

29.
17



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

**Implantación de Programas
Software para Realizar Técnicas
de Escalamiento Multidimensional.**

T E S I S

QUE PARA OBTENER EL TITULO DE:

A C T U A R I O

P R E S E N T A :

Manuel Gerardo Irazos López

México, D. F.

1984.



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis está 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.

PREFACIO

El objetivo de la presente tesis es el exponer los principales tipos de procedimientos para escalamiento:

El método de escalamiento clásico , y

El método de escalamiento ordinal

Las técnicas de escalamiento encaran esencialmente el problema de encontrar una configuración de puntos (o individuos u objetos) a partir de la información acerca de las distancias entre los puntos, las cuales a su vez pueden estar sujetas a error estocástico. Una característica importante de este tipo de procedimientos, es el poco uso de un modelo estocástico inherente. - Para algunos expertos en métodos de análisis estadístico de datos, las técnicas de escalamiento son métodos de naturaleza exploratoria, cuyo fin principal puede enmarcarse dentro del marco de generación de hipótesis, más que de la prueba de ellas. El tipo de enfoque que proveen estos procedimientos se describe generalmente como "Análisis de Datos", más que "Estadística".

Otro aspecto que caracteriza estas técnicas y las hace diferir en forma importante de otras técnicas multivariadas como Análisis Factorial, Análisis de Correlación Canónica, etc., es el hecho de que operan con una función de los puntos dados, i.e., las $n(n-1)/2$ distancias, disimilaridades o similitudes, más que con las coordenadas de los n puntos mismos. Es decir, las técnicas de escalamiento multidimensional operan con una matriz de $n \times n$ cuyos elementos comparan todos los pares de individuos u objetos de interés, ya sea al medir "similitud" o "disimilaridad". El capítulo 1 trata con cierto detalle las medidas de similitud, disimilaridad y distancia utilizadas más frecuentemente.

El capítulo 2 se ocupa del escalamiento clásico, el cual es esencialmente un método algebraico de reconstrucción de coordenadas que supone a las disimilaridades como distancias euclidianas. En el caso de que estas distancias estén sujetas a error, el comportamiento de este procedimiento es aún bastante satisfactorio. Originalmente propuesto por Torgerson (1952, 1968) - fué popularizado por Gower (1966), quien le dió el nombre de "Análisis de Coordenadas Principales". Se expone también en este capítulo la relación de esta técnica con el método llamado "Análisis de Componentes Principales".

El capítulo 3 expone los fundamentos del escalamiento no métrico o escalamiento ordinal . Cuando las disimilaridades observadas son de tal naturaleza que sus valores numéricos o magnitudes son de poco interés, y el rango de orden de éstas es la única información útil , es conveniente la utilización de una técnica alternativa desarrollada por R.N. Shepard y J.B. - - Kruskal a principios de la década de los sesentas, la cual utiliza solo las propiedades ordinales de las disimilaridades. Los desarrollos y alternativas propuestas hasta hoy, siguen los lineamientos originales propuestos por estos dos autores.

La comparación de dos configuraciones obtenidas por diferentes métodos o aun bajo el mismo pero utilizando una métrica diferente, es expuesto bajo el enfoque de una novedosa técnica debida a Gower llamada Rotación de Procrustes, de la cual se ocupará el capítulo 4.

El tema del capítulo 5 son los métodos numéricos utilizados para la implantación automática de estas técnicas que han sido desarrollados utilizando el lenguaje de programación Algol y la computadora B7800 de la UNAM. Algunos ejemplos de aplicación práctica que ilustran el uso de los procedimientos expuestos en los capítulos anteriores es lo tratado en el capítulo 6.

Agradezco al Dr. Alfonso Hernández Avila, coordinador de la Maestría en Estadística del I.I.M.A.S. , por su valiosa ayuda y apoyo en la realización de esta tesis.

Al Dr. Francisco Aranda Ordaz, Investigador del I.I.M.A.S., por la revisión y corrección del texto así como sus atinadas sugerencias.

A mis tíos Arq. Abel Iraizos , así como a su esposa Sra. Ma. de la Luz Ochoa de Iraizos por su apoyo durante la elaboración de este trabajo.

A la Srta. Ma. del Carmen Hernández por su paciente labor al mecanografiar el texto.

C O N T E N I D O

PREFACIO	i
CAPITULO 1	
MEDIDAS DE SIMILARIDAD Y DISIMILARIDAD	1
CAPITULO 2	
ESCALAMIENTO CLASICO	9
CAPITULO 3	
ESCALAMIENTO ORDINAL	17
CAPITULO 4	
ROTACION DE PROCRUSTES	31
CAPITULO 5	
METODOS NUMERICOS	34
CAPITULO 6	
EJEMPLO NUMERICO	49
APENDICE A	61
APENDICE B	63
APENDICE C	65
APENDICE D	68
BIBLIOGRAFIA	69

CAPITULO 1

MEDIDAS DE SIMILARIDAD Y DISIMILARIDAD

En algunas aplicaciones los datos a analizar provienen de una matriz D ($n \times n$) que consiste de similitudes o disimilitudes entre puntos, más que de una matriz X ($n \times p$) de coordenadas de los puntos, antes de continuar, es necesario definir precisamente lo que se considera una medida de similitud.

DEFINICION

Una medida razonable de similitud $s(A,B)$, debe tener las siguientes propiedades

$$i) s(A,B) = s(B,A)$$

$$ii) s(A,B) \geq 0$$

iii) $s(A,B)$ se incrementa a medida que la similitud entre A y B se incrementa, donde A y B son puntos que pueden representar medidas x , y en dos objetos distintos. Las medidas de similitud son frecuentemente llamadas coeficientes de similitud. Las medidas de similitud y disimilitud están cercanamente relacionadas en forma inversa. Supongamos que tenemos una función s , definida en cada par de individuos, que mide similitud. Una posible medida de disimilitud puede construirse como $d = \text{constante} - s$. Tanto las medidas de similitud como las de disimilitud pueden considerarse como funciones $P \times P$ (donde P es un conjunto de objetos) a los números reales. Jardine y Sibson (1971) definen una clase de funciones de disimilitud, llamados "coeficientes de disimilitud". Un coeficiente de disimilitud es una función d de $P \times P$ a los números reales positivos tal que

$$i) d(A,B) \geq 0 \text{ para todo } A, B \text{ en } P$$

$$ii) d(A,A) = 0 \text{ para todo } A \text{ en } P$$

$$iii) d(A,B) = d(B,A) \text{ para todo } A, B \text{ en } P.$$

Muchos de los procedimientos de clasificación automática y escalamiento pueden ser considerados como métodos en los cuales coeficientes de similitud o disimilitud son convertidos uno en otro (i.e., similitud a disimilitud, o, disimilitud a similitud). Un ejemplo es proporcionado por las técnicas de escalamiento multidimensional no métrico, las cuales son métodos para la transformación de una similitud o disimilitud numérica u ordinal, en un

arreglo métrico de distancias representable en un número pequeño de coordenadas euclidianas. Una función de valores reales $d(A,B)$ es una distancia si tiene las siguientes propiedades

- i) $d(A,B) = d(B,A)$ simetría
- ii) $d(A,B) > 0$ no negatividad
- iii) $d(A,A) = 0$ identificación

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

- iv) $d(A,B) = 0$ si y solo si $A = B$
- v) $d(A,B) < d(A,C) + d(B,C)$ desigualdad del triángulo.

Si las propiedades i) - v) se satisfacen d es llamada una métrica. Para algunos propósitos es suficiente considerar funciones de distancia que satisfagan solamente i) - iii), sin embargo es conveniente en general considerar funciones de distancia para las cuales i) - v) son satisfechas. Es de esperarse que $d(P,Q)$ se incremente mientras la disimilaridad entre P y Q se incremente; por lo que $d(A,B)$ es descrito también como un coeficiente de disimilaridad, aun cuando no cumpla las propiedades de una métrica.

DISTANCIA EUCLIDEANA

Posiblemente la medida de disimilaridad más común sea la distancia euclidiana. Sea X una matriz de datos de $n \times p$ (que representa las coordenadas espaciales de cada uno de los n puntos) con filas x_i , $i=1, \dots, n$. Entonces la distancia euclidiana entre los puntos x_i y x_j es d_{ij} , donde

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} = |x_i - x_j|$$

La función d satisface las propiedades i) - v) además de verificarse:

a) Sea $A = (-\frac{1}{2} d_{ij}^2)$ Entonces HAH es positiva

semidefinida, donde $H = I - n^{-1} 11'$, es la matriz de "centrado" (centering - matrix), $1' (1 \times n) = (1, 1, \dots, 1)$

MÉTRICA DE MINKOWSKY

Una clase importante de funciones de distancia son las llamadas métricas en o métricas r de Minkowsky. Para cualquier $r > 1$, se define la r -distancia entre los puntos x_i, y_j como

$$d_r(x_j, x_i) = \left\{ \sum_{s=1}^n |x_{is} - x_{js}|^r \right\}^{1/r}$$

d_r es una métrica (i.e., satisfacen i) - v). Las métricas r de Minkowsky comparten algunas propiedades con la métrica euclídeana. En particular:

$$i) d_r(x_i, x_j) = d_r(x_i + x_k, x_j + x_k)$$

$$ii) d_r(kx_i, kx_j) = k d_r(x_i, x_j)$$

es decir, estas métricas son invariantes bajo traslaciones y deformaciones uniformes. Un punto de divergencia crucial entre la métrica euclídeana y las métricas l_p , consiste en que la distancia bajo la métrica euclídeana permanece invariable cuando se involucran rotaciones rígidas, mientras que las únicas rotaciones que dejan las distancias sin cambio bajo las métricas l_p , son aquellas que transforman ejes coordenados en ejes coordenados (i.e., aquellas transformaciones lineales cuya matriz asociada es unitaria), por lo que los ejes coordenados tienen una significancia para las métricas l_p que no tienen para la distancia euclídeana.

Existen numerosas medidas de disimilaridad de probada utilidad que no satisfacen la desigualdad métrica. Tal es el caso del cuadrado de la distancia euclídeana y la medida D^2 (la función de distancia de Mahalanobis). Es en consecuencia un requerimiento no necesario para una medida de disimilaridad - el cumplimiento de la desigualdad métrica para ser considerada de utilidad, - muy frecuentemente es de mayor importancia. Las propiedades de orden que posea la medida en cuestión, eg. Sibson (1972) define dos coeficientes d_1, d_2 como de orden global equivalente si

$$d_1(w, x) < d_1(y, z) \text{ si y solo si } d_2(w, x) < d_2(y, z)$$

Si d_1, d_2 son de orden global equivalente entonces d_1, d_2 serán medidas de disimilaridad con el mismo rango de orden. Un ejemplo de este tipo de medidas de disimilaridad está dado por la distancia euclídeana y el cuadrado de ésta.

Hasta aquí consideramos solo medidas o coeficientes de disimilaridad. A continuación se darán algunos ejemplos de coeficientes de similitud.

MEDIDAS DE SIMILARIDAD

Variables Cualitativas

Sean la presencia o la ausencia de p atributos en dos objetos P y Q de notada (x_1, \dots, x_p) , (y_1, \dots, y_p) donde $x_i = 1$ o 0 dependiendo si el i -ésimo atributo está respectivamente ausente o presente en el objeto P .

Hacemos:

$$a = \sum x_i y_i$$

$$b = \sum (1-x_i) y_i$$

$$c = \sum x_i (1-y_i)$$

$$d = \sum (1-x_i) (1-y_i)$$

estos es: a, b, c, d son las frecuencias de $(1,1)$, $(0,1)$, $(1,0)$ y $(0,0)$. La medida más simple de similitud entre p y Q es $s_1(P, Q) = a / p$. Una alternativa es el coeficiente de apareamiento simple (simple matching coefficient).

$$s_2(P, Q) = \frac{a+d}{p}$$

donde $p = a+b+c+d$.

Una dificultad inherente a este coeficiente es la que acarrea la consideración de la ausencia en ambos objetos de cierta característica, en otras palabras, la presencia de un atributo en ambos individuos puede decir más acerca del parecido de estos, que la ausencia del atributo. El coeficiente de Jacquard fué diseñado en forma tal que salve el inconveniente mencionado.

$$S_3(P, Q) = \frac{a}{a+b+c}$$

Sea $X_{(n \times p)}$ una matriz de datos que contenga la información ausencia/ presencia de p atributos para n objetos. La matriz X se llama matriz de incidencia ya que $x_{ij} = 1, 0$.

Considerando el coeficiente de similaridad S_1 , la matriz de similaridades correspondiente es simplemente

$$S_1 = \frac{XX^T}{p}$$

mientras que la matriz de similaridades basada en S_2 es:

$$S_2 = \frac{XX^T + (J - X)(J - X)^T}{p}$$

donde $J = 11^T$

$$S_2 = s_{ij}^2$$

$$1^T = (1, \dots, 1)$$

VARIABLES MIXTAS

Si hay variables tanto cualitativas como cuantitativas, Gower (1971a) ha propuesto el siguiente coeficiente de similaridad entre el i -ésimo y j -ésimo puntos.

$$S^{(3)} = 1 - \frac{1}{p} \sum_{k=1}^p w_k |x_{ik} - x_{jk}|$$

donde $w = 1$ si k es cualitativa, $w = \frac{1}{R_k}$

si k es cuantitativa y donde R_k es el rango de la k -ésima variable.

Existen muchas otras medidas de las cuales no nos ocuparemos ya que no es tal el tema principal de esta exposición. Finalmente podemos citar a Goodman (1954) quien en su artículo clásico dentro de las medidas de asociación argumenta que: "idealmente, cada trabajo de investigación debería contar con una medida de asociación desarrollada para sus necesidades únicas". A través de los años los científicos han desarrollado muchas diferentes medidas. Por lo tanto, cuando se busca una medida de similaridad o asociación, hay raramente necesidad de desarrollar una nueva. Más que otra cosa el problema consiste en escoger sabiamente de entre las medidas ya existentes.

ESCALAMIENTO CLASICO

El escalamiento clásico es un método de reconstrucción algebraico para encontrar una configuración de puntos a partir de las disimilaridades entre los mismos. Este enfoque es particularmente apropiado cuando las disimilaridades son exactas, o, al menos aproximadamente euclidianas.

Supongamos que tenemos una configuración X en un espacio euclideo p -dimensional. Las distancias entre los puntos de la configuración se pueden calcular directamente a partir de la matriz $X = x_{ij}$ $i=1, \dots, n$; $j=1, \dots, n$. Una alternativa más conveniente es proporcionada por el uso de la matriz de $(n \times n)$ $B = XX^T$, la matriz de suma de productos y de cuadrados (la matriz de productos escalares entre vectores puntos) la matriz B , i.e., b_{rs} está dado por la siguiente expresión

$$b_{rs} = \sum_{k=1}^p x_{rk} x_{sk} \quad (1)$$

La matriz D $(n \times n)$ de cuadrados de distancias euclidianas de acuerdo a lo anterior, está dado por los elementos:

$$d_{rs} = \sum_{j=1}^p (x_{rj} - x_{sj})^2$$

$$= \sum_{j=1}^p x_{rj}^2 - 2 \sum_{j=1}^p x_{rj} x_{sj} + \sum_{j=1}^p x_{sj}^2$$

Es decir las distancias d_{ij} se pueden representar en función de los elementos de B o de las coordenadas x_{ij} .

