### A picture containing text Description automatically generatedUniversidad de Granada

### Escuela Internacional de Posgrado

### Máster en Estadística Aplicada

### Materia: Análisis de datos. Técnicas aplicadas a datos de proximidad.

### Alumno: Francisco Javier Márquez Rosales

# **Tema 4: Diferencias individuales en MDS.**

# **Ejercicios:**

Noviembre, 2022

Ejercicios de Tema 4:

Ejercicio 4.1. La siguiente Tabla contiene las distancias entre los 45 pares de

colores formados con 10 colores, emitidas por 11 individuos normales (Helm,

1959).

Leer los datos con SPSS.

1. Realizar el análisis de los datos usando PROXSCAL SPSS para el

modelo identidad.

2. Realizar el análisis de los datos con PROXSCAL de SPSS para el

modelo de diferencias individuales.

3. Compara los resultados anteriores.

Ejercicio 4.2: Si consideramos e=, estudiar la relación entre y

,.

**Solución:**

**Preparación de los datos**

Inicialmente, para crear el conjunto de datos llevamos la tabla que se encuentra en el PDF a un conjunto .txt el cual es leído desde el spss. A continuación, un resumen del conjunto y del código utilizado en la importación.

**Sintaxis empleada:**

PRESERVE.

SET DECIMAL DOT.

GET DATA /TYPE=TXT

/FILE="C:\Users\franm\OneDrive\...\data\_helm\_color2.txt"

/ENCODING='UTF8'

/DELCASE=LINE

/DELIMITERS=" "

/ARRANGEMENT=DELIMITED

/FIRSTCASE=2

/DATATYPEMIN PERCENTAGE=95.0

/VARIABLES=

A AUTO

C AUTO

E AUTO

G AUTO

I AUTO

K AUTO

M AUTO

O AUTO

Q AUTO

S AUTO

/MAP.

RESTORE.

CACHE.

EXECUTE.

DATASET NAME DataSet2 WINDOW=FRONT.

**Resumen de la lectura**

Data written to the working file.10 variables and 110 cases written.Variable: A Type: Number Format : F4.1Variable: C Type: Number Format : F4.1Variable: E Type: Number Format : F4.1Variable: G Type: Number Format : F4.1Variable: I Type: Number Format : F4.1Variable: K Type: Number Format : F4.1Variable: M Type: Number Format : F4.1Variable: O Type: Number Format : F3.1Variable: Q Type: Number Format : F3.1Variable: S Type: Number Format : F3.1

Luego, se comprueba la lectura haciendo un análisis descriptivo de las variables obtenidas.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Descriptive Statistics** | | | | | |
|  | N | Minimum | Maximum | Mean | Std. Deviation |
| A | 110 | .0 | 18.8 | 8.209 | 4.1737 |
| C | 110 | .0 | 17.3 | 8.114 | 4.6987 |
| E | 110 | .0 | 16.1 | 7.162 | 5.2037 |
| G | 110 | .0 | 17.3 | 5.195 | 4.9591 |
| I | 110 | .0 | 15.8 | 4.334 | 4.8813 |
| K | 110 | .0 | 15.1 | 3.322 | 4.4433 |
| M | 110 | .0 | 13.9 | 2.304 | 3.7536 |
| O | 110 | .0 | 7.5 | 1.154 | 2.3639 |
| Q | 110 | .0 | 5.2 | .356 | 1.0892 |
| S | 110 | .0 | .0 | .000 | .0000 |
| Valid N (listwise) | 110 |  |  |  |  |

**Ejercicio 4.1.1 Realizar el análisis de los datos usando PROXSCAL SPSS para el modelo identidad.**

A continuación, se ejecuta el análisis definiendo el modelo de la siguiente forma, se asume que los datos son del tipo intervalo:

Graphical user interface, application

Description automatically generated

El resultado obtenido se presenta a continuación:

|  |  |  |
| --- | --- | --- |
| **Case Processing Summary** | | |
| Cases | | 110 |
| Sources | | 11 |
| Objects | | 10 |
| Proximities | Total Proximities | 495a |
| Missing Proximities | 0 |
| Active Proximitiesb | 495 |
| a. Sum over sources of all strictly lower-triangular proximities. | | |
| b. Active proximities include all non-missing proximities. | | |

El resumen del procesamiento nos muestra como leímos de forma adecuada los datos desde 11 fuentes distintas (individuos) para los diez objetos (grupos de color).

**Goodness of Fit**

|  |  |
| --- | --- |
| **Stress and Fit Measures** | |
| Normalized Raw Stress | .02488 |
| Stress-I | .15774a |
| Stress-II | .45030a |
| S-Stress | .09363b |
| Dispersion Accounted For (D.A.F.) | .97512 |
| Tucker's Coefficient of Congruence | .98748 |
| PROXSCAL minimizes Normalized Raw Stress. | |
| a. Optimal scaling factor = 1.026. | |
| b. Optimal scaling factor = .992. | |

El resultado de la prueba de la bondad del ajuste indica que el estrés bruto normalizado es 0.024, este valor por ser cercano a cero (criterio común aceptado menor a 0.1) indica una buena calidad del modelo.

