



DATA DRIVEN ANALYSIS ON SPREAD OF CORONAVIRUS IN INDIA - A TIME DEPENDENT NON-PARAMETRIC MATHEMATICAL APPROACH

Geetha Narayanan Kannaiyan¹, Bridjesh Pappula²

¹Department of Science and Humanities, Sri Krishna College of Engineering
and Technology, Coimbatore, Tamilnadu, India

²Department of Mechanical Engineering, MLR Institute of Technology,
Hyderabad, Telangana, India

¹nkgeeth@gmail.com, ²meetbridjesh@gmail.com

Corresponding Author: Geetha Narayanan Kannaiyan

<https://doi.org/10.26782/jmcms.2020.04.00006>

Abstract

Statistical analysis is a qualitative research used to quantify data adapting a statistical tool. The present study is to device a time dependent non-parametric mathematical model to analyze the spread of COVID-19 in INDIA based on the statistics available. As the medicine to treat COVID-19 is not invented yet, the best possible way to break the chain of spreading virus is, "Personal Hygiene and Social Distancing".

Keywords : COVID-19, Statistical analysis, Non-parametric analysis,

I. Introduction

The most talked word of this day across the globe is Coronavirus disease (COVID-19). This is an infectious disease caused by coronavirus. World Health Organization (WHO) proclaims that COVID-19 is pandemic. The possible ways of spreading of COVID-19 are through either saliva droplets or nose discharge from an infected person [I]. COVID-19, being a highly contagious disease shall also be transmitted to others by touch virus contaminated objects. The incubation period for this novel coronavirus differs from person to person. However, Centers for Disease Control and Prevention (CDC) claims that the incubation period is between 2-14 days of exposure [II]. Liu et al reported that transmissibility of COVID-19 is very high than other viruses but the mortality rate is low as compared with other viruses [III]. The situation report 1 published by WHO dated: 20 January 2020 states that 282 people got infected by COVID-19 from four countries such as china, Thailand, Japan and Republic of Korea [IV] of which, casualties were 6. The situation report 66 published by WHO dated: 26 March 2020 states that the number of confirmed cases across the globe was 4,62,684. Whereas, the statistics about India reported by WHO as per situation update report -8 revealed that 360 were confirmed as infected with

Copyright reserved © J. Mech. Cont.& Math. Sci.

Geetha Narayanan Kannaiyan et al

COVID-19 and 7 casualties as on 22 March, 2020 [V]. As per the statistics published in [VI], the number of COVID-19 confirmed cases as on 01 February, 2020 was one but by 26 March, 2020 the number of confirmed cases escalated to 735 and the number of casualties was 16. These numbers sound an alarm unto mankind, the catastrophe about to fall.

Statistical analysis is a qualitative research used to quantify data adapting a statistical tool. Statistical analysis brings life to lifeless data by adapting suitable model for analysis wherein, the inferred results shall be precise and valid [VII]. Depending on the data available either parametric or non-parametric test are conducted to analyze the data. Parametric tests are used when the data are distributed normally [VIII] whereas, non-parametric tests are conducted when data are skewed [IX].

The present study is to device a time dependent non-parametric mathematical model to analyze the spread of COVID-19 in INDIA based on the statistics available.

II. Methods

Regression function and forecasting

The following mathematical approach is established based on the data collected from sources and the details of data are presented in Appendix.

Let A be the number of people affected by COVID-19, t be the time duration in days and k is a constant.

As on 01 February, 2020 the number counts as one (A=1). The generalized equation for estimating number of people affected is given as,

$$\frac{dA}{dt} = kt \quad (1)$$

Then,

$$A = Ce^{kt} \quad (2)$$

Where, C is a constant.

Considering 01-02-2020 as day zero, the number of affected people is 1. Such that,

Substituting t=0 and A=1 in eq. (2), we get C=1.

For a period of 50 days, i.e when t=50, A=403 (Refer Appendix Table 1)

Substituting t=50 and A=403 and C=1 in eq. (2), we get,

$$403 = 1xe^{k \times 50},$$

Then, k=0.1199

Hence the generalized formula shall be established as,

$$A = 1xe^{0.1199t} \quad (3)$$

The correlation coefficient is calculated as 0.935. Likewise the number of people affected on a day to day basis shall be calculated.

Error function

For a generalized equation of form,

$$Y = X\beta + e \quad (4)$$

Let Y is response vector, e is error vector. Where,

$$e = Y - X\beta \quad (5)$$

The sum of squared errors shall be computed as,

$$e^T e = (Y - X\beta)^T (Y - X\beta) \quad (6)$$

The root mean squared error was calculated as 0.91

III. Results and Discussion

Figure 1 shows the graph between number of people infected and days. It can be seen from the figure that number of people affected by COVID-19 has deviated from the normal curve on 04-03-2020. This number has escalated from this day in an exponential manner. If this trend persists, the projected number of people infected increases and is presented in Figure 2.

