

Original Research Paper

Prediction and Analysis of COVID-19 Cases using Regression Models: A Descriptive Case Study of India

Ismail Keshta

Department of Computer Science and Information Systems, Al Maarefa University, Riyadh, Saudi Arabia

Article history

Received: 15-07-2022

Revised: 23-09-2022

Accepted: 30-09-2022

Email: imohamed@mcst.edu.sa

Abstract: The Coronavirus (SARS-CoV-2) is a respiratory illness that emerged in Wuhan, China, on December 31, 2019, according to reports by the World Health Organization (WHO). Hospital records in India indicate that COVID-19 hospitalizations in the second wave of COVID-19 more than doubled those in the first wave. Limited studies have been conducted to establish the extent to which the second wave of COVID-19 increased the infection status and determine the effectiveness of the intervention strategies employed by the Indian government. This study employed regression models to establish the extent to which government interventions helped reduce the prevalence of COVID-19 in India, focusing specifically on the Kanyakumari, Tirunelveli, Thoothukudi, and Tenkasi districts. The researcher relied on daily Ministry of Health reports on COVID-19 and generated data for further analysis using Statistical Package for Social Sciences (SPSS) software. Findings from the regression analysis show an abnormal rise in the number of COVID-19 cases as well as in the number of deaths. However, government interventions such as stay-at-home orders, social distancing, vaccination, and prohibition of social gatherings, among others, helped to significantly reduce the number of COVID-19 cases and deaths in the country. This study thus recommends that the healthcare sector in India should create long-term interventions to improve safety and well-being during emergencies.

Keywords: COVID-19, Virus, Pandemic, India

Introduction

A. Background Information

Coronavirus (SARS-CoV-2) is a respiratory illness that emerged in Wuhan, China, on December 31, 2019, according to reports by the World Health Organization (WHO) (United Nations Development Programme, 2021). The Coronavirus Disease (COVID-19) was later declared a global health emergency and eventually a pandemic by the WHO on March 11, 2020, due to its rapid spread across many countries (Maddock, 2020). The United Nations Development Program (2021) suggested that it is the greatest challenge the world has faced since World War II, with the infection spreading to every continent. Currently, Worldometer (2021) shows that the number of infected patients is approaching 24 million, while critical cases amount to more than 108,000. India has been a region of concern since it was one of the areas severely affected by the second wave of COVID-19. Furthermore, a significant number of infections in India resulted in deaths; more than 500,000 people died and consequently had confirmed novel coronavirus lab results.

The downside of COVID-19 is that there is a deficiency of effective antiviral medication and vaccines that can counter its rapid spread by providing passive immunity. Additionally, several strains of the virus seem to have surfaced, which causes a tremendous challenge in vaccine compliance. As a result, efforts to mitigate the illness in India became strained due to pandemic fatigue (Frontiersin, 2021). Healthcare practitioners on the frontlines were exhausted from treating numerous patients who were infected and re-infected. Communities in India suffered from pandemic fatigue, attributed to prolonged exposure to mitigation measures, including staying indoors and wearing masks (Frontiersin, 2021). Moreover, citizens saw their close friends and families die daily due to the disease with no signs of improvement, thus causing stress and fatigue not only to COVID-19 patients but also to the medical system as a whole.

Literature confirms that the recent outbreak of COVID-19 in various countries across the globe has exhibited similar characteristics to those exhibited during the SARS outbreak that occurred in 2003 in China and the

Middle East Respiratory Syndrome (MERS) outbreak of 2012 in Saudi Arabia. COVID-19 has been linked to the SARS and MERS viruses because their effects appear in the respiratory tract. Based on these previous outbreaks, statisticians and scholars have the statistician tools for understanding the necessary relationships using the existing signs and symptoms of COVID-19.

Prediction and forecasting new infections and the characteristics of the various related outbreaks are essential for planning within each nation's healthcare and financial sectors. Therefore, regression models are vital in the study of COVID-19 in India, to comprehensively understand the infection status, the effectiveness of interventions, and the outcomes of the government's planning and decision-making for future prevention strategies.

It is important to note that even though the first case of COVID-19 in India was reported in January 2020 (Ketu and Mishra, 2022; Kumar and Kumar, 2022), it took more than a year for the initial wave of infection to spread throughout the entire country. In comparison to many other nations, India had a relatively low number of daily confirmed cases per million people throughout the first wave. However, this began to change in March 2021, when a sharp increase in the number of COVID-19-positive cases was reported across the country. The first wave of COVID-19 positive cases/day in India began in March 2020, reached its peak in September 2020 with more than 90,000 confirmed cases/day, and gradually decreased in intensity until it reached 10,000 confirmed cases/day in February 2021, according to the repository known as "our world in data" (Kumar and Kumar, 2022; Mohan *et al.*, 2022). The first wave of COVID-19 appeared in most countries and continents before August 2020, except for a few countries, notably India. The second wave of COVID-19 began to emerge in August-September 2020 (Guleria *et al.*, 2022; Mohan *et al.*, 2022), while the third wave appeared in March 2021 and is still active (Mohan *et al.*, 2022).

