

UNIVERSIDAD DE GRANADA

Doctorado en Matemáticas y Estadística



Departamento de Estadística e Investigación Operativa

TESIS DOCTORAL

Un modelo de Ecuaciones Estructurales para el
escalamiento multidimensional de datos
asimétricos

Christian David Rivera Ramírez

Granada, España

Editor: Universidad de Granada. Tesis Doctorales
Autor: Christian David Rivera Ramírez
ISBN: 978-84-9163-876-6
URI: <http://hdl.handle.net/10481/51653>

A Angie y Marco David

Para Suyim especialmente

Agradecimientos

Al Dr. J. Fernando Vera por su paciencia y su amistad y por haberme ayudado a conocerme un poco más. Todo tuvo su peso específico: las discusiones académicas, el café de media mañana y sobre todo las tertulias de primavera ante ese testigo al que Neruda describió con pies de púrpura y sangre de topacio.

A Rodrigo y a Cony, su amistad ha sido un hermoso regalo.

A la Dra. Graciela Farías y a los trabajadores del CI-MAT Monterey, México, por proveerme de un espacio de mucha cordialidad y riqueza intelectual para mi estancia.

A la Univerisdad de Los Andes, mi Alma Mater.

Índice general

1. Introducción	7
1.1. Conceptos Básicos del MDS	10
1.1.1. Los datos en MDS	15
1.1.2. Naturaleza métrica de las disimilaridades	22
1.1.3. Modelos en MDS	25
1.1.4. Modelos métricos de MDS	27
2. MDS para datos asimétricos	35
2.1. Modelos basados en producto escalar	35
2.1.1. Modelo DEDICOM	36
2.1.2. Modelo GIPSCAL	40
2.1.3. Modelo GIPSCAL Generalizado	42
2.2. Modelos basados en distancias	45
2.2.1. Preliminares	45
2.2.2. Modelo I de distancia generalizada	48
2.2.3. Modelo II de distancia generalizada	51
2.2.4. Modelos del tipo Distancia-Densidad	56
2.2.5. El modelo Slide-Vector	60
3. Modelos de Ecuaciones Estructurales	63
3.1. Introducción	63
3.1.1. Características de los modelos SEM	64
3.1.2. Tipos de modelos SEM	65
3.1.3. Lógica de los modelos de ecuaciones estructurales	67
3.1.4. El modelo de ecuaciones estructurales general	69
3.1.5. Identificación	74
3.1.6. Estimación y ajuste	78

3.1.7. Estimación de los valores de las variables latentes (factor scores)	84
4. El Modelo SEM-MDS	89
4.1. Introducción	89
4.1.1. Optimización en MDS: El algoritmo SMACOF	92
4.2. El modelo	95
4.3. Errores de medida no modelados	101
4.4. Resultados: Un algoritmo SEM-MDS alternante	106
4.5. Aplicaciones	109
5. Conclusiones	115
6. Summary	117
Bibliografía	

Objetivos y metodología

El objetivo general de la investigación cuyos resultados se recogen en esta tesis doctoral es el de diseñar y validar un modelo para el análisis de matrices de proximidades asimétricas de un modo, que combine las ventajas metodológicas y procedimentales de los modelos de ecuaciones estructurales con variables latentes para la estimación de parámetros, con los procedimientos del escalamiento multidimensional de estructuras mediante proximidades bajo transformaciones admisibles, para la representación espacial de los datos en espacios de baja dimensión.

Objetivos específicos

1. Especificar el modelo de ecuaciones estructurales (SEM) que exprese teóricamente las relaciones entre los datos y su representación espacial.
2. Lograr la identificación de los parámetros del modelo según la satisfacción de condiciones necesarias y suficientes para la identificación o con base a procedimientos algebraicos.
3. Diseñar un algoritmo alternante basado en el algoritmo SMACOF que en una fase estime los parámetros del modelo estructural y en la otra fase estime el espacio MDS que explique lo mejor posible las disimilitudes transformadas.
4. implementar un experimento simulado que recoja los escenarios más comunes de la investigación empírica en el ámbito teórico del modelo que permita evaluar su desempeño comparativo e identificar sus fortalezas y debilidades.

5. Evaluar el desempeño del modelo con datos reales con el objeto de construir un juicio informado acerca de las ocasiones de investigación en las que el modelo pueda resultar de utilidad.

Estrategias metodológicas

El logro de los objetivos específicos requirió el diseño y desarrollo de las estrategias metodológicas que se exponen a continuación.

Estrategias desarrolladas para lograr el objetivo 1

1. Definición del concepto de especificación que se ajustara a nuestra perspectiva investigativa: Es el primer paso en el desarrollo de un modelo como una herramienta intelectual para aproximarse a la comprensión de una realidad particular de interés científico para el investigador, y consiste en decidir cuáles variables pueden cuantificar y/o cualificar de manera mas sistemática y sistémica las distintas dimensiones de la realidad de interés y recoge la perspectiva teórica del investigador en un conjunto de relaciones entre dichas variables.
2. Revisión exhaustiva de la literatura científica referida a los modelos MDS para datos asimétricos, considerando cuáles podían ser aplicados en el ámbito de nuestro interés. Análisis y discusión de las fortalezas y debilidades de esos modelos, reportadas en la literatura científica.
3. Establecimientos de la direccionalidad de las relaciones del componente estructural del modelo y las restricciones necesarias para que este componente pueda funcionar conjuntamente con el componente MDS.

Estrategias desarrolladas para lograr el objetivo 2

1. Revisión de la literartura para establecer las restricciones teoricamente plausibles en el contexto de la investigación, para lograr un modelo sobreidentificado.
2. desarrollo de procedimientos algebraicos para establecer la identificación del modelo.

Estrategias desarrolladas para lograr el objetivo 3

1. Realización de las comprobaciones matemáticas y procedimentales a los fines de modificar el algoritmo SMACOF de manera de dar cabida a la representación de matrices no simétricas en el algoritmo SEMDS propuesto, sin perder las capacidades de decrecimiento y convergencia.
2. Análisis de las distintas funciones para la estimación de parámetros en SEM de manera de seleccionar la más conveniente en el sentido de que conjuntamente con la función de pérdida de mínimos cuadrados de MDS, produzca una sucesión de valores del STRESS decreciente y de convergencia rápida.
3. Comprobación de funcionamiento del algoritmo con datos reales y artificiales fijando los parámetros del componente estructural para obtener los mismos resultados de SMACOF

Estrategias desarrolladas para lograr el objetivo 4

1. Revisión exhaustiva de la comunicación científica en el ámbito de la representación de disimilaridades asimétricas, con el fin de determinar las situaciones más comunes de investigación empírica que permitiera definir los parámetros adecuados para las pruebas de desempeño del modelo con datos artificiales.
2. Realizar el experimento de simulación con todos los escenarios posibles derivados de los resultados obtenidas de la estrategia anterior, con un número de replicas similar a los estudios de simulación reportados en la literatura científica.

Estrategias desarrolladas para lograr el objetivo 5

1. revisión exhaustiva de la literatura científica para determinar varios conjuntos de datos reales en el contexto de aplicabilidad del modelo SEMDS propuesto, para realizar análisis comparativos del desempeño del modelo SEMDS.
2. Análisis detallado de los resultados del desempeño comparativo del modelo MDSEM, para determinar debilidades, fortalezas y requerimientos para su aplicación.

Capítulo 1

Introducción

Esta investigación trata el problema de la representación de datos asimétrica con técnicas de análisis multidimensional de estructuras mediante proximidades o Multidimensional Scaling (MDS).

El insumo básico del MDS es una matriz simétrica cuyas entradas corresponden a alguna medida de proximidad, $p_{i,j}$, entre los objetos i y j . El MDS intenta representar tales proximidades como distancias entre puntos en un espacio de baja dimensión, por lo general un espacio Euclidiano conocido como el espacio MDS, de tal forma que los pares de objetos percibidos como muy similares son representados mediante puntos cercanos entre si. La simetría aparece al suponer que la proximidad del objeto i al objeto j es la misma que la del objeto j al objeto i . Este supuesto permite representar completamente a las proximidades como distancias entre los puntos en el espacio MDS (Borg y Groenen, 2005 pág. 495). Varios enfoques han sido desarrollados para el análisis de medidas de proximidad simétricas, algunos de los cuales son presentados de manera sucinta en este trabajo.

Sin embargo, en muchas ocasiones la simetría no está presente cuando se recolectan los datos y se construyen las proximidades. Es posible que las desviaciones de la simetría obedezcan a fluctuaciones estadísticas que generan errores aleatorios. En este caso, la práctica común es promediar las entradas (i,j) y (j,i) de la matriz de proximidades original para construir una matriz simétrica sobre la cual realizar un análisis MDS simétrico. Así, si la matriz original es $A = (a_{ij})$, donde por lo general $a_{ij} \neq a_{ji}$, entonces la matriz sobre la cual se llevaría a cabo el análisis podría ser:

$$\Delta = (\delta_{ij})$$

con

$$\delta_{ij} = \frac{a_{ij} + a_{ji}}{2}.$$

Es posible también que las asimetrías aparezcan como producto de alguna clase de sesgo en las mediciones de las proximidades, como ocurre en algunos experimentos en el contexto de la Teoría de los Niveles Conceptuales (CLT) (Trope et.al., 2007). Si la asimetría es el resultado de errores sistemáticos, podría ser evitada con diseños experimentales convenientes o eliminada en la fase de análisis preliminar de los datos.

En otras situaciones sin embargo, la asimetría tiene un efecto significativo sobre la estructura interna de los datos y en consecuencia no puede ser descartada como el resultado de errores de medición. Saito y Yadohisa (2005, pág. 104) identifica dos casos en los que la asimetría es de interés para explicar la distribución de las proximidades y debería ser considerada en la construcción del espacio MDS.

En el primero de los casos, la asimetría es una característica intrínseca a la naturaleza de los datos. Un ejemplo típico surge en estudios de mercado cuando se analiza la probabilidad de que un consumidor cambie entre bienes sustitutivos. En tales casos, una medida común de proximidad entre los bienes i y j , se construye sobre el conteo de consumidores que compran el bien j cuando en la ocasión anterior compraron el bien i . Aquí, $p_{i,j} \neq p_{j,i}$ debido a que, en general, el número de consumidores que migran del bien i al bien j no es igual al número de consumidores que migran en sentido contrario.

Otro ejemplo surge en el ámbito de la sociometría al medir relaciones interpersonales en un grupo social, a través de la matriz de valoración sociométrica (sociometric rating). En las filas de esta matriz aparecen los individuos como jueces mientras que en las columnas aparecen siendo juzgados. Así, el individuo i podría expresar un mayor agrado (elección) o desagrado (rechazo) por el individuo j que al contrario, con lo cual volvería a darse la relación $p_{i,j} \neq p_{j,i}$. Muchos otros ejemplos aparecen también en estudios de tasas de migración entre países, citaciones entre revistas científicas, flujos comerciales entre países o flujos de distribución insumo/producto entre industrias. En casos como estos, no cabe esperar que la matriz de proximidades sea simétrica con lo cual, no es apropiado considerar la no simetría como resultado de errores aleatorios o sistemáticos. Por el contrario, las asimetrías observadas son importantes para analizar la estructura interna de los datos y deben ser consideradas para representar las proximidades en el espacio MDS.

El segundo caso aparece en experimentos diseñados para estudiar dife-

rentes tipos de juicios psicológicos de proximidad, tales como medidas de confusión estímulo-estímulo donde los sujetos juzgan si dos estímulos presentados ante ellos son iguales o diferentes, o confusión estímulo-respuesta donde los sujetos deben identificar al estímulo presentado. En el primer tipo, la medida de confusión es el porcentaje de respuestas en las que un par de estímulos diferentes, dados en un orden específico, son considerados iguales (Shepard 1963, Wish 1967). Dados los estímulos i y j , el número de veces en las que los sujetos los consideran iguales cuando son presentados en el orden (i, j) , por lo general es diferente al número de veces en las que los sujetos los consideran iguales cuando son presentados en el orden (j, i) . Si consideramos la tasa de confusión como una medida de proximidad, entonces $p_{ij} \neq p_{ji}$. Si el juicio requerido es del tipo estímulo-respuesta, segundo tipo considerado arriba, la medida de confusión es el porcentaje de veces en las que los individuos responden j cuando el estímulo presentado es i ($i \neq j$) (Zilman y Heiser, 1996). El número de veces en las que los sujetos confunden al estímulo i con el estímulo j es, por lo general, diferente al número de veces en las que los sujetos confunden al estímulo j con el estímulo i . Con lo cual, una vez más, $p_{ij} \neq p_{ji}$.

Tradicionalmente, las desviaciones de la simetría incluidas en el segundo caso han sido tratadas como errores aleatorios y en consecuencia, desestimada al momento de construir la representación. Sin embargo, varios autores han señalado la importancia de la asimetría en la investigación de los aspectos complejos de los juicios de disimilaridad. Zeilman y Heiser (1996) consideran que las confusiones aportan información acerca de la similaridad de los estímulos. Otros han propuesto que las similaridades y disimilaridades dependen de la atención (Shepard, 1964), del contexto y el marco de referencia (Torgerson 1965, Sjöberg 1972) y que en general están conectadas con las tareas cognitivas (Rosh, 1975).

Así, las confusiones que surgen en los juicios de similaridad y que son la fuente de las asimetrías en las proximidades no deben ser consideradas a priori como debidas solamente a ruido. Tales asimetrías, podrían ser el resultado de aspectos de naturaleza sistemática que surgen en las relaciones entre los datos analizados e ignorarlas podría resultar inapropiado para la construcción del espacio MDS.

En este contexto, se propone un modelo de escalamiento multidimensional para analizar datos de disimilaridad no simétricos que incorpora metodologías de los modelos de ecuaciones estructurales, SEM por sus siglas en inglés, para estimar una matriz simétrica latente de proximidades como una función

de las disimilaridades no simétricas observadas, al propio tiempo que los objetos son representados en un espacio de baja dimensión. La matriz simétrica latente estimada, pondera de manera óptima las asimetrías de la matriz de disimilaridades observadas y considera tal ponderación en la construcción del espacio MDS. La estimación de la matriz de disimilaridad simétrica latente y la representación de los estímulos en un espacio de baja dimensión son el resultado de un procedimiento alternante que converge al espacio MDS. Durante el proceso, las submatrices triangulares de la matriz asimétrica observada se consideran indicadores efecto de la matriz simétrica latente. El modelo ha sido denominado Modelo MDSSEM.

Este capítulo continúa con una introducción a los fundamentos del MDS con referencias breves a los tipos de datos y modelos propuestos para analizar datos simétricos. El Capítulo 2 describe con cierto detalle los modelos propuestos hasta la fecha para analizar proximidades asimétricas. El Capítulo 3 trata el tema de los modelos de ecuaciones estructurales. En el Capítulo 4 se describe en detalle el modelo propuesto mientras que el Capítulo 4.5 discute algunas aplicaciones tanto para datos simulados como para datos reales.

1.1. Conceptos Básicos del MDS

El MDS, es una denominación bajo la cual se agrupa a un conjunto de técnicas que han sido diseñadas para desvelar la estructura interna de un conjunto de datos a partir de su representación en un espacio geométrico de baja dimensión, por lo general Euclidiano, conocido como el espacio MDS. Los datos representan alguna medida de proximidad entre objetos, individuos, sujetos o estímulos, donde objeto se refiere a cosas inanimadas como revistas; individuo y sujeto suele referirse a personas o animales mientras que estímulo normalmente se refiere a entidades intangibles como percepciones o juicios (Cox y Cox, 2001).

Una proximidad, que puede ser de similaridad o disimilaridad, es un número que indica cuan similares o diferentes son los objetos o estímulos que analizamos, tal como ellos existen en la realidad o como son percibidos por los sujetos. Desde un punto de vista general, el término *proximidad* se asocia al concepto de cercanía espacial, perceptual, temporal, cultural, económica o en cualquier otro contexto. Las proximidades pueden ser correlaciones entre respuestas a ítems, datos de comercio entre países, valoraciones de similaridad entre candidatos políticos, entre muchas más posibilidades en ámbitos

como educación, sociología, economía, comercio, entre muchas otras.

La representación de las proximidades en el espacio MDS busca dar respuesta a preguntas del tipo: “¿Cómo de similares o diferentes son o nos parecen los objetos que analizamos?”, “¿Qué factores determinan la estructura de similaridad observada?” o “¿Qué características pueden deducirse de la representación de tales objetos en un espacio geométrico, que ayude a entender cómo lo sujetos construyen sus juicios de similaridad?”.

Cuando se trata de estudios de percepción, el cual ha sido un dominio tradicional del MDS, básicamente se busca determinar el número y la importancia relativa de las dimensiones o atributos que subyacen a la evaluación de proximidad que hacen los individuos de un conjunto de estímulos y cómo éstos se relacionan perceptualmente cuando las medidas de proximidad son proyectadas como distancias entre puntos en un espacio geométrico q -dimensional. Así, el objetivo del MDS en este contexto es descubrir, más que imponer, las dimensiones que subyacen a los juicios de proximidad. Por esta razón, no se especifican los atributos con respecto a los cuales los individuos deben evaluar los estímulos, por el contrario, se espera que esta información sea deducible del espacio MDS correspondiente. No obstante, se pueden requerir clases específicas de similaridad como similaridad política o cultural, entre otras (Kruskal y Wish, 1978).

Un ejemplo del uso del MDS para explorar dimensiones subyacentes que ayuden a explicar los juicios de proximidad, es la investigación de Wish (1971) en la que se le pidió a un grupo de 18 individuos que establecieran una taza de similaridad global entre pares de países empleando para ello una escala de 9 puntos, donde 1 equivale a “muy diferentes” y 9 equivale a “muy similares”. Se presentaron 12 naciones y no se instruyó a los individuos sobre los atributos que deberían emplear para construir la taza de similaridad entre ellas. La Figura 1.1 muestra el espacio MDS para los datos de este estudio ¹. Lo que vemos en esta figura es una configuración geométrica bidimensional en la que cada punto representa a un estímulo, en este caso un país, y las distancias entre ellos representan las medidas de similaridad entre los países construidas por los sujetos. Mientras mayor sea la similaridad entre los países, menor será la distancia entre los puntos que los representan.

Una primera aproximación a la interpretación de la solución que es de uso frecuente en las aplicaciones, es examinar visualmente la configuración obtenida en busca de patrones que ayuden a entender la distribución de

¹Más adelante se explica cómo puede ser construido el espacio MDS

los objetos en el espacio MDS. La recomendación general, es trazar líneas sobre la configuración de manera que los objetos proyectados en extremos opuestos de una línea, difieran entre ellos de una manera fácil de describir. Las diferencias se extraen del conocimiento sustantivo que se tiene de los objetos representados por los puntos en el espacio MDS.

Las líneas punteadas en la Figura 1.1, son producto de la interpretación que hacen Kruskal y Wish (1978) para ayudar a entender la configuración obtenida siguiendo la recomendación general dada arriba. Ellos propusieron dos ejes que denominaron “Alineación Política”, eje $\overrightarrow{PS.PO}$ en la figura (PS por Pro Soviéticos, PO por Pro Occidentales) y “Desarrollo Económico”, eje $\overrightarrow{ED.D}$ (ED por En Desarrollo, D por Desarrollados). Efectivamente, los países que para la fecha del estudio eran considerados pro-soviéticos o pro-comunistas, tales como China, antigua Yugoslavia y Cuba, se proyectan hacia el extremo PS del eje, mientras que los considerados pro-occidentales

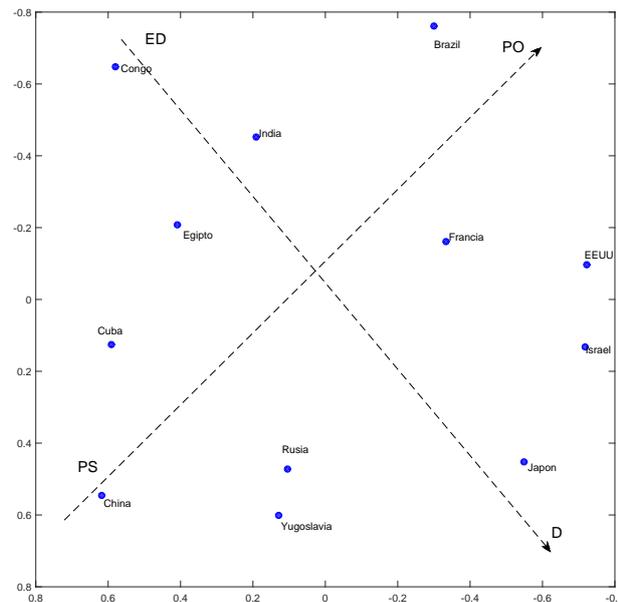


Figura 1.1: Espacio MDS para las tazas de similaridad entre 12 países.

o pro-capitalistas, tales como Israel, Brasil y Francia, se ubican hacia el extremo PO. Análogamente, los países considerados para la fecha como en vías de desarrollo se ubican hacia el extremo ED del eje $\overrightarrow{ED.D}$, mientras que los considerados como desarrollados se ubican hacia el extremo opuesto (D). Así, los ejes propuestos parecen explicar, en buena medida, la distribución de los estímulos en el Espacio MDS y revelan al menos dos de las dimensiones probablemente empleadas por los sujetos para construir la tasa de similaridad global entre las 12 naciones ². En esta ilustración, las líneas rectas resultaron de utilidad para analizar la solución; en otros casos, han sido más útiles curvas suaves (p.ej. círculos) o superficies, así como particiones apropiadas del espacio solución (ver por ejemplo, Borg y Groenen, 2005 pág. 80).

Este ejemplo ilustra otro punto importante relacionado con la interpretación del espacio. Debido a que la información que contiene el espacio MDS en relación con los estímulos analizados está asociada con las distancias entre los puntos, la solución es invariante ante cualquier transformación del espacio que preserve las distancias. Formalmente, si la configuración de puntos dada por P_1, \dots, P_n donde cada P_i tiene coordenadas dadas por $\mathbf{x}_i^t = (x_{i1}, \dots, x_{ip})$, con $i = 1, \dots, n$, es una solución MDS en p dimensiones y si \mathbf{A} es una matriz ortogonal y \mathbf{b} un vector cualquiera, entonces

$$\mathbf{y}_i = \mathbf{A}\mathbf{x}_i + \mathbf{b} \quad i=1, \dots, n$$

es también una solución.

Es decir, la solución es la misma ante cualquier isometría. De hecho, no hay una posición rotacional incorrecta para la configuración. Sin embargo, las coordenadas de los puntos sobre los ejes principales impresos por el programa usado para el escalamiento, cambian drásticamente con cada rotación. Por esta razón, los ejes principales no son susceptibles de interpretación directa y no tienen mayor significación que otra línea trazada en dirección arbitraria.

Para determinar la solución, el MDS opera bajo el supuesto de que dado un conjunto de objetos, individuos o estímulos sobre los cuales se tiene una medida de proximidad, existe una configuración de puntos en un espacio geométrico de baja dimensión donde cada punto representa a un objeto, de tal forma que las distancias entre pares de ellos pueden representar a las proximidades adecuadamente. La expresión *adecuadamente* significa que la configuración obtenida minimiza una función de la diferencia entre las proximidades y las distancias.

²La interpretación completa puede verse en Kruskal y Wish (1978), pág. 30

A modo de introducción al proceso de construcción del espacio MDS, supóngase que disponemos de una matriz $N \times N$ con las distancias entre N ciudades y estamos interesados en conocer su posición geográfica ³. El problema fue resuelto para una dimensión arbitraria por Schoenberg (1935) y Young & Houselholder (1938), en lo que se conoce como la solución clásica de MDS. Antes de presentar este resultado recuérdese que una matriz cuadrada $\mathbf{D} = (d_{rs})$ es una matriz de distancias si es simétrica y si $d_{rs} \geq 0 \forall rs$ (la igualdad se cumple solo si $r = s$). Además, \mathbf{D} es una matriz de distancias Euclidianas, o simplemente Euclidiana, si existen puntos $\mathbf{x}_1, \dots, \mathbf{x}_n$ en un espacio euclidiano E^p tales que $d_{rs}^2 = (\mathbf{x}_r - \mathbf{x}_s)^t(\mathbf{x}_r - \mathbf{x}_s)$. Es decir, \mathbf{D} es una matriz Euclidiana si sus entradas corresponden a las distancias entre pares de puntos de alguna configuración en un espacio Euclidiano.

Teorema 1.1.1 *Sea \mathbf{D} una matriz de distancias entre pares de puntos de una configuración que existe en un espacio de dimensión k . Sea $\mathbf{B} = \mathbf{H}\mathbf{A}\mathbf{H}$ con $\mathbf{H} = \mathbf{I} - n^{-1}\mathbf{1}\mathbf{1}^t$ y $\mathbf{A} = (a_{rs})$ con $a_{rs} = -\frac{1}{2}d_{rs}^2$. Entonces, \mathbf{D} es Euclidiana si y solo si \mathbf{B} es una matriz semidefinida positiva. En particular, se tiene que:*

1. *Si \mathbf{D} es la matriz de distancias Euclidianas entre pares de puntos de una configuración $\mathbf{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_n)'$, entonces $\mathbf{B} = (\mathbf{H}\mathbf{Z})(\mathbf{H}\mathbf{Z})^t$. Es decir,*

$$b_{rs} = (\mathbf{z}_r - \bar{\mathbf{z}})^t(\mathbf{z}_s - \bar{\mathbf{z}}), \quad \forall r, s = 1, \dots, n,$$

Así, $\mathbf{B} \geq 0$ y puede ser interpretada como la matriz de producto interno centrada de la configuración \mathbf{Z} .

2. *Recíprocamente, si \mathbf{B} es semidefinida positiva de rango k ($r(\mathbf{B}) = k$), entonces puede construirse una configuración correspondiente a \mathbf{B} de la manera siguiente: Sean $\lambda_1 > \dots > \lambda_k$ los autovalores positivos de \mathbf{B} y sean $\mathbf{X}_{(n \times k)} = (\mathbf{x}_{(1)}, \dots, \mathbf{x}_{(k)})$ los correspondientes autovectores normalizados por*

$$\mathbf{x}_{(i)}' \mathbf{x}_{(i)} = \lambda_i, \quad \forall i = 1, \dots, k$$

Así, los puntos P_r en \mathbb{R}^k con coordenadas $\mathbf{x}_r = (x_{r1}, \dots, x_{rk})'$ (\mathbf{x}_r es la r -ésima fila de \mathbf{X}) tienen distancias entre pares de ellos dadas por \mathbf{D} .

³Nótese que este problema es el inverso a determinar las distancias entre ciudades a partir de un mapa

Además, esta configuración tiene como centro de gravedad $\bar{\mathbf{x}} = \mathbf{0}$ y \mathbf{B} representa la matriz del producto interno para esta configuración.

El teorema 1.1.1, cuya demostración puede verse, por ejemplo, en Mardia, et. al. (1979), establece la existencia de una solución única para distancias Euclidianas en un espacio de dimensión $k = r(\mathbf{B})$, excepto por transformaciones de los datos que preserven las distancias. Además, debido a que $\mathbf{1}$ es un autovalor de \mathbf{B} asociado al valor propio 0, k a lo sumo será igual a $n - 1$. No obstante, el problema cuya solución es de mayor importancia práctica y que es resuelto por el MDS, es el de la recuperación de la configuración en un espacio de dimensión menor que k (Por ejemplo, $k \leq 3$), o la construcción de la configuración a partir de datos que no son distancias Euclidianas sino pseudo distancias o más generalmente, proximidades. Para ello, el MDS colapsa la solución en un espacio de baja dimensión a partir de modelos que especifican cómo los datos, dados como coeficientes de similitud o disimilitud, son mapeados como distancias entre puntos de una configuración \mathbf{X} con una mínima pérdida de información. Los tipos de datos requeridos y los distintos modelos disponibles en MDS son los temas tratados en las dos secciones que siguen.

1.1.1. Los datos en MDS

El objetivo del MDS es analizar relaciones entre objetos, individuos o estímulos a partir de datos de proximidad. Tales datos pueden ser de similitud, ψ_{ij} , o de disimilitud, δ_{ij} .

Clasificación de los datos en MDS

Como es sabido, las variables pueden ser clasificadas de acuerdo con la escala en la que han sido medidas y en MDS, tal clasificación conduce a distintos modelos para analizar los datos. Las distintas escalas de medida son:

- *Escala nominal.* Los datos medidos en esta escala son categóricos, con lo cual, los valores asignados a los estímulos (sujetos, etc.) solo establecen pertenencia a una, de un conjunto de clases o categorías mutuamente excluyentes y colectivamente exhaustivas. Por ejemplo, la variable *Género* es categórica con dos categorías: hombre - mujer. Es común

asignarle valores a los individuos según la pertenencia a una de ellas, por ejemplo, 1 si es hombre, 2 si es mujer. Sin embargo, no tiene sentido realizar operaciones aritméticas con estos valores ni están sujetos a relaciones de orden (mujer no es mayor que hombre). Así, los datos que surgen de una medición en la escala nominal no son cuantitativos.

- *Escala Ordinal.* Los datos ordinales tienen características análogas a los datos nominales, con la diferencia de que, en este caso, las categorías son ordenables. Así, el valor asignado a un estímulo (sujeto, etc.) perteneciente a una categoría puede ser mayor o menor que el valor asignado a otro perteneciente a una categoría diferente, aunque no pueda determinarse cuanto mayor o menor sea uno del otro. Por ejemplo, la medición de la variable *Calidad* sobre un lote de productos podría asignarles las etiquetas A, B y C como pertenecientes, respectivamente, a las categorías: *Calidad baja*, *Calidad media* y *Calidad alta*. En tal caso, podría decirse que $A < B$ pero no podría determinarse cuanta más calidad posee B respecto de A. Así, los datos medidos en escala ordinal tampoco son cuantitativos.
- *Escala de intervalo.* La escala de intervalo es numérica y la distancia entre las unidades de medida es uniforme, con lo cual, las diferencias entre pares de mediciones pueden compararse significativamente. Así, en esta escala, además de tener cabida las relaciones de orden, también se pueden realizar operaciones de suma y resta. Sin embargo, la escala no posee un cero absoluto, mas bien, el cero es un punto arbitrario de la escala y por tanto las operaciones de multiplicación y división no tienen sentido. Un ejemplo de una variable medida en esta escala, es la temperatura medida en grados centígrados o Fahrenheit. Aquí, 0° no indica ausencia de calor o de energía sino que es un punto arbitrario de la escala lo que hace que las proporciones entre sus valores no son equivalentes si la temperatura se mide en grados centígrados a si se miden en grados Fahrenheit. No obstante, la escala de intervalo se considera una escala cuantitativa.
- *Escala de razón.* Esta escala tiene propiedades similares a la escala de intervalo, con el añadido de que posee un cero absoluto, con lo cual, sus valores además de sumarse y restarse, pueden también multiplicarse y dividirse. Ejemplos de datos medidos en esta escala son las variables *Peso* y *Altura*.

Ahora bien, mas allá de la escala en la que han sido medida las variables que dan lugar a las proximidades, es común en MDS agrupar a los datos según la siguiente clasificación:

- *Numero de modos.* En MDS, se denomina *modo* a cada conjunto de objetos, estímulos, etc., que subyace a los datos. Como una ilustración, supóngase que se quiere comparar distintas marcas de vino provenientes de bodegas diferentes a través de las pruebas que realiza un único catador. El procedimiento podría ser conformar todas las parejas de vino posibles y solicitarle al catador que emita un juicio de disimilaridad entre parejas lo cual podría hacerse, por ejemplo, asignándole el valor de 0 a las parejas cuya diferencia no pueda ser establecida por el catador por que son muy similares en sus características apreciables, y 10 a las parejas que le parecen al catador totalmente diferentes. En este caso, la disimilaridad correspondiente se denotaría por δ_{ij} , donde i y j hacen referencia a los vinos mientras que no hay referencia al catador en la nomenclatura. Así, los datos de esta ilustración son de un modo. Por otro lado, si más de un catador participa en el ejercicio y cada uno compara todas las parejas de vino, entonces la disimilaridad se denotaría por $\delta_{ij,r}$, con r haciendo referencia al catador y los datos serían de dos modos, un modo serían los vinos y otro modo los catadores.
- *Número de vías* Se denomina *vía* a cada índice en la medida entre objetos, estímulos, etc. En la ilustración de los vinos con un solo catador, las proximidades, δ_{ij} , son de una vía debido a que son derivadas a partir de un único conjunto: los vinos. El caso con mas de un catador, se tienen dos conjuntos, los vinos y los catadores, para derivar las proximidades, así que éstas son de dos vías: $\delta_{ij,r}$.

Proximidades

Ya ha sido dicho que el término proximidad, que puede referirse a disimilaridad o similaridad, significa cercanía entre objetos (estímulos, etc.) y que la cercanía puede ser espacial, pero también puede ser cultural, económica, perceptual, entre muchos otros contextos.

Más formalmente, sea O el conjunto que contiene a los objetos, estímulos, etc. Entonces, una proximidad es una función de $O \times O$ en \mathbb{R} que vendrá dada en términos de disimilaridad δ_{ij} , o de similaridad ψ_{ij} entre los objetos i y j . Aunque dada una de ellas es posible obtener la relación en términos de

la contraria, entre ambos conceptos, son las disimilaridades las proximidades que de forma natural están relacionadas con las distancias. Por lo general, $\delta_{ij} \geq 0$ y $\psi_{ij} \geq 0$. Además, $\delta_{ii} = 0$, con lo cual, la disimilaridad de un objeto consigo mismo es 0 lo que conduce a $\delta_{ii} \leq \delta_{ij}$. Por otro lado, en ψ_{ii} la función alcanza un máximo, así que $\psi_{ii} \geq \psi_{ij}$. Es común escalar las similaridades de manera que $\psi_{ij} = 1$. Este escalamiento permite una transformación monótona de las similaridades en disimilaridades preservando la estructura interna de los datos. Tal transformación es útil debido a que algunos modelos de MDS requieren que los datos sean disimilaridades.

Algunos métodos para transformar similaridades en disimilaridades son:

$$\begin{aligned}\delta_{ij} &= 1 - \psi_{ij} \\ \delta_{ij} &= c - \psi_{ij} \text{ para alguna constante } c \\ \delta_{ij} &= \{2(1 - \psi_{ij})\}^{\frac{1}{2}}\end{aligned}\tag{1.1}$$

Las proximidades pueden obtenerse directamente, solicitando a los sujetos que emitan juicios de disimilaridad/similaridad sobre pares de objetos, estímulos, etc., o indirectamente, a partir de puntuaciones sobre atributos asociados con cada uno de los objetos. Entre los métodos para obtener proximidades directamente está el de *comparación por pares (pairwise comparison)* donde cada sujeto tasa a cada par de estímulos de acuerdo con una escala de disimilaridad. En otros contextos tal como en estudios de mercado, la comparación se hace con base al ordenamiento de los estímulos según su disimilaridad global, para lo cual, cada par de objetos es presentado sobre una tarjeta que los individuos deben ordenar colocándolas una sobre otra de tal manera que el par más similar queda en la cima y el menos similar se ubica en la base. Hay versiones de este método que reduce el tiempo y el esfuerzo para lograr un ordenamiento confiable de las tarjetas, tales como el *Q-sort* y el *free sorting*. Una descripción más detallada de estos y otros métodos para obtener proximidades directas, puede verse en Borg & Groenen (2005, pág. 112).

Cuando las proximidades son construidas indirectamente, se denominan *coeficientes de proximidad* y se obtienen a partir de matrices de datos asociados con los objetos. Existen distintas medidas para construir proximidades a partir de matrices de datos. Cox & Cox (2001, pág.21) resume los resultados obtenidos por varios autores acerca de distintas medidas de similaridad/disimilaridad y sus problemas asociados. Algunas de estas medidas se

presentan a continuación.

