

## Analysis of Crime Data in the Limpopo Province

Kwabena A. Kyei

University of Venda, Limpopo province, South Africa  
Kyei61@gmail.com, Kwabena.Kyei@univen.ac.za

**Abstract:** South Africa has a very high rate of murders, assaults, rapes and other crimes compared to most countries. Most South Africans who immigrate to other countries cite crime as the major reason. Crime has become a concern for all, the police, private security industry, real estate developer, car manufacturers, businessmen, etc. There is a high demand for crime prevention; and this calls for a continuous use of new, advanced and reliable methods to prevent crime. How bad is the level of crime in Limpopo and what are the major crime types? This study uses secondary data from the 2011 Census conducted by Statistics South Africa; and tries to look at the composition of crime in the province and the variables that influence crime, in order to propose measures to tackle and minimize crime in the province. Multivariate statistical analysis has been employed, and the study shows that the following variables; the population size, number of households, youth unemployment, growth rate, and dependency ratio; have a positive influence in the occurrence of crime. The study recommends the slowing down of population growth rate, decrease in household size and the reduction of youth unemployment to curb crime in the province.

**Keywords:** *Crime type, principal component analysis, discriminant analysis*

### 1. Introduction

Crime is rampant and very high in South Africa. The rates of murder, rape (adult, child and infant) and assault, for example, are very high compared to those of most countries. Most South Africans who immigrate to other countries cite crime as the biggest factor that influences their decision to leave the country. Crime rate is often considered to be extremely important index to judge the welfare and the quality of living within a particular area. People value the safety issue when they move or decide to purchase a real estate and maybe to relocate for better career opportunities. We can already imagine many different causes of crime and in fact many studies have been conducted all over the world to strategize how to bring down criminal activities. It is a constant endeavour of governments and policing organizations all over the world to bring down crime rates so that the world becomes a safer place to live in. The fight against crime is not a new one to humanity and security services have, since the establishment of society, tried to bring crimes down. The South African Police Service is responsible for managing 1115 police stations across the country which is fighting against crime in the country (Statistics South Africa, 2011, 2013). This study intends to look at the composition of crime in the Limpopo province and the variables that influence crime, in order to propose measures to minimizing crime in the province. The study aims not only to make the police departments have a sense of dangerous municipalities in order to pay more attention to them, but also to assist the local governments in finding out what variables that they need to manipulate in order to make the province a better and safer place to live in.

### 2. Literature Review

Crime is a global crisis affecting almost all countries but with different intensity and tempo. The causes of crime have been the investigative topic of many social researchers in the past, and they have been identified to be similar across the world, although they differ from place to place (Simpson, 1998). The current economic turmoil has led governments at all levels to reflect on their approaches to fight crime. According to studies by the Institute for Security Studies (ISS) in 2001, crime in South Africa began to increase in the mid-1980s. Since the 1990s and especially since 1996, the increase has been dramatic. Violent crimes increased at a greater rate than the rest of the crimes. The trends of crime in South Africa's major cities have kept on increasing since 1994. When we compare crime rates among cities, Johannesburg has the highest volume of serious crime, followed by Pretoria, Cape Town and Durban. Crime levels in all these urban centres with the exception of Johannesburg increased between 1994 and 1999 (ISS, 2001). Youth unemployment, poverty and proliferation of guns are cited as major contributors of crime. But according to the Centre for the Study of Violence and Reconciliation' survey (CSV) in 2006, unemployment was regarded as the greatest

concern(33%) with crime in second place (30%), HIV/AIDS in third (15%) and poverty in fourth place (9%). On the other hand, among people in the highest income group (household income of more than R8 000 per month (US\$600)), however, crime was regarded as the highest priority among 42%, with unemployment regarded as the highest priority by 24% (CSV, 2008, 2009).

