.269
	.061	.093	.125	.279
	.038	.128	.155	.321
	.010	.022	.051	.083
c_j	.183	.412	.405	

Tabla (3.6) Matriz de correspondencia

A partir de P , se pueden calcular las masas de renglón r_i y las de columna c_j realizando las sumas por renglón o columna:

$$\begin{aligned} r_i &= \sum_j p_{ij} & i=1, \dots, r \\ c_j &= \sum_i p_{ij} & j=1, \dots, c \end{aligned}$$

Las masas r_i y c_j del ejemplo se muestran en la tabla (3.6). De esta manera, considerando la matriz P , los valores de S son:

$$s_{ij} = \frac{(p_{ij} - r_i c_j)}{\sqrt{r_i c_j}}$$

La matriz de dispersión S del ejemplo es:

$$S = \begin{bmatrix} .077 & .016 & -.068 \\ .040 & .109 & -.136 \\ .044 & -.065 & .036 \\ -.086 & -.012 & .069 \\ -.042 & -.066 & .095 \end{bmatrix}$$

Tabla (3.7) Matriz de dispersión

Por lo tanto, con los valores de S se puede calcular la inercia total:

$$\frac{\chi^2}{n} = \sum_i \sum_j s_{ij}^2 = 0.078$$

La matriz S representa la base para hallar el subespacio óptimo que contiene a los puntos de perfil y que explican las variaciones entre ellos. Esta matriz contiene los valores que conforman a la inercia total, por lo tanto, si S es descompuesta en matrices más sencillas, éstas deben explicar el mismo comportamiento de las dispersiones. Esto se obtiene a partir de la DVS de la matriz S , donde además de que las matrices obtenidas permiten conocer las transformaciones necesarias para explicar la dispersión de los puntos, éstas representan la naturaleza geométrica de los subespacios óptimos sobre los que se despliegan los datos.

Descomposición en Valores Singulares

Siguiendo el procedimiento de la DVS para una matriz de $m \times n$, se tiene que, para $S_{5 \times 3}$, la DVS obtenida es:

$$S = P \Delta Q' = \begin{bmatrix} -.356 & .446 \\ -.690 & -.306 \\ .171 & .651 \\ .365 & -.530 \\ .484 & .063 \end{bmatrix} \begin{bmatrix} .256 & 0 \\ 0 & .113 \end{bmatrix} \begin{bmatrix} -.391 & -.505 & .769 \\ .816 & -.577 & .035 \end{bmatrix}$$

Tabla (3.8) Descomposición en Valores Singulares de S

En este procedimiento se calculan las eigenestructuras de $S'S$ y de SS' , las cuales son análogas a la matriz $C(X)$ y representan la dispersión original sobre los ejes de cada espacio considerado.

Denotando a P como U y a Q como V, entonces U son los eigenvectores principales de SS' y V son los eigenvectores principales de $S'S$. La matriz Δ contiene dos valores singulares $\alpha_k = \sqrt{\lambda_k}$ por lo que se han obtenido dos valores propios diferentes mayores a cero ($\lambda_1=0.065$ y $\lambda_2=0.013$) y el subespacio óptimo es bidimensional. Los valores λ_i representan los valores de las inercias principales, es decir, la inercia total está determinada por la suma de los valores propios de $S'S$:

$$\frac{\chi^2}{n} = \sum_k \lambda_k = 0.065 + 0.013 = 0.078$$

Los valores singulares están ordenados descendientemente, esto significa que $\lambda_1=0.065$ explica la mayor cantidad de la inercia total y, por lo tanto, se considera como la primera inercia principal correspondiente al primer eje principal; $\lambda_2=0.013$ explica la mayor cantidad de la inercia restante y es la segunda inercia principal y así sucesivamente. De aquí que un valor de $\lambda_i=0$ no represente aportación alguna a la inercia principal y su eje respectivo no sea considerado para determinar la dimensión.

En este punto se observa la vinculación entre un análisis de renglón y uno de columna ya que, suponiendo que en lugar de ser 3 las columnas de la matriz S fueran 5 (es

decir, transponiendo la matriz S o, equivalentemente, realizando un análisis por columna), el número de valores propios (y por lo tanto de valores singulares) mayores a cero también es dos y, más aún, los valores para λ_1 y λ_2 son los mismos en ambos casos, representando los mismos valores para las inercias principales, obteniendo exactamente la misma dimensión en ambos espacios.

Las matrices U y V son rotaciones de las bases de los espacios originales de dimensión 4 y 2 respectivamente (la DVS no es completa), es decir, si se consideraran los renglones como vértices del espacio, entonces al calcular la eigenestructura de SS' se obtienen únicamente dos eigenvalores que son suficientes para representar la inercia de los puntos; los eigenvectores correspondientes a estos valores son los que se muestran en U . De la misma forma, si los vértices son las columnas entonces los eigenvalores de $S'S$ son dos (los mismos que para SS') y los eigenvectores correspondientes están determinados por V . Estas rotaciones representan las posiciones de los ejes principales con los cuales se minimizan las distancias entre éstos y los puntos y maximizan la variación dada por el valor de la inercia considerando ya la reducción de la dimensión, por lo que se refieren a los ejes del espacio o subespacio óptimo.

Los valores de la matriz U corresponden a las nuevas coordenadas que cada perfil de renglón tiene con respecto a los ejes principales. La matriz V contiene los valores de las coordenadas de los perfiles de columna al realizar la rotación de los ejes óptimos. Cada columna de las matrices se refiere a cada una de las k dimensiones consideradas y representadas por el k -ésimo eje principal.

Los vectores de U y V no están ponderados por las masas respectivas, por lo que, para determinar los vértices en términos de distancias χ^2 y obtener las matrices de coordenadas estándar Φ y Γ para los vértices de renglón y columna se calculan, respectivamente:

$$\phi_{ik} = \frac{u_{ik}}{\sqrt{r_i}} \quad \text{Coordenadas estándar de renglón}$$

$$\gamma_{jk} = \frac{v_{jk}}{c_j} \quad \text{Coordenadas estándar de columna}$$

En este caso, las matrices son:

$$\Phi = \begin{bmatrix} -1.68 & 2.00 \\ -1.32 & -.629 \\ .311 & 1.26 \\ .653 & -.894 \\ 1.667 & .129 \end{bmatrix} \quad \Gamma = \begin{bmatrix} -.924 & 1.907 \\ -.782 & -.903 \\ 1.209 & .061 \end{bmatrix}$$

Para obtener las coordenadas principales, se aplica el factor de escala α_k a las matrices Φ y Γ para obtener las matrices F de coordenadas principales de renglón y G de coordenadas principales de columna:

$$f_{ik} = \alpha_k \phi_{ik} \quad \text{Coordenadas principales de renglón}$$

$$g_{jk} = \alpha_k \gamma_{jk} \quad \text{Coordenadas principales de columna}$$

Por lo tanto:

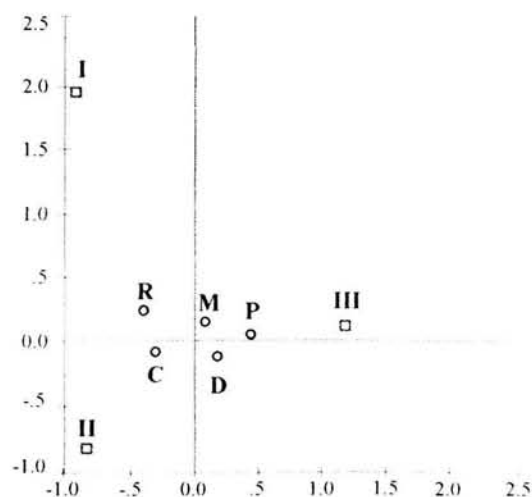
$$F = \begin{bmatrix} -.431 & .227 \\ -.338 & -.071 \\ .079 & .143 \\ .167 & -.101 \\ .427 & .015 \end{bmatrix} \quad G = \begin{bmatrix} -.236 & .216 \\ -.200 & -.102 \\ .309 & .007 \end{bmatrix}$$

La matriz F son los perfiles de renglón que se despliegan en el mismo espacio de Γ . La matriz G representa los perfiles de columna posicionados en el espacio donde también se muestran las coordenadas Φ .

Las primeras k columnas para cada matriz dan las coordenadas de los perfiles y de los vértices con respecto a los primeros k ejes principales. Con estas cuatro matrices se pueden obtener los distintos mapas referentes a cada tipo de análisis para el ejemplo:

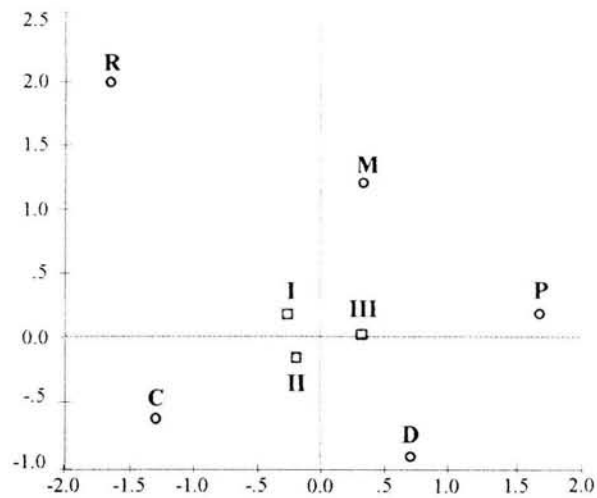
Análisis por renglón:

Con las matrices Γ y F



Análisis por columna:

Con las matrices Φ y G



Inercias y Contribuciones:

- En cuanto a la inercia total para el ejemplo anterior se obtiene su valor sumando los valores propios de $S'S$, sumando los valores de la diagonal de la misma matriz, o bien, sumando los cuadrados de los elementos de S .

$$\frac{\chi^2}{n} = \sum_k \lambda_k = \text{tr}(S'S) = \sum_i \sum_j s_{ij}^2 = 0.078$$

- La contribución de cada celda es simplemente el cuadrado del valor de la celda de S correspondiente (s_{ij}^2).

- Las inercias de renglón y columna se obtienen sumando los cuadrados de los valores de S para cada renglón y/o columna, es decir:

$$\sum_j s_{ij}^2 \text{ para el renglón } i$$

$$\sum_i s_{ij}^2 \text{ para la columna } j$$

por ejemplo: inercia del renglón 2 = 0.032 (41% de la inercia total)

inercia de la columna 1 = 0.019 (24% de la inercia total)

- Las inercias principales son los valores de $\lambda_1, \lambda_2, \dots$, etc. Para el ejemplo:

$$\lambda_1 = 0.065, \lambda_2 = 0.013, \lambda_3 = 0$$

λ_1 contribuye con el 83% de la inercia total

λ_2 contribuye con el 17% de la inercia total

- La contribución de los perfiles de renglón o columna al k-ésimo eje principal (contribución absoluta) es, en porcentajes:

$$\frac{r_{ik}^2}{\lambda_k} \quad \text{para renglones}$$

$$\frac{c_{jk}^2}{\lambda_k} \quad \text{para columnas}$$

por ejemplo: la contribución del perfil de renglón 1 al eje 2 es: 0.19

la contribución del perfil de columna 3 al eje 1 es: 0.59

- La contribución del k-ésimo eje principal al i-ésimo renglón o la j-ésima columna (contribuciones relativas) es:

$$\frac{r_{ik}^2}{\sum_j s_{ij}^2} \quad \text{para renglones}$$

$$\frac{c_{jk}^2}{\sum_i s_{ij}^2} \quad \text{para columnas}$$

La contribución relativa del eje 1 al perfil de renglón 4 es: 0.73

La contribución relativa del eje 2 al perfil de columna 2 es: 0.20

- La calidad del mapa obtenido en una solución de k dimensiones es la suma de las k primeras contribuciones relativas, es decir:

$$\sum_k \frac{r_i f_{ik}^2}{\sum_j s_{ij}^2} \quad \text{para el punto del } i\text{-ésimo renglón}$$

$$\sum_k \frac{c_j g_{jk}^2}{\sum_i s_{ij}^2} \quad \text{para el punto de la } j\text{-ésima columna}$$

si k es igual a la dimensión total del subespacio, entonces la suma es igual a 1.

En este caso todos los puntos están representados al 100%, por lo que la calidad del mapa es 1.

En el capítulo siguiente se ejemplificará y explicará más detalladamente la obtención y significado de estos valores.

MAPAS SIMÉTRICOS Y ASIMÉTRICOS

Los mapas que se han mostrado hasta el momento reflejan las posiciones de los puntos de perfil; el despliegue de los vértices sirve como referencia para la interpretación de los perfiles. Los mapas que muestran el conjunto de los perfiles (coordenadas principales) y el conjunto de los vértices (coordenadas estándar) son mapas asimétricos, o bien, mapas con escalas asimétricas, por lo que, para el mismo conjunto de datos existen mapas asimétricos relacionados dependiendo si se trata de un análisis de renglón o un análisis de columna, como se ve en los mapas de la figura (3.11).

Un mapa simétrico, o mapa con escalas simétricas, es aquel en el que los conjuntos de los perfiles de renglón y los perfiles de columna se despliegan simultáneamente, es decir, ambos están en coordenadas principales. Este es el tipo más común en el Análisis de Correspondencia, donde los mapas de los perfiles de renglón y de columna que se generan por separado y que tienen la misma dimensión se sobreponen para obtener un solo mapa que represente a ambos. Al contrario de los mapas asimétricos donde los perfiles se agrupan al centro del mapa, en los simétricos los dos conjuntos de puntos están igualmente dispersos en la gráfica, independientemente del grado de asociación entre los renglones y columnas

determinado por el factor de escala. El mapa simétrico de la tabla (3.4) se muestra en la figura (3.14).