Sea $\mathbf{X} = (\mathbf{x}_r^t)$ una matriz de datos referida a n objetos sobre p variables, donde \mathbf{x}_r es el vector de observaciones del r -th objeto. Las expresiones siguientes corresponden a métodos para obtener disimilaridades para datos medidos en escala de razón o intervalo.

- Distancia Euclídea

$$\delta_{ij} = \left\{ \sum_r (x_{ir} - x_{jr})^2 \right\}^{\frac{1}{2}} \quad (1.2)$$

- Distancia Euclídea Ponderada

$$\delta_{ij} = \left\{ \sum_r w_r (x_{ir} - x_{jr})^2 \right\}^{\frac{1}{2}} \quad (1.3)$$

- Distancia de Mahalanobis

$$\delta_{ij} = \{(\mathbf{x}_i - \mathbf{x}_j)^t \Sigma^{-1} (\mathbf{x}_i - \mathbf{x}_j)\}^{\frac{1}{2}} \quad (1.4)$$

- Métrica de Minkowski

$$\delta_{ij} = \left\{ \sum_r w_r |x_{ir} - x_{jr}|^\lambda \right\}^{\frac{1}{\lambda}}, \quad \forall \lambda \geq 1 \quad (1.5)$$

- Distancia de Bhattacharyya

$$\delta_{ij} = \left\{ \sum_r (x_{ir}^{\frac{1}{2}} - x_{jr}^{\frac{1}{2}})^2 \right\}^{\frac{1}{2}} \quad (1.6)$$

- Separación Angular

$$\delta_{ij} = 1 - \frac{\sum_r x_{ir} x_{jr}}{\left[\sum_r x_{ir}^2 \sum_r x_{jr}^2 \right]^{\frac{1}{2}}} \quad (1.7)$$

Si las p variables son todas binarias con los números 1 y 0, por ejemplo, indicando la pertenencia a la primera y segunda categoría respectivamente, los estímulos i y j pueden ser comparados con un coeficiente de similaridad construidos sobre la base de los números a , b , c y d , donde a es el número de variables que toman el valor 1 sobre ambos estímulos, b es el número de variables que toman el valor 1 sobre el estímulo i y 0 sobre el estímulo j , c es el número de variables que toman el valor 1 sobre j y 0 sobre i , y d es el número de variables que toman el valor 0 sobre ambos estímulos. Esta medida puede ser transformada luego en un coeficiente de disimilaridad. La siguiente lista es un resumen de la presentada por Hubálek (1982) de coeficientes de similaridad para datos binarios.

- Braun, Blanque

$$s_{ij} = \frac{a}{\text{máx}\{(a+b), (a+c)\}} \quad (1.8)$$

- Czekanowski, Sorensen, Dice

$$s_{ij} = \frac{2a}{2a+b+c} \quad (1.9)$$

- Coeficiente de Jaccard

$$s_{ij} = \frac{a}{a+b+c} \quad (1.10)$$

- Kulczynski

$$s_{ij} = \frac{1}{2} \left(\frac{a}{a+b} + \frac{a}{a+c} \right) \quad (1.11)$$

- Mountford

$$s_{ij} = \frac{2a}{a(b+c) + 2bc} \quad (1.12)$$

- Phi

$$s_{ij} = \frac{ad - bc}{[(a+b)(a+c)(b+d)(c+d)]^{\frac{1}{2}}} \quad (1.13)$$

- Russell, Rao

$$s_{ij} = \frac{a}{a+b+c+d} \quad (1.14)$$

- Coeficiente de concordancia simple

$$s_{ij} = \frac{a + d}{a + b + c + d} \quad (1.15)$$

Los datos pueden ser también una mezcla de variables categóricas, ordinales y cuantitativas. En tales casos, los coeficientes de disimilaridad y similitud listados arriba son inapropiados. Gower (1971) resolvió el problema introduciendo un coeficiente de similitud general, dado por:

$$s_{ij} = \frac{\sum_{r=1}^p \omega_{ijr} s_{ijr}}{\sum_{r=1}^p \omega_{ijr}},$$

donde s_{ijr} es la similitud entre los objetos i y j medida solamente sobre la r -ésima variable y ω_{ijr} vale uno si los estímulos i y j son comparables sobre la r -ésima variable y 0 en caso contrario. De esta forma, s_{ij} es un promedio de todas las similitudes posibles entre los objetos i y j .⁴

Otro tema de interés relativo a las proximidades es la determinación de la significación estadística de los valores obtenidos en relación con los valores esperados. Como es sabido, esto requiere conocer la forma de la distribución del coeficiente de disimilaridad/similitud empleado. Si se puede asumir que el vector de datos sigue una distribución normal multivariante, entonces el problema para algunos coeficientes de disimilaridad de datos cuantitativos, puede ser resuelto. Por ejemplo, si δ_{ij} es la distancia de Mahalanobis, entonces, bajo el supuesto de normalidad, $\delta_{ij} \sim 2\chi^2$.

Para datos categóricos con solo dos categorías, Goodall (1967) halló la media y la varianza para el coeficiente de concordancia simple bajo el supuesto de independencia de las variables:

$$E[s_{ij}] = \mu = p^{-1} \sum_{r=1}^p (p_r^2 + (1 - p_r)^2)$$

y

$$Var[s_{ij}] = \sigma_{ij}^2 = p^{-1} \left\{ \mu(1 - \mu) - p^{-1} \sum_{i=1}^p (p_i + (1 - p_i)^2 - \mu^2)^2 \right\}.$$

⁴Ver Cox & Cox (2001), pág. 21 y siguientes, para más detalles de este y otros coeficientes presentados en esta sección.

Snijders et al. (1990) estudian los momentos para el coeficiente de Jaccard y el coeficiente de Dice y extienden los resultados al caso de variables dependientes. A partir de los momentos de primer orden, pueden obtenerse distribuciones aproximadas de los coeficientes para evaluar la significación de los valores obtenidos.

1.1.2. Naturaleza métrica de las disimilaridades

Las disimilaridades/similaridades deben satisfacer algunas propiedades en orden a ser representadas completamente como distancias entre puntos en el espacio MDS. Estas propiedades, están vinculadas a las propiedades definitorias de una métrica.

Definición 1.1.1 Sea $\Delta = (\delta_{ij})_{n \times n}$ una matriz de disimilaridades tal que $\delta_{ii} = 0 \forall i$. Se dice que Δ es un matriz de disimilaridades métrica, si $\delta_{ij} \leq \delta_{ki} + \delta_{kj} \forall i, j, k$

De la definición se deduce que $\delta_{ii} \leq \delta_{ki} + \delta_{ki}$ con lo cual $\delta_{ki} \geq 0 \forall i, k$. Además, para $k = j$, $\delta_{ij} \leq \delta_{ki} + \delta_{kj} \Rightarrow \delta_{ij} \leq \delta_{ji}$ y para $k = i$, $\delta_{ji} \leq \delta_{kj} + \delta_{ki} \Rightarrow \delta_{ji} \leq \delta_{ij}$. En consecuencia, toda matriz de disimilaridades métrica es simétrica con elementos no negativos. Por otro lado, si $\delta_{ij} = 0$ entonces

$$\delta_{ik} \leq \delta_{ij} + \delta_{kj} \Rightarrow \delta_{ik} \leq \delta_{kj}$$

y

$$\delta_{kj} \leq \delta_{ik} + \delta_{ij} \Rightarrow \delta_{kj} \leq \delta_{ik}$$

así que $\delta_{ik} = \delta_{jk}$. Esta propiedad básica de las métricas puede usarse para establecer que si dos estímulos, digamos i y j , son similares, es decir, $\delta_{ij} \approx 0$, entonces la relación de cualquier otro estímulo, digamos k , con cada uno de ellos será similar. Esto es, $\delta_{ik} \approx \delta_{jk}$.

Gower & Legendre (1986) estudian detalladamente las propiedades métricas de varios coeficientes de disimilaridad. A continuación, se reproduce aquí un resumen que aparece en Cox & Coz (2001), de los resultados más importantes reportados en dicho estudio.

- Si $\Delta^* = (\delta_{ij}^*)$ es una matriz de disimilaridades no métrica, entonces la matriz Δ con elementos $\delta_{ij} = \delta_{ij}^* + c$, con $i \neq j$, es métrica, donde $c \geq \max_{i,j,k} |\delta_{ij}^* + \delta_{ik}^* - \delta_{jk}^*|$.

- Sea $i \neq j$ y c cualquier constante. Si $\Delta = (\delta_{ij})$ es una matriz de disimilaridades métrica, entonces también lo son las matrices con elementos dados por:
 - (i) $\delta_{ij} + c^2$
 - (ii) $\delta_{ij}^{\frac{1}{\lambda}}$ con $\lambda \geq 1$
 - (iii) $\frac{\delta_{ij}}{(\delta_{ij} + c^2)}$
- Si Ψ es una matriz de similaridades semi definida positiva con elementos $0 \leq \psi_{ij} \leq 1 \forall i, j$ y $\psi_{ii} = 1 \forall i$, entonces, la matriz de disimilaridades $\Delta = (\delta_{ij} = (1 - \psi_{ij})^{\frac{1}{2}})$ es Euclidiana.
- Si Δ es una matriz de disimilaridades, entonces existe una constante h tal que la matriz cuyos elementos están dados por $(\delta_{ij}^2 + h)^{\frac{1}{2}}$ es una matriz Euclidiana, donde $h \geq \lambda_n$ con λ_n el menor de los valores propios de $\Lambda_1 = \mathbf{H}\Lambda\mathbf{H}$ y \mathbf{H} es la matriz centrada $\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^t$
- Si Δ es una matriz de disimilaridades, entonces existe una constante k tal que la matriz con elementos dados por $(\delta_{ij} + k)$ es una matriz Euclidiana, donde $k \leq \mu_n$ y μ_n es el mayor de los valores propios de la matriz

$$\begin{pmatrix} \mathbf{0} & 2\Lambda_1 \\ -\mathbf{I} & -4\Lambda_2 \end{pmatrix},$$

con $\Lambda_2 = [-\frac{1}{2}\delta_{ij}]$ y $\Lambda_1 = \mathbf{H}\Lambda\mathbf{H}$ y \mathbf{H} es la matriz centrada $\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^t$.

Los últimos dos resultados resuelven un problema clásico en MDS conocido como el problema de la constante aditiva (Ver Cox & Cox 2001, cap. 2 para más detalles).

En ocasiones, son las variables, más que los estímulos, las que resultan de interés para ser analizadas con técnicas de MDS. Una manera de definir disimilaridades sobre las variables es invertir los roles entre estímulos y variables y emplear alguno de los coeficientes presentados hasta ahora. Sin embargo, es posible definir disimilaridades más apropiadas para las variables. Una posibilidad es emplear el coeficiente de correlación muestral, dado por:

$$r_{ij} = \frac{\sum_r (x_{ir} - \bar{x}_i)(x_{jr} - \bar{x}_j)}{\sqrt{\sum_r (x_{ir} - \bar{x}_i)^2 (x_{jr} - \bar{x}_j)^2}}$$

el cual es un coeficiente de similaridad, mientras que $1 - r_{ij}$ es un coeficiente de disimilaridad. Otra posibilidad, es la separación angular de los vectores de observaciones para los N estímulos sobre las variables i y j , dado por:

$$\delta_{ij} = 1 - \frac{\sum_r x_{ir}x_{jr}}{(\sum_r x_{ir}^2 \sum_r x_{jr}^2)^{\frac{1}{2}}}.$$

Para cada escala de medida, existen transformaciones admisibles que son aquellas que no cambian las relaciones entre los datos y las distancias de la configuración MDS. Es decir, una transformación admisible de los datos no tiene consecuencia sobre la solución del problema porque las distancias, antes y después de la transformación, son las mismas. Por ejemplo,

$$u_{ij} = \left(\frac{1}{n} \sum_r x_{rj}^2 \right)^{-\frac{1}{2}} x_{ij}$$

es una transformación admisible para datos medidos en escala de razón, mientras que

$$u_{ij} = \left(\frac{1}{n-1} \sum_r (x_{rj} - \bar{x}_j)^2 \right)^{-\frac{1}{2}} (x_{ij} - \bar{x}_j)$$

es una transformación admisible para datos medidos en escala de intervalo.

Bajo el supuesto de que el coeficiente de similaridad es invariante bajo transformaciones admisibles, Zegers and ten Berge (1985), Zegers (1986) and Fagot and Mazo (1989) consideran coeficientes de similaridad general para variables que han sido medidas en diferentes escalas. Uno de tales coeficientes basado en la media al cuadrado de la diferencia o media cuadrática, es:

$$s_{ij} = 1 - cn^{-1} \sum_r (u_{ir} - u_{jr})$$

y la constante c se obtiene de manera que $s_{ij} = s_{ji}$, resultando:

$$s_{ij} = \frac{2 \sum_r u_{ir}u_{jr}}{(\sum_r u_{ir}^2 + \sum_r u_{jr}^2)} \quad (1.16)$$

Este coeficiente es usado como una alternativa al coeficiente de correlación.

1.1.3. Modelos en MDS

Sea N el número de objetos, estímulos, etc. bajo consideración y, como antes, sea δ_{ij} la medida de disimilaridad entre los objetos i y j ⁵. El objetivo del MDS es encontrar una matriz \mathbf{X} tal que la distancia entre sus filas, $d_{ij}(\mathbf{X})$, sea aproximadamente igual a las correspondientes disimilaridades, δ_{ij} . La matriz \mathbf{X} es de orden $N \times p$, donde p representa la dimensión del espacio solución la cual es siempre especificada por adelantado por el usuario. La fila i de \mathbf{X} contiene las p coordenadas del objeto i y $d_{ij}(\mathbf{X})$ es por lo general la distancia Euclidiana dada en la ecuación (1.2)

El problema de determinar la matriz \mathbf{X} se formula en un contexto de mínimos cuadrados, mediante la función *Stress bruto* denotada por $\sigma^2(X)$ y cuya expresión es:

$$STRESS = \sum_{i < j} \left(\hat{d}_{ij}^* - d_{ij}(\mathbf{X}^*) \right)^2. \quad (1.17)$$

Esta función fue originalmente definida por Kruskal (Kruskal, 1964a, 1964b) como una manera formal de medir el ajuste de un modelo MDS. El término w_{ij} en (1.17) es una ponderación no negativa que, por lo general, toma el valor de 0 para disimilaridades faltantes. Como se verá mas adelante, además de la expresión en (1.17) hay otras maneras de medir el ajuste del modelo. Todas tienen en común, el que constituyen una medida de la diferencia entre las disimilaridades/similaridades y las distancias que las representan, razón por la cual se les denomina, genéricamente, *función de pérdida*.

Los valores que se obtienen al emplear el *Stress bruto* dependen de la escala en la que están dados los datos, lo cual complica la comparación de modelos corridos sobre datos dados en diferentes escalas (p.ej. centímetros Vs. metros). Esto puede evitarse normalizando el *Stress bruto* por la suma de las distancias al cuadrado; Esto es,

$$STRESS = \frac{\sum_{i < j} \left(\hat{d}_{ij}^* - d_{ij}(\mathbf{X}^*) \right)^2}{\sum_{i < j} d_{ij}^2(\mathbf{X})}. \quad (1.18)$$

En MDS las distancias $d_{ij}(\mathbf{X})$ siempre son desconocidas. Así, el objetivo es encontrar una configuración \mathbf{X} cuyas distancias representen a las disimilaridades tan bien como sea posible. Es decir, se espera que las diferencias

⁵Por simplicidad, en esta sección se hace referencia solo a datos de un modo y dos vías.

entre las disimilaridades y las distancias sean razonablemente pequeñas, lo cual implica minimizar la función en (1.17) o (1.18) para los parámetros contenidos en \mathbf{X} .

Ahora bien, la relación entre las disimilaridades y las distancias está dada por una función de representación

$$f : \delta_{ij} \mapsto d_{ij}(\mathbf{X}) \quad (1.19)$$

Se asume, entonces, que las disimilaridades dadas, después de alguna transformación f , son iguales a las distancias entre los puntos de una configuración \mathbf{X} . Es decir,

$$f(\delta_{ij}) = d_{ij}(\mathbf{X}) \quad (1.20)$$

la forma de la función monótona f especifica el modelo MDS. Si f es paramétrica, el modelo pertenece a la clase de modelos MDS métricos, mientras que si solo se exige que f sea monótona, el modelo pertenece a la clase MDS no métricos. Ambas clases se exponen más adelante en esta misma sección.

En la mayoría de las aplicaciones f no está totalmente definida ya que, aunque pudiera estar restringida a algún tipo de función (exponencial, lineal, etc.), sus parámetros no están determinados a priori. Un caso de interés es el MDS de intervalo, donde

$$f(\delta_{ij}) = a + b\delta_{ij} \quad \forall i, j \quad (1.21)$$

y los parámetros a y b deben ser estimados. En otras ocasiones, lo que interesa es preservar el orden de las disimilaridades. En tal caso, la función f puede ser arbitraria pero debe satisfacer la restricción de monotonía:

$$\delta_{ij} \leq \delta_{rs} \Rightarrow d_{ij}(\mathbf{X}) \leq d_{rs}(\mathbf{X}) \quad \forall i, j, r, s = 1, \dots, N \quad (1.22)$$

Los parámetros de la función f deben estimarse simultáneamente con la matriz \mathbf{X} . La solución a este problema de estimación simultánea, se resuelve a partir de la regresión de las proximidades sobre las distancias calculadas entre las filas de la matriz \mathbf{X} . En MDS métrico se emplea regresión lineal, mientras que en MDS no métrico se emplea regresión monótona. En cada

caso, la regresión produce proximidades transformadas, $f(p_{ij})$, donde f es una transformación admisible de los datos. Las proximidades transformadas reciben el nombre de pseudo-distancias o disparidades y se denotan por \widehat{d}_{ij} .

Las disparidades son incorporadas a la función *Stress* como:

$$\begin{aligned}
\sigma_r(\widehat{\mathbf{d}}, \mathbf{X}) &= \sum_{i < j} w_{ij} (\widehat{d}_{ij} - d_{ij}(\mathbf{X}))^2 \\
&= \sum_{i < j} w_{ij} \widehat{d}_{ij}^2 + \sum_{i < j} w_{ij} d_{ij}^2(\mathbf{X}) - 2 \sum_{i < j} w_{ij} \widehat{d}_{ij} d_{ij}(\mathbf{X}) \\
&= \eta_{\widehat{\mathbf{d}}}^2 + \eta^2(\mathbf{X}) - 2\rho(\widehat{\mathbf{d}}, \mathbf{X})
\end{aligned} \tag{1.23}$$

donde $\widehat{\mathbf{d}}$ denota al vector $s \times 1$ de disparidades con $s = N(N - 1)/2$. Esto es, las disparidades se construyen para todo $i < j$. El proceso para hallar disparidades es conocido como *Escalamiento Óptimo* y fue introducido por Kruskal (1964a, 1964b). El objetivo de estimar las disparidades y las distancias, se alcanza minimizando la función (??) sobre $\widehat{\mathbf{d}}$ y sobre \mathbf{X} . Sin embargo, debido a que (??) es no negativa, puede alcanzarse una solución degenerada en $\widehat{\mathbf{d}} = 0$ y $\mathbf{X} = 0$. La manera de evitarla es imponer una condición de longitud sobre las disparidades de manera que su norma sea igual a una constante fija. De Leeuw (1977) propuso la condición

$$\sum_{i=2}^N \sum_{j=1}^{i-1} w_{ij} \widehat{d}_{ij}^2 = N(N - 1)/2. \tag{1.24}$$

1.1.4. Modelos métricos de MDS

Sea O un conjunto de objetos, estímulos, etc., y δ_{ij} la disimilaridad del par $(i, j) \in O \times O$. Sea $\mathbf{X} = \{x_i : i \in O\}$ el recorrido de una función arbitraria $\phi : O \rightarrow E$ tal que $\phi(i) = x_i$. E es por lo general, un espacio euclidiano aunque no necesariamente. Sea d_{ij} la distancia entre $x_i, x_j \in \mathbf{X}$. El MDS métrico intenta encontrar una función ϕ tal que $d_{ij} \approx f(\delta_{ij})$, para una transformación admisible f paramétrica, monótona y continua.

Los dos modelos de MDS métrico más populares son el escalamiento clásico y el escalamiento de mínimos cuadrados.

Escalamiento Clásico

Este modelo asume que las disimilaridades son distancias entre puntos en un espacio euclidiano y encuentra una configuración \mathbf{X} cuyas coordenadas preservan las distancias. Contribuyeron a su origen y desarrollo Schoenberg (1935) y Young and Householder (1938) quienes derivaron un método para encontrar las coordenadas a partir de las distancias en un espacio euclidiano (Teorema 1.1.1), y posteriormente Torgerson (1952) quien empleó la técnica en el ámbito del escalamiento. Sin embargo, lo que es de mayor interés por su uso práctico es encontrar una configuración de puntos para un conjunto de disimilaridades δ_{ij} , más que para un conjunto de distancias d_{ij} . Esto puede lograrse según el procedimiento que se expone a continuación⁶

Sea $\mathbf{x}_r = (x_{r1} \dots x_{rp})^t$ con $r = 1, \dots, N$, las coordenadas que representan al objeto r en un espacio euclidiano p -dimensional. Así, la distancia Euclidiana entre los puntos \mathbf{x}_r y \mathbf{x}_s viene dada por:

$$d_{rs}^2 = (\mathbf{x}_r - \mathbf{x}_s)^t(\mathbf{x}_r - \mathbf{x}_s) \quad (1.25)$$

Por otro lado, sea $\mathbf{B} = (b_{rs})$ con $b_{rs} = \mathbf{x}_r^t \mathbf{x}_s$ la matriz de producto interno. Así, \mathbf{B} puede obtenerse a partir del cuadrado de las distancias d_{rs} y, a partir de \mathbf{B} , se pueden determinar las coordenadas desconocidas.

Para obtener \mathbf{B} , primero se ubica el centroide de la configuración en el origen para evitar la indeterminación de la solución debido a traslaciones. Así, $\sum_{r=1}^n x_{ri} = 0$ con $i = 1, \dots, p$. Con esta restricción y dado que (1.25) se puede escribir

$$d_{rs}^2 = \mathbf{x}_r^t \mathbf{x}_r + \mathbf{x}_s^t \mathbf{x}_s - 2\mathbf{x}_r^t \mathbf{x}_s,$$

de forma que

$$\frac{1}{n^2} \sum_{r=1}^n \sum_{s=1}^n d_{rs}^2 = \frac{2}{n} \sum_{r=1}^n \mathbf{x}_r^t \mathbf{x}_r. \quad (1.26)$$

Sea $A = (a_{ij})$ donde $a_{ij} = -\frac{1}{2}d_{ij}^2$. La matriz de producto escalares puede obtenerse como

⁶Aquí se presenta un resumen del procedimiento. El desarrollo completo puede verse en Cox & Cox (2001), pág. 45 y siguientes, de donde este resumen ha sido extraído.

$$\mathbf{B} = \mathbf{H}\mathbf{A}\mathbf{H} \quad (1.27)$$

donde $\mathbf{H} = \mathbf{I} - n^{-1}\mathbf{1}\mathbf{1}^t$, de donde \mathbf{B} es una matriz doblemente centrada.

Para recuperar las coordenadas a partir de \mathbf{B} , se tiene que

$$\mathbf{B} = \mathbf{X}\mathbf{X}^t \quad (1.28)$$

donde $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]^t$ es la matriz de coordenadas. Así, el rango de \mathbf{B} , $r(\mathbf{B})$, es

$$r(\mathbf{B}) = r(\mathbf{X}\mathbf{X}^t) = r(\mathbf{X}) = p$$

y \mathbf{B} es simétrica, semidefinida positiva y de rango p , con lo cual, posee p valores propios no negativos y $n - p$ valores propios nulos.

Considerando la descomposición espectral de \mathbf{B} viene dada por

$$\mathbf{B} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^t \quad (1.29)$$

con $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n)$ y λ_i valor propio de \mathbf{B} , para todo i . Además, $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_n]$ son los correspondientes vectores propios, normalizados de forma que $\mathbf{v}_i^t \mathbf{v}_i = 1$.

Debido a que \mathbf{B} tiene $n - p$ valores propios iguales a cero, (1.29) puede escribirse como

$$\mathbf{B} = \mathbf{V}_1 \mathbf{\Lambda}_1 \mathbf{V}_1^t, \quad (1.30)$$

donde $\mathbf{\Lambda}_1 = \text{diag}(\lambda_1, \dots, \lambda_p)$ y $\mathbf{V}_1 = [\mathbf{v}_1, \dots, \mathbf{v}_p]$. Por lo tanto, partiendo de (1.28) se obtiene

$$\begin{aligned} \mathbf{X}\mathbf{X}^t &= \mathbf{B} \\ &= \mathbf{V}_1 \mathbf{\Lambda}_1 \mathbf{V}_1^t \\ &= \left(\mathbf{V}_1 \mathbf{\Lambda}_1^{\frac{1}{2}} \right) \left(\mathbf{\Lambda}_1^{\frac{1}{2}} \mathbf{V}_1^t \right) \end{aligned}$$

con lo cual

$$\mathbf{X} = \mathbf{V}_1 \mathbf{\Lambda}_1^{\frac{1}{2}}, \quad (1.31)$$

donde $\mathbf{\Lambda}_1^{\frac{1}{2}} = \text{diag}(\lambda_1^{\frac{1}{2}}, \dots, \lambda_p^{\frac{1}{2}})$. De esta manera se recuperan las coordenadas a partir de las distancias.

Si los datos originales no son distancias, sino disimilaridades, usando la descomposición espectral es posible obtener una configuración en un espacio euclidiano p -dimensional de manera que las distancias correspondientes sean tales que $d_{rs} = \delta_{rs}$, si y solo si, \mathbf{B} es semidefinida positiva de rango p . Una demostración de este resultado puede verse, por ejemplo, en de Leeuw & Heiser (1982).

Si la matriz de disimilaridades \mathbf{B} no es semidefinida positiva, puede hacerse la transformación

$$\delta'_{rs} = \delta_{rs} + c(1 - \delta^{rs}), \quad (1.32)$$

donde c es una constante apropiada y δ^{rs} es el delta de Kronecker dado por

$$\delta^{rs} = \begin{cases} 1 & \text{si } r = s \\ 0 & \text{en otro caso} \end{cases}$$

Esta transformación hace que la matriz \mathbf{B} con las δ'_{rs} , sea semidefinida positiva y así, es posible encontrar una configuración en un espacio euclidiano tal que $d_{rs} = \delta'_{rs}$ según el procedimiento expuesto arriba.

La transformación en la ecuación (1.32) es conocida como el problema de la constante aditiva que es estudiado, por ejemplo, en Cailliez(1983).

Finalmente, dado que $\mathbf{B}\mathbf{1} = \mathbf{H}\mathbf{A}\mathbf{H}\mathbf{1} = 0$, entonces \mathbf{B} tienen al menos un valor propio igual a 0, por tanto, si \mathbf{B} es semidefinida positiva siempre es posible encontrar una configuración en un espacio euclidiano de dimensión $n - 1$, tal que las distancias asociadas sean iguales a las correspondientes disimilaridades. Si \mathbf{B} no es semidefinida positiva, la dimensión del espacio es el número de valores propios positivos. En ambos casos, esas son las dimensiones máximas. Sin embargo, la solución debe estar dada en un espacio de baja dimensión ($p = 2$ o $p = 3$), para que tenga sentido práctico, para lo cual bastaría tomar, por ejemplo, los dos o tres primeros valores propios de \mathbf{B} .

Puede construirse una medida de variación explicada al usar solamente p valores propios, como sigue: Si \mathbf{B} es semidefinida positiva, por (1.26), se tiene que

$$\frac{1}{2} \sum_{r=1}^n \sum_{s=1}^n d_{rs}^2 = n \sum_{r=1}^n \mathbf{x}_r^t \mathbf{x}_r \quad (1.33)$$

pero

$$n \sum_{r=1}^n \mathbf{x}_r^t \mathbf{x}_r = n \operatorname{tr}(\mathbf{B}) = n \sum_{r=1}^{n-1} \lambda_i,$$

así que

$$\frac{1}{2} \sum_{r=1}^n \sum_{s=1}^n d_{rs}^2 = n \sum_{r=1}^{n-1} \lambda_i. \quad (1.34)$$

El lado derecho de (1.34) es una medida de la variación explicada al considerar $n-1$ dimensiones. Por tanto, una medida de la variación explicada al usar los primeros p valores propios para la representación, está dada por

$$\frac{\sum_{r=1}^p \lambda_i}{\sum_{r=1}^{n-1} \lambda_i}. \quad (1.35)$$

Si \mathbf{B} no es semidefinida positiva, puede emplearse

$$\frac{\sum_{r=1}^p \lambda_i}{\sum_{r=1}^{n-1} |\lambda_i|}. \quad (1.36)$$

Escalamiento métrico de mínimos cuadrados

El escalamiento de mínimos cuadrados encuentra una configuración \mathbf{X} en un espacio de baja dimensión, minimizando una función de pérdida S que incluye una transformación de las disimilaridades, $f(\delta_{rs})$, monótona y posiblemente continua.

Sammon (1969), propuso minimizar la función

$$S = \frac{\sum_{r < s} \delta_{rs}^{-1} (d_{rs} - \delta_{rs})^2}{\sum_{r < s} \delta_{rs}}, \quad (1.37)$$

donde d_{rs} es la distancia euclidiana entre los puntos r y s . La ponderación δ_{rs}^{-1} tiene el efecto de asignarle un peso mayor a las disimilaridades pequeñas en relación a las grandes. El denominador elimina el efecto de la escala en el cálculo de S . El mínimo se obtiene según el procedimiento estándar para funciones de varias variables (ver Cox & Cox, 2001, pág. 60).

Sammon emplea un método de pendiente máxima, de tal forma que si $x_{tk}^{(m)}$ es la m -ésima iteración en el proceso de minimización, entonces

$$x_{tk}^{(m+1)} = x_{tk}^{(m)} - \text{MF} \frac{\partial S}{\partial x_{tk}} / \left| \frac{\partial^2 S}{\partial x_{tk}^2} \right|$$

donde MF es un factor de aceleración de la convergencia que usualmente toma el valor de 0,3 o 0,4. A este fin, otros han empleado un método de relajación heurística (Chang & Lee, 1973) o un tamaño de paso óptimo (Niemann and Weiss, 1979). Sin embargo, la velocidad de la convergencia no es un problema hoy en día.

La función de pérdida de *Mínimos residuos absolutos*, LAR por sus siglas en inglés, dada por

$$\text{LAR} = \sum_{r < s} w_{rs} |d_{rs} - \delta_{rs}| \quad (1.38)$$

fue propuesta por Heisser (1988). Esta función no se ve afectada por valores extremos, tanto como las funciones que consideran el cuadrado de los residuos. Para minimizar LAR, Heisser empleó un algoritmo de mayorización iterativa basado en el trabajo de De Leeuw (1977). Una presentación detallada de este algoritmo y su aplicación en la minimización del Stress, puede verse en Borg & Groenen (2005), cap. 8.

Klein and Dubes (1989) usaron la función de pérdida

$$S = \sum \frac{1}{\sqrt{\sum \delta_{rs}}} \sum \frac{|d_{rs} - \delta_{rs}|}{\delta_{rs}}$$

y el método de optimización denominado *enfriamiento simulado* (simulated annealing). Este método usa cadenas de Markov lo que lo hace bastante eficiente para evitar los mínimos locales. En efecto, la cadena de Markov se deja correr en simulación hasta que alcance el estado estable que corresponde con el mínimo de la función. El algoritmo permite que la función S pueda crecer, con lo cual, se pueden salvar los mínimos locales. Los algoritmos basados en métodos de máxima pendiente siempre tienen como objetivo disminuir la función S en cada paso, así que cuando la función experimenta un incremento el proceso se detiene y puede quedar atrapado en un mínimo local.

Cuando la distribución de los residuos, $\delta_{rs} - d_{rs}$, es conocida, es posible construir regiones de confianza para los parámetros y contrastar hipótesis para estudiar la significación del ajuste de diferentes modelos, entre otras posibilidades de inferencia estadística. Las distribuciones más usadas son la distribución normal y la lognormal que permiten hacer estimaciones en el contexto de máxima verosimilitud, asumiendo además que los errores son independientes. Un ejemplo es la función de pérdida MULTISCALE de Ramsay (1977), dada por

$$\delta_{MU}(\mathbf{X}) = \sum_{i < j} [\log(d_{ij}(X)) - \log(\delta_{ij})]^2.$$

MDS no métrico

El MDS no métrico persigue, básicamente, el mismo objetivo que el MDS métrico. Este es, encontrar una configuración de puntos en un espacio de baja dimensión, por lo general euclidiano, de tal manera que las distancias asociadas representen, tan bien como sea posible, a las disparidades. La formalización de este objetivo es análoga a la construida para el MDS métrico. La diferencia está en que el MDS no métrico busca representar solo las propiedades ordinales de los datos. Así, la función f es arbitraria y solo se requiere que satisfaga la condición de monotonía:

$$\delta_{ij} < \delta_{rs} \Rightarrow \hat{d}_{ij} \leq \hat{d}_{rs} \tag{1.39}$$

Por lo general, d_{ij} es la distancia euclídea, aunque pueden usarse otras como la métrica de Minkowski. Una vez elegidos el espacio y la métrica junto con la manera de calcular las disparidades, el problema de MDS no métrico es encontrar un algoritmo apropiado para minimizar la función de pérdida.

La función de pérdida más comúnmente empleada en MDS no métrico, fue presentada por Kruskal (1964a, 1964b) junto con un algoritmo de minimización desarrollado a partir de resultados previamente obtenidos por Shepard (1962a, 1962b). La forma de esta función es

$$S = \sqrt{\frac{\sum_{r < s} (d_{rs} - \hat{d}_{rs})^2}{\sum_{r < s} d_{rs}^2}} \quad (1.40)$$

Para una revisión del algoritmo de Kruskal, así como de otros procedimientos para MDS no métrico, ver, por ejemplo, Cox & Cox (2001) y Kruskal & Wish (1978).

Capítulo 2

MDS para datos asimétricos

Las disimilaridades pueden no ser simétricas y por lo tanto, no siempre pueden ser representadas adecuadamente como distancias entre puntos en un espacio MDS. Las desviaciones de la simetría pueden ser producto de errores aleatorios, caso en el cual los modelos para datos simétricos aún pueden ser utilizados, o la asimetría puede contener información significativa que debe considerarse para la construcción del espacio MDS, caso en el cual se requieren modelos especiales para analizar los datos. Algunos de los modelos en este último grupo, se basan en la descomposición de la matriz de datos no simétrica en un componente simétrico y un componente antisimétrico. La parte simétrica puede ser analizada, por ejemplo, con MDS clásico (Torgerson 1958; Gower 1966) y la parte antisimétrica por métodos como el *Skew-Symmetry Analysis* (SSA) propuesto por Gower (1977) basado en la descomposición en valores singulares de la matriz antisimétrica. Otros métodos proveen un análisis conjunto de las relaciones simétricas y antisimétricas, como el procedimiento de Escoufier & Grolud (1980) basado en la descomposición espectral de una matriz Hermitiana.

El propósito de este capítulo es presentar algunos de los modelos disponibles en MDS para el análisis de datos asimétricos.

2.1. Modelos basados en producto escalar

En el Capítulo 1 se describen algunos modelos MDS para datos simétricos desarrollados en términos de distancias. También se han desarrollado mode-

los basados en el producto escalar¹. Dada una matriz de producto escalar, puede determinarse una configuración \mathbf{X} que aproxime la matriz de datos, Δ , resolviendo la ecuación $\mathbf{B} = \mathbf{X}\mathbf{X}^t$ empleando (1.29) y (1.31). Si en vez de \mathbf{X} empleamos una transformación ortogonal dada por $\mathbf{Y} = \mathbf{X}\mathbf{T}$ con $\mathbf{T}\mathbf{T}^t = \mathbf{I}$, se tiene que

$$\mathbf{Y}\mathbf{Y}^t = (\mathbf{X}\mathbf{T})(\mathbf{X}\mathbf{T})^t = \mathbf{X}\mathbf{T}\mathbf{T}^t\mathbf{X}^t = \mathbf{X}\mathbf{X}^t$$

Así, \mathbf{X} puede ser rotada y/o reflectada sin afectar la calidad de la solución.

