International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 04 Issue: 04 | April -2020 ISSN: 2582-3930

Prediction of Day wise Positive Corona Cases in India Using Statistical Model

Deepali S. Akolkar, Asmita S. Sutar, Ankita G.Deshmukh, Rohini M. Jadhav

Abstract:

Covid-19 also called as corona virus is the infectious disease caused by the most recently discovered virus. This new virus and disease were unknown before the outbreak began in Wuhan, China, in December 2019.The current population of India is 1,376,751,999 as based on Worldometer elaboration of the latest United Nations data. In this article as compare to India's population we study number of corona positive cases seen in last month and try to predict daily probably positive cases will be arrive in next month. Using Statistical model non-linear method of regression .we predict the possibility of actual positive corona cases. Using available data, Sources, best fitting curve i.e. second degree curve or Polynomial we evaluate possible positive cases from 14th Apr. to 20th Apr.2020 so that to preventing or handle the spread of present situations.

Key words: *Ms-Excel, Nonlinear regression method, Line graph, Scatter Plot, Simple bar Chart, Correlation.*

Introduction:

India is the second most populated country in the world with nearly a fifth of the world's population. Government of India is taking all necessary steps to ensure that we are prepared well to face the challenge and threat posed by the growing pandemic of COVID 19 – the Corona Virus. Official names have been announced for the virus responsible for COVID-19 (previously known as "2019 novel corona virus") and the disease it causes **by Virus Severe Acute Respiratory Syndrome corona virus 2**(SARS-CoV-2)With active support of the people of India, we have been able to contain the spread of the Virus in our country. The most important factor in preventing the spread of the Virus locally is to empower the citizens with the right information and taking precautions as per the advisories being issued by Ministry of Health & Family Welfare.

In this method we have collected data of one month from 3march2020to 3April2020 (we have started the case studies from 3 march because on the date the first corona case has been registered in India and after this the cases started to increasing with huge and unknown count) from the references on Google and using statistical analysis method of second degree curve, we estimate parameters or constants and got predicted results.

In this paper we try to study the correlation between number of days increases and active cases per day. Also by using past data we predict how many active cases will present on 14 April by using regression analysis. For that we collect data from 3 march to 3 April day wise with active cases and death per day.

Methodology:

In this paper try to find relation between number of days increases and active cases per day by using correlation. Correlation is extent of linear relationship between two variable. After that we use regression analysis. Regression is the technique of prediction on the basis of correlation. there are two regression

model, one is linear regression model another is non-linear regression model. Sometimes there is no linear relation between two variable but curvilinear relation is found such as second degree curve and exponential curve.

Both types of models can fit curves to your data so that's not the defining characteristic. The difference between linear and non-linear regression models isn't as straightforward as it sounds. You'd think that linear equations produce straight lines and nonlinear equations model curvature. If a regression equation doesn't follow the rules for a linear model, then it must be a nonlinear model. That means A non-linear model is literally not linear.

The added flexibility opens the door to a huge number of possible forms. Consequently, nonlinear regression can fit an enormous variety of curves. However, because there are so many candidates, you may need to conduct some research to determine which functional form provides the best fit for your data.

After that we fit the regression line for linear regression and second degree curve(polynomial). Calculate adjusted R^2 it is shown that adjusted R^2(polynomial) is greater than adjusted R^2(linear).from this it is clear that second degree curve(polynomial) fitting is better fitting than linear fitting.

Curve fitting: least squares methods

Fitting of Second Degree Regression Curve:

Let {(Xi, Yi), I = 1……n } be a sample taken on two variable X and Y.

