



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

FACULTAD DE CIENCIAS

“Escalamiento Multidimensional No Métrico
en Comunidades Ecológicas ”

T E S I S

QUE PARA OBTENER EL TÍTULO DE :

ACTUARIA

P R E S E N T A :

NARANJO ALBARRÁN LIZBETH

TUTORA: MAT. MARGARITA ELVIRA CHÁVEZ CANO

MEXICO, D.F.

2006



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

Hoja de Datos del Jurado

Alumna:

Naranjo Albarrán Lizbeth

tel. 56 42 12 26

Universidad Nacional Autónoma de México

Facultad de Ciencias

Actuaría

Tutora:

Matemática

Margarita Elvira Chávez Cano

Sinodal 1:

Maestro en Ciencias

Jesús Agustín Cano Garcés

Sinodal 2:

Maestro en Ciencias

Gerardo Rivas Lechuga

Sinodal 3:

Actuario

Jaime Vázquez Alamilla

Sinodal 4:

Doctor

Eduardo Arturo Gutiérrez Peña

Tesis

“Escalamiento Multidimensional No Métrico en Comunidades Ecológicas”

157 páginas

2006

Agradecimientos

A Dios por iluminar mi camino y por estar siempre a mi lado.

A mi papá y mi mamá por el gran Amor que les tengo. Gracias por estar a mi lado en todo momento, por todo lo que me han enseñado y por que sin ustedes esto no hubiera sido posible.

A mis hermanos Israel y Karina por su Amor y Cuidados.

A mi prima Rocío por los momentos que hemos pasado juntas.

A mis abuelos, tios y primos y a mi sobrino por ser parte importante en mi vida.

A mis amigas Abigail, Alma y Cristina por su amistad y apoyo incondicional.

A mis amigos Armando, Claudia y Cynthia por estar siempre conmigo.

A mis amigas de la Facultad por que sin ustedes la Universidad no hubiera sido la misma.

A mi madrina Erika por su cariño, comprensión y sabiduría.

A la Mat. Margarita Chávez Cano con respeto, admiración y agradecimiento por su apoyo y dirección en este trabajo.

A la UNAM por ser la mejor Universidad.

Índice general

Introducción	ix
1. Escalamiento Multidimensional Métrico	1
1.1. Escalamiento Clásico	2
1.1.1. Disimilaridades como Distancias Euclidianas	6
1.1.2. Similaridades	12
1.1.3. Propiedades Óptimas de la Solución Clásica y Bondad de Ajuste	14
1.1.4. Un Algoritmo Práctico para el Escalamiento Clásico . .	19
1.2. Dimensionalidad	20
1.3. Rotación	22
1.4. Interpretación	24
1.4.1. Método Subjetivo	25
1.4.2. Ajuste de un Vector Característico	25
1.4.3. Análisis de Correlación Canónica	27
1.4.4. Análisis de Conglomerados	28
1.4.5. Análisis con Datos Externos	28
1.5. Ejemplo	29
2. Escalamiento Multidimensional No Métrico	43
2.1. Espacios \mathcal{R}^p y la Métrica de Minkowski	44
2.2. Medidas de Bondad de Ajuste	46
2.2.1. <i>Ties</i> (empates)	47
2.2.2. Stress	47
2.2.3. S-Stress	49
2.2.4. Coeficiente de Alienación	49
2.2.5. Rsq	50
2.3. Algoritmo	50

2.3.1.	Configuración Inicial	51
2.3.2.	Coordenadas y Distancias Estandarizadas	52
2.3.3.	Fase No Métrica	54
2.3.4.	Fase Métrica	57
2.3.5.	Fase No Métrica Alternativa	58
2.3.6.	Observaciones	59
2.4.	Aproximación de Kruskal	59
2.4.1.	Minimizando S con Respecto a las Disparidades	60
2.4.2.	Una Configuración con Mínimo Stress	63
2.4.3.	Técnica Iterativa de Kruskal	64
2.5.	Aproximación de Guttman	66
2.6.	Dimensionalidad, Rotación e Interpretación	68
2.6.1.	Dimensionalidad	68
2.6.2.	Rotación	69
2.6.3.	Interpretación	70
2.7.	Diagramas de Shepard	71
2.8.	Problemas de Cálculo	71
2.8.1.	Mínimo Local	73
2.8.2.	Solución Degenerada	74
2.8.3.	Convergencia	75
2.9.	Ejemplo	76
3.	Datos y Proximidades	91
3.1.	Tipos de Datos	91
3.2.	Formas de Datos	94
3.3.	Estandarización y Transformación de Datos	95
3.4.	Tipos de Variables y Escalas de Medición	97
3.4.1.	Tipos de Variables	97
3.4.2.	Escalas de Medición	98
3.5.	Construcción de Proximidades	98
3.5.1.	Similaridades	99
3.5.2.	Disimilaridades	99
3.5.3.	Distancias	99
3.5.4.	Medidas de Proximidad Directas	100
3.5.5.	Medidas de Proximidad Derivadas	102
3.5.6.	Transformando de Similaridades a Disimilaridades	107
3.6.	Propiedades de las Medidas de Asociación	108

4. Aplicación	111
4.1. Comunidades y Estadística	111
4.2. Área Estudiada	113
4.3. Metodología	115
4.4. Resultados	116
4.4.1. Análisis de los Cuadros	117
4.4.2. Análisis de las Especies	140
Conclusiones	153

Introducción

Introducción Histórica

La técnica que hoy en día se conoce como escalamiento multidimensional tiene sus inicios en los primeros modelos desarrollados por la Psicología a finales del siglo XIX. Tales modelos, que hoy forman parte de una rama conocida como Psicofísica, pretendían estudiar la relación entre la intensidad física de los estímulos y la sensación subjetiva que éstos logran obtener en los sujetos.

Todos los modelos de escalamiento suponían que el continuo psicológico sobre el que se situaba al estímulo objeto de estudio era unidimensional. Esto es, que los juicios subjetivos acerca del estímulo se basan en un único atributo o dimensión del mismo. La extensión de los modelos de escalamiento al caso multidimensional gracias al trabajo de Torgerson (1958) permitió al escalamiento (ahora multidimensional) salirse del campo de la Psicología y convertirse en una técnica de análisis de datos de aplicación universal, a pesar de que los subsiguientes desarrollos teóricos y matemáticos (publicados casi todos ellos en la revista *Psychometrika*) se mantuvieron dentro del campo de la Psicología.

En la actualidad, el escalamiento multidimensional (MDS, por su nombre en inglés *Multidimensional Scaling*) puede considerarse como una familia muy amplia de procedimientos de análisis multivariado, aptos para tratar gran cantidad de tipos diferentes de datos de entrada (tablas de contingencia, matrices de proximidad, correlaciones, datos de perfil, etc.). Todos los programas de análisis estadístico de propósito general disponibles hoy en día (como SPSS, SYSTAT y STATISTICA) han incorporado algún procedimiento de MDS a su repertorio de técnicas multivariantes.

No obstante lo anterior, el MDS continúa siendo subutilizado, especialmente en la investigación desarrollada en los países de habla hispana.

Escalamiento Multidimensional

El escalamiento multidimensional es una técnica de análisis multivariado que permite representar las proximidades entre un conjunto de objetos o estímulos como distancias en un espacio de baja dimensionalidad (generalmente 2 o 3 dimensiones). Podemos establecer una analogía muy sencilla entre lo que hace el MDS y el uso de un mapa. Si tomamos un mapa de México y medimos las distancias entre diez capitales (Mexicali, La Paz, Tuxtla Gutiérrez, Chihuahua, Chilpancingo, Guadalajara, Monterrey, Jalapa, Mérida y Zacatecas) obtendremos una matriz de distancias como la que se muestra en la tabla 1.

Cada celda de la matriz indica la distancia entre dos ciudades. La diagonal de la matriz contiene ceros, dado que en la diagonal se encontrarían las distancias de cada ciudad consigo misma. Asimismo, sólo se muestran las distancias correspondientes a la mitad triangular inferior, esto se debe a que sabemos que la distancia que hay, por ejemplo, de Jalapa a Mérida es la misma que de Mérida a Jalapa, por lo que nos basta indicar una de ellas. La utilidad de esta matriz de distancias es que nos permite conocer directamente la distancia que separa una ciudad de otra sin necesidad de calcularla a partir del mapa.

Podríamos decir que el MDS lleva a cabo el procedimiento inverso. Es decir, partiendo de una matriz de proximidades (en nuestro ejemplo, distancias en kilómetros), nos proporciona un mapa de las capitales. Si utilizásemos la matriz de distancias anterior como entrada para un análisis MDS, obtendríamos la solución que se muestra en la figura 1.

La solución nos muestra las posiciones relativas de las diez capitales en un espacio de dos dimensiones, representadas por los ejes horizontal y vertical que aparecen en la figura 1. Para facilitar la visualización de la configuración se muestra un mapa de México en la figura 2. Las propiedades geométricas del mapa permanecen intactas, puesto que las capitales se encuentran donde deberían estar. Este fenómeno constituye una de las peculiaridades del MDS.

Se ha dicho que el MDS permite representar proximidades entre objetos como distancias entre puntos en un espacio de varias dimensiones. Existe una serie de transformaciones del espacio que podemos realizar sin alterar las distancias entre los objetos, por lo que estas transformaciones son aplicables también a cualquier solución obtenida a partir del MDS.

Tabla 1. Matriz de distancias en kilómetros entre diez capitales mexicanas.

			1	2	3	4	5	6	7	8	9	10
1	Baja California	Mexicali	0									
2	Baja California Sur	La Paz	1075	0								
3	Chiapas	Tuxtla Gutiérrez	2831	1938	0							
4	Chihuahua	Chihuahua	1013	663	1844	0						
5	Guerrero	Chilpancingo	2288	1306	694	1381	0					
6	Jalisco	Guadalajara	1800	819	1125	925	500	0				
7	Nuevo León	Monterrey	1631	975	1250	613	906	613	0			
8	Veracruz	Jalapa	2325	1438	506	1344	356	638	769	0		
9	Yucatán	Mérida	2838	2119	600	1825	1113	1388	1219	781	0	
10	Zacatecas	Zacatecas	1681	813	1163	738	644	250	313	656	1313	0

Las transformaciones admisibles son las siguientes:

1. *Rotación*. Podemos girar un conjunto de puntos simultáneamente sin que la configuración formada por todos ellos sufra ningún cambio.
2. *Traslación*. Podemos trasladar un conjunto de puntos a otro lugar en el espacio sin alterar la configuración.
3. *Escala*. Podemos aumentar o disminuir la escala de medida de las distancias entre los objetos sin alterar la configuración de los puntos.
4. *Reflexión*. La imagen reflejada de una configuración de puntos tampoco altera las posiciones relativas de los mismos.

Dado que todas estas transformaciones son admisibles, también lo es utilizar varias de ellas simultáneamente. En el MDS es posible cambiar el sistema de coordenadas de referencia si ello facilita o simplifica la interpretación de los resultados, o en dado caso suprimir los ejes de coordenadas si éstos no son necesarios.

El MDS puede parecer una técnica ingeniosa, pero de dudosa aplicabilidad. Sin embargo su utilidad es enorme:

- Los psicólogos, quienes fueron el instrumento para desarrollar la técnica, utilizan el MDS para entender las percepciones y las evaluaciones de estímulos auditivos (tales como el idioma y tonos musicales), estímulos visuales (tales como colores y muecas), y entidades sociales (tales como factores de personalidad y situaciones sociales), entre otras.
- En política se usa frecuentemente el MDS para entender porqué los candidatos políticos son percibidos como similares o disimilares y encontrar las diferencias entre ellos, y así poder desarrollar indicadores que puedan medirse en futuras elecciones.
- Los sociólogos utilizan el MDS para determinar la estructura de grupos y organizaciones, basados en las percepciones interpersonales y sus comportamientos.
- En antropología se utiliza el MDS para estudiar las diferencias culturales de varios grupos, basados en creencias, lenguaje e información sobre artefactos.

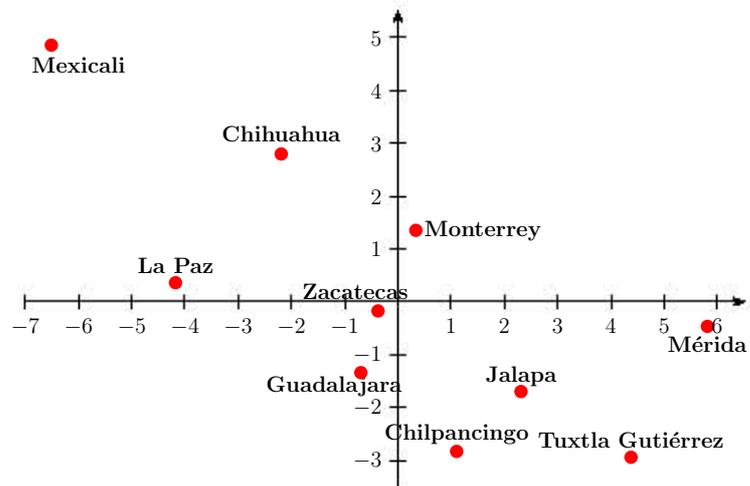


Figura 1: MDS de 10 capitales de México



Figura 2: Mapa de México

- Los economistas y mercadotecnistas utilizan el MDS para investigar la manera en que los clientes evalúan los productos y para entender la relación entre las características de los mismos.
- Los investigadores educacionales usan el MDS para estudiar la estructura de la inteligencia y del comportamiento en el salón.
- Los urbanistas han identificado las similitudes de varias ciudades, pueblos o regiones en términos de su posición en un espacio de dimensión reducida derivado de datos demográficos, fiscales y económicos.
- Los ecólogos usan el MDS para estudiar la estructura de las diferentes comunidades de especies.

El objetivo de este trabajo es mostrar la teoría del Escalamiento Multidimensional Métrico y del No Métrico, y aplicar éste último como una herramienta, en un caso particular, en el análisis de la estructura de las comunidades ecológicas.

El texto consta de cuatro capítulos. El primero de ellos construye y desarrolla el escalamiento multidimensional métrico. El capítulo 2 explica el escalamiento multidimensional no métrico y los elementos utilizados para un mejor análisis. El capítulo 3 explica el tipo de datos y proximidades utilizadas en ecología para el análisis de escalamiento multidimensional. El capítulo 4 muestra un ejemplo de la aplicación del escalamiento multidimensional no métrico en comunidades ecológicas, particularmente, en un estudio sobre la abundancia de las especies de plantas en un territorio específico.

Capítulo 1

Escalamiento Multidimensional Métrico

El escalamiento multidimensional es una técnica de análisis multivariado que permite representar las proximidades entre un conjunto de objetos o estímulos como distancias en un espacio de baja dimensionalidad (generalmente 2 ó 3 dimensiones).

Supóngase que hay n objetos con disimilaridades $\{\delta_{rs}\}$. El escalamiento multidimensional métrico¹ tiene como objetivo encontrar una configuración de un conjunto de puntos en un espacio donde cada punto representa uno de los objetos, y las distancias entre los puntos $\{d_{rs}\}$ son tales que:

$$d_{rs} \approx f(\delta_{rs}),$$

donde f es una función paramétrica monótona continua.

El modelo de MDS métrico supone que los datos están medidos en una escala de intervalo o de razón. De aquí que existe una forma funcional exacta (lineal, cuadrática, etc.) que relaciona las medidas de proximidad con las distancias. El análisis de MDS métrico supone que la relación entre las medidas de proximidad y las distancias son de tipo lineal: $d_{rs} = a + b\delta_{rs}$. Una relación de este tipo utiliza las propiedades métricas de las proximidades, es por ello que se le denomina escalamiento multidimensional métrico.

El primer procedimiento de MDS métrico se debe (cf. Kruskal y Wish, 1978) a Richarson (1938) quien se basó en un teorema de Young y Hausholder (1938), según el cual a partir de una matriz de distancias entre n

¹A partir de aquí MDS métrico. Su nombre en inglés es *Metric Multidimensional Scaling* o *Parametric Multidimensional Scaling*.

objetos se puede derivar una matriz de productos escalares entre vectores. A su vez, es posible descomponer la matriz de productos escalares en el producto XX^t , donde $X_{n \times p}$ es la matriz de coordenadas de los n objetos en p dimensiones de tal manera que las distancias se preservan. Torgerson (1952, 1958), apoyado por Ledyard y Tucker y trabajando sobre las investigaciones de Young y Householder, propuso uno de los primeros algoritmos del escalamiento multidimensional. Gower (1966, 1982) discutió y extendió los resultados de Torgerson (1952).

Torgerson (1958) refinó el procedimiento llevando a cabo una transformación de la matriz de proximidades $\Delta_{n \times n}$ en una matriz de distancias $D_{n \times n}$ tal que respetase las propiedades de la función de distancia euclidiana. El cumplimiento de estos axiomas, que se muestran en la tabla 2, es de gran importancia para la aplicación del MDS.

Tabla 2. Las tres propiedades de la distancia euclidiana.

1. No negatividad	$d_{rs} \geq d_{rr} = 0$
2. Simetría	$d_{rs} = d_{sr}$
3. Desigualdad del triángulo	$d_{rs} \leq d_{rt} + d_{ts}$

De las tres propiedades, las dos primeras (no-negatividad y simetría) son fáciles de cumplir, basta con utilizar como entrada una matriz de proximidades simétrica. El problema planteado para garantizar el cumplimiento de la tercer propiedad (desigualdad del triángulo), es conocido como el problema de la “estimación de la constante aditiva”. Se trata de buscar un valor c que, sumado a las proximidades originales (δ_{rs}) nos proporcione las distancias ($d_{rs} = \delta_{rs} + c$). Torgerson determinó que el valor mínimo c que satisface la desigualdad del triángulo ($d_{rs} \leq d_{rt} + d_{ts}$) para toda terna de objetos (r, s, t) se define como:

$$c_{\min} = \max_{(r,s,t)} (\delta_{rs} - \delta_{rt} - \delta_{ts})$$

1.1. Escalamiento Clásico

Sean x_r ($r = 1, \dots, n$) las coordenadas de n puntos en un espacio euclidiano p dimensional, donde $x_r = (x_{r1}, \dots, x_{rp})^t$. Entonces el cuadrado de la distancia euclidiana entre el r -ésimo y el s -ésimo puntos está dada por

$$d_{rs}^2 = \sum_{i=1}^p (x_{ri} - x_{si})^2 = (x_r - x_s)^t (x_r - x_s)$$

Sea $B_{n \times n}$ la matriz de producto interno, donde

$$[B]_{rs} = b_{rs} = x_r^t x_s$$

La matriz de producto interno B se encuentra a partir de las distancias conocidas $\{d_{rs}\}$.

Matriz de producto interno centrado B

Para evitar soluciones indeterminadas debido a traslaciones, el centroide de la configuración de puntos es puesto en el origen, de aquí

$$\sum_{r=1}^n x_{ri} = 0 \quad (i = 1, \dots, p)$$

Para encontrar B , sabemos que

$$\begin{aligned} d_{rs}^2 &= (x_r - x_s)^t (x_r - x_s) \\ &= x_r^t x_r - x_r^t x_s - x_s^t x_r + x_s^t x_s \\ &= x_r^t x_r - 2x_r^t x_s + x_s^t x_s \end{aligned} \tag{1.1}$$

y de aquí

$$\begin{aligned} \frac{1}{n} \sum_{r=1}^n d_{rs}^2 &= \frac{1}{n} \sum_{r=1}^n x_r^t x_r + x_s^t x_s \\ \frac{1}{n} \sum_{s=1}^n d_{rs}^2 &= x_r^t x_r + \frac{1}{n} \sum_{s=1}^n x_s^t x_s \\ \frac{1}{n^2} \sum_{r=1}^n \sum_{s=1}^n d_{rs}^2 &= \frac{2}{n} \sum_{r=1}^n x_r^t x_r \end{aligned} \tag{1.2}$$

Para obtener b_{rs} en términos de distancias cuadradas utilizamos (1.1) y (1.2):

$$\begin{aligned}
b_{rs} &= x_r^t x_s \\
&= \frac{x_r^t x_r}{2} + \frac{x_s^t x_s}{2} - \frac{d_{rs}^2}{2} \\
&= -\frac{1}{2} \left(d_{rs}^2 - \frac{1}{n} \sum_{s=1}^n d_{rs}^2 + \frac{1}{n} \sum_{s=1}^n x_s^t x_s - \frac{1}{n} \sum_{r=1}^n d_{rs}^2 + \frac{1}{n} \sum_{r=1}^n x_r^t x_r \right) \\
&= -\frac{1}{2} \left(d_{rs}^2 - \frac{1}{n} \sum_{s=1}^n d_{rs}^2 - \frac{1}{n} \sum_{r=1}^n d_{rs}^2 + \frac{2}{n} \sum_{r=1}^n x_r^t x_r \right) \\
&= -\frac{1}{2} \left(d_{rs}^2 - \frac{1}{n} \sum_{s=1}^n d_{rs}^2 - \frac{1}{n} \sum_{r=1}^n d_{rs}^2 + \frac{1}{n^2} \sum_{r=1}^n \sum_{s=1}^n d_{rs}^2 \right) \\
&= a_{rs} - a_{r.} - a_{.s} + a_{..}
\end{aligned}$$

donde

$$a_{rs} = -\frac{1}{2} d_{rs}^2$$

y

$$a_{r.} = \frac{1}{n} \sum_{s=1}^n a_{rs}, \quad a_{.s} = \frac{1}{n} \sum_{r=1}^n a_{rs}, \quad a_{..} = \frac{1}{n^2} \sum_{r=1}^n \sum_{s=1}^n a_{rs}.$$

Definimos a la matriz $A_{n \times n}$ como $[A]_{rs} = a_{rs}$ y de aquí la matriz de producto interno B es

$$B = HAH$$

donde H es la matriz de centrado,

$$H_{n \times n} = I_n - n^{-1} \mathbf{1}_n \mathbf{1}_n^t$$

con I_n la matriz identidad de orden $n \times n$ y $\mathbf{1}_n = (1, \dots, 1)^t$, un vector de unos de orden n .

Proposición 1. $B=HAH$

Demostración.

$$\begin{aligned}
HAH &= (I - n^{-1}11^t) A (I - n^{-1}11^t) \\
&= IAI - n^{-1}IA11^t - n^{-1}11^tAI + n^{-2}11^tA11^t \\
&= \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} - \frac{1}{n} \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \\
&\quad - \frac{1}{n} \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \\
&\quad + \frac{1}{n^2} \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \\
&= \begin{pmatrix} a_{11} - a_{1.} - a_{.1} + a_{..} & \cdots & a_{1n} - a_{1.} - a_{.n} + a_{..} \\ \vdots & \ddots & \vdots \\ a_{n1} - a_{n.} - a_{.1} + a_{..} & \cdots & a_{nn} - a_{n.} - a_{.n} + a_{..} \end{pmatrix} \\
&= B
\end{aligned}$$

■

Configuración de X

La matriz de producto interno B puede ser expresada como

$$B = XX^t$$

donde $X_{n \times p} = [x_1, \dots, x_n]^t$ es la matriz de coordenadas.

B es simétrica, semidefinida positiva y de rango p , y de aquí que tiene p valores propios no negativos y $n - p$ valores propios iguales a cero. La matriz B es escrita en términos de su descomposición espectral

$$B = V\Lambda V^t$$

donde $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, la matriz diagonal de valores propios (raíces caracterísctas) $\{\lambda_i\}$ de B , y $V = [v_1, \dots, v_n]$, la matriz de vectores propios

correspondientes, normalizados tal que $v_i^t v_i = 1$. Por conveniencia los valores propios de B están etiquetados de tal manera que $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$.

Como B tiene $n - p$ valores propios iguales a cero, puede ser reescrita como

$$B = V_1 \Lambda_1 V_1^t$$

donde $\Lambda_1 = \text{diag}(\lambda_1, \dots, \lambda_p)$ y $V_1 = [v_1, \dots, v_p]$.

De aquí que, como $B = X X^t$, la matriz de coordenadas X está dada por

$$X = V_1 \Lambda_1^{\frac{1}{2}}$$

donde $\Lambda_1^{\frac{1}{2}} = \text{diag}(\lambda_1^{\frac{1}{2}}, \dots, \lambda_p^{\frac{1}{2}})$, y así las coordenadas de los puntos son obtenidas a partir de las distancias entre los puntos.

1.1.1. Disimilaridades como Distancias Euclidianas

En la práctica generalmente se necesita encontrar la configuración de los puntos a partir de un conjunto de disimilaridades $\{\delta_{rs}\}$ en lugar de las distancias euclidianas entre los puntos $\{d_{rs}\}$.

Suponga que las disimilaridades $\{\delta_{rs}\}$ son usadas en lugar de las distancias $\{d_{rs}\}$ para definir la matriz A , la cual está doblemente centrada para producir B como se describió². Es interesante preguntarnos bajo qué circunstancias B puede dar una configuración de puntos en un espacio euclidiano, usando la descomposición espectral, de tal manera que las distancias asociadas $\{d_{rs}\}$ sean tales que $d_{rs} = \delta_{rs}$ para toda r, s . La respuesta es que si B es semidefinida positiva de rango p , entonces se puede encontrar una configuración en un espacio euclidiano p -dimensional. Este resultado se muestra en el siguiente teorema.

Teorema 1. *Sea $D = [d_{rs}]$ una matriz de distancias y defínanse las matrices: $A = [a_{rs}]$ con $a_{rs} = -\frac{1}{2}d_{rs}^2$ y $B = HAH$ donde $H = I_n - n^{-1}1_n 1_n^t$ con $1_n = (1, \dots, 1)^t$, de modo que $b_{rs} = a_{rs} - a_{r.} - a_{.s} + a_{..}$. Entonces D es euclidiana si y sólo si B es semidefinida positiva.*

En particular el teorema establece lo siguiente:

²Una matriz doblemente centrada es aquella en la cual la media de los elementos en cada renglón y la media de los elementos en cada columna es igual a 0.

a) Si D es la matriz de distancias euclidianas entre los puntos de una configuración $Z = [z_1, \dots, z_n]^t$ entonces $b_{rs} = (z_r - z)^t (z_s - z)$, lo que equivale a decir que $B = (HZ)(HZ)^t$ y B es semidefinida positiva.

B es la matriz de producto interno centrada para la configuración Z .

b) Inversamente, si B es semidefinida positiva de rango p entonces se puede construir una configuración correspondiente a D como sigue:

Sean $\lambda_1 > \lambda_2 > \dots > \lambda_p$ los valores propios positivos de B con sus correspondientes vectores propios $X = [x_{(1)}, x_{(2)}, \dots, x_{(p)}]$ normalizados de modo que $x_{(k)}^t x_{(k)} = \lambda_k$ para $k = 1, \dots, p$. Entonces los puntos x_r con coordenadas $(x_{r1}, \dots, x_{rp})^t$, $r = 1, \dots, n$, definen una configuración donde las distancias entre ellos están dadas por los elementos de D . Además esta configuración tiene centro de gravedad $x = 0$ y B representa su matriz de producto interno.

Demostración a). Se probará que $b_{rs} = (z_r - z)^t (z_s - z)$ utilizando las definiciones de b_{rs} y a_{rs} . Sabemos que

$$d_{rs}^2 = -2a_{rs} = (z_r - z_s)^t (z_r - z_s)$$

entonces

$$\begin{aligned} -2a_{rs} &= z_r^t z_r - z_r^t z_s - z_s^t z_r + z_s^t z_s \\ &= z_r^t z_r - 2z_r^t z_s + z_s^t z_s \end{aligned}$$

Sumando sobre $s = 1, \dots, n$ y multiplicando por $\frac{1}{n}$ se obtiene

$$\begin{aligned} -2a_{r.} &= \frac{1}{n} \sum_{s=1}^n z_r^t z_r - \frac{2}{n} \sum_{s=1}^n z_r^t z_s + \frac{1}{n} \sum_{s=1}^n z_s^t z_s \\ &= z_r^t z_r - 2z_r^t z + \frac{1}{n} \sum_{s=1}^n z_s^t z_s \end{aligned}$$

Sumando sobre $r = 1, \dots, n$ y multiplicando por $\frac{1}{n}$ se obtiene

$$\begin{aligned} -2a_{.s} &= \frac{1}{n} \sum_{r=1}^n z_r^t z_r - \frac{2}{n} \sum_{r=1}^n z_r^t z_s + \frac{1}{n} \sum_{r=1}^n z_s^t z_s \\ &= \frac{1}{n} \sum_{r=1}^n z_r^t z_r - 2z^t z_s + z_s^t z_s \end{aligned}$$

8 CAPÍTULO 1. ESCALAMIENTO MULTIDIMENSIONAL MÉTRICO

Sumando sobre $r, s = 1, \dots, n$ y multiplicando por $\frac{1}{n^2}$ se obtiene

$$\begin{aligned}
 -2a_{..} &= \frac{1}{n^2} \sum_{r=1}^n \sum_{s=1}^n z_r^t z_r - \frac{2}{n^2} \sum_{r=1}^n \sum_{s=1}^n z_r^t z_s + \frac{1}{n^2} \sum_{r=1}^n \sum_{s=1}^n z_s^t z_s \\
 &= \frac{1}{n} \sum_{r=1}^n z_r^t z_r - \frac{2}{n} \sum_{r=1}^n z_r^t z + \frac{1}{n} \sum_{s=1}^n z_s^t z_s \\
 &= \frac{2}{n} \sum_{r=1}^n z_r^t z_r - 2z^t z
 \end{aligned}$$

Substituyendo en b_{rs}

$$\begin{aligned}
 b_{rs} &= a_{rs} - a_{r.} - a_{.s} + a_{..} \\
 &= -\frac{1}{2} (z_r^t z_r - z_r^t z_s - z_s^t z_r + z_s^t z_s) + \frac{1}{2} \left(z_r^t z_r - 2z_r^t z + \frac{1}{n} \sum_{s=1}^n z_s^t z_s \right) \\
 &\quad + \frac{1}{2} \left(\frac{1}{n} \sum_{r=1}^n z_r^t z_r - 2z^t z_s + z_s^t z_s \right) - \frac{1}{2} \left(\frac{2}{n} \sum_{r=1}^n z_r^t z_r - 2z^t z \right) \\
 &= -\frac{1}{2} z_r^t z_r + \frac{1}{2} z_r^t z_s + \frac{1}{2} z_s^t z_r - \frac{1}{2} z_s^t z_s + \frac{1}{2} z_r^t z_r - z_r^t z + \frac{1}{2n} \sum_{s=1}^n z_s^t z_s \\
 &\quad + \frac{1}{2n} \sum_{r=1}^n z_r^t z_r - z^t z_s + \frac{1}{2} z_s^t z_s - \frac{1}{n} \sum_{r=1}^n z_r^t z_r + z^t z \\
 &= z_r^t z_s - z_r^t z - z^t z_s + z^t z \\
 &= (z_r - z)^t (z_s - z)
 \end{aligned}$$

De esto se obtiene que $B = XX^t \geq 0$ (B es semidefinida positiva) donde $X = [z_1 - z, \dots, z_n - z]^t$.

Así la parte a) queda demostrada. ■

Demostración b). Supóngase que $B \geq 0$ (semidefinida positiva) es de rango p y sea $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$ entonces por el teorema de la descomposición espectral existe una matriz ortogonal $V = [v_{(1)}, v_{(2)}, \dots, v_{(p)}]$ tal que

$$B = V\Lambda V^t = \left(V\Lambda^{\frac{1}{2}}\right) \left(V\Lambda^{\frac{1}{2}}\right)^t = XX^t.$$

De aquí se obtiene que $b_{rs} = x_r^t x_s$ y B representa la matriz de producto interno de esta configuración.

En lo siguiente se prueba que los elementos de D determinan las distancias entre los puntos de la configuración.

$$\begin{aligned}
 d_2^2(x_r, x_s) &= (x_r - x_s)^t (x_r - x_s) \\
 &= x_r^t x_r - x_r^t x_s - x_s^t x_r + x_s^t x_s \\
 &= x_r^t x_r - 2x_r^t x_s + x_s^t x_s \\
 &= b_{rr} - 2b_{rs} + b_{ss} \\
 &= (a_{rr} - a_{r.} - a_{.r} + a_{..}) - 2(a_{rs} - a_{r.} - a_{.s} + a_{..}) \\
 &\quad + (a_{ss} - a_{s.} - a_{.s} + a_{..})
 \end{aligned}$$

como $a_{rs} = a_{sr}$, entonces $a_{r.} = a_{.r}$ ya que

$$a_{r.} = \frac{1}{n} \sum_{s=1}^n a_{rs} = \frac{1}{n} \sum_{s=1}^n a_{sr} = a_{.r}$$

y se tiene

$$\begin{aligned}
 d_2^2(x_r, x_s) &= a_{rr} - 2a_{rs} + a_{ss} \\
 &= -\frac{1}{2}d_{rr}^2 - 2\left(-\frac{1}{2}d_{rs}^2\right) - \frac{1}{2}d_{ss}^2 \\
 &= d_{rs}^2
 \end{aligned}$$

Finalmente nótese que

$$\begin{aligned}
 H1_n &= (I_n - n^{-1}1_n 1_n^t) 1_n \\
 &= 1_n - n^{-1}1_n 1_n^t 1_n \\
 &= 1_n - n^{-1}1_n n \\
 &= 1_n - 1_n \\
 &= 0_n
 \end{aligned}$$

Por lo que $B1_n = HAH1_n = 0_n = 0(1_n)$ y 1_n es un vector propio de B correspondiente al valor propio $\lambda = 0$. Esto implica que 1_n es ortogonal a las columnas de X :

$$x_{(k)}^t 1_n = 0 \quad \text{para } k = 1, \dots, p$$

De esto se deduce que

$$nx = \sum_{r=1}^n x_r = X^t \mathbf{1}_n = [x_{(1)}^t \mathbf{1}_n, \dots, x_{(p)}^t \mathbf{1}_n]^t = \mathbf{0}_p$$

Así que el centro de gravedad de esta configuración está en el origen. Y con esto termina la demostración. ■

La siguiente pregunta será saber cuántas dimensiones son requeridas en general para producir una configuración de puntos a partir de una matriz semidefinida positiva B de disimilaridades. Como se vió en el teorema 1, B tiene al menos un valor propio cero. Hasta aquí siempre puede ser encontrada una configuración de puntos en un espacio euclidiano $n-1$ dimensional, cuyas distancias asociadas son iguales a las disimilaridades.

Si las disimilaridades dan una matriz B que no sea semidefinida positiva, puede agregarse una constante a todas las disimilaridades (excepto las disimilaridades de tipo δ_{rr} para $r = 1, \dots, n$) haciendo a B semidefinida positiva. Formando nuevas disimilaridades $\{\delta_{rs}^*\}$ como $\delta_{rs}^* = \delta_{rs} + c(1 - \delta^{rs})$, donde c es una constante apropiada y δ^{rs} es la delta de Kronecker ($\delta^{rs} = 1$ si $r = s$ y $\delta^{rs} = 0$ en otro caso). Una vez que B se ha hecho semidefinida positiva, se puede encontrar un espacio euclidiano como se hizo anteriormente donde las distancias d_{rs} son exactamente iguales a las disimilaridades δ_{rs}^* .

Proposición 2. Sea $\delta_{rs}^* = \delta_{rs} + c(1 - \delta^{rs})$ donde δ^{rs} es la delta de Kronecker y

$$c \geq c_{\min} = \max_{(r,s,t)} (\delta_{rs} - \delta_{rt} - \delta_{ts}).$$

Entonces B es semidefinida positiva.

Demostración. Por el teorema 1, hay que demostrar que $D = [\delta_{rs}^*]$ es euclidiana, es decir, que cumple con las tres propiedades de la distancia euclidiana presentadas en la tabla 2.

No negatividad:

$$\delta_{rr}^* = \delta_{rr} + c(1 - \delta^{rr}) = \delta_{rr} = 0$$

$$\delta_{rs}^* = \delta_{rs} + c(1 - \delta^{rs}) = \delta_{rs} + c \geq 0$$

Simetría:

$$\delta_{rs}^* = \delta_{rs} + c(1 - \delta^{rs}) = \delta_{sr} + c(1 - \delta^{sr}) = \delta_{sr}^*$$

Desigualdad del triángulo:

Si $r = s$ se cumple la desigualdad. Si $r \neq s$ $\delta_{rs}^* = \delta_{rs} + c$. Como

$$\max_{(r,s,t)} (\delta_{rs} - \delta_{rt} - \delta_{ts}) - (\delta_{rs} - \delta_{rt} - \delta_{ts}) \geq 0$$

entonces

$$c - (\delta_{rs} - \delta_{rt} - \delta_{ts}) \geq 0$$

$$c + \delta_{rt} + \delta_{ts} \geq \delta_{rs}$$

$$\delta_{rt}^* + \delta_{ts}^* \geq \delta_{rs}^*$$

■

Observaciones

1. La matriz X puede ser visualizada como sigue, en términos de los vectores propios de B y los puntos correspondientes

$$\begin{array}{cccc}
 & \lambda_1 & \lambda_2 & \dots & \lambda_p \\
 P_1 & \left(\begin{array}{cccc} x_{11} & x_{12} & \dots & x_{1p} \end{array} \right) & x_1^t \\
 P_2 & \left(\begin{array}{cccc} x_{21} & x_{22} & \dots & x_{2p} \end{array} \right) & x_2^t \\
 \vdots & \left(\begin{array}{cccc} \vdots & \vdots & \ddots & \vdots \end{array} \right) & \vdots \\
 P_n & \left(\begin{array}{cccc} x_{n1} & x_{n2} & \dots & x_{np} \end{array} \right) & x_n^t \\
 & x_{(1)} & x_{(2)} & \dots & x_{(p)}
 \end{array}$$

Centro de gravedad: $x_i = 0$ para $i = 1, \dots, p$ y $x = \frac{1}{n} \sum_{r=1}^n x_r = 0_p$.

En resumen, el i -ésimo renglón de X contiene las coordenadas del i -ésimo punto de la configuración, mientras que la j -ésima columna de X contiene al vector propio correspondiente al valor propio λ_j .

2. Geométricamente, si B es la matriz de producto interno centrado para la configuración Z , entonces $b_{rr}^{1/2}$ es igual a la distancia entre z_r y z y $b_{rs}/(b_{rr}b_{ss})^{1/2}$ es igual al coseno del ángulo subtendido desde z entre z_r y z_s .
3. El vector 1_n es vector propio de B sea D euclidiana o no.
4. Nótese que la solución que se obtiene no es única, un cambio del origen y una rotación o una reflexión no cambiaría las distancias entre los puntos, por ejemplo, supóngase que T es una matriz ortogonal de $(p \times p)$, entonces Tx_r también proporciona una solución

$$\begin{aligned}
d_2^2(Tx_r, Tx_s) &= (Tx_r - Tx_s)^t (Tx_r - Tx_s) \\
&= (x_r^t T^t - x_s^t T^t) (Tx_r - Tx_s) \\
&= x_r^t T^t Tx_r - x_r^t T^t Tx_s - x_s^t T^t Tx_r + x_s^t T^t Tx_s \\
&= x_r^t x_r - x_r^t x_s - x_s^t x_r + x_s^t x_s \\
&= (x_r - x_s)^t (x_r - x_s) \\
&= d_2^2(x_r, x_s)
\end{aligned}$$

Sin embargo, el problema del cambio de origen se soluciona al hacer que el centro de gravedad sea igual a cero; y la posibilidad de rotar o reflejar la solución es favorable para la interpretación de la misma.

5. Este resultado fue probado primero por Schoenberg (1935) y Young y Householder (1938). Su utilización como base del MDS se debe a Torgerson (1958) y sus ideas fueron sustancialmente amplificadas por Gower (1966)³.

1.1.2. Similaridades

En algunas situaciones no se comienza con una matriz de distancias sino con una de similaridades $C = [c_{rs}]$. Recordemos que una matriz $C_{n \times n}$ es llamada matriz de similaridades si $c_{rs} = c_{sr}$ y si $c_{rs} \leq c_{rr} \forall r, s = 1, \dots, n$. Para utilizar los procedimientos del escalamiento multidimensional clásico se

³Ver Mardia 1979.

deben transformar las similaridades en distancias, la llamada transformación estándar es muy útil y está dada por

$$d_{rs} = (c_{rr} - 2c_{rs} + c_{ss})^{\frac{1}{2}}$$

Si la matriz de similaridades es semidefinida positiva entonces la matriz de distancias que se obtiene resulta ser euclidiana, como lo establece el siguiente teorema.

Teorema 2. Si $C \geq 0$ (semidefinida positiva) entonces $D = [d_{rs}]$ definida por $d_{rs} = (c_{rr} - 2c_{rs} + c_{ss})^{\frac{1}{2}}$ es euclidiana con matriz de producto interno centrado $B = HCH$.

Demostración: Primero obsérvese que: como $C \geq 0$ entonces $x^t C x \geq 0$ para toda $x \in \mathfrak{R}^n$. En particular si x es un vector con un 1 en el lugar r -ésimo, un -1 en el lugar s -ésimo y ceros en los demás lugares, se tiene:

$$d_{rs}^2 = c_{rr} - 2c_{rs} + c_{ss} = x^t C x \geq 0$$

Con lo que se prueba que d_{rs} está bien definida.

Ahora, sean las matrices A y B como en el teorema 1. Dado que HCH también es semidefinida positiva es suficiente probar que $B = HCH$ para deducir que D es euclidiana con matriz de producto interno centrado HCH .

