UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

ESCUELA NACIONAL DE ESTUDIOS PROFESIONALES
"A C A T L Á N"

ESCALAMIENTO MULTIDIMENSIONAL. ANÁLISIS DE ASIGNACIÓN DE NIVELES SOCIOECONÓMICOS A HOGARES

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

A C T U A R I A

P R E S E N T A :

JULIA ADELA PALACIOS ROMÁN

ASESOR: ACT. ALBERTO SÁNCHEZ ALDANA.

ACATLÁN, ESTADO DE MÉXICO. JULIO DE 2003.

A mis Padres por todo su apoyo

INDICE

INTRODUCCIÓN

En una economía de constante competencia como la nuestra, las organizaciones públicas y privadas, requieren de información más detallada acerca del consumidor o del mercado en general; es por esto, que la industria de investigación de mercados ha tenido un auge considerable en los últimos años.

Una de las variables más empleadas por las agencias de investigación de mercados es el nivel socioeconómico, razón por la cual, la Asociación Mexicana de Agencias de Investigación de Mercado y Opinión Pública (AMAI), se ha preocupado y ocupado desde su creación en definir un conjunto de niveles socioeconómicos que se conviertan en el estándar de la industria.

El objetivo de esta investigación es analizar las definiciones de niveles socioeconómicos de AMAI y proponer las técnicas de escalamiento multidimensional como un método complementario más que alternativo en la asignación de niveles socioeconómicos.

La muestra empleada en este estudio se obtuvo de una empresa de investigación de mercados, que por confidencialidad de la misma, no se menciona en este trabajo. Dicha muestra de hogares es de la Ciudad de México y contiene toda la información suficiente para realizar el análisis objeto de esta investigación, sin embargo, el tamaño de muestra empleado no fue calculado para obtener resultados representativos de la Ciudad de México, por lo que los resultados aquí obtenidos son representativos únicamente de la muestra empleada y no a un universo mayor.

Las técnicas de escalamiento multidimensional son un caso particular de estadística multivariada para obtener una inspección visual -fácil de entender- ante información suficientemente numerosa sin necesidad de inferencia estadística. Sin embargo, todo el proceso que involucra el obtener una buena representación gráfica con estas técnicas es muy enriquecedor ya que la lógica que aquí se requiere es aplicable a muchas otras técnicas de inspección visual.

Las técnicas de escalamiento multidimensional son descriptivas, pero el proceso que involucra obtener el resultado no es tan simple, especialmente cuando la información de la que se dispone es muy variada y compleja como es el caso de los niveles socioeconómicos que define la Asociación Mexicana de Agencias de Investigación de Mercados (AMAI), objetivo de esta investigación.

En el primer capítulo se hace una disección total de conceptos que involucran el escalamiento multidimensional y en particular, conceptos relacionados con el término "proximidad", así como algunos métodos interesantes de obtener dichas proximidades.

En el segundo capítulo se describen las técnicas de escalamiento multidimensional métrico, donde, se parte del supuesto que las proximidades insumo son distancias.

En el tercer capítulo se describen diferentes técnicas de escalamiento donde el supuesto de métrica ya no es necesario; haciendo énfasis en el escalamiento de Kruskal.

En el cuarto capítulo se describen ciertas técnicas como Diferencias Individuales, etc. Dichas técnicas son un tanto más complejas que las de los capítulos segundo y tercero, pero en la práctica son las más empleadas e interesantes.

Finalmente, en el quinto y último capítulo se presentan variadas aplicaciones al problema de asignación de niveles socioeconómicos a hogares, partiendo de las definiciones AMAI y concluyendo con una nueva asignación de niveles socioeconómicos a hogares, de acuerdo a los resultados obtenidos con estas técnicas.

Las primeras aplicaciones aquí presentadas, son en cierta medida, ejemplos de configuraciones obtenidas con las técnicas de escalamiento multidimensional, pero poco enriquecedoras al problema de asignación de niveles socioeconómicos a hogares. En las últimas aplicaciones se obtienen resultados interesantes, que no son necesariamente los resultados esperados, pero que dejan un tema abierto a futuras investigaciones.

CAPÍTULO 1

CONCEPTOS BÁSICOS DE ESCALAMIENTO MULTIDIMENSIONAL

CAPÍTULO 1

CONCEPTOS BÁSICOS DE ESCALAMIENTO MULTIDIMENSIONAL

La visualización de los datos consiste en explorar, analizar y presentar los mismos. Una representación gráfica es generada por el contenido de los datos y analizada por un observador, con sentido humano y habilidad para reconocer patrones. Por ejemplo, la correlación que presenten dos variables puede observarse con una gráfica de dispersión (scatter plot), sin embargo, una visualización efectiva puede ser difícil para una colección abstracta de datos como una base de datos con muchos atributos, dado que no hay una manera obvia inmediata de ordenar a los objetos basándonos en su contenido. Afortunadamente, la semejanza o disimilaridad entre pares de elementos de tal colección puede medirse y obtener una representación gráfica de manera que los objetos semejantes estén cercanos y los más disimilares más lejanos.

Un método estándar para modelar relaciones de proximidades (semejanza o disimilaridad) es el escalamiento multidimensional. Generalmente, el resultado es una configuración de puntos en dos o tres dimensiones donde cada punto representa a un solo elemento de la colección.

Algunos autores emplean el término de escalamiento multidimensional a una gran variedad de técnicas de análisis de exploración de datos que incluyen el análisis por conglomerados[1] y análisis por factores[2]. Otros autores dan una definición más limitada, refiriéndose a una clase de técnicas matemáticas que permiten al investigador descubrir la "estructura oculta" de los datos encontrando una representación espacial de proximidades como distancias entre puntos en un espacio multidimensional de pocas dimensiones. Este trabajo aborda el tema en el sentido más limitado: Encontrar una representación espacial de objetos en un espacio de pocas dimensiones empleando proximidades como distancias.

En este capítulo se abordarán los diferentes objetivos que un investigador puede tener al emplear las técnicas de escalamiento multidimensional, se definen algunos conceptos necesarios para familiarizarse con las técnicas y se abordarán algunos métodos para obtener información apropiada para un buen análisis de escalamiento multidimensional.

[1]El análisis por conglomerados (Cluster Analysis) es una técnica multivariada empleada para clasificar individuos o unidades experimentales en subgrupos definidos de manera única. Este análisis trata de los problemas de clasificación cuando no se sabe de antemano de cuáles subgrupos se originan las observaciones.

[2]El análisis por factores (Factor Analysis) es una técnica multivariada que se emplea frecuentemente para crear nuevas variables que resuman toda la información de la que podría disponerse en las variables originales. Un objetivo básico del análisis por factores es estudiar la estructura de correlación de las variables en un conjunto de datos.

1.1 EL PROPÓSITO DEL ESCALAMIENTO MULTIDIMENSIONAL

El objetivo principal de una ciencia, además de una descripción empírica del fenómeno, es establecer, a través de leyes y teorías, principios generales con los que el fenómeno empírico puede ser explicado; como se puede observar, la medida es una de las cosas que permite que este proceso se realice.

El problema de la medición consiste en encontrar procedimientos que nos permitan asignar un número a cierta cantidad de atributo (tipo particular de la propiedad de un objeto), de tal manera que se reflejen relaciones de analogía entre los números y las cantidades del atributo, donde objeto se refiere a todo aquello que posee un atributo y el término cantidad se refiere al monto particular o grado que el objeto posee el atributo, de esta manera, un objeto posee un gran número de atributos: color, tamaño, textura, peso, etc.

La asignación de un número o un conjunto de números a las cantidades del atributo o atributos es análoga a la especificación de la posición de los puntos en un espacio multidimensional. El término dimensión se refiere a la caracterización de un atributo particular, por ejemplo, si hablamos del atributo "costo" de un objeto, las dimensiones podrían ser caro y barato.

Cuando la medida es unidimensional, el atributo corresponde a una línea recta y la cantidad a un punto sobre una línea recta. Cuando la medición es multidimensional, el atributo corresponde a un espacio n-dimensional y la cantidad a un punto en ese espacio. El proceso de asignar conjuntos de números corresponde a la posición de los puntos en un espacio multidimensional en términos de un conjunto de relaciones entre los puntos especificados por un modelo geométrico particular.

Los modelos geométricos que pueden emplearse son, por supuesto, de muchas variedades diferentes, como espacios métricos[1], tal como la familia de modelos euclideanos y no euclideanos y varios modelos no métricos, dependiendo de la naturaleza del problema.

En el escalamiento, se establecen reglas de correspondencia; se establece el significado de los números, elementos y propiedades del modelo a datos observables que convierten al modelo en una teoría que puede ser demostrada. Si se verifica la teoría, se asignan números a las cantidades del atributo multidimensional. Una vez que se tiene esto, hemos escalado o "medido" el atributo multidimensional en cuestión.

El problema clásico que se tiene en las técnicas de escalamiento multidimensional se establece de la siguiente manera: Dado un conjunto de objetos que varían respecto a un número desconocido de atributos (dimensiones) determinar:

1.- La dimensionalidad mínima del conjunto, es decir, desechar aquellas dimensiones que tienen una importancia relativamente insignificante para dejar únicamente aquellas dimensiones clave necesarias para obtener una inspección visual relativamente fácil.

2.- Encontrar una representación espacial de los puntos representando a cada uno de los objetos de tal manera que las distancias entre objetos se asemeje lo mejor posible a las disimilaridades originalmente percibidas.

[1]Véase Pág. 7, subtema 1.2.2 Proximidades y su naturaleza métrica.

Los autores pioneros en desarrollar y utilizar estas técnicas de escalamiento multidimensional, tratan concretamente los problemas con los que se han enfrentado y han presentado una gran variedad de aplicaciones.

Algunas aplicaciones del escalamiento multidimensional se presentan a continuación:

En psicología:

Psicólogos han utilizado estas técnicas para entender la percepción y evaluación de estímulos auditivos, tales como el habla y tonos musicales, estímulos visuales como colores y estímulos sociales como rasgos de personalidad y situaciones sociales.

En sociología:

Sociólogos han utilizado los métodos para determinar estructura de grupos y organizaciones, basados en percepciones entre sus miembros y sus interacciones.

En antropología:

Antropólogos han usado estos métodos para comparar diferentes grupos culturales, basados en sus creencias, lenguajes y artificios.

En economía:

Economistas han utilizado estos métodos para investigar reacciones de los consumidores a una gran variedad de clases de productos.

En mercadotecnia:

En investigación de mercados donde se busca una segmentación del mercado, es decir, diferenciar el mercado total de un producto o servicio, en un cierto número de elementos (personas u organizaciones) homogéneos entre sí y diferentes de los demás, en cuanto a hábitos, necesidades y gustos de sus componentes, que denominan segmentos, de manera que se puedan emplear a cada segmento las estrategias de marketing más adecuadas para lograr los objetivos establecidos por la empresa.

En Seguros:

Los actuarios podrían emplear estas técnicas para clasificar su cartera de clientes de acuerdo a los riesgos y así calcular tarifas o descuentos, de manera que por ejemplo en seguro de autos, las personas que nunca hayan chocado y lleven tiempo manejando y tengan alarma en su automóvil obtengan un descuento en la tarifa de su seguro.

Como se puede observar, los propósitos que puede tener un investigador al utilizar las técnicas de escalamiento multidimensional pueden ser muy variados, sin embargo, se pueden resumir en 4 propósitos fundamentales:

1. Como técnica de exploración

El escalamiento multidimensional es generalmente empleado como método que representa datos de proximidades como distancias en un espacio de pocas dimensiones de tal manera que estos datos sean accesibles a una inspección visual y exploración.

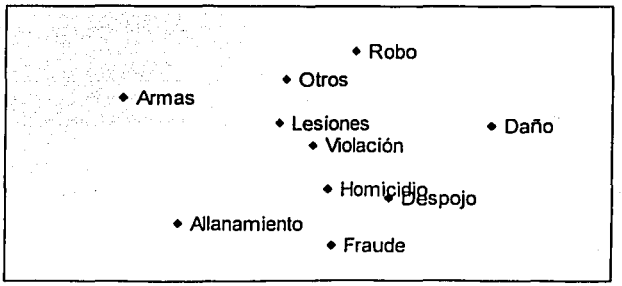
El análisis exploratorio de los datos es utilizado para estudiar teóricamente la forma de los datos, esto es, datos que no se encuentran ligados a una teoría explícita que predice sus magnitudes o patrones.

Considerando el siguiente ejemplo de estadísticas publicadas por INEGI[1], surge la siguiente interrogante: ¿se puede suponer una tasa alta de ocurrencia de homicidios si se sabe que hay una tasa alta de ocurrencia de delitos de despojo? Una respuesta parcial a esta interrogante se puede obtener calculando las correlaciones de las tasas de criminalidad en las entidades federativas de la República Mexicana (tabla 1), sin embargo, aunque se tenga una pequeña tabla de correlaciones es difícil interpretar la estructura de estos coeficientes. Este problema resulta más sencillo si se tiene una ilustración gráfica de esta información (figura 1)[2]. La gráfica es una representación de escalamiento multidimensional en dos dimensiones donde cada punto representa a cada uno de los delitos.

Tabla 1.
Correlaciones de la tasa de criminalidad entre las entidades federativas de la República Mexicana, 2000.

	Robo	Lesiones	Daño	Homicidio	Fraude	Despojo	Violación	Allanamiento	Armas	Otros
Robo	1.000	0.617	0.697	0.531	0.334	0.482	0.673	0.272	0.597	0.795
Lesiones	0.617	1.000	0.577	0.715	0.626	0.593	0.734	0.472	0.568	0.855
Daño	0.697	0.577	1.000	0.623	0.589	0.534	0.413	0.346	0.270	0.417
Homicidio	0.531	0.715	0.623	1.000	0.789	0.750	0.819	0.694	0.402	0.617
Fraude	0.334	0.626	0.589	0.789	1.000	0.644	0.602	0.624	0.387	0.445
Despojo	0.482	0.593	0.534	0.750	0.644	1.000	0.863	0.460	0.409	0.393
Violación	0.673	0.734	0.413	0.819	0.602	0.863	1.000	0.561	0.450	0.786
Allanamiento	0.272	0.472	0.346	0.694	0.624	0.460	0.561	1.000	0.552	0.419
Armas	0.597	0.568	0.270	0.402	0.387	0.409	0.450	0.552	1.000	0.530
Otros	0.795	0.855	0.417	0.617	0.445	0.393	0.786	0.419	0.530	1.000

Figura 1
Representación de escalamiento multidimensional en dos dimensiones de las correlaciones de la tabla 1



[1]Fuente: INEGI: Estadísticas sociodemográficas 2000
[2]Representación en dos dimensiones empleando la técnica de escalamiento multidimensional métrico clásico

Los puntos en la gráfica se ubican de acuerdo a sus distancias correspondientes a las correlaciones, lo que significa, que dos puntos se encuentran cercanos (como homicidio y despojo) si sus tasas de criminalidad son altamente correlacionadas y, de la misma manera, dos puntos se encuentran lejanos (como armas prohibidas y daño a las cosas) si sus tasas de criminalidad son poco correlacionadas.

2. Para probar hipótesis estructurales

Cuando se tiene más información del fenómeno en estudio, los métodos de exploración pueden ser menos importantes y el interés se centra en pruebas de hipótesis, es decir, se quiere probar si cierto criterio por el cual se puede distinguir diferencias en los objetos de interés corresponde a las diferencias empíricas apreciadas.

3. Para explorar estructuras psicológicas

Cuando el interés es descubrir las dimensiones (caracterización de los atributos de los objetos) de los juicios de similitud o disimilitud. Este tipo de modelos se emplea en estudios de investigación de mercados donde el interés es conocer la apreciación del consumidor acerca de un producto o varios productos.

4. Como modelo de juicios de semejanza

Las matemáticas del escalamiento multidimensional pueden servir como modelo de juicios de similitud o disimilitud en términos de una regla que refleja la distancia psicológica en un tipo de función de distancia en particular, es decir, cuando a una persona se le preguntan juicios de similitud o disimilitud entre pares de objetos, actúa como si calculara una distancia en su "espacio psicológico" y esta distancia puede expresarse como una función de distancia particular.

1.2 TIPOS DE DATOS Y PROXIMIDADES

Los modelos de escalamiento multidimensional emplean proximidades como insumo, y algunas veces puede resultar muy interesante encontrar una medida de proximidad que nos pueda ayudar a explicar el fenómeno y ser analizado por escalamiento multidimensional. En gran parte, la medida de proximidad depende del tipo de datos o variables que se emplean en el estudio.

1.2.1 TIPOS DE DATOS[1]

Las técnicas de escalamiento multidimensional miden atributos percibidos o diferencias percibidas entre atributos. Los datos que se emplean son categóricos y de acuerdo a su escala de medición, los datos pueden ser nominales, ordinales o numéricas.

Nominales:
Los datos nominales pueden clasificarse y solo son distinguibles entre clases de atributos, como el color del cabello, el color de los ojos, etc.

Ordinales:
Los datos ordinales de una categoría pueden ser ordenados relativamente a lo datos de otra categoría, pero no son datos cuantitativos. Por ejemplo en un estudio de investigación de mercados, puede saberse si un producto alimenticio es juzgado como más sabroso que otro producto alimenticio de su mismo tipo.

Numéricos:
Los datos de una categoría están funcionalmente relacionados a los datos de otra categoría. Los datos numéricos pueden ser de intervalo o de razón, por ejemplo, en el caso de datos de intervalo: la temperatura en grados Celsius tiene diferente significado antes (categoría 1) y después de hacer ejercicio (categoría 2).

En cuanto al proceso de medición, es decir, la naturaleza permisible de la relación entre datos de la misma categoría, se tienen dos procesos de medición:

Discreto: Todos los datos en la categoría son representados por un número.
Continuo: Todos los datos en la categoría son representados por un intervalo de números.

[1]Fuente: Young F.W. & Hammer, R.M, 1987

La condicionalidad de la medición, es decir, la naturaleza permisible de la relación de datos entre conjuntos de categorías, puede ser:

Incondicional: Todos los datos son comparables.
Condicional a renglón/columna: Únicamente los datos entre renglones/columnas de una matriz son comparables.
Condicional a matriz: Solo los datos entre matrices son comparables (generalmente se tiene este caso).

El escalamiento multidimensional trabaja con datos relacionados a objetos, individuos, sujetos o estímulos. La medida más común que refleja las relaciones entre objetos, individuos, etc., es la medida de proximidad.

1.2.2 PROXIMIDADES Y SU NATURALEZA MÉTRICA

Una proximidad significa, literalmente, cercanía en espacio, tiempo, etc. Es un número que indica que tan similares o diferentes son o son percibidos dos objetos. La "cercanía" entre objetos, individuos o estímulos necesita una definición y medida antes de proceder al análisis estadístico. En algunas situaciones esto es una regla, sin embargo, en otras situaciones esto puede ser difícil y controvertido.

Las medidas de proximidad son de dos tipos: similaridad (s_{ij}) y disimilaridad (δ_{ij}) con su respectiva interpretación de medida de que tan similar o diferentes son los objetos entre sí.

Al referirnos a una medida de que tan diferentes o similares son dos objetos entre sí, surge la idea de distancia, sin embargo, esta medida puede definirse tanto en espacios métricos como no métricos. A continuación se mencionan ciertas definiciones necesarias para formar posibles estructuras de proximidad.

Métrica:

$d_{ij} : \mathbf{O} \mathbf{x} \mathbf{O} \rightarrow \mathbf{R}^{+} \cup \{0\}$ es una métrica si:

(1) simetría $d_{ij} = d_{ji}$ para todo $1 \leq i, j \leq n$,

(2) $d_{ij} = 0$ si y solo si $i=j$

(3) $d_{ik} \leq d_{ij} + d_{jk}$ para todo $1 \leq i,j,k \leq n$.

Espacio métrico:

Un espacio métrico es un par $(\mathbf{O}, d)$ donde $\mathbf{O}$ es un conjunto y d es una métrica sobre $\mathbf{O}$.

Espacio euclideano:

El conjunto $\mathbf{R}^n$ formado por las n-adas de números reales

$$\mathbf{R}^n = \{(x_1,...,x_n) \mid x_i \in \mathbf{R}, i = 1,...,n\}$$

es el espacio euclideano de n dimensiones.

De esta manera, ejemplos de posibles estructuras de proximidad definidas en un espacio métrico son:

Distancia euclideana $\delta_{ij} = \left\{ \sum_k (x_{ik} - x_{jk})^2 \right\}^{1/2}$

Métrica city block $\delta_{ij} = \sum_k |x_{ik} - x_{jk}|$

Ejemplo de posibles estructuras de proximidad definidas en un espacio no métrico son:

Proximidad tricotómica que divide al conjunto de objetos en tres subconjuntos:
- Objetos similares
- Objetos diferentes
- Los pares restantes

Como puede observarse, las estructuras posibles de proximidad son muy variadas y su elección depende del problema que se tenga y de la información disponible.

Las proximidades son necesarias para un modelo de escalamiento multidimensional y la manera en que las proximidades son generadas tiene implicaciones para la elección del modelo, es por esto que se estudiará a mayor detalle esta sección.

Las proximidades pueden obtenerse directamente por juicio de que tan semejantes o diferentes son pares de objetos o pueden derivarse de un vector de atributos asociado a cada uno de los objetos.

1.2.2.1 PROXIMIDADES DIRECTAS[1]

El ejemplo clásico en la mayoria de la literatura de escalamiento multidimensional consiste en construir un mapa con principales ciudades, de donde se tiene como insumo una matriz de proximidades con las distancias medidas directamente de un atlas.
Las proximidades directas generalmente se obtienen al preguntar a la gente su juicio de "distancia psicológica o perceptual" (o cercania) entre objetos o estimulos. Con la finalidad de descubrir más que imponer las dimensiones, los atributos por los cuales el objeto puede ser juzgado, generalmente no es especificado con anterioridad.

El método más popular para recolectar proximidades directas es pedir a los individuos que califiquen a cada par de objetos en una escala de disimilaridad o semejanza (0= no son semejantes,..., 5= idénticos) y las proximidades se pueden obtener promediando o contando esos juicios para todos los diferentes pares de objetos.

[1]Fuente: Cox & Cox. 1994

En la investigación de mercados, se generan tarjetas con parejas de objetos y se le pide al encuestado que ordene las tarjetas de los pares más semejantes a los pares más diferentes.

Un método útil para conjuntos de objetos muy grandes (de 50 a 100 objetos) es pedir al encuestado ordenar a los objetos o conjuntos de objetos en categorías exhaustivas y disjuntas de manera que los objetos en la misma categoría son más semejantes entre ellos que con los de las otras categorías y se puede generar una matriz de proximidades entre los objetos contando las veces en que cada par de objetos se pusieron en la misma categoría, este método es llamado "ordenamiento libre".

Otro método también para conjuntos de objetos extensos, es el método de "ordenamiento-Q de Stephenson" que consiste en pedirle a los individuos que formen dos grupos, no necesariamente del mismo tamaño, uno con las parejas de objetos semejantes y otro con las parejas de objetos diferentes y luego se le pide que repita el proceso con cada grupo, y así sucesivamente. Al conjunto con los objetos más semejantes se le da una calificación de 1 y al siguiente conjunto con objetos semejantes de 2, etc.

El método de ancla consiste en elegir a un objeto A de una colección de n objetos y comparar a los n-1 objetos restantes con A, proporcionando así, juicios de semejanza. Cada objeto de la colección es elegido como ancla, de esta manera se tienen n conjuntos con n-1 proximidades. Las proximidades que resultan del método de ancla son condicionales y comparar la proximidad entre el elemento ancla A y X con la proximidad entre el elemento ancla B y Y no tiene sentido, por lo que las proximidades derivadas de este método requieren métodos de escalamiento multidimensional particulares, con una función de error más débil [2], sin embargo, emplear datos condicionales tiene la ventaja de que se requiere calificar menos datos al mismo tiempo, en vez de calificar las n combinaciones de 2 parejas de objetos, solo se requiere calificar a n-1 objetos al mismo tiempo.

1.2.2.2 PROXIMIDADES DERIVADAS[1]

En la práctica, las proximidades directas son casos atípicos, generalmente son índices derivados de otra información, como correlaciones o distancias calculadas para pares de variables X y Y. Estos índices dependen de la escala de medición de las variables X y Y.

Suponiendo que $X=[x_n]$ es una matriz de datos obtenida para n objetos en p variables. El vector de observaciones para el r-ésimo objeto se denota por x_r.

Variables nominales:

Si para la i-ésima variable nominal, los objetos *r* y *s* comparten la misma categorización, entonces s_{rsi} =1 y 0 en otro caso. La medida de similaridad es:

$$p^{-1}\sum_i s_{rsi}$$

[1]Fuente: Borg & Groenen, 1997
[2]Se profundizará en la función de error en lo subsiguiente.

Si se cuenta con mayor información de la relación entre varias categorías, entonces se tendría que dar una medida apropiada para s_{rsl}. Por ejemplo, en el caso de la variable "forma de la botella de una bebida" tiene las siguientes categorías: estándar (1), chica y cilíndrica (2), larga y delgada (3) o cuadrada (4), se podría emplear la siguiente matriz:

		Botella r			
		1	2	3	4
Botella s	1	1	0.5	0.5	0
	2	0.5	1	0.3	0
	3	0.5	0.3	1	0
	4	0	0	0	1

De esta manera, si la botella s es estándar y la botella r es larga y delgada, s_{rsl} =0.5.

Variables ordinales:

Si la variable i-ésima es ordinal con k categorías, entonces existen k-1 variables indicadoras que pueden emplearse para representar estas categorías y así obtener medidas de semejanza. Por ejemplo, en el caso de la variable "gusto por un producto", se tienen las categorías: me gustó muchísimo, me gustó, ni me gustó ni me desagradó, no me gustó del todo, no me gustó para nada. Se podría emplear lo siguiente:

	Ind 1	Ind 2	Ind 3	Ind 4
Me gustó muchísimo	0	0	0	0
Me gustó	1	0	0	0
Ni me gustó ni me desagradó	1	1	0	0
No me gustó del todo	1	1	1	0
No me gustó para nada	1	1	1	1

Con la matriz anterior y aplicando el coeficiente de aparejamiento simple siguiente:

Tabla 2

		Objeto s	
		1	0
Objeto r	1	a	b
	0	c	d

$$s_{rs} = \frac{a+d}{a+b+c+d}$$, donde a, b, c y d son las frecuencias.

Variables numéricas:

La Tabla 3 proporciona una lista con medidas de disimilaridad para datos cuantitativos que son en particular, continuos, posiblemente discretos, pero no binarios. Las correlaciones entre variables generalmente se calculan entre individuos, es decir, los datos se encuentran en la forma típica personas o compañías *x* variables. Suponiendo que una compañía de seguros quiere encontrar una medida de semejanza entre un conjunto de automóviles de acuerdo a su riesgo; una forma de hacerlo podría ser preguntar a n compañías de

seguros la siniestralidad de cada automóvil. Las proximidades pueden calcularse correlacionando las siniestralidades entre las compañías, pero solo se estaría considerando una característica y tendría más sentido cuestionar a las compañías acerca del lugar donde ocurrió el siniestro, la hora del siniestro, presencia de alarma anti-robo, iluminación del lugar de siniestro, etc. En este caso se tendría una matriz de la forma compañías *x* variables *x* automóviles.

Una posible forma de obtener medidas de semejanza puede ser calculando las correlaciones de los atributos para los automóviles, obteniendo una matriz por compañía o correlacionando los automóviles entre compañías y atributos, obteniendo una sola matriz global. Otra alternativa podría ser calculando las distancias entre los vectores de atributos, por ejemplo, si X es una matriz de la forma automóviles *x* atributos, que contenga el promedio de atributo de n compañías. Una distancia simple en los renglones de X es la distancia "city block".

Variables binarias:

Cuando todas las variables son binarias, la medida de semejanza en el objeto r y el objeto s se basa en la tabla 2 y en la tabla 4 de la página 13.

