Correspondence Analysis as a Strategy to Explore the Association between Different Categories of Crime in Yobe State, Nigeria

Nicholas Pindar Dibal, Ibrahim Adamu Usman

Department of Mathematical Sciences, University of Maiduguri, Maiduguri, Nigeria

Email address: pdibal@gmail.com (N. P. Dibal), ibbiusman94@gmail.com (I. A. Usman)

To cite this article: Nicholas Pindar Dibal, Ibrahim Adamu Usman. Correspondence Analysis as a Strategy to Explore the Association between Different Categories of Crime in Yobe State, Nigeria. *American Journal of Theoretical and Applied Statistics*. Vol. 7, No. 3, 2018, pp. 118-125.
doi: 10.11648/j.ajtas.20180703.14

Received: February 22, 2018; Accepted: March 20, 2018; Published: April 13, 2018

Abstract: With the rapid increase in population at both rural and urban areas and the increasing rate of unemployment, criminal activities in Nigeria have increased dramatically in dimension, frequency and sophistication. The study examine the relationships between crimes committed by criminals that were caught and reported to the police by individuals or apprehended by the police and their offences documented in police stations across Yobe State for the period of ten years using correspondence analysis. The crimes analyzed were; murder, robbery, rape, theft, house breaking, kidnapping, grievous hurt and wounding, assault, store breaking, fraud and cheating. The analyzed complex relationships between the different crimes and the local government areas showed significant relationships between the different categories of the variables. The gained information is useful in developing action plans for identified high-risk areas, building a strong database of habitual offenders and, building crime and intelligence database for national security, law enforcement agencies and the business community.

Keywords: Crime, Offenders, High-Risk Areas, Correspondence Analysis, National Security, Law Enforcement Agencies

1. Introduction

Over the years, Nigeria has witnessed dramatic increase in crimes rate with high level of sophistication particularly in high density areas like Lagos, Port Harcourt, Warri, Kano and other places characterized by rapid population growth. Crime is a moral threat which has been with man since creation, and have posed great threats and challenges to individuals, the society, properties and constituted authorities [17]. It has been reported globally that in every five (5) years, up to 60 percent of city dwellers have reported being victims of at least one type of crime or the other while over half of these cases involve personal crimes such as; arson, rape, robbery, burglary, cheating, fraud and others [3]. Crime does not have boundary nor does it depend on developmental status of communities or countries, it varies only in frequency, dimension and sophistication. Criminal records show that multiple crimes are committed by individuals making it difficult for proactive approaches in fighting crimes due to complexities.

In most cases, crimes results as a combination of complex interactions between several intervening variables of human behavior such as; genetics, environmental, social, economic, political, psychological and cultural factors. It has been observed that different types of crimes are committed by different individuals, at different times, in different places, and under different circumstances [4]. Violations of norms or what are considered as crime varies depending on locations where they were committed, since norms does not have a universal definition [26]. Crimes varies in types, frequency and dimension as some crimes are peculiar to particular countries, communities or societies, for instance; kidnapping, homicides, armed robbery and killing involving the use of armed weaponry are more common in developed countries. Suicide bombing, rape, homicides, religious extremism and massacre are widely reported in the Arab countries. Kidnapping, ethnic fighting, tribal wars, corruption, rioting, religious clashes etc., are common occurrences in Africa [20]. The diverse differences in geographical areas, population density, demographic characteristics, natural
breeding grounds for criminals by guaranteeing the much needed anonymity for personal crimes as well as organized and specialized criminal gangs to hide. Crime is inevitable in the society and cannot be totally eradicated but managed to bearable limits [6].

