

Deterring Gang Street: A Study on Logistic Analysis

Gabriel Joseph Mukungu¹
Gwangyong Gim²

¹Doctoral Student, Department of Business Administration, Soongsil University.

²Professor, Department of Business Administration, Soongsil University

Abstract— The adolescence and early adulthood period are crucial time in an individual's life. This is a time when many young people become grown up in skills, and subsequently, they start to face new challenges in their lives. In the midst of these challenges, many adolescence most likely becomes involved in gang activities.

This paper, employed the logistic analysis model as an investigation tool for Tanzanian police force to overcome the emerging gang street.

The study collected 200 data of adolescence, which consisted with gang street (criminals) and non criminals from the Tanzania police force database of which 140 were taken as the gang streets (criminal) and non criminals were 60.

By means of random sampling the results implicated that 83.0% were the overall percentage of the gang street (criminals) correctly predicted, 87.8% were sensitivity; 72.1% were specificity; 27.8% were false positive and 12.2% were false-negative rate.

Keywords—Gang street, logistic analysis model

1. INTRODUCTION

1.1. What is adolescent

An adolescent is simply a young person through the process of developing from a child into an adult. Adolescence is said to be the period during which the child achieves sexual maturity but not yet taken over the roles and responsibilities, or the rights, that be an adjunct to full adult status [1]. The adolescence is a very influential time during the life of a young person. It is a time of change and possible insecurity, accompanied by low self-esteem and emphasis on peer approval [2], [3], [4]. Some researchers, elucidate that, it is the effects of physical change, increased intellectual capacity, the development of sexual impulses, and social pressure to achieve independence [5].

Other researchers state that, it is a transition period between childhood and adulthood, gaining skills deemed necessary to live separately from one's parents or guardians encompass. During this stage of growth, the bodies change, grow and the thinking capacity becomes more concrete and logical; emotions to become more mature, and moral confidence internalized [5]. This may be the reason that so many dangerous habits are developed during the adolescence period, for instance, the gang street activities.

1.2. What is a gang street.

The term gang street can be easily defined as the group of criminals or hooligans that band together to commit crimes like robbery, armed robberies, house breaking, burglaries, money laundering, etc. These offences may be committed by gang streets either at a day or night hours. On the other side, the community needs safety and security from gang activities. Thus, the preventions against the gang street (crime) is a necessity to make people live in safer and peaceful. For that case, the criminology analysis is an important concept that can assist in crime prevention, and crime controls efforts [6].

1.3. Current situation

Tanzania crime scenario through police records in the period between 2007 to 2010 indicated that by the end with the year 2007 were (88, 527), 2008 (102, 092) and 2009 (103, 669) [7]. These figures show the slight increase of a crime rate due to many factors. One of the main factors is the traditional crime-solving tools against with the emerging gang streets.

The mainly traditional crime-solving tools used by the Tanzania police force are:

- Unfocused random motorized and foot patrols.
- Follow-up investigations by police detectives.
- Increase number of police ability to detect and arrest offenders, etc.

1.4. Statement of the Problem

The traditional crime-solving tools used to prevent gang streets (criminals) are not satisfactory to assist the law enforcers like police to overcome the emerging streets' crimes. Over time, these techniques are difficult and are achieved with limited success. Moreover, they do not solve crimes in real time and consequently; they are so highly cost [8].

1.5. Motivation of the Study

The logistic analysis model is expected to deter the gang street in real time and less cost.