Tabla 3

Distancia Euclideana	$\delta_{ij} = \left\{ \sum_k (x_{ik} - x_{jk})^2 \right\}^{1/2}$
Distancia Euclideana ponderada	$\delta_{ij} = \left\{ \sum_k w_k (x_{ik} - x_{jk})^2 \right\}^{1/2}$
Distancia de Mahalanobis	$\delta_{ij} = \{(x_i - x_j)^T \Sigma^{-1} (x_i - x_j)\}^{1/2}$
Métrica city block	$\delta_{ij} = \sum_k \lvert x_{ik} - x_{jk} \rvert$
Métrica de Minkoski	$\delta_{ij} = \left\{ \sum_k \lvert x_{ik} - x_{jk} \rvert^{\lambda} \right\}^{1/\lambda} \quad \lambda \geq 1$
Métrica de Canberra	$\delta_{ij} = \sum_k \lvert x_{ik} - x_{jk} \rvert / (x_{ik} + x_{jk})$
Bray-Curtis	$\delta_{ij} = \sum_k \lvert x_{ik} - x_{jk} \rvert / p \sum_k (x_{ik} + x_{jk})$
Distancia de Bhattacharyya	$\delta_{ij} = \left\{ \sum_k x_{ik}^{1/2} - x_{jk}^{1/2} \right\}^{1/2}$
Correlación	$\delta_{ij} = 1 - \dfrac{\sum_k (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)}{\left\{ \sum_k (x_{ik} - \bar{x}_i)^2 \sum_k (x_{jk} - \bar{x}_j)^2 \right\}^{1/2}}$
Separación Angular	$\delta_{ij} = \dfrac{\sum_k x_{ik} x_{jk}}{\left[\sum_k x_{ik}^2 \sum_k x_{ij}^2 \right]^{1/2}}$

Tabla 4

Czekanowski, Sorensen, Dice	$s_{ij} = \frac{2a}{2a+b+c}$
Hamman	$s_{ij} = \frac{a-(b+c)+d}{a+b+c+d}$
Coeficiente de Jaccard	$s_{ij} = \frac{a}{a+b+c}$
Kulezynski	$s_{ij} = \frac{a}{a+b}$
Mountford	$s_{ij} = \frac{2a}{a(b+c)+2bc}$
Mozley, Margalef	$s_{ij} = \frac{a(a+b+c+d)}{(a+b)(a+c)}$
Ochiai	$s_{ij} = \frac{a}{[(a+b)(a+c)]^{1/2}}$
Phi	$s_{ij} = \frac{ad-bc}{[(a+b)(a+c)(b+d)(c+d)]^{1/2}}$
Rogers, Tanimoto	$s_{ij} = \frac{a+d}{a+2b+2c+d}$
Russell, Rao	$s_{ij} = \frac{a}{a+b+c+d}$
Coeficiente de aparejamiento simple	$s_{ij} = \frac{a+d}{a+b+c+d}$
Yule	$s_{ij} = \frac{ad-bc}{ad+bc}$

Fuente: Cox & Cox, 1994

1.2.3 TRANSFORMACIONES DE PROXIMIDADES[1]

Generalmente, los coeficientes de similaridad necesitan transformarse en coeficientes de disimilaridad. Las transformaciones posibles son las siguientes y su elección depende del problema que se tenga.

$$\delta_{ij} = 1 - s_{ij}$$

$$\delta_{ij} = c - s_{ij}, \text{ para alguna constante } c$$

$$\delta_{ij} = \left[2(1 - s_{ij})\right]^{1/2}$$

Sea D la matriz de disimilaridades $\{\delta_{ij}\}$, entonces D es una matriz métrica si δ_{ij} es una métrica y D es euclideana si n puntos pueden fijarse en un espacio euclideano tal que la distancia euclideana entre el punto i y el punto j sea δ_{ij}, para todo1 $\leq$ i, j $\leq$ n.

Si D es una matriz no métrica con elementos $\delta_{ij} + c$ (i $\neq$ j) es métrica cuando $c \geq \max_{i,j,k} \left|\delta_{ij} + \delta_{ik} - \delta_{jk}\right|$

Si D es métrica, entonces:

i) $\delta_{ij} + c^2$

ii) $\delta_{ij}^{1/\lambda}$ $(\lambda \geq 1)$

iii) $\dfrac{\delta_{ij}}{\delta_{ij} + c^2}$

Son métricas.

Sea Λ=[-(1/2) d_{ij}^2], entonces, D es euclideana si y solo si la matriz $(\mathbf{I} - \mathbf{1s}^T)\Lambda(\mathbf{I} - \mathbf{1s}^T)$ es positiva semidefinida[2], donde I es la matriz identidad, **1** es un vector de unos y **s** es un vector tal que $\mathbf{s}^T\mathbf{1}$=1.

Si S es una matriz positiva semidefinida de similitud con elementos $0 \leq s_{ij} \leq 1$ y s_{ij}=1, entonces la matriz de disimilaridad con elementos $d_{ij} = (1 - s_{ij})^{1/2}$ es euclideana.

Si D es una matriz de disimilaridades, entonces existe una constante *h* tal que la matriz con elementos $(\delta_{ij}^2 + h)^{1/2}$, es euclideana cuando $h \geq -\lambda_n$, el menor eigenvalor de Λ_1=HΛH, donde H= (I – 11^T/n)

Si D es una matriz de disimilarides, entonces existe una constante *k* tal que la matriz con elementos $(\delta_{ij} + k)$ es euclideana, donde $k \geq \lambda_n$, el mayor eigenvalor de

$$\begin{bmatrix} O & 2\Lambda_1 \\ -I & -4\Lambda_2 \end{bmatrix}, \text{donde } \Lambda_2 = [\tfrac{1}{2} d_{ij}]$$

[1] Mardia et al, 1979

[2] Una matriz simétrica A de n x n es positiva semidefinida si todo eigenvalor de A es no negativo.

1.3 TIPOS DE ESCALAMIENTO MULTIDIMENSIONAL

Los datos que se emplean en un escalamiento multidimensional se describen por su número de factores, los niveles correspondientes a cada factor y el número de modos. Un factor como en la teoría de diseño de experimentos, se refiere al número de condiciones experimentales manipulables, los niveles se refieren al número de los diferentes valores que puede tomar cada factor y modo se refiere al número de factores únicos en el diseño. Por ejemplo, si la información que se dispone para un escalamiento multidimensional consiste en 50 estudiantes con un cuestionario de 20 preguntas, se tendrían {50x20} datos. Estos datos son de 2 factores, 2 modos, 50x20. Hay 50 niveles para el factor estudiantes y 20 niveles para el factor juicio. Si se tuvieran 15 juicios de una sola persona, los datos serían de la forma: 2 factores, 1 modo, 15x15. Suponiendo que se quiere comparar a 5 compañías aseguradoras, de acuerdo a 20 características de los siniestros de robo de automóviles de 10 marcas de automóviles; en este caso, se tendían {5x5x20x10} datos. Estos datos son de 4-factores, 3-modos. Hay 5 niveles para las compañías, 20 niveles para las características de los siniestros de robo y 10 niveles para los automóviles.

De esta manera, las disimilaridades de la forma $\delta_{ij,k}$ son de dos modos, un modo son los objetos y el otro modo los juicios a esos objetos. Cada índice en la medición entre objetos es el tratamiento o factor, por lo que las disimilaridades $\delta_{ij,k}$ son de tres factores.

Suponiendo un conjunto de n objetos con disimilaridades δ_{ij} medidas entre todas las parejas posibles de los objetos. Una configuración de los n puntos representando a los objetos sería en un espacio de p dimensiones. Suponiendo que las distancias entre pares de objetos, no necesariamente euclideanas, son $\{d_{ij}\}$, entonces el propósito del escalamiento multidimensional es encontrar una configuración tal que las distancias $\{d_{ij}\}$, correspondan lo mejor posible a las disimilaridades $\{\delta_{ij}\}$. Esta correspondencia da lugar a los diferentes tipos de modelos de escalamiento multidimensional.

Escalamiento Clásico

Si las distancias del espacio de configuración son euclideanas y $d_{ij} = \delta_{ij}$ i,j=1,.....,n (1), entonces las disimilaridades son distancias euclideanas y es posible encontrar una configuración de acuerdo a (1). El escalamiento clásico trabaja con las disimilaridades δ_{ij} directamente como distancias euclideanas y aplica el teorema de la descomposición espectral que en el siguiente capítulo se tratará a mayor detalle.

Escalamiento por mínimos cuadrados

El escalamiento por mínimos cuadrados encuentra distancias $\{d_{ij}\}$ que correspondan a disimilaridades transformadas $\{f(\delta_{ij})\}$, donde f es una función monótona continua. La función f es tal que se encuentre una configuración que mejor ajuste las distancias ajustadas por mínimos cuadrados a $\{f(\delta_{ij})\}$. Por ejemplo una configuración puede encontrarse tal que la función de error siguiente sea minimizada.

$$\frac{\sum_i \sum_j (d_{ij} - (a + b\delta_{ij}))^2}{\sum_i \sum_j d_{ij}^2}$$

El escalamiento clásico y el escalamiento por mínimos cuadrados son casos de **"escalamiento métrico"**, donde el término métrico se refiere al tipo de transformaciones a las disimilaridades y no al espacio en el cual los puntos son configurados.

Escalamiento no métrico

Si se abandona la naturaleza métrica de las transformaciones a las disimilaridades, se emplea el escalamiento no métrico. La trasformación f puede ser arbitraria siempre y cuando: $\delta_{ij} < \delta_{i'j'} \Rightarrow f(\delta_{ij}) \leq f(\delta_{i'j'})$ para todo $1 \leq i, j, i', j' \leq n$.

Análisis de procrustes[1]

Suponiendo que el escalamiento multidimensional se ha trabajado con datos de disimilaridad empleando dos métodos distintos y obteniendo así, dos configuraciones de puntos que representan a los mismos objetos. El análisis de procrustes dilata, traslada, rota, etc., una de las configuraciones de puntos tal que correspondan, lo mejor posible, a la otra configuración, pudiendo realizar comparaciones entre configuraciones.

Desdoblamiento (Unfolding)

Suponiendo que se tiene n juicios de m tipos de bebidas refrescantes, cada uno ordena a las bebidas refrescantes de acuerdo a su preferencia personal. El método de desdoblamiento produce una configuración de puntos donde cada punto representa uno de los juicios junto con otra configuración de puntos en el mismo espacio, representando a las bebidas refrescantes.

Diferencias individuales

Suponiendo de nuevo que se tienen n juicios comparando todos los pares de bebidas refrescantes, se pueden realizar n análisis de escalamiento multidimensional por separado. El modelo de diferencias individuales produce una configuración general de puntos representando las bebidas refrescantes, el cual es llamado el espacio de grupos de objetos o estímulos, junto con una configuración de puntos representando a los juicios en un espacio diferente llamado espacio del individuo.

Análisis de correspondencia

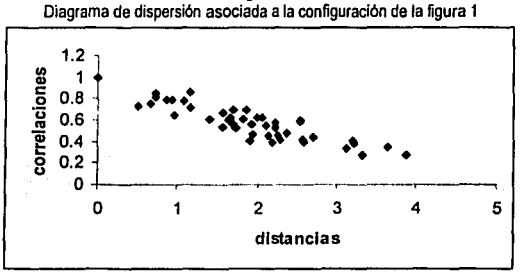
Los datos de la forma de una tabla de contingencia de dos factores pueden analizarse por análisis de correspondencia, donde se encuentra un espacio en el cual se configuran los renglones y otro donde se configuran las columnas.

[1]Procrustes no tiene traducción al español. Su nombre se debe a la mitología Griega. Procrustes o Damastes, vivía a la orilla del camino a Atenas y cualquier viajero que pasaba por ese camino era invitado de Damastes quien los sorprendía por su hospitalidad y la cama para dormir del invitado. Si su invitado no cabía en la cama, Damastes ajustaba la cama, haciéndola más pequeña o larga según sea el caso. Por tal motivo, a Damastes lo llamaron Procrustes que significa extendimiento o bastidor

1.4. DEFINICIÓN DE LA FUNCIÓN OBJETIVO

El concepto central del escalamiento multidimensional es que las distancias d_{ij} entre los puntos correspondan a las proximidades δ_{ij}. Una buena manera de visualizar esta correspondencia es mediante un diagrama de dispersión o diagrama de Shepard. En el ejemplo del apartado 1.1. con las estadísticas de los delincuentes, se obtuvo el siguiente diagrama de dispersión (figura 2).

Figura 2.
Diagrama de dispersión asociada a la configuración de la figura 1



Suponiendo que las proximidades son disimilaridades, entonces disimilaridades pequeñas corresponderían a distancias pequeñas y grandes disimilaridades a grandes distancias, es decir, se tendría un patrón creciente en el diagrama de dispersión y en el caso de similitudes como en la figura anterior (figura 2) se tendría un patrón decreciente. La manera tradicional de describir este patrón o relación lineal es por medio de una función, tal como $d = a + b\delta$, $d = b\delta$, o de manera más general $d = f(\delta)$ y se pueden realizar procedimientos numéricos para encontrar dicha relación y cada tipo de relación corresponde a diferentes tipos de modelos de escalamiento multidimensional. De esta manera la función objetivo, función de error o función de bondad de ajuste se define de la siguiente manera:
Para cualquier conjunto de datos y para cualquier configuración, la función objetivo es un número que muestra que tan bien (o que tan pobre) los datos se ajustan a la configuración.

Una función objetivo natural sería $f(\delta_{ij}) - d_{ij}$. Dado que las diferencias positivas y negativas son igualmente indeseables, se toma la suma de los cuadrados de estas diferencias y se divide por el factor de escala de manera que sean relativas a su medición. Esta función es llamada "función de Stress"

$$\sqrt{\frac{\sum_{i,j}(d_{ij} - f(\delta_{ij}))^2}{\sum_{i,j} d_{ij}^2}}$$

Existen muchos procedimientos para encontrar dicha función que minimice el stress, como regresión lineal de mínimos cuadrados y si tenemos el caso en que $f(\delta)=\delta$ con f una función monótona, entonces se utilizarán procedimientos de regresión isotónica o regresión monótona de mínimos cuadrados. Existen muchos otros métodos de expresar la función de Stress y métodos para encontrar la función que minimice el stress como redes neuronales, método de Newton, método de Newton-Raphson, búsqueda de Tabú, algoritmos genéticos, mayorización iterativa, simulación anidada, etc. Algunos de estos métodos se tratarán en el capítulo 3.

CAPÍTULO 2

ESCALAMIENTO MULTIDIMENSIONAL MÉTRICO

CAPÍTULO 2

ESCALAMIENTO MULTIDIMENSIONAL MÉTRICO

Suponiendo que se tienen n objetos con disimilaridades $\{\delta_{ij}\}$. El escalamiento multidimensional métrico busca un conjunto de puntos en un espacio donde cada punto represente uno de los objetos y las distancias entre esos puntos $\{d_{ij}\}$ sean tales que:

$$d_{ij} \approx f(\delta_{ij})$$

donde f es una función paramétrica continua y monótona. La función f puede ser la función identidad o una función que transforme las disimilaridades en forma de distancia.

Sea $\{a_1,...,a_n\}$ el conjunto de n objetos en un espacio $\mathbf{O}$ con $\binom{n}{2}$ disimilaridades entre pares de objetos definidas en el espacio $\mathbf{O}\mathbf{x}\mathbf{O}$. Sea $\phi : \mathbf{O} \rightarrow \mathbf{E}$ una transformación que mapea a $\mathbf{O}$ en $\mathbf{E}$, donde $\mathbf{E}$ es generalmente un espacio euclideano, en el cual los puntos representan a los objetos.

De esta manera, sea $\phi(a_i) = x_i$ ($a_i \in \mathbf{O}$, $x_i \in \mathbf{E}$), y sea $\mathbf{X} = \{ x_i / a_i \in \mathbf{O} \}$, el conjunto imagen. Sea d_{ij} la distancia entre dos puntos x_i y x_j, entonces, el objetivo es encontrar la transformación ϕ, para el cual las distancias d_{ij} sean aproximadamente iguales a $f(\delta_{ij})$, para todo $a_i, a_j \in \mathbf{O}$.

Los métodos más importantes de escalamiento multidimensional métrico son el escalamiento clásico y el escalamiento por mínimos cuadrados. En este capítulo se describen los dos métodos.

2.1 ESCALAMIENTO CLÁSICO[1]

El escalamiento clásico fue creado en 1938 por Young y Householder, donde mostraron que partiendo de una matriz de distancias entre puntos en un espacio euclideano, se puede encontrar las coordenadas de los puntos de manera que las distancias se preserven. Torgerson y Gower le dieron popularidad a esta técnica llamándole escalamiento o escalamiento de Torgerson o de Torgerson-Gower.

El procedimiento consiste en encontrar una matriz producto interno a partir de una matriz de distancias conocida y a partir de la matriz producto interno, se encuentran las coordenadas desconocidas de los puntos. Este método se ha hecho popular debido a que proporciona una solución analítica sin necesidad de iteraciones.

Las coordenadas de los n puntos en un espacio euclideano de p dimensiones son de la forma $\mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{ip})^T$ y la distancia euclideana cuadrada entre el i-ésimo y el j-ésimo punto esta dada por:

$$d_{ij}^2 = (\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{x}_i - \mathbf{x}_j)$$

Sea $\mathbf{B}$ la matriz producto interno $\mathbf{B} = \mathbf{X}\mathbf{X}^T$, donde

$$[\mathbf{B}]_{ij} = b_{ij} = \mathbf{x}_i^T \mathbf{x}_j \text{ y rango}(\mathbf{B}) = p$$

centrando la configuración de puntos al origen se tendría:

$$\sum_{i=1}^{n} x_{ij} = 0 \quad (j = 1, ..., p)^2$$

Para encontrar la matriz producto interno a partir de la matriz de distancias cuadradas se tiene:

$$d_{ij}^2 = \mathbf{x}_i^T \mathbf{x}_i + \mathbf{x}_j^T \mathbf{x}_j - 2\mathbf{x}_i^T \mathbf{x}_j,$$

Centrando doblemente la expresión anterior se tiene:

$$\frac{1}{n}\sum_{j=1}^{n}\frac{1}{n}\sum_{i=1}^{n} d_{ij}^2 = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{x}_i^T \mathbf{x}_i + \mathbf{x}_j^T \mathbf{x}_j - 2\mathbf{x}_i^T \mathbf{x}_j\right) = \frac{1}{n}\sum_{j=1}^{n}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i^T \mathbf{x}_i + \mathbf{x}_j^T \mathbf{x}_j\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i^T \mathbf{x}_i + \frac{1}{n}\sum_{j=1}^{n}\mathbf{x}_j^T \mathbf{x}_j = \frac{2}{n}\sum_{i=1}^{n}\mathbf{x}_i^T \mathbf{x}_i$$

[1] fuente: Mardia et al., 1979
[2] Véase Apéndice A, apartado A.1

Partiendo de la función de distancia se obtiene la matriz producto interno $\mathbf{B}$ de la siguiente forma:

$$b_{ij} = \mathbf{x}_i^T \mathbf{x}_j = -\frac{1}{2}\left(d_{ij}^2 - \frac{1}{n}\sum_{i=1}^{n} d_{ij}^2 - \frac{1}{n}\sum_{j=1}^{n} d_{ij}^2 + \frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{n} d_{ij}^2 \right) = a_{ij} - a_{i.} - a_{.j} + a_{..},$$

donde $a_{ij} = -\frac{1}{2} d_{ij}^2$, $a_{i.} = \frac{1}{n}\sum_{j=1}^{n} a_{ij}$, $a_{.j} = \frac{1}{n}\sum_{i=1}^{n} a_{ij}$, $a_{..} = \frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{n} a_{ij}$.

Definiendo a la matriz $\mathbf{A}$ como $[\mathbf{A}]_{ij} = a_{ij}$, se tiene que la matriz producto interno $\mathbf{B}$ es

$$\mathbf{B} = \mathbf{H}\,\mathbf{A}\,\mathbf{H}$$

Donde $\mathbf{H}$ es la matriz que centra, $\mathbf{H} = \mathbf{I} - n^{-1}\mathbf{1}\mathbf{1}^T$, con 1= (1,1,...,1)T, un vector de n unos.

La matriz producto interno $\mathbf{B}$ es una matriz simétrica, positiva semidefinida y de rango p, por lo que tiene p eigenvalores no negativos y n-p eigenvalores cero[1].

La matriz $\mathbf{B}$ se puede expresar en términos de su descomposición espectral:

$$\mathbf{B} = \mathbf{V}\,\Lambda\,\mathbf{V}^T,$$

donde Λ =diag($\lambda_1,...,\lambda_n$), la matriz diagonal de los eigenvalores de $\mathbf{B}$, y $\mathbf{V} = [\mathbf{v}_1,...,\mathbf{v}_n]$, la matriz de los eigenvectores correspondientes, normalizados tal que $\mathbf{v}_i\mathbf{v}_i^T$=1. Por conveniencia los eigenvalores de $\mathbf{B}$ se ordenan de la siguiente manera $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq ... \geq \lambda_n \geq 0$. Como se tienen n – p eigenvalores cero, $\mathbf{B}$ se puede re expresar de la siguiente manera:

$$\mathbf{B} = \mathbf{V}_1\,\Lambda_1\,\mathbf{V}_1^T,$$

donde Λ_1=diag($\lambda_1,...,\lambda_p$), $\mathbf{V}_1 = [\mathbf{v}_1,...,\mathbf{v}_p]$ y la matriz de coordenadas X está dada por

$$\mathbf{X} = \mathbf{V}_1\,\Lambda_1^{1/2}.$$

En la práctica, se necesita encontrar una configuración de puntos a partir de un conjunto de disimilaridades $\{\delta_{ij}\}$ y no de distancias euclideanas entre puntos $\{d_{ij}\}$. Suponiendo que empleamos disimilaridades en lugar de distancias para encontrar la matriz $\mathbf{A}$, que se centra doblemente para encontrar la matriz $\mathbf{B}$ anteriormente descrita. Si las disimilaridades son tales que $d_{ij} = \delta_{ij}$, entonces $\mathbf{B}$ genera una configuración de puntos en un espacio euclideano si $\mathbf{B}$ es una matriz positiva semidefinida de rango p[2].

[1]Consultar Apéndice A, apartado A.5

La matriz **B** tiene al menos un eigenvalor cero, ya que **B 1 = H A H 1= 0.** Por tanto, siempre se puede encontrar una configuración de puntos en un espacio euclideano de n-1 dimensiones con distancias asociadas iguales a las disimilaridades.

Si las disimilaridades dan lugar a una matriz **B** que no sea positiva semidefinida, se puede agregar una constante a las disimilaridades[1] (excepto a las disimilaridades entre los mismos objetos $\{\delta_{ij}\}$) de manera que la matriz **B** se vuelva positiva semidefinida. Así tomando a las distancias de la forma:

$$d_{ij} = \delta_{ij} + c\,(1 - \delta_{ij}^{KR})$$

donde c es una constante apropiada y δ_{ij}^{KR} es la delta de kronecker.

La matriz **B** es de rango n-1, por lo que se necesitarían n-1 dimensiones para la configuración, sin embargo, esto no reduce significativamente las dimensiones que para nuestro estudio necesitaríamos.

La configuración obtenida puede rotarse en sus ejes principales en el sentido de componentes principales, es decir, las proyecciones de los puntos de la configuración en su primer eje principal tienen la máxima variación posible, las proyecciones de los puntos en su segundo eje principal tienen la máxima variación posible pero debe ser ortogonal al primer eje principal, y así sucesivamente. De esta manera, solo los primeros p ejes (p < n-1) se eligen para representar la configuración. Sin embargo, no se tienen que encontrar esos p ejes principales (componentes principales) en el procedimiento ya que **X** contiene los puntos referidos a sus ejes principales.

De esta manera, escalamiento clásico, análisis de componentes principales y escalamiento métrico son sinónimos, la diferencia es que el análisis por componentes principales emplea una matriz multivariada de tamaño n x m, mientras que el escalamiento métrico emplea una matriz de n x n disimilaridades. Cuando la matriz de disimilaridades se construye a partir de una matriz multivariada, definiendo las disimilaridades como distancias euclideanas entre vectores de atributos, ambos producen la misma configuración de puntos.

En la descomposición espectral de la matriz **B** , las distancias entre los puntos en un espacio euclideano de n-1 dimensiones está dada por:

$$d_{ij}^2 = \sum_{k=1}^{n-1} \lambda_k \left(x_{ik} - x_{jk}\right)^2 ,$$

y por tanto, si muchos eigenvalores son "pequeños", entonces su contribución a las distancias cuadradas son mínimas. Si solo nos quedamos con los p eigenvalores significativamente más grandes, entonces se puede encontrar un espacio euclideano de p dimensiones formado por los primeros p eigenvalores y x_i truncado con sus primeros p elementos para representar así a los n objetos. Se espera que p fuera pequeño, de preferencia 2 o 3 para su representación gráfica[2].

[1]Problema de constante aditiva. Caillez, 1983
[2]Fuente: Cox & Cox., 1994

Los eigenvalores indican el número de dimensiones requeridas para representar las disimilaridades. Si $\mathbf{B}$ es positiva semidefinida, entonces el número de eigenvalores diferentes de cero, sería el número de dimensiones requeridas. Si $\mathbf{B}$ no es positiva semidefinida, entonces el número de eigenvalores positivo sería un número apropiado para el número de dimensiones. Sin embargo, como se describió anteriormente, se busca un número pequeño de dimensiones. Eligiendo los primeros p eigenvalores y eigenvectores (p= 2 o 3) proporciona un espacio de pocas dimensiones.

La suma de las distancias cuadradas entre puntos en el espacio completo es:

$$\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}^2 = n\sum_{i=1}^{n} x_i^T x_i = n\,tr\mathbf{B} = n\sum_{i=1}^{n-1}\lambda_i .$$

Una medida de la proporción de la variación explicada empleando solo p dimensiones es:

$$\frac{\sum_{i=1}^{p}\lambda_i}{\sum_{i=1}^{n-1}\lambda_i}$$

Cuando se realiza la descomposición en el escalamiento métrico se pueden obtener eigenvalores negativos pero no si la matriz es una matriz de distancias euclideanas. En el escalamiento clásico, los eigenvalores negativos son simplemente ignorados, de esta manera, si $\mathbf{B}$ no es positiva semidefinida, la medida se modifica a:

$$\frac{\sum_{i=1}^{p}\lambda_i}{\sum \text{eigenvalores positivos}}$$

2.2 ESCALAMIENTO POR MÍNIMOS CUADRADOS[1]

Hasta mediados de los 70, el escalamiento por mínimos cuadrados no jugaba un papel importante en la teoría del escalamiento multidimensional. El escalamiento por mínimos cuadrados permite una transformación continua monótona a las disimilaridades $f(\delta_{ij})$ antes de que se encuentre una configuración empleando mínimos cuadrados.
La configuración de puntos es tal que la siguiente función se minimice.

$$S = \frac{\sum_{i\neq j} w_{ij}(d_{ij} - f(\delta_{ij}))^2}{\sum_{i\neq j} d_{ij}^2},$$

[1]Fuente: Cox & Cox, 1994. Capítulo 2
Borg & Groenen, 1997. Capítulo 12
Young & Hammer, 1987. Capítulo 3

donde w_{ij} son los pesos elegidos apropiadamente. Las distancias $\{d_{ij}\}$ no necesariamente son euclideanas y la minimización de S puede hacerse numéricamente o por el método del gradiente. La elección de la función $f(\delta_{ij})$ depende del problema particular y las formas más comunes son las siguientes:

Escalamiento absoluto (clásico):

$$f(\delta_{ij}) = \hat{d}_{ij}$$

En este caso hablar de minimizar S no tiene sentido, ya que las disimilaridades son exactamente las distancias euclideanas.