Ahora supongamos que conocemos las distancias (y por consiguiente los cuadrados de estas) pero no las coordenadas de los puntos que las generan. Podemos imaginar el proceso de obtención de coordenadas como el problema inverso al que acabamos de exponer. La determinación consiste en dos etapas:

- i) La construcción de la matriz B
- ii) La factorización de la matriz B en la forma $B = XX^T$

Es evidente que a partir de (1) es posible encontrar b_{ij} en términos de las distancias conocidas, i.e., d_{ij} . Como en muchos otros problemas, la obtención de una solución única depende enteramente en la imposición de ciertas restricciones. En nuestro caso particular, tal restricción es del tipo de localización, y consiste en poner el centro de gravedad \bar{X} en el origen. - Tal restricción puede reformularse en la siguiente condición

$$\sum_{r=1}^n x_{rj} = 0$$

Ahora bajo esta restricción tenemos :

$$\sum_{s=1}^n b_{rs} = \sum_{s=1}^n \sum_{j=1}^p x_{rj} x_{sj} = \sum_{j=1}^p \sum_{r=1}^n x_{rj} x_{sj} = \sum_{j=1}^p x_{sj} \sum_{r=1}^n x_{rj} = 0$$

Esto es, la suma de términos en cualquier columna o fila de B es nula. Sumando (1) sobre r, sobre s y sobre r y s, tenemos:

$$\sum_r d_{rs}^2 = (b_{rr} + b_{ss} - 2b_{rs}) = T + nb_{ss}$$

$$\sum_s d_{rs}^2 = (b_{rr} + b_{ss} - 2b_{rs}) = T + nb_{rr}$$

$$\sum_{r,s} d_{rs}^2 = (b_{rr} + b_{ss} - 2b_{rs}) = nT + nT = 2nT$$

donde T es la traza de la matriz B.

Definimos

$$d_{r.}^2 = \frac{1}{n} \sum_{s=1}^n d_{rs}^2$$

$$d_{s.}^2 = \frac{1}{n} \sum_{r=1}^n d_{rs}^2$$

$$d_{..}^2 = \frac{1}{n} \sum_{r=1}^n \sum_{s=1}^n d_{rs}^2$$

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

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

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

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

Si D consiste en cuadrados de distancias euclideanas exactas, es posible demostrar que B es una matriz simétrica, positiva semidefinida. Si B es de rango $k < n$, entonces tendrá k valores propios distintos de cero. Sean l_1, l_2, \dots, l_p tales valores, los vectores propios de longitud unitaria $\{e_i\}$ son normalizados de tal manera que la suma de los cuadrados de sus entradas sea igual l_i . Esto se logra haciendo $f_i = l_i e_i$. Mostraremos que una posible matriz de coordenadas está dada por

$$X = | f_1, f_2, \dots, f_k |$$

La matriz X es de dimensión $n \times k$, así que en consecuencia tenemos una configuración en k dimensiones. El hecho de que una matriz positiva semidefinida pueda ser factorizada en la forma XX' depende del teorema de factorización de Young-Householder. El hecho de que X es de la forma $X = | f_1, \dots, f_k |$ depende solamente en mostrar que si X tiene esta forma, entonces para $i=1, \dots, k$

$$(XX^*)e_i = X \begin{array}{c} f^* \\ 1 \\ : \\ : \\ : \\ : \\ f^* \\ k \end{array} = X \begin{array}{c} e_i \\ = X \\ 0 \\ : \\ 1 \\ i \\ : \\ 0 \end{array} = X f_i = l_i e_i^*$$

-(i-ésima fila)

Ya que e_i son ortonormales y $f_i = l_i e_i$. Pero $Be_i = l_i e_i$ por ser e_i un vector propio; para $i=1, \dots, k$. Ya que los e_i forman una base ortonormal se concluye que $B = XX^*$.

En resumen, la matriz X puede ser vista en terminos de los vectores propios de B y los puntos correspondientes

Vectores Propios

		l_1	l_2	l_p
u	P_1	x_{11}	$x_{12} \dots$	x_{1p}
n	P_2	x_{21}	$x_{22} \dots$	x_{2p}
o	P_3	x_{31}	$x_{33} \dots$	x_{3p}
s	:	:	:	:
	P_n	x_{n1}	$x_{n2} \dots$	x_{np}

Es decir, la r -ésima fila de X contiene las coordenadas del r -ésimo punto mientras que la r -ésima columna de X contiene al vector propio asociado a l_i . Todo lo dicho hasta aquí es válido mientras B no presente valores propios negativos, debido a la imposibilidad de normalizar un vector que tenga norma cuadrada negativa.

MEDIDAS DE BONDAD DE AJUSTE

Una vez encontrada la configuración asociada a una matriz de distancia D , esperaremos que al escoger una dimensión baja para el espacio en el cual se encuentra, las distancias entre los puntos de ésta, se asemejarán cercanamente a las entradas de la matriz considerada D .

Consideremos una configuración X en R^p , un enfoque particularmente apropiado al visualizar una configuración en k dimensiones, es considerarla como la proyección de la configuración sobre un espacio de R^p generada por las columnas de una matriz ortogonal L_1 , siendo más explícitos

$$\text{Sea } L_{(p \times p)} = (L_1, L_2), \quad L_1_{(p \times k)}$$

$$\hat{X} = XL_1$$

al ser L ortogonal, las distancias entre las filas de X y aquellas de XL , serán iguales i.e.

$$d_{rs}^2 = \sum (x_{ri} - x_{si})^2 = \sum (x'_r l_{(i)} - x'_s l_{(i)})^2$$

Sea \hat{D} la matriz de distancias entre las filas de XL_1 , entonces

$$\hat{d}_{rs} = \sum_{l=1}^k (x'_r l_{(l)} - x'_s l_{(l)})^2$$

de aquí se sigue en forma directa la desigualdad

$$\hat{d}_{rs} \leq d_{rs}$$

Es un hecho digno de consideración darse cuenta de la importancia de este resultado, pues podríamos considerar la discrepancia entre la configuración original y la configuración ajustada como cuantificada por

$$\emptyset = \sum (d_{rs} - \hat{d}_{rs})^2$$

Es posible demostrar que de entre todas las proyecciones XL_1 de X sobre espacios k -dimensionales de R^p , \emptyset se minimiza cuando X se proyecta sobre -

sus coordenadas principales en k dimensiones, obteniendo que el valor correspondiente para esta proyección en coordenadas principales

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

Cuando D es no necesariamente euclídeana es más conveniente trabajar con la matriz $B = HAH$, $H = I - n^{-1} 1'1$, $1 = (1, \dots, 1)$. Si \hat{X} es una configuración ajustada con matriz asociada \hat{B} entonces una medida de discrepancia entre B y \hat{B} está dada por

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

encontrándose que al igual que en el caso anterior que para una k fija la medida Ψ es minimizada sobre todas las configuraciones X en k dimensiones - cuando X es la solución clásica al problema del escalamiento multidimensional, hallándose que el valor mínimo de Ψ está dado por

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

De aquí que la discrepancia total entre la configuración en cuestión X y cualquier otra X ajustada está dada al considerar la suma total de los valores característicos en cada caso y no a partir de k+1.

Lo anterior sugiere posibles medidas de concordancia para "la proporción de matriz de distancia explicada". Suponiendo $\lambda_k > 0$, Mardia (1970) propone

$$\alpha_{1,k} = \left(\sum_{i=1}^k \lambda_i / \sum_{i=1}^n \lambda_i \right) \times 100 \%$$

$$\alpha_{2,k} = \left(\sum_{i=1}^k \lambda_i^2 / \sum_{i=1}^n \lambda_i^2 \right) \times 100 \%$$

RELACION ENTRE EL ESCALAMIENTO CLASICO Y EL
ANALISIS DE COMPONENTES PRINCIPALES

En esta sección se describe la conexión entre el escalamiento clásico y el análisis de componentes principales, suponiendo que una matriz $X_{n \times n}$ corregida por la media es dada, en lugar de una matriz de disimilaridad (para una definición de componentes principales ver el apéndice B). Como se vió en la sección anterior, el problema del escalamiento clasico, se reduce básicamente a encontrar los valores propios y los correspondientes vectores propios - asociados a la matriz XX' . Para efectuar un análisis de componentes principales es necesario encontrar los valores y vectores propios de la matriz de covarianza muestral $X'X$. Una pregunta natural que surge, es si existe alguna - conexión entre los valores y vectores característicos de XX' y $X'X$, si alguna relación existe, entonces habrá una liga interesante entre estas técnicas.

Sean λ_i, a_i los valores y vectores propios de $X'X$ entonces $X'Xa_i = \lambda_i a_i$

Premultiplicando por X , encontramos

$$(XX')Xa_i = \lambda_i Xa_i$$

Lo anterior nos permite visualizar que los valores característicos de XX' son los mismos que aquellos de $X'X$, mientras que los vectores característicos de $X'X$, a_i estarán relacionados por una transformación lineal simple a los de XX' , pues e_i debe ser proporcional a Xa_i .

Es importante observar que e_i es un vector de orden $(n \times 1)$ mientras a_i es de orden $(p \times 1)$. Si suponemos que tanto a_i como e_i han sido normalizados de tal forma que la suma de cuadrados de sus entradas sea unitaria, Xa_i es un vector que da los puntajes (scores) para cada individuo en relación a la i -ésima componente, y la suma de cuadrados de Xa_i se sabe que es λ_i . Por lo tanto debemos tener

$$\lambda_i a_i = Xa_i$$

(posiblemente despues de cambiar el signo de e_i o a_i).

Esto tiene como consecuencia inmediata que, es posible obtener los punta

jes (scores) de las componentes directamente de los vectores propios XX' en lugar de transformar los vectores propios de $X'X$. La fórmula que relaciona lo anterior es:

$$f_i = \frac{1}{\lambda_i} e_i$$

Por lo tanto, los resultados de un análisis de componentes principales son equivalentes (salvo posiblemente una reflexión) a los obtenidos del escalamiento clásico. Gower (1966) muestra que los resultados anteriores se aplican también a una matriz de datos que consista de ceros y unos, si las distancias se calculan con el coeficiente de apareamiento simple (simple matching coefficient). La equivalencia del análisis de componentes principales y el escalamiento clásico significa que, si queremos reducir la dimensionalidad efectiva de una matriz de datos al transformar a nuevos ejes coordenados, entonces, no tiene sentido efectuar ambos análisis. Esto implica que si $n > p$, entonces un análisis de componentes principales se preferirá generalmente, ya que es ciertamente más fácil encontrar los valores propios de una matriz $(X'X)_{p \times p}$ que de una matriz $(XX')_{n \times n}$.

El escalamiento clásico es sorprendentemente robusto a alejamientos de la distancia euclídeana. Errores aleatorios comparativamente grandes pueden ser añadidos a las distancias euclídeanas, sin distorsionar seriamente a la configuración resultante; aun si tomásemos una función monótona de las distancias, obtendríamos una configuración razonable. Sin embargo, si las disimilaridades no son ni siquiera aproximadamente euclídeanas, entonces el escalamiento clásico no es recomendable. Esta situación puede ser indicada al encontrar que B tiene un valor propio negativo "grande" o al encontrar que B tiene varios valores característicos positivos de tamaño medio, así que la solución trata de dispersarse en un número grande de dimensiones. En estas situaciones es preferible una técnica que utiliza el rango de orden de las disimilaridades, la cual será descrita en el capítulo siguiente.

CAPITULO 3

ESCALAMIENTO NO METRICO

Definición del problema

El objetivo del escalamiento no métrico en forma general, es el de encontrar n puntos tales que, las distancias entre ellos se asemejen en algún sentido a las disimilaridades experimentales entre n objetos, utilizando exclusivamente el conocimiento del rango de orden de las disimilaridades observadas entre los objetos de interés. No existe razón por la cual no sea posible el empleo de medidas de proximidad o disociación de cualquier otro tipo. Tales medidas incluirían por ejemplo: coeficientes de correlación, probabilidades de confusión, relaciones de interacción entre grupos, etc. El hecho de que un valor grande implique cercanía o lejanía, es algo que no tiene relevancia alguna. Lo que es esencial para el empleo de esta técnica, es la existencia de una relación monótona, ya sea creciente o decreciente, entre las medidas experimentales y las distancias en la configuración.

UN POCO DE HISTORIA

Durante la época de aparición del libro de Torgerson (1958), los procedimientos de escalamiento en uso corriente podrían describirse como:

i) Procedimientos generales que pueden desglosarse en dos etapas. La primera emplea una técnica de escalamiento unidimensional que tiene como objeto convertir las similaridades o disimilaridades en distancias. La segunda etapa tiene como meta encontrar puntos cuyas distancias entre sí, tengan aproximadamente los valores encontrados en la primera etapa. Un gran avance en este campo fué la aparición de una técnica debida a Shepard, que marcaría el camino a seguir en investigaciones posteriores, que pueden describirse como:

ii) Procedimientos generales que adoptan como característica principal, la obtención de una relación monótona entre las disimilaridades o similaridades experimentales y las distancias en la configuración. Los procedimientos mencionados en i), debido a la naturaleza de las técnicas de escalamiento unidimensional disponibles en aquella época, aceptaban ya sea disimilaridades promediadas o alguna transformación fija de ellas como distancias, o usaban la variabilidad de los datos como elemento crítico en la formación de las distancias.

Los procedimientos en ii) establecen que la satisfactoriedad de una solución debe ser juzgada por el grado en que la condición de monotonía es satisfecha. Shepard muestra que simplemente pidiendo un alto grado de satisfactoriedad en este sentido, y, sin hacer uso de la variabilidad en ninguna forma, es posible obtener soluciones fuertemente restringidas, recuperando al mismo tiempo la forma de las supuestas, pero desconocidas relaciones monótonas; es decir, para la determinación de la solución es necesario solamente conocer el rango de orden de las disimilaridades. No está por demás recalcar que lo anterior convierte a esta clase de procedimientos, ajenos a cualquier suposición distribucional, así como de suponer que las disimilaridades y las distancias están relacionadas por alguna fórmula fija. Esta solución se conoce con el nombre de escalamiento multidimensional.

A pesar de la consideración que se hace acerca de la calidad de la solución a este problema (el de el escalamiento multidimensional), no se incorpora la medida de alejamiento de la condición de monotonía, como criterio principal para la terminación de los procedimientos.

Shepard (1962) describe e implementa un procedimiento iterativo con la ayuda de una computadora cuyos lineamientos generales se ajustan aquellos mencionados en ii).

Kruskal (1964) en un intento por fundamentar y perfeccionar las ideas de Shepard propone una técnica similar a la de regresión por mínimos cuadrados, la que a su vez se inscribe dentro del marco central de una relación monótona entre las disimilaridades y las distancias.

Brevemente la técnica de Kruskal consiste en lo siguiente:

Dada una configuración, efectúa una regresión monótona de distancia con disimilaridad y usa la varianza residual, como una medida cuantitativa de alejamiento de la relación monótona, a la que llama stress. Define una solución el problema de escalamiento multidimensional como la configuración que minimiza el stress, dando de esta manera una definición matemática explícita de una solución, lo cual había hasta ese momento permanecido extrañamente descuidado por muchos autores, incluyendo al propio Shepard.

Sea $\&_{ij}$ la disimilaridad obtenida experimentalmente entre los objetos i y j . Suponemos que el procedimiento experimental es inherentemente simétrico de manera que $\&_{ij} = \&_{ji}$ ignorando las autodisimilaridades (i.e, $\&_{ii}$).

Consideremos n objetos, entonces existen $n(n-1)/2$ números $\&_{ij}$. Suponemos además la siguiente condición: $\&_{ij} \neq \&_{kl}$ para toda $i \neq k$ y para toda $j \neq l$

La condición anterior nos habilita para ordenar las disimilaridades - en estricto orden creciente:

$$\psi \quad \&_{i(1)j(1)} < \dots < \&_{i(M)j(M)}$$

donde $M = n(n-1)/2$

Deseamos representar los n objetos como n puntos en un espacio t -dimensional donde por el momento suponemos que sabemos el valor apropiado de t . Más tarde se abordará la cuestión de la determinación de tal valor.

