

Research Article

Identifying and Predicting Major Factors affecting the Suicides in Sri Lanka

ABSTRACT

Aims: Identifying factors related to suicide and the prediction of future suicides are very important because suicide becomes a significant factor that engaged with education, social status, age, gender and many other factors. Therefore, the main objective of this study is to find the civil and education factors effecting on suicidal attempts in Sri Lanka and propose a model to predict the future suicides.

Study design: Statistical analysis with descriptive analysis and proposing models for predicting future suicides.

Place and Duration of Study: Data collected from the Department of Police, Sri Lanka, between January 2006 and December 2016.

Methodology: Data set has separated into two categories namely 'civil data' and 'educational data'. We modeled the data from 2006 to 2011 and the data from 2014 to 2016 were used for model validation purposes. Quasi Poisson and negative binomial regression models were fitted to identify the major factors affecting suicide in both categories. Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values were used to select the best model. Further, the Mean Absolute Percentage Deviation (MAPD) and Symmetric Mean Absolute Percent Error (SMAPE) were calculated to find the prediction accuracy of the proposed models.

Results: For both regression models, the variables age, gender and level of education are significant for the models fitted for educational data, and civil status and gender are significant in the civil status dataset. According to the analysis, highest suicides were recorded for the age groups 21-30 and over 61 males, minimally-educated and married people. By considering the MAPD values, the prediction accuracy of both Quasi Poisson models and Negative binomial models were above 99%. But the negative binomial model is the best model because of the comparable high accuracy than the other model. A considerable reduction in suicides was obtained in 2010, due to the peaceful situation in Sri Lanka after the civil war. It is observed that by paying special attention to teenagers, old-aged and married people can reduce the number of suicides.

Keywords: Mean Absolute Percentage Deviation, Negative Binomial, Quasi Poisson, Suicides, Symmetric Mean Absolute Percent Error.

1. INTRODUCTION

Nowadays suicide is one of the main social problems experienced by Sri Lanka which is a significant factor in health problems[1]. It can be considered as a complex event which it has the implication of social, biology and psychology [2]. Almost 4,000 people die by suicide in Sri Lanka every year. Besides Sri Lanka has a very high rate of suicides in the world in the 20th century [3]. Further, Sri Lanka is a developing country and higher suicide rates are very

harmful to the development of the country. Because of the recorded high suicide rates of Sri Lanka, investors are afraid to invest. Also, it is affected to the manpower for the development of the country and many others ways. As a country, we should pay our attention to reduce the suicides. Not only Sri Lanka but also the whole world is being affected because of the suicides in various ways [4]. When considering globally, suicide is the 13th leading reason for the death [5]. According to the reports of the World Health Organization, suicides are increased by 60% in the period of 2000-2009 compared to the previous years[6,7]. Many people thought that suicide is an entity that people kill themselves [8]. Although suicide has not one universal definition, in general, it is defined as intentionally self-inflicted death [9]. That means to end their life themselves, because they do not want to live any more [10]. Most of the suicide cases, the suicidal person has some mental situations that lead to the suicide [11]. However, suicide is a social problem rather than considered as an individual illness [2]. Reason for the suicide is not only the characteristics of the person but also the influence of the society [12].

Suicides became a significant factor that engaged with education, social status, age, gender and many other factors. Hence, it has emphasized the importance of identifying various factors affecting suicide. Therefore, the main objective of this study is to find civil and education factors effecting on suicidal attempts in Sri Lanka and model to predict the number of suicides according to the categories of significant factors. Suicide is a rare case and hence it is difficult to predict the future suicides manually without a statistical model. Also this paper related to the current data set in Sri Lanka and tries to give a statistical and mathematical interpretation for the future suicides. It is very important to reduce the suicides as it is important for the development of the country[13,14,15]. Therefore we need to pay attention to factors affecting suicides. Using the proposed model, we can observe the how major factors affecting suicides and the number of future suicides can be predicted. In this study we show the categories related to each factor that has the most risk to suicides. Besides, we try to reduce the number of suicides by identifying those categories clearly.

