ISSN 2581-3463

AJMS

Asian Journal of Mathematical Sciences

RESEARCH ARTICLE

Mapping and Cluster Analysis of Crime Rates in Nigeria

Saleh Ibrahim Musa, Mustafa Babagana Abubakar

Department of Statistics, Federal University of Lafia, PMB 146 Lafia, Nigeria

Received: 25-02-2021; Revised: 30-03-2021; Accepted: 10-04-2021

ABSTRACT

In this study, attempt is made to use the multidimensional scaling analysis in conjunction with Cluster Analysis to given visual representation of the crime rate of some selected cities in Nigeria graphically in a two-dimensional multidimensional scaling space, since visual inspection can greatly aid interpretations. Multidimensional scaling is an explanatory technique used to visualize proximities in a low dimensional space. Interpretation of the dimension can lead to an understanding of the Processes underlining the perceived nearness of entities. From the result of findings; cities like Lagos, Port Harcounrt, Warri, Benin City, Onitsha, Kaduna and Kano which has high crime rate can be labeled as hot spots, cause of availability of certain environmental factors. Unlike cities like Sokoto, Jimeta, lafia, Ilorin, and Jos which has low crime rate.

Key words: Cluster analysis, mapping, crimes, multidimensional scaling

INTRODUCTION

Crime mapping and analysis have evolved significantly over the past 30 years. In the beginning, many agencies utilized city maps with colored pins to visualize individual crime events and crime-plagued areas. Today, with the rapid advancement of technology, computer-based techniques for exploring, visualizing, and explaining the occurrences of criminal activity have been essential. One of the more influential tools facilitating the exploration of the spatial distribution of crime has been the Global Positioning System (GPS).[1,2-4] As Murray *et al*. (2001) note, it is the ability to combine spatial information with other data that makes GPS so valuable.[4] Furthermore, the sheer quantity of information available to most analysts necessitates an intelligent computational system, able to integrate a wide variety of data and facilitate the identification of patterns with minimal effort.

Fundamental to the explanation of criminal activities in a spatial context are certain environmental factors, such as the physical layout of an area, proximity to various services, and land use mixes - all of which are likely to influence criminal behavior.[5,6] Issues of access, exposure, opportunity, and the availability of targets are also important elements in helping explain crime from an environmental perspective.[4,7-9] Not surprisingly, research indicates that certain areas are more prone to higher concentrations of crime. Widely labeled as "hot spots," such areas are often targets of increased manpower from law enforcement agencies in an effort to reduce crime. Where resources are concerned, the identification of hot spots is helpful because most police departments are understaffed. As such, the ability to prioritize intervention through a geographic lens is appealing.[3,10]

Operationally, the delineation of hot spot boundaries is somewhat arbitrary. As Levine (1999a) notes, crime density is measured over a continuous area. What does become clear in previous work on hot spot detection is that combining cartographic visualization of crime events with statistical tools provides valuable insight for detecting areas of concern.[10,11] Results of the Crime Mapping Research Center (1998) study suggest that good approaches for detecting hot spots are tests of spatial autocorrelation that are implemented in a variety of statistical packages.

Address for correspondence:
Saleh Ibrahim Musa
salehibro22@gmail.com

The purpose of this article is to explore and provide a visual representation of the crime rates of some selected cities in Nigeria graphically in a two-dimensional Multidimensional scaling (MDS) space since visual inspection can greatly aid interpretations.[12]

Given an input distance matrix of some selected Crime Rates of selected cities in Nigeria, this study hope to accomplish the following objectives; to give a visual representation and distribution of the crime rate of some selected cities in Nigeria graphically in a two dimensional MDS space and to show how Cluster analysis can be used in conjunction with MDS analysis.

MATERIALS AND METHODS

After multivariate data are collected, observations are grouped using cluster analysis. MDS analysis is then used on the resulting groups to discover the linear structure of the measures.

MDS

This encompasses a collection of methods that allow gaining insight in the underlining structure of the relation between entities by providing a geometrical representation of these relations.