En adelante, usaremos $\Psi = (\psi_{jk})$ para denotar una matriz asimétrica de similaridades poblacional (no observada) y $\mathbf{O} = (o_{jk})$ para referenciar la observación en una muestra.

2.1.1. Modelo DEDICOM

El modelo DEDICOM (DEcomposition into DIrectional COMponents) fue propuesto por Harshman (1978), según la expresión

$$\Psi = \mathbf{X}\mathbf{R}\mathbf{X}^t + \mathbf{E} \quad (2.1)$$

donde $\Psi = (\psi_{jk})$ es una matriz $N \times N$ de similaridades asimétricas, $\mathbf{X} = (x_{jt})$ denota una matriz $N \times p$ de coeficientes que relacionan los objetos con los p aspectos o dimensiones que subyacen a los objetos, $\mathbf{R} = (r_{ts})$ es una matriz asimétrica $p \times p$ que representa relaciones direccionales no simétricas entre las dimensiones y \mathbf{E} es una matriz de residuos. Dada una dimensión p , el procedimiento estima \mathbf{X} y \mathbf{R} que minimicen $\text{tr}(\mathbf{E}^t\mathbf{E})$ en un contexto de mínimos cuadrados. En términos de las componentes matriciales, el modelo puede también escribirse como

$$\psi_{jk} = \sum_{s=1}^p \sum_{t=1}^p x_{jt}x_{ks}r_{ts} + e_{ij}. \quad (2.2)$$

Las dimensiones pueden verse como aspectos con los cuales los objetos se relacionan y la matriz \mathbf{X} contiene los pesos de cada aspecto para cada objeto.

¹para una revisión de cómo construir matrices de producto escalar asociadas a juicios de similaridad, ver por ejemplo Borg & Groenen (2005), cap. 7 y 18

Indeterminación del modelo

Sea \mathbf{T} una matriz $p \times p$ invertible. Sea \mathbf{X}_* la matriz resultado de transformar \mathbf{X} por \mathbf{T} . Esto es, $\mathbf{X}_* = \mathbf{X}\mathbf{T}$ y sea $\mathbf{R}_* = \mathbf{T}^{-1}\mathbf{R}\mathbf{T}^{-1t}$. Así

$$\begin{aligned}\mathbf{X}\mathbf{R}\mathbf{X}^t &= \mathbf{X}_*\mathbf{T}^{-1}\mathbf{T}\mathbf{R}_*\mathbf{T}^t\mathbf{T}^{-1t}\mathbf{X}_*^t \\ &= \mathbf{X}_*\mathbf{R}_*\mathbf{X}_*^t.\end{aligned}\tag{2.3}$$

Por lo tanto, el modelo tiene una indeterminación rotacional para un p dado. Para resolverla, se debe normalizar \mathbf{R} o \mathbf{X} . Si se normaliza \mathbf{R} , se pueden estudiar a través de las columnas de \mathbf{X} las cargas relativas que un objeto tiene sobre diferentes dimensiones. Si se normaliza \mathbf{X} , se puede imponer una de las siguientes condiciones:

$$\sum_{j=1}^n x_{jt} = 1 \quad t = 1, \dots, p\tag{2.4}$$

$$\sum_{j=1}^n x_{jt}^2 = 1 \quad t = 1, \dots, p\tag{2.5}$$

$$\mathbf{X}^t\mathbf{X} = \mathbf{I}_p.\tag{2.6}$$

Tales condiciones pueden ser impuestas antes o después de obtener la solución, en función del algoritmo de estimación empleado. La condición (2.4) es útil cuando $x_{jt} > 0 \quad \forall j = 1, \dots, n; \forall t = 1, \dots, p$.

A partir de (2.2) y bajo la condición (2.4), se obtiene

$$\sum_{j=1}^n \sum_{k=1}^n \psi_{jk} = \sum_{s=1}^p \sum_{t=1}^p r_{st}.\tag{2.7}$$

Por tanto, \mathbf{R} puede ser visto como una versión comprimida de Ψ . Por otro lado y bajo la condición (2.4),

$$\sum_{j=1}^n \sum_{k=1}^n x_{jt}x_{ks}r_{ts} = r_{st},\tag{2.8}$$

con lo cual, r_{st} es la suma de las influencias de los aspectos de t a s . Cuando se impone la condición (2.5), la estimación de \mathbf{X} y \mathbf{R} se hace utilizando un algoritmo alternante. De (2.1) se sigue que

$$\begin{aligned}
 \text{tr}(\Psi) &= \text{tr}(\mathbf{X}\mathbf{R}\mathbf{X}^t) \\
 &= \text{tr}(\mathbf{R}^t(\mathbf{X}^t\mathbf{X})) && \text{por propiedades del operador } \text{tr}(\cdot) \\
 &= \text{tr}(\mathbf{R}^t\mathbf{I}_p) && \text{por la condición (2.6)} \\
 &= \text{tr}(\mathbf{R}).
 \end{aligned}$$

Por lo tanto

$$\sum_{j=1}^n \psi_{jj} = \sum_{t=1}^p r_{tt} \tag{2.9}$$

Transformación del modelo

Es conocido que toda matriz puede descomponerse en la suma de una matriz simétrica y una matriz antisimétrica. Si

$$\mathbf{R}_s = \frac{1}{2}(\mathbf{R} + \mathbf{R}^t) \quad \text{matriz simétrica}$$

y

$$\mathbf{R}_a = \frac{1}{2}(\mathbf{R} - \mathbf{R}^t) \quad \text{matriz antisimétrica}$$

entonces

$$\mathbf{R} = \mathbf{R}_s + \mathbf{R}_a.$$

En consecuencia, el lado derecho de (2.1), prescindiendo del error, puede escribirse como

$$\mathbf{X}\mathbf{R}\mathbf{X}^t = \mathbf{X}\mathbf{R}_s\mathbf{X}^t + \mathbf{X}\mathbf{R}_a\mathbf{X}^t, \tag{2.10}$$

con lo cual, el modelo consiste de un modelo factorial oblicuo que constituye la parte simétrica de la representación, y un modelo antisimétrico. La descomposición en valores singulares de \mathbf{R}_a , puede escribirse como

$$\mathbf{R}_a = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^t, \quad (2.11)$$

con $\mathbf{\Lambda}$ matriz diagonal conteniendo los valores propios de \mathbf{R}_a en la diagonal, con la última entrada nula si p es impar. En virtud de la especificación del modelo, p es mucho más pequeño que n y por lo tanto es razonable asumir que \mathbf{R}_a posee un solo valor propio nulo (Saito & Yadohisa, 2005, cap. 4). Además, \mathbf{V} contiene en sus columnas los correspondientes vectores propios de \mathbf{R}_a normalizados de forma que $\mathbf{v}_i^t \mathbf{v}_i = 1$. La matriz ortogonal \mathbf{V} puede usarse como una matriz de rotación para \mathbf{X} . Sea $\mathbf{X}_* = \mathbf{X}\mathbf{V}$ y $\mathbf{R}_{s*} = \mathbf{V}^t \mathbf{R}_s \mathbf{V}$. Al sustituir (2.11) en (2.10), se obtiene

$$\mathbf{\Psi} = \mathbf{X}_* \mathbf{R}_{s*} \mathbf{X}_*^t + \mathbf{X}_* \mathbf{\Lambda} \mathbf{X}_*^t. \quad (2.12)$$

Pueden obtenerse representaciones gráficas de manera análoga al caso *SSA* (Gower, 1977). Sin embargo, el modelo DEDICOM no da una representación gráfica simple para las relaciones no simétricas entre los objetos (Kiers & Takane, 1994).

Estimación de los parámetros en el modelo DEDICOM

Para la matrix de datos observada, \mathbf{O} , el modelo (2.1) se escribe como

$$\mathbf{O} = \mathbf{A}\mathbf{R}\mathbf{A}^t + \mathbf{E}. \quad (2.13)$$

Bajo la restricción (2.6) y para un p específico, la expresión

$$\phi(\mathbf{X}, \mathbf{R}) = \text{tr}(\mathbf{E}^t \mathbf{E}) = \|\mathbf{X} - \mathbf{A}\mathbf{R}\mathbf{A}^t\|^2, \quad (2.14)$$

constituye un criterio de mínimos cuadrados para estudiar el ajuste del modelo. Para \mathbf{X} fija, el mínimo de ϕ se alcanza en $\mathbf{R} = \mathbf{A}^t \mathbf{X} \mathbf{A}$. Mientras que para \mathbf{R} fija, minimizar ϕ sobre \mathbf{X} equivale a maximizar

$$f(\mathbf{X}|\mathbf{R}) = \text{tr}(\mathbf{X}^t \mathbf{O} \mathbf{X} \mathbf{X}^t \mathbf{O}^t \mathbf{X}). \quad (2.15)$$

Diferentes procedimientos de ajuste se han propuesto para este modelo (e.g. Kiers, 1989 y Kiers et al., 1990). Una vez obtenidas las estimaciones $\hat{\mathbf{X}}$

y $\widehat{\mathbf{R}}$, la bondad de ajuste puede estudiarse examinando por separado la parte simétrica y la parte antisimétrica (Bove and Critchley, 1993).

En efecto, sea $\mathbf{O} = \mathbf{O}_s + \mathbf{O}_a$ y $\mathbf{R} = \mathbf{R}_s + \mathbf{R}_a$ con $\mathbf{O}_s, \mathbf{R}_s$ simétricas y $\mathbf{O}_a, \mathbf{R}_a$ antisimétricas. Entonces, de (2.13) se tiene que

$$\begin{aligned} \text{tr}(\mathbf{E}^t \mathbf{E}) &= \|\mathbf{O} - \mathbf{X} \mathbf{R} \mathbf{X}^t\|^2 \\ &= \|\mathbf{O}_s + \mathbf{O}_a - \mathbf{X}(\mathbf{R}_s + \mathbf{R}_a) \mathbf{X}^t\|^2 \\ &= \|\mathbf{O}_s - \mathbf{X} \mathbf{R}_s \mathbf{X}^t\|^2 + \|\mathbf{O}_a - \mathbf{X} \mathbf{R}_a \mathbf{X}^t\|^2 \end{aligned}$$

La última igualdad es producto de que

$$\|(\mathbf{X}_s - \mathbf{A} \mathbf{R}_s \mathbf{A}^t)(\mathbf{X}_a - \mathbf{A} \mathbf{R}_a \mathbf{A}^t)\| = 0.$$

Por lo tanto, el problema de ajustar el modelo a los datos presentado en (2.14), se trata como dos problemas separados de ajuste por mínimos cuadrados: un ajuste por mínimos cuadrados para el componente simétrico de la descomposición de \mathbf{O} dado por $\|\mathbf{O}_s - \mathbf{X} \mathbf{R}_s \mathbf{X}^t\|^2$ y un ajuste por mínimos cuadrados para el componente antisimétrico dado por $\|\mathbf{O}_a - \mathbf{X} \mathbf{R}_a \mathbf{X}^t\|^2$.

2.1.2. Modelo GIPSCAL

Chino (1978, 1990) propuso el modelo GIPSCAL (Generalized Inner Product SCALing) para analizar simultáneamente la parte simétrica y la parte antisimétrica de una matriz de similaridades no simétrica. El modelo también provee una representación gráfica de las relaciones no simétricas. En notación matricial, el modelo GIPSCAL puede escribirse como

$$\mathbf{\Psi} = \alpha \mathbf{X} \mathbf{X}^t + \beta \mathbf{X} \mathbf{R}_I \mathbf{X}^t + \gamma \mathbf{1} \mathbf{1}^t + \mathbf{E} \quad (2.16)$$

donde \mathbf{X} es de orden $N \times p$ y contiene las coordenadas de los puntos que representan a los estímulos en p dimensiones, $\mathbf{R}_I = (r_{jk})$ es una matriz antisimétrica fija tal que

$$r_{jk} = \begin{cases} (-1)^{j+k-1} & \text{si } j < k \\ -r_{kj} & \text{si } j \neq k \\ 0 & \text{si } j = k. \end{cases} \quad (2.17)$$

Por ejemplo, si $p = 4$ entonces

$$\mathbf{R}_I = \begin{pmatrix} 0 & 1 & -1 & 1 \\ -1 & 0 & 1 & -1 \\ 1 & -1 & 0 & 1 \\ -1 & 1 & -1 & 0 \end{pmatrix}$$

Además, $\mathbf{E}_{N \times N}$ es una matriz de residuos y α, β y γ son escalares que expresan la importancia de cada uno de los tres componentes del modelo. El modelo expresado en (2.16) es una variante del modelo DEDICOM. Chino (1990) replanteó el modelo a partir de la transformación $\mathbf{Y} = \mathbf{TX}$ con $\mathbf{T} = \alpha \mathbf{I}_p + \beta \mathbf{R}_I$, con lo cual (2.16) puede escribirse en términos de productos escalares como

$$\Psi = \mathbf{XY}^t + \gamma \mathbf{11}^t + \mathbf{E}. \quad (2.18)$$

Indeterminación y estimación

El modelo tiene una indeterminación de unidades de escala para α, β y \mathbf{X} debido a que

$$\Psi = \frac{\alpha}{c} \sqrt{c} \mathbf{X} (\sqrt{c} \mathbf{X})^t + \frac{\beta}{c} \sqrt{c} \mathbf{X} \mathbf{R}_I (\sqrt{c} \mathbf{X})^t + \gamma \mathbf{11}^t, \quad (2.19)$$

con $c > 0$.

Chino desarrolló un algoritmo de mínimos cuadrados alternante que permite estimar la matriz \mathbf{X} y los parámetros α, β y γ que minimizan $\mathbf{E}^t \mathbf{E}$.

Si $\widehat{\mathbf{X}} = (\widehat{\mathbf{x}}_j)$, $\widehat{\alpha}$, $\widehat{\beta}$ y $\widehat{\gamma}$ son las estimaciones y $\widehat{\mathbf{x}}_j^t$ es vector fila de \mathbf{X} , entonces los términos $\widehat{\alpha} \widehat{\mathbf{x}}_j^t \widehat{\mathbf{x}}_k + \widehat{\gamma}$ y $\widehat{\mathbf{x}}_j^t \mathbf{R}_I \widehat{\mathbf{x}}_k$, explican el componente simétrico y antisimétrico de la matriz de datos, respectivamente. De esta forma, la parte simétrica es explicada como en MDS clásico por el producto interno de los vectores de coordenadas para i y j .

La explicación de la parte antisimétrica requiere de estructuras geométricas bidimensionales, de tal forma que para cada par de dimensiones se calcula el área del triángulo definido por el origen y las proyecciones de i y j sobre esas dimensiones. Así, el elemento (i, j) es representado por la suma de estas áreas para todos los pares de dimensiones posibles. Evidentemente, para $p > 2$ el esfuerzo es considerable. Kiers y Takane (1994) propusieron una simplificación de este modelo y, basados en ella, presentan una generalización del modelo GIPSCAL.

2.1.3. Modelo GIPSCAL Generalizado

En razón de lo relativamente complicado que puede ser la interpretación del componente asimétrico de la solución en el modelo GIPSCAL, Kiers y Takane (1994) propusieron una simplificación del modelo basados en a una reparametrización de las matrices \mathbf{X} y \mathbf{R}_I . Sea $\mathbf{R}_I = \mathbf{U}\mathbf{M}\mathbf{U}^t$ la descomposición en valores singulares de \mathbf{R}_I , donde \mathbf{M} es una matriz diagonal por bloques con matrices $\begin{pmatrix} 0 & \lambda_t \\ -\lambda_t & 0 \end{pmatrix}$, a lo largo de la diagonal y un cero en la última posición, si n es impar. De (2.17) se sigue que los elementos de \mathbf{R}_I son fijos, así que \mathbf{M} es a su vez una matriz fija. \mathbf{U} es una matriz ortogonal conformada con los correspondientes vectores propios de \mathbf{R}_I .

A partir de aquí, se define $\tilde{\mathbf{X}} = \mathbf{X}\mathbf{U}$ con lo cual, el modelo en (2.16) puede escribirse, usando la referencia muestral, como

$$\mathbf{O} = \alpha\tilde{\mathbf{X}}\tilde{\mathbf{X}}^t + \beta\tilde{\mathbf{X}}\tilde{\mathbf{M}}\tilde{\mathbf{X}}^t + \gamma\mathbf{1}\mathbf{1}^t + E, \quad (2.20)$$

de modo que con estas modificaciones, el componente antisimétrico, \mathbf{O}_a , de la matriz de datos es explicado por $\beta\tilde{\mathbf{X}}\tilde{\mathbf{M}}\tilde{\mathbf{A}}^t$.

Esta representación permite examinar la contribución de cada par de dimensiones por separado (Ver Kiers y Takane para detalles adicionales). El modelo en (2.20) es una forma parametrizada de (2.16). Sin embargo, esta formulación del GIPSCAL, con valores fijos en la matriz \mathbf{R}_I y por lo tanto en \mathbf{M} , no simplifican la interpretación de la solución. En este sentido, Kiers y Takane proponen su modelo GIPSCAL *generalizado* en el cual los elementos no nulos en \mathbf{M} no son fijos sino que deben ser estimados junto con los demás parámetros, aunque se mantiene el requerimiento de que \mathbf{M} debe tener bloques 2×2 antisimétricos a lo largo de la diagonal. Esta modificación del modelo GIPSCAL, se obtiene sustituyendo la matriz \mathbf{R}_I por una matriz antisimétrica arbitraria. Bajo esta consideración, los escalares α y β pueden ser eliminados del modelo al incluir $\sqrt{\alpha}$ en \mathbf{A} siempre que $\alpha > 0$, y β en \mathbf{M} . De esta manera, el modelo (2.20) puede escribirse, una vez eliminadas las tildes, como

$$\mathbf{O} = \mathbf{X}\mathbf{X}^t + \mathbf{X}\mathbf{M}\mathbf{X}^t + \gamma\mathbf{1}\mathbf{1}^t + E, \quad (2.21)$$

que constituye la expresión del modelo GIPSCAL *generalizado*.

Estimación de parámetros del modelo GIPSCAL generalizado

Los parámetros del modelo GIPSCAL *generalizado* son, por un lado, la matriz \mathbf{X} y el coeficiente γ y, por otro lado, los elementos $\delta_1, \dots, \delta_q$ de la matriz \mathbf{M} . Tales parámetros se estiman en un contexto de mínimos cuadrados, de manera que minimicen la expresión

$$\sigma_g(\mathbf{X}, \delta_1, \dots, \delta_q, \gamma) = \|\mathbf{O} - \mathbf{X}\mathbf{X}^t - \mathbf{X}\mathbf{M}\mathbf{A}^t - \gamma\mathbf{1}\mathbf{1}^t\|^2, \quad (2.22)$$

donde \mathbf{M} es una matriz con bloques 2×2 del tipo $\begin{pmatrix} 0 & \delta_l \\ -\delta_l & 0 \end{pmatrix}$ en la diagonal y cero en la última entrada si p es impar. Los δ_l , con $l = 1, \dots, q$, dan cuenta de la importancia de cada par de dimensiones para explicar el componente antisimétrico de \mathbf{O} . Cuando solo hay un par de dimensiones, caso en el que $p = 2$, no hay diferencia entre GIPSCAL y GIPSCAL *generalizado*, la diferencia aparece cuando hay más de un par de dimensiones lo cual ocurre cuando $p \geq 4$.

Para minimizar (2.22) se ha propuesto un algoritmo alternante de mínimos cuadrados que minimiza σ_g alternativamente sobre \mathbf{X} , \mathbf{M} y γ . Si \mathbf{X} y \mathbf{M} son fijas, minimizar (2.22) sobre γ equivale a minimizar una forma cuadrática en γ , cuya solución es

$$\gamma = \mathbf{1}^t(\mathbf{O} - \mathbf{X}\mathbf{X}^t - \mathbf{X}\mathbf{M}\mathbf{X}^t)\mathbf{1}/n^2 \quad (2.23)$$

Para \mathbf{M} y γ fijas, se sigue un procedimiento que minimiza (2.22) sobre cada fila de \mathbf{X} asumiendo en cada caso, que las filas restantes son constantes. La expresión a minimizar es

$$g(\mathbf{x}_i^t) = (o_{ii} - \mathbf{x}_i^t \mathbf{x}_i)^2 + \|\boldsymbol{\phi} - \mathbf{F}\mathbf{x}_i\|^2, \quad (2.24)$$

donde \mathbf{x}_i^t es la i -ésima fila de \mathbf{X} y o_{ii} denota el i -ésimo elemento de la diagonal de $(\mathbf{O} - \gamma\mathbf{1}\mathbf{1}^t)$. Además, $\boldsymbol{\phi} = \begin{pmatrix} \mathbf{o}_{r(i)} \\ \mathbf{o}_{c(i)} \end{pmatrix}$

$$\mathbf{F} = \begin{pmatrix} \mathbf{x}_{-i}(\mathbf{I} - \mathbf{M}^t) \\ \mathbf{x}_{-i}(\mathbf{I} - \mathbf{M}) \end{pmatrix},$$

con $\mathbf{o}_{r(i)}$ y $\mathbf{o}_{c(i)}$ denotando la i -ésima fila y la i -ésima columna de $(\mathbf{O} - \gamma\mathbf{1}\mathbf{1}^t)$ respectivamente, habiendo eliminado el i -ésimo elemento de la diagonal, y \mathbf{X}_{-i} es la matriz \mathbf{X} sin la i -ésima fila. Este problema fue resuelto por Ten Berge (1991).

Si \mathbf{X} y γ se dejan fijas, se puede minimizar (2.22) sobre $\delta_1, \dots, \delta_q$ tomando $\mathbf{M} = \mathbf{K}\mathbf{D}$ donde \mathbf{K} es una matriz fija de la misma forma que \mathbf{M} con los

elementos $\delta_1, \dots, \delta_q$ remplazados por 1s y 1 también en la última entrada de la diagonal si p es impar, y \mathbf{D} es la matriz diagonal que tiene los elementos $\delta_1, \delta_1, \delta_2, \delta_2, \dots, \delta_q, \delta_q$ en la diagonal y cero al final si p es impar. Así, minimizar (2.22) sobre \mathbf{M} con \mathbf{X} y γ fijos, se reduce a minimizar

$$\begin{aligned}
\sigma_g(\mathbf{D}) &= \|(\mathbf{O} - \mathbf{X}\mathbf{X}^t) - \mathbf{X}\mathbf{K}\mathbf{D}\mathbf{X}^t\|^2 \\
&= \|Vec(\mathbf{O} - \mathbf{X}\mathbf{X}^t) - Vec(\mathbf{X}\mathbf{K}\mathbf{D}\mathbf{X}^t)\|^2 \\
&= \|Vec(\mathbf{O} - \mathbf{X}\mathbf{X}^t) - (\mathbf{X} \otimes \mathbf{X}\mathbf{K})Vec(\mathbf{D})\|^2 \\
&= \|Vec(\mathbf{O} - \mathbf{X}\mathbf{X}^t) - (\mathbf{X} \times (\mathbf{X}\mathbf{K}))\mathbf{d}\|^2, \tag{2.25}
\end{aligned}$$

donde $\mathbf{X} \times (\mathbf{X}\mathbf{K})$ denota la matriz $n^2 \times p$ que contiene los productos Kronecker, (\otimes) , de las columnas de \mathbf{X} y $\mathbf{X}\mathbf{K}$ correspondientes (ver Carrol & Chang, 1970, pág. 286), y \mathbf{d} contiene los elementos de la diagonal de \mathbf{D} .

Así, el problema se reduce a minimizar $\|Vec(\mathbf{O} - \mathbf{X}\mathbf{X}^t) - \mathbf{B}\mathbf{w}\|$ con $\mathbf{B}_{n \times q}$ la matriz construida sumando columnas consecutivas de $(\mathbf{X} \times (\mathbf{X}\mathbf{K}))$ sin considerar la última columna si p es impar, y $\mathbf{w} = (\delta_1, \dots, \delta_q)^t$. La solución es

$$\mathbf{w} = (\mathbf{B}^t\mathbf{B})^{-1}\mathbf{B}^tVec(\mathbf{O} - \mathbf{X}\mathbf{X}^t). \tag{2.26}$$

Actualizando alternativamente γ , \mathbf{x} y M se obtiene un algoritmo monótonamente convergente para el modelo GIPSCAL *generalizado*.

En el ámbito de la generalización del producto escalar, se han propuesto otros modelos además de los descritos hasta ahora. Entre ellos está el modelo ASYMSCAL propuesto por Young (1975) basado en la distancia Euclidiana ponderada dada en (1.3) (ver Chino, 1978), y el modelo de Chino & Shiraiwa (1993) basado en una generalización del teorema Young & Householders (1938).

Un modelo propuesto por Holman (1979), se basa en la relación de transitividad dada por

$$(\psi_{ij} > \psi_{ji}) \wedge (\psi_{jk} > \psi_{kj}) \Rightarrow (\psi_{ik} > \psi_{ki}), \tag{2.27}$$

donde $\psi_{..}$ es una medida de similaridad asimétrica entre dos estímulos. Los modelos basados en (2.27) han sido denominados *Modelos de sesgo débil*. En términos del componente antisimétrico, τ_{ij} , de la descomposición de $\psi_{ij} = \gamma_{ij} + \tau_{ij}$ donde γ_{ij} es el componente simétrico, (2.27) puede escribirse como

$$(\tau_{ij} > \tau_{ji}) \wedge (\tau_{jk} > \tau_{kj}) \Rightarrow (\tau_{ik} > \tau_{ki}). \quad (2.28)$$

Holman presentó un caso especial del modelo de sesgo débil que denominó *modelo de similaridad y sesgo*, para el cual

$$\psi_{ij} = F(\gamma_{ij} + r_i + c_j), \quad (2.29)$$

donde F es una función monótona general, γ_{ij} es una similaridad simétrica, r_i es una función de sesgo sobre la i -ésima fila y c_j es una función de sesgo sobre la j -ésima columna. El modelo (2.29) puede formularse como un modelo de sesgo en las filas de Ψ según la expresión

$$\psi_{ij} = F(\lambda_i + p_{ij}), \quad (2.30)$$

con $\lambda_i = r_i - c_i$ y $p_{ij} = c_i + c_j + \gamma_{ij}$. Análogamente, si $\mu_j = c_j - r_j$ y $q_{ij} = r_i + r_j + \gamma_{ij}$ entonces, (2.29) puede formularse en términos del sesgo en las columnas como

$$\psi_{ij} = F(\mu_j + q_{ij}), \quad (2.31)$$

Para un estudio más detallado del modelo de similaridad y sesgo y una revisión extensa de los modelos MDS asimétricos, ver por ejemplo Zielman & Heisser (1996).

2.2. Modelos basados en distancias

Sea $\Delta = (\delta_{ij})$ una matriz $N \times N$ de disimilaridades asimétricas. En adelante, se empleará Δ tanto en la formulación de los modelos como en los algoritmos de estimación.

2.2.1. Preliminares

Saito & Yadohisa, siguiendo a Saito (1991, 1993a) y Saito & Takeda (1990), presentan una clasificación de matrices asimétricas y consideran algunas condiciones asociadas.

Tipos de matrices de disimilaridades asimétricas

T1: $\delta_{ii} \neq 0$ para algún i

T2: $\delta_{ii} \neq 0$ para algún i y no todas son idénticas.

T3: $\delta_{ii} = 0 \quad \forall i$

Los modelos asociados con cada una de estos tipos se describirán bajo las denominaciones Modelo I para las matrices del tipo T1, Modelo II para las matrices del tipo T2 y el modelo Slide-Vector para las matrices del tipo T1. También se describirán las particularidades de los modelos que admiten matrices del tipo T3.

Para las matrices del tipo T1 y T2 se definen las cantidades siguientes:

$$\delta_{ij}^- = \frac{1}{2}(\delta_{ij} + \delta_{ji} - \delta_{ii} - \delta_{jj}) \quad (2.32)$$

$$\delta_{ij}^+ = \frac{1}{2}(\delta_{ij} - \delta_{ji} + \delta_{ii} + \delta_{jj}) \quad (2.33)$$

$$\tilde{\delta}_{ij} = \frac{1}{2}(\delta_{ij} - \delta_{ji} + \delta_{ii} - \delta_{jj}) \quad (2.34)$$

$$\delta_{ij}^* = \delta_{ij} - \delta_{.j} - \delta_{i.} + \delta_{..}, \quad (2.35)$$

donde

$$\delta_{i.} = \frac{1}{n} \sum_{j=1}^n \delta_{ij}, \quad \delta_{.j} = \frac{1}{n} \sum_{i=1}^n \delta_{ij}, \quad \delta_{..} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \delta_{ij}, \quad (2.36)$$

y

$$\delta_{i.}^+ = \frac{1}{n} \sum_{j=1}^n \delta_{ij}^+, \quad \delta_{.j}^+ = \frac{1}{n} \sum_{i=1}^n \delta_{ij}^+, \quad \delta_{..}^+ = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \delta_{ij}^+. \quad (2.37)$$

De (2.33) y (2.34) se sigue que

$$\begin{aligned} \tilde{\delta}_{ij} &= \frac{1}{2}(\delta_{ij} - \delta_{ji} + \delta_{ii} - \delta_{jj}) \\ &= \frac{1}{2}(\delta_{ij} - \delta_{ji} + \delta_{ii} + \delta_{jj} - \delta_{jj} - \delta_{jj}) \\ &= \frac{1}{2}(\delta_{ij} - \delta_{ji} + \delta_{ii} + \delta_{jj}) - \frac{1}{2}(\delta_{jj} + \delta_{jj}) \\ &= \delta_{ij}^+ - \delta_{jj}, \end{aligned}$$

pero $\delta_{jj} = \delta_{jj}^+$, con lo cual

$$\tilde{\delta}_{ij} = \delta_{ij}^+ - \delta_{jj}^+. \quad (2.38)$$

Análogamente,

$$\delta_{j.}^+ + \delta_{.j}^+ = \delta_{jj}^+ - \delta_{..}^+. \quad (2.39)$$

La media de los elementos de la diagonal, $\bar{\delta}_{ii}$, viene dada por

$$\bar{\delta}_{ii} = \frac{1}{n} \sum_{i=1}^n \delta_{ii} = \frac{1}{n} \sum_{i=1}^n \delta_{ii}^+ = \delta_{..}^+. \quad (2.40)$$

Para las ternas ordenadas (k, i, j) y los pares ordenados (i, j) , se tienen las siguientes condiciones para datos asimétricos:

- (i) $\delta_{ki} + \delta_{ij} - \delta_{kj} \geq \delta_{jj}$ Desigualdad triangular generalizada
- (ii) $\tilde{\delta}_{ki} + \tilde{\delta}_{ij} = \tilde{\delta}_{kj}$ Aditividad
- (iii) $\delta_{ij}^* = \delta_{ji}^*$ Simetría centrada
- (iv) $\frac{\delta_{ij} - \delta_{ji}}{\delta_{ii} - \delta_{jj}} = c$ ($\delta_{ii} \neq \delta_{jj}$) Razón constante.

La condición de aditividad puede expresarse como

$$\tilde{\delta}_{ij} = \tilde{\delta}_{i.} - \tilde{\delta}_{.j} = \tilde{\delta}_{.j} - \tilde{\delta}_{i.}, \quad (2.41)$$

con lo cual, bajo aditividad, $\tilde{\Delta} = (\tilde{\delta}_{ij})$ es una matriz de rango 2.

Como un resultado importante para el estudio de modelos de distancia para matrices de disimilaridad asimétricas, Saito (1991) estableció que si Δ es una matriz de disimilaridades asimétricas del tipo T1, las condiciones de aditividad y simetría centrada son equivalentes. Además, si se satisface la desigualdad triangular generalizada, entonces δ_{ij}^- en (2.32) es una métrica.

2.2.2. Modelo I de distancia generalizada

Sea $\Delta = (\delta_{ij})$ una matriz asimétrica de disimilaridades del tipo T1 con $\delta_{ij} \geq 0$ y sea $d_{ij}(\mathbf{X})$ la métrica de Minkowsky dada en (1.5). El modelo

$$\delta_{ij} = d_{ij}(\mathbf{X}) + \theta_j + \phi_k + \gamma \quad \forall i, j = 1, \dots, n, \quad (2.42)$$

fue propuesto por Saito (1991) y denominado Modelo I de distancia generalizada o GDM-I, por sus siglas en inglés. En este modelo, los residuos $r_{ij} = \delta_{ij} - d_{ij}(\mathbf{X})$ pueden ser aproximados por la estructura aditiva $r_{ij} = \theta_j + \phi_k + \gamma$. Los parámetros de (2.42) son \mathbf{X} , θ_i , ϕ_j y γ . Zielman & Heisser (1996) definieron los términos

$$\begin{aligned} b_i &= \frac{1}{2}(\theta_i + \phi_i) \\ c_i &= \frac{1}{2}(\theta_i - \phi_i) \\ g_{ij} &= d_{ij} + b_i + b_j + \gamma. \end{aligned}$$

Así, el modelo puede ahora ser escrito como

$$\delta_{ij} = g_{ij} + c_i - c_j, \quad (2.43)$$

donde g_{ij} es un componente simétrico y $c_i - c_j$ es un componente antisimétrico. El ajuste del modelo GDM-I por mínimos cuadrados, conduce a dos procedimientos de mínimos cuadrados para el ajuste de los componentes simétrico y antisimétrico, respectivamente. Sin embargo, algunos problemas pueden presentarse si se requiere que g_{ij} sea una métrica (ver Saito & Yadohisa, pág. 132).

Indeterminación del modelo

El modelo en (2.42), posee una indeterminación por traslación para los parámetros de la parte aditiva, debido a que

$$\begin{aligned} \delta_{ij}^+ &= (\theta_i + c\gamma) + (\phi_j + c\gamma) + \gamma \\ &= (\theta_i - a) + (\phi_j - b) + (\gamma + a + b), \end{aligned} \quad (2.44)$$

con a, b y c constantes desconocidas. Una solución es hacer $\gamma = 0$. Sin embargo, para retener el parámetro γ la expresión en (2.42) puede descomponerse en la forma

$$\delta_{ij}^- = d_{ij} \quad (2.45)$$

$$\delta_{ij}^+ = \theta_i + \phi_j + \gamma. \quad (2.46)$$

También

$$\tilde{\delta}_{ki} = \theta_i - \theta_j, \quad (2.47)$$

lo cual, comparado con (2.41), conduce a $\theta_i = \tilde{\delta}_i + \alpha$ con α constante.

Una condición necesaria y suficiente para que el modelo GDM-I consiga una configuración \mathbf{X} para las disimilaridades asimétricas δ_{ij} dadas, es que $\Delta = (\delta_{ij})$ satisfaga tanto la desigualdad triangular generalizada como la aditividad.

