

Statistical Analysis of Future Prospective in COVID-19 Cases in India

Ibrahim Abubakar Sadiq¹, Sanjeev Kumar Sharma², Rabilu Ado³

¹Department of Statistics, Ahmadu Bello University, Zaria, Nigeria.

^{1,2,3}Department of Mathematics and Statistics, Mewar University, Chittorgarh

2. DATA ANALYSIS

Abstract:- In this study, we have done regression analysis to get confirmed positive cases of Covid-19 in the peripheral of Indian. We have applied 5% level of significance to get an ideal mathematical model to reach on conclusion about positive cases of COVID-19. We have used two relevant statistical software R-Studio and Gretl to find residual plot of the fitted modeled, Actual confirmed cases, Active cases, Recovered cases, Forecasting about Actual fitted deceased cases. This statistical precise calculation may be useful to predict actual data of all above mentioned variables so that doctors and government give them proper treatment.

Keywords: COVID-19, R-Studio, Gretl, Model

1. INTRODUCTION

COVID-19 is described as infection by a novel corona virus. Now it is identified as rigorous acute respiratory syndrome corona virus two (SARS-CoV-2). It was acknowledged as 2019-nCoV formerly. Firstly symptom of this respiratory illness was seen in Wuhan City, Hubei Province, China. It was primarily notified by the World Health Organization on December 31, 2019. On January 30, 2020, the WHO announced the COVID-19 outbreak a world health emergency. On March 11, 2020, the WHO declared COVID-19 a world pandemic, it was first come in existence since declaring H1N1 respiratory tract infection as pandemic in 2009.

When a new virus exists, for the survival purpose, we often have to discover much more knowledge about it. How does COVID-19 transmit from one person to another person? As we know through printed, electronic media and medical staff, majority of respiratory viruses are used to spread by huge droplets that come out during human being cough as well as sneeze and remain in the open sky air generally for about five to six feet. They come down on materials surfaces, and transmitted one person to another person by either touching or holding those specified surfaces, like iron or any other hard surfaces. Some research also says that COVID-19 viruses have possibility to spread in the surrounding air and remain there for a longer time in the form of tiny droplet nuclei.

It's a technology era so we have got statistical data of COVID-19 positive patients from various reputed online sources like Ministry of Health and Family Welfare, a unit of Government of India, Worldometers, Covid19india.org, Biospectrumindia.com, CoronaTracker.Com, Mygov.in, Hindustan Times and Times of India.

2.1 Regression analysis

COVID-19 positive patients in India are increasing continue on a daily basis even with implementation of physical distancing of nationwide lockdown. As per the record of health ministry of India, developed and industrialist state Maharashtra, Gujarat and Delhi continue to remain the most affected states by the Covid-19 pandemic than others.

We have developed a precise statistical model concerned to confirmed, active, recovered and deaths cases in India due to this pandemic up to 13th May, 2020. The fitted model is **Confirmed Cases = -4.737e-14 + 1.000e+00(Active) + 1.000e+00(Recovered) + 1.000e+00 (Deceased)**

Descriptive Measures for Residuals:

| Min | 1Q | Median | 3Q | Max |
|------------|------------|-----------|-----------|-----------|
| -3.715e-13 | -2.282e-13 | 1.416e-14 | 8.212e-14 | 4.887e-13 |

Table 1

Coefficients:

| | Estimated Standard Error | t-value | P (> t) | |
|-------------|--------------------------|-----------|------------|------------|
| (Intercept) | -4.737e-14 | 4.762e-14 | -9.950e-01 | 0.327 |
| Active | 1.000e+00 | 4.126e-17 | 2.424e+16 | <2e-16 *** |
| Recovered | 1.000e+00 | 9.581e-17 | 1.044e+16 | <2e-16 *** |
| Deceased | 1.000e+00 | 6.490e-16 | 1.541e+15 | <2e-16 *** |

