Volume 65, Issue 9, 2021
Journal of Scientific Research
of
The Banaras Hindu University

Spatio-Temporal Aspect of Suicide and Suicidal Ideation: An Application of SaTScan to Detect Hotspots in Four Major Cities of Tamil Nadu

Anjali, B. Rushi Kumar*, and Jitendra Kumar

Department of Mathematics, School of Advanced Sciences, VIT, Vellore-632014, TN
E-Mail: anjali.t@vit.ac.in, *E-Mail: rushikumar@vit.ac.in, E-Mail: jitendra.kumar@vit.ac.in

Abstract—**The history of suicide in India and particularly in Tamil Nadu has been witnessed since the existence of human settlement through ages, from pre-historic to ancient, medieval, modern and even to contemporary time. The primary objective of the present work is to identify the hotspots of suicide prevailing among four major cities of Tamil Nadu, namely, Chennai, Coimbatore, Madurai and Tiruchirappalli, for the time frame of 2012-2019. In addition to this, we are interested in descriptive and inferential analysis to find out some prominent attributes using multivariate statistical techniques. The present work analysed the Spatio-Temporal and longitudinal time series data acquired from National Crime Record Bureau (NCRB). Suicide hotspots due to various factors are detected using SaTScan. Despite metropolitan and cosmopolitan characteristics, high and dense urban population, proportionately higher cluster of illegal emigrants and settlements, Chennai being a capital city has not been identified as a hotspot of suicide for any factors under study. This study is useful and relevant for control and monitoring bodies in planning and curbing such a precious loss of human lives in the state of Tamil Nadu.**

Index Terms—**Hotspot, Longitudinal Data, Multiple Regression, Spatio-Temporal Analysis**

I. INTRODUCTION

According to the World Health Organization, suicide is the act of deliberately killing oneself. Suicide is among the top twenty leading causes of death worldwide, with more deaths due to suicide than to Malaria, breast cancer, or war and homicide [Lancet Global Health (2018)]. Close to 8, 00,000 people die by suicide every year [WHO (2019)]. Suicidal behaviour is a global cause of death and disability [Annu (2016)].

The Government of India classifies a death as suicide (intentionally ending life) if it meets the following three criteria: (a) It is an unnatural death, (b) The intent to die originated within the person, and (c) There is a reason for the person to end his or her life. The reason may have been specified in a suicide note or unspecified [ADSI (1995)]. Each human life is precious for humanity and the personal tragedy leads to social agony, therefore, Suicide is a prohibited and penalizes offence according to the Indian Penal Code Section-305(abetment of Suicide) and section-309(attempt of Suicide), as it is a misdeed and detrimental act for the individual, society, and for the future generations. The history of suicide in India has been reported many thousand years back. In Indian mythological works viz. Vedic, Puranic and Dharmashastric texts, there are many references/stories of Suicide. And in the contemporary scenario, it is estimated that over 1,00,000 people die by suicide in India every year; India alone contributes to more than 10% of suicides in the world [ADSI (2019)]. According to World Health Organization (WHO), *India's ranking* is 21 in the World with the 16.3/100k suicide rate, and in South East Asia, India has the highest suicide rate [WHO (2019)]. A total of 1,39,123 suicides were reported in India during 2019 showing an increase of 3.4% in comparison to the year 2018 and the second highest proportions of suicide were attempted in the state of Tamil

DOI: 10.37398/JSR.2021.650902

Nadu with 13,493 cases that are accounting for 9.7% of total suicide in India [ADSI (2019)].

II. REVIEW OF LITERATURE AND RELATED WORK

Nkhoma et.al (2004) examined the Spatio-Temporal clusters of accidental poisoning mortality among Texas counties, U.S. during 1980-2001 and found that the geographic and temporal persistence of clusters was found to vary by racial group, gender and race/gender combinations.

In India, Manoranjitham et.al (2010) determines the risk factors for suicide in rural south India using matched case-control design and psychological autopsy and finally concluded that psychosocial stress and social isolation, rather than psychiatric morbidity, are risk factors for suicide in rural South India. Another study, Nath et.al (2011) estimated the prevalence of suicidal behaviours among Indian college students and identify potential risk factors for suicide ideation and attempts using Logistic Regression analysis and found that the prevalence of lifetime suicide ideation and lifetime suicide attempts was 11.7% and 4.0%, respectively. Further, college students from low socio-economic classes who faced economic difficulties, and students who experienced distress as a result of caste discrimination or caste conflict, and communal unrest, were at a higher risk for suicidal behaviour.

Bando et. al (2012) identifies the possible significant clusters in the city of Sou Paulo using spatial scan statistics, and then verify their statistical association with socio-economic and cultural characteristics, and concluded that this risk protection profile is in accordance with the interpretation that, as a social phenomenon, suicide is related to social isolation.

Sathian et.al (2015) forecasted the suicide incidence of India up to 2020 during time frame 1989-2011. Using statistical modelling, it is forecasted that suicide incidence of India has an increasing trend. In 2020, it is estimated that the suicide incidence of India will be 109814 with CI [86593, 133034] for male, 76224 with CI [55151, 97297] for female and 186038 with CI [145605, 226471] for total [both male and female]. Orellana et. al (2016) examined the Spatio-temporal trends and risk of suicide in Central Brazil, and found that the suicide risk among the indigenous population was 8.1 (95%CI 7.2-9.0) times higher than in the non-indigenous population. Sukhija et. al (2017) determined the areas having a high intensity of crime (Criminal Hotspots) in Haryana using spatial visualization approach during time frame 20102014 and concluded that Gurgaon has been consistently found to be the primary hotspot whereas secondary hotspots come from the district of Rewari, Faridabad, Ambala, and Hissar. Kirtee & Nutan (2017) briefly overviewed the recent trends in the number of suicides and study risk factors for suicides, means of suicides in India. Results showed that the leading risk factors for suicides are family problems, illness, drug addiction, failure in examination, etc. Hanging and poisoning are the common means adopted by males and females while females most frequently commit suicide by fire/self immolation. Fontanella et.al (2017) identified the high-risk clusters of suicide in Ohio and assess area level correlates of these clusters and estimated spatially smoothed standardized mortality ratios (SMR) using Bayesian conditional autoregressive models (CAR) for the period 2004 to 2013. And found that the risk of suicide was up to 2.1 times higher in high risk clusters than in areas outside of the clusters (relative risks ranged from 1.22 to 2.14 ($p \prec 0.01$)). Gutirrez-Barroso et. al (2018) analyzed the phenomenon of suicide in the European Union from a sociological point of view, by creating a multivariate model. The findings of this study suggest that European women might make suicide attempts with higher risk of lethality. Hajela et. al (2019) proposed a Spatio-Temporal crime prediction technique based on machine learning coupled with 2-Dimensional Hotspot analysis and Performance of the proposed model is compared when it used state of the art classification techniques without hotspot analysis and with hotspot analysis and it is found that model with hotspot analysis achieves better performance.