In general, the goal of the analysis is to detect meaningful dimensions that allow the researcher to explain observed similarities or dissimilarities (distances) between the investigated objects.

MDS Basic Algorithm

1. For N items, there will be M=N (N-1)/2 similarities (distances) between pairs of different items. These similarities constitute the basic data.
   Assuming no ties, the similarities can be arranged in a strictly ascending order as

   $$S_{i1k1} < S_{i2k2} < \ldots\ldots < S_{imkm} \tag{1}$$

   (That is the distances are ordered from largest to smallest. if similarities (distances) cannot be computed, the rank orders must be specified). Si1k1 is the smallest of them similarities.
2. We want to find a q-dimension configuration of the N items such that the distances, $d^{(p)}_{ik}$ between pairs of items match the ordering in 1 above, and this is done, a perfect match occurs when $d^{(p)}_{i1k1} > d^{(p)}_{i2k2} > \ldots\ldots > d^{(p)}_{imkm}$
   That is, the descending ordering of the distances in q dimensions is exactly analogous to the ascending ordering of the initial similarities.
3. Using the $d^{(p)}_{ik}$'s, (which are frequently determined within scaling computer programs using regression methods design to produce monotone fitted distances), the points are moved around to obtain an improved configuration, which is determined by a general function minimization procedure applied to the stress, defined as;

$$\text{Stress }(q) = \left[\frac{\sum_{i=j}^{n}\sum_{j>i}^{n}(\delta_{ij}-d_{ij})^2}{\sum_{i=j}^{n}\sum_{j>i}^{n}d_{ij}^2}\right]^{1/2} \tag{2}$$

A new configuration will have new $d^{(p)}_{ik}$, and smaller stress. The process is repeated until the best (minimum stress) representation is obtained.

Plot minimum stress (q) versus q and choose the best number of dimensions from an examination of this plot. This idea is to find a representation of the items as points in q-dimensions such that the stress is as small as possible. Kruskal suggests the stress be informally interpreted according to the following guidelines;

Table 1 shows the goodness of fit at different percentage level of stress, the higher the stress level the poorer the goodness of fit and the lower the stress level the better the goodness of fit.

Cluster Analysis

- Crime mapping and analysis have evolved significantly over the past 30 years. In the beginning, many agencies utilized city and precinct maps with colored pins to visualize individual crime events and crime-plagued areas. Today, with the rapid advancement of technology, computer-based techniques for exploring, visualizing, and explaining the

Table 1: Goodness of fit and percentage of stress

Stress (%)	Goodness of Fit
20	Poor
10\	Fair
5	Good
2.5	Excellent
0	Perfect

occurrences of criminal activity have been essential

- What does become clear in previous work on hot spot detection is that combining cartographic visualization of crime events with statistical tools provides valuable insight for detecting areas of concern
- Statistical approaches for cluster analysis are widely available in a number of software packages, including *CrimeStat*, *SAS*, *SPSS*, *Systat*, and *SPlus.*

Steps in Doing a Cluster Analysis

A common approach to doing a cluster analysis is to first create a table of relative similarities or differences between all objects and second to use this information to combine the objects into groups.[13] The table of relative similarities is called a proximities matrix. The method of combining objects into groups is called a clustering algorithm. The idea is to combine objects that are similar to one another into separate groups.[14]

The Proximities Matrix

Cluster analysis starts with a data matrix, where objects (usually people in the social sciences) are rows and observations are columns. From this beginning, a table is constructed where objects are both rows and columns and the numbers in the table are measures of similarity or differences between the two observations.