It has been observed that criminal activities have increased dramatically worldwide [27], and crimes such as theft, fraud, arson, forgery and cheating have become means of survival to many criminals thereby putting lives and property at greater risk. The international crime victims survey conducted from the six major world regions including Africa, Asia, central and eastern Europe, Latin America, and western Europe (1989 – 1996) shows that more than half of the urban residents have reported being victim at least once regardless of which part of the world they inhabit [1]. Property and property-related crimes have consistently dominated Nigeria’s crime scene. According to [19], the worsening crime situation and the ineffectiveness of the crime control apparatuses, Nigeria is currently facing serious security challenges. With the multiple and complex nature of crimes, good tools are required to adequately analyze the crime data. [23] suggested that due to the complexity involved in criminal investigation, collaborative approach should be promoted in crime investigation using combination of forensic methods that can convey substantial information and statistical methods to clearly map the relationships among different crimes.

Multivariate statistical analysis have made significant contributions in the analysis and interpretation of many crime data ([2, 16]). Many studies on crimes were conducted using multivariate methods; [27] investigated crime rates in Sokoto State, Nigeria using principal component analysis where they found that assault and, grievous hurt and wounding were the most frequently crimes committed in the state. [14] analyzed crime data for Katsina State, Nigeria using principal component analysis where no significant relationships among crimes against persons was found. Cluster and multivariate correspondence analysis were also used to analyze crimes reported to the Integrated Women's Aid Center in Rio de Janeiro, Brazil, specifically to investigate associations between injuries and define criteria for classifying aggressions against women [18]. Oyo State crime data was analyzed using principal component analysis, and strong positive relationship between store breaking and stealing was revealed. Other studies indicates that there is significant relationship among different categories of crimes with one crime leading to another [20]. Yobe State have high crime rates as indicated by the records, hence the need to investigate the association between the different crimes committed in the State. Correspondence analysis technique is used to analyze the crime data to measure the level of correspondence between rows (crimes) and columns (local government areas) of the data matrix.

Correspondence analysis (CA) provides visual components that explains the vital information contained in cross classified multivariate data, it is known as multi-dimensional scaling or bivariate network analysis. It is a descriptive exploratory technique for multivariate statistical analysis which allows for clear definition of the nature and structure of the relationships between qualitative variables measured in either nominal or ordinal scales ([5, 24]). Correspondence analysis produces two or three dimensional plot based on relationships among two or more row and column categories of given data, and it is more flexible in data requirements ([10, 15]). It simplifies complex data by identifying systematic relations between variables and provides detail description of every bit of information in the data yielding simple, yet exhaustive analysis [25]. The correspondence map reduces the number of analyzed space dimensions by choosing low-dimensional subspace in which the chi-square distances between points are shown with the best accuracy [13]. Distinct feature of correspondence analysis is the transformation of tables containing numerical information into graphical display of row and column points in bi-plots helping to detect structural relationships among the variable categories and objects [25]. The process uses the singular value decomposition (SVD) algorithm of the matrix decomposition with respect to specific values [22]. The graphical displays facilitate the detection of relationships and, it allows for easy and intuitive reasoning on the relationships between the analyzed categories of variables [21].
2. Methodology

Yobe State crime data was studied to investigate the relationships and associations amongst different crimes on the records of the state police command using correspondence analysis. The data consists of activities of criminals that were caught by individuals or apprehended by the police and their offences documented in police stations across Yobe State. Breaches of private law (torts and contract) are not automatically punished by the state, but are enforced through civil procedures, hence crimes with low incidence rates were considered outliers and excluded from the study. Based on the stated exclusion criteria, 684 reported cases of different crimes obtained from divisional police stations across Yobe State for ten (10) years up to 2016 was analyzed. The crimes analyzed are; murder, robbery, rape, theft, house breaking, kidnapping, grievous hurt and wounding, assault, store breaking, fraud and cheating, also incidences of multiple crimes committed by an individual were also documented and studied. The usual procedure for analyzing cross-tabulated data to determine the probability of global association between rows and columns tested by chi-square provides no further information about the significance of individual associations between row-column pairs of the data matrix. Correspondence analysis which shows not only the existence of relationships but how categories of the variables are related to each other was used [8].