In the literature survey, we found that prior studies have explored spatial pattern of crime but not for suicide in India.

Though some statistical work on suicide has been done, but there is no exclusive work done on application of Scan statistics on suicide in India. As such the objective of the present work is (a) to identify the hotspots of suicide prevailing among four major cities of Tamil Nadu, namely, Chennai, Coimbatore, Madurai and Tiruchirappall,

(a) to analyze the seriousness of the cases and their adverse effect on society,

(b) to develop statistical model to fit the suicide cases in Tamil Nadu,

(c) to find out some prominent attributes that enable us to improve our decision-making process using descriptive and inferential analysis.

Thus the present work will be of great societal benefits and it offers an opportunity for further research in this direction. As such the scope of further research has been realised. The relevance of the work seems to be high for the government, judiciaries, policing agencies and societies to monitor, control and operate their plan to deal with the situation and enable them to curb such national loss.

III. METHODOLOGY

A. Data Design and Methods

The present work analysed the suicide cases due to various factors and sub factors in state of Tamil Nadu, India. On the basis of logical and functional relation, we clubbed the sixteen variables and constructed the five main variables (Table-I). Taking into account the final merged variables, the total number of suicide (yearly) was considered as the dependent variable, whereas other variable were regarded as the independent variables. For the purpose of computational procedure, the analyses were performed by using the software namely SPSS, STATA, MS Excel SOLVER. We develop a multiple linear regression model to analyse the relationships between the total number of suicide cases (dependent variable) and five explanatory variables (namely, economical factors, socio-cultural factors, health factors, personal and professional factors and miscellaneous factors). Further, error analysis is performed to diagnose the violation of assumptions and then we checked for the over fitting or best fitting of the model and the trend fitting of suicide cases. Subsequently, we conducted test of multi-collinearity, autocorrelation, linearity, normality, constant variance using various methods like variance inflammatory factors (VIF), Durbin-Watson test, error analysis, normal probability plot etc. (Table-I) shows the set of variables under study.

B. Statistical Analysis

1) Spatio-Temporal Analysis: SaTscan [M. Kulldorff (2010)] is a software that analyse spatial, temporal and space-time data using the spatial, temporal and space-time scan statistics. It is designed to achieve the interrelated purposes: (a) perform geographical surveillance of suicide to detect spatial or space-time suicide clusters. (b) Examine the statistically significance of suicide hotspots (c) Determine whether suicide hotspots is randomly distributed over space, over time or over space and time. (d) Analyze prospective real-time or time-periodic suicide surveillance for the early detection of outbreaks.

In order to determine the geographical hotspots of suicide, we used Spatio-temporal scan statistics to obtain location and to evaluate the statistical significance of suicide hotspots. SaTScan can be used for discrete as well as continuous scan statistics. In our study, we employed the discrete poison model to analyze and obtain the suicide hotspots distributed among the four major cities of Tamil Nadu.

Table I. List of Clubbed Variables

S.No.	**Types of Variables**	**Sub-Variates**
1	Economical Factors($E_F = X_1$)	• Bankruptcy(X_{11}), • Poverty(X_{12}), • Unemployment(X_{13}), • Property Dispute(X_{14})
2	Socio-Cultural factors($SC_F = X_2$)	• Marriage Related issue(X_{21}), • Family Issue(X_{22}), • Death of Dear Person(X_{23}), • Physical Abuse(X_{24}), • Fall in Social Reputation(X_{25}), • Ideological/ Heroic cause(X_{26}),
3	Health factors($H_F = X_3$)	• Illness(X_{31}), • Drug and Alcohol(X_{32}),
4	Personal and Professional factors($PP_F = X_4$)	• Failure in Examination(X_{41}), • Love Affairs(X_{42}), • Professional Issues(X_{42}),
5	Miscellaneous factors($M_F = X_5$)	• Cause not Known(X_{51}), • other Causes(X_{52}),

2) Statistical Modelling: Multiple Regression model has been developed to estimate the relationship between a single dependent metric variable and five independent (regressors) metric variables. The model is shown below:

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\beta_4X_4+\beta_5X_5+\epsilon \qquad (1)$$

Where, Y represents the total number of suicide (dependent variable), whereas X_1, X_2, X_3, X_4, and X_5 represents the independent variables (namely, E_F, SC_F, H_F, PP_F, and M_F respectively), and denotes the random error in the model. In order to improve the model, we have applied the power series transformation to the original model (1).

The transformed model using Box-Cox (power series) has been applied by taking log of dependent as well as independent variables,

$$\log Y = \beta_0+\beta_1 logX_1+\beta_2 logX_2+\beta_3 logX_3+\beta_4 logX_4+\beta_5 logX_5+\epsilon \qquad (2)$$

So, the transformed model is,

$$\log y = \beta'_0+\beta'_1x_1+\beta'_2x_2+\beta'_3x_3+\beta'_4x_4+\beta'_5x_5+\epsilon' \qquad (3)$$

Where, $y = logY$, $x_1 = logX_1$, $x_2 = logX_2$, $x_3 = logX_3$, $x_4 = logX_4$, and $x_5 = X_5$

2.1 Model Validation

Algorithm adopted is being explained through following steps:

Step1: Setting H_0 and H_1, where H_0 is the null hypothesis and H_1 is the alternative hypothesis.

H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$

H_1: At least one of these coefficients is not equal to zero,

Or, H_1: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 \neq 0$

Step2: Selection of $\propto$, the level of significance, let $\propto = 0.5$

Step3: Test statistics

$p - value$, associated with F-statistics

Step4: Decision Criteria,

If p-value<$\propto$-value, we reject H_0.

Step 5: Drawing inference, writing report and making correct decision.

2.2 Assumption Validation

2.2.1 Error Analysis:

In a multiple linear regression, a Normal probability plot graph is used to validate the assumption of normality with the decision criteria that a straight line indicates the assumptions of normality of the error term. Further, plot of standard residual versus the fitted value provides an overall picture, while the plots of the standard residual versus each explanatory variable may help identify any violation of assumption of the regression model, specifically related to an individual explanatory variable.

Decision Criteria [Statistics for Business and Economics (Eleventh Edition)]:

(a) The systematic pattern observed in a residual plot suggests a violation of the assumption of linearity.

(b) Further, in a residual plot, Hetroscedasticity (non-constant variance) is often identified by a cone shaped pattern.