1.6. Logistic Analysis model

The logistics analysis model is a classification method. It returns the probability that the binary dependent variable may be predicted from the independent variables. The logistic analysis model is a form of regression, which is used when the dependent is a dichotomy, and the independent is of any type. It predicts a dependent variable into the basis of independent's variables to determine the percent of variance in the dependent variable being clarified by the independents from the community. Logistic analysis model applies the maximum likelihood estimation after transforming the dependent into a logit variable (the natural logs of the odds of the dependent

2. LITERATURE REVIEW

2.1. Logistic analysis model

A key concept for understanding the tests used for logistic analysis model (and many other procedures using maximum likelihood methods) is that of log likelihood. The likelihood just means probability, though it tends to be used by statisticians of a Bayesian orientation. It always means the probability under a specified hypothesis. In thinking about logistic analysis model, two hypotheses are likely to be of interest: the null hypothesis, which is that all the coefficients for the regression equation take the value zero, and the hypothesis that the model currently under consideration is accurate. So, many researchers have been used a logistic analysis model to perform criminal prediction analyses [9].

2.2. The advantages of the logistic analysis model.

The logistic analysis model has several advantages of certain occurring events. One of the most important advantages of using the logistic analysis model is that it allows estimation of both relative and absolute risk properties on a risk assessment scale referred to as discrimination and calibration, respectively [10].

To make obvious, the use and advantages of this method, this study presents analyses of gang street (criminals) versus non-criminal's records obtained from the Tanzania police force database.

The logistic model is believed to be relatively robust; i.e., many types of underlying assumptions lead to the same logistic formulation [11].

Another advantage of the logistic analysis model relates to its use of an alternative to contingency table analysis. Moreover, the logistic analysis model (that is, use with the logit function which is called the canonical link function. The canonical link function parameter estimates under logistic regression with fully efficient, and tests on those parameters are better behaved for small samples.

2.3. The disadvantages of the logistic model

On the other hand, the logistic analysis model has the following disadvantages:

It is based on a special assumption on the data set [12]. The relationship between the mean of dependent variable and a set of independent variables follows a logistic distribution and that the errors are binomially distributed. In reality, however, this assumption may not be true. The assumption of a binomial distribution for the error term may be invalid due to over dispersion in the dependent variable [13].

Some data points, termed influential values, may have an undue influence on the overall fit into the model, either on the set of parameter estimates or on a single parameter estimate [14]. Sometimes, an observation may have undue influence so that including a term in the model based upon a likelihood ratio test is due solely to that observation.

Several numerical problems like zero cell count in a contingency table, collinearity between covariate, or a set of covariate that separates the dependent variable completely, and the maximum

likelihood estimates do not exist and might be encountered during the logistic analysis model is fitted [12].

Another disadvantage is when the data set contains a large number of variables of which a problem with these data over fitting may occur. Variable selection is not easy and is time consuming, especially when interactions are taken into account.

Another disadvantage is based on some software packages that do not take into account the multiple comparison problems and may select noise variables that cause a decrease in the predictive power [15], [16].

3. METHODOLOGY

3.1. Sample size

Originally, the information on the crime is recorded at any police stations followed by the immediate action of arrests for the suspect. Once, the suspect is arrested by police, then his or her details would have to be recorded and subsequently apprehended before the courts of law to prove beyond reasonable doubts. In fact, upon proofs against his or her allegations, then he or she will be convicted and jailed, and thus recorded at court and police station's databases respectively as a criminal.

For the purpose of this study, 200 data were the total number of the adolescence; gang streets (criminal) were 140 and non criminals were 60.

3.2. Data Sources

The aim of this study was to deter gang street (criminals) by using the logistic analysis model implications within the community of Tanzania. The sources of these data were taken as primary data from the Tanzania police force database to be processed by the logistic analysis model to predict the gangs' effects to the societies.

3.3. Data designing

In order the collected data to be computed by the logistic analysis model was designed in the form of dichotomous by the aid of training set and learning algorithms set to develop the suitable models. The explanatory variables were gang street (criminal) and non-criminal who were coded with 0 and 1 respectively.

Thus, the adolescent is either a gang street (criminal) or non criminal. Think of the process of randomly selecting the adolescent and recording the values of 1 or 0 and whether or not the person is the gang street (criminal).

3.4. The logistic analysis model

The gang and non-criminal were designed to be computed as dependent and covariate variables respectively.