Significant-Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual Standard Error: 2.365e-13 on 32 degree of freedom
Adjusted R-Squared: 1, Multiple R-Squared: 1, F-Statistic: 2.965e+33 on 3 as well as 32 DF, P-value: < 2.2e-16

Above mentioned P-values (2e-16) for the front explanatory variable is far less compare to the alpha 5% level of significance which are adequately explained and predict about the future control of corona virus confirmed cases in India as at 13th May, 2020. The Residual standard error 2.365e-13 on 32 degrees of freedom, Multiple R-squared 1, Adjusted R squared 1, F-statistic: 2.965e+33 on 3 as well as 32 DF, p-value: < 2.2e-16 are supposed to give accurate and justified results and our results are same to them. Figure 1 depicts the residual plot of the fitted modeled with the help of R-studio which points lies between +4e-14 to -4e-14. Any fluctuations in confirmed cases are directly proportional to active cases, recovered and death cases respectively. They will increase or decrease by 1.000e+00, while rest variables remained constants.

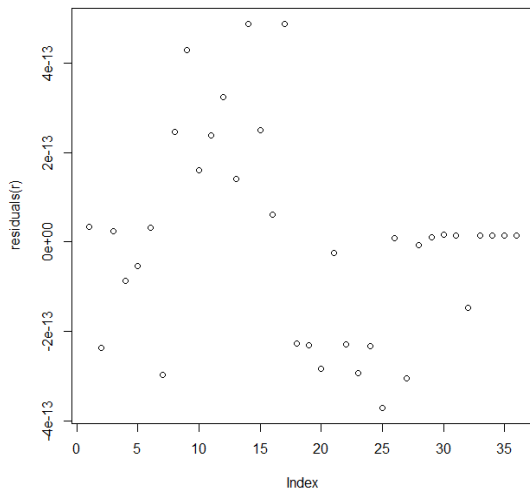


Fig. 1. Residual Plot of the Fitted Modeled

Table2
Dependent Variable (Confirmed Cases)

| | Coeff. | t-ratio | Standard Error | p-value |
|------------|---------|---------|----------------|------------|
| Constant | 0.58571 | 2.221 | 0.263752 | 0.0331** |
| State Unit | 0.95032 | 76.45 | 0.0124311 | <0.0001*** |

| Mean Dependent Variable | Sum Squared Residual | R-Squared | F(1, 34) | Log-likelihood | Schwarz Criterion |
|-------------------------|----------------------|--------------------|-------------|------------------|-------------------|
| 18.1666 | 20.4121 | 0.9942 | 5844.18 | -40.8687 | 88.9045 |
| S.D. Dependent Variable | S.E. of Regression | Adjusted R-Squared | P-value (F) | Akaike Criterion | Hannan-Quinn |
| 10.0413 | 0.7748 | 0.9940 | 1.24e-39 | 85.73749 | 86.84287 |

Table3
Confidence limits

| Variable | Coefficient | 95 confidence interval |
|------------|-------------|------------------------|
| Constant | 0.585714 | (0.0497060, 1.12172) |
| State Unit | 0.950322 | (0.925059, 0.975585) |

2.2 OLS Method for Confirmed Cases

The least squares fitted model for the COVID-19 provide the relation between confirmed cases and state wise units of corona virus is estimated to be Confirmed = 0.586 + 0.95(states) in thousands of reported cases per each state. Table 2 presents the model coefficient and its respective probability (p-values 0.0331 and < 0.0001). The statistical calculation of t-ratio is 2.221 and 76.45 according to table 2. We also has got testing the hypothesis of t ratio i.e., $H_0: \beta_0 = 0$. This statistic with $\beta_{0,0} = 0$, as $t_0 = 2.221$. Null hypothesis is rejecter due to less than 5% level of significance of p-value. The statistical value of t-test is large into the critical section, implying that $H_0: \beta_1 = 0$ must be also rejected. There is strong mathematical proof to maintain this claim; meanwhile the P-value is < 0.0001. We know that the t-statistical precise value for this slope is 76.45 and P-value is 0.0001. It is now evidently sufficient significance statistics for the well accurate and adequacy of the estimated confirmed cases modeled. The