This study is a modest effort to analyze COVID-19 infections in India based on the infection rates, their effects, and the prevention measures put in place by local and national governments. This study used regression tools to explore various daily COVID-19 infection rates in 2020.

B. Justification for the Study

Although numerous studies have focused on understanding India's infection rate through the analysis of the dynamics of COVID-19, few studies exist that analyze the spread and effects of the virus at the state level (Pinchoff *et al.*, 2020). Based on India's population diversity, population density, and geographical conditions, the study of India as a country provides unsatisfactory results for understanding the actual status of the spread and effects of the virus. Individual analysis of the Indian states by the district is vital since

the large population in India compared to other parts of the world is crucial to consider in making future decisions about the containment of the pandemic.

C. Broad Objective

To present an overview of COVID-19 in India using regression models

D. Specific Objectives

Based on the above broad objective, the following constitute the specific objectives:

- [1] To identify the COVID-19 infection status in India
- [2] To present challenges and concerns regarding different consequences of COVID-19 in India
- [3] To establish the effectiveness of the intervention strategies employed by the Indian government

E. Hypothesis

H₀: There is no significant relationship between the COVID-19 infection status in India and the intervention strategies employed by the Indian government

H₁: There is a significant relationship between the COVID-19 infection status in India and the intervention strategies employed by the Indian government

Literature Review

A. Overview of the current state of India

India is among the top countries seriously affected by the COVID-19 pandemic, due to high infection and daily death rates (Kotwal *et al.*, 2020). With a shocking increase in the death toll during the second wave, the pandemic swept through India in 2021 to the extent that it staggered scientists across the globe. With approximately 273,810 new infections daily, the Indian government has worked constantly to ensure that all the necessary medical resources are available and practices are enforced. The discovery of the other vaccines approved by the WHO, such as AstraZeneca, Moderna, Johnson & Johnson's, and Pfizer, led to the protection of a majority of the Indian population, especially the elderly, decreasing infections as well as the severity of the virus's effects (Changotra *et al.*, 2021). However, the virus's unpredictability has led to the emergence of more deadly waves with more lethal effects.

COVID-19 symptoms include high fever, body pain, dry cough, and respiratory distress, with the most severe conditions resulting in death (Jaipuria *et al.*, 2021). Its incubation period varies, ranging from 2 to 14 days. The virus is transmitted through respiratory droplets and contact transmission, making it one of the deadliest contagious infections ever, throughout the world. In addition, some infected people are asymptomatic and do

not readily show any signs of disease. These patients, the silent spreaders of COVID-19, are the most dangerous and most challenging to reach for the enforcement of appropriate healthcare measures. This has been a major challenge in India. The percentage of infected patients has been increasing, with the highest death tolls recorded in 2021 when the second wave of the virus, more contagious and impactful than the first wave, emerged.

B. Measures Taken by the Government to Curb the Spread

Random testing aids in containing the spread of the virus at both community and societal levels in India. To ensure that silent spread effects are less rampant, India undertakes random testing at a cluster level. 10,500 people have been tested in the city of Chennai; 30,500 have been tested in Mumbai and 10,300 have been tested in Ahmedabad (Kotwal *et al.*, 2020). Moreover, to further contain the spread of the virus and prevent human-to-human transmission, the federal government announced and enforced a nationwide lockdown for 21 days in March, followed by subsequent lockdowns imposed based on the rates of confirmed infection in the country (Kotwal *et al.*, 2020). The lockdowns had devastating effects on the economy, due to pressure from small and large-scale producers (Kumar *et al.*, 2020). The government's attempts to recover from the impact on the economy while containing the effects and spread of the virus saw the introduction of numerous packages, totaling approximately \$20 trillion, which aimed to support various groups negatively affected by the pandemic. In January 2021, the country started COVID-19 vaccination programs, aiming to shield vulnerable populations such as the elderly from the effects of the virus. Vaccination exercises were conducted across the country, along with other measures such as hand hygiene, compulsory use of masks, public gathering restrictions, social distancing, increased sample testing, and improvement of quarantine facilities. The effectiveness of COVID-19 vaccines in reducing the death rate has been insignificant, due to the low vaccination rate and the continuous transformation of the virus, as indicated by the different waves in India (Thayer *et al.*, 2021). However, they have been instrumental in reducing the overall effects of the virus, hence aiding in the minimization of its impact.