3.5. Data processing

These data were computed on the training data to the learning algorithms to build the desirable models. The training set data, learning set data and the output results were obtained through logistic analysis to develop the model.

4. MODELS DEVELOPMENT AND HYPOTHESES

4.1. The development of models

For the purpose of this study, the goodness of data fit as well as validities were very useful for the model built hypotheses from the logistic model to predict the gang street against non criminal's respect with their time of committing the offences (i.e., night hours or day hours).

4.1.1. Logistic model

The gang (criminal) and non- criminal were computed as dependent and covariate variables respectively through training data into testing data to build desirable modules for high accuracy.

4.1.1.1. Odds

The odds are simply the ratio of the proportions for the two possible outcomes.

The precision of odds is estimated by the 95% confidence interval (CI). A large CI indicates a low level of precision of the odds, whereas a small CI indicates a higher precision of the odds.

$$\ln(\text{ODDS}) = \ln\left(\frac{\hat{y}}{1-\hat{y}}\right) = a + bx \dots \dots \dots (1)$$

Where a is the constant, b is the intercept, \hat{y} is the predicted probability to the event which is coded with 0 (to predict gang street) whereas 1 is coded to predict non- criminal and x is the predictor variable of "gang street."

From equation (1), the output variable's regression's equation model is given by $\ln(\text{ODDS}) = -1.971 + 2.922 \text{ Gang street}$ "TABLE 5-5."

4.2. Hypotheses

In regard with this study, it is hypothesized that the:

- i The logistic analysis model has the positive effect over the traditional crime- solving tools used by Tanzanian police force to predict gang street (criminals).
- ii The adolescent gang street (criminals) has positive effects with-

TABLE [5-3] MODEL SUMMARY

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	175.462 ^a	.297	.420

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001. in the Tanzanian community.

5. THE VALIDATION OF LOGISTIC MODEL.

The training data and learning algorithm were mainly used for building the model to verify the accuracy of the predictive data test. When the input variables are many then, the possible ways of combining them is necessary in order to build the model "fit" or the variable data of this method should be over trained. In other words, the model should reflect natural noise to the input data more than actual statistical between the input and output data predictions. To overcome this problem, it is adversely to split these data as set into training set and a testing set so that the output variable will be well-

known for each set [14]. However, this method fitted the model to the training data where the model was then applied in the data tests to observe whether it had about the same predictive accuracy as on the training data or not. This effect gave the accuracy of the model being generated.

5.1. Results of hypotheses

The analyzed results with their corresponding outputs were resulted from the logistic analyses.

5.1.1. Hypothesis 1

The logistic analysis has the positive effect over the traditional crime- solving tools used by the Tanzanian police force to predict gang street (criminals).

TABLE [5-1] VARIABLES IN THE EQUATION

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 0 Constant	-.824	.154	28.757	1	.000	.439

The intercept in this model is $\ln(\text{odds}) = -.824$. If both sides are exponentiated, then the predicted odds $[\text{Exp}(B)] = .439$. That is, the predicted gang street is .439 "TABLE 5-1."

TABLE [5-2] OMNIBUS TESTS OF MODEL COEFFICIENTS

	Chi-square	df	Sig.
Step 1 Step	70.555	1	.000
Block	70.555	1	.000
Model	70.555	1	.000

The Omnibus tests of model coefficients give a Chi-Square of 70.555 on 1df, significant beyond .000 "TABLE 5-2." This is a significant test of the null hypothesis. Adding the gang street variable to the model has not significantly increased because $p < 0.05$.

From "TABLE 5-3," indicated that, the -2 Log Likelihood statistic is 175.462. This meant that the statistic was a good model to predict the decisions since the smaller statistic the better the model. The Cox & Snell R^2 can be interpreted like R^2 in multiple regressions, but cannot reach a maximum value of 1. The Nagelkerke R^2 can reach a maximum of 1. In other words, the Nagelkerke R^2 should always be greater than Cox & Snell R^2 .