Los elementos de $B = HAH$ se pueden escribir, utilizando su definición y la de d_{rs} , como sigue

$$\begin{aligned} b_{rs} &= a_{rs} - a_{r.} - a_{.s} + a_{..} \\ &= -\frac{1}{2}d_{rs}^2 - \frac{1}{n} \sum_{s=1}^n \left(-\frac{1}{2}d_{rs}^2 \right) - \frac{1}{n} \sum_{r=1}^n \left(-\frac{1}{2}d_{rs}^2 \right) + \frac{1}{n^2} \sum_{r=1}^n \sum_{s=1}^n \left(-\frac{1}{2}d_{rs}^2 \right) \end{aligned}$$

de donde se tiene que

$$\begin{aligned}
-2b_{rs} &= d_{rs}^2 - \frac{1}{n} \sum_{s=1}^n d_{rs}^2 - \frac{1}{n} \sum_{r=1}^n d_{rs}^2 + \frac{1}{n^2} \sum_{r=1}^n \sum_{s=1}^n d_{rs}^2 \\
&= (c_{rr} - 2c_{rs} + c_{ss}) - \frac{1}{n} \sum_{s=1}^n (c_{rr} - 2c_{rs} + c_{ss}) \\
&\quad - \frac{1}{n} \sum_{r=1}^n (c_{rr} - 2c_{rs} + c_{ss}) + \frac{1}{n^2} \sum_{r=1}^n \sum_{s=1}^n (c_{rr} - 2c_{rs} + c_{ss}) \\
&= c_{rr} - 2c_{rs} + c_{ss} - c_{rr} + 2c_{r.} - \frac{1}{n} \sum_{s=1}^n c_{ss} \\
&\quad - \frac{1}{n} \sum_{r=1}^n c_{rr} + 2c_{.s} - c_{ss} + \frac{1}{n} \sum_{r=1}^n c_{rr} - 2c_{..} + \frac{1}{n} \sum_{s=1}^n c_{ss} \\
&= -2c_{rs} + 2c_{r.} + 2c_{.s} - 2c_{..}
\end{aligned}$$

finalmente se obtiene que

$$b_{rs} = c_{rs} - c_{r.} - c_{.s} + c_{..}$$

y

$$B = HCH \geq 0$$

Queda entonces demostrado el teorema 2. ■

1.1.3. Propiedades Óptimas de la Solución Clásica y Bondad de Ajuste

Algunas veces la matriz de distancia D no es euclidiana, y entonces algunos de los valores propios de B son negativos. En este caso no se puede escribir Λ como $\Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}}$ y el teorema fundamental no se cumple.

Por otro lado, aún cuando la matriz D sea euclidiana, la dimensión del espacio en el que puede ser representada, generalmente resulta demasiado grande para ser de interés práctico.

En ambos casos, se puede construir una configuración en un espacio \mathfrak{R}^k cuyas coordenadas están determinadas por k vectores propios de B , correspondientes a los primeros k valores propios de B , suponiendo que éstos son

positivos y comparativamente grandes y que los restantes $p-k$ valores propios son cercanos a cero (positivos o negativos).

Cuando se tiene una configuración de esta manera, a los puntos también suele llamárseles las coordenadas principales de X en k dimensiones. Y, a esta configuración se le conoce como la solución clásica en k dimensiones del análisis de escalamiento multidimensional.

Dada una matriz de distancias $D = [d_{rs}]$ el objetivo del análisis de escalamiento multidimensional clásico es encontrar una configuración \widehat{X} en un espacio euclidiano de dimensión pequeña, de manera que las distancias entre los puntos estén dadas por:

$$\widehat{d}_{rs}^2 = (\widehat{x}_r - \widehat{x}_s)^t (\widehat{x}_r - \widehat{x}_s) = d^2(r, s).$$

El símbolo “ $\widehat{}$ ” se utiliza en esta parte para indicar las distancias \widehat{D} de la configuración \widehat{X} que ajustan a las verdaderas distancias D . Sea $\widehat{B} = \widehat{X}\widehat{X}^t$ la matriz de producto interno centrado ajustada.

Sea X la configuración en \mathfrak{R}^p y $T = [T_1, T_2]_{p \times p}$ una matriz ortogonal, con T_1 de orden $(p \times k)$ y T_2 de orden $(p \times (p - k))$. Entonces XT representa una proyección de la configuración X en el subespacio de \mathfrak{R}^p expandido por las columnas de T . Se puede pensar en $\widehat{X} = XT_1$ como la configuración ajustada de dimensión k .

Como T es ortogonal, las distancias entre los renglones de X son las mismas que entre los renglones de XT :

$$d_{rs}^2 = \sum_{h=1}^p (x_{rh} - x_{sh})^2 = \sum_{h=1}^p (x_r^t 1_h - x_s^t 1_h)^2$$

donde 1_h es un vector con un 1 en el lugar h -ésimo y ceros en los demás lugares.

Si se denotan las distancias entre los renglones de XT_1 por \widehat{D} entonces

$$\widehat{d}_{rs}^2 = \sum_{h=1}^k (x_r^t 1_h - x_s^t 1_h)^2$$

De donde se observa que $\widehat{d}_{rs} \leq d_{rs}$, lo que significa que proyectando una configuración se reducen las distancias entre los puntos.

Una medida de discrepancia entre la configuración original X y la configuración proyectada \widehat{X} está dada por

$$\phi = \sum_{r=1}^n \sum_{s=1}^n (d_{rs}^2 - \widehat{d}_{rs}^2)$$

Entonces, la solución clásica del análisis de escalamiento multidimensional en k dimensiones tiene las siguientes propiedades óptimas.

Teorema 3. *Sea D una matriz de distancias euclidiana correspondiente a la configuración X en \mathbb{R}^p y sea k un número fijo ($1 \leq k \leq p$). Entonces de entre todas las proyecciones XT de X , en subespacios de dimensión k de \mathbb{R}^p la cantidad ϕ se minimiza cuando X se proyecta sobre sus coordenadas principales en k dimensiones.*

Demostración: Usando la definición de d_{rs}^2 y de \widehat{d}_{rs}^2 , se puede escribir a ϕ como sigue

$$\begin{aligned}
\phi &= \sum_{r=1}^n \sum_{s=1}^n \left(d_{rs}^2 - \widehat{d}_{rs}^2 \right) \\
&= \sum_{r=1}^n \sum_{s=1}^n \left[\sum_{h=1}^p (x_r^t 1_h - x_s^t 1_h)^2 - \sum_{h=1}^k (x_r^t 1_h - x_s^t 1_h)^2 \right] \\
&= \sum_{r=1}^n \sum_{s=1}^n \sum_{h=k+1}^p (x_r^t 1_h - x_s^t 1_h)^2 \\
&= \text{tr} \left[T_2^t \left(\sum_{r=1}^n \sum_{s=1}^n (x_r - x_s)(x_r - x_s)^t \right) T_2 \right] \\
&= 2n^2 \text{tr} (T_2^t S T_2)
\end{aligned}$$

Esto último porque

$$\begin{aligned}
\sum_{r,s=1}^n (x_r - x_s)(x_r - x_s)^t &= \sum_{r,s=1}^n (x_r x_r^t - x_r x_s^t - x_s x_r^t + x_s x_s^t) \\
&= \sum_{r,s=1}^n x_r x_r^t - \sum_{r,s=1}^n x_r x_s^t - \sum_{r,s=1}^n x_s x_r^t + \sum_{r,s=1}^n x_s x_s^t \\
&= 2n \sum_{r=1}^n x_r x_r^t - 2n \sum_{r=1}^n x x_r^t \\
&= 2n \sum_{r=1}^n (x_r x_r^t - x x_r^t - x_r x^t + x_r x^t)
\end{aligned}$$

$$\begin{aligned}
&= 2n \sum_{r=1}^n (x_r - x) (x_r - x)^t \\
&= 2n (nS)
\end{aligned}$$

Sean $\lambda_1, \dots, \lambda_p$ los valores propios de nS con sus correspondientes vectores propios estandarizados $V_{p \times p} = [v_{(1)}, \dots, v_{(p)}]$, entonces ϕ se puede escribir como

$$\phi = 2n \operatorname{tr} (U_2^t \Lambda U_2)$$

con $U_2 = V^t T_2$ matriz columna ortogonal de orden $p \times (p - k)$.

Utilizando que ϕ se minimiza cuando $U_2 = [0_{(p-k) \times k}, I_{p-k}]^t$, esto es, cuando $T_2 = [v_{(k+1)}, \dots, v_{(p)}]$. Se tiene entonces que las columnas de T_1 expanden el espacio de los primeros k vectores propios de nS ; XT_1 representa las coordenadas principales de X en k dimensiones.

Nótese que para esta proyección de coordenadas principales se tiene entonces que

$$\phi = 2n (\lambda_{(k+1)}, \dots, \lambda_{(p)})$$

■

Cuando D no necesariamente es euclidiana es más conveniente trabajar con la matriz $B = HAH$. Si \hat{X} es una configuración ajustada con matriz de producto interno centrado \hat{B} , entonces la medida de discrepancia entre B y \hat{B} está dada por

$$\psi = \sum_{r=1}^n \sum_{s=1}^n (b_{rs} - \hat{b}_{rs})^2 = \operatorname{tr} (B - \hat{B})^2$$

Para esta medida también se puede probar que la solución clásica del escalamiento multidimensional es óptima.

Teorema 4. *Si D es una matriz de distancias (no necesariamente euclidianas), entonces para un número k fijo, ψ se minimiza sobre todas las configuraciones \hat{X} en k dimensiones cuando se obtiene de la solución clásica del análisis de escalamiento multidimensional.*

Demostración: Sean $\lambda_1 \geq \dots \geq \lambda_n$ los valores propios de B , algunos de los cuales pueden ser negativos con sus correspondientes vectores propios estandarizados $V = [v_{(1)}, \dots, v_{(n)}]$. Por simplicidad supóngase que $\lambda_k > 0$.

Si $\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_n \geq 0$ son los valores propios de \hat{B} . Por el teorema de la descomposición espectral se puede escribir

$$V^t \hat{B} V = G \hat{\Lambda} G^t$$

con G ortogonal, entonces

$$\begin{aligned} \psi &= \text{tr} (B - \hat{B})^2 = \text{tr} [V^t (B - \hat{B}) V V^t (B - \hat{B}) V] \\ &= \text{tr} (V^t B V - V^t \hat{B} V) (V^t B V - V^t \hat{B} V) \\ &= \text{tr} (\Lambda - G \hat{\Lambda} G^t) (\Lambda - G \hat{\Lambda} G^t) \end{aligned}$$

Obsérvese que para $\hat{\Lambda}$ fija, ψ se minimiza para $G = I$ de tal manera que

$$\psi = \sum_{h=1}^n (\lambda_h - \hat{\lambda}_h)^2$$

Como cada renglón de \hat{X} está en \Re^k , $B = H \hat{X} \hat{X}^t H$ tendrá k valores propios diferentes que sean positivos, y se puede ver que ψ se minimiza si

$$\hat{\lambda}_h = \begin{cases} \lambda_h & \text{si } h = 1, \dots, k \\ 0 & \text{si } h = k + 1, \dots, n \end{cases}$$

De donde $\hat{B} = V_1 \Lambda_1 V_1^t$ con $V_1 = [v_{(1)}, \dots, v_{(k)}]$ y $\Lambda_1 = \text{diag}(\lambda_1, \dots, \lambda_k)$. Entonces \hat{X} puede ser tomada igual que $V_1 \Lambda_1^{1/2}$, la solución clásica en k dimensiones del escalamiento multidimensional.

Nótese que el valor mínimo de ψ está dado por

$$\psi = \lambda_{k+1}^2 + \dots + \lambda_p^2$$

■

Los dos teoremas anteriores sugieren una posible medida de ajuste para la proporción explicada de una matriz de distancias por la solución clásica en k dimensiones.

Supóngase que $\lambda_k > 0$, entonces quedan definidas dos medidas como

$$\alpha_{1k} = \left(\frac{\sum_{h=1}^k \lambda_h}{\sum_{h=1}^n |\lambda_h|} \right) \times 100 \%$$

$$\alpha_{2k} = \left(\frac{\sum_{h=1}^k \lambda_h^2}{\sum_{h=1}^n \lambda_h^2} \right) \times 100 \%$$

Se necesita valor absoluto o elevar al cuadrado porque los valores propios más pequeños pueden ser negativos.

Si se desea una solución en cierto número de dimensiones p^* ($p^* \leq k$), entonces se puede examinar el conjunto de vectores propios asociados con los p^* valores propios más grandes. La medida de ajuste de la representación obtenida estaría dada de manera similar con los índices citados, sustituyendo el valor de k por el de p^* .

1.1.4. Un Algoritmo Práctico para el Escalamiento Clásico

La mecánica del análisis métrico se puede resumir con el siguiente algoritmo:

1. Obtener las disimilaridades $\{\delta_{rs}\}$.
2. Construir la matriz $A = [-\frac{1}{2}\delta_{rs}^2]$.
3. Construir la matriz $B = HAH$ donde $H = I_n - \frac{1}{n}1_n1_n^t$ y $1_n = (1, \dots, 1)^t$ de modo que los elementos de B quedan definidos por:

$$b_{rs} = a_{rs} - a_{r.} - a_{.s} + a_{..}$$

4. Obtener los valores propios $\lambda_1, \dots, \lambda_{n-1}$ y vectores propios v_1, \dots, v_{n-1} , donde los vectores propios están normalizados de tal manera que $v_i^t v_i = \lambda_i$. Si B no es semidefinida positiva (algunos de los valores propios son negativos), se pueden (1) ignorar los valores negativos y continuar con el análisis, o (2) añadir una constante apropiada c a las disimilaridades, $\delta_{rs}^* = \delta_{rs} + c(1 - \delta^{rs})$, donde δ^{rs} es la delta de Kronecker⁴ y $c \geq \max_{(r,s,t)} (\delta_{rs} - \delta_{rt} - \delta_{ts})$ y regresar al paso 2.

⁴La delta de Kronecker está definida como $\delta^{rs} = 1$ si $r = s$ y $\delta^{rs} = 0$ en otro caso.

5. Elegir un número apropiado de dimensiones p . Posiblemente usar

$$\frac{\sum_{i=1}^p \lambda_i}{\sum (\text{valores propios positivos})}$$

para legir la dimensión p , de acuerdo a la proporción de variación explicada.

6. Las coordenadas de los n puntos en el espacio p dimensional euclidiano están dadas por $x_{ri} = v_{ir}$ ($r = 1, \dots, n; i = 1, \dots, p$).

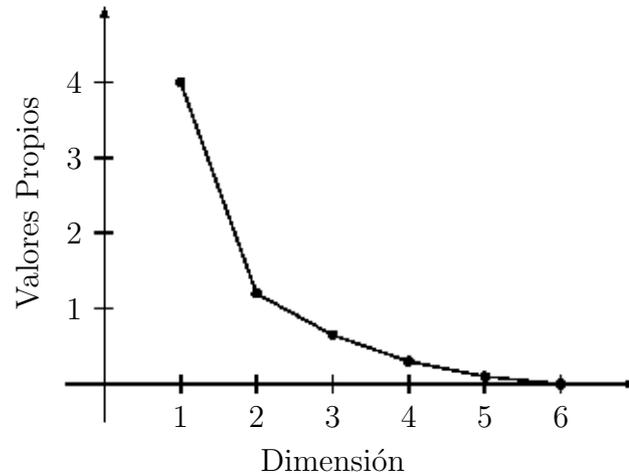
1.2. Dimensionalidad

Si B es semidefinida positiva entonces el número de valores propios distintos de cero da el número de dimensiones requerido. Si B no es semidefinida positiva entonces el número de valores propios positivos es el número apropiado de dimensiones. Esto es el máximo número de dimensiones que el espacio requiere. Sin embargo, en la práctica, el número de dimensiones del espacio elegido necesita ser pequeño. Como las coordenadas obtenidas están en función de sus coordenadas principales, entonces se eligen los primeros k valores propios y vectores propios de B (digamos $k=2$ o 3) dando un espacio de dimensión pequeña.

Como el MDS casi siempre es usado como un modelo descriptivo para la representación e interpretación de los datos, es recomendable obtener varias soluciones en diferentes dimensiones y elegir entre ellas con base en algunos criterios como: ajuste de los datos, reproductibilidad, interpretabilidad y proporción de la variación explicada. Davison (1983) sugiere que si k^* es el valor a priori que el investigador considera apropiado, entonces se deben obtener soluciones en cada dimensión de $k^* - 3$ a $k^* + 3$; si $k^* - 3 \leq 0$, entonces en cada dimensión de 1 a $k^* + 3$.

Ajuste de los datos

La solución clásica del análisis de escalamiento multidimensional es uno de los métodos en los cuales la medida de ajuste no juega un rol muy importante en la decisión sobre la dimensionalidad adecuada. Sin embargo, los valores propios $\{\lambda_i\}$ si juegan un rol importante para saber cuántas dimensiones son requeridas para representar las disimilaridades $\{\delta_{rs}\}$. Cada valor propio

Figura 1.1: λ vs. k

está asociado con una dimensión en la solución y éste es simplemente la suma de cuadrados de los valores de las coordenadas sobre esa dimensión, es decir

$$X^t X = \left(V \Lambda^{\frac{1}{2}} \right)^t \left(V \Lambda^{\frac{1}{2}} \right) = \Lambda^{\frac{1}{2}} V^t V \Lambda^{\frac{1}{2}} = \Lambda$$

entonces

$$\lambda_k = \sum_{r=1}^n x_{rk}^2$$

Una gráfica muy útil para determinar la dimensionalidad adecuada es llamada gráfica de codo o *scatter diagram*, la cual consiste en establecer sobre un plano cartesiano los puntos cuyas abscisas son los valores de las dimensiones y cuyas ordenadas son sus correspondientes valores propios (figura 1.2).

Si los datos satisfacen completamente el modelo, entonces la gráfica mostrará un levantamiento exactamente sobre la dimensión $k + 1$. En otras palabras, aparece un “codo” en la gráfica una unidad adelante de la dimensión k , el valor apropiado de la dimensionalidad. En los datos que no conforman exactamente el modelo o en los cuales hay una gran cantidad de medidas y muestran error, un codo puede ser difícil de identificar. En tales casos, una gráfica de los valores propios puede no ser suficiente para determinar el número correcto de dimensiones.

Reproductibilidad

La reproductibilidad puede ser usada como un criterio sólo cuando hay dos o más muestras de la población de objetos. La idea básica consiste en obtener una solución para cada muestra y si hay k dimensiones que aparecen en cada muestra, entonces la solución final debe contener k dimensiones.

Interpretabilidad

La interpretabilidad como un criterio para la elección de la dimensionalidad adecuada requiere algunos juicios subjetivos. Establece que se debe conservar el espacio en el cual aparezcan todas las características importantes de los objetos.

Esto significa que una solución de dimensión alta es preferible si provee una mejor interpretación, es decir, si en ella aparecen características importantes de los objetos que no aparecen en una solución de menor dimensión. Inversamente, es preferible una solución de dimensión pequeña si en la dimensión mayor no aparecen características significativas.

Proporción de la variación explicada

Una medida de la proporción de la variación explicada usando sólo k dimensiones es:

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

Si B no es semidefinida positiva esta medida se modifica a:

$$\frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^{n-1} |\lambda_i|} \quad \text{ó} \quad \frac{\sum_{i=1}^k \lambda_i}{\sum_{\{i \in I; \lambda_i > 0\}} \lambda_i}$$

La dimensión k es preferible si la proporción de variación explicada usando k dimensiones es suficientemente grande (digamos 75 %).

1.3. Rotación

En general resulta interesante, una vez obtenida la matriz de coordenadas X , rotar la solución para mejorar la interpretabilidad del resultado.

La rotación de los ejes no altera las distancias entre los objetos, por lo que es posible multiplicar la matriz de coordenadas $X_{n \times p}$ por una matriz de transformación ortogonal $T_{p \times p}$. Si X es una solución, $B = XX^t$, entonces $Y = XT$ también es solución, ya que

$$YY^t = (XT)(XT)^t = XTT^tX^t = XX^t = B$$

En la decisión sobre rotar una solución para hacerla más interpretable, hay tres opciones básicas. Si la solución no rotada es interpretable, entonces se conserva de esta manera. Si la solución no es fácilmente interpretable, entonces se puede realizar una rotación objetiva o una rotación a mano.

Una rotación objetiva es un algoritmo matemático para encontrar una solución más interpretable. Tales rotaciones fueron diseñadas primero para utilizarse en el análisis de factores y se usan ocasionalmente en MDS. En algunas aplicaciones una rotación objetiva como Varimax (Kaiser, 1958) o Equimax (Saunders, 1960) proveen una solución altamente interpretable, sin embargo, no se debe suponer que tales algoritmos de rotación objetiva producirán automáticamente la solución más interpretable.

Una rotación a mano es realizada por el investigador y se basa en una inspección visual de la configuración no rotada. En la práctica, algunas veces se puede ver, literalmente, cuál rotación de la solución podría hacerla más interpretable.

Además de la rotación, existen otras transformaciones que podemos realizar en las soluciones proporcionadas por MDS sin alterar para nada sus propiedades. Una de estas transformaciones es la reflexión. Es posible reflejar las escalas de los objetos en cualquiera de las dimensiones, o en varias de ellas simultáneamente, si con ello se facilita la interpretación de los resultados. Para reflejar la configuración de objetos en una dimensión determinada bastará con cambiar el signo a todas las coordenadas de esa dimensión (convirtiendo los valores positivos en negativos, y viceversa). La segunda transformación posible consiste en reescalar la configuración, multiplicando los valores originales de las coordenadas por una constante. Esto es posible debido a que las distancias son valores en escala de razón, por lo que una transformación del tipo $Y = aX$ es perfectamente válida. Una tercera transformación posible (excepto en modelos vectoriales) es trasladar el origen de coordenadas de su posición actual, sumando un valor constante a los valores de coordenadas originales, $Y = X + b$.

En resumen, si X es una solución del MDS en p dimensiones con coordenadas $x_r = (x_{r1}, \dots, x_{rp})^t$, $r = 1, \dots, n$, entonces $y_r = aAx_r + b$ también

es solución, donde $A_{p \times p}$ es una matriz ortogonal que permite la rotación o reflexión, $b_{p \times 1}$ es un vector que permite la traslación y a es una constante que permite el reescalamiento de los objetos. Esto es válido porque se preserva la proporción de las distancias:

$$\begin{aligned}
d_2^2(y_r, y_s) &= (y_r - y_s)^t (y_r - y_s) \\
&= y_r^t y_r - y_r^t y_s - y_s^t y_r + y_s^t y_s \\
&= (aAx_r + b)^t (aAx_r + b) - (aAx_r + b)^t (aAx_s + b) \\
&\quad - (aAx_s + b)^t (aAx_r + b) + (aAx_s + b)^t (aAx_s + b) \\
&= a^2 x_r^t A^t A x_r + a x_r^t A^t b + a b^t A x_r + b^t b \\
&\quad - a^2 x_r^t A^t A x_s - a x_r^t A^t b - a b^t A x_s - b^t b \\
&\quad - a^2 x_s^t A^t A x_r - a x_s^t A^t b - a b^t A x_r - b^t b \\
&\quad + a^2 x_s^t A^t A x_s + a x_s^t A^t b + a b^t A x_s + b^t b \\
&= a^2 (x_r^t x_r - x_r^t x_s - x_s^t x_r + x_s^t x_s) \\
&= a^2 (x_r - x_s)^t (x_r - x_s) \\
&= a^2 d_2^2(x_r, x_s)
\end{aligned}$$

1.4. Interpretación

Una vez que se ha decidido el valor de la dimensionalidad adecuada, la configuración de puntos que representa a los objetos debe ser interpretada. Se puede simplemente permitir que la posición de los puntos en el espacio determina la interpretación, lo que se llama método subjetivo, o seguir un método más objetivo como ajustar un vector característico o realizar un análisis de correlación canónica.

Una manera en que se interpretan habitualmente las gráficas de MDS es a partir de la posición de los objetos en las dimensiones: en qué se diferencian los objetos situados a la derecha de los situados a la izquierda, o los situados arriba de los situados abajo. De este modo utilizamos las dimensiones como continuos que nos permiten resumir la información proporcionada por la representación espacial. Dado que podemos rotar, reflejar, trasladar o reescalar la configuración del MDS, alguna de ellas servirá para mejorar la interpretabilidad de los datos.

1.4.1. Método Subjetivo

El método subjetivo para interpretar la configuración recae únicamente en la posición de los puntos en el espacio. Se trata de encontrar “características significativas”, en otras palabras, ordenamientos o agrupamientos de los puntos.

Un agrupamiento significativo es un conjunto de puntos que aparece en una región del espacio multidimensional y cuyos objetos representados poseen atributos en común.

Un ordenamiento es un arreglo ordenado de los puntos con respecto a cierta característica de los objetos que representan. El primer paso para identificar un ordenamiento es descubrir los puntos que se encuentran en posiciones extremas y observar las características que poseen los objetos que representan. De estas características, se elige la que mejor explica la posición relativa de los puntos en el espacio.

1.4.2. Ajuste de un Vector Característico

Este tipo de método objetivo para interpretar la configuración obtenida se basa en el siguiente razonamiento: supóngase que se conoce una variable que mide cierta característica de los objetos que se sospecha que tienen una relación sistemática con la posición de los puntos en el espacio. En este caso resulta conveniente utilizar dicha variable para tratar de explicar la configuración.

La figura 1.2 muestra una gráfica de lo que tratamos de hacer.

Esencialmente lo que se hace es buscar una dirección a través del espacio, la cual corresponda al incremento en la variable que mide la característica elegida. Geométricamente, significa insertar una línea, que se denotará por L , con la propiedad de que la proyección de cada punto sobre esta línea corresponda tanto como sea posible al grado en que el objeto representado posee la característica en cuestión. A la línea L se le llama vector característico o vector atributo.

En la figura hay una alta correspondencia entre los puntos x_1, x_2, \dots, x_6 que dan el grado que posee el atributo actual por cada uno de los seis objetos, y los puntos $X_{1L}, X_{2L}, \dots, X_{6L}$, que equivalen a las estimaciones.

Si la variable está fuertemente relacionada con la configuración, entonces los valores reales que poseen los objetos corresponderán considerablemente con las proyecciones de los puntos en el espacio y habrá una correlación alta

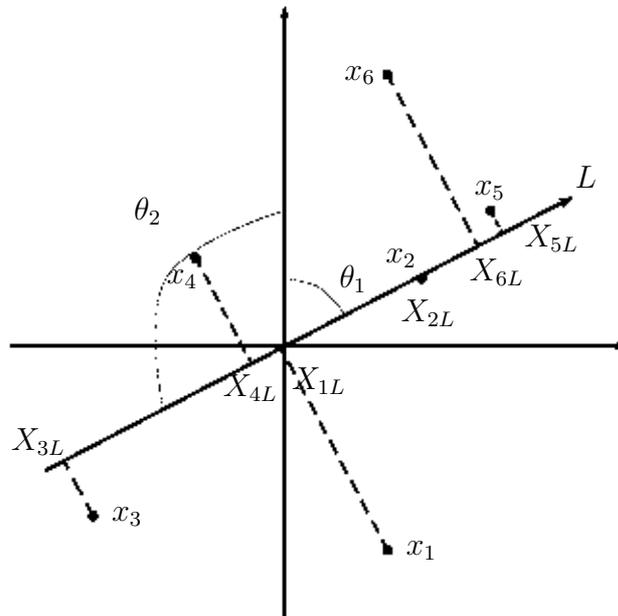


Figura 1.2: Vector característico

entre ellos. Si por el contrario, la variable no está muy relacionada con la configuración entonces la correlación será muy baja.

El procedimiento para encontrar el vector característico hace uso del análisis de regresión múltiple. Sea a_i el valor específico de la característica que posee el objeto i , con $i = 1, \dots, n$, y sean $x_{i1}, x_{i2}, \dots, x_{ik}$ las coordenadas del punto que lo representa, entonces la ecuación de regresión múltiple ordinaria es:

$$a_i \approx \hat{a}_i = b_0 + b_1 x_{i1} + \dots + b_k x_{ik}$$

A los valores b_1, b_2, \dots, b_k se les llama coeficientes de regresión y al término b_0 la intercepción. El valor de \hat{a}_i es el mejor estimador para la proyección del objeto con coordenadas x_{ij} , $j = 1, \dots, k$, sobre el vector característico.

Se calcula el coeficiente de correlación múltiple entre las proyecciones y los valores de la característica, y si su valor es bajo entonces se puede concluir con seguridad que los sujetos no utilizaron la variable en cuestión cuando hicieron sus juicios de similitud.

Para graficar el vector característico, primero se calculan los valores estimados de los coeficientes de regresión, denotándolos por $\beta_1, \beta_2, \dots, \beta_k$. En

seguida, se determina en el espacio de la configuración el punto

$$\beta^* = (\beta_1, \beta_2, \dots, \beta_k).$$

Finalmente, suponiendo que la media de las coordenadas en cada dimensión es cero, se dibuja una línea que parte del origen y pasa por el punto β^* . Se acostumbra que la longitud de la línea sea proporcional al cuadrado del coeficiente de regresión (aunque en un sentido estricto la longitud es arbitraria), y poner una flecha en el punto final del vector.

Los coeficientes $\beta_1, \beta_2, \dots, \beta_k$ son los cosenos directores del vector característico. Después de una transformación inversa del coseno se convierten en los ángulos (θ_1 y θ_2) entre el vector característico L y los ejes coordenados.

Estos ángulos se pueden utilizar para hacer una rotación de la solución, de manera que el vector característico quede como una de las dimensiones del espacio.

1.4.3. Análisis de Correlación Canónica

En la sección previa se consideró ajustar un vector característico en el espacio de la configuración. La discusión se limita al caso que una sola variable puede ser utilizada para interpretar la solución. En la mayoría de las aplicaciones es razonable esperar que se pueda obtener un mayor número de variables para hacerlo.

Cuando se tienen varias características para interpretar la solución, se puede obtener un vector característico para cada una, pero de esta manera se ignoran las interrelaciones entre ellas. En tal caso, se requiere un procedimiento que permita relacionar simultáneamente varias variables con los puntos de la configuración. Tal procedimiento se llama análisis de correlación canónica.

Dados dos conjuntos de variables X y Y , el análisis de correlación canónica busca una asociación lineal entre los dos conjuntos. El proceso consiste en determinar dos combinaciones lineales, una para X y otra para Y , de tal manera que la correlación entre las dos combinaciones sea tan grande como sea posible.

En el contexto del análisis de escalamiento multidimensional, el conjunto Y consta de los valores de cada una de las variables para cada objeto y el conjunto X contiene las coordenadas de los puntos en cada dimensión. La correlación es entre la suma ponderada de los valores de las variables y las proyecciones de los puntos sobre el vector característico canónico.

1.4.4. Análisis de Conglomerados

Un procedimiento de gran utilidad para interpretar las soluciones del MDS es buscar los agrupamientos de objetos. Estos agrupamientos indicarían conjuntos de objetos muy semejantes entre sí y diferentes a los demás, y pueden ser de utilidad si la finalidad principal de nuestro análisis es la clasificación. El análisis de conglomerados genera agrupamientos jerárquicos de los objetos en función de su proximidad. Esta agrupación se hace de tal modo que aquellos objetos más similares entre sí formarán parte de un mismo conglomerado. A medida que la proximidad vaya disminuyendo, otros objetos u otros conglomerados se irán uniendo a esta estructura jerárquica hasta que, finalmente, todos los objetos pertenezcan a un único conglomerado. Si completamos la información proporcionada por el MDS con la información sobre agrupamientos proporcionada por el análisis de conglomerados, nos será más sencillo identificar grupos de objetos con características semejantes, así como el número de grupos que existen.

1.4.5. Análisis con Datos Externos

También puede interpretarse la configuración de objetos averiguando si existe alguna relación entre la disposición espacial de los objetos y alguna o algunas características de éstos que hayan sido medidas independientemente. Estas características se utilizarían como “datos externos” (Arce, 1993) que pueden ponerse en relación con las coordenadas de los objetos en las distintas dimensiones mediante regresión múltiple. Cada característica sería utilizada como variable de respuesta en un análisis de regresión múltiple, mientras que las coordenadas de los objetos en las distintas dimensiones serían las variables predictoras. Aquellas características relacionadas con la solución proporcionada por el MDS mostrarán un coeficiente de regresión múltiple elevado y significativo, y un peso importante y significativo en alguna o algunas de las dimensiones, lo que indicaría una elevada relación entre esa característica y la dimensión o las dimensiones correspondientes. Este procedimiento puede resultar de ayuda cuando no resulta sencillo interpretar las posiciones de los objetos en el espacio directamente, o bien cuando existen varias posibilidades que compiten en la interpretación.

1.5. Ejemplo

La tabla 3 contiene diversos índices socioeconómicos y demográficos de 14 países (Alemania, Australia, España, Estados Unidos, Francia, Grecia, Islandia, Italia, Japón, Nueva Zelanda, Portugal, Reino Unido, Suecia y Turquía). Para cada país se han obtenido 7 índices socioeconómicos distintos: densidad de población (depo), porcentaje de empleo agrícola (emag), ingresos nacionales per cápita (inna), inversión del rendimiento del capital en maquinaria (inrc), tasa de mortalidad infantil (moin), consumo de energía por cada 100 habitantes (ener), y número de aparatos de televisión por cada 100 habitantes (aptv).

Tabla 3. Índices socioeconómicos y demográficos de 14 países.

Países	depo	emag	inna	inrc	moin	ener	aptv
Alemania	247	6	12.4	9.1	15	4.6	33
Australia	2	6	8.4	10.1	12	5.2	36
España	74	19	5.3	6.9	15	2.0	21
Estados Unidos	24	4	10.6	7.3	13	8.7	62
Francia	97	9	10.7	9.2	10	3.7	28
Grecia	72	31	4.1	8.1	19	1.7	12
Islandia	2	13	11.0	6.6	11	5.8	25
Italia	189	15	5.7	7.9	15	2.5	22
Japón	311	11	8.7	10.9	8	3.3	24
Nueva Zelanda	12	10	6.8	8.0	14	3.4	26
Portugal	107	31	2.1	5.5	39	1.1	9
Reino Unido	229	3	7.2	9.3	13	3.9	39
Suecia	18	6	12.8	7.2	7	6.3	37
Turquía	56	61	1.6	8.8	153	0.7	6

Dado que nuestros datos originales no son proximidades, tendremos que obtener a partir de ellos un coeficiente de disimilaridades entre países. Un procedimiento sencillo consiste en obtener la matriz de correlaciones entre los países, y luego transformar éstas en disimilaridades.

La matriz de correlaciones R se muestra en la tabla 4. Si a_{rs} corresponde al índice r del país s , la correlación entre el país r y el país s , r_{rs} , es

$$r_{rs} = \frac{\sum_{i=1}^7 (a_{ri} - a_{r.})(a_{si} - a_{s.})}{\sqrt{\sum_{i=1}^7 (a_{ri} - a_{r.})^2} \sqrt{\sum_{i=1}^7 (a_{si} - a_{s.})^2}}$$

Tabla 4. Matriz de correlaciones entre países.

Países	Al	Au	Es	E.U.	Fr	Gr	Is	It	Ja	N.Z.	Po	R.U.	Su	Tu
Al	1	-0,265	0,970	0,220	0,991	0,908	-0,430	0,997	0,998	-0,344	0,914	0,998	0,275	0,114
Au	-0,265	1	-0,183	0,867	-0,146	-0,340	0,908	-0,293	-0,316	0,950	-0,366	-0,222	0,817	-0,173
Es	0,970	-0,183	1	0,274	0,976	0,966	-0,294	0,979	0,967	-0,203	0,952	0,971	0,316	0,232
E.U.	0,220	0,867	0,274	1	0,334	0,074	0,705	0,187	0,168	0,756	0,048	0,263	0,979	-0,195
Fr	0,991	-0,146	0,976	0,334	1	0,897	-0,314	0,986	0,984	-0,229	0,891	0,995	0,390	0,081
Gr	0,908	-0,340	0,966	0,074	0,897	1	-0,375	0,934	0,917	-0,295	0,977	0,901	0,113	0,356
Is	-0,430	0,908	-0,294	0,705	-0,314	-0,375	1	-0,440	-0,470	0,946	-0,436	-0,398	0,678	-0,100
It	0,997	-0,293	0,979	0,187	0,986	0,934	-0,440	1	0,998	-0,353	0,936	0,995	0,239	0,153
Ja	0,998	-0,316	0,967	0,168	0,984	0,917	-0,470	0,998	1	-0,390	0,918	0,994	0,226	0,108
N.Z.	-0,344	0,950	-0,203	0,756	-0,229	-0,295	0,946	-0,353	-0,390	1	-0,318	-0,306	0,686	0,076
Po	0,914	-0,366	0,952	0,048	0,891	0,977	-0,436	0,936	0,918	-0,318	1	0,906	0,065	0,471
R.U.	0,998	-0,222	0,971	0,263	0,995	0,901	-0,398	0,995	0,994	-0,306	0,906	1	0,313	0,102
Su	0,275	0,817	0,316	0,979	0,390	0,113	0,678	0,239	0,226	0,686	0,065	0,313	1	-0,300
Tu	0,114	-0,173	0,232	-0,195	0,081	0,356	-0,100	0,153	0,108	0,076	0,471	0,102	-0,300	1

Como se observa en la tabla 4, existen correlaciones altas entre algunos países, como la existente entre Alemania y Reino Unido (0.998), o entre Estados Unidos y Suecia (0.979), y correlaciones casi nulas, como la existente entre Estados Unidos y Portugal (0.048), o entre Francia y Turquía (0.081). Sin embargo, resulta muy difícil interpretar los resultados directamente a partir de la matriz de correlaciones, en esto nos ayudará el modelo de MDS.

Debemos transformar las correlaciones en disimilaridades. Una de las más conocidas transformaciones esta dada por

$$\delta_{rs} = \sqrt{2(1 - r_{rs})}$$

La matriz de disimilaridades D se muestra en la tabla 5.

A partir de la matriz de disimilaridades D obtenemos la matriz de producto interno centrado B , la cual se muestra en la tabla 6. Recordemos que los elementos b_{rs} de la matriz de producto interno centrado B se obtienen de la siguiente manera:

$$b_{rs} = -\frac{1}{2}(\delta_{rs}^2 - \delta_{r.}^2 - \delta_{.s}^2 + \delta_{..}^2)$$

donde

$$\begin{aligned}\delta_{r.}^2 &= \frac{1}{14} \sum_{s=1}^{14} \delta_{rs}^2 \\ \delta_{.s}^2 &= \frac{1}{14} \sum_{r=1}^{14} \delta_{rs}^2 \\ \delta_{..}^2 &= \frac{1}{(14)(14)} \sum_{r=1}^{14} \sum_{s=1}^{14} \delta_{rs}^2\end{aligned}$$

Sabemos que la matriz de coordenadas de los objetos X es tal que $B = XX^t$. La matriz B puede ser definida en términos de su descomposición espectral

$$B = V\Lambda V^t$$

donde $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{14})$, matriz diagonal de valores propios $\{\lambda_i\}$ de B , y $V = [v_1, v_2, \dots, v_{14}]$, matriz de vectores propios de B , normalizados tal que $v_i^t v_i = 1$. Por conveniencia los valores propios de B serán tales que $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{14} \geq 0$.

Tabla 5. Matriz de disimilaridades entre países.

Países	Al	Au	Es	E.U.	Fr	Gr	Is	It	Ja	N.Z.	Po	R.U.	Su	Tu
Al	0	1,591	0,244	1,248	0,130	0,428	1,691	0,071	0,062	1,639	0,414	0,048	1,203	1,331
Au	1,591	0	1,538	0,514	1,514	1,637	0,427	1,608	1,622	0,313	1,653	1,563	0,604	1,531
Es	0,244	1,538	0	1,204	0,216	0,259	1,609	0,200	0,255	1,551	0,308	0,238	1,168	1,238
E.U.	1,248	0,514	1,204	0	1,154	1,360	0,767	1,274	1,289	0,698	1,379	1,214	0,203	1,546
Fr	0,130	1,514	0,216	1,154	0	0,452	1,621	0,163	0,177	1,568	0,465	0,097	1,104	1,355
Gr	0,428	1,637	0,259	1,360	0,452	0	1,658	0,361	0,405	1,609	0,213	0,443	1,331	1,134
Is	1,691	0,427	1,609	0,767	1,621	1,658	0	1,697	1,715	0,328	1,694	1,672	0,802	1,483
It	0,071	1,608	0,200	1,274	0,163	0,361	1,697	0	0,060	1,645	0,356	0,099	1,233	1,300
Ja	0,062	1,622	0,255	1,289	0,177	0,405	1,715	0,060	0	1,667	0,403	0,101	1,243	1,335
N.Z.	1,639	0,313	1,551	0,698	1,568	1,609	0,328	1,645	1,667	0	1,623	1,616	0,791	1,358
Po	0,414	1,653	0,308	1,379	0,465	0,213	1,694	0,356	0,403	1,623	0	0,433	1,367	1,027
R.U.	0,048	1,563	0,238	1,214	0,097	0,443	1,672	0,099	0,101	1,616	0,433	0	1,171	1,339
Su	1,203	0,604	1,168	0,203	1,104	1,331	0,802	1,233	1,243	0,791	1,367	1,171	0	1,613
Tu	1,331	1,531	1,238	1,546	1,355	1,134	1,483	1,300	1,335	1,358	1,027	1,339	1,613	0

Tabla 6. Matriz de producto interno centrado B .

Países	Al	Au	Es	E.U.	Fr	Gr	Is	It	Ja	N.Z.	Po	R.U.	Su	Tu
Al	0,346	-0,554	0,275	-0,314	0,302	0,277	-0,628	0,345	0,361	-0,613	0,288	0,333	-0,267	-0,152
Au	-0,554	1,076	-0,513	0,697	-0,469	-0,606	1,075	-0,579	-0,587	1,046	-0,627	-0,523	0,639	-0,074
Es	0,275	-0,513	0,263	-0,301	0,246	0,293	-0,534	0,286	0,289	-0,514	0,285	0,264	-0,267	-0,075
E.U.	-0,314	0,697	-0,301	0,584	-0,235	-0,437	0,626	-0,345	-0,348	0,605	-0,458	-0,283	0,555	-0,343
Fr	0,302	-0,469	0,246	-0,235	0,276	0,231	-0,547	0,300	0,313	-0,534	0,230	0,294	-0,187	-0,220
Gr	0,277	-0,606	0,293	-0,437	0,231	0,391	-0,551	0,305	0,303	-0,543	0,373	0,258	-0,407	0,112
Is	-0,628	1,075	-0,534	0,626	-0,547	-0,551	1,256	-0,637	-0,652	1,131	-0,607	-0,608	0,590	0,088
It	0,345	-0,579	0,286	-0,345	0,300	0,305	-0,637	0,350	0,363	-0,621	0,312	0,331	-0,301	-0,110
Ja	0,361	-0,587	0,289	-0,348	0,313	0,303	-0,652	0,363	0,381	-0,642	0,310	0,346	-0,299	-0,140
N.Z.	-0,613	1,046	-0,514	0,605	-0,534	-0,543	1,131	-0,621	-0,642	1,114	-0,560	-0,588	0,527	0,194
Po	0,288	-0,627	0,285	-0,458	0,230	0,373	-0,607	0,312	0,310	-0,560	0,401	0,268	-0,449	0,233
R.U.	0,333	-0,523	0,264	-0,283	0,294	0,258	-0,608	0,331	0,346	-0,588	0,268	0,322	-0,241	-0,175
Su	-0,267	0,639	-0,267	0,555	-0,187	-0,407	0,590	-0,301	-0,299	0,527	-0,449	-0,241	0,567	-0,456
Tu	-0,152	-0,074	-0,075	-0,343	-0,220	0,112	0,088	-0,110	-0,140	0,194	0,233	-0,175	-0,456	1,120

Por tanto la matriz de coordenadas de objetos X puede estar definida como

$$X = V\Lambda^{\frac{1}{2}}$$

Los valores propios asociados a la matriz de productos escalares están dados en la tabla 7.