Escalamiento de razón:

$$f(\delta_{ij}) = b\delta_{ij}$$

Minimizar la función S consiste en minimizar $\sum_{i\neq j} w_{ij}(d_{ij} - f(\delta_{ij}))^2$ conocida como Stress.

$$\sum_{i\neq j} w_{ij}(d_{ij} - f(\delta_{ij}))^2 = \sum_{i<j} w_{ij} d_{ij}^2 + b^2 \sum_{i<j} w_{ij}\delta_{ij}^2 - 2b\sum_{i<j} w_{ij} d_{ij}^2\delta_{ij}^2$$

$$= \eta^2 + b^2\eta_\delta^2 - 2b\rho$$

Para minimizar la función con respecto a b, se deriva la función de stress y se iguala a cero.

$$\frac{\partial Stress}{\partial b} = 2b\eta_\delta^2 - 2\rho = 0,$$

de donde $b = \frac{\rho}{\eta_\delta^2}$.

Escalamiento de intervalo:

$$f(\delta_{ij}) = a + b\delta_{ij} f$$
$$f(\delta_{ij}) = a + b\log\delta_{ij}$$
$$f(\delta_{ij}) = a + b\exp(\delta_{ij})$$
$$f(\delta_{ij}) = a + b\delta_{ij} + c\delta_{ij}^2$$

De la misma manera que se minimizó la función de stress para el escalamiento de razón, se minimizarían las funciones de stress para este tipo de escalamientos de intervalo, es por eso que se le llama escalamiento de mínimos cuadrados.

CAPÍTULO 3

ESCALAMIENTO MULTIDIMENSIONAL NO MÉTRICO

CAPITULO 3

ESCALAMIENTO MULTIDIMENSIONAL NO MÉTRICO

Al igual que con el escalamiento multidimensional métrico se supone que se tienen n objetos con disimilaridades $\{\delta_{ij}\}$ y el procedimiento consiste en encontrar una configuración de n puntos en un espacio, generalmente euclideano, donde cada punto represente a cada objeto de manera que las distancias entre pares de puntos reflejen lo mejor posible a las disimilaridades originales.

Sea $\{a_1,...,a_n\}$el conjunto de n objetos en un espacio $\mathbf{O}$ con disimilaridades definidas en $\mathbf{OxO}$. Sea $\phi:\mathbf{O}\rightarrow\mathbf{X}$, donde $\mathbf{X}$ es un subconjunto del espacio empleado para representar los objetos. Sea $d_{x_i x_j}$, la distancia entre los puntos x_i, x_j en $\mathbf{X}$. Entonces la disparidad $\hat{d}$ definida en $\mathbf{OxO}$, es una medida de que tan bien la distancia $d_{\phi(i)\phi(j)}$ se ajusta a la disimilaridad δ_{ij}. El objetivo es encontrar la función ϕ de manera que la distancia $d_{\phi(i)\phi(j)}$ sea aproximadamente igual a $\hat{d}_{ij}$ y generalmente se encuentra por medio de una función de error (Stress). El conjunto X regularmente es un subconjunto de $\mathbf{R}^p$

Sea $x_i=(x_{i1},x_{i2},...,x_{ip})^T$ las coordenadas del i-ésimo punto en $\mathbf{X}$. Sea d_{ij} la métrica de Minkowski que mide la distancia entre el punto i-ésimo y el j-ésimo en $\mathbf{X}$, donde:

$$d_{ij}=\left[\sum_{k=1}^{p}\left|x_{ik}-x_{jk}\right|^{\lambda}\right]^{1/\lambda},\lambda\geq 1, x_i\neq x_j$$

Espacio euclideano: $\lambda=2$,
Espacio city-block: $\lambda=1$,
Espacio de dominancia: $\lambda=\infty$.

Se definen las disparidades $\{\hat{d}_{ij}\}$ como función de las disimilaridades $\{\delta_{ij}\}$ por:

$$\hat{d}_{ij}=f(\delta_{ij}),$$

donde f es una función monótona tal que: $\hat{d}_{ij}\leq\hat{d}_{kl}$ siempre que $\delta_{ij}<\delta_{kl}$ ....(Condición C1)

Los modelos de escalamiento multidimensional requieren que cada disimilaridad o disparidad corresponda a su distancia en la configuración final. Esto proporciona la noción de error. Sin embargo, las proximidades empíricas siempre contienen cierto ruido debido a la imprecisión de su medición, efectos muestrales, etc. Por lo tanto, es casi imposible que las disparidades correspondan exactamente a sus distancias.

Una vez elegidos el espacio y el método de cálculo de disparidades, el problema en el escalamiento multidimensional no métrico se centra en encontrar un algoritmo apropiado para minimizar la función de error.

Una posible representación de error es la siguiente:

$$S_R = \sum_{i,j} (f(\delta_{ij}) - d_{ij})^2$$

Esta función de error, es conocida como stress renglón (row stress), pero no es muy informativa ya que valores grandes no necesariamente implican un mal ajuste, ya que es muy sensible a la escala de medición.

Shepard fue el primero en producir un algoritmo para el escalamiento multidimensional no métrico sin emplear una función de error y Kruskal mejoró las ideas de Shepard incluyendo la función de error (1) llamada Stress1 que es minimizada con respecto a $\{d_{ij}\}$, es decir, con respecto a $\{x_{ik}\}$ y con respecto a $\{\hat{d}_{ij}\}$ empleando regresión isotónica[1].

$$S = \left\{ \frac{\sum_{i,j} (d_{ij} - \hat{d}_{ij})^2}{\sum_{i,j} d_{ij}^2} \right\}^{1/2} \quad \text{....(1)}$$

La función de error puede expresarse en términos de las coordenadas $\{x_{ik}\}$ reemplazando las distancias $\{d_{ij}\}$, y la función de error se pude derivar parcialmente con respecto a $\{x_{ik}\}$ en busca del mínimo. Las disparidades generalmente son funciones complicadas no diferenciables de distancias que son funciones de $\{x_{ik}\}$, lo que significa que la función de error no puede diferenciarse completamente con respecto a las coordenadas $\{x_{ik}\}$.

En este capítulo se describen algunos métodos para minimizar la función de error y las ventajas y desventajas de cada uno de ellos.

[1]Véase Pág 26, subtema 3.1.1 Regresión Isotónica

3.1. MÉTODO DE KRUSKAL

Como se mencionó en el apartado anterior, Kruskal mejoró el algoritmo de Shepard para el escalamiento multidimensional no métrico empleando regresión isotónica. En el siguiente apartado se describe el método de regresión isotónica para minimizar la función de Stress1.

3.1.1 REGRESIÓN ISOTÓNICA

La regresión isotónica busca el minorante convexo mayor de las sumas acumuladas de las distancias y de esta manera se encuentran disparidades que minimizan la función de Stress1. Para ilustrar el método de regresión isotónica, considérese el siguiente ejemplo:

Suponiendo que se tiene cuatro objetos con $\frac{1}{2}n(n-1) = 6$ disimilaridades:

$$\delta_{12} = 2.1 \quad \delta_{13} = 3.0 \quad \delta_{14} = 2.4 \quad \delta_{23} = 1.7 \quad \delta_{24} = 3.9 \quad \delta_{34} = 3.2$$

y la configuración de puntos representando los cuatro objetos con distancias:

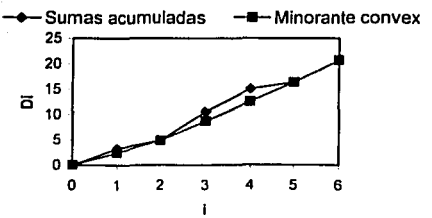
$$d_{12} = 1.6 \quad d_{13} = 4.5 \quad d_{14} = 5.7 \quad d_{23} = 3.3 \quad d_{24} = 4.3 \quad d_{34} = 1.3$$

Ordenando a las disimilaridades con sus distancias asociadas se tiene:

$$\begin{aligned} \delta_1 &= 1.7 & d_1 &= 3.3 \\ \delta_2 &= 2.1 & d_2 &= 1.6 \\ \delta_3 &= 2.4 & d_3 &= 5.7 \\ \delta_4 &= 3.0 & d_4 &= 4.5 \\ \delta_5 &= 3.2 & d_5 &= 1.3 \\ \delta_6 &= 3.9 & d_6 &= 4.3 \end{aligned}$$

Sea $D_i = \sum_{j=1}^{i} d_j$ (i = 1,...,N) la suma acumulada de $\{d_i\}$. En la siguiente figura se grafica la suma acumulada de las distancias contra i.

figura 3. Regresión Isotónica



El minorante convexo mayor de las sumas acumuladas es la gráfica del supremo de todas las funciones cuyas gráficas se encuentran por debajo de la gráfica de las sumas acumuladas.

Las disparidades $\{\hat{d}_i\}$ que minimizan S (ecuación (1)) están dadas por el minorante convexo mayor , donde la disparidad i-ésima $\hat{d}_i$ es el valor del minorante en i. En la figura 2.1 algunos valores de $\hat{d}_i$ son iguales a d_i y S=0. $\hat{d}_i = \hat{D}_i - \hat{D}_{i-1}$, es decir, la pendiente de la línea que une el i-ésimo con el (i-1)-ésimo punto. Si $\hat{D}_i < D_i$, entonces $\hat{d}_i = \hat{d}_{i-1}$.

Demostrando que $\{\hat{d}_i\}$ minimizan S, sea $\{d^*_i\}$ un conjunto arbitrario de números reales que satisfacen la condición C1. Se tiene que demostrar que:

$$\sum_{i=1}^{N}(d_i - d_i^*)^2 \geq \sum_{i=1}^{N}(d_i - \hat{d}_i)^2 \text{ .... (Desigualdad D1)}$$

Desarrollando la parte izquierda de la desigualdad anterior, se tiene:

$$\sum_{i=1}^{N}(d_i - d_i^*)^2 = \sum_{i=1}^{N}\{(d_i - \hat{d}_i) + (\hat{d}_i - d_i^*)\}^2$$

$$= \sum_{i=1}^{N}(d_i - \hat{d}_i)^2 + \sum_{i=1}^{N}(\hat{d}_i - d_i^*)^2 + 2\sum_{i=1}^{N}(d_i - \hat{d}_i)(\hat{d}_i - d_i^*).$$

Ahora, aplicando la fórmula de Abel: $\sum_{i=1}^{N}a_i b_i = \sum_{i=1}^{N-1}A_i(b_i - b_{i+1}) + A_N b_N$, donde $A_i = \sum_{j=1}^{i}a_j$ y Sea $D_i^* = \sum_{j=1}^{i}d_j^*$, $\hat{D}_i = \sum_{j=1}^{i}\hat{d}_j$. Ahora,

$$\sum_{i=1}^{N}(d_i - \hat{d}_i)(\hat{d}_i - d_i^*) = \sum_{i=1}^{N-1}(D_i - \hat{D}_i)(\hat{d}_i - \hat{d}_{i+1}) - \sum_{i=1}^{N-1}(D_i - \hat{D}_i)(d_i^* - d_{i+1}^*) + (D_N - \hat{D}_N)(\hat{d}_N - d_N^*).$$

Dado que el último punto y el último punto de minorante convexo mayor coinciden, $D_N - \hat{D}_N = 0$. Considerando $(D_i - \hat{D}_i)(\hat{d}_i - \hat{d}_{i+1})$. Si el i-ésimo punto de la gráfica coincide con el i-ésimo minorante convexo mayor, entonces, la expresión anterior se hace cero y si $\hat{D}_i < D_i$, entonces $\hat{d}_i = \hat{d}_{i+1}$ y la expresión anterior nuevamente se hace cero. Dado que $D_i - \hat{D}_i \geq 0$ y por la condición C1, $d_i^* < d_{i+1}^*$, el término $-\sum_{i=1}^{N-1}(D_i - \hat{D}_i)(d_i^* - d_{i+1}^*)$ se hace positivo y por lo tanto la desigualdad D1 se cumple. En el ejemplo anterior se tiene una S=0.14[1].

[1]Fuente: Cox & Cox , 1994. p. 45-52.

3.1.2 MÉTODO DEL DESCENSO MÁS RÁPIDO

La minimización del Stress (S) no es un problema fácil. El primer paso consiste en poner todas las coordenadas de los puntos de X en un vector $\mathbf{x} = (x_{11},...,x_{1p},...x_{np})$ con np elementos. La función de Stress es vista como una función de **x** y es minimizada con respecto a **x** de una manera iterativa, empleando el método de **"descenso más rápido"**[1], de manera que si $\mathbf{x}_m$ es el vector de coordenadas después de la m-ésima iteración

$$\mathbf{x}_{m+1} = \mathbf{x}_m - \alpha \frac{\frac{\partial S}{\partial x}}{\left|\frac{\partial S}{\partial x}\right|},$$

donde el valor de α se discutirá más adelante.

Ahora:

$$\frac{\partial S^2}{\partial x_{he}} = \frac{\partial}{\partial x_{he}}\left(\frac{\sum_i\sum_j (d_{ij} - \hat{d}_{ij})^2}{\sum_i\sum_j d_{ij}^2}\right),$$

Dado que $B = \sum_i\sum_j d_{ij}^2$ es constante , se tiene que

$$\frac{\partial S^2}{\partial x_{he}} = \frac{1}{B}\frac{\partial}{\partial x_{he}}\left[\sum_i\sum_j (d_{ij} - \hat{d}_{ij})^2\right] = \frac{1}{B}\sum_i\sum_j \frac{\partial}{\partial x_{he}}(d_{ij} - \hat{d}_{ij})^2$$
$$= \frac{2}{B}\sum_i\sum_j (d_{ij} - \hat{d}_{ij})\frac{\partial}{\partial x_{he}}(d_{ij} - \hat{d}_{ij}) = \frac{2}{B}\sum_i (d_{ih} - \hat{d}_{ih})\frac{\partial}{\partial x_{he}} d_{ih}$$

Kruskal definió a la distancia empleando la métrica de Minkowski y la derivada de la distancia relativa a la coordenada x_{he} de la siguiente manera:

$$\frac{\partial d_{ih}}{\partial x_{he}} = \partial\left[\sum_{a=1}^{p}|x_{ia} - x_{ha}|^{\lambda}\right]^{1/\lambda} = \frac{1}{\lambda}\left[\sum_a |x_{ia} - x_{ha}|^{\lambda}\right]^{1/\lambda - 1}\frac{\partial}{\partial x_{he}}\left[\sum_a |x_{ia} - x_{ha}|^{\lambda}\right]$$
$$= \frac{1}{\lambda} d_{ih}^{(1-\lambda)}\sum_a \frac{\partial}{\partial x_{he}}|x_{ia} - x_{ha}|^{\lambda} = d_{ih}^{(1-\lambda)}\sum_a |x_{ia} - x_{ha}|^{\lambda - 1}\frac{\partial}{\partial x_{he}}[x_{ia} - x_{ha}]$$

[1]El método del descenso más rápido es mayormente conocido en su traducción al inglés "Steepest descent".

Las derivadas de x_{ha} son cero, excepto cuando $a = e$, por lo tanto, la ecuación anterior es igual a:

$$d_{ih}^{1-\lambda}\left|x_{ie}-x_{he}\right|^{\lambda-1}\operatorname{sgn}\left[x_{ie}-x_{he}\right]\frac{\partial}{\partial x_{he}}\left[x_{ie}-x_{he}\right]=-\frac{1}{d_{ih}^{\lambda-1}}\left|x_{ie}-x_{he}\right|^{\lambda-1}\operatorname{sgn}\left[x_{ie}-x_{he}\right]$$

Ahora, sustituyendo la derivada de la distancia en la derivada de la función de Stress al cuadrado (S^2), se tiene:

$$\frac{\partial S^2}{\partial x_{he}}=-\frac{2}{B}\left[\sum_i\frac{(d_{ih}-\hat{d}_{ih})}{d_{ih}^{\lambda-1}}\left|x_{ie}-x_{he}\right|^{\lambda-1}\operatorname{sgn}\left[x_{ie}-x_{he}\right]\right].$$

Se necesita empezar con una configuración dada x_0 elegida anteriormente. Esta configuración inicial puede ser una configuración arbitraria. Los puntos pueden localizarse en los vértices de un sistema de p dimensiones o puede generarse como realización de procesos Poisson en una región de $\mathbf{R}^p$, es decir, generar las n coordenadas independientemente a partir de una distribución uniforme [-1,1] y la configuración se normaliza teniendo el centroide en el origen y con distancia media cuadrada de los puntos al origen la unidad, otra alternativa es generar una configuración inicial empleando escalamiento multidimensional métrico.

Existen otros métodos sugeridos para elegir una configuración inicial. Guttman y Lingoes[1968] y Roskman[1973] sugirieron el siguiente procedimiento para encontrar una configuración inicial. Sea **C** una matriz definida como:

$$[\mathbf{C}]_{ij}=c_{ij}$$

$$\text{donde, } c_{ij}=\begin{cases}1+\sum_k\frac{\rho_{ik}}{N} & (\mathrm{i}=\mathrm{j})\\ 1-\frac{\rho_{ij}}{N} & (\mathrm{i}\neq\mathrm{j})\end{cases}$$

donde N es el total de disimilaridades y ρ_{ij} es el rango de la disimilaridad δ_{ij} en orden numérico de las dimilaridades $\{\delta_{ij}\}$. Los componentes principales de **C** se encuentran y la configuración inicial está dada por los eigenvectores de los primeros p componentes principales, ignorando el de eigenvector constante.

La técnica iterativa de Kruskal empleada para encontrar una configuración con mínimo Stress se resume en los siguientes puntos[1]:

I. Elegir una configuración inicial.
II. Normalizar la configuración de manera que tenga su centroide en el origen y distancia cuadrada unitaria a partir del origen.
III. Encontrar $\{d_{ij}\}$ a partir de la configuración normalizada.
IV. Ajustar las disparidades $\{\hat{d}_{ij}\}$, que consiste en encontrar el minorante convexo mayor por medio de regresión isotónica antes mencionada.
V. Encontrar el gradiente $\frac{\partial S}{\partial x}$. Si $\left|\frac{\partial S}{\partial x}\right| < \xi$, donde ξ es un número pequeño, elegido anteriormente, entonces, se ha encontrado una configuración con mínimo stress. Esta configuración puede darnos un mínimo local para el Stress y no un mínimo global.
VI. Encontrar la nueva α

$$\alpha_{actual} = (\alpha_{anterior})\beta_1\beta_2\beta_3$$

donde,

$\beta_1 = 4.0^{\cos^3\theta}$, θ = ángulo entre el gradiente actual y el gradiente anterior.

$$\beta_2 = \frac{1.3}{1+(5h)^5}, \quad h = min\left[1, \left(\frac{\text{stress actual}}{\text{stress de 5 iteraciones previas}}\right)\right]$$

$$\beta_3 = min\left[1, \frac{\text{stress actual}}{\text{stress anterior}}\right]$$

VII. Encontrar la nueva configuración

$$\mathbf{x}_{m+1} = \mathbf{x}_m - \alpha \frac{\frac{\partial S}{\partial x}}{\left|\frac{\partial S}{\partial x}\right|},$$

VIII. Regresar al paso II.

Fuente: Cox & Cox, 1994. p. 45-52.
Burden, 1996. p. 614-620.

3.1.3 MÉTODO DE GUTTMAN[1]

Guttman definió una función de error llamada Coeficiente de alineación básicamente equivalente a la función de Stress de Kruskal, pero con un algoritmo diferente de minimización.

Sea δ el vector con elementos δ_i $(i = 1,\ldots, N)$, las disimilaridades ordenadas y $\mathbf{d}$ el vector de distancias asociadas a las disimilaridades ordenadas. Sea $\mathbf{E}$ una matriz de NxN que permuta los elementos de $\mathbf{d}$ en orden ascendente. Las disparidades $\mathbf{d}^*$ están dadas por $\mathbf{d}^* = \mathbf{E}\,\mathbf{d}$.

El coeficiente de continuidad, μ, para la configuración está dado por:

$$\mu = \sqrt{\frac{(\sum d_i d_i^*)^2}{\sum d_i^2 \sum d_i^{*2}}}$$

que valdría 1 si se tuviera un ajuste perfecto. Con el objetivo de encontrar la mejor configuración, el coeficiente de alineación K dado por

$$K = \sqrt{1-\mu^2}$$

es minimizado empleando el método de **descenso más rápido.**

Guttman sugirió que un coeficiente de alineación menor a 0.15 implica una solución con precisión aceptable.

Para cada dimensión p, la configuración que tenga el menor stress es llamada la mejor configuración ajustada en p dimensiones. La elección de la correcta dimensión busca el menor número de dimensiones para la configuración con menor stress posible. Generalmente se busca una dimensión p=2 y una regla de dedo según Kruskal de tolerancia del Stress es la siguiente:

Stress ≥ 20%, ajuste pobre
Stress=10% ajuste justo
Stress ≤ 5% buen ajuste
Stress=0 ajuste perfecto.

Muchos autores han estudiado la función de stress con mayor detalle y en los últimos años se han aplicado diversos métodos numéricos para minimizar tal función, en este capítulo se describen únicamente los más utilizados.

[1]Fuente: Cox & Cox, 1994. Capítulo 3.

3.2. OPTIMIZACIÓN EN EL ESCALAMIENTO MULTIDIMENSIONAL NO MÉTRICO

Como se mencionó anteriormente, el escalamiento multidimensional no métrico se centra en emplear un algoritmo apropiado para minimizar la función de error o Stress. La búsqueda del mejor algoritmo que minimice la función de Stress se ha convertido en un problema de optimización continuo. Los algoritmos determinísticos como los mencionados anteriormente, son los más empleados.

3.2.1 MÉTODO DE SPENCE Y LEWADOWSKY (TUFSCAL)[1]

Spence y Lewadowsky (1989) consideraron el efecto de atipicidades en el escalamiento multidimensional sugiriendo un método de estimación paramétrico robusto y un índice de ajuste robusto.

Supóngase una configuración de n puntos en un espacio euclideano de p dimensiones con distancias asociadas $\{d_{ij}\}$ representado a las disimilaridades $\{\delta_{ij}\}$. Concentrándonos en la coordenada x_{ij} se tienen n-1 distancias y n-1 discrepancias $f(x_{ij})$ entre la disimilaridad y la distancia,

$$f_s(x_{ij}) = \delta_{is} - \{\sum_{k=1}^{p}(x_{ik} - x_{sk})^2\}^{1/2} \qquad i \neq j, \quad s = 1,\ldots,n$$

$$f_s(x_{i1}) \equiv f_s(x_{i2}) \equiv \ldots \equiv f_s(x_{in}).$$

Sean $\{x_{ij}^t\}$ las coordenadas en la t-ésima iteración en la búsqueda de la configuración óptima, y sean $\{d_{ij}^t\}$, las distancias asociadas. Aplicando el método de Newton-Raphson para encontrar raíces de ecuaciones, se tiene:

$$x_{ij}^{t+1} = x_{ij}^t - \frac{f_s(x_{ij}^t)}{f_s'(x_{ij}^t)}$$

$$= x_{ij}^t + \frac{(\delta_{is} - d_{is}^t)d_{is}^t}{x_{ij}^t - x_{sj}^t} = x_{ij}^t + {}_s g_{ij}^t$$

Las correcciones ${}_s g_{ij}^t$ pueden estar altamente influenciadas por atipicidades y Spence y Lewandowsky sugirieron emplear la mediana.

Así,

$$x_{ij}^{t+1} = x_{ij}^t + {}_M g_{ij}^t, \text{ donde } {}_M g_{ij}^t = \underset{i \neq s}{mediana}({}_s g_{ij}^t)$$

[1] El término TUFSCAL son las siglas del inglés Tunnelling Function SCALing

También, sugirieron la siguiente modificación al método de Newton-Raphson:

$$x_{ij}^{t+1} = x_{ij}^{t} + \beta^{t}{}_{M}\, g_{ij}^{t}$$

donde, $\beta^{t} = \dfrac{\alpha^{t}}{g^{t}}$, $\alpha^{t+1} = \alpha^{t}\left\{\dfrac{\sum_{i,j}(x_{ij}^{t-1} - x_{ij}^{t-2})^2}{\sum_{i,j}(x_{ij}^{t} - 2x_{ij}^{t-1} + x_{ij}^{t-2})^2}\right\}^{1/2}$, $g^{t} = \left\{\dfrac{\sum_{i,j}({}_{M}g_{ij}^{t})^2}{\sum_{i,j}(x_{ij}^{t})^2}\right\}^{1/2}$

En este algoritmo, el índice de ajuste o medida de error es el índice llamado TUF, donde

$$\text{TUF} = \underset{i}{mediana}\ \underset{s\neq i}{mediana}\left|\frac{\delta_{is} - d_{is}}{\delta_{is}}\right|,$$

Su interpretación es el porcentaje mediano de discrepancia entre las disimilaridades y las distancias ajustadas.

Spence y Lewandowsky realizaron numerosas simulaciones concluyendo que los métodos no métricos son más resistentes a las atipicidades que los métricos y que su método era el más resistente.

3.2.2 MAYORIZACIÓN ITERATIVA (SMACOF)[1]

El método SMACOF se basa en el algoritmo iterativo, llamado mayorización iterativa propuesto inicialmente por De Leeuw (1977). El algoritmo de mayorización iterativa (iterative majorization) busca minimizar una función complicada $f(x)$, empleando una función auxiliar manejable $g(x,y)$. La función g es seleccionada para $x \in D_f$ y $f(x) \leq g(x,y)$ para un valor particular de $y \in D_g$, tal que $f(y) = g(y,y)$. De esta manera, la gráfica de la función g está siempre por encima de la función f y las funciones f y g se interceptan en el punto $x = y$. Tal función g es llamada "función mayorizante" de f

Uno de los hechos principales del algoritmo de mayorización iterativa es que genera una sucesión de valores de funciones monótonas no crecientes. Si la función es inferiormente acotada, generalmente se tiene un punto estacionario como mínimo local.

Sea x^* el mínimo de $g(x,y)$ sobre x, entonces:

$$f(x^*) \leq g(x^*, y) \leq g(y,y) = f(y) \quad \text{(Desigualdad del Sandwich de De Leeuw)}$$

[1]Se prefirió emplear el término en inglés, ya que no existe una traducción precisa del término en la bibliografía existente. El término SMACOF en inglés significa Scaling by Maximizing a Convex Function, traducido al español: Escalamiento maximizando una función convexa.

Este algoritmo puede emplearse tanto para los métodos métricos como los no métricos. Considerando la función de stress

$$S = \sum_{i<j} w_{ij}(\delta_{ij} - d_{ij})^2 ,$$

para el caso métrico y reemplazando la disimilaridad por una disparidad en el caso no métrico, se tiene:

$$S = \sum_{i<j} w_{ij}\delta_{ij}^2 + \sum_{i<j} w_{ij}d_{ij}^2 - 2\sum_{i<j} w_{ij}\delta_{ij}d_{ij}$$
$$= \eta_\delta^2 + \eta^2 - 2\rho$$

En su forma matricial, se tiene:

$$\eta^2 = \sum_{i<j} w_{ij}(x_r - x_s)^T(x_r - x_s) = \mathbf{tr}(\mathbf{X}^T\mathbf{V}\mathbf{X})$$

donde,

$$[\mathbf{V}]_{ij} = \begin{cases} \sum_{i \neq s} w_{is} \text{ si } i = j \\ -w_{ij} \text{ si } i \neq j \end{cases}$$

$$\rho = \sum_{i<j} \frac{w_{ij}\delta_{ij}}{d_{ij}} d_{ij}^2 = \sum_{i<j} \frac{w_{ij}\delta_{ij}}{d_{ij}} (x_i - x_j)^T(x_i - x_j) = \mathbf{tr}(\mathbf{X}^T\mathbf{B}\mathbf{X}).$$

donde,

$$[\mathbf{B}]_{ij} = \frac{w_{ij}\delta_{ij}}{d_{ij}} \text{ si } d_{ij} \neq 0$$
$$= 0 \quad \text{si } d_{ij} = 0$$

Por lo tanto, $S = \eta_\delta^2 + \mathbf{tr}(\mathbf{X}^T\mathbf{V}\mathbf{X}) - 2\mathbf{tr}(\mathbf{X}^T\mathbf{B}\mathbf{X})$

Una función $\mathbf{T}$ (función mayorizante), para la función de Stress está dada por:

$$\mathbf{T}(\mathbf{X}, \mathbf{Y}) = \eta_\delta^2 + \mathbf{tr}(\mathbf{X}^T\mathbf{V}\mathbf{X}) - 2\mathbf{tr}(\mathbf{X}^T\mathbf{B}(\mathbf{Y})\mathbf{Y})$$

Ahora, el problema se centra en minimizar $\mathbf{T}$,

$$\frac{\partial \mathbf{T}}{\partial \mathbf{Y}} = 2\mathbf{V}\mathbf{X} - 2\mathbf{B}(\mathbf{Y})\mathbf{Y} = \mathbf{0}.$$

V tiene rango $n-1$, ya que sus renglones suman cero y por el teorema de la inversa de Moore-Penrose[1], se tiene:

$$\mathbf{X} = \mathbf{V}^+\mathbf{B}(\mathbf{Y})\mathbf{Y} \text{ ....(Transformación de Guttman)}$$

[1]Consultar Apéndice A, apartado A.6.