2.2.2 Test of Constant Variance:

Several tests are available for the validation of non-constant variance but we performed J. Szroeters, which tends to be better at detecting non-constant variance. The test is conducted in following steps:

Step1: setting null hypothesis (H_0) and alternative hypothesis (H_1),

H_0: Variance is constant,

i.e., $V(\in_i) = \sigma^2$; i= 1, 2… n

H_1: Variance is not constant,

i.e., $V(\in_i) \neq \sigma^2$; i= 1, 2,… n

Step2: J. Szroeters test statistics

$$Q=\left(\sqrt{\frac{6n}{n^2-1}}\right)\left(h-\frac{n+1}{2}\right),$$

where, n is the number of observations in the sample and h is computed as

$$h=\left[\frac{\sum_{i=1}^{n} i\hat{e}_i}{\sum_{i-1}^{n} \hat{e}_i^{\,2}}\right],\ i=1, 2\ldots n$$

where, $\hat{e}_i$ is the residual from i^{th} observation in the regression

Step3: Decision Criteria:

If $Q > Z_\propto$, then reject H_0, and

If $Q \leq Z_\propto$, then do not reject H_0

where, $z_\propto$ is chosen from the standard normal table with upper tail area $\propto$, which is always positive value.

2.2.3 Auto-Correlation:

The assumption of independence of disturbance (AutoCorrelation) is being tested using Durbin-Watson test and computed by first using least square to estimate the regression and by computing the residuals.

$$\hat{e}_i = y_i - \hat{y}_i, \tag{4}$$

Where, y_i represents number of suicide cases and $\hat{y}_i$ represents the corresponding predicted values

Step1: Setting null hypothesis (H_0) and alternative hypothesis (H_1)

H_0: disturbance is independent

H_1: disturbance is not independent

Step2: D-W test statistics

$$d = \sum[\hat{e}_i - \hat{e}_{i-1}]^2, \tag{5}$$

where $\hat{e}_i$ represents the residual.

Step3: Decision Criterion:

If $d < d_L(\propto, n, K)$, then reject H_0, and

If $d > d_U(\propto, n, K)$, then do not reject H_0.

Where, $d_L(\propto, n, K)$ and $d_U(\propto, nK)$ are critical values which can be found from D-W statistics table and the critical value depends upon the level of significance ($\propto$), the sample size (n) and the number of explanatory variables in the equation (K).

Note: There is a range of values for the test statistics when the test is said to be inconclusive $i.e.$ $[d_L \leq d \leq d_U]$.

2.2.4 Multi-Collinearity:

Algorithm adopted:

Step1: Test statistics

$$VIF_j = \frac{1}{1-R_j^{\,2}},\ j=1, 2\ldots n \tag{6}$$

Where, $R_j^{\,2}$ represents the R^2 value of the corresponding regression model.

Step2: For the present work, we have five independent variables(X_1, X_2, X_3, X_4, X_5), i.e., K=5 and n=8 and p=6 (number of parameters i.e. $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$) to perform regression of X_j (j=1, 2, 3, 4, 5), on the remaining (K-

1)explanatory variables and call the coefficient from the regression (R_j^2).

Step3: Decision Criteria

(a) For any individual VIF_j larger than 10 indicates that MC exists and it may be influencing the least square estimate of the regression coefficients.

(b) If the average of VIF_j, i.e., $\overline{VIF_j} = \sum \frac{VIF_j}{K}$ is considerably larger than 1 then serious MC problem exists.

IV. RESULTS AND DISCUSSION

We prepared master descriptive statistics (Table-II) to characterized and describe the statistics attributes of the variables. Now, it enable us to summarize the description of the variables in terms of Mean, Standard deviation (SD), Standard Error (SE), Lowest and Highest value, Skewness and Kurtosis, which is essential for understanding the distribution of the datasets. Factors (Love Affairs, Failure in Examination and Illness), have skewness values (-2.340, -0.694 and -1.846) respectively which shows that their distribution is negatively skewed whereas other variables are positively skewed and distribution of none of the variable is symmetric (*i.e.*$S_K = 0$). The mean and SD value of the variables will be used for various computation and analysis, for example- we have computed coefficient of variation to find out the consistency in the factors, (Table-III) based on the consistency we do ranking as per statistical principle of lowest CV value means more consistent and higher rank assign to that factors. We determined the rankings for the purpose of comparison, finding out the relationship among variables and drawing reference about related measures rather than the absolute measures.

Our data consists of suicide cases for the time period 20112019. Therefore, we must visualize the pattern trend line as shown by fig.1, which shows straight line with negative slope in the state of Tamil Nadu, and fig.2 to 5 representing the time trend line for the major cities of Tamil Nadu, namely Chennai, Coimbatore, Madurai and Tiruchirappalli respectively.