estimates increase in reported confirmed cases each state of at least at the range of 950 across each state nationwide. The average statistical calculation of dependent variable is 18.16667 for the confirmed cases state wise. The variation across the state is 10.04134 which occur to be large sources of variation of the confirmed cases across the state. The sum of squares residuals and mean square error of regression serves the precise ANOVA for statistical testing H_0 (null hypothesis): $\beta_1 = 0$ is reviewed in the Table 2. The test statistics is $F(1, 34) = 5844.181$, for which that the P-value (F) = 1.24E-39, so we briefed that β_1 is not zero. The Log-likelihood is -40.86875, Akaike criterion is 85.73749, Schwarz criterion is 88.90453, Hannan-Quinn is 86.84287, Rho is 0.664243 and Durbin-Watson is 0.691710 shows statistical significances. Table 3 presents the confidence interval, so there is rigid proof (at $\alpha = 0.05$) that the gotten slope and intercept is not zero. Figure 2 is the graph of actual and fitted confirmed cases verses state unit which shows an increase trend with which in the long run will declined. Figure 3 is the graph representing the confirmed cases, forecast and 95 percent confidence limits for the forecast.

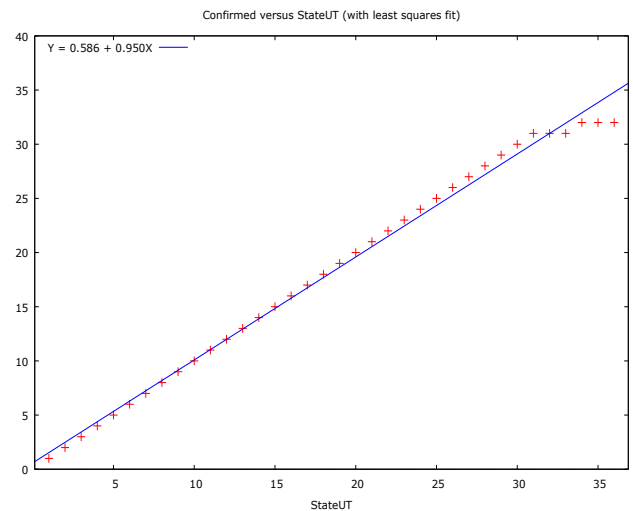


Fig. 2. Actual and Fitted Confirmed Cases verses State Unit

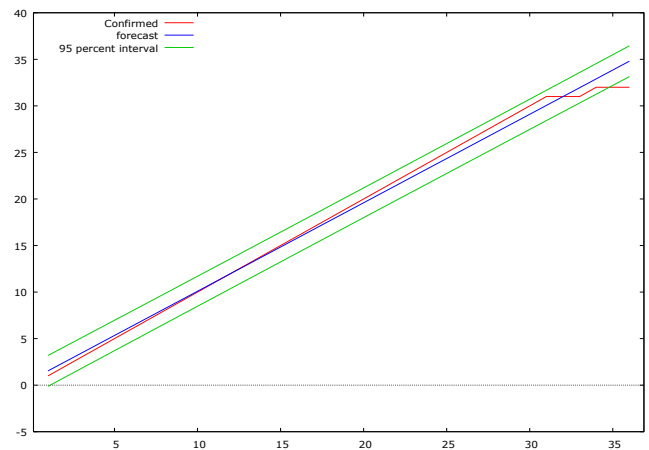


Fig 3. Confirmed Cases, Forecast and 95% Confidence limits

Table 4
Dependent variable (Active)