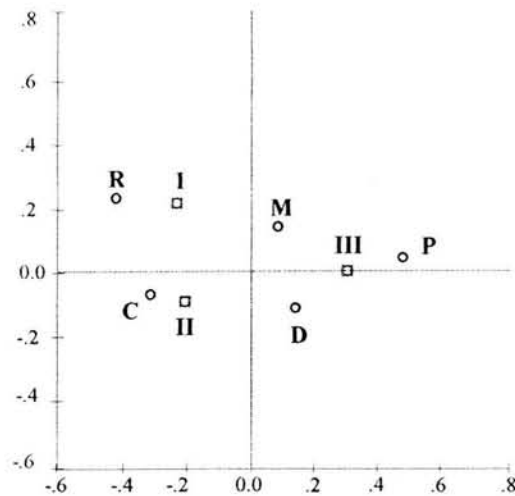


Fig. (3.14) Mapa simétrico, desplegando los perfiles de renglón y los perfiles de columna

En la interpretación de un mapa debe considerarse si éste es asimétrico o simétrico. En los asimétricos es posible aproximar distancias entre los puntos de perfil desplegados e incluso entre los puntos de perfil y los vértices, donde las distancias están relacionadas inversamente a los elementos del perfil para cada vértice, sin embargo, la distancia entre vértices no tiene una interpretación relevante.

Por otro lado, en un mapa simétrico sólo es posible aproximar las distancias entre perfiles de renglón, o bien, entre perfiles de columna, no así las distancias entre los perfiles de renglón y de columna. Esto se debe a que los mapas de los que proviene cada conjunto son diferentes y se generan por separado, lo que no sucede en los mapas asimétricos donde todos los puntos (perfiles y vértices) provienen del mismo espacio. La interpretación de una distancia entre un perfil de renglón y uno de columna en un mapa simétrico no está necesariamente relacionado con el grado de asociación entre ellos y puede conducir a conclusiones erróneas, por lo que debe tenerse en cuenta el tipo de mapa al que se está refiriendo.

De esta manera, pueden obtenerse los análisis de renglón y columna que generan mapas asimétricos, o bien, generar el mapa simétrico de coordenadas principales.

Para el mapa asimétrico de un análisis de renglón, se hace referencia a la matriz Γ (vértices de columna) y a la matriz F (perfiles de renglón).

Para el mapa asimétrico de un análisis de columna, son necesarias las matrices Φ (vértices de renglón) y la matriz G (perfiles de columna).

Para un mapa simétrico, se consideran las matrices F y G (aunque las coordenadas no están en el mismo espacio), tal como se muestra en la figura (3.14).

CAPÍTULO IV

EJEMPLO DE APLICACIÓN

El resultado gráfico que se deriva de los procedimientos que se han mencionado en los capítulos anteriores es el aspecto al que se le da mayor importancia en el Análisis de Correspondencia. Las gráficas son una representación que sintetiza la información sobre la afinidad o interacción entre las variables de renglón y columna como resultado del análisis de tablas de contingencia. Sin embargo, son igualmente importantes los valores numéricos sobre los que se basan dichas representaciones porque son los que determinan la validez y confiabilidad de la interpretación de éstas.

En la práctica, la mayoría de las veces los resultados se despliegan en espacios bidimensionales formados por el primer y segundo ejes principales ya que en ellos resulta sencillo observar las categorías que son similares a otras. En gran parte de las aplicaciones estos espacios son el resultado de reducir la dimensión del espacio original generado por una tabla de contingencia con un gran número de renglones y columnas.

Con la finalidad de integrar los principales conceptos expuestos en los capítulos anteriores, en el siguiente ejemplo se plantea una situación en donde se tienen datos obtenidos a partir de una encuesta aplicada a una muestra de 1890 personas que pertenecen a una población total definida por personas que cuentan con servicios contratados en alguna de las instituciones bancarias que se mencionan.

Entre los aspectos analizados en dicha encuesta, se interrogó a las personas sobre cuál es el banco principal con el que tienen servicios contratados y cuál es la fuente principal de sus ingresos personales. Con estas variables, "Banco principal" y "Fuente principal de ingresos", se construye una tabla de contingencia que desagrega el total de los casos considerando las distintas opciones posibles y mostrando las frecuencias observadas para cada categoría (Tabla (4.1)); los renglones de la tabla representan los distintos bancos considerados, las columnas representan la distribución de las principales fuentes de ingresos de los clientes.

BANCOS	FUENTES DE INGRESO							Margen activo
	Jornal	Honorarios Profesionales	Ganancia Negocios	Rentas	Sueldo Quincenal/ Mensual	Pensiones	Otras	
BANCRECER	6	38	78	28	20	26	10	206
SCOTIBANK	2	4	26	2	24	8	2	68
BANORTE	12	50	98	42	58	34	10	304
BANAMEX	6	30	82	70	52	16	8	264
BBVA	20	44	94	18	52	18	22	268
BANCOMER								
SANTANDER	6	22	50	30	68	24	10	210
IXE	2	12	28	10	22	10	4	88
HSBC	0	24	68	34	46	22	18	212
CITIBANK	4	10	22	8	14	4	24	86
OTROS	4	22	74	16	40	4	24	184
Margen activo	62	256	620	258	396	166	132	1890

Tabla (4.1) Tabla de Contingencia

Ambas variables son categóricas; sus respectivas categorías mostradas en la tabla son exhaustivas y mutuamente exclusivas, sin indicar un orden sino únicamente clasificando a las personas de acuerdo con las características analizadas, por lo que las variables son medidas en una escala nominal.

Los totales marginales muestran que existe un mayor número de clientes de BANORTE, seguido de BBVA BANCOMER y BANAMEX con cantidades muy similares, y posteriormente HSBC, SANTANDER y BANCRECER principalmente. En cuanto a los totales de columna destaca el ingreso por “ganancia de negocios” marcando una diferencia considerable con el resto de las categorías, después de ésta sobresalen los clientes con ingreso mediante “sueldo quincenal o mensual” y le siguen “rentas” y “honorarios profesionales” con cantidades similares.

El Análisis de Correspondencia genera distintas representaciones gráficas que ilustran las relaciones entre las variables y sus categorías dependiendo del enfoque del análisis. En este ejemplo, considérese que el interés principal es observar cómo difieren los bancos entre sí, esto significa que debe realizarse un análisis por renglón.

Mediante el Análisis de Correspondencia y, en este caso, en un análisis por renglón se pretende hallar una configuración en donde las distancias Euclidianas entre los puntos de perfil en el espacio multidimensional original sean igual a las distancias χ^2 entre los

renglones de la tabla de contingencia, mientras que en un espacio reducido las distancias Euclidianas son sólo una aproximación de las distancias χ^2 .

El espacio óptimo original de la tabla (4.1) consta de 6 dimensiones; reduciendo la dimensión para visualizar los resultados. el mapa asimétrico correspondiente a la tabla de contingencia anterior es el de la figura (4.1), obtenido mediante el paquete estadístico SPSS.