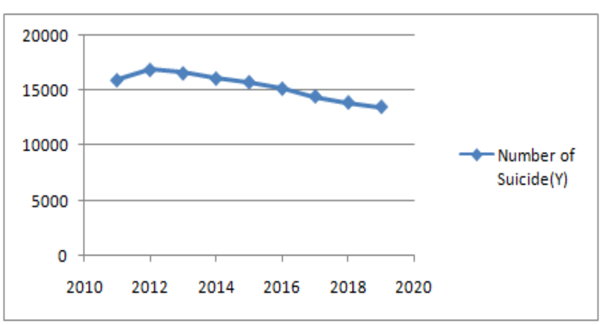

Fig. 1 Time Trend Line for Tamil Nadu

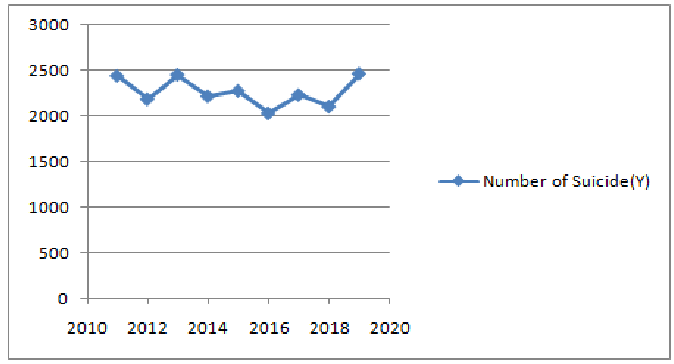

Fig. 2 Time Trend Line for Chennai

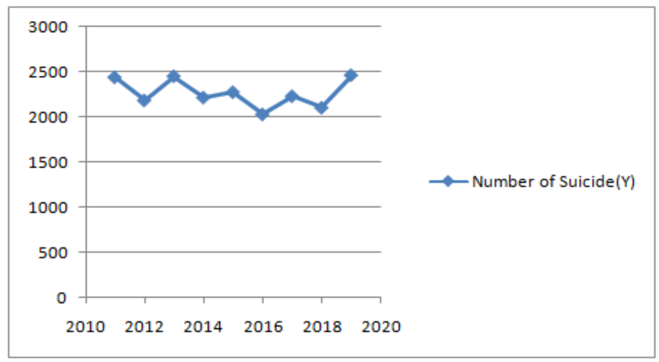

Fig. 3 Time Trend Line for Coimbatore

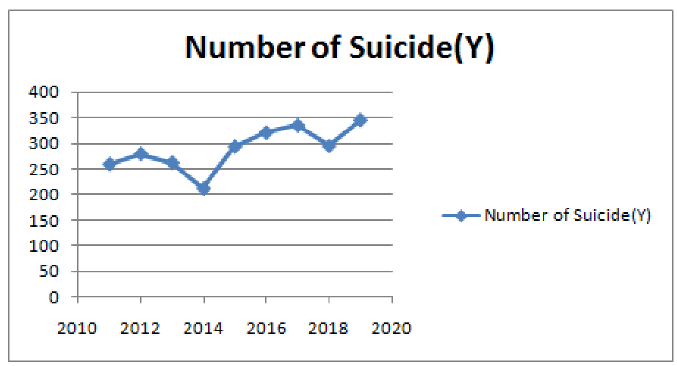

Fig. 4 Time Trend Line for Madurai

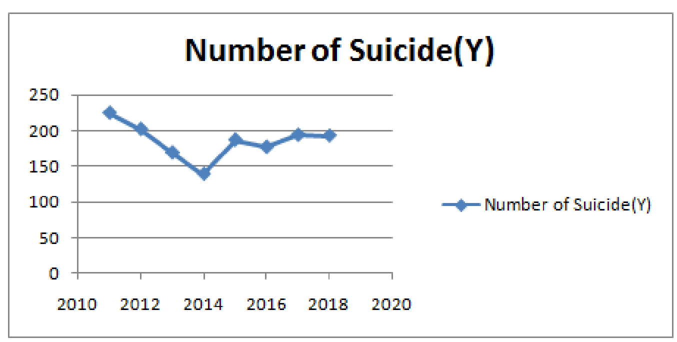

Fig. 5 Time Trend Line for Tiruchirappalli

Table II. Descriptive Statistics

*Variables S.No.	Range Statistic	Minimum Statistic	Maximum Statistic	Sum	Mean Statistic	Mean SE	SD	Variance	Skewness Statistic	Skewness SE	Kurtosis Statistic	Kurtosis SE
1	363	39	402	1736	217	37.242	105.338	11096	0.077	0.752	1.261	1.481
2	381	286	667	3196	399.50	41.163	116.428	13555.429	2.048	0.752	5.067	1.481
3	144	178	322	1944	243	15.714	44.446	1975.429	0.433	0.752	0.384	1.481
4	4096	3281	7377	44830	5603.75	433.009	1224.733	1499971.93	-0.694	0.752	1.190	1.481
5	3633	879	4512	27411	3426.38	402.534	1138.538	1296267.7	-1.846	0.752	4.157	1.481
6	151	87	238	1270	158.75	19.286	54.549	2975.643	0.151	0.752	-1.250	1.481
7	799	242	1041	5270	658.75	104.095	294.425	86686.214	0.112	0.752	-1.363	1.481
8	485	0	485	513	64.13	60.136	170.09	28930.696	2.826	0.752	7.991	1.481
9	23	0	23	62	7.75	2.743	7.760	60.214	1.153	0.752	1.004	1.481
10	391	165	556	3688	461	44.418	125.633	15783.714	-2.340	0.752	5.848	1.481
11	126	24	150	481	60.13	13.951	39.459	1556.982	2.031	0.752	4.658	1.481
12	203	211	414	2281	285.13	24.820	70.203	4928.411	1.002	0.752	0.021	1.481
13	140	11	151	555	69.38	15.227	43.068	1854.839	0.939	0.752	1.028	1.481
14	125	0	125	205	25.63	16.683	47.186	2226.554	1.802	0.752	2.313	1.481
15	211	85	296	1343	167.88	29.183	82.543	6813.268	1.089	0.752	-0.506	1.481
16	5995	544	6539	21081	2635.13	708.439	2003.769	4015091.554	1.116	0.752	0.858	1.481

*Variables Used in (Table-II): 1-Bankruptcy; 2-Marriage Related Issues; 3-Family Problem; 4-Failure in Examination; 5-Illness; 6-Death of Dear person; 7-Drug/Alcohol; 8-Fall in Social Reputation; 9-Ideological Cause; 10-Love Affairs; 11-Poverty; 12-Unemployment; 13-Property Dispute; 14-Physical Abuse; 15-Professional problems; 16-Miscellaneous;

Table III. CV Based Ranking

Rank	Factor
1	Failure in Examination
2	Family Problems
3	Unemployment
4	Love Affairs
5	Marriage related issues
6	Illness
7	Death of Dear Person
8	Drug/Alcohol
9	Bankruptcy
10	Profession
11	Property
12	Poverty
13	Miscellaneous
14	Ideological cause
15	Physical Abuse
16	Fall in Social Reputation

A. *Regression computation*

Since our objective is to develop and fit a statistical model for the purpose of forecasting and planning of cause of suicide to benefit the society. We applied multiple linear regressions and for that purpose variable selection has been done carefully to satisfy our statistical need. In order to have an idea about outliers, missing values and any other irrelevant information's suitable data processing has been done. Firstly, the box plot is obtained for each variable, then out of sixteen variables four has been found to be statistically significant, namely, family problems, illness, drug /alcohol and miscellaneous. The mean difference are showing an effective distribution patterns (Figure-6) and secondly we performed multivariate factor analysis, and found only six out of sixteen namely, bankruptcy, marriage related issues, family problems, failure in examination, illness, death of dear person, and drug/alcohol) found to be suitable for the analysis based on their eigen values (Figure-7). Further on comparing the results obtained through box plot and factor analysis, we found that factor analysis results in possible loss of substantial information. To avoid the loss of information, we categorized the explanatory factors and clubbed related sub-factors and assigned the name as main factor (as mentioned in the Table-1). Further, this enables us to reduce the risk of over-fitting of the model.