Tabla 7. Valores propios asociados a B .

λ_1	6,48943	λ_8	0,00011
λ_2	1,67727	λ_9	0,00009
λ_3	0,16901	λ_{10}	0,00005
λ_4	0,07131	λ_{11}	0,00002
λ_5	0,03740	λ_{12}	-0,00004
λ_6	0,00828	λ_{13}	-0,00008
λ_7	0,00015	λ_{14}	-0,00011

El problema es que algunos de los valores propios son negativos. Hay dos formas de resolver esto: la primera es ignorar los valores propios y por consiguiente sus vectores propios correspondientes, o la segunda forma es agregar una constante apropiada c a las disimilaridades, $\delta_{rs}^* = \delta_{rs} + c(1 - \delta^{rs})$ (con $\delta^{rs} = 1$ si $r = s$ y $\delta^{rs} = 0$ si $r \neq s$), y definir nuevamente la matriz de productos escalares a partir de la nueva matriz de disimilaridades.

Lo que utilizaremos aquí será sumar la constante:

$$c = \max_{(r,s,t)} (\delta_{rs} - \delta_{rt} - \delta_{ts}) = 1,7151,$$

de tal manera que las distancias d_{rs} serán exactamente iguales a las disimilaridades δ_{rs}^* , la cual se muestra en la tabla 8.

La nueva matriz de producto interno centrado B^* esta en la tabla 9. Las tablas 10 y 11 muestran las matrices de valores propios y vectores propios asociados a la matriz de producto interno B^* :

$$B^* = V_1\Lambda_1V_1^t$$

donde Λ_1 es la matriz diagonal de valores propios asociados a la matriz de productos escalares y V_1 es la matriz de los vectores propios correspondientes.

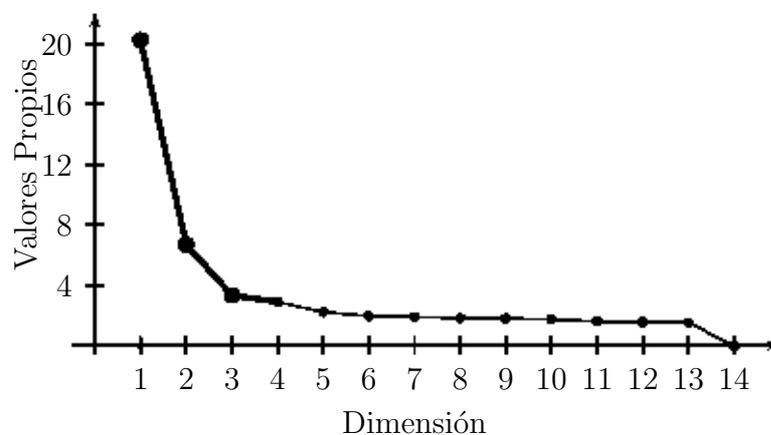
La figura 1.3 muestra claramente un codo al final de la dimensión 3, lo cual implica que la dimensión adecuada es $k = 2$.

Tabla 8. Matriz de distancias entre países.

Países	Al	Au	Es	E.U.	Fr	Gr	Is	It	Ja	N.Z.	Po	R.U.	Su	Tu
Al	0	3,306	1,959	2,963	1,845	2,143	3,406	1,786	1,777	3,354	2,129	1,763	2,198	3,046
Au	3,306	0	3,253	2,229	3,229	3,352	2,142	3,323	3,337	2,028	3,368	3,279	2,319	3,246
Es	1,959	3,253	0	2,919	1,931	1,974	3,324	1,915	1,970	3,266	2,023	1,953	2,884	2,954
E.U.	2,963	2,229	2,919	0	2,869	3,075	2,482	2,989	3,004	2,413	3,094	2,929	1,918	3,261
Fr	1,845	3,229	1,931	2,869	0	2,167	3,336	1,878	1,892	3,283	2,181	1,812	2,819	3,070
Gr	2,143	3,352	1,974	3,075	2,167	0	3,373	2,076	2,120	3,325	1,928	2,158	3,046	2,849
Is	3,406	2,142	3,324	2,482	3,336	3,373	0	3,412	3,430	2,043	3,410	3,387	2,517	3,198
It	1,786	3,323	1,915	2,989	1,878	2,076	3,412	0	1,775	3,360	2,071	1,814	2,948	3,015
Ja	1,777	3,337	1,970	3,004	1,892	2,120	3,430	1,775	0	3,382	2,118	1,816	2,958	3,050
N.Z.	3,354	2,028	3,266	2,413	3,283	3,325	2,043	3,360	3,382	0	3,338	3,331	2,506	3,073
Po	2,129	3,368	2,023	3,094	2,181	1,928	3,410	2,071	2,118	3,338	0	2,148	3,082	2,743
R.U.	1,763	3,279	1,953	2,929	1,812	2,158	3,387	1,814	1,816	3,331	2,148	0	2,886	3,054
Su	2,918	2,319	2,884	1,918	2,819	3,046	2,517	2,948	2,958	2,506	3,082	2,886	0	3,328
Tu	3,046	3,246	2,954	3,261	3,070	2,849	3,198	3,015	3,050	3,073	2,743	3,054	3,328	0

Tabla 9. Matriz de producto interno centrado B^* .

Países	Al	Au	Es	E.U.	Fr	Gr	Is	It	Ja	N.Z.	Po	R.U.	Su	Tu
Al	2,628	-1,734	0,658	-1,184	0,879	0,499	-1,853	1,030	1,094	-1,842	0,540	1,053	-1,063	-0,706
Au	-1,734	4,832	-1,612	1,822	-1,529	-1,720	2,754	-1,794	-1,792	2,829	-1,762	-1,665	1,608	-0,235
Es	0,658	-1,612	2,528	-1,105	0,667	0,797	-1,626	0,741	0,682	-1,600	0,710	0,650	-1,012	-0,479
E.U.	-1,184	1,822	-1,105	3,784	-0,954	-1,354	1,446	-1,264	-1,260	1,450	-1,402	-1,103	1,934	-0,807
Fr	0,879	-1,529	0,667	-0,954	2,538	0,402	-1,662	0,818	0,837	-1,650	0,383	0,920	-0,822	-0,826
Gr	0,499	-1,720	0,797	-1,354	0,402	2,965	-1,573	0,638	0,593	-1,574	1,115	0,446	-1,276	0,040
Is	-1,853	2,754	-1,626	1,446	-1,662	-1,573	5,269	-1,875	-1,888	3,017	-1,686	-1,808	1,348	0,138
It	1,030	-1,794	0,741	-1,264	0,816	0,638	-1,875	2,623	1,096	-1,862	0,658	0,960	-1,152	-0,616
Ja	1,094	-1,792	0,682	-1,26	0,837	0,593	-1,888	1,096	2,720	-1,889	0,609	1,004	-1,135	-0,672
N.Z.	-1,842	2,829	-1,600	1,450	-1,650	-1,574	3,017	-1,862	-1,889	4,941	-1,609	-1,784	1,210	0,365
Po	0,540	-1,762	0,710	-1,402	0,383	1,115	-1,686	0,658	0,609	-1,609	2,986	0,479	-1,374	0,350
R.U.	1,053	-1,665	0,650	-1,103	0,920	0,446	-1,808	0,960	1,004	-1,784	0,479	2,588	-0,989	-0,752
Su	-1,063	1,608	-1,012	1,934	-0,822	-1,276	1,348	-1,152	-1,135	1,210	-1,374	-0,989	3,763	-1,037
Tu	-0,706	-0,235	-0,479	-0,807	-0,826	0,040	0,138	-0,616	-0,672	0,365	0,350	-0,752	-1,037	5,238

Figura 1.3: λ vs. k

Por lo tanto, la matriz de coordenadas de los países estará dada por

$$X = V_1 \Lambda_1^{\frac{1}{2}}$$

la cual se muestra en la tabla 12.

Tabla 10. Valores propios asociados a B^* .

λ_1	20,2492	λ_8	1,8253
λ_2	6,7055	λ_9	1,8073
λ_3	3,3457	λ_{10}	1,7473
λ_4	2,8993	λ_{11}	1,6097
λ_5	2,2326	λ_{12}	1,5672
λ_6	1,9696	λ_{13}	1,5358
λ_7	1,9140	λ_{14}	0

En la figura 1.4 podemos apreciar que aquellos países entre los que existían altas correlaciones se encuentran próximos entre sí, como es el caso Alemania y Reino Unido o Estados Unidos y Suecia. Por su parte, aquellos países cuyas correlaciones eran casi nulas se encuentran alejados entre sí, como es el caso de Estados Unidos y Portugal o Francia y Turquía.

Tabla 11. Matriz de vectores propios asociados a la matriz de producto interno centrado B^* .

V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14
-0,229	-0,124	-0,081	0,268	0,013	0,130	-0,117	0,058	-0,022	0,031	0,394	-0,110	0,759	0,267
0,396	-0,061	-0,142	0,066	-0,598	0,362	0,496	0,008	0,083	-0,008	0	0,003	0,004	0,267
-0,202	-0,036	-0,043	-0,225	-0,005	-0,510	0,400	-0,152	-0,055	-0,588	0,154	-0,126	0	0,267
0,270	-0,278	0,489	-0,069	-0,041	-0,032	-0,153	-0,225	-0,671	0,077	-0,022	-0,008	0,006	0,267
-0,198	-0,160	-0,041	0,195	0,039	-0,341	0,237	-0,293	0,178	0,521	-0,497	-0,035	0,100	0,267
-0,209	0,131	-0,055	-0,599	-0,014	-0,011	0,122	0,550	-0,178	0,390	0,039	0,026	0,012	0,267
0,406	0,080	-0,442	-0,059	0,685	0,188	0,132	-0,149	-0,110	0,003	0,007	0,004	-0,004	0,267
-0,235	-0,088	-0,093	0,149	0	0,137	-0,109	0,158	-0,105	-0,314	-0,329	0,748	-0,006	0,267
-0,237	-0,107	-0,100	0,221	0,014	0,255	-0,178	0,262	-0,110	-0,220	-0,328	-0,623	-0,279	0,267
0,397	0,123	-0,332	0,024	-0,284	-0,485	-0,550	0,121	0,097	0,001	-0,005	-0,007	0	0,267
-0,215	0,189	0,022	-0,460	-0,135	0,340	-0,326	-0,567	0,259	-0,037	-0,026	-0,027	-0,006	0,267
-0,220	-0,140	-0,066	0,272	0,006	-0,015	-0,018	-0,088	0,036	0,251	0,594	0,142	-0,578	0,267
0,248	-0,333	0,487	-0,098	0,254	0,007	-0,037	0,264	0,601	-0,099	0,023	0,014	-0,007	0,267
0,029	0,806	0,399	0,313	0,066	-0,027	0,105	0,052	-0,002	-0,008	-0,005	0	-0,001	0,267

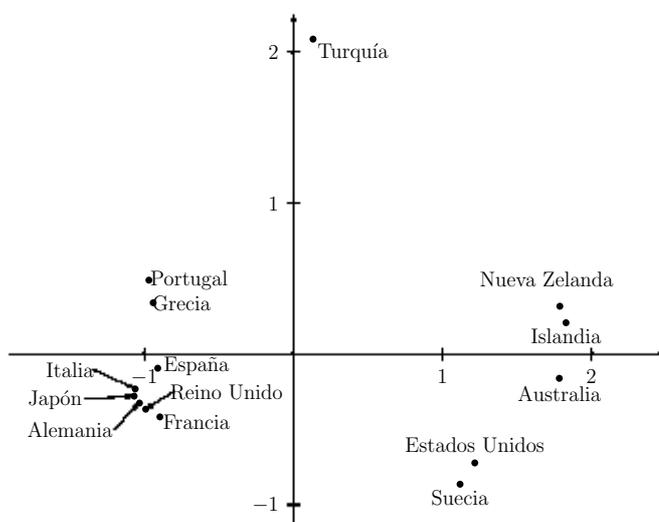


Figura 1.4: Configuración del MDS

Tabla 12. Coordenadas de los países.

Países	Dimensión I	Dimensión II
Alemania	-1,0324	-0,3234
Australia	1,7844	-0,1600
España	-0,9103	-0,0939
Estados Unidos	1,2176	-0,7202
Francia	-0,8951	-0,4163
Grecia	-0,9415	0,3412
Islandia	1,8288	0,2087
Italia	-1,0613	-0,2300
Japón	-1,0688	-0,2785
Nueva Zelanda	1,7885	0,3187
Portugal	-0,9692	0,4913
Reino Unido	-0,9910	-0,3645
Suecia	1,1178	-0,8623
Turquía	0,1325	2,0892

Se observan claramente dos agrupaciones de países. El primero está formado por Australia, Estados Unidos, Islandia, Nueva Zelanda y Suecia. El

segundo está formado por Alemania, España, Francia, Grecia, Italia, Japón, Portugal y Reino Unido. Turquía aparece aislado, separado de ambos grupos. Un segundo examen nos revelará que los dos grupos de países se alinean a lo largo de la dimensión I, mientras que ambos grupos se separan de Turquía a lo largo de la dimensión II.

Si examinamos los índices socioeconómicos de cada país veremos que los países del primer grupo tienen baja densidad de población, mientras que el segundo corresponde a países con alta densidad de población, por lo que la dimensión I parece corresponder con esta variable. Los países situados del lado positivo de esta dimensión tienen menores densidades poblacionales, mientras que los situados del lado negativo están más densamente poblados.

Por lo que se refiere a la dimensión II, si de nuevo examinamos los índices socioeconómicos, veremos que los países situados del lado positivo tienen mayor número de personas empleadas en la agricultura, menor renta per cápita, menor consumo de energía, mayores tasas de mortalidad infantil, así como menor número de aparatos de TV. Todas estas variables están relacionadas con el desarrollo económico y tecnológico, de tal modo que los países situados en el lado positivo de la dimensión II se encuentran menos desarrollados que los países situados en el lado negativo.

La figura 1.5 muestra los vectores característicos asociados a cada uno de los siete índices socioeconómicos y demográficos. Estos vectores nos pueden ayudar con la interpretación de la configuración obtenida por el MDS. Por ejemplo, si observamos el vector asociado a MOIN (tasa de mortalidad infantil), notaremos que Turquía tiene una alta tasa de mortalidad infantil pues existe una fuerte correlación con el vector asociado, mientras que Suecia tiene una baja tasa de mortalidad infantil.

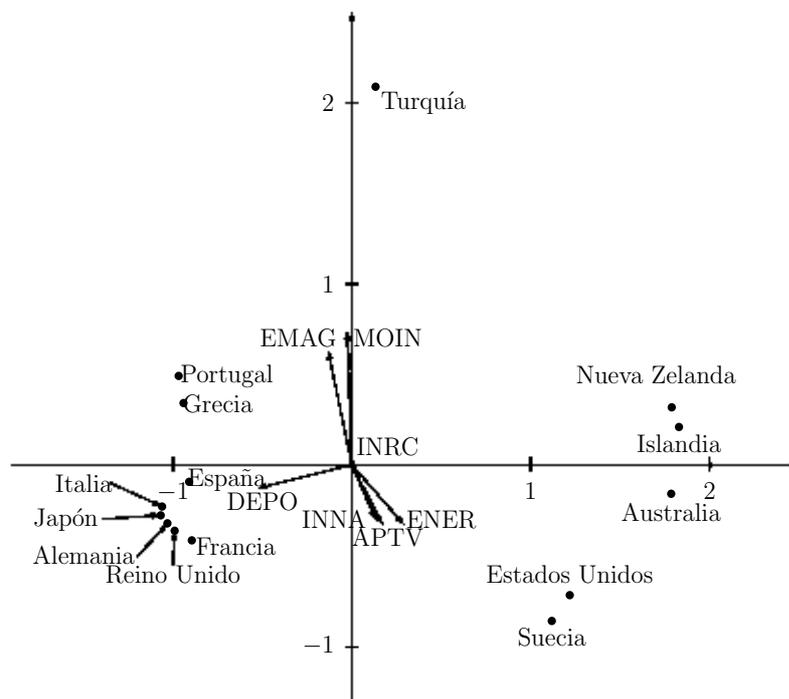


Figura 1.5: Configuración del MDS

Capítulo 2

Escalamiento Multidimensional No Métrico

Shepard (1962) propuso un método para estimar coordenadas de objetos bajo supuestos menos restrictivos que los de Torgerson (1952). Posteriormente Kruskal (1964) mejoró el modelo de Shepard, proporcionando el fundamento teórico del análisis de escalamiento multidimensional no métrico¹. Fue precisamente Kruskal quien le dió el nombre a este análisis.

El MDS no métrico supone que los datos están medidos en una escala ordinal. De aquí que la hipótesis fundamental del MDS no métrico es que las medidas de proximidad están relacionadas con las distancias entre los puntos mediante una función monótona (ordenada), a diferencia del modelo métrico que supone una forma funcional exacta. En otras palabras, a diferencia del modelo métrico, el modelo de escalamiento multidimensional no métrico no presupone una relación de tipo lineal entre las proximidades y distancias, sino simplemente una relación de tipo monótona creciente, que puede corresponder a una línea recta, curva o quebrada. La única condición es que, a medida que aumentan los valores de las proximidades, los de las distancias también aumentan o, en todo caso, no disminuyen. Es decir, la relación no se establece en forma de sumas o productos, sino simplemente en términos de la preservación del orden de las proximidades. Por este motivo se etiqueta este modelo como no métrico.

El algoritmo de MDS no métrico calcula las estimaciones de las coordena-

¹A partir de aquí MDS no métrico. Su nombre en inglés es *Nonmetric Multidimensional Scaling* o *Nonparametric Multidimensional Scaling*.

das de los objetos \hat{x}_r en un número preespecificado de dimensiones p , así que cuando las distancias estimadas $\{\hat{d}_{rs}\}$ sean calculadas a partir de esas coordenadas estimadas, el orden del rango de las distancias estimadas concuerde con el orden del rango de los datos originales $\{\delta_{rs}\}$ tanto como sea posible.

El modelo de Shepard sólo requiere que los datos estén relacionados monótonamente a las distancias, es decir, el método de Shepard para estimar las coordenadas de los objetos supone que las medidas de disimilaridad y las distancias están relacionadas mediante una función monótona. Específicamente, su algoritmo supone que

$$\delta_{rs} = f(d_{rs}) = f\left(\left[\sum_{i=1}^p (x_{ri} - x_{si})^2\right]^{1/2}\right)$$

Aquí f es una función monótona tal que

$$d_{rs} < d_{r's'} \Rightarrow f(d_{rs}) < f(d_{r's'}) \quad \text{para todo } r, r', s, \text{ y } s'.$$

Ejemplos de este tipo de funciones son las lineales, potencia, exponencial, y logarítmica.

A diferencia del algoritmo de Shepard, el de Kruskal permite que las distancias sean euclidianas o no, siempre que pertenezcan a la métrica de Minkowski.

2.1. Espacios \mathfrak{R}^p y la Métrica de Minkowski

Aunque el MDS no métrico puede ser aplicado en espacios abstractos, la mayoría de los análisis de MDS se aplican con X un subconjunto de \mathfrak{R}^p , y con $p = 2$ en particular. Se busca una configuración de puntos en \mathfrak{R}^p la cual representa los objetos originales, tal que las distancias entre los puntos $\{d_{rs}\}$ estén igualmente ordenadas, tanto como sea posible, como las disimilaridades originales $\{\delta_{rs}\}$.

La medida de distancia para X es la métrica de Minkowski o función de distancia de Minkowski, y así para los puntos r y s en X ,

$$d_{rs} = \left(\sum_{i=1}^p |x_{ri} - x_{si}|^\lambda\right)^{1/\lambda} \quad (2.1)$$

Varios casos especiales de esta función son de interés particular, y cada caso corresponde a un valor diferente de λ . Primero, si $\lambda = 2$, entonces la ecuación (2.1) se reduce a la función de distancia euclidiana estándar:

$$d_{rs} = \left(\sum_{i=1}^p |x_{ri} - x_{si}|^2 \right)^{1/2} \quad (2.2)$$

El segundo caso especial es cuando λ se aproxima a ∞ , la ecuación (2.1) se aproxima a la métrica *dominance* en la cual la distancia entre los objetos r y s está determinada por la diferencia entre las coordenadas sólo a lo largo de una dimensión, dimensión en la cual este valor $|x_{ri} - x_{si}|$ es el mayor. Esto es,

$$d_{rs} = \max_i |x_{ri} - x_{si}| \quad (2.3)$$

En general, una dimensión determina la distancia d_{rs} que varía de un par de objetos a otro.

El tercer caso de interés especial es la métrica *city-block* en la cual $\lambda = 1$. En el modelo de *city-block*,

$$d_{rs} = \sum_{i=1}^p |x_{ri} - x_{si}|. \quad (2.4)$$

Esta es llamada la métrica *city-block* porque, en dos dimensiones, la distancia entre los puntos r y s equivale a la distancia recorrida desde el punto que corresponde al objeto r al punto que corresponde al objeto s a lo largo de las calles de una ciudad recorridas paralelamente a los ejes coordenados.

Si el espacio es unidimensional para un conjunto de objetos, d_{rs} será igual sin importar qué valor de λ se emplee en la función de distancia de Minkowski. En más de dos dimensiones, las mismas coordenadas producirán diferentes distancias dependiendo de si esas distancias se basan en el modelo euclidiano de la ecuación (2.2), el modelo *dominance* de la ecuación (2.3), o el modelo *city-block* de la ecuación (2.4).

El algoritmo de Kruskal (1964) permite al usuario estimar las coordenadas de los objetos a partir de los datos de modo que:

$$\delta_{rs} = f(d_{rs}) = f \left(\left(\sum_{i=1}^p |x_{ri} - x_{si}|^\lambda \right)^{1/\lambda} \right)$$

donde f es otra vez una función monótona que satisface

$$d_{rs} < d_{r's'} \Rightarrow f(d_{rs}) < f(d_{r's'}) \quad \text{para todo } r, r', s, \text{ y } s'. \quad (2.5)$$

El valor de λ no es estimado por el algoritmo y en general lo especifica el usuario.

¿Qué valor de λ especifican los usuarios? Algunos investigadores usan las medidas de ajuste para guiar su elección de λ . Estos autores ajustan modelos para varias distancias de Minkowski a sus datos, variando sólo los valores de λ , y deciden a favor del modelo (valor de λ) que proporcione el mejor ajuste de los datos. Cuando la dimensión p es conocida, entonces la medida ajustada puede ser usada para indicar el mejor valor de λ . Si p es desconocida, la medida estimada puede ser un indicador que lleve a conclusiones erróneas del mejor valor de λ .

2.2. Medidas de Bondad de Ajuste

Una característica importante del algoritmo de Kruskal es la medida de bondad de ajuste que él propuso para calcular el grado para el cual las coordenadas estimadas de los objetos \hat{x}_r y \hat{x}_s reproducen el orden del rango de los datos $\{\delta_{rs}\}$ (orden de las disimilaridades). La mayoría de los algoritmos no métricos, incluyendo el de Kruskal, calculan tres conjuntos de parámetros sobre los que se basa la medida de bondad de ajuste. El primero y más importante contiene las coordenadas estimadas \hat{x}_r de los puntos de la configuración. El segundo conjunto contiene las distancias estimadas entre los puntos, calculadas a partir de las coordenadas estimadas como sigue:

$$\hat{d}_{rs} = \left(\sum_{i=1}^p |\hat{x}_{ri} - \hat{x}_{si}|^\lambda \right)^{1/\lambda}$$

El tercer conjunto contiene parámetros “dummy” que Guttman (1968) llama imágenes de rango de los datos, y que otros llaman pseudodistancias o disparidades. La frase de Guttman, imágenes de rango, describe con mayor precisión estos parámetros porque reflejan el orden del rango de los datos. Las disparidades $\{\hat{\delta}_{rs}\}$ son valores calculados tan próximos a las distancias estimadas $\{\hat{d}_{rs}\}$ como sea posible, sujetas a la restricción de que se relacionen

monótonamente con los datos originales. Esto es, se igualan a las distancias estimadas tanto como sea posible, sujetas a la siguiente restricción monótona:

$$\delta_{rs} < \delta_{r's'} \Rightarrow \widehat{\delta}_{rs} \leq \widehat{\delta}_{r's'} \quad \forall r, r', s, s' \quad (2.6)$$

Este método no trabaja directamente con las disimilaridades $\{\delta_{rs}\}$ originales, sólo utiliza su relación de orden a través de las disparidades $\{\widehat{\delta}_{rs}\}$.

2.2.1. Ties (empates)

La expresión (2.6) no especifica la relación que debe haber entre las disparidades $\widehat{\delta}_{rs}$ y $\widehat{\delta}_{r's'}$, cuando sus correspondientes disimilaridades están empatadas, es decir, cuando $\delta_{rs} = \delta_{r's'}$. Hay dos formas de plantear la condición de monotonicidad en caso de que haya empates entre los datos. La primera forma es llamada comúnmente planteamiento primario (o *untied*) y consiste en no restringir la relación de orden de las disparidades, en este caso puede ocurrir que $\widehat{\delta}_{rs} < \widehat{\delta}_{r's'}$, $\widehat{\delta}_{rs} = \widehat{\delta}_{r's'}$, ó $\widehat{\delta}_{rs} > \widehat{\delta}_{r's'}$. La segunda forma es llamada planteamiento secundario (o *tied*) y consiste en restringir las disparidades a ser iguales en caso de que sus proximidades correspondientes lo sean, es decir, que se cumpla la condición adicional:

$$\delta_{rs} = \delta_{r's'} \Rightarrow \widehat{\delta}_{rs} = \widehat{\delta}_{r's'} \quad \forall r, r', s, s'.$$

La mayoría de los autores recomiendan el planteamiento primario por encima del planteamiento secundario, porque éste da un mejor ajuste para los datos. Si las disparidades son calculadas de acuerdo al planteamiento primario o secundario para empates, éstas jugarán un rol prominente en la definición de las tres medidas de ajuste empleadas por la mayoría de los algoritmos de MDS no métrico. Estas medidas de ajuste son llamadas Stress, S-Stress, y el coeficiente de alienación.

2.2.2. Stress

Kruskal usó las disparidades y las distancias para definir una medida de bondad de ajustada que él llamó Stress, pero que a menudo es llamada Stress fórmula uno (S_1) para distinguirla de una variación de esta misma medida llamada Stress fórmula dos (S_2):

$$S_1 = \left(\frac{\sum_{(r,s)} (\hat{\delta}_{rs} - \hat{d}_{rs})^2}{\sum_{(r,s)} \hat{d}_{rs}^2} \right)^{1/2} \quad (2.7)$$

$$S_2 = \left(\frac{\sum_{(r,s)} (\hat{\delta}_{rs} - \hat{d}_{rs})^2}{\sum_{(r,s)} (\hat{d}_{rs} - \hat{d}_{..})^2} \right)^{1/2} \quad (2.8)$$

donde

$$\hat{d}_{..} = \frac{1}{n^2} \sum_{(r,s)} \hat{d}_{rs}$$

en otras palabras, $\hat{d}_{..}$ es la media aritmética de las distancias estimadas.

S_1 y S_2 difieren sólo en la constante de normalización que aparece en el denominador en ambas fórmulas. S_1 usa la suma de cuadrados de las distancias estimadas. S_2 usa una cantidad proporcional a la varianza de las distancias estimadas, la suma de cuadrados de las desviaciones con respecto a la distancia media estimada.

Cuando la matriz de datos es una matriz simétrica que contiene las disimilaridades (o similaridades) entre los objetos, entonces es posible reducir varios problemas computacionales empleando la fórmula uno. Por otro lado, si los datos son preferencias, entonces la fórmula dos dará ventajas para algunos problemas.

El Stress mide cuánto se aproximan las distancias obtenidas a partir de X a las disparidades obtenidas de la transformación de esas distancias. Cuanto menor sea el valor de Stress, mejor será el ajuste encontrado entre distancias y disparidades. Es decir, el Stress no es propiamente un índice de bondad de ajuste, sino de “maldad” de ajuste. Su valor mínimo se encontrará, por tanto, en 0, cuando no exista diferencia entre distancias y disparidades. Su valor máximo no es estable, pero se sabe que su límite superior, para un número n de objetos es:

$$\sqrt{1 - \frac{2}{n}}$$

El Stress es invariante bajo cambios de escala (alargamientos o contracciones uniformes) y bajo movimientos rígidos (rotaciones, traslaciones y reflexiones), debido a que está normalizado.

2.2.3. S-Stress

Otra medida de bondad de ajuste es el S-Stress. Al igual que el Stress, el S-Stress tiene dos variaciones llamadas fórmula uno y dos. El S-Stress fórmula uno está definido como sigue:

$$SS_1 = \left[\frac{\sum_{(r,s)} \left(\widehat{\delta}_{rs}^2 - \widehat{d}_{rs}^2 \right)^2}{\sum_{(r,s)} \left(\widehat{d}_{rs}^2 \right)^2} \right]^{1/2} \quad (2.9)$$

El S-Stress fórmula uno y el Stress fórmula uno difieren sólo en que SS_1 , ecuación (2.9), está definida en términos de distancias cuadradas y disparidades cuadradas. S_1 en la ecuación (2.7), por el otro lado, está definida en términos de distancias y disparidades.

El S-Stress fórmula dos está definido como:

$$SS_2 = \left[\frac{\sum_{(r,s)} \left(\widehat{\delta}_{rs}^2 - \widehat{d}_{rs}^2 \right)^2}{\sum_{(r,s)} \left(\widehat{d}_{rs}^2 - \widehat{d}_{..}^2 \right)^2} \right]^{1/2} \quad (2.10)$$

donde

$$\widehat{d}_{..}^2 = \frac{1}{n^2} \sum_{(r,s)} \widehat{d}_{rs}^2$$

ésta es la distancia cuadrada media estimada. La definición del S-Stress fórmula dos, ecuación (2.10), es paralela al Stress fórmula dos, ecuación (2.8). Lo anterior está definido en términos de cuadrados para las distancias y disparidades.

La fórmula uno es recomendada cuando los datos son disimilaridades (o similaridades) pero no cuando son preferencias.

2.2.4. Coeficiente de Alienación

Guttman (1968) propuso una tercera medida de ajuste para el análisis no métrico, el coeficiente de alienación. Para definir el coeficiente de alienación, primero definió un coeficiente de monotonicidad,

$$\mu = \frac{\sum_{(r,s)} \widehat{\delta}_{rs} \widehat{d}_{rs}}{\left[\left(\sum_{(r,s)} \widehat{\delta}_{rs}^2 \right) \left(\sum_{(r,s)} \widehat{d}_{rs}^2 \right) \right]^{1/2}}$$

Esencialmente μ mide el grado de asociación ordinal entre las medidas originales y las distancias, siendo también una medida de ajuste.

El coeficiente de alienación está definido en términos de μ como sigue:

$$\kappa = (1 - \mu^2)^{1/2}$$

Al igual que el Stress y el S-Stress, un valor grande de κ indica un mal ajuste, un valor pequeño indica un buen ajuste. Al mejorar el ajuste del modelo no métrico para los datos, κ será bajo; aquí κ es una medida de “maldad” de ajuste.

2.2.5. Rsq

Además del Stress y S-Stress existe otro índice de ajuste del modelo a nuestros datos, la Rsq. Este índice es una correlación cuadrática entre las disparidades derivadas a partir de los datos originales, y las disparidades derivadas por el modelo de escalamiento, de modo que puede ser interpretado como la proporción de varianza en las disparidades que es explicada por las distancias. Formalmente:

$$Rsqr = \frac{\left[\sum_{(r,s)} (\hat{\delta}_{rs} - \hat{\delta}_{..}) (\hat{d}_{rs} - \hat{d}_{..}) \right]^2}{\left[\sum_{(r,s)} (\hat{\delta}_{rs} - \hat{\delta}_{..})^2 \right] \left[\sum_{(r,s)} (\hat{d}_{rs} - \hat{d}_{..})^2 \right]}$$

donde los subíndices punteados ($\hat{\delta}_{..}$ y $\hat{d}_{..}$) representan la media en el elemento correspondiente del subíndice.

Su interpretación es mucho más sencilla y directa que la del Stress y del S-Stress, y sus límites son fijos (mínimo de cero y máximo de uno).

2.3. Algoritmo

Hay gran variación en los algoritmos actualmente empleados para realizar escalamientos no métricos. El algoritmo presentado aquí da una idea de cómo operan los algoritmos.

Una vez elegida la medida de bondad de ajuste S , se define una solución en k dimensiones como la configuración que minimiza su valor. El problema de encontrar esta configuración óptima se puede abordar de tres formas:

- a) Intuitivamente se describe un método de aproximación. Comenzando con una configuración inicial, se mueven un poco los puntos para mejorar el ajuste, repitiendo el procedimiento hasta que la configuración no pueda ser mejorada.
- b) Formalmente significa acercar los puntos x_r y x_s si $\hat{\delta}_{rs} < \hat{d}_{rs}$ y alejarlos en caso contrario, de forma que $\hat{\delta}_{rs}$ se parezca cada vez más a \hat{d}_{rs} .
- c) A nivel teórico, se trata del problema de minimizar una función de varias variables. De hecho $S = S(x_1, x_2, \dots, x_n)$ es una función de nk variables, cada x_r tiene k coordenadas. El problema se resuelve con un método iterativo de análisis numérico llamado método del gradiente o de “*steepest descent*” (descenso por pendiente máxima). A nivel práctico se han desarrollado algoritmos de computadora que realizan este método exitosamente. Aunque hay una gran variación entre ellos, a continuación se presenta uno que da la idea general de cómo funcionan, suponiendo sin pérdida de generalidad que la medida de bondad de ajuste es el Stress fórmula uno.

La mayoría de los algoritmos consisten de cuatro fases. En la primera fase se obtiene una configuración inicial. En la segunda fase, las coordenadas estimadas y las distancias entre los objetos se estandarizan. La tercera y cuarta fase serán llamadas fases no métrica y métrica, respectivamente. El principal propósito de la fase no métrica es estimar las disparidades. El principal propósito de la fase métrica es estimar las coordenadas de los objetos.

Después de calcular la configuración inicial, cada iteración consiste de una estandarización, una fase no métrica, y una fase métrica. Al finalizar cada iteración se calcula la medida de bondad de ajuste previamente seleccionada. Las iteraciones continúan hasta que el cambio en la medida de bondad de ajuste de una iteración a la siguiente sea menor que algún valor establecido, tal como 0.001. En cada una de las siguientes cuatro secciones se discuten cada uno de los pasos de un algoritmo no métrico: obteniendo una configuración inicial, estandarizando las distancias y las coordenadas de los objetos, la fase no métrica, y la fase métrica.

2.3.1. Configuración Inicial

Cualquier algoritmo del análisis de escalamiento no métrico comienza con la obtención de una configuración inicial, un paso realizado sólo una vez.

Una posibilidad para una configuración inicial es simplemente usar una configuración arbitraria, por ejemplo, elegir n puntos aleatoriamente a partir de una distribución uniforme o normal. Esta configuración se normaliza de manera usual para tener el centroide en el origen y la distancia cuadrada media de los puntos desde el origen igual a la unidad.

Los mejores métodos utilizan la solución clásica del escalamiento multidimensional para obtener una configuración inicial.

En general, las primeras k componentes principales de la matriz de productos escalares constituyen la configuración inicial para una solución k -dimensional. Sean $\hat{x}_1^o, \hat{x}_2^o, \dots, \hat{x}_n^o$ los puntos de la configuración inicial, las distancias estimadas entre ellos se calculan de manera usual como:

$$\hat{d}_{rs}^o = \left[\sum_{i=1}^k (\hat{x}_{ri}^o - \hat{x}_{si}^o)^2 \right]^{\frac{1}{2}} \quad (2.11)$$

En la ecuación (2.11) \hat{x}_{ri}^o y \hat{x}_{si}^o se refieren a la i -ésima coordenada estimada de la configuración inicial de los objetos r y s , respectivamente.

Guttman (1968) y Lingoes y Roskam (1973) sugieren lo siguiente para encontrar una configuración inicial. Sea la matriz C definida como $[C]_{rs} = c_{rs}$, donde

$$c_{rs} = \begin{cases} 1 + \sum_r \frac{\rho_{rs}}{N} & (r = s) \\ 1 - \frac{\rho_{rs}}{N} & (r \neq s) \end{cases}$$

donde N es el número total de disimilaridades $\{\delta_{rs}\}$, y ρ_{rs} es el rango de δ_{rs} en el ordenamiento numérico de $\{\delta_{rs}\}$. Las componentes principales de C son encontradas y la configuración inicial está dada por los vectores propios de las primeras k componentes principales, pero ignorando la primera con vector constante.

2.3.2. Coordenadas y Distancias Estandarizadas

Después de que la configuración inicial ha sido obtenida, la primera iteración comienza. En el comienzo de cada iteración, muchos algoritmos emplean alguna estandarización de las coordenadas estimadas (\hat{x}_{ri}) y de las distancias (\hat{d}_{rs}). Comenzando la primera iteración, las estimaciones actuales son obtenidas a partir de la configuración inicial. Para cada iteración después de la primera, la estimación actual se obtiene en la fase métrica de la iteración previa.

La experiencia indica que esta estandarización de las distancias reduce la probabilidad de encontrar una solución degenerada. Si el Stress fórmula uno es usado, es conveniente para estandarizar las distancias que su suma de cuadrados sea igual a 1,0: $\sum_{(r,s)} \widehat{d}_{rs}^2 = 1,0$. Cuando las distancias son así estandarizadas, S_1 se reduce a:

$$S_1 = \left[\sum_{(r,s)} \left(\widehat{\delta}_{rs} - \widehat{d}_{rs} \right)^2 \right]^{1/2} \quad (2.12)$$

Aquí minimizar S_1 equivale a minimizar S :

$$S = \sum_{(r,s)} \left(\widehat{\delta}_{rs} - \widehat{d}_{rs} \right)^2 \quad (2.13)$$

El algoritmo descrito arriba supone que las distancias estimadas están estandarizadas, de esta forma su suma de cuadrados se iguala a 1,0. Esta estandarización puede realizarse, por ejemplo, multiplicando cada distancia estimada no estandarizada de la c -ésima iteración, \widehat{d}_{rs}^c , por una constante adecuada, ce , de manera que su suma de cuadrados sea igual a 1,0. Para expresar las coordenadas estimadas sobre la misma escala que las distancias, cada coordenada estimada de la c -ésima iteración, \widehat{x}_{ri}^c , deberá ser multiplicada por la misma constante. Típicamente, esta estandarización ocurrirá una vez en cada iteración justo antes de la fase no métrica.

Note que como deseamos que $\sum_{(r,s)} \widehat{d}_{rs}^2 = 1,0$, donde \widehat{d}_{rs} son las distancias estandarizadas, si hacemos

$$\widehat{d}_{rs} = \widehat{d}_{rs}^c = \frac{\widehat{d}_{rs}^{(c-1)}}{\left[\sum_{(r,s)} \left(\widehat{d}_{rs}^{(c-1)} \right)^2 \right]^{1/2}}$$

donde \widehat{d}_{rs}^c es la distancia entre el objeto r y el objeto s de la c -ésima iteración,

para $c = 1, \dots, C$, se cumplirá la restricción, ya que

$$\begin{aligned}
 \sum_{(r,s)} \left(\hat{d}_{rs}^c \right)^2 &= \sum_{(r,s)} \left\{ \frac{\hat{d}_{rs}^{(c-1)}}{\left[\sum_{(r,s)} \left(\hat{d}_{rs}^{(c-1)} \right)^2 \right]^{1/2}} \right\}^2 \\
 &= \sum_{(r,s)} \frac{\left(\hat{d}_{rs}^{(c-1)} \right)^2}{\left[\sum_{(r,s)} \left(\hat{d}_{rs}^{(c-1)} \right)^2 \right]} \\
 &= \frac{\sum_{(r,s)} \left(\hat{d}_{rs}^{(c-1)} \right)^2}{\left[\sum_{(r,s)} \left(\hat{d}_{rs}^{(c-1)} \right)^2 \right]} \\
 &= 1
 \end{aligned}$$

De esta manera la constante (ce) por la que tenemos que multiplicar las distancias y las coordenadas para estandarizar será:

$$ce = \frac{1}{\left[\sum_{(r,s)} \left(\hat{d}_{rs}^{(c-1)} \right)^2 \right]^{1/2}}$$

Como las fases no métrica y métrica operan sólo sobre coordenadas estandarizadas y distancias estimadas, los términos “distancias estimadas” y “coordenadas estimadas” en la discusión se refieren a valores estandarizados.

2.3.3. Fase No Métrica

La fase no métrica usa los datos de disimilaridades y distancias estandarizadas de la iteración previa (o en dado caso, de la configuración inicial) para calcular las disparidades. La parte principal de la fase no métrica está constituida por una serie de pasos por medio de los datos (descritos anteriormente) para una simple iteración. En esta sección se describe primero el arreglo de los datos de los puntos en orden ascendente, y entonces se procede a una descripción de los pasos. Note que cada iteración contiene una serie completa de estos pasos. La fase no métrica no ajusta las estimaciones de las coordenadas

de objetos o las distancias estimadas. Sólo las disparidades $\{\widehat{\delta}_{rs}\}$ cambian en la fase no métrica. Antes de describir la fase no métrica, se define la notación.