The rate of murder is increasing. For the first time in 20 years the number of murders and the murder rate has increased for a second consecutive year (CSV, 2009; Statistics South Africa, 2013).The murder rate is regarded as one indicator of a country’s stability – the higher it is, the less stable a country is regarded to be. [The murder rate refers to the number of people who are murdered per 100 000 of the population.]A recent editorial in the Palm Beach Post stated that pressure is on the Criminal Justice Commission (CJC) to demonstrate that cherished programmes, such as the youth empowerment centres, are effective in reducing crime (International Centre for the Prevention of Crime, (ICPC) 2010). Crime prevention implies programmes to train and incapacitate young people with some skills because an adage goes like: “an idle hand causes damage”. Crime prevention has been described as “any initiative or policy which reduces, avoids or eliminates victimization by crime or violence (Homel, 2005; CSV, 2009). It includes governmental and non-governmental initiatives to reduce the fear of crime as well as lessen the impact of crime on victims” (ICPC, 2010).

**Crime** can be prevented through social prevention, situational crime prevention and legal sanction strategies, among others. Social prevention involves neighbours forming organised watch-dog programmes; situational crime prevention involves measures that tighten access control and surveillance making the opportunities for criminals to work extremely hard, and the risk for them to be apprehended equally very high. Legal sanctions, on the other hand, aim at incarcerating offenders with long-term sentences to serve as a deterrent effect (Schlossman et al., 1984; Clarke, 1977; Homel, 2005). There is clear evidence that well-planned crime prevention strategies prevent crime and victimization, and equally promote community safety and contribute to the sustainable development of countries. Effective crime prevention enhances the quality of life of all citizens. It has long-term benefits in terms of reducing the cost associated with the formal criminal justice system, as well as other social costs that result from crime. Crime prevention offers opportunities for a humane and more cost-effective approach to the problems of crime (Clarke and Homel, 1997; UN, 2002).

### 3. Methodology

Data based on 13 variables that fall under several general characteristics and categories such as population characteristics, economic characteristics, social characteristics, and housing characteristics all from the 2011 Census have been used (Statistics SA, 2012, 2013).The variables are: population size (POP), population density (POPDEN), number of households (NOH), unemployment rate (UMPR), youth unemployment rate (YUMPR), growth rate (GRTR), youth (YTH), Working age (WKA), elderly (ELDERLY), sex ratio (SRT), average household size (AHS),female headed households (FHH) and dependency ratio (DPR).The dataset totally has 25 entries; each entry represents information of a particular local municipality in the Limpopo province.

**Table 1: Glance of the first five entries of the dataset**

<b>Local municipalities</b>	<b>CRIME rate</b>	<b>POP</b>	<b>YTH %</b>	<b>WKA %</b>	<b>ELDERL Y %</b>	<b>DPR</b>	<b>SEX RATIO</b>	<b>GRTR %</b>
Polokwane	1.6828	628999	30.1	64.8	5.1	54.3	92.5	2.13
Thulamela	1.9729	618462	35.2	58.8	6	70.1	82	0.62
Musina	0.4658	68359	28.7	69.2	2.6	44.5	101.9	5.53
Molemole	0.1359	108321	35.6	56.9	7.6	75.9	85.4	-0.1
Modimolle	0.6786	68513	30.8	63.8	5.4	56.8	102.5	-0.07

**Sampling adequacy:** Before using any statistical procedure, there is always a need to test the relevance of such procedure for the data in question. Factor Analysis (FA) or precisely, Principal Component Analysis (PCA) has been used in this study therefore Table 2 tests its suitability. Table 2 shows the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and the Bartlett’s test of sphericity. The KMO statistic varies between 0

and 1. When it is zero then the sum of partial correlations is large relative to the sum of correlations, which means there is some diffusion in the pattern of correlations and therefore the data is likely to be inappropriate for Factor Analysis (Everitt and Dunn, 2001; Everitt and Hothorn, 2010).