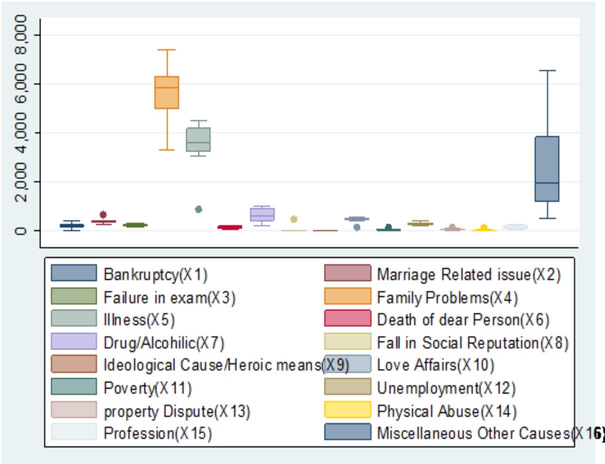

Fig. 6 Box plot of original 16 variables

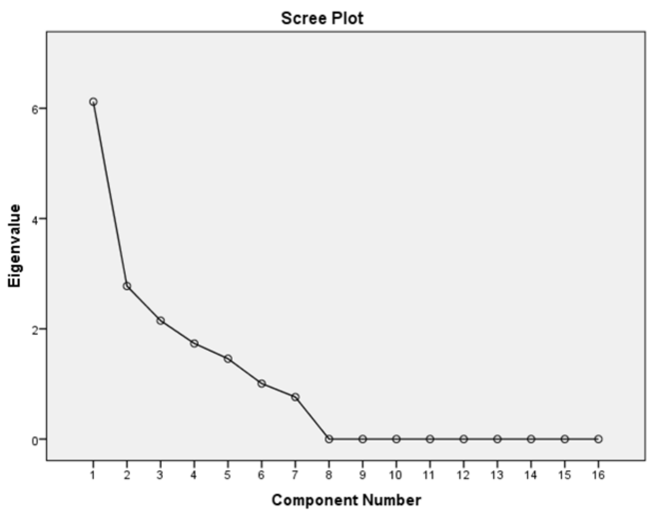

Fig. 7 Factor analysis of original 16 variables

In our findings of regression, we found that except PP_F factor, all remaining factors under study have positive coefficient values that mean a unique change in the value of these variables will change the overall suicide cases positively. The R^2 and Adjusted R^2 values are 0.99075 and 0.96763 respectively which seems to be little higher than expected that may be due to the effect of multi-collinearity in the model. Moreover, the Standard error value is found to be statistical significant (i.e. 0.00652). Since, p-value associated with F-statistics is 0.022962 which is less than α-value (0.05), therefore, we decide to reject the null hypothesis, which validate our model. Further, if we conduct F-test using critical value approach, we get the same result, i.e. our model is validated ($F_{calculated}$ value is 42.84885 which is larger than $F_{tabulated}$, i.e. $F_{(\alpha=0.05,\, n1=5,\, n2=2)} = 19.2964$. This implies that there is enough evidence in the sample to reject the null hypothesis). Since, the lower and upper confidence interval computed on 95% as well as 99% for each factors(table-regression), contains the hypothesized value of β_i's = 0,i=1,2,3,4,5. Therefore, the suicide model has been found to be statistically significant.Hence, the fitted model is:

$$y = 32.651 + 0.0045x_1 + 0.250x_2 + 23.74x_3 + (-206.95)x_4 + 0.112x_5 + \in' \qquad (7)$$

where, $\beta_0 = 32.651$, $\beta_1 = 0.0045$, $\beta_2 = 0.250$, $\beta_3 = 23.74$, $\beta_4 = -206.95$, $\beta_5 = 0.112$

We have examined the assumptions of linearity, normality, homoscedasticity and independence of the error terms in the model by doing suitable error analysis and plots. Since, normal probability plot gives a straight line pattern; hence normality seems to be preserved, as shown by the Fig-8 to 12.