2.1. Patterns of Inter-dependence

Correspondence analysis employs chi-square distances to calculate the similarities or dissimilarities between frequencies in each cells of the contingency tables. Pairs of cells having the same observed and expected values are said to be independent of each other. The frequency data in the contingency table are standardized and transformed using the chi-square distances from the row and column profiles. The correspondence analysis method reduces the complexities in the row and column profiles by creating low-dimensional representation by factoring the basic structure of the chi-square distance matrix using singular value decomposition to produce sets of row vectors, column vectors and singular values ([12]; [28]).

2.2. The Correspondence Matrix

The data is first presented in a contingency table format and the correspondence matrix Z is derived from the table where each element in matrix Z is obtained by dividing each element of the matrix z_{ij} by N, that is z_{ij}/N, where N and z_{ij} are the total number of elements and the individual elements in the contingency table respectively. Both the row and column profiles are computed by dividing the elements of Z by row and column totals respectively, i.e.

Row profiles: Z is divided by the row total i.e. z_{ij} /row total.

Column profiles: Z is divided by the column total i.e. z_{ij} /column total.

2.3. The Weighted Chi-square Distance (D)

The weighted chi-square distances are computed from the following relation;

\[ D = D_r^{-1/2} (Z - rc^T) D_c^{-1/2}, \]  

where \( r = \) row matrix, \( c = \) column matrix, \( D_r = \) diagonal matrix whose elements are the row masses and \( D_c = \) diagonal matrix whose diagonal elements are the column masses. This gives the chi-square distance (D) which is a measure of the level of relationship between row and column variables.

2.4. Singular Value Decomposition (SVD)

Problems involving square matrices can be solved using eigenvalues and eigenvectors decomposition, however for non-square matrix, singular value decomposition is used to partition the matrix D into three components; \( U_{pxk} V_{qxk} \) and \( S_{kxk} \) diagonal matrix with elements \( s_1 \geq s_2 \geq ... \geq s_k > 0 \) where k is the reduced dimension, \( s_k \) is the singular value for the \( k^{th} \) principal component, and \( s_k^2 \) is the eigenvalue of the \( k^{th} \) principal component. The Eigenvalue is a measure of the amount of variance of the original data set captured by the \( k^{th} \) principal component (pc). The computed matrix \( DD^T \) where both \( D_{pxp} \) and \( D_{qxp} \) are \( p \times p \) matrices have \( \lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p \) as eigenvalues and the singular values \( s_j \) are the \( \sqrt{\lambda_j}, \) and the determinant \( |DD^T - \lambda I| = 0 \)

2.5. Row and Column Analyses

The row/column analysis of the matrix consists of putting the row/column profiles in a multidimensional space and finding the low-dimensional subspace which comes closer to the profile points. The row/column profiles are projected onto the subspace for interpretation of the inter-profile positions. Analyzing the rows yields the columns output and vice versa, thus the maximum number of eigenvalues that can be extracted from a two-way table equals the minimum of either; \( (columns-1) \) and \( (rows-1) \) to extract all the information contained in the table. Identifying the significant correspondences is difficult when each of the categories have many elements e.g., eleven (11) different crimes spread across seventeen (17) local government areas in this study. The strongest correspondences within given dataset were measured and projected onto multidimensional space to enable better visualization and interpretation.

3. Results and Discussions

The chi-square test of independence between the two variables for rows (crimes), and columns (local government areas) is 408.3278 with p-value = 1.201765e^{-23} ( < 0.05 ) indicating significant relationship between the row and column variables, further analysis was then conducted to measure the level of associations between all categories of the row and column variables.
Acronyms
Murder, Robbery, Rape, Grievous Hurt and Wounding (GHW), THEFT, House Breaking (H/BRK), Kidnapping (KNP), Assault, Store Breaking (S/BRK), Cheating, Fraud Tarnuwa (TMW), Potiskum (PKM), Fune (FUN), Jakusko (JAK), Yunusari (YUN), Karasuwa (KRS), Machina (MCN), Nangere (NNR), Bursari (BUR), Bade (BDE), Gaidam (GDM), Damaturu (DTR), Gujba (GUJ), Yusufari (YSF), Fika (FKA), Gashua (GSH), Nguru (NGU)

Figure 1. Row profile showing relationships between pairs of crimes.

