Identificación de Coordenadas de las Ciudades Capitales de Bolivia utilizando el Escalamiento Multidimensional

Autor: Lic. Fernando O. Rivero Suguiura

1. Introducción

El "escalamiento multidimensional" es una técnica multivariante que tiene el objetivo de identificar la representación gráfica perceptual de un conjunto de objetos, cuando se conocen las similitudes entre dichos objetos.

Shepard demuestra empíricamente, que si se conoce una ordenación de las distancias entre objetos, podría encontrarse las coordenadas de estos objetos en un espacio euclidiano, cuyas interdistancias reproducen prácticamente la ordenación original. Las ideas de Shepard fueron después mejoradas por Kruskal en los años 60 y desarrolladas por otros autores como Guttman y Lingoes donde su aplicación se inicia en la Psicología y después a muchas áreas del que hacer científico.

El presente artículo presenta una aplicación de este método como parte del Análisis Multivariante de manera práctica y didáctica, en un ejemplo de la medición de distancias euclídeas entre las nueve ciudades capitales de nuestro país (Sucre, La Paz, Cochabamba, Oruro, Potosí, Tarija, Santa Cruz, Trinidad y Cobija), con el objetivo de identificar, luego las coordenadas de las ciudades nombradas sobre un plano, de tal manera que se pueda comprender mejor lo que hace metodológicamente el análisis de Escalamiento Multidimensional.

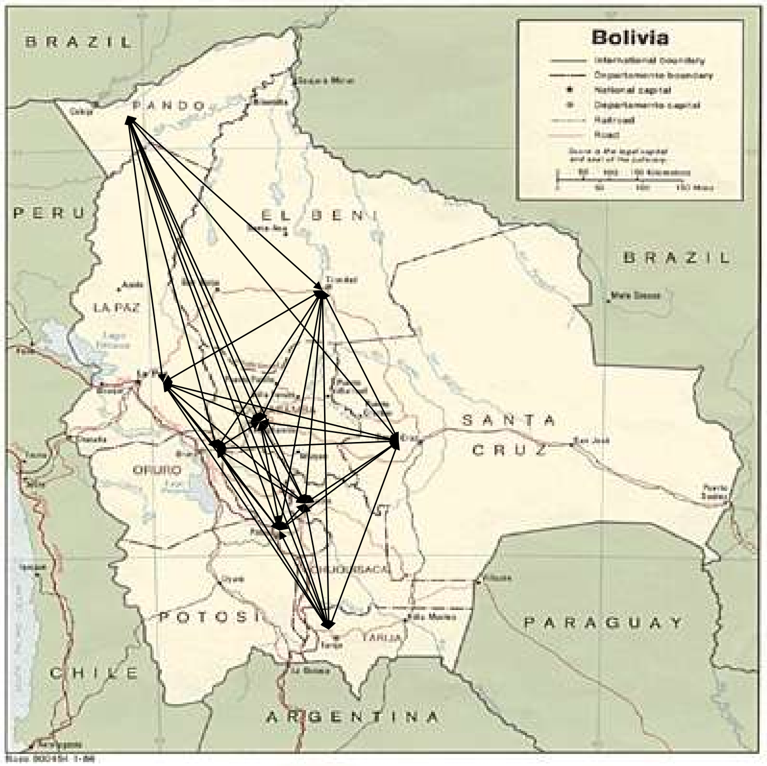
Sea la matriz D de distancias euclídeas de las nueve ciudades capitales de Bolivia, medido en centímetros, obtenidas de un mapa del país a escala 1:2'500'000, representado en menor escala en el gráfico Nº 1. La matriz de distancias, es:

Ciudades	Sucre	La Paz	Cochabamba	Oruro	Potosí	Tarija	Santa Cruz	Trinidad	Cobija
Sucre	-	28,5	14,2	15,9	5,4	19	18,4	32,1	64,5
La Paz		-	16	13,2	29	45,1	37,7	28	40,3
Cochabamba			-	8,2	16,6	32,8	22,3	22,1	50,6
Oruro				-	15,7	32,1	29,2	29	52,3
Potosí					-	16,6	23,3	36,4	66,5
Tarija						-	30,5	50,4	83
Santa Cruz							-	25,3	64,6
Trinidad								-	40,7
Cobija									-

Fuente: Propia. Medición en centímetros de ciudad a ciudad de un mapa de Bolivia a escala 1:2'500'000.