**Table 2: KMO and Bartlett's Test**

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.	.730
Bartlett's Test of Sphericity	Approx. Chi-Square
	707.254
	Df
	78
	Sig.
	.000

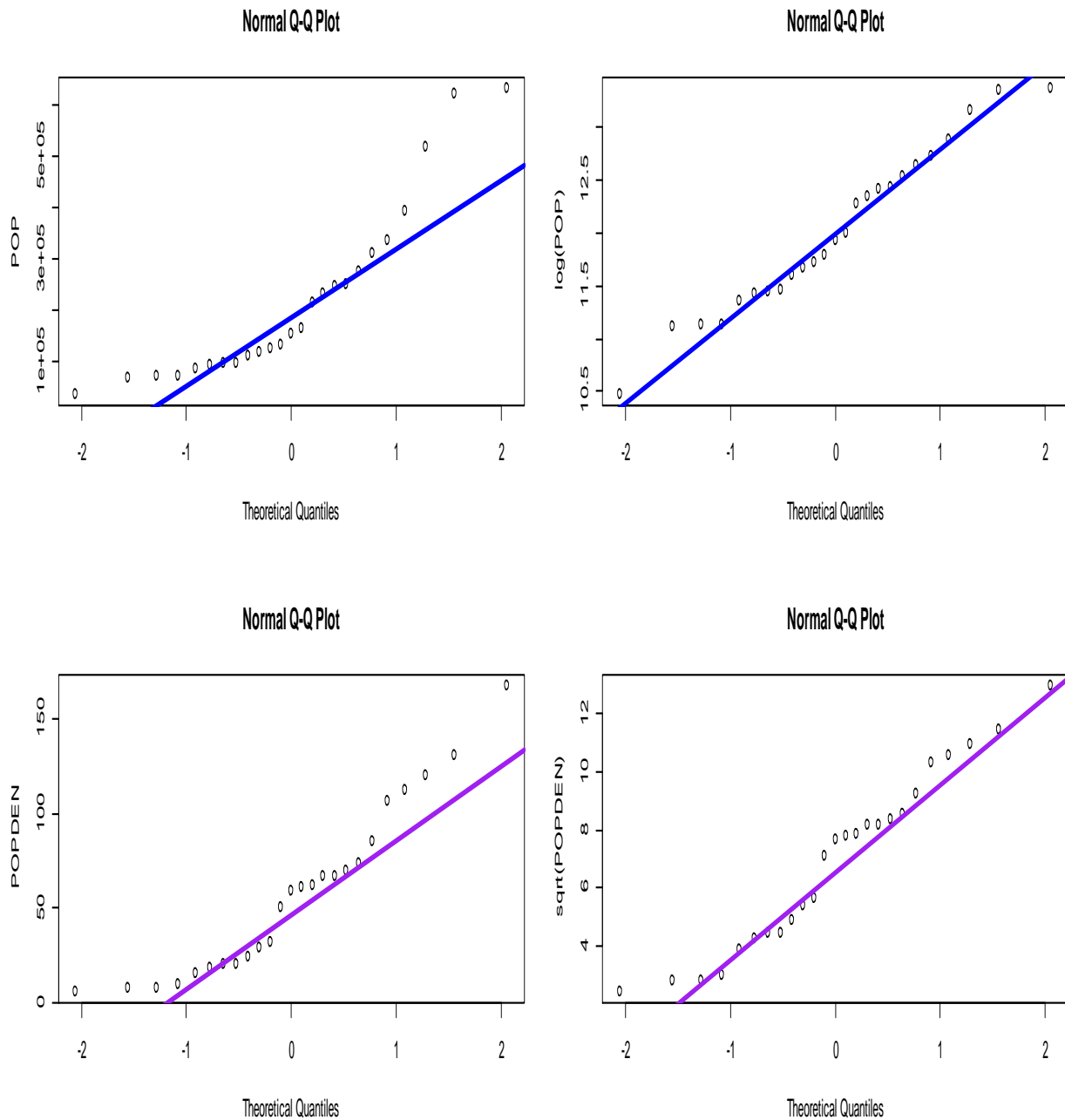
On the other hand when KMO is close to 1 then the patterns of correlations are relatively compact and so Factor Analysis should yield distinct and reliable factors. It is recommended that values greater than 0.5 are acceptable. Values between 0.5 and 0.7 are mediocre, between 0.7 and 0.8 are good, between 0.8 and 0.9 are great and those greater than 0.9 are superb (see Hutcheson and Sofronion, 1999, pp. 224-225). For these data, the value is 0.730, which falls into the range of being good, so we should be confident that FA or more precisely, PCA is appropriate to be used. Distribution assumptions' verification is crucial before doing any statistical modelling. Multivariate normal distribution is one of the most frequently made distributional assumptions when using multivariate statistical analysis or techniques like Principal Component Analysis and Discriminant Analysis. We know that if  $X=(X_1, X_2, \dots, X_p)$  follows the multivariate normal distribution, then its individual components  $X_1, X_2, \dots, X_p$  are all normally distributed. Therefore, we need to examine normality of each  $X_i$  to guarantee that  $X=(X_1, X_2, \dots, X_p)$  is normally distributed (Everitt and Hothorn, 2010; Johnson and Wichern, 2014).