Sea $\widehat{\delta}_{rs}^{(c+1)}$ la disparidad obtenida para el par de objetos (r, s) , calculada en la $(c+1)$ -ésima iteración. Sean \widehat{x}_{ri}^c ($i = 1, \dots, k$) las coordenadas estimadas en la iteración c ($c = 0, \dots, C$). Si $c = 0$, entonces estas coordenadas estimadas constituyen la configuración inicial. Finalmente, sea \widehat{d}_{rs}^c una distancia estimada en la c -ésima iteración:

$$\widehat{d}_{rs}^c = \left[\sum_{i=1}^k (\widehat{x}_{ri}^c - \widehat{x}_{si}^c)^2 \right]^{\frac{1}{2}} \quad (2.14)$$

En la fase no métrica las disparidades $\{\widehat{\delta}_{rs}^{(c+1)}\}$ se calculan de modo que constituyan una transformación monótona de las disimilaridades originales $\{\delta_{rs}\}$. Esto es, las disparidades satisfacen la expresión (2.6). El algoritmo no aplica simplemente una función monótona continua bien conocida (tal como una función logaritmo o una potencia) a los datos. En lugar de esto, la fase no métrica consiste en una serie de pasos en los cuales cada disparidad $\widehat{\delta}_{rs}^{(c+1)}$ es igual a la correspondiente distancia estimada \widehat{d}_{rs}^c , o es igual al promedio de varias distancias estimadas. Cuando las disparidades son calculadas en esta forma, minimizan la siguiente suma de cuadrados de discrepancias:

$$\sum_{(r,s)} \left(\widehat{\delta}_{rs}^{(c+1)} - \widehat{d}_{rs}^c \right)^2.$$

Las distancias estimadas de la iteración c son las disparidades iniciales estimadas para la iteración $(c+1)$.

El primer paso de la fase no métrica es arreglar las disimilaridades y sus correspondientes disparidades en forma ascendente. Surge una pequeña complicación cuando hay empates en los datos, es decir, cuando $\delta_{rs} = \delta_{uv}$ para algún par de puntos. Si $\delta_{rs} = \delta_{uv}$ y $\widehat{d}_{rs}^c = \widehat{d}_{uv}^c$, entonces no hay problema en cual precede. Si, por el contrario, $\delta_{rs} = \delta_{uv}$ y $\widehat{d}_{rs}^c \neq \widehat{d}_{uv}^c$, entonces la disparidad asociada con la distancia más pequeña estimada precede. En el primer paso de esta fase en una iteración, las disparidades son iguales a las distancias obtenidas en la configuración inicial o en la iteración anterior. En los pasos siguientes en esa misma iteración, las disparidades son las calculadas en el paso previo.

Cada paso después del primero comienza con la división del conjunto de disparidades en bloques con disparidades iguales, si no las hay entonces cada uno constituye su propio bloque.

El resto de cada paso consiste de comparar bloques adyacentes. Sea m ($m = 1, \dots, M$) un subíndice que distingue a los bloques, desde el primero que contiene las disparidades más pequeñas ($m=1$), hasta el último que contiene las disparidades más grandes ($m = M$). Comenzando con $m=1$, los elementos del m -ésimo bloque se comparan con los elementos del $(m+1)$ -ésimo bloque. Si los elementos del m -ésimo bloque son menores que los del $(m+1)$ -ésimo bloque, entonces simplemente se prosigue con la comparación de los siguientes dos bloques. Si por el contrario, los elementos del m -ésimo bloque son mayores que los del $(m+1)$ -ésimo, entonces se sustituyen todos los elementos en el m -ésimo y el $(m+1)$ -ésimo bloques por la media aritmética de los elementos de ambos bloques. Después de sustituir los elementos en el m -ésimo y el $(m+1)$ -ésimo bloques, todos los elementos en ambos bloques serán iguales, y de aquí se pueden reagrupar dentro de un mismo bloque formando el nuevo m -ésimo bloque. Habiendo reagrupado los bloques, se procede a comparar este nuevo m -ésimo con el siguiente, el nuevo $(m+1)$ -ésimo bloque. El paso finaliza una vez que todos los bloques adyacentes hayan sido comparados. El resultado de los pasos es un nuevo conjunto de disparidades estimadas. Si los bloques no fueran agrupados en el último paso, entonces la fase no métrica de la iteración finaliza.

Si por el contrario, algunos bloques fueran reagrupados, entonces da inicio un nuevo paso. Cada paso comienza con un agrupamiento de la actual disparidad estimada dentro de los bloques de valores iguales. El resto de los pasos consiste en comparar pares de bloques adyacentes y reagruparlos cuando sea necesario. Los pasos continúan hasta que no sea requerido reagrupar. Las disparidades estimadas obtenidas en el último paso de la fase son las disparidades estimadas de la $(c+1)$ -ésima iteración, $\widehat{\delta}_{rs}^{(c+1)}$. Cuando termina esta fase las disparidades satisfacen la condición de monotonidad. La fase no métrica es una aplicación de la regresión monótona de Kruskal de las disimilaridades contra las distancias. Las disparidades estimadas de la última iteración son las disparidades estimadas finales $\widehat{\delta}_{r,s}$ del análisis.

2.3.4. Fase Métrica

La fase métrica sigue de la fase no métrica. Esta fase utiliza las disparidades calculadas en la fase no métrica recién concluida $\{\widehat{\delta}_{rs}^{(c+1)}\}$, las coordenadas estimadas de la iteración previa $\{\widehat{x}_{ri}^c\}$ y las distancias estimadas de la iteración previa $\{\widehat{d}_{rs}^c\}$, para calcular las nuevas coordenadas estimadas $\{\widehat{x}_r^{(c+1)}\}$, para las cuales se obtienen las nuevas distancias estimadas $\{\widehat{d}_{rs}^{(c+1)}\}$. Las disparidades permanecen sin cambio en esta fase. Primero describimos la función que la fase métrica tiene designada a minimizar, y entonces se procede a una descripción de cómo calcular las nuevas coordenadas estimadas.

Habiendo obtenido las disparidades $\{\widehat{\delta}_{rs}^{(c+1)}\}$, se inicia la fase métrica para obtener las nuevas estimaciones de las coordenadas $\{\widehat{x}_r\}$. Como las distancias están estandarizadas, su suma de cuadrados es igual a 1,0, entonces el Stress fórmula uno se reduce a la ecuación (2.12). Minimizando S_1 entonces es equivalente a minimizar S en la ecuación (2.13).

Lingoes y Roskam (1973) dan la siguiente forma para las coordenadas estimadas $\widehat{x}_{ri}^{(c+1)}$:

$$\widehat{x}_{ri}^{(c+1)} = \widehat{x}_{ri}^c - \frac{1}{n} \sum_{s=1}^n \left(1 - \frac{\widehat{\delta}_{rs}^{(c+1)}}{\widehat{d}_{rs}^c} \right) (\widehat{x}_{ri}^c - \widehat{x}_{si}^c). \quad (2.15)$$

La ecuación (2.15) da la igualdad requerida para calcular las nuevas coordenadas de objetos. Para evitar dividir entre cero, la razón $\left(\frac{\widehat{\delta}_{rs}^{(c+1)}}{\widehat{d}_{rs}^c} \right)$ se iguala a 1,0 si $\widehat{d}_{rs}^c = 0,0$.

El lado derecho de la ecuación (2.15) utiliza las coordenadas estimadas de la iteración previa $\{\widehat{x}_{ri}^c\}$, las disparidades estimadas de la iteración actual $\{\widehat{\delta}_{rs}^{(c+1)}\}$, y las distancias estimadas de la iteración previa $\{\widehat{d}_{rs}^c\}$.

Después de la estandarización de las nuevas coordenadas y las distancias estimadas, comienza la segunda iteración. La iteración continúa hasta que mejora el Stress de una iteración a la siguiente quedando la diferencia debajo de algún valor tal como 0,001.

La evaluación del ajuste proporcionado por el valor del Stress para una solución determinada debe hacerse teniendo en cuenta varios factores:

1. El valor del Stress suele ser más alto cuanto mayor sea el número de objetos, debido a que cuando tenemos pocos objetos, el número de proximidades a ajustar en la solución será también pequeño, pero a medida que aumenta el número de objetos, el número de proximidades a ajustar se incrementa rápidamente.
2. El valor del Stress será siempre más alto para soluciones de menor dimensionalidad, e irá bajando a medida que la solución contenga un mayor número de dimensiones. Cuando el número de dimensiones es igual al número de objetos menos 2 ($n - 2$), el ajuste será siempre perfecto. El objetivo en este caso será buscar un valor suficientemente bajo del Stress (buen ajuste) unido a una dimensionalidad también baja (representación parsimoniosa de los datos).
3. Cuando se permite desempatar los empates presentes en los datos (disimilaridades), el valor del Stress obtenido suele ser menor que el obtenido cuando no se permite desempatar.

2.3.5. Fase No Métrica Alternativa

Después del artículo original de Shepard (1962) sobre MDS no métrico, otros autores han propuesto varios algoritmos no métricos (Guttman, 1968; Kruskal, 1964).

La fase no métrica propuesta por Guttman (1968) es una alternativa de la fase no métrica de Kruskal (1964) ya descrita anteriormente, y ha jugado un papel importante en la literatura del MDS que será descrito aquí. Semejante a la fase no métrica de Kruskal (1964), Guttman (1968) utiliza las disimilaridades originales y las distancias calculadas en la iteración anterior para calcular las nuevas disparidades de cada iteración, pero ejecuta los cálculos de manera diferente.

En la fase no métrica de Guttman (1968), cada disparidad de la iteración $(c + 1)$ se iguala a una de las distancias de la iteración c . Específicamente, si el par de objetos (r, s) corresponde a la i -ésima disimilaridad más pequeña δ_{rs} , entonces la disparidad correspondiente $\widehat{\delta}_{rs}^{(c+1)}$ es igual a la i -ésima distancia más pequeña estimada (que no necesariamente es \widehat{d}_{rs}^c). Esto es, si el par de objetos (r, s) corresponde a la disimilaridad más pequeña δ_{rs} , entonces $\widehat{\delta}_{rs}^{(c+1)}$ es igual a la distancia estimada más pequeña. Si el par de objetos (r, s) corresponde a la segunda disimilaridad más pequeña δ_{rs} , entonces $\widehat{\delta}_{rs}^{(c+1)}$

es igual a la segunda distancia estimada más pequeña. Si el par de objetos (r, s) corresponde a la tercera disimilaridad más pequeña, entonces la correspondiente disparidad es igual a la tercera distancia estimada más pequeña, y así sucesivamente hasta llegar al par de objetos (r, s) que corresponda a la mayor disimilaridad. La correspondiente disparidad es igual a la mayor distancia.

Las fases no métricas de Guttman y Kruskal son semejantes en un aspecto. Ambas usan las disimilaridades originales $\{\delta_{rs}\}$ y las distancias estimadas en la iteración previa $\{\widehat{d}_{rs}^c\}$ para calcular las nuevas disparidades de la iteración actual, $\{\widehat{\delta}_{rs}^{(c+1)}\}$. Estas difieren, sin embargo, en la forma en que usan las distancias estimadas en los cálculos. En la fase no métrica de Kruskal, cada disparidad $\widehat{\delta}_{rs}^{(c+1)}$ es igual a la correspondiente distancia estimada \widehat{d}_{rs}^c , o igual al promedio de las distancias estimadas en caso de empates. En la fase no métrica de Guttman, la disparidad para el par de objetos (r, s) , que corresponde a la i -ésima disimilaridad más pequeña, es igual a la i -ésima distancia estimada más pequeña.

2.3.6. Observaciones

En general, el análisis del escalamiento multidimensional no métrico, donde sólo se utiliza la relación de orden, provee espacios de dimensión baja mejor ajustados que su contraparte métrico, debido principalmente al proceso iterativo y a que utiliza una medida de bondad de ajuste en cada iteración. A pesar de que las soluciones obtenidas utilizando cada uno de los modelos pueden ser muy similares, la elección del modelo métrico o no métrico, es un tema de controversia tanto a nivel teórico como práctico.

2.4. Aproximación de Kruskal

Sean $\{\delta_{rs}\}$ las disimilaridades originales. Sea x_r el r -ésimo punto en X con coordenadas $(x_{r1}, \dots, x_{rp})^t$. Sea la medida de distancias para X la métrica de Minkowski, y así para los puntos r y s en X ,

$$d_{rs} = \left(\sum_{i=1}^p |x_{ri} - x_{si}|^\lambda \right)^{1/\lambda} \quad \lambda > 0$$

Definimos las disparidades $\{\widehat{\delta}_{rs}\}$, como una función de las distancias $\{d_{rs}\}$, por

$$\widehat{\delta}_{rs} = f(d_{rs})$$

donde f es una función monótona tal que

$$\widehat{\delta}_{rs} \leq \widehat{\delta}_{uv} \quad \text{siempre que} \quad \delta_{rs} < \delta_{uv} \quad (\text{condición } C_1).$$

Así las disparidades “preservan” el orden de las disimilaridades originales pero permite posibles empates en las disparidades.

Sea S la medida de bondad de ajuste y sea

$$S = \sqrt{\frac{S^*}{T^*}}$$

donde $S^* = \sum_{(r,s)} (d_{rs} - \widehat{\delta}_{rs})^2$, y $T^* = \sum_{(r,s)} d_{rs}^2$. Note que, como $\delta_{rs} = \delta_{sr}$, la suma se toma bajo $1 = r < s = n$ para toda r, s . La medida de bondad de ajuste se minimiza con respecto a $\{d_{rs}\}$, es decir, con respecto a $\{x_{ri}\}$, las coordenadas de la configuración, y también con respecto a $\{\widehat{\delta}_{rs}\}$ usando regresión isotónica.

2.4.1. Minimizando S con Respecto a las Disparidades

Sean las disimilaridades $\{\delta_{rs}\}$ reetiquetadas por $\{\delta_i : i = 1, \dots, N\}$ y suponga que han sido acomodadas en orden numérico y que no hay empates, es decir $\delta_i < \delta_{i+1}$ $i = 1, \dots, N - 1$. Reetiquete las distancias $\{d_{rs}\}$ como $\{d_i : i = 1, \dots, N\}$ donde d_i corresponde a la disimilaridad δ_i . Para ilustrar la demostración se emplea un ejemplo.

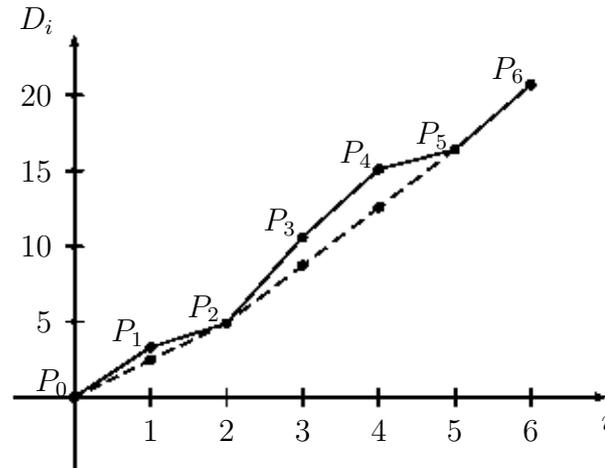
Ejemplo

Suponga que hay sólo cuatro objetos con disimilaridades

$$\delta_{12} = 2,1, \quad \delta_{13} = 3,0, \quad \delta_{14} = 2,4, \quad \delta_{23} = 1,7, \quad \delta_{24} = 3,9, \quad \delta_{34} = 3,2$$

y una configuración de puntos representa los cuatro objetos con distancias

$$d_{12} = 1,6, \quad d_{13} = 4,5, \quad d_{14} = 5,7, \quad d_{23} = 3,3, \quad d_{24} = 4,3, \quad d_{34} = 1,3$$

Figura 2.1: D_i vs. i

Entonces las disimilaridades ordenadas con la nueva notación, junto con sus distancias asociadas son

$$\delta_1 = 1,7, \delta_2 = 2,1, \delta_3 = 2,4, \delta_4 = 3,0, \delta_5 = 3,2, \delta_6 = 3,9$$

$$d_1 = 3,3, d_2 = 1,6, d_3 = 5,7, d_4 = 4,5, d_5 = 1,3, d_6 = 4,3$$

Minimizar S equivale a minimizar $S' = \sum_i (d_i - \widehat{\delta}_i)^2$, otra vez usando la nueva notación del sufijo. Sean las sumas acumulativas de $\{d_i\}$

$$D_i = \sum_{j=1}^i d_j \quad (i = 1, \dots, N),$$

y considere una gráfica de D_i contra i , cuyos puntos son P_0, P_1, \dots, P_N donde el origen está etiquetado por P_0 . La figura 2.1 muestra la gráfica para el ejemplo. Note que la pendiente de la línea que une P_{i-1} y P_i es justo d_i . El mínimo convexo mayor de las sumas acumulativas es la gráfica del supremo de todas las funciones convexas cuya gráfica se extiende por debajo de las sumas acumulativas. El mínimo convexo mayor para el ejemplo también se muestra en la figura 2.1. Las $\{\widehat{\delta}_i\}$ que minimizan S' están dadas por el mínimo convexo mayor, donde $\widehat{\delta}_i$ es la i -ésima abscisa más pequeña. De la

figura, se ve que algunos de los valores $\widehat{\delta}_i$ son iguales a d_i , y obviamente $S' = 0$ si $\widehat{\delta}_i = d_i$, para toda i . Note que $\widehat{\delta}_i = \widehat{D}_i - \widehat{D}_{i-1}$ es la pendiente de la línea. Así si $\widehat{D}_i < D_i$ entonces $\widehat{\delta}_i < \widehat{\delta}_{i+1}$.

Para mostrar que estas disparidades $\{\widehat{\delta}_i\}$ verdaderamente minimizan S' , sea $\{d_i^*\}$ un conjunto arbitrario de valores reales que satisface la condición C_1 . Vamos a mostrar que

$$\sum_{i=1}^N (d_i - d_i^*)^2 \geq \sum_{i=1}^N (d_i - \widehat{\delta}_i)^2.$$

Sea

$$D_i^* = \sum_{j=1}^i d_j^* \quad \widehat{D}_i = \sum_{j=1}^i \widehat{\delta}_j$$

La suma de Abel dada por $\sum_{i=1}^N a_i b_i = \sum_{i=1}^{N-1} A_i (b_i - b_{i+1}) + A_N b_N$, donde $A_i = \sum_{j=1}^i a_j$ son sumas parciales, será necesaria en lo siguiente.

Escribimos

$$\begin{aligned} \sum_{i=1}^N (d_i - d_i^*)^2 &= \sum_{i=1}^N \left\{ (d_i - \widehat{\delta}_i) + (\widehat{\delta}_i - d_i^*) \right\}^2 \\ &= \sum_{i=1}^N (d_i - \widehat{\delta}_i)^2 + \sum_{i=1}^N (\widehat{\delta}_i - d_i^*)^2 + 2 \sum_{i=1}^N (d_i - \widehat{\delta}_i) (\widehat{\delta}_i - d_i^*) \end{aligned}$$

Ahora

$$\begin{aligned} \sum_{i=1}^N (d_i - \widehat{\delta}_i) (\widehat{\delta}_i - d_i^*) &= \sum_{i=1}^{N-1} (D_i - \widehat{D}_i) (\widehat{\delta}_i - \widehat{\delta}_{i+1}) \\ &\quad - \sum_{i=1}^{N-1} (D_i - \widehat{D}_i) (d_i^* - d_{i+1}^*) \\ &\quad + (D_N - \widehat{D}_N) (\widehat{\delta}_N - d_N^*) \end{aligned}$$

Ahora $D_N - \widehat{D}_N = 0$ porque el último punto del mínimo convexo mayor y P_N coinciden. Considere $(D_i - \widehat{D}_i) (\widehat{\delta}_i - \widehat{\delta}_{i+1})$. Si el i -ésimo punto sobre el mínimo convexo mayor coincide con P_i entonces $D_i = \widehat{D}_i$ y así el término es cero. Por otro lado, si $\widehat{D}_i < D_i$ entonces $\widehat{\delta}_i = \widehat{\delta}_{i+1}$ y así el término es otra vez

cero. De aquí, como $D_i - \widehat{D}_i \geq 0$ y por la condición C_1 , $d_i^* < d_{i+1}^*$ el término $-\sum_{i=1}^{N-1} (D_i - \widehat{D}_i) (d_i^* - d_{i+1}^*)$ es positivo. De aquí

$$\sum_{i=1}^N (d_i - d_i^*)^2 \geq \sum_{i=1}^N (d_i - \widehat{\delta}_i)^2 + \sum_{i=1}^N (\widehat{\delta}_i - d_i^*)^2$$

y así

$$\sum_{i=1}^N (d_i - d_i^*)^2 \geq \sum_{i=1}^N (d_i - \widehat{\delta}_i)^2$$

Las disparidades $\{\widehat{\delta}_{rs}\}$ dan el mínimo de S' , y por lo tanto el mínimo de S . En la literatura de MDS, la regresión isotónica se refiere a regresión de mínimos cuadrados de $\{d_{rs}\}$ sobre $\{\delta_{rs}\}$.

Así para ilustrar el ejemplo

$$\widehat{\delta}_1 = \widehat{\delta}_2 = 2,45, \quad \widehat{\delta}_3 = \widehat{\delta}_4 = \widehat{\delta}_5 = 3,83, \quad \widehat{\delta}_6 = 4,3$$

note que $\widehat{\delta}_1, \widehat{\delta}_2$ son la media de d_1 y d_2 ; $\widehat{\delta}_3, \widehat{\delta}_4, \widehat{\delta}_5$ son la media de d_3, d_4 y d_5 ; $\widehat{\delta}_6$ es igual a d_6 . El valor de S es 0.14.

2.4.2. Una Configuración con Mínimo Stress

Con $\{\widehat{\delta}_{rs}\}$ definida como una regresión de mínimos cuadrados monótona de $\{d_{rs}\}$ sobre $\{\delta_{rs}\}$, S es entonces el Stress de la configuración; S^* es llamado el Stress puro (*raw stress*). El numerador T^* en la fórmula para Stress es usado como un factor de normalización permitiendo al Stress ser de dimensión libre.

Ahora se busca una configuración que minimice el Stress S . La minimización del Stress no es una tarea fácil. El primer paso es poner todas las coordenadas de los puntos de X en un vector $x = (x_{11}, \dots, x_{1p}, \dots, x_{n1}, \dots, x_{np})^t$, vector con np elementos. El Stress S se estima como una función de x , y se minimiza con respecto a x de una manera iterativa. Se usa el método de *steepest descent*, así que si x_m es el vector de coordenadas después de la m -ésima iteración

$$x_{m+1} = x_m - \frac{\frac{\partial S}{\partial x} \times sl}{\left| \frac{\partial S}{\partial x} \right|},$$

donde sl es la longitud del paso discutido previamente.

Ahora

$$\begin{aligned}
\frac{\partial S}{\partial x_{ui}} &= \frac{1}{2} \sqrt{\frac{T^*}{S^*}} \left(\frac{\partial S^*}{\partial x_{ui}} - \frac{\partial T^*}{\partial x_{ui}} \right) \\
&= \frac{1}{2} S \left(\frac{1}{S^*} \frac{\partial S^*}{\partial x_{ui}} - \frac{1}{T^*} \frac{\partial T^*}{\partial x_{ui}} \right) \\
\frac{\partial S^*}{\partial x_{ui}} &= 2 \sum_{r,s} (d_{rs} - \hat{\delta}_{rs}) \frac{\partial d_{rs}}{\partial x_{ui}} \\
\frac{\partial T^*}{\partial x_{ui}} &= 2 \sum_{r,s} d_{rs} \frac{\partial d_{rs}}{\partial x_{ui}}
\end{aligned}$$

Por la métrica de Minkowski tenemos que

$$\frac{\partial d_{rs}}{\partial x_{ui}} = d_{rs}^{1-\lambda} \sum_{r,s} (x_{ri} - x_{si})^{\lambda-1} (\delta^{ru} - \delta^{su}) \text{signo}(x_{ri} - x_{si})$$

y de aquí

$$\frac{\partial S}{\partial x_{ui}} = S \sum_{r,s} (\delta^{ru} - \delta^{su}) \left[\frac{d_{rs} - \hat{\delta}_{rs}}{S^*} - \frac{d_{rs}}{T^*} \right] \times \frac{|x_{ri} - x_{si}|^{\lambda-1}}{d_{rs}^{\lambda-1}} \text{signo}(x_{ri} - x_{si})$$

Se necesita elegir una configuración inicial x_0 . Una posibilidad es generar n puntos de acuerdo a un proceso Poisson en una región de \mathbb{R}^p . En su forma simple, esto quiere decir simular cada coordenada de manera individual para cada punto, independientemente, a partir de una distribución uniforme sobre $[0,1]$. Hay otros métodos sugeridos para elegir una configuración tal como los mencionados en secciones anteriores.

Una vez que x_0 ha sido elegida, el método de *steepest descent* puede ser empleado para encontrar una configuración con Stress mínimo usando el siguiente algoritmo.

2.4.3. Técnica Iterativa de Kruskal

Los siguientes puntos resumen la técnica iterativa usada para encontrar una configuración con mínimo Stress.

1. Elegir una configuración inicial.

2. Normalizar la configuración para tener su centroide en el origen y la distancia cuadrada media desde el origen igual a la unidad. Esto se hace porque el Stress es invariante bajo traslación, dilatación uniforme, y otra manera de iteraciones sucesivas de los procedimientos que pueden tener las configuraciones.
3. Encontrar $\{d_{rs}\}$ a partir de la configuración normalizada.
4. Ajustar $\{\widehat{\delta}_{rs}\}$. Se observó que la regresión monótona de mínimos cuadrados de $\{d_{rs}\}$ sobre $\{\delta_{rs}\}$ particiona $\{\delta_{rs}\}$ en bloques, en los cuales los valores de δ_{rs} son constantes, igual a la media de los correspondientes valores d_{rs} . Para encontrar la partición apropiada de $\{\delta_{rs}\}$, primero la partición que es usada tiene N bloques en el que cada uno contiene una δ_i , usando la notación alternativa. Si esta partición inicial tiene $d_1 \leq d_2 \leq \dots \leq d_N$, entonces $\widehat{\delta}_i = d_i$ y esta partición es la final. De otra manera dos bloques consecutivos son reagrupados de tal manera que si $\delta_i > \delta_{i+1}$, entonces $\widehat{\delta}_i = \widehat{\delta}_{i+1} = (d_i + d_{i+1})/2$. Los bloques son reagrupados continuamente y las nuevas disparidades $\widehat{\delta}_i$ encontradas, hasta que la partición requerida sea alcanzada. La partición requerida también puede ser encontrada considerando la gráfica de las sumas acumulativas, D_i , y encontrar el mínimo convexo mayor. La pendiente, s_i , de D_i desde el origen es D_i/i . El punto con la pendiente menor tiene que estar sobre el mínimo convexo mayor. Todos los puntos precedentes a este punto no están sobre el mínimo y sus pendientes pueden ser removidas a partir de otra consideración. El punto con la siguiente pendiente más pequeña se encuentra a partir de las pendientes que permanecen. Este punto está sobre el mínimo, pero los puntos entre el punto mínimo precedente y éste no están. Sus pendientes son descartadas. Este procedimiento continúa hasta que el N -ésimo punto es alcanzado. Una vez que el mínimo convexo mayor ha sido establecido es una tarea fácil encontrar $\{\widehat{\delta}_i\}$.
5. Encontrar el gradiente $\frac{\partial S}{\partial x}$. Si $|\frac{\partial S}{\partial x}| < \varepsilon$, donde ε es un número preseleccionado muy pequeño, entonces ha sido encontrada una configuración con un mínimo Stress y el proceso iterativo puede terminar. Note que esta configuración podrá estar dando un mínimo local para el Stress, y no el mínimo global.

6. Encontrar la longitud del paso nuevo sl . Kruskal recomienda que sl cambie en cada paso de acuerdo a

$$sl_{\text{presente}} = sl_{\text{previo}} \times (\text{angle factor}) \\ \times (\text{relaxation factor}) \times (\text{good luck factor})$$

donde

$$\begin{aligned} \text{angle factor} &= 4,0^{\cos^3 \theta}, \\ \theta &= \text{ángulo entre el gradiente presente y el previo}, \\ \text{relaxation factor} &= \frac{1,3}{1 + (5 \text{ step ratio})^5}, \\ 5 \text{ step ratio} &= \min \left[1, \left(\frac{\text{Stress actual}}{\text{Stress de 5 iteraciones anteriores}} \right) \right], \\ \text{good luck factor} &= \min \left[1, \frac{\text{Stress actual}}{\text{Stress previo}} \right]. \end{aligned}$$

7. Encontrar la nueva configuración

$$x_{n+1} = x_n - sl \left. \frac{\partial S}{\partial x} \right|_{\frac{\partial S}{\partial x}}$$

8. Ir al paso 2.

2.5. Aproximación de Guttman

Guttman tomó una aproximación diferente a la de Kruskal para ajustar el MDS no métrico. Definió la función de alienación que es básicamente equivalente a la función de Stress de Kruskal, pero que tiene un algoritmo diferente para la minimización. Su aproximación se describe brevemente.

Sean las disimilaridades ordenadas por rango $\{\delta_{rs}\}$ puestas en un vector δ con elementos δ_r ($r = 1, \dots, N$). Sean las distancias $\{d_{rs}\}$ de una configuración puestas en un vector d en orden correspondiente a $\{\delta_r\}$. Sea E una matriz de permutación $N \times N$ que coloca a los elementos de d en orden ascendente. Las disparidades se definen por la imagen del rango d^* de d , dada por

$$d^* = Ed$$

El coeficiente de continuidad, μ , para la configuración se define como

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

que toma el valor uno para un ajuste perfecto. Para encontrar una mejor configuración, el coeficiente de alienación, κ , dado por

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

se minimiza usando el método de *steepest descent*.

Ejemplo

Suponga que hay sólo tres objetos, con disimilaridades

$$\delta_{12} = 4, \delta_{13} = 1, \delta_{23} = 3$$

con $\delta_{rr} = 0$. Sea una configuración particular cuyas distancias entre sus puntos son

$$d_{12} = 2, d_{13} = 4, d_{23} = 5$$

Entonces en la nueva notación, ordenando las disimilaridades, tenemos que

$$\delta_1 = 1, \delta_2 = 3, \delta_3 = 4$$

$$d_1 = 4, d_2 = 5, d_3 = 2$$

La matriz de permutaciones es

$$E = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

dando $d^* = (2, 4, 5)^t$.

El coeficiente de continuidad, μ es 0.84, y el coeficiente de alienación κ es 0.54.

El trabajo de Guttman es mucho más detallado que esta simple exposición, y se refiere a la monotonicidad fuerte y débil, así como a los empates en los datos. Puede mostrarse que minimizar κ es equivalente a minimizar el Stress S .

2.6. Dimensionalidad, Rotación e Interpretación

El dato de entrada de un programa de MDS no métrico es una matriz de disimilaridades. Los datos de salida contienen los siguientes elementos: primero, una medida de bondad de ajuste, cualquiera de los dos Stress, S-Stress, o el coeficiente de alienación, y segundo, una matriz $\hat{X}_{n \times k}$ estimada de las coordenadas de los objetos. La decisión acerca de la dimensionalidad, rotación e interpretación se basan primero sobre estos dos elementos.

2.6.1. Dimensionalidad

Los puntos aquí discutidos se desarrollaron suponiendo que el número de dimensiones de la configuración buscada, k , es conocido y está fijo. En la práctica, la dimensionalidad debe determinarse como parte del mismo análisis. Para elegir k en tales casos, se obtienen soluciones en distintas dimensiones. Si k^* es la más usada en un supuesto a priori como la dimensionalidad, se puede proceder para obtener soluciones en cada dimensión desde $k^* - 3$ hasta $k^* + 3$. Si $k^* \leq 3$, entonces se desearía obtener soluciones desde 1 hasta $k^* + 3$ dimensiones.

Obteniendo las soluciones en varias dimensiones, el usuario escogería la más aproximada. Otra vez, los criterios de interpretabilidad, reproductibilidad, y ajuste proporcionan las bases para la selección. El criterio de interpretabilidad establece que el número de dimensiones será la dimensión más pequeña en la cual aparecen todas las características de los objetos importantes (objetos ordenados o agrupados). La reproductibilidad establece que la solución se compone de dimensiones que emergen consistentemente a través de los subgrupos.

La decisión sobre el número de dimensiones de la solución puede hacerse simple al elegir la dimensionalidad que proporcione el valor de Stress más pequeño. Sin embargo, antes de usar ciegamente el Stress como un indicador de la dimensión “verdadera” o “correcta”, se deben comprender ciertos aspectos de esta medida.

1. La forma exacta de la medida de Stress puede variar con el programa de computadora empleado para el análisis de escalamiento multidimensional. Aunque las diversas medidas son similares, un valor numérico

dado puede reflejar un ajuste bueno o excelente para una pero pobre o pésimo para otra.

2. La terminación prematura del proceso de iteración antes de la convergencia o la terminación en un mínimo local, puede producir valores del Stress más altos que el verdadero valor mínimo para esa dimensión.
3. Los valores del Stress muy cercanos a cero (por ejemplo 0,01 o menos) pueden significar una solución total o parcialmente degenerada. Esta situación se da cuando la posición de los objetos en el espacio ajustado sugieren grupos o conglomerados naturales, es decir, los objetos se encuentran agrupados en un número reducido de regiones, y no se cumple la condición de monotonicidad.
4. La interpretación del valor del Stress es sensible al número de objetos y de dimensiones.

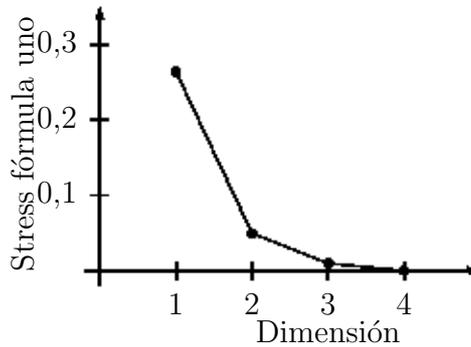
Por estas razones, se deben tomar algunas medidas de precaución. Los problemas de convergencia, del mínimo local y de la solución degenerada, así como los diagramas de Shepard son muy útiles para detectar algunas otras anomalías.

No obstante las cuestiones presentadas, el valor del Stress es un buen elemento para determinar la dimensionalidad adecuada de la solución. Para este propósito se construye una gráfica con los valores del Stress sobre el eje de las ordenadas y las dimensiones correspondientes sobre el eje de las abscisas en un plano cartesiano como la que se muestra en la figura 2.2. Al igual que la gráfica de valores propios contra dimensiones del modelo métrico, esta gráfica de valores del Stress contra dimensiones es llamada “*scatter diagram*”.

Suponiendo que no existe un error de muestreo o de medición demasiado grande, la gráfica debe mostrar un codo sobre el valor apropiado k de la dimensionalidad. Obsérvese que a diferencia de la gráfica del modelo métrico, el codo debe aparecer directamente sobre el valor k y no sobre el valor $k + 1$ ($\lambda = 0$ es valor propio de B). Nuevamente, si la gráfica no muestra claramente el valor de la dimensionalidad apropiada, o aún en ese caso, no deben olvidarse los otros criterios de elección.

2.6.2. Rotación

Si se ha empleado alguna métrica no euclidiana, entonces no hay problemas de rotación. Si X es una solución, entonces en general $X^* = XT$ con

Figura 2.2: Stress vs. k

T una matriz ortogonal, no será una solución. De hecho $X^* = XT$ será una solución sólo si T es diagonal con elementos iguales a 1 o -1 .

Cuando ha sido empleada una métrica euclidiana, el problema de rotación no es un problema serio en la configuración o aplicaciones de reducción de datos cuando la solución contiene sólo dos dimensiones, las características de los objetos importantes serán identificables independientemente de la rotación.

La rotación es un problema serio cuando hay más dimensiones ya que las coordenadas pueden no corresponder con las características significativas de los objetos y sería difícil interpretar las dimensiones. Como en el caso de una solución métrica, si la dimensión no rotada corresponde a objetos ordenados interpretables, entonces la rotación es innecesaria. Si no, algunas rotaciones objetivo pueden ser favorables. Si esto también falla para producir una solución interpretable, entonces el usuario se verá forzado a rotar la solución a mano.

2.6.3. Interpretación

En MDS no métrico, la interpretación de la solución procede de la misma manera que en MDS métrico. Esto consiste en gran parte de identificar importantes objetos agrupados u ordenados. Además, una etiqueta o descripción de los atributos de los objetos correspondientes para cada ordenamiento, y una más para describir las características en común para cada objeto agrupado.

2.7. Diagramas de Shepard

Existen tres gráficas ampliamente discutidas en la literatura del análisis de escalamiento multidimensional que proporcionan aspectos importantes del modelo no métrico.

La primera es llamada por Guttman diagrama de imagen. Las distancias $\{\widehat{d}_{rs}\}$ se toman como las abscisas y las disparidades $\{\widehat{\delta}_{rs}\}$ como las ordenadas de los puntos en un plano cartesiano. Hay un punto en la gráfica por cada par de objetos (aunque puede haber menos puntos debido a que los valores \widehat{d}_{rs} y $\widehat{\delta}_{rs}$ de algún par de objetos pueden ser exactamente iguales a los de algún otro par).

Si los datos satisfacen perfectamente el modelo, en cuyo caso la medida de bondad de ajuste será igual a cero, los puntos caerán sobre la recta identidad (la recta con pendiente igual a uno, que pasa por el origen). Mientras más puntos se desvíen de esta línea, peor será el ajuste del modelo a los datos. Los puntos que se desvíen más marcadamente, los llamados “*outliers*”, son los que menos satisfacen el modelo no métrico. Una cuidadosa inspección de éstos puntos ayuda a descubrir desviaciones del modelo teóricamente importantes.

Los puntos de la segunda gráfica tienen como abscisas los valores de las disimilaridades originales $\{\delta_{rs}\}$ y como ordenadas los de las disparidades $\{\widehat{\delta}_{rs}\}$. Nuevamente hay un punto por cada par de objetos (aunque puede haber menos debido a que dos pares de objetos tengan valores exactamente iguales). Esta gráfica es una representación gráfica de la función monótona f que relaciona las disimilaridades con las distancias.

La tercera gráfica también provee una representación gráfica de la función f . Se construye estableciendo las disimilaridades $\{\delta_{rs}\}$ como las abscisas de los puntos y las distancias $\{\widehat{d}_{rs}\}$ como las ordenadas. Hay un punto en la gráfica por cada par de objetos, y puede haber menos.

A estas dos últimas gráficas (δ_{rs} vs. $\widehat{\delta}_{rs}$ y δ_{rs} vs. \widehat{d}_{rs}) se les llama, más generalmente, diagramas de Shepard.

2.8. Problemas de Cálculo

Cualquier algoritmo del análisis de escalamiento multidimensional no métrico comienza con el cálculo de una configuración inicial y la va modificando a través de varias iteraciones. En cada una de ellas las coordenadas de la

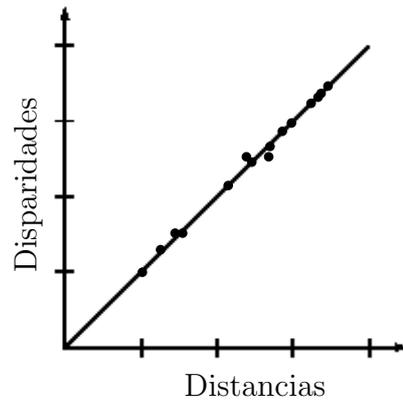


Figura 2.3: Disparidades vs. distancias

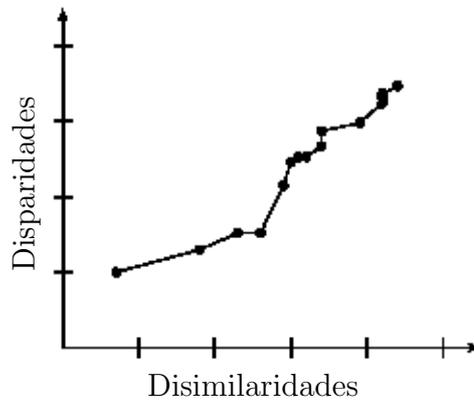


Figura 2.4: Disparidades vs. disimilaridades

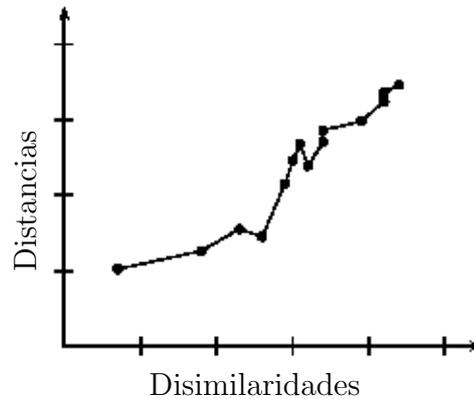


Figura 2.5: Distancias vs. disimilaridades

configuración se ajustan de tal manera que se reduzca el valor de la medida de bondad de ajuste. Si todo va bien, las iteraciones continúan hasta que se obtiene una solución óptima.

Desafortunadamente, hay algunas cosas que pueden salir mal en el proceso iterativo. Estos problemas potenciales son llamados problema del mínimo local, problema de la solución degenerada y problema de convergencia. A continuación se analiza brevemente en que consiste cada uno.

2.8.1. Mínimo Local

Una configuración para la cual la medida de bondad de ajuste es mínima es por definición un mínimo local, sin embargo puede ser que éste no sea un mínimo global. El método de “*steepest descent*” genera soluciones que proporcionan mínimos locales. El problema del mínimo local es el de asegurar que la solución encontrada corresponda efectivamente al mínimo global.

Una primera forma de solucionar el problema es utilizando diferentes configuraciones iniciales para el algoritmo. En principio cada configuración inicial puede derivar en un mínimo local diferente. Considerando el valor más pequeño de éstos como el mínimo global, se tendría como solución la configuración que le corresponde. Aunque esto no garantiza que el mínimo sea un mínimo global.

Otra forma, y al parecer la mejor, de asegurar que la solución obtenida corresponda al mínimo global y no a un mínimo local es utilizar una bue-

na configuración inicial. Algunas variantes de la solución clásica proveen los mejores métodos para obtener tal configuración. Aunque varían, los diversos programas estadísticos comienzan aplicando alguna transformación monótona a las disimilaridades para después aplicarles el análisis clásico y obtener la solución en k dimensiones.

La configuración obtenida por alguna variante de la solución clásica es llamada configuración inicial racional. Los algoritmos que utilizan configuraciones iniciales racionales resultan ser menos propensos al problema del mínimo local, que aquéllos que utilizan cualquier otra configuración inicial no racional.

2.8.2. Solución Degenerada

Una solución se dice degenerada cuando el número de puntos distintos en la configuración es pequeño comparado con el número de objetos a representar. La solución es degenerada porque algunos objetos se colapsan en un sólo punto en la gráfica que los representa.

Se deben revisar las soluciones por degeneración, particularmente aquéllas en una o dos dimensiones. Shepard (1974) establece que las soluciones con un valor de Stress muy cercano o igual a cero son generalmente soluciones degeneradas, por lo cual se debe hacer un esfuerzo extra al revisar alguna solución con un valor de ajuste sospechosamente bajo.