Figure 2. Column profile showing relationship between local government areas.
Figure 1 shows the relationships between rows (crimes) points, where rows with similar profiles are grouped together. Negatively correlated rows are positioned on opposite sides of the plot origin (opposite quadrants). The distance between row points and the origin measures the quality of the row points on the factor map. Crimes appearing in the same quadrant indicates the presence of relationships, and the strength of the relationship is measured by the distance between crime points and the origin, the closer the points are, the stronger the relationship. It can be seen that theft, grievous hurt and wounding, and assault are related, while strong relationships exists between; house and store breaking, fraud and cheating, cheating and robbery, and robbery and rape. Weak relationship exists between; store breaking and kidnapping, kidnapping and fraud. The results further indicate that murder is not related to any of the crimes committed in Yobe State as it is located in the negative quadrant of the graph. Figure 2 similarly shows the relationship between columns (local government areas) points where columns with similar profiles are grouped together. Negatively correlated columns are positioned on opposite sides of the plot origin (opposite quadrants). The distance between column points and the origin measures the quality of the column points on the factor map.

**Figure 3. Joint graph of row and column profiles.**

With the crimes map imbedded on the local government areas bi-plot, the relationship between crimes and places where they are frequently committed can easily be understood, these relationships makes the fight against crime much more easier and effective. The position of the column profile points shown in Figure 3 remains unchanged, while the distances of the row (crime) points from the centroid relates to the level of contribution that each row category gives the principal axes ([9, 11]). The closer arrows are (in terms of angular distance) to an axis (or to a pole thereof), the greater the contributions of the row category on that axis relative to the other axes. If the arrow is halfway between two axes, its row category contributes to both axes to the same extent. The size of the angle between pairs of crimes measures the degree of relationship or association of crimes and location, the smaller the angle the stronger the level of relationship or association. The scatter plot shows which pole of the dimension the row categories are actually contributing to, it is evident that row categories assault and, store and house breaking have significant contributions to the positive pole of the first dimension while the row category murder contributes to the two axes to the same extent. The symmetric plot shown in Figure 3 indicates the global pattern within the crime data with rows denoted by blue arrows and columns by red triangles. Distances between any pair of row or column points measures their similarity or dissimilarity. Row points with similar profiles are close on the bi-plot, and the same also holds for column points. The relationship between crime type and location shows that; murder was associated with Tarmuwa, Potiskum and Fune local government areas. Relationships also exists between robberies and, Jakusko and Yunusari local government areas, Karasuwa and Nangere are related with cheating. There are also similarities between fraud and Machina, Nangere, Karasuwa, Bursari and Bade local government areas. Also, kidnapping, store and House breakings are associated with...
Gaidam, Damaturu, and Gujba local government areas. Theft, grievous hurt and wounding, and assault are related to Yusufari and Damaturu, while rape is a common occurrence in Fika local government area.

The data plots in Figures 1-3 shows clearly the relationships and linkages within the different crime categories and the local government areas, and between the two variables. However, further confirmation on the validity of the data analysis is necessary. The eigenvalues correspond to the amount of variance retained by each axis. Dimensions are ordered decreasingly and listed according to the amount of variance explained in the solution. Table 1 shows that dimension 1 explains the most variance in the solution with 59.96%, followed by dimension 2 with 14.28%, dimension 3 with 7.87% and dimension 4 with 6.25%. Cumulatively dimensions 1, 2, 3 and 4 were able to explain 88.6% of the total variation within the original dataset. A good dimension reduction was achieved as it can be seen that the first four axes explains 88.64% of the total variation which is acceptably a large percentage.