C. Challenges Faced in COVID-19 Prevention Measures

The research on the challenges facing India amid the COVID-19 pandemic relates to other research on the implications of the pandemic on individual countries, cities, municipalities, and communities. Folliot analyzed the challenges that local and national governments face in combating COVID-19 in Middle Eastern countries, including India (2020). The author pointed out that the

central governments in these countries have limited financial resources to fund emergencies of this magnitude. Research shows that most Indian economies are too weak to contribute the money required to effectively implement underlying response strategies. The intervention strategies to halt the spread of COVID-19 have financial, cultural, social, and economic implications on all populations. Indian local authorities face the challenge of balancing the implementation of intervention strategies with the maintenance of a functional society. For instance, the WHO's recommended pandemic intervention strategies include shutting down cities, with adverse case reporting; the secession of movement; and isolation, quarantine, and treatment of positive cases (Worldometer, 2021). These measures are not popular with the public, who cite health challenges like fatigue, poor psychological health, and related personal well-being concerns.

The Indian government has rolled out national strategies in line with the WHO's recommendations to limit the spread of COVID-19 (2020). According to Anwar *et al.* (2020), local authorities implement national government strategies and policies (2020). The Indian Ministry of Health spearheads the implementation of the COVID-19 management plan in respective administrative units. Intervention strategies like regular spraying of bus stops and related amenities have relatively high costs. The Indian government has the limited financial capability to fund these strategies and cannot hence implement them. The installation of hand washing points in Indian cities and at strategic points is delegated to local authorities under central government funding. Additionally, the WHO recommends installing automatic body sanitization units, especially in crowded municipalities (2020). This recommendation is inadequately implemented in local administrative unities in India due to limited financial capacities (Zhang *et al.*, 2020). Furthermore, measures like installing handwashing sites in public places present other challenges to local authorities. Zhang *et al.* (2020) note that the world struggles to safely dispose of wastewater from sanitization and handwashing sites. The wastewater could be infected with COVID-19 and hence poor disposal could lead to further spread.

According to Rowan and Laffey (2020), local hospitals and community health centers are on the frontlines in the acquisition of COVID-19 infection statistics in any population. Local health centers fall under the management of local councils in the Indian administration hierarchy. The local authorities should ensure an adequate supply of PPE and drugs to treat COVID-19 symptoms. However, the world has witnessed an acute shortage of PPE, including body covers, safety gloves, and masks (Wang *et al.*, 2020). The shortage has also affected the local administrations in India, inhibiting the acquisition of COVID-19 statistics and the effective treatment of diagnosed patients. Health center workers in local health facilities are regularly exposed to the virus, increasing the risk of spread.

Maddock (2020) stated that such measures as mandatory curfews and closures reduced the infection rates in different districts in India. Social gathering places, including bars and restaurants, were gradually reopened due to decreasing cases of the virus. However, the reopening of schools and social gathering premises, as argued by Frontiersin (2021), caused the second wave of COVID-19, which surpassed the first wave. According to Young (2020), when the second wave hit, most hospitals in India saw more than 1,000 cases in a month; whereas only 485 patients were hospitalized the previous month. Consequently, many people had to be put on ventilators as a result of the relaxation of public health measures (Young, 2020).

Stress on healthcare practitioners in India was among the factors that hindered the mitigation of COVID-19 (Young, 2020). They had a higher exposure frequency to the virus and opted not to take additional shifts due to burnout and fear of infection. As a result, Malavika *et al.* (2021) argued that healthcare centers were understaffed and patients were not well managed, which led to a surge in cases.

Pandemic fatigue was another factor that hindered the mitigation of COVID-19 in India. The second wave left people tired and overworked, therefore letting down their guard concerning the recommended measures to reduce the spread of the virus (Young, 2020). Public gatherings have increased recently, despite the dangers that they pose for the spread of the virus and often people are spotted without face masks at public gatherings. This phenomenon has greatly hindered the mitigation of COVID-19.

Furthermore, the continuous movement of people in and out of towns in India hindered the mitigation of COVID-19. The airport, bus stations, and railway stations in India are busy, with people coming in and out. The continuous movement of people has further rendered the state vulnerable to the virus (Folliot, 2020). Additionally, Young (2020) argued that poor community participation impeded COVID-19 mitigation in most parts of the country. For a disaster to be mitigated, the public must be involved. In India, however, community participation in the fight against COVID-19 at the beginning of the pandemic was not at its ideal level (Young, 2020). Some residents ignored the safety measures intended to mitigate the disease.