Frecuentemente una solución degenerada significa que la solución correcta debe buscarse en una dimensionalidad mayor, por lo que aquella debe ser desechada.

Por ejemplo, la figura 2.6 muestra una solución degenerada en dos dimensiones. Notemos que los objetos A, B, y C tienen las mismas coordenadas en ambas dimensiones; de aquí que en la gráfica éstos se colapsan en el mismo punto. Similarmente, D, E, y F colapsan dentro de un mismo punto porque tienen exactamente las mismas coordenadas. Los últimos dos objetos, G y H, también colapsan dentro de un mismo punto. Consecuentemente, hay sólo tres puntos distintos en la gráfica aunque hay ocho objetos. Esta solución es degenerada porque el número de puntos distintos en la gráfica (tres) es menor comparado con el número de objetos (ocho).

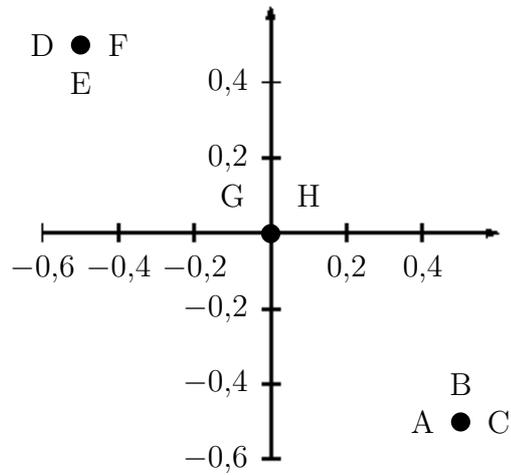


Figura 2.6: Solución degenerada

$$X = \begin{array}{cc} & \begin{array}{cc} I & II \end{array} \\ \begin{pmatrix} 0,5 & -0,5 \\ 0,5 & -0,5 \\ 0,5 & -0,5 \\ -0,5 & 0,5 \\ -0,5 & 0,5 \\ -0,5 & 0,5 \\ 0,0 & 0,0 \\ 0,0 & 0,0 \end{pmatrix} & \begin{array}{l} A \\ B \\ C \\ D \\ E \\ F \\ G \\ H \end{array} \end{array}$$

2.8.3. Convergencia

De los tres problemas de cálculo, el de la falta de convergencia es el más fácil de solucionar, pero probablemente el más frecuente.

Debido a que se debe establecer el número máximo de iteraciones que se pueden llevar a cabo, en algunos casos no hay convergencia, lo que significa que el número de iteraciones requeridas para alcanzar la solución excede al establecido. La mayoría de los programas muestra un mensaje cuando esto ha ocurrido. El problema puede resolverse incrementando el número de iteraciones.

2.9. Ejemplo

En la siguiente discusión, se supondrá que la medida de bondad de ajuste es Stress fórmula uno y se dará una solución euclidiana en dos dimensiones.

Los datos de la tabla 13 son de un estudio realizado por Rounds et al. (1979). Esta tabla muestra las correlaciones entre seis escalas de el Inventario de Preferencias Vocacionales (Holland, 1965) en una muestra de 1234 hombres. Cada escala corresponde a una de seis tipos de vocaciones en la teoría de Holland (1973) de la elección de carrera.

Tabla 13. Correlaciones del inventario de preferencias vocacionales.

	R	I	A	S	E	C
Realista	1	0,46	0,16	0,21	0,30	0,36
Investigador	0,46	1	0,34	0,30	0,16	0,16
Artista	0,16	0,34	1	0,42	0,35	0,11
Social	0,21	0,30	0,42	1	0,54	0,38
Emprendedor	0,30	0,16	0,35	0,54	1	0,68
Convencional	0,36	0,16	0,11	0,38	0,68	1

Para hacer un análisis utilizando disimilaridades en vez de similaridades, las correlaciones en la tabla 13 serán convertidas en disimilaridades usando la siguiente transformación:

$$\delta_{rs} = (1 - r_{rs})^{\frac{1}{2}} \quad (2.16)$$

Puede usarse alguna otra ecuación que invierta el orden de los rangos en lugar de la ecuación (2.16). La tabla 14 muestra la matriz de disimilaridades obtenidas.

Tabla 14. Disimilaridades.

	R	I	A	S	E	C
Realista	0					
Investigador	0,734	0				
Artista	0,916	0,812	0			
Social	0,888	0,836	0,761	0		
Emprendedor	0,836	0,916	0,806	0,678	0	
Convencional	0,800	0,916	0,943	0,787	0,565	0

Configuración inicial

Para obtener la configuración inicial se aplicó a los datos de las disimilitudes en la tabla 14 el algoritmo de Torgerson. La matriz etiquetada como “Distancias estimadas no estandarizadas” en la tabla 17 muestra las distancias euclidianas estimadas calculadas a partir de las coordenadas estimadas no estandarizadas de la tabla 16. Estas distancias estimadas fueron calculadas como:

$$\hat{d}_{rs}^o = \left[\sum_{i=1}^k (\hat{x}_{ri}^o - \hat{x}_{si}^o)^2 \right]^{1/2} \quad (2.17)$$

En la ecuación (2.17) \hat{x}_{ri}^o y \hat{x}_{si}^o se refieren a las coordenadas estimadas de la configuración inicial, y \hat{d}_{rs}^o se refiere a la distancia estimada inicial.

Tabla 15. Productos escalares.

	R	I	A	S	E	C
Realista	0,305					
Investigador	0,041	0,316				
Artista	-0,105	-0,009	0,323			
Social	-0,119	-0,068	-0,005	0,245		
Emprendedor	-0,089	-0,153	-0,055	0,001	0,215	
Convencional	-0,031	-0,125	-0,147	-0,051	0,083	0,271

Tabla 16. Coordenadas no estandarizadas.

	I	II
Realista	-0,2015	0,4168
Investigador	-0,4522	0,1016
Artista	-0,2034	-0,4348
Social	0,1105	-0,2653
Emprendedor	0,3619	-0,0380
Convencional	0,3847	0,2198

Tabla 17. Distancias estimadas no estandarizadas.

	R	I	A	S	E	C
Realista	0					
Investigador	0,402	0				
Artista	0,851	0,591	0			
Social	0,750	0,671	0,356	0		
Emprendedor	0,724	0,826	0,690	0,338	0	
Convencional	0,618	0,845	0,880	0,557	0,258	0

Coordenadas y distancias estandarizadas

La tabla 19 muestra las distancias estimadas estandarizadas obtenidas multiplicando cada distancia estimada no estandarizada en la tabla 17 por la constante de estandarización de manera que la suma de cuadrados de las distancias estimadas se igualará a 1,0. Se multiplica cada coordenada estimada no estandarizada en la tabla 16 por la misma constante dada, las coordenadas estimadas estandarizadas están en la tabla 18.

En este caso la constante de estandarización está dada por:

$$\begin{aligned}
 ce &= \frac{1}{\left(\sum_{(r,s)} d_{rs}^2\right)^{1/2}} = \frac{1}{(6,4257)^{1/2}} \\
 &= \frac{1}{2,5349} = 0,39449
 \end{aligned}$$

Tabla 18. Coordenadas estandarizadas.

	I	II
Realista	-0,0795	0,1644
Investigador	-0,1784	0,0401
Artista	-0,0803	-0,1715
Social	0,0436	-0,1047
Emprendedor	0,1428	-0,0150
Convencional	0,1518	0,0867

Tabla 19. Distancias estandarizadas.

	R	I	A	S	E	C
Realista	0					
Investigador	0,158	0				
Artista	0,336	0,233	0			
Social	0,395	0,265	0,140	0		
Emprendedor	0,285	0,325	0,272	0,133	0	
Convencional	0,244	0,333	0,347	0,219	0,102	0

Fase no métrica

En nuestro ejemplo, las disparidades estimadas para la primera iteración de la fase no métrica son las distancias estandarizadas estimadas en la tabla 19. La columna 4 de la tabla 20 contiene las distancias estimadas $\{\widehat{d}_{rs}^o\}$ correspondientes a cada punto $\{\widehat{\delta}_{rs}\}$.

El primer paso de la fase no métrica es arreglar las disimilaridades en una columna, así que los datos en la columna están en orden ascendente. La columna 3 de la tabla 20 contiene las disimilaridades de la tabla 14 ordenadas en forma ascendente.

En la tabla 20, δ_{SI} ha sido puesta antes de δ_{ER} a pesar de que los dos datos son idénticos ($\delta_{SI} = \delta_{ER} = 0,836$) porque la distancia estimada (columna 4) para el primero, $\widehat{d}_{SI}^o = 0,265$, es menor que para el segundo, $\widehat{d}_{ER}^o = 0,285$. Similarmente δ_{EI} precede a δ_{CI} que precede a δ_{AR} en la tabla 20 aunque las tres disimilaridades son iguales a 0,916 porque $\widehat{d}_{EI}^o < \widehat{d}_{CI}^o < \widehat{d}_{AR}^o$.

Después de arreglar las disimilaridades en orden ascendente, el algoritmo comienza una serie de pasos por medio de los datos. La columna 4 de la tabla 20 contiene las distancias estandarizadas estimadas a partir de la configuración inicial (tabla 19) que son las disparidades que inician el primer paso de la fase no métrica para la primera iteración. El paso comienza con la división de las disparidades estimadas en la tabla 20, no hay valores de empate, así cada disparidad constituye su propio bloque.

El resto de cada paso consiste en comparar bloques adyacentes. Sea m ($m = 1, \dots, 15$) un subíndice designado al bloque del menor ($m = 1$) al mayor ($m = 15$). La designación del bloque en la tabla 20 se muestra en la columna 5. Comenzando con $m = 1$, el elemento en el m -ésimo bloque se compara a los elementos en el $(m + 1)$ -ésimo bloque.

Tabla 20.

			Antes de agrupar			Después de un agrupamiento		Después de dos agrupamientos		Distancias
			Disparidades	Bloque	Disparidades	Bloque	Disparidades	Bloque		
C	E	0,5657	0,1021	1	0,1021	1	0,1021	1	0,1028	
E	S	0,6782	0,1337	2	0,1337	2	0,1337	2	0,1312	
I	R	0,7348	0,1589	3	0,1498	3	0,1498	3	0,1550	
S	A	0,7616	0,1407	4	0,1498	3	0,1498	3	0,1387	
C	S	0,7874	0,2198	5	0,2198	4	0,2198	4	0,2175	
C	R	0,8000	0,2440	6	0,2440	5	0,2440	5	0,2457	
E	A	0,8062	0,2725	7	0,2725	6	0,2529	6	0,2685	
A	I	0,8124	0,2332	8	0,2332	7	0,2529	6	0,2377	
S	I	0,8367	0,2650	9	0,2650	8	0,2650	7	0,2667	
E	R	0,8367	0,2857	10	0,2857	9	0,2857	8	0,2880	
S	R	0,8888	0,2959	11	0,2959	10	0,2959	9	0,2956	
E	I	0,9165	0,3259	12	0,3259	11	0,3259	10	0,3267	
C	I	0,9165	0,3334	13	0,3334	12	0,3334	11	0,3327	
A	R	0,9165	0,3360	14	0,3360	13	0,3360	12	0,3356	
C	A	0,9434	0,3472	15	0,3472	14	0,3472	13	0,3438	

En la columna 4 de la tabla 20, los elementos en el primer bloque son más pequeños que los elementos en el segundo bloque, así nosotros podemos proseguir a una comparación del segundo y el tercer bloque. El elemento en el segundo bloque es más pequeño que en el tercer bloque, así continuamos con los bloques 3 y 4. El elemento en el bloque 3 es mayor que el del 4, así estos dos elementos se reemplazan por su media aritmética, $(0,158 + 0,140)/2 = 0,149$. La columna 6 muestra el reemplazo. Estos dos elementos llegan a ser un nuevo bloque (bloque 3), y el resto de los bloques se enumera a partir de éste. Este paso continúa sobre los elementos en la columna 6.

Los elementos en el nuevo bloque 3, columna 6, son menores que en el bloque 4, el bloque 4 es menor que el bloque 5, y el bloque 5 es menor que el bloque 6. El bloque 6, sin embargo, es mayor que el bloque 7. Los elementos en los bloques 6 y 7 se reemplazan por su media aritmética, $(0,272 + 0,233)/2 = 0,252$, y se agrupan dentro de un nuevo bloque. Esta nueva mezcla del bloque 6 aparece en la columna 8. El paso continúa otra vez con los elementos en la columna 8.

Los elementos en el nuevo bloque 6 son más pequeños que los elementos en el 7, el bloque 7 es más pequeño que el 8, y así sucesivamente. La comparación del resto de los bloques adyacentes revela que no se requieren más agrupamientos. Completando la comparación entre los últimos dos bloques adyacentes, 12 y 13, el paso llega a su fin.

Los números en la columna 8 son las disparidades estimadas finales para la $(c + 1)$ -ésima iteración (primera iteración) y serán utilizadas dentro de la fase no métrica de esta iteración.

Cuando el paso se completa, las disparidades satisfacen la expresión (2.6).

Fase métrica

La tabla 18 contiene las coordenadas estandarizadas estimadas $\{\hat{x}_{ri}^o\}$. La columna 8 de la tabla 20 contiene las disparidades estimadas, $\{\hat{\delta}_{rs}^1\}$, y la columna 4 contiene las distancias estimadas, $\{\hat{d}_{rs}^o\}$. Por tanto, las nuevas coordenadas estimadas $\{\hat{x}_{ri}^1\}$ pueden ser calculadas usando la ecuación (2.15) y la información en las tablas 18 y 20.

Por ejemplo, las nueva coordenada estimada a lo largo de la dimensión I para la escala Realista es:

$$\begin{aligned}
\hat{x}_{R1}^1 &= \hat{x}_{R1}^o - \frac{1}{6} \sum_{j=1}^6 \left(1 - \frac{\hat{\delta}_{Rj}^1}{\hat{d}_{Rj}^o} \right) (\hat{x}_{R1}^o - \hat{x}_{j1}^o) \\
&= -0,079 - \frac{1}{6} \left(\begin{array}{l} + \left[1 - \frac{0}{0} \right] * [-0,079 - (-0,079)] \\ + \left[1 - \frac{0,154}{0,158} \right] * [-0,079 - (-0,178)] \\ + \left[1 - \frac{0,336}{0,333} \right] * [-0,079 - (-0,080)] \\ + \left[1 - \frac{0,295}{0,295} \right] * [-0,079 - 0,043] \\ + \left[1 - \frac{0,285}{0,285} \right] * [-0,079 - 0,142] \\ + \left[1 - \frac{0,244}{0,244} \right] * [-0,079 - 0,151] \end{array} \right) \\
&= -0,0804
\end{aligned}$$

La tabla 21 muestra las nuevas coordenadas estimadas $\{\hat{x}_{ri}^1\}$ obtenidas en la fase métrica de la primera iteración. La columna 10 de la tabla 20 muestra las nuevas distancias estimadas $\{\hat{d}_{rs}^1\}$ calculadas a partir de las coordenadas de la tabla 21 de acuerdo a la ecuación (2.15). Para la primera iteración, el Stress fórmula uno se calcula substituyendo las disparidades de la columna 8 y las distancias estimadas de la columna 10 de la tabla 20 en la ecuación (2.14). S_1 es igual a 0,023.

Tabla 21. Coordenadas de la primera iteración.

	Dimensión I	Dimensión II
Realista	-0,0812	0,1648
Investigador	-0,1781	0,0439
Artista	-0,0761	-0,1708
Social	0,0447	-0,1026
Emprendedor	0,1431	-0,1026
Convencional	0,1518	0,0867

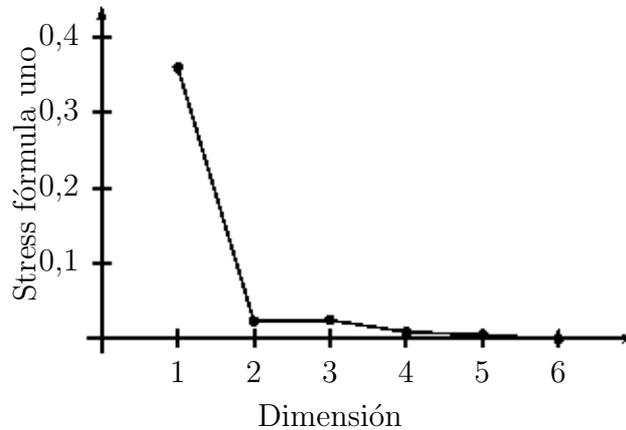
Fase no métrica alternativa

Tabla 22. Fase no métrica de Guttman.

r	s	δ_{rs} Disimilaridades	\hat{d}_{rs}^0 Distancias	$\hat{\delta}_{rs}^1$ Disparidades
E	C	0,5657	0,1021	0,1021
S	E	0,6782	0,1337	0,1337
R	I	0,7348	0,1589	0,1407
A	S	0,7616	0,1407	0,1589
S	C	0,7874	0,2198	0,2198
R	C	0,8000	0,2440	0,2332
A	E	0,8062	0,2725	0,2440
I	A	0,8124	0,2332	0,2650
I	S	0,8367	0,2650	0,2725
R	E	0,8367	0,2857	0,2857
R	S	0,8880	0,2959	0,2959
I	E	0,9165	0,3259	0,3259
I	C	0,9165	0,3334	0,3334
R	A	0,9165	0,3360	0,3360
A	C	0,9434	0,3472	0,3472

La tabla 22 muestra los resultados de la fase no métrica de Guttman. Para ilustrar el método de Guttman, otra vez considere los datos en la tabla 20. La columna 3 contiene las disimilaridades $\{\delta_{rs}\}$ y la columna 4 contiene las distancias estimadas $\{\hat{d}_{rs}^0\}$ a partir de las cuales se calculan las disparidades en la primera iteración. Los puntos más pequeños en la columna 3 corresponden al par de vocaciones (C, E): $\delta_{CE} = 0,565$. La distancia más pequeña en la columna 4 es 0,102, así en la primera iteración $\hat{\delta}_{CE}^1$ será igual a 0,102.

La segunda disimilaridad más pequeña en la columna 3 corresponde al par de vocaciones (E, S), 0,678. La segunda distancia más pequeña estimada en la columna 4 es 0,133, así $\hat{\delta}_{ES}^1$ es igual a 0,133. La tercer disimilaridad más pequeña en la columna 3 corresponde al par de vocaciones (I, R), 0,734. La tercer distancia más pequeña estimada en la columna 4 es 0,140, así la disparidad $\hat{\delta}_{IR}^1$ será igual a 0,140 en la primera iteración, y así sucesivamente hasta llegar a la disimilaridad mayor en la columna 3 que corresponde al par de vocaciones (C, A), 0,943. La mayor distancia estimada en la columna 4 es

Figura 2.7: Stress vs. k

0,347, así $\hat{\delta}_{CA}$ es igual a 0,347.

Otra manera de describir cómo trabaja la fase no métrica de Guttman en este ejemplo es permutar las distancias estimadas de la columna 4 de la tabla 20 de manera que estén en orden ascendente; después de la permutación, la columna 4 se leerá como: 0,102, 0,133, 0,140, 0,158, 0,219, 0,233, 0,244, 0,265, 0,272, 0,285, 0,295, 0,325, 0,333, 0,336, 0,347. Estas distancias permutadas estimadas constituyen las disparidades calculadas de la primera iteración.

Dimensionalidad

En la figura 2.7, se presenta un codo en la dimensión 2, sugiriendo que la solución apropiada puede ser una configuración en dos dimensiones.

Rotación

La tabla 23 muestra una solución en dos dimensiones de las escalas del Inventario de Preferencias Vocacionales. Estas coordenadas están graficadas en la figura 2.8. En esta configuración hexagonal es fácil visualizar una rotación y aunque en este ejemplo la rotación no es un resultado importante, aplicaremos una rotación de 45° .

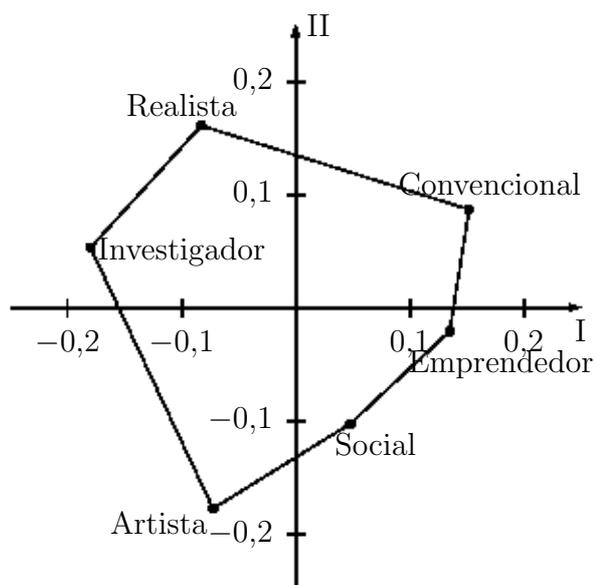


Figura 2.8: Configuración del MDS

Tabla 23. Coordenadas estimadas X .

	Dimensión I	Dimensión II
Realista	-0,0828	0,1609
Investigador	-0,1797	0,0530
Artista	-0,0724	-0,1775
Social	0,0480	-0,1024
Emprendedor	0,1349	-0,0209
Convencional	0,1520	0,0868

$$\begin{aligned}
 T &= \begin{matrix} I \\ II \end{matrix} \begin{matrix} 1 & 2 \\ \cos(45^\circ) & \cos(135^\circ) \\ \cos(315^\circ) & \cos(45^\circ) \end{matrix} \\
 &= \begin{matrix} I \\ II \end{matrix} \begin{matrix} 1 & 2 \\ \cos(45^\circ) & -\text{sen}(45^\circ) \\ \text{sen}(45^\circ) & \cos(45^\circ) \end{matrix} \\
 &= \begin{pmatrix} 0,71 & -0,71 \\ 0,71 & 0,71 \end{pmatrix}
 \end{aligned}$$

$$\begin{aligned}
\hat{X}^* &= XT \\
&= \begin{pmatrix} -0,0828 & 0,1609 \\ -0,1797 & 0,0530 \\ -0,0724 & -0,1775 \\ 0,0480 & -0,1024 \\ 0,1349 & -0,0209 \\ 0,1520 & 0,0868 \end{pmatrix} \begin{pmatrix} 0,71 & -0,71 \\ 0,71 & 0,71 \end{pmatrix} \\
&= \begin{pmatrix} 0,0554 & 0,1730 \\ -0,0899 & 0,1652 \\ -0,1774 & -0,0746 \\ -0,0386 & -0,1067 \\ 0,0809 & -0,1106 \\ 0,1696 & -0,0463 \end{pmatrix}
\end{aligned}$$

En la figura 2.9 se presenta la configuración rotada 45° y sus coordenadas se muestran en la tabla 24.

Tabla 24. Coordenadas rotadas 45° .

	Dimensión 1	Dimensión 2
Realista	0,0554	0,1730
Investigador	-0,0899	0,1652
Artista	-0,1774	-0,0746
Social	-0,0386	-0,1067
Emprendedor	0,0809	-0,1106
Convencional	0,1696	-0,0463

Interpretación

Haciendo una inspección de la figura 2.8, cuyas coordenadas se presentan en la tabla 23, los puntos muestran el patrón de un polígono hexagonal, cuyos lados no son del mismo tamaño. La teoría de Holland es que el interés de las personas en los tipos de ocupaciones adyacentes en el hexágono es más similar que el interés de las personas en los tipos de ocupaciones más distantes; los tipos de ocupaciones opuestas en el hexágono son menos similares.

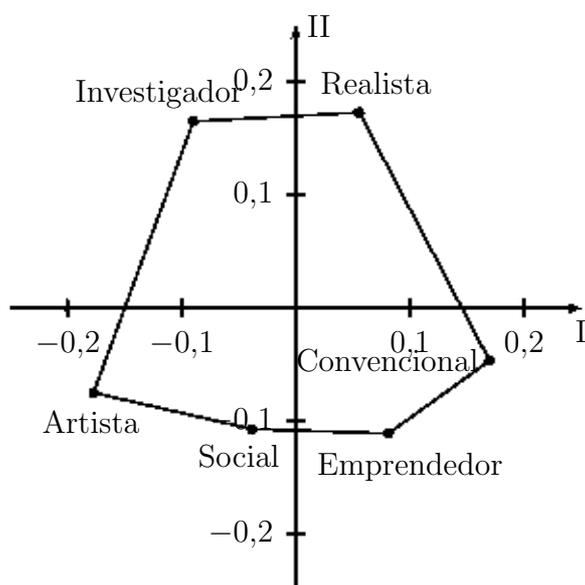


Figura 2.9: Configuración rotada 45°.

Diagramas de Shepard

Tabla 25. Disimilaridades, disparidades y distancias.

		Disimilaridades	Disparidades	Distancias
I	R	0.7348	0.1459	0.1450
A	R	0.9165	0.3384	0.3385
A	I	0.8124	0.2530	0.2543
S	R	0.8888	0.2941	0.2939
S	I	0.8367	0.2750	0.2757
S	A	0.7616	0.1417	0.1419
E	R	0.8367	0.2838	0.2836
E	I	0.9165	0.3234	0.3232
E	A	0.8062	0.2605	0.2598
E	S	0.6782	0.1201	0.1191
C	R	0.8000	0.2461	0.2462
C	I	0.9165	0.3335	0.3335
C	A	0.9434	0.3467	0.3467
C	S	0.7874	0.2162	0.2159
C	E	0.5657	0.1086	0.1091

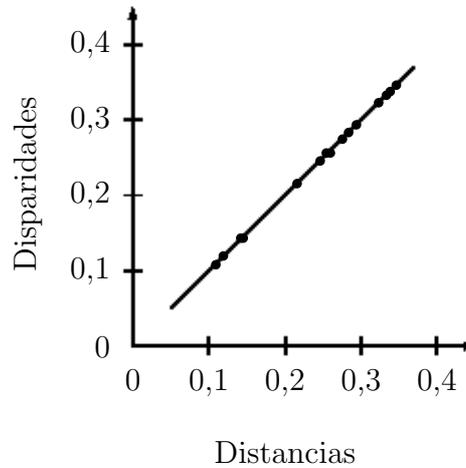


Figura 2.10: Disparidades vs. distancias

Los resultados en la tabla 25 han sido obtenidos de las correlaciones de la escala del Inventario de Preferencias Vocacionales. La columna uno y dos indican qué par de la escala corresponde a cada renglón de la tabla 25. Las columnas 3, 4 y 5, respectivamente, son las disimilaridades, las disparidades, y las distancias de cada par de la escala.

La figura 2.10 muestra los puntos de las distancias contra las disparidades, los datos satisfacen el modelo perfectamente, en cuyo caso la medida de bondad de ajuste converge a cero.

El segundo diagrama de puntos de interés es una gráfica de las disimilaridades contra disparidades, la figura 2.11 muestra los puntos. El tercer tipo de diagrama se muestra en la figura 2.12, disimilaridades contra distancias. Se ve una relación monótona creciente, lo cual es lo que deseamos.

Ahora calculamos el Stress fórmula uno:

$$\begin{aligned}
 S_1 &= \left[\frac{\sum_{(i,j)} (\hat{\delta}_{ij} - \hat{d}_{ij})^2}{\sum_{(i,j)} \hat{d}_{ij}^2} \right]^{\frac{1}{2}} \\
 &= \left(\frac{0,0000221}{0,999} \right)^{\frac{1}{2}} = 0,0047
 \end{aligned}$$

el cual nos indica que hay un buen ajuste de los datos.

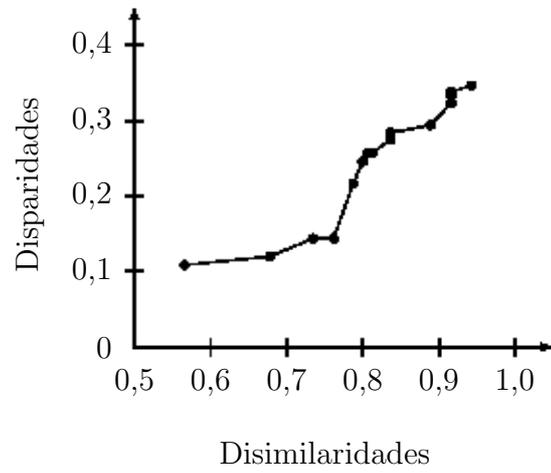


Figura 2.11: Disparidades vs. disimilaridades

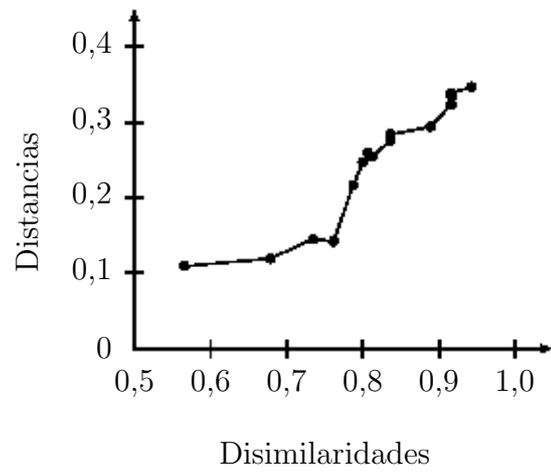


Figura 2.12: Distancias vs. disimilaridades

Capítulo 3

Datos y Proximidades

3.1. Tipos de Datos

Uno de los tipos más comunes de registro para los datos utilizado por los ecólogos es contar el número de veces que ocurre un evento. Esta cuenta puede ser el número de individuos de una especie particular que ocurre en un cuadro¹, o el número de veces que una especie se encuentra en la vecindad de otras especies. En el primer caso el conteo es una medida de abundancia, mientras que en el segundo caso es una medida de asociación. Otros tipos de medidas de abundancia son frecuentemente usadas en estudios botánicos como biomasa², densidad y cobertura; estas medidas de abundancia están generalmente expresadas en términos de la unidad del área y son independientes del tamaño de muestra. Otra medida es la frecuencia de ocurrencia, es decir, el porcentaje de la ocurrencia de la muestra, que se relaciona con el tamaño de la muestra. La escala Braun-Blanquet intenta probar una estimación combinada única de la cobertura de plantas y el número de plantas individuales. Otras escalas se basan en su porcentaje de cobertura o número de individuos y algunas veces usan intervalos de clase logarítmica (tabla 26).

¹Los cuadros, en inglés *quadrat*, son áreas de tamaño y forma fija. Lugar dentro del marco muestral que algunas veces es aleatorio y algunas veces es sistemático.

²La biomasa es la cantidad total de materia viviente en una unidad ecológica determinada, por unidad de superficie o volumen.

Tabla 26. Tres escalas usadas en botánica: la escala de cobertura-abundancia de Braun-Blanquet usa aproximadamente intervalos de clase aditiva para categorizar el % de cobertura, mientras que la escala H-S-D (Hult-Sernander-Du Rietz) adopta intervalos de clase logarítmica. Lowe (1984) similarmente adopta una escala de clase logarítmica en la determinación de abundancia de insectos.

Escala	Braun-Blanquet	H-S-D	Lowe
r	Uno o nuevos individuos.		
+	Ocasionales y menores que 5 % de la cobertura del total del area.		
1	Abundantes y con muy poca cobertura, o menos abundantes con gran cobertura, pero siempre menores que 5 % del total del area.	< 6 % de cobertura (1/16 area)	1 individuo.
2	Muy abundantes y menor que 5 % de cobertura, o 5-25 % de cobertura.	6-12.5 % de cobertura.	2-3 individuos.
3	25-50 % de cobertura, prescindiendo de el número de individuos.	12.5-25 % de cobertura.	4-7 individuos
4	50-75 % de cobertura.	25-50 % de cobertura.	8-15 individuos.
5	75-100 % de cobertura.	50-100 % de coberura.	16-31 individuos.
6	—	—	32-63 individuos.

Una alternativa para registrar la abundancia es considerar sólo los rangos de acuerdo a la abundancia de las especies. El rango de las especies también puede basarse sobre alguna medida de abundancia, y medidas diferentes pueden producir rangos diferentes cuando especies de diferentes tamaños o conductas están incluidas. Podríamos considerar la formación de un registro con base en la presencia o ausencia de cada especie en cada sitio.

La elección del tipo de datos coleccionados dependerá del costo relativo de la colección y el nivel de precisión requerido en el análisis. Para estudios biogeográficos de escalas grandes, donde las muestras son muy heterogéneas, un registro de la presencia o ausencia será suficiente; pero para la comparación local de tipos de vegetación una medida cuantitativa de abundancia será más eficiente. Otra vez, la elección de la medida cuantitativa dependerá del objetivo de la medición y la técnica factible.

En estudios de comunidades, los primeros datos usualmente serán abundancias de especies, pero los ecólogos también pueden incorporar otros datos de tipo morfológicos³, conductuales o ambientales relacionados con abundancias. Por ejemplo, si una especie ocurre sólo en una porción de sitios en la muestra, ésta es de interés para averiguar si estos sitios tienen alguna asociación ambiental. Similarmente, si un número de especies se encuentra agrupado de acuerdo a la distribución de su hábitat, este agrupamiento puede estar relacionado a las similitudes en especies morfológicas. Estos datos auxiliares pueden ser de varios tipos diferentes. Pueden ser cuantitativos, como pH del suelo, gradiente⁴ y orientación de un sitio en declive, o a veces desde un disturbio ambiental, pero tienen escalas de medidas diferentes, así que puede ser deseable alguna forma de estandarización antes del análisis. Alternativamente, pueden ser cualitativos o nominales que caen en un número de estados discretos que no tienen orden definido, tal como suelo o tipo de hábitat, método de dirección, patrón de alas o color de flor. Un caso especial de un atributo cualitativo es el que tiene sólo dos estados observables, como sexo. Hay los llamados atributos binarios y son convencionalmente cifrados como 0 y 1.

Los datos ausentes requieren consideraciones especiales y pueden ser de diferentes tipos. Un dato puede no haber sido registrado por descuido, o el registro puede haber sido subsecuentemente perdido. Segundo, en inspecciones grandes algunos registros pueden ser más detallados que otros; por ejemplo no todos los campos de trabajo pueden ser capaces de identificar todos los niveles de especies. En un tercer caso, un carácter puede no ser registrado porque éste depende de la presencia de otro carácter, por ejemplo, en una clasificación taxonómica⁵, un insecto puede ser clasificado teniendo o

³La morfología se encarga del estudio de las formas externas e internas y de la estructura de los seres orgánicos, sus transformaciones y las causas y leyes que las determinan.

⁴El gradiente en ecología es la relación de la intensidad metabólica entre un lugar y otro de un organismo.

⁵La Taxonomía es la parte de la Historia Natural que trata de la clasificación sistemática

no teniendo alas, esos insectos con alas pueden ser adicionalmente caracterizados de acuerdo al patrón de alas, pero este segundo carácter es inaplicable para aquellos insectos sin alas y algunas veces son tratados como ausentes. La abundancia de especies algunas veces se considera como un caso especial de una variable dependiente, ya que ésta sólo es valorada cuando la especie aparece en la muestra. La ausencia de las muestras no necesariamente implica la ausencia de parte de la comunidad ya que la especie puede estar presente en una frecuencia tan pequeña que no aparece en una fracción de la muestra. Ésta es una consideración importante cuando se usan las muestras para comparar las comunidades de diferentes tamaños. Pueden aparecer diferencias en el perfil de las especies pueden aparecer simplemente porque las muestras grandes recolectan las especies raras.

3.2. Formas de Datos

En la mayoría de las situaciones el conjunto de datos ecológicos se presenta por medio de arreglos que forman una tabla o matriz de, digamos, especies por sitios. Entonces la abundancia de la r -ésima especie en el s -ésimo sitio, denotado por x_{rs} , está localizada en la celda en el r -ésimo renglón y la s -ésima columna de la matriz $X_{n \times p}$. La matriz X tiene n renglones correspondientes al número de especies, y p columnas correspondientes al número de sitios. Cuando se tienen datos ambientales para un número de variables en varios sitios, podemos construir una matriz Z donde z_{rs} es el valor del dato para la s -ésima medida de la variable ambiental en el r -ésimo sitio, por convención, el renglón ahora corresponde a los sitios y las columnas a las variables.

En estos dos ejemplos, los renglones y columnas de la matriz de datos han sido clasificados por diferentes conjuntos de objetos: especies y sitios para los datos de abundancia, y sitios y variables para los datos ambientales. Como una alternativa, los datos pueden estar en forma de una matriz cuadrada donde los renglones y columnas están clasificados por el mismo conjunto de objetos. Tales datos son medidas de asociación entre pares de objetos, por ejemplo, el número de veces que dos especies de plantas particulares ocurren conjuntamente en un sitio; aquí la matriz contiene información para todos los pares de especies.

Como este último ejemplo, la mayoría de tales conjuntos de datos, en que la matriz de asociación es simétrica, la asociación entre la r -ésima y la s -ésima de los seres, basándose en las diferencias que existen entre ellos.

s -ésima especie es la misma que la asociación entre la s -ésima y la r -ésima especie: $x_{rs} = x_{sr}$.

En ecología, las matrices de asociación no se obtienen directamente de la muestra observada. Sin embargo, son particularmente usadas para ciertos tipos de análisis de datos y son por tanto frecuentemente construidas a partir de una matriz de datos rectangular. Por ejemplo, una matriz de datos de abundancia de especies por sitios puede usarse para construir una matriz simétrica de las asociaciones de las especies.

3.3. Estandarización y Transformación de Datos

Las matrices de datos ecológicas algunas veces requieren alguna forma de estandarización interna de manera que algún efecto dominante que no sea de interés primario pueda ser removido antes del análisis. En algunos casos una estandarización está implícita en el análisis particular, pero usualmente esto queda a decisión de los ecólogos.

La forma más común de estandarización en una matriz de especies por sitios está dada por las diferencias en las medias de especies o sitios. Estas son usualmente de menor interés que las interrelacionadas entre especies y sitios y pueden ser simplemente artefactos de alguna elección arbitraria de medida de abundancia o método de muestreo. Por ejemplo, el número de especies diferentes de polillas en una trampa de luz puede decirnos más acerca de las diferentes fotosíntesis de la especie que acerca de su frecuencia relativa en el sitio. Similarmente, la eficiencia del método muestral puede variar de un sitio a otro, bajo las diferentes condiciones ambientales: por ejemplo, los dispositivos del muestreo del bentos⁶ tales como tubos colectores se sabe que varían en eficiencia cuando se usan sobre diferentes sustratos. Estos efectos pueden ser removidos centrando los valores de los datos alrededor de su media de especies o sitios,

$$x_{rs} \longrightarrow x_{rs} - x_r, \quad x_{rs} \longrightarrow x_{rs} - x_{.s},$$

o centrando para especies y sitios,

$$x_{rs} \longrightarrow x_{rs} - x_r - x_{.s} + x_{..}$$

⁶El bentos es el conjunto de los seres que viven en el fondo del mar.

aquí $x_{r.} = \frac{1}{p} \sum_{s=1}^p x_{rs}$, $x_{.s} = \frac{1}{n} \sum_{r=1}^n x_{rs}$ y $x_{..} = \frac{1}{np} \sum_{r=1}^n \sum_{s=1}^p x_{rs}$

Alternativamente, si los cambios en la abundancia de la población se suponen primero multiplicados que sumados, la abundancia puede ser expresada como proporciones estandarizando las especies o sitios totales,

$$x_{rs} \longrightarrow \frac{x_{rs}}{px_{r.}} \quad x_{rs} \longrightarrow \frac{x_{rs}}{nx_{.s}}.$$

Una estandarización algunas veces aplicada después de centrar para asegurar que todas las columnas de la matriz de datos tienen varianzas iguales:

$$x_{rs} \longrightarrow \frac{x_{rs} - x_{.s}}{(\sum_r (x_{rs} - x_{.s})^2 / n)^{1/2}}.$$

Esto puede ser necesario cuando trabajamos con datos ambientales o morfológicos que son diferentes, y algunas veces tienen escalas de medida arbitrarias. Esta no puede ser recomendada y es raramente justificada cuando se trabajan con datos de abundancia. Considere, por ejemplo, dos especies, una de las cuales muestra variación geográfica considerable en su abundancia, mientras la otra no y ocurren con números similares en todos los sitios, si estas dos especies muestran variación similar dentro de los sitios se esperará que la primera especie sea un mejor indicador de las diferencias entre los sitios y se deseará que se tome en cuenta en el análisis, sin embargo, estandarizando a una varianza constante bajo los sitios elimina diferencias reales efectivas en la capacidad discriminante de las especies.

Para asociar la ocurrencia natural se requiere alguna forma de estandarización. Una medida de asociación que está menos afectada por la abundancia bajo todos los datos puede ser obtenida estandarizando a_{rs} (el número de ocurrencia conjunta de especies r con especies s) con, por ejemplo, el número promedio de ocurrencia de las dos especies,

$$a_{rs} \longrightarrow \frac{2a_{rs}}{a_{rr} + a_{ss}}.$$

La medida de estandarización de asociación ahora varía de 0, para especies que nunca ocurren conjuntamente en el mismo terreno, a 1 (o 100%), para especies que siempre ocurren conjuntamente, sin hacer caso del número actual de terrenos en las que éstas ocurren.

Una de las estandarizaciones interesantes es aquella en la cual los datos son independientes de la escala, sin embargo, puede lograrse (excepto por

una constante aditiva) sin recurrir a información interna transformando los valores originales a logaritmos. La transformación logaritmo,

$$x_{rs} \longrightarrow \log_b(x_{rs} + c),$$

tiene un lugar predominante en ecología cuantitativa: tiene el efecto de comprimir los valores extremos de la escala de medida y así reducir la importancia de los valores relativamente grandes a valores pequeños en la matriz de datos. La base de los logaritmos que generalmente se elige es 2, e, o 10. La constante c usualmente se elige para tomar un valor positivo pequeño así como para reducir diferencias entre valores pequeños y, particularmente, para poder trabajar con valores cero. Para los datos de abundancia en los que X es un conteo o un porcentaje de frecuencia, usualmente se toma la constante $c = 1$ o $c = \frac{1}{2}$.