For example, given the following data matrix:

X_1	X_2	X_3	X_4	X_5	
Q_1	x_1q_1	x_2q_1	…	…	x_5q_1
Q_2	x_1q_2	x_2q_2	…	…	x_5q_2
Q_3	x_1q_3	x_2q_3	…	…	x_5q_3
Q_4	x_1q_4	x_2q_4	…	…	x_5q_4
Q_5	x_1q_5	x_2q_5	…	…	x_5q_5

A proximities matrix would appear as follows:

O_1	O_2	O_3	O_4	O_5	
O_1	q_1q_1	q_2q_1	…	…	q_5q_1
O_2	q_1q_2	q_2q_2	…	…	q_5q_2
O_3	q_1q_3	q_2q_3	…	…	q_5q_3
O_4	q_1q_4	q_2q_4	…	…	q_5q_4
O_5	q_1q_5	q_2q_5	…	…	q_5q_5

The difference between a proximities matrix and a correlation matrix in cluster analysis is that a correlation matrix contains similarities between variables (X_1, X_2) while the proximities matrix contains similarities between observations (O_1, O_2).

The second problem is how to combine multiple measures into a single number, the similarity between the two observations. This is the point where univariate and multivariate cluster analysis separate. Univariate cluster analysis groups are based on a single measure, while multivariate cluster analysis is based on multiple measures.

DATA ANALYSIS AND DISCUSSION OF RESULTS

In this section, the data collected were presented, and the various MDS analysis was carried out and the results were also presented and discuss.

Data Input

The data used for the analysis was collected from the record book in the crime department of the Nigeria Police Force H/Q, Abuja, as seen in Table 2.[12]

Computational Approach

MDS arrives at a configuration that best approximates the observed distances. It actually moves objects around in the space defined by the requested number of dimensions and checks how well the distances between objects can be reproduced by the new configuration.

Table 2 shows the crime rates for selected Cities in Nigeria recorded from different types of crime incidents across cities and categorized as follows:

Table 3 above shows the proximities matrix for crime rates from major selected Cities in Nigeria, the cities naturally lies in a two-dimensional space, that is a nearly level part of the curved surface of the MDS with q=2 will locate these cities about as they occur with respect to crime rates distributions. Note if the distances in Table 3 are ordered from largest to smallest, the first position is occupied by $d_{(\text{Lagos, Sokoto})} = 6845.40$.

Figure 1 shows the dimension of the stress function for crime rates in selected Major Cities in Nigeria, whereas the stress level started at the high percent and sharply dropped to about 0.0005 with a later raise.

Figure 2 shows a multidimensional scaling plot for q=2 dimensions, the axes lies along the sample principal components of the scatter plot. Therefore,

Table 2: Displays the crime rates for selected cities in nigeria

State	Murder and non-negligent manslaughter	Forcible rape	Robbery	Aggravated assault	Burglary	Larceny-theft	Motor Vehicle Theft	Violent crime 1	Property crime 2	Total
LAGOS	35.80	51.10	578.50	716.90	1,712.90	3,627.20	2,119.00	1,382.30	7,459.10	8,841.40
BENIN CITY	33.10	45.10	544.60	700.20	1,077.30	3,025.00	1,596.90	1,323.00	5,699.20	7,022.20
WARRI	29.90	60.90	529.70	656.70	1,034.40	3,147.90	1,572.50	1,277.20	5,754.80	7,032.00
PORT HARCOURT	37.20	63.00	534.40	579.40	1,190.40	3,527.20	2,072.90	1,214.00	6,790.50	8,004.50
IFE	5.40	38.00	172.40	495.50	956.10	2,773.30	450.20	711.30	4,179.60	4,890.90
IBADAN	4.50	35.70	76.50	183.60	836.60	3,279.00	515.60	300.30	4,631.20	4,931.50
ONITSHA	51.10	82.30	559.30	587.90	978.70	3,097.00	718.30	1,280.60	4,794.00	6,074.60
ILORIN	8.90	54.60	108.30	315.50	546.80	1,735.70	315.20	487.30	2,597.70	3,085.00
MAKURDI	5.30	44.20	187.70	363.30	810.20	2,874.10	367.70	600.50	4,052.00	4,652.50
LAFIA	6.90	17.00	84.70	96.80	540.00	1,325.00	201.00	205.40	2,266.00	2,471.40
ABUJA	7.40	40.90	180.10	357.60	991.20	2,246.00	969.50	586.00	4,206.70	4,792.70
MINNA	5.60	38.50	133.40	249.10	986.00	2,729.50	309.60	426.60	4,025.10	4,451.70
JOS	6.20	27.40	137.90	276.20	884.80	2,659.40	316.00	447.70	3,860.20	4,307.90
KADUNA	12.70	35.80	345.40	444.90	1,204.50	2,969.20	836.60	838.80	5,010.30	5,849.10
JIMETA	4.50	10.40	66.30	73.40	731.10	1,034.00	308.40	154.60	2,073.50	2,228.10
ZARIA	8.40	32.40	86.20	364.10	1,093.50	2,683.80	235.80	491.10	4,013.10	4,504.20
BAUCHI	5.90	27.60	149.80	501.80	1,020.30	2,866.80	419.40	685.10	4,306.50	4,991.60
MAIDUGURI	6.20	25.70	115.20	343.40	703.30	2,750.20	450.00	490.50	3,903.50	4,394.00
KANO	10.60	51.40	224.80	179.70	1,157.40	3,252.80	782.60	466.50	5,192.80	5,659.30
SOKOTO	1.40	6.10	25.10	46.80	301.10	1,472.70	142.80	79.40	1,916.60	1,996.00
										100,180.60

