International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 07 | July 2021 www.irjet.net p-ISSN: 2395-0072

REGRESSION ANALYSIS ON PREDICTION OF NUMBER OF DECEASED CASES COVID-19

Prof. Manishkumar Jaiswal

Department of Mathematics and Statistics, Shri Chinai College of Commerce & Econoimics, Andheri East, Mumbai-4000-69, Maharashtra, India

---

Abstract – ***Background:*** *The fastest-growing virus which has not only affected one or two countries but the whole world is nonother but coronavirus also known as COVID-19 (SARS-CoV-2). This virus had lead to a huge amount of human as well as economic loss in many countries. Coronavirus disease (COVID-19) is an infectious disease caused by a newly discovered coronavirus and it is a viral disease. By observing the impact due to disease, the researcher has decided to study statistics of confirmed cases, active cases, recovered cases, and deceased cases in India to give a prediction on future cases that might help to decide precautionary measures and infrastructure for COVID-19 patients.*

Methods: *The researcher has used regression analysis on secondary data collected from the www.COVID19india.org website about confirmed cases, active cases, recovered cases, and deceased cases in India.*

Findings: The researcher has found that the feasibility of the formula is precise and could help in the prediction of deceased cases in India caused by COVID-19.

Interpretation: *From the research, it is interpreted the multiple correlation coefficient very strong positive correlation between COVID-19 deceased cases and confirmed cases, active cases and recovered cases. The researcher has also interpreted that the overall regression analysis model is good fit.*

Key Words: **COVID-19, Correlation, Regression and COVID-19, Coronaviruses, Confirmed Cases, Active Cases, Recovered Cases, Deceased Cases**.

1. INTRODUCTION

The COVID-19 pandemic in India is part of the global pandemic of coronavirus sickness 2019 (COVID-19) resulting from extreme acute respiration syndrome coronavirus (SARS-CoV-2). The first case of COVID-19 in India, which originated from China, became said on 30 January 2020. Currently, India has the biggest range of showed instances in Asia. As of 12 June 2021, India has the second-maximum range of showed instances within side the world (after the US) with 29.three million said instances of COVID-19 contamination and the third-maximum range of COVID-19 deceased (after the US and Brazil) at 367,081 deceased. Coronavirus which is commonly called COVID-19 is an Infectious disease. This is a new virus and its spread ratio is very fast. COVID-19 was first identified in December 2019 in the wet market of Wuhan city in China. But now, this virus has spread to almost 200 countries of the world. On 11th March 2020, the World Health Organization(WHO) declared the COVID-19 outbreak pandemic symptoms- The most common symptoms of coronavirus are fever, cough, and shortness of breath, some patients may have aches and pains, nasal congestion, runny nose, loss of smell and sore throat, most people recover from the disease without needing special treatment. Older people and those with underlying medical problems like high blood pressure, heart problems, or diabetes are more likely to develop serious illnesses. COVID-19 refers to the disease caused in humans due to contraction of the Novel Human Coronavirus (Severe Acute Respiratory Syndrome CoV-2) which is established to be a respiratory pathogen. In human hosts, it causes symptoms that vary from dry cough and fever to ARDS-induced respiratory failure. This virus is of zoonotic origin, which essentially means that it started with an animal to human transmission or jumping. This puts into context its place of origin, which was early in December 2019 in the Wuhan Live Seafood Market, in the Hubei province of China, where trading of live animals including exotic mammals ran rampant and unchecked.

Then, this study aimed to achieve the following objectives:

- To predict COVID-19 deceased cases by daily regression tests in a linear regression analysis.
- To predict COVID-19 deceased cases by the number of days.

2. LITERATURE REVIEW

Rath, S., Tripathy, A., & Tripathy, A. R. (2020) reveals an effective way to forecast the next number of daily active cases. The researcher has also found that there is a strong relation between active cases and deceased and recovered cases.

Ogundokun, R., Lukman, A., & Kibria, G. (2020) aims to predict the prevalence of COVID-19 in Nigeria using a linear regression model. Also, to measure the impact of traveling history and contact on COVID-19 confirmed cases. The researcher concluded that traveling history and contacts increase people's chances of being infected with COVID-19 by 85% and 88% respectively and the government should enforce the right policy for the containment of COVID-19.