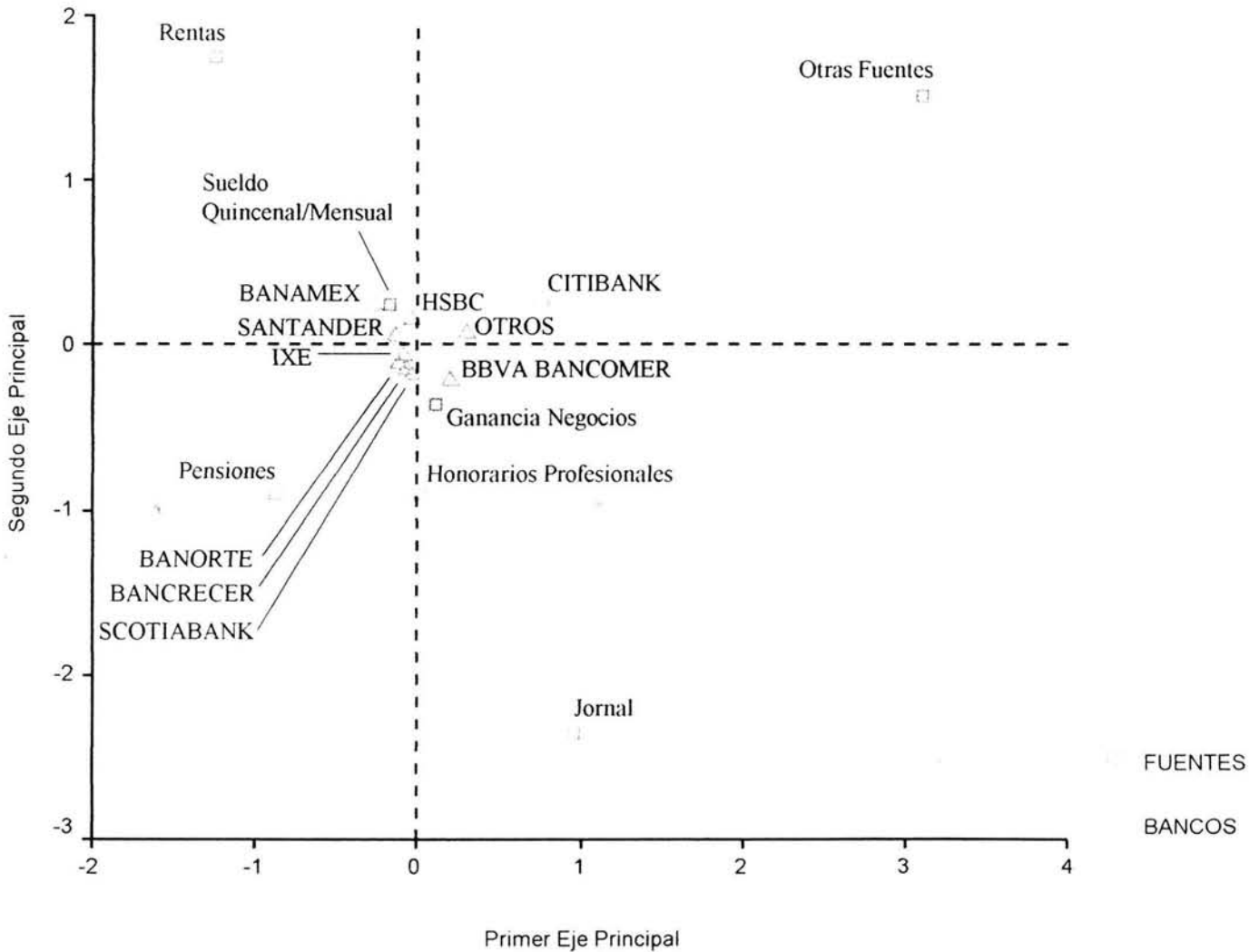


Fig. (4.1) Puntos de renglón y columna mediante análisis por renglón

El mapa asimétrico del análisis por renglón es una de las mejores representaciones por contener los puntos que pertenecen a un mismo espacio y, además, brinda mayor certeza por ser el que menos conduce a interpretaciones erróneas.

Generalmente sobre el primer eje principal se observan las diferencias más grandes entre los puntos (hay mayor dispersión) y para explicar las similitudes o diferencias entre ellos también es necesario considerar las posiciones de los puntos que representan a los vértices, los cuáles sirven como referencia para su interpretación.

De este modo, en el mapa se muestra la asociación existente entre los bancos así como la relación entre éstos y las fuentes principales de ingresos. Las principales variaciones de los perfiles se dan entre las categorías extremas relativas al primer eje principal, es decir, las correspondientes a los bancos BANAMEX en el extremo izquierdo y a CITIBANK en el extremo derecho; los demás bancos se encuentran entre estas dos categorías. En el caso de los vértices, la mayor variación puede notarse entre la categoría de "rentas" que se ubica en el extremo izquierdo y la categoría "Otras fuentes" que se encuentra en el extremo opuesto. En cuanto al segundo eje principal, los perfiles no se encuentran tan dispersos como en el primero, sin embargo también existen diferencias entre los puntos; las categorías de los perfiles que se oponen respecto a este eje son CITIBANK en la parte superior y BBVA BANCOMER en la parte inferior; los vértices extremos son "rentas" hacia la parte superior del mapa y "jornal" en la parte inferior.

La intersección de los ejes principales del mapa representa el perfil promedio de renglón y con respecto a él se consideran las desviaciones de los puntos. La distancia Euclidiana entre un punto de perfil y el origen de la gráfica aproxima la distancia χ^2 entre la categoría del renglón correspondiente y el perfil promedio de renglón, es decir, la distancia Euclidiana ponderada por la masa respectiva; de la misma manera, las distancias Euclidianas entre cualquier par de puntos de perfil aproximan las distancias χ^2 entre las categorías de los renglones mostradas en la tabla de contingencia.

La interpretación del mapa es sencilla, los puntos de perfil que son similares tienen posiciones similares. En este ejemplo, de forma muy general, los perfiles de los bancos que se encuentran en el primer cuadrante del mapa están más asociados con "Otras fuentes de ingresos", entre más se encuentren hacia la parte superior y al extremo izquierdo la principal fuente de ingresos son las "rentas" y entre más se muevan hacia la parte inferior se tendrá predominio hacia ingresos por "jornal".