Ahora obtenemos el resultado gráfico del escalamiento multidimensional.

img.emf

La forma de las categorías en el grafico sugiere una dispersión balanceda entre las opiniones de los individuos a los grupos de color. La primera dimensión, principalmente caracterizada por las diferencias de los grupos de colores {A,S,Q} versus el grupo {G,I,K} mientras que el segundo eje con los grupos {C,E} versus {O,M}. Podemos sugerir finalmente los siguientes grupos afines de acuerdo a la opinión de los individuos: {A,S,Q}, {G,I,K},{C,E} y {O,M}.

**Ejercicio 4.1.2. Realizar el análisis de los datos con PROXSCAL de SPSS para el modelo de diferencias individuales.**

A continuación, se ejecuta el análisis definiendo el modelo de la siguiente forma, se asume que los datos son del tipo intervalo:

Graphical user interface, application

Description automatically generated

|  |  |  |
| --- | --- | --- |
| **Case Processing Summary** | | |
| Cases | | 110 |
| Sources | | 11 |
| Objects | | 10 |
| Proximities | Total Proximities | 495a |
| Missing Proximities | 0 |
| Active Proximitiesb | 495 |
| a. Sum over sources of all strictly lower-triangular proximities. | | |
| b. Active proximities include all non-missing proximities. | | |

El resumen del procesamiento nos muestra, al igual que en el análisis anterior, como leímos de forma adecuada los datos desde 11 fuentes distintas (individuos) para los diez objetos (grupos de color).

**Goodness of Fit**

|  |  |
| --- | --- |
| **Stress and Fit Measures** | |
| Normalized Raw Stress | .01394 |
| Stress-I | .11806a |
| Stress-II | .32223a |
| S-Stress | .03838b |
| Dispersion Accounted For (D.A.F.) | .98606 |
| Tucker's Coefficient of Congruence | .99301 |
| PROXSCAL minimizes Normalized Raw Stress. | |
| a. Optimal scaling factor = 1.014. | |
| b. Optimal scaling factor = 1.001. | |

El resultado de la prueba de bondad de ajuste indica que el estrés bruto normalizado es 0.013 (menor que para el modelo identidad), este valor por ser cercano a cero (criterio común aceptado menor a 0.1) indica una buena calidad del modelo.

Ahora obtenemos el resultado gráfico del escalamiento multidimensional.

img.emf

La forma de las categorías en el grafico sugiere las mismas conclusiones observadas en el análisis anterior, de hecho el gráfico es idéntico. La primera dimensión, principalmente caracterizada por las diferencias de los grupos de colores {A,S,Q} versus el grupo {G,I,K} la segunda dimensión con los grupos {C,E} versus {O,M}. Se observan los siguientes grupos afines de acuerdo a la opinión de los individuos: {A,S,Q}, {G,I,K},{C,E} y {O,M}.

Se nota que la distribución en la misma que el análisis anterior pero la escala es menor, por tanto los puntos son más cercanos. Ahora obtenemos los pesos enla dimensiones

|  |  |  |
| --- | --- | --- |
| **Individual Spaces**  **Dimension Weights** | | |
| Source | Dimension | |
| 1 | 2 |
| SRC\_1 | .514 | .470 |
| SRC\_2 | .673 | .515 |
| SRC\_3 | .489 | .443 |
| SRC\_4 | .469 | .417 |
| SRC\_5 | .457 | .432 |
| SRC\_6 | .483 | .448 |
| SRC\_7 | .532 | .484 |
| SRC\_8 | .435 | .411 |
| SRC\_9 | .475 | .463 |
| SRC\_10 | .405 | .394 |
| SRC\_11 | .439 | .437 |

img.emf

Para todos los individuos, los pesos presentan valores muy parecidos en cuanto a la importancia que otorga cada uno de los dos ejes, se puede ver como hacia la dimensión 1 los valores están cercanos a 0.5 mientras que frente a la dimensión 2 cercanos a 0.4. El segundo individuo destaca porque su proyección sobre ambos ejes es mayor al resto, 0.67 primera dimensión y 0.51 segunda. Sugieriendo para esta caso mayor importancia de la primera.

**4.1.3. Compara los resultados anteriores**

El modelo identidad ofrece una bondad del ajuste que indica que el modelo es apropiado para describir el comportamiento de las categorías, con un estrés bruto normalizado cercano a cero y con los valores de DAF y de coeficiente de Tucker cercanos a 1. En el modelo de diferencias individuales estos valores son aún más favorables en cuanto al ajuste del modelo.

La distribución de los puntos en el plano en ambos análisis es similar, la diferencia está en la espala que sugiere que para el modelo de diferencias individuales son menores distancias.