**Table 3 New Variables which are after transformation**

Abbrev	Variables	Transformed Variables
POP	Population	Log(POP)
YTH	Young(0 - 14)	Sqrt(sqrt(YTH))
WKA	Working age(15 - 64)	log(WKA)
ELDERLY	Elderly(64+)	sqrt(ELDERLY)
DPR	Dependency ratio	sqrt(DPR)
SRT	Sex ratio	log(log(SEX.RATIO))
GRTR	Growth rate	log(GRTR)
POPDEN	Population dependency	log(POPDEN)
UMPR	Unemployment rate	Log(sqrt(UMPR))
YUMPR	Youth unemployment rate	log(YUMPR)
NOH	Number of households	sqrt(NOH)
AHS	Average household size	log(AHS)
FHH	Female headed household	sqrt(FHH)

Here, the quantile - quantile plot (QQ plot) has been used to assess normality of data. In QQ plot, we compare the real standardized values of the variables against the standard normal distribution. The correlation between the sample data and normal quantiles measures how well the data is modelled by normal distribution. For normal data, the points plotted should fall approximately on a straight line in the QQ plot. If not, data transformation is applied to make the data to appear more closely normally distributed. For transformation, we can use a variety of methods including the logarithm, square root, power transformation, and/or scale function (Bartholomew et al., 2008). Our QQ plot of each variable revealed that all the 14 variables do not follow normal distribution. We then try different forms of transformation on all the variables to obtain the substitute variables which perform better on normality. **Table 3** lists the specific forms used to get the new variables. **Figure 1** is the QQ plot of POP and POPDEN before- transformation and after-transformation. We can see the effectiveness of data transformation method because data tends to be normally distributed after transformation (Everitt and Hothorn, 2010; Johnson and Wichern, 2014).

Figure 1: Normality check samples



In general, a large number of variables in any study make it difficult to do analysis because of multicollinearity between variables (Fajnzylber et al., 2002; Yan, 2011).