Los bancos que tienen un comportamiento similar según sus posiciones en el mapa son BANORTE, BANCRECER, SCOTIABANK e IXE, esto significa que las proporciones de sus clientes con respecto a las distintas fuentes de ingresos son muy parecidas y por lo tanto se localizan muy cerca unas de otras. Por otro lado, CITIBANK se encuentra alejado de BANAMEX por lo que las diferencias en su comportamiento son más notorias. Las posiciones de los vértices de columna permiten identificar las tendencias de los puntos de perfil. Para conocer la afinidad entre los puntos de perfil con respecto a los vértices supóngase que se traza una línea desde uno de los vértices hasta el origen del mapa; si los puntos de perfil (bancos) se proyectan perpendicularmente sobre esta línea o sobre su prolongación entonces la distancia que exista desde las intersecciones hasta el vértice indicará la asociación de los perfiles con respecto al vértice determinado; entre menor sea la distancia el punto de perfil estará más asociado con la categoría que representa el vértice. Así, por ejemplo, si se traza una línea a partir del vértice “rentas” hasta el origen del mapa y los puntos de perfil se proyectan ortogonalmente sobre esta línea entonces se observa que BANAMEX se encuentra más cerca del vértice indicando que en cuanto a esta fuente de ingresos específica este banco es el que tiene una mayor asociación que el resto, es decir, BANAMEX está más relacionado con clientes que perciben rentas como fuente principal de ingresos, en segundo lugar se encuentra HSBC y el más alejado es BBVA BANCOMER por lo que una de las características de los clientes de este banco es que son los que presentan la menor asociación con este tipo de ingresos; este último banco, por otro lado, es el más cercano a “honorarios profesionales”, seguido por SCOTIABANK, y el banco que se encuentra más alejado de este vértice es CITIBANK, siendo el que menor relación tiene con clientes que perciben sus ingresos de la forma mencionada.

Al interpretar los resultados del análisis deben considerarse, además, las limitaciones existentes en las representaciones gráficas y examinar los resultados numéricos que muestran la información referente a los puntos de perfil, la inercia y sus contribuciones, obtenidos también mediante el paquete *SPSS*.

Perfiles de Renglón y Columna

En las tablas (4.2a) y (4.2b) se muestran los perfiles de renglón y de columna respectivamente, así como las masas correspondientes. El mapa de la figura (4.1) es, entonces, el plano óptimo sobre el que se proyectan los perfiles de renglón así como los

vértices representados por las columnas de la tabla de contingencia, los cuales, como se mencionó, forman un espacio de seis dimensiones originalmente.

BANCOS	FUENTES DE INGRESO							Margen activo
	Jornal	Honorarios Profesionales	Ganancia Negocios	Rentas	Sueldo Quincenal/Mensual	Pensiones	Otras	
BANCRECER	.029	.184	.379	.136	.097	.126	.049	1.000
SCOTIBANK	.029	.059	.382	.029	.353	.118	.029	1.000
BANORTE	.039	.164	.322	.138	.191	.112	.033	1.000
BANAMEX	.023	.114	.311	.265	.197	.061	.030	1.000
BBVA	.075	.164	.351	.067	.194	.067	.082	1.000
BANCOMER								
SANTANDER	.029	.105	.238	.143	.324	.114	.048	1.000
IXE	.023	.136	.318	.114	.250	.114	.045	1.000
HSBC	.000	.113	.321	.160	.217	.104	.085	1.000
CITIBANK	.047	.116	.256	.093	.163	.047	.279	1.000
OTROS	.022	.120	.402	.087	.217	.022	.130	1.000
MASA	.033	.135	.328	.137	.210	.088	.070	

Tabla (4.2a) Perfiles de renglón

BANCOS	FUENTES DE INGRESO							MASA
	Jornal	Honorarios Profesionales	Ganancia Negocios	Rentas	Sueldo Quincenal/Mensual	Pensiones	Otras	
BANCRECER	.097	.148	.126	.109	.051	.157	.076	.109
SCOTIBANK	.032	.016	.042	.008	.061	.048	.015	.036
BANORTE	.194	.195	.158	.163	.146	.205	.076	.161
BANAMEX	.097	.117	.132	.271	.131	.096	.061	.140
BBVA	.323	.172	.152	.070	.131	.108	.167	.142
BANCOMER								
SANTANDER	.097	.086	.081	.116	.172	.145	.076	.111
IXE	.032	.047	.045	.039	.056	.060	.030	.047
HSBC	.000	.094	.110	.132	.116	.133	.136	.112
CITIBANK	.065	.039	.035	.031	.035	.024	.182	.046
OTROS	.065	.086	.119	.062	.101	.024	.182	.097
Margen activo	1.000	1.000	1.000	1.000	1.000	1.000	1.000	

Tabla (4.2b) Perfiles de columna

Los valores de las masas indican la forma en la que se clasifica el total de las personas encuestadas; por ejemplo, de las 1890 personas, el 3.3% está clasificado en la categoría que percibe sus ingresos por medio de "jornal", el 13.5% lo percibe mediante "honorarios profesionales", etc. De esto se observa que la principal fuente de ingresos es por "ganancia de negocios" con un 32.8%, seguida por "sueldo quincenal o mensual" con un 21% del total, mientras que las personas cuyo ingreso principal proviene de "jornal" representan la menor proporción. Así mismo, en cada renglón se tienen los porcentajes para

cada categoría, por ejemplo, de las 212 personas que tienen servicios contratados en HSBC el 32.1% tiene "ganancia de negocios", el 21.7% recibe un "sueldo quincenal o mensual", mientras que ninguno de ellos percibe "jornal".

En los renglones se aprecian algunas diferencias entre los valores de perfil y los valores correspondientes del perfil promedio; por ejemplo, el porcentaje que se esperaría encontrar en cada renglón para la categoría "sueldo quincenal o mensual" es del 21%, pero en el renglón SCOTIABANK el porcentaje es del 35.3% que representa una diferencia considerable respecto al promedio, lo que se verá reflejado en el valor de la inercia.

Por su parte, en los perfiles de columna se observa que, de los que perciben "ganancias de negocios", se tiene casi la misma proporción de clientes de BANORTE y BBVA BANCOMER y que del total de los que reciben "rentas" la mayor proporción son clientes de BANAMEX, entre otros datos.

Inercia

En cuanto a la dispersión de los puntos, se tiene que la inercia total tiene un valor de .129, de la cual el primer eje principal explica la mayor proporción con un valor de .056 para la primera inercia principal, correspondiente al 43.5% del total. Los valores de las inercias principales (λ_k) se muestran en la columna "Inercia" de la tabla (4.3), donde también se indican los valores singulares α (segunda columna) que representan los ajustes que se realizan sobre cada eje principal.