| | Coefficient | t-ratio | Standard Error | p-value |
|------------|-------------|---------|----------------|------------|
| Constant | 2.41905 | 3.203 | 0.755233 | 0.0029*** |
| State Unit | 0.765637 | 21.51 | 0.0355954 | <0.0001*** |

| Mean Dependent Variable | Sum Squared Residual | R-Squared | F(1, 34) | Log-likelihood | Schwarz criterion | Rho |
|-------------------------|----------------------|--------------------|-------------|------------------|-------------------|---------------|
| 16.58333 | 167.3625 | 0.931542 | 62.6553 | -78.74137 | 164.6498 | 0.664243 |
| S.D. Dependent Variable | S.E. of Regression | Adjusted R-squared | P-value (F) | Akaike criterion | Hannan-Quinn | Durbin-Watson |
| 8.357631 | 2.218655 | 0.929529 | 2.24e-21 | 161.4827 | 162.5881 | 0.691710 |

Table 5

Confidence limits

| Variable | Coefficient | 95 confidence interval |
|------------|-------------|------------------------|
| Constant | 2.41905 | (0.884230, 3.95387) |
| State Unit | 0.765637 | (0.693298, 0.837976) |

2.3 OLS method for Active cases

The least squares fitted model for the India active cases and states units of corona virus is estimated to be Active = 2.41905+ 0.766(states) in thousands of reported cases per each state. Table 4 presents the model coefficient and its respective probability (p-values 0.0029 as well as < 0.0001). The reference value of t-ratio is 3.203 and 21.51. The analysis also state that t-statistics for precise testing of null hypothesis is $H_0: \beta_0 = 0$. This statistic stand with $\beta_{0,0} = 0$, for $t_0 = 3.203$. Obviously, this hypothesis which intercept is 0 that is rejected. The statistical value of test is very away towards the critical region, involving that $H_0: \beta_1 = 0$ must be rejected. The required P-value for statistical test is < 0.0001. Notice that the t-statistical precise value for this incline is 21.51 and report P-value is P = 0.0001. It is now evidently sufficient significance statistics for the well accurate and adequacy of the estimated active cases modeled. The estimates tends to increase in reported active cases each state of at least at the range of 766 across each state nationwide. The average measure of dependent variable is 16.58333 location of distribution centre for the active cases state wise. The variation across the state is 8.357631 which occur to be large sources of variation of the active cases across the state. The sum of squares residuals and mean square error of regression serves the precise calculation of variance for this testing $H_0: \beta_1 = 0$ is briefed in the table 4. The statistical test is $F(1, 34) = 462.6553$, for which that the P-value (F) = 1.24E-21, so we reached on the conclusion that β_1 is not zero. The Log-likelihood is -78.74137, Akaike criterion is 161.4827, Schwarz criterion is 164.6498, Hannan-Quinn is 162.5881, Rho is 0.664243 and Durbin-Watson is 0.691710 shows statistical significances. Table 5 presents the confidence interval, so here is strong proof (at $\alpha = 0.05$) that the gotten slope and intercept is not 0. Figure 4 is the graph of actual and fitted active cases verses state unit which shows an increase trend with which in the long run will declined. Figure 5 is the graph representing the active cases, forecast and 95 percent confidence limits for the forecast.