2. MATERIAL AND METHODS / EXPERIMENTAL DETAILS / METHODOLOGY

The data considered for this study were gathered by the Sri Lanka Police, and the descriptions of the variables are listed in Table 1.

Table 1. Description of Variables

Variable	Variable Name	Variable Type
Y	Number of suicides	Numerical
X ₁	Age Category	Categorical
X ₂	Gender	Categorical
X ₃	Educational Level	Categorical
X ₄	Civil Status	Categorical

The number of suicides is modeled with their age category, civil status, educational levels and gender from 2006 to 2011 and data from 2014 to 2016 were used for validation purposes. The data for the two years (i.e., 2012 and 2013) are not available in the data source. The number of suicides per year was considered as the response variable and age category, gender, educational levels and the civil status as the independent variables. The total size of the dataset is 31887 observations which consist of the entire total suicides occurred in Sri Lanka during the period of 2006 to 2016, except in 2012 and 2013. All the independent variables are categorical and dependent variable in numerical. The dependent variable is represented as count data. The dataset was divided to two main categories; one is called as 'educational data' which consists of the variables gender, educational level, age categories and number of suicides. Other one is with gender, civil status, age categories and

number of suicides called 'civil status data'. First, the dataset was prepared for the analysis by dividing categories to the independent variables according to the data. Then a preliminary analysis was done in order to get an idea about the dataset. After that Quasi Poisson and Negative Binomial regression models were fitted to identify the major factors affecting suicide. Both models are used to over dispersed count data and a generalized version of the linear regression models. When the assumption of mean equals to variance is violated for count data, Quasi Poisson regression is used. Quasi regression model is characterized by the first two moments and long with the log link function. When considering the negative binomial regression model it is more similar to the multiple regression except the dependent variable is count data. It is based on Poisson gamma mixture distribution which can be used with both numerical and categorical data. This model is so popular because it is allowed to use gamma distribution to modeling for Poisson heterogeneity. Significant factors are identified using the p-values of the variables of the fitted model. In order to select the best model Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values are used. Further, Mean Absolute Percentage Deviation (MAPD) and Symmetric Mean Absolute Percent Error (SMAPE) are calculated to find the prediction accuracy of the proposed model. The accuracy of the model was calculated as $(100 - \text{MAPD}) \%$. The best model is selected using the minimum values of AIC, BIC, MAPD and SMAPE values. For the prediction purposes, we use only significant variables from the training data set and then predicted the number of suicides. Then, the graph of actual number of suicides vs. predicted suicides was plotted. Moreover, MAPD and SMAPE values were calculated using observed and estimated suicides and the accuracy of the fitted models was calculated. All statistical analysis performed in this study was performed using the R statistical software.

3. RESULTS AND DISCUSSION

Some data visualization techniques were used to get an idea about the data and to explain about the how factors are affecting the suicides. Figure 1 and 2 shows how the total numbers of suicides are changing according to the civil status and the educational levels respectively.

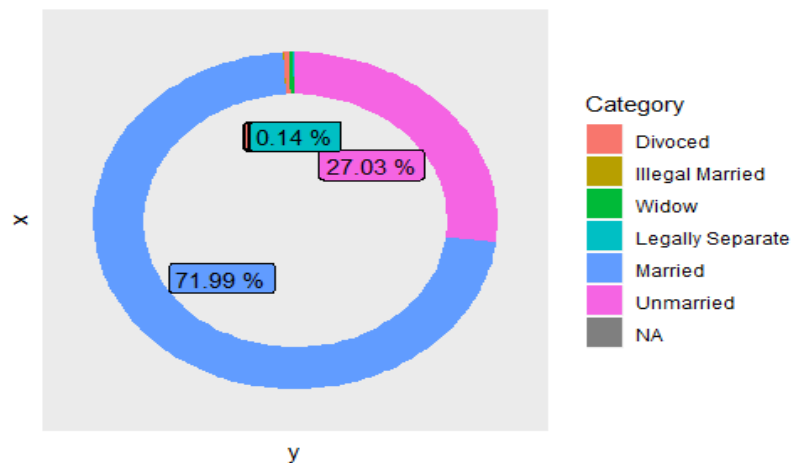


Fig 1. Percentage of suicides according to the civil status.