Ahora bien, si está dada una matriz de disimilaridades asimétricas $\Delta = (\delta_{ij})$ que satisface la desigualdad triangular generalizada y la aditividad, es posible estimar los parámetros $\mathbf{X}, \theta_i, \phi_j$ y γ , para un $\lambda = 2$ de la métrica de Minkowsky dada en (1.5). La descomposición del modelo dada en (2.45) a (2.47), indica que \mathbf{X} y los demás parámetros pueden ser tratados separadamente. La matriz $\mathbf{X} = (x_{ij})$ que representa las coordenadas de los estímulos en un espacio euclidiano p -dimensional, puede ser determinada, más allá de la indeterminación traslacional y rotacional, por MDS métrico (Torgerson, 1952). Si se utiliza la distancia Euclidiana ponderada en lugar de la métrica de Minkowsky, la solución puede obtenerse a partir del procedimiento de Schönemann (1972). Para la estimación de θ_i, ϕ_j y γ pueden considerarse los siguientes casos (ver Bove & Critchley, 1993):

- *Caso 1.* Bajo la restricción

$$\sum_{i=1}^n \theta_i = \sum_{i=1}^n \phi_i = 0 \quad (2.48)$$

la solución es

$$\theta_i = \delta_{i.}^+ - \delta_{..}^+, \quad \phi_i = \delta_{.i}^+ - \delta_{..}^+, \quad \gamma = \delta_{..}^+ \quad j = 1, \dots, n \quad (2.49)$$

- *Caso 2.* Bajo la restricción

$$\sum_{i=1}^n \theta_i = \sum_{i=1}^n \phi_i \quad y \quad \gamma = 0 \quad (2.50)$$

la solución es

$$\theta_i = \delta_{i.} - \delta_{.i} + \delta_{ii}, \quad \phi_i = \delta_{.i} - \delta_{i.} + \delta_{ii}, \quad (2.51)$$

- *Caso 3.* Bajo la restricción

$$\sum_{i=1}^n \theta_i \quad y \quad \gamma = 0 \quad (2.52)$$

la solución es

$$\begin{aligned} \theta_i &= \frac{1}{2}(\delta_{i.} - \delta_{.i} + \delta_{ii} - \bar{\delta}_{ii}) \\ \phi_i &= \frac{1}{2}(\delta_{.i} - \delta_{i.} + \delta_{ii} + \bar{\delta}_{ii}), \end{aligned} \quad (2.53)$$

Estimación de parámetros en el modelo GMD-I

Al incorporar el error $\mathbf{E} = (e_{ij})$, el modelo GDM-I definido en (2.42) puede escribirse como

$$\delta_{ij} = d_{ij}(\mathbf{X}) + \theta_i + \phi_j + \gamma + e_{ij} \quad \forall i, j = 1, \dots, n, \quad (2.54)$$

donde

$$d_{ij}(\mathbf{X}) = \left\{ \sum_r w_r |x_{ir} - x_{jr}|^\lambda \right\}^{\frac{1}{\lambda}}, \quad \forall \lambda \geq 1. \quad (2.55)$$

Si $d_{ij}(\mathbf{X})$ es la distancia Euclidiana, la estimación de los parámetros puede hacerse siguiendo un algoritmo de dos etapas. Primero se aplica MDS métrico a Δ^- y se obtiene una estimación de mínimos cuadrados para \mathbf{X} . Obtenida $\hat{\mathbf{X}}$, se obtienen también las distancias estimadas $\hat{\mathbf{d}}_{ij}$. Sea $a_{ij} = \delta_{ij} - \hat{\mathbf{d}}_{ij}$. La estimación de los parámetros de la estructura aditiva del modelo, se obtiene minimizando

$$Q = \sum_i^n \sum_j^n (a_{ij}(\widehat{\mathbf{X}}) - \theta_i - \phi_j - \gamma)^2, \quad (2.56)$$

y la solución es

$$\widehat{\gamma} = a_{..}, \quad \widehat{\theta}_i = a_{i.} - a_{..}, \quad \widehat{\phi}_i = a_{.i} - a_{..}, \quad (2.57)$$

con $j = 1, \dots, n$ y $a_{i.}, a_{.i}$ y $a_{..}$ los promedios por fila, columna y total de la matriz $\mathbf{A} = (a_{ij})$, respectivamente.

Si el modelo se ajusta directamente a los datos, las estimaciones pueden obtenerse minimizando la función

$$L_1(\mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\phi}, \gamma) = \sum_{i=1}^n \sum_{j=1}^n (\delta_{ij} - (d_{ij}(\mathbf{X}) + \theta_i + \phi_j + \gamma))^2, \quad (2.58)$$

sujeta a las restricciones (2.48) a (2.53). Si $d_{ij}(\mathbf{X})$ es Euclidiana, las estimaciones dadas en (2.57) pueden ser consideradas estimaciones condicionadas asociadas a una matriz arbitraria \mathbf{X} no nula, lo que da paso a la segunda función de pérdida dada solamente en términos de \mathbf{X} por,

$$L_2(\mathbf{X}) = \sum_{i=1}^n \sum_{j=1}^n (\delta_{ij} - (d_{ij}(\mathbf{X}) + \theta_i(\mathbf{X}) + \phi_j(\mathbf{X}) + \gamma(\mathbf{X})))^2. \quad (2.59)$$

Esta función se minimiza sin restricciones para obtener los parámetros estimados, para una dimensión p establecida previamente.

2.2.3. Modelo II de distancia generalizada

Este modelo, conocido como GDM-II por sus siglas en inglés, fue propuesto por Saito and Takeda (1990) como una variante del GDM-I.

Sea $\boldsymbol{\Delta} = (\delta_{ij})$ una matriz de disimilaridades asimétrica del tipo T2, con $\delta_{ij} \geq 0$. La formulación del modelo es

$$\delta_{ij} = d_{ij}(\mathbf{X}) + a\theta_i + b\theta_j + \gamma \quad \forall i, j = 1, \dots, n, \quad (2.60)$$

donde d_{ij} es una métrica. Condicionar a la matriz de disimilaridades a ser del tipo T2, es equivalente a establecer que $a + b \neq 0$ y no todos los θ_i son iguales. Así, puede definirse una constante c como

$$c = \frac{a - b}{a + b}, \quad (2.61)$$

para la cual se satisface la condición de razón constante dada por

$$\frac{\delta_{ij} - \delta_{ji}}{\delta_{ii} - \delta_{jj}} = c. \quad (2.62)$$

Ahora, dado que

$$\delta_{ii} = (a + b)\theta_i + \gamma, \quad (2.63)$$

el modelo GDM-II dado en (2.60) puede escribirse de manera equivalente como

$$\delta_{ij} = d_{ij}(\mathbf{X}) + \frac{1}{2}(1 + c)\delta_{ii} + \frac{1}{2}(1 - c)\delta_{jj}, \quad (2.64)$$

con lo cual el modelo puede explicar las asimetrías en la matriz Δ , a partir únicamente del parámetro c .

Indeterminación del modelo

Si el modelo (2.60) es correcto, puede descomponerse como

$$\delta_{ij}^- = d_{ij} \quad (2.65)$$

$$\delta_{ij}^+ = a\theta_i + b\theta_j + \gamma. \quad (2.66)$$

Así, el modelo implica las siguientes relaciones

$$\delta_{ij}^+ = \frac{1}{2}(1 + c)\delta_{ii} + \frac{1}{2}(1 - c)\delta_{jj} \quad (2.67)$$

$$\tilde{\delta}_{ij} = a(\theta_i + \theta_j). \quad (2.68)$$

Una condición necesaria y suficiente para que el modelo en (2.60) encuentre una representación de las disimilaridades, es que se satisfagan las condiciones de aditividad, de razón constante y la desigualdad triangular generalizada para la matriz $\mathbf{\Delta} = (\delta_{ij})$.

De la condición de razón constante, se sigue que

$$\frac{\delta_{i.} - \delta_{.i}}{\delta_{ii} - \bar{\delta}_{ii}} = c \quad \text{con } i = 1, \dots, n. \quad (2.69)$$

Para la identificación de los parámetros, es necesario imponer algunas restricciones. La parte aditiva del modelo se reformula en los términos siguientes

$$a\theta_i + b\theta_j + \gamma = a\theta'_i + b\theta'_j = \phi_i + \lambda\phi_j, \quad (2.70)$$

con

$$\theta'_i = \theta_i + \frac{\gamma}{a+b}, \quad \phi_i = a\theta'_i, \quad \text{y } \lambda = \frac{b}{a} \neq 1, \quad (2.71)$$

lo que muestra que existe una indeterminación de traslación y multiplicación para la parte aditiva. Sustituyendo λ en (2.61) se obtiene

$$c = \frac{1 - \lambda}{1 + \lambda}. \quad (2.72)$$

Ahora bien, si $\mathbf{\Delta}$ satisface la desigualdad triangular generalizada, la aditividad y la condición de razón constante, entonces se pueden estimar los parámetros del modelo GDM-II especificando $d_{ij}(\mathbf{X})$ como distancia Euclidiana. Análogamente al GDM-I, la descomposición del modelo GDM-II dada en (2.65) y (2.66) permite tratar a \mathbf{X} independiente de los parámetros aditivos. La matriz de representación \mathbf{X} puede ser obtenida por MDS métrico. Las dos restricciones dadas en los casos 1 y 2 siguientes, permiten tratar la indeterminación de los parámetros aditivos. Sea

$$\sigma_\delta = \left(\sum_{i=1}^n (\delta_{ii} - \bar{\delta}_{ii})^2 \right)^{\frac{1}{2}} \quad \text{y} \quad \delta_{dr} = \left(\sum_{i=1}^n \delta_{ii}^2 \right)^{\frac{1}{2}}. \quad (2.73)$$

- *Caso 1.* Bajo la restricción

$$\sum_{i=1}^n \theta_i^2 = 1 \quad \text{y} \quad \sum_{i=1}^n \theta_i = 0 \quad (2.74)$$

la solución es

$$\theta_i = \pm \frac{1}{\sigma_\delta} (\delta_{ii} - \bar{\delta}_{ii}), \quad \gamma = \bar{\delta}_{ii}, \quad (2.75)$$

$$a = \pm \frac{1}{2} (1 + c) \sigma_\delta \quad \text{y} \quad b = \pm \frac{1}{2} (1 - c) \sigma_\delta \quad (2.76)$$

- *Caso 2.* Bajo la restricción

$$\sum_{i=1}^n \theta_i^2 = 1 \quad \text{y} \quad \gamma = 0 \quad (2.77)$$

la solución es

$$\theta_i = \pm \frac{\delta_{ii}}{\delta_{dr}}, \quad (2.78)$$

$$a = \pm \frac{1}{2} (1 + c) \delta_{dr} \quad \text{y} \quad b = \pm \frac{1}{2} (1 - c) \delta_{dr} \quad (2.79)$$

Estimación de parámetros en el GDM-II

Debido a la indeterminación de la parte aditiva, el modelo GDM-II puede reescribirse como

$$\delta_{ij} = d_{ij}(\mathbf{X}) + \theta_i + \lambda \theta_j + \gamma + e_{ij} \quad (2.80)$$

$$= d_{ij}(\mathbf{X}) + \phi_i + \lambda \phi_j + e_{ij}, \quad (2.81)$$

donde

$$\phi_i = \theta_i + \frac{\gamma}{1 + \lambda} \quad (2.82)$$

y

$$d_{ij}(\mathbf{X}) = \left\{ \sum_r w_r |x_{ir} - x_{jr}|^\lambda \right\}^{\frac{1}{\lambda}}, \quad \forall \lambda \geq 1. \quad (2.83)$$

Si $\Delta^- = (\delta_{ij}^-)$ es una métrica, entonces $d_{ij}(\mathbf{X})$ puede ser especificada a tener una estructura particular. Si $d_{ij}(\mathbf{X})$ es distancia Euclidiana, entonces c , o equivalentemente λ , puede estimarse y entonces encontrar \mathbf{X} . Para ello, puede tomarse (2.69) como \hat{c} y usarla para obtener $\hat{\lambda}$ a partir de (2.72). Luego, asumiendo que se satisfacen (2.65) y (2.66), se tiene que

$$\delta_{ij}^- = d_{ij}(\mathbf{X}) \quad (2.84)$$

$$\delta_{ij} = d_{ij}(\mathbf{X}) + \frac{1}{2}(1+c)\delta_{ii} + \frac{1}{2}(1-c)\delta_{jj} \quad (2.85)$$

$$\delta_{ij}^+ = \phi_i + \lambda\phi_j, \quad (2.86)$$

y \mathbf{X} puede obtenerse aplicando MDS métrico a (2.84). Otra manera de obtener a \mathbf{X} es calcular $d_{ij}(\mathbf{X})$ sustituyendo \hat{c} en (2.85) y aplicar MDS métrico a las distancias derivadas. Si se especifica $d_{ij}(\mathbf{X})$ y se requiere $p \neq 2$, es necesario el uso de estimación no lineal sobre (2.85) para obtener $\hat{\mathbf{X}}$ y \hat{c} .

Finalmente, con $\hat{\mathbf{X}}$ y \hat{c} conocidas, se obtiene $\hat{\lambda}$ a partir de (2.72) y se sustituye en (2.86) para obtener una estimación de mínimos cuadrado de ϕ_i .

Ahora bien, si $\Delta = (\delta_{ij})$ es del tipo T3, es decir $\delta_{ii} = 0$ para todo i , las definiciones dadas en (2.32) a (2.35) se reducen a

$$\delta_{ij}^- = \frac{1}{2}(\delta_{ij} + \delta_{ji}) \quad (2.87)$$

$$\delta_{ij}^+ = \frac{1}{2}(\delta_{ij} - \delta_{ji}) = \tilde{\delta}_{ij}, \quad (2.88)$$

y δ_{ij}^* es la misma que en (2.35). La desigualdad triangular generalizada es ahora la desigualdad triangular. Análogamente a una matriz de tipo T1, para una matriz del tipo T3 la desigualdad triangular se satisface si y solo si, se satisface la aditividad. Además, si se satisface la desigualdad triangular entonces δ_{ij}^- satisface los axiomas de una métrica.

Si $\Delta = (\delta_{ij})$ es de tipo T3, entonces para el modelo GDM-I dado por $\delta_{ij} = d_{ij} + \theta_i + \phi_i + \gamma$, se tiene que $\delta_{ii} = \theta_i + \phi_i + \gamma = 0$ así que $\gamma = -\theta_i - \phi_i$, sustituyendo esta expresión en el modelo se obtiene

$$\delta_{ij} = d_{ij} - \phi_j - \phi_i, \quad (2.89)$$

o, de manera equivalente,

$$\delta_{ij} = d_{ij} + \theta_i - \theta_j. \quad (2.90)$$

Por otro lado, para el GDM-II dado por $\delta_{ij} = d_{ij} + a\theta_i + b\theta_j + \gamma$, se tiene que $\delta_{ii} = a\theta_i + b\theta_i + \gamma = 0$ con lo cual $\gamma = -a\theta_i - b\theta_i$. Al sustituir esta expresión en el modelo, se obtiene

$$\delta_{ij} = d_{ij} + b\theta_j - b\theta_i, \quad (2.91)$$

o, de forma equivalente

$$\delta_{ij} = d_{ij} + a\theta_i - a\theta_j. \quad (2.92)$$

Sustituyendo $a\theta_i$ por θ_i se obtiene la misma expresión que en (2.90). La descomposición de este modelo está dada por

$$\delta_{ij}^- = d_{ij}(\mathbf{X}) \quad (2.93)$$

$$\delta_{ij}^+ = \theta_j - \theta_i = \tilde{\delta}_{ij}. \quad (2.94)$$

El modelo cuya formulación está dada por (2.90) provee una representación \mathbf{X} para Δ en un espacio de dimensión p , si y solo si, Δ satisface la desigualdad triangular generalizada y la condición de aditividad.

2.2.4. Modelos del tipo Distancia-Densidad

En el estudio de la representación MDS de proximidades asimétricas, también se ha propuesto una clase de modelos en los que se asume que las proximidades observadas no solo son una función de las distancias entre los puntos de la configuración \mathbf{X} , sino que también dependen de la densidad espacial de otros puntos en la región circundante a la configuración. En este sentido, Krumhansl (1978) propuso un modelo desarrollando la noción sugerida por Birnbaum (1974), Parducci (1973), Parducci & Marshall (1961) y Parducci & Perrett (1971), acerca del efecto de la densidad espacial sobre las

medidas de proximidad. El supuesto central de tales modelos, conocidos como modelos de distancia-densidad, es que dos puntos en una región densa del espacio de los estímulos, tendrían una medida de proximidad menor que dos puntos con igual distancia entre ellos pero localizados en una región menos densa del espacio. Aquí se expone una versión métrica propuesta por Saito & Yadohisa (2005) del modelo de distancia-densidad original de Krumhansl.

La formulación del modelo para una matriz de disimilaridades $\Delta = (\delta_{ij})$ del tipo T2, es

$$\delta_{ij} = d_{ij}(\mathbf{X}) + \alpha\rho_i(\mathbf{X}) + \beta\rho_j(\mathbf{X}) + \gamma, \quad (2.95)$$

donde α y β son parámetros positivos y ρ es una función de densidad espacial para la cual han sido propuestas dos definiciones:

$$\rho_i(\mathbf{X}) = \sum_{i \neq j}^n g(d_{ij}(\mathbf{X})), \quad (2.96)$$

donde $g(\cdot)$ es una función decreciente de la distancia, tal como $g(d) = d^{-h}$ o $\exp(-hd)$, y

$$\rho_i(\mathbf{X}) = \sum_{j=1}^n \varepsilon_{ij}, \quad (2.97)$$

donde $\varepsilon_{ij} = 1$ si $d_{ij}(\mathbf{X}) \leq \rho$ y $\varepsilon_{ij} = 0$ en otro caso. Así, $\rho_i(\mathbf{X})$ indica el número de puntos dentro de un radio fijo ρ .

Haciendo $i = j$ en (2.95) se obtiene

$$\delta_{ii} = (\alpha + \beta)\rho_i + \gamma. \quad (2.98)$$

Dado que $\alpha, \beta > 0$ se sigue que $\rho_i > \rho_j \Leftrightarrow \delta_{ii} > \delta_{jj}$. Por otro lado,

$$\begin{aligned} \delta_{ij} - \delta_{ji} &= (d_{ij}(\mathbf{X}) + \alpha\rho_i(\mathbf{X}) + \beta\rho_j(\mathbf{X}) + \gamma) \\ &\quad - (d_{ji}(\mathbf{X}) + \alpha\rho_j(\mathbf{X}) + \beta\rho_i(\mathbf{X}) + \gamma) \\ &= \alpha(\rho_i - \rho_j) + \beta(\rho_j - \rho_i) \\ &= (\alpha - \beta)(\rho_i - \rho_j). \end{aligned} \quad (2.99)$$

Cuando $\alpha > \beta$, entonces $\delta_{ij} > \delta_{ji} \Leftrightarrow \delta_{ii} > \delta_{jj}$, y cuando $\alpha < \beta$, entonces $\delta_{ij} < \delta_{ji} \Leftrightarrow \delta_{ii} > \delta_{jj}$. En consecuencia, los elementos de la diagonal están relacionados con la dirección de la asimetría.

Si ρ_i es reemplazado por θ_i , el modelo (2.95) se transforma en el modelo GDM-II dado en (2.60) y satisface (2.69), con lo cual una condición necesaria y suficiente para que el modelo de distancia-densidad dado en (2.95) pueda representar adecuadamente a las disimilaridades observadas, es que para la matriz $\Delta = (\delta_{ij})$ se satisfagan la desigualdad triangular generalizada, la aditividad y la condición de razón constante. El modelo (2.95) se descompone como

$$\delta_{ij}^- = d_{ij}(\mathbf{X}) \quad (2.100)$$

$$\delta_{ij}^+ = \alpha\rho_i(\mathbf{X}) + \beta\rho_j(\mathbf{X}) + \gamma. \quad (2.101)$$

Para estimar los parámetros del modelo, es necesario especificar las funciones $d_{ij}(\mathbf{X})$ y $\rho_i(\mathbf{X})$ en orden a evitar que (2.95) constituya un conjunto de ecuaciones no lineales. Si $d_{ij}(\mathbf{X})$ es la distancia Euclidiana, entonces \mathbf{X} puede estimarse a partir de (2.92) usando MDS ordinario. La estimación de los parámetros α , β y γ se hace a partir de (2.93). Para ello, se obtiene c , como ha sido definida en (2.61), usando (2.62) o (2.69). Puede fijarse $\gamma = 0$ debido a que ρ_i no debería tener indeterminación traslacional. Una vez conocido el valor de la constante c y con $\gamma = 0$, α y β pueden estimarse bajo alguna de las dos situaciones siguientes:

- *Caso 1.* Fijar α o β a 1. Por ejemplo, si $\alpha = 1$ entonces, a partir de (2.61) con $a = \alpha$ y $b = \beta$, se obtiene que

$$\beta = \frac{1 - c}{1 + c} \quad (2.102)$$

con lo cual, de (2.93)

$$\rho_i(\mathbf{X}) = \frac{1}{1 + \beta}\delta_{ii} = \frac{1}{2}(1 + c)\delta_{ii} \quad (2.103)$$

- *Caso 2.* Para determinar los valores relativos de la densidad, hacemos

$$\sum_{i=1}^n (\rho_i(\mathbf{X}))^2 = 1 \quad (2.104)$$

y así, por (2.73)

$$\rho_i(\mathbf{X}) = \frac{\delta_{ii}}{\delta_{dr}} \quad (2.105)$$

$$\alpha = \frac{1}{2}(1+c)\delta_{dr} \quad \text{y} \quad \beta = \frac{1}{2}(1-c)\delta_{dr} \quad (2.106)$$

En ambos casos,

$$\alpha\rho_i(\mathbf{X}) = \frac{1}{2}(1+c)\delta_{ii} \quad \text{y} \quad \beta\rho_i(\mathbf{X}) = \frac{1}{2}(1-c)\delta_{ii}. \quad (2.107)$$

En este caso el parámetro c es suficiente para explicar la asimetría.

El modelo de Weeks y Bentler

Sea $\mathbf{\Delta} = (\delta_{ij})$ una matriz de disimilaridades asimétricas con $\delta_{ij} \geq 0$ y $\delta_{ii} = c$. Para una matriz tal, Weeks & Bentler (1982) propusieron un modelo dado por

$$\delta_{ij} = \beta d_{ij} + \eta_i - \eta_j + \gamma \quad \text{con} \quad i, j = 1, \dots, n, \quad (2.108)$$

donde $\beta > 0$ y

$$d_{ij}(\mathbf{X}, \mathbf{W}) = ((x_i - x_j)^t \mathbf{W} (x_i - x_j))^{\frac{1}{2}}, \quad (2.109)$$

donde \mathbf{W} es una matriz $p \times p$ definida positiva. Para este modelo, la desigualdad triangular generalizada y la condición de aditividad constituyen una condición necesaria y suficiente.

Weeks & Bentler propusieron sendos procedimientos de estimación de mínimos cuadrados, para tratar por separado la parte simétrica y la parte antisimétrica del modelo. Estos son

$$Q_1(\mathbf{X}, \mathbf{W}, \gamma) = \sum_{i=1}^n \sum_{j=1}^n (\delta_{ij}^s - \beta d_{ij}(\mathbf{X}, \mathbf{W}) + \gamma)^2, \quad (2.110)$$

y

$$Q_2(\boldsymbol{\eta}) = \sum_{i=1}^n \sum_{j=1}^n (\delta_{ij}^a - \eta_i + \eta_j)^2, \quad (2.111)$$

donde δ_{ij}^s y δ_{ij}^a corresponden a la parte simétrica y antisimétrica respectivamente, de la descomposición $\delta_{ij} = \delta_{ij}^s + \delta_{ij}^a$ derivada, a su vez, de la descomposición $\Delta = \Delta_s + \Delta_a$, donde $\Delta_s = (\Delta + \Delta^t)/2$ y $\Delta_a = (\Delta - \Delta^t)/2$. Si en (2.110), d_{ij} es Euclidiana y $\beta = 1$, la estimación de \mathbf{X} se reduce a un problema de MDS métrico con una constante aditiva. Por otro lado, (2.111) conduce a aproximar a Δ_a por una matriz antisimétrica de rango 2.

2.2.5. El modelo Slide-Vector

Zielman & Heisser (1993) propusieron el modelo *slide-vector* para analizar matrices de disimilaridad asimétricas del tipo T1. El modelo representa las asimetrías por un desplazamiento uniforme del vector diferencia entre los dos puntos de un espacio p -dimensional. El modelo se formula como

$$\delta_{ij}^2(\mathbf{X}, \mathbf{z}) = \sum_{t=1}^p (x_{it} - x_{jt} + z_t)^2. \quad (2.112)$$

Como siempre, \mathbf{X} es una matriz de coordenadas $n \times p$ y \mathbf{z} es el slide-vector. De acuerdo con esto,

$$\delta_{ii} = \left(\sum_{t=1}^p z_t^2 \right)^{\frac{1}{2}} = l_z \neq 0. \quad (2.113)$$

Para consideraciones gráficas del modelo, ver, por ejemplo el cap. 4 de Saito & Yadohisa. El modelo postula que a mayor disimilaridad entre los estímulos, existe un mayor grado de asimetría. Además, los estímulos posicionados en la misma dirección del slide-vector dominan a los otros en las relaciones asimétricas.

El desarrollo de (2.114) produce

$$\delta_{ij}^2 = \sum_{t=1}^p (x_{it} - x_{jt})^2 + \sum_{t=1}^p (z_t)^2 + 2 \sum_{t=1}^p z_t (x_{it} - x_{jt}). \quad (2.114)$$

Indeterminación del modelo

Sea

$$\delta_{ij}^2 = s_{ij} + a_{ij}, \quad (2.115)$$

donde

$$s_{ij} = \frac{1}{2}(\delta_{ij}^2 + \delta_{ji}^2) \quad (2.116)$$

$$a_{ij} = \frac{1}{2}(\delta_{ij}^2 - \delta_{ji}^2). \quad (2.117)$$

La expresión en (2.115) constituye la descomposición del cuadrado de las disimilaridades en un componente simétrico y uno antisimétrico, lo cual corresponde a la descomposición del modelo en la forma

$$s_{ij} = d_{ij}^2 + l_z^2 \quad (2.118)$$

$$a_{ij} = 2\mathbf{z}^t(\mathbf{x}_{(i)} - \mathbf{x}_{(j)}), \quad (2.119)$$

donde

$$d_{ij}^2 = \sum_{t=1}^p (x_{(i)} - x_{(j)})^2. \quad (2.120)$$

Para ver la indeterminación de los parámetros, sea \mathbf{T} una matriz ortogonal y sea $\boldsymbol{\xi}$ un vector de traslación. Entonces, bajo las transformaciones $\tilde{\mathbf{X}} = \mathbf{X}\mathbf{T} + \boldsymbol{\xi}$ y $\tilde{\mathbf{z}} = \mathbf{T}^t\mathbf{z}$ se tiene que

$$d_{ij}(\tilde{\mathbf{X}}) = d_{ij}(\mathbf{X}) \quad \text{y} \quad z^t z = \tilde{\mathbf{z}}^t \tilde{\mathbf{z}} = l_z^2, \quad (2.121)$$

con lo cual s_{ij} y a_{ij} resultan invariantes bajo rotaciones y traslaciones y se requieren algunas restricciones para tratar con tal indeterminación.

La matriz $\mathbf{A} = (a_{ij})$ del componente antisimétrico de $\boldsymbol{\Delta}$ cuyos elementos están dados en (2.108), se declara de tal forma que

$$\mathbf{A} = 2(\mathbf{X}\mathbf{z}\mathbf{1}^t - \mathbf{1}\mathbf{z}^t\mathbf{X}^t) = 2(\boldsymbol{\tau}\mathbf{1}^t - \mathbf{1}\boldsymbol{\tau}^t), \quad (2.122)$$

donde $\boldsymbol{\tau} = \mathbf{X}\mathbf{z}$, con lo cual \mathbf{A} es de rango 2 y $a_{ij} = 2(\tau_i - \tau_j)$. A partir de (2.124) puede derivarse la expresión

$$\mathbf{A}\mathbf{A}\mathbf{1} = \lambda\mathbf{1}, \quad (2.123)$$

con

$$\lambda = 4(\theta_1^2 - n\theta_2)\mathbf{1}, \quad \theta_1 = \mathbf{1}^t\boldsymbol{\tau} \quad \text{y} \quad \theta_2 = \boldsymbol{\tau}^t\boldsymbol{\tau} \quad (2.124)$$

Por otro lado, definiendo

$$q_{ij} = \frac{1}{2}(\delta_{ij}^2 + \delta_{ji}^2 + \delta_{jj}^2 + \delta_{ii}^2), \quad (2.125)$$

si q_{ij} es no negativo, entonces de (2.125) se sigue que

$$\delta_{ij}^2 + \delta_{ji}^2 \geq \delta_{jj}^2 + \delta_{ii}^2, \quad (2.126)$$

por lo tanto, si se satisface la desigualdad anterior entonces q_{ij} es una métrica.

Sea $\mathbf{Q} = (q_{ij})$ y \mathbf{H} una matriz centrada y sea

$$\mathbf{B} = \mathbf{H}\mathbf{Q}\mathbf{H}. \quad (2.127)$$

Serán necesarias las siguientes tres condiciones para que el modelo slide-vector pueda representar a una matriz de disimilaridades $\mathbf{\Delta}$ del tipo T1:

- *Condición 1:* la matriz $\mathbf{\Delta}$ satisface la aditividad.
- *Condición 2:* La matriz \mathbf{A} es de rango 2. El gráfico bidimensional de los estímulos en términos de los vectores singulares de \mathbf{A} revela un patrón lineal sobre el plano. El espaciamiento de los puntos sobre la línea corresponde a la proyección de los estímulos sobre el slide-vector.
- *Condición 3:* $q_{ij} \geq 0 \forall i, j = 1, \dots, n$. \mathbf{B} es semidefinida positiva de rango p .

Saito (2003) muestra que, bajo algunos supuestos, las condiciones 1 y 3 son necesarias y suficientes para el modelo. Para la estimación de los parámetros del modelo slide-vector, Zielman & Heisser sugirieron un algoritmo de unfolding restringido en dos etapas, una etapa mejora la localización de los puntos y la otra resuelve el problema de la proyección métrica. Un paso adicional provee el escalamiento de las disimilaridades.

Capítulo 3

Modelos de Ecuaciones Estructurales

3.1. Introducción

En las últimas décadas del siglo pasado, los modelos de ecuaciones estructurales, SEM por sus siglas en inglés, emergieron como una versátil herramienta estadística para el análisis de datos multivariantes en una gran variedad entre las cuales están la sociología, psicología, educación, estudios de mercado, finanzas y ciencia política. Sin embargo, las raíces de SEM se remontan a la primera mitad del siglo pasado cuando Spearman (1904, 1927) desarrolló el análisis factorial y Wright (1918, 1921, 1934, 1960) introdujo el análisis de trayectorias para modelar relaciones complejas entre variables observadas, empleando coeficientes de correlación y análisis de regresión. En la década de los 60, Karl Jöreskog contribuye a desarrollar el Análisis Factorial Confirmatorio (CFA por sus siglas en inglés) basado en los trabajos de Howe (1955), Anderson & Rubin (1956) y Lawley (1958). CFA es empleado con frecuencia para probar si la variabilidad observada entre un conjunto de ítems está determinada por un constructo subyacente y para examinar patrones de interrelaciones entre constructos. Estos, son variables hipotéticas o latentes y por lo tanto no observables, que subyacen a los ítems y pueden explicar la mayor parte de su variabilidad conjunta. SEM, en esencia, combina el análisis de trayectorias y el CFA en una herramienta de análisis estadístico que permite el empleo simultáneo de variables observadas y variables latentes. Un punto de inflexión en el desarrollo de los modelos SEM, se

produjo cuando Jöreskog expuso que las variables observadas en los modelos econométricos están sujetas a errores de medición y propuso sustituirlas por variables latentes.

Un impulso notable que condujo a la popularización de los modelos SEM, también denominados análisis de estructura de covarianza, proviene de los trabajos de Jöreskog (1969, 1973), Keesling (1972) y Wiley (1973) cuyos resultados se conocieron en principio como el modelo JKW y se popularizó luego como el modelo de relaciones estructurales lineales (LISREL: LInear Structural RELation). Este mismo nombre fue dado al primer programa computacional para analizar modelos estructurales desarrollado en la década de los 70 por Jöreskog & Van Thillo (1973) y Jöreskog (1977). En sus inicios, el uso de LISREL requirió la programación en un lenguaje de comandos matricial. Posteriormente, en la década de los 90, LISREL se hizo más amigable como resultado de los trabajos de Jöreskog & Sörborn (1993a, 1993b) y la introducción del lenguaje SIMPLIS (SIMple LISrel) y más tarde, con la primera versión interactiva. Desde entonces, varios programas han sido desarrollados para analizar modelos estructurales y han contribuido al uso extendido de SEM en una gran cantidad de áreas de investigación. Entre ellos, EQS (Bentler, 1995), Amos (Arbuckle, 1995), MPLUS (Muthén & Muthén, 1998), SAS-PROC CALIS (SAS Institute, 1989) y RAMONA (Browne & Mels, 1994). Referencias históricas más extensas acerca del desarrollo de los modelos de ecuaciones estructurales pueden ser revisadas, por ejemplo, en Bentler, P.M. (1980, 1986) y Shumacker & Lomax (2010).

Este capítulo se presenta como una muy breve introducción a los modelos con ecuaciones estructurales. Solo expone algunas ideas subyacentes a la metodología y por razones de espacio y propósitos del trabajo no trata varias temas importantes. No obstante, a través del texto se ofrecen amplias referencias cuya lectura subsana exhaustivamente las carencias de esta exposición.

3.1.1. Características de los modelos SEM

Las siguientes son algunas de las características de los modelos SEM.

1. Permiten especificar relaciones entre variables observadas, entre variables latentes, y entre variables latentes y observadas en un mismo modelo, lo que le otorga una gran flexibilidad a la metodología.

2. Permiten especificar errores de medición para todas las variables observadas, tanto predictores como predichas, y modelar correlaciones entre errores a través de las ecuaciones. Esta característica es una ampliación a los modelos multiecuacionales clásicos los cuales solo modelan errores para las variables predichas y los errores no deben estar correlacionados.
3. El ajuste del modelo deriva de minimizar una función de la diferencia entre las covarianzas observadas y las covarianzas predichas, a diferencia de los modelos clásicos basados en regresión los cuales minimizan una función de la diferencia entre los valores predicho y observado para cada caso. El empleo de matrices de covarianza en lugar de observaciones individuales, hace que modelos como regresión, ANOVA y los modelos econométricos entre otros, constituyan casos especiales de los modelos SEM (Bollen, 1989).
4. La metodología es útil para evaluar la validez de teorías sustantivas con datos empíricos. En este sentido, SEM es una metodología eminentemente confirmatoria, aun cuando en ocasiones ha sido usada con fines exploratorios.
5. Estadísticamente, los modelos SEM representan una generalización del modelo lineal generalizado (GLM). Así que los modelos SEM son modelos lineales, aunque han sido extendidos para analizar términos de interacción y relaciones no lineales.
6. El desarrollo y la expansión de SEM es rápida y continua (ver por ejemplo Carpita, Brentari & Qannari, 2015; Montfort, Oud & Satorra 2010, Hancock & Samuelsen, 2008; Skrondal & Rabe-Hesketh, 2004).

Una revisión más extensa acerca de las características de los modelos de ecuaciones estructurales, puede encontrarse en Marcoulides & Moustaki (2002) y Marcoulides & Schumacker (2001).

3.1.2. Tipos de modelos SEM

De manera similar a MDS, el término SEM se refiere a un conjunto de modelos diseñados para el análisis estadístico de procesos complejos. A continuación se describen brevemente los más comúnmente usados.

- *Modelos de variables observadas.* Este modelo es también conocido como análisis de trayectorias (Path analysis model). No incluye variables latentes y asume que todas las variables son medidas sin error. Los errores considerados en el modelo, son errores en las ecuaciones o perturbaciones, no errores de medición, que afectan solo a las variables criterio. Por esta razón, algunos investigadores no consideran a los modelos de variables observadas como modelos SEM típicos (Raycov & Marcoulides, 2010). Sin embargo, el análisis de trayectorias es una generalización del modelo de regresión múltiple. El modelo involucra múltiples ecuaciones de regresión que son estimadas simultáneamente, con la particularidad de que una variable puede ser criterio y predictor simultáneamente¹. También es posible analizar relaciones bidireccionales entre variables. Modelos de esta clase son comunes en las ciencias sociales y son, en realidad, un caso particular del modelo de ecuaciones estructurales general.