Por tanto, al emplear el método de mayorización iterativa para encontrar el stress mínimos, simplemente se tiene que estar actualizando la transformación de Guttman hasta que el stress se minimice.

3.2.3 ESCALAMIENTO DE MÍNIMOS CUADRADOS ALTERNANTE (ALSCAL)[1]

El escalamiento de mínimos cuadrados alternante (ALSCAL), posiblemente es uno de los algoritmos más populares. Este algoritmo fue desarrollado por Takane, Young y De Leeuw (1977). El Algoritmo ALSCAL puede analizar los siguientes tipos de datos:

i) nominales, ordinales, de intervalo o de razón
ii) observaciones completas o con observaciones faltantes
iii) simétricos o asimétricos
iv) condicionales o incondicionales
v) replicados o no replicados
vi) continuos o discretos (caja de Pandora).

Supóngase que las disimilaridades son de la forma $\{\delta_{ij,s}\}$ que puede ser de los tipos antes mencionados (i)-(vi). Los subíndices ij representan a los objetos y el subíndice s a los individuos. El problema de escalamiento busca una correspondencia ϕ de las disimilaridades $\{\delta_{ij,s}\}$ a las disparidades $\{\hat{d}_{ij,s}\}$,

$$\phi[\delta_{ij,s}^2] = \hat{d}_{ij,s}^2$$

Donde $\{\hat{d}_{ij,s}^2\}$ son las estimaciones de mínimos cuadrados de $\{d_{ij,s}^2\}$ resultantes de minimizar la función de SStress, denotada por **SS**.

$$SS = \sum_i \sum_j \sum_s (d_{ij,s}^2 - \hat{d}_{ij,s}^2)^2$$

La función ϕ que mapea a las disimilaridades $\{\delta_{ij,s}\}$ en disparidades $\{\hat{d}_{ij,s}\}$, tiene restricciones que dependen del modelo en particular y del tipo de datos. Existen tres tipos de restricciones: **restricciones de proceso, restricciones de nivel y restricciones condicionales.**

Una restricción de proceso se emplea cuando los datos son binarios y otra restricción de proceso cuando los datos son continuos. Para el caso de datos discretos, las observaciones dentro de una categoría particular deben estar representados por el mismo valor real bajo el mapeo ϕ.

$$\phi:\ \delta_{ij,s} \sim \delta_{i'j',s'} \ \Rightarrow\ \hat{d}_{ij,s} = \hat{d}_{i'j',s'}$$ [2]

[1] ALSCAL son las siglas en inglés de Alternating Least squares SCALing, que traducido al español significa: Escalamiento de mínimos cuadrados alternante.
[2] El símbolo ~ significa que son miembros de la misma categoría

Cuando los datos son continuos, estos se discretizan de manera de hacer a los datos categóricos (por intervalos $[l,u)$). De esta manera, la restricción de continuidad se establece de la siguiente manera:

$$\phi:\ \delta_{ij,s} \sim \delta_{i'j',s'} \ \Rightarrow l \le \hat{d}_{ij,s}, \hat{d}_{i'j',s'} \le u$$

Una restricción de nivel se emplea cuando los datos son ordinales, nominales o cuantitativos. En el caso de que los datos sean de tipo ordinal, la restricción de nivel se establece de la siguiente manera:

$$\phi:\ \delta_{ij,s} \prec \delta_{i'j',s'} \ \Rightarrow \hat{d}_{ij,s} \le \hat{d}_{i'j',s'}$$

Cuando los datos son cuantitativos, la relación es lineal,

$$\phi: \hat{d}_{ij,s} = \mathrm{a}_0 + \mathrm{a}_1\ \delta_{ij,s}$$

Con $a_0 = 0$ para el caso de datos de razón. La linealidad puede reemplazarse por una relación polinomial.

Cuando los datos son nominales, no se requiere de una restricción de nivel, si ya se ha considerado la restricción de proceso.

Restricción condicional: Diferentes situaciones experimentales dan lugar a diferentes condiciones sobre las disimilaridades. Si las mediciones efectuadas por diferentes individuos son comparables, se tiene el caso incondicional. Cuando estas observaciones no son comparables, entonces se imponen unas matrices condicionales donde todas las disimilaridades dentro cada matriz individual son comparables y no lo son entre matrices. Esto implica que ϕ está compuesta por N mapeos $\{\phi_s\}$. De manera similar, la condición a renglón, está dada por $\{\phi_{ij}\}$, donde las disimilaridades entre los renglones de una matriz con otra son comparables pero no entre renglones de la misma matriz. Por ejemplo N personas juzgan el sabor de una bebida de p diferentes tipos de bebida.

MINIMIZACIÓN DEL STRESS CON ALSCAL[1]

Cada iteración del algoritmo ALSCAL[1] tiene dos fases: la fase de escalamiento óptimo y la fase de estimación del modelo. Denotando a la función de STRESS como $SS(\mathbf{X},\mathbf{W},\hat{\mathbf{D}})$, donde **X** es la matriz de coordenadas, **W** es la matriz de pesos y $\hat{\mathbf{D}}$ la matriz de disparidades, entonces la fase de escalamiento óptimo calcula las disparidades de mínimos cuadrados con **X** y **W** fijas. En la fase de estimación calcula las nuevas coordenadas **X** y los pesos **W** para $\hat{\mathbf{D}}$ fija.

[1]Fuente: Cox & Cox., 1994. Cap. 10

Fase de escalamiento óptimo

Primero se calculan las distancias actuales $\{d_{ij,k}^2\}$ a partir de la matriz de coordenadas originales **X** y **W**. Si todas las disparidades se localizan en vector $\hat{\mathbf{d}}$, y de manera similar, las distancias en **d**, entonces

$$\hat{\mathbf{d}} = \mathbf{E}\mathbf{d}, \text{ donde } \mathbf{E} = \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T.$$

Si los datos son de intervalo o de razón Z es el vector de disimilaridades cuadradas $\{\delta_{ij,k}^2\}$. Otra manera, es reemplazar las disimilaridades cuadradas en la función de STRESS por $a + bd\delta_{ijs}^2$ y a y b se estiman por mínimos cuadrados.

Si los datos son ordinales o nominales, Z es la matriz de variables "dummy" indicando que distancias deben tratarse para satisfacer las condiciones de medición. Esto ocurre, cuando para objetos distintos se tiene la misma disparidad, es decir, $\hat{d}_i^2 = \hat{d}_j^2$.

Ahora, la función de STRESS puede expresarse de la siguiente manera:

$$SS = \mathbf{d}^T(\mathbf{I} - \mathbf{E})\mathbf{d}$$

y la función de STRESS normalizada:

$$SS = \mathbf{d}^T(\mathbf{I} - \mathbf{E})\mathbf{d} / \mathbf{d}^T\mathbf{d}$$

El último paso en la fase de escalamiento óptimo es normalizar la solución, primero con respecto a la configuración y pesos y otros parámetros y después con respecto al STRESS.

Fase de estimación del modelo

En la fase de estimación del modelo, se estima la matriz de pesos **W** por medio de mínimos cuadrados dadas las disparidades y las coordenadas de los puntos **X** de puntos en el espacio de objetos. Después se estiman las coordenadas de X por mínimos cuadrados.

Para la primera minimización, las $\frac{1}{2}n(n-1)$ $(x_{it} - x_{jt})^2$ cantidades correspondientes a la t-ésima columna de la matriz **Y**. De manera similar se compone la matriz **D*** de disparidades cuadradas de tamaño $\frac{1}{2}n(n-1)$ x p.

De esta manera, el STRESS se puede expresar de la siguiente manera:

$$SS = tr(\mathbf{D}^* - \mathbf{W}\mathbf{Y}^T)^T(\mathbf{D}^* - \mathbf{W}\mathbf{Y}^T)$$

y por tanto

$$\mathbf{W} = \mathbf{D}^*\mathbf{Y}(\mathbf{Y}^T\mathbf{Y})^{-1}$$

Para la segunda minimización, la función de STRESS es minimizada con respecto a las coordenadas X. Igualando las derivadas parciales a cero, se obtiene una serie de ecuaciones cúbicas que pueden resolverse por el método de newton.

3.2.4 OTROS MÉTODOS DE ESCALAMIENTO

Las transformaciones monótonas son parte de cualquier algoritmo de escalamiento no métrico. Hasta este momento, únicamente se ha empleado la regresión isotónica para encontrar tales disparidades que minimicen el error, sin embargo, los métodos que se pueden emplear son muy extensos.

Por lo general, en los algoritmos de escalamiento no métrico las distancias son normalizadas, es decir, $\sum_{i<j} d_{ij}^2 = 1$ y el problema se centra en minimizar la función de error conocida como Row Stress S_R. El Stress está en función de las disparidades y de las distancias. En los algoritmos se re-estiman las disparidades de manera que reflejen las distancias lo mejor posible sujetos a la condición (C1). En el método de Kruskal, esto se hace con regresión isotónica y a esta etapa se le llama "etapa no métrica". Los métodos anteriormente descritos, se centran en minimizar la función de error en la "etapa métrica", es decir, se ajustan las posiciones de las coordenadas de manera que las distancias correspondan lo mejor posible a las disparidades.

Uno de los métodos que se encontró en la bibliografía fue el de redes neuronales, sin embargo este tema está fuera del alcance de esta investigación, pero podemos resaltar lo siguiente:

El algoritmo de redes neuronales que se emplea es el algoritmo de propagación de error hacia atrás, que se centra en la parte no métrica del escalamiento multidimensional no métrico.

Esta red neuronal toma a las disimilaridades como entradas y las disparidades como salidas y una capa de unidades ocultas con funciones de transferencia no lineales. La unidad de salida emplea la función identidad como función de transferencia.[1]

[1]Para mayor información véase Apéndice, apartado II

CAPÍTULO 4

MODELOS ESPECIALES DE ESCALAMIENTO MULTIDIMENSIONAL

CAPÍTULO 4

MODELOS ESPECIALES DE ESCALAMIENTO MULTIDIMENSIONAL

En los capítulos anteriores se describieron los conceptos básicos de las técnicas de escalamiento multidimensional y los diferentes tipos de escalamiento multidimensional. En este capítulo se describen brevemente algunos modelos especiales que son los más frecuentemente utilizados.

4.1. PROCRUSTES

Algunas veces es necesario comparar una configuración de puntos en un espacio euclideano con algún otro, donde existe una correspondencia uno a uno de un conjunto de puntos a otro. A la técnica que hace esta correspondencia entre una configuración con otra y produce una medida de esta correspondencia se le llama análisis de Procrustes[1].

El análisis de Procrustes hace dilataciones y transformaciones rígidas. Las trasformaciones rígidas son las trasformaciones que preservan la distancia. Bajo una transformación rígida, el plano se mueve sin deformación como si fuese un cuerpo rígido. Estas transformaciones rígidas son traslaciones, rotaciones alrededor de un punto, y una rotación de 180° alrededor de una recta del plano. Esta última transformación es llamada reflexión.

Suponiendo una configuración de n puntos en un espacio euclideano de q dimensiones, con coordenadas $\mathbf{X}$ de nxq. Dicha configuración necesita coincidir de manera óptima con otra configuración de n puntos en un espacio euclideano de p ($p \geq q$) dimensiones, con matriz de coordenadas $\mathbf{Y}$. El primer paso consiste en agregar p-q columnas de ceros a la matriz $\mathbf{X}$, de manera que $\mathbf{X}$ y $\mathbf{Y}$ tengan las mismas dimensiones. La suma de las distancias entre los puntos en el espacio $\mathbf{Y}$ y los puntos correspondientes en el espacio $\mathbf{X}$ está dada por:

$$R^2 = \sum_{i=1}^{n} (\mathbf{y}_i - \mathbf{x}_i)^T (\mathbf{y}_i - \mathbf{x}_i) \text{ ..Coeficiente 1}$$

[1]Procrustes no tiene traducción al español. Su nombre se debe a la mitología Griega. Procrustes o Damastes, vivía a la orilla del camino a Atenas y cualquier viajero que pasaba por ese camino era invitado de Damastes quien los sorprendía por su hospitalidad y la cama para dormir del invitado. Si su invitado no cabía en la cama, Damastes ajustaba la cama, haciéndola más pequeña o larga según sea el caso. Por tal motivo, a Damastes lo llamaron Procrustes que significa extendimiento o bastidor.

Donde $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_n]^T$, $\mathbf{Y} = [\mathbf{y}_1, ..., \mathbf{y}_n]^T$, y $\mathbf{x}_i$, $\mathbf{y}_i$ son los vectores coordenadas del i -ésimo punto en los dos espacios.

Aplicando las transformaciones de dilatación, traslación, rotación y reflexión a los puntos del espacio $\mathbf{X}$, se obtienen las nuevas coordenadas $\mathbf{x}_i^{'}$, donde

$$\mathbf{x}_i^{'} = \rho \mathbf{A}^T \mathbf{x}_i + \mathbf{b}.$$

La matriz $\mathbf{A}$ es ortogonal proporcionando la rotación rígida, el vector $\mathbf{b}$ es la traslación rígida del vector y ρ es la dilatación. Estos movimientos se calculan de manera que la suma de distancias entre puntos se minimice, así:

$$R^2 = \sum_{i=1}^{n} (\mathbf{y}_i - \rho \mathbf{A}^T \mathbf{x}_i - \mathbf{b})^T (\mathbf{y}_i - \rho \mathbf{A}^T \mathbf{x}_i - \mathbf{b})$$

4.1.1 Traslación Óptima

Sean $\mathbf{x}_0$ y $\mathbf{y}_0$ los centroides de las dos configuraciones, es decir:

$$\mathbf{x}_0 = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i, \quad \mathbf{y}_0 = \frac{1}{n} \sum_{i=1}^{n} \mathbf{y}_i$$

Midiendo a $\mathbf{x}_i$, $\mathbf{y}_i$ relativos a sus centroides, se tiene:

$$R^2 = \sum_{i=1}^{n} ((\mathbf{y}_i - \mathbf{y}_0) - \rho \mathbf{A}^T (\mathbf{x}_i - \mathbf{x}_0) + \mathbf{y}_0 - \rho \mathbf{A}^T \mathbf{x}_0 - \mathbf{b})^T ((\mathbf{y}_i - \mathbf{y}_0) - \rho \mathbf{A}^T (\mathbf{x}_i - \mathbf{x}_0) + \mathbf{y}_0 - \rho \mathbf{A}^T \mathbf{x}_0 - \mathbf{b})$$

$$= \sum_{i=1}^{n} ((\mathbf{y}_i - \mathbf{y}_0) - \rho \mathbf{A}^T (\mathbf{x}_i - \mathbf{x}_0))^T ((\mathbf{y}_i - \mathbf{y}_0) - \rho \mathbf{A}^T (\mathbf{x}_i - \mathbf{x}_0)) + n(\mathbf{y}_0 - \rho \mathbf{A}^T \mathbf{x}_0 - \mathbf{b})^T (\mathbf{y}_0 - \rho \mathbf{A}^T \mathbf{x}_0 - \mathbf{b})$$

Como el último término de la expresión anterior no es negativo, con el objetivo de minimizar la expresión anterior con respecto a la traslación, se tiene:

$$\mathbf{b} = \mathbf{y}_0 - \rho \mathbf{A}^T \mathbf{x}_0 \text{ y por tanto: } \mathbf{x}_i^{'} = \rho \mathbf{A}^T (\mathbf{x}_i - \mathbf{x}_0) + \mathbf{y}_0$$

Haciendo esto, el centroide de la configuración $\mathbf{X}$ es igual al centroide de la configuración $\mathbf{Y}$.

4.1.2 Dilatación Óptima

Asumiendo que $\mathbf{x}_0 = \mathbf{y}_0 = \mathbf{0}$, entonces:

$$R^2 = \sum_{i=1}^{n} (\mathbf{y}_i - \rho \mathbf{A}^T \mathbf{x}_i)^T (\mathbf{y}_i - \rho \mathbf{A}^T \mathbf{x}_i)$$

$$= \sum_{i=1}^{n} \mathbf{y}_i^T \mathbf{y}_i + \rho^2 \sum_{i=1}^{n} \mathbf{x}_i^T \mathbf{x}_i - 2\rho \sum_{i=1}^{n} \mathbf{x}_i^T \mathbf{A}^T \mathbf{y}_i$$

$$= tr(\mathbf{Y}\mathbf{Y}^T) + \rho^2 tr(\mathbf{X}\mathbf{X}^T) - 2\rho tr(\mathbf{X}\mathbf{A}\mathbf{Y}^T) \text{...Ecuación (E1)}$$

Derivando parcialmente la función anterior con respecto a ρ e igualando a cero, se tiene:

$$\hat{\rho} = \frac{tr(\mathbf{X}\mathbf{A}\mathbf{Y}^T)}{tr(\mathbf{X}\mathbf{X}^T)} = \frac{tr(\mathbf{A}\mathbf{Y}^T\mathbf{X})}{tr(\mathbf{X}\mathbf{X}^T)}$$

4.1.3 Rotación Óptima:

Sibson derivó la matriz de rotación en una forma elegante sin requerir de la derivada matricial de R^2. El valor de R^2 de la ecuación (E1) se minimiza si $tr(\mathbf{X}\mathbf{A}\mathbf{Y}^T) = tr(\mathbf{A}\mathbf{Y}^T\mathbf{X})$ se maximiza. Sea $\mathbf{C} = \mathbf{Y}^T\mathbf{X}$, con descomposición en valores singulares:

$$\mathbf{C} = \mathbf{U}\,\Lambda\,\mathbf{V}^T,$$

Donde $\mathbf{U}$ y $\mathbf{V}$ son matrices ortonormales y Λ es la matriza diagonal de valores singulares. Entonces:

$$tr(\mathbf{A}\mathbf{C}) = tr(\mathbf{A}\mathbf{U}\,\Lambda\,\mathbf{V}^T) = tr(\mathbf{V}^T\mathbf{A}\mathbf{U}\,\Lambda)$$

Ahora, $\mathbf{V}$, $\mathbf{A}$ y $\mathbf{U}$ son matrices ortonormales y por tanto, $\mathbf{V}^T\mathbf{A}\mathbf{U}$ también lo es. La matriz diagonal Λ es ortonormal y por lo tanto, no puede tener ningún elemento mayor a la unidad,

$$tr(\mathbf{A}\mathbf{C}) = tr(\mathbf{V}^T\mathbf{A}\mathbf{U}\,\Lambda) \le tr(\Lambda).$$

y por lo tanto, $\mathbf{R}^2$ se minimiza cuando $tr(\mathbf{A}\mathbf{C}) = tr(\Lambda)$ y $\mathbf{V}^T\mathbf{A}\mathbf{U}\,\Lambda = \Lambda$ ..Ecuación (E2) de donde la solución de la matriz de rotación óptima está dada por:

$$\mathbf{A} = \mathbf{V}\mathbf{U}^T$$

Ahora, pre-multiplicando y post-multiplicando por $\mathbf{V}$ y $\mathbf{V}^T$ en la ecuación (E2), se tiene:

$$\mathbf{A}\mathbf{U}\,\Lambda\mathbf{V}^T = \mathbf{V}\Lambda\,\mathbf{V}^T,$$

y por tanto, $\mathbf{AC} = \mathbf{V\Lambda V^T} = (\mathbf{V\Lambda^2 V^T})^{1/2} = (\mathbf{V\Lambda UU^T\Lambda V^T})^{1/2} = (\mathbf{C^TC})^{1/2}$

Por lo tanto, la matriz de rotación óptima está dada por:

$(\mathbf{C^TC})^{1/2}\mathbf{C^{-1}} = (\mathbf{X^TYY^TX})^{1/2}(\mathbf{Y^TX})^{-1}$ si $\mathbf{Y^TX}$ es no singular y por $(\mathbf{C^TC})^{1/2}$ en otro caso.

4.1.4 Coeficiente de Congruencia

Tucker[1951] introdujo otro coeficiente de congruencia entre dos vectores **x** y **y** con las siguiente expresión:

$$\Gamma(\mathbf{x},\mathbf{y}) = \frac{\mathbf{x^Ty}}{(\mathbf{x^Txy^Ty})^{1/2}}$$ ..Coeficiente 2

El valor máximo de este coeficiente es la unidad cuando $\mathbf{x} = \lambda y,\ \lambda > 0$.

Otra expresión para medir la congruencia está dada por :

$$c(\mathbf{X},\mathbf{Y}) = \frac{\sum_{i<j} d_{ij}(\mathbf{X})d_{ij}(\mathbf{Y})}{(\sum_{i<j} d^2{}_{ij}(\mathbf{X})^{1/2})(\sum_{i<j} d^2{}_{ij}(\mathbf{Y})^{1/2})}$$ Coeficiente 3

4.2 DESDOBLAMIENTO

El modelo de desdoblamiento (unfolding[1]) es un es un modelo de elección preferencial. Asume que los individuos perciben varios objetos de su elección en la misma manera pero difieren con respecto a lo que cada uno considera la combinación ideal de los atributos de los objetos. En unfolding, los datos generalmente son preferencias o rankings de diferentes individuos para un conjunto de objetos. Los individuos son presentados como "puntos ideales" en la configuración del escalamiento multidimensional de tal manera que las distancias de los individuos a los objetos correspondan a las preferencias individuales y de esta manera, los puntos relativos a objetos más alejados del punto ideal deben preferirse menos. Los modelos unfolding pueden ser unidimensionales o multidimensionales, métricos o no métricos.

Coombs (1950) introdujo la escala J y la escala I en el caso unidimensional no métrico.

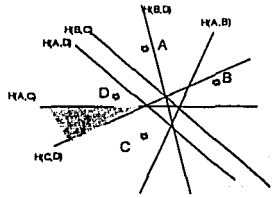
[1] "Unfolding" traducido al español significa "Desdoblamiento". Supongamos que se tiene un escalmiento con n objetos y m individuos, es decir, con n+m puntos impreso en un pañuelo. Si este pañuelo se toma con dos dedos en un punto representando a un individuo y doblando con la otra mano, entonces los puntos representando a los objetos más alejados de forma vertical son los menos preferidos por ese individuo. El escalamiento multidimensional es el proceso inverso al del doblar el pañuelo, es decir, el desdoblamiento (unfolding) de las preferencias en distancias. En este trabajo se empleará el término Unfolding.

4.2.1 DESDOBLAMIENTO NO MÉTRICO

Suponiendo que se tiene m objetos ordenados de acuerdo a la preferencia de n individuos. La preferencia ordenada de los individuos es llamada escala I y la línea sobre la cual son colocados los puntos para los n juicios o individuos con sus m objetos es llamada escala J.

Benett y Hays (1960) generalizaron el modelo unfolding unidimensional de Coombs al caso multidimensional. La localización de los puntos, equidistantes del objeto A y B, es el hiperplano H(A,B) de dimensión p-1. El hiperplano H(A,B) divide el espacio en dos mitades, donde los individuos localizados en una de las mitades prefieren al objeto A más que al B y los individuos localizados en la otra mitad, prefieren al B más que al A. Estas mitades divididas por los hiperplanos son llamadas *regiones isotónicas*. Los puntos localizados en la región sombreada de la figura 4.1 es la región isotónica de los individuos cuyo orden de preferencia es D-C-A-B.

Figura 4
Regiones isotónicas para cuatro objetos en dos dimensiones



4.2.2 DESDOBLAMIENTO MÉTRICO

Schönemann (1970) encontró una solución algebraica al unfolding métrico. Suponiendo que se tiene a n individuos o juicios, y que el r-ésimo individuo produce disimilaridades $\{\delta_{ri}\}$ para los m objetos. La configuración resultante, sería un espacio euclideano de p dimensiones con n+m puntos representando a los objetos y a los individuos. Sea $\mathbf{x}_i$ (i=1,..n), la coordenada del punto que representa al i-ésimo individuo y $\mathbf{y}_j$ (j=1,..,m) la coordenada del punto que representa al j-ésimo objeto. Sea d_{ij} la distancia entre el i-ésimo individuo y el j-ésimo objeto. El unfolding métrico busca una configuración de puntos tal que las distancias representen lo mejor posible a las disimilaridades.

Sea $\mathbf{X} = [\mathbf{x}_1,...,\mathbf{x}_n]^T$, $\mathbf{Y} = [\mathbf{y}_1,...,\mathbf{y}_m]^T$. Sea $\mathbf{D}(\mathbf{X},\mathbf{Y})$ la matriz de distancias cuadradas entre puntos de individuos y puntos de objetos. Por tanto,

$$[D(\mathbf{X},\mathbf{Y})]_{ij} = (\mathbf{x}_i - \mathbf{y}_j)^T(\mathbf{x}_i - \mathbf{y}_j)$$

Sea $\mathbf{D}$, la matriz de disimilaridades cuadradas, es decir, $\mathbf{D} = [\delta_{ij}^2]$. El unfolding métrico busca que $\mathbf{D} = \mathbf{D}(\mathbf{X},\mathbf{Y})$. Las matrices $\mathbf{D}$ y $\mathbf{D}(\mathbf{X},\mathbf{Y})$ se centran doblemente de manera que $\mathbf{C} = \mathbf{H}\,\mathbf{D}\,\mathbf{H}$ y $\mathbf{C}(\mathbf{X},\mathbf{Y}) = \mathbf{H}\,\mathbf{D}\,(\mathbf{X},\mathbf{Y})\,\mathbf{H}$, donde $\mathbf{H} = \mathbf{I} - n^{-1}\mathbf{1}\mathbf{1}^T$, la matriz que centra. De esta manera, el problema de unfolding se puede rescribir de la siguiente forma:

El primer paso consiste en encontrar las coordenadas $(\mathbf{X},\mathbf{Y})$ que satisfagan $\mathbf{C}(\mathbf{X},\mathbf{Y}) = \mathbf{C}$ y el segundo paso es que satisfagan:

$$\mathbf{D}(\mathbf{X},\mathbf{Y})_{i.} = \mathbf{D}_{i.}\ \ i = 1,..., n$$
$$\mathbf{D}(\mathbf{X},\mathbf{Y})_{.j} = \mathbf{D}_{.j}\ \ j = ,..., m$$

Este algoritmo es conocido como algoritmo de Schönemann. Greenacre y Browne (1986) introdujeron otro algoritmo incorporando los residuales $\{\varepsilon_{ij}\}$ donde

$$\delta_{ij}^2 = d_{ij}^2 + \varepsilon_{ij}$$

El unfolding busca minimizar $\sum_i \sum_j \varepsilon_{ij}^2 = tr(\mathbf{R}\mathbf{R}^T)$, donde $[\mathbf{R}]_{ij} = \varepsilon_{ij}$

Sea

$$f = \sum_{i=1}^{n} \sum_{j=1}^{m} \{\delta_{ij}^2 - (\mathbf{x}_i - \mathbf{y}_j)^T(\mathbf{x}_i - \mathbf{y}_j)\}^2 \text{ ...Ecuación (E3)}$$

Entonces derivando parcialmente la ecuación (E3) respecto a las coordenadas de los individuos e igualando a cero, se tiene:

$$\frac{\partial f}{\partial \mathbf{x}_i} = 4\sum_{j=1}^{m}\{\delta_{ij}^2 - (\mathbf{x}_i - \mathbf{y}_j)^T(\mathbf{x}_i - \mathbf{y}_j)\}(\mathbf{x}_i - \mathbf{y}_j) = 0$$

De donde,

$$\sum_{j=1}^{m}\{\delta_{ij}^2 - (\mathbf{x}_i - \mathbf{y}_j)^T(\mathbf{x}_i - \mathbf{y}_j)\}\mathbf{x}_i = \sum_{j=1}^{m}\{\delta_{ij}^2 - (\mathbf{x}_i - \mathbf{y}_j)^T(\mathbf{x}_i - \mathbf{y}_j)\}\mathbf{y}_j$$

De otra manera: $\sum_{j=1}^{m}[\mathbf{R}]_{ij}\mathbf{y}_j = \sum_{j=1}^{m}[\mathbf{R}]_{ij}\mathbf{x}_j$ y por tanto

$$\mathbf{R}\mathbf{Y} = \text{diag}(\mathbf{R}\mathbf{J}^T)\,\mathbf{X} \text{ .. Ecuación (E4)}$$

De manera similar:

$$\mathbf{R}^T\mathbf{X} = \text{diag}(\mathbf{R}^T\mathbf{J})\,\mathbf{Y} \text{ . Ecuación (E5)}$$

Donde J es una matriz *n x m* de unos. Las ecuaciones (E4) y (E5) se resuelven numéricamente.