Gupta, A. K., Singh, V., Mathur, P., & Travieso-Gonzalez, C. (2020) the purpose of research to predict the active rate, the deceased rate, and the cured rate in India by analyzing the data of COVID-19. be the best predictive method for predicting active rate, deceased rate, and cured rate compared to SVM and Linear Regression when the vast uncertain and small data sets.

3. RESEARCH METHODOLOGY

Data Collection: The present research is based on secondary data collected from www.COVID19india.org. The date-wise data of confirmed cases, active cases, recovered cases, and deceased cases were collected from the period of 20th May 2021 to 20th June 2021 (for the 32 days). The collected data were analyzed by regression analysis to draw a research conclusion.

Table -1: CONFIRMED CASES, ACTIVE CASES, RECOVERED CASES, AND DECEASED CASES FROM THE PERIOD OF 20TH MAY 2021 TO 20TH JUNE 2021 (FOR THE 32 DAYS).

Date	Confirmed Cases	Active cases	Recovered cases	Deceased cases
20-May-21	259242	102171	357173	4209
21-May-21	257299	104595	357625	4194
22-May-21	240897	118063	355138	3739
23-May-21	222834	83904	302253	4454
24-May-21	195857	134479	326737	3509
25-May-21	208992	90474	295264	4160
26-May-21	211510	75500	283054	3843
27-May-21	186075	88712	271002	3659
28-May-21	174083	114950	285332	3611
29-May-21	165282	102705	264483	3463
30-May-21	153396	87442	237664	3130
31-May-21	126883	131096	255125	2783
1-Jun-21	133152	101493	231397	3205
2-Jun-21	134044	80800	211890	2898
3-Jun-21	132424	77062	206722	2712
4-Jun-21	120454	80762	197763	3372
5-Jun-21	114488	77617	189374	2682
6-Jun-21	101209	75431	174156	2444

7-Jun-21	85804	99141	182866	2107
8-Jun-21	92787	71800	162356	2222
9-Jun-21	93883	61373	149022	6139
10-Jun-21	91849	46652	135329	3414
11-Jun-21	84574	42148	122685	3996
12-Jun-21	80525	55500	132664	3300
13-Jun-21	71001	52574	119574	3922
14-Jun-21	60008	60152	117276	2733
15-Jun-21	62215	48118	107776	2540
16-Jun-21	67289	38965	103900	2329
17-Jun-21	62436	27673	88500	1591
18-Jun-21	60765	38760	97854	1645
19-Jun-21	58615	30589	87608	1574
20-Jun-21	52978	26659	78189	1424

Source: www.COVID19india.org.

The above table 1. reveals the number of confirmed cases, active cases, recovered cases, and deceased cases in India from 20th May 2021 to 20th June 2021 (for the 32 days)

4. RESULTS AND DISCUSSION

4.1. CORRELATION ANALYSIS

Table-2: CORRELATION ANALYSIS

	Confirmed Cases	*Active cases*	*Recovered cases*	*Deceased Cases*
Confirmed Cases	1			
Active cases	0.740172322	1		
Recovered cases	0.973409724	0.87439129	1	
Deceased Cases	0.576080851	0.389000838	0.540353489	1

The correlation analysis shows that:

- There is a significantly strong positive correlation between COVID-19 deceased cases and confirmed cases.
- There is a significantly moderate positive correlations between COVID-19 deceased cases and active cases.
- There is a significantly strong positive correlations between COVID-19 deceased and recovered cases.

- There is a significantly very strong positive correlations between COVID-19 recovered cases and confirmed cases.
- There is a significantly very strong positive correlations between COVID-19 recovered cases and active cases.
- There is a significantly very strong positive correlations between COVID-19 active cases and confirmed cases.

4.2. REGRESSION ANALYSIS

Regression analysis is a collection of statistical processes that determine the linkage between a dependent variable (otherwise referred to as" outcome variable") and a single or more independent (often referred to as "predictors") variable. Analysis of regression generates a regression equation where coefficient represents the relationship between each independent variable and the dependent variable. To make predictions, you can also use the equation. The benefit of regression analysis is that it helps to decide which variables are most relevant, which can be ignored and how those variables to relate to each other. In the present study, Multiple Linear Regression analysis is used to find the relation between a dependent and an independent variable.