Sea $X = \{ x_i \mid i=1, \dots, n; x_i \text{ en } R^t \}$ llamamos a X una configuración. Esperamos encontrar la configuración que mejor se ajuste a un conjunto dado de puntos, por lo que es necesario la construcción de un criterio que nos permita juzgar la bondad de cualquier configuración. Tal criterio se desarrolla a continuación:

EL STRESS

Sea d_{ij} la distancia que existe entre x_i y x_j (donde por el momento se considerara la distancia euclídeana y más tarde al igual que con las restricciones hechas a las disimilaridades, serán sustituidas por un esquema de suposiciones más generales).

Intuitivamente la aproximación perfecta a las disimilaridades a partir - de las distancias, se daría en el caso de que se observase lo siguiente:

$$\&_{i(K)j(K)} < \&_{i(1)j(1)} \text{ implica } d_{i(K)j(K)} < d_{i(1)j(1)}$$

en otros términos la aproximación perfecta se da cuando las distancias entre

los puntos de la configuración presentan el mismo orden que las correspondientes disimilaridades

$$d_{i(1)j(1)} < \dots < d_{i(M)j(M)}$$

Examinemos el diagrama que se presenta en la figura 3.1. Como pueden observarse, consta de puntos con abscisa d_{ij} y ordenada \hat{d}_{ij} . Este tipo de diagramas son conocidos como diagramas de dispersión (scatter diagrams). Podemos hacer uso de los diagramas de dispersión para visualizar que también las distancias se asemejan a las disimilaridades. En términos de diagramas de dispersión, el concepto intuitivo de aproximación perfecta, se traduce que al trazar cada punto $(d_{i(k)j(k)}, \hat{d}_{i(k)j(k)})$ empezando con aquel con menor ordenada y continuando de esta forma hasta llegar a aquel con ordenada mayor, nuestra dirección de trazo sea siempre a la derecha, nunca a la izquierda, fig. 3.2. Para medir el alejamiento de una configuración de la aproximación perfecta, es natural ajustar una curva que represente la aproximación perfecta al conjunto de puntos $(d_{i(k)j(k)}, \hat{d}_{i(k)j(k)})$ $k = 1, \dots, n(n-1)/2$ y después medir la desviación de cada punto a la curva. Las desviaciones entre la curva y los puntos se medirán a lo largo del eje de las distancias. Es conveniente hacer una reflexión en cuanto a lo anterior. ¿Que implicaría medir las desviaciones entre la curva y los puntos a lo largo del eje de las disimilaridades?. Implicaría ciertamente, dar un valor numérico a los rangos, es decir, hacer uso de otra información que solo el rango de orden de las disimilaridades. Lo anterior, en consecuencia, no sería la base para la construcción de una medida de bondad de ajuste, en la que no hay efecto de distorsión monótona en el eje de las disimilaridades. La forma en que se medirán las desviaciones implica que de la curva ajustada se necesitan solo M puntos, i.e.,

$$(d_{i(k)j(k)}, \hat{d}_{i(k)j(k)})$$

De aquí que ajustar la curva se reduce a ajustar los valores de $\hat{d}_{i(k)j(k)}$

Por supuesto, es necesario recalcar que los números d_{ij} no son distancias. No existe necesariamente una configuración de puntos cuyas distancias entre sí sean \hat{d}_{ij} . Estos números son meramente una sucesión monótona de números que se escogen tan cerca de las d_{ij} como sea posible, los cuales se usan como referencia de la extensión en la cual los números d_{ij} se alejan de la relación de monotonía. La definición exacta de los números \hat{d}_{ij} se hará más tarde con el -

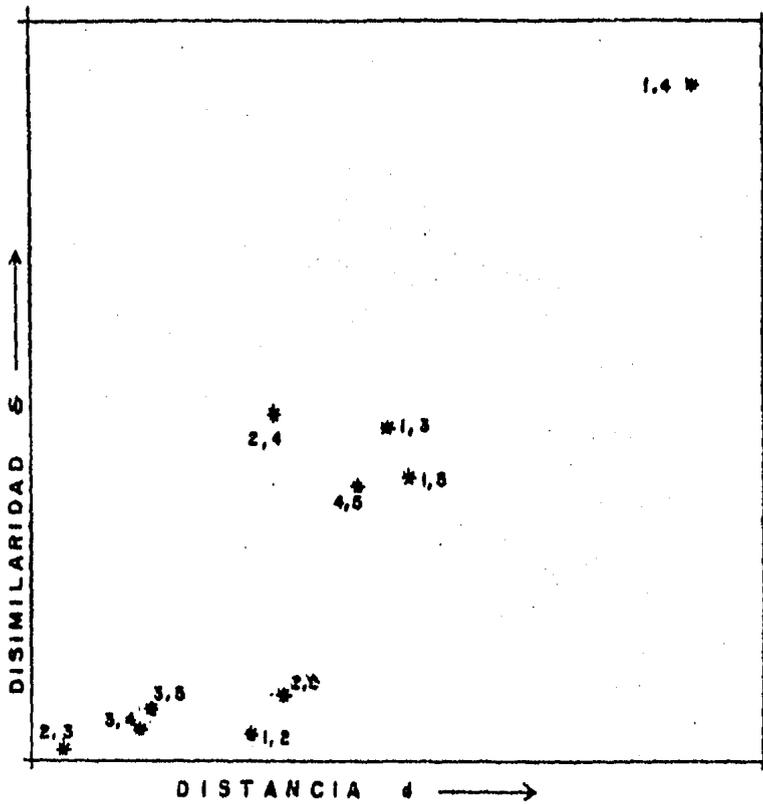


FIG. 3.1

DIAGRAMA DE DISPERSION

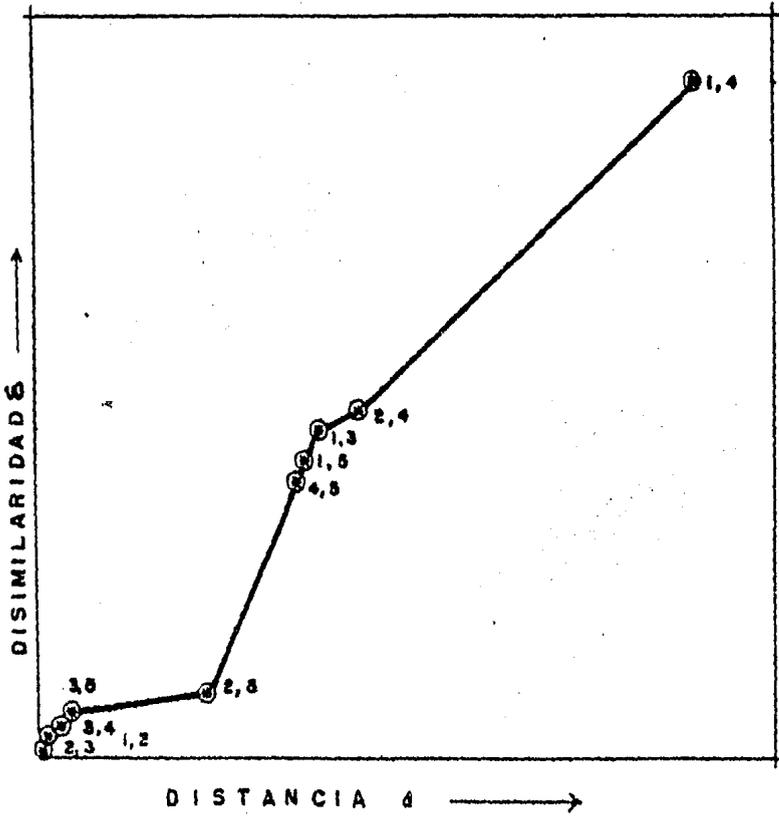


FIG. 32

RELACION ASCENDENTE PERFECTA

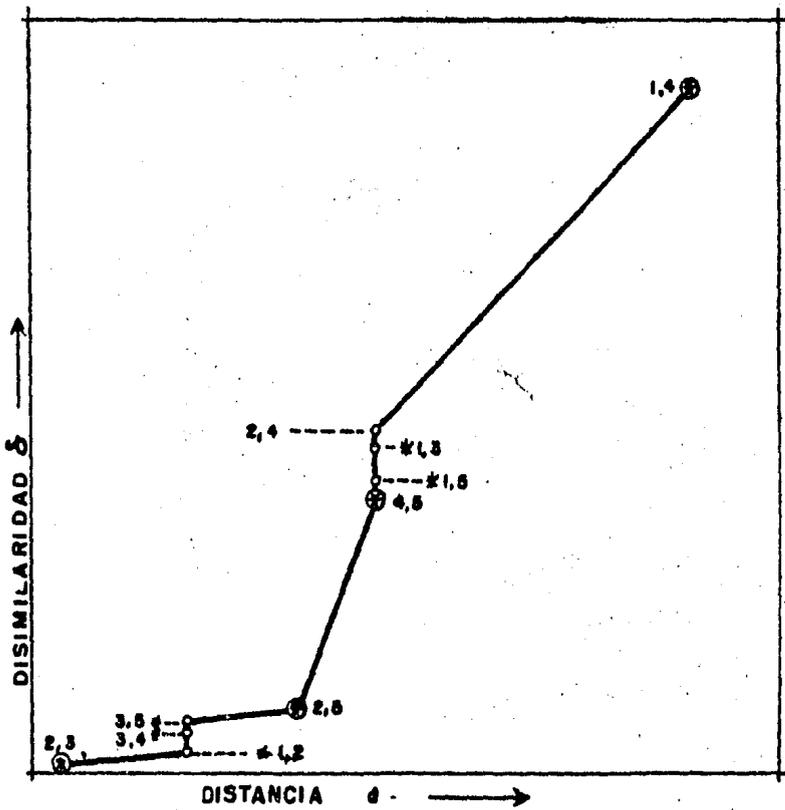


FIG. 3.2 (b)

RELACION ASCENDENTE IMPERFECTA

fin de simplificar la exposición. La curva ajustada es desde luego de carácter ascendente, esto es, con las mismas características que las expuestas en términos del diagrama de dispersión para una configuración con aproximación perfecta. Haciendo referencia a los números \hat{d}_{ij} , esto significa que cuando son arreglados en el orden normal, i.e.,

$$d_{i(1)j(1)}, \dots, d_{i(M)j(M)}$$

entonces se verificará la siguiente relación

$$\hat{d}_{i(1)j(1)} < \dots < \hat{d}_{i(M)j(M)}$$

De acuerdo a lo anterior el diagrama de la figura 3.2(b) muestra una relación ascendente imperfecta, debido a que el conjunto de distancias no guarda la misma relación ascendente que las correspondientes disimilaridades, en otras palabras, $\&$ y d no son de orden global equivalente (Sibson (1972)). Un ejemplo lo proporciona el par de puntos 1,2 ; 3,4 correspondientes a $(d_{12}, \&_{12})$, $(d_{34}, \&_{34})$. Para este par en particular observamos

$$\&_{12} < \&_{34} \text{ no implica } d_{12} < d_{34}$$

La misma situación se hace evidente al observar los pares de puntos $(d_{24}, \&_{24}), (d_{15}, \&_{15}) ; (d_{12}, \&_{12}), (d_{35}, \&_{35})$. En la figura 3.2 (b) se muestra además la curva ajustada por regresión monótona a los puntos $(d_{i(k)j(k)}, \&_{i(k)j(k)})$. Los puntos $(d_{i(k)j(k)}, \&_{i(k)j(k)})$ se han representado en el diagrama por medio de un *, y los puntos $(\hat{d}_{ij}, \hat{\&}_{ij})$ por medio de un pequeño círculo. En caso de que el punto ajustado coincida con el punto observado, se representa por un asterisco encerrado en un círculo. La curva ajustada es finalmente representada por un conjunto de líneas rectas que unen los puntos ajustados. la figura 3.2 muestran una relación ascendente perfecta, donde, de acuerdo a la simbología usada, se aprecia que los puntos ajustados coinciden perfectamente con los puntos observados, esto quiere decir simplemente que, dado cualquier par de disimilaridades $\&_{i(K)j(K)}$, $\&_{i(1)j(1)}$ tales que $\&_{i(K)j(K)} < \&_{i(1)j(1)}$ las correspondientes distancias cumplen $d_{i(K)j(K)} < d_{i(1)j(1)}$ esto es $\&$ y d son de orden global equivalente o están relacionadas monótonamente. Cuando cualquier conjunto de números satisface las desigualdades anteriores, se dice que están relacionados monótonamente a las distancias d_{ij} . Supongamos que tenemos los valores ajustados

\hat{d}_{ij} , definimos

$$S^* = \sum_{i < j} (d_{ij} - \hat{d}_{ij})^2$$

como nuestra medida tentativa de bondad de ajuste (stress no normalizado), que no es otra cosa que la suma de cuadrados residuales asociados a muchas técnicas de ajuste. Esta medida presenta dos particularidades especiales

1) el medir desviaciones en el eje de las distancias

2) la curva ajustada se escoge no de una familia de curvas paramétricas (i.e., polinomios o series trigonométricas), sino de una familia de curvas no paramétricas: la familia de curvas monótonas ascendentes. Esta medida es invariante bajo movimientos rígidos de la configuración (i.e., rotaciones, traslaciones y reflexiones), sin embargo, no es invariante bajo deformaciones uniformes, esto es si afectamos a cada punto de la configuración por el escalar k ($f(x_i) = kx_i$), el stress cambia de S^* a $k^2 S^*$ pues los números d_{ij} , \hat{d}_{ij} se afectan por el mismo factor. Cualquier deformación uniforme (expansión, reflexión o contracción) al no afectar la relación de orden entre las distancias, debería por lo tanto mantener inalterada la medida de bondad de ajuste. Una forma obvia de remediar este defecto en el stress, es dividirlo por un factor de escala, que tenga la misma dependencia cuadrática en la escala que éste. Un factor adecuado es:

$$T^* = \sum_{i < j} d_{ij}^2$$

Por lo tanto

$$\frac{S^*}{T^*} = \frac{\sum_{i < j} (d_{ij} - \hat{d}_{ij})^2}{\sum_{i < j} d_{ij}^2}$$

es una medida que posee las propiedades originales de S^* con las ventajas de ser invariante bajo cambios de escala. Finalmente, es deseable utilizar la raíz cuadrada de esta expresión, lo cual es equivalente a utilizar la desviación standard en lugar de la varianza, con lo que la definición de stress nor-

malizado es :

$$S^2 = \frac{S^*}{T^*} = \frac{\sum_{i < j} (d_{ij} - \hat{d}_{ij})^2}{\sum_{i < j} d_{ij}^2}$$

Esta medida nos representa la bondad con la que una configuración representa los datos. Mientras menor stress, mayor ajuste. Un stress nulo representa un ajuste perfecto dentro de nuestro contexto particular.

La definición de los números \hat{d}_{ij} que fué aplazada hasta este momento, será dada en terminos de la función stress a la que acabamos de llegar. Las cantidades \hat{d}_{ij} son aquellas que minimizan S (y en consecuencia S^*) sujetas a la restricción de monotonía. Utilizando la anterior es posible condensar la definición de stress en la siguiente fórmula:

$$S^2(x_1, \dots, x_n) = \min_D \frac{\sum_{i < j} (d_{ij} - \hat{d}_{ij})^2}{\sum_{i < j} d_{ij}^2}$$

$$D = \{ d_{ij} \mid d_{i(1)j(1)} < \dots < d_{i(M)j(M)} \}$$

Esta minimización es lograda por un algoritmo rápido y eficiente que se describe en el capítulo 6. No está por demás recalcar que, la definición de stress nos provee de medios para la evaluación de cualquier configuración. La configuración que se considera solución al problema del escalamiento multidimensional es obviamente aquella a la cual está asociada el stress mínimo para un valor particular de t (la dimensión del espacio modelo). Teóricamente la obtención de la configuración con mínimo stress, es equivalente a minimizar la función $S(x_1, x_2, \dots, x_n)$ sujeta a la restricción

$$d_{ij} < d_{i \cdot j} \text{ cuando } \&_{ij} < \&_{i \cdot j}.$$

Este problema será abordado en el capítulo siguiente.