- *Análisis factorial confirmatorio.* Estos modelos son usados para analizar patrones de interrelaciones entre variables latentes o constructos. Los constructos son, por lo general, medidos con indicadores múltiples. En un modelo de análisis factorial confirmatorio se supone que no hay relaciones direccionales entre los constructos, solo se asume que correlacionan entre si.

- *Modelo de ecuaciones estructurales general.* Este modelo, también denominado modelo de regresión estructural, es una síntesis del análisis de trayectorias y el análisis factorial confirmatorio. Permite evaluar la relación entre los constructos y sus indicadores y al propio tiempo, contrastar hipótesis acerca del efecto direccional o causal de unas variables latentes sobre otras. También es posible modelar la influencia de una variable observada sobre una variable latente, así como la influencia entre indicadores de distintas variables latentes.

¹En análisis de regresión clásico, los predictores son variables independientes en el sentido que ellas no son predichas o explicadas por otras variables en el modelo. En análisis de trayectorias y en los modelos SEM en general, una variable puede ser predictor para otra variable y al propio tiempo ser explicada por otras variables del modelo. Por esta razón, en el contexto SEM es común emplear los términos *variable endógena* para referir a una variable que es explicada en el modelo aun cuando ella pueda ayudar a explicar a otras variables, y *variable exógena* para las variables que no son explicadas en el modelo

En la primera década de este siglo, la investigación en el contexto SEM ha estado dirigida al desarrollo de modelos para datos longitudinales con variables latentes. Estos desarrollos incluyen modelos de tiempo continuo para datos tipo panel y modelos de crecimiento latente. Para una amplia revisión de estos avances, ver por ejemplo Montfort, et. al. (2010).

3.1.3. Lógica de los modelos de ecuaciones estructurales

Los modelos descritos arriba, tienen en común que la evaluación de la bondad de ajuste parte de contrastar la hipótesis

$$\Sigma = \Sigma(\theta), \quad (3.1)$$

donde Σ es la matriz de covarianzas poblacional de las variables observadas, θ es un vector que contiene los parámetros del modelo y $\Sigma(\theta)$ es la matriz de covarianzas escrita como una función de θ .

Una forma de entender la ecuación anterior es pensar que Σ es una síntesis de las relaciones entre las variables observadas tal como existen en una realidad de interés para el investigador. Para intentar explicar tal realidad, el investigador asume una perspectiva teórica que conduce a un modelo teórico específico para el cual las relaciones propuestas se resumen en $\Sigma(\theta)$. Así, la ecuación (3.1) postula que la perspectiva teórica resumida en $\Sigma(\theta)$ explica las relaciones entre las variables observadas resumidas en Σ . En este sentido, $\Sigma(\theta)$ es la matriz de covarianzas implicada por el modelo. La ecuación (3.1) postula que la matriz de covarianzas poblacional de las variables observadas es función de los parámetros del modelo. Por lo tanto, si el modelo es correcto y los parámetros en θ son conocidos, entonces la matriz Σ puede ser exactamente reproducida por la matriz implicada por el modelo.

No es común que $\Sigma(\theta)$ sea exactamente igual a Σ y como veremos más adelante (sección 3.1.6), tampoco es de interés estadístico. Lo que realmente interesa, es que la diferencia entre ellas no sea significativamente distinta de cero. Para chequearlo, hay varias funciones de ajuste disponibles que producen un estadístico para contrastar la hipótesis $H_0 : \Sigma = \Sigma(\theta)$ (ver sección 3.1.6).

A modo de ilustración, considérese la ecuación de regresión simple

$$y = \gamma x + \zeta, \quad (3.2)$$

donde y , x y γ son variables aleatorias, γ es el coeficiente de regresión y ζ es el término de error en la ecuación o término de perturbación que no correlaciona con x y tal que su valor esperado, $E(\zeta)$, es cero. La matriz de covarianzas poblacional de las variables observadas x e y , incluyendo solamente los términos no redundantes, es

$$\Sigma = \begin{pmatrix} \sigma_y^2 & \\ \sigma_{xy} & \sigma_x^2 \end{pmatrix}, \quad (3.3)$$

donde σ_y^2 y σ_x^2 denotan la varianza de y y x , respectivamente y σ_{xy} es la covarianza de x con y . El vector de parámetros para este modelo es $\theta = [\sigma_x^2 \ \sigma_\zeta^2 \ \gamma]^t$.

Para escribir Σ en función de θ , nótese que a partir de (3.2) se obtiene

$$\sigma_y^2 = \gamma^2 \sigma_x^2 + \sigma_\zeta^2 \quad (3.4)$$

$$y \quad \sigma_{xy} = \gamma \sigma_x^2, \quad (3.5)$$

así que

$$\Sigma(\theta) = \begin{pmatrix} \gamma^2 \sigma_x^2 + \sigma_\zeta^2 & \\ \gamma \sigma_x^2 & \sigma_x^2 \end{pmatrix}. \quad (3.6)$$

La tarea sería entonces, encontrar estimaciones para los elementos de θ que minimicen una función apropiada de la diferencia entre Σ y $\Sigma(\theta)$. El modelo en (3.2) podría modificarse para incluir más variables explicativas, algunas de las cuales podrían ser variables latentes o factores, medidos con indicadores múltiples y sus respectivas cargas factoriales. Podría permitirse además, correlaciones entre términos de perturbación y entre términos de varianza única creando así un modelo bastante general. Aun así, podría probarse que el modelo resultante es un caso especial de la ecuación de estructura de covarianza dada en (3.1).

En la sección (3.1.4), se examina el modelo de ecuaciones estructurales general considerando el proceso de especificación y la construcción de la matriz de covarianza implicada. En esa sección también se muestra que el análisis de trayectorias y el análisis factorial confirmatorio son especializaciones del modelo general. El tema de la identificación del modelo y los parámetros, se trata en la sección (3.1.5) y la estimación y el ajuste en la sección (3.1.6). El capítulo finaliza con la sección (3.1.7) dedicada a los procedimientos para estimar las puntuaciones de los factores.

3.1.4. El modelo de ecuaciones estructurales general

El modelo de ecuaciones estructurales general o modelo de regresión estructural, es una síntesis del modelo de variables observadas y el modelo de análisis factorial confirmatorio. Consiste de un componente de medida que especifica la relación entre variables observadas y variables latentes en el sentido del análisis factorial confirmatorio y un componente latente, que analiza las relaciones direccionales entre variables latentes².

Una variable latente es una variable cuya realización no es observable directamente, en contraste con las variables manifiestas cuyas realizaciones son observadas. Por lo general, representan conceptos que varían en su grado de abstracción. Por ejemplo, inteligencia, clase social y expectativas son conceptos altamente abstractos mientras que el ingreso, es un concepto menos abstracto. Los conceptos, también llamados constructos hipotéticos (Cronbach, 1971), se miden indirectamente a través de un conjunto de ítems o indicadores, los cuales son variables observadas o manifiestas. Los indicadores contienen errores de medida aleatorios o sistemáticos, pero la variable latente está libre de ellos (Bollen, 1989). Una variable latente también puede representar una variable continua medida con error, de tal forma que el verdadero valor, que no es observado, se define como el valor esperado de la variable manifiesta (Skrondal & Rabe-Hesketh, 2004). En cualquier caso, las variables latentes requieren, por lo general, indicadores múltiples para ser medidas y la relación entre los indicadores y las variables latentes constituyen el componente de medida del modelo estructural general.

Las variables, latentes o manifiestas, se denominan endógenas si son explicadas en el modelo, aun cuando al propio tiempo pueden contribuir a la explicación de otras variables, o exógenas si solo se usan para explicar a otras variables. El apartado siguiente revisa cómo las variables entran en un modelo de regresión estructural.

Especificación del modelo

La especificación de un modelo estructural general requiere, en primer lugar, una estructura específica entre las variables latentes exógenas y endógenas. Tal estructura constituye un conjunto de hipótesis acerca de relaciones direccionales entre los constructos y representa el componente latente

²Bollen (1989) llama modelos a estos componentes y se refiere a ellos como el modelo de medición y el modelo de variables latentes, respectivamente.

del modelo. En segundo lugar, requiere especificar el componente de medida que indica cómo se han de medir las variables latentes. Ambos componentes constituyen el modelo estructural general.

La formulación general del componente latente está dada por

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \quad (3.7)$$

donde $\boldsymbol{\eta}$ es el vector $m \times 1$ de variables latentes endógenas, $\boldsymbol{\xi}$ es el vector $n \times 1$ de variables latentes exógenas, \mathbf{B} es una matriz $m \times m$ cuyos elementos son los coeficientes que miden el efecto de variables latentes endógenas sobre otras variables latentes endógenas, $\boldsymbol{\Gamma}$ es una matriz $m \times n$ de coeficientes para los efectos de $\boldsymbol{\xi}$ sobre $\boldsymbol{\eta}$ y $\boldsymbol{\zeta}$ es el término de perturbación estocástica para las relaciones planteadas por el modelo, también llamado término de error en las ecuaciones. Para una observación i , ζ_i incluye a las variables que afectan a η_i y no están incluidas de manera explícita en la ecuación correspondiente.

El modelo en (3.7) asume que

$$E(\boldsymbol{\eta}) = 0 \quad E(\boldsymbol{\xi}) = 0 \quad E(\boldsymbol{\zeta}) = 0 \quad (3.8)$$

$$\sigma_{\zeta\xi} = 0 \quad |\mathbf{I} - \mathbf{B}| = 0. \quad (3.9)$$

Las dos primeras expresiones en (3.8), indican que η_i en $\boldsymbol{\eta}$ y ξ_i en $\boldsymbol{\xi}$ entran al modelo desviadas de su media. Esto es, si η_i^* y ξ_i^* son las variables originales, entonces $\eta_i = \eta_i^* - E(\eta_i^*) \quad \forall i = 1, \dots, m$, y $\xi_i = \xi_i^* - E(\xi_i^*) \quad \forall i = 1, \dots, n$. Esta transformación simplifica los cálculos sin afectar la covarianza entre las variables. Si la matriz $\mathbf{I} - \mathbf{B}$ es no singular, como se asume en (3.9), entonces la ecuación en (3.7) puede escribirse en su forma reducida. En efecto, de (3.7) se sigue que

$$\boldsymbol{\eta} - \mathbf{B}\boldsymbol{\eta} = \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta},$$

con lo cual

$$\boldsymbol{\eta} = (\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}). \quad (3.10)$$

Las matrices \mathbf{B} y $\boldsymbol{\Gamma}$ son matrices paramétricas. Sus elementos deben ser estimados en orden a determinar las propiedades del modelo. No obstante, algunos de sus elementos pueden fijarse a un valor constante para efectos de identificación o por razones teóricas, o restringirse a ser una función de otros

parámetros. Se asume además que los ζ_i con $i = 1 \dots m$, son homoscedásticos y no autocorrelacionados. Es decir, para una misma ecuación las perturbaciones asociadas a dos observaciones cualesquiera no están correlacionadas y tienen la misma varianza. Sin embargo, el modelo estructura general es de ecuaciones múltiples y las perturbaciones para dos ecuaciones distintas pueden tener correlación no nula y varianzas distintas.

Para el componente latente se definen también las matrices

$$\Phi = (\phi_{ij})_{n \times n} = E(\xi \xi^t) \quad (3.11)$$

y

$$\Psi = (\psi_{ij})_{m \times m} = E(\zeta \zeta^t). \quad (3.12)$$

Así, Φ es la matriz de covarianzas de las variables latentes exógenas y Ψ es la matriz de covarianzas de las perturbaciones. Si los errores en las ecuaciones no están correlacionados, entonces Ψ es diagonal. Estas matrices, al igual que \mathbf{B} y $\mathbf{\Gamma}$, son matrices paramétricas.

El componente de medición del sistema general, está dado por

$$\mathbf{y} = \Lambda_y \boldsymbol{\eta} + \boldsymbol{\epsilon} \quad (3.13)$$

$$\mathbf{x} = \Lambda_x \boldsymbol{\xi} + \boldsymbol{\delta}, \quad (3.14)$$

donde \mathbf{y} es un vector $p \times 1$ de indicadores de $\boldsymbol{\eta}$ y \mathbf{x} es un vector $q \times 1$ de indicadores de $\boldsymbol{\xi}$. La matriz Λ_y de orden $p \times m$ y la matriz Λ_x de orden $q \times n$, contienen ambas los parámetros λ_i que vinculan a las variables latentes con las variables manifiestas. En términos del análisis factorial, los λ_i representan las cargas factoriales. Por otro lado, $\boldsymbol{\epsilon}$, vector $p \times 1$ y $\boldsymbol{\delta}$, vector $q \times 1$, son los errores de medición o errores de medida de \mathbf{y} y \mathbf{x} , respectivamente. Las matrices Λ_y y Λ_x son matrices paramétricas, lo que significa que poseen elementos libres que deben ser estimados. Además, se definen las matrices

$$\Theta_\epsilon = (\theta_{ij})_{p \times p} = E(\boldsymbol{\epsilon} \boldsymbol{\epsilon}^t) \quad (3.15)$$

y

$$\Theta_\delta = (\theta_{ij})_{q \times q} = E(\boldsymbol{\delta} \boldsymbol{\delta}^t). \quad (3.16)$$

que son las matrices de covarianzas de los errores de medida. Θ_ϵ contiene las varianzas del error y sus covarianzas para el vector \mathbf{y} , mientras que Θ_δ contiene las varianzas del error y sus covarianzas para el vector \mathbf{x} . Si los errores no están correlacionados, entonces Θ_ϵ y Θ_δ son, cada una de ellas, una matriz diagonal. En (3.13) y (3.14) se asume que

$$E(\boldsymbol{\eta}) = E(\boldsymbol{\xi}) = E(\boldsymbol{\epsilon}) = E(\boldsymbol{\delta}) = 0.$$

Además, $\boldsymbol{\epsilon}$ y $\boldsymbol{\delta}$ no correlaciona entre ellos ni con $\boldsymbol{\xi}$ y $\boldsymbol{\eta}$.

Matriz de covarianza implicada por el modelo

La hipótesis fundamental del modelo SEM general es que la matriz de covarianzas poblacional de las variables observadas, $\boldsymbol{\Sigma}$ puede escribirse en función de los parámetros del modelo (ecuación 3.1). La relación de $\boldsymbol{\Sigma}$ a $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ es básica para entender los procesos de identificación, estimación y ajuste del modelo (Bollen, 1989). $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ puede escribirse como

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\Sigma}_{yy}(\boldsymbol{\theta}) & \\ \boldsymbol{\Sigma}_{xy}(\boldsymbol{\theta}) & \boldsymbol{\Sigma}_{xx}(\boldsymbol{\theta}) \end{pmatrix}, \quad (3.17)$$

donde $\boldsymbol{\Sigma}_{yy}(\boldsymbol{\theta})$, $\boldsymbol{\Sigma}_{xx}(\boldsymbol{\theta})$ y $\boldsymbol{\Sigma}_{xy}(\boldsymbol{\theta})$ corresponden a la matriz de covarianza de \mathbf{y} , la matriz de covarianza de \mathbf{x} y la matriz de covarianza de \mathbf{y} con \mathbf{x} , respectivamente, escritas en función de los parámetros del modelo contenidos en el vector $\boldsymbol{\theta}$. A partir de (4.9), se tiene que

$$\begin{aligned} \boldsymbol{\Sigma}_{yy}(\boldsymbol{\theta}) &= E(\mathbf{y}\mathbf{y}^t) \\ &= E[(\boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\epsilon})(\boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\epsilon})^t] \\ &= E[(\boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\epsilon})(\boldsymbol{\eta}^t\boldsymbol{\Lambda}_y^t + \boldsymbol{\epsilon}^t)] \\ &= \boldsymbol{\Lambda}_y E(\boldsymbol{\eta}\boldsymbol{\eta}^t)\boldsymbol{\Lambda}_y^t + \boldsymbol{\theta}_\epsilon. \end{aligned} \quad (3.18)$$

Usando (3.10), puede escribirse

$$E(\boldsymbol{\eta}\boldsymbol{\eta}^t) = (\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}^t + \boldsymbol{\Psi})[(\mathbf{I} - \mathbf{B})^{-1}]^t, \quad (3.19)$$

y sustituyendo (3.19) en (3.18), se obtiene

$$\boldsymbol{\Sigma}_{yy}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}_y(\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}^t + \boldsymbol{\Psi})[(\mathbf{I} - \mathbf{B})^{-1}]^t\boldsymbol{\Lambda}_y^t + \boldsymbol{\theta}_\epsilon. \quad (3.20)$$

Analogamente, se prueba que

$$\Sigma_{yx}(\boldsymbol{\theta}) = \Lambda_y(\mathbf{I} - \mathbf{B})^{-1}\Gamma\Phi\Lambda_x^t, \quad (3.21)$$

y que

$$\Sigma_{xx}(\boldsymbol{\theta}) = \Lambda_x\Gamma\Lambda_x^t + \theta_\delta, \quad (3.22)$$

de tal forma que

$$\Sigma(\boldsymbol{\theta}) = \begin{pmatrix} \Sigma_{yy}(\boldsymbol{\theta}) & \Sigma_{yx}(\boldsymbol{\theta}) \\ \Sigma_{xy}(\boldsymbol{\theta}) & \Sigma_{xx}(\boldsymbol{\theta}) \end{pmatrix} = \begin{pmatrix} \Lambda_y(\mathbf{I} - \mathbf{B})^{-1}(\Gamma\Phi\Gamma^t + \Psi)[(\mathbf{I} - \mathbf{B})^{-1}]^t\Lambda_y^t + \theta_\epsilon & \Lambda_y(\mathbf{I} - \mathbf{B})^{-1}\Gamma\Phi\Lambda_x^t \\ \Lambda_x\Phi\Gamma^t[\mathbf{I} - \mathbf{B}]^{-1}]^t\Lambda_y^t & \Lambda_x\Phi\Lambda_x^t + \Theta_\delta \end{pmatrix}, \quad (3.23)$$

es la matriz implicada por el modelo estructural general, donde cada elemento de Σ está escrito en función de parámetros del modelo. Nótese que $\Sigma_{xy}(\boldsymbol{\theta}) = (\Sigma_{yx}(\boldsymbol{\theta}))^t$.

El modelo de ecuaciones estructurales con variables observadas, asume que todas las variables son medidas sin error. Con lo cual (3.13) y (3.14) se transforman en $\mathbf{y} = \boldsymbol{\eta}$ y $\mathbf{x} = \boldsymbol{\xi}$, respectivamente. Esto conduce a que $\Lambda_y = \mathbf{I}_p$, $\Lambda_x = \mathbf{I}_q$ y Θ_ϵ y Θ_δ son matrices nulas. Sustituyendo estos resultados en (4.12) se obtiene para el modelo de variables observadas dado por

$$\mathbf{y} = \mathbf{B}\mathbf{y} + \Gamma\mathbf{x} + \boldsymbol{\zeta}, \quad (3.24)$$

que la matriz de covarianza implicada es

$$\Sigma(\boldsymbol{\theta}) = \begin{pmatrix} (\mathbf{I} - \mathbf{B})^{-1}(\Gamma\Phi\Gamma^t + \Psi)[(\mathbf{I} - \mathbf{B})^{-1}]^t & (\mathbf{I} - \mathbf{B})^{-1}\Gamma\Phi \\ \Phi\Gamma^t[\mathbf{I} - \mathbf{B}]^{-1}]^t & \Phi \end{pmatrix}. \quad (3.25)$$

Por otra parte, el modelo factorial confirmatorio se formula como en las ecuaciones (3.13) y (3.14). Dado que no plantea hipótesis acerca de relaciones direccionales entre las variables $\boldsymbol{\eta}$ y $\boldsymbol{\xi}$, entonces \mathbf{B} y Γ son nulas. Por la misma razón, no hay variables latentes endógenas así que $\Lambda_y = 0$, $\Theta_\epsilon = 0$ y $\Psi = 0$. A partir de estos resultados, los términos $\Sigma_{yy}(\boldsymbol{\theta})$, $\Sigma_{yx}(\boldsymbol{\theta})$ y $\Sigma_{xy}(\boldsymbol{\theta})$ de (4.12) no son necesarios, así que la matriz de covarianza implicada por este modelo viene dada por

$$\Sigma(\boldsymbol{\theta}) = \Lambda_x \Phi \Lambda_x^t + \Theta_\delta. \quad (3.26)$$

Los procesos que conducen a (3.25) y a (3.26), muestran que el modelo de variables observadas y el análisis factorial confirmatorio son casos especiales del modelo de ecuaciones estructurales general.

3.1.5. Identificación

Una vez que el modelo ha sido especificado, el paso siguiente es establecer la identificación antes de la estimación de los parámetros. La identificación es crucial debido a que las estimaciones de parámetros no identificados no son consistentes y los índices de ajuste global de un modelo no identificado no pueden ser interpretadas (Rigdon, 1995). La identificación se refiere a si existe suficiente información en la matriz de covarianzas de las variables observadas, para estimar de manera única los elementos libres (no fijos y no restringidos) de las matrices paramétricas contenidos en el vector $\boldsymbol{\theta}$.

Un parámetro $\theta \in \boldsymbol{\theta}$ está identificado si puede escribirse como una función de uno o varios de los elementos no redundantes de Σ y esta función conduce a una solución única. Si todos los elementos de $\boldsymbol{\theta}$ están identificados, el modelo está identificado. Para determinar la identificación del modelo a partir de esta definición, cada elemento de $\boldsymbol{\theta}$ debe ser resuelto en términos de los elementos de Σ a partir de la ecuación $\Sigma(\boldsymbol{\theta}) = \Sigma$. Si se tienen k variables observadas, el número de ecuaciones que se derivan de la igualdad anterior es $\frac{1}{2}k(k+1)$ con lo cual, si t es el número de parámetros tres situaciones son posibles:

1. $t = \frac{1}{2}k(k+1)$. En este caso, la información contenida en Σ es suficiente para estimar los parámetros y existe una solución única para los elementos en $\boldsymbol{\theta}$. Un modelos que satisfaga esta condición se denomina **saturado**.
2. $t < \frac{1}{2}k(k+1)$. Aquí, hay más información en Σ de la requerida para estimar los parámetros y el modelo se denomina **sobre identificado**. En este caso, el sistema determinado por $\Sigma(\boldsymbol{\theta}) = \Sigma$ posee infinitas soluciones. En el proceso de estimación, se escoge la solución que minimice una función de pérdida (ver sección 3.1.6). En este sentido, la escogencia es óptima.

3. $t > \frac{1}{2}k(k+1)$, caso en el cual la información contenida en Σ es insuficiente para estimar los parámetros y el estatus del modelo es **no identificado** o **sub identificado**.

En ocasiones el álgebra matricial permite establecer la identificación de parámetros en modelos sencillos (ver Vera & Rivera, 2014 para un ejemplo). Para modelos con muchos parámetros este enfoque resulta tedioso en extremo y propenso a errores, por lo que ha sido definido un conjunto de reglas para ayudar a establecer la identificación. Antes de pasar a examinar algunas de estas reglas, es conveniente determinar correctamente los parámetros del modelo que deben ser estimados. En párrafos anteriores se ha sugerido que los parámetros a estimar son los elementos libres de las 8 matrices paramétricas del modelo general. Un *parámetro libre* es aquel que no es fijo ni restringido. Un *parámetro fijo* es un parámetro al que se le asigna el valor de una constante conocida y su valor no cambia como resultado del proceso de estimación. Un *parámetro restringido* es aquel que se postula igual a otro parámetro, no fijo, o que se escribe como función de otros parámetros libres. La siguiente lista es un resumen de la propuesta por Bentler (1995) y es útil para identificar los parámetros del modelo.

1. Todas las varianzas y covarianzas de las variables exógenas son parámetros del modelo.
2. Todas las cargas factoriales que conectan a las variables latentes con sus indicadores son parámetros del modelo.
3. Todos los coeficientes de regresión entre variables observadas o entre variables latentes son parámetros del modelo.
4. Las varianzas y covarianzas entre variables endógenas y las covarianzas entre endógenas y exógenas, nunca son parámetros del modelo.

las siguientes reglas son de uso común para ayudar a establecer el estatus de identificación de modelos SEM.

Regla t

La información disponible para la identificación corresponde a los elementos no redundantes de Σ . Debido a que hay p variables \mathbf{y} y q variables \mathbf{x} , la matriz Σ es de orden $(p+q) \times (p+q)$ con lo cual, por la simetría de la matriz

de covarianza, hay $\frac{1}{2}(p+q)(p+q+1)$ elementos no redundantes. En consecuencia, para escribir los parámetros en Θ en función de los elementos de Σ se dispone de $\frac{1}{2}(p+q)(p+q+1)$ ecuaciones. Así, una condición necesaria, aunque no suficiente, para la identificación de los parámetros es que

$$t \leq \frac{1}{2}(p+q)(p+q+1), \quad (3.27)$$

donde t es el número de elementos de Θ . Esta regla es bastante general en el sentido de que puede aplicarse a cualquier modelo SEM. Su mayor utilidad es que permite identificar rápidamente un modelo no identificado. Aun cuando su cumplimiento no garantiza que el modelo pueda identificarse por cuanto la condición no es suficiente, su incumplimiento indicaría que el modelo no es identificable por cuanto la condición es necesaria.

Regla de los dos pasos

La aplicación de esta regla se hace en dos etapas. En la primera se ignoran las matrices \mathbf{B} , $\mathbf{\Gamma}$ y $\mathbf{\Psi}$, con lo cual no se consideran las relaciones direccionales entre las variables latentes y solo es de interés las varianzas y covarianzas contenidas en Φ . Así, el modelo queda reformulado como un análisis factorial confirmatorio y aplican las siguientes reglas.

- *Regla de los tres indicadores* Un modelo de análisis factorial confirmatorio está identificado cuando tiene (1) al menos tres indicadores por variable latente, (2) cada fila de $\Lambda_{\mathbf{x}}$ con uno y solo un elemento distinto de cero y (3) Θ_{δ} diagonal.
- *Regla de los dos indicadores* Dos indicadores por variable latente es suficiente para la identificación, siempre que Θ_{δ} es diagonal, cada fila de $\Lambda_{\mathbf{x}}$ posee uno y solo un elemento distinto de cero y ningún elemento en Φ es nulo.

Si estas reglas establecen la identificación del modelo de análisis confirmatorio, se pasa a la segunda etapa. En caso contrario, la regla de los dos pasos no es aplicable. La segunda etapa de la regla considera el modelo como un modelo estructural de variables observadas, ignorando el componente de medida, y evalúa la identificación según las siguientes reglas.

- *Regla \mathbf{B} nula* Si el modelo de variables observadas no formula relaciones direccionales entre variables endógenas, entonces la matriz \mathbf{B} es nula y constituye una condición suficiente para la identificación.
- *Regla recursiva* Una condición suficiente para que un modelo de variables observadas se pueda identificar es que \mathbf{B} sea una matriz triangular inferior y Ψ una matriz diagonal.

En resumen, si la primera etapa muestra que los parámetros del componente de medida están identificados y la segunda etapa muestra que los parámetros del componente latente están identificados, entonces esto es suficiente para identificar el modelo completo.

Por otro lado, las variables latentes en un modelo deben poseer una escala de medida o el modelo no podrá ser identificado. Tales variables no poseen una escala universal por lo que deben establecerse algunas restricciones en orden a asignarle una escala en el contexto del modelo propuesto. Lo típico en SEM, es fijar las varianzas de esas variables a 1 o asignarle a la variable latente la misma escala de alguno de sus indicadores. Esto último se logra fijando el parámetro correspondiente λ_{ij} a 1.

Otras reglas de identificación han sido propuestas distintas a las ya mencionadas. Entre ellas, la regla para modelos MIMIC (ver por ejemplo Bollen, 1989) y las propuestas por Bollen & Davis (2009) denominadas regla de 2 o más trayectorias emitidas y la regla de X exógena. La primera de ellas es una condición necesaria y la segunda una condición suficiente. Rigdon (1995), propuso una técnica gráfica denominada sistema de clasificación directa como una condición necesaria y suficiente para la identificación de cierto tipo de modelos recursivos. Bekker et al. (1994), introdujo un conjunto de técnicas heurísticas de identificación basadas en la evaluación del rango de una matriz Jacobiana aumentada implicada por la estructura del modelo. Otra vía importante para determinar el estatus de identificación, son las pruebas de identificación empírica que involucran la evaluación de algunos resultados del análisis del modelo tal como la matriz de información o la correlación entre parámetros estimados (ver por ejemplo, McDonald, 1982).

Estas reglas utilizan el concepto de identificación global y local. Un vector de parámetros θ se dice que está globalmente identificado si no existe θ^* en el espacio paramétrico, con $\theta \neq \theta^*$, tal que $\Sigma(\theta) = \Sigma(\theta^*)$. Una definición de identificación local puede establecerse restringiendo la definición global a un entorno del parámetro cuya identificación se analiza (ver por ejemplo, Skrondal & Rabe-Hesketh, 2004).

Las pruebas empíricas permiten abordar el problema de la sub identificación empírica (Kenny, 1979), el cual ocurre cuando se estiman parámetros no identificados aun cuando se tienen pruebas de identificación previas al análisis. Para una revisión más extensa del tema, puede consultarse por ejemplo Brito & Pearl (2002); Davis (1993); O'Brien (1994), Reilly & O'Brien (1996), Rigdon (1995) y Bollen (1989).

3.1.6. Estimación y ajuste

Los procesos de estimación derivan de la relación entre la matriz de covarianza de las variables observadas y los parámetros del modelo. Como las matrices Σ y $\Sigma(\theta)$ son desconocidas, en la práctica se requiere una muestra apropiada para obtener la matriz de covarianza muestral de las variables observadas, \mathbf{S} , como una estimación de Σ . Así, los parámetros en θ son estimados de manera que la matriz de covarianza implicada, $\Sigma(\hat{\theta})$, esté tan próxima a \mathbf{S} como sea posible. La evaluación de tal proximidad requiere, como es común, minimizar una función de la matriz de residuos $(\mathbf{S} - \Sigma)$. Así, los parámetros en θ se estiman de manera que minimicen una función de ajuste del tipo $F(\mathbf{S}, \Sigma)$ basada en \mathbf{S} y en Σ .

Las propiedades deseables para $F(\mathbf{S}, \Sigma)$ son

1. $F(\mathbf{S}, \Sigma)$ es un escalar.
2. $F(\mathbf{S}, \Sigma) \geq 0$.
3. $F(\mathbf{S}, \Sigma) = 0$ si y solo si $\Sigma = \mathbf{S}$
4. $F(\mathbf{S}, \Sigma)$ es continua en \mathbf{S} y Σ .

Minimizar una función de ajuste que satisfaga tales condiciones, conduce a estimadores consistentes de θ (Browne, 1984, 1966). Entre las funciones más importantes en SEM están: máxima verosimilitud (ML), mínimos cuadrados no ponderados (ULS), mínimos cuadrados generalizados (GLS) y la distribución asintóticamente libre (ADS), también llamada mínimos cuadrados ponderados.

Función de máxima verosimilitud

El método de máxima verosimilitud (ML) es el más popular entre los usuarios de SEM. La función adopta la expresión

$$F_{ML} = \log|\Sigma(\boldsymbol{\theta})| + \text{tr}(\mathbf{S}\Sigma^{-1}(\boldsymbol{\theta})) - \log|\mathbf{S}| - (p + q), \quad (3.28)$$

donde $\Sigma(\boldsymbol{\theta})$ y \mathbf{S} se asumen definidas positivas, con lo cual $|\Sigma(\boldsymbol{\theta})|$ y $|\mathbf{S}|$ son ambos mayores que cero y así, todos los términos del lado derecho de (3.28) están bien definidos. Las estimaciones para los parámetros son los valores donde F_{ML} alcanza el mínimo absoluto.

La maximización de F_{ML} conduce a estimadores consistentes, eficientes e insesgados de $\boldsymbol{\theta}$ siempre que \mathbf{y} y \mathbf{x} sigan una distribución normal multivariante y el modelo esté sobre identificado. La consistencia asegura que al aumentar el tamaño de la muestra, el estimador converge al verdadero valor del parámetro. Por otra parte, la eficiencia indica que la distribución de $\hat{\boldsymbol{\theta}}$ tiene varianza mínima, mientras que al ser insesgado, se tiene que $E(\hat{\boldsymbol{\theta}}) = \boldsymbol{\theta}$, lo que asegura que el estimador se distribuye alrededor del parámetro. El supuesto de normalidad multivariante de las variables observadas, por lo general, no se satisface completamente con datos reales. Para una revisión del proceso para chequear la normalidad multivariante ver, por ejemplo, Tabachnick & Fidell (2001) y Khattree & Naik (1999). Las pruebas estadísticas de adecuación del modelo se ven afectadas por la no normalidad (Kaplan, 2000). Para abordar este problema, Satorra & Bentler (1988, 1984, 2001, 2010) desarrollaron pruebas estadísticas para corregir la prueba de bondad de ajuste por el sesgo producido por las desviaciones de la normal multivariante. No obstante, algunas investigaciones (por ejemplo, Raykov & Widaman, 1995) han mostrado que F_{ML} puede ser empleada aun con leves desviaciones de la normalidad, en particular, si el interés primario es la estimación de parámetros.

La distribución asintótica de $(N - 1)\underset{\boldsymbol{\theta}}{\text{mín}}(F_{ML})$ es χ^2 con $\frac{1}{2}(p + q)(p + q + 1) - t$ grados de libertad. Esto provee una prueba de ajuste global del modelo, conocida como prueba de bondad de ajuste χ^2 , para contrastar la hipótesis $H_0 : \Sigma = \Sigma(\boldsymbol{\theta})$. Con modelos saturados, se tiene $t = \frac{1}{2}(p + q)(p + q + 1)$ por lo que no habrían grados de libertad para la prueba. Por esta razón, el modelo saturado, aunque es de interés para asuntos de comparación de modelos y la construcción de algunos índice de bondad de ajuste, no es e interés como modelo teórico.

Debido a su carácter asintótico, $((N - 1)\underset{\boldsymbol{\theta}}{\text{mín}}(F_{ML}))$ sigue una distribución χ^2 para valores grandes de N . Como $((N - 1)\underset{\boldsymbol{\theta}}{\text{mín}}(F_{ML}))$ es proporcional a N , valores grandes de N impulsan al estadístico hacia la zona de rechazo con

lo cual, la prueba puede rechazar modelos que explican adecuadamente las relaciones observadas en los datos. Esto ha conducido al desarrollo de una batería de índices de ajuste que complementan el análisis de adecuación del modelo. Algunos de estos índices se presentan más adelante.

Función de mínimos cuadrados no ponderados

Esta función está dada por

$$F_{ULS} = \frac{1}{2} \text{tr} [(\mathbf{S} - \Sigma(\boldsymbol{\theta}))^2], \quad (3.29)$$

donde mediante F_{ULS} se minimiza la suma de cuadrados de los elementos de la matriz de residuos $(\mathbf{S} - \Sigma(\boldsymbol{\theta}))$ de una manera análoga a la regresión de mínimos cuadrados ordinarios (OLS). Así, F_{ULS} es la versión de la función OLS para el caso en que los residuos consisten en las diferencias entre varianzas y covarianzas muestrales y las predichas por el modelo, en lugar de las diferencias entre los valores muestrales de las variables observadas y los predichos por el modelo.

La minimización de F_{ULS} produce una estimación consistente de $\boldsymbol{\theta}$ sin el supuesto de una distribución particular para las variables observadas, siempre que $\boldsymbol{\theta}$ esté identificado. Sin embargo, los estimadores no son eficientes.