The results presented in Table 2 shows for the first dimension that murder has the highest contribution with 64.8% to the total chi-square value and has a stronger association with the column categories (LGA) than the other row (crimes) categories. In the row contributions table, murder (73.96%) contributes the most to the inertia of component 1, while rape (37.27%), grievous hurt and wounding (15.34%); and assault (15.43%) contributes the most to the inertia of component 2. Similarly, Fune has the highest contribution with 22.14% to the total chi-square value and have stronger association with the row (crime) categories than the other column (LGA) categories. In this column contributions table, Fune (21.95%) contributes the most to the inertia of component 1 while Fika (28.04%) contributes the most to the inertia of component 2.

### 4. Conclusion

The analyzed crime data for Yobe State provides vital information on the linkages and relationships amongst different crimes, local government areas, and between the rows (crimes) and columns (LGAs). The relationships and associations reveals the groupings of individuals having tendencies to commit similar types of crimes, and the likely places where such crimes are commonly committed. The results reveals that murder is not related to any other crimes committed, that is criminals with tendencies to commit murder are not likely to engage seriously in other types of crimes. Theft, assault and, grievous hurt and wounding have strong relationships, also related are the pairs; house and store breaking, rape and robbery, while moderate relationships exists between fraud and cheating, kidnaping and fraud. Weak associations was reported between assaults, kidnaping, house breaking and cheating. The analyzed crime data shows that murder is frequently reported in Tarmuwa, Potiskum, and Fune local government areas. Robbery is commonly reported in Jakusko and Yunusari while in Karasuwa and Nangere cheating is more common. The results further indicates that there are similarities
between fraud with Machina, Nangere, Karasuwa, Bursari and Bade local government areas. Kidnapping, store breaking, and house breaking are common occurrences in Gaidam, Damaturu, and Gujba local government areas. Theft, grievous hurt and wounding, and assault are related to Yusufari, and Damaturu, while rape is common occurrence in Fika local government area.

The information gained from the data provides vital input for policy makers and law enforcement agencies to effectively control and manage crimes in the State. The synthesized data provides vital information for the identification of crime patterns, series and trends to enable policy makers take informed decisions in terms of resource allocation, enactment of relevant laws and, design and implementation of programs aimed at reducing criminal activities in the state. The developed crime map shows hot spots, disorders and incivilities making it much easier for law enforcement agencies to respond appropriately to challenges and threats. The available information obtained helps to develop action plans for identified high-risk areas, build a strong database of habitual offenders and, build crime and intelligence database for national security, law enforcement agencies and the business community. With collaborative efforts among law enforcement agencies, policy makes and data analyst, proactive measures could be put in place to effectively tackle the threats and challenges posed by criminals.

References

[1] Ackerman, W. V. and Murray, A. T. (2004). Assessing Spatial Patterns of Crime in Lima, Ohio U. S. A Cities, 21(5): 423-437.

[2] Alberi, G. (2015). CAinterprTools: An R package to help interpreting Correspondence Analysis results. SoftwareX 1–2: 26–31.

[3] Aremu, M. A. and Ahmed, Y. A. (2011). An Investigation of Security and crime in Management

[4] Dambazau, A. B. (2007). Criminology Justice. 2nd edition. Ibadan: Spectrum Books Limited.

[5] Deschamps, R. (2017). Correspondence Analysis for Historical Research with R. Postdoctoral Tutorial, University of Waterloo, Canada. The Programming Historian.

[6] Durkheim, E. (1933). The division of Labour in Society.

[7] Fajemirokun, F., Adewale, O., Idowu, T., Oyewusi, A., and Maiyegun, B. (2006). A GIS Approach to crime mapping and management in Nigeria: A case study of Victoria Island Lagos. www.0icrf.org

[8] Fellenberg, K., Hauser, N. C., Brors, B., Neutzen, A. Hoheisel, J. D., and Vingron, M. (2001). Correspondence analysis applied to microarray data. Proceedings of the National Academy of Sciences, 98, 10781-10786.