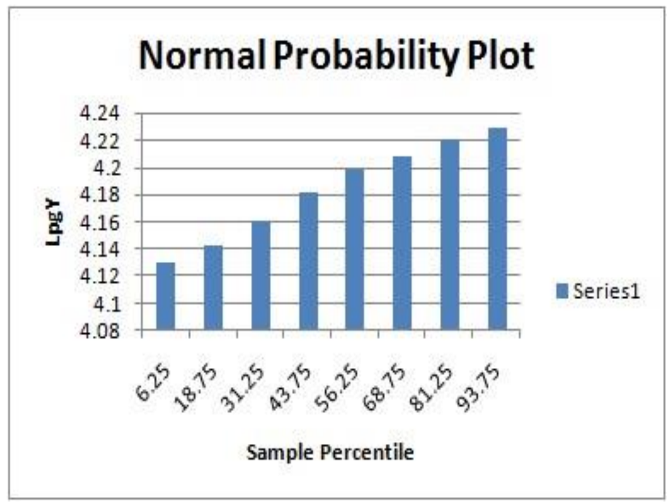

Fig. 8 Normal Probability Plot for Tamil Nadu

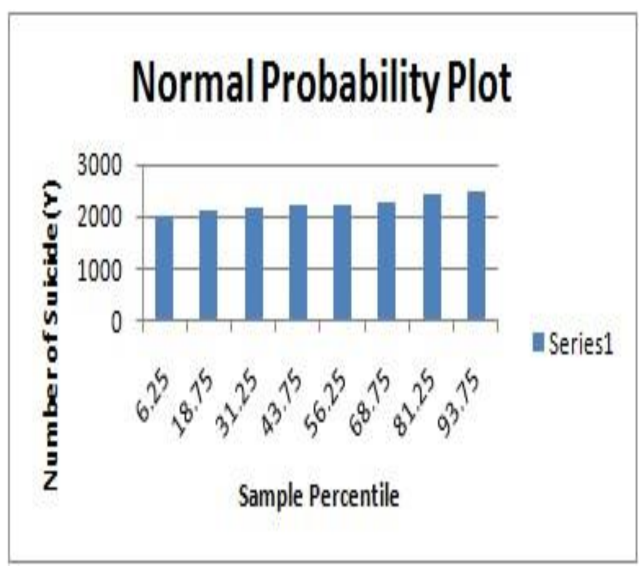

Fig. 9 Normal Probability Plot for Chennai

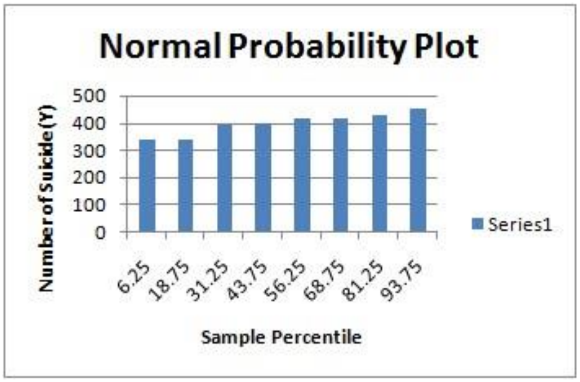

Fig. 10 Normal Probability Plot for Coimbatore

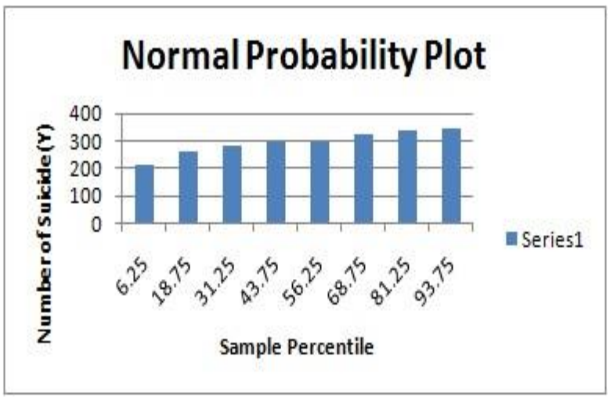

Fig. 11 Normal Probability Plot for Madurai

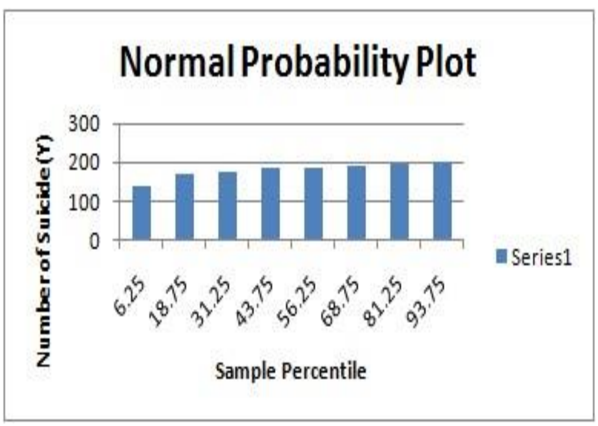

Fig.12 Normal Probability Plot for Tiruchirappalli

The error plot of the standard residual versus the fitted value provides an overall picture of the standardized error plots shown by Fig-13 to 17, enables us to examine and investigates the assumptions involved. There is no systematic pattern being observed that ensures random distribution of the errors and further there is complete absent of cone shaped pattern of the error. These two validates the assumption of randomness and constant variance (homoscedasticity) of the error terms in the model. Further, we have conducted the J. Szroeter's test for the test of constant variance and found that Q-statistic value ≤ Zα (which is for upper tail value of standard normal Z-table, which is always positive ≥ 0), therefore, we do not reject the null hypothesis. Thus it can be concluded that the variance is constant for the model that means constant variance (homoscedasticity) assumption is validated for the model.

To check the assumption of independence of disturbance, we have performed the Durbin-Watson test and we obtained the value of D-W statistics d= 1.6361 (i.e. approximately near to the value 2), that implies that the problem of auto-correlation do not exists for our model. Moreover, the problem of multi-collinearity occurs when explanatory variables are correlated with one another. To check the presence of multi-collinearity in our model we have examined and computed the value of variance inflammatory factor (VIF), which is 108.1221, that indicates that the variables under study are seems to be inter-related and the statistical assumption of independence of errors seems to be violated which confirms the multi-collinearity effect in the model.

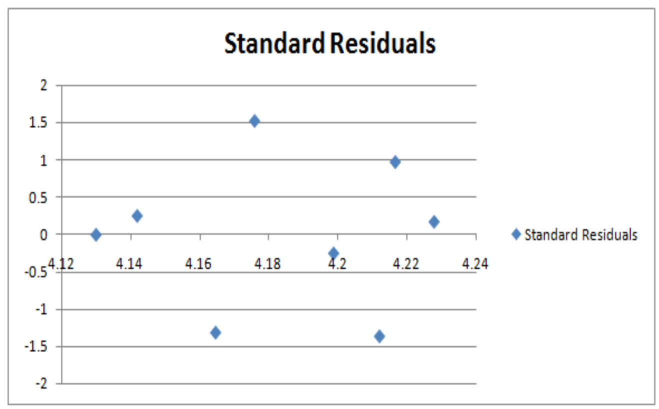

Fig. 13 Standard Residual Plot for Tamil Nadu

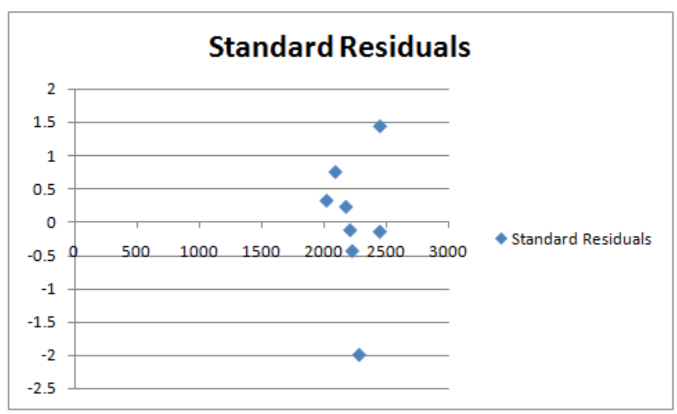

Fig. 14 Standard Residual Plot for Chennai

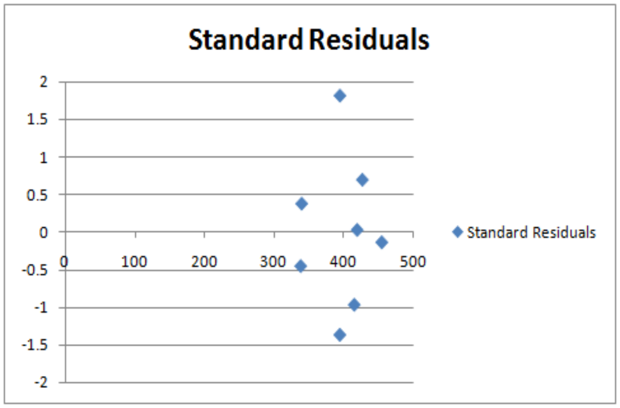

Fig.15 Standard Residual Plot for Coimbatore

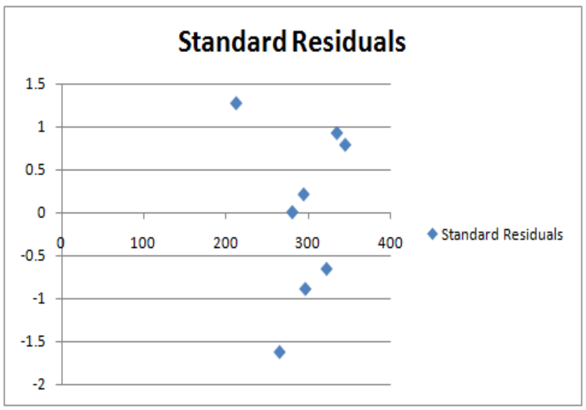

Fig. 16 Standard Residual Plot for Madurai

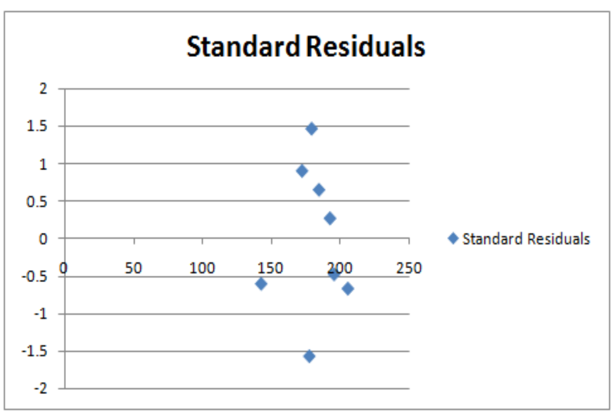

Fig. 17 Standard Residual Plot for Tiruchirappalli

B. Hotspot identification

In our present work we did hotspot identification using two sets of data, one with amalgamated and merged variables (E_F,SC_F,H_F,PP_F,M_F) (Table-IV) and another data set we choose exact factor available and projected through NCRB portal and conducted factor analysis to obtain the top six factors out of sixteen variables (Table-V)(reference-fig.7). In both approaches we determined the hotspots and performed a comparative study to identify the hotspot (most likely cluster) with respect to the time frame and with the different factors under the study. Though, we have found hotspot for each year respective of each factor but for comparative visualization, we represent through the google map only for the year 2019 and a complete table (Table-IV) representing the hotspot for each year respective of each factor obtained through the SaTScan for better understanding.