Por ejemplo, 28.5 cm. corresponde a la distancia en línea recta entre las ciudades de Sucre y La Paz. 36.4 cm. es la distancia entre las ciudades de Potosí y Trinidad, y así sucesivamente.

Gráfico Nº 1
Bolivia: Trazado de distancias entre las capitales de Departamento y medidas en centímetros
(Escala inferior a 1:2’500’000).



2. Metodología

El método consiste en determinar la matriz de distancias euclidianas D a partir de una matriz de proximidades Δ.

Sea la matriz $A=-\frac{1}{2}D^{(2)}$ definida a partir de la matriz D de distancias euclídeas al cuadrado. Sea $B=H'AH$ de dimensión 9x9 denominada “matriz de productos escalares”, con H “matriz de centrado” donde $H=I-n^{-1}11'$. I matriz identidad, n número de objetos en la muestra, y 1 vector columna de “unos”. La matriz de centrado es simétrica e idempotente de rango $n-1$, que al aplicarla sobre un vector de datos lo transforma en sus desviaciones respecto a la media de sus componentes. Es decir, cumple con las siguientes propiedades:

i) $H'=H$
ii) $H^2=H$
iii) $H1=1'H=0$
iv) los valores propios de H están entre 0 y 1

Entonces, la matriz B queda como:

$$
\begin{aligned}
B &= H'AH \\
&= (I-n^{-1}11')'A(I-n^{-1}11') \\
&= A-n^{-1}A\,11'-n^{-1}11'A+n^{-2}11'A11'
\end{aligned}
$$

Es semidefinida positiva, puesto que:

$$B=HAH=-\frac{1}{2}HD^{(2)}H=XX'\geq 0$$

Donde X es matriz de dimensión $n * p$ de "coordenadas" en el Análisis de Escalado Multidimensional, siendo p la dimensión de las coordenadas.

Por la descomposición espectral de una matriz, se tiene que:

$$B = P\Delta P' = XX' \geq 0$$

Δ es la matriz de valores propios normalizados de B. Con $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > \lambda_{p+1} = 0$, y la matriz P de dimensión $n * p$, es la que contiene los vectores propios asociados a los valores propios correspondientes. Es decir:

$$B = P\Delta P' = \left(P\Delta^{1/2}\right)\left(P\Delta^{1/2}\right)'$$

Con $X = P\Delta^{1/2}$. Las coordenadas principales resultantes de los objetos en estudio, tienen dimensión $q < p$.

Si $q = 2$, las dos primeras coordenadas de X proporcionan la representación gráfica en el plano. El gráfico Nº 2, muestra la salida en SPSS de las coordenadas en un plano, resultante del ejemplo, siendo los primeros dos valores y vectores propios representativos, de la matriz B, los siguientes:

$\lambda_1 = 4.50$

$v_1 = (-0.21 \quad 0.17 \quad -0.00 \quad -0.02 \quad -0.24 \quad -0.49 \quad -0.17 \quad 0.20 \quad 0.75)$

$\lambda_2 = 0.83$

$v_2 = (0.01 \quad 0.42 \quad 0.05 \quad 0.32 \quad 0.17 \quad 0.13 \quad -0.62 \quad -0.54 \quad 0.06)$

En resumen el procedimiento consiste en transformar:

Δ (Proximidades) $\rightarrow D$ (Distancias) $\rightarrow B$ (Productos escalares) $\rightarrow X$ (coordenadas)

Una medida que nos informa de la bondad del Modelo de Escalamiento Multidimensional en un plano, es el estadístico *Stress* que Kruskal definió como:

$$Stress = \sqrt{\frac{\sum_{i,j}\left(d_{ij} - \hat{d}_{ij}\right)^2}{\sum_{i,j} d^2{}_{ij}}}$$

Mientras mayor sea la diferencia entre las disparidades y las distancias, mayor será el *Stress* y por tanto peor será el modelo. Kruskal (1964) sugiere las siguientes interpretaciones del *Stress*:

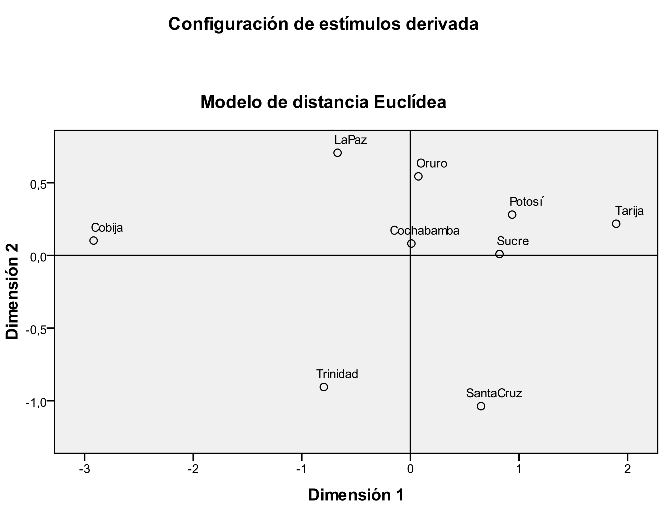
0.200	→ Pobre
- 0.100	→ Regular
- 0.050	→ Bueno
- 0.025	→ Aceptable
- 0.010	→ Excelente

El resultado del ejemplo, da como estadístico de *Stress* igual a:

For matrix
Stress = ,00170 RSQ = ,99999

De acuerdo a la clasificación anterior, está entre aceptable y excelente.

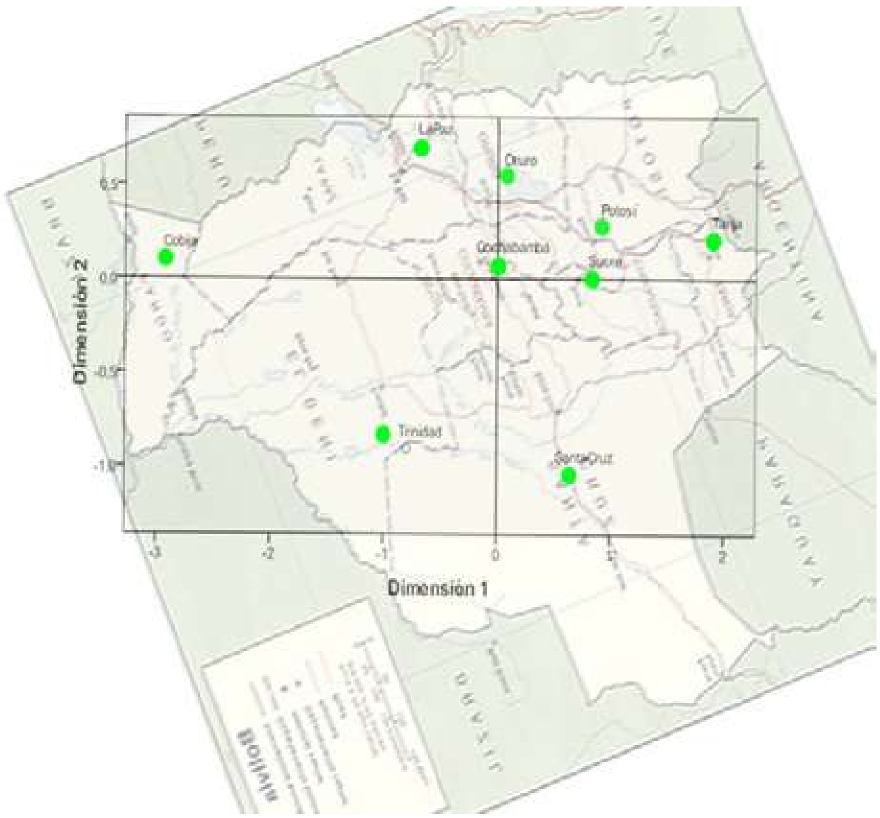
Gráfico Nº 2
Diagrama de dispersión de puntos en el plano euclídeo, de los dos primeros vectores propios de la matriz ***B***, (salida en SPSS).



Gráficamente, este es el resultado de identificación de coordenadas de las ciudades capitales de Bolivia en un plano, según el método de Escalamiento Multidimensional.

Para comprender mejor su ubicación, se puede observar el Gráfico Nº 3, donde se traslada y rota el mapa de Bolivia, haciendo coincidir con el Gráfico Nº 2.

Gráfico Nº 3
Diagrama de dispersión de puntos en el plano euclídeo, de los dos primeros vectores propios de la matriz ***B***, sobreponiendo el mapa de Bolivia ajustado al resultado.