Figure1 shows that highest total numbers of suicides in 2006 to 2016 are recorded from married people.

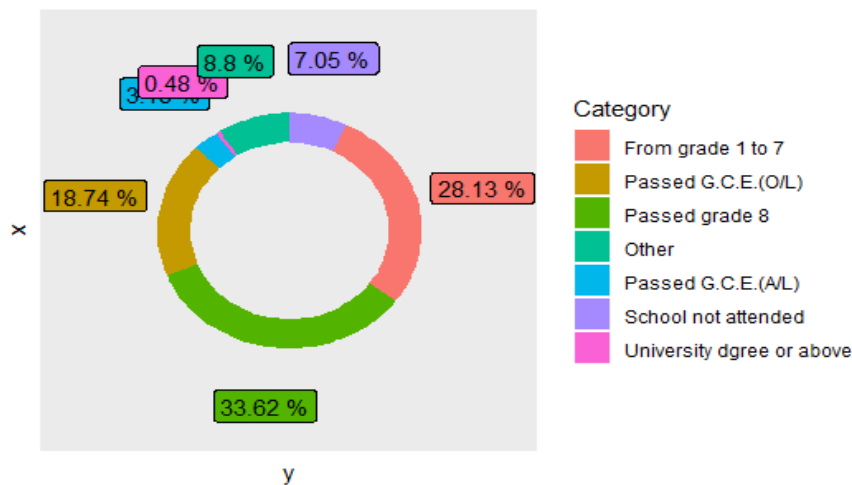


Figure 2. Percentage of suicides according to the educational level.

When considering the Figure 2, the highest number of suicides in educational level is recorded as the 'Passed Grade 8' category. Further 'Grade 1-7' category has the 2nd highest number of suicides while 'Degree or above' category shows the lowest number of suicides. Figure 3 show how the percentage of suicides are changing from 2006 to 2016 according to the age categories. The age category of 21-30 years has the highest percentage of suicides from 2006 to 2011 and age over 61 years category has the highest percentage of suicides from 2014 to 2016.

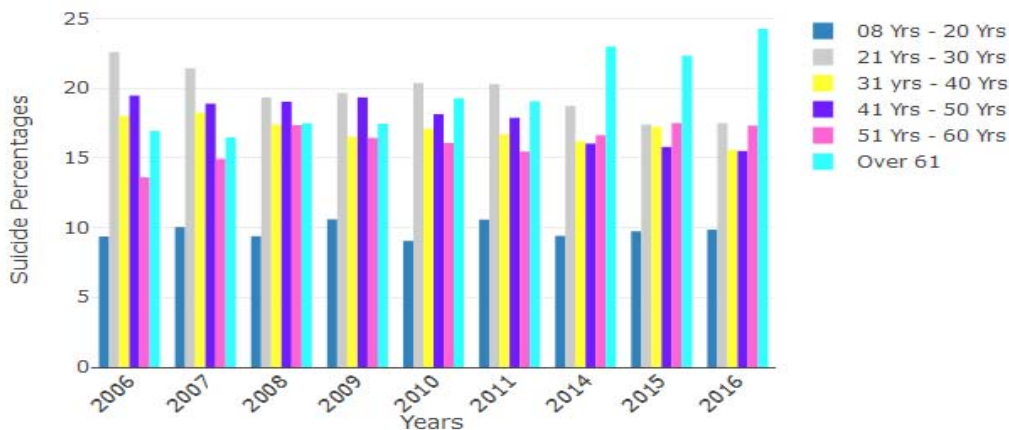


Figure 3. Suicide percentages from 2006 – 2016 with the age category.