Healy y Taylor (1962) consideran una familia general de transformaciones potencia

$$x \longrightarrow x^{1-k/2}$$

donde $k \neq 2$ es elegida para estabilizar las varianzas de los valores transformados cuando se argumentan con su media. El caso especial $k = 2$ se representa por una transformación logaritmo, mientras que $k = 1$ da una transformación de raíz cuadrada y $k = 4$ una transformación del inverso del valor. En trabajos ecológicos, la elección de un valor no entero de k generalmente estará fuera de lugar en vista del refinamiento del análisis subsecuente; por cierto, alternativas para la transformación logaritmo son poco usadas. Muchas escalas para la valoración rápida de abundancia usan intervalos de clase logarítmica y no es necesario una transformación adicional. Cuando la abundancia se describe por orden de rangos, puede ser apropiado usar una transformación *rankit* (Fisher y Yates, 1963). Esta tiene el efecto de extender la escala a sus extremos del rango y así separar en forma creciente a las especies entre las de menor rango y entre las especies de mayor rango.

3.4. Tipos de Variables y Escalas de Medición

3.4.1. Tipos de Variables

Las variables categóricas son las que clasifican a cada una de las unidades experimentales dentro de cierta categoría. Sus diferentes niveles (categorías)

difieren en cualidad, no en cantidad. Las categorías deben ser exhaustivas y mutuamente excluyentes.

Las variables nominales son variables categóricas en donde no existe una relación de orden entre sus categorías. Las variables ordinales son variables categóricas cuyas categorías están ordenadas, además las distancias absolutas entre categorías son desconocidas.

Las variables cuantitativas son aquellas que surgen de un proceso de medición o conteo sobre las unidades experimentales.

Una variable cuantitativa es continua si entre dos posibles valores existe una infinidad de valores. Una variable cuantitativa es discreta si entre dos posibles valores consecutivos no existe otro valor (surgen, en general, de un proceso de conteo).

3.4.2. Escalas de Medición

Las variables nominales se dicen medidas en escala nominal y las variables ordinales se dicen medidas en escala ordinal.

Una variable cuantitativa se dice medida en escala de intervalo si se conoce la distancia que separa a dos observaciones, independientemente del lugar donde el cero es colocado (el cero es arbitrario). Una variable cuantitativa se dice medida en escala de razón si se conoce la distancia que separa a dos observaciones, sólo que para éstas se tiene un origen verdadero (el cero no es arbitrario).

La diferencia primordial entre la escala de intervalo y la escala de razón, es que en la primera no se conservan las proporciones si se cambia de escala mientras que en la segunda sí. Por ello, el realizar cocientes no tiene sentido cuando la variable corresponde a una escala de intervalo.

3.5. Construcción de Proximidades

El análisis de escalamiento multidimensional se basa en la comparación de objetos. Con el término “objetos” se denomina a las cosas o eventos que se encuentran bajo interés y/o estudio (especies, zonas geográficas, estímulos, productos, etc.)

Muchos términos son usados para describir la asociación entre pares de objetos: similaridad, proximidad, disimilaridad, distancia, así como asociación entre ellos mismos.

Las medidas de proximidad son un conjunto de números que indican el grado de semejanza (similaridad) o diferencia (disimilaridad) que guarda cada par de objetos, con relación a cierto número de características cualitativas y cuantitativas.

Tales medidas, reunidas en una matriz, son los elementos básicos para realizar un análisis de escalamiento multidimensional.

3.5.1. Similaridades

Sea A el conjunto de n objetos, indexado por $\{1, 2, \dots, j, \dots, n\}$. Se llama similaridad a la medida de proximidad que indica el grado de semejanza o similitud entre el objeto r y el objeto s , y se denota por $s(r, s)$. Un valor pequeño de $s(r, s)$ indica poca similitud entre r y s .

La mayoría son considerados como simétricos, es decir, tales que $s(r, s) = s(s, r)$, y son acomodados de tal manera que sean positivos y tengan un límite superior igual a 1, aunque algunos cumplen la condición: $-1 \leq s(r, s) \leq 1$.

3.5.2. Disimilaridades

Sea A el conjunto de n objetos, indexados por $\{1, 2, \dots, j, \dots, n\}$. Se llama disimilaridad a la medida de proximidad que indica el grado de diferencia entre el objeto r y el objeto s , denotándolo por $\delta(r, s)$.

Cuanto mayor es la diferencia entre los objetos r y s , mayor es el valor de $\delta(r, s)$. Un valor pequeño de $\delta(r, s)$ indica poca diferencia entre los objetos.

Asociada a cada medida de similaridad limitada por cero y uno, existe una medida de disimilaridad definida por: $\delta(r, s) = 1 - s(r, s)$, la cual es simétrica y no negativa.

Un grado mayor de similaridad entre los objetos r y s incrementa el valor de $s(r, s)$ y disminuye el valor de $\delta(r, s)$.

3.5.3. Distancias

Sean r y s dos puntos (que pueden representar a dos objetos). Una función real denotada por $d(r, s)$ es una función de distancia si cumple con las propiedades:

- I) $d(r, s) = d(s, r)$ (simetría)
- II) $d(r, s) \geq 0$ (no negatividad)
- III) $d(r, r) = 0$

Para algunas funciones de distancia también se cumplen las siguientes propiedades:

IV) $d(r, s) = 0$ si y sólo si $r = s$

V) $d(r, s) \leq d(r, t) + d(t, s)$ (desigualdad del triángulo)

VI) $d(r, s) \leq \max_t \{d(r, t), d(t, s)\}$ (desigualdad ultramétrica)

VII) Existe un espacio euclidiano \mathfrak{R}^m y dos puntos $P_r, P_s \in \mathfrak{R}^m$ de coordenadas $P_r = (x_{r1}, \dots, x_{rm})$ y $P_s = (x_{s1}, \dots, x_{sm})$, tales que: $d(r, s) = [\sum_{k=1}^m (x_{rk} - x_{sk})^2]^{1/2} = d_2(P_r, P_s)$ llamada distancia euclidiana.

Una función de distancia recibe diferentes denominaciones, según las propiedades que verifica:

Denominación	Propiedades
Disimilaridad	I, II, III
Distancia Métrica	I, II, III, IV, V
Distancia Ultramétrica	I, II, III, VI
Distancia Euclidiana	I, II, III, IV, VII

3.5.4. Medidas de Proximidad Directas

Este término se aplica a las medidas de proximidad originadas al establecer directamente una relación binaria de similaridad entre los objetos. Para esto se elige a N sujetos y a cada uno se le presentan los objetos por pares y se le pide que emita su juicio de similaridad (o disimilaridad) sobre cada par.

Los métodos más comunes para obtener estos datos son los siguientes:

Evaluación categórica (puntaje)

La técnica de evaluación categórica es el método más común utilizado para la obtención de disimilaridades. Este método requiere que cada sujeto evalúe la similaridad o disimilaridad total de cada par de objetos, y asigne un puntaje de acuerdo al siguiente esquema:

	Objeto r - Objeto s	
Alta Similaridad	×	Alta Disimilaridad
	0 1 2 3 4 5 6 7 8 9	

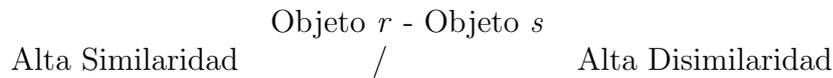
Generalmente la escala contiene de seis a nueve categorías.

El valor que recibe el par de objetos (r, s) es el número de la categoría que fue marcada por el sujeto al evaluar este par.

Calculando la media aritmética de los valores asignados por cada uno de los sujetos al par (r, s) se obtiene una medida de disimilaridad $\delta(r, s)$.

Evaluación gráfica (puntaje gráfico)

Ésta es una ligera variación de la evaluación categórica. El sujeto debe marcar sobre una línea continua (de preferencia de longitud 10 cm.) al evaluar cada par de objetos, de acuerdo al esquema siguiente:



A cada par se le asigna la longitud del segmento que queda a la izquierda de la marca. Se obtiene una medida de disimilaridad $\delta(r, s)$, calculando la media aritmética (o con menor frecuencia la mediana) de los valores asignados por todos los sujetos, al par (r, s) .

Evaluación condicional (objeto ancla)

Se selecciona un objeto-ancla y cada uno de los $n - 1$ objetos restantes son comparados uno por uno con éste, de acuerdo a su similitud o diferencia. Se pueden utilizar los métodos antes descritos para obtener las disimilaridades.

Cada objeto sirve, en turno, de objeto-ancla. Esto da n conjuntos con $n - 1$ medidas de proximidad cada uno.

Este método tiene la ventaja de que se necesitan hacer menos comparaciones a un tiempo; en lugar de evaluar $\binom{n}{2} = \frac{n(n-1)}{2}$ pares de objetos, el método del objeto-ancla sólo necesita $n - 1$. Sin embargo, las medidas así obtenidas son condicionales y menos comparables.

Clasificación de objetos

En este método se les pide a los N sujetos que clasifiquen los objetos en diferentes grupos, en función de su similaridad. El número de grupos puede ser determinado por el investigador o dejar a los sujetos decidirlo. Si N_{rs} es el número de sujetos que han incluido al objeto r y al s en el mismo grupo, entonces se define una medida de disimilaridad $\delta(r, s) = \frac{N - N_{rs}}{N}$.

Clasificación de pares

Al sujeto se le pide que clasifique los pares de objetos en diferentes grupos de tal manera que los pares formados por objetos muy similares deben quedar en el primer grupo y los pares formados por los objetos más disimilares deben quedar en el último. A cada grupo se le asigna un número $\{1, 2, \dots, k\}$, comenzando por el primero y a cada par se le asigna el número del grupo al que pertenece.

Se define una medida de disimilaridad $\delta(r, s)$ como la media aritmética de los valores asignados al par de objetos r y s , por todos los sujetos.

Ordenamiento de pares

A cada sujeto se le pide que ordene los pares de objetos de acuerdo a su similaridad o disimilaridad, de tal manera que queden acomodados en una pila, donde el par de objetos más similares quede en la base y el de los más disimilares en el tope de la pila. A cada par se le asigna un número del conjunto $\{1, 2, \dots, \binom{n-1}{2}\}$ comenzando por la base de la pila.

Se define una medida de disimilaridad $\delta(r, s)$ como la media aritmética de los valores asignados al par de objetos r y s , por todos los sujetos.

Observaciones

Si el número de objetos es muy grande, la tarea de evaluar todos los pares es muy demandante para los sujetos. Una alternativa es emplear un diseño de datos incompletos, en el cual los pares de objetos se dividen en subconjuntos no ajenos y cada individuo evalúa uno. Cada par de objetos es evaluado por igual número de sujetos pero no por todos.

3.5.5. Medidas de Proximidad Derivadas

En la práctica, generalmente las medidas de proximidad no se obtienen de juicios directos de similaridad, sino más bien son índices derivados de otra información sobre los objetos, dando lugar a las medidas de proximidad derivadas.

Describiremos cómo una matriz X de $n \times p$ puede ser usada para construir una matriz simétrica $n \times n$ de asociación entre los n renglones de X . Se pueden intercambiar renglones y columnas de X y así construir la matriz

simétrica $p \times p$ de asociación entre las p columnas de X . Consideraremos la formulación de medidas de asociación para tres tipos de datos ecológicos discutidos aquí: binarios, cualitativos, y cuantitativos. También se necesita considerar qué hacer cuando los datos son una mezcla de este tipo y cómo poder continuar con valores ausentes.

Datos binarios (variables dicotómicas)

Cuando todos los datos son binarios, la información sobre el grado de asociación entre algún par de objetos r y s puede mostrarse como una tabla de contingencia 2×2 :

		Objeto s		
		presente (1)	ausente (0)	total
Objeto r	presente (1)	a	b	$a + b$
	ausente (0)	c	d	$c + d$
	total	$a + c$	$b + d$	$p = a + b + c + d$

donde el número total de variables p se particiona en aquellos en los cuales ambos objetos están presentes (a), en los que sólo uno está presente (b o c), y en los que ambos están ausentes (d). Los siguientes son coeficientes de similaridad basados sobre las cantidades a , b , c y d .

$$s_{rs} = \frac{a}{\max\{(a+b), (a+c)\}} \quad \text{Braun, Blanque}$$

$$s_{rs} = \frac{2a}{2a+b+c} \quad \text{Czekanowski, Sorensen, Dice}$$

$$s_{rs} = \frac{a - (b+c) + d}{a+b+c+d} \quad \text{Hamman}$$

$$s_{rs} = \frac{a}{a+b+c} \quad \text{Jaccard}$$

$$s_{rs} = \frac{a}{b+c} \quad \text{Kulczynski}$$

$$s_{rs} = \frac{1}{2} \left(\frac{a}{a+b} + \frac{a}{a+c} \right) \quad \text{Kulczynski}$$

$s_{rs} = \frac{4(ad - bc)}{\{(a + d)^2 + (b + c)^2\}}$	Michael
$s_{rs} = \frac{2a}{a(b + c) + 2bc}$	Mountford
$s_{rs} = \frac{a(a + b + c + d)}{(a + b)(a + c)}$	Mozley, Margalef
$s_{rs} = \frac{a}{[(a + b)(a + c)]^{1/2}}$	Ochiai
$s_{rs} = \frac{ad - bc}{[(a + b)(a + c)(b + d)(c + d)]^{1/2}}$	Phi
$s_{rs} = \frac{a + d}{a + 2b + 2c + d}$	Rogers, Tanimoto
$s_{rs} = \frac{a}{a + b + c + d}$	Russell, Rao
$s_{rs} = \frac{a + d}{a + b + c + d}$	Simple matching coefficient
$s_{rs} = \frac{a}{\min\{(a + b), (a + c)\}}$	Simpson
$s_{rs} = \frac{a}{a + 2(b + c)}$	Sokal, Sneath, Anderberg
$s_{rs} = \frac{ad - bc}{ad + bc}$	Yule

Datos cualitativos (variables categóricas)

Este tipo de datos cualitativos son datos en escala nominal y ordinal.

Si, para la i -ésima variable nominal, los objetos r y s aparecen en la misma categorización, definimos $s_{rsi} = 1$, y 0 en otro caso. Una medida de similitud es

$$s_{rs} = \frac{\sum_i s_{rsi}}{p}.$$

Si una variable es ordinal con k categorías, entonces $k - 1$ variables indicadoras pueden ser usadas para representar estas categorías. Las variables indicadoras pueden estar sujetas a coeficientes de similitudes para dar un valor a s_{rsi} .

Datos cuantitativos (variables cuantitativas)

Para datos cuantitativos es usual operar con disimilaridades en lugar de similitudes. La siguiente lista da posibles medidas de disimilaridades para datos cuantitativos que son en particular, continuos, posiblemente discretos, pero no binarios.

$\delta_{rs} = \left\{ \sum_i (x_{ri} - x_{si})^2 \right\}^{\frac{1}{2}}$	Distancia Euclidiana
$\delta_{rs} = \left\{ \sum_i w_i (x_{ri} - x_{si})^2 \right\}^{\frac{1}{2}}$	Medida Euclidiana
$\delta_{rs} = \left\{ (x_r - x_s)^t \Sigma^{-1} (x_r - x_s) \right\}^{\frac{1}{2}}$	Distancia Mahalanobis
$\delta_{rs} = \sum_i x_{ri} - x_{si} $	Métrica City block
$\delta_{rs} = \frac{1}{p} \sum_i \frac{ x_{ri} - x_{si} }{r_i}$	Métrica City block (Manhattan)
$\delta_{rs} = \left\{ \sum_i w_i x_{ri} - x_{si} ^\lambda \right\}^{\frac{1}{\lambda}}, \quad \lambda \geq 1$	Métrica Minkowski
$\delta_{rs} = \sum_i \frac{ x_{ri} - x_{si} }{(x_{ri} + x_{si})}$	Métrica Canberra
$\delta_{rs} = \frac{1}{p} \sum_i \frac{(x_{ri} - x_{si})^2}{(x_{ri} + x_{si})^2}$	Divergencia
$\delta_{rs} = \frac{1}{p} \frac{\sum_i x_{ri} - x_{si} }{\sum_i (x_{ri} + x_{si})^2}$	Bray-Curtis
$\delta_{rs} = \frac{\sum_i x_{ri} - x_{si} }{\sum_i \max(x_{ri}, x_{si})}$	Sorgel
$\delta_{rs} = \left\{ \sum_i \left(x_{ri}^{1/2} - x_{si}^{1/2} \right)^2 \right\}^{\frac{1}{2}}$	Distancia Bhattacharyya

$$\begin{aligned} \delta_{rs} &= \frac{1}{p} \sum_i \left(1 - \frac{\min(x_{ri}, x_{si})}{\max(x_{ri}, x_{si})} \right) && \text{Wave-Hedges} \\ \delta_{rs} &= 1 - \frac{\sum_i x_{ri} x_{si}}{\{\sum_i x_{ri}^2 \sum_i x_{si}^2\}^{1/2}} && \text{Separación Angular} \\ \delta_{rs} &= 1 - \frac{\sum_i (x_{ri} - x_r)(x_{si} - x_s)}{\{\sum_i (x_{ri} - x_r)^2 \sum_i (x_{si} - x_s)^2\}} && \text{Correlación} \end{aligned}$$

Similaridad/disimilaridad para datos mixtos (forma general del coeficiente de similaridad)

Cuando los datos son de tipo mixtos las medidas de similaridad y disimilaridad anteriores no pueden ser aplicadas directamente. Gower propuso una forma general del coeficiente de similaridad, la idea básica es definir un coeficiente de similaridad s_{rsi} entre el r -ésimo y el s -ésimo objeto dada la i -ésima variable, y entonces promediando s_{rsi} bajo todas las variables i forman una medida de similaridad s_{rs} :

$$s_{rs} = \frac{\sum_{i=1}^p \omega_{rsi} s_{rsi}}{\sum_{i=1}^p \omega_{rsi}}$$

donde $\omega_{rsi} = \omega(x_{ri}, x_{si})$ es una función de x_{ri} y x_{si} ; $\omega_{rsi} = 1$ si el r -ésimo y el s -ésimo objetos pueden ser comparados sobre la i -ésima variable, y $\omega_{rsi} = 0$ en otro caso, es decir, cuando algún valor de x_{ri} o x_{si} están ausentes o cuando $x_{ri} = x_{si} = 0$.

Gower sugiere los siguientes valores para s_{rsi} y ω_{rsi} para variables binarias que miden presencia/ausencia.

x_{ri}	x_{si}	s_{rsi}	ω_{rsi}
1	1	1	1
1	0	0	1
0	1	0	1
0	0	1	0

Para variables nominales, Gower sugiere $s_{rsi} = 1$ si los objetos r y s particionan la misma categorización para la variable i , $s_{rsi} = 0$ en otro caso.

Para variables cuantitativas,

$$s_{rsi} = 1 - \frac{|x_{ri} - x_{si}|}{R_i},$$

donde R_i es el rango de las observaciones para la variable i .

El coeficiente de Gower puede ser generalizado usando los pesos $\{w_i\}$ para las variables a

$$s_{rsi} = \frac{\sum_i s_{rsi} \omega_{rsi} w_i}{\sum_i \omega_{rsi} w_i}.$$

3.5.6. Transformando de Similaridades a Disimilaridades

Es posible convertir las similaridades en disimilaridades y viceversa, mediante transformaciones sencillas de los datos contenidos en la matriz de proximidades. Las transformaciones más utilizadas son las que convierten las similaridades en disimilaridades, ya que éstas últimas suelen ser las más solicitadas por la mayoría de los programas. Las transformaciones más utilizadas son:

- $\delta_{rs} = 1 - s_{rs}$. Suele utilizarse cuando las similaridades están medidas en una escala de 0 (mínima similaridad) a 1 (máxima similaridad), como las matrices cuyos elementos son en realidad proporciones. Las disimilaridades resultantes estarán medidas en una escala que va de 0 (mínima disimilaridad) a 1 (máxima disimilaridad).
- $\delta_{rs} = c - s_{rs}$. Suele utilizarse cuando las similaridades están medidas en una escala de 0 (mínima similaridad) a c (máxima similaridad), como las matrices cuyos elementos son abundancias o frecuencias. Las disimilaridades resultantes estarán medidas en una escala que va de 0 (mínima disimilaridad) a c (máxima disimilaridad).
- $\delta_{rs} = \{2(1 - s_{rs})\}^{\frac{1}{2}}$. Muy utilizada con índices de similaridad tales como correlaciones, cuyos valores van de -1 (relación perfecta e inversa), a $+1$ (relación perfecta directa). Las disimilaridades correspondientes recibirán los valores de 2 (máxima disimilaridad) a 0 (mínima disimilaridad).

3.6. Propiedades de las Medidas de Asociación

Gower y Legendre (1986) discutieron en detalle la métrica y las propiedades euclidianas de muchos coeficientes de disimilaridad.

Sean las disimilaridades $\{\delta_{rs}\}$ puestas en una matriz D , la matriz de disimilaridades. Sean $\{s_{rs}\}$ las similaridades puestas en una matriz S . Entonces D es métrica si δ_{rs} es una métrica. La matriz D es euclidiana si n puntos pueden estar en un espacio euclidiano tal que la distancia euclidiana entre el r -ésimo y el s -ésimo puntos es δ_{rs} , para todo $1 \leq r, s \leq n$.

Recordemos que una función $\delta : X \times X \rightarrow \mathfrak{R}$ con $\delta_{rs} = \delta(x, y)$ es una métrica si:

- (M1) $\delta(x, y) = 0$ si y sólo si $x = y$
- (M2) $\delta(x, y) = \delta(y, x)$ para todo x y y
- (M3) $\delta(x, y) \leq \delta(x, z) + \delta(z, y)$

Una función $d : X \times X \rightarrow \mathfrak{R}$ es una distancia euclidiana, si es una métrica tal que

$$d(x, y) = \|x - y\| = \sqrt{\sum_i (x_i - y_i)^2}$$

Proposición 3. Si D es no métrica entonces la matriz con elementos $\delta_{rs} + c$ ($r \neq s$) es métrica donde $c \geq \max_{(r,s,t)} |\delta_{rs} - \delta_{rt} - \delta_{st}|$.

Proposición 4. Si D es métrica entonces también lo son las matrices con elementos (i) $\delta_{rs} + c^2$, (ii) $\delta_{rs}^{1/\lambda}$ ($\lambda \geq 1$), (iii) $\delta_{rs}/(\delta_{rs} + c^2)$ para alguna constante real c , y $r \neq s$.

Sea la matriz $\Lambda = \left[-\frac{1}{2}d_{rs}^2\right]$.

Proposición 5. D es euclidiana si y sólo si la matriz $(I - 1s^t)\Lambda(I - s1^t)$ es semidefinida positiva, donde I es la matriz identidad, 1 es un vector de unos, y s es un vector tal que $s^t1 = 1$.

Proposición 6. Si S es una matriz de similaridad semidefinida positiva con elementos $0 \leq s_{rs} \leq 1$ y $s_{rr} = 1$, entonces la matriz de disimilaridades con elementos $d_{rs} = (1 - s_{rs})^{\frac{1}{2}}$ es euclidiana.

Proposición 7. Si D es una matriz de disimilaridades, entonces existe una constante h tal que la matriz con elementos $(\delta_{rs}^2 + h)^{\frac{1}{2}}$ es euclidiana, donde

$h \geq -\lambda_n$, el valor propio más pequeño de $\Lambda_1 = H\Lambda H$, H será la matriz centrada $(I - \frac{1}{n}11^t)$.

Proposición 8. Si D es una matriz de disimilaridades, entonces existe una constante k tal que la matriz con elementos $(\delta_{rs} + k)$ es euclidiana, donde $k \geq \mu_n$, el valor propio más grande de

$$\begin{bmatrix} 0 & 2\Lambda_1 \\ -I & -4\Lambda_2 \end{bmatrix}$$

donde $\Lambda_2 = [-\frac{1}{2}d_{rs}]$.

Para variables binarias, Gower y Legendre definen

$$S_\theta = \frac{a+d}{a+d+\theta(b+c)} \quad T_\theta = \frac{a}{a+\theta(b+c)}$$

Entonces, para una elección apropiada de θ , pueden obtenerse los coeficientes de similaridad. Gower y Legendre muestran:

Proposición 9. Para $\theta \geq 1$, $1 - S_\theta$ es métrica; $\sqrt{1 - S_\theta}$ es métrica para $\theta \geq \frac{1}{3}$; si $\theta < 1$ entonces $1 - S_\theta$ puede ser no métrico; si $\theta < \frac{1}{3}$ entonces $\sqrt{1 - S_\theta}$ puede ser no métrico. Hay resultados similares cuando S_θ se reemplaza por T_θ .

Proposición 10. Si $\sqrt{1 - S_\theta}$ es euclidiana entonces también lo es $\sqrt{1 - S_\phi}$ para $\phi \geq \theta$, con un resultado similar para T_θ .

Proposición 11. Para $\theta \geq 1$, $\sqrt{1 - S_\theta}$ es euclidiana; para $\theta \geq \frac{1}{2}$, $\sqrt{1 - T_\theta}$ es euclidiana. Sin embargo, $1 - S_\theta$ y $1 - T_\theta$ pueden ser no euclidianas.

Gower y Legendre dan una tabla de varios coeficientes de similaridad y disimilaridad y usan estos resultados para establecer que los coeficientes son métricas y que también son distancias euclidianas.

	Métrica	
Coefficiente de similaridad	$\delta_{rs} = (1 - s_{rs})$	$\delta_{rs} = \sqrt{(1 - s_{rs})}$
<i>Simple matching</i>	sí	sí
Jaccard	sí	sí
Czekanowski	no	sí
Ochiai	no	sí
Mozley	?	?
Mountford	?	?
Medida de distancia	$\delta_{rs} = d_{rs}$	$\delta_{rs} = \sqrt{d_{rs}}$
Euclidiana	no	sí
<i>City-block</i>	sí	sí
Bray-Curtis	no	?
Canberra	sí	sí
Mahalanobis	no	sí
Correlación	no	sí

	Euclidiana	
Coefficiente de similaridad	$\delta_{rs} = (1 - s_{rs})$	$\delta_{rs} = \sqrt{(1 - s_{rs})}$
<i>Simple matching</i>	no	sí
Jaccard	no	sí
Czekanowski	no	sí
Ochiai	no	sí
Mozley	?	?
Mountford	?	?
Medida de distancia	$\delta = d_{rs}$	$\delta_{rs} = \sqrt{d_{rs}}$
Euclidiana	no	sí
<i>City-block</i>	no	?
Bray-Curtis	no	?
Canberra	no	?
Mahalanobis	no	sí
Correlación	no	sí

Capítulo 4

Aplicación

En este capítulo se presenta un breve marco de referencia sobre las comunidades y su manejo estadístico, además se maneja un ejemplo en el que se utilizan métodos del análisis multivariado no métrico, es decir, técnicas que reconocen la importancia de la matriz de similaridades entre muestras y cuyas inferencias son obtenidas a partir de sus rangos. Algunas características del análisis multivariado clásico no tienen analogía obvia en este ajuste basado en similaridades.

Los datos aquí estudiados fueron obtenidos del artículo “*Ecological indicator values as an aid in the interpretation of ordination diagrams*”, realizado por Persson (1981), sobre la medida de abundancia de las especies de plantas en una Reserva Natural de Suecia.

4.1. Comunidades y Estadística

El intento de definir tipos de paisajes y tipos de vegetación, data de tiempos muy remotos. La tipificación y clasificación de los fenómenos naturales responde a una necesidad de la mente humana y probablemente es conveniente para el progreso científico.

Promediar de manera significativa las características de una serie de comunidades plantea un problema estadístico. Es más difícil sentar criterios para tomar decisiones en cuanto al número de grupos a formar y a cómo formarlos, con base en una colección de comunidades o de la expresión abreviada de las mismas. Usualmente, los datos que se poseen sobre una comunidad consisten en una lista de las especies presentes (incompleta) del

resultado del censo.

Las comunidades son entidades con muchos atributos, que requieren la aplicación de una estadística de muchas variables. También son entidades con muchas variables los organismos y, sin embargo, en ellos, los problemas de tipificación y clasificación no son tan complejos como los que generan las comunidades. Sin embargo la taxonomía numérica y la sistematización de comunidades comparten muchos errores de enfoque. La principal diferencia estriba en que, en los organismos, amplios grupos de caracteres muestran una fuerte correlación entre sí. Diversos caracteres correlacionados entre sí valen como uno solo, de manera que, a pesar de la gran complicación de los organismos, el procedimiento seguido para su clasificación es relativamente simplificable y encuentra una expresión cómoda en las claves dicotómicas habituales de las obras de clasificación. En las comunidades, sin embargo, la correlación entre las diversas características es mucho más débil.

La tipificación y clasificación de las comunidades ha sido inspirada por la definición de especies en los organismos y su clasificación.

En la práctica, se utilizan dos tipos de unidades: asociaciones y biocenosis. Las primeras resultan sencillamente de la técnica de muestreo y elaboración ulterior de las muestras. Puede decirse que es una unidad operacional, como se deduce de los siguientes ejemplos: se analiza a simple vista la vegetación en cuadros de 10 m de lado y se expresa el resultado en forma de un inventario, con los nombres de las especies identificadas y una estimación de la cobertura, más otras posibles indicaciones sobre agregación o sociabilidad; se obtienen muestras de sedimento con un cilindro de una pulgada de diámetro y se cuentan todos los organismos vivos presentes en las capas superficiales; se realizan pescas con una red de plancton de malla de 60 μm y 100 cm² de abertura de boca, remolcada a dos metros de profundidad sobre un trayecto de 1 km.

En los tres casos presentados como ejemplos, las características de las muestras se definen explicando la operación, y precisando los grupos taxonómicos de organismos que se identifican y cuentan. De esta forma son repetibles y los resultados obtenidos de distintas muestras se pueden comparar y reunir en un colectivo estadístico, que generalmente tiene la forma de una tabla, con una columna para cada una de las muestras. El ecólogo calculará en el conjunto, por ejemplo, un valor medio de cobertura, dando a cada especie un valor de cobertura.

Es habitual referirse a cada unidad de este tipo con el nombre de asociación. Una muestra es representativa de una asociación, y un conjunto de

muestras que operacionalmente son equivalentes pueden servir de material para definir tipos de asociaciones o asociaciones abstractas.

Una asociación corresponderá a la descripción de un segmento de ecosistema, razonablemente uniforme para que las distintas muestras puedan considerarse como repeticiones casi aleatorias, de modo que el conjunto se puede representar adecuadamente por sus valores medios.

En el segundo tipo de unidad se trata de combinar diferentes procedimientos de muestreo (que puede servir de base para describir otros tantos sistemas de asociaciones) para tener una idea de la estructura completa del ecosistema.

Se entiende por biocenosis una comunidad formada por plantas y animales que se condicionan mutuamente, se mantienen en un estado estacionario dinámico, en virtud de la reproducción propia, y sólo dependen del ambiente inanimado exterior a la biocenosis (o exterior al biótomo, que es el ambiente físico coextensivo con la biocenosis en cuestión); pero no, o de manera no esencial, de organismos vivos exteriores a la biocenosis. Así, para limitar la biocenosis, se escogen como límites aquellos a través de los cuales es mínimo el intercambio de organismos o de sus producciones.

Se comprende que la unidad práctica de trabajo, para el ecólogo, será generalmente la asociación y que la biocenosis tiene la categoría de una construcción teórica. Raramente la biocenosis será sencilla, o el muestreo abarcará tanto sin perder calidad, como para que la biocenosis quede reflejada en una sola asociación. Si se permite una analogía de valor limitado, que quizás aclare estos conceptos, se puede comparar la biocenosis con un organismo y la asociación con un tejido. Situar varias asociaciones en una biocenosis significa reconocer las relaciones funcionales que ligan a aquéllas.

4.2. Área Estudiada

El área estudiada es la Reserva Natural de Steneryd en la provincia de Blekinge, al sur de Suecia (56°08'N, 15°50'E), con un área alrededor de 6 hectáreas. El área fue en un tiempo una pradera para segar y pastar con árboles que eran seleccionados para el forraje de invierno. La siega cesa probablemente antes de 1900, aunque la selección continuó hasta 1940. En 1965 ambas prácticas fueron reanudadas en un intento para recrear un ecosistema antiguamente extenso en Suecia. Al mismo tiempo el registro de la vegetación comenzó y a continuación se presentan los datos usados aquí.

4.3. Metodología

En 1965 diecisiete terrenos de 5×4 m. fueron elegidos, subjetivamente, para cubrir lo que parece ser la extensión de los tipos de vegetación. Cada terreno fue sistemáticamente muestreado con nueve cuadros de 0.5×0.5 m. ordenados en tres filas de tres con espacios entre ellas.

La tabla anterior muestra la abundancia de 50 especies de plantas en los 153 cuadros. Cada uno tiene asignado un rango de 0 a 5, donde el 0 representa la ausencia de la especie y el 5 representa la mayor abundancia de la especie en dicho cuadro.

Los análisis numéricos fueron hechos en PRIMER v. 5 para windows. A partir de una matriz de similaridades (obtenida con el coeficiente de Bray-Curtis, aplicando la transformación $\log(a+1)$ y estandarizando) se aplicaron las siguientes técnicas del análisis multivariado:

- a) MDS no métrico (*non-metric multidimensional scaling*) para exhibir el patrón de las comunidades por medio de la ordenación y el agrupamiento en una configuración de dos dimensiones.
- b) Análisis de agrupamiento (*cluster analysis*) para clasificar las especies en grupos, por medio de aglomeraciones jerárquicas con enlazamiento de promedios grupales.
- c) SIMPER (*similarity percentages*) que identifica las especies cuya abundancia da una mayor aportación a los patrones multivariados observados. Cuando un análisis de agrupamiento divide un conjunto de muestras en digamos dos grupos, puede ser importante qué especies son las que contribuyen principalmente a esta división. El análisis SIMPER compara las abundancias promedio y examina la contribución de cada especie al promedio de las disimilaridades de Bray-Curtis entre dos grupos definidos a priori de las muestras.
- d) ANOSIM (*analysis of similarities*) es un procedimiento no paramétrico de permutaciones que puede ser usado para probar la significancia en las diferencias multivariadas entre grupos que han sido definidos a priori. Este método utiliza el rango de similaridades entre muestras, y determina si las similaridades entre las muestras de los diferentes grupos son significativamente más pequeñas que las similaridades entre las muestras de dentro de los grupos. La hipótesis nula del ANOSIM es:

H_0 : No hay diferencias entre los grupos muestrales.

La prueba de ANOSIM realiza lo siguiente:

1. Asigna el rango a todos los pares de combinaciones de las especies con base en su similaridad, por lo tanto el rango 1 significa el más similar;
2. divide los pares de combinaciones en dos tipos: *entre* los grupos y *dentro* de los grupos;
3. calcula la media de los rangos de cada tipo; el rango más pequeño equivale a un rango más similar;
4. se calcula la estadística de prueba R :

$$R = \frac{\left\{ \begin{array}{l} \text{media de los rangos} \\ \text{entre los grupos} \end{array} \right\} - \left\{ \begin{array}{l} \text{media de los rangos} \\ \text{dentro de los grupos} \end{array} \right\}}{\left\{ \begin{array}{l} \text{factor de corrección para el} \\ \text{número de combinaciones} \end{array} \right\}}$$

si no hay efecto de los grupos se espera que $R = 0$, si *dentro* de los grupos son más similares que *entre* los grupos se espera que $R > 0$.

4.4. Resultados

A partir de la tabla de las medidas de abundancia de las 50 especies en los 153 cuadros se realizó una transformación logarítmica $\log(a + 1)$ y se obtuvo la matriz de similaridades de los cuadros y la matriz de similaridades de las especies utilizando el coeficiente de similaridad de Bray-Curtis y estandarizando:

$$s_{rs} = 100 - \frac{\sum_{i=1}^m 100 |a_{ri} - a_{si}|}{\sum_{j=1}^m (a_{ri} + a_{si})}$$

Para ambas matrices se realizaron los análisis de MDS no métrico, análisis de conglomerados, ANOSIM y SIMPER, y además se obtuvieron las gráficas de la abundancia de cada especie en la configuración del MDS no métrico para el caso de la matriz de similaridades entre los terrenos y la abundancia de cada terreno en la configuración del MDS no métrico de la matriz de similaridades entre las especies.

4.4.1. Análisis de los Cuadros

A partir del gráfico de análisis de agrupamiento realizado sobre la matriz de similitudes entre cada cuadro se pueden establecer dos grupos de muestras con el 30 % de similitud:

- **Grupo 1:** terrenos 1, 2, 3, 4, 5 y 6.
- **Grupo 2:** terrenos 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 y 17.

O tres grupos de muestras con el 35 % de similitud:

- **Grupo 1:** terrenos 1, 2, 3, 4, 5 y 6.
- **Grupo 2:** terrenos 7, 8, 9, 10, 12 y 13.
- **Grupo 3:** terrenos 14, 15, 16 y 17.

Aquí la muestra 11 aparece dentro del Grupo 1 y el Grupo 2.

Para el caso del MDS no métrico se pueden visualizar los mismos grupos, los cuales obtuvieron una conformación significativa (valor $p < 0.05$) a partir del análisis de similitudes, ya sea en dos o tres grupos.

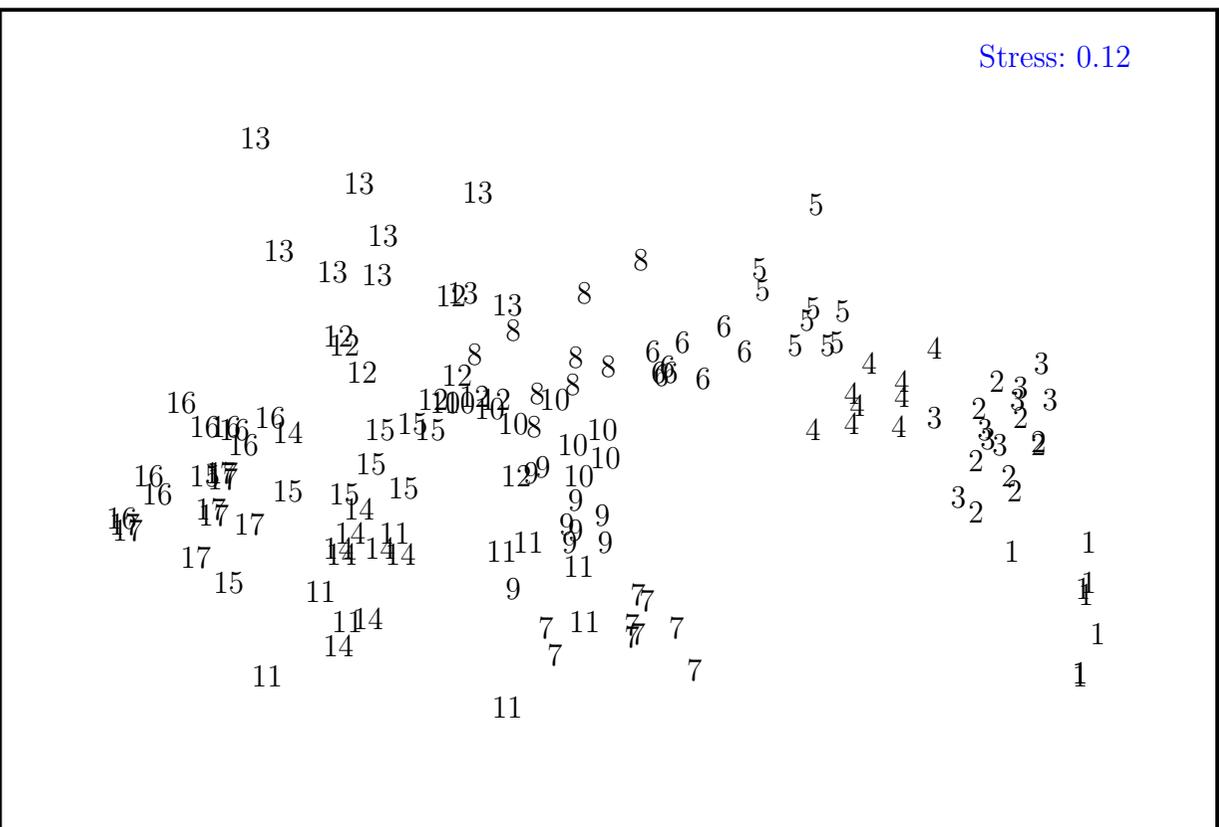
Para este caso el MDS no métrico permitió establecer grupos de manera más práctica y con la ventaja adicional de que fue posible evaluar mediante una prueba estadística (ANOSIM) la conformación de los grupos, los cuales, de forma complementaria pueden sugerirse a partir de análisis de agrupamientos.

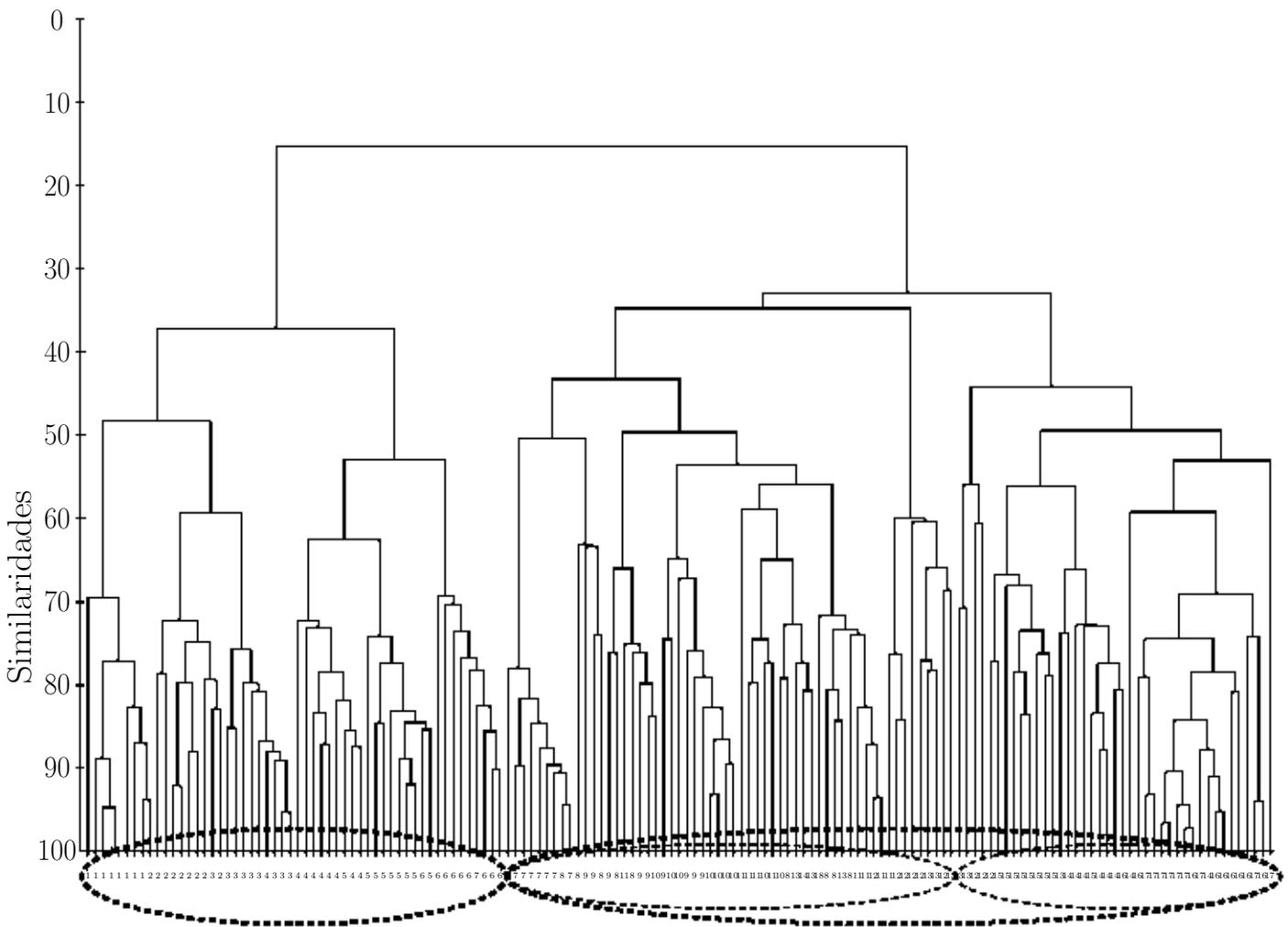
Las tablas muestran el porcentaje de similaridad que cada especie aporta utilizando el análisis SIMPER, a partir de dos grupos a priori.