4.3 ANÁLISIS DE CORRESPONDENCIA

El análisis de Correspondencia es un método multivariado popular y está relacionado con otras técnicas estadísticas como Análisis de correlación canónico, análisis de componentes principales, escalamiento dual, etc. Existe numerosa bibliografía que trata este tema. En este trabajo, el método de análisis de correspondencia es visualizado como un método de escalamiento multidimensional métrico, donde los datos son renglones y columnas de una matriz o tabla de contingencia con valores no negativos.

El análisis de correspondencia encuentra dos espacios vectoriales, uno para los datos de renglones de la tabla de contingencia y otro para las columnas, de tal manera que ambos vectores se traten de igual manera.

Se estandarizan los renglones de $\mathbf{X}$ de manera que la suma del renglón i-ésimo sea 1. De igual manera, se estandarizan las columnas de $\mathbf{X}$ de manera que la suma de la columna j-ésima sea 1. Así, la matriz de renglones estandarizados es $\mathbf{D}_i^{-1}\mathbf{X}$ donde $\mathbf{D}_i$ es la matriz diagonal que estandariza los renglones de $\mathbf{X}$ y la matriz de columnas estandarizadas es $\mathbf{D}_j^{-1}\mathbf{X}^T$, donde $\mathbf{D}_j$ es la matriz diagonal que estandariza las columnas de $\mathbf{X}$.

Sea $\mathbf{X} = \mathbf{A}\mathbf{D}_\lambda\mathbf{B}^T$..(1) la descomposición en valores singulares de $\mathbf{X}$, donde

$$\mathbf{A}^T\mathbf{D}_i^{-1}\mathbf{A} = \mathbf{B}^T\mathbf{D}_j^{-1}\mathbf{B} = \mathbf{I}$$

Donde $\mathbf{A}$ es una base ortonormal para la columnas de $\mathbf{X}$, normalizadas con respecto a $\mathbf{D}_i^{-1}$ y $\mathbf{B}$ es una base ortonormal para los renglones de $\mathbf{X}$, normalizado respecto a $\mathbf{D}_j^{-1}$.

Ahora, para los renglones, a partir de (1) se tiene:

$$\mathbf{D}_i^{-1}\mathbf{X} = \mathbf{D}_i^{-1}\mathbf{A}\mathbf{D}_\lambda\mathbf{B}^T,$$

con

$$(\mathbf{D}_i^{-1}\mathbf{A})^T\mathbf{D}_i(\mathbf{D}_i^{-1}\mathbf{A}) = \mathbf{B}^T\mathbf{D}_j^{-1}\mathbf{B} = \mathbf{I}.$$

Sea $\mathbf{U} = \mathbf{D}_i^{-1}\mathbf{A}$, entonces, $\mathbf{D}_i^{-1}\mathbf{X} = \mathbf{U}\mathbf{D}_\lambda\mathbf{B}^T$, $\mathbf{U}^T\mathbf{D}_i\mathbf{U} = \mathbf{B}^T\mathbf{D}_j^{-1}\mathbf{B} = \mathbf{I}$...Ecuación (E6)

De manera similar, para las columnas: $\mathbf{D}_j^{-1}\mathbf{X}^T = \mathbf{D}_j^{-1}\mathbf{B}\mathbf{D}_\lambda\mathbf{A}^T$,

con

$$\mathbf{A}^T\mathbf{D}_i^{-1}\mathbf{A} = (\mathbf{D}_j^{-1}\mathbf{B})^T\mathbf{D}_j(\mathbf{D}_j^{-1}\mathbf{B}) = \mathbf{I}$$

Sea $\mathbf{V} = \mathbf{D}_j^{-1}\mathbf{B}$, entonces, $\mathbf{D}_j^{-1}\mathbf{X}^T = \mathbf{V}\mathbf{D}_\lambda\mathbf{A}^T$, $\mathbf{A}^T\mathbf{D}_i^{-1}\mathbf{A} = \mathbf{V}^T\mathbf{D}_j\mathbf{V} = \mathbf{I}$...Ecuación (E7)

La ecuación (E6) muestra que los renglones pueden expresarse en el espacio $\mathbf{Y}\,\Delta_\lambda$, con $\mathbf{B}$ la matriz de rotación que transforma a los renglones en puntos en el espacio $\mathbf{Y}\,\Delta_\lambda$.

Para una representación en un espacio de pocas dimensiones, la descomposición en valores singulares permite las primeras k columnas de $\mathbf{Y}\,\Delta_\lambda$ como la mejor aproximación de mínimos cuadrados en un espacio de k dimensiones..

De manera similar, la ecuación (E7) muestra que las columnas pueden representarse en el espacio $\mathbf{Y}\,\Delta_\lambda$, con $\mathbf{A}$ la matriz de rotación.

Inercia:

Sean

$x_{i.} = \sum_j x_{ij}$, la suma de los renglones de $\mathbf{X}$,

$x_{.j} = \sum_i x_{ij}$, la suma de las columnas de $\mathbf{X}$,

$n = \sum_{i,j} x_{ij}$, la suma total.

La distancia euclideana ponderada entre el renglón i-ésimo y el renglón j-ésimo está dada por:

$$d_{ij} = \left(\sum_k \frac{\left(\frac{x_{ik}}{x_{i.}} - \frac{x_{jk}}{x_{j.}} \right)^2}{\frac{x_{.k}}{n}} \right)^{1/2},$$

y la distancia euclideana ponderada entre el renglón i-ésimo y el renglón promedio z está dada por:

$$d_{iz} = \left(\sum_k \frac{\left(\frac{x_{ik}}{x_{i.}} - \frac{x_{.k}}{n} \right)^2}{\frac{x_{.k}}{n}} \right)^{1/2}.$$

Las distancias anteriores, son llamadas distancias χ^2 (ji-cuadrada), ya que:

$$\sum_i \frac{x_{i.}}{n} d_{iz}^2 = n^{-1} \left(\sum_{i,k} x_{i.} \frac{\left(\frac{x_{ik}}{x_{i.}} - \frac{x_{.k}}{n} \right)^2}{\frac{x_{.k}}{n}} \right)$$

$$= n^{-1} \left(\sum_{i,k} \frac{\left(x_{ik} - \frac{x_{i.} x_{.k}}{n} \right)^2}{\frac{x_{i.} x_{.k}}{n}} \right) = n^{-1} \chi^2$$

La cantidad anterior $n^{-1}\chi^2$ es llamada "inercia total". La inercia total es una medida de dispersión de los renglones.

La expresión anterior puede expresarse de la siguiente forma:

$$n^{-1}\mathbf{X}^2 = \sum_i \left\{ x_{i.} \sum_k \frac{\frac{x_{ik}}{x_{i.}} - \frac{x_{.k}}{n}}{x_{.k}} \right\},$$

$$= \sum_i x_{i.}(\mathbf{i}-\mathbf{z})^T \mathbf{D}_I^{-1}(\mathbf{i}-\mathbf{z}) = \mathbf{I},$$

Donde $\mathbf{i}$ es el i-ésimo renglón expresado en forma vectorial, $\mathbf{z}$ es el renglón promedio formado por la suma de columnas. Esta es la inercia total para los renglones. Es una suma ponderada por las distancias ponderadas cuadradas entre los renglones y el promedio para los renglones.
Intercambiando renglones por columnas, se tiene la inercia total para los renglones

$$= \sum_j x_{.j}(j-\mathbf{z})^T \mathbf{D}_J^{-1}(j-\mathbf{z}) = \mathbf{I},$$

Donde j es la j-ésima columna expresada en forma vectorial y $\mathbf{z}$ es la columna promedio formada por la suma de los renglones. Por simetría de χ^2, la expresión anterior es igual a la inercia total para los renglones.

Ahora, $\mathbf{D}_I^{-1}\mathbf{X}\mathbf{1} = \mathbf{D}_J^{-1}\mathbf{X}^T\mathbf{1} = \mathbf{1}$, por lo que siempre hay un eigenvalor 1 asociado al eigenvector $\mathbf{1}$, proporcionando la dimensión trivial y puede omitirse de los cálculos, por lo que se tendría:

$\mathbf{D}_I^{-1}\mathbf{X} - \mathbf{1}J_\Sigma^T$ y $\mathbf{D}_J^{-1}\mathbf{X}^T - \mathbf{1}I_\Sigma^T$, donde I_Σ y J_Σ son los vectores de la suma de los renglones y columnas respectivamente.

Ahora $\mathbf{I}$, puede expresarse de la siguiente forma:

$$\mathbf{I} = tr(\mathbf{D}_I(\mathbf{D}_I^{-1}\mathbf{X} - \mathbf{1}J_\Sigma^T)\mathbf{D}_J^{-1}(\mathbf{D}_I^{-1}\mathbf{X} - \mathbf{1}J_\Sigma^T)^T)$$

Donde $\mathbf{D}_I^{-1}\mathbf{X} - \mathbf{1}J_\Sigma^T$ es la matriz para de renglones estandarizados eliminando la dimensión trivial.

Asumiendo que $\mathbf{D}_I^{-1}\mathbf{X} - \mathbf{1}J_\Sigma^T = \mathbf{D}_I^{-1}\mathbf{X}$, es decir, que la dimensión trivial ha sido eliminada, se tendría:

$$\mathbf{I} = tr(\mathbf{D}_I(\mathbf{D}_I^{-1}\mathbf{X})\mathbf{D}_J^{-1}(\mathbf{D}_I^{-1}\mathbf{X})^T)$$

$$= tr((\mathbf{A}\mathbf{D}_\lambda\mathbf{B})\mathbf{D}_J^{-1}(\mathbf{B}\mathbf{D}_\lambda^1\mathbf{A}^T)\mathbf{D}_I^{-1})$$

$$= tr(\mathbf{A}\mathbf{D}_\lambda^2\mathbf{A}^T\mathbf{D}_I^{-1}) = tr(\mathbf{D}_\lambda^2\mathbf{A}^T\mathbf{D}_I^{-1}\mathbf{A})$$

$$= tr(\mathbf{D}_\lambda^2)$$

Por lo tanto, la inercia total es la suma de los valores singulares cuadrados. La dimensión requerida para los renglones y columnas puede ser juzgada por la contribución de la inercia total. De esta manera, si se elige un espacio de k dimensiones, su contribución a la inercia total está dada por:

$$\frac{\sum_{i=1}^{k} \lambda_i^2}{\sum_{i=1}^{n} \lambda_i^2}$$

Donde n es el número total de valores singulares (diferentes a la unidad).

COMPARACIÓN DE ANÁLISIS DE CORRESPONDENCIA Y ESCALAMIENTO MULTIDIMENSIONAL

El análisis de correspondencia tiene numerosas propiedades comunes con el escalamiento multidimensional, pero difiere en otros aspectos. Ambas técnicas realizan una representación gráfica de los objetos en un espacio de pocas dimensiones. El escalamiento multidimensional métrico clásico es una técnica de un modo, mientras que análisis de correspondencia es una técnica de dos modos (como es el caso del modelo unfolding). Los datos en tabla de contingencia son estrictamente no negativos, mientras que el escalamiento multidimensional no tiene esta restricción. El escalamiento multidimensional acepta cualquier medida de disimilaridad mientras que el análisis de correspondencia únicamente emplea distancias ji-cuadrada.

En el escalamiento multidimensional, se puede interpretar directamente las distancias entre todos lo puntos, mientras que en el análisis de correspondencia, únicamente se pueden interpretar directamente las distancias entre renglones o las distancias entre columnas.

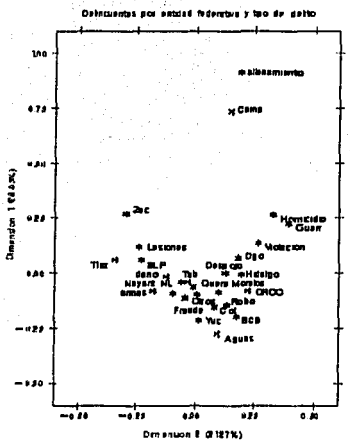
Existe una gran relación entre el escalamiento multidimensional clásico y análisis de correspondencia. Supongamos de $\mathbf{D}$ es la matriz de distancias ji-cuadradadas entre renglones, entonces, reemplazando H, la matriz que centra empleada en el escalamiento multidimensional métrico, por $\mathbf{H}_w = \mathbf{I} - (\mathbf{1}^T \mathbf{D}_r \mathbf{1})^{-1} \mathbf{1}\mathbf{1}^T \mathbf{D}_r$, entonces la descomposición de:

$$\mathbf{B} = \mathbf{H}_w \ \mathbf{A} \ \mathbf{H}_w, \text{ donde } \mathbf{A} = -\frac{1}{2}\mathbf{D}^2,$$

sería la misma solución para los renglones, empleando escalamiento multidimensional métrico y análisis de correspondencia.

Considérese la tabla de contingencia (Tabla 4.1 del I anexo, apartado 4.1) de los delincuentes sentenciados INEGI[2000], por entidad federativa y por tipo de delito.

El resultado obtenido con el paquete de cómputo SAS del análisis de correspondencia en dos dimensiones fue:



La contribución a la inercia obtenida en la representación gráfica en dos dimensiones fué de 47.2% (Anexo I , apartado 4.1), siendo muy pobre.

4.4 DIFERENCIAS INDIVIDUALES

Hasta este momento, el tipo de datos que se han analizado son de dos factores y de uno o dos modos. Este modelo emplea datos de tres factores y dos modos, es decir, cuando las disimilaridades son de la forma $\{\delta_{ij,s}\}$ donde los subíndices ij se refieren a un conjunto de objetos y s a otro conjunto de objetos.

Suponiendo el caso en que se tiene a N individuos que realizan juicios acerca de n objetos, para el cual se forman N matrices de disimilaridades con disimilaridades $\{\delta_{ij}\}$ entre el i y j-ésimo objeto.

Existen dos formas de analizar este tipo de datos. La primera consiste en promediar entre individuos y la segunda es comparar los resultados individuales. Tucker y Messick (1963) atacaron las dos formas anteriores ya que en el primer caso, al promediar entre individuos se pierde mucha información y en el segundo caso, el análisis pierde sentido ya que sería muy difícil comparar muchos escalamientos diferentes. Tucker y Messick propusieron colocar a las disimilaridades $\{\delta_{ij,s}\}$ en una matriz $\mathbf{X}$, con renglones formados por todas los $\frac{1}{2}n(n-1)$ parejas posibles de objetos y N columnas representando a los individuos.

De esta manera, la descomposición en valores singulares de X se encuentra de la siguiente forma:

$$\mathbf{X} = \mathbf{U}\,\mathbf{\Lambda}\,\mathbf{V}^T,$$

y por tanto, la aproximación de mínimos cuadrados en p dimensiones de X, es:

$$\hat{\mathbf{X}}_p = \mathbf{U}_p\,\mathbf{\Lambda}_p\,\mathbf{V}_p^T$$

La matriz $\mathbf{U}_p$ es la matriz de componentes principales en el espacio de las parejas de estímulos y la matriz $\mathbf{\Lambda}_p\,\mathbf{V}_p^T$ es la matriz de componentes principales para el espacio de los individuos.

Carroll y Chang (1970) propusieron un modelo métrico empleando dos espacios: el primer espacio para el conjunto de objetos y el otro para los individuos, ambos de dimensión p. Los puntos que representan a los objetos se colocan en un espacio y a los puntos que representan a los individuos se colocan en el espacio de los objetos. Las coordenadas para cada individuo son los pesos requeridos en las distancias euclideanas ponderadas entre los puntos del espacio de objetos y son los valores que mejor representen las disimilaridades correspondientes a cada individuo. A este modelo le llamaron INDSCAL[1].

Sean x_{it} *(i=1,...,n; t=1,..,p)* los puntos en el espacio de objetos. Sean w_{st} *(s=1,...N; t=1,...,p)* los puntos en el espacio de individuos. Entonces, la distancia euclideana ponderada entre el objeto i y el objeto j para el s-ésimo individuo:

$$d_{ij,s} = \left\{ \sum_{t=1}^{p} w_{st} (x_{it} - x_{jt})^2 \right\}^{1/2}.$$

Lo que se busca es que los pesos y las coordenadas de los objetos representen lo mejor posible la igualdad $\{d_{ij,s}\} = \{\delta_{ij,s}\}$.

ALGORITMO DE ESCALAMIENTO DE DIFERENCIAS INDIVIDUALES (INDSCAL)[1]

Como en el caso del escalamiento métrico del capítulo 2, las disimilaridades $\{\delta_{ij,s}\}$ se convierten en distancias estimadas $\{d_{ij,s}\}$ y la x_{it} y w_{st} se encuentran por mínimos cuadrados.

Las distancias asociadas a cada uno de los individuos se centran doblemente, obteniendo las matrices $\mathbf{B}_s$, donde

$$[\mathbf{B}_s]_{ij} = b_{ij,s} = \sum_{t=1}^{p} w_{st} x_{it} x_{jt}$$

$$= -\frac{1}{2}\left(d_{ij,s}^2 - \frac{1}{n}\sum_{i=1}^{N} d_{ij,s}^2 - \frac{1}{n}\sum_{j=1}^{N} d_{ij,s}^2 + \frac{1}{n^2}\sum_{i=1}^{N}\sum_{j=1}^{N} d_{ij,s}^2 \right) = \mathbf{H}\mathbf{A}_s\mathbf{H}$$

Donde $[\mathbf{A}_s]_{ij} = a_{ij,s} = -\frac{1}{2} d_{ij,s}^2$. Las estimaciones de mínimos cuadrados de $\{w_{st}\}$ y $\{x_{it}\}$ se encuentran minimizando:

$$S = \sum_{i,j,s}\left(b_{ij,s} - \sum_{t=1}^{p} w_{st} x_{it} x_{jt} \right)^2$$

El algoritmo de Carroll y Chang emplea mínimos cuadrados recursivos. En la ecuación anterior se agregan los índices L y R para marcar la diferencia entre unas coordenadas y otras. Primero, la función anterior se minimiza con respecto a los pesos $\{w_{st}\}$, mantenido $\{x^L_{it}\}$ y de $\{x^R_{jt}\}$ fijos.

$$S = \sum_{i,j,s}\left(b_{ij,s} - \sum_{t=1}^{p} w_{st} x^L_{it} x^R_{jt} \right)^2$$

Si $\{x^L_{it} x^R_{jt}\}$ forman una matriz $\mathbf{G}$ de n^2 x p, donde $[\mathbf{G}]_{\alpha t} = x^L_{it} x^R_{jt}$ con α=n (i-1) + j y $\{b_{ij,s}\}$ forma una matriz $\mathbf{F}$ de $N \times n^2$, donde $[\mathbf{F}]_{s\alpha} = b_{ij,s}$. Sea W una matriz de $N \times p$, donde $[\mathbf{W}]_{st} = w_{st}$. Entonces la estimación de mínimos cuadrados de W está dada por :

$$\hat{\mathbf{W}} = \mathbf{F}\mathbf{G}(\mathbf{G}^T\mathbf{G})^{-1}$$

[1]INDSCAL significa en inglés, INdividual Differences SCALing, en español, escalamiento de diferencias individuales. Se empleará el término INDSCAL en lo subsiguiente.

Después, se encuentran los valores estimados por mínimos cuadrados de $\{x^L_{it}\}$ para $\{w_{st}\}$ y $\{x^R_{jt}\}$ fijos.
Sea $\mathbf{G}$ una matriz de *Nn x p*, donde $[\mathbf{G}]_{\alpha t} = w_{st} x^R_{jt}$ y $\alpha = n(s-1)+j$.
Sea $\mathbf{F}$ una matriz de *n x Nn*, donde $[\mathbf{F}]_{\alpha\beta} = b_{ij,s}$, donde $\alpha = i$, $\beta = n(s-1)+j$. Sea $\mathbf{X}^L$ una matriz de tamaño n x p, donde $[\mathbf{X}^L]_{it} = x^L_{it}$. Entonces la estimación de mínimos cuadrados de $\mathbf{X}^L$ es

$$\hat{\mathbf{X}}^L = \mathbf{FG}(\mathbf{G}^T\mathbf{G})^{-1}.$$

De la misma manera, se encuentra la estimación de mínimos cuadrados de $\mathbf{X}^R$ manteniendo $\{w_{st}\}$ y $\{x^L_{it}\}$ fijos.
El proceso se repite hasta encontrar la convergencia de $\hat{\mathbf{X}}^L$ y $\hat{\mathbf{X}}^R$. Esta convergencia se asegura si:

$$\mathbf{X}^L = \mathbf{X}^R\mathbf{C},$$

donde $\mathbf{C}$ es una matriz diagonal de tamaño p x p con entradas no negativas. La convergencia se asegura si $\mathbf{X}^L = \mathbf{X}^R\mathbf{C}$, ya que:

$\sum_{t=1}^{p} w_{st} x_{it} x_{jt} = \sum_{t=1}^{p} (w_{st}/c_t) x^L_{it} (x_{jt} c_t)$ y por lo tanto, el mínimo de la suma de los cuadrados no es afectado por $\{c_t\}$.

El paso final es igualar $\hat{\mathbf{X}}^L = \hat{\mathbf{X}}^R$ y de esta manea calcular $\hat{W}$.

CAPÍTULO 5

APLICACIONES

CAPÍTULO 5

APLICACIONES

El problema de asignación de niveles socioeconómicos a hogares como una medida de segmentación de la población no deja de preocupar a la mayoría de las agencias de investigación de mercados en nuestro país. La mayoría de los estudios de consumo y de bienestar social buscan una segmentación de su población por nivel de ingreso y de esta manera tomar decisiones más acertadas acerca del perfil de sus consumidores o beneficiarios. Sin embargo, la variable ingreso en Latinoamérica no es una información confiable, debido al gran sesgo que esta variable puede arrojar al ser cuestionada a un hogar. De esta manera, la variable nivel socioeconómico es una alternativa a la variable ingreso que permite clasificar a los hogares o a sus integrantes con menor error que la variable ingreso.

Debe considerarse que el concepto de nivel socioeconómico no tiene una definición precisa, sin embargo, se busca analizar la variable nivel socioeconómico como una característica del hogar que permita discriminar de alguna manera más "económica" que "social" a los hogares de nuestro país. El nivel socioeconómico que se asigna a un hogar es el resultado de la evaluación del hogar en base a ciertas características fáciles de medir.

En este capítulo se busca analizar el problema de asignación de niveles socioeconómicos a hogares mediante las técnicas de escalamiento multidimensional y de esta manera proponer a las técnicas de escalamiento multidimensional como un método complementario en la selección de variables y valuación de resultados obtenidos con otras técnicas multivariadas para obtener una asignación más acertada de niveles socioeconómicos.

5.1 NIVEL SOCIOECONÓMICO

La AMAI[1] es la primera y única asociación de Investigación de Mercado y Opinión Pública en México. AMAI se fundó en septiembre de 1992 y su principal objetivo es establecer códigos de ética y estándares de calidad que promuevan la excelencia profesional en los servicios ofrecidos por las agencias de investigación.

Una de las cuestiones más preocupantes para AMAI desde su creación ha sido la homogeneización de los Niveles socioeconómicos (NSE). Indudablemente la principal variable de investigación de mercados es el Nivel Socioeconómico. Anteriormente a la creación de AMAI, todas las agencias tenían diferentes definiciones y criterios de medición para clasificar esta variable. La AMAI se encargó entonces de unificar criterios para lograr una compatibilidad en los datos y estudios de las agencias.

Así, el primer proyecto metodológico AMAI fue la creación de un índice para definir y clasificar el NSE.

El objetivo principal del estudio de NSE en México fue el de obtener un método para clasificar, a través de un índice, a cualquier hogar urbano (hogares en poblaciones de más de 50,000 habitantes) de la República Mexicana en un NSE específico.

Para ello, fue necesario encontrar un conjunto de variables que permitieran la mejor discriminación posible entre los distintos niveles socioeconómicos; así como estimar el tamaño relativo de cada NSE y ejemplificar con resultados su conformación.

Las características que busca AMAI de las variables para que permitieran la mejor discriminación posible entre los distintos niveles socioeconómicos son:

1. Las variables deben estar relacionadas con aspectos de la vivienda, de la posesión de durables, y sociales del jefe de familia.
2. Ser fácilmente medibles
3. Poder ser aplicadas a respondientes de cualquier edad y escolaridad
4. No depender de la observación física de la vivienda, de tal manera que puedan ser aplicadas tanto en entrevistas por intercepción como telefónicas.
5. Ser válidas (tener sentido) en cualquier contexto urbano (localidades de 50,000 y más habitantes)
6. Ser poco intrusivas (para no crear desconfianza en el respondiente y minimizar la no respuesta)

[1]AMAI: Asociación Mexicana de Agencias de Investigación de Mercado y Opinión Pública

Desde 1994 a la fecha, la AMAI, a través de su comité de NSE, ha venido haciendo estudios y pruebas que condujeron a generar dos diferentes reglas: La primera, que fue anunciada en 1994[1] y que prevaleció hasta mediados de 1997, estaba basada en una **combinación lineal**[2] de variables cuyo resultado final se expresaba en un cierto porcentaje para cada caso y dependiendo de puntaje alcanzado el caso se asignaba a un nivel socioeconómico particular. A esta forma de regla de asignación se le conoció como el Índice de Niveles Socioeconómicos AMAI. La segunda, que fue anunciada públicamente durante el seminario de actualización profesional AMAI en 1997[3] adoptó la forma de un **árbol de decisión**, construido a partir de la utilización de la técnica estadística llamada CHAID[4].

El siguiente cuadro sintetiza algunos de los aspectos más relevantes difundidos en las ponencias presentadas por el Comité de Niveles Socioeconómicos durante los pasados seminarios:

	1994	1995	1996	1997	1998	1999	2000	2001	2002
Enfoque de Regla de Asignación	Combinación Lineal de Variables			Árbol de decisión					
Herramienta estadística	Discriminante Regresión múltiple Conglomerados			CHAID					
Variables predictoras	Último año de estudios del Jefe de Familia								
	Nivel de mando del jefe de Familia								
	Número de focos								
	Número de habitaciones (sin incluir baño)								
	Número de sirvientes								
	Aspiradora								
	Tostador de pan								
	Boiler								
	Tv's color								
				Número de baños con regadera dentro de la vivienda					
				Posesión de autos en la vivienda (sean propios o no)					
				Tipo de piso					
				Computadora					
				Videocassettera					
				Horno de microondas					
				Lavadora de ropa					

A partir de 1998, la AMAI considera 2 reglas para la determinación de los niveles socioeconómicos (NSE): La regla 6x4 (6 variables para 4 niveles socioeconómicos) y la regla 13x6 (13 variables para 6 niveles socioeconómicos).