Figure 2: Multicollinearity

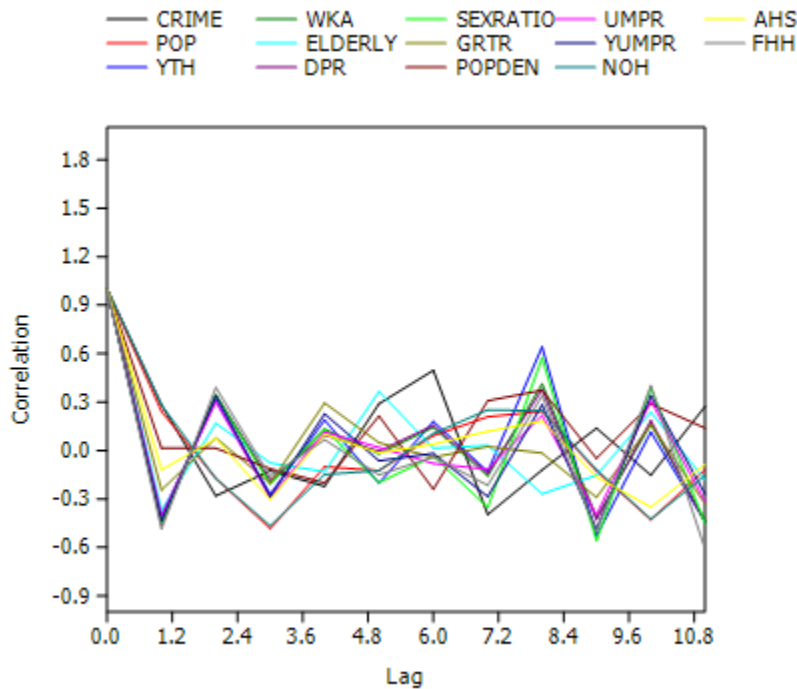


Figure 2 shows the multicollinearity between the variables. Therefore, we introduce Principal Component Analysis in order to rotate the variables matrix  $X=(X_1, X_2, \dots, X_p)$  to achieve orthogonality to decipher patterns more easily and at the same time reduce the dimension of the data for simpler data process (Everitt and Hothorn, 2010). With the assistance of PCA, we obtain a clear pattern of our municipalities profile data, without much loss of variables to catch up with the whole data information.

#### 4. Results

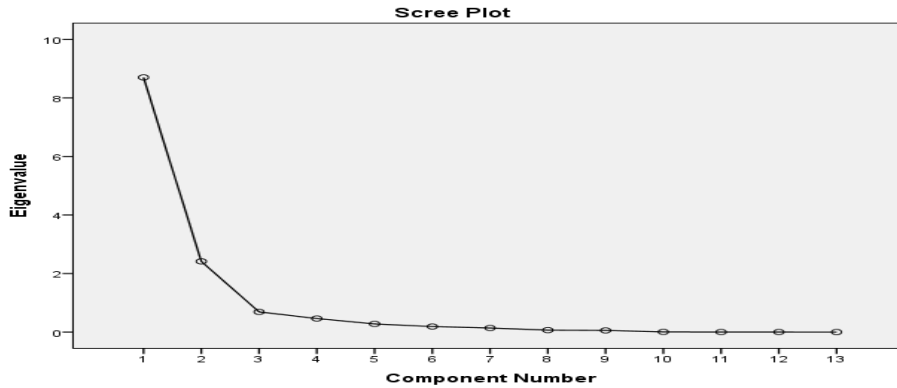
Table 4: Total Variance Explained

Component	Initial Eigen values			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	8.700	66.926	66.926	8.700	66.926	66.926	8.189	62.993	62.993
2	2.414	18.572	85.497	2.414	18.572	85.497	2.926	22.504	85.497
3	.689	5.299	90.796						
4	.462	3.555	94.351						
5	.278	2.135	96.487						
6	.189	1.450	97.937						
7	.139	1.068	99.006						
8	.066	.510	99.515						
9	.056	.429	99.944						
10	.005	.040	99.984						
11	.001	.010	99.994						
12	.001	.006	100.000						
13	5.582E-5	.000	100.000						

Extraction Method: Principal Component Analysis.

The following tables and figures give the results from the principal component analysis. Only two factors were retained, the first could be called *ratio* because it consists of rates or percentages and the second factor called *size*. Table 4 shows that factor 1 explains 66.9% of the variance while factor 2 explains 18.6% and the two factors account for over 85% of the variation. Figure 3 shows the scree plot which proves that indeed only two factors are eligible to be retained.