DIMENSIÓN	VALOR PROPIO	INERCIA	CHI-CUADRADO	SIG.	PROPORCIÓN DE INERCIA	
					Explicada	Acumulada
1	.237	.056			.435	.435
2	.167	.028			.216	.652
3	.160	.026			.198	.850
4	.100	.010			.078	.928
5	.093	.009			.066	.994
6	.028	.001			.006	1.000
Total		.129	243.777	.000 ^a	1.000	1.000

a. 54 grados de libertad

Tabla (4.3) Resumen

El valor de la inercia total no es muy grande (el valor máximo que puede alcanzar es 6) por lo que los puntos de perfil no se encuentran muy dispersos en el mapa y tienden a

agruparse hacia la parte central de éste (perfil promedio). La descomposición de la inercia con respecto a los ejes principales permite conocer la calidad de los espacios obtenidos para una solución de k dimensiones ya que las inercias principales cuantifican la dispersión de los puntos de perfil a lo largo de los ejes principales. En este ejemplo, la proporción de inercia acumulada que se explica en el mapa bidimensional es del 65.2%, correspondiente al 43.5% explicada por el primer eje principal y al 21.6% explicada por el segundo eje principal. Estos valores se muestran en las columnas de "Proporción de inercia" de la tabla anterior indicando las contribuciones de cada dimensión. Si se considera el tercer eje principal, entonces éste aumenta la calidad del espacio, ya que se estaría explicando el 85% del total de la inercia; así, conforme aumenta la dimensión del espacio se explica en mayor medida la inercia total, aunque, en la práctica, lo ideal es que con un número pequeño de dimensiones se explique la mayor cantidad posible de ésta. Si con dos o tres dimensiones se considera que es suficiente la proporción de inercia desplegada entonces puede omitirse el estudio de las dimensiones mayores si éstas no representan un valor significativo de la inercia total.

En la tabla (4.3) también se muestra el valor del estadístico χ^2 y el nivel de significancia con 54 grados de libertad. En el análisis por renglón la suma de las distancias cuadradas entre los puntos de perfil y el perfil promedio de renglón ponderadas por los totales marginales es igual al valor del estadístico χ^2 . De manera análoga, en el análisis de columna sucede lo mismo con los perfiles de columna y el perfil promedio de columna. El valor χ^2 dividido entre el total general n es igual al valor de la inercia total $\left(\frac{243.777}{1890}\right)$, o bien, a la suma de las inercias principales.

En tablas de contingencia reales tiene sentido aplicar la prueba χ^2 en donde se asume que los n elementos del total general de la tabla de contingencia conforman una muestra aleatoria de la población y que cada uno de ellos pertenece a una sola celda de la tabla.

Para interpretar un mapa, sin embargo, la prueba χ^2 y la significancia estadística no son requisitos necesarios, ya que el mapa es sólo una forma de representación que ayuda a comprender el comportamiento de los datos de la tabla de contingencia.

La proporción de inercia que es explicada en el mapa bidimensional da una idea general de qué tan adecuada es la representación de los puntos en éste y cómo es su comportamiento. Sin embargo, esto no significa que para todos los puntos mostrados en el mapa se tenga la misma certeza al interpretarlos, sino que es necesario analizar las contribuciones de cada uno de ellos para realizar un análisis más profundo.

Las tablas (4.4a) y (4.4b) contienen otros resultados importantes del Análisis de Correspondencia que proporcionan un diagnóstico útil de los puntos de perfil y los vértices desplegados en el mapa. La segunda columna de cada tabla contiene las masas de renglón y columna respectivamente; las masas de renglón corresponden a los valores del perfil promedio de columna de la tabla (4.2b) y las masas de columna son los elementos del perfil promedio de renglón mostrados en la tabla (4.2a).

BANCOS	MASA	PUNTUACIÓN EN LA DIMENSIÓN		INERCIA	CONTRIBUCIÓN				
		1	2		De los puntos a la inercia de la dimensión		De la dimensión a la inercia del punto		Total
					1	2	1	2	
BANCRECER	.109	-.076	-.150	.012	.011	.088	.053	.205	.258
SCOTIBANK	.036	-.041	-.184	.010	.001	.043	.006	.126	.132
BANORTE	.161	-.128	-.119	.006	.047	.082	.465	.401	.867
BANAMEX	.140	-.271	.237	.022	.182	.281	.457	.351	.807
BBVA	.142	.189	-.220	.015	.090	.247	.343	.466	.809
BANCOMER									
SANTANDER	.111	-.135	.052	.012	.036	.011	.166	.024	.191
IXE	.047	-.088	-.064	.001	.006	.007	.247	.132	.379
HSBC	.112	-.031	.152	.005	.002	.093	.020	.488	.508
CITIBANK	.046	.755	.278	.032	.462	.126	.820	.111	.931
OTROS	.097	.305	.081	.014	.161	.023	.649	.045	.695
Total activo	1.000			.129	1.000	1.000			

Tabla (4.4a) Examen de los puntos de renglón

FUENTES	MASA	PUNTUACIÓN EN LA DIMENSIÓN		INERCIA	CONTRIBUCIÓN				
		1	2		De los puntos a la inercia de la dimensión		De la dimensión a la inercia del punto		Total
					1	2	1	2	
Jornal	.033	.957	-2.351	.013	.030	.181	.132	.396	.528
Honorarios Profesionales	.135	.014	-.898	.007	.000	.109	.000	.416	.416
Ganancia Negocios	.328	.110	-.361	.007	.004	.043	.033	.179	.212
Rentas	.137	-1.259	1.755	.028	.216	.420	.433	.419	.852
Sueldo Quin./ Mensual	.210	-.180	.248	.018	.007	.013	.021	.019	.040
Pensiones	.088	-.877	-.912	.012	.068	.073	.306	.164	.470
Otras	.070	3.110	1.515	.043	.675	.160	.875	.103	.979
Total activo	1.000			.129	1.000	1.000			

Tabla (4.4b) Examen de los puntos de columna

Coordenadas Principales y Coordenadas Estándar

Los valores de las columnas denominadas "Puntuación en la dimensión" son las coordenadas principales de renglón (Tabla (4.4a)) y las coordenadas estándar de columna (Tabla (4.4b)) utilizadas para graficar los puntos del mapa de la figura (4.1). Estas columnas corresponden a las matrices F y Γ respectivamente; a partir de estos valores se pueden obtener también las posiciones de los vértices de renglón (matriz Φ) y/o los perfiles de columna (matriz G) aplicando los factores de escala (α_k) correspondientes a cada dimensión y que se muestran en la segunda columna de la tabla (4.3).

Inercias de renglón y columna

En estas tablas, las columnas de "Inercia" contienen las inercias de renglón y columna respectivamente. Con estos valores se conocen las contribuciones que cada renglón y columna hacen a la inercia total; por ejemplo, en el caso de los renglones, el que más contribuye es el renglón CITIBANK con un valor de .032 y el que menos contribuye es IXE con un valor de .001. Del mismo modo, la columna "Otras fuentes" es la que mayor dispersión presenta y contribuye con un valor de .043 a la inercia total; mientras que las columnas "honorarios profesionales" y "ganancia de negocios" son las que aportan menos.