[1]Ver ponencia titulada: *Estudio de Niveles Socioeconómicos en México. Primera versión.* Comité de niveles socioeconómicos, AMAI. Memoria del I Seminario de Actualización Profesional. Junio, 1994.
[2]Análisis resultante de un Análisis de Discriminate canónico, correlaciones y regresiones con la variable ingreso como variable dependiente.
[3]Ver ponencia titulada: *Un Nuevo Enfoque en la Determinación de Niveles Socioeconómicos.* Comité de niveles socioeconómicos, AMAI. Memoria del IV Seminario de Actualización Profesional. Agosto, 1997.
[4]Chi-squared Automatic Interaction Detector. Es uno de los métodos más antiguos de árboles de clasificación basado en pruebas ji-cuadrada para hacer cortes

Las variables que actualmente considera AMAI son las siguientes:

Variables	Regla 6x4	Regla 13x6
Número de baños completos (numbanc)	●	●
Escolaridad del jefe de familia (edujef)	●	●
Número de focos (focos)	●	●
Número de piezas (npieza)	●	●
Video (vcr)		●
Tostador de pan (tosta)		●
Microondas (micro)		●
Número de automóviles (nauto)	●	●
Lavadora programable (lavarropa)		●
Tipo de piso (piso)		●
Boiler (boiler)	●	●
Aspiradora (aspirado)		●
Computadora (pc)		●

Los niveles socioeconómicos definidos por AMAI son los siguientes:

NSE	Descripción	Regla 6x4	Regla 13x6
Nivel A/B	Población con el más alto nivel de vida e ingresos del país.		●
Nivel C+	Población con ingresos o nivel de vida ligeramente superior al medio.		●
Nivel A/B/C+	Población con el nivel de vida e ingresos superior al medio.	●	
Nivel C	Población con ingresos o nivel de vida medio.	●	●
Nivel D+	Población con ingresos o nivel de vida ligeramente por debajo del nivel medio.	●	●
Nivel D	Población con un nivel de vida austero y bajos ingresos.		●
Nivel E	Población con menores ingresos y nivel de vida de las zonas urbanas de todo el país.		●
Nivel D/E	Población con menores ingresos y nivel de vida austero de las zonas urbanas de todo el país.	●	

5.2. EXPLORACIÓN DE NIVELES SOCIOECONÓMICOS DE ACUERDO A LA REGLA AMAI 13X6

La regla más empleada por la mayoría de las agencias de investigación de mercados para la asignación de niveles socioeconómicos es la Regla AMAI 13x6, debido a sus posibles agrupaciones, tales como Alta=A/B, Media=C+/C y Baja=D+/D/E ó Alta=AB/C+, Media Alta=C, Media Baja=D+ y Baja=D/E.

En este apartado se pretende visualizar a los niveles socioeconómicos calculados por la regla AMAI 13x6, mediante diferentes técnicas de escalamiento multidimensional considerando una muestra de 600 hogares (100 hogares por nivel socioeconómico) con información acerca de las 13 variables de la Regla AMAI 13x6.

Se calcularon los promedios por variable para cada nivel socioeconómico obteniendo la tabla 5.2.1 siguiente:

Tabla 5.2.1

NSE	numbanc	focos	npleza	edujef	nauto	lavaropa	tosta	boiler	vcr	micro	aspirado	pc	piso
A/B	2.8	28.41	7.72	10.77	2.24	0.79	0.87	1	0.95	0.93	0.8	0.73	1
C+	1.46	12.83	5.57	9.09	1.48	0.72	0.63	0.94	0.88	0.75	0.3	0.42	1
C	1.25	9.62	5.34	6.45	1	0.49	0.44	0.9	0.73	0.42	0.09	0.15	1
D+	1.23	7.83	4.31	3.98	0.29	0.22	0.13	0.81	0.42	0.13	0.03	0.08	0.98
D	0.4	5.09	3.1	3.52	0.13	0.02	0.05	0.11	0.24	0.04	0.01	0	1
E	0	2.55	1.74	2.42	0.02	0.01	0.01	0	0.11	0.02	0	0	0.65

Como se puede observar en la tabla anterior, los promedios más altos se obtuvieron para el nivel A/B y luego para el nivel C+ y así consecutivamente hasta el nivel E con los promedios más bajos para todas las variables excepto la variable piso, en donde el nivel D tiene mayor promedio que el nivel D+.

Para tener una visualización de la diferenciación entre los niveles que resuman la información de la tabla 5.2.1 se calculó una matriz de distancias euclideanas ponderadas[1] de la siguiente forma:

$$d_{ij} = \left(\sum_{k=1}^{13} \frac{1}{\max(x_k)^2} (x_{ik} - x_{jk})^2 \right)^{1/2}, \quad i,j = 1,\ldots,6 \text{ y } k = 1,\ldots 13$$

De esta forma, se obtiene la matriz de distancias euclideanas ponderadas tabla 5.2.2

Tabla 5.2.2

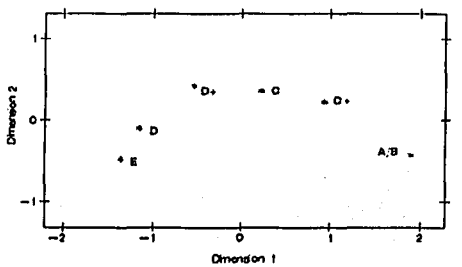
NSET	A/B	C+	C	D+	D	E
A/B	0					
C+	1.2	0				
C	1.86	0.79	0			
D+	2.5	1.53	0.80	0		
D	3.03	2.09	1.45	0.87	0	
E	3.29	2.35	1.74	1.17	0.48	0

[1]Debido a que la información de la tabla 5.2.1 es muy sensible a la escala de medición de cada una de las variables, la distancia euclideana ponderada resulta más apropiada.

En la matriz de distancias de la tabla 5.2.2 se pueden apreciar las diferencias entre los niveles socioeconómicos, donde la diferencia más significativa se encuentra entre el nivel A/B y el nivel E. Para facilitar la interpretación de la tabla 5.2.2, se realizó un escalamiento multidimensional métrico clásico obteniendo la siguiente configuración (Figura 5.2.1):

En la figura siguiente (figura 5.2.1) se muestran graficados los niveles socioeconómicos de acuerdo a la tabla de distancias 5.2.2 empleando un escalamiento métrico clásico[1] y obteniendo un stress S=.0235. El programa SAS y los resultados se encuentran en el anexo, apartado I.1.1.

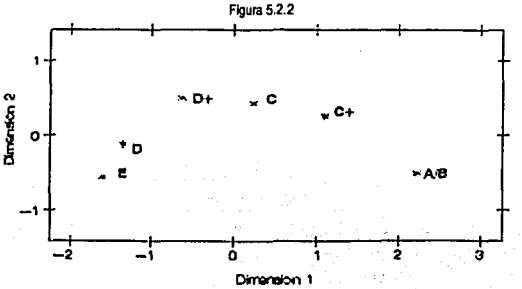
Figura 5.2.1



Como se puede apreciar en la configuración anterior, la menor distancia se encuentra entre los niveles D y E, debiéndose probablemente a que basándose en la información de cada variable para esos niveles, había menos diferencia que con respecto a la información de las variables de los demás niveles socioeconómicos.

Empleando el algoritmo de escalamiento métrico por mínimos cuadrados con transformaciones de la forma: $f(\delta_{ij})=b\,\delta_{ij}$, se obtuvo la configuración de la figura 5.2.2 siguiente.

Figura 5.2.2

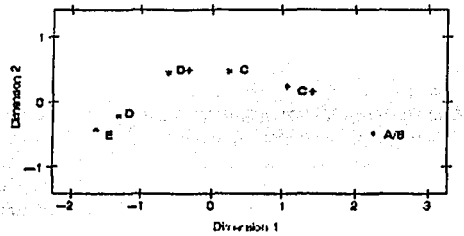


[1]Los datos de la tabla 5.2.2 son de la forma 2-factores, 2-modos 6x6

En la configuración 5.2.2 se obtuvo un stress S= 0.177 , mayor error que en la primera configuración (figura 5.2.1), sin embargo, la configuración resultante 5.2.2 es muy similar a la obtenida en la figura 5.2.1. los resultados se pueden consultar en el anexo , apartado I.1.2.

Empleando el algoritmo kruskal para de escalamiento no métrico, se obtuvo la configuración de la figura 5.2.3 siguiente.

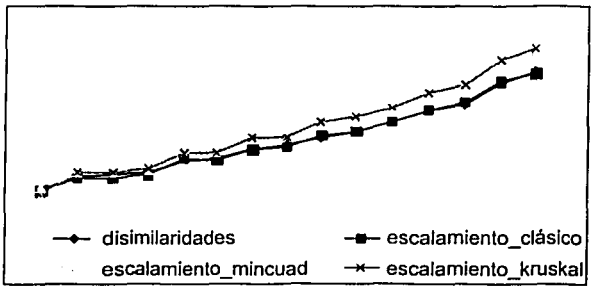
Figura 5.2.3



El Stress de la configuración anterior fue S= .178 y los resultados se pueden consultar en el anexo, apartado A.1.3.

Con el objetivo de comparar las tres configuraciones anteriores, se graficaron las distancias obtenidas con cada uno de los tres escalamientos anteriores y las disimilaridades originales de la tabla 5.2.2 (Figura 5.2.4) para ver cual fue el mejor ajuste.

Figura 5.2.4



Al observar la figura 5.2.4, se aprecia que el mejor ajuste se obtuvo con el escalamiento multidimensional métrico clásico y luego con el escalamiento de mínimos cuadrados, sin embargo, todas las configuraciones anteriores proporcionaron configuraciones similares, de las cuales se puede concluir que los niveles que presentan menor diferencia son los niveles D y E.

5.3 ANÁLISIS ESTRUCTURAL DE LAS REGLAS AMAI

En este apartado se pretende probar mediante una visualización gráfica de hogares, si el criterio que AMAI emplea para asignar niveles socioeconómicos es apropiada para los hogares de una muestra de la Ciudad de México[1]. Para obtener la representación gráfica se emplea la técnica de diferencias individuales.

5.3.1. REGLA 13X6

En el análisis de la aplicación anterior, se calcularon los promedios de los hogares para generar la tabla 5.1.1, sin embargo, las gráficas resultantes no proporcionan mucha información para su análisis, además de que al promediar la información de los hogares estamos perdiendo mucha información. En este apartado se busca graficar una muestra de 10 hogares por nivel socioeconómico (60 observaciones) de tal manera que las distancias entre hogares en la representación gráfica se parezcan lo mejor posible a las disimilaridades obtenidas entre hogares. Los hogares se van a etiquetar con su nivel socioeconómico calculado con la regla 13x6 de AMAI y se esperaría encontrar grupos homogéneos si la regla AMAI es correcta para esta muestra.

Las variables que emplea AMAI son variables ordinales y binomiales, por lo que se va a calcular una matriz de similaridades entre las parejas de hogares por cada una de las 13 variables. Visto de esta manera, el insumo para el análisis de escalamiento es de la forma: 3-factores, 2-modos, 13x60x60.

Las variables ordinales son transformadas en variables categóricas de la siguiente forma:
Número de baños completos (numbanc):

Numbanc0:Los hogares con 0 baños completos
Numbanc1:Los hogares con 1 baño completo
Numbanc2:Los hogares con 2 baños completos
Numbanc3:Los hogares con más de 2 baños completos

Número de focos (focos):

Focos1: Los hogares con 1 a 3 focos
Focos2: Los hogares con 4 a 10 focos
Focos3: Los hogares con 11 a 30 focos
Focos4: Los hogares con más de 30 focos

Número de piezas (npieza):

Npieza1: Los hogares con 1 o 2 piezas
Npieza2: Los hogares con 3 piezas
Npieza3: Los hogares con 4 o 5 piezas
Npieza4: Los hogares con 6 o 7 piezas
Npieza5: Los hogares con más de 7 piezas

[1]La muestra empleada no fue calculada para ser representativa de la Ciudad de México. El tamaño de muestra empleado fue seleccionado con fines prácticos.

Educación del jefe de familia (edujef):

- Edujef0: Los hogares con jefe de familia sin educación
- Edujef1: Los hogares con educación del jefe de familia primaria incompleta o terminada.
- Edujef2: Los hogares con educación del jefe de familia secundaria incompleta o terminada.
- Edujef3: Los hogares con educación del jefe de familia carrera comercial o técnica.
- Edujef4: Los hogares con educación del jefe de familia preparatoria incompleta o terminada.
- Edujef5: Los hogares con educación del jefe de familia universidad incompleta o terminada.
- Edujef6: Los hogares con educación del jefe de familia diplomado, maestría o doctorado

Número de automóviles (nauto):

- Nauto0: Hogares sin auto
- Nauto1: Hogares con 1 auto
- Nauto2: Hogares con 2 autos
- Nauto3: Hogares con más de 2 autos

Para las variables anteriores, se emplearon las medidas de similaridad (para variables ordinales) sugeridas en la Pág.10 del capítulo 1, quedando de la siguiente forma:

	numbanc0	numban1	numbanc2	numbanc3
numbanc0	1			
numbanc1	0.667	1		
numbanc2	0.333	0.667	1	
numbanc3	0.000	0.333	0.667	1

La matriz de similaridades anterior es para la variable numbanc, sin embargo, las variables focos y nauto, emplean la misma matriz de similaridades ya que tienen el mismo número de categorías.

	npieza1	npieza2	npieza3	npieza4	npieza5
npieza1	1				
npieza2	0.75	1			
npieza3	0.5	0.75	1		
npieza4	0.25	0.5	0.75	1	
npieza5	0	0.25	0.5	0.75	1

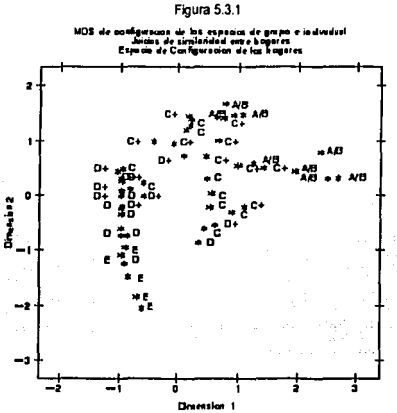
	edujef0	edujef1	edujef2	edujef3	edujef4	edujef5	edujef6
edujef0	1						
edujef1	0.83	1					
edujef2	0.67	0.83	1				
edujef3	0.50	0.67	0.83	1			
edujef4	0.33	0.50	0.67	0.83	1		
edujef5	0.17	0.33	0.50	0.67	0.83	1	
edujef6	0.00	0.17	0.33	0.50	0.67	0.83	1

Para las variable binomiales: boiler, aspirado, vcr, micro, piso, pc, lavoropa y tosta, la medida de similaridad entre el hogar i-ésimo y el hogar j-ésimo es 1 si comparten la misma categoría y 0 si no la comparten.

De esta manera, se calcula una matriz de similaridad para cada una de las variables. Lo que se busca es una configuración de los hogares de acuerdo al peso que tiene cada una de las variables. La técnica de escalamiento multidimensional que se va a emplear es la técnica de diferencias individuales con el algoritmo ALSCAL, el programa SAS con el que se ejecutó esta aplicación está en el anexo apartado I.2.1.

El espacio de configuración de los hogares, considerando el peso que tiene cada una de las variables es la figura 5.3.1 siguiente:

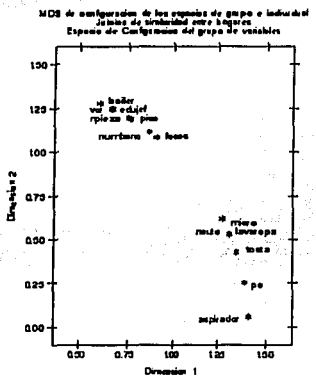
Figura 5.3.1



En la figura anterior se graficaron a los hogares, de acuerdo a cada una de las matrices de similaridades correspondientes a cada variable y se etiquetó cada punto con el nivel socioeconómico calculado con la regla de AMAI. Los grupos que se pueden apreciar son:
{E}, {D, D+} {C+, A/B}; los hogares {C} no se distinguen con un grupo, están con D+ y C+. Si la regla de AMAI aplicara perfectamente a los datos, entonces se tendrían los 6 grupos de niveles socioeconómicos perfectamente definidos.

En la figura 5.3.2, se muestra la configuración obtenida para las variables:

Figura 5.3.2

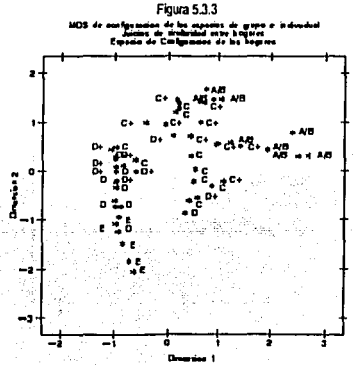


En la figura anterior (Figura 5.3.2), se aprecian los siguientes grupos de variables: {boiler, vcr, edujef, npieza, piso}, {numbanc, focos}, {micro, nauto, lavaropa, tosta}, y {pc, aspirador}. La configuración anterior son los pesos a cada una de las dos dimensiones que cada variable aporta a la configuración de hogares. Los grupos anteriormente formados implican que los pesos de esas variables a la configuración final son muy parecidos.

En la configuración de la figura 5.3.1, se obtuvo un Stress de .24 y las variables que tuvieron mejor ajuste en la representación fueron: aspirador, boiler, focos, lavarropa, micro, nauto y tosta. Los resultados se pueden consultar en el anexo, apartado I.2.1.

Repitiendo el análisis pero con las variables con mejor ajuste, se obtuvo la siguiente configuración:

Figura 5.3.3



En la configuración de la figura 5.3.3, se obtuvo un mejor stress de .13; los resultados se pueden consultar en el anexo, apartado I.2.2

Se puede concluir de las representaciones anteriores que de las 13 variables que emplea AMAI en su regla para asignar 6 niveles socioeconómicos, se pueden desechar 6 variables y únicamente con las variables boiler, focos, nauto, micro, lavarropa, tosta y aspirador, se obtendría la misma clasificación de niveles socioeconómicos.

5.3.2. REGLA 6X4

Al igual que en la aplicación anterior (apartado 5.3.1) , se quiere probar si el criterio de la regla 6x4 que AMAI emplea para asignar niveles socioeconómicos es apropiado para los hogares que se tienen de muestra.

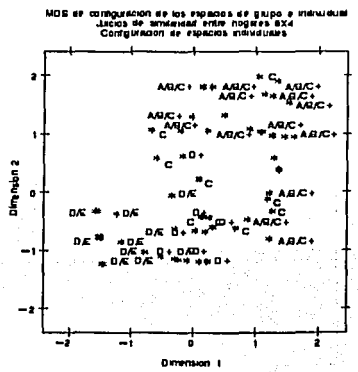
En esta aplicación, se emplea la misma muestra de hogares de la aplicación anterior y las 6 variables de la regla 6x4 (numbac, edujef, focos, npieza, nauto y boiler), transformando las variables de la misma manera que en la aplicación 5.3.1.

El insumo para el análisis de escalamiento es de la forma: 3-factores, 2-modos, 6x60x60.

La técnica de escalamiento multidimensional que se va a emplear es la técnica de diferencias individuales con el algoritmo ALSCAL, el programa SAS con el que se ejecutó esta aplicación está en el anexo, apartado I.2.3

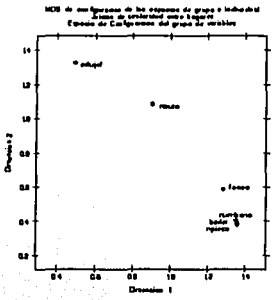
El espacio de configuración de los hogares, considerando el peso que tiene cada una de las variables es la figura 5.4.1 siguiente:

Figura 5.3.4



En la figura anterior se graficaron a los hogares, de acuerdo a cada una de las matrices de similaridades correspondientes a cada variable y se etiquetó cada punto con el nivel socioeconómico calculado con la regla 6x4 de AMAI. Se puede observar que los niveles A/B/C+ y D/E forman grupos separados uno del otro y los niveles C y D+ se encuentran mezclados con los otros dos grupos.

En la figura 5.3.4 se muestra la configuración obtenida para las variables



En la configuración de la figura 5.3.4, se obtuvo un Stress de .20 y las variables que tuvieron peor ajuste en la representación fueron: npieza y nauto. Los resultados se pueden consultar en el anexo, apartado I.2.3.

5.4 PROPUESTA DE ASIGNACIÓN DE NIVELES SOCIOECONÓMICOS

En esta aplicación se quiere formar grupos de niveles socioeconómicos basándose en una serie de variables (variables AMAI y otras) para establecer así un criterio en la asignación de niveles socioeconómicos.

Las variables que se van a emplear en esta aplicación son las variables AMAI excepto la variable piso, en vez de esa variable se va a emplear la variable tipopiso que a continuación se describirá y las descritas a continuación:

Tipopiso: 0 si el piso es de tierra, 1 si el piso es de cemento y 2 si el piso es otro acabado

Agua: 1 si tiene agua dentro de la vivienda y 0 si no es así

Sirv: Número total de sirvientes, ya sea de entrada por salida o no.

Videocam: Número de videocámaras con las que cuentan en la vivienda

Modular: 1 si en el hogar tienen modular y 0 si no.

Batidora: 1 si tiene batidora, 0 si no tiene batidora.

Secaropa: 1 si tienen secadora de ropa y 0 si no

Celular: 1 si tiene celular algún miembro del hogar o no.

Tvsign:1 si en el hogar cuentan con tv por cable, multivisión, sky, directv o parabólica.

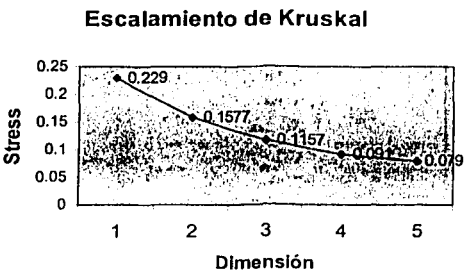
Internet: 1 si en el hogar cuentan con conexión a internet y 0 si no.

En esta aplicación se tomó una muestra de 120 hogares de la Ciudad de México[1], con información de cada una de las variables descritas anteriormente. El primer paso consiste en encontrar una medida de proximidad entre todos los hogares en base a todas la variables. Se estandarizaron las variables, de manera que el valor máximo que pueda tomar cada variable sea 1. De esta manera, se calcula una matriz de 120x120 de distancias euclideanas entre los hogares y se realizó un escalamiento multidimensional no métrico de kruskal.

[1]La muestra empleada no fue calculada para ser representativa de la Ciudad de México. El tamaño de muestra empleado fue seleccionado con fines prácticos.

En la siguiente figura (figura 5.4.1) se muestran los niveles de Stress obtenidos para diferentes dimensiones:

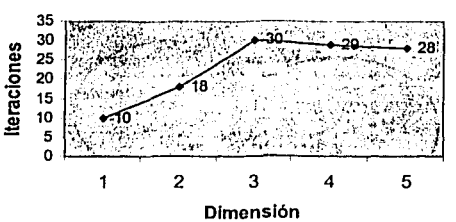
Figura 5.4.1



Un escalamiento en dos dimensiones tendría un Stress de .1577, lo que significa que es un error aceptable.

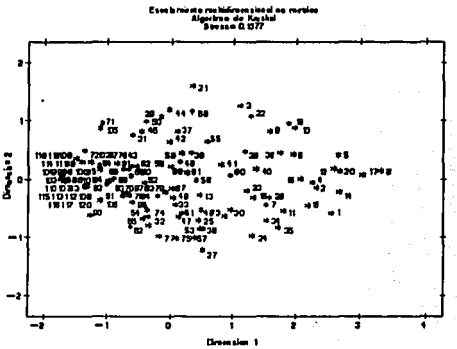
En la siguiente figura (figura 5.4.2) se muestra el número de iteraciones en la ejecución del algoritmo de kruskal para cada dimensión.

Figura 5.4.2



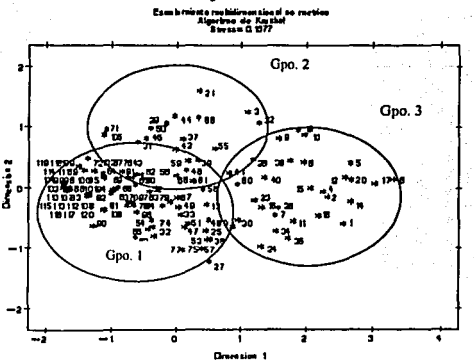
La configuración resultante en dos dimensiones con el escalamiento de kruskal se muestra en la figura 5.4.3 siguiente:

Figura 5.4.3



En la figura anterior se pueden apreciar tres grupos (figura 5.4.4)

Figura 5.4.4

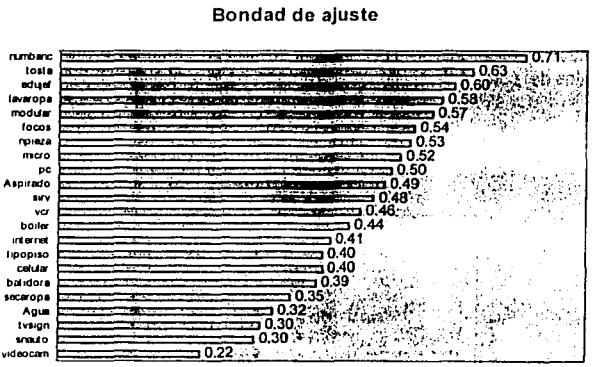


El número de variables que se emplearon para obtener la configuración anterior fue de 22, es probable que algunas variables no proporcionan una discriminación para los hogares. Una observación interesante se obtendría si se conoce la relación que tiene cada una de las variables con la configuración anteriormente obtenida.

La relación que tiene cada una de las variables con la configuración obtenida se encuentra por medio de una regresión multivariada, donde las coordenadas de los puntos de la configuración son las variables independientes y la variable es la variable dependiente.

La bondad de ajuste de la regresión para cada una de las variables fue la siguiente:

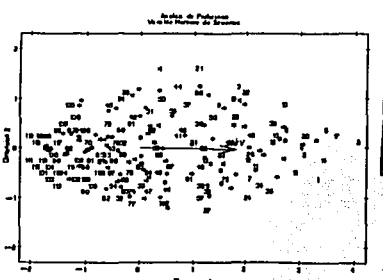
Figura 5.4.5



La medida de bondad de ajuste de la regresión para cada una de las variables es pobre para algunas variables. La variable videocam es la variable que tiene peor ajuste lineal para la configuración.

Para el caso de la variable sirv, la recta estimada fue: sirv = (0.12003)*dim1 – (0.00169)*dim2.
la recta de la variable sirv se va a proyectar en dos dimensiones. Las proyecciones se emplean para determinar el orden de preferencia con representación vectorial; las proyecciones son los puntos definidos por líneas perpendiculares del objeto al vector.

Figura 5.4.6



En la figura 5.4.6 se muestra el resultado de una regresión vectorial para la variable número de sirvientes en el hogar (sirv) . La dirección del vector indica que los hogares que se encuentran en la dirección del vector son los que tienen mayor número de sirivientes y los que se encuentra en dirección contraria al vector son los hogares que tiene menor número de sirvientes o que no tienen.

Se va a realizar un escalamiento no métrico de Kruskal para cada uno de los grupos de la figura 5.4.4, para poder indagar sobre esos grupos y ver si se pueden desagregar en más grupos.

Para el grupo 1 se obtuvo la configuración de la figura 5.4.7 con un Stress de 0.144 y en la figura 5.4.8. se muestran los coeficientes de bondad de ajuste de la regresión de la configuración 5.4.7 con cada una de las variables.