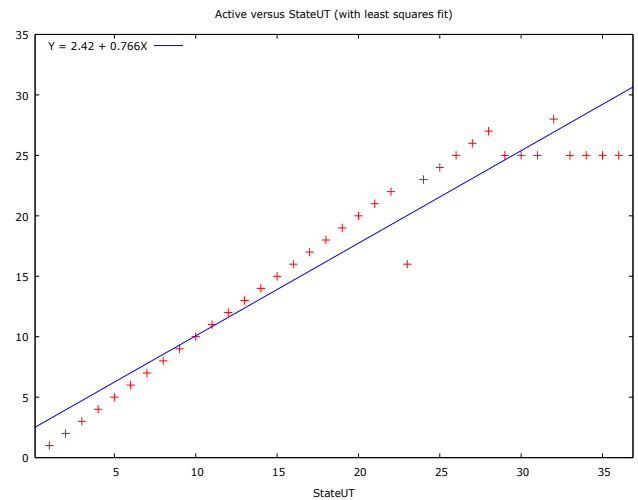


Fig. 4. Actual and Fitted Active cases verses State Unit

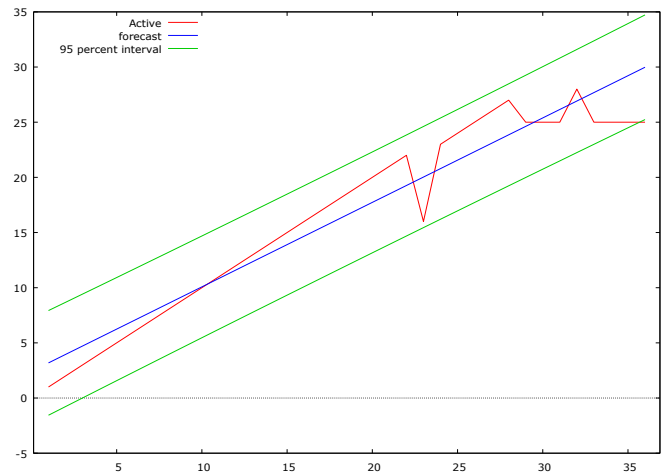


Fig. 5. Active cases, forecast and 95 percent confidence limits

Table 6
Dependent variable (Recovered)

| | Coefficient | t-ratio | Standard Error | p-value |
|------------|-------------|---------|----------------|------------|
| Constant | 1.08095 | 1.827 | 0.591600 | 0.0765* |
| State Unit | 0.901030 | 32.31 | 0.0278831 | <0.0001*** |

| Mean Dependent Variable | Sum Squared Residual | R-Squared | F(1, 34) | Log-likelihood | Schwarz Criterion | Rho |
|-------------------------|----------------------|--------------------|-------------|------------------|-------------------|---------------|
| 17.75000 | 102.6959 | 0.968467 | 1044.227 | -69.95034 | 147.0677 | 0.060889 |
| S.D. Dependent variable | S.E. of Regression | Adjusted R-squared | P-value (F) | Akaike Criterion | Hannan-Quinn | Durbin-Watson |
| 9.646243 | 1.737949 | 0.967539 | 4.16e-27 | 143.9007 | 145.0061 | 1.814611 |

Table 7

Confidence limits

| Variable | Coefficient | 95 confidence interval |
|------------|-------------|------------------------|
| constant | 1.08095 | (-0.121324, 2.28323) |
| State Unit | 0.901030 | (0.844364, 0.957695) |

2.4 OLS method for Active cases

The least squares fitted model for the India recovered cases and states units of corona virus is estimated to be Recovered = 1.081+ 0.901(states) in thousands of reported cases per each state. Table 6 presents the model coefficient and its respective probability (p-values 0.0765 as well as < 0.0001). The reference value of t-ratio is 1.827 and 32.31. The analysis also state that t-statistic for precise testing the null hypothesis $H_0: \beta_0 = 0$. This precise statistic with $\beta_{0,0} = 0$, for $t_0 = 1.827$. Obviously, the null hypothesis of the statistical intercept is 0 and not rejected. The numerical value of the statistical test is away within the critical area, indicating that null hypothesis $H_0: \beta_1 = 0$ must fail to accept. The statistical P-value for this precise test is < 0.0001. Notice that the statistical t-value for the incline is 32.31 and that the observed P- value is P = 0 0001. It is now evidently sufficient significance statistics for the well accurate and adequacy of the estimated recovered cases modeled. The estimates tends to increase in reported recovered cases each state of at least at the range of 901 across each state nationwide. The average measure of dependent variable is 17.75000 location of distribution centre for the recovered cases state wise. The variation across the state is 9.646243 which occur to be large sources of variation of the active cases across the state. The sum of squares residuals and mean square error of regression serves the precise analysis of variance for null hypothesis $H_0: \beta_1 = 0$ is briefed in the table 6. The statistical test is $F(1, 34) = 1044.227$, for which that the P-value (F) = 4.16E-27, so we conclude that β_1 is not zero. The Log-likelihood is -69.95034, Akaike criterion is 143.9007, Schwarz criterion is 147.0677, Hannan-Quinn is 145.0061, Rho is 0.060889 and Durbin-Watson is 1.814611 shows statistical significances. Table 7 presents the confidence interval, so here is big proof (at $\alpha = 0.05$) that the gotten slope and intercept is not 0. Figure 6 is the graph of actual and fitted recovered cases verses state unit which shows an increase trend with which in the long run will declined. Figure 7 is the graph representing the recovered cases, forecast and 95 percent confidence limits for the forecast.