For three independent variables, The Multiple Linear regressions equation is

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$$

Where,

Dependent variable (Y): Deceased cases.

Independent Variable X_1, X_2, X_3,: Confirmed Cases, Active Cases, and Recovered Cases respectively.

β_0 = Intercept, β_1 = Coefficient of Confirmed cases, β_2 = Coefficient of active cases and β_3 = Coefficient of recovered cases

Table-3: REGRESSION STATISTICS

Regression Statistics	
Multiple R	0.99822533
R Square	0.99645381
Adjusted R Square	0.996073861
Standard Error	62.94317499
Observations	32

Results showed that the multiple correlation coefficient R^2 is 0.99645381. This indicates that the strong positive correlation between COVID-19 deceased cases and confirmed cases, active cases, and recovered cases

The coefficient of determination, R^2 is 99.64% of the variation in the deceased cases variable is explained by confirmed cases, active cases, and recovered cases.

The standard error of the regression is 62.9431, which is an estimate of the variation of the observed recent number of deceased about the regression line.

Table-4: ANOVA TABLE

	Df	*SS*	*MS*	*F*	*Significance F*
Regression	3	31170979.61	10390326.54	2622.59908	2.15194E-34
Residual	28	110931.6118	3961.843278		

Total	31	31281911.22

F-Test statistics: F-test statistics is a test of significance for overall regression analysis

To test Hypothesis at $\alpha = 0.05$ Level of significance

Null hypothesis: $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ (There is no significant linear relationship between deceased cases and confirmed cases, active cases, and recovered cases)

Alternative hypothesis H_1 : At least one of the β_i Coefficients is not equal to 0 (i = 1, 2, 3)

The dependent variable (deceased cases) is regressed on the predicted variable of confirmed cases, active cases, and recovered cases. The independent variables significantly predict deceased cases, F (3,28) = 2622.59908, Significance F value(p-value) $<$0.05, Which is highly significant. Rejecting the null hypothesis at $\alpha = 0.05$ level of significance i.e accepting the alternative hypothesis at $\alpha = 0.05$ level of significance, it is an evidence that at least one of the independent variable (confirmed cases, active cases, and recovered cases) has a significant linear relationship between dependent variable (deceased cases)

This indicates that at least one of the independent variable (confirmed cases, active cases, and recovered cases) under study has a significant impact on deceased cases.

Table-5: REGRESSION COEFFICIENT

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	35.14032994	41.79419882	0.840794439	0.407586215	-50.47120541	120.751865
Confirmed Cases	0.985200037	0.013496324	72.99765781	1.70311E-33	0.95755407	1.01284600
Active cases	0.984548443	0.013673297	72.00519923	2.49375E-33	0.956539965	1.01255692
Recovered cases	-0.98512026	0.013631046	-72.2703362	2.25108E-33	-1.013042195	-0.95719833

t -test statistics: t-test statistics is a test of significance for each of individual independent variable.

To test Hypothesis at $\alpha = 0.05$ level of significance.

Null hypothesis: $H_0 : \beta_1 = 0$ (There is a no significant linear relationship between confirmed cases and deceased cases)

Alternative hypothesis $H_1 : \beta_1 \neq 0$ (There is a significant linear relationship between confirmed cases and deceased cases)

From the table of Regression Coefficient.

t-statistics value = 72.99765781, P-value(1.70311E-33) $<$0.05,

Rejecting the null hypothesis at $\alpha = 0.05$ level of significance i.e accepting the alternative hypothesis at $\alpha = 0.05$ level of significance, it is evidence that there is a significant linear relationship between confirmed cases and deceased cases. This is indicates that there is positive linear relationship between confirmed cases and deceased cases. Through it can be concluded:

β_1 coefficient for confirmed cases (X_1) is equal to 0.985200037. This means that for each single cases, increase in confirmed cases there is an increase in deceased cases by 98.20 per cent.