Source: Nigeria Police Force H/Q, ABUJA. (1) Violent crimes are the offenses of murder, forcible rape, robbery, and aggravated assault. (2) Property crimes are the offenses of burglary, larceny-theft, and motor vehicle theft

Figure 1: Stress function for crime rates for selected major cities in Nigeria

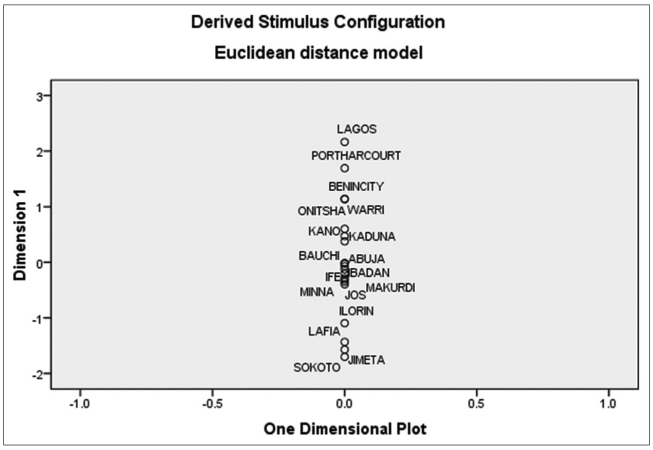

Figure 2: One dimension geometrical representation of crime rates for selected major cities in Nigeria, produced by multidimensional scaling

the plot in Figure 1 indicates q=2 is the best choice for the dimension of the final configuration.

The Figure 3 above indicates scaling pattern of crime rate of some selected cities where distances are the original distances for any two points in the input matrix.

Figure 4 shows the Scatter Plot of Linear Fit (Shepard diagram): This plot shows the reproduced distances (disparities) plotted on the vertical (y) axis versus the original similarities plotted on the horizontal (x) axis. In a perfect model, the distances and disparities for any two points are equal. Consequently, the more the scatter plot of linear fit forms a straight 45-degree line, the better the fit of the MDS model to the data, and this can be seen from Figure 4 which shows that our output data has a better fit of the MDS.