EL PROBLEMA DE LA DIMENSION

En la práctica es por lo general desconocido el número de dimensiones a emplear. Es finalmente el criterio científico del investigador en donde reside la determinación del número de coordenadas a recuperar a través de los datos. Kruskal (1964) sugiere dos lineamientos generales para la determinación de t o que suponen cierta ayuda para tal fin. Estos dos criterios son intuitivamente obvios y de sentido común.

El primer criterio sugiere efectuar varios análisis utilizando diferentes valores de t , la construcción de una gráfica que muestre la dependencia del stress mínimo y la dimensión. Un punto importante debe ser tomado en cuenta en este paso. A medida que el valor de t aumenta, el stress asociado disminuye. Para $t > n-1$, el stress asociado es nulo. (Siempre se puede lograr un ajuste perfecto con n puntos en $n - 1$ dimensiones). Es razonable establecer un compromiso entre el valor de t y el stress asociado, consistente en la elección de un valor de t aceptablemente pequeño, para el cual, valores mayores de t no reduzcan en forma significativa el stress. Datos cualitativamente buenos exhiben un "codo" notable en la curva, apuntando el valor apropiado de t . El segundo criterio descansa enteramente en la interpretación de las coordenadas. Si la solución t -dimensional provee una interpretación satisfactoria, pero si la solución $(t+1)$ -dimensional no muestra más una estructura, puede ser entonces conveniente usar solamente la solución t -dimensional.

VIOLACION A LAS HIPOTESIS Y DATOS FALTANTES

El desarrollo anterior se hizo bajo ciertas hipótesis básicas en cuanto a las disimilaridades. Se expone a continuación el procedimiento a seguir en el caso de violación a la simetría e igualdad de disimilaridades.

NO SIMETRIA

Supongamos que el procedimiento de medida no es inherentemente simétrico i.e., $d_{ij} \neq d_{ji}$. Si suponemos que la diferencia observada se debe simplemente a fluctuación estadística, podemos proceder utilizando dos enfoques naturales:

i) Formar medidas simétricas promediando los valores de las disimilaridades iguales $\&_{ij}$, $\&_{ji}$

ii) Generalizar la definición de stress dejando correr las sumatorias tanto en el numerador como en el denominador sobre $i \neq j$ (en lugar de solo $i < j$) En algunos casos pueden tener significado las autodisimilaridades y ser deseable la inclusión del caso $i=j$.

DISIMILARIDADES IGUALES

Los números \hat{d}_{ij} fueron definidos como aquellos que minimizan S sujetos a la restricción que estén monótonamente relacionados a las disimilaridades $\&_{ij}$. El problema de interpretación de esta restricción tiene dos enfoques, a saber primario y secundario.

a) Enfoque Primario

En el caso $\&_{ij} = \&_{kl}$, el valor que tomen las correspondientes distancias d_{ij} , d_{kl} no es considerado relevante, en consecuencia, no afectará la configuración al ocurrir $d_{ij} \neq d_{kl}$, y es natural pensar que el stress no debe reflejar tal situación. La forma de lograr esto, es no restringiendo \hat{d}_{ij} , \hat{d}_{kl} , por lo que los términos pueden ser cero, excepto por otras restricciones. Resumiendo, para el caso de enfoque primario las únicas restricciones sobre \hat{d}_{ij}

$$\text{Cuando } \&_{ij} < \&_{kl} \text{ entonces } \hat{d}_{ij} < \hat{d}_{kl}$$

Nótese que en este enfoque se consideran los rangos de todas las disimilaridades y esto es algunas veces conocido como escalamiento global.

b) Enfoque Secundario

El caso $\&_{ij} = \&_{kl}$ es evidencia de que la d_{ij} debe ser igual a d_{kl} y no afectar la configuración si este no es el caso. El stress por su parte debe reflejar esta situación. La forma de lograr esto es imponiendo la restricción $\hat{d}_{ij} = \hat{d}_{kl}$. Si $d_{ij} \neq d_{kl}$ los términos no pueden ser cero y reflejan directamente nuestro malestar por la desigualdad de d_{ij} y d_{kl} .

Resumiendo, para el caso de enfoque secundario las restricciones sobre las \hat{d}_{ij} son:

Cuando $\&_{ij} < \&_{kl}$ entonces $\hat{d}_{ij} < \hat{d}_{kl}$

Cuando $\&_{ij} = \&_{kl}$ entonces $\hat{d}_{ij} = \hat{d}_{kl}$

Este enfoque es demasiado restrictivo y conduce frecuentemente a problemas de convergencia.

La implantación para el uso opcional de ambos enfoques es muy simple.

La parte en la que la diferencia en estos dos enfoques tiene lugar es dentro del algoritmo para el ajuste de las \hat{d}_{ij} .

DATOS FALTANTES

En algunos casos no todas las disimilaridades son susceptibles de ser obtenidas por causas de diversa índole (diseño, error, etc.) Para valores grandes de n esto puede ser bastante común. Ante esta situación se hace necesario replantear la definición original de stress, generalizándola de tal manera que cubra esta contingencia. La forma natural que surge para la solución de este problema, es la simple omisión tanto en el numerador S^* , como en el denominador T^* , de los términos correspondientes a las disimilaridades faltantes.

Si no existen suficientes disimilaridades el método expuesto fallará. - Esto es, habrá una configuración con stress cero que no tendrá ninguna relación real con los datos. Es importante tener presente que la cuestión del número considerado de disimilaridades, no es determinante en la obtención de resultados con significado real, sino depende de que disimilaridades se consideren.

En muchos casos de importancia práctica, la consideración de un medio o cuarta parte, o aun menos, del total de disimilaridades son más que suficientes, si están distribuidas propiamente en la matriz de disimilaridades.

DISTANCIA EUCLIDEANA

Otra posible generalización en el planteamiento original, es como se -

mencionó, la utilización de funciones de distancia más generales que la euclídeana. En principio, no parece existir razón por la que la definición de stress, no pueda ser usada con cualquier distancia. La única restricción al parecer, es la dificultad de implantación en un programa de computadora.

EL PROCEDIMIENTO DE SHEPARD

La técnica expuesta, tiene su origen en los intentos de Kruskal por racionalizar y perfeccionar el procedimiento iterativo desarrollado por Shepard. La comparación de estos dos procedimientos es un punto interesante que puede ser descrito dentro del contexto de esta exposición.

Sea d_{ij} la m -ésima distancia en orden creciente, definimos \hat{d}_{ij} como la m -ésima disimilaridad. En lugar de hacer la influencia de x_i sobre x_j proporcional a $d_{ij} - \hat{d}_{ij}$, Shepard la hace proporcional a $\hat{d}_{ij} - \hat{d}_{ij}$. No parece posible describir bajo estas circunstancias su procedimiento como aquel que minimice alguna medida particular de no monotonía.

ROTACION DE PROCRUSTES

El problema de la evaluación de una configuración obtenida por alguno de los dos métodos expuestos, ha quedado resuelto por las medidas propuestas por Mardia de proporción de matriz de distancia explicada para el escalamiento clásico y por el stress de Kruskal para el escalamiento ordinal.

El carácter individual de estas medidas dado principalmente por el enfoque diferente necesario para la solución del mismo problema siguiendo un distinto esquema inicial de premisas o aun bajo el mismo método, pero utilizando alguna variación permitida dentro del mismo, hace evidente la conveniencia de establecer algún método de evaluación para la bondad de ajuste al comparar dos configuraciones.

De contarse con alguna forma de evaluar dos configuraciones, el método en cuestión no estará limitado para su uso en el escalamiento, sino se aplicaría en cualesquiera métodos de naturaleza semejante.

(Green 1952; Gower 1971) han desarrollado una medida de bondad de ajuste para comparar dos configuraciones. Consideremos a la matriz $X_{(n \times p)}$ de coordenadas de n puntos obtenida de la matriz D por una técnica.

Supongamos que $Y_{(n \times q)}$ es la matriz de coordenadas de otro conjunto de puntos obtenidos por otra técnica, o utilizando otra medida de distancia. Podemos considerar que las matrices son del mismo orden, al añadir columnas de ceros a la matriz de orden menor.

La medida de bondad de ajuste adoptada es obtenida al mover los puntos y_r con respecto a los puntos x_r hasta que la suma de cuadrados residuales

$$\sum_{r=1}^n (x_r - y_r)' (x_r - y_r)$$

sea mínimo. Podemos mover y_r con respecto a x_r a través de rotación, reflexión y traslación, i.e., por,

$$A'y_r + b \quad r=1, \dots, n$$

donde $A_{(p \times p)}$ es una matriz ortogonal, por lo que necesitamos resolver

$$R^2 = \min_{A,b} \sum_{r=1}^n (x_r - A'y_r - b)'(x_r - A'y_r - b)$$

para A y b.

Es posible demostrar que los valores de A y b que minimizan la medida están dados por

$$b = \bar{X} - A'\bar{Y}, \quad A = VU'$$

donde las matrices V, U' provienen de la descomposición en valores singulares de la matriz $Z = Y'X = UV'$

La obtención de estos valores particulares, vienen de resolver para A y b la expresión equivalente, (suponiendo que \bar{X} , \bar{Y} son ceros)

$$R^2 = \min \text{tr} (X - A'Y)'(X - A'Y) = \text{tr}XX' + \text{tr}YY' - \max \text{tr}X'YA$$

Hemos supuesto que los vectores medias de X, Y son ceros. Entonces la mejor rotación de Y con respecto a X es $\hat{Y}A$, donde \hat{A} está definida como anteriormente, y \hat{A} es llamada la Rotación de Procrustes de Y con respecto a X.

FACTOR DE ESCALA

Si las escalas de las dos configuraciones son diferentes, entonces la transformación debe ser de la forma

$$cA'y_r + b \quad \text{con } c > 0.$$

Puede demostrarse que

$$\hat{c} = \frac{\text{tr} \Gamma}{\text{tr}YY'} \quad (1)$$

y los otros estimadores permanecen igual.

Esta transformación se llama rotación de Procrustes con escalamiento de Y con respecto a X. La nueva mínima suma de cuadrados residuales está dada por

$$R^2 = \text{tr}(XX') + \hat{c}^2 \text{tr}(YY') - 2 \hat{c} \text{tr}(\Gamma)$$

donde c está dada por (1). Este procedimiento no es simétrico con respecto a X Y . La simetría puede ser obtenida al seleccionar el escalamiento de tal forma que

$$\text{tr}(XX') = \text{tr}(YY').$$

Escalamiento Clásico

Sea D una matriz de distancia que represente aproximadamente las distancias entre puntos de una configuración en un espacio euclideo de dimensión baja, teniendo en cuenta que D no necesariamente debe ser euclidea. Como sabemos la configuración asociada a esta matriz por el método clásico se obtiene al tomar los primeros k vectores propios de la matriz B . Si los primeros k valores propios de B son "grandes" y los restantes son cercanos a cero, entonces las distancias entre los puntos de la configuración se aproximarán cercanamente a las entradas de la matriz D .

El método numérico utilizado para la obtención de la configuración asociada a la matriz D es bastante simple, la única dificultad es de hecho, el conseguir una buena rutina para encontrar los valores propios y los correspondientes vectores propios de una matriz simétrica. Los cálculos involucrados en el método numérico pueden considerarse comprendidos en 4 etapas.

a) La construcción a partir de la matriz de distancia D de la matriz A - cuyo elemento genérico es $(-\frac{1}{2} d_{rs}^2)$.

En caso de presentarse una matriz de similitudes, será necesario transformar las similitudes a distancias. Sea C la matriz de similitudes, puede demostrarse que la transformación de similitudes a distancia está dada - por $d_{rs} = (c_{rr} - 2c_{rs} + c_{ss})^{1/2}$

b) La obtención de la matriz B con elementos $b_{rs} = a_{rs} - a_{r.} - a_{.s} + a_{..}$

c) La obtención de los k primeros valores propios de la matriz B $\lambda_1 > \lambda_2 > \dots > \lambda_k$, con correspondientes vectores propios $X = (x_{(1)}, \dots, x_{(k)})$

d) La normalización de los vectores $x_{(i)}$, en forma tal que $x'_{(i)} x_{(i)} = 1$ (suponiendo que los primeros k valores propios son positivos)

e) Los puntos requeridos son de la forma $x_r = (x_{r1}, \dots, x_{rk})'$, $r = 1, \dots, k$, las filas de X .

ESCALAMIENTO ORDINAL

El metodo numérico expuesto en esta sección tiene como finalidad, la obtención de la configuración de mejor ajuste para un valor fijo de t (dimensión del espacio de configuración), esto es, la configuración que minimice el stress

$$S^2 = \frac{S^*}{T^*} = \frac{\sum_{i < j} (d_{ij} - \hat{d}_{ij})^2}{\sum_{i < j} d_{ij}^2}$$

donde el mínimo es tomado sobre d_{ij}^{mon} tales que $d_{ij}^{mon} \sim \hat{d}_{ij}$ donde \sim significa relacionado monótonamente con. Un problema secundario de interés independiente, es la obtención de \hat{d}_{ij} a partir de los valores fijos de d_{ij} , este es el problema computacional de regresión monótona.

ENTRADAS FALTANTES

Este método es fácilmente adaptable en la situación en la que existen valores faltantes. Simplemente se omiten las disimilaridades en el orden

$$\hat{d}_{i(1)j(1)} < \dots < \hat{d}_{i(M)j(M)}$$

y se suprimen los términos correspondientes del numerador y denominador. Mientras no haya demasiados valores faltantes, el método parece todavía trabajar bien.

DISTANCIA

La posibilidad de utilización de una función de distancia más general que la euclídeana se incluye en este método. Específicamente, las técnicas numéricas y las formulas dadas cubren una clase de funciones introducidas en el capítulo 1, conocidas como métricas L_p o L_p , ocasionalmente llamadas métricas r de Minkowsky.

METODO DE GRADIENTES

Entramos ahora a la descripción del método computacional de obtención de la configuración que minimice el stress para valores fijos de r y t , es decir, para una métrica particular, en un espacio t -dimensional.

Hasta aquí se ha descrito una configuración como n puntos en un espacio t -dimensional; es conveniente, e igualmente válido, considerar tal configuración como un solo punto en un espacio nt -dimensional, i.e.,

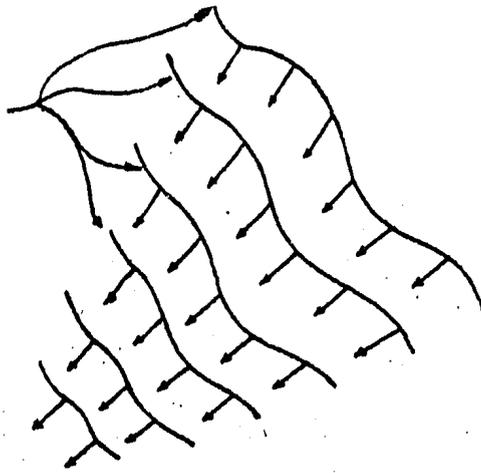
$$(x_{11}, \dots, x_{1t}, \dots, x_{n1}, \dots, x_{nt})$$

Nos referimos al espacio nt -dimensional como el espacio de configuración y al espacio t -dimensional como espacio modelo.

Supongamos que están dados los valores ϵ_{ij} . Para cada configuración existe un valor de S asociado. En forma equivalente $S : R^{nt} \rightarrow R$ es una función definida en cada punto del espacio de configuración. El problema a resolver es la obtención del punto que minimice a S . Nos enfrentamos con un problema común en análisis numérico, minimizar una función de varias variables. El método adoptado aquí es el método de gradientes.