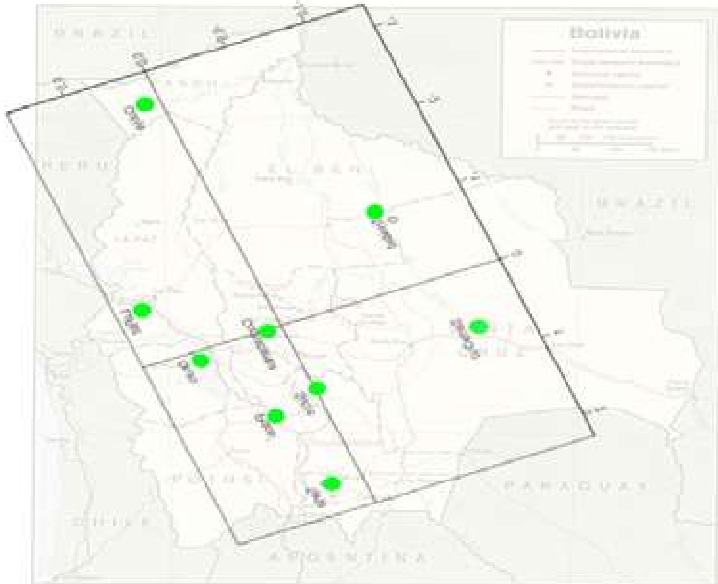
Como se puede notar en el Gráfico Nº 3, el punto de la ciudad de La Paz, aparece en la parte superior del cuadrante 2. Cerca al punto La Paz, aparece Oruro y luego el punto Cochabamba, Sucre, Potosí, Tarija, y así sucesivamente. Se puede observar cierta coincidencia y precisión en la ubicación de los puntos.

El Gráfico Nº 4, muestra la rotación y traslación del dispersograma, hasta ubicar de forma cercana y precisa, los puntos de las ciudades capitales en el mapa de Bolivia.

"No existe la suerte. Sólo hay preparación adecuada o inadecuada para hacer frente a una estadística".
Robert Heinlein

Gráfico Nº 4

Mapa de Bolivia sobreponiendo el diagrama de dispersión de puntos en el plano euclídeo, de los dos primeros vectores propios de la matriz $\boldsymbol{B}$ ajustado al resultado.



3. Conclusiones y Recomendaciones

El artículo es una demostración sencilla de un ejemplo didáctico, que da a comprender lo que realmente hace el Escalamiento Multidimensional con las distancias observadas de ciudad a ciudad y con el resultado de la identificación final de las posiciones o coordenadas de cada una de las nueve ciudades con alta aproximación. Para ello se utilizó un mapa de Bolivia a escala 1:2’500’000 representada a escala menor en el Gráfico Nº 1.

A partir de este ejemplo, el analista se da cuenta de lo importante y aplicado que es este método para otro tipo de investigación. Por ejemplo, reagrupar gráficamente en un mapa perceptual, municipios de Bolivia que presentan niveles de pobreza altos, medios y bajos. Ayudaría bastante a comprender el agrupamiento que se da de estos municipios por una característica como es el nivel de pobreza.

El método requiere que la medición de las similitudes entre objetos o elementos, cumplan la condición de distancia euclidiana. Si no fuese así, se puede cometer errores en la aplicación del Escalamiento Multidimensional, y es que la matriz B de productos escalares puede no ser semidefinida positiva, y complicar el análisis. Especialmente, las similitudes pueden no cumplir la propiedad de la desigualdad triangular $d_{ij} \leq d_{ij'} + d_{j'j}$. Sin embargo, existe un método denominado “Estimación de la Constante Aditiva” debido a Torgerson que soluciona el problema.

En la próxima versión de la Revista Varianza, se puede realizar un artículo aplicativo a un campo de investigación como el mencionado.

4. Referencias

[1] S. JAMES PRESS (1948): “Applied Multivariate Analysis”, University of Chicago. New York.

[2] K. .V. MARDIA (1979): “Multivariate Analysis”, Academic Press. Inc. London.

[3] DONALD F. MORRISON (1967): “ Multivariate Statistical Methods”, Mc Graw Hill Book Company. New York.

[4] CUADRAS, CARLES M. (2006): “Métodos de Análisis Multivariante”. Barcelona, España.

[5] EZEQUIEL URIEL – JOAQUIN ALDAS (2005): “Análisis Multivariante Aplicado”. Thomson. España.