Let us take X as independent variable and Y as dependent variable. The equation of second degree curve is as,

Let **y= a+bx+cx2**

Be the second-degree parabola to be fitted to bivariate data where a, b and c are the constants to be determine by using principle of least squares. It involves the minimization of sum of squares of error in the estimation,

For X = xi the regression estimate of Y is

Yi. = a+b*Xi + c*Xi2

x=	No. of Days	Independent Variable
y=	New Positive corona cases	Dependent Variable

Observation Table:

The following table gives values of new corona cases day wise. for further calculations we add columns

Days	X	X^2	Y
03-Mar	1	1	4
04-Mar	2	4	22
05-Mar	3	9	1
06-Mar	4	16	1
07-Mar	5	25	3
08-Mar	6	36	5
09-Mar	7	49	6
10-Mar	8	64	15

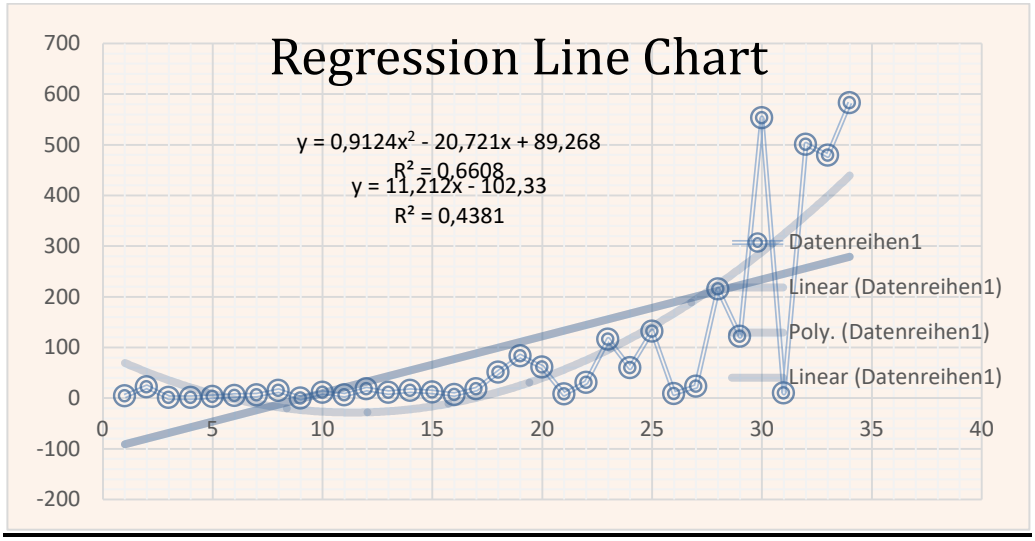
11-Mar	9	81	0
12-Mar	10	100	11
13-Mar	11	121	6
14-Mar	12	144	18
15-Mar	13	169	11
16-Mar	14	196	15
17-Mar	15	225	12
18-Mar	16	256	7
19-Mar	17	289	18
20-Mar	18	324	51
21-Mar	19	361	83
22-Mar	20	400	61
23-Mar	21	441	8
24-Mar	22	484	31
25-Mar	23	529	116
26-Mar	24	576	60
27-Mar	25	625	132
28-Mar	26	676	9
29-Mar	27	729	23
30-Mar	28	784	215
31-Mar	29	841	122
01-Apr	30	900	553
02-Apr	31	961	10
03-Apr	32	1024	501
04-Apr	33	1089	479
05-Apr	34	1156	583

***Statistical analysis:**

Correlation

Correlation between x and y= 0.661859

It is the positive correlation between No. of Days and new positive corona cases



Graph 1: here R^2(polynomial) 0.66 > R2(linear)0.438. Therefore Second degree curve is best fitting

Here by graph, we get Coefficient of determination R^2 for linear line is 0.43 and Coefficient of determination R^2 for non-linear line is 0.660.so for fitting line Non-linear line is best fitted. we can analyse data for polynomial i.e Non linear line.

***Non-Linear Regression**

I) c b a
0.912377 -20.721 89.26771
here a,b,c, are constant or parameter used in below equation,
Y = a+bx+cx^2

* **On 14th(43rd day) April there will be 885 new positive caseswill arises.**
If X=43 then Y(corona positive is
Y=885.2511

* **On 15th(44th day) April there will be 944 new positive cases will arises.**
If X(day)=44 then Y(corona positive) is
Y= 943.907

* **On 16th(45th day) April there will be 1004 new positive cases will arises.**
If X(day)=45 then Y(corona positive) is
Y= 1004.388