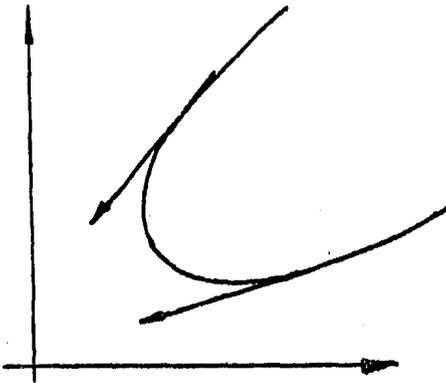
Este método inicialmente considera un punto arbitrario en el espacio de configuración. El paso siguiente consiste en intentar mejorar un poco la configuración moviéndola ligeramente. El procedimiento adoptado por el método para lograr este fin, es seguir la dirección en el espacio de configuración en la cual S decrece más rápido. Tal dirección es proporcionada por el gradiente negativo de la función S (ver apéndice C). El gradiente, por supuesto, cambia de dirección de punto en punto, pero si pudiésemos seguir la dirección del gradiente negativo local, seríamos guiados al mínimo local. Ya que todas las direcciones cercanas conducen al mismo local, no es necesario seguir el gradiente negativo exactamente, podemos ser muy descuidados en este aspecto y todavía esperar el mínimo local.

Curvas de nivel



Gradientes.

El método de gradiente simple tiene la desventaja de que cerca del mínimo, el gradiente negativo apunta solo débilmente (debido al error de redondeo) al mínimo.



Sin embargo esta debilidad no debe ser muy exagerada; si todas las tangentes son pequeñas, hay probablemente poca ganancia al refinar aun más la estimación de la posición del mínimo. El método en su forma más sencilla consiste en tomar

$$x_j^{(i+1)} := x_j^{(i)} - h \frac{\partial S}{\partial x_j}$$

el superíndice indica el número de iteración , y h el tamaño del paso.

El primer problema es adivinar un paso razonable h (tal punto es abordado

más adelante) , teniendo en cuenta que no necesitamos ser exactos, pues todas las trayectorias cercanas conducen al mismo mínimo local.

Un punto importante en la discusión de cualquier metodo iterativo, ya sea de minimización o de cualquier otra índole, es el criterio de terminación del proceso. Las respuestas existentes en lo referente a este método en particular, no son satisfactorias. Existen varios criterios, entre los más populares están:

Sea $f(x)$ la función a minimizar, para cualquier real positivo ϵ tenemos

$$\begin{array}{l}
 \text{i) } \sum_{j=1}^n \left| \frac{\partial f}{\partial x_j} \right| < \epsilon \\
 \text{ii) } \sum_{j=1}^n \left| f^{(i+1)}(x)_j - f^{(i)}(x)_j \right| < \epsilon
 \end{array}$$

Como en el caso de medidas de similitud, el criterio de terminación de un proceso debe idealmente ser diseñado en forma específica para cada problema. Kruskal sugiere una forma modificada del criterio del cuadrado de la norma del gradiente i). El criterio en cuestión relaciona la norma del gradiente y la norma de la configuración normalizada, es decir, lo que Kruskal llama la magnitud relativa del gradiente.

Lo anterior es compatible con los criterios de obtención de mínimos de funciones vectoriales de valores reales dados en los textos standard de Cálculo Vectorial. Específicamente, se afirma que dado un subconjunto de un espacio euclideo U , y una función $f : U \rightarrow R$ y $x \in R^n$, es un extremo local si, y solo si, $\text{grad}(x_0) = 0$. La demostración de este hecho y la justificación acerca de la dirección en la cual decrece con mayor rapidez una función de valores reales, es dada en los apéndices.

LA DIFICULTAD DE MINIMOS LOCALES

Una dificultad general que padecen los métodos numéricos de minimización - incluyendo por supuesto del que nos ocupamos, es el de arribar a un mínimo local

en lugar de un mínimo global. En el problema de minimización que tratamos, esta dificultad no llega a alcanzar dimensiones de consideración, por las razones que se exponen a continuación.

i) Es posible iniciar el método de gradientes con configuraciones iniciales diferentes, las cuales pueden conducir a diferentes mínimos locales. Mientras sólo el menor de estos puede ser el mínimo verdadero, podríamos preguntarnos por mínimos más pequeños. Generalmente la mayoría de las configuraciones conducen al mismo mínimo local, el cual es mucho menor que los restantes que encontremos. Parece innecesario preocuparse cuando esto sucede.

ii) La configuración de mínimo local no es en sí el producto final del análisis, la cual deba ser aceptada ciegamente. En la mayoría de los casos la configuración de mínimo local es de interés solo si tiene algún sentido o pueda ser interpretada. Si la configuración posee una de estas propiedades, no es muy probable que sea un error considerarla.

iii) A menos que el stress de la supuesta configuración con mínimo local sea lo suficientemente bajo, esta configuración tendrá algún valor. Kruskal sugiere la siguiente clasificación de configuraciones con respecto a su stress asociado.

stress	bondad de ajuste
20 %	pobre
10 %	aceptable
5 %	buena
2 %	excelente
0 %	perfecta

A la fecha esta consideración acerca de la calidad de la configuración se considera demasiado simple ya que el stress depende de n y p .

iv) Existe la posibilidad de hacer varias pruebas de la configuración y los datos, así como de análisis separados de subconjuntos de datos.

DESCRIPCION DE LA IMPLANTACION

La técnica empleada para la minimización del stress es de carácter itera

tivo. En terminos generales consiste en empezar con una configuración arbitra-
ria , calcular el gradiente negativo, mover la configuración en esta direc-
ción una distancia conveniente y repetir estos dos pasos un número suficiente
de veces. Lo anterior se puede lograr dividiendo la tarea en siete subtareas
principales:

- 1) Creación de una configuración inicial
- 2) Normalización de la configuración
- 3) Cálculo de las distancias entre los puntos de la configuración
- 4) Ajuste de una regresión monótona creciente a las distancias
- 5) Cálculo del gradiente de la función stress
- 6) Cálculo de la magnitud en que se moverá la configuración
- 7) Creación de una nueva configuración

La descripción detallada de cada subtask se hará a continuación

1) La configuración inicial que se construya debe reunir solo dos cuali-
dades, siendo arbitraria de otra manera.

a) La configuración no puede estar en un espacio de dimensión menor al -
que se haya escogido para el análisis.

Una buena elección para la configuración inicial es aquella que resulta
del escalamiento clásico en la dimensión requerida, si no hay configuración -
disponible conveniente , una configuración arbitraria debe ser generada. Una
forma satisfactoria de conseguirlo es tomar los n primeros puntos de la forma

$$\begin{aligned} &(1,0,\dots,0), \\ &(0,1,\dots,0), \\ &\cdot \\ &\cdot \\ &\cdot \\ &\cdot \\ &(0,0,\dots,1), \\ &(2,0,\dots,0), \\ &(0,2,\dots,0), \text{etc.} \end{aligned}$$

Otra solución igualmente satisfactoria es la generación de los puntos en
la configuración con la ayuda de un generador de números aleatorios.

2) La normalización de la configuración se logra en dos pasos.

a) Trasladando el centro de gravedad de la configuración al origen.

b) Haciendo que la raíz cuadrada del promedio de las distancias de los puntos al origen sea unitaria.

La normalización de acuerdo a lo anterior implica la transformación de los puntos originales en puntos de la forma

$$x = \frac{x - \bar{x}}{fn}$$

donde

$$fn = \frac{1}{n} \sum_{i,s} x_{is}^2$$

3) El cálculo de la distancia entre los puntos en la configuración se hará utilizando la métrica de Minkowsky para un valor dado de r , el parámetro de esta métrica, la fórmula a utilizar, como es bien sabido es la siguiente:

4) Ajuste de una regresión monótona creciente a las distancias. Supongamos que las disimilaridades $d_{i(k)j(k)}$ presentan las siguientes relaciones

$$d_{i(1)j(1)} < \dots < d_{i(M)j(M)}$$

Los valores correctos de \hat{d}_{ij} pueden ser descritos de la siguiente manera:

Existe una partición de las disimilaridades en bloques consecutivos $b_i, i=1, \dots, u$ tal que, dentro de cada bloque el valor de \hat{d}_{ij} es constante, este valor común \hat{d}_{bk} es el promedio de las d_{ij} en ese bloque. Es por lo tanto,

suficiente encontrar la partición correcta con el fin de encontrar los -- números \hat{d}_{ij} . En un principio consideramos la partición en bloques más fina -- posible, uniendo según sea preciso, paso a paso los bloques apropiados hasta encontrar la partición correcta. La partición más fina posible consiste -- obviamente, de M bloques cada uno conteniendo una sola disimilaridad.

Supongamos que tenemos cualquier partición en bloques consecutivos. -- Denotemos \hat{d}_b como el promedio de las d_{ij} en el bloque b. Sean $b-1, b, b+1$ tres bloques consecutivos en orden creciente.

Definición

Se dice que el bloque b es satisfecho superiormente (up-satisfied) si -- $\hat{d}_b < \hat{d}_{b+1}$ o si es el último elemento en orden creciente de la partición.

Definición

Se dice que el bloque b es satisfecho inferiormente (down-satisfied) si -- $\hat{d}_b < \hat{d}_{b-1}$ o si es el primer bloque en orden creciente en partición.

En cada paso del algoritmo tenemos una partición en bloques, de los cua -- les solo uno se considera activo. Este bloque puede ser ya sea activo supe -- rior (up-active) o activo inferior (down-active). Al principio el primer blo -- que en la partición en orden ascendente, consistente solo de $d_{i(1)j(1)}$, es ac -- tivo superior. El algoritmo procede posteriormente de la manera siguiente:

Si el bloque activo, es activo superior, verifica si es satisfecho supe -- riormente. En caso afirmativo la partición en curso permanece inalterada; en -- caso contrario el bloque activo es unido al siguiente bloque superior, alte -- rando la partición. En ambos casos el bloque resultante se considera ahora ac -- tivo inferior.

Si el bloque activo, es activo inferior, verifica si es satisfecho infe -- riormente. En caso afirmativo la partición en curso permanece inalterada; en -- caso contrario, el bloque activo se une al siguiente bloque inferior, alteran -- do la partición. En ambos casos, el bloque resultante se considera ahora acti --

vo superior.

Esta alternación de estados, resulta eventualmente en un bloque activo el cual es simultáneamente satisfecho superior e inferiormente. Cuando esto ocurre, ninguna unión adicional puede hacerse por este procedimiento. El procedimiento se repite hasta que el último bloque en la partición en orden creciente es simultáneamente satisfecho superior e inferiormente. El caso de disimilaridades iguales puede ser manejado enteramente por el algoritmo, el cual sólo debe ser modificado por un preproceso. En caso de haberse adoptado el enfoque primario, el preproceso consiste en arreglar las disimilaridades de tal manera que las distancias formen una sucesión creciente en cada bloque donde aparezcan disimilaridades iguales. En caso de que se haya adoptado el enfoque secundario, es decir, si restringimos a los números \hat{d}_{ij} a que sean iguales cuando los correspondientes d_{ij} son iguales, el preproceso consiste en formar bloques de tal manera que el bloque que contenga a d_{ij} contenga además a todas las disimilaridades que igualen a esta. De ésta forma el algoritmo no empezará con la partición más fina sino con la partición formada por bloques con disimilaridades iguales.

4) Procederemos a encontrar la forma explícita del gradiente de la función stress, para el caso particular de la función de distancia de Minkowsky.

$$S^2 = \frac{S^*}{T^*} = \frac{\sum_{i < j} (d_{ij} - \hat{d}_{ij})^2}{\sum_{i < j} d_{ij}^2}$$

el stress asociado a la configuración.

$$S : R^{nt} \rightarrow R$$

La componente k_l del gradiente de S es:

$$\frac{\partial S}{\partial x_{kl}} = S \sum_{i,j} (d_{ij}^{kl} - \hat{d}_{ij}^{kl}) \left(\frac{d_{ij} - \hat{d}_{ij}}{S^*} - \frac{d_{ij}}{T^*} \right) \frac{|x_{i1} - x_{j1}|^{r-1}}{d^{r-1}} \text{signum}(x_{i1} - x_{j1})$$

donde δ^{ki} es la delta de Kronecker

$$\delta^{ki} = 1,0 \text{ si } k=i, k \neq i \text{ respectivamente}$$

La derivación completa puede ser encontrada en el apéndice A

5) La magnitud h ("tamaño de paso" en que se moverá la configuración a lo largo del gradiente) se varía en cada iteración. La magnitud h escogida en cada iteración, no afecta la solución obtenida al final. Lo que si es -- afectado, es el número de iteraciones requeridas para llegar a la solución, lo que en efecto es, un factor computacional importante.

El procedimiento para el calculo de h , provee pasos grandes durante las primeras iteraciones, y, pasos pequeños al final. El valor inicial de h para una configuración inicial arbitraria debe ser aproximadamente 0.02. Despues, el valor h es determinado por la formula:

$$h := h * f_a * f_r * f_b \quad \text{donde}$$

$$f_a = \frac{\cos^3 \theta}{4}$$

θ = ángulo entre el gradiente actual y el gradiente anterior

$$f_r = \frac{1.3}{1 + c^{\frac{5}{5}}}$$

$$c_5 = \min \left\langle 1, \frac{S}{S_5} \right\rangle$$

$$f_b = \min \left\langle 1, \frac{S}{S} \right\rangle$$

S_5 = Stress de la configuración obtenida hace cinco iteraciones

S = Stress de la configuración obtenida en la iteración anterior

Si no han sido calculadas todavía 5 iteraciones, S_5 puede ser tomado como el primer stress calculado. Artificios similares pueden ser utilizados en relación a S' y θ en la primera iteración. Recordando el significado geo

métrico del gradiente de una función de valores reales f como el vector

$$\text{grad } f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right)$$

El ángulo entre $\text{grad } S$ y $\text{grad } S$ está dado por :

$$\cos \theta = \frac{\langle \text{grad } S, \text{grad } S \rangle}{|S| |S|}$$

6) Supongamos que hemos llegado a una configuración X en R^S , entonces — formamos la nueva configuración siguiendo la dirección del vector $X + t \text{ grad}(S)$ donde

$$t = - \frac{h}{\text{mag}(\text{grad}(S))}$$

i.e., la componente x_{is} de la nueva configuración es la dada por

$$x_{is} := x_{is} + \frac{h}{\text{mag}(\text{grad}(S))} \frac{\partial S}{\partial x_{is}}$$

Si suponemos que X está normalizada

$$\text{mag}^2(\text{grad } S) = \frac{1}{n} \sum_{is} \left\langle \frac{\partial S}{\partial x_{is}}, \frac{\partial S}{\partial x_{is}} \right\rangle \quad 2$$

La ejecución de los procedimientos descritos en los párrafos anteriores, constituye el aspecto computacional central de la implantación del método descrito en el capítulo anterior. Mientras este proceso se desarrolla se logran sucesivamente, valores más pequeños de la función stress. Eventualmente el valor del stress desciende en forma significativa, de tal forma que, iteraciones adicionales resultan en una mejora nula o insignificante. Es conveniente contar con criterios adicionales para la terminación del proceso, como puede ser el número de iteraciones, etc. Para una configuración arbitraria, la condición de mínimo local sugerida es aproximadamente del 2% (i.e., $\text{mag}(\text{grad}(S)) = 0.02$). Este criterio dependerá en algunos casos de las condiciones específicas de los datos. Para datos con gran variación estadística (i.e., $n=40, t=2, o, n=15, t=5$) — un valor mayor es apropiado. Con más precisión lo anterior significa: un valor grande del criterio podría representar tanto como $e=0.05$ o posiblemente tan alto como $e=0.10$. Un valor bajo podría ser $e=0.005$ o cualquier valor menor, incluyen

do por supuesto cero.