Null hypothesis: $H_0 : \beta_2 = 0$ (There is a no significant linear relationship between confirmed cases and deceased cases)

Alternative hypothesis $H_1 : \beta_2 \neq 0$ (There is a significant linear relationship between active cases and deceased cases)

From the table of Regression Coefficient.

t-statistics value = 72.00519923, P-value(2.49375E-33) $<$0.05,

Rejecting the null hypothesis at $\alpha = 0.05$ level of significance i.e accepting the alternative hypothesis at $\alpha = 0.05$ level of significance, it is evidence that there is a significant linear relationship between confirmed cases and deceased cases. This is indicates that there is positive linear relationship between active cases and deceased cases. Through it can be concluded:

β_2 coefficient for active cases (X_2) is equal to 0.98454843. This means that for each single cases, increase in active cases there is an increase in deceased cases by 98.45 per cent.

Null hypothesis: $H_0 : \beta_3 = 0$ (There is a no significant linear relationship between recovered cases and deceased cases)

Alternative hypothesis $H_1 : \beta_3 \neq 0$ (There is a significant linear relationship between recovered cases and deceased cases)

From the table of Regression Coefficient.

t-statistics value = -72.27033624, P-value(2.25108E-33) $<$0.05,

Rejecting the null hypothesis at $\alpha = 0.05$ level of significance i.e accepting the alternative hypothesis at $\alpha = 0.05$ level of significance, it is evidence that there is a significant linear relationship between recovered cases and deceased cases. This is indicates that there is negative linear relationship between recovered cases and deceased cases. Through it can be concluded:

β_3 coefficient for recovered cases (X_3) is equal to -0.985120263. This means that for each one cases increase in recovered cases, there is decrease in deceased cases by 98.51 per cent.

5. FINDINGS:

It is an evidence that confirmed cases, active cases, recovered cases have coefficient of 0.98520003, 0.98454843 and -0.985120263 respectively and the following regression equation can be derived from the available data for predicting deceased cases.

Regression analysis equations for finding predicted deceased cases.

$$\hat{Y} = 35.14032994 + (0.985200037)X_1 + (0.98454843)X_3 + (-0.985120263)X_3$$

Where, $\hat{Y}$ = Prediction of deceased cases , X_1 = confirmed cases, X_2 = active cases, X_3 = recovered cases

OR

Prediction of deceased cases = 35.14032994+0.985200037(confirmed cases) + 0.984548443 (active cases) – 0.985120263(recovered cases)

Table-6: RESIDUAL TABLE

Observation	Predicted Deceased Cases	Residuals	Standard Residuals
1	4174.307475	34.69252474	0.579948049
2	4201.334872	-7.334871786	-0.122615596
3	3751.9764	-12.97640034	-0.216923909
4	4423.20298	30.79701992	0.514827669
5	3519.31459	-10.31459041	-0.17242696

6	4139.552865	20.44713497	0.341810696
7	3905.97658	-62.97657998	-1.052766986
8	3727.937093	-68.93709329	-1.152407704
9	3629.226939	-18.22693916	-0.304696124
10	3441.458095	21.54190519	0.360111753
11	3124.147907	5.85209343	0.097828284
12	2781.832166	1.16783428	0.019522454
13	3187.397232	17.60276843	0.294261985
14	2909.6757	-11.67569983	-0.195180356
15	2724.511079	-12.51107934	-0.20914523
16	3400.188319	-28.18831884	-0.47121853
17	2690.253934	-8.253933735	-0.137979372
18	2447.119915	-3.119915014	-0.052155
19	2033.35945	73.6405499	1.231034454
20	2199.288914	22.7110859	0.379656715
21	6148.775124	-9.775124304	-0.163408813
22	3140.592379	273.4076208	4.570500924
23	3994.716531	1.283468752	0.021455492
24	3320.817292	-20.8172923	-0.347998543
25	3952.207643	-30.20764299	-0.504975171
26	2846.618109	-113.6181087	-1.89933137
27	2531.541123	8.458877456	0.141405375
28	2337.200347	-8.200346584	-0.137083566
29	1609.355599	-18.35559942	-0.306846912
30	1663.959988	-18.95998785	-0.316950353
31	1594.576795	-20.57679497	-0.343978197
32	1450.576565	-26.57656491	-0.44427516

6. Conclusion

In this research paper, the researcher has used regression analysis for predicting number of deceased cases due to COVID-19. The researcher has found that the multiple correlation coefficient very strong positive correlation between COVID-19 deceased cases and confirmed cases, active cases and recovered cases.