**Figure 3: Screen plot**



Earlier on, we mentioned three methods which can be used to determine the number of factors (PCs) to retain; above is the Kaiser’s criterion which says we should retain only the factors (PCs) with eigenvalues greater than 1. Figure 3 is a scree plot, the largest change in the slope occurs on PC3, which is the “elbow” of the plot, thus we should retain the first two PCs (Everitt and Hothorn, 2010; Johnson and Wichern, 2014).

**Interpretation:** Table 5 shows that the variables that load highly onto factor1 are WKA%, DPR, YTH%, FHH, AHS, UMPR%, YUMPR%, SEX RATIO, ELDERLY% and GRTR%. We call factor1 **ratio**; whereas, the variables that load highly onto factor2 are NOH, POP and POPDEN. We call factor 2, **size**.

**Table 5: Rotated Component Matrix<sup>a</sup>**

	<b>Component</b>	
	<b>1(ratio)</b>	<b>2 (size)</b>
WKA %	-.981	
DPR	.981	
YTH %	.944	
FHH %	.928	
AHS	.890	
UMPR %	.883	
YUMPR %	.880	
SEX RATIO	-.877	
ELDERLY %	.848	
GRTR %	-.751	
NOH		.978
POP		.973
POPDEN		.826

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.<sup>a</sup>  
 Rotation converged in 3 iterations.

Figure 4: PCA plot

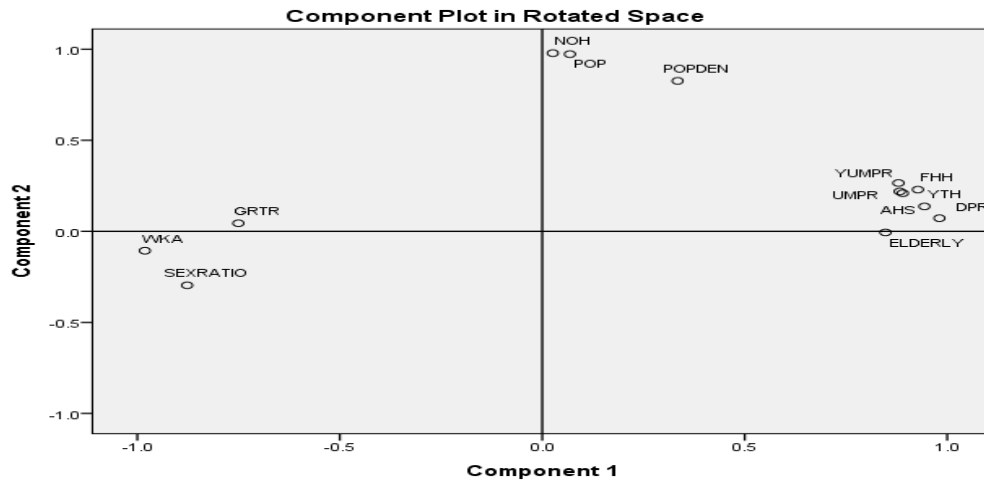


Figure 4 shows that GRTR%, WKA% and SEX RATIO are negatively correlated to the rest of the variables that load onto factor1 (ratio) with WKA% and SEXRATIO also negatively correlated to GRTR%. Table 6 shows the extraction method from PCA.

Table 6: Communalities

	Initial	Extraction
POP	1.000	.951
YTH %	1.000	.909
WKA %	1.000	.974
ELDERLY %	1.000	.718
DPR	1.000	.967
SEX RATIO	1.000	.857
GRTR %	1.000	.565
POPDEN	1.000	.794
UMPR %	1.000	.827
YUMPR %	1.000	.844
NOH	1.000	.958
AHS	1.000	.836
FHH %	1.000	.913

Extraction Method: Principal Component Analysis.