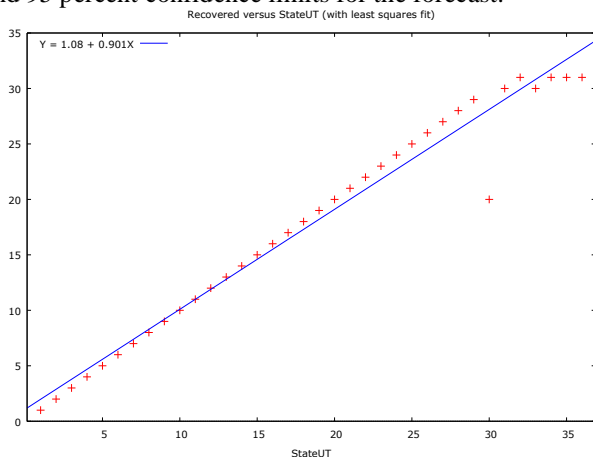


Figure 6. Actual and fitted recovered cases verses state unit

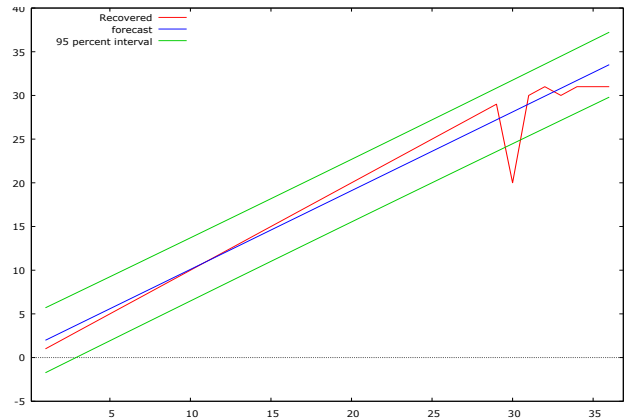


Fig 7. Recovered cases, forecast and 95% confidence limit

Table 8

Dependent variable (Deceased)

| | Coefficient | t-ratio | Standard Error | p-value |
|------------|-------------|---------|----------------|------------|
| Constant | 206.200 | 4.707 | 43.8067 | <0.0001*** |
| State Unit | -7.97928 | -3.865 | 2.06468 | 0.0005*** |

| Mean Dependent Variable | Sum Squared Residual | R-Squared | F(1, 34) | Log-likelihood | Schwarz criterion | Rho |
|-------------------------|----------------------|--------------------|-------------|------------------|-------------------|---------------|
| 58.5833 | 563089.1 | 0.3052 | 4.9355 | -224.919 | 157.0069 | 0.417425 |
| S.D. Dependent Variable | S.E. of Regression | Adjusted R-squared | P-value (F) | Akaike Criterion | Hannan-Quinn | Durbin-Watson |
| 52.1693 | 128.6913 | 0.284773 | 0.000476 | 153.8398 | 154.9452 | 0.564198 |