TABLE [5-4] CLASSIFICATION TABLE^a

	Observed	Predicted			
		Gang street		Percentage Correct	
		Yes	No		
Step 1	Gang street	Yes	122	17	87.8
		No	17	44	72.1
	Overall Percentage				83.0

a. The cut value is .500

From the computation of 200 sample size, "TABLE 5-4" indicated 139 as gang street of which 122 from the prediction were correctly occurred i.e. they have a fitted probability (calculated from our model) of less than 0.05. Similarly; 44 out of 61 were not correctly predicted as the gang street.

In this study 166/200 = 83%) were correctly classified as the overall percentage of the gang street for the better the model.

5.1.2. Hypothesis 2

The time for crime commission has the positive effects over the Tanzanian community.

Moreover, based on error rates in classification "TABLE 5-4" showed that;

The decision rule predicted gang street 61 times; however, correctly not all occurred.

- The false- positive results predicted to occur when, in fact, did not occur. The prediction was wrong 17 times, for a false- positive rate of $17/61 = 27.8\%$.

- The decision rule again did not predict 200 times, but some of them correctly occurred.

The false- negative did not predict to occur when, in fact, it did occur. The false- negative prediction was wrong 17 times, for the false- negative rate of $17/139 = 12.2\%$.

These results are simply known as error rates in classification.

TABLE [5-5] VARIABLES IN THE EQUATION

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for	
							EXP(B)	
							Lower	Upper
Step 1 ^a Gang street	2.922	.385	57.460	1	.000	18.574	8.726	39.538
Constant	-1.971	.259	57.954	1	.000	.139		

a. Variable(s) entered on step 1: Gang street.

The "TABLE 5-5" shows that, the Wald statistic with the chi-square significance values less than .05 for gang street enough to reject the null hypothesis as the variable does make a significant contribution.

The estimated values yield the model:

The 95% confidence interval for exp(B) was from 8.726 to 39.538 indicating that the gang streets (criminals) are between 8.726 to 39.538 times as likely to commit crimes than non- criminals.

6. CONCLUSION

From this study, it was observed that the logistic analysis model had used to analyze the calculated probabilities, odds and relative odds from the collected data. Some theories were introduced to show on how the logistic analysis model framework was good way to investigate some associations in multi-way dimensions of the two categories. Apart from these analyses, this study has some limitations followed by the discussions and recommendations.

6.1. Limitations of the study

Among the limitations from this study's finding are listed below:

One of the main limitations in this study was based upon the application of logistic analysis by law enforcers like police, whether it is effectively used for investigation purposes or not and merely used for exchange the intelligence information about the gang streets' prevention among the police departments.

Another limitation was based upon the sample of data used in this study. The sample of data was small and merely collected from police nowhere else.

Moreover, the method used in this test was only based on the logistic analysis to predict gang street (criminal), and the test results were based on the dependent and independent random sample selection of collected data from police stations.

6.2. Discussion

Detecting gang street (criminals) and crime- solving is not an easy task at all by the law enforcement like police. With the ever-increasing sophistication in computer technology, the logistic analysis models can now be used in order to speed up the process of crimes- solving by tracking gang street (criminals) and their activities from the community.

Logistic analysis model plays the most important roles in counter criminals by predicting them, identifying and tracking suspicious criminal activities. According to [17] about the application of logistic analysis model technology, he emphasized through the use with regression model based upon the poison distribution as a tool for solving general problems by analyzing the aggregate crime rates. When it comes to the population size of an aggregate unit seemed to be small compared to the actual offense rate, then the crime rates were calculated from a small number of observed offenses.

Some research show that, in criminology, the logistic analysis model has been used successfully by many police departments to predict and crackdown criminals having the ability of dichotomous power to output dependent variable [9].