**Discriminant Analysis (DA) on Principal Components:** Since our PCs capture the majority of the information from the original data, and at the same time, the number of PCs are much less than that of original variables, it is easy to think of using the factors computed from PCA as input factors for DA algorithm. The new factors which are a linear combination of the original variables have the following advantageous properties:

- The interpretation of them allows us to detect patterns in the initial data space.
- Reduce very large factors into a smaller number of factors; hence we can remove noise from the dataset by using only the most relevant factors.
- Algorithms such as Linear Discriminant Analysis (LDA) could have a better behaviour because PCs come from an algorithm basis.
  - Principal Components Analysis is successfully applied into our dataset by extracting two PCs out of the 13 original variables, which implies a great dimensional reduction. In addition, these PCs account for 85.497% variance of the original dataset, thus we did not lose much information. As mentioned earlier Factor1 consists of WKA%, DPR, YTH%, FHH, AHS, UMPR%, YUMPR%, SEXRATIO, ELDERLY and GRTR% which are highly correlated with each other; and we call that

- Factor1 the ratio. The Component Plot shows that GRTR%, WKA% and SEXRATIO are negatively correlated to the rest of the variables in factor1 with WKA% and SEXRATIO also negatively correlated to GRTR%.
- Factor2 consists of NOH, POP and POPDEN, we call Factor2 the size. These two factors can be used to classify a municipality as a safe one or unsafe one.
  - Discriminant Analysis is applied to classify a Local Municipality by the composition of crime even if there are no data records of crime. For our case, we applied Linear Discriminant Analysis on the original variables and to the Principal Components respectively (Bian, 2005).
  - We used the unstandardized coefficients to construct the actual prediction equation which can be used to classify new cases (Johnson & Wichern, 2007). Our analysis gives the following model:

$$D_i = -1.877 + 2.981\text{factor1} - 0.012\text{factor2} + \epsilon$$

From the model we can see that to minimize the value of  $D_i$ , we have to increase more of factor2 which is **size** and minimize factor1 which is **ratio**

**Discussion:** The foregoing analysis shows that population size influences crime. Population increase has to be considered seriously because it triggers a quantum effect on the society with negative consequence. It leads to the creation of more people with some form of frustration or resentment towards the society such that they end up engaging in criminal activities, (Statistics SA, 2011, 2013). Increased population leads to congestion (excess population), competition and jealousy; and this is one of the biggest causes of crime and much of the challenges that the world faces. Reduction in the number of households and population density are seen as important factors to fight and bring crime down. Sex ratio is seen to be negatively related to elderly (advanced in age) because when people advance in age, usually the men die faster than the women (their wives) therefore decreasing the sex ratio. The questions again are, what factors may be related to safety? Are there some variables that we can observe to predict the safety and security of communities? What are the strategies needed to reduce crime rates? These are some of the questions that have prompted this study.

## 5. Conclusion

The study shows that several variables have an influence on crime. Notably, the population size, population density, number of households, unemployment rate, youth unemployment rate, growth rate, and dependency ratio have a positive influence in the prevalence of crime. These variables are all related to the population size of a particular place. Crime increases when population increases, therefore the local authorities should take note of that. Excess population results in high youth unemployment in a shrinking economy like South Africa's, therefore to control crime in the Limpopo Province in particular and in the country as a whole, local, provincial and national governments have an obligation to fight crime by creating job opportunities to reduce unemployment, especially that of youth, and also provide educational facilities to the youth to equip them with skills so that they can get employments or they can be busy doing some trade. The government can also fight crime by increasing the severity of punishment for offenders. Provide educational centres where people are taught about crime and how they can protect themselves from being victimized.

**Limitation of the study:** The variables chosen are based on the historical study of possible factors causing crime, besides the availability and accessibility of the variables. If data resources were sufficient, additional variables such as alcohol-drug consumption, educational factors and other factors could have been considered in this analysis.