La representación de la abundancia de cada especie de planta en la configuración del MDS permite ver cómo se distribuye cada especie dentro de los cuadros obtenidas.

Note que en la configuración obtenida por el MDS no métrico los cuadros no muestran un ordenamiento lineal, sino en forma de herradura o arco, a éste efecto se le llama “efecto herradura”, ésto se debe a que los cuadros adyacentes son más similares, y los cuadros que no son adyacentes no son tan disimilares como para ordenarlos en una línea recta, sino que también tienen un similaridad grande; además se muestra un ordenamiento de manera decreciente sobre el eje de las abscisas.

Escalamiento multidimensional no métrico





Análisis de Similaridades (ANOSIM)**Evaluación de 2 grupos (30 % de similitud):**

- **Grupo 1:** 1, 2, 3, 4, 5 y 6.
- **Grupo 2:** 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 y 17.

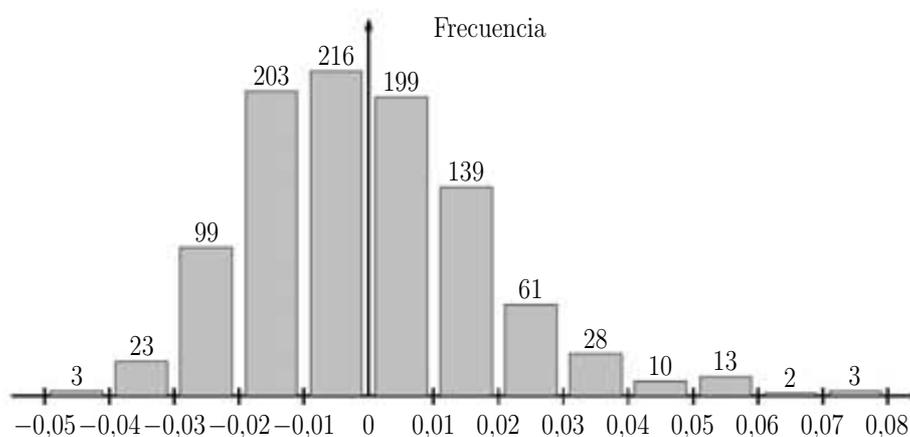
Prueba Global:

Estadística de prueba (correlación R global): 0.774

Nivel mínimo de significancia: 0.001

Número de permutaciones: 999 (de $\binom{17}{6}$ permutaciones posibles).

Número de permutaciones mayores o iguales a R: 0



Evaluación de 3 grupos (35 % de similitud):

- **Grupo 1:** 1, 2, 3, 4, 5 y 6.
- **Grupo 2:** 7, 8, 9, 10, 12 y 13.
- **Grupo 3:** 14, 15, 16 y 17.

Prueba Global:

Estadística de prueba (correlación R global): 0.776

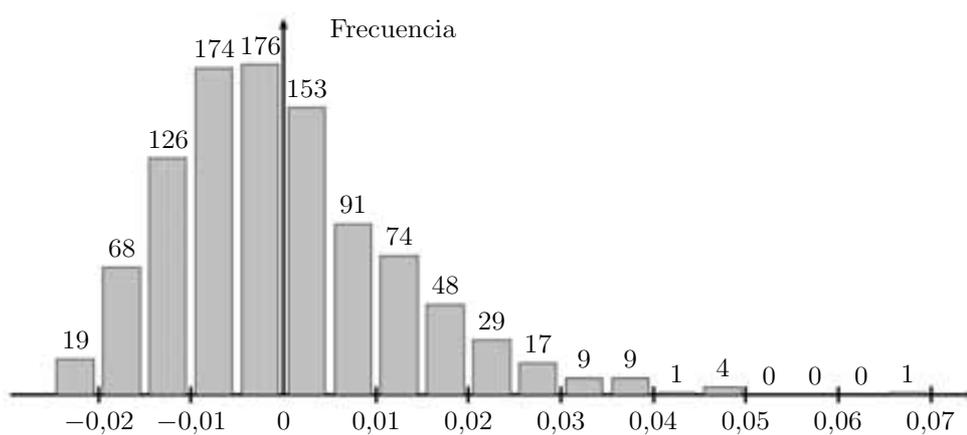
Nivel mínimo de significancia: 0.001

Número de permutaciones: 999 (de $\binom{16}{6}\binom{10}{6}$ permutaciones posibles).

Número de permutaciones mayores o iguales a R: 0

Pruebas por pares:

Grupos	Estadística de prueba R	Nivel mínimo de significancia	Permutaciones posibles	Permutaciones actuales	No. de permutaciones \geq a R
1 y 2	0.76	0.001	$\binom{12}{6}$	999	0
1 y 3	0.963	0.001	$\binom{10}{6}$	999	0
2 y 3	0.613	0.001	$\binom{10}{6}$	999	0



Porcentaje de Similaridades - Contribución de las Especies SIMPER

Datos estandarizados. Transformación: $\log(a+1)$.

- **Grupo 1:** 1, 2, 3, 4, 5 y 6.
- **Grupo 2:** 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 y 17.

Separando los grupos 1 y 2

Similaridad promedio:

$$s = \sum_{j=1}^{153} \sum_{k=1}^{153} \frac{s_{jk}}{(153)(153)}$$

donde

$$s_{jk} = \sum_{i=1}^{50} s_{jk}(i) \quad \text{y} \quad s_{jk}(i) = \frac{200 \text{ mín}(y_{ij}, y_{ik})}{\sum_{i=1}^{50} (y_{ij} - y_{ik})}$$

En este caso $s_{jk} = 100 - \delta_{jk}$

y_{ij} es la abundancia (transformada) de la i -ésima especie en el j -ésimo cuadro para el grupo 1.

En las siguientes tablas tenemos las columnas:

$y_{1i} = \sum_{j=1}^{54} \frac{y_{ij}}{(6)(9)}$ es la abundancia promedio de la i -ésima especie en el grupo 1.

$y_{2i} = \sum_{j=55}^{153} \frac{y_{ij}}{(11)(9)}$ es la abundancia promedio de la i -ésima especie en el grupo 2.

$s_i = \sum_{j=1}^{153} \sum_{k=1}^{153} \frac{s_{jk}(i)}{(153)(153)}$ es la contribución promedio de la i -ésima especie en la similaridad total.

$SD(i) = \left[\sum_{j=1}^{153} \sum_{k=1}^{153} (s_{jk}(i) - s_i)^2 \right]^{1/2} / 153$ es la desviación estándar de la i -ésima especie.

$s_i \% = \frac{s_i}{\sum_{i=1}^{50} s_i}$ es el porcentaje de contribución de la i -ésima especie.

$\sum s_i \% = \sum_{h=1}^i \frac{s_h}{\sum_{i=1}^{50} s_i}$ es el porcentaje de la contribución acumulada hasta la i -ésima especie.

Grupo 1: $\bar{s} = 49,08$

Especies	\bar{y}_{1i}	\bar{s}_i	$\bar{s}_i/SD(i)$	$\bar{s}_i\%$	$\sum \bar{s}_i\%$
<i>Festuca ovina</i>	3.24	7.77	1.84	15.82	15.82
<i>Agrostis tenuis</i>	1.50	6.29	2.76	12.81	28.64
<i>Luzula campestris</i>	0.85	3.73	1.32	7.61	36.24
<i>Plantago lanceolata</i>	1.00	3.44	1.24	7.01	43.25
<i>Hieracium pilosella</i>	0.93	3.38	0.96	6.89	50.14
<i>Achillea millefolium</i>	0.78	2.40	0.82	4.90	55.04
<i>Ranunculus bulbosus</i>	0.69	2.14	0.87	4.35	59.39
<i>Saxifraga granulata</i>	0.70	2.05	0.83	4.18	63.57
<i>Rumex acetosa</i>	0.65	1.95	0.78	3.97	67.54
<i>Galium verum</i>	0.56	1.48	0.53	3.02	70.56
<i>Trifolium repens</i>	0.76	1.21	0.53	2.47	73.03
<i>Filipendula vulgaris</i>	0.48	1.07	0.53	2.19	75.22
<i>Anthoxanthum odoratum</i>	0.48	1.07	0.54	2.19	77.41
<i>Cynosurus cristatus</i>	0.67	1.00	0.50	2.03	79.44
<i>Taraxacum sp.</i>	0.48	0.94	0.48	1.91	81.35
<i>Vicia angustifolia</i>	0.35	0.79	0.36	1.60	82.95
<i>Poa pratensis</i>	0.41	0.78	0.43	1.59	84.53
<i>Rumex tenuifolius</i>	0.31	0.76	0.31	1.54	86.07
<i>Cerastium fontanum</i>	0.41	0.75	0.44	1.54	87.61
<i>Campanula rotundifolia</i>	0.39	0.74	0.41	1.50	89.11
<i>Stellaria graminea</i>	0.39	0.73	0.38	1.49	90.61
<i>Anemone nemorosa</i>	0.39	0.58	0.39	1.18	91.78
<i>Veronica chamaedrys</i>	0.37	0.56	0.39	1.15	92.93
<i>Viscaria vulgaris</i>	0.43	0.52	0.27	1.06	93.99
<i>Teesdalia nudicaulis</i>	0.22	0.45	0.22	0.92	94.90
<i>Trifolium pratense</i>	0.33	0.41	0.30	0.84	95.74
<i>Alchemilla sp.</i>	0.39	0.38	0.30	0.77	96.51
<i>Aira praecox</i>	0.19	0.33	0.18	0.67	97.18
<i>Festuca rubra</i>	0.39	0.33	0.28	0.66	97.85
<i>Trifolium dubium</i>	0.19	0.18	0.18	0.37	98.22
<i>Hypericum maculatum</i>	0.39	0.17	0.20	0.34	98.56
<i>Lathyrus pratensis</i>	0.17	0.11	0.16	0.22	98.78
<i>Allium sp.</i>	0.17	0.09	0.16	0.19	98.97
<i>Dactylis glomerata</i>	0.15	0.07	0.14	0.15	99.13
<i>Fraxinus excelsior (juv.)</i>	0.15	0.07	0.14	0.15	99.28
<i>Campanula persicifolia</i>	0.15	0.07	0.14	0.15	99.43
<i>Hepatica nobilis</i>	0.15	0.07	0.14	0.15	99.58
<i>Lathyrus montanus</i>	0.13	0.06	0.12	0.12	99.70
<i>Geum urbanum</i>	0.13	0.05	0.12	0.11	99.81
<i>Anthriscus silvestris</i>	0.13	0.04	0.10	0.08	99.89
<i>Stellaria holostea</i>	0.11	0.04	0.10	0.08	99.97
<i>Viola riviniana</i>	0.07	0.02	0.06	0.03	100.00

Grupo 2: $\bar{s} = 41,88$

Especies	\bar{y}_{2i}	\bar{s}_i	$\bar{s}_i/SD(i)$	$\bar{s}_i\%$	$\sum \bar{s}_i\%$
<i>Anemone nemorosa</i>	1.71	9.18	2.50	21.92	21.92
<i>Stellaria holostea</i>	1.62	5.89	1.22	14.07	35.99
<i>Fraxinus excelsior</i> (juv.)	0.61	3.20	0.73	7.63	43.62
<i>Ranunculus ficaria</i>	1.20	2.80	0.49	6.68	50.31
<i>Poa pratensis</i>	0.75	2.71	0.70	6.47	56.77
<i>Veronica chamaedrys</i>	0.64	2.57	0.70	6.14	62.92
<i>Dactylis glomerata</i>	0.68	2.53	0.66	6.04	68.96
<i>Rumex acetosa</i>	0.49	1.61	0.54	3.84	72.81
<i>Geum urbanum</i>	0.41	1.39	0.41	3.32	76.12
<i>Viola riviniana</i>	0.41	1.19	0.44	2.84	78.96
<i>Mercurialis perennis</i>	1.04	1.17	0.26	2.80	81.76
<i>Campanula persicifolia</i>	0.38	1.09	0.40	2.60	84.37
<i>Lathyrus montanus</i>	0.40	0.96	0.38	2.30	86.67
<i>Hepatica nobilis</i>	0.37	0.96	0.36	2.29	88.96
<i>Allium</i> sp.	0.33	0.95	0.34	2.27	91.23
<i>Deschampsia flexuosa</i>	0.62	0.78	0.29	1.85	93.08
<i>Anthriscus silvestris</i>	0.32	0.56	0.26	1.35	94.43
<i>Agrostis tenuis</i>	0.49	0.44	0.24	1.05	95.47
<i>Saxifraga granulata</i>	0.23	0.36	0.23	0.87	96.34
<i>Festuca rubra</i>	0.30	0.28	0.21	0.68	97.02
<i>Festuca ovina</i>	0.31	0.24	0.16	0.58	97.60
<i>Primula veris</i>	0.23	0.22	0.18	0.53	98.12
<i>Veronica hederifolia</i>	0.13	0.16	0.13	0.39	98.52
<i>Luzula campestris</i>	0.14	0.13	0.13	0.30	98.81
<i>Moehringia trinervia</i>	0.25	0.11	0.10	0.26	99.08
<i>Anthoxanthum odoratum</i>	0.13	0.11	0.13	0.26	99.33
<i>Convallaria majalis</i>	0.10	0.06	0.09	0.15	99.48
<i>Poa nemoralis</i>	0.10	0.06	0.09	0.14	99.63
<i>Hypericum maculatum</i>	0.10	0.05	0.09	0.13	99.75
<i>Lathyrus pratensis</i>	0.09	0.04	0.09	0.10	99.85
<i>Ranunculus bulbosus</i>	0.07	0.02	0.06	0.05	99.90
<i>Trifolium repens</i>	0.15	0.02	0.05	0.05	99.95
<i>Taraxacum</i> sp.	0.04	0.01	0.04	0.02	99.97
<i>Cynosurus cristatus</i>	0.04	0.01	0.04	0.02	99.99
<i>Filipendula vulgaris</i>	0.03	0.00	0.02	0.01	99.99
<i>Campanula rotundifolia</i>	0.02	0.00	0.01	0.00	100.00
<i>Achillea millefolium</i>	0.02	0.00	0.01	0.00	100.00

Uniando los grupos 1 y 2

Disimilaridad promedio:

$$\delta = \sum_{j=1}^{153} \sum_{k=1}^{153} \frac{\delta_{jk}}{(153)(153)} = 84,66$$

donde

$$\delta_{jk} = \sum_{i=1}^{50} \delta_{jk}(i) \quad \text{y} \quad \delta_{jk}(i) = \frac{100 |y_{ij} - y_{ik}|}{\sum_{i=1}^{50} (y_{ij} + y_{ik})}$$

y_{ij} es la abundancia (transformada) de la i -ésima especie en el j -ésimo cuadro.

En la siguiente tabla tenemos las columnas:

$y_{1i} = \sum_{j=1}^{54} \frac{y_{ij}}{(6)(9)}$ es la abundancia promedio de la i -ésima especie en el grupo 1.

$y_{2i} = \sum_{j=55}^{153} \frac{y_{ij}}{(11)(9)}$ es la abundancia promedio de la i -ésima especie en el grupo 2.

$\delta_i = \sum_{j=1}^{153} \sum_{k=1}^{153} \frac{\delta_{jk}(i)}{(153)(153)}$ es la contribución promedio de la i -ésima especie en la disimilaridad total.

$SD(i) = \left[\sum_{j=1}^{153} \sum_{k=1}^{153} (\delta_{jk}(i) - \delta_i)^2 \right]^{1/2} / 153$ es la desviación estandar de la i -ésima especie.

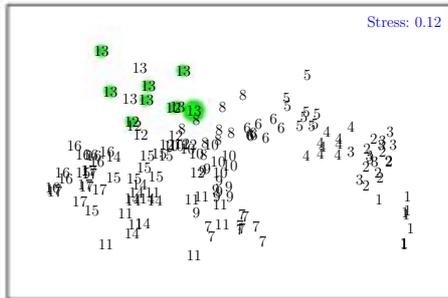
$\delta_i \%$ = $\frac{\delta_i}{\sum_{i=1}^{50} \delta_i}$ es el porcentaje de contribución de la i -ésima especie.

$\sum \delta_i \%$ = $\sum_{h=1}^i \frac{\delta_h}{\sum_{i=1}^{50} \delta_i}$ es el porcentaje de la contribución acumulada hasta la i -ésima especie.

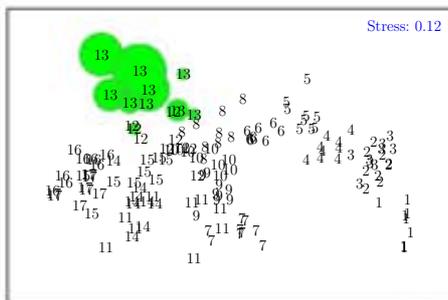
Especies	y_{1i}	y_{2i}	$\bar{\delta}_i$	$\bar{\delta}_i/SD(i)$	$\bar{\delta}_i \%$	$\sum \bar{\delta}_i \%$
<i>Festuca ovina</i>	3.24	0.31	4.75	1.77	5.61	5.61
<i>Anemone nemorosa</i>	0.39	1.71	3.70	1.66	4.38	9.98
<i>Stellaria holostea</i>	0.11	1.62	3.60	1.61	4.25	14.23
<i>Agrostis tenuis</i>	1.50	0.49	3.39	1.85	4.01	18.24
<i>Hieracium pilosella</i>	0.93	0.00	2.95	1.35	3.48	21.72
<i>Ranunculus ficaria</i>	0.00	1.20	2.76	0.86	3.27	24.98
<i>Plantago lanceolata</i>	1.00	0.00	2.76	1.75	3.26	28.24
<i>Luzula campestris</i>	0.85	0.14	2.63	1.62	3.11	31.35
<i>Fraxinus excelsior</i> (juv.)	0.15	0.61	2.40	1.15	2.84	34.19
<i>Achillea millefolium</i>	0.78	0.02	2.31	1.26	2.73	36.92
<i>Dactylis glomerata</i>	0.15	0.68	2.22	1.10	2.63	39.54
<i>Poa pratensis</i>	0.41	0.75	2.19	1.15	2.59	42.13

Especies	y_{1i}	y_{2i}	δ_i	$\delta_i/SD(i)$	$\delta_i\%$	$\sum \delta_i\%$
<i>Veronica chamaedrys</i>	0.37	0.64	2.09	1.13	2.47	44.60
<i>Ranunculus bulbosus</i>	0.69	0.07	2.00	1.32	2.37	46.97
<i>Saxifraga granulata</i>	0.70	0.23	1.99	1.21	2.35	49.32
<i>Rumex acetosa</i>	0.65	0.49	1.94	1.12	2.29	51.61
<i>Mercurialis perennis</i>	0.00	1.04	1.89	0.58	2.23	53.84
<i>Galium verum</i>	0.56	0.00	1.87	0.92	2.21	56.05
<i>Trifolium repens</i>	0.76	0.15	1.81	0.93	2.13	58.18
<i>Geum urbanum</i>	0.13	0.41	1.72	0.83	2.03	60.21
<i>Viola riviniana</i>	0.07	0.41	1.55	0.83	1.83	62.04
<i>Campanula persicifolia</i>	0.15	0.38	1.55	0.84	1.83	63.87
<i>Cynosurus cristatus</i>	0.67	0.04	1.53	0.89	1.81	65.68
<i>Anthoxanthum odoratum</i>	0.48	0.13	1.52	0.97	1.80	67.47
<i>Allium</i> sp.	0.17	0.33	1.48	0.78	1.75	69.22
<i>Hepatica nobilis</i>	0.15	0.37	1.48	0.80	1.74	70.97
<i>Filipendula vulgaris</i>	0.48	0.03	1.47	0.93	1.73	72.70
<i>Lathyrus montanus</i>	0.13	0.40	1.43	0.81	1.69	74.39
<i>Rumex tenuifolius</i>	0.31	0.00	1.41	0.64	1.67	76.06
<i>Taraxacum</i> sp.	0.48	0.04	1.41	0.88	1.66	77.72
<i>Deschampsia flexuosa</i>	0.00	0.62	1.41	0.61	1.66	79.38
<i>Vicia angustifolia</i>	0.35	0.00	1.33	0.72	1.57	80.95
<i>Festuca rubra</i>	0.39	0.30	1.31	0.76	1.55	82.50
<i>Stellaria graminea</i>	0.39	0.00	1.25	0.74	1.48	83.97
<i>Campanula rotundifolia</i>	0.39	0.02	1.25	0.78	1.47	85.45
<i>Anthriscus silvestris</i>	0.13	0.32	1.24	0.65	1.47	86.92
<i>Viscaria vulgaris</i>	0.43	0.00	1.21	0.60	1.43	88.35
<i>Cerastium fontanum</i>	0.41	0.00	1.19	0.81	1.40	89.75
<i>Teesdalia nudicaulis</i>	0.22	0.00	1.13	0.52	1.33	91.08
<i>Aira praecoꝝ</i>	0.19	0.00	0.98	0.46	1.16	92.24
<i>Trifolium pratense</i>	0.33	0.00	0.93	0.63	1.10	93.34
<i>Alchemilla</i> sp.	0.39	0.01	0.91	0.64	1.08	94.42
<i>Hypericum maculatum</i>	0.39	0.10	0.88	0.58	1.04	95.46
<i>Primula veris</i>	0.00	0.23	0.68	0.46	0.80	96.25
<i>Lathyrus pratensis</i>	0.17	0.09	0.64	0.54	0.76	97.02
<i>Trifolium dubium</i>	0.19	0.00	0.62	0.47	0.74	97.75
<i>Veronica hederifolia</i>	0.00	0.13	0.59	0.38	0.70	98.45
<i>Moehringia trinervia</i>	0.00	0.25	0.58	0.33	0.69	99.14
<i>Convallaria majalis</i>	0.00	0.10	0.36	0.31	0.43	99.57
<i>Poa nemoralis</i>	0.00	0.10	0.36	0.31	0.43	100.00

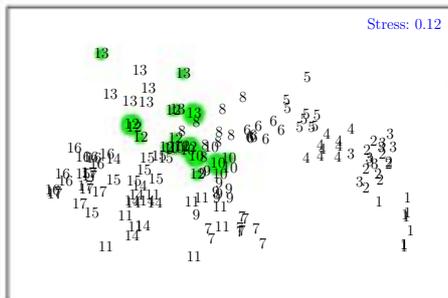
1. *Poa nemoralis*



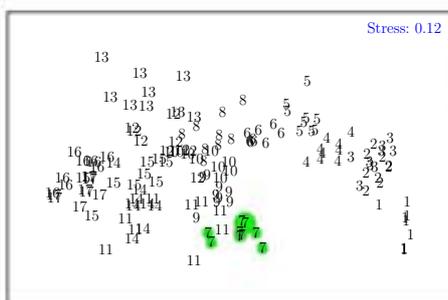
2. *Moehringia trinervia*

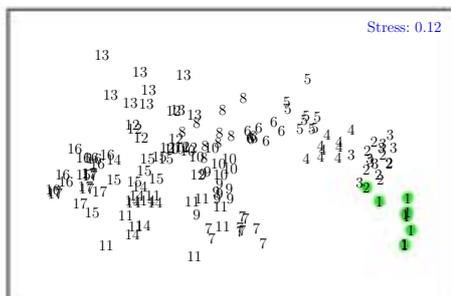
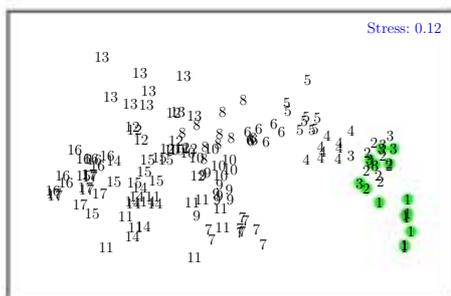
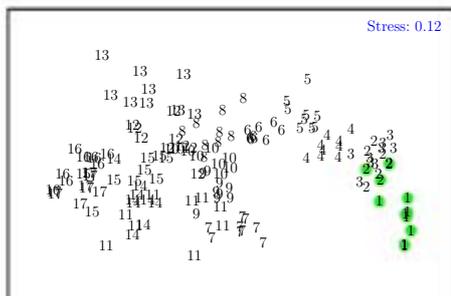
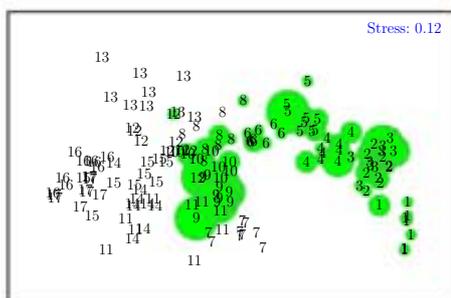


3. *Primula veris*

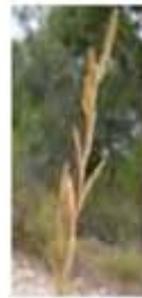
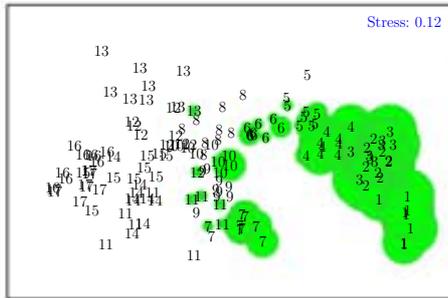


4. *Convallaria majalis*

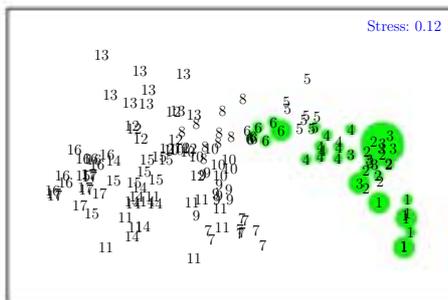


5. *Aira praecox*6. *Rumex tenuifolius*7. *Teesdalia nudicaulis*8. *Agrostis tenuis*

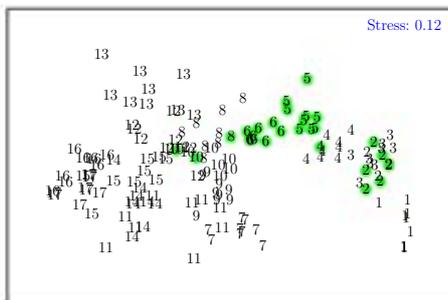
9. *Festuca ovina*



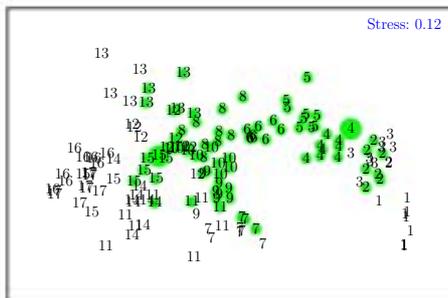
10. *Hieracium pilosella*

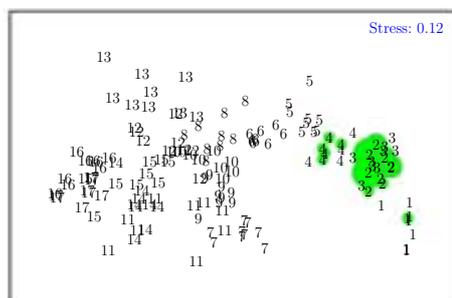
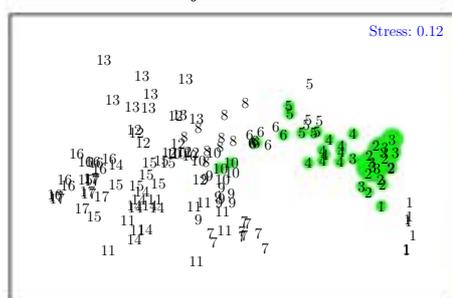
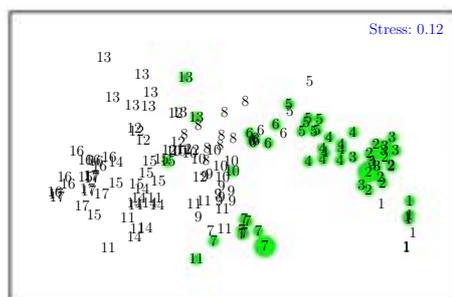
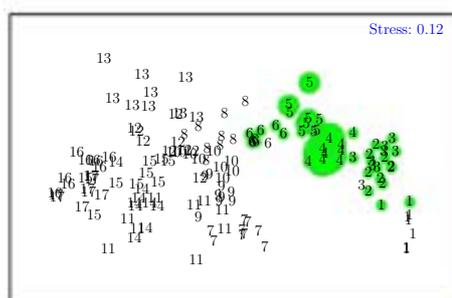


11. *Filipendula vulgaris*

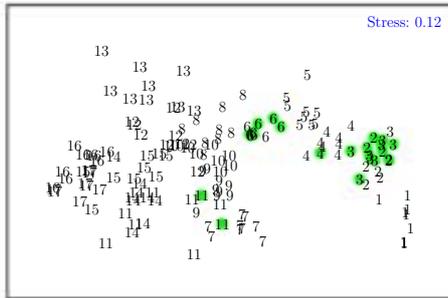


12. *Rumex acetosa*

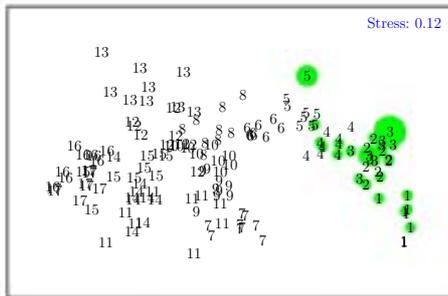


13. *Viscaria vulgaris*14. *Achillea millefolium*15. *Luzula campestris*16. *Plantago lanceolata*

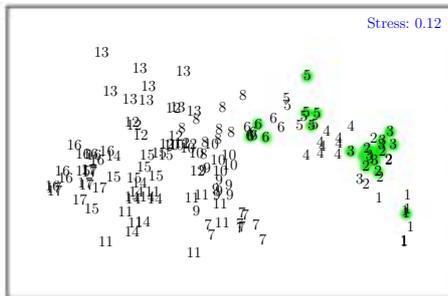
17. *Campanula rotundifolia*



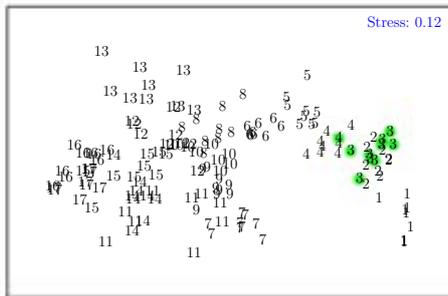
18. *Galium verum*

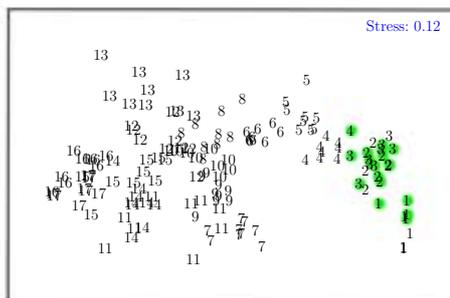
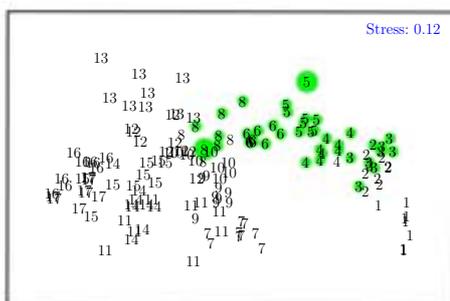
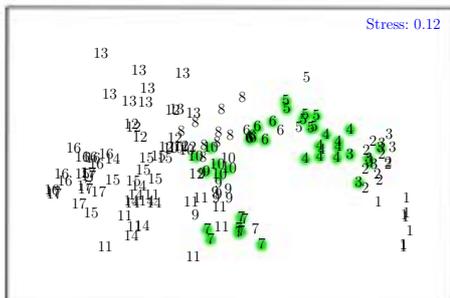
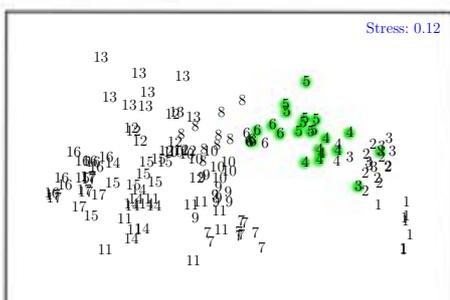


19. *Stellaria graminea*

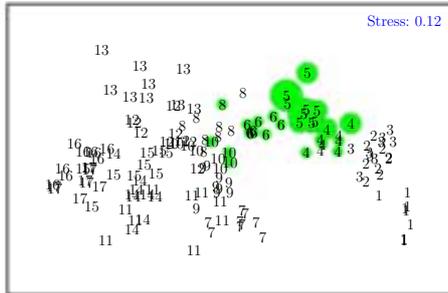


20. *Trifolium dubium*

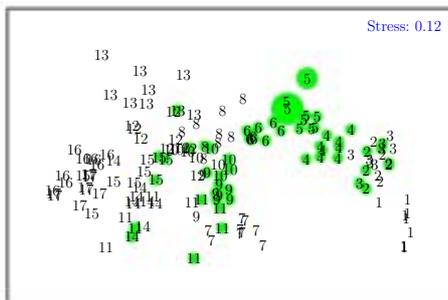


21. *Vicia angustifolia*22. *Ranunculus bulbosus*23. *Anthoxanthum odoratum*24. *Cerastium fontanum*

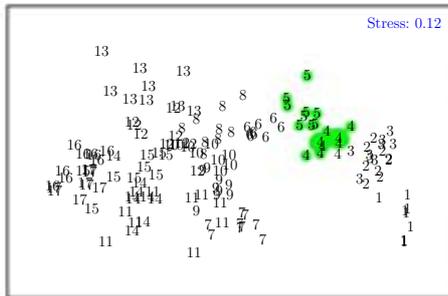
25. *Cynosurus cristatus*



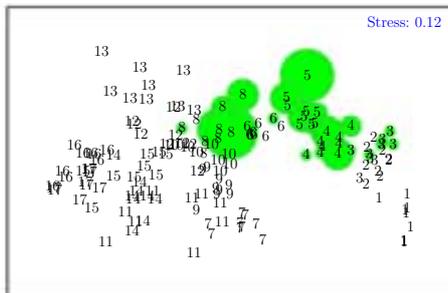
26. *Saxifraga granulata*

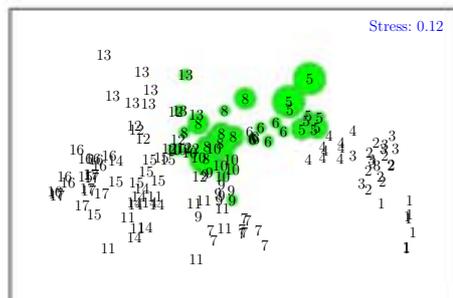
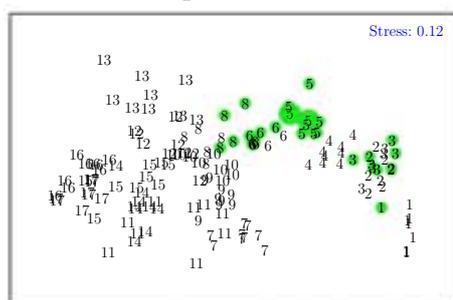
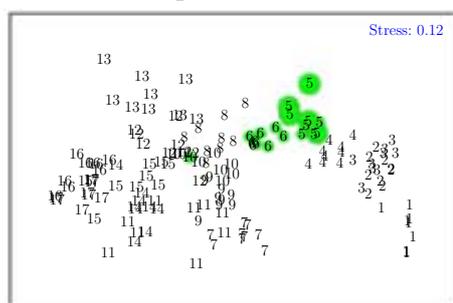
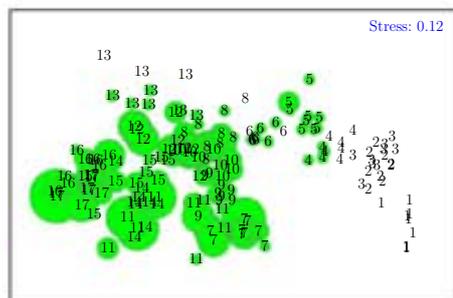


27. *Trifolium pratense*

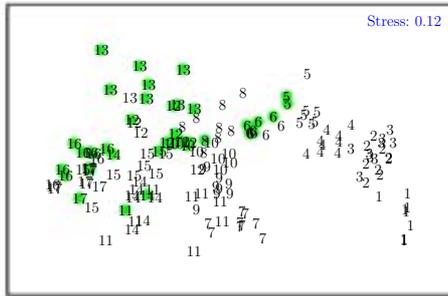


28. *Trifolium repens*

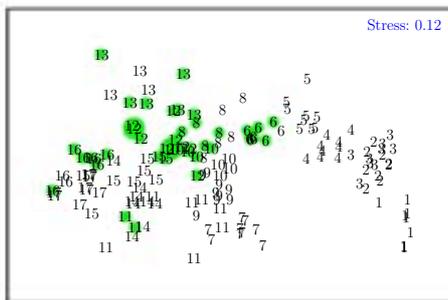


29. *Festuca rubra*30. *Taraxacum* sp.31. *Alchemilla* sp.32. *Anemone nemorosa*

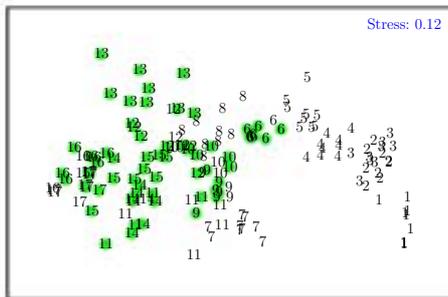
33. *Allium* sp.



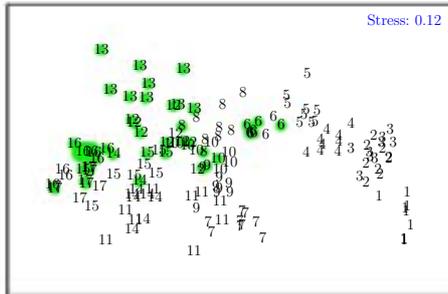
34. *Hepatica nobilis*



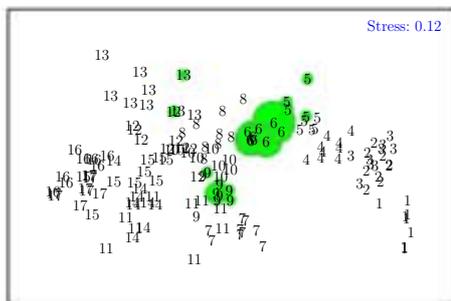
35. *Fraxinus excelsior* (juv.)