In (Table-IV), we found that in 2012 except health related factor (H_F), remaining all factors have their primary hotspot in Coimbatore that implies further that Coimbatore is the epicentre of suicide in 2012. To the contrary of year 2012, in 2013 Madurai is the epicentre of suicide as it occurs in all factors except the SC_F that means we can say that there is a clear change in hotspot dynamics from 2012 to 2013. The special finding in 2014 is a point to mention that Tiruchirappalli has emerged as a hotspot for the two factors i.e. SC_F and PP_F for the first time. For the year 2015 and 2016, there is no specific region which is emerging in all the factors, Tiruchirrappllli, Madurai and Coimbatore is emerged as the hotspot in these years.

In 2017, there is a sudden change in the hotspot dynamics, where Madurai has emerged as a hotspot for all the factors except H_F. For the year 2018 and 2019, there has been a various serious finding emerges out of the present work that Coimbatore has been re-emerged as a hotspot after 2015.

On comparing these results of Table (4) with the Table (5), we can examine that for the year 2012 and 2013, Coimbatore and Madurai has been emerged as the epicentre of suicide respectively in both datasets but there is a variation for the year 2018 and 2019 hotspot. We found that Chennai despite being the capital of Tamil Nadu and metropolitan city too having high urban density is not the hotspot in any of the year which appears to be very interesting findings.

TABLE IV
HOTSPOT RESPECTIVE TO FIVE AMALGAMATED FACTORS

Year	Factors	Hotspot Cities	Relative Risk	Log-likelihood Ratio	p-Value
	E_F	Coimbatore	7.77	13.140598	0.00000020
	SC_F	Coimbatore	2.90	62.617827	$\prec$ 0.00000000000000001
2012	H_F	Madurai	3.55	69.251755	$\prec$ 0.00000000000000001
	PP_F	Coimbatore	7.69	27.866302	$\prec$ 0.00000000000000022
	M_F	Coimbatore	0.21	109.137058	$\prec$ 0.00000000000000001
2013	E_F	Madurai	0	10.560988	0.00000082
	SC_F	Coimbatore	1.62	17.035699	0.000000000053
	H_F	Madurai	2.42	38.521774	$\prec$ 0.00000000000000001
	PP_F	Madurai	0.69	0.747877	0.58
	M_F	Madurai	0	91.306807	$\prec$ 0.00000000000000001
2014	E_F	Madurai	0.19	7.189795	0.000083
	SC_F	Tiruchirappalli	1.94	14.390221	0.0000000026
	H_F	Madurai	1.62	10.965202	0.00000027
	PP_F	Tiruchirappalli	2.40	4.923804	0.0021
	M_F	Madurai	0	36.617349	$\prec$ 0.00000000000000001
2015	E_F	Tiruchirappalli	0	17.051502	0.000000000096
	SC_F	Tiruchirappalli	1.81	18.996941	0.0000000000034
	H_F	Coimbatore	1.30	40.653273	0.0032
	PP_F	Madurai	3.03	12.531899	0.000000045
	M_F	Madurai	0	37.378832	$\prec$ 0.00000000000000001
2016	E_F	Tiruchirappalli	1.99	4.376527	0.012
	SC_F	Tiruchirappalli	0.027	63.736353	$\prec$ 0.00000000000000001
	H_F	Madurai	1.17	1.547581	0.199
	PP_F	Tiruchirappalli	1.88	2.707318	0.053
	M_F	Madurai	0	45.916234	$\prec$ 0.00000000000000001
2017	E_F	Madurai	0.085	15.435736	0.00000000073
	SC_F	Madurai	2.68	46.148530	$\prec$ 0.00000000000000001
	H_F	Tiruchirappalli	0.47	13.035547	0.000000012
	PP_F	Madurai	0.34	6.211159	0.00045
	M_F	Madurai	0	91.045964	$\prec$ 0.00000000000000001
2018	E_F	Madurai	0.12	14.513444	0.000000044
	SC_F	Madurai	1.61	15.864591	0.00000000032
	H_F	Coimbatore	1.38	5.798077	0.0004
	PP_F	Coimbatore	2.77	10.186486	0.0000020
	M_F	Coimbatore	0	52.298088	$\prec$ 0.00000000000000001
2019	E_F	Madurai	0	26.997887	0.000000000000000011
	SC_F	Coimbatore	1.62	19.227597	0.0000000000021
	H_F	Coimbatore	0.47	18.858588	0.0000000000041
	PP_F	Coimbatore	2.64	14.990970	0.0000000017
	M_F	Madurai	0	35.304929	$\prec$ 0.00000000000000001

Table V. Hotspot respective to factor analysis variables

Year	Hotspot Cities	Relative Risk	Log likelihood Ratio	p-Value
2012	Coimbatore	3.33	123.430756	$\prec$ 0.00000000000000001
2013	Madurai	1.78	32.388704	$\prec$ 0.00000000000000001
2014	Madurai	1.31	6.350098	0.00014
2015	Tiruchirappalli	1.24	3.838316	0.015
2016	Tiruchirappalli	0.55	15.638816	0.00000000014
2017	Madurai	1.57	24.636126	0.00000000000000022
2018	Madurai	1.24	5.196170	0.0010
2019	Madurai	1.20	4.477294	0.0028

Fig. 18 Primary Hotspot during year 2019 respective of socio-cultural factors

Fig. 19 Primary Hotspot during year 2019 respective of Health Factor

V. CONCLUSION

As the main objective of our paper is to identify the hotspots among the four major cities under the study, we found that Coimbatore, Madurai and Tiruchirappalli has been emerged as the hotspots (most likely clusters) during the time period of the year 2012-19 with respective to major factors and six sub-factors causing suicidal cases. It is quite enthralling to mention that the capital city Chennai despite having metropolitan and cosmopolitan characteristics with very high population and high urban density, not being identified as suicidal hotspot in any of the year irrespective of any factors. This finding is supported by Durkheims theory [Suicide: A Study in Sociology (1897)] in which he states that Suicide is an indication of disorganization of both individual and society and further he believes that Suicide is not an individual act but the real cause for suicide is the social factors. Hence, we may say that suicide cases may not be the directly proportional to urban conurbation, population density and size of the city; it is somehow more related to psychological and socio-cultural factors. Secondly fitting of suicidal modelling using main factors, namely, economical ($E_F = X_1$), socio-cultural ($SC_F = X_2$), health related ($H_F = X_3$), personal and professional ($PP_F = X_4$), and miscellaneous ($M_F = X_5$) is also being supported by various other theories mentioned in our literature survey. Model fitting procedure is being supported by Bhatia et. al (2021). Though, the present work seems to be a case study but its not, as findings can be generalized for any location of larger dimension, like the whole state, region and the country. Thus our findings are expected to be relevant and useful for the people and society of Tamil Nadu in general and respective government and officials in particular.