Contribuciones absolutas

Finalmente, las tablas muestran los valores de las contribuciones absolutas y relativas (columna "Contribución"). De datos anteriores se sabe que el plano de la figura (4.1) contabiliza el 65.2% de la inercia total, con las contribuciones absolutas es posible conocer los renglones y columnas que han determinado en mayor medida la orientación de los primeros dos ejes principales. Como se mencionó en el capítulo anterior, la contribución de un punto a la inercia de los ejes principales es la distancia cuadrada ponderada de la proyección del punto sobre un eje al origen (r_{ik}) dividida entre la inercia de la dimensión (λ_k); es decir, con los datos de la tabla (4.4a) se puede calcular, por ejemplo, la contribución absoluta del punto BANCRECER a la inercia del primer eje principal mediante:

$$\frac{r_{ik}^2}{\lambda_k} = \frac{r_{11}^2}{\lambda_1} = \frac{(.109)(-.076)^2}{.056} = .011$$

Examinando las columnas de las contribuciones de los puntos a la inercia de la dimensión se observa que para el primer eje principal el renglón que más influye es el de la categoría CITIBANK ya que el 46.2% de la inercia desplegada sobre la primera dimensión se debe a este punto, seguido de BANAMEX que contribuye con un 18.2% y así

sucesivamente. En cuanto a las columnas, para el primer eje, los puntos que más influyen son "Otras fuentes" aportando el 67.5% de la primera inercia principal y "rentas" con un 21.6%. Estos valores se reflejan en el mapa, al ser estos puntos los que mayor dispersión muestran en él y que más influyen en la orientación del eje. Estas contribuciones dependen de la distancia de los puntos al origen y también de sus masas respectivas.

De la misma manera, los puntos que determinan en mayor grado la posición del segundo eje principal son los de las categorías BANAMEX y BBVA BANCOMER de los renglones y las categorías "rentas" y "jornal" de las columnas.

Contribuciones relativas

Con los ejes principales determinados, las contribuciones relativas permiten saber cuáles puntos son los que están mejor explicados en el plano. En la mayoría de los casos, aquellos puntos que más determinan la orientación de los ejes son los que tienen mayor correlación con éstos y que son mejor explicados, aunque también algunos puntos que contribuyen en menor medida pueden ser bien explicados por los ejes. En las columnas de contribución de la dimensión a la inercia del punto se muestran las contribuciones relativas, por ejemplo, el punto CITIBANK que contribuye principalmente a la inercia del primer eje tiene el 82% de su inercia explicada por ese eje calculada mediante:

$$\frac{r_i f_{ik}^2}{\sum_j s_{ij}^2} = \frac{r_{91} f_{91}^2}{\sum_j s_{9j}^2} = \frac{(.046)(.755)^2}{.032} = .82$$

donde r_i es la masa del renglón i , f_{ik} es la coordenada de la proyección del punto i sobre el eje k y $\sum_j s_{ij}^2$ es la inercia del renglón i ; mientras que el punto SCOTIBANK que contribuye relativamente poco al mismo eje es explicado por éste en un 0.6%. Estas contribuciones proporcionan la correlación entre los puntos y los ejes si se calculan sus raíces cuadradas y, gráficamente, corresponden a los cosenos de los ángulos formados entre los puntos y los ejes; entre más pequeño sea el ángulo la contribución relativa tiende a 1 y la correlación es mayor; si el ángulo se aproxima a 90° la contribución relativa y la correlación con el eje respectivo tienden a cero. Así, entre más cerca esté un perfil de un eje, este último explicará en mayor medida la inercia del perfil.

En la tabla, el perfil que mejor explica el primer eje principal es CITIBANK (.82); la correlación de este punto con el eje es $\sqrt{.82} = .906$ que corresponde a un ángulo de 25.1°

entre ellos, por el contrario, el perfil SCOTIABANK tiene una contribución relativa de .006, es decir, su correlación es .077 y el ángulo que forma con el primer eje principal es de 85.5° . Lo mismo sucede para los valores de las dimensiones sucesivas; así, los puntos que mejor explican el segundo eje principal son HSBC (.488), BBVA BANCOMER (.466) y BANORTE (.401).

Calidad del mapa

Si se suman las contribuciones de las dimensiones a la inercia de cada punto (contribuciones relativas), se obtiene la calidad con la que cada punto es representado en el plano. Estos valores se muestran en la última columna de las tablas (4.4a) y (4.4b) e indican los puntos que se encuentran más alejados del plano y por lo tanto sus posiciones en el mapa deben interpretarse más cuidadosamente. Por ejemplo, los puntos que están mejor representados son CITIBANK, BANORTE, BBVA BANCOMER y BANAMEX ya que los dos primeros ejes explican respectivamente el 93.1%, 86.7%, 80.9% y 80.7% de la inercia de los puntos. Sin embargo, los puntos SCOTIABANK y SANTANDER no están representados adecuadamente porque caen lejos del plano y tienden a estar más cerca de los ejes principales que conforman las dimensiones mayores.

Con la información proporcionada en las tablas vistas hasta el momento resulta más precisa la interpretación del mapa considerando que algunos de los puntos desplegados no están lo suficientemente explicados para poder hacer una interpretación adecuada de ellos y aunque las posiciones de algunos puntos sean muy similares debe tomarse en cuenta la calidad con la que están siendo representados.

Tercera Dimensión

En este ejemplo es posible analizar también un espacio de mayor dimensión que permita explorar los datos más profundamente. Las tablas (4.5a) y (4.5b) muestran las contribuciones considerando el tercer eje principal cuya proporción de inercia explicada es del 19.8% dando un total acumulado por las tres dimensiones del 85% de la inercia total.

En las tablas se observa que la calidad del mapa y de los puntos aumenta y son mejor representados en un espacio tridimensional, el cual consiste en el espacio bidimensional óptimo anterior añadiéndole un eje perpendicular a éste (anidamiento de ejes principales) lo que genera una mayor dispersión de los puntos representados en el mapa.

BANCOS	MASA	PUNTUACIÓN EN LA DIMENSIÓN			INERCIA	CONTRIBUCIÓN						
		1	2	3		De los puntos a la inercia de la dimensión			De la dimensión a la inercia del punto			
						1	2	3	1	2	3	Total
BANCRECER	.109	-.076	-.150	.249	.012	.011	.088	.264	.053	.205	.565	.823
SCOTIBANK	.036	-.041	-.184	-.444	.010	.001	.043	.277	.006	.126	.733	.865
BANORTE	.161	-.128	-.119	.049	.006	.047	.082	.015	.465	.401	.069	.935
BANAMEX	.140	-.271	.237	.117	.022	.182	.281	.075	.457	.351	.085	.892
BBVA BANCOMER	.142	.189	-.220	.022	.015	.090	.247	.003	.343	.466	.005	.813
SANTANDER	.111	-.135	.052	-.274	.012	.036	.011	.328	.166	.024	.686	.877
IXE	.047	-.088	-.064	-.115	.001	.006	.007	.024	.247	.132	.428	.807
HSBC	.112	-.031	.152	-.030	.005	.002	.093	.004	.020	.488	.019	.527
CITIBANK	.046	.755	.278	.066	.032	.462	.126	.008	.820	.111	.006	.938
OTROS	.097	.305	.081	-.024	.014	.161	.023	.002	.649	.045	.004	.699
Total activo	1.000				.129	1.000	1.000	1.000				