Table 3: Displays the proximities matrix for crime rates for major selected cities in nigeria

	LAGOS	BENIN CITY	WARRI	PORT HARCOURT	IFE	IBADAN	ONITSHA	ILORIN	MAKURDI	LAFIA
LAGOS	0									
BENIN CITY	1819.2	0								
WARRI	1809.4	101.4	0							
PORT HARCOURT	836.9	1200.3	1098.9	0						
IFE	3950.5	2131.3	2141.1	3113.6	0					
IBADAN	3909.9	2090.7	2100.5	3073	862.6	0				
ONITSHA	2766.8	947.6	964.2	2063.1	1183.7	1143.1	0			
ILORIN	5756.4	3937.2	3947	4919.5	1805.9	2220.5	2989.6	0		
MAKURDI	4188.9	2369.7	2379.5	3352	238.4	879.4	1422.1	1567.5	0	
LAFIA	6370	4550.8	4560.6	5533.1	2419.5	2460.1	3603.2	613.6	2181.1	0
ABUJA	4048.7	2229.5	2239.3	3211.8	152.4	710.2	1281.9	1707.7	169.2	2321.3
MINNA	4389.7	2570.5	2580.3	3552.8	439.2	732.4	1622.9	1488.1	200.8	1980.3
JOS	4533.5	2714.3	2724.1	3696.6	583	918.4	1766.7	1302.1	344.6	1836.5
KADUNA	2992.3	1173.1	1182.9	2155.4	958.2	917.6	658.1	2764.1	1196.6	3377.7
JIMETA	6613.3	4794.1	4803.9	5776.4	2662.8	2703.4	3846.5	856.9	2424.4	243.3
ZARIA	4337.2	2518	2527.8	3500.3	386.7	808.9	1570.4	1419.2	148.3	2032.8
BAUCHI	3849.8	2030.6	2040.4	3012.9	153.1	709.5	1083	1906.6	339.1	2520.2
MAIDUGURI	4447.4	2628.2	2638	3610.5	496.9	917.9	1680.6	1309	258.5	1922.6
KANO	3182.1	1362.9	1372.7	2345.2	1258	727.8	1212.9	2615.9	1274.8	3187.9
SOKOTO	6845.4	5026.2	5036	6008.5	2894.9	2935.5	4078.6	1089	2656.5	475.4

	ABUJA	MINNA	JOS	KADUNA	JIMETA	ZARIA	BAUCHI	MAIDUGURI	KANO	SOKOTO
LAGOS										
BENIN CITY										
WARRI										
PORT HARCOURT										
IFE										
IBADAN										
ONITSHA										
ILORIN										
MAKURDI										
LAFIA										
ABUJA	0									
MINNA	341	0								
JOS	484.8	186	0							
KADUNA	1056.4	1397.4	1541.2	0						
JIMETA	2564.6	2223.6	2079.8	3621	0					
ZARIA	288.5	76.5	196.3	1344.9	2276.1	0				
BAUCHI	198.9	539.9	683.7	857.5	2763.5	487.4	0			
MAIDUGURI	398.7	185.5	86.1	1455.1	2165.9	110.2	597.6	0		
KANO	1105.6	1207.6	1351.4	554.8	3431.2	1204.3	1104.9	1313.3	0	
SOKOTO	2796.7	2455.7	2311.9	3853.1	232.1	2508.2	2995.6	2398	3663.3	0

Interpretation of Results

A plot of stress (q) versus q is shown in Figure 1. Since stress (1) =2.25E-3 x 100%=0.225%, a representation of the cities in one dimension, that is, along a single axis, as shown in Figure 2 is not reasonable. The "elbow" of the stress function occurs at q=2 [Figure 1], here stress (2) =5.0E-4 × 100%=0.05%, which shows that the fit is perfect. A multidimensional scaling plot for q=2 dimensions is shown in Figure 1, the axes lies along the sample principal components of the scatter plot. Therefore, the plot in Figure 2 indicates q=2 is the best choice for the dimension of the final configuration. Which shows that cities to the far right cities such as Lagos, Port Harcounrt, Warri,