Table9

Confidence limits

| Variable | Coefficient | 95 confidence interval |
|------------|-------------|------------------------|
| Constant | 206.200 | (117.174, 295.226) |
| State Unit | -7.97928 | (-12.1752, -3.78334) |

2.5 OLS method for Active cases

The least squares fitted model for the India death cases and states units of corona virus is estimated to be death = 206.20 – 7.98(states) in thousands of reported cases per each state. Table 8 presents the model coefficient and its respective probability (p-values <0.0001 and 0.0005). The reference value of t-ratio is 43.8067 and 2.06468. The precise calculation also state that the t-statistic for testing the null hypothesis $H_0: \beta_0 = 0$. This precise statistic with $\beta_{0,0} = 0$, for $t_0 = 43.8067$. It's clear, the null hypothesis that the statistical intercept is 0 which is rejected. The numerical value of the statistical test is far within this critical region, indicating that $H_0: \beta_1 = 0$ must be rejected. The statistical P-value for this precise test is 0.0005. Notice that the statistical t-value for the incline is 2.06468 and that the examined P-value is P = 0 0005. It is now evidently sufficient significance statistics for the well accurate and adequacy of the estimated death cases modeled. The estimates tends to decrease in reported death cases each state of at least at the range of 797 across each state nationwide. The average measures of dependent variable are 58.58333location of distribution centre for the death cases state wise. The variation across the state is 152.1693 which

occur to be highly large sources of variation of the death cases across the state. The sum of squares residuals & mean square error of regression serves the precise ANOVA for null hypothesis $H_0: \beta_1 = 0$ is briefed in the table 8. The precise test statistic is $F(1, 34) = 14.93551$, for which that the P-value (F) = 0.000476, so we conclude that β_1 is not zero. The Log-likelihood is -224.9199, Akaike criterion is 453.8398, Schwarz criterion is 457.0069, Hannan-Quinn is 454.9452, Rho is 0.417425 and Durbin-Watson is 0.564198 shows statistical significances. Table 9 presents the confidence interval, so there is big proof (at $\alpha = 0.05$) that the slope and intercept are not zero. Figure 8 is the graph of actual and fitted death cases verses state unit which shows a declining trend. Figure 9 is the graph representing the death cases, forecast and 95 percent confidence limits for the forecast.

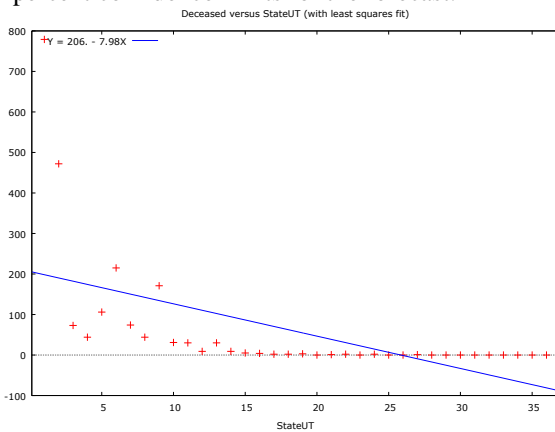


Fig. 8. Actual and fitted deceased cases verses state unit

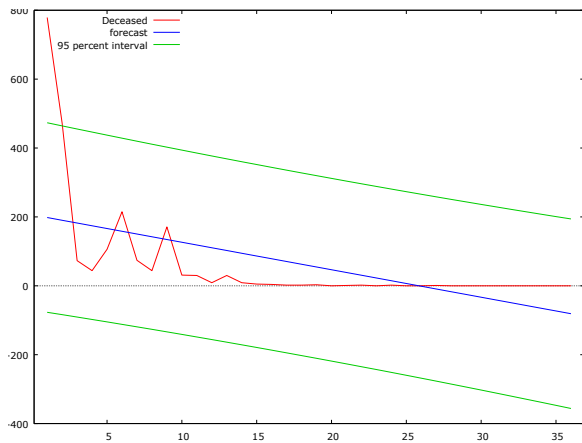


Fig 9. Deceased cases, forecast and 95 % confidence limits.