Figura 5.4.7

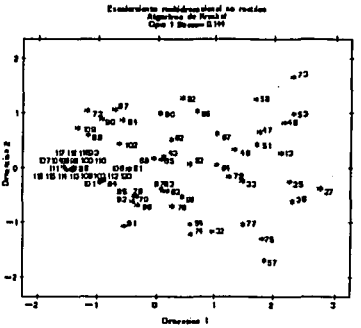
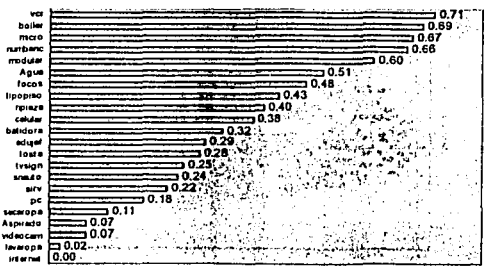


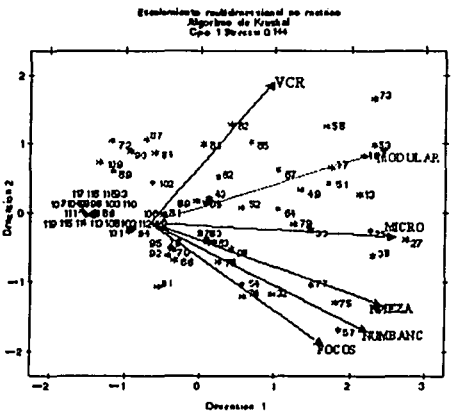
Figura 5.4.8



En la figura 5.4.8 se aprecia que las variables que no tienen buen ajuste con la configuración son: secaropa, aspirado, videocam, lavarropa e internet, esto ocurre porque ninguno o casi ninguno de los hogares tiene secadora de ropa, aspiradora, videocámara, lavadora de ropa ni internet; este grupo es el grupo de nivel socioeconómico bajo.

Para interpretar mejor la figura 5.4.7. se grafican los vectores correspondientes a las variables con mejor ajuste, resultantes de la regresión multivariada de las coordenadas de puntos como variables explicativas y la variable como explicada. En la figura 5.4.9 se muestran dichas relaciones.

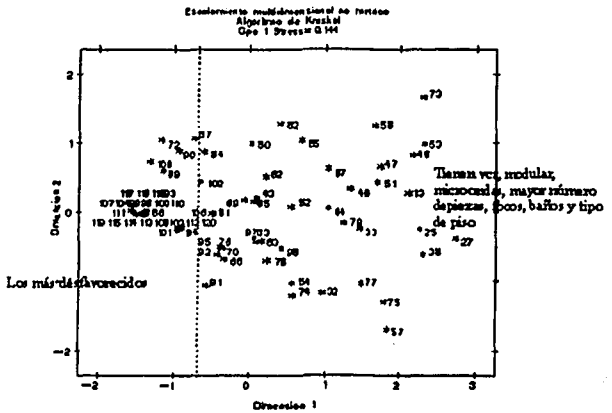
Figura 5.4.9
Análisis de preferencias



En la figura anterior no se graficaron los vectores correspondientes a las variables agua, boiler y tipopiso ya que tienen un comportamiento muy similar a la variable numbanc. En la parte derecha de la configuración se encuentran los hogares que tienen mayor número de focos, baños completos y piezas, algunos tienen microondas, modular, vcr, agua y boiler, Ninguno de estos hogares cuenta con intenet y muy pocos con lavadora de ropa, videocámara, aspiradora, secadora de ropa, pc y sirvientes.

En la figura 5.4.10, se muestra el resumen de la figura 5.4.9:

Figura 5.4.10



Para el grupo 2 se obtuvo la configuración de la figura 5.4.11 con un Stress de 0.2064 y en la figura 5.4.12. se muestran los coeficientes de bondad de ajuste de la regresión de la configuración 5.4.11 con cada una de las variables.

Figura 5.4.11

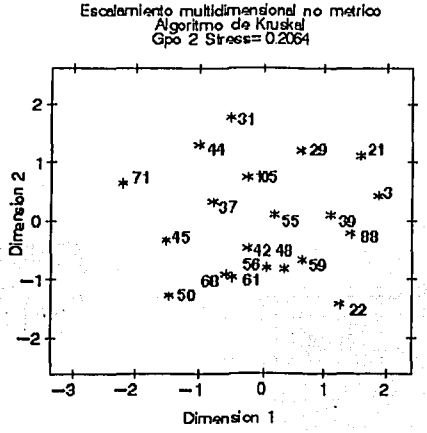
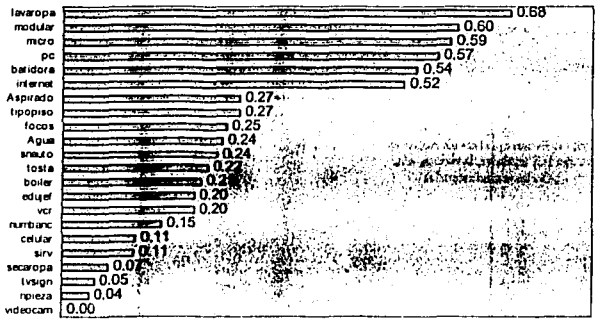


Figura 5.4.12

Bondad de ajuste Gpo2

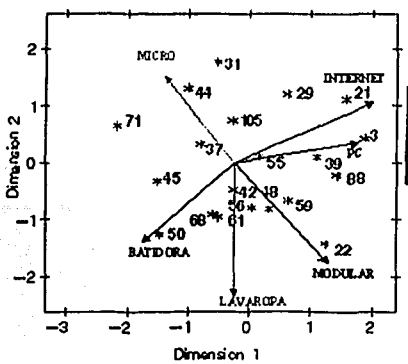


En la figura 5.4.12 la variable videocam no tiene correlación con la configuración ya que ningún hogar de este grupo cuenta con cámara de video. La variable npieza tiene muy baja bondad de ajuste ya que estos hogares varían de 2 a 10 piezas. Las variables tvsign, secaropa, sirv, celular también tienen poca bondad de ajuste ya que pocos hogares cuentan con estos servicios y en el caso de tvsign los hogares que tienen es este servicio tienen cable visión, parabólica o multivisión.

En la figura 5.4.13 se muestra la relación de las variables con mejor ajuste a la configuración de la figura 5.4.11.

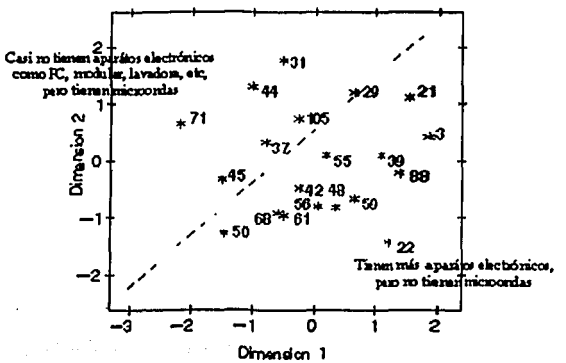
Figura 5.4.13

Escalamiento multidimensional no metrico
Algoritmo de Kruskal
Gpo 2 Stress= 0.2064



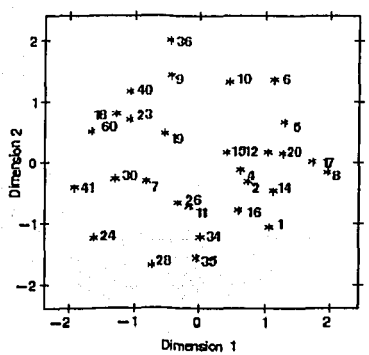
Como se puede apreciar en la figura anterior (figura 5.4.13) este grupo es menos homogéneo, en la siguiente figura (figura 5.4.14) se muestra el resumen de la figura 5.4.13.

Figura 5.4.14
Escalamiento multidimensional no metrico
Algoritmo de Kruskal
Gpo 2 Stress= 0.2064



Para el grupo 3 se obtuvo la configuración de la figura 5.4.15 con un Stress de 0.2217

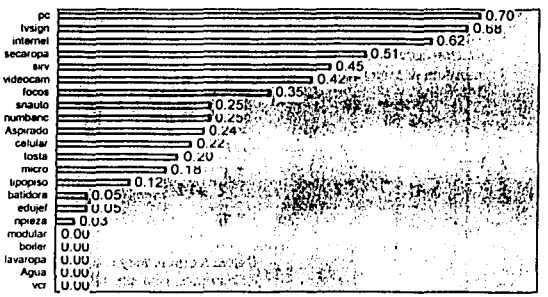
Figura 5.4.15
Escalamiento multidimensional no metrico
Algoritmo de Kruskal
Gpo 3 Stress= 0.2217



Las medidas de bondad de ajuste de la regresión calculada para las variables y la configuración anterior es la siguiente:

Figura 5.4.16

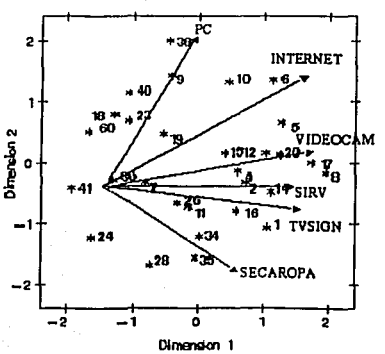
Bondad de ajuste Gpo 3



En este grupo de hogares, todos tienen agua, boiler, lavadora de ropa, vcr y modular. Estos hogares tienen de 4 a 16 piezas, casi todos tienen batidora, ninguno tiene tipo de piso tierra, la mayoría tiene microondas, tostador de pan y celular. La educación del jefe de familia es muy variable.

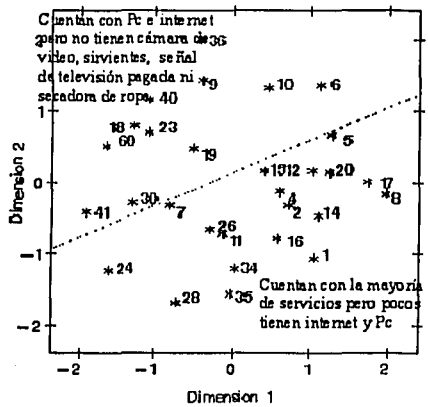
En la siguiente figura (figura 5.4.17) se muestra la relación de las variables con mejor ajuste a la configuración 5.4.15.

Figura 5.4.17
Escalamiento multidimensional no metrico
Algoritmo de Kruskal
Gpo 3 Stress= 0.227



En la siguiente figura se muestra el resumen de la figura anterior.

Figura 5.4.18
Escalamiento multidimensional no metrico
Algoritmo de Kruskal
Gpo 3 Stress= 0.2217



En la siguiente tabla se muestran los porcentajes de algunas de las variables para los tres grupos.

Tabla 5.4.1

Variable	Gpo1 Bajo	Gpo2 Medio	Gpo3 Alto
Agua	70%	95%	100%
Aspiradora	4%	35%	72%
Batidora	23%	40%	90%
Boiler	52%	90%	100%
Cámara de video	3%	0%	48%
Celular	13%	20%	66%
Computadora	4%	50%	62%
Internet	0%	20%	31%
Lavadora de ropa	4%	40%	100%
Microondas	28%	20%	79%
Modular	49%	50%	100%
Secadora de ropa	3%	15%	48%
Señal de tv pagada	11%	60%	55%
Totador de Pan	10%	10%	93%
Vcr	41%	85%	100%

En la tabla anterior se aprecia que las variables más discriminantes entre grupos son: agua, aspiradora, batidora, boiler, celular, computadora e internet, lavadora de ropa, secadora de ropa, señal de tv pagada y vcr.

En la siguiente tabla se muestran los promedios de algunas variables por grupo.

Tabla 5.4.2

Variable	Gpo1 Bajo	Gpo2 Medio	Gpo3 Alto
Tipo de piso	2.3	2.6	2.8
Promedio de baños completos	0.6	1.6	2.4
Promedio de piezas	3	5	7
Promedio de autos	0.4	0.8	1.4
promedio de focos	5	12	25
promedio de educación	4.1	6.8	10.0
promedio de sirvientes	0.1	0.4	1.1

En la tabla anterior (tabla 5.4.2) se aprecia que la variable tipo de piso no es muy discriminante entre grupos para esta muestra, esto puede deberse a que la muestra empleada se obtuvo de población urbana y este resultado hubiera sido diferente en el caso de población rural.

5.5 CONCLUSIONES

Como resultado obtenido en las diferentes aplicaciones mencionadas en este capítulo, se puede concluir que la regla de AMAI 6x4 y 13x6 no es tan clara para los hogares empleados en este estudio. Las variables piso y tostador de pan que emplea AMAI no es discriminante en los resultados de la aplicación 3.

Una de las variables que resultaron más discriminantes en la aplicación 3 fue Internet, donde se aprecia que en el grupo 1 ningún hogar cuenta con este servicio y muy pocos hogares tienen computadora.

Al hacer este análisis de asignación de niveles socioeconómicos, una de las variables que se desearía tener es el nivel de ingreso, sin embargo, la variable ingreso está sujeta a muchos errores sistemáticos al emplearla en una muestra de esta naturaleza, la variable ingreso es una variable declarada por lo que el entrevistado se puede encontrar en situaciones de desconfianza y está sujeta a no ser real en el momento de la entrevista. Sin embargo, las variables que se pueden relacionar mejor con el nivel de ingreso son las variables de presencia de algún artículo de mayor costo como computadora, señal pagada, lavadora y secadora de ropa, automóvil, aspiradora, sirvientes y celular, que son bienes costosos y no de primera necesidad.

La variable agua, que si bien es cierto, discrimina en cierta medida a los tres grupos de la aplicación 3, es un bien necesario que como se apreció en la aplicación 3 no correlaciona muy bien en los tres grupos.

Como consecuencia de las aplicaciones aquí presentadas, se sugiere emplear tres niveles socioeconómicos en vez de 4 o 6 como sugiere AMAI y eliminar la variable tostador de pan y piso y hasta cierto punto analizar a mayor detalle si la variable microondas discrimina o no. En vez de emplear esas variables se podría analizar la viabilidad de emplear computadora, señal de televisión pagada e Internet y número de sirvientes.

Los grupos encontrados en la aplicación 3 tienen las siguientes características:

Grupo 1: Son hogares con nivel de vida austero, con un nivel de educación de jefe de familia de secundaria, no cuentan con Internet ni televisión pagada, que habitan en casas pequeñas con un promedio de 3 piezas con un baño completo o con ninguno

Grupo 2: Son hogares con nivel de vida medio, con un nivel de educación de jefe de familia de preparatoria o licenciatura, que pueden contar con Internet y computadora, habitan en casas medianas con un promedio de 5 piezas y de 1 a 2 baños completos y cuentan con automóvil.

Grupo 3: Son hogares con nivel de vida alto, con un nivel de educación de licenciatura o postrado, tienen computadora y sus casas son grandes con 7 piezas en promedio y más de dos baños completos, tienen sirvientes y con uno o más automóviles.

Este trabajo no tiene por objetivo establecer reglas de asignación de niveles socioeconómicos a hogares, existen técnicas multivariadas más avanzadas como árboles de clasificación tales como la técnica C&ART (Classification and Regression Trees), CHAID, etc y análisis discriminante, entre otros que permiten analizar muestras mucho más grandes que las aquí presentadas. Sin embargo, siempre que se use alguna de estas técnicas es aconsejable emplear alguna técnica alternativa que permita una visualización de la información analizada y obtener conclusiones más robustas. En este trabajo se sugiere emplear las técnicas de escalamiento multidimensional como un método complementario a las técnicas empleadas para encontrar una regla de asignación de niveles socioeconómicos a hogares.

Las técnicas de escalamiento multidimensional permiten un análisis de la información mediante una inspección visual fácil de interpretar siempre y cuando el número de observaciones a analizar sea pequeño.

Para el problema de asignación de niveles socioeconómicos, se recomienda emplear éstas técnicas a una submuestra pequeña de hogares que validen la regla de asignación obtenido por alguna otra técnica.

CONCLUSIONES

CONCLUSIONES

Las técnicas de escalamiento multidimensional permiten una visualización de datos con muchos atributos en un espacio de pocas dimensiones que permiten al investigador obtener conclusiones sin necesidad de inferencia estadística.

En esta investigación se obtuvo por medio de inspección visual una serie de aplicaciones enfocadas al problema de asignación de niveles socioeconómicos a hogares mediante el empleo de escalamiento multidimensional.

La muestra empleada en esta investigación es pequeña, debido a que al emplear mayor observaciones para el análisis provocaría una inspección de los datos más difícil y con un mayor error o stress; las técnicas de escalamiento multidimensional permiten el empleo de numerosas variables, pero sus alcances en número de objetos a visualizar en una configuración es limitado.

En las aplicaciones se encontró que en la muestra empleada en este trabajo, las definiciones de AMAI de niveles socioeconómicos no son tan claras como se hubiera esperado, algunas de las variables que AMAI sugiere en sus reglas 6x4 y 13x6 no discriminan a los hogares de la muestra, como son piso y tostador de pan y se sugiere el empleo de otras variables como señal de televisión pagada, computadora, internet, entre otras.

En la última aplicación de este trabajo, se definieron 3 niveles socioeconómicos a hogares que se encontraron a partir de incluir más variables al análisis por medio de una inspección visual a una muestra de 120 hogares.

Actualmente, empresas de investigación de mercados emplean 3 niveles socioeconómicos en sus reportes a clientes, sin embargo, dichas empresas se basan en las definiciones y en las variables que AMAI sugiere. Dados los resultados en esta investigación se sugiere que se cuestione el empleo de dichas variables y definiciones, para encontrar una definición de estratos más heterogéneos entre sí y homogéneos dentro de sus grupos.

Los grupos encontrados en esta investigación no se observan muy definidos en la configuración resultante, y se tendría que hacer un análisis más detallado de aquellos puntos frontera que conviven entre grupos, sin embargo, los resultados obtenidos en cada uno de los grupos son bastante consistentes.

Para poder crear una teoría que tuviera un mayor alcance a una población en especial como Ciudad de México sería necesario entonces, emplear técnicas estadísticas alternativas o de inferencia y una muestra diseñada para ese análisis.

APÉNDICE

APÉNDICE

A. ALGUNOS RESULTADOS DE ÁLGEBRA LINEAL

A.1 Ecuaciones lineales homogéneas

Teorema 1.1 Sea $a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = 0$

$$\vdots \qquad\qquad \vdots$$

$$a_{p1}x_1 + a_{p2}x_2 + \ldots + a_{pn}x_n = 0$$

un sistema de p ecuaciones lineales con n incógnitas y supóngase que n>p. Entonces el sistema tiene una solución no trivial.

A.2 Eigendescomposición

Cualquier matriz cuadrada A de números reales puede descomponerse en el producto de varias matrices. Ahora se considera el caso particular de la eigendescomposición, que se puede construir para la mayoría de las matrices, pero siempre para las matices simétricas. formalmente:

$$\mathbf{A} = \mathbf{Q}\,\mathbf{\Lambda}\,\mathbf{Q}^T, \qquad \ldots(1)$$

con Q un vector ortonormal y $\mathbf{\Lambda}$ una matriz diagonal. La ecuación (1) generalmente se expresa de la siguiente manera:

$$\mathbf{A}\mathbf{q}_i = \lambda_i \mathbf{q}_i \text{ con } \mathbf{q}_i \neq 0 \ (\text{i=1,}\ldots\text{n})$$

una sucesión de vectores columna que pueden representarse en:

$$\mathbf{AQ} = \mathbf{Q}\,\mathbf{\Lambda}$$

los vectores $\mathbf{q}_i$ son llamados eigenvectores de A y las λ_i's los eigenvalores de A. Generalmente se ordenan los eigenvalores y sus respectivos eigenvectores de la siguiente manera: $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots \geq \lambda_n$.

A.3 Descomposición espectral

Sea A una matriz n x n simétrica, con eigenvalores $\{\lambda_i\}$ y eigenvectores asociados $\{\mathbf{q}_i\}$ tales que $\mathbf{q}_i^T\mathbf{q}_i = 1$. Entonces A puede expresarse como:

$$\mathbf{A} = \mathbf{Q}\,\Lambda\,\mathbf{Q}^T = \sum_{i=1}^{n}\lambda_i\mathbf{q}_i\mathbf{q}_i^T,$$

donde Λ =diag($\lambda_1,...,\lambda_n$), Q=[$\mathbf{q}_1,...,\mathbf{q}_n$].

La matriz $\mathbf{Q}$ es ortonormal, tal que $\mathbf{QQ}^T = \mathbf{Q}^T\mathbf{Q} = \mathbf{I}$. Si A es no singular $\mathbf{A}^m = \mathbf{Q}\,\Lambda^m\,\mathbf{Q}^T$ con Λ^m =diag($\lambda_1^m,...,\lambda_n^m$) para cualquier entero m.

A.4 Algunas propiedades de la descomposición espectral

1) No todas las matrices reales cuadradas poseen una eigendescomposición real, aunque los eigenvectores sean ortonormales.
2) Los eigenvectores no son únicos.
3) El rango de la matriz es igual al número de eigenvalores distintos de cero.
4) Si la matriz es simétrica, sus eigenvalores y eigenvectores son siempre reales.
5) Si la matriz es simétrica, sus eigenvectores son ortonormales.

A.5 Descomposición en valores singulares (SVD)

Si A es una matriz n x p con rango r, entonces A se puede expresar como

$$\mathbf{A} = \mathbf{U}\,\Lambda\,\mathbf{V}^{\mathrm{T}},$$

donde $\Lambda = diag(\lambda_1,\lambda_2,...,\lambda_r)$, con $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq ... \geq \lambda_r \geq 0$, $\mathbf{U}$ es una matriz ortonormal de n x r, y $\mathbf{V}$ es una matriz ortonormal de r x r, es decir, $\mathbf{VV}^{\mathrm{T}} = \mathbf{UU}^{\mathrm{T}} = \mathbf{I}$. El conjunto de valores $\{\lambda_i\}$ son llamados valores singulares de $\mathbf{A}$. Si $\mathbf{U}$ y $\mathbf{V}$ se expresan en términos de vectores columna, $\mathbf{U} = [\mathbf{u}_1,...,\mathbf{u}_r]$, $\mathbf{V} = [\mathbf{v}_1,...,\mathbf{v}_r]$, etonces $\{\mathbf{u}_i\}$ son los vectores singulars izquierdos de A y $\{\mathbf{v}_i\}$ son los vectores singulars derechos. La matriz A puede expresarse de la siguiente manera:

$$\mathbf{A} = \sum_{i=1}^{r}\lambda_i\mathbf{u}_i\mathbf{v}_i^T \quad .$$

Los eigenvalores $\{\lambda_i^2\}$ son los eigenvalores distintos de cero de la matriz simétrica $\mathbf{AA}^T$ y de $\mathbf{A}^T\mathbf{A}$. Los vectores $\{\mathbf{u}_i\}$ son los eigenvectores normalizados de $\mathbf{AA}^T$ y los vectores $\{\mathbf{v}_i\}$ son los eigenvectores normalizados de $\mathbf{A}^T\mathbf{A}$.

A.6 Inversa de Moore-Penrose

Considerando la ecuación matricial:

$$\mathbf{AX} = \mathbf{B},$$

donde A es una matriz n x p , X una matriz de p x n , y B una matriz de n x n. La matriz x que minimiza la suma de cuadrados $tr(\mathbf{AX}-\mathbf{B})^T tr(\mathbf{AX}-\mathbf{B})$, tiene el menor valor de de $tr(\mathbf{X}^T\mathbf{X})$ de entre todas las soluciones de mínimos cuadrados dada por:

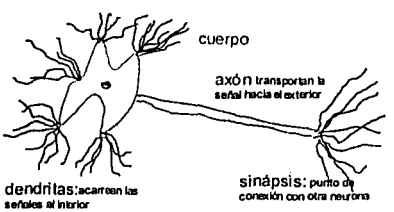
$$\mathbf{X} = \mathbf{A}^+\mathbf{B},$$

donde $\mathbf{A}^+$ es la inversa única generalizada de Moore-Penrose p x n de A, definida por las ecuaciones:

$$\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}$$
$$\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+$$
$$(\mathbf{A}\mathbf{A}^+)^* = \mathbf{A}\mathbf{A}^+$$
$$(\mathbf{A}^+\mathbf{A})^* = \mathbf{A}^+\mathbf{A}$$

B. ESCALAMIENTO CON REDES NEURONALES

Una neurona es una célula viva y, como tal, contiene los mismos elementos que forman parte de todas las células biológicas. En general, una neurona consta de un cuerpo celular de donde sale una rama principal llamada "axón" y varias ramas más cortas llamas "dendritas". A su vez, el axón puede producir ramas en torno a su punto de arranque, y con frecuencia se ramifica extensamente cerca de su extremo.



Una de las características que distinguen a las neuronas es su capacidad de comunicarse. En términos generales, las dendritas y el cuerpo celular reciben señales de entrada; el cuerpo celular las combina e integra señales de salida que a su vez, son transportadas por el axón a los terminales axónicos, que se encargan de distribuir información a un nuevo conjunto de neuronas. Por lo general, una neurona recibe información de miles de otras neuronas y, a su vez, envía información a miles de neuronas más.

Las redes neuronales artificiales son modelos que intentan reproducir el comportamiento del cerebro. Como tal modelo, realiza una simplificación, averiguando cuáles son los elementos relevantes del sistema, ya sea porque la cantidad de información de que se dispone es excesiva, o bien porque es redundante. Una elección adecuada de sus características, más una estructura conveniente, es el procedimiento convencional para construir redes capaces de realizar determinada tarea.

[1]Van Wezeel, Michael C et al., 2001

Generalmente, se pueden encontrar tres tipos de neuronas:

1) Aquellas que reciben estímulos externos, relacionados con el aparato sensorial, que tomarán la información de entrada. (*unidades de entrada*)
2) Dicha información se transmite a ciertos elementos internos que se ocupan de su procesamiento. Es en la sinapsis y neuronas correspondientes a este nivel donde se genera cualquier tipo de representación interna de la información. Puesto que no tiene relación directa con la información de entrada ni con la de salida, estos elementos se denominan *unidades ocultas*.
3) Una vez finalizado el período de procesamiento, la información llega a las *unidades de salida*, cuya misión es dar la respuesta del sistema.

Una *capa* o *nivel* es un conjunto de neuronas cuyas entradas provienen de la misma fuente y cuyas salidas se dirigen a un mismo destino.

Cada neurona i-ésima está caracterizada en cualquier instante por un valor numérico denominado *estado de activación* $a_i(t)$; asociado a cada unidad, existe una *función de salida*, f_i, que trasforma el estado actual de activación en una *señal de salida*, y_i. Dicha señal es enviada a través de los canales de comunicación unidireccionales a otras unidades de la red; en estos canales la señal se modifica de acuerdo con la sinapsis (*el peso*, w_{ji}) asociada a cada uno de ellos según una determinada regla. Las señales moduladas que han llegado a la unidad j-ésima se combinan entre ellas, generando así la *entrada total*, Net_j.

$$Net_j = \sum_i y_i w_{ji}$$

Una *función de activación* F, determina el nuevo estado de activación $a_j(t+1)$ de la neurona, teniendo en cuenta la entrada total calculada y el anterior estado de activación $a_j(t)$.

Existen numerosos modelos de redes neuronales diferenciados por las siguientes características:

a) *Topología*: La topología o arquitectura de las redes neuronales consiste en la organización y disposición de las neuronas en la red formando *capas* o agrupaciones de neuronas mas o menos alejadas de la entrada y salida de la red. En este sentido, los parámetros fundamentales de la red son: el número de capas, el número de neuronas por capa, el grado de conectividad y el tipo de conexiones entre neuronas.

b) *Mecanismo de aprendizaje:* El aprendizaje es el proceso por el cual una red neuronal modifica sus pesos en respuesta a una información de entrada. Los cambios que se producen durante el proceso de aprendizaje se reducen a la destrucción, modificación y creación de conexiones entre las neuronas. En los sistemas biológicos existe una continua creación y destrucción de conexiones. En los modelos de redes neuronales artificiales, la creación de una nueva conexión implica que el peso de la misma pasa a tener un valor distinto de cero. De la misma forma, una conexión se destruye cuando su

peso pasa a ser cero. Existen modelos de redes neuronales con aprendizaje supervisado y no supervisado. Lo que distingue a una red con aprendizaje supervisado con una de aprendizaje no supervisado es la existencia de un agente externo (supervisor) que controle el proceso de aprendizaje de la red. Otro criterio que se puede utilizar para diferenciar las reglas de aprendizaje se basa en considerar si la red puede aprender durante su funcionamiento habitual o si el aprendizaje supone la desconexión de la red. En el primer caso, se trataría de un aprendizaje ON LINE, mientras que el segundo es lo que se conoce como aprendizaje OFF LINE. En el caso de las redes con aprendizaje OFF LINE, los pesos de las conexiones permanecen fijos después que termina la etapa de entrenamiento de la red.

c) *Tipo de asociación de las informaciones de entrada y salida y la forma de representación de esta información.* Las redes neuronales almacenan cierta información aprendida; esta información se registra de forma distribuida en los pesos asociados a las conexiones entre neuronas. Existen dos formas de realizar la asociación entre entrada/salida que se corresponden con la naturaleza de la información almacenada en la red. Una primera es la *heteroasociación*, que se refiere al caso en que la red aprende parejas de datos [(A1,B1),..(An,Bn)], de tal manera que cuando se presente cierta información de entrada Ai, deberá responder generando la correspondiente salida asociada Bi. La segunda se conoce como *autoasociación* de tal forma que cuando se le presenta una información de entrada realiará una autocorrelación respondiendo con uno de los datos almacenados, el más parecido al de entrada-

d) *Señales de entrada/salida.* Las redes neuronales pueden clasificarse en función de la forma en que se representan las informaciones de entrada y las respuestas o datos de salida. Estas representaciones pueden ser analógicas, es decir, son valores reales continuos normalmente normalizados y su valor absoluto será menor que la unidad. Otra forma es que los datos sean valores discretos o binarios.