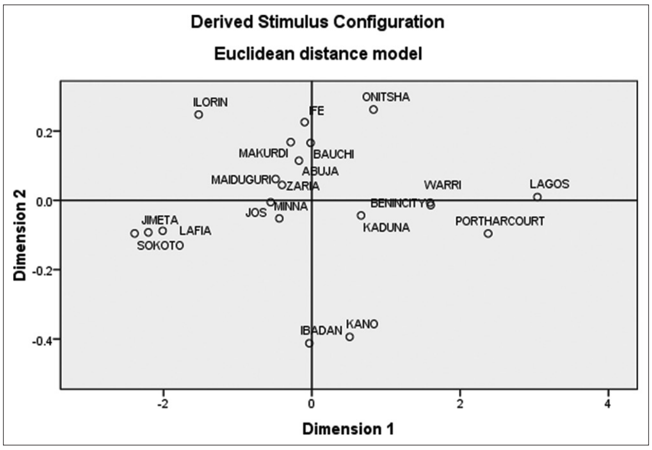

Figure 3: A two dimension geometrical representation of crime rates for selected major cities in Nigeria produced by multidimensional scaling

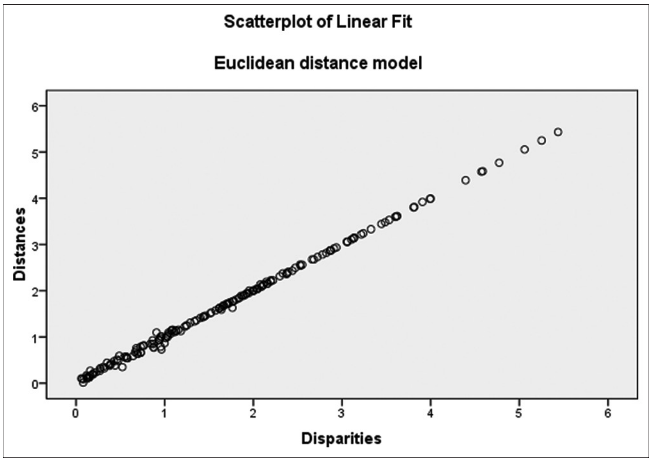

Figure 4: Scatter plot of linear fit for a two dimension geometrical representation of crime rates for selected major cities in Nigeria produced by multidimensional scaling

Benin City, Onitsha, Kaduna, and Kano which has high crime rate can be labeled as hot spots, unlike cities such as Sokoto, Jimeta, lafia, Ilorin, and Jos which has low crime rate.

From the Scatter plot of linear fit (Shepard diagram) in Figure 4, distances are the original distances for any two points in the input matrix. Since in a perfect model, the distances and disparities for any two points are equal. Consequently, the more the scatter plot of linear fit forms a straight 45-degree line, the better the fit of the MDS model to the data, and this can be seen from Figure 4, which shows that our output data has a better fit of the MDS.

CONCLUSION

Nigeria covers an area of 923,768 sq km (356,669 sq mi). At its greatest expanse, it measures about 1200 km (about 750 mi) from east to west and about 1050 km (about 650 mi) from North to South. Nigeria is bounded by Cameroon to the east, Chad to the northeast, Niger to the North, Benin to the west, and the Gulf of Guinea on the Atlantic Ocean to the South.

In 2007 Nigeria's estimated population was 135,031,164, yielding an average density of 148 persons per sq km (384/sq mi). With a birth rate of 40.2/1000 and a death rate of 16.7/1000, Nigeria's population is growing at an average of 2% annually-a rapid pace and little changed from the 1970s. The average Nigerian woman gives birth five times in her lifetime, although among more educated women the rate is somewhat lower. Nearly half of Nigerians are younger than 15 years. By 2025 the population is projected to grow to 206 million.

Fundamental to the explanation of criminal activities in a spatial context are certain environmental factors, such as the physical layout of an area, proximity to various services - all of which are likely to influence criminal behavior. Issues of access, exposure, opportunity, and the availability of targets are also important elements in helping explain crime from an environmental perspective. Not surprisingly, research indicates that certain areas are more prone to higher concentrations of crime. Widely labeled as "hot spots," such areas are often targets of increased manpower from law enforcement agencies in an effort to reduce crime.