Función de mínimos cuadrados generalizados

La forma general de esta función es

$$F_{GLS} = \frac{1}{2} \text{tr} \left\{ [(\mathbf{S} - \Sigma(\boldsymbol{\theta})) \mathbf{W}^{-1}]^2 \right\}, \quad (3.30)$$

donde la matriz de pesos \mathbf{W}^{-1} es aleatoria y converge en probabilidad a una matriz definida positiva cuando $N \rightarrow \infty$, o es una matriz definida positiva de constantes. La matriz \mathbf{W}^{-1} pondera los elementos de la matriz de residuos de acuerdo a sus varianzas y covarianzas con otros elementos para corregir por heterocedasticidad. Si $\mathbf{W}^{-1} = \mathbf{I}$, F_{GLS} deriva en F_{ULS} . Bajo las condiciones dadas para \mathbf{W}^{-1} , el estimador de $\boldsymbol{\theta}$ obtenido al minimizar (3.30) es consistente, bajo el supuesto de distribución multinormal de las variables observadas. De hecho, los estimadores producidos al minimizar F_{GLS} , poseen las mismas propiedades que los obtenidos por F_{ML} , bajo supuestos de multinormalidad menos rigurosos. La distribución asintótica de $\hat{\boldsymbol{\theta}}$ es multinormal con matriz

de covarianza asintótica conocida, así que se pueden construir pruebas estadísticas de significación (ver Browne 1982, 1984). No todas las selecciones de \mathbf{W}^{-1} son apropiadas. Bajo algunos supuestos sobre los elementos de \mathbf{s} , \mathbf{W}^{-1} debería seleccionarse de manera que $p \lim \mathbf{W}^{-1} = c\mathbf{\Sigma}^{-1}$ (ver Bollen, 1989).

La selección más común es $\mathbf{W}^{-1} = \mathbf{S}^{-1}$ y $c = 1$, con lo cual

$$\begin{aligned} F_{GLS} &= \frac{1}{2} \text{tr} \left\{ [(\mathbf{S} - \mathbf{\Sigma}(\boldsymbol{\theta}))\mathbf{S}^{-1}]^2 \right\} \\ &= \frac{1}{2} \text{tr} \left\{ (\mathbf{I} - \mathbf{\Sigma}(\boldsymbol{\theta})\mathbf{S}^{-1})^2 \right\}. \end{aligned} \quad (3.31)$$

Distribución asintóticamente libre

La función de ajuste para la distribución asintóticamente libre es

$$F_{ADF} = [\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})]^t \mathbf{W}^{-1} [\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})], \quad (3.32)$$

donde \mathbf{S} es el vector conformado con los $\frac{1}{2}(p+q)(p+q-1)$ elementos no redundantes de \mathbf{S} , $\boldsymbol{\sigma}(\boldsymbol{\theta})$ es el vector correspondiente con los elementos de $\mathbf{\Sigma}(\boldsymbol{\theta})$ y $\boldsymbol{\theta}$ es el vector de parámetros libres del modelo. \mathbf{W}^{-1} es una matriz $\frac{1}{2}(p+q)(p+q-1) \times \frac{1}{2}(p+q)(p+q-1)$ de ponderaciones definida positiva. Los valores de $\boldsymbol{\theta}$ se seleccionan de manera que minimicen la suma ponderada del cuadrado de los residuos, de manera análoga a la regresión de mínimos cuadrados ponderados, solo que aquí los residuos vienen de las diferencias entre varianzas y covarianzas predichas y observadas y no entre valores individuales. El estimador obtenido al minimizar (3.32) es consistente y, si \mathbf{W} es la matriz de covarianza asintótica de \mathbf{S} o un estimador insesgado de ella, también es asintóticamente eficiente como ha sido establecido por Browne (1982, 1984). Por lo general, \mathbf{W} es la matriz de covarianzas de las covarianzas muestrales (Bollen, 1989, pág. 426).

Índices de bondad de ajuste

Para medir la discrepancia entre la matriz de covarianza muestral \mathbf{S} y la matriz de covarianzas predicha por el modelo $\mathbf{\Sigma}(\hat{\boldsymbol{\theta}})$, están disponibles en SEM varios índices de naturaleza inferencial y descriptiva. Para que un modelo sea considerado plausible para describir y explicar los datos, debería tener valores aceptables en la mayoría de estos índices y simultáneamente, la magnitud y el

signo de los parámetros debe corresponderse con consideraciones sustantivas e investigación previa.

En este apartado se mencionan algunos de estos índices y se remite al lector a Bentler (2004), Bollen (1989), Byrne (1998), Jöreskog & Sörbom (1993a, 1993b), Muthén & Muthén (2004). Schumacker & Lomax (2010), para una revisión exhaustiva.

- **Índice χ^2 .** Este índice es una prueba estadística de bondad de ajuste para contrastar la hipótesis nula $\mathbf{H}_0 : \Sigma = \Sigma(\theta)$. El estadístico de prueba es

$$T = (N - 1)\underset{\theta}{\text{mín}}(F_{(\cdot)}), \quad (3.33)$$

donde N denota el tamaño de la muestra y $F_{(\cdot)}$ la función usada para la estimación de los parámetros (ML, GLS, ULS o ADF). Como ya ha sido dicho, T se distribuye asintóticamente χ^2 con $\frac{1}{2}(p + q)(p + q + 1) - t$ grados de libertad. La importancia del índice de bondad de ajuste χ^2 estriba en ser una prueba inferencial, lo que significa que las conclusiones sobre el ajuste del modelo obtenida sobre los datos muestrales pueden ser generalizados a la población. Además, la mayoría de los otros índices de bondad de ajuste son funciones de χ^2 . Tiene el problema de su dependencia del tamaño de muestra señalado anteriormente.

Un valor significativo del estadístico de prueba, indica que las matrices \mathbf{S} y $\Sigma(\hat{\theta})$ difieren y por lo tanto, el modelo no se ajusta adecuadamente a los datos. Un valor no significativo indica que el modelo es plausible. Este criterio de decisión, se deriva de la hipótesis a contrastar dada por $H_0 : \Sigma = \Sigma(\theta)$ que establece que un modelo es plausible si no hay evidencia en los datos para rechazar la hipótesis nula. El índice χ^2 es también sensible a las desviaciones de la normalidad multivariante de las variables observadas.

- **Índice *GFI*.** El índice de bondad de ajuste GFI, mide la proporción de varianzas y covarianzas en las variables observadas que es predicha por el modelo. En este sentido, es análogo al R^2 del análisis de regresión. La expresión para el índice es

$$\text{GFI} = 1 - [\chi_{\text{modelo}}^2 / \chi_{\text{nulo}}^2], \quad (3.34)$$

donde χ_{modelo}^2 es valor de χ^2 para el modelo teórico y χ_{nulo}^2 es el valor para el modelo independiente o nulo que no formula relación direccional alguna entre las variables. Así, el modelo nulo tendría el peor ajuste a los datos. De acuerdo con esto, $\chi_{modelo}^2 \leq \chi_{nulo}^2$ y así, $GFI \in [0, 1]$. Valores cercanos a 1 indican un buen ajuste. Sin embargo, no hay una norma estricta que señale un valor de GFI por arriba del cual el modelo pueda considerarse una descripción razonable de las relaciones reales entre las variables observadas. Por esta razón, GFI es empleado con frecuencia como un índice para la comparación de modelos. No obstante, como una regla general Hu & Bentler (1999) sugieren considerar valores mayores o iguales a 0.95 como indicadores de un buen ajuste.

- **Indice RMSEA.** RMSEA corresponden a las siglas en inglés de *raíz cuadrática media del error de aproximación*, un índice de ajuste global que ha ganado popularidad debido a que es el índice de ajuste menos dependiente del tamaño de muestra. Se calcula a través de la expresión

$$\text{RMSEA} = \sqrt{\frac{\chi_{modelo}^2 - df_{modelo}}{(N - 1)df_{modelo}}} \quad (3.35)$$

donde df_{modelo} son los grados de libertad del modelo propuesto. Se sugiere que un valor inferior a 0.05 indica un modelo plausible (Browne & Cudeck, 1993). Este índice también considera la complejidad del modelo al incorporar los grados de libertad del modelo teórico. Como es sabido, mientras mayor es el número de grados de libertad mayor es la probabilidad de rechazar el modelo. Así, entre dos modelos plausibles para analizar el mismo conjunto de datos, es más confiable el modelo con más grados de libertad. Esto es debido a que un modelo con más grados de libertad, lo cual significa un menor número de parámetros, debe reproducir con mayor precisión la matriz Σ , a objeto de no conducir al rechazo de H_o . Esta es la base del concepto de parsimonia, ampliamente discutido en SEM.

- **Indice TLI.** TLI se refiere a Tucker-Lewis-Index (Tucker & Lewis, 1973) que es una medida usada para comparar modelos alternativos o un modelo teórico contra un modelo nulo. Su expresión está dada por

$$\mathbf{TLI} = \frac{\chi_{nulo}^2/df_{nulo} - \chi_{modelo}^2/df_{modelo}}{\chi_{nulo}^2/df_{nulo} - 1}, \quad (3.36)$$

donde $\mathbf{TLI} \in [0, 1]$ y valores superiores a 0.9 indican un buen ajuste del modelo a los datos.

El problema de la evaluación de los modelos de ecuaciones estructurales, que como ya se ha dicho se refiere a la medida en la cual un modelo teórico puede ser reconocido como una perspectiva razonable para explicar las varianzas y covarianzas observadas en un conjunto de variables, es un tema complejo y no exento de controversias (Raycov & Marcoulides, 2006). La subjetividad para decidir el valor a partir del cual un índice señala un buen ajuste, siempre está presente.

Además de los descritos, otros índices de ajuste se han formulado para complementar la prueba de bondad de ajuste χ^2 que dan al investigador una batería que usada conjuntamente con los aspectos sustantivos de la teoría y la investigación previa, ayudan a la toma de decisiones acerca de la utilidad de un modelo propuesto.

3.1.7. Estimación de los valores de las variables latentes (factor scores)

Una vez que un modelo ha sido estimado y resulta aceptable para explicar las relaciones observadas, es posible que el investigador requiera conocer los valores de las variables latentes para observaciones individuales con el propósito de usarla para análisis posteriores. Por ejemplo, en algunas ocasiones el investigador necesita conocer la ubicación de los individuos con relación al factor, o usar la información para conocer si y cómo los factores difieren entre grupos, o incorporar las puntuaciones de los factores como parte de un análisis de regresión. Para ello es necesario estimar las puntuaciones que representan la ubicación de cada individuo sobre cada factor. Existen dos métodos generales para estimar las puntuaciones de los factores: los métodos no refinados y los métodos refinados (DiStefano et al, 2009).

Métodos no refinados

Estos métodos son fáciles de usar y sus resultados fáciles de interpretar. Entre los más usados están:

- *Suma de puntuaciones por factor.* EL método consiste en sumar los valores brutos para cada individuo sobre todos los indicadores del factor (Comrey & Lee, 1992). La suma es algebraica con relación al signo de la carga factorial correspondiente al indicador. Así, si un ítem tiene carga factorial con signo positivo su valor se suma, en caso contrario, se resta. Es conveniente promediar tales valores brutos para retener la escala y para facilitar comparaciones entre factores con distinto número de indicadores. Además, las puntuaciones de los factores obtenidas por este método, preservan la variabilidad de los datos originales. Hair et al (2006, p. 240) sugieren emplear este método en los casos en los que las escalas usadas para obtener los datos originales tengan poca validez o poca confiabilidad.

Como aspectos negativos del método está el que todos los indicadores de un factor reciben la misma ponderación, con independencia de las diferencias en las cargas factoriales. Además, si los ítems están medidos en escalas diferentes, con diferentes variabilidad, los factores pueden resultar menos confiables.

- *Suma de puntuaciones por arriba de un valor de específico.* Según este método, solo se suman los indicadores del factor cuya carga factorial supere un valor preestablecido (cut-off value). La selección del valor es siempre subjetiva. Además, este procedimiento no preserva la variabilidad de los datos originales.

Una variante de estos métodos es estandarizar las variables observadas antes de realizar la suma. Este procedimiento adicional es útil cuando existe una marcada diferencia en la variabilidad de los ítems.

- *Suma ponderada.* Según este método, los ítems se ponderan por las respectivas cargas factoriales. De esta forma se logra que los ítem con mayor carga factorial tengan un mayor peso en la determinación de las puntuaciones de los factores. El problema está en que las cargas factoriales pueden no representar las diferencias entre los factores, debido al método seleccionado de extracción y rotación de los factores.

Más acerca de métodos no refinados para la determinación de las puntuaciones de los factores, puede encontrarse, por ejemplo, en Grice & Harris (1998) y en Glass & Maguire (1966).

Métodos refinados

Con los métodos refinados, las puntuaciones de los factores se obtienen como combinación lineal de las variables observadas o indicadores. Tales combinaciones incorporan la varianza compartida entre el ítem y el factor y el error de medida o término de varianza única.

- *Método de regresión.* El primero en usar la regresión de mínimos cuadrados para predecir las puntuaciones de un factor, fue Thurstone (1935). La expresión para estimar las puntuaciones es

$$\hat{\xi} = \hat{\Phi} \hat{\Lambda}_x^t \hat{\Sigma}^{-1} \mathbf{x}, \quad (3.37)$$

donde el factor $\hat{\Phi} \hat{\Lambda}_x^t \hat{\Sigma}^{-1}$ es el estimador de mínimos cuadrados ordinarios del coeficiente de la regresión hipotética de ξ sobre \mathbf{x} . Para la estimación de las puntuaciones, el método considera la covarianza entre los factores vía la matriz $\hat{\Phi}$, también la asociación entre el factor y sus indicadores vía $\hat{\Lambda}_x$ y las asociaciones entre las variables observadas vía $\hat{\Sigma}$.

- *Método de Bartlett.* Este método, formulado por Bartlett (1937), estima las puntuaciones del factor de manera similar al método de regresión pero con una mayor probabilidad de que los valores estimados sean iguales a las puntuaciones verdaderas del factor. La expresión para la estimación es

$$\hat{\xi} = [(\hat{\Lambda}_x^t \hat{\Theta}_\delta^{-1} \hat{\Lambda}_x)^{-1}]^t \hat{\Lambda}_x^t (\hat{\Theta}_\delta^{-1})^t \mathbf{x}. \quad (3.38)$$

El método de Bartlett solo considera la asociación entre el factor y sus indicadores y el término de varianza única. Un problema con el método es que puede producir factores puntuaciones de factores correlacionados aun cuando los factores no lo estén.

- *Método de Anderson-Rubin.* El método propuesto por Anderson and Rubin (1956) es una variante del método de Bartlett. Aquí, la regresión de mínimos cuadrados se ajusta de forma que los factores estimados no correlacionen cuando los factores sean oblicuos o no correlacionados. La ecuación de regresión es

$$\widehat{\boldsymbol{\xi}} = (\mathbf{UVU}^t)^{-1/2} \widehat{\boldsymbol{\Lambda}}_x^t (\widehat{\boldsymbol{\Theta}}_\delta^{-1})^t \mathbf{x}, \quad (3.39)$$

donde \mathbf{UVU}^t es la descomposición en valores singulares de $\widehat{\boldsymbol{\Theta}}_\delta$.

En la práctica, los distintos métodos producen estimaciones de los factores altamente correlacionadas (Bollen, 1989 p. 305). Independientemente del método usado para obtener los factores estimados, siempre $\widehat{\boldsymbol{\xi}} = \boldsymbol{\xi}$. Esta indeterminación es debido a que la suma de variables latentes y errores de medida es mayor que el número de variables observadas. Así, pueden formarse muchas estimaciones del factor y las puntuaciones individuales probablemente no se ordenen de la misma manera en cada estimación (Bollen, 1989 p.305). De acuerdo con esto, el uso de las puntuaciones estimadas debe ser cuidadoso.

Capítulo 4

El Modelo SEM-MDS

4.1. Introducción

En el capítulo 2 han sido examinados algunos de los modelos para datos asimétricos disponibles en MDS. La importancia del desarrollo de estructuras para el análisis de proximidades asimétricas, se fundamenta en la existencia de una gran diversidad de áreas en las que, con frecuencia, se generan este tipo de datos. Entre ellas, el área de investigación de mercado, la sociología, las ciencias de la conducta, ecología. Las asimetrías, por lo general, contienen información relevante que debe ser expuesta para lograr una comprensión más completa de los datos analizados.

En algunos casos, la asimetría es considerada no relevante y puede ser descartada con alguna transformación simple de los elementos de la matriz de proximidades asimétricas. El método de Kruskal (1964), por ejemplo, simetriza la matriz asimétrica original promediando las entradas correspondientes a ambos lados de la diagonal y asignándole este valor a todas las entradas de la matriz. Así, la matriz resultante puede ser representada por un modelo MDS simétrico. La asimetría también puede ser evitada con diseños experimentales apropiados, cuando se asume que es producto de alguna clase de sesgo en las respuestas. Sin embargo, en otros casos la asimetría tiene un efecto significativo sobre la estructura de los datos y en consecuencia, tiene importancia para la explicación de la distribución espacial de las proximidades.

Como se expuso en el capítulo introductorio, las asimetrías son importantes en el estudio de los aspectos complejos de los juicios de disimilaridad. En este contexto, las confusiones aportan información valiosa para analizar

el efecto de la atención y de los aspectos contextuales y referenciales de las tareas cognitivas.

Los modelos MDS para datos asimétricos, pueden clasificarse de manera bastante general en dos grandes tipos. Los modelos en el primer grupo, se basan en la descomposición de la matriz de datos en un componente simétrico y un componente antisimétrico. Algunos de los métodos de este tipo, representan el componente simétrico con técnicas de MDS mientras que el componente antisimétrico es representado por técnicas de visualización basadas en su descomposición en valores singulares. (ver por ejemplo, Constantine & Gower, 1978; Gower, 1977). Otros procedimientos de este grupo, representan simultáneamente ambos componentes, simétrico y antisimétrico, basados en la descomposición espectral de una matriz Hermitiana de tal forma que el componente simétrico es representado como puntos en un espacio MDS y la parte antisimétrica se muestra como vectores direccionales en el mismo espacio (Escoufier & Ground, 1980).

Un segundo grupo, incluye modelos que resultan de la generalización de los modelos de distancia y de producto escalar. Todos estos modelos fueron revisados parcialmente en el capítulo 2 y más extensamente en Saito & Yadohisa (2005 cap. 4).

El modelo para el análisis de disimilaridades asimétricas de un modo que se propone en esta investigación, aborda una perspectiva diferente producto de la incorporación de herramientas metodológicas de los modelos de ecuaciones estructurales. El supuesto fundamental del modelo propuesto, es que la matriz de disimilaridades asimétricas observadas es una medición imperfecta de una matriz de disimilaridades simétrica no observada. Además, se asume que las desviaciones de la simetría son debidas a errores de medición. Así, la matriz subyacente es medida teóricamente por dos indicadores, uno correspondiente a la parte triangular inferior de la matriz asimétrica observada y otro correspondiente a la parte triangular superior. La matriz simétrica latente se estima de manera que explique la mayor cantidad de varianza de los indicadores. Así, el modelo propuesto toma en consideración las asimetrías observadas como función de las varianzas de los errores de medida y las cargas factoriales, en una estructura que contribuye a mejorar la configuración MDS estimada. Como resultado, las disimilaridades asimétricas pueden ser completamente representadas como distancias entre los puntos de la configuración obtenida.

En el proceso de estimación se emplea un algoritmo de mínimos cuadrados alternante basado en SMACOF (de Leeuw & Heiser, 1980). El procedimiento

es tal que en cada iteración se obtiene una configuración MDS y, al propio tiempo, la matriz de disimilaridades simétricas desconocida se estima en un problema de estructura de covarianzas (ver Vera, Macias & Heiser, 2009a, 2009b, para tales procedimientos de MDS combinados en un contexto de clasificación).

El modelo propuesto podría ser de interés cuando las disimilaridades son obtenidas en experimentos diseñados para estudiar diferentes tipos de juicios psicológicos de proximidad. Algunas de las proximidades surgidas en este contexto, se expresan como medidas de confusión estímulo-estímulo o confusión estímulo-respuesta. En el primer caso, los sujetos juzgan si dos estímulos presentados ante ellos, en un orden específico, son iguales o diferentes. En el segundo caso, los sujetos deben identificar al estímulo presentado ante ellos. En ambas situaciones, las matrices de disimilaridad generadas son, por lo general, asimétricas.

En Miller & Nicely (1995) y Rothkopf (1957), pueden verse ejemplos de como obtener matrices de disimilaridad en el contexto señalado. Miller & Nicely realizaron un análisis articulatorio para examinar ciertos tipos de confusiones perceptuales. Los tipos analizados surgen cuando los sujetos escuchan palabras que son pronunciadas bajo condiciones de distorsión de frecuencia o ruido de enmascaramiento. El análisis articulatorio estudia puntuaciones que son función del porcentaje de palabras que el individuo escucha correctamente. En las situaciones experimentales reportadas por Miller & Nicely, los individuos son forzados a adivinar los sonidos emitidos por sistemas de comunicación de voz, sujetos a distorsiones de frecuencia y ruido. Así, el análisis de percepción de Miller & Nicely es una situación de confusión estímulo-respuesta. En tal situación, la asimetría aparece debido a que el número de veces en la que los sujetos confunden el estímulo i con el estímulo j , es diferente al número de veces en las que el estímulo j es confundido con el estímulo i . En el capítulo (4.5), los datos de confusiones perceptuales son analizados con el modelo SEMDS propuesto y los resultados son comparados con los obtenidos por Miller & Nicely (1995). La comparación muestra que el procedimiento propuesto produce una mejor configuración en términos del STRESS.

Rothkopf (1957) llevó a cabo un conocido estudio para medir la inteligibilidad del código Morse. El propósito del estudio fue mostrar cuáles de los códigos de este sistema de representación de letras y números, son más fáciles de confundir. Cada código Morse consiste de un conjunto de señales cortas (0.05 seg.) y/o largas (0.15 seg.). Cuando se utilizan señales cortas y largas

en una misma representación, éstas van separadas por períodos de silencio de 0.05 seg. de duración. El objetivo del estudio fue establecer cuáles de los códigos Morse son más fáciles de confundir. Rothkopf presentó a cada uno de los 598 individuos de la muestra todas las combinaciones diferentes de pares de códigos, o estímulos según el argot en MDS. Cada par de estímulos fue emitido uno después del otro con una separación de 0.05 seg., y los respondiente debían decidir rápidamente si ambos estímulos eran los mismos o eran diferentes. En esta situación de confusión estímulo-estímulo, se produce una matriz asimétrica debido a que el porcentaje de veces en las que los respondientes consideran que los estímulos i y j , presentados en ese orden, (i, j) , son los mismos, varía con respecto a si el mismo par de estímulos es presentado en el orden inverso, (j, i) . En pruebas preliminares pudo verse que en general el modelo SEMDS se desempeña mejor que el algoritmo SMACOF aplicado sobre la simetrización clásica, $\Delta = (\Delta_1 + \Delta_2)/2$. Sin embargo, hay pocas diferencias en los resultados obtenidos por ambos procedimientos lo cual probablemente se deba a la alta correlación ($r > 0,8$) entre las partes triangular superior e inferior de la matriz de confusión original del código Morse.

4.1.1. Optimización en MDS: El algoritmo SMACOF

La minimización de una función de pérdida tal como (??) no siempre puede resolverse analíticamente. Para ello, se han programado algoritmos numéricos con procesos iterativos para hallar \hat{d} y \mathbf{X} simultáneamente, que minimicen la función. Uno de estos algoritmos es SMACOF propuesto por De Leeuw y Heisser (1977, 1980) y De Leeuw (1977, 1988), basado en el principio de mayorización iterativa.

La idea central del método de optimización por mayorización, es reemplazar iterativamente la función original que debe ser minimizada y que es, por lo general, compleja, digamos $\varphi(x)$, por una función auxiliar, digamos $\hat{\varphi}(x, y)$, donde y en $\hat{\varphi}(x, y)$ es un valor fijo. La función g debe satisfacer las siguientes condiciones en orden a ser usada como función mayorante en el proceso de minimización de $\varphi(x)$:

1. $\hat{\varphi}(x, y)$ debe ser más simple de minimizar que $\varphi(x)$.

2. $\varphi(x) \leq \hat{\varphi}(x, y) \quad \forall x, y$

3. $\hat{\varphi}(x, y)$ debe tocar a $\varphi(x)$ en el punto y , llamado punto de apoyo. Es decir,

$$\varphi(y) = \hat{\varphi}(y, y)$$

El algoritmo puede resumirse de la siguiente manera:

1. $y_0 \rightarrow y$, donde y_0 es un valor inicial.
2. Hallar x^+ para el cual $\hat{\varphi}(x^+, y) = \min_x \hat{\varphi}(x, y)$.
3. si $\varphi(y) - \varphi(x^+) < \epsilon$ el proceso termina. (ϵ es una constante pequeña y positiva).
4. $x^+ \rightarrow y$ e ir al paso 2.

El algoritmo produce una sucesión no creciente de valores de la función, garantizando la convergencia del proceso.

Así, la función Stress dada en (??) es mayorizada por la función

$$\sigma_r^2(\hat{d}, X, Z) = \eta_\delta^2 + tr(X'VX) - 2tr(X'B(Z)Z), \quad (4.1)$$

que es una función cuadrática en X . Si el gradiente respecto de \mathbf{X} es igual a 0, se obtiene el punto donde la función alcanza el mínimo, dado por

$$\mathbf{X}^u = \mathbf{V}^+\mathbf{B}(\mathbf{Z})\mathbf{Z}, \quad (4.2)$$

Donde V^+ es la inversa de Moore-Penrose de la matriz $V = (v_{ij})$, con $v_{ij} = -w_{ij}$ if $i \neq j$ and $v_{ii} = \sum_{j=1, j \neq i}^n w_{ij}$.

De Leeuw and Heiser (1980) llamaron a (4.2) la transformada de Guttman en reconocimiento a Guttman (1968). Ver Borg & Groenen (2005) para más detalles del algoritmo.

Dentro de las clases de modelos MDS, el modelo propuesto es un modelo de intervalo. En los modelos de tipo intervalo, la relación entre las distancias y las disimilaridades está dada por una transformación lineal tipo: $d_{ij} = b\delta_{ij} + a$ en donde $b > 0$ (si las proximidades son similaridades b debe ser menor que cero. ver Borg & Groenen, 2005 pág. 201) y a una constante tal que se cumpla la desigualdad triangular para todas las ternas de disimilaridades. Si la

desigualdad triangular no se satisface para todas las ternas posibles de disimilaridades, es posible sumar a todas ellas una constante a suficientemente grande para resolver el problema. Si, por ejemplo, la mayor violación de este requerimiento se da para la terna δ_{rs}, δ_{st} y δ_{rt} , de tal forma que $\delta_{rs} + \delta_{st} \leq \delta_{rt}$, debería seleccionarse a de tal forma que $\delta_{rs} + a + \delta_{st} + a \geq \delta_{rt} + a$, para lo cual basta tomar $a \geq \delta_{rt} - \delta_{rs} - \delta_{st}$. La transformación $b\delta_{ij} + a$ produce las disparidades \widehat{d}_{ij} que son incorporadas al STRESS (ver 1.1.3) como

$$\sigma_r(\widehat{\mathbf{d}}, \mathbf{X}) = \sum_{i < j}^{i-1} (\widehat{d}_{ij} - d_{ij}(\mathbf{X}))^2, \quad (4.3)$$

así, el algoritmo SMACOF es

1. Se hace $\mathbf{Z} = \mathbf{X}^{[0]}$, donde $\mathbf{X}^{[0]}$ es alguna configuración inicial aleatoria, o no. Se inicia el contador de iteraciones $k = 0$ y se fija ϵ a una constante positiva pequeña.
2. Se estiman las disparidades óptimas para las distancias de la configuración inicial $d_{ij}(\mathbf{X}^{[0]})$.
3. Las disparidades óptimas \widehat{d}_{ij} se estandarizan de tal forma que $\eta_d^2 = n(n-1)/2$ (ver ecuación 1.23).
4. Se calcula $\sigma_r^{[0]} = \sigma_r(\widehat{\mathbf{d}}, \mathbf{X}^{[0]})$, tomando $\sigma_r^{[-1]} = \sigma_r^{[0]}$.
5. Mientras que $k = 0$ o $(\sigma_r^{[k-1]} - \sigma_r^{[k]})$ y $k \leq$ máximo número de iteraciones) hacer
6. $k = k + 1$.
7. Calcular la transformada de Guttman $\mathbf{X}^{[k]}$ por (4.2).
8. Estimar \widehat{d}_{ij} para $d_{ij}(\mathbf{X}^{[k]})$.
9. Estandarizar \widehat{d}_{ij} de forma que $\eta_d^2 = n(n-1)/2$. item Calcular $\sigma_r(\widehat{\mathbf{d}}, \mathbf{X}^{[k]})$.
10. Hacer $\mathbf{Z} = \mathbf{X}^{[k]}$

4.2. El modelo

Considérese el conjunto de n objetos dado por $O = \{ o_i \mid i = 1, \dots, n \}$, y sea $\Delta = (\delta_{ij})$, $i, j = 1 \dots n$, una matriz asimétrica de disimilaridades, con Δ_1 y Δ_2 las partes triangular superior e inferior de Δ , respectivamente, escritas en forma de vector. En este contexto, Δ_1 y Δ_2 son indicadores de una matriz simétrica latente $\Delta^* = (\delta_{ij}^*)$. El objetivo del MDS es encontrar una configuración, \mathbf{X} , de n puntos x_i , $i = 1, \dots, n$, en un espacio Euclidiano de dimensión k , de tal forma que las distancias entre cada par de puntos de \mathbf{X} , sean la mejor aproximación posible para las disimilaridades en Δ^* . Es decir, si $\mathbf{D}(\mathbf{X}) = (d_{ij}(\mathbf{X}))$ es la matriz de distancias Euclidianas dada por

$$d_{ij}(\mathbf{X}) = \left(\sum_{k=1}^K (x_{ik} - x_{jk})^2 \right)^{1/2} .$$

entonces, $\mathbf{D}(\mathbf{X}) = (d_{ij}(\mathbf{X}))$ es la mejor aproximación de Δ^* .

El problema MDS puede ser formulado en un contexto de mínimos cuadrados, minimizando,

$$STRESS = \sum_{i < j} \left(\hat{d}_{ij} - d_{ij}(\mathbf{X}) \right)^2 , \quad (4.4)$$

donde \hat{d}_{ij} values, $i, j = 1, \dots, n$, representan disimilaridades simétricas o, en general, disimilaridades transformadas denominadas también disparidades. (ver Vera, Heiser and Murillo, 2007 para un algoritmo MDS global en cualquier métrica de Minkowski).

Debido a que en el contexto de MDS asimétrico, la matriz simétrica de disimilaridades latente, Δ^* , es una matriz no conocida, entonces el vector cuyos elementos son las entradas de la parte triangular de Δ^* , puede ser considerado como una variable latente que denotaremos por δ^* que puede ser estimada en el contexto de un modelo de estructura de covarianzas.

En general, si se tienen solo una variable endógena, η , y una variable exógena, ξ , entonces el modelo de ecuaciones estructurales puede formularse como

$$\eta = \gamma\xi + \zeta, \quad (4.5)$$

donde el coeficiente γ mide el efecto directo de ξ sobre η y ζ es un término de perturbaciones aleatorias o término de error en la ecuación, que da cuenta

de los factores que pudieran afectar a η y que no han sido considerados explícitamente en el modelo. Se asume que el valor esperado de ζ es cero y no correlaciona con ξ .

En la situación particular donde la variable exógena η es medida por solo un indicador, el modelo de medida puede expresarse por,

$$y = \lambda_y \eta + e, \quad (4.6)$$

$$\mathbf{x} = \boldsymbol{\lambda}_x \xi + \boldsymbol{\epsilon}, \quad (4.7)$$

donde $\mathbf{x} = (x_1, \dots, x_p)^T$ es el vector de variables indicadoras de la variable exógena ξ . El coeficiente λ_y y el vector $\boldsymbol{\lambda}_x = (\lambda_1, \dots, \lambda_p)^T$ representan los efectos esperados de η sobre y y de ξ sobre \mathbf{x} , respectivamente. La variable e y el vector $\boldsymbol{\epsilon}$ son los términos de error de medida para las variables η y ξ , respectivamente, los cuales no correlacionan entre ellos ni con las variables latentes η y ξ . Sin pérdida de generalidad, se asume que los términos de error tienen media cero y que \mathbf{x} , y , η y ξ se escriben desviadas de sus respectivas medias.

La expresión dada en (4.5), es una particularización del modelo dado en (3.7):

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}, \quad (4.8)$$

mientras que las expresiones dadas en (4.6) y (4.7) son casos particulares de los modelos dados en (4.9) y (4.10):

$$\mathbf{y} = \boldsymbol{\Lambda}_y \boldsymbol{\eta} + \boldsymbol{\epsilon} \quad (4.9)$$

$$\mathbf{x} = \boldsymbol{\Lambda}_x \boldsymbol{\xi} + \boldsymbol{\delta}, \quad (4.10)$$

Si $\boldsymbol{\theta}$ denota al vector de parámetros del modelo, entonces la matriz de covarianza implicada por el modelo en las ecuaciones (4.5) a (4.7) adopta la expresión,

$$\boldsymbol{\Sigma}_E(\boldsymbol{\theta}) = \begin{bmatrix} \lambda_y^2(\gamma^2\sigma_\xi^2 + \sigma_e^2) + \sigma_e^2 & \\ \gamma\lambda_y\sigma_\xi^2\boldsymbol{\lambda}_x & \sigma_\xi^2\boldsymbol{\lambda}_x\boldsymbol{\lambda}_x^T + \boldsymbol{\Sigma}_\epsilon \end{bmatrix}, \quad (4.11)$$

donde σ_ξ^2 y σ_ζ^2 son las varianzas de ξ and ζ respectivamente, mientras que σ_e^2 , and $\boldsymbol{\Sigma}_\epsilon$ son las varianzas de e y la matriz de covarianza de los elementos de $\boldsymbol{\epsilon}$, respectivamente.

la matriz en (4.11) es derivada de la expresión general de la matriz de covarianza implicada para el modelo completo, dada por

$$\Sigma(\boldsymbol{\theta}) = \begin{pmatrix} \Lambda_y(\mathbf{I} - \mathbf{B})^{-1}(\Gamma\Phi\Gamma^t + \Psi)[(\mathbf{I} - \mathbf{B})^{-1}]^t\Lambda_y^t + \theta_\epsilon & \Lambda_y(\mathbf{I} - \mathbf{B})^{-1}\Gamma\Phi\Lambda_x^t \\ \Lambda_x\Phi\Gamma^t[\mathbf{I} - \mathbf{B}]^{-1}]^t\Lambda_y^t & \Lambda_x\Phi\Lambda_x^t + \Theta_\delta \end{pmatrix}. \quad (4.12)$$

Las consideraciones anteriores, permiten formular el problema MDS de estimar la configuración \mathbf{X} para la matriz desconocida Δ^* , como un problema de estructura de covarianza. Así, en un contexto MDS bajo transformaciones admisibles y siguiendo la metodología para la estimación en el ámbito de los modelos de estructura de covarianza, el modelo puede formularse como

$$d^* = b \delta^* + \zeta \quad (4.13)$$

$$d(\mathbf{X}) = \lambda_d d^* + e \quad (4.14)$$

$$\delta_1 = \lambda_1 \delta^* + \epsilon_1 \quad (4.15)$$

$$\delta_2 = \lambda_2 \delta^* + \epsilon_2. \quad (4.16)$$

donde d^* es la variable latente de las distancias desconocidas que mejor aproximan a los proximidades en Δ^* y cuya relación es modelada por la ecuación (4.13). Así, d^* es la variable latente de los elementos de la parte triangular de la matriz simétrica desconocida \mathbf{D}^* , la cual es medida en cada iteración por la variable observada $d(\mathbf{X})$ tal como está dado en la ecuación (4.14). los valores de $d(\mathbf{X})$ se obtienen de una estimación previa de la configuración \mathbf{X} . Por otro lado, δ^* es la variable latente de los elementos de la parte triangular de la matriz simétrica desconocida Δ^* , medida según las ecuaciones (4.15) y (4.16). A su vez, δ_1 y δ_2 corresponden a las parte triangular superior e inferior, respectivamente, de la matriz asimétrica de disimilaridades observadas Δ .