The researcher also found that there is positive linear relationship between confirmed cases and deceased cases that means for each single cases increase in confirmed cases, there is an increase in deceased cases by 98.20 per cent. There is positive linear relationship between active cases and deceased cases that means for each single cases increase in active cases, there is an increase in deceased cases by 98.45 per cent and there is negative linear relationship between recovered cases and deceased cases that means for each single cases increase in recovered cases, there is decrease in deceased cases by 98.51 per cent.

The conclusion through F-test, the researcher has found that the overall regression analysis model is good fit. The researcher also found that the predicted number of deceased cases of COVID-19 through linear regression analysis model is very close to actual number of deceased of COVID-19 according data available in www.COVID19india.org. and it is revealed in the following table:

Date	Actual Deceased Cases	Predicted Deceased Cases
20-May-21	4209	4174.307475
21-May-21	4194	4201.334872
22-May-21	3739	3751.9764
23-May-21	4454	4423.20298
24-May-21	3509	3519.31459
25-May-21	4160	4139.552865
26-May-21	3843	3905.97658
27-May-21	3659	3727.937093
28-May-21	3611	3629.226939
29-May-21	3463	3441.458095
30-May-21	3130	3124.147907
31-May-21	2783	2781.832166
1-Jun-21	3205	3187.397232
2-Jun-21	2898	2909.6757
3-Jun-21	2712	2724.511079
4-Jun-21	3372	3400.188319
5-Jun-21	2682	2690.253934
6-Jun-21	2444	2447.119915
7-Jun-21	2107	2033.35945
8-Jun-21	2222	2199.288914
9-Jun-21	6139	6148.775124
10-Jun-21	3414	3140.592379
11-Jun-21	3996	3994.716531
12-Jun-21	3300	3320.817292

13-Jun-21	3922	3952.207643
14-Jun-21	2733	2846.618109
15-Jun-21	2540	2531.541123
16-Jun-21	2329	2337.200347
17-Jun-21	1591	1609.355599
18-Jun-21	1645	1663.959988
19-Jun-21	1574	1594.576795
20-Jun-21	1424	1450.576565

The above research will help the stakeholders to predict the number of deceased cases through COVID-19 in India accordingly they can plan precautionary measure to overcome from this unexpected state of COVID-19 Pandemic.

REFERENCES

[1] Gupta, A. K., Singh, V., Mathur, P., & Travieso-Gonzalez, C. (2020). Prediction of COVID-19 pandemic measuring criteria using support vector machine, prophet and linear regression models in Indian scenario. *Journal of Interdisciplinary Mathematics*, *24*(1), 89–108. https://doi.org/10.1080/09720502.2020.1833458

[2] Kaya Uyanık, Gülden & Güler, Neşe. (2013). A Study on Multiple Linear Regression Analysis. Procedia - Social and Behavioral Sciences. 106. 234–240. 10.1016/j.sbspro.2013.12.027.

[3] Kumari, Khushbu & Yadav, Suniti. (2018). Linear regression analysis study. Journal of the Practice of Cardiovascular Sciences. 4. 33. 10.4103/jpcs.jpcs_8_18.

[4] Ogundokun, R., Lukman, A., & Kibria, G. (2020). Predictive modelling of COVID-19 confirmed cases in Nigeria. *Infectious Disease Modelling*, *5*, 543–548.

[5] Rath, S., Tripathy, A., & Tripathy, A. R. (2020). Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, *14*, 1467–1474.

[6] S. Shaikh, J. Gala, A. Jain, S. Advani, S. Jaidhara and M. Roja Edinburgh, "Analysis and Prediction of COVID-19 using Regression Models and Time Series Forecasting," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2021, pp. 989-995, doi: 10.1109/Confluence51648.2021.9377137.

BIOGRAPHIES

Prof. Manishkumar Jaiswal

Department of Mathematics and Statitstics, Shri Chinai College of Commerce & Economics, Andheri (E), Mumbai-400-069, Maharashtra, India.