La complicada naturaleza de la rutina iterativa arroja como resultado, - que la cantidad de cómputo involucrado en la obtención de una configuración - por el método ordinal, es mucho mayor que la necesaria utilizando el método - clásico. Para un conjunto típico de datos, el escalamiento ordinal podría tomar de 10 a 15 veces más tiempo que el escalamiento clásico para la obtención de una configuración asociada a una matriz dada.

ROTACION DE PROCRUSTES

Abordamos ahora la descripción del método numérico utilizado en la técnica Procrustes.

Podemos considerar que la única dificultad en la implementación de esta técnica, reside en la descomposición de la matriz $Z=Y'X$ en la forma $Z=U\Gamma V'$ (descomposición de valores singulares). Un excelente algoritmo puede encontrarse en Wilkinson (1970) bajo el nombre de SVD:

Para propósitos computacionales sumaremos los cálculos involucrados.

Sean X, Y dos configuraciones

a) A partir de X, Y construir la matriz $Z=Y'X$.

b) Encontrar las matrices U, Γ, V tales que $Z=U\Gamma V'$ donde U es una matriz con columnas ortonormalizadas, V es una matriz ortogonal y Γ es una matriz diagonal de valores positivos.

c) Calcular el factor de escala c

$$c = \frac{\text{tr}(\Gamma)}{\text{tr}(YY')}$$

d) Calcular la medida de bondad de ajuste

$$R^2 = c^2 \text{tr}(XX') + \text{ctr}(YY') - 2\text{ctr}(\Gamma)$$

e) Encontrar la matriz de Rotación A

$$A = VU'$$

f) Encontrar el vector de rotación b

$$b = \bar{X} - A'\bar{Y}$$

g) Rotar y trasladar la configuración Y i.e

$$y_r^* = cA'y_r + b$$

CAPITULO 6

Es el propósito de esta sección el mostrar los reportes impresos de los subprogramas que conforman MDS, las tablas que componen la subsección siguiente fueron en su totalidad generadas por MDS, y servirán de base en la discusión del ejemplo numérico que ilustrará el empleo práctico de los métodos descritos en los capítulos anteriores.

Ejemplo Numérico

	1	2	3	4	5	6	7	8	9	10	11	12
1	0											
2	244	0										
3	218	350	0									
4	284	77	369	0								
5	197	167	347	242	0							
6	312	444	94	463	441	0						
7	215	221	150	236	279	245	0					
8	469	583	251	598	598	169	380	0				
9	166	242	116	257	269	210	55	349	0			
10	212	53	298	72	170	392	168	531	190	0		
11	253	325	57	340	359	143	117	264	91	273	0	
12	270	168	284	164	277	378	143	514	173	111	256	0

La tabla se utilizó como matriz de prueba para los subprogramas Escalamiento Clásico y Escalamiento Ordinal. Los datos utilizados corresponden a las distancias por carretera entre doce poblaciones de la Gran Bretaña.

Escalamiento Clásico

* * * * ESCALAMIENTO CLASICO * * * *

No	VALOR PROPIO	PROPORCION DE MATRIZ DE DISTANCIA EXPLICADA	PROPORCION DE MATRIZ DE DISTANCIA EXPLICADA
1	394472.79	79.73	97.25
2	63634.37	92.59	99.78
3	13544.07	95.33	99.90
4	10245.31	97.40	99.96
5	2466.68	97.90	99.97
6	1450.27	98.19	99.97
7	500.89	98.29	99.97
8	-0.00	98.29	99.97
9	-16.94	98.30	99.97
10	-214.08	98.34	99.97
11	-1141.22	98.57	99.97
12	-7063.42	100.00	100.00

La tabla muestra los valores propios asociados a la matriz B así como las medidas de bondad de ajuste sugeridas por Mardia para la proporción de matriz de distancia explicada por una configuración obtenida por el método de escalamiento clásico, la expresión numérica correspondiente a cada medida es respectivamente

$$w_{1,s} = \frac{\sum_{r=1}^s |1_r|}{n}$$

$$w_{2,s} = \frac{\sum_{r=1}^s 1_r^2}{n}$$

el empleo de cada medida depende del carácter de la matriz de datos considerada, recordando que la segunda medida es más conveniente en el caso de que la matriz de distancia D no sea necesariamente Euclídeana. Un examen a la tabla revela enseguida el carácter bidimensional de la configuración

obtenida. Ambas medidas reportan una proporción superior al 90% y en el caso concreto de la segunda medida, la configuración en dos dimensiones explica prácticamente la totalidad de la matriz de distancia.

La configuración obtenida utilizando esta técnica se muestra en la tabla siguiente

* * * * ESCALAMIENTO CLASICO * * * *

CONFIGURACION DE PUNTOS, SOLUCION CLASICA

1	2
0.249325	1.921810
1.117395	-0.242210
-0.762621	0.424228
1.168113	-1.047268
1.042611	1.925208
-1.290233	0.426846
-0.042854	-0.691027
-2.104742	-0.353027
-0.174756	-0.075266
0.842838	-0.369057
-0.661968	-0.462649
0.616892	-1.457587

Escalamiento Ordinal

* * * * ESCALAMIENTO ORDINAL * * * *

ESCALAMIENTO ORDINAL,
CORRIDO BAJO LAS SIGUIENTES OPCIONES

MATRIZ DE DATOS SIMETRICA
AJUSTE BAJO ENFOQUE PRIMARIO
CONDICION DE MINIMO LOCAL = 0.000100
MAXIMO DE ITERACIONES = 100

DIMENSION REQUERIDA	=	2
METRICA DE MINKOWSKY	=	=2.000000

PARAMETROS ASOCIADOS A LA CONFIGURACION OBTENIDA		
ITERACION EN CURSO	=	101
MAGNITUD DE GRADIENTE	=	=0.000126
STRESS	=	=0.010643

El metodo de escalamiento ordinal fué ahora utilizado para la obtención de una configuración asociada a la matriz de datos de la tabla. La tabla nos muestra algunos datos correspondientes a la configuración obtenida impresos por el subprograma ESCALAMIENTO ORDINAL. En específico se informa el tipo de matriz de datos analizada, el criterio de ajuste en caso de medidas experimentales con el mismo valor numérico, las condiciones que de cumplirse resultarán en la terminación del proceso iterativo, i.e., la magnitud relativa del gradiente de la función stress especificado como condición de mínimo local y el número máximo de iteraciones permitidas para la búsqueda de la configuración con stress mínimo. A continuación se muestran las condiciones que resultaron al término del proceso, i.e., la iteración en curso y la magnitud relativa del gradiente de la función stress.

De acuerdo con el reporte mostrado, el método no convergió, pues la terminación del proceso fué causada por excederse en el número máximo de iteraciones permitido. Esto se debió al criterio tan restringido pedido como condición de mínimo local (0.000100). Por otra parte, el stress asociado a la configuración obtenida por este metodo nos lleva a clasificarla según el criterio de Kruskal (considerado hoy en día demasiado simple) como excelente (stress = 1.06 %). La siguiente tabla muestra la configuración de puntos a la que se llegó por el metodo que tratamos

* * * * ESCALAMIENTO ORDINAL * * * *

* CONFIGURACION DE PUNTOS, SOLUCION ORDINAL *

0.723483	0.336022
-0.270878	1.035944
0.275689	-0.650876
-0.630301	0.964504
0.477298	1.145919
0.432855	-1.123869
-0.221228	-0.107837
0.149550	-1.907132
0.062437	-0.105478

-0.285246	0.734465
0.001730	-0.650266
-0.715389	0.328602

En cierto sentido esto podría parecer una muestra inútil del uso de las técnicas, pues se conoce la configuración real que dió origen a la matriz - analizada, sin embargo, servirá para mostrar algunos aspectos de interés - inherentes a los métodos de escalamiento.

Las observaciones hechas sobre las configuraciones obtenidas ilustran el hecho evidente acerca de la carencia completa de significado particular de los ejes resultantes, por lo que es perfectamente factible la búsqueda - de otros que conduzcan a una interpretación obvia. El problema de encontrar dimensiones con significado es uno de los propósitos principales del escala - miento, Kruskal y Wish (1978) sugieren que una regresión lineal múltiple - entre alguna dirección en la configuración y varias características de los puntos, puede ser útil al tratar de interpretar las dimensiones.

Otra forma en la que la gráfica de una configuración puede ser usada - es para localizar grupos o conglomerados de puntos que estén cercanos. Si - se encuentra una configuración en una dimensión mayor a dos, los problemas de interpretación son considerablemente más complejos. Puede ser de ayuda - graficar las dimensiones dos a la vez, una regresión lineal múltiple puede ser también de utilidad.

COMPARACION DE CONFIGURACIONES

Las configuraciones obtenidas por cada uno de los metodos tratados fue - ron comparadas por el subprograma COMPARACION DE CONFIGURACIONES utilizando la técnica de Rotación de Procrustes. En este ejemplo en particular al con - tarse con la configuración real, fué posible la comparación de cada configu - ración generada con dicha configuración.

El subprograma COMPARACION DE CONFIGURACIONES reportó en cada caso la suma de cuadrados residuales.

$$\sum_{i=1}^n (x_i - (Ay_i + b))^2 (x_i - (Ay_i + b))^2$$

donde x_1 es un punto de la configuración con respecto a la cual se compara la configuración Y de la cual y_1 forma parte.

Asimismo se imprimió el factor de escala correspondiente, la matriz de Rotación A y el vector de traslación b que minimizan la suma de cuadrados residuales. Por último el subprograma en base al factor de escala, la matriz de rotación y el vector de traslación calculados, escala rota y -- traslada la configuración que se ha comparado. Los reportes impresos por este subprograma se muestra a continuación en forma condensada.

* * * * ROTACION DE PROCRUSTES * * * *

COMPARACION DE CONFIGURACION CLASICA CON CONFIGURACION VERDADERA

* SUMA DE CUADRADOS RESIDUALES = 4.0381E+06 *
 * FACTOR DE ESCALA = 1.7557E+02 *

*MATRIZ DE ROTACION *

1	2
0.32	-0.95
-0.95	-0.32

*VECTOR DE TRASLACION *

1
 424.25
 383.42

* * * * ROTACION DE PROCRUSTES * * * *

COMPARACION DE CONFIGURACION ORDINAL CON CONFIGURACION VERDADERA

* SUMA DE CUADRADOS RESIDUALES = 3.9357E+06 *
 * FACTOR DE ESCALA = 2.6492E+02 *

* MATRIZ DE ROTACION *

1	2
-0.99	0.13
-0.13	-0.99

*VECTOR DE TRASLACION *

1
 424.25
 383.42

Un examen a la suma de cuadrados residuales correspondiente a cada comparación en este ejemplo en particular muestra una superioridad del método ordinal sobre el clásico, aún cuando tal superioridad es casi insignificante pues al hacer el cociente de los cuadrados residuales este resultado ser casi 1 (0.9746).

La última tabla muestra por una parte a la configuración real y por otra a las configuraciones obtenidas por el método clásico y el método ordinal en ese orden. El subprograma COMPARACION DE CONFIGURACIONES imprime opcionalmente la configuración que ha sido comparada escalada, rotada y trasladada con respecto a la otra configuración.

* * * * R O T A C I O N D E P R O C R U S T E S * * * *

257.00	279.00	118.90	233.36	222.36	321.07
529.00	104.00	527.66	211.37	458.33	101.80
339.00	554.00	310.63	486.20	375.14	564.12
629.00	142.00	664.35	248.44	555.24	107.71
292.00	90.00	163.17	101.31	258.01	99.67
259.00	665.00	280.38	573.76	350.79	693.90
508.00	433.00	536.69	429.59	486.18	403.82
265.00	842.00	363.98	753.22	453.15	889.39
433.00	438.00	426.28	416.72	411.63	413.34
533.00	183.00	533.23	264.18	472.88	180.42
420.00	563.00	463.74	519.60	447.03	554.17
627.00	308.00	701.40	363.26	600.30	271.59

Comparación de Configuraciones

Compara dos configuraciones por la técnica Rotación de Procrustes desarrollada por Gower.

Escalamiento Clásico

Obtiene la configuración asociada a una matriz de datos por el método de escalamiento clásico debido a Torgerson y Gower.

Escalamiento Ordinal

Obtiene la configuración asociada a una matriz de datos por el método de escalamiento ordinal desarrollado por Kruskal.

Ejecución

Transmitir RUN MDS en respuesta al prompt de CANDE. El programa mostrará una lista de subprogramas en la terminal. El usuario puede seleccionar uno a la vez. El control será transferido al subprograma seleccionado, el cual a su vez desplegará en la terminal una lista de opciones o información requerida para el proceso deseado. Cualquier número de requerimientos pueden ser ejecutados en el mismo o diferentes archivos, así como en los archivos creados por el programa. Cuando la ejecución del subprograma seleccionado haya sido completada, el subprograma devolverá el control al programa principal, mostrará nuevamente el menú en la pantalla y esperará el siguiente requerimiento. Si el requerimiento no puede ser completado se mostrará el mensaje de error apropiado y se devolverá el control al programa principal en espera del siguiente requerimiento. En esta descripción cualquier referencia a un archivo debe ser un identificador de archivo válido asociado a un archivo existente en el sistema.

Requerimiento

Comparación de Configuraciones

Definición

Compara dos configuraciones por la técnica de Rotación de Procrustes desarrollada por Gower.

Discusión

La exposición detallada de esta técnica se encuentra en el capítulo 4 de este trabajo. Este subprograma debe ser ejecutado cuando se tengan dos configuraciones 'X', 'Y' obtenidas por diferentes técnicas, o con una diferente función de distancia y se desee comparar la configuración 'Y' con respecto a la configuración 'X'.

Entrada y Salida

El subprograma requiere dos archivos. Cada uno conteniendo una configuración. El formato requerido es (8F10.6). El subprograma al término de su ejecución provee la facilidad de imprimir o crear un archivo en disco con la configuración 'Y' rotada y trasladada, el formato será también (8F10.6)

Opciones o Información

Requerida

Dimensión

Indicará la dimensión del espacio en que se encuentra cada configuración, la cual podrá diferir para cada configuración.

Tamaño de Muestra

Precisará el número de puntos en cada configuración. Deberá tenerse en cuenta que este número deberá coincidir en ambas configuraciones.

Requerimiento

Escalamiento Clásico

Definición

Obtiene la configuración asociada a una matriz de datos por el método de escalamiento clásico debido a Torgerson y Gower.

Discusión

La fundamentación teórica de esta técnica puede ser encontrada en el capítulo 2 de esta exposición. Este subprograma debe ser usado cuando se tenga una matriz de distancia que sea exacta o al menos aproximadamente euclídeana, o en el caso de una matriz de similaridad.

Entrada y Salida

El subprograma requiere solo un archivo de datos, el cual contendrá la matriz de datos analizar. El formato esperado es (8F10.6). Al completar la tarea el programa está en posibilidad de imprimir o crear un archivo en disco con la configuración obtenida por este método, el formato utilizado será también (8F10.6).

Opciones o Información Requerida

Espacio de Observaciones

Proporcionará el total de puntos considerados los cuales deben coincidir con la dimensión de la matriz de datos.

Espacio Modelo

Indicará el número de coordenadas o dimensiones requeridas en la configuración por obtener asociada a la matriz de datos en cuestión.

Matriz de Datos

El carácter de la matriz de datos determinará el tipo de proceso a seguir según el caso. Si no se proporciona dato al respecto se asumirá una matriz de distancia, la otra opción disponible es considerar una matriz de similaridad.

Requerimiento

Escalamiento Ordinal

Definición

Obtiene una configuración asociada a una matriz de datos por el método ordinal debido a Kruskal.

Discusión

Una exposición detallada de esta técnica puede ser encontrada en el capítulo 3 de este trabajo.

Este subprograma deberá utilizarse cuando las disimilaridades observadas son tales que sus valores - numéricos son de poco significado y el rango de orden de éstas es la única información considerada - relevante .