La ecuación (4.13) modela la relación estructural espacial entre las distancias desconocidas \mathbf{D}^* y las disimilaridades simétricas desconocidas Δ^* . En el ámbito de MDS, esta relación puede ser percibida como una relación de disparidad con un término de error, donde las distancias son determinadas por un procedimiento de estimación alternante. b es un coeficiente de regresión que describe el efecto de las disimilaridades sobre las distancias, y ζ es

el término de error referido a la relación espacial con media cero y varianza σ_ζ^2 , y no correlacionado con δ^* . Así, $\widehat{\mathbf{D}} = \widehat{b}\widehat{\mathbf{\Delta}}^*$, donde b es también estimado en la fase SEM, representa las disparidades óptimas en MDS. Después que las disparidades son estimadas, las distancias pueden ser estimadas considerándolas predictores de las disparidades simétricas, en un procedimiento de estimación alternante usando la transformada de Guttman (de Leeuw & Heiser, 1980).

El modelo de medida para d^* dado en la ecuación (4.14), está expresado en términos de un solo indicador, $d(\mathbf{X})$. Para resolver el problema de identificación que aparece cuando la variable latente es media con un sólo indicador, la varianza del error, σ_ζ^2 , se expresa como proporcional a la varianza de las distancias conocidas, $\sigma_{d(\mathbf{X})}^2$, de tal forma que $\sigma_\zeta^2 = k\sigma_{d(\mathbf{X})}^2$ donde k es un número entre 0 y 1. El valor de k puede estar predeterminado o puede ser establecido sobre la base del valor de disminución del STRESS para cada iteración. También, el valor de λ_d se fija a 1.

El modelo de medida para δ^* está dado en las ecuaciones (4.15) y (4.16). El modelo expresa que las partes triangulares de la matriz asimétrica de disimilaridades $\mathbf{\Delta}$, son indicadores efecto de la matriz simétrica de disimilaridades no observada $\mathbf{\Delta}^*$. El modelo asume que los términos de error ϵ_1 y ϵ_2 no correlacionan entre ellos ni con δ^* , d^* , e ni ζ . Además, se asume que los términos de error ϵ_1 y ϵ_2 tienen igual varianza, lo cual es coherente con el ámbito usual de MDS.

Para el objetivo de la identificación global del modelo, la escala de δ^* se fija tomando $\sigma_{\delta^*}^2 = 1$. Así, la matriz de covarianza implicada por el modelo dada en la ecuación (4.11) puede escribirse como

$$\Sigma_E(\boldsymbol{\theta}) = \begin{bmatrix} b^2 + \sigma_\zeta^2 + k\sigma_{d(\mathbf{X})}^2 & \\ b\boldsymbol{\lambda}_x & \boldsymbol{\lambda}_x\boldsymbol{\lambda}_x^T + \Sigma_\epsilon \end{bmatrix}, \quad (4.17)$$

donde $\boldsymbol{\lambda}_x = (\lambda_1, \lambda_2)^T$, y $\Sigma_\epsilon = \sigma_\epsilon^2\mathbf{I}$ denota la matriz de covarianza del vector $\boldsymbol{\epsilon}$.

Así, $\boldsymbol{\theta} = (b, \lambda_1, \lambda_2, \sigma_\epsilon^2, \sigma_\zeta^2)^t$ y dado que el número de elementos no redundantes de Σ es $p(p+1)/2 = 6$, donde p es el número de variables observadas, se sigue que el modelo propuesto satisface la regla t . Sin embargo, esta regla es apenas una condición necesaria y algunas características del modelo ponen en duda su identificación. En particular, una variable latente con solo dos indicadores emite una trayectoria hacia otra variable latente con un

solo indicador. Modelos con esta característica pudieran no estar globalmente identificados (p.ej. Bollen & Davis, 2009; Davis, 1993). Sin embargo, la sencillez del modelo permite establecer su identificación algebraicamente.

Partiendo de la ecuación $\Sigma = \Sigma(\theta)$, con

$$\Sigma = \begin{bmatrix} \text{VAR}(d^*) & & \\ \text{COV}(\delta_1, d^*) & \text{VAR}(\delta_1) & \\ \text{COV}(\delta_2, d^*) & \text{COV}(\delta_2, \delta_1) & \text{VAR}(\delta_2) \end{bmatrix}, \quad (4.18)$$

y $\Sigma(\theta)$ como está dada en (4.17), surge el sistema de ecuaciones

$$\text{VAR}(d^*) = b^2 + \sigma_\zeta^2 + k\sigma_{d(\mathbf{x})}^2 \quad (4.19)$$

$$\text{COV}(\delta_1, d^*) = \lambda_1 b \quad (4.20)$$

$$\text{COV}(\delta_2, d^*) = \lambda_2 b \quad (4.21)$$

$$\text{VAR}(\delta_1) = \lambda_1^2 + \sigma_\epsilon^2 \quad (4.22)$$

$$\text{COV}(\delta_2, \delta_1) = \lambda_1 \lambda_2 \quad (4.23)$$

$$\text{VAR}(\delta_2) = \lambda_2^2 + \sigma_\epsilon^2. \quad (4.24)$$

De (4.20) y (4.21) se sigue que

$$\lambda_1 = \frac{\lambda_2 \text{COV}(\delta_1, d^*)}{\text{COV}(\delta_2, d^*)}, \quad (4.25)$$

de (4.23) se obtiene

$$\lambda_2 = \frac{\text{COV}(\delta_1, \delta_2)}{\lambda_1}, \quad (4.26)$$

así, una expresión que establece la identificación de λ_1 se obtiene al sustituir (4.26) en (4.25), con lo cual

$$\lambda_1 = \left(\frac{\text{COV}(\delta_1, \delta_2) \text{COV}(\delta_1, d^*)}{\text{COV}(\delta_2, d^*)} \right)^{1/2}. \quad (4.27)$$

Análogamente, se prueba que

$$\lambda_2 = \left(\frac{\text{COV}(\delta_1, \delta_2) \text{COV}(\delta_2, d^*)}{\text{COV}(\delta_1, d^*)} \right)^{1/2}. \quad (4.28)$$

Luego, σ_ϵ^2 puede ser identificado a partir de (4.22) o (4.24) y b a partir de (4.20) o (4.21). Finalmente, sustituyendo la expresión obtenida para b en (4.19) se demuestra que ζ también está identificado.

Una vez que los parámetros estructurales han sido estimados, el modelo propuesto requiere la estimación de las entradas de la matriz simétrica latente no observada. Tal estimación es necesaria porque las disparidades deben ser conocidas en orden a estimar la configuración \mathbf{X} . El problema de estimar las entradas de la parte triangular de la matriz simétrica, es el problema de estimar las puntuaciones del factor en análisis factorial. Existen distintos métodos para la estimación de las puntuaciones de los factores en el contexto del análisis factorial. Los métodos asumen que los verdaderos valores de los parámetros son conocidos. En la práctica, solo se tienen estimaciones con lo que el supuesto significa que las estimaciones son iguales a los valores verdaderos. Así, las fluctuaciones de muestro no son consideradas. El método de Thomson (1935), tal vez el más usado para la recuperación de las puntuaciones de los factores, se describe a continuación.

El modelo unifactorial puede escribirse como

$$\mathbf{x} = \mathbf{\Lambda}_x f + \mathbf{e}, \quad (4.29)$$

donde f es el factor, \mathbf{x} es el vector de indicadores del factor, $\mathbf{\Lambda}_x$ es el vector de cargas factoriales y \mathbf{e} es un vector que en el contexto del análisis factorial se denominan términos de varianza única, y el contexto de los modelos de medida en SEM, son los errores de medida. Así, la matriz de covarianzas de las variables observadas, $\mathbf{\Sigma}$, está dada por

$$\begin{aligned} \mathbf{\Sigma} &= \text{E}(\mathbf{xx}^t) \\ &= \text{E}[(\mathbf{\Lambda}_x f + \mathbf{e})(\mathbf{\Lambda}_x f + \mathbf{e})^t] \\ &= \mathbf{\Lambda}_x \text{E}(f f^t) \mathbf{\Lambda}_x^t + \text{E}(\mathbf{e} \mathbf{e}^t) \\ &= \mathbf{\Lambda}_x \mathbf{\Phi} \mathbf{\Lambda}_x^t + \mathbf{\Psi}. \end{aligned} \quad (4.30)$$

y se busca una función lineal de las observaciones que provea un buen predictor de f . Sea

$$f = \mathbf{a}^t \mathbf{x} = \mathbf{x}^t \mathbf{a}, \quad (4.31)$$

\mathbf{a} es un vector que debe ser seleccionado de manera que minimice la varianza de la diferencia entre los valores estimados y los verdaderos valores del factor. Tal varianza viene dada por

$$E(\widehat{f} - f)^2 = E(\mathbf{x}^t \mathbf{a} - f)^2. \quad (4.32)$$

Para minimizar (4.32), se iguala su derivada con respecto al vector \mathbf{a} , a cero. La derivada es,

$$E[2\mathbf{x}(\mathbf{x}^t \mathbf{a} - f)] = 2(\boldsymbol{\Sigma} \mathbf{a} - \boldsymbol{\lambda}), \quad (4.33)$$

donde $\boldsymbol{\lambda} = \boldsymbol{\Lambda}$ para un modelo unifactorial. Igualando a cero la expresión en (4.33), se obtiene,

$$\boldsymbol{\Sigma} \mathbf{a} = \boldsymbol{\lambda}, \quad (4.34)$$

o

$$\mathbf{a} = \boldsymbol{\Sigma}^{-1} \boldsymbol{\lambda}, \quad (4.35)$$

así, las puntuaciones estimadas vienen dadas por

$$\widehat{f} = \widehat{\boldsymbol{\lambda}}^t \widehat{\boldsymbol{\Sigma}}^{-1} \mathbf{x}. \quad (4.36)$$

En términos del modelo propuesto, (4.36) se escribe como

$$\widehat{\boldsymbol{\Delta}}^* = \widehat{\boldsymbol{\lambda}}_x^T \widehat{\boldsymbol{\Sigma}}^{-1} \widetilde{\boldsymbol{\Delta}}^T, \quad (4.37)$$

donde $\widetilde{\boldsymbol{\Delta}} = [\delta_1 \ \delta_2]$ es la matriz por bloques de dimensión $n(n-1)/2 \times 2$ conformada por los valores de las disimilaridades observadas, $\widehat{\boldsymbol{\lambda}}_x = (\widehat{\lambda}_1, \widehat{\lambda}_2)^t$, y $\widehat{\boldsymbol{\Sigma}}^{-1}$ denota la matriz inversa de los valores estimados de $\widehat{\boldsymbol{\lambda}}_x \widehat{\boldsymbol{\lambda}}_x^T + \widehat{\boldsymbol{\Sigma}}_\epsilon$.

4.3. Errores de medida no modelados

La incorporación de los errores de medida en el modelo SEMDS es un importante beneficio derivado de la metodología SEM para la estimación de los parámetros. Distintos trabajos han examinado el efecto de los errores de medida no modelados sobre la estimación de parámetros en análisis de regresión (p. ej. Rigdon, 2004; Oliver et al, 1999; Bollen, 1989). Lo más común en la investigación práctica es que la medición de las variables observadas esté sujeta a errores aleatorios. Sin embargo, algunos de los modelos clásicos utilizados con frecuencia en distintas áreas tales como regresión y los modelos multiecuacionales no poseen facilidades metodológicas para modelarlos.

Ignorar los errores de medida tiene distintos efectos de acuerdo con la estructura del modelo utilizado. En el modelo de regresión múltiple clásico tal como

$$\eta = \beta_{11}\xi_1 + \beta_{12}\xi_2 + \cdots + \beta_{1n}\xi_n + \zeta, \quad (4.38)$$

donde el modelo de medida está dado por

$$\begin{aligned} y &= \eta \\ x_i &= \xi_i \quad \text{con } i = 1, \dots, n, \end{aligned}$$

cada variable observada se considera una medida perfecta de su correspondiente constructo con lo cual se asume que la varianza del error es cero. Si, contradiciendo este supuesto, una de las variables observadas, digamos x_1 , tiene asociado un error de medida no incluido en el modelo, entonces β_{11} , el parámetro asociado con ξ_1 , se verá atenuado en su estimación con respecto a su verdadero valor (Johnston, 1984). Es decir, $\hat{\beta}_{11}$ subestimarán a β_{11} .

Por otro lado, si la correlación de ξ_1 con otro constructo, digamos ξ_j ($j \neq 1$), es significativa, entonces β_{1j} también experimentará el efecto de atenuación aun cuando x_j no esté sujeto a errores de medida. La magnitud de la atenuación depende positivamente de la varianza del error y de la correlación entre los constructos (Griliches & Ringstad, 1971). Si ambas variables son medidas con error, β_{1j} estará sesgado pero la dirección del sesgo variará de una manera difícil de establecer (Maddala, 1988). Así, el error de medida de x_1 hace que la influencia de ξ_1 sobre η sea apenas parcialmente explicada y esto conduce a estimaciones sesgadas de los otros parámetros, tal como ocurre en un problema de variable omitida (Rigdon, 2004). En modelos con múltiples regresores correlacionados, los errores de medida no modelados en una variable pueden producir sesgos de distinto signo en la estimación de los parámetros para las otras variables.

Se han propuesto diferentes soluciones para encontrar buenas estimaciones ante la presencia de errores de media aleatorios. Entre ellas está el uso de regresión invertida (Goldberger, 1984) y el uso de variables instrumentales (Bartlett, 1949; Durbin, 1954; Pakes, 1982; Wald, 1940). En Psicometría se han desarrollado fórmulas para corregir las estimaciones de los parámetros por atenuación. Sin embargo, tales soluciones han sido cuestionadas por

ejemplo por Pedhazur & Schmelkin (1991) y a pesar de su larga tradición, el tema de las medidas correctivas en psicometría sigue siendo controvertido (Oliver et al, 1999).

En el contexto SEM, los errores no son considerados en el modelo estructural con variables observadas que considera que cada variable es una medida perfecta de una variable latente. Para este modelo se han reportado distintos efectos sobre las estimaciones. Podría haber un efecto de atenuación sobre las estimaciones de los parámetros en la matriz $\mathbf{\Gamma}$ mientras que los elementos de \mathbf{B} podrían experimentar sesgo positivo (Gillespie y Fox, 1980). Es decir, por un lado los $\hat{\gamma}$ podrían subestimar a los γ y por otro lado $\hat{\beta}$ podrían sobre estimar a los β , lo cual, a su vez, podría inducir una atenuación de las varianzas de los errores. En este sentido, los modelos de ecuaciones estructurales con variables latentes son una manera alternativa de disminuir el sesgo debido a la no inclusión de los errores de medida. Al modelar los errores, es posible analizar relaciones entre constructos eliminando sesgos debido a la atenuación (Huba y Harlow, 1987), así como otros efectos producidos por las correlaciones entre las variables (Reuterberg y Gustafsson, 1992).

Como ilustración de lo expuesto hasta ahora en esta sección, veamos el caso para un modelo de regresión simple siguiendo la exposición de Bollen (1989, pág. 154). Se asume que el modelo correcto es el modelo de variables latentes y se compara el coeficiente de regresión para este modelo con el coeficiente para el modelo de variables observadas que no considera los errores de medida. Considérese que el verdadero modelo para la relación entre las variables incluidas es

$$\begin{aligned}\eta &= \gamma\xi + \zeta \\ x &= \lambda_1\xi + \delta \\ y &= \lambda_2\eta + \epsilon,\end{aligned}\tag{4.39}$$

se asume que las variables están desviadas de su media. λ_1, λ_2 y γ son constantes distintas de cero y $E(\delta) = E(\epsilon) = E(\zeta) = 0$. δ y ϵ son los errores de medida para x e y respectivamente que no están correlacionados entre ellos, con ζ ni con las variables latentes. También, ζ y ξ están incorreladas. A fin de asignarle una escala a las variables latentes lo cual es una condición necesaria para la identificación, se fija $\lambda_1 = \lambda_2 = 1$. El objetivo es estimar la magnitud de la relación entre η y ξ medida por γ . Una opción es suponer

que las variables observadas son idénticas a las variables latentes que ellas miden con lo cual, el modelo en (4.39) se reduce a

$$y = \gamma^* x + \zeta^*. \quad (4.40)$$

El asterisco en γ y ζ en (4.40) los distinguen de los correspondientes parámetros para el modelo con variables latentes. Según lo expuesto anteriormente $\gamma^* \neq \gamma$ y $\zeta^* \neq \zeta$. En efecto,

$$\begin{aligned} COV(\xi, \eta) &= COV(\xi, \gamma\xi + \zeta) \\ &= \gamma VAR(\xi) + COV(\xi, \zeta) \\ &= \gamma\phi, \end{aligned} \quad (4.41)$$

donde ϕ es la varianza de ξ . Por otro lado, a partir del modelo de medida en (4.39) y dado que $\lambda_1 = \lambda_2 = 1$, se obtiene

$$\begin{aligned} COV(x, y) &= COV(\xi + \delta, \eta + \epsilon) \\ &= COV(\xi + \delta, \gamma\xi + \zeta + \epsilon) \\ &= \gamma VAR(\xi) \\ &= \gamma\phi, \end{aligned} \quad (4.42)$$

así, las covarianzas de las variables latentes y las covarianzas poblacionales de las variables observadas son iguales. Por otro lado, de (4.40) se obtiene

$$\begin{aligned} COV(x, y) &= COV(x, \gamma^* x + \zeta^*) \\ &= \gamma^* VAR(x) + COV(x, \zeta^*) \\ &= \gamma^* VAR(x), \end{aligned} \quad (4.43)$$

con lo cual

$$\gamma^* = \frac{COV(x, y)}{VAR(x)}, \quad (4.44)$$

sustituyendo (4.42) en (4.44) se obtiene

$$\gamma^* = \gamma \left[\frac{\phi}{VAR(x)} \right]. \quad (4.45)$$

El resultado obtenido en (4.45) dice que el coeficiente de regresión γ^* del modelo de variables observadas que no considera los errores de medida, será igual al coeficiente de regresión γ del modelo con variables latentes que sí considera los errores de medida, si y solo si $VAR(x) = \phi$. De (4.39), para $\lambda_1 = 1$, se tiene que $VAR(x) = VAR(\xi) + VAR(\delta) = \phi + VAR(\delta)$. Así, $\gamma^* = \gamma$ solo si la varianza del error de medida para x es cero, es decir en ausencia del error de medida. En caso contrario se tiene que $\phi < \phi + VAR(\delta)$ y así, el término entre corchetes en (4.45) es menor que uno y $\gamma^* < \gamma$. De tal modo que en presencia de errores de medida para la variable exógena, el coeficiente para el modelo de variables observadas experimenta un efecto de atenuación relativo al coeficiente para el modelo de variables latentes. De hecho, γ^* es un estimador inconsistente de γ (ver Bollen, 1989 para una revisión más extensa del efecto de los errores de medida no modelados).

Gillespie & Fox (1980) examina los efectos de errores de medida presentes pero no modelados en un sistema estructural con dos ecuaciones. Gillespie & Fox encontraron que los errores de medida no incorporados al modelo tiene un efecto sobre los parámetros que relacionen a las variables exógenas con las variables endógenas, y el efecto se extiende a los parámetros que miden las relaciones entre variables endógenas así como a las estimaciones de las varianzas y covarianzas de los errores en las ecuaciones. Por otro lado, la covarianza que está asociada con la influencia de las variables exógenas, ante errores no modelados es realmente explicada por las variables endógenas, lo que hace que los estimadores $\hat{\beta}$ sobre estimen los valores de los parámetros lo que podría conducir a sesgo negativo en la estimación de la covarianza de los términos de error en las ecuaciones o errores estructurales. Para una revisión detallada de los resultados obtenidos por Gillespie & Fox (1980) ver por ejemplo, Rigdon (2004). Efectos sobre datos longitudinales pueden verse en Fergusson & Horwood, (1984).

4.4. Resultados: Un algoritmo SEM-MDS alternante

En MDS de mínimos cuadrados, cuando se considera la posibilidad de admitir una transformación de las disimilaridades o disparidades, el proceso de estimación involucra dos fases. En la primera fase los parámetros del modelo son estimados asumiendo las disparidades conocidas. En la segunda fase, se estiman las disparidades con los demás parámetros desconocidos. El procedimiento continúa hasta minimizar el STRESS. Así, el objetivo en del modelo MDS de razón de mínimos cuadrados propuesto es determinar una configuración \mathbf{X}^* , tal que las distancias $d_{ij}^* = d_{ij}(\mathbf{X}^*)$ se aproximen tanto como sea posible a las disparidades simétricas $\hat{d}_{ij}^* = \hat{b}\delta_{ij}^*$, como resultado de minimizar

$$STRESS = \sum_{i < j} \left(\hat{d}_{ij}^* - d_{ij}(\mathbf{X}^*) \right)^2. \quad (4.46)$$

En presencia de variables latentes o errores de medida en el análisis de regresión, la metodología SEM produce estimaciones de los parámetros al menos tan buenas como las producidas por el procedimiento de mínimos cuadrados ordinarios (Bentler, 1983; Joreskog, 1978). Así, en términos del modelo propuesto las disparidades pueden ser estimadas en un contexto de estructura de covarianzas a través de un procedimiento de estimación alternante que minimiza el (4.46) usando SMACOF (de Leeuw & Heiser, 1980). Una vez establecidos los valores iniciales para los parámetros, se estima la configuración \mathbf{X} a partir de la minimización de (4.46). Luego, se estiman los valores de la matriz simétrica $\hat{\Delta}^*$, a partir de las distancias calculadas sobre la configuración obtenida en la fase previa. Esta fase también logra disminuir el STRESS.

Para la r -ésima interacción, la matriz de distancias $\mathbf{D}(\widehat{\mathbf{X}}^{(r-1)})$ es conocida cuando se entra en la fase SEM. Así, la matriz de covarianza muestral, $\mathbf{S}^{(r)}$, se puede escribir como

$$\mathbf{S}^{(r)} = \begin{bmatrix} \text{VAR}(\mathbf{D}(\widehat{\mathbf{X}}^{(r-1)})) & & \\ \text{COV}(\mathbf{D}(\widehat{\mathbf{X}}^{(r-1)}), \Delta_1) & \text{VAR}(\Delta_1) & \\ \text{COV}(\mathbf{D}(\widehat{\mathbf{X}}^{(r-1)}), \Delta_2) & \text{COV}(\Delta_1, \Delta_2) & \text{VAR}(\Delta_2) \end{bmatrix}. \quad (4.47)$$

La función de ajuste para la estimación de los parámetros del modelo estructural, es

$$F_{ULS} = \frac{1}{2} \text{tr} [(\mathbf{S} - \Sigma(\boldsymbol{\theta}))^2], \quad (4.48)$$

Esta función de ajuste es adecuada en este contexto debido a que no se hacen supuestos distribucionales para los datos de disimilaridad (Finney & DiStefano, 2006). Cuando las estimaciones de los parámetros son conocidas, se estiman las disimilaridades simétricas con la ecuación (4.37). Este es el procedimiento de regresión de mínimos cuadrados ordinarios para estimar las puntuaciones de los factores que provee puntuaciones altamente correlacionadas con el factor (Grice, 2001). Ambos procedimientos están en concordancia con la minimización del STRESS que se hace también en un contexto de mínimos cuadrados ordinarios, una vez que las disparidades han sido estimadas.

Los parámetros son estimados en cada iteración del algoritmo. Para reflejar esta importante característica del procedimiento, la ecuación (4.48) puede escribirse como

$$F(\boldsymbol{\theta}^{(r)}) = \frac{1}{2} \text{tr} \left[\left(\mathbf{S}^{(r)} - \Sigma(\boldsymbol{\theta}^{(r)}) \right)^2 \right]. \quad (4.49)$$

Así, el objetivo de la fase SEM es minimizar esta función la cual provee estimadores consistente para los elementos de $\boldsymbol{\theta}$ (Bollen, 1989; Browne, 1982). El procedimiento continúa hasta minimizar el STRESS. El algoritmo propuesto de un modelo SEM para el escalamiento multidimensional de disimilaridades asimétricas (SEMDS), puede resumirse como sigue:

1. Se promedian Los valores Δ_1 and Δ_2 correspondientes a la matriz de disimilaridades asimétricas observadas para encontrar una configuración inicial $\mathbf{X}^{(0)}$ por MDS clásico. Luego, la matriz de disimilaridades simétrica óptima, $\widehat{\Delta}^{*(0)}$, y el coeficiente $\widehat{b}^{(0)}$ se estiman minimizando (4.49). Luego, se calculan las disparidades iniciales, $\widehat{\mathbf{D}}^{(0)}$, y se normalizan de forma que $\sum_{ij} \widehat{d}_{ij}^{(0)2} = n(n-1)/2$ y se calcula el valor inicial del STRESS según (4.4).
2. **Fase de configuración óptima:** Se estima la configuración $\widehat{\mathbf{X}}^{(r)}$ en la r -ésima iteración empleando la transformada de Guttman (ver por ejemplo Borg and Groenen, 2005) sobre $\widehat{\mathbf{D}}^{(r-1)}$, según la expresión

$$\widehat{\mathbf{X}}^{(r)} = \frac{1}{n} \mathbf{B}(\widehat{\mathbf{X}}^{(r-1)}) \widehat{\mathbf{X}}^{(r-1)}, \quad (4.50)$$

donde los elementos de la matriz \mathbf{B} están dados por, $b_{ii} = -\sum_{j \neq i} b_{ij}$, y

$$b_{ij} = \begin{cases} \frac{\hat{d}_{ij}^{(r-1)}}{d_{ij}(\widehat{\mathbf{X}}^{(r-1)})}, & \text{for } i \neq j, \text{ y } d_{ij}(\widehat{\mathbf{X}}^{(r-1)}) \neq 0, \\ 0, & \text{for } i \neq j, \text{ y } d_{ij}(\widehat{\mathbf{X}}^{(r-1)}) = 0. \end{cases}$$

3. **Fase SEM:** se obtienen las disimilaridades simétricas $\widehat{\Delta}^{(r)}$ y los valores de $\hat{b}^{(r)}$ minimizando (4.49), usando los valores Δ_1 , Δ_2 y las distancias obtenidas en la iteración previa $\mathbf{D}(\widehat{\mathbf{X}}^{(r-1)})$. Para minimizar (4.49) o equivalentemente,

$$\begin{aligned} F(\boldsymbol{\theta}) = & \frac{1}{2} (S_{11}^{(r)} - \theta_1^2 - \theta_5 - \kappa \sigma_{d(\hat{\mathbf{X}}^{(r-1)})}^2 S_{11}^{(r)})^2 + (S_{21}^{(r)} - \theta_1 \theta_2)^2 \\ & + \frac{1}{2} (S_{22}^{(r)} - \theta_2^2 - \theta_4)^2 + (S_{31}^{(r)} - \theta_1 \theta_3)^2 + (S_{32}^{(r)} - \theta_2 \theta_3)^2 \\ & + \frac{1}{2} (S_{33}^{(r)} - \theta_3^2 - \theta_4)^2. \end{aligned} \quad (4.51)$$

puede emplearse un algoritmo de región de confianza reflectivo para mínimos cuadrado no lineales (para más detalles ver Moré & Sorensen, 1983, o Coleman, Branch & Grace, 1999).

Así, las disparidades óptimas $\widehat{\mathbf{D}}^{(r)}$ se calculan y normalizan de tal forma que $\sum_{ij} \hat{d}_{ij}^{(r)2} = n(n-1)/2$, a fin de evitar soluciones triviales; luego, se calcula el STRESS de la r -ésima iteración.

4. Se calcula la diferencia entre dos valores de STRESS consecutivos y el algoritmo retorna a la fase MDS para repetir el procedimiento hasta que se alcance el criterio de convergencia.

El algoritmo se detiene cuando el criterio de convergencia haya sido logrado; es decir, si la diferencia entre dos valores consecutivos del STRESS sea menor a una constante positiva seleccionada o cuando se alcance el número máximo de iteraciones previamente establecido.

4.5. Aplicaciones

En este capítulo se analiza el comportamiento del modelo propuesto SEMDS para la representación de objetos en base a una matrix asimétrica de disimilaridades. El procedimiento podría ser útil, por ejemplo, cuando las disimilaridades son recolectadas en experimentos de confusión de estímulos. En tales situaciones, las asimetrías en las disimilaridades se asocian con aspectos vinculados a la atención, al entrenamiento, al contexto y al marco de referencia, como ha sido referenciado en el capítulo introductorio. Podría pensarse que en ausencia de estos factores que ejercen influencia sistemática sobre las disimilaridades observadas, éstas podrían ser simétricas. Así, hay argumentos sustantivos que apoyan el supuesto de que a la matriz de disimilaridades asimétricas observadas, subyace una matriz simétrica de la cual, aquella es una medida imperfecta.

En la situación descrita, el procedimiento típico es simetrizar la matriz observada promediando los elementos correspondientes a ambos lados de la diagonal y entonces buscar una representación con algún modelo MDS simétrico. Debido a que el procedimiento SEMDS, estima una matriz simétrica para la cual una configuración es determinada, pudiera ser interpretado como otro proceso de simetrización de la matriz de disimilaridades observada. Sin embargo, hay diferencias importantes en las dos metodologías. En el procedimiento SEMDS, la estimación de la matriz simétrica se hace en cada iteración y emplea en cada caso información obtenida en la iteración anterior. Así, el proceso de estimación de la matriz simétrica contribuye a disminuir la diferencia entre disimilaridades y distancias. Por otro lado, la contribución de cada parte triangular de la matriz asimétrica observada para la estimación de la matriz simétrica, no son seleccionados de manera arbitraria (por ejemplo $\lambda_1 = \lambda_2 = 1/2$). La contribución es estimada junto con otros parámetros del modelo, en un procedimiento que involucra minimizar una función de mínimos cuadrados que produce estimaciones consistentes de los parámetros.

Para probar el desempeño del modelo, se obtuvieron 8100 conjunto de datos artificiales generando verdaderas distancias de una configuración y agregando perturbaciones aleatorias para simular errores de medida según (4.15) y (4.16), siguiendo una metodología similar a la propuesta por Weeks and Bentler (1979). En primer lugar, se generaron n puntos aleatoriamente de una distribución uniforme en dos dimensiones, para obtener la matriz de configuración, para tamaños de $n = 15, 25, 50$. Luego, para cada configuración se calculó la distancias Euclidianas entre las filas y se organizaron en

forma de vector, para obtener muestras de tamaño 105, 300 y 1225, respectivamente. Los errores se seleccionaron de una distribución normal con media cero y varianza proporcional a la varianza de las distancias verdaderas, con factor de proporcionalidad p igual a .05, .10, .15, .25, .50, 1.00, 1.50, 2.00, h 3.00. Para los casos en los que esta metodología produjo valores negativos, se sumó una constante de manera que el valor más pequeño fuese 0. Los valores posibles para las cargas factoriales fueron (0.5,0.5),(0.3,0.7) y (1,2). el proceso se repitió 100 veces para cada combinación. Los 8100 conjuntos generados se analizaron con el procedimiento SEMDS propuesto y con SAMCOF considerando la matriz $\Delta = (\Delta_1 + \Delta_2)/2$, como las fuentes de las similitudes simétricas.

Para cada conjunto de datos simulados, se calculó la correlación entre las distancias perturbadas Δ_1 y Δ_2 y para la configuración dada, se calcularon los valores del STRESS normalizado y del estadístico Procrustes escalado, tanto para la configuración original como para la recuperada. Las correlaciones para cada par de cargas factoriales (λ_1, λ_2) y para cada nivel p de la varianza del error, se promediaron a través de las réplicas correspondiendo a todos los valores de n . Como se esperaba, para valores bajos de λ_1 and λ_2 se encontraron valores de correlación alto para valores de p de 0.05 a 0.10, pero para valores altos de las cargas factoriales la correlación aumentó para todos los valores de p , como se muestra en la Figura 4.1.

Las diferencias promedio entre los resultados obtenidos por el algoritmo SMACOF aplicado sobre la simetrización dada por $(\Delta_1 + \Delta_2)/2$, y el algoritmo SEMDS aplicado al conjunto de datos asimétricos, en términos del STRESS y el valor del índice Procrustes de bondad de ajuste, se resumen en la Figura 4.2 para cada término de la varianza del error y cada par de cargas factoriales. Valores negativos de esas diferencias reflejan el desempeño superior del modelo propuesto. En general, el procedimiento SEMDS recupera correctamente las cargas factoriales originales en todas las situaciones y, para conjunto de datos grandes produce los mejores resultados para todas las cargas factoriales y todos los niveles de la varianza del error, tanto en STRESS como en el índice Procustes entre las configuración original y la recuperada.

Para conjunto de datos más pequeños, los dos procedimientos producen resultados similares en el STRESS cuando las cargas factoriales son (0,5, 0,5) y para valores de p hasta 0.5 correspondiendo a un coeficiente de correlación promedio de 0.32. Para las cargas factoriales (0,3, 0,7), el procedimiento SEMDS logró mejores resultados con un valor de p de hasta 0.5 y $n = 15$

correspondiendo a un coeficiente de correlación promedio de 0.27, y de $p = 2$ para $n = 2$ correspondiendo a un coeficiente de correlación promedio de 0.87. Para niveles de error más altos y bajas correlaciones, SMACOF produjo mejores resultados en el STRESS para conjunto de datos pequeños. Sin embargo, en términos del Procrustes los resultados fueron bastante diferentes; en todas las situaciones, excepto para la combinación $n = 2, p = 2$ y cargas factoriales de (0,5, 0,5), el procedimiento SEMDS produjo la mejor configuración.

Las dos metodologías también se compararon con datos reales. Se utilizaron los datos clásicos de confusión perceptual de Miller & Nicely (1955). Los datos son confusiones entre 16 consonantes del idioma Inglés. En esta prueba, un grupo de mujeres escuchan a una mujer leer sílabas consonante-vocal formadas por el emparejamiento de las consonantes /b,d,g,p,t,k,v,ð,z,ʒ,f,θ,s,ʃ,m,n/ con la vocal /a/. A los sujetos se les pidió escribir la consonante que escucharon cuando cada sílaba fue pronunciada. Se compilaron matrices de confusión o errores de identificación bajo 17 diferentes condiciones experimentales. Las 4 primeras tablas de 16×16 dan un resumen de los datos obtenidos cuando

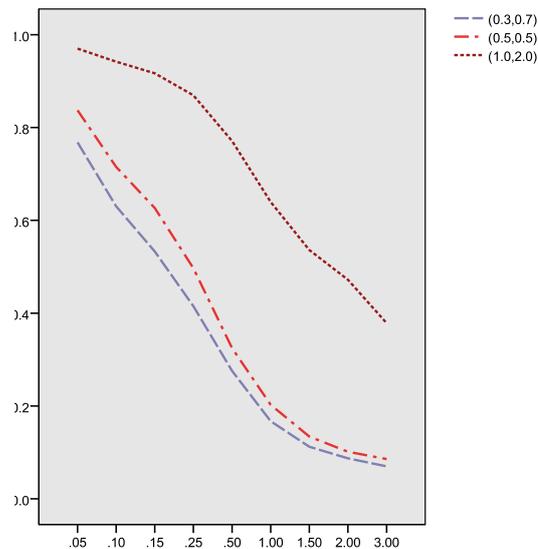


Figura 4.1: Valores de Correlación promedio entre los valores simulados de δ_1 y δ_2 a través de todas las réplicas y tamaños, para cada par de cargas factoriales y nivel de la varianza del error.