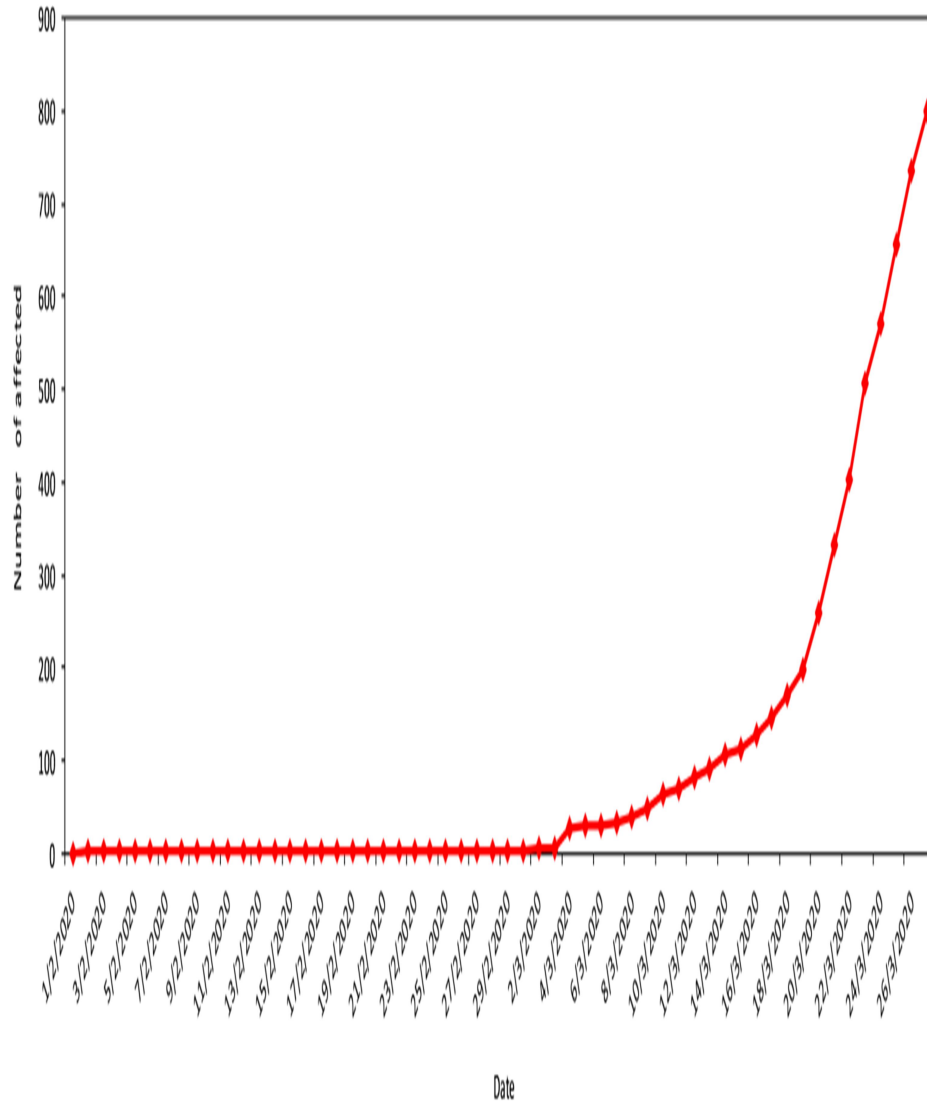


Fig.1 Number of affected from 01-02-2020 to 26-03-2020

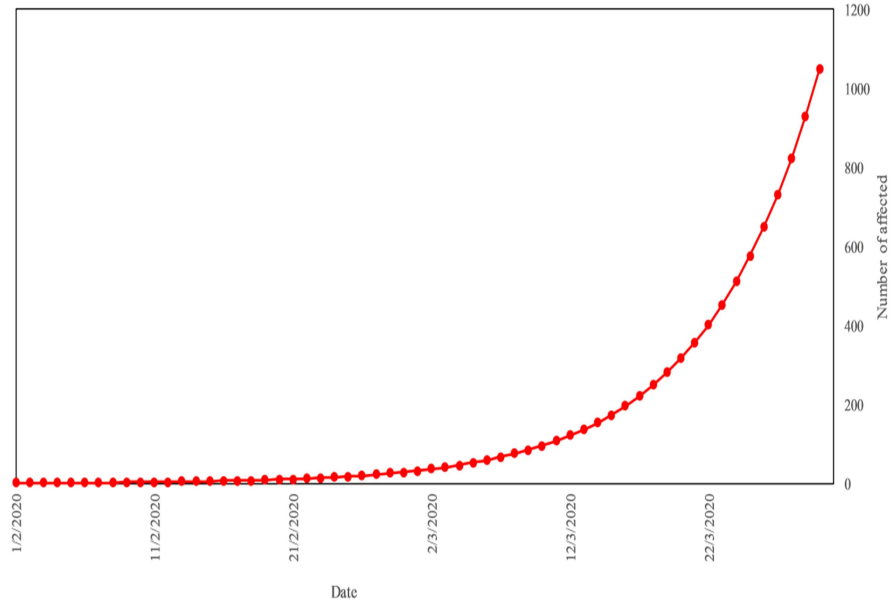


Fig. 2 Projected number of affected till 30-03-2020

Minitab was used to project polynomial equations for analyzing the number of affected for verification. The polynomial equation derived for the given data is,

$$Y=0.5277X^2-46314X+1E+09. \quad (7)$$

The same methodology shall be adapted to analyze the count of recovered and casualties too.

IV. Conclusion

The study dealt with the statistical analysis of COVID-19 infected number of people. Unless the proactive and preventive measures are taken to control the pandemic disease, it shall be a catastrophe to mankind. As the medicine to treat COVID-19 is not invented yet, the best possible way to break the chain of spreading virus is, “Personal Hygiene and Social Distancing”.

References

- I. www.who.int/health-topics/coronavirus#tab=tab_1 (accessed on 26-03-2020)
- II. www.cdc.gov/coronavirus/2019-ncov/about/symptoms.html (accessed on 26-03-2020)
- III. Y. Liu, A. A. Gayle, A. Wilder-Smith, J. Rocklöv, “The reproductive number of COVID-19 is higher compared to SARS coronavirus”, *J Travel Med*, 27(2), 2020.
- IV. www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf?sfvrsn=20a99c10_4 (accessed on 26-03-2020)
- V. www.who.int/docs/default-source/wrindia/situation-report/india-situation-report-8bc9aca340f91408b9efbedb3917565fc.pdf?sfvrsn=5e0b8a43_2 (accessed on 26-03-2020)
- VI. www.covid19india.org/ (accessed on 26-03-2020)