Entrada y Salida

Este subprograma opera con un archivo de datos conteniendo la matriz a procesar. El formato esperado es el usual (8F10.6). Al término del procesamiento de la matriz el programa podrá a elección del usuario, imprimir o crear un archivo en disco con la - configuración obtenida por este método. En este último caso el formato a utilizar será (8F10.6)

Opciones o Información Requerida

Dimensión

La dimensión de la matriz de datos es necesaria para la lectura de esta, por lo que respecta al espacio de observaciones. Para el espacio modelo, la - dimensión del espacio en el que se encontrará la - configuración a obtener, lo cual indicará el núme- ro de coordenadas requeridas en la configuración - buscada.

Configuración

Informará al programa de la disponibilidad o ausencia de una configuración inicial. Si esta opción - no es considerada se supondrá la ausencia de ésta.

Medida experimental

Definirá el carácter de la medida experimental en- contrada en la matriz de datos, esto es, si se trata de un coeficiente de disimilaridad o de un coe-ficiente de similaridad. Al ignorar esta opción se

asumirá que la medida utilizada corresponde a un -
coeficiente de disimilaridad.

Distancia

Seleccionará un elemento de la familia de funciones de distancia Minkowsky al especificar el parámetro R de esta clase de métricas. Esta función de distancia se utilizará en el cálculo de distancias entre los puntos de la configuración. La métrica default es la euclideana (correspondiente a $R=2$).

Ajuste

Determinará el enfoque a utilizar en el caso de encontrar medidas que coincidan en valor numérico dentro de las entradas de la matriz de datos. Los enfoques a elegir son el primario o secundario, teniendo por default el primero de estos.

Criterio de Terminación

Determinará el criterio de terminación del proceso iterativo. Se elegirá entre un número máximo de iteraciones o la magnitud relativa del gradiente de la función Stress. El proceso finalizará cuando alguna de las dos condiciones se satisfagan. Los valores - default considerados por el subprograma son de 100 iteraciones y 0.025 como condición de mínimo local.

EL GRADIENTE DE LA FUNCION STRESS

$$\nabla S^* = \sum_{i,j} (d_{ij} - \hat{d}_{ij})^2$$

$$\nabla S^* = 2 \sum_{i,j} (d_{ij} - \hat{d}_{ij}) D (d_{ij} - \hat{d}_{ij})$$

$$\nabla S^* = 2 \sum_{i,j} (d_{ij} - \hat{d}_{ij}) D (d_{ij} - \hat{d}_{ij})$$

$$\nabla S^* = 2 \sum (d_{ij} - \hat{d}_{ij}) D (d_{ij})$$

$$D^* = \sum_{i,j} d_{ij}^2$$

$$\nabla D^* = 2 \sum_{i,j} d_{ij} D (d_{ij})$$

$$\nabla S = \frac{1}{2} S \left\{ \frac{\nabla (S^*)}{S^*} - \frac{\nabla (T^*)}{T^*} \right\} = \frac{1}{2} S \left\{ \frac{2 \sum (d_{ij} - \hat{d}_{ij}) D (d_{ij})}{S^*} - \frac{2 \sum d_{ij} D (d_{ij})}{T^*} \right\}$$

$$\nabla S = \sum_{i,j} S \left\{ \frac{d_{ij} - \hat{d}_{ij}}{S^*} - \frac{d_{ij}}{T^*} \right\} D (d_{ij})$$

$$\begin{aligned} D (d_{ij}) &= D \left(\left(\sum_{s=1}^t |x_{is} - x_{js}|^r \right)^{\frac{1}{r}} \right) \\ &= \frac{1}{r} \left(\sum_{s=1}^t |x_{is} - x_{js}|^r \right)^{\frac{1}{r} - 1} D \left(\sum_{s=1}^t |x_{is} - x_{js}|^r \right) \\ &= \frac{1}{r} \left(\sum_{s=1}^t |x_{is} - x_{js}|^r \right)^{\frac{1}{r} - 1} \sum_{s=1}^t D (|x_{is} - x_{js}|^r) \end{aligned}$$

$$D (|x_{is} - x_{js}|^r) = r |x_{is} - x_{js}|^{r-1} D (|x_{is} - x_{js}|)$$

$$D (|x_{is} - x_{js}|) = \frac{|x_{is} - x_{js}|}{(x_{is} - x_{js})} D (x_{is} - x_{js})$$

$$\left(\sum |x_{is} - x_{js}|^r \right)^{\frac{1}{r} - 1} = \frac{\left(\sum_{s=1}^t |x_{is} - x_{is}| \right) d_{ij}}{\left(\sum_{s=1}^t |x_{is} - x_{js}| \right)^r d_{ij}^r} = \frac{1}{d_{ij}^{r-1}}$$

$$\frac{|x_{is} - x_{js}|}{|x_{is} - x_{js}|} = \text{signum}(x_{is} - x_{js})$$

$$s_{ij} = \frac{1}{r} \frac{1}{d_{ij}^{r-1}} r \sum_{s=1}^t |x_{is} - x_{js}|^{r-1} \text{signum}(x_{is} - x_{js}) \cdot D(x_{is} - x_{js})$$

De acuerdo a lo anterior la componente k l del gradiente de S es

$$= S \sum (\delta^{kl} - \delta^{kj}) \left(\frac{\hat{d}_{ij} - d_{ij}}{S^*} - \frac{d_{ij}}{r^*} \right) \frac{|x_{il} - x_{jl}|^{r-1}}{d_{ij}^{r-1}} \text{signum}(x_{il} - x_{jl})$$

EL ANALISIS DE COMPONENTES PRINCIPALES

La relación que existe entre el método de escalamiento clásico y el análisis de componentes principales se discutió en el capítulo 3. Tratando de evitar una complicación innecesaria en la exposición se ha diferido hasta aquí la presentación de los fundamentos de tal técnica.

Definición

Si x es un vector aleatorio con media u y matriz de covarianza S , entonces la transformación de componentes principales es la transformación

$$T : x \rightarrow y = G'(x - u)$$

donde t es ortogonal, $G'SG = L$ es diagonal y $l_1 > l_2 > \dots > l_p > 0$. El tener valores de l_i no negativos está garantizado si S es positiva definida. La representación de S se sigue del teorema de descomposición espectral. La i -ésima componente principal puede ser definida como el i -ésimo elemento del vector y , denotado como:

$$y_i = g_{(i)}'(x - u)$$

Aquí $g_{(i)}$ es la i -ésima columna de G , y puede ser llamado el i -ésimo vector de cargas de componentes principales.

Las propiedades principales que poseen las nuevas variables generadas vía la transformación T son enunciadas por el siguiente teorema:

Teorema

Sea T la transformación definida arriba; entonces se cumple

- i) $E(Y_i) = 0$
- ii) $V(Y_i) = 1$
- iii) $C(Y_i, Y_j) = 0, i \neq j$
- iv) $V(y_1) > V(y_2) > \dots > V(y_p) > 0$

$$v) \sum_{i=1}^p V(y_i) = \text{tr } S$$

$$vi) \prod_{i=1}^p V(y_i) = |S|$$

las propiedades i) a iv) se siguen a partir de la definición de la transformación T y las propiedades del operador Esperanza. La propiedad v) se sigue a partir de ii) y el hecho de que $\text{tr } S$ es la suma de los valores característicos, vi) se sigue de ii) y del hecho de que $|S|$ es el producto de los valores característicos.

TEOREMA DE DESCOMPOSICION ESPECTRAL

Cualquier matriz simétrica $A_{p \times p}$ puede ser escrita como

$$A = GLG^T = \sum l_i g_{(i)} g_{(i)}^T$$

donde L es una matriz diagonal de valores propios de A , y G es una matriz ortogonal cuyas columnas son vectores propios estandarizados.

Demostración

Supongamos que podemos encontrar vectores ortonormales $g_{(1)}, \dots, g_{(p)}$ tales que $A g_{(i)} = l_i g_{(i)}$ para algunos números l_i . Entonces

$$g_{(i)}^T A g_{(i)} = l_i g_{(i)}^T g_{(i)} = \begin{cases} l_i & i = j \\ 0 & i \neq j \end{cases}$$

o en forma matricial $G^T A G = L$ por lo que A y L tienen los mismos valores propios, así que los elementos de L son exactamente los valores propios de A con la misma multiplicidad. Por lo tanto debemos encontrar una base ortonormal de vectores propios. Hay que notar que si $l_i \neq l_j$ son valores propios distintos con vectores propios asociados x y y respectivamente, entonces

$$l_i x^T y = x^T A y = y^T A x = l_j y^T x, \text{ así que } y^T x = 0$$

De aquí que para una matriz simétrica, vectores propios correspondientes a valores propios distintos son ortogonales entre si. Supongamos que hay k valores propios de A distintos con correspondientes espacios propios H_1, \dots, H_k de dimensiones r_1, \dots, r_k

$$\text{Sea } r = \sum_{j=1}^k r_j$$

Ya que espacios propios distintos son ortogonales, existe un conjunto ortonormal de vectores e_1, \dots, e_r tal que los vectores

$$\begin{array}{c}
 j-1 \qquad \qquad j \\
 \Sigma r \quad \dots, \Sigma r \\
 i=1 \quad i+1 \qquad \qquad i=1 \quad i
 \end{array}$$

forman una base para H_j , de tal manera que r_j es menor o igual a la multiplicidad del valor propio correspondiente. De aquí que reordenando los valores propios l_i si es necesario, podemos suponer

$$A e_i = l_i e_i, \quad i = 1, \dots, r \text{ y } r \leq p$$

Si $r = p$, hagamos $g_i = e_i$ y el teorema está probado.

Si $r < p$ entonces nos lleva a una contradicción y por lo tanto no puede aparecer tal situación. Supongamos sin pérdida de generalidad que todos los valores propios de A son estrictamente positivos (si no, podemos reemplazar A por $A + aI$ para un escalar a apropiado, debido a que A y $A + aI$ tiene los mismos valores propios.

$$\text{Sea } B = A - \sum_{i=1}^r l_i e_i e_i'$$

entonces

$$\text{tr } B = \text{tr } A - \sum_{i=1}^r l_i (e_i' e_i) = \sum_{i=r+1}^p l_i > 0$$

ya que $r < p$. Por lo tanto B tiene al menos un valor propio no nulo, digamos f .

Sea $x \neq 0$ el vector propio correspondiente, entonces para $1 \leq j \leq r$

$$f e_j' x = e_j' B x = (l_j e_j' - \sum_{i=1}^r l_i (e_j' e_i) e_i') x = 0$$

así que es ortogonal a e_j , $j = 1, \dots, r$

Entonces

$$fx = Bx = (A - \sum l_i e_i e_i')x = Ax - \sum l_i (e_i' x) e_i = Ax$$

así que x es vector propio de A también. Por lo tanto $f = l_i$ para alguna i y x es una combinación lineal de las e_i , lo cual contradice la ortogonalidad entre las x y las e_i .

Teorema

Si UCR^n es abierto y $f: UCR^n \rightarrow R$ es diferenciable y $x_0 \in U$ es un extremo local, entonces $\nabla f(x_0) = \underline{0}$.

Demostración

Supongamos que f alcanza mínimo local en x_0 . Entonces para cualquier $h \in R^n$, la función $g(t) = f(x_0 + th)$, $t \in R$ tiene un mínimo local en $t = 0$, esto implica que $g'(0) = 0$. Por otra parte por la regla de la cadena

$$g'(0) = \nabla f(x_0) \cdot h = 0$$

Por lo tanto $\nabla f(x_0) \cdot h = 0$, para cualquier h , por lo que $\nabla f = \underline{0}$

Teorema

Supongamos que $\nabla f \neq \underline{0}$. Entonces $-\nabla f$ apunta en la dirección en que f decrece más rápidamente.

Demostración

Si n es un vector unitario, la rapidéz de crecimiento de f en la dirección n es $\langle -\nabla f, n \rangle = \|\nabla f\| \cos \theta$ donde θ es el ángulo entre n y $-\nabla f$. Este es máximo cuando $\theta = 0$; esto es cuando n y $-\nabla f$ son paralelos. En otras palabras, si se desea moverse en la dirección en que f decrece más rápidamente se debe proceder en la dirección de $-\nabla f$.

B I B L I O G R A F I A

Chattfield, C. and Collins, A. J. (1980)

Introduction to Multivariate Analysis

London: Chapman and Hall.

Gower, J.C. (1966)

Some distance properties of latent roots and vector methods used in multivariate analysis

Biometrika 53, 325-8.

Hamming, R. W. (1971)

Introduction to Applied Numerical Analysis

New York: McGraw-Hill.

Jardine, N. and Sibson, R. (1971)

Mathematical Taxonomy

London: Wiley Interscience.

Kruskal, J. B. (1964)

Multidimensional Scaling by optimizing goodness of fit to a nonmetric hypothesis

Psychometrika, 29, 1-27.

Kruskal, J. B. (1964)

Nonmetric Multidimensional Scaling: A Numerical Method

Psychometrika, 29, 115-29.

Kruskal, J. B. and Wish, M. (1978)

Multidimensional Scaling

London, Sage University Paper Series on Quantitative Applications in the Social Sciences.

Mardia, K. V., Kenneth, J.T. and Bibby, J. M. (1979)

Multivariate Analysis

London: Academic Press.

Sibson, R. (1972)

Order Invariant Methods for Data Analysis

J. R. Statist. Soc., B, 34, 311-49.

Torgerson, W. S. (1958)

Theory and Methods of Scaling

New York: Wiley.

Shepard, R. N. (1962)

The Analysis of Proximities: Multidimensional Scaling with an unknown distance function.

Psychometrika, 27, 125-139, 219-46.