* **On 17th(46th day) April there will be 1066 new positive cases will arises.**

If X(day)=46 then Y(corona positive) is

Y= 1066.693

* **On 18th(47th day) April there will be 1130 new positive cases will arises.**

If X(day)=47 then Y(corona positive) is

Y= 1130.823

***On 19th(48th day) April there will be 1197 new positive cases will arises.**

If X(day)=48 then Y(corona positive) is

Y= 1196.776

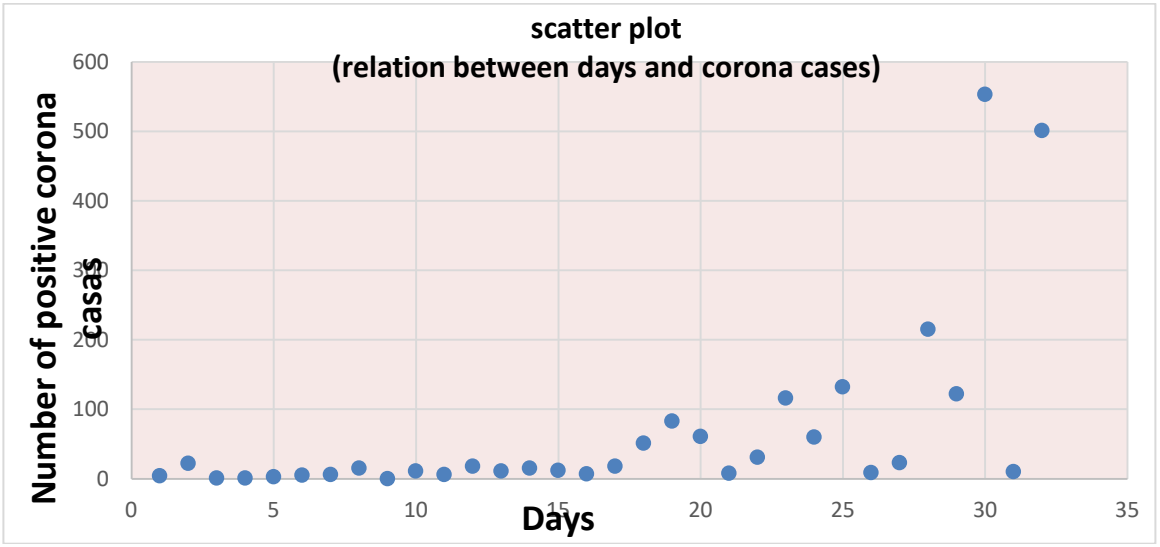
***On 20th(49th day) April there will be 1265 new positive cases will arises.**

If X(day)=49 then Y(corona positive) is Y= 1264.55

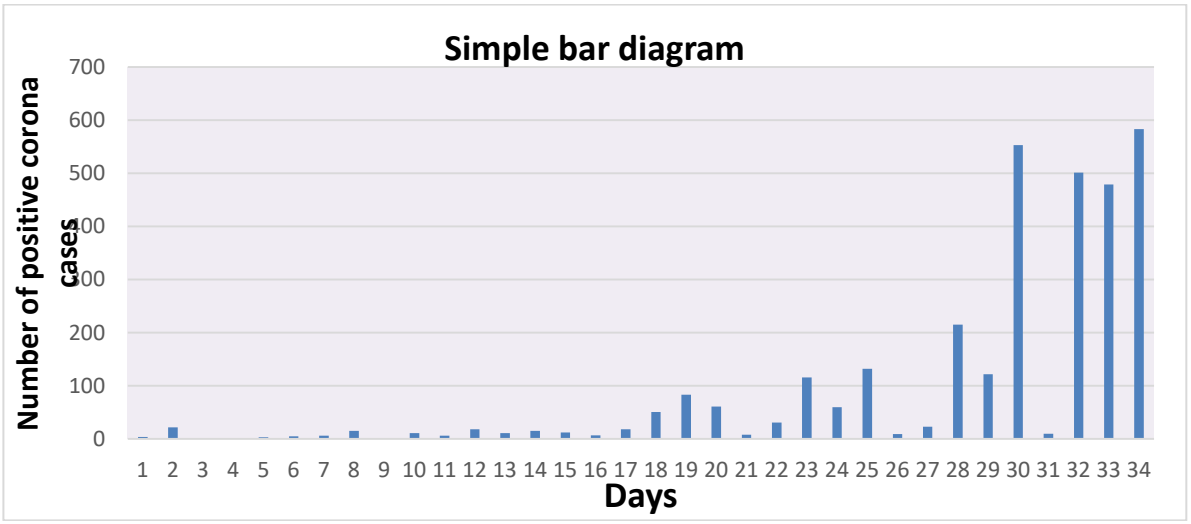
***The observation table of estimation of new positive corona cases**:

Days	Number of New positive corona patients
14/04/2020	885
15/04/2020	944
16/04/2020	1004
17/04/2020	1066
18/04/2020	1130
19/04/2020	1197
20/04/2020	1165

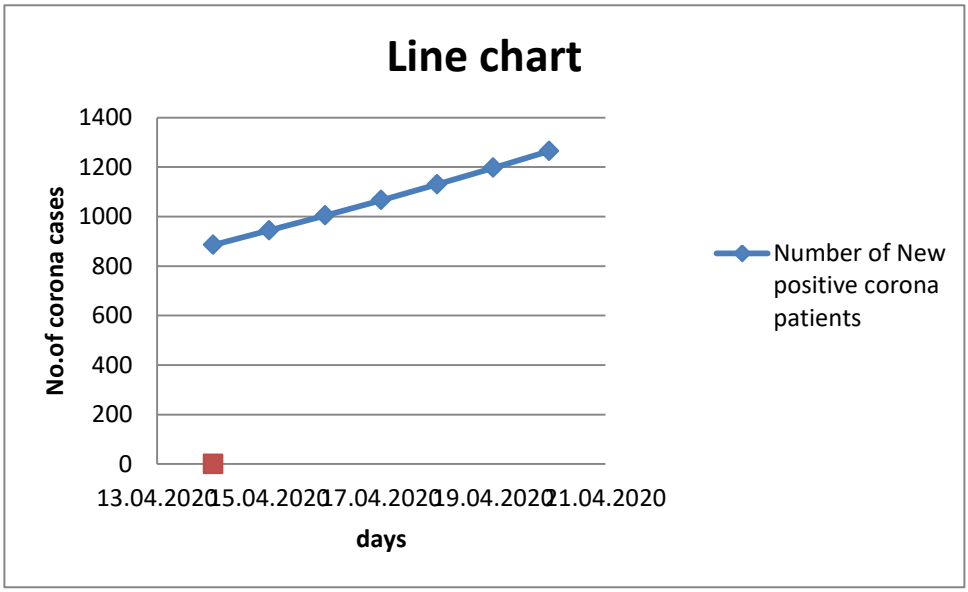
*Graphical representation:



Graph 2:hereabove scatter plot represent there is positive correlation between two variables.



Graph 3:here simple bar diagram represent no.of positive corona cases are increased as day by day.



Graph 4: above chart represent the estimated new corona patients/cases.

Conclusion:

By calculation and by graph 2,we conclude that there is no highly positive correlation but considerable correlation. We can also conclude that depending upon data we can fit non- linear polynomial model (second degree curve) for estimation or prediction purpose.
We cannot fit exponential curve as the growth of positive corona cases(y) is not larger rate with respect to days(x). Using Statistical model of polynomial we predicted the probable positive corona cases from 14 April to 18 April.

Overall, we can say that the days are increase as the number of positive cases also increases. As study of COVID 19 ,the above predicted/estimated figures for upcoming days will be high.

References:

1. https://www.who.int/news-room/q-a-detail/q-a-coronaviruses
2. https://www.worldometers.info/coronavirus/country/india/
3. https://en.wikipedia.org/wiki/Simple_linear_regression
4. https://timesofindia.indiatimes.com/india/coronavirus-in-india-live-updates-total-number-of-corona-cases-in-india-rises-to-3374-and-death-toll-climbs-to-77/liveblog/74975998.cms
5. www.techmaxbooks.com
6. https://en.wikipedia.org/wiki/Polynomial