[9] Francois, H., Le, S. and Pagès, J. (2017), Exploratory Multivariate Analysis by Example Using R. 2nd ed. Boca Raton, Florida: Chapman; Hall/CRC. http://factominer.free.fr/bookV2/index.html.

[10] Gatnar, E. and Walesiak, M. (2004), Metody statystycznej analizy wielowymiarowej w bada- niach marketingowych, Wydawnictwo Akademii Ekonomicznej, Wrocław.

[11] Greenacre, M. (2013). “Contribution Biplots.” Journal of Computational and Graphical Statistics 22 (1): 107–22. http://dx.doi.org/10.1080/10618600.2012.702494.

[12] Greenacre, M. (1984). Theory and applications of correspondence analysis. London: Academic Press.

[13] Greenacre, M., and Hastie T. (1987). The Geometric Interpretation of Correspondence Analysis, Journal of the American Statistical Association. 82 (398).

[14] Gulumbe, S. U., Dikko, H. G. and Bello, Y. (2008), Analysis of Crime Data using Principal Component Analysis: A case study of Katsina State. CBN Journal of Applied Statistics. 3 (2); 39

[15] Hoffman, D. L., and Franke, G. R. (1986). Correspondence Analysis: Graphical Representation of Categorical Data in Marketing Research. Journal of Marketing Research, 23, 213-227.

[16] Kpédeko, G. M. C. and Arya, P. L. (1981). Social and Economic Statistics for Africa. George Allen and Unwin, London.

[17] Louis, S., Cookie, W. S., Louis, A. Z. and Sheldon, R. E. (1981). Human Response to Social Problems. The Dorsey Press, Illinois.

[18] Mota, J. C., Vasconcelos, A. G. G. and Assis, S. G. (2008). Correspondence analysis: a method for classifying similar patterns of violence against women. Cad. Saúde Pública. 24 (6). http://dx.doi.org/10.1590/S0102-311X2008000600020

[19] Odekonle, F. (1981) Crime and Control in Nigeria, a seminar paper presented at National Institute of Policy and Strategic Studies, Kuru, Nigeria, 24–27.

[20] Olufolabi, O. O., Akintande, O. J and Ekum M. I. (2015), analyzing the distribution of crime in Oyo State (Nigeria) using principal component analysis. Journal of mathematics. 11.

[21] Panek, T. (2009), Statystyczne metody wielowymiarowej analizy porównawczej, SGH, Warszawa, p. 247-276.

[22] Press, W. H., Teukolsky, S. A., Vetterling W. T. and Flannery B. P. (2007), Numerical Recipes: The Art of Scientific Computing (3rd ed.), New York, Cambridge University Press.

[23] Rossy, O. and Ribaux, O. (2014). A collaborative approach for incorporating forensic case data into crime investigation using criminal intelligence analysis and visualization. Science and Justice 54 (2): 146–153.

[24] Stanisz A. (2007), Przystępny kurs statystyki z zastosowaniem pakietu STATISTICA PL na przykładach z medycyny, t. 3, Analizy wielowymiarowe, Statsoft, Kraków.

[25] Storti, D. (2010). Correspondence Analysis, from Unesco. Retrieved from http://www.unesco.org/webworld/idams/advguide/Chapt6_5.htm.

[26] Sutanto, H. T. (2015). Correspondence analysis between the place where the Crimes occur and the variety of crimes in Indonesia. Proceeding of International Conference on Research, Implementation and Education of Mathematics and Sciences Yogyakarta State University, 207-214.
[27] Usman, U., Yakubu, M. and Bello, A. Z. (2012). An Investigation on the Rate of Crime in Sokoto State Using Principal Component Analysis. *Nigerian Journal of Basic and Applied Science*. 20 (2): 152-160.

[28] Weller, S. C., and Romney, A. K. (1990). *Metric Scaling: Correspondence Analysis*. Newbury Park, CA: Sage.