Wilkinson, J. H. and Reinsch, C. (1971)

Linear Algebra

Handbook for Automatic Computation Vol II

New York, Springer Verlag.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200

```

BEGIN
  FILE
  TTY(KIND=REMOTE, MAXRECSIZE=22, MYUSE=10, PAGESIZE=25),
  PRN(KIND=PRINTER, PAGESIZE=55),
  DATAFILE(KIND=DISK, FILETYPE=7),
  OUTFILE(KIND=DISK, MAXRECSIZE=14, BLOCKSIZE=426);
  SWITCH FILE F:=TTY, PRN, OUTFILE;
  PROCEDURE STOPPAGE(DDV);
  INTEGER DDV;
  IF DDV = 1 THEN
    BEGIN
      FORMAT
      DISP(" * TRANSKITA CUALQUIER CARACTER PARA CONTINUAR *");
      STRING C;
      WRITE(TTY, DDV, DISP);
      READ(TTY, <A>, C);
      WHILE C = EMPTY DO
        BEGIN
          WRITE(TTY, DISP);
          READ(TTY, <A>, C);
        END;
      END;
    END;
  PROCEDURE QUICKSORT(D, IND, N);
  ARRAY D, IND(1);
  INTEGER N;
  BEGIN
    PROCEDURE SORT(L, R);
    INTEGER L, R;
    BEGIN
      INTEGER I, J;
      REAL X, U, Z;
      I:=L;
      J:=R;
      X:=D[(L+R) DIV 2];
      DO
        BEGIN
          WHILE D[I] > X DO
            I:=I + 1;
          WHILE X > D[J] DO
            J:=J - 1;
          IF I < J THEN
            BEGIN
              D:=D[I];
              Z:=IND[I];
              D[J]:=D[J];
              IND[J]:=Z;
              D[J]:=D[I];
              I:=I + 1;
              J:=J - 1;
            END;
        END;
      END;
      UNTIL I > J;
      IF I < J THEN SORT(L, J);
      IF I < R THEN SORT(I, R);
    END;
    SORT(L, N);
  END;

```

0000001
0010000
0010010
0010020
0010030
0010040
0010050
0010060
0010070
0010080
0010090
0010100
0010110
0010120
0010130
0010140
0010150
0010160
0010170
0010180
0010190
0010200
0010210
0010220
0010230
0010240
0010250
0010260
0010270
0010280
0010290
0010300
0010310
0010320
0010330
0010340
0010350
0010360
0010370
0010380
0010390
0010400
0010410
0010420
0010430
0010440
0010450
0010460
0010470
0010480
0010490
0010500
0010510
0010520
0010530
0010540
0010550
0010560

27100
27101
27102
27103
27104
27105
27106
27107
27108
27109
27110
27111
27112
27113
27114
27115
27116
27117
27118
27119
27120
27121
27122
27123
27124
27125
27126
27127
27128
27129
27130
27131
27132
27133
27134
27135
27136
27137
27138
27139
27140
27141
27142
27143
27144
27145
27146
27147
27148
27149
27150
27151
27152
27153
27154
27155
27156
27157
27158
27159
27160
27161
27162
27163
27164
27165
27166
27167
27168
27169
27170
27171
27172
27173
27174
27175
27176
27177
27178
27179
27180
27181
27182
27183
27184
27185
27186
27187
27188
27189
27190
27191
27192
27193
27194
27195
27196
27197
27198
27199
27200

```
FOR I=L + 1 STEP 1 UNTIL K DO
  BEGIN
    G:=E[I];
    V:=Q[I];
    H:=C * G;
    Z:=SQRT(F * F + H * H);
    C:=F / Z;
    S:=H / Z;
    G:=G * C + S * S;
    H:=H * C - S * G;
    FOR J=1 STEP 1 UNTIL N DO
      BEGIN
        V[J] := - 1;
        Z:=V[J];
        V[J] := X * C + Z * S;
        V[J] := X * S + Z * C;
        Q[I] := Z := SQRT(F * F + H * H);
        C:=F / Z;
        S:=H / Z;
        G:=G * C + S * V;
        H:=H * C - S * G;
        FOR J=1 STEP 1 UNTIL N DO
          BEGIN
            V[J] := - 1;
            Z:=V[J];
            V[J] := Y * C + Z * S;
            V[J] := Y * S + Z * C;
          END;
        END;
      END;
    END;
  END;
  BEGIN
    EKI:=F;
    EKI:=F;
    GO SPLITTING;
    CONVERGENCE :=
    IF Z < 0 THEN
      BEGIN
        EKI ES HECHO NO NEGATIVO -----
        FOR J=1 STEP 1 UNTIL N DO
          FOR I=L + 1 STEP 1 UNTIL K DO
            V[J,I] := - V[J,I];
          END;
        END;
      END;
    END;
  END;
  BEGIN
    PRMT(X37," * * * * ROTACION DE PROCRUSTES * * * *");
    PRMT(X8," * * * * CONFIGURACION 'Y' ROTADA Y TRASLADADA * * * *");
    PRMT(X37," * * * * ROTACION DE PROCRUSTES * * * *");
    PRMT(X37," * * * * CONFIGURACION 'X' * * * *");
    PRMT(X37," * * * * CONFIGURACION 'Y' * * * *");
    PRMAIN(X37," * * * * COMPARACION DE CONFIGURACIONES * * * *");
    X37," * * * * NIVEL DE SELECCION * * * *";
    X37," * * * * TAMANO DE MUESTRA * * * *";
    X37," * * * * DIMENSION * * * *";
    RPTF(X37," * * * * SELECCION * * * *");
    X37," * * * * ROTACION EN CADA CONFIGURACION * * * *";
    RDNF(X37," * * * * NIVEL DE SELECCION * * * *");
    X37," * * * * CONFIGURACION * * * *";
    X37," * * * * SELECCION OPCION * * * *";
    RDXF(X37," * * * * NIVEL DE SELECCION * * * *");
    X37," * * * * DIMENSION * * * *";
    RDXF(X37," * * * * NIVEL DE SELECCION * * * *");
    X37," * * * * DIMENSION DEL ESPACIO DE LA CONFIGURACION * * * *");
    X37," * * * * NIVEL DE SELECCION * * * *";
    RNF(X37," * * * * MATRIZ DE ROTACION * * * *");
    X37," * * * * TAMANO DE MUESTRA * * * *";
    RNF(X37," * * * * MATRIZ DE TRASLACION * * * *");
    X37," * * * * TAMANO DE CUADRADOS RESIDUALES * * * *";
    X37," * * * * FACTOR DE ESCALA * * * *";
  END;
  BEGIN
    DO, 2 DIMENSION DE LA CONFIGURACION 'X'
    DY, 2 DIMENSION DE LA CONFIGURACION 'Y'
    DX, X DIMENSION MAXIMA DE LAS CONFIGURACIONES DADAS
    NY, Y NUMERO DE PUNTOS EN LAS CONFIGURACIONES
  LABEL PRG1;
  PROCEDURE RDYF(COP);
  REAL OP;
  GET(OP,6);
  PROCEDURE RDXF(OP);
  REAL OP;
  GET(OP,6);
  PROCEDURE RDNF(OP);
  REAL OP;
  CASE OP OF
    BEGIN
      1 : COMANDO(RDXF, RDAP);
      2 : COMANDO(RDYF, RDYF);
    ELSE
      ERROR;
    END;
  END;
  PROCEDURE RPTF(OP);
  REAL OP;
  GET(OP,6);
  PROCEDURE PRMAIN(COP);
  REAL OP;
  CASE OP OF
    BEGIN
      1 : COMANDO(RPTF, RPTF);
      2 : COMANDO(RDNF, RDNF);
    ELSE
      ERROR;
    END;
  END;
  PROCEDURE XTX(N);
  ARRAY X(1:N,1);
  INTEGER N;
  BEGIN
    REAL AUX;
    INTEGER I;
    FOR I=1 STEP 1 UNTIL N DO
      FOR J=1 STEP 1 UNTIL I - 1 DO
        BEGIN
          AUX:=X(I,J);
          X(I,J):=X(J,I);
          X(J,I):=AUX;
        END;
      END;
    END;
  END;
  PROCEDURE XTY(X,Y,Z,N,F,N);
  ARRAY X(1:N,1);
  ARRAY Y(1:N,1);
  INTEGER N;
  BEGIN
```

00 57100
00 57101
00 57102
00 57103
00 57104
00 57105
00 57106
00 57107
00 57108
00 57109
00 57110
00 57111
00 57112
00 57113
00 57114
00 57115
00 57116
00 57117
00 57118
00 57119
00 57120
00 57121
00 57122
00 57123
00 57124
00 57125
00 57126
00 57127
00 57128
00 57129
00 57130
00 57131
00 57132
00 57133
00 57134
00 57135
00 57136
00 57137
00 57138
00 57139
00 57140
00 57141
00 57142
00 57143
00 57144
00 57145
00 57146
00 57147
00 57148
00 57149
00 57150
00 57151
00 57152
00 57153
00 57154
00 57155
00 57156
00 57157
00 57158
00 57159
00 57160
00 57161
00 57162
00 57163
00 57164
00 57165
00 57166
00 57167
00 57168
00 57169
00 57170
00 57171
00 57172
00 57173
00 57174
00 57175
00 57176
00 57177
00 57178
00 57179
00 57180
00 57181
00 57182
00 57183
00 57184
00 57185
00 57186
00 57187
00 57188
00 57189
00 57190
00 57191
00 57192
00 57193
00 57194
00 57195
00 57196
00 57197
00 57198
00 57199
00 57200


```

3 JCC:=L;
4 END;
5 END;
6 IF M=0 = J THEN
7 BEGIN
8   TIE(TTY, <M * MATRIZ DE DATOS VACIA * M);
9   TAREA ABORTADA;
10 GO OSEOT;
11 END;
12 FOR K:=1 STEP 1 UNTIL N DO
13   DEALLOCATE (X(K), Y(K));
14 END;
15 PROCEDURE INITCODE;
16 IF CT = 1 THEN * CONFIGURACION INICIAL DISPONIBLE
17 BEGIN
18   INPUT DESK (OSFPT);
19   TPUT DATA (X, N, T, CSEOT);
20 END
21 ELSE
22 BEGIN
23   INTEGER T, J;
24   REAL SEED;
25   SEED:=TIME(11);
26   FOR I:=1 STEP 1 UNTIL N DO
27     FOR J:=1 STEP 1 UNTIL T DO
28       X(I, J) := RANDOP(SEED);
29     END;
30   END;
31 PROCEDURE TILS;
32 X---- PARA CADA ENTRADA DE LA MATRIZ DE DATOS, ENCUENTRA EL NUMERO-----
33 X---- DE FUERA DE LA MATRIZ A ESTE TIE(K) = 0 SI TIE(K) = 0 SI TIE(K) = 0 SI TIE(K) = 0 SI
34 X---- DISSIM(K) <> DISSIM(L), PARA TODA K <> L-----
35 BEGIN
36   INTEGER K, L;
37   DO
38     BEGIN
39       K := K + 1;
40       DISSIM(K) := 1;
41       WHILE DISSIM(K) = DISSIM(L) DO
42         K := K + 1;
43       TIE(K) := L - K;
44     END;
45   UNTIL K > N;
46 END;
47 PROCEDURE DISTANCES;
48 BEGIN
49   REAL EXP;
50   EXP := 1 / S;
51   FOR L:=1 STEP 1 UNTIL M DO
52     BEGIN
53       DIST(L) := 0;
54       FOR S:=1 STEP 1 UNTIL T DO
55         DIST(L) := * + RES(X(I(L), S) - X(J(L), S)) * * R;
56       DIST(L) := * + EXP;
57     END;
58 END;
59 PROCEDURE NORMALIZE;
60 BEGIN
61   INTEGER T, S;
62   REAL MEAN(S);
63   MEAN(S) := * / N;
64   FOR S:=1 STEP 1 UNTIL N DO
65     FOR I:=1 STEP 1 UNTIL T DO
66       X(I, S) := * - MEAN(S);
67       X(I, S) := * + X(I, S) * * 2;
68     END;
69   SE := SE * (SE / N);
70   FOR I:=1 STEP 1 UNTIL N DO
71     FOR J:=1 STEP 1 UNTIL T DO
72       X(I, J) := * / SE;
73     END;
74 END;
75 PROCEDURE PRIMARY;
76 BEGIN
77   INTEGER K, NT, P;
78   K := 1;
79   P := IF (COUNT = 0) THEN -1 ELSE 1;
80   WHILE K <= N DO
81     IF K >= N THEN
82       BEGIN
83         D := INDEFINITE(TIE(K));
84         INDEFINITE(U);
85         S := K + TIE(K) - 1;
86         NT := S - K;
87         TIE(K) := S;
88         FOR I:=K STEP 1 UNTIL S DO
89           BEGIN
90             U := * + 1;
91             DIST := DIST(L) * P;
92             INDCO := INDCO;
93           END;
94           SORT(CO, IND, NT);
95           FOR L:=K STEP 1 UNTIL S DO
96             BEGIN
97               U := * + 1;
98               DIST := DIST(L) := DCU * * P;
99               INDCO := INDCO;
100            END;
101          END;
102          TIE(K) := NT;
103          K := S + 1;
104        END;
105      ELSE
106        BEGIN
107          DIST(K) := DIST(K);
108          DIST(K) := *;
109          K := * + 1;
110        END;
111      END;
112 END;
113 PROCEDURE SECONDARY;
114 BEGIN
115   INTEGER K;
116   K := 1;
117   DO
118     BEGIN
119       IF (K) = 1 THEN 1 ELSE TIE(K);
120       IF (K) < 2 THEN
121         TIE(K) := S(C);
122       END;
123     END;
124   UNTIL K >= N;
125 END;

```

```

2110 DOOTHERK:=0;
2120 K:=K+1;
2130 END
2140 BEGIN
2150 INTEGER L;
2160 REAL AC;
2170 INTEGER S;
2180 S:=K+TIE(K)-1;
2190 DOOTHERK:=K;
2200 FOR L:=K STEP 1 UNTIL S DO
2210 AC:=*+DSTEL;
2220 DHATE(K)=AC/TIE(K);
2230 DHATE(K+1):=AC;
2240 W:=S+1;
2250 END;
2260 UNTIL K > 4;
2270 END;
2280 PROCEDURE MONOTONE;
2290 BEGIN
2300 INTEGER FIRST,K,LAST;
2310 ACCELO STATE UP DOWN;
2320 INTEGER PROCEDURE EPB(L);
2330 INTEGER L;
2340 IF DOOTHER(L) = 0 THEN
2350 DO:=1;
2360 ELSE
2370 DO:=DOOTHER(L);
2380 REAL PROCEDURE FIT(L);
2390 INTEGER L;
2400 IF DOOTHER(L) = 0 THEN
2410 FIT:=DHATE(L);
2420 ELSE
2430 FIT:=DHATE(L+1);
2440 INTEGER PROCEDURE INDICE(L);
2450 INTEGER L;
2460 IF DOOTHER(L) = 0 THEN
2470 INDICE:=L;
2480 ELSE
2490 INDICE:=DOOTHER(L);
2500 PROCEDURE JOIN;
2510 BEGIN
2520 LACT:=FIRST+(IF DOOTHER(FIRST) = 0 THEN 0
2530 ELSE DOOTHER(FIRST)-1);
2540 DHATE(LACT)=EPB(K)+EPB(FIRST);
2550 DOOTHER(KACT)=EPB(K)+EPB(FIRST);
2560 DHATE(K)=DHATE(K+1)/DOOTHER(K);
2570 DOTHER(LACT):=K;
2580 END;
2590 DEFACTIVE = TRUE #;
2600 DOWNACTIVE = FALSE#;
2610 W:=1;
2620 DO
2630 BEGIN
2640 STATE:=UPACTIVE;
2650 DOWN:=FALSE;
2660 DO
2670 IF STATE THEN
2680 BEGIN
2690 IF FIRST=(K+(IF DOOTHER(K)=0 THEN 0
2700 ELSE DOOTHER(K)-1)) = W THEN
2710 UP:=TRUE;
2720 IF ACT UP:=DHATE(K) < DHATEFIRST:=*+1 THEN
2730 JOIN;
2740 STATE:=DOWNACTIVE;
2750 END;
2760 BEGIN
2770 IF K = 1 THEN
2780 DOWN:=TRUE;
2790 IF ACT DOWN:=DHATE(INDICE(K)-1) < DHATE(K) THEN
2800 FIRST:=K;
2810 K:=INDICE(K-1);
2820 JOIN;
2830 END;
2840 STATE:=UPACTIVE;
2850 UNTIL UP AND DOWN;
2860 K:=*+(IF DOOTHER(K) = 1 THEN 1 ELSE DOOTHER(K));
2870 END;
2880 UNTIL K > W;
2890 K:=1;
2900 BEGIN
2910 REAL FITTED;
2920 INTEGER L,M;
2930 S:=K+TIE(K)-1;
2940 FIT:=DHATE(K);
2950 FOR L:=K STEP 1 UNTIL S DO
2960 DHATE(L):=FITTED;
2970 K:=S+1;
2980 END;
2990 UNTIL K > W;
3000 END;
3010 PROCEDURE FIT;
3020 BEGIN
3030 IF IT = 0 THEN
3040 PRIMARY;
3050 ELSE
3060 SECONDARY;
3070 MONOTONE;
3080 END;
3090 PROCEDURE TRACE;
3100 BEGIN
3110 IF IT = 1 THEN
3120 BEGIN
3130 PSI:=FIAS:=S;
3140 FILL SAVE[4] WITH S(0);
3150 END;
3160 IF IT > 5 THEN
3170 BEGIN
3180 INTEGER I;
3190 I:=IT MOD 5;
3200 FIAS:=SAVE[I];
3210 SAVE[I]:=S;
3220 END;
3230 ELSE
3240 SAVE[IT MOD 5]:=S;
3250 END;
3260 PROCEDURE STRESS;
3270 BEGIN
3280 INTEGER P;
3290 SASTK:=;
3300 TASTK:=;

```