## References

- Bian, H. (2005). SPSS Discriminant Function Analysis. Office for Faculty Excellence, Houston.
- Bartholomew, D. J., Steele, F., Moustaki, I., Galbraith, L. (2008). *Analysis of Multivariate Social Data; (2<sup>nd</sup> edition)*.CRC Press.
- Centre for the Study of Violence and Reconciliation (CSVR).(2008). Adding insult to injury: How exclusion and inequality drive South Africa's problem of Violence, Johannesburg: CSVR.
- Centre for the Study of Violence and Reconciliation (CSVR).(2009). Why does South Africa have such high rates of violent crime? Johannesburg: CSVR.
- Clarke, R. V. (1997). Introduction. In RV Clarke (ed.): *Situational Crime Prevention: Successful Case Studies*. Guiderland, NY: Harrison and Heston.
- Clarke, R. V. & Homel, R.(1997). Revised Classification of Situational Crime Prevention. In S. P. Lab (ed.): *Crime Prevention at a Crossroads*. Cincinnati, OH: Anderson.
- Demombynes, G. & Özler, B. (2005).Crime and local inequality in South Africa. *Journal of Development Economics*, 76(4), 265-292.
- Everitt, B. S. & Dunn, G. (2001).*Applied Multivariate Data Analysis*(2<sup>nd</sup> edition). Arnold.
- Everitt, B. S. & Hothorn, T. (2010). An introduction to Applied Multivariate Analysis using R. Springer.
- Fajnzylber, P., Lederman, D. & Loayza, N. (2002). What causes violent Crime? *European Economic Review*, 46, 1323-1357.
- Homel, R. (2005). Developmental crime prevention. In Nick Tilley (Ed.), *Handbook of crime prevention and community safety* (pp. 71-106). Cullumpton, Devon, UK: Willan Publishing.
- Hutcheson, G. & Sofroniou, N. (1999). *The multivariate social scientist: introductory statistics using generalized linear models*. London: Sage Publication.
- Institute of Security Studies (ISS).(2001). South African Crime. *The South African Crime Quarterly (SACQ) Journal*.
- International Centre for the Prevention of Crime (ICPC) (2010). International Report 2010 on Crime Prevention. Available at [www.crime-prevention-intl.org/nc](http://www.crime-prevention-intl.org/nc). Accessed on January 15, 2014.
- Johnson, R. & Wichern, D. (2007).*Applied Multivariate Statistical Analysis*. Prentice Hall Upper Saddle River, NJ.
- Johnson, R. & Wichern, D. (2014).*Applied Multivariate Statistical Analysis*. Pearson Education Limited, London.
- Schlossman, S., Zellman, G., Shavelson, R. & Cobb, J. (1984). Delinquency prevention in South Chicago: a fifty year assessment of the Chicago Area Project, Santa Monica, Calif.: Rand.
- Simpson, G.(1998). Urban Crime and Violence in South Africa. In Petty, C. & Brown, M. (eds), *Justice for Children: Challenges for policy and practice in Sub-Saharan Africa*, p. 66-71, London: Save the Children, 1998.
- Statistics South Africa (Stats SA).(2011). available at <http://www.beta2.statssa.gov.za> and <http://www.statssa.gov.za/crime>.Accessed December 13, 2013.
- Statistics South Africa (Stats SA).(2012).Income and Expenditure of households 2005/2006. Unit record, Pretoria: Statistics South Africa.
- Statistics South Africa (Stats SA). (2013).Methodological Report on the development of the poverty lines for statistical reporting. Technical report DO 300. Pretoria: Statistics South Africa.
- Yan, F.(2011). *Multivariate Methods Analysis of Crime Data in Los Angeles Communities*. University of California, Los Angeles.
- <http://en.wikipedia.org/wiki/crime> Accessed December 15, 2014.
- [http://law.jrank.org/pages/12004/causes\\_crime.html](http://law.jrank.org/pages/12004/causes_crime.html). Accessed October 12, 2013.
- [http://topyaps.com/top\\_10\\_causes\\_of\\_crime](http://topyaps.com/top_10_causes_of_crime). Accessed November 15, 2014.