- VII. Z. Ali, S. Balabhaskar, “Basic statistical tools in research and data analysis”, *Indian J Anaesth*, 60(9), 662–669, 2016
- VIII. D.G. Altman, J. M. Bland, “Parametric v non-parametric methods for data analysis”, *BMJ* 338, 3167-3173, 2009

Appendix

Table 1:Data of COVID-19 in INDIA [6]

COVID-19 Cases in INDIA				
Day count	Date	Confirmed	Recovered	Deceased
0	01/02/2020	1	0	0
1	02/02/2020	2	0	0
2	03/02/2020	3	0	0
3	04/02/2020	3	0	0
4	05/02/2020	3	0	0
5	06/02/2020	3	0	0
6	07/02/2020	3	0	0
7	08/02/2020	3	0	0
8	09/02/2020	3	0	0
9	10/02/2020	3	0	0
10	11/02/2020	3	0	0
11	12/02/2020	3	0	0
12	13/02/2020	3	1	0
13	14/02/2020	3	1	0
14	15/02/2020	3	1	0
15	16/02/2020	3	2	0
16	17/02/2020	3	2	0
17	18/02/2020	3	2	0
18	19/02/2020	3	2	0
19	20/02/2020	3	3	0
20	21/02/2020	3	3	0
21	22/02/2020	3	3	0
22	23/02/2020	3	3	0
23	24/02/2020	3	3	0
24	25/02/2020	3	3	0
25	26/02/2020	3	3	0
26	27/02/2020	3	3	0
27	28/02/2020	3	3	0
28	29/02/2020	3	3	0
29	01/03/2020	3	3	0
30	02/03/2020	5	3	0
31	03/03/2020	6	3	0
32	04/03/2020	28	3	0

Copyright reserved © J. Mech. Cont.& Math. Sci.
Geetha Narayanan Kannaiyan et al

33	05/03/2020	30	3	0
34	06/03/2020	31	3	0
35	07/03/2020	34	3	0
36	08/03/2020	39	3	0
37	09/03/2020	48	3	0
38	10/03/2020	63	4	0
39	11/03/2020	70	4	0
40	12/03/2020	82	4	1
41	13/03/2020	91	10	1
42	14/03/2020	107	10	2
43	15/03/2020	113	13	3
44	16/03/2020	127	14	2
45	17/03/2020	146	15	3
46	18/03/2020	171	15	3
47	19/03/2020	199	20	4
48	20/03/2020	258	23	4
49	21/03/2020	334	23	4
50	22/03/2020	403	23	7
51	23/03/2020	505	25	9
52	24/03/2020	571	40	10
53	25/03/2020	657	50	16
54	26/03/2020	735	50	16