Figure 4 shows the total number of suicides according to the gender in each year. According to Figure 4 the number of suicides in males is higher than that of the females in each year. With the cultural situations in Sri Lanka, males have more responsibility in day to day life, especially in the financial status of the family. Thus, the pressure with this responsibility can be a reason for more suicides in males. Moreover, a significant drop of the total number of suicides is observed in 2010, because of the peaceful situation after the civil war.

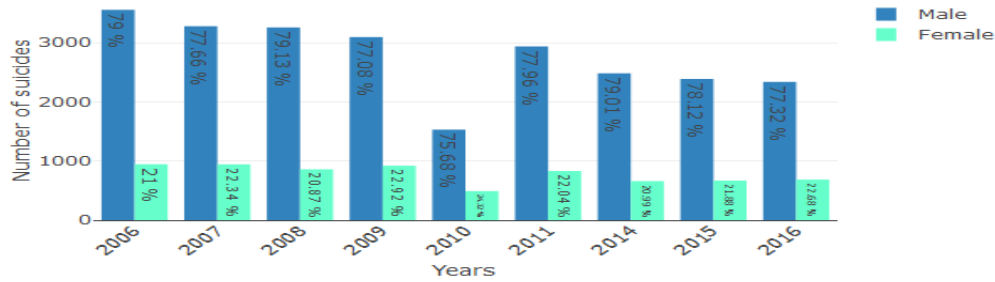


Figure 4. Number of suicides from 2006 – 2016 with gender.

Table 2 shows the summary results of the fitted Quasi Poisson model and the Negative Binomial model.

Table 2. Summary results of the best fitted models.

Dataset	Model	Variable	Coefficients	Stand. Error	Pr(> t)
Civil status data	Quasi Poisson	Intercept	6.30494	0.1569	2e-16 ***
		Gender	-1.18092	0.1064	2e-16 ***
		Civil Status	-0.40434	0.0395	2e-16 ***
	Neg. Binomial	Intercept	7.27661	0.1700	2e-16 ***
		Gender	-1.12547	0.0943	2e-16 ***
		Civil Status	-0.94793	0.0322	2e-16 ***
Educational data	Quasi Poisson	Intercept	5.494959	0.1309	2e-16 ***
		Age	-0.009181	0.0174	0.294
		Gender	-1.172217	0.0772	2e-16 ***
		Educational Level	-0.197399	0.0087	2e-16 ***
	Neg. binomial	Intercept	6.096085	0.1379	2e-16 ***
		Age	-0.056974	0.0089	1.64e-10 ***
		Gender	-1.221591	0.0163	2e-16 ***
		Educational Level	-0.259403	0.0653	2e-16 ***

According to the p-values of Table 2, educational data is significant with gender, age and educational levels and Quasi Poisson model is significant with only the gender and civil status variables.

Table 3. Summary result of fitted models.

Data set	Model type	BIC value	AIC value	MAPE	SMAPE	Deviance	Model Accuracy
Civil Status data	Quasi Poisson	5776.24	5687.25	0.9158	1.1258	36567.82	99.08%
	Neg.	5663.232	5645.4	0.8747	0.8812	709.42	99.13%

	binomial						
Educational data	Quasi	7512.12	7441.2	0.7630	0.8183	19690.37	99.24%
	Poisson		3				
	Neg.	7403.33	7379.2	0.7442	0.9712	1011.50	99.26%
	Binomial						

Best model was selected using minimum values of the BIC, AIC, deviance, MAPD and SMAPE. The prediction accuracies of the selected Negative binomial models and Quasi Poisson models were above 99% for the both civil status and educational data categories. The negative binomial model was considered as the best model, as its slightly higher accuracy compared to the Quasi Poisson model.