The largest Nigerian cities are Lagos and Kano. Lagos, one of the world's largest cities, grew as colonial Nigeria's capital and leading port. Kano grew to prominence as the center of a prosperous agricultural district and as a major terminus of trans-Saharan trade. It remains a major commercial, transportation, industrial, and administrative center. Other important cities include the Yoruba centers of Ogbomosho, Oyo, and Ife; the Hausa cities of Zaria, Katsina, and Sokoto; and the newer, colonial-era cities of Kaduna, Jos, and Enugu.

Crime in Nigeria rose in the mid-1990s as a result of unemployment, economic decline, and social inequality, which are abetted by inefficient and corrupt police and customs forces. More than half of all offenses are thefts, burglaries, and break-ins, although armed robberies are also prominent. Nigeria is a major conduit for drugs moving from Asia and Latin America to markets in Europe and North America. Large-scale Nigerian fraud rings have targeted business people in other parts of the world. The business people are invited to help transfer large sums of money out of Nigeria, with the promise of a share of the transferred money. Advance fees are requested to expedite this transfer, but the advanced money routinely disappears. Although there have been periodic campaigns to root out corrupt politicians and attack crime, they have had a little lasting effect.

In this study, from the result of findings; cities such as Lagos, Port Harcounrt, Warri, Benin City, Onitsha, Kaduna, and Kano which has high crime rate can be labeled as hot spots, cause of availability of certain environmental factors. Unlike cities like Sokoto, Jimeta, lafia, Ilorin, and Jos which has low crime rate. Where resources are concerned, the identification of hot spots is helpful because most police departments are understaffed. As such, the ability to prioritize intervention through a geographic lens is appealing.

ACKNOWLEDGMENTS

We express our profound gratitude to late Danesi Jubril Hussein of Auchi Polytechnic Auchi, Nigeria for his assistance, advice, and guidance throughout this research.

REFERENCES

1. Cox MF, Cox MA. Multidimensional Scaling. London: Chapman and Hall; 2001.
2. Morrison DF. Multivariate Statistical Methods. 3rd ed. New York: McGraw-Hill; 1990.
3. Stevens J. Applied Multivariate Statistics for the Social Sciences. Hillsdale, NJ: Erlbaum; 1986.
4. Tatsuoka MM. Multivariate Analysis. New York: Wiley; 1971.
5. Anderson TW. An Introduction to Multivariate Statistical Analysis. 2nd ed. New York: Wiley; 1984.
6. Borg I, Groenen P. Modern Multidimensional Scaling. Theory and Applications. New York: Springer; 1997.
7. Tukey PA, Tukey JW. Graphical display of data sets in 3 or more dimensions. In: Barnett V, editor. Interpreting Multivariate Data. Chichester, UK: Wiley; 1981.
8. Young FW, Hamer RM. Multidimensional Scaling: History, Theory, and Applications. Hillsdale, NJ: Lawrence Erlbaum; 1987.
9. Multidimensional scaling. Wikipedia Online Encyclopedia; 2008. Available from: https://www.en.wikipedia.org/wiki/Multidimensional_scaling.
10. Kim JO, Mueller CW. Factor Analysis: Statistical Methods and Practical Issues. Beverly Hills, CA: Sage Publications; 1978a.
11. Sireci SG. Using Cluster Analysis to Solve the Problem of Standard Setting. New York: Paper Presented at the Meeting of the American Psychological Assocation; 1995.
12. Canter PR. Baltimore County's autodialer system. In: La Vigne N, Wartell J, editors. Crime Mapping Case Studies: Successes in the Field. Washington, DC: Police Executive Research Forum; 1998. p. 81-92.
13. Finn JD. A General Model for Multivariate Analysis. New York: Holt, Rinehart and Winston; 1974.
14 Timm NH. Multivariate Analysis with Applications in Education and Psychology. Monterey, CA: Brooks, Cole; 1975.