In this study, the first hypothesis was to examine whether the logistic analysis model had the positive effect over the traditional crime- solving tools used by the Tanzanian police force to predict gang street (criminals). And the second hypothesis was to examine whether the time for crime commission had the positive effects to the Tanzanian community. The results showed that 83.0% were the overall percentage gang street (criminals); 87.8% were the sensitivity; 72.1% were specificity; 27.8% were false- positive and 21.2% were false- negative rate.

6.3. Recommendations

The law enforcement to be better effective in their daily internal and external crime operations, this study elucidates the following recommendations;

Law enforcers like police should be equipped with computer programs like logistic crime- solving analysis in order to crack down gang street (criminals) in real time and at low cost [18].

The government should build up a big data within the law enforcement like police where various data should be linked and extracted. For example, the data of voter's registration, national identity, phone numbers, bank accounts, motor vehicle registrations, national insurances, etc., moreover, the police should utilize professionally the existing data algorithms and develop them better capable of detecting more serious crimes easily.

And the community as a part of the community policing for a crime-solving should be more corporate with police by reporting any case of irregular transaction pattern in real-time basis.

REFERENCE

- [1] LeFrançois, G. (1996). Physical and cognitive development: Adolescence. Brabant, S., Mann, A. (Eds.), *The lifespan* (pp. 315-341). New York: Wadsworth Publishing Company.
- [2] Bergman, M. M. & Scott, J. (2001). Young adolescents' wellbeing and health-risk behaviours: Gender and socio-economic differences. *Journal of Adolescence*, 24, 183-197.
- [3] Brack, C. J., Orr, D. P., & Ingersoll, G. (1988). Pubertal maturation and adolescent self-esteem. *Journal of Adolescent Health Care*, 9, 280-285.
- [4] McGee, R. & Williams, S. (2000). Does low self-esteem predict health compromising behaviours among adolescents? *Journal of Adolescence*, 23, 569-582.
- [5] Gordon, C. (1971). Social characteristics of early adolescence. *Daedalus*, 100.
- [6] Boba, R. (2005) *Crime Analysis and Crime Mapping*, Sage, USA.
- [7] Takwimu (2014) retrieved on 31th January 2014 from <http://www.policeforce.go.tz/index.php/sw/pakua/category/7-takwimu>.
- [8] Weisburd, D., & Eck, J. E. (2004). What can the police do to reduce crime, disorder, and fear? *The Annals of the American Academy of Political and Social Science*, 593, 42-65.
- [9] Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage Publications.
- [10] Gail, M. H., & Pfeiffer, R. M. (2005). On criteria for evaluating models of absolute risk. *Biostatistics*, 6, 227-239. doi: 10.1093/biostatistics/kxi005.
- [11] Klain, E. J., Davies, H. J., & Hicks, M. A. (2001). *Child pornography: The criminal-justice-system response*. Alexandria, VA: American Bar Association Center on Children and the Law for the National Center for Missing & Exploited Children.
- [12] Hosmer, D.W., Lemeshow, S., (2000) *Applied Logistic Regression*. Second ed., New York: Wiley.
- [13] Collett, D. (2003) *Modelling Binary Data*. Second ed., Boca Raton: Chapman & Hall/CRC.
- [14] Anderson, J.A. (1972), "Separate Sample Logistic Discrimination," *Biometrika*, 59, 19-35.
- [15] Harrell, F.E., Lee, K.L., Califf, R.F., et al, (1984) *Regression modeling strategies for improved prognostic prediction*. *Stat. Med.*, 3: p. 143-152.
- [16] Harrell, F.E., Lee, K.L., Mark, D.B., et al, (1996) *Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors*. *Stat. Med.*, 15: p. 361-387.
- [17] Osgood, D. Wayne (2000) "Poisson-Based Regression Analysis of Aggregate Crime Rates". *Journal of Quantitative Criminology*, 16(1), 21-43.
- [18] Adderley, R., and Musgrove, P.B., (1999) *Data Mining at the West Midlands Police: A Study of Bogus Official Burglaries*, *BCS Specialist Group on Expert Systems, ES99*, London, Springer – Verlag, pp191-203, 1999.