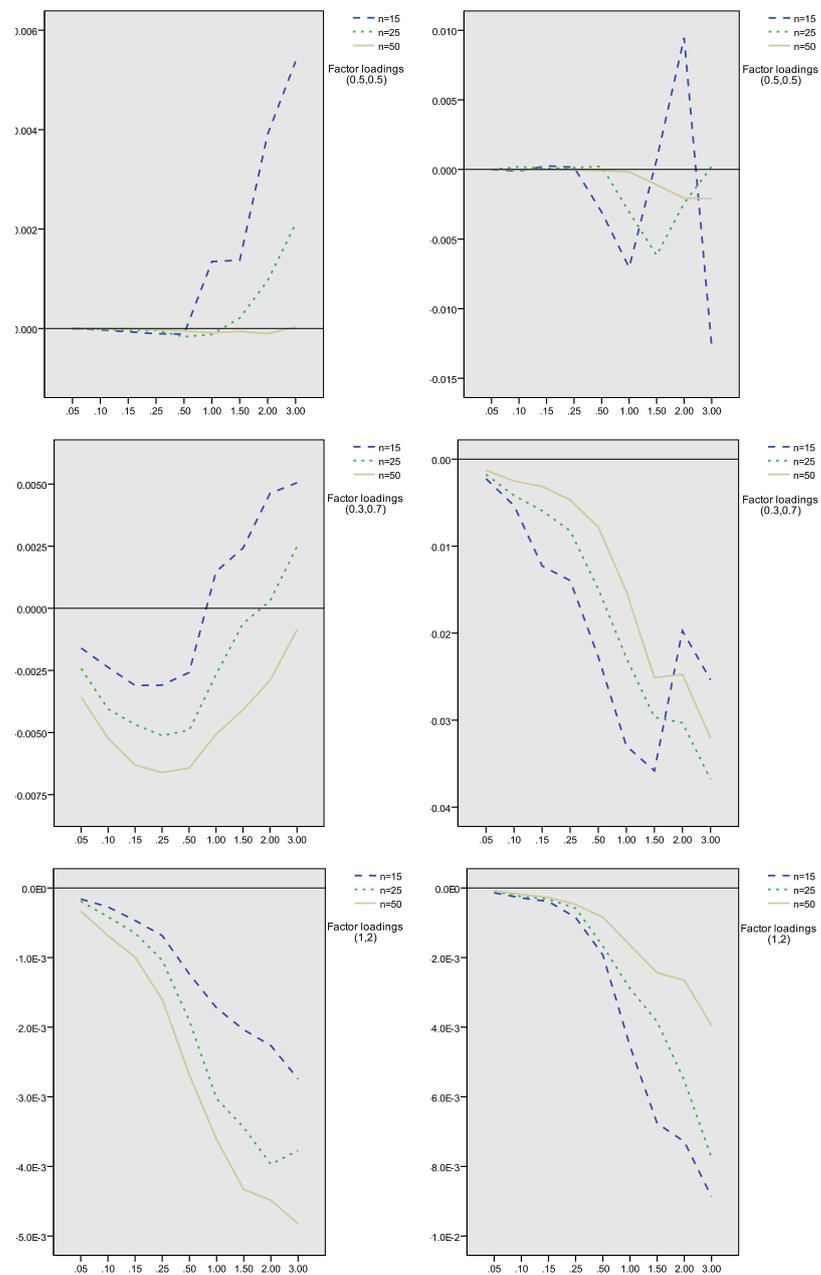


Figura 4.2: Diferencias promedio para el STRESS (panel de la izquierda) y para el índice de bondad de ajuste Procrustes (panel de la derecha) entre los resultados obtenidos por SMACOF y por el algoritmo propuesto para el conjunto de datos simulados, para cada término de la varianza del error y para cada par de cargas factoriales.

condiciones de ruido atenuado producen varios ratios discurso-a-ruido (D/R), con la adición de ruido aleatorio a diferentes niveles. En esta aplicación solo se analizan las primeras 4 tablas.

Las similitudes originales fueron transformadas en disimilitudes considerando el procedimiento de normalización descrito por Hubert (1972) para este conjunto de datos. Así, las probabilidades condicionales se calcularon dividiendo cada entrada de la matriz original por el total de la fila correspondiente. Luego se obtienen las disimilitudes asimétricas como $\delta_{ij} = 1 - s_{ij}/(\max(s_{ij}) + c)$, donde la constante c toma el valor de 0.001, para evitar entradas cero, mientras se mantiene el valor absoluto de las entradas (ver Arabie and Soli, 1979).

Cada parte triangular de la matriz de disimilitudes asimétricas, Δ_1 and Δ_2 , puede ser considerada una medida diferente de una relación simétrica subyacente entre los pares de señales que incluyen 120 entradas. La matriz de disimilitudes asimétrica fue analizada por la metodología SEMDS propuesta, y los resultados comparados con los obtenidos por el procedimiento SMACOF usual para MDS de razón, con las disimilitudes simetrizadas por $(\Delta_1 + \Delta_2)/2$.

SMACOF produjo un valor del STRES bruto de 9,0052 y un valor del intercepto $b = 1,6083$. Para el procedimiento SEMDS el valor del STRESS fue de 8,8448 asociado con valores de los parámetros de $\theta = (0.3632, 0.0970, 0.0614, 0.0158, 2.2463e-14)$, con una correlación de 0,16 entre las dos partes triangulares de la matriz de disimilitudes asimétricas. Así, la mejor configuración en términos del stress fue alcanzada por el procedimiento SEMDS. Además, ambos procedimientos mostraron diferencias en la configuración obtenidas siguiendo una transformación Procrustes, especialmente en las consonantes oclusivas y fricativas b y z, como puede verse en la Figura 4.3

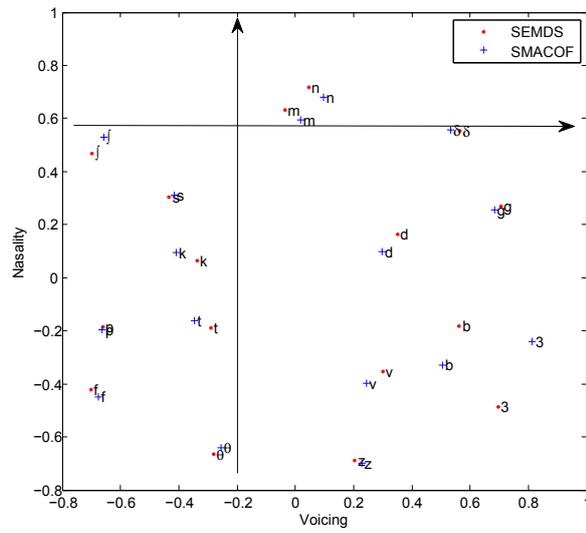


Figura 4.3: Representación de la configuración óptima por el procedimiento SEMDS (puntos), y la configuración Procrustes obtenida con SMACOF cuando las disimilaridades fueron primero simetrizadas (cruces).

Capítulo 5

Conclusiones

Los datos de proximidad asimétricos no son raros en muchas áreas de investigación. Esto se ve reflejado en la gran cantidad de enfoques disponibles para analizar proximidades no simétricas y la importante cantidad de investigación vinculada al tema. Sin embargo, la mayoría de esta investigación pareciera dirigida a situaciones donde la asimetría es intrínseca a la naturaleza de los datos, en el sentido que se expuso en el capítulo introductorio.

El modelo SEMDS propuesto, pudiera ser adecuado para analizar asimetrías en un contexto diferente. Es aquel que examina proximidades como juicios psicológicos de confusión de estímulos. Las asimetrías que tienen como fuente las medidas de confusión, tradicionalmente han sido modeladas en MDS como debidas a errores aleatorios. Esta tradición podría haber incidido para que la literatura sobre esta línea de trabajo sea relativamente escasa. Más allá de los trabajos de Zeilman y Heiser (1996), Shepard, (1964), Torgerson (1965), Sjöberg (1972) y Rosh, (1975), Poco trabajos aparecen en la literatura dirigidos a estudiar y modelar este tipo de asimetrías aunque, paradójicamente, estos investigadores han señalado que las confusiones aportan información acerca de las similitudes de los estímulos (Zeilman y Heiser) o que los niveles de atención se asocian con las medidas de confusión de estímulos (Shepard), y con el contexto y el marco de referencia (Torgerson y Sjöberg) y que en general están conectadas con las tareas cognitivas (Rosh).

El modelo SEMDS mostró, tanto en su aplicación a datos artificiales como a datos reales, un mejor desempeño global que el procedimiento típico de simetrizar la matriz de disimilaridades original. La simetrización asigna valores iguales a los aportes de cada parte triangular asimétrica en la determinación de la matriz simétrica que finalmente es representada, y esos valores son los

mismos para todos los casos, mientras que el procedimiento propuesto estima de manera óptima esas ponderaciones a través de un procedimiento que asegura estimadores consistentes para los parámetros. Parte de la diferencia en el desempeño de ambos enfoques descansa en esta importante característica del SEMDS.

El modelo propuesto alcanzó el mejor desempeño en el experimento simulado cuando los tamaños de muestra fueron los más grandes considerados ($N > 100$). Esto es congruente con los requerimientos de tamaños de muestras grandes que tienen los modelos de estructura de covarianza, aunque aquí no se hicieron estudios de ajuste global con pruebas de distribución asintótica.

Cuando los tamaños de muestra son pequeños los resultados indican un comportamiento similar de ambos procedimientos aunque los valores del STRESS y del índice Procrustes fueron levemente superiores para el SSEMDS.

Cuando la correlación entre las partes triangular inferior y triangular superior fueron altas, los resultados también fueron similares. Esto pudiera ser un resultado esperado por cuanto en esa situación, las partes triangulares tendrían un peso similar en la construcción de la matriz simétrica que es el supuesto que asume el método de simetrización, a todo evento.

Finalmente, un valor agregado del modelo SEMDS que contribuye por un lado a mejorar la precisión en la estimación de la matriz simétrica que mejor explique la variabilidad de los datos y por otro lado, a la velocidad de aproximación entre las disimilaridades transformadas y las distancias, es la inclusión de los errores de medida. Es una característica que pudiera contribuir a representar completamente las asimetrías como distancias entre puntos en el espacio MDS

Capítulo 6

Summary

For one-mode proximity data Multidimensional Scaling (MDS) usually states a monotone relation between dissimilarities and distances, according to which the symmetry constitutes an important subjacent hypothesis in the formulation of the model. Nevertheless there are practical situations for which the symmetry is not manifest when proximity data are compiled, and the MDS subjacent ideal relation between dissimilarities and distances is then weakened. Several models have been proposed to analyze data not symmetrical. These models fall into two categories. The first of these categories includes models that are based on the decomposition of the proximity matrix into two components, one symmetric and the other skew-symmetric. Some models in this category represents the symmetrical component by MDS, discussing separate specialized visualization techniques for the skew-symmetric component through the singular value decomposition of the skew-symmetric matrix. Others deal with both components simultaneously by applying the spectral decomposition of a Hermitian matrix. The second category includes models of generalized distances, or based on extensions of scalar product models, among others possibilities. In this research we adopt a different perspective from the structural equation model (SEM) framework, by which the asymmetry is taken into account as an outcome of measurement errors in MDS, so that asymmetric proximities are fully represented by distances between points in a MDS space. Thus, the upper and lower triangular parts of the observed asymmetric dissimilarity matrix are assumed to be imperfect measurements of an unobserved subjacent symmetric dissimilarity matrix (or latent matrix), to which each triangular matrix (or effect indicator) is linearly related. In this situation, the principal aim of the SEM stage is to

deal with the unknown symmetric dissimilarity matrix that best explains the observed variability in a covariance structure estimation problem. The effective incorporation of the error terms in the model is an important benefit derived from the covariance structure methodology for the parameter estimation, which helps improve the MDS estimated configuration.

A least squares alternating estimation procedure using SMACOF (de Leeuw and Heiser, 1980) is developed such that in each iteration the MDS configuration is attained, while the unknown symmetric dissimilarities are estimated in a covariance structural problem. When transformations are allowed in the MDS framework, and following the usual covariance structure estimation methodology, it can be assumed that the unknown distances arranged in vector form constitute a latent variable denoted by d^* , which can be measured through a coefficient λ_d using an indicator variable $d(\mathbf{X})$, whose values are obtained from a previous estimation of the configuration matrix \mathbf{X} .

In terms of a covariance structure model, the problem can be formulated by means of the system of structural equations given by,

$$d^* = b \delta^* + \zeta \quad (6.1)$$

$$d(\mathbf{X}) = \lambda_d d^* + e \quad (6.2)$$

$$\delta_1 = \lambda_1 \delta^* + \epsilon_1 \quad (6.3)$$

$$\delta_2 = \lambda_2 \delta^* + \epsilon_2. \quad (6.4)$$

Equation 4.13 models the structural spatial relation between the unknown distances \mathbf{D}^* (measured by $\mathbf{D}(\mathbf{X})$) and the unknown symmetric dissimilarities $\mathbf{\Delta}^*$, denoted as the latent variables d^* and δ^* respectively.

In the MDS framework this relation can be perceived as a disparity relation with an error term, where the distances are determined by an alternating estimation procedure. Coefficient b is a regression coefficient which describes the effects of the dissimilarities on the distances, and ζ is the error term referring to the spatial relationship of mean zero and variance σ_ζ^2 , which is uncorrelated with δ^* . Thus, $\widehat{\mathbf{D}} = \widehat{b} \widehat{\mathbf{\Delta}}^*$, where b is also estimated in the SEM phase, represents the optimal disparities in MDS. After the disparities are estimated, the distances can be estimated by considering them as predictors of the symmetric disparities in an alternating estimation procedure using the Guttman transform (de Leeuw and Heiser, 1980).

In the usual least squares MDS framework, two main alternate estimation phases are involved when disparities are taken into account. In the first, the parameters of the model are estimated assuming the disparities are known; secondly, the disparities are estimated when the remaining parameters are known. Thus, for the overall estimation procedure in this context, the disparities can be estimated in the covariance structure framework in an alternating estimation procedure that minimizes the STRESS using SMACOF (de Leeuw and Heiser, 1980); from the previous values of the parameter estimators, the configuration is first estimated by minimizing the STRESS from the estimated disparities, and then the values of the symmetric dissimilarity matrix $\widehat{\Delta}^*$ are estimated in a structural equation model of three observed variables, assuming the configuration \mathbf{X} and thus the distance matrix $\mathbf{D}(\mathbf{X})$ is known, such that the STRESS is also decreased.

The proposed Structural Equation Multidimensional Scaling (SEMDS) procedure can be summarized as follows:

1. Using the average of Δ_1 and Δ_2 , the initial configuration $\mathbf{X}^{(0)}$ is achieved by classical MDS. Then, the optimal symmetric dissimilarity matrix $\widehat{\Delta}^{*(0)}$ and coefficient $\hat{b}^{(0)}$ are calculated in the SEM framework by minimizing (4.49). The initial disparities $\widehat{\mathbf{D}}^{(0)}$ are then calculated and normalized such that $\sum_{ij} \hat{d}_{ij}^{(0)2} = n(n-1)/2$; the initial STRESS value is calculated using (4.4).
2. **Optimal configuration phase:** In the r -th iteration, the configuration $\widehat{\mathbf{X}}^{(r)}$ is estimated by the Guttman transform (see e.g. Borg and Groenen, 2005) from $\widehat{\mathbf{D}}^{(r-1)}$,

$$\widehat{\mathbf{X}}^{(r)} = \frac{1}{n} \mathbf{B}(\widehat{\mathbf{X}}^{(r-1)}) \widehat{\mathbf{X}}^{(r-1)}, \quad (6.5)$$

where the elements of matrix \mathbf{B} are given by, $b_{ii} = -\sum_{j \neq i} b_{ij}$, and

$$b_{ij} = \begin{cases} \frac{\hat{d}_{ij}^{(r-1)}}{d_{ij}(\widehat{\mathbf{X}}^{(r-1)})}, & \text{for } i \neq j, \text{ and } d_{ij}(\widehat{\mathbf{X}}^{(r-1)}) \neq 0, \\ 0, & \text{for } i \neq j, \text{ and } d_{ij}(\widehat{\mathbf{X}}^{(r-1)}) = 0. \end{cases}$$

3. **SEM phase:** The symmetric dissimilarities $\widehat{\Delta}^{*(r)}$ and the value of $\hat{b}^{(r)}$ are then obtained by minimizing (4.49), using the values of Δ_1 , Δ_2 and of the last given distances $D(\widehat{\mathbf{X}}^{(r-1)})$. A common trust-region-reflective algorithm for nonlinear least-squares problem can be employed (see e.g. Moré and Sorensen, 1983, or Coleman, Branch and Grace, 1999 for further details) to minimize (4.49), or equivalently,

$$\begin{aligned}
F(\boldsymbol{\theta}) = & \frac{1}{2}(S_{11}^{(r)} - \theta_1^2 - \theta_5 - \kappa\sigma_{d(\widehat{\mathbf{X}}^{(r-1)})}^2 S_{11}^{(r)})^2 + (S_{21}^{(r)} - \theta_1\theta_2)^2 \\
& + \frac{1}{2}(S_{22}^{(r)} - \theta_2^2 - \theta_4)^2 + (S_{31}^{(r)} - \theta_1\theta_3)^2 + (S_{32}^{(r)} - \theta_2\theta_3)^2 \quad (6.6) \\
& + \frac{1}{2}(S_{33}^{(r)} - \theta_3^2 - \theta_4)^2.
\end{aligned}$$

Then, the optimal disparities $\widehat{\mathbf{D}}^{(r)}$ are calculated and normalized such that $\sum_{ij} \hat{d}_{ij}^{(r)2} = n(n-1)/2$, to avoid trivial solutions; the STRESS in the r -th iteration is then calculated.

4. The difference between two consecutive STRESS values is obtained and the algorithm returns to the MDS phase until the convergence criterium is attained.

The algorithm stops if the overall convergence criterium is achieved. For artificial data, en general, the SEMDS procedure correctly recovers the original factor loadings in all situations, and for large data sets, it produces the best results for all factor loadings and error levels both in terms of STRESS and of the Procrustes goodness of fit criterium between the recovered and the original configurations.

Bibliografía

- Arbuckle, J. L. (1995). *Amos user's guide*. Chicago: Smallwaters.
- Anderson, T. W., & Rubin, H. (1956). Statistical inference in factor analysis. In J. Neyman (Ed.), *Proceedings of the third Berkeley symposium on mathematical statistics and probability, Vol. V* (pp. 111-150). Berkeley: University of California Press.
- Bekker, P. A., Merckens, A., & Wansbeek, T. J. (1994). Identification, equivalent models, and computer algebra. San Diego, CA: Academic.
- Bentler, P.M. (1980). Multivariate analysis with latent variables: Causal modeling. *Annual Review of Psychology, 31*, 419-456.
- Bentler, P.M. (1986). Structural modeling and Psychometrika: An historical perspective on growth and achievements. *Psychometrika, 51*(1), 35-51.
- Bentler, P. M. (1995). EQS structural equations program manual. Encino, CA: Multivariate Software
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Bollen, K.A. & Davis W.R. (2009). Two rules of identification
- Birnbaum, M. H. (1974). Using contextual effects to derive psychophysical scales. *Perception & Psychophysics, 15*, 89-96.
- Brito, C., & Pearl, J. (2002). A new identification condition for recursive models with correlated errors. *Structural Equation Modeling, 9*, 459-474.

Borg I., & Groenen, P.J.F. (2005). *Modern Multidimensional Scaling: Theory and Applications* (2nd ed.). New York, NY: Springer.

Bove, G. & Critchley, F. (1993). Metric multidimensional scaling for asymmetric proximities when the asymmetry is one-dimensional. In: Steyer, R., Wender, K. F., Widaman, K. F., eds. *Psychometric Methodology: Proceedings of the 7th (1991) European Meeting of the Psychometric Society*. Germany: Gustav Fischer Verlag.

Browne, M.W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37,62-83. item|| Browne, M. W., & Mels, G. (1994). *RAMONA PC user's guide*. Columbus, OH: Department of Psychology, Ohio State University.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (136-162). Newbury Park, CA: Sage.

Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Mahwah, NJ: Erlbaum.

Cailliez, F. (1983). The analytical solution of the additive constant problem. *Psychometrika*, 48, 305-308.

Carpita M., Brentari E., & Qannari E. M. (2015). *Advances in Latent Variables*. Springer: N.Y.

Carroll, J.D. and Chang, J.J (1970) Analysis of individual differences in multidimensional scaling via an n-way generalization of Eckart-Young decomposition. *Psychometrika*, 35, 283-319.

Chang C.L., & Lee, R.C.T. (1973). A heuristic relaxation method for nonlinear mapping in cluster analysis. *IEEE Transactions on System, Man and Cybernetics*,3, 197-200.

Chino, N. (1978) A graphical technique for representing asymmetric relationships between 11 objects. *Behaviormetrika*, 5, 23-40.

- Chino, N. (1990) A generalized inner product model for the analysis of asymmetry. *Behaviormetrika*, 27, 25-46.
- Chino, N., and Shiraiwa, K. (1993). Geometrical structures of some non-distance models for asymmetric MDS. *Behaviormetrika*, 20, 35-47.
- Cox, T. F., & Cox, M. A. A. (2000). *Multidimensional Scaling*. (2nd ed.). New York, NY: Chapman & Hall/CRC.
- Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Davis, W. R. (1993). The FC1 rule of identification for confirmatory factor analysis: A general sufficient condition. *Sociological Methods & Research*, 21, 403-437.
- De Leeuw J. 1977. Applications of convex analysis to multidimensional scaling. In J. R. Barra, F. Brodeau, G. Romier, B. van Cutsem (Eds.), *Recent developments in statistics*. Amsterdam, The Netherlands; North-Holland, 133-145.
- De Leeuw J. 1988. Convergence of the majorization method for multidimensional scaling. *Journal of Classification*, 5; 163-180.
- De Leeuw J, Heiser WJ. 1977. Convergence of correction-matrix algorithms for multidimensional scaling. In J.C. Lingoes, E.E Roskam, I. Borg (Eds.), *Geometric representations of relational data* . 735-752. Ann Arbor, MI: Mathesis Press.
- De Leeuw J, Heiser WJ. 1980. Multidimensional scaling with restrictions on the configuration. In *Multivariate analysis, Vol. V*, P.R. Krishnaiah (Ed.); Amsterdam: North-Holland; 501-522.
- Escoufier, Y. & Grolud, A. (1980). Analyses factorielle des matrices carrées non symmetriques, en *Data Analysis and Informatics*, Eds., E. Diday, L. Lebart, J.P. Pagés & R. Tomassone, Amsterdam: North Holland, 263-276.
- Fagot, R. & Mazo, R. (1989). Association coefficients of identity and proportionality for metric scales. *Psychometrika*, 54(1), 93-104.

Goodall, D. W. (1967). The distribution of simple matching coefficient. *Biometrics* 23, 647-656.

Gower, J.C. (1966) Some distance properties of latent root and vector methods in multivariate analysis. *Biometrika*, 53, 325-338.

Gower J. C. (1971). A General Coefficient of Similarity and Some of Its Properties. *Biometrics* 27(4), 857-871.

Gower, J.C. (1977) The analysis of asymmetry and orthogonality. In Barra, J.R. et al. (eds.), *Recent Developments in Statistics*, Amsterdam: North Holland.

Hair Jr. J. F., Black W., Babin B., & Anderson R.I. (2009). *Multivariate Data Analysis* (7/E). N.Y.: Prentice Hall.

Grice, J. W. & Harris, R. J. (1998). A comparison of regression and loading weights for the computation of factor scores. *Multivariate Behavioral Research*, 33(2), 221-247.

Glass, G. V., & Maguire, T. O. (1966). Abuses of factor scores. *American Educational Research Journal*, 3(4), 297-304.

Hancock, G.R. & Samuelson, K.M. (2008). *Advances in Latent Variable Mixture Models*. Charlotte: N.C.

Harshman, R.A. (1978) Models for analysis of asymmetrical relationships among N objects or stimuli. Paper presented at the First Joint Meeting of the Psychometric Society and the Society of Mathematical Psychology, Hamilton, Ontario.

Heiser, W. J. (1988) Multidimensional scaling with least absolute residuals. In Bock, H.H. *Classification and Related Methods of Data Analysis*, Amsterdam: North Holland, 455-462.

Holman, E. W. (1979). Monotonic models for asymmetric proximities. *J. Mathemat. Psychol.* 20: 1-15.

Howe, W. G. (1955). *Some contributions to factor analysis* (Report No. ORNL-1919). Oak Ridge National Laboratory, Oak Ridge, Tennessee.

- Hubalek Z. (1982). Coefficients of association and similarity, based on binary (presence-absence) data: An evaluation. *Biological Reviews of the Cambridge Philosophical Society* 57 (4), 669-689.
- Hu, L-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34, 183-202.
- Jöreskog, K. G. (1970). A general method for analysis of covariance structures. *Biometrika* 57: 239-251.
- Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 85-112). New York: Seminar.
- Jöreskog, K.G. (1977). Structural equation models in the social sciences: Specification, estimation and testing. In P.R. Krishnaiah (Ed.), *Application of Statistics* (pp. 265-287). Amsterdam: North-Holland.
- Jöreskog, K.G., and Sorbom, D. (1993a). *LISREL 8 User's Reference Guide*. Chicago: Scientific Software International.
- Jöreskog, K.G., and Sorbom, D. (1993b) . *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*. Chicago: Scientific Software.
- Jöreskog, K.G., & Van Thillo, M. (1973). LISREL-A general computer program for estimating a linear structural equation system involving multiple indicators of unmeasured variables. Research Report 73-5, Department of Statistics, Uppsala University, Sweden.
- Kaplan, D. (2000). *Structural equation modeling: Foundation and extensions*. Thousand Oaks, CA: Sage Publications.
- Kaplan, D. (1990). Evaluating and modifying covariance structure models: A review and recommendation. *Multivariate Behavioral Research*, 25(2), 137-155.

- Khatree, R., & Naik, D. N. (1999). *Applied multivariate statistics*. Cary, NC: SAS Institute.
- Keesling, J. W. (1972). *Maximum likelihood approaches to causal flow analysis*. Unpublished doctoral dissertation. Chicago: University of Chicago.
- Kiers, H.A.L. (1989) An alternating least squares algorithm for fitting the two- and three-way DEDICOM model and the IDIOSCAL model. *Psychometrika*, 54, 515-521.
- Kiers, H.A.L., ten Berge, J.M.F., Takane, Y. and de Leeuw, J. (1990). A generalization of Takane's algorithm for DEDICOM. *Psychometrika*, 55, 151-158.
- Kiers, H.A.L. and Takane, Y. (1994) A generalization of GIPSCAL for the analysis of non-symmetric data. *J. Classification*, 11, 79-99.
- Klein, R.W. and Dubes, R.C. (1989) Experiments in projection and clustering by simulated annealing. *Pattern Recognition*, 22, 213-220.
- Krumhansl, C. L. (1978). Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Psychol. Rev.* 85: 445-463.
- Kruskal JB. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29;1-27.
- Kruskal JB. (1964b). Nonmetric multidimensional scaling: A numerical method. *Psychometrika* 29; 28-42.
- Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling*. Beverly Hills, CA: Sage.
- Lawley, D. N. (1958). Estimation in factor analysis under various initial assumptions. *British Journal of Statistical Psychology*, 11, 1-12.
- Marcoulides, G. A., & Moustaki, I. (Eds.) (2002). *Latent Variable and Latent Structure Models*. Mahwah, NJ: Lawrence Erlbaum.

- Marcoulides, G. A., & Schumacker, R. E. (2001). *New developments and techniques in structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum.
- Micko, H. C., & Fischer, W. (1970). The metric of multidimensional psychological spaces as a function of the differential attention to subjective attributes. *Journal of Mathematical Psychology*, 7, 118-143.
- Montfort, K, Oud, J.H.L. & Satorra, A. (Eds.) (2010). *Longitudinal Research with Latent Variables*. Springer Verlag, New York.
- Muthén, B.O., & Muthén, L.K. (2004). Mplus user's guide. Muthén & Muthén.
- Muthén, B., & Muthén, L. (1998). *MPLUS user's guide*. Los Angeles: Muthén & Muthén.
- Niemann, H. & Weiss, J.(1979). A fast-converging algorithm for nonlinear mapping of high-dimensional data to a plane. *IEEE Trans. Comput.*, 28, 142-147.
- O'S'Brien, R. M. (1994). Identification of simple measurement models with multiple latent variables and correlated errors. *Sociological Methodology*, 24, 137-170.
- Parducci, A. (1973). Contextual effects: A range-frequency analysis. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. 2). New York: Academic Press.
- Parducci, A., & Marshall, L. M. (1961). Context effects in judgments of length. *American Journal of Psychology*, 74, 576-583.
- Parducci, A., & Perrett, L. F. (1971). Category rating scales: Effects of relative spacing and frequency of stimulus values. *Journal of Experimental Psychology*, 89, 427-452.
- Ramsay, J.O. (1977). Maximum likelihood estimation in multidimensional scaling. *Psychometrika*, 42, 241-266.
- Raycov, T. & Marcoulides G.A. (2010). *Introduction to Psychometric Theory*. N.Y.: Routledge. 347 pages.

- Raykov, T., & Widaman, K. F. (1995). Issues in structural equation modeling research. *Structural Equation Modeling*, 2, 289-318.
- Raykov T. & Marcoulides G.A. (2006). *A First Course in Structural Equation Modeling*. London:LEA.
- Reilly, T., & O'S'Brien, R. M. (1996). Identification of confirmatory factor analysis models of arbitrary complexity: The side-by-side rule. *Sociological Methods & Research*, 24, 473-491.
- Rigdon, E. E. (1995). A necessary and sufficient identification rule for structural models estimated in practice. *Multivariate Behavioral Research*, 30, 359-383.
- Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104, 192-233.
- Satorra, A. (2000). Scaled and adjusted restricted tests in multisample analysis of moment structures. In D. D. H. Heijmans, D. S. G. Pollock, & A. Satorra (Eds.), *Innovations in multivariate statistical analysis: A Festschrift for Heinz Neudecker* (pp. 233-247). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Satorra, A., & Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. *ASA 1988 Proceedings of the Business and Economic Statistics, Section* (308-313). Alexandria, VA: American Statistical Association.
- Satorra, A., & Bentler, P.M. (1986). Some robustness properties of goodness of fit statistics in covariance structure analysis. *Proceedings of the Business & Economic Statistics Section*, American Statistical Association, 549-554.
- Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399-419). Thousand Oaks, CA:Sage.
- Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66, 507-514.

- Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled chi-square test statistic. *Psychometrika*, *75*, 243-248.
- Saito, T. (1991). Analysis of asymmetric proximity matrix by a model of distance and additive terms. *Behaviormetrika* *29*: 45-60.
- Saito, T. (1993a). Multidimensional scaling for asymmetric proximity data. In: Steyer, R., Wender, K. F., Widaman, K. F., eds. *Psychometric Methodology: Proceedings of the 7th European Meeting of the Psychometric Society*, Germany: Gustav Fischer Verlag, pp. 452-456.
- Saito, T. (1993b). Multivariate analysis of environmental data with asymmetrical relationship among locations. *Hokkaido Behavioral Science Report M-22*: 1-12.
- Saito, T., Takeda, S. (1990). Multidimensional scaling for asymmetric proximities: Model and method. *Behaviormetrika* *28*: 49-80.
- Saito T., & Yadohisa T. (2005). *Data Analysis of Asymmetric Structures: Advanced Approaches in Computational Statistics*. New York, NY: Dekker.
- Sammon J.W. (1969). A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers*, *18*, 401-402.
- SAS Institute. (1989). *SAS PROC CALIS user's guide*. Cary, NC: Author.
- Shepard, R.N. (1962a). The analysis of proximities: multidimensional scaling with an unknown distance function I. *Psychometrika*, *27*, 125-140.
- Shepard, R.N. (1962b) The analysis of proximities: multidimensional scaling with an unknown distance function II. *Psychometrika*, *27*, 219-246.
- Shepard, R. N. (1963). Analysis of proximities as a technique for the study of information processing in man. *Human Factors*, *5*, 33-48.
- Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, *1*, 54-87.

- Schönemann, P. H. (1972). An algebraic solution for a class of subjective metrics models. *Psychometrika* 37: 441-451.
- Shumacker, R. E. & Lomax, R. G. (2010). *A beginner's guide to Structural Equation Modeling*. New York: Taylor & Francis Group.
- Sjöberg, L. (1972). A cognitive theory of similarity. Göteborg: Psychological Reports, 2(10).
- Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multinivel, Longitudinal, and Structural Equation Models*. Chapman & Hall/CRC: N.Y.
- Snijders T., Dormaar M., van Schuur W., Dijkman-Caes C., and Driesen G. (1990). Some Similarity Coefficients for Dyadic Binary Data in the Case of Associated Attributes. *Journal of Classification* 7, 5-31.
- Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, 15, 72-101.
- Spearman, C. (1927). *The abilities of man*. New York: Macmillan.
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics*. Boston, MA: Allyn and Bacon.
- Thomson GH. (1935) The definition and measurement of g (general intelligence) *Journal of Educational Psychology*, 26:241-262.
- Torgerson, W.S. (1952). Multidimensional scaling: Theory and method. *Psychometrika*, 17, 401-419.
- Torgerson, W.S. (1958) *Theory and Method of Scaling*, New York: Wiley.
- Torgerson, W. S. (1965). Multidimensional scaling of similarity. *Psychometrika*, 30, 333-367.
- Trope Y., Liberman N., & Waslak C. (2007). Construal Levels and Psychological Distance: Effects on Representation, Prediction, Evaluation, and Behavior. *Journal of Consumer Psychology*, 17(2): 83-95.

- Tversky, A. (1975). Features of similarity. *Psychological Review*, 89, 1233-154.
- Vera, J. F., Macías, R., & Heiser, W. J. (2009a). A dual latent class unfolding model for two-way two-mode preference rating data. *Computational Statistics and Data Analysis*, 53, 3231-3244.
- Vera, J. F., Macías, R., & Heiser, W. J. (2009b). A latent class multidimensional scaling model for two-way one-mode continuous rating dissimilarity data. *Psychometrika*, 74, 297-315.
- Weeks, D. G., & Bentler, P. M. (1982). Restricted multidimensional scaling models for asymmetric proximities. *Psychometrika* 47: 201-208.
- Wiley, D. E. (1973). The identification problem for structural equation models with unmeasured variables. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 69-83). New York: Seminar.
- Wish, M. (1967). A model for the perception of Morse code like signals. *Human Factors*, 9, 529-539.
- Wish, M. (1971). Individual differences in perceptions and preferences among nations. In C. W. King & D. Tigert (Eds.) *Attitud research reaches new heights*. Chicago: American Marketing Association.
- Wright, S. (1918). On the nature of size factors. *Genetics*, 3, 367-374.
- Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, 20, 557-585.
- Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics*, 5, 161-215.
- Wright, S. (1960). Path coefficients and path regression: Alternative or complementary concepts? *Biometrics*, 16, 189-202.
- Young, F. W. (1975). An asymmetric Euclidian model for multi-process asymmetric data. *Paper presented at U.S.-Japan Seminar on MDS*, San Diego, U.S.A., 79-88.

Young, G., & Householder, A. S. (1938). Discussion of a set of points in terms of their mutual distances. *Psychometrika* 3: 19-22.

Zegers, F.E. (1986). A Family of Chance-corrected Association Coefficients for Metric Scales. *Psychometrika*, 51, 559-562.

Zegers, F.E., & Ten Berge, J.M.F. (1985). A Family of Association Coefficients for Metric Scales. *Psychometrika*, 50, 17-24.

Zielman, B., & Heiser, W. J. (1993). Analysis of asymmetry by a slide-vector. *Psychometrika* 58: 101-114.

Zielman, B., & Heiser, W. J. (1996). Models for asymmetric proximities. *British Journal of Mathematical and Statistical Psychology*, 49, 127-146.