4. CONCLUSION

From the proposed models, it is observed that socioeconomic factors such as gender, age, educational level and civil status are the most influential factors affecting suicides in Sri Lanka. The Negative binomial regression model is considered as the best model to describe the factors affect for the suicides with the highest accuracy. We should pay more attention to males, teenagers and minimally educated people because the number of suicides was considerably high in each year for those categories. Hence by paying more attention, we can identify the related issues and situations for suicides. Also, after end of the civil war in 2009, a significant reduction of number of suicides was observed, and hence peaceful status of the country is necessary to reduce suicides. From 2012, a considerable increase of number of suicides of eldest people was observed compared to the previous years. With the development of the country after the civil war, peoples are busy with their jobs and other day to day activities and hence they missed looking after their parents. It can be a reason to increment of the suicides of eldest people. Moreover, in order to reduce the future suicides in Sri Lanka, more attention should be given for the major factors affecting suicides while taking necessary actions towards preventing suicides.

REFERENCES

1. Thalagala N. Suicide trends in Sri Lanka 1880-2006; Social Demographic and Geographical Variations. Journal of the College of community physicians of Sri Lanka, pp. 2009;24-32.
2. Rathnaweera RHA. Trends of suicides in Galle-Sri Lanka. Medico-Legal Journal of Sri Lanka. 2014;2(1),5-11.
3. Widger T. Reading Sri Lanka's Suicide Rate. Modern Asian studies.2013;1-35.
4. Aslam H,10 Countries with highest Suicide rate in 2016. [Online] Available at: <http://www.strongestinworld.com/highest-suicide-rate>
5. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study. 2010. The Lancet. 2012;380:2095–128.

- 182 6. World Health Organization. Suicide Rates Per 100,000 by Country, Year and Sex
183 (Table). Geneva: World Health Organization. 2011.
- 184 7. World Health Organization. Figures and facts about suicide. Geneva; 1999.2011.
185 http://whqlibdoc.who.int/hq/1999/WHO_MNH_MBD_99.1.pdf.
- 186 8. Ghoreishi SA, Mousavinasab N. Systematic review of researches on suicide and
187 suicide attempt in Iran. Iran J Psychaitry Clin Psychol.2008;14:115–21.
- 188 9. Masango SM, Rataemane ST, Motojesi AA. Suicide and suicide risk factors: A
189 literature review. SA Fam Prac. 2008;50(6):25-28.
- 190 10. Bandara UNJ, Nawarathna LS, Socioeconomic factors affecting suicides in Sri
191 Lanka. Biometrics & Biostatistics International Journal. 2018;7(5):405–411.
- 192 11. Ibiloglu AO, Atli A, Demir S, Gunes M, Kaya MC, Bulut M, et. al. The investigation of
193 factors related to suicide attempts in Southeastern Turkey. Neuropsychiatric
194 Disease and Treatment. 2016;12, 407–416.
- 195 12. Knipe DW, Gunnell D, Pearson M, Jayamanne S, Pieris R, Priyadarshana C, et. al.
196 Attempted suicide in Sri Lanka – An epidemiological study of household and
197 community factor. Journal of Affective Disorders. 2018;232:177-184.
- 198 13. Junsoo R, Johngeon P, Jinsuk L, et al. Factors That Affct Suicidal Attempt Risk
199 Among Korean. J Prev Med Public Health. 2015;48(1):28–37.
- 200 14. Haghighat M, Moravej H, and Moatamedi M. Epidemiology of pediatric acute
201 poisoning in southern Iran: A hospital-based study. Bull Emerg Trauma. 2013;1:28–
202 33.
- 203 15. Eautrais AL. A case control study of suicide and attempted suicide in older adults.
204 Suicide Life Threat Behav. 2002;32(1):1-9.
205