VI. LIMITATIONS

It has been observed that multi-collinerity is likely to affect the regression coefficient estimates; probably it could lead to higher value of coefficient of determination R^2 and adjusted R^2 value. So far, there is no single remedial methods that can be used to remove the effects of multi-collinearity from the model completely. However, it does not affect the availability to obtain a good fit, nor does it affect the quality of forecast and prediction from the regression.

VII. GLOSSARIES

Accidental Deaths and Suicides in India (ADSI), Autocorrelation, Multi-collinearity, Log-likelihood ratio, p-value, Q-statistic, d-statistic, Variance Inflation Factor (VIF), Factor analysis, Multiple linear regression, Scan statistic

REFERENCES

World Health Organization, Suicide in the World (2019), (data: 2016), Global Health Estimates

Lancet Global Health 2018; published online Sept 12. $http://dx.doi.org/10.1016/S2214-109X(18)30407-8$

Annu. Rev. Clin. Psychol. (2016), 12:30730.TheAnnual Review of Clinical Psychology 10.1146/annurev-clinpsy021815-093204

Accidental Deaths and Suicide in India (1995), National Crime Record Bureau, Ministry of Home Affairs, Government of India

Accidental Deaths and Suicide in India (2019), National Crime Record Bureau, Ministry of Home Affairs, Government of India

Niedzwiedz, C., Haw, C., Hawton, K., Platt, S., (2014). The definition and epidemiology ofclusters of suicidal behavior: a systematic review. Suicide Life-Threat. 44 (5),569581

Ella T Nkhoma, Chiehwen Ed Hsu, Victoria I Hunt andAnn Marie Harris, Detecting spatio-temporal clusters of accidental poisoning mortality among Texas counties, U.S., (1980 2001), International Journal of Health Geographics (2004), 3:25. https://doi.org/10.1186/1476-072X-3-2

S. D. Manoranjitham, A. P. Rajkumar, P. Thangadurai, J. Prasad, R. Jayakaran and K. S. Jacob,Risk factors for suicidein rural south India The British Journal of Psychiatry (2010) 196, 2630. https://doi.org/10.1192/bjp.bp.108.063347

Yogini Nath, Joel Paris, Brett Thombs and Laurence Kirmayer, Prevalence and social determinants of suicidal behaviours among college youth in India International Journal of Social Psychiatry(2011) 58(4) 393399. https://doi.org/10.1177/0020764011401164

Daniel H Bando, Rafael S Moreira, Julio CR Pereira andLigia V Barrozo, Spatial clusters of suicide in the municipality ofSo Paulo (19962005): an ecological study BMC Psychiatry (2012), 12:124.

Brijesh Sathian, Asis De, Edwin van Teijlingen, Padam-Simkhada, Indrajit Banerjee, Bedanta Roy, HS Supram, Suresh Devkota, and Rajesh E, Time Trend of the Suicide Incidence in India: a Statistical Modelling. American Journal of Public Health Research, vol. 3, no. 5A (2015): 80-87. https://doi.org/10.12691/ajphr-3-5A-17

Jesem D. Orellana, Antonio A. Balieiro, Fernanda R. Fonseca, Paulo C. Basta, Maximiliano L. Ponte de Souza, Spatial-temporal trends and risk of suicide in Central Brazil: an ecological study contrasting indigenous and non-indigenous populations RevistaBrasileira de Psiquiatria. (2016) 38:222230.Associaca o Brasileira de Psiquiatria https://doi.org/10.1590/1516-4446-2015-1720

Khushboo Sukhija, Shailendra Narayan Singh, Jitendra Kumar, Spatial visualization approach for detecting criminal hotspots: An analysis of total cognizable crimes in the state of Haryana, 2nd IEEE International Conference on Recent Trends in Electronics Information and Communication Technology (RTEICT), May 19-20, 2017, India https://doi.org/10.1109/RTEICT.2017.8256761

Kirtee K. Kamalja and Nutan V. Khangar, A statistical study of suicidal behaviour of Indians, Kamalja and Khangar Egyptian Journal of Forensic Sciences (2017) 7:12 https://doi.org/10.1186/s41935-017-0007-9

Cynthia A. Fontanella, Daniel M. Saman, John V. Campo, Danielle L. Hiance-Steelesmith,Jeffrey A. Bridge, Helen Anne Sweeney, Elisabeth D. Root, Mapping suicide mortality in Ohio: A spatial epidemiological analysis ofsuicide clusters and area level correlates. https://doi.org/10.1016/j.ypmed.2017.10.033

Josu Gutirrez-Barroso, Fernando Barragn-Medero and David Prez-Jorge, Suicide in Europe Countries: A Multivariate Approach Analysis Global Journal of Health Science; Vol. 10, No. 4; 2018. https://doi.org/10.5539/gjhs.v10n4p12

Gaurav Hajela, Dr.MeenuChawla, Dr.AkhtarRasool, A Clustering Based Hotspot Identification Approach For Crime Prediction, International Conference on Computational Intelligence and Data Science (ICCIDS 2019), Procedia Computer Science 167 (2020) 14621470. https://doi.org/10.1016/j.procs.2020.03.357

M. Kulldorff, SaTScan v9.0: Software for the spatial and space-time scan statistics, Information Management Services, (2010).

Textbook by Emile Durkheim, Suicide: A Study in Sociology (1897), by Routledge and Kegan Paul Ltd, London and New York

Udbhav Bhatia, Tshering, Jitendra Kumar and Dilip Kumar Chaubey, Drowsiness Image Detection using Computer Vision,(July 2021), Springer, Soft Computing Theories and Applications, pp 667-683, Advances in Intelligent Systems and Computing, Book Series(AISC, Volume 1380). *Link* : *https* : *//link.springer.com/chapter/*10.1007/978 − 98/ − 16 − 1740 − $9_5$5

David R. Anderson, Dennis J. Sweeney, Thomas A. Williams : Statistics for Business and Economics, Eleventh Edition, Joe Sabatino, 2011, 2008 South-Western, Cengage Learning.

***