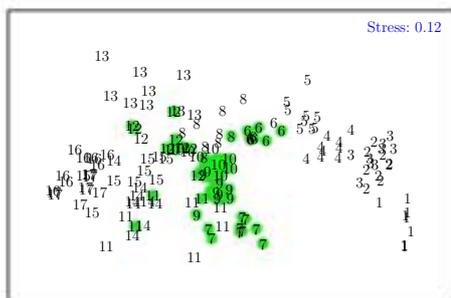
36. *Geum urbanum*



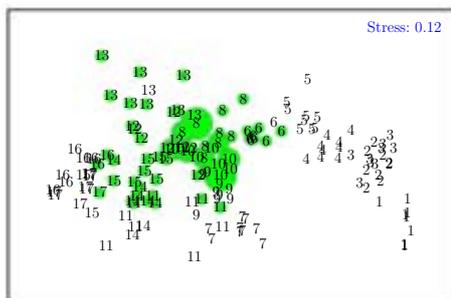
37. *Hypericum maculatum*



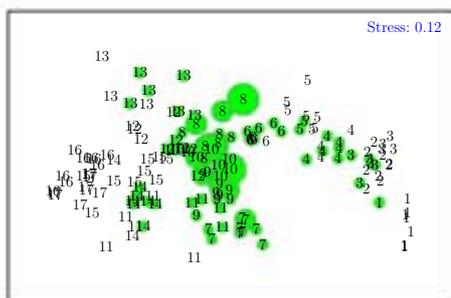
38. *Lathyrus montanus*



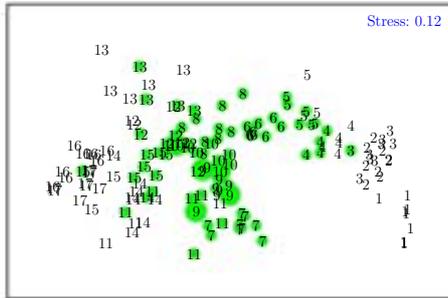
39. *Dactylis glomerata*



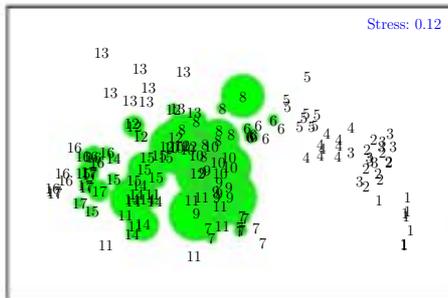
40. *Poa pratensis*



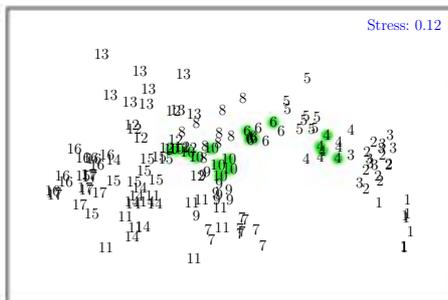
41. *Veronica chamaedrys*



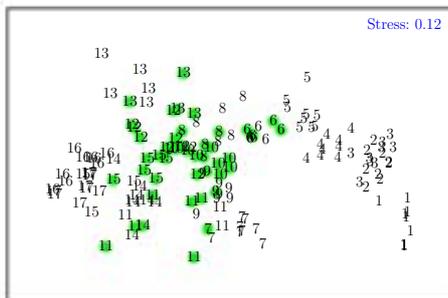
42. *Stellaria holostea*



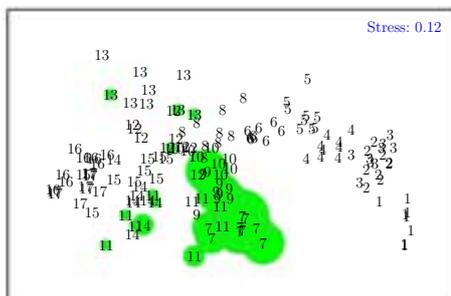
43. *Lathyrus pratensis*



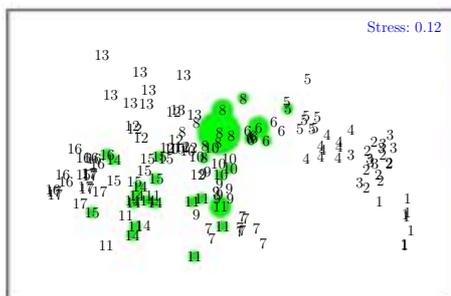
44. *Viola riviniana*



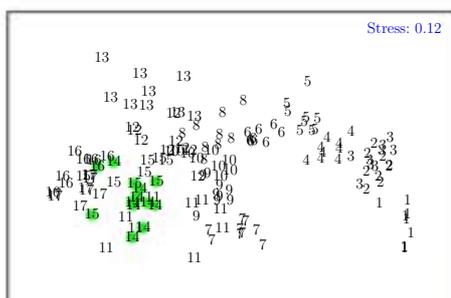
45. *Deschampsia flexuosa*



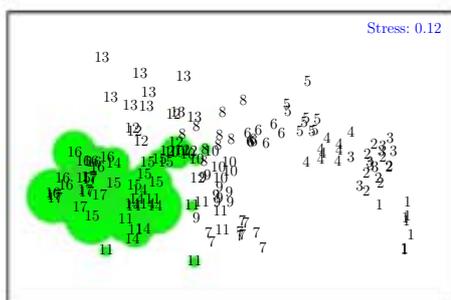
46. *Anthriscus silvestris*



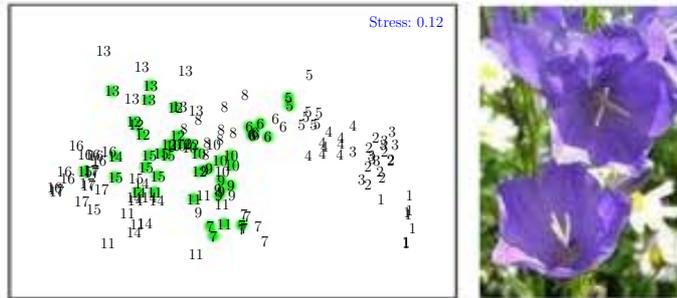
47. *Veronica hederifolia*



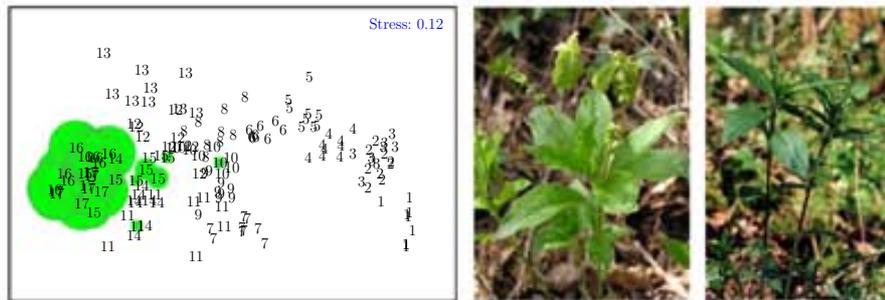
48. *Ranunculus ficaria*



49. *Campanula persicifolia*



50. *Mercurialis perennis*



4.4.2. Análisis de las Especies

A partir del gráfico de análisis de agrupamiento realizado sobre la matriz de similitudes entre cada especie se pueden establecer dos grupos de especies con el 12% de similitud (indicadas con la numeración de la tabla de la abundancia de las especies):

- **Grupo 1:** especies 5, 6, 7, 8, 9, 10, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 y 31.
- **Grupo 2:** especies 1, 2, 3, 4, 12, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 y 50.

O tres grupos de especies con el 14% de similitud:

- **Grupo 1:** especies 5, 6, 7, 8, 9, 10, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 y 31.
- **Grupo 2:** especies 4 y 45.
- **Grupo 3:** especies 1, 2, 3, 12, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 46, 47, 48, 49 y 50.

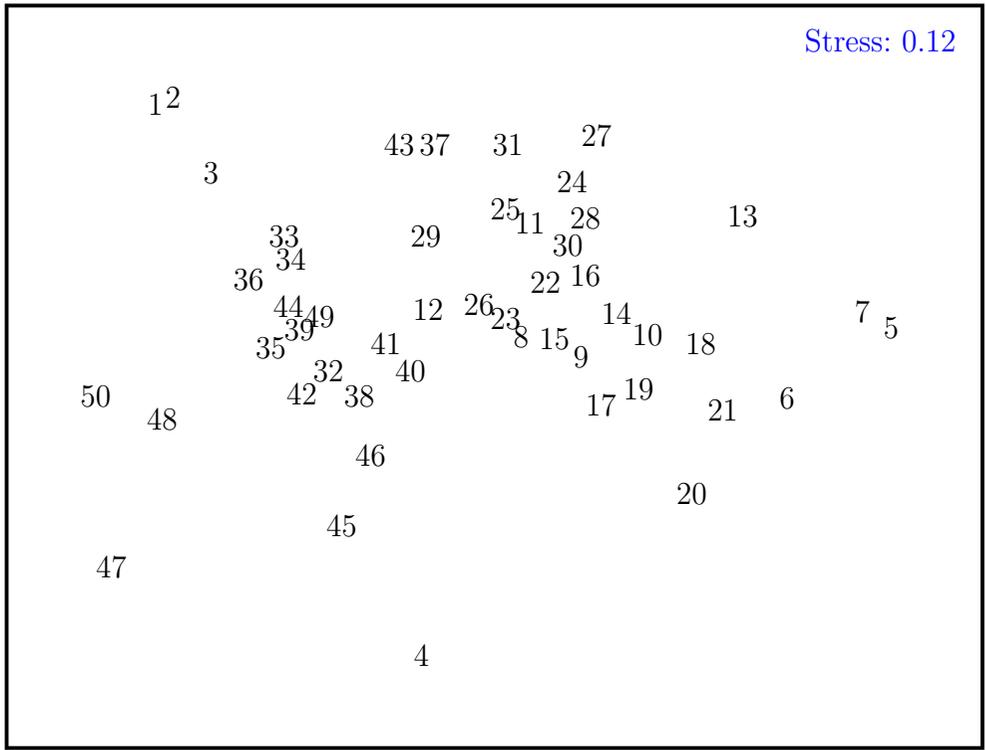
Para el caso del MDS no métrico se pueden visualizar los mismos grupos de especies, los cuales obtuvieron una conformación significativa a partir del análisis de similitudes, ya sea en dos o tres grupos ($p < 0.05$).

Para este caso el MDS no métrico permitió establecer grupos de manera más práctica y con la ventaja adicional de que fue posible evaluar mediante una prueba estadística (ANOSIM) la conformación de los grupos, los cuales, de forma complementaria pueden sugerirse a partir de análisis de agrupamientos.

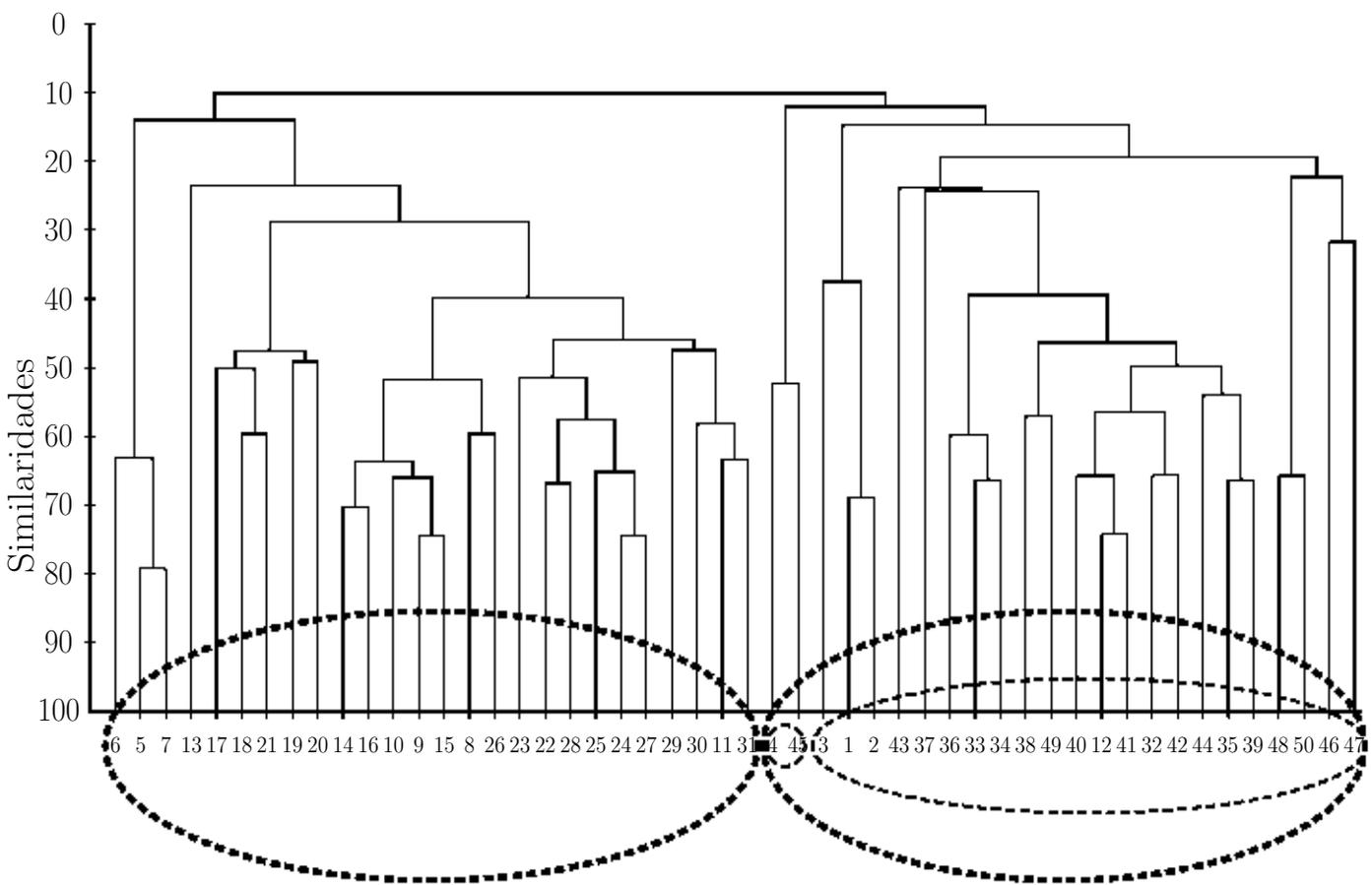
Las tablas muestran el porcentaje de similitud que cada terreno (conjunto de 9 cuadros) aporta utilizando el análisis SIMPER, a partir de dos grupos a priori.

La representación de la abundancia de las especies en cada uno de los terrenos en la configuración del MDS permite ver cómo se distribuye cada terreno dentro de las especies obtenidas.

Escalamiento multidimensional no métrico



Análisis de conglomerados



Análisis de Similaridades (ANOSIM)**Evaluación de 2 grupos (12 % de similitud):**

- **Grupo 1:** 5, 6, 7, 8, 9, 10, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 y 31.
- **Grupo 2:** 1, 2, 3, 4, 12, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 y 50.

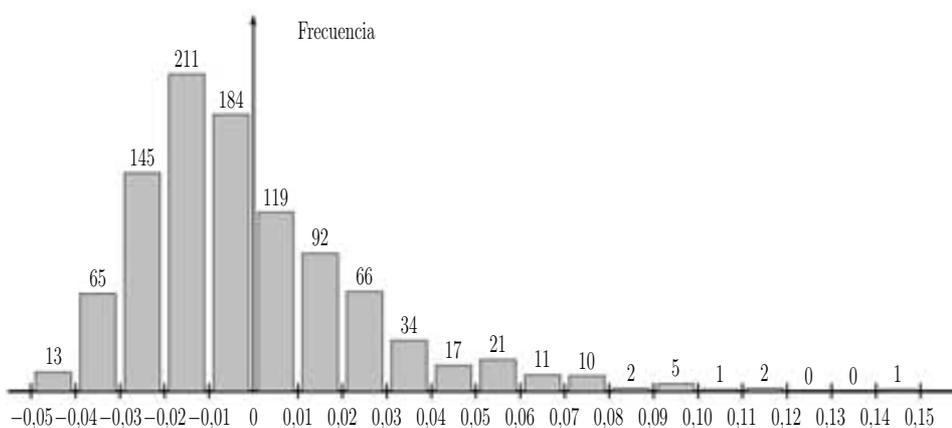
Prueba Global:

Estadística de prueba (correlación R global): 0.599

Nivel mínimo de significancia: 0.001

Número de permutaciones: 999 (de $\binom{50}{26}$ permutaciones posibles).

Número de permutaciones mayores o iguales a R: 0



Evaluación de 3 grupos (14 % de similitud):

- **Grupo 1:** 5, 6, 7, 8, 9, 10, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 y 31.
- **Grupo 2:** 4, 45
- **Grupo 3:** 1, 2, 3, 12, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 46, 47, 48, 49 y 50.

Prueba Global:

Estadística de prueba (correlación R global): 0.635

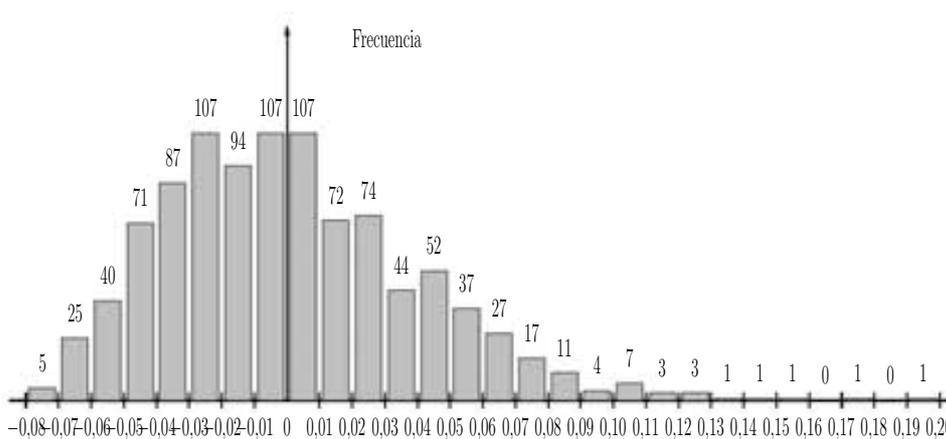
Nivel mínimo de significancia: 0.001

Número de permutaciones: 999 (de $\binom{50}{26} \binom{24}{2}$ permutaciones posibles).

Número de permutaciones mayores o iguales a R: 0

Pruebas por pares:

Grupos	Estadística de prueba R	Nivel mínimo de significancia	Permutaciones posibles	Permutaciones actuales	No. de permutaciones \geq a R
1 y 2	0.853	0.003	$\binom{28}{2} = 378$	378	1
1 y 3	0.628	0.001	$\binom{48}{26}$	999	0
2 y 3	0.542	0.033	$\binom{24}{2} = 276$	276	9



Porcentaje de Similaridades - Contribución de las Especies

SIMPER

Datos estandarizados. Transformación: $\log(a+1)$.

- **Grupo 1:** 5, 6, 7, 8, 9, 10, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 y 31.
- **Grupo 2:** 1, 2, 3, 4, 12, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 y 50.

Separando los grupos 1 y 2

Similaridad promedio:

$$s = \sum_{j=1}^{50} \sum_{k=1}^{50} \frac{s_{jk}}{(50)(50)}$$

donde

$$s_{jk} = \sum_{i=1}^{17} s_{jk}(i) \quad y \quad s_{jk}(i) = \frac{200 \min(y_{ij}, y_{ik})}{\sum_{i=1}^{17} (y_{ij} - y_{ik})}$$

En este caso $s_{jk} = 100 - \delta_{jk}$

y_{ij} es la abundancia (transformada) del i -ésimo terreno en la j -ésima especie para el grupo 1.

En la siguientes tablas tenemos las columnas:

$y_{1i} = \sum_{\{Grupo\ 1\}} \frac{y_{ij}}{26}$ es la abundancia promedio del i -ésimo terreno en el grupo 1.

$y_{2i} = \sum_{\{Grupo\ 2\}} \frac{y_{ij}}{24}$ es la abundancia promedio del i -ésimo terreno en el grupo 2.

$s_i = \sum_{j=1}^{50} \sum_{k=1}^{50} \frac{s_{jk}(i)}{(50)(50)}$ es la contribución promedio del i -ésimo terreno en la similaridad total.

$SD(i) = \left[\sum_{j=1}^{50} \sum_{k=1}^{50} (s_{jk}(i) - s_i)^2 \right]^{1/2} / 50$ es la desviación estandar del i -ésimo terreno.

$s_i \%$ = $\frac{s_i}{\sum_{i=1}^{17} s_i}$ es el porcentaje de contribución del i -ésimo terreno.

$\sum s_i \%$ = $\sum_{h=1}^i \frac{s_h}{\sum_{i=1}^{17} s_i}$ es el porcentaje de la contribución acumulada hasta el i -ésimo terreno.

Grupo 1: $s = 47,22$

Terrenos	y_{1i}	s_i	$s_i/SD(i)$	$s_i\%$	$\sum s_i\%$
5	7.35	8.97	0.90	19.00	19.00
4	6.31	8.69	0.95	18.40	37.39
3	7.42	8.52	0.85	18.04	55.43
2	5.77	8.39	0.81	17.78	73.21
6	4.54	6.84	0.86	14.50	87.70
1	3.73	3.19	0.48	6.75	94.46
10	1.54	1.29	0.37	2.73	97.19
8	2.15	0.87	0.27	1.85	99.04
9	1.42	0.12	0.13	0.26	99.30
7	1.38	0.11	0.10	0.24	99.54
11	0.46	0.11	0.13	0.24	99.78
13	0.23	0.07	0.12	0.15	99.92
12	0.23	0.02	0.09	0.05	99.97
15	0.15	0.01	0.06	0.03	100.00

Grupo 2: $s = 41,77$

Terrenos	y_{2i}	s_i	$s_i/SD(i)$	$s_i\%$	$\sum s_i\%$
10	6.79	6.18	1.07	14.79	14.79
12	5.25	5.74	0.86	13.74	28.53
13	4.50	4.91	0.68	11.74	40.27
6	5.08	4.74	0.80	11.35	51.62
15	4.50	3.75	0.72	8.97	60.59
11	3.75	3.20	0.80	7.65	68.25
14	4.04	2.84	0.66	6.81	75.05
9	5.63	2.72	0.57	6.50	81.55
8	4.46	2.17	0.58	5.20	86.75
16	4.96	1.99	0.45	4.77	91.52
7	4.33	1.62	0.37	3.88	95.40
17	5.42	0.96	0.30	2.29	97.68
5	1.50	0.69	0.32	1.65	99.34
4	1.17	0.27	0.18	0.64	99.97
3	0.25	0.01	0.06	0.03	100.00

Uniando los grupos 1 y 2

Disimilaridad promedio:

$$\delta = \sum_{j=1}^{50} \sum_{k=1}^{50} \frac{\delta_{jk}}{(153)(153)} = 81,78$$

donde

$$\delta_{jk} = \sum_{i=1}^{17} \delta_{jk}(i) \quad \text{y} \quad \delta_{jk}(i) = \frac{100 |y_{ij} - y_{ik}|}{\sum_{i=1}^{17} (y_{ij} + y_{ik})}$$

y_{ij} es la abundancia (transformada) del i -ésimo terreno en la j -ésima especie.

En la siguiente tabla tenemos las columnas:

$y_{1i} = \sum_{\{Grupo\ 1\}} \frac{y_{ij}}{26}$ es la abundancia promedio del i -ésimo terreno en el grupo 1.

$y_{2i} = \sum_{\{Grupo\ 2\}} \frac{y_{ij}}{24}$ es la abundancia promedio del i -ésimo terreno en el grupo 2.

$\delta_i = \sum_{j=1}^{50} \sum_{k=1}^{50} \frac{\delta_{jk}(i)}{(50)(50)}$ es la contribución promedio del i -ésimo terreno en la disimilaridad total.

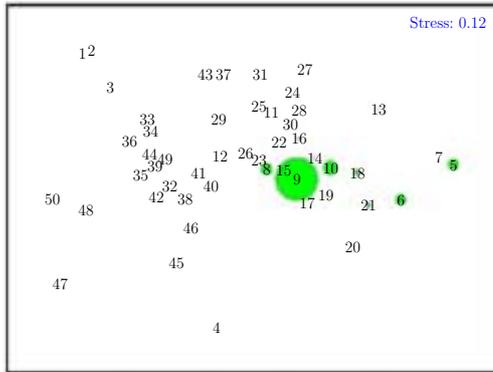
$SD(i) = \left[\sum_{j=1}^{50} \sum_{k=1}^{50} (\delta_{jk}(i) - \delta_i)^2 \right]^{1/2} / 50$ es la desviación estandar del i -ésimo terreno.

$\delta_i \% = \frac{\delta_i}{\sum_{i=1}^{17} \delta_i}$ es el porcentaje de contribución del i -ésimo terreno.

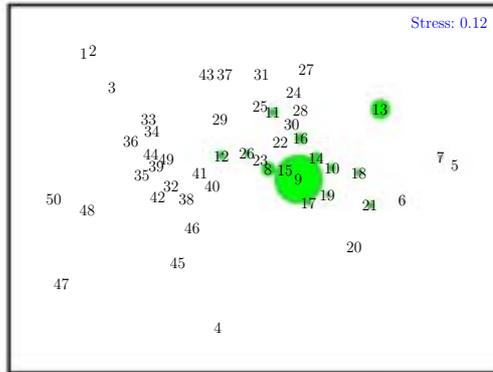
$\sum \delta_i \% = \sum_{h=1}^i \frac{\delta_h}{\sum_{i=1}^{17} \delta_i}$ es el porcentaje de la contribución acumulada hasta la i -ésimo terreno.

Terrenos	y_{1i}	y_{2i}	δ_i	$\delta_i/SD(i)$	$\delta_i\%$	$\sum \delta_i\%$
3	0.25	7.42	6.91	1.16	8.45	8.45
2	0.29	5.77	6.76	1.05	8.26	16.71
5	1.50	7.35	6.53	1.18	7.99	24.70
4	1.17	6.31	6.25	1.20	7.65	32.35
12	5.25	0.23	5.62	1.18	6.87	39.22
13	4.50	0.23	5.59	0.89	6.84	46.06
10	6.79	1.54	5.19	1.20	6.35	52.41
6	5.08	4.54	5.12	1.02	6.26	58.68
15	4.50	0.15	4.28	1.08	5.23	63.91
1	0.04	3.73	4.21	0.74	5.15	69.06
9	5.63	1.42	3.99	0.98	4.88	73.94
8	4.46	2.15	3.93	1.02	4.80	78.74
14	4.04	0.04	3.90	0.88	4.77	83.52
7	4.33	1.38	3.86	0.62	4.72	88.23
11	3.75	0.46	3.71	1.17	4.53	92.77
16	4.96	0.00	3.41	0.77	4.17	96.93
17	5.42	0.00	2.51	0.59	3.07	100.00

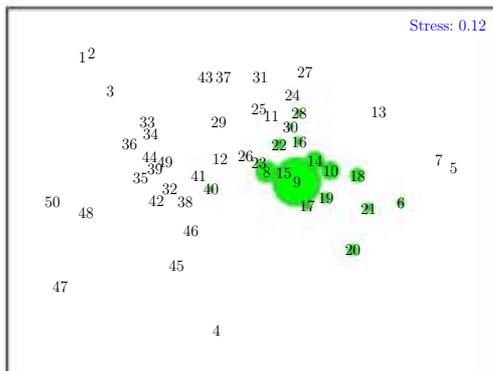
Terreno 1



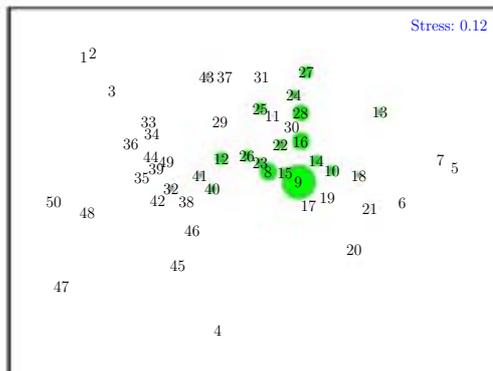
Terreno 2



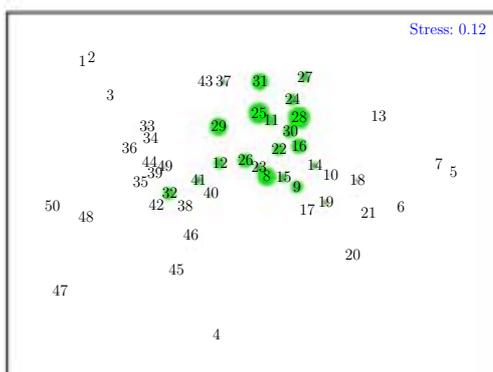
Terreno 3



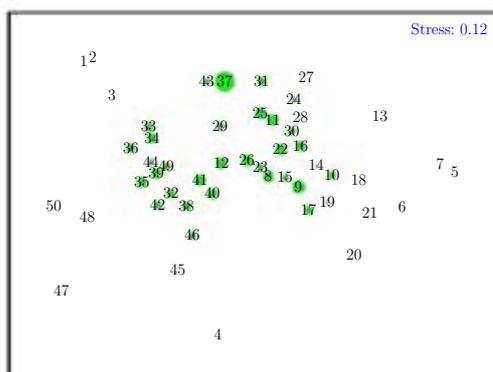
Terreno 4



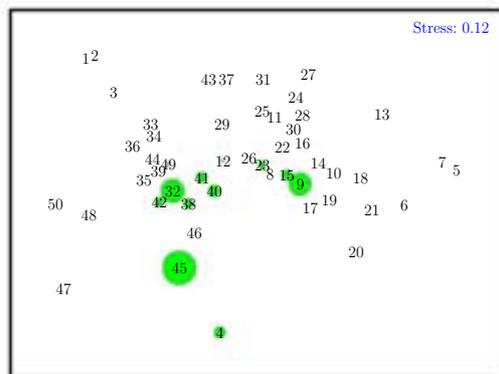
Terreno 5



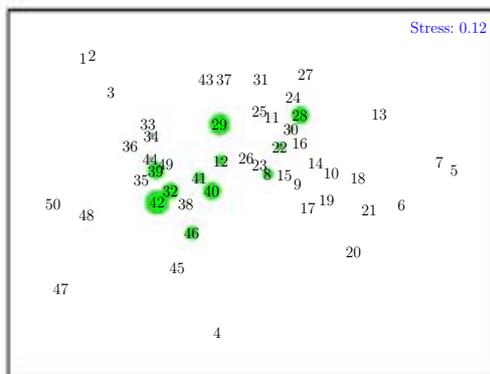
Terreno 6



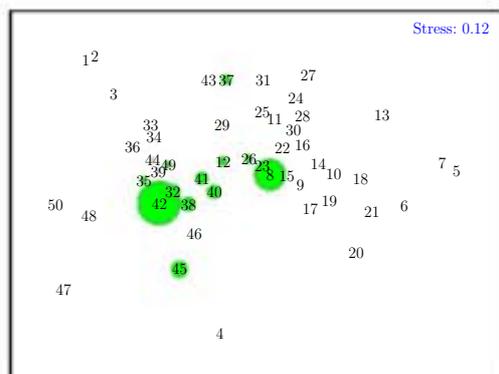
Terreno 7



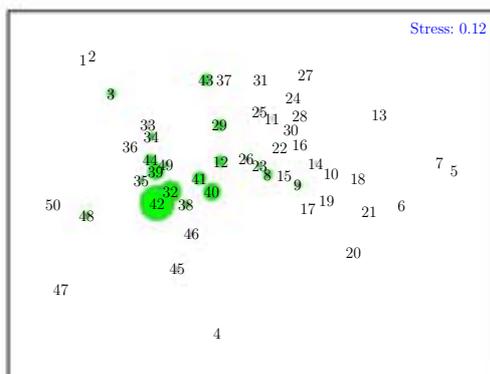
Terreno 8



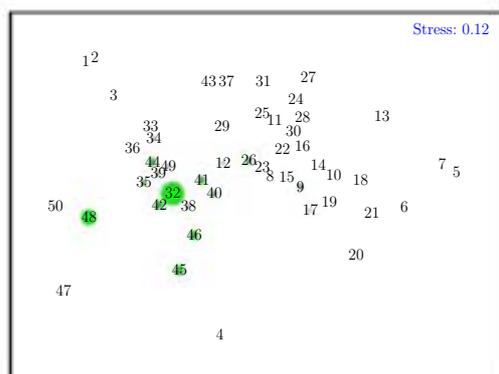
Terreno 9



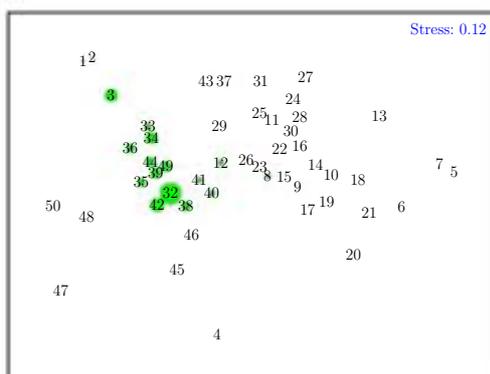
Terreno 10



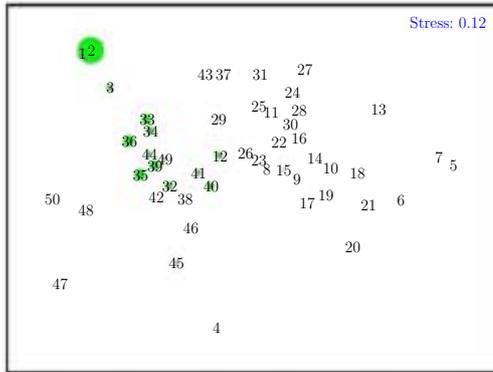
Terreno 11



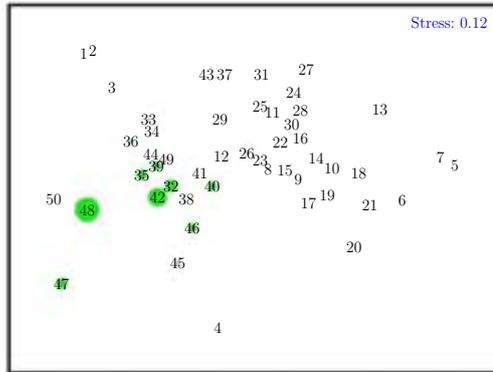
Terreno 12



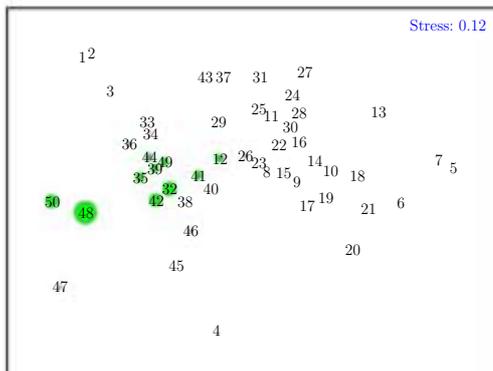
Terreno 13



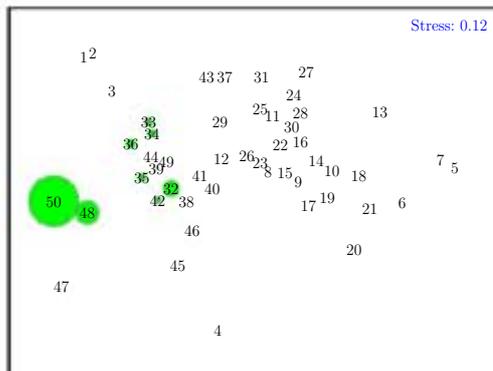
Terreno 14



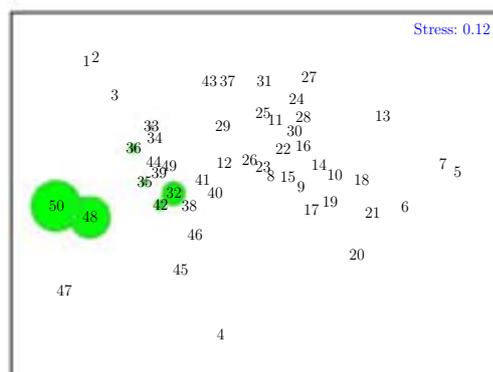
Terreno 15



Terreno 16



Terreno 17



Conclusiones

El MDS no métrico, como un método de la estadística no paramétrica, nos ha permitido visualizar en un plano de dos dimensiones, a partir de una tabla de abundancia de las especies en cada uno de los cuadros, la estructura que tienen los cuadros a partir de una matriz de similaridades entre los 153 cuadros; además el programa PRIMER tiene un análisis adicional, que quizá muchos paquetes estadísticos no contienen, que es la representación de la abundancia de cada una de las especies en la configuración del MDS no métrico, esto nos permite visualizar los cuadros que contienen mayor abundancia de una determinada especie y el parecido que dos o más especies pueden tener según su distribución en este sentido. Además hemos podido visualizar en un plano de dos dimensiones la estructura de las especies a partir de una matriz de similaridades entre las 50 especies y con la representación de la abundancia de cada terreno en la configuración del MDS no métrico hemos podido ver la relación y distribución que tienen.

El análisis de conglomerados ha permitido formar los grupos de cuadros y especies de acuerdo a la medida de abundancia, aunque a partir de una configuración del MDS se pueden identificar estos grupos, un análisis de conglomerados permite identificarlos con mayor precisión. El análisis SIMPER nos ha dado las especies y los cuadros que contribuyen principalmente a la división de los grupos obtenidas por el análisis de conglomerados. El ANOSIM ha permitido probar la significancia que tienen estos grupos definidos a partir del análisis de conglomerados.

Estos métodos de estadística no paramétrica permiten visualizar la estructura de las muestras y verificar la confianza en la conformación de grupos. Quizá podría pensarse que esto se puede realizar a partir de una simple comparación entre los valores de la tabla de abundancias, lo cual para tablas pequeñas es fácil de hacer; sin embargo, como en este caso, para tablas con gran cantidad de datos es necesario e indispensable utilizar los métodos

estadísticos como los que se han aplicado aquí, ya que la gran cantidad de datos hace difícil una simple comparación.

En particular el Escalamiento Multidimensional No Métrico puede ser más útil que el Métrico, ya que tiene menos restricciones, y por tanto permite una gran variedad de tipos de datos de entrada para obtener una configuración. Quizá las desventajas que tiene el MDS No Métrico es que da distintas configuraciones finales a partir de diferentes configuraciones iniciales dado que es un método iterativo y que sólo utiliza la relación de orden entre los datos, sin embargo, esto quizá ayude a dar una mejor configuración de acuerdo al índice de bondad de ajuste que se este utilizando.

El Escalamiento Multidimensional No Métrico puede ser de gran ayuda para observar la cercanía que tienen los objetos; en particular a los ecólogos les permite observar la estructura y distribución de las muestras y de las especies, además éstas configuraciones pueden ayudar a estudiar casos particulares que sucedan en el medio ambiente. Aunque la aplicación dada en este trabajo es un caso particular de los muchos que hay en ecología, y que hay en otras áreas de estudio, este método se puede aplicar en cada una de esos casos de manera análoga.

Bibliografía

- [1] ARCE, C. (1993): *Escalamiento Multidimensional. Una Técnica Multivariante para el Análisis de Datos de Proximidad y Preferencia*. PPU, Barcelona.
- [2] CLARKE, K. R. (1993): “Non-parametric multivariate analyses of changes in community structure”. *Australian Journal of Ecology*, 18, 117-143.
- [3] CLARKE, K. R. y WARWICK, R. M. (1994): “Similarity-based testing for community pattern: the two-way layout with no replication”. *Marine Biology*, 118, 167-176.
- [4] CLARKE, K. R. y WARWICK, R. M. (1998): “Quantifying structural redundancy in ecological communities”. *Oecología*, 113, 278-289.
- [5] COX, Trevor F. y COX, Michael A.A. (2000): *Multidimensional Scaling*. Chapman and Hall, Londres.
- [6] DAVISON, Mark L. (1983): *Multidimensional Scaling*. Wiley, Nueva York.
- [7] DIGBY, P. G. N. y KEMPTON, R. A. (1987): *Multivariate Analysis of Ecological Communities*. Chapman and Hall, Londres.
- [8] DILLON, William R. y GOLDSTEIN, Matthew (1984): *Multivariate Analysis. Methods and Applications*. Wiley, Nueva York.
- [9] EHRICH, S. y STRANSKY, C. (1999): “Fishing effects in northeast Atlantic shelf seas: patterns in fishing effort, diversity and community structure. VI. Gale effects on vertical distribution and structure of a fish assemblage in the North Sea”. *Fisheries Research*, 40, 185-193.

- [10] GOWER, J. C. (1966): "Some distance properties of latent root and vector methods used in multivariate analysis". *Biometrika*, 53, 325-388.
- [11] GOWER, J. C. (1982): "Euclidean distance geometry". *Mathematical Scientist*, 7, 1-14.
- [12] GOWER, J. C. y LEGENDRE, P. (1986): "Metric and Euclidean properties of dissimilarity coefficients". *J. Classification*, 3, 5-48.
- [13] GUTTMAN, L. (1968): "A general nonmetric technique for finding the smallest coordinate space for a configuration of points". *Psychometrika*, 33, 469-504.
- [14] HOLLAND, J. L. (1965): *Manual for Vocational Preference Inventory*. 6ª edición, Consulting Psychologists Press, Palo Alto, CA.
- [15] HOLLAND, J. L. (1973): *Making vocational choices: A theory of careers*. Prentice-Hall, Englewood Cliffs, NJ.
- [16] KAISER, H. F. (1958): "The varimax criterion for analytic rotation in factor analysis". *Psychometrika*, 23, 187-200.
- [17] KRUSKAL, J. B. (1964): "Multidimensional scaling by optimizing goodness-of-fit to a nonmetric hypothesis". *Psychometrika*, 29, 1-28, 115-129.
- [18] KRUSKAL, J. B. y WISH, M. (1978): *Multidimensional Scaling*. CA; Sage, Newbury Park.
- [19] KRUSKAL, J. B. y WISH, Myron (1981): *Multidimensional Scaling*. Sage Publications, Beverly Hills, Londres.
- [20] LINGOES, J. C. y ROSKAM, E. E. (1973): "A mathematical and empirical study of two multidimensional scaling algorithms". *Psychometrika Monograph Supplement*, 38.
- [21] MARDIA K. V., KENT J. T. y BIBBY J. M. (1979): *Multivariate Analysis*. Academic Press, Londres.
- [22] MARGALEF, Ramón (1977): *Ecología*. 2ª edición, Ediciones Omega, Barcelona.

- [23] PERSSON, S. (1981): "Ecological indicator values as an aid in the interpretation of ordination diagrams". *Journal of Ecology*, 69, 71-84.
- [24] REAL DEUS, José Eulogio (2001): *Escalamiento Multidimensional*. La Muralla, Madrid.
- [25] RICHARDSON, M. W. (1938): "Multidimensional psychophysics". *Psychological Bulletin*, 35, 659-660.
- [26] ROUNDS, J. B., Jr., DAVISON, M. L., y DAWIS, R. V. (1979): "The fit between Strong-Campbell Interest Inventory general occupational themes and Holland's hexagonal model". *Journal of Vocational Behavior*, 15, 303-315.
- [27] SAUNDERS, D. R. (1960): "A computer program to find the best-fitting orthogonal factors for a given hypothesis". *Psychometrika*, 25, 207-210.
- [28] SCHOENBERG, I. J. (1935): "Remarks to Maurice Fréchet's article "Sur la définition axiomatique d'une classe d'espaces vectoriels distanciés applicables vectorialement sur l'espaces de Hilbert". *Ann. Math.*, 36, 724-732.
- [29] SHEPARD, R. N. (1962): "The analysis of proximities: Multidimensional scaling with an unknown distance function". *Psychometrika*, 27, 125-140, 219-246.
- [30] SOMERFIELD, P. J., CLARKE, K. R. y OLSGARD, F. (2002): "A comparison of the power of categorical and correlational test applied to community ecology data from gradient". *Journal of Animal Ecology*, 71, 581-593.
- [31] TORGERSON, W. S. (1952): "Multidimensional Scaling: I. Theory and method". *Psychometrika*, 17, 401-419.
- [32] TORGERSON, W. S. (1958): *Theory and methods of scaling*. Wiley, Nueva York.
- [33] YOUNG, G. y HOUSEHOLDER, A. S. (1938): "Discussion of a set of points in terms of their mutual distances". *Psychometrika*, 3, 19-22.