Tabla (4.5a) Examen de los puntos de renglón en tres dimensiones

BANCOS	MASA	PUNTUACIÓN EN LA DIMENSIÓN			INERCIA	CONTRIBUCIÓN						
		1	2	3		De los puntos a la inercia de la dimensión			De la dimensión a la inercia del punto			
						1	2	3	1	2	3	Total
Jornal	.033	.957	-2.351	.400	.013	.030	.181	.005	.132	.396	.011	.539
Honorarios Profesionales	.135	.014	-.898	1.013	.007	.000	.109	.139	.000	.416	.484	.900
Ganancia Negocios	.328	.110	-.361	.319	.007	.004	.043	.033	.033	.179	.128	.340
Rentas	.137	-1.259	1.755	.981	.028	.216	.420	.131	.433	.419	.120	.972
Sueldo Quincenal/ Mensual	.210	-.180	.248	-1.799	.018	.007	.013	.678	.021	.019	.938	.978
Pensiones	.088	-.877	-.912	-.321	.012	.068	.073	.009	.306	.164	.019	.488
Otras	.070	3.110	1.515	.231	.043	.675	.160	.004	.875	.103	.002	.981
Total activo	1.000				.129	1.000	1.000	1.000				

Tabla (4.5b) Examen de los puntos de columna en tres dimensiones

De esta manera, considerando las coordenadas para el tercer eje (columna 3 de "Puntuación en la dimensión"), los puntos que más se oponen son BANCRECER y SCOTIABANK así como los vértices "honorarios profesionales" y "sueldo quincenal o mensual", mientras que los puntos que contribuyen en mayor medida a la orientación del eje son SANTANDER y "sueldo quincenal o mensual", con contribuciones absolutas de 32.8% y 67.8% respectivamente.

Bajo estas condiciones, se puede considerar que los puntos están representados adecuadamente ya que las contribuciones relativas indican, en su mayoría, valores de más del 80% para las categorías de renglón, por ejemplo, BANCRECER aumentó su calidad de 25.8% a 82.3% con la contribución del tercer eje y el punto mejor explicado es CITIBANK con un 93.8%. En las categorías de columna también la calidad aumentó considerablemente

y los puntos mejor representados son "Otras fuentes" y "sueldo quincenal o mensual" con un 98.1% y 97.8% respectivamente.

Los puntos que aún son explicados de forma inadecuada tienden a estar más cerca del cuarto, quinto y sexto ejes principales y, es posible, numéricamente, obtener los datos relativos a estas dimensiones para realizar un mejor diagnóstico. Los puntos pueden ser interpretados con mayor confianza una vez que se analizan las posiciones con respecto a un mayor número de ejes aunque no sea posible visualizarlos gráficamente, sin embargo, por cuestiones de facilitar la interpretación, el número de dimensiones más conveniente es dos, siempre y cuando se tome en cuenta la calidad de los puntos desplegados en el mapa obtenido.

Los resultados mostrados en este capítulo son los que finalmente interesan al investigador, ya que le permiten tomar las decisiones que considere más adecuadas de acuerdo a los objetivos que persiga. Es así como los procedimientos descritos en los capítulos anteriores conforman un método que puede aplicarse en casos específicos de análisis de datos.

Considerando lo anterior, se observa que el Análisis de Correspondencia es un medio para transformar información numérica en representaciones gráficas que presenten los datos de una forma más fácil de interpretar; no es, estrictamente, un medio que permita generar conclusiones determinantes, sino que brinda al investigador un panorama más detallado y más fácil de entender que le proporciona una idea suficientemente acertada y válida del comportamiento de los eventos que le sean de interés.

CONCLUSIONES

Durante el desarrollo de este trabajo no sólo fue posible explicar con mayor profundidad la estructura del Análisis de Correspondencia y conocer algunos aspectos que generalmente se omiten al utilizarlo en la práctica, sino que también destaca la importancia de conocer la aplicación de conceptos que en principio resultan abstractos y que por su sentido teórico no parecen tener mayor relevancia.

La relativa facilidad con la que finalmente se interpretan los resultados del Análisis de Correspondencia a través de los mapas que genera lo hacen, en la práctica, un método muy accesible y de fácil manipulación cuando se conoce el tipo de información que se requiere y la forma en que se realiza la interpretación final.

Debido a que el método ofrece la ventaja de trabajar con datos categóricos existe un número importante de disciplinas que se ven beneficiadas al tener una alternativa más para la realización de análisis especializados, por lo que resulta importante contribuir a la difusión de este tipo de métodos estadísticos.

La aplicación real de los conceptos mencionados en este trabajo permitió darles un sentido más práctico y útil. En su mayoría, estos conceptos son parte del Álgebra Lineal, aunque también intervienen aspectos de Geometría y Estadística, haciendo patente la vinculación entre distintas materias impartidas en la carrera, integrándose no sólo para conformar varios métodos que forman parte del análisis estadístico sino dando a cada uno de ellos características especiales, como en el caso del Análisis de Correspondencia, para ser aplicados en determinadas condiciones.

Esta es sólo una parte de la gran cantidad de aplicaciones que pueden encontrarse al estudiar las distintas asignaturas de la carrera de Matemáticas Aplicadas y Computación, cuya principal ventaja es la interacción que puede tener con diversas disciplinas, generando una fuente muy rica de investigación y desarrollo.

REFERENCIAS BIBLIOGRÁFICAS

Bloom, David. Linear Algebra and Geometry. Cambridge University Press. London. 1979. 625 p.

Carroll, J. Douglas. Mathematical Tools for Applied Multivariate Analysis. Academic Press. San Diego, California. 1997. 367 p.

Greenacre, Michael J. Correspondence Analysis in Practice. Academic Press. San Diego, California. 1993. 195 p.

Hair, Joseph F. Multivariate Data Analysis. Prentice Hall. 5a. Edición. New Jersey, USA. 1998. 730 p.

Hopkins, Kenneth D. Estadística Básica para las Ciencias Sociales y del Comportamiento. Prentice Hall. 3ª. Edición. México, D.F. 1996. 406 p.

Kachigan, Sam. Statistical Analysis, an Interdisciplinary Introduction to Univariate and Multivariate Methods. Radius Press. New York. 1986. 589 p.

Malhotra, Naresh K. Marketing Research, an Applied Orientation. Prentice Hall. 2a. edición. New Jersey, USA. 1996. 890 p.

Morgan, George A. Easy Use an Interpretation of SPSS for Windows. Lawrence Erlbaum Associates Publishers. New Jersey. 1998. 276 p.

Myers, James H. Segmentation And Positioning for Strategic Marketing Decisions. American Marketing Association. Chicago, USA. 1996. 358 p.

Nakos, George. Álgebra Lineal con Aplicaciones. International Thompson. México, D.F. 1999. 666p.

Tabachnick, Barbara. Using Multivariate Statistics. 3a. edición. Harper Collins. New York. 1996. 880 p.

Tinsley, Howard. Handbook of Applied Multivariate Statistics and Mathematical Model. Academic Press. San Diego, California. 2000. 721 p.