CONCLUSION:

Thus, as of May 18, 2020, with the consequences of statistical and precise data processing, it is tried to predict the future spread increment or reduction variables of this pandemic in India during lockdown. This paper describes it to look all the constraints and limited mass gathering in precise manner. To get systematic and precise result, R-studio and Gretl software

were used to find testing variables like confirm, active, recover and deceased COVID-19 cases in India at 5% level of significance. With farsighted, India had imposed lockdown and containment at early stage consequently till today we have 96000 COVID-19 patients. Being countable patients in thousands, doctors are also giving full of attention on them consequently recovery rate is high till now. If lockdown was not declared on the appropriate time, condition of India was going to very worst. Community spreading was going to become common; doctors can't do widespread testing to all, being limited sources. Consequently, numbers of positive cases were going to become more than 1,800000. Our research says, more lockdown are mandatory in India especially in red and orange zone till vaccination although green zone may be open with full vigilance, rigid social distancing, thermal temperature detector, proper hand wash and mask must be compulsory. For the further observations and research, this study may be very beneficial to scholars, Bio-researchers and scientists.

REFERENCES:

- [1] Ministry of Health and Family Welfare, a unit of Government of India, 18 May 2020.
- [2] Read, M., Bridgen, R., Cummings, A., & Jewell, C. P. (2020). nCoV: Early Estimation of epidem. Para. and epidemic pred. <https://doi.org/10.1101/2020.01.23.20018549>
- [3] Koh, G. (2020). F1000Prime recomm. of report of the WHO-China joint mission on
- [4] coronavirus disease 2019 (COVID-19). F1000 - Post-publication peer review of the biomedical
- [5] lit. <https://doi.org/10.3410/f.737509210.793572110>
- [6] Maier, F., & Brockmann, D. (2020). Effective containment explains sub-expon. growth in conf. cases of recent COVID-19 outbreak in mainland china
- [7] Danon, L., Brooks E., Bailey, M., & Keeling, J. (2020). A spatial model of COVID-19 trans. in England and Wales: Early spread and peak timing. <https://doi.org/10.1101/2020.02.12.20022566>
- [8] Jin, H., Liu, J., Cui, M., & Lu, L. (2020). Novel C.V. pneumonia emergency in Zhuhai: Impact and challenges. Journal of Hospital Infection. <https://doi.org/10.1016/j.jhin.2020.02.005>
- [9] Topcuoglu, N. (2020). Public health emergency of international concern: Coronavirus disease 2019 (COVID-19). The Open Dentistry Journal, 14(1), 71-72. <https://doi.org/10.2174/1874210602014010071>
- [10] Niederberger, C. (2020). Re: Rapid sperm capture: high-throughput Flagellar waveform
- [11] analysis. Journal of Urology. <https://doi.org/10.1097/ju.0000000000000722>
- [12] Erasmus, Z. (2020). "Who was here first?"; or "Who lives here now?": Indigeneity, a difference like no other. Persistence of Race, 155-165. <https://doi.org/10.18820/9781928480457/09>
- [13] Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., Wang, B., Xiang, H., Cheng, Z., Xiong, Y., Zhao, Y., Li, Y., Wang, X., & Peng, Z. (2020). Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. JAMA, 323(11), 1061. <https://doi.org/10.1001/jama.2020.1585>
- [14] Koh, G. (2020). F1000 Prime recommend. of report of the WHO-China joint mission on
- [15] CV disease 2019 (COVID-19). F1000 - Post-publication peer review of the biomedical
- [16] literature. <https://doi.org/10.3410/f.737509210.793572110>.