MODELO DE PROPAGACIÓN DE ERROR HACIA ATRÁS

El modelo de red neuronal empleado en el escalamiento multidimensional no métrico es llamado, modelo de propagación de error hacia atrás (*Backpropagation*). La topología de este modelo es de *N capas*, su aprendizaje es *supervisado* por corrección de error y los pesos de las conexiones permanecen fijos después que termina la etapa de entrenamiento (*OFF LINE*) . El algoritmo de aprendizaje por corrección de error lo constituye la regla delta generalizada conocida como error de mínimos cuadrados. El tipo de asociación entre las informaciones entrada/salida es *heteroasociativo*. La información de entrada y salida es de tipo *analógico*.

El funcionamiento de una red de propagación de error hacia atrás consiste en un aprendizaje de un conjunto predefinido de pares de entradas-salidas empleando un ciclo de propagación-adaptación de dos fases.

1) Se aplica un patrón de entrada como estímulo para la primera capa de las neuronas de la red, se va propagando a través de todas las capas superiores hasta generar una salida. Se compara el

resultado obtenido en las neuronas de salida con la salida que se desea obtener y se calcula un valor de error para cada neurona de salida.

2) Estos errores se transmiten hacia atrás, partiendo de la capa de salida, hacia todas las neuronas de la capa intermedia que contribuyan directamente a la salida, recibiendo el porcentaje de error aproximado a la participación de la neurona intermedia en la salida original.
3) El proceso se repite capa por capa hasta que todas reciban un error que describa su aportación relativa al error total.
4) Se reajustan los pesos de conexión de cada neurona, de manera que la siguiente vez que se presente el patrón, la salida esté más cercana a la deseada.

MINIMIZACIÓN DE LA FUNCIÓN DE STRESS CON REDES NEURONALES

Esta red neuronal toma a las disimilaridades como entradas y las disparidades como salidas y una capa de unidades ocultas con funciones de transferencia no lineales (tangente hiperbólica). La unidad de salida emplea la función identidad como función de transferencia. Sean a, b y c los índices de unidades de entrada, ocultas y de salida respectivamente.

Sea $y_{k/ij}$ la salida de la unidad k-ésima (de entrada, oculta o de salida) cuando se tiene como entrada la disimilaridad δ_{ij}. Sea $g_k(.)$ la función de transferencia para la unidad k. De esta manera, la función de Stress se denota de la siguiente manera:

$$S = \sum_{i \neq j} (y_{c/ij} - d_{ij})^2 = \sum_{i \neq j} S^{ij}$$

Como las disimilaridades contienen un error intrínseco, se introduce un sesgo. Así, el sesgo para la unidad k-ésima se denota como θ_k. La entrada total a la unidad k-ésima se denota por $net_{k/ij}$, de esta manera, $y_{k/ij} = g_k(net_{k/ij})$. Debido a que las funciones individuales de transferencia son monónotas crecientes, la condición (C1) se satisface.

$net_{c/ij} = \sum_b w_{bc}^2 y_{b/ij} + w_{ac}^2 y_{a/ij} + \theta_c$ ... la entrada total a la unidad de salida .... Ecuaciones (2)

$net_{b/ij} = \sum_a w_{ab}^2 y_{a/ij} + \theta_b$ ... la entrada total a la unidad oculta

Derivando parcialmente la función de Stress con respecto a los pesos y empleando los subíndices k y l para denotar a cualquier unidad, se tiene:

$$\frac{\partial S^{ij}}{\partial w_{kl}} = \frac{\partial S^{ij}}{\partial net_{l/ij}} \frac{\partial net_{l/ij}}{\partial w_{kl}}$$

Por otra parte,

$$\frac{\partial S^{ij}}{\partial \theta_i} = \frac{\partial S^{ij}}{\partial net_{l\,l\,ij}} \frac{\partial net_{l\,l\,ij}}{\partial \theta_i}$$

A partir de las ecuaciones (2) se tiene: $\frac{\partial net_{l\,l\,ij}}{\partial w_{kl}} = 2 w_k y_{k\,l\,ij}$ y $\frac{\partial net_{l\,l\,ij}}{\partial \theta_l} = 1$

Y

Si se denota, $\Delta_{l\,l\,ij} = \frac{\partial S^{ij}}{\partial net_{l\,l\,ij}}$, se tiene: $\frac{\partial S^{ij}}{\partial w_{kl}} = 2 w_k y_{k\,l\,ij} \Delta_{l\,l\,ij}$ y $\frac{\partial S^{ij}}{\partial \theta_l} = \Delta_{l\,l\,ij}$

Para la unidad de salida con función de activación la identidad, los valores de Δ son fácilmente calculados:

$$\Delta_{c\,l\,ij} = \frac{\partial S^{ij}}{\partial net_{c\,l\,ij}} = 2(a_{c\,l\,ij} - d_{ij}).$$

Para las unidades ocultas, con función de activación tanh sugerida en Wezel[2001], los valores de Δ son:

$$\Delta_{b\,l\,ij} = \frac{\partial S^{ij}}{\partial net_{b\,l\,ij}} = \Delta_{c\,l\,ij} w_{bc}^2 (1 - g_b(a_{b\,l\,ij})^2).$$

Dadas las ecuaciones anteriores, el algoritmo de entrenamiento para una red neuronal monótona como esta, se describe de la siguiente manera: Primero, se eligen pares de objetos i,j de manera aleatoria de todos los posibles pares de objetos. Las disimilaridades entre este par de objetos es empleado como entrada en la red. La salida correspondiente es calculada y los valores de Δ se calculan para todas las unidades y se calculan las derivadas parciales descritas anteriormente. Estas derivadas parciales son empleadas para actualizar los pesos de la red.

Los algoritmos mencionados anteriormente han sido determinísticos y son los más frecuentemente utilizados para minimizar la función de Stress ya que su convergencia es rápida, sin embargo, tienden a un mínimo local.

Técnicas estocásticas como simulated annealing[1] tratan a las coordenadas como variables aleatorias y evitan un mínimo local pero consumen mucho tiempo/máquina.. En simulated annealing, las coordenadas esperadas en escalamiento multidimensional son estimadas empleando la distribución de Gibbs mediante el método de Monte Carlo.

Desde la investigación de Kruskal, se han efectuado numerosas investigaciones de la función de Stress empleando métodos de Monte Carlo. Una de las conclusiones en numerosas investigaciones fue que entre mayor sea la dimensión de la configuración final menor es el stress y en cuanto mayor sea n, el número de objetos, mayor es el stress.

Se sugieren varias configuraciones iniciales en los algoritmos determinísticos para evitar mínimos locales.

[1]Se prefirió emplear el término en inglés, ya que no existe una traducción precisa del término en la bibliografía existente.

ANEXO

ANEXO

RESULTADOS DE APLICACIONES

I.1.1 Exploración de Niveles Socioeconómicos de acuerdo a la regla AMAI 13x6

Escalamiento multidimensional métrico de la figura 5.2.1

Programa SAS:

```
proc mds data=tabla_5_2
level=absolute /*no hace transformaciones a las disimilaridades*/
pineigval pineigvec /*imprime los eigenvectores y los eigenvalores de B*/
out=salidal; /*imprime las coordenadas resultantes*/
outres=salida2_res; /*imprime las distancias originales y las distancias
resultantes*/
id nset; /*Etiqueta los puntos con NSE*/
run;
%plotit(data=salidal, datatype=mds, color=black, colors=red,labelvar=nset,
vtoh=1.75, labfont=swissb);
run; /*grafica la configuracion resultante*/
```

Resultado:

NSET	Dim1	Dim2
A/B	1.891308	-0.42749
C+	0.933369	0.221177
C	0.206838	0.36115
D+	-0.53881	0.419944
D	-1.14629	-0.10111
E	-1.34642	-0.47366

I.1.2 Exploración de Niveles Socioeconómicos de acuerdo a la regla Amai 13x6

Escalamiento multidimensional métrico de mínimos cuadrados de la figura 5.2.2 del capítulo 5. .

Programa SAS:

```
proc mds data=tabla_5_2
level=ratio /*hace transformaciones de la forma f(d)=ad*/
out=salida2 /*imprime las coordenadas resultantes*/
outres=salida2_res; /*imprime las distancias originales y las distancias
resultantes*/
id nset; /*Etiqueta los puntos con NSE*/
run;
```

```
%plotit(data=salida2, datatype=mds, color=black, colors=red,labelvar=nset,
vtoh=1.75, labfont=swissb);
run; /*grafica la configuracion resultante*/
```

Resultado:

NSET	Dim1	Dim2
A/B	2.223637	-0.50261
C+	1.097371	0.260026
C	0.243185	0.424633
D+	-0.63349	0.493733
D	-1.3477	-0.11889
E	-1.583	-0.55689

I.1.3 Exploración de Niveles Socioeconómicos de acuerdo a la regla AMAI 13x6

Escalamiento multidimensional no métrico con método de Kruskal, de la figura 5.2.3 del capítulo 5.

Programa SAS:

```
proc mds data=tabla_5_2
level=oridal /*hace regresion isotonica*/
out=salida3/*imprime las coordenadas resultantes*/
outres=salida3_res; /*imprime las distancias originales y las distancias
resultantes*/
id nset; /*Etiqueta los puntos con NSE*/
run;

%plotit(data=salida3, datatype=mds, color=black, colors=red,labelvar=nset,
vtoh=1.75, labfont=swissb);
run;
```

Resultado:

NSET	Dim1	Dim2
A/B	2.247542	-0.48491
C+	1.077669	0.232278
C	0.244649	0.46527
D+	-0.6107	0.435652
D	-1.32328	-0.21995
E	-1.63588	-0.42834

I.2.1 Análisis estructural de las reglas AMAI

Modelo de diferencias individuales con ALSCAL de la figura 5.3.1 y de la figura 5.3.2

Programa SAS:

```
title  'MDS de configuracion de los espacios de grupo e indivudual';
title2 'Juicios de similaridad entre hogares';

Proc mds fit=squared /*Hace transformaciones cuadradas a las dsimilaridades*/
data=tes2.matriz
similar=2  /*Considera al insumo, como similaridades y las transforma a
disimilaridades*/
level=ORDINAL UNTIE /*Realiza un escalamiento no metrico*/
CONDITION=MATRIX  /*La restriccion condicional es a nivel matriz, ya que los
datos son comparables  dentro de cada matriz, pero no entre matrices*/
ALTERNATE=MATRIX  /*Ajusta todos los parametros para la primera variable
y despues para la segunda, etc. y al final ajusta todos los parametros
que no tenga que ver con las variables como coordenadas y transformaciones
incondicionales*/
coef=diagonal  /*produce distancias euclideanas ponderadas por cada
variable, puede tener diferentes pesos para las dimensiones*/
out=salida
pfit; /*imprime los criterios de ajuste y varios tipos de correlaciones
entre los datos y los valores ajustados para cada matriz y para la muestra
completa*/
var    dis_1-dis_60;
MATRIX variable;
id nset;
title3 'Analisis de Diferencias individuales';
run;

data config; /*archivo con las coordenadas de los hogares*/
set salida;
if _type_="CONFIG";
run;
/*grafica a los hogares en el espacio de hogares*/
%plotit(data=config,datatype=mds,color=black,colors=red,labelvar=nse,   vtoh=1.75,
labfont=swissb);
run;

data diagconfig; /*archivo con las coordenadas o pesos de las variables*/
set salida;
if _type_="DIAGCOEF";
RUN;

title3 'Espacio de Configuracion del grupo de variables';
/*grafica a las variables en el espacio de variables*/
%plotit(data=diagconfig,datatype=mds,color=black,colors=blue,
labelvar=variable, vtoh= 1.75, labfont=swissb);
run;
```

```
proc sql;
create table inspace as  /* archivo con los espacios individuales */
select outa.dim1*outb.dim1 as indim1, /* muliplica las
outa.dim2*outb.dim2 as indim2,  configuraciones*/
outa.nse as NSE ,
outb.variable as variable
from salida as outa, salida as outb
where outa._type_='CONFIG' and /* selecciona la configuracion de los hogares */
outb._type_='DIAGCOEF'/* selecciona la configuracion de las variables*/
order by variable; /*ordenado por matriz de variable */
```

Resultados:

	Dim1	Dim2
aspiradora	1.412931	0.060215
boiler	0.607219	1.277218
edujef	0.666539	1.247287
focos	0.905792	1.086067
lavaropa	1.309796	0.533324
micro	1.271464	0.619176
nauto	1.311683	0.528666
npieza	0.753569	1.196718
numbanc	0.867044	1.117244
pc	1.391328	0.253392
piso	0.765972	1.188818
tosta	1.347757	0.428428
vcr	0.669354	1.245779

hogar	nse	Dim1	Dim2	hogar	nse	Dim1	Dim2
1	A/B	1.244927	0.587408	31	D+	-0.96897	-0.24926
2	A/B	1.05752	1.470519	32	D+	-0.83439	0.11323
3	A/B	0.626993	1.41231	33	D+	-0.98001	0.289562
4	A/B	2.381482	0.785017	34	D+	-0.97945	0.073477
5	A/B	0.757436	1.667876	35	D+	-0.96961	0.243748
6	A/B	0.90364	1.463497	36	D+	-0.5962	-0.01496
7	A/B	0.903639	1.463497	37	D+	-0.96023	0.473133
8	A/B	2.683291	0.312864	38	D+	0.100429	0.718981
9	A/B	1.941635	0.43859	39	D+	-0.97408	0.0617
10	A/B	2.496621	0.303585	40	D+	0.597577	-0.54306
11	C+	-0.42084	0.985999	41	D	-0.95837	-0.20591
12	C+	0.970276	0.537431	42	D	0.352387	-0.85868
13	C+	1.102297	-0.22128	43	D	-0.97536	-0.01738
14	C+	1.006347	0.556816	44	D	-0.92148	-1.24053
15	C+	0.732946	1.410218	45	D	-0.87589	-0.73877
16	C+	-0.06051	0.946518	46	D	-0.98873	-0.61534
17	C+	0.459448	0.709597	47	D	-0.92148	-1.24053
18	C+	0.174708	1.449143	48	D	-0.97577	-0.73844
19	C+	1.421279	0.504469	49	D	-0.92148	-1.24053
20	C+	0.655235	0.99933	50	D	-0.9676	-0.34834

21	C	0.475178	0.298847	51	E	-0.85031	-1.4869
22	C	0.213559	1.386069	52	E	-0.85031	-1.4869
23	C	0.878465	-0.31102	53	E	-0.85031	-1.4869
24	C	0.143348	1.2007	54	E	-0.96542	-1.0864
25	C	-1.04402	0.433002	55	E	-0.6257	-2.05618
26	C	0.549863	0.035053	56	E	-0.91209	-0.94478
27	C	0.210986	1.277956	57	E	-0.72145	-1.84501
28	C	0.524344	-0.21736	58	E	-0.6257	-2.05618
29	C	-0.60613	0.220564	59	E	-0.85031	-1.4869
30	C	0.43667	-0.60628	60	E	-0.85031	-1.4869

Analisis de Diferencias individuales
Multidimensional Scaling: Data=TES2.MATRIZ
Shape=TRIANGLE Condition=MATRIX Level=ORDINAL UNTIE
Coef=DIAGONAL Dimension=2 Formula=1 Fit=2

variable	Number of Nonmissing Data	Weight	Badness-of-Fit Criterion	Distance Correlation	Uncorrected Distance Correlation	Fit Correlation
aspirador	1770	0.08	0.15	0.96	0.98	0.98
boiler	1770	0.08	0.16	0.95	0.99	0.98
edujef	1770	0.08	0.41	0.80	0.95	0.82
focos	1770	0.08	0.19	0.94	0.99	0.96
lavaropa	1770	0.08	0.16	0.96	0.99	0.98
micro	1770	0.08	0.18	0.94	0.98	0.97
nauto	1770	0.08	0.19	0.95	0.98	0.97
npieza	1770	0.08	0.31	0.88	0.97	0.90
numbanc	1770	0.08	0.27	0.91	0.98	0.92
pc	1770	0.08	0.23	0.93	0.97	0.96
piso	1770	0.08	0.29	0.93	0.98	0.91
tosta	1770	0.08	0.15	0.96	0.99	0.98
vcr	1770	0.08	0.30	0.87	0.97	0.91
All -	23010	1.00	0.24	0.92	0.98	0.94

variable	Uncorrected Fit Correlation
aspirador	0.99
boiler	0.99
edujef	0.91
focos	0.98
lavaropa	0.99
micro	0.98
nauto	0.98
npieza	0.95
numbanc	0.96
pc	0.97
piso	0.96
tosta	0.99
vcr	0.95
- All -	0.97

I.2.2 Análisis estructural de las reglas AMAI

Modelo de diferencias individuales con ALSCAL de la figura 5.3.3

Resultados:

MDS de configuracion de los espacios de grupo e indivudual 21
Juicios de similaridad entre hogares 16:16 Saturday, April 1, 2000
Analisis de Diferencias individuales

Multidimensional Scaling: Data=WORK.MATRIZ2
Shape=TRIANGLE Condition=MATRIX Level=ORDINAL UNTIE
Coef=DIAGONAL Dimension=2 Formula=1 Fit=2

variable	Number of Nonmissing Data	Weight	Badness-of-Fit Criterion	Distance Correlation	Uncorrected Distance Correlation	Fit Correlation
aspirador	1770	0.14	0.10	0.98	0.99	0.99
boiler	1770	0.14	0.13	0.97	0.99	0.99
focos	1770	0.14	0.13	0.97	0.99	0.99
lavaropa	1770	0.14	0.11	0.97	0.99	0.99
micro	1770	0.14	0.11	0.97	0.99	0.99
nauto	1770	0.14	0.20	0.95	0.99	0.98
tosta	1770	0.14	0.11	0.97	0.99	0.99
- All -	12390	1.00	0.13	0.97	0.99	0.99

variable	Uncorrected Fit Correlation
aspirador	1.00
boiler	0.99
focos	0.99
lavaropa	0.99
micro	0.99
nauto	0.98
tosta	0.99
- All -	0.99

I.2.3 Análisis estructural de las reglas AMAI

Modelo de diferencias individuales con ALSCAL de la figura 5.3.4 y 5.3.5

Resultados:

Multidimensional Scaling: Data=WORK.MATRIZ2
Shape=TRIANGLE Condition=MATRIX Level=ORDINAL UNTIE
Coef=DIAGONAL Dimension=2 Formula=1 Fit=2

variable	Number of Nonmissing Data	Weight	Badness-of-Fit Criterion	Distance Correlation	Uncorrected Distance Correlation	Fit Correlation
boiler	1770	0.17	0.12	0.97	0.99	0.99
edujef	1770	0.17	0.16	0.97	0.99	0.98
focos	1770	0.17	0.16	0.98	0.99	0.97
nauto	1770	0.17	0.34	0.85	0.97	0.86
npieza	1770	0.17	0.21	0.95	0.99	0.96
numbanc	1770	0.17	0.15	0.97	0.99	0.98
- All -	10620	1.00	0.20	0.95	0.99	0.96

variable	Uncorrected Fit Correlation
boiler	0.99
edujef	0.99
focos	0.99
nauto	0.94
npieza	0.98
numbanc	0.99
- All -	0.98

I.3.1 Propuesta de asignación de niveles socioeconómicos

Escalamiento de Kruskal de la aplicación 3

Programa SAS:

```
proc mds data=insumo
dim=2
level=ordinal
out=salida
outres=salida_res;
id hogar;
run;

/*uso insumo*/
proc sql; create table stress as
select distinct sum((a.fitdata-a.distance)**2 )as Num_stress,
sum(a.distance**2) as den_stress, (calculated num_stress/calculated
den_stress)**.5 as Stress
from salida_res a;

title 'Escalamiento multidimensional no metrico';
title2 'Algoritmo de Kruskal';
title3 'Stress=0.1577';

%plotit(data=salida, datatype=mds, color=black, colors=red,labelvar=hogar,
vtoh=2.645, labfont=swissb);
run;


proc plot data=salida;
   plot dim2 * dim1 $ hogar;
   where _type_='CONFIG';
run;

%let var=ssirv;

proc sql; create table tabla as
select distinct a.hogar, a.dim1, a.dim2, b.&var. as &var
from salida a, datos b
where a._type_="CONFIG" and a.hogar=b.hogar;


*---Compute Endpoints for MPG and Reliability Vectors---;
proc transreg data=tabla;
   Model identity(&var.)=identity(dim1 dim2);
   output tstandard=center coordinates replace out=res1;
   id hogar;
   title2 'Analisis de Preferencias';
run;

proc print; run;

data res1;
set res1;
```

```
if _name_="&var." then prin1=dim1*10 and prin2=dim2*10;
prin1=dim1;
prin2=dim2;
run;

data res1;
set res1;
if _type_="M COEFFI" then prin1=dim1*15;
IF _TYPE_="M COEFFI" then prin2=dim2*15;
run;
TITLE 'Analisis de Preferencia';
TITLE2 'Variable Numero de Sirvientes';

%plotit(data=res1, datatype=vector ideal, color=black, antiidea=1,vtoh=2.9);

proc mds data=insumo
level=ordinal
out=salida
outres=salida_res;
id hogar;
run;

proc sql; create table stress as
select distinct sum((a.fitdata-a.distance)**2 )as Num_stress,
sum(a.distance**2) as den_stress, (calculated num_stress/calculated
den_stress)**.5 as Stress
from salida_res a;

title 'Escalamiento multidimensional no metrico';
title2 'Algoritmo de Kruskal';
title2 'Stress=0.1522';

%plotit(data=salida, datatype=mds, color=black, colors=red,labelvar=hogar,
vtoh=2.645, labfont=swissb);
run;
```

I.4.1 Análisis de correspondencia del Capítulo 4, pág 50

Insumo:

Tabla 4.1

NSET	Numbanc	focos	npieza	edujef	nauto	lavaropa	tosta	boiler	vcr	micro	aspirado	pc	piso
A/B	2.64	26.10	7.58	10.91	2.36	0.87	0.84	0.97	0.87	0.88	0.81	0.67	0.97
C+	1.55	14.66	5.89	8.68	1.47	0.68	0.79	0.97	0.84	0.61	0.45	0.42	1.00
C	1.35	10.89	5.43	6.78	1.15	0.41	0.28	0.93	0.67	0.39	0.07	0.22	0.98
D+	1.15	8.35	4.66	4.42	0.32	0.24	0.21	0.78	0.41	0.22	0.02	0.11	1.00
D	0.65	4.84	2.94	3.39	0.33	0.12	0.08	0.16	0.12	0.06	0.00	0.02	1.00
E	0.00	2.87	1.73	2.20	0.07	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.73

Programa SAS:

```
proc corresp data=aplicacion1_1 out=Results short;
   var numbanc focos npieza edujef nauto
       piso lavaropa
       tosta boiler vcr micro aspirado pc;
   id nse;
run;

%plotit(data=Results, datatype=corresp, plotvars=Dim1 Dim2) ;
```

Resultado:

Singular Value	Principal Inertia	Chi-Square	Percent	Cumulative Percent	14 28 42 56 70
0.18442	0.03401	5.69120	71.38	71.38	**********
0.08890	0.00790	1.32248	16.59	87.97	******
0.05991	0.00359	0.60048	7.53	95.50	***
0.03474	0.00121	0.20191	2.53	98.03	*
0.03062	0.00094	0.15693	1.97	100.00	*
Total	0.04765	7.97300	100.00		

Row Coordinates

	Dim1	Dim2
A/B	-0.1755	0.0683
C+	-0.0779	-0.0294
C	0.0510	-0.0967
D+	0.1559	-0.1119
D	0.3075	0.0540
E	0.4868	0.2383

Column Coordinates

	Dim1	Dim2
numbanc	-0.0998	-0.1365
focos	-0.0786	0.0514
npieza	0.1845	-0.0518
edujef	0.0830	0.0145
nauto	-0.2737	0.0087
piso	0.5937	0.1173
lavaropa	-0.2566	-0.0999
tosta	-0.3341	-0.0650
boiler	-0.0386	-0.3852
vcr	-0.0888	-0.2049
micro	-0.3211	-0.0893
aspirado	-0.6839	0.2764
pc	-0.4409	0.0137

BIBLIOGRAFÍA

AMAI.
Memoria del I Seminario de Actualización Profesional.
Junio, 1994.

AMAI.
Memoria del II Seminario de Actualización Profesional.
Junio, 1995.

AMAI.
Memoria del III Seminario de Actualización Profesional.
Agosto, 1996.

AMAI.
Memoria del IV Seminario de Actualización Profesional.
Agosto, 1997.

AMAI.
Memoria del V Seminario de Actualización Profesional.
Agosto, 1998

AMAI.
Memoria del VI Seminario de Actualización Profesional.
Agosto, 1999.

AMAI.
Memoria del IX Seminario de Actualización Profesional.
Agosto, 2002

Apóstol, Tom M.
Análisis Matemático
Segunda Edición
Ed. Reverté, 1996.

Borg & Groenen.
Modern Multidimensional Scaling. Theory and Applications.
Springer, 1997.

Burden, Richard L & Farres, J. D
Análisis Numérico
Grupo Editorial Iberoamérica, 1996

Caillez, F & Pages, J
Introduction à l'Analyse des Données
SMASH, 1976

Cox, Trevor F. and Cox, Michael A.A.
Multidimensional Scaling
Monographs on Statistics and Applied Probability 59
Chapman & Hall, 1994.

Golledge, Reginald G. & Rayner, John N.
Proximity and Preference: Problems in the Multidimensional Analysis of Large Data Sets.
University of Minnesota Press, 1982.

Graybill, Franklin A.
An Introduction to Lineal Statistical Models Vol1
Ed. McGraw-Hill, 1961

Haaser, La Salle, Sullivan
Análisis Matemático Vol2
Ed. Trillas, 1999.

Hilera, J.R y Martínez, V. J.
Redes Neuronales artificiales: Fundamentos, modelos y aplicacioenes
Ra-Ma, 1995

Hoffman, Kenneth y Kunze Ray
Álgebra Lineal
Prentice Hall, 1961

Jonson, Dallas E.
Métodos multivariados aplicados al análisis de datos.
International Thomson Editores, 2000.

Krazanowski, W.J.
Principles of Multivariate Analysis: A users perspective.
Oxford University Press, 1990.

Kruskal, Joseph B & Wish, Mirón
Multidimensional Scaling.
Series: Quantitative Applications in the Social Science.
Sage University Paper, 1978.

Lang Serge
Introducción al Álgebra lineal
Addison Wesley Iberoamericana, 1990

Mardia et al.
Multivariate Analysis
Academic Press, Second Edition, 1979

Torgerson, Warren S.
Theory & Methods of Scaling
John Willey & Sons, 1958.

Van Wezel, Michael C., Josters, Walter A., Van der Putten, Peter.
Nonmetric Multidimensional Scaling with Neural Networks.
LIACS, 2001.

Young F.W. & Hammer, R.M.
Multidimensional Scaling: History, Theory and Applications.
Eribaum, New York, 1987.