Materials and Methods

A. Research Design

A quantitative design was applied in the present study to establish statistical relationships between the dependent variables and the dependent variables. As stated by Malavika *et al.* (2021), the quantitative approach is the best study design when the researcher is interested in comparing relationships between variables or performing cause-effect analysis. Maddock (2020) studied the impact of COVID-19 on the economy using the quantitative approach. The same design has been adopted for the

current study. The association between the prevalence of COVID-19 and the effects such prevalence has on general well-being and economic performance in India can only be effectively established through a study in which quantitative data is involved. This will help answer in terms of numbers and figures, illustrating the extent of the association concerning the relevant numerical strength.

While COVID-19 has been scientifically confirmed to spread in various ways, new factors influencing its spread are still being discovered. Prediction of the space and the exact effects of the virus is highly related to feature-reproduction rates. Thus, data science can be applied to ensure the tracking of the crucial features used for predicting any number of features. The available traditional quantitative approaches, such as the Pearson correlation coefficient and Chi-square, provide essential features about others (Bandeekar and Ghosh, 2022). The selection of reliable and effective methods dramatically relies on the desired role of the information, the effectiveness of the tool in the provision of the data needed, and the type of data used in the analysis. More importantly, feature selection reduces the under-fitting and over-fitting problems, time, and computational costs.

B. Data Collection and Sampling Procedure

It is important to note that this study specifically used secondary data generated from Ministry of Health and Family Welfare records. As required by the World Health Organization, the India Ministry of Health maintains updated health records regarding the prevalence of COVID-19 and new cases daily. Moreover, the government also reports on the different strategies undertaken by the Ministry of Health to contain the further spread of the virus. As a result, the researcher considered it sufficient to use such data to establish an overview of COVID-19 in India through the use of regression models. Specifically, this method allowed the researcher to identify the COVID-19 infection status in India, present challenges and concerns regarding the consequences of COVID-19 in India, and establish the effectiveness of the intervention strategies employed by the Indian government.

The reproduction rate prediction is instrumental, as it is associated with COVID-19 infection status. Feature ranking was undertaken using various tools such as the Random Forest Regression, XGBoost, and Gradient Boosting (Senapati *et al.*, 2021). The study considers five factors within the regression models: The total number of recovered cases, the number of new daily infections, the death rate, the number of tests conducted, and the death toll within the country. Since researching a population of approximately 1.43 billion people is a tedious process and may result in wrongful prediction, a sample representation of the Indian population in its four significant districts was vital for understanding the characteristics of the larger population. Thus, the Kanyakumari, Tirunelveli,

Thoothukudi, and Tenkasi districts represent the larger Indian population; data were collected through a random sampling method in 2020.

C. Variables

This study aims to provide an overview of the COVID-19 pandemic in India through the use of regression models. The cumulative number of confirmed cases is used as the predictor and the number of recovered cases is used as the outcome variable of the analysis. Other variables of interest were the measures undertaken by the government, like social distancing, the use of personal protective equipment like face masks and stay-at-home orders, and the impacts that these measures had on the economy and the general well-being of the people of India.

D. Data Analysis

The present study applied secondary data. Data was specifically generated from the Ministry of Health and Family Welfare records and analyzed using different steps, as per the research objectives.

Step 1:

A trend analysis of the prevalence and incidence of COVID-19 across different regions in India was done by entering the daily statistics provided by the government into SPSS computer software. Descriptive statistics comprised of frequencies, median, and mean regarding COVID-19 infection status were generated at this level of analysis.

Step 2:

A correlation and regression analysis was undertaken to establish the extent to which COVID-19 has impacted the well-being of people in India and determine the effectiveness of various government interventions to abate the spread of the virus among the population. The linear and polynomial regression models were considered appropriate statistical tools for analyzing the pandemic data for India and the real-time COVID-19 infection rates (Bandekar and Ghosh, 2022). The predictions derived from these regression models are as reliable as the available data and indications from any trends that may emerge in the coming days. In addition, short-term forecasts of the new infection rates were estimated using polynomial regression (Takele, 2020). Both short and long-term forecasts are instrumental in identifying any sensitive, unusual trends in infection rates throughout the country that may indicate new waves of COVID-19. Moreover, this short-term perspective on the virus provides information on the effectiveness of the various health measures put in place by the Indian government to limit the spread of COVID-19. Therefore, the linear and polynomial regression model characterizes the common aspects of the pandemic concerning the current status and infection rates.

Regression model analysis of the pandemic in which case:

$$Y_{it} = \alpha + \beta_1 X_{it} + V_{it} \quad (1)$$

Y_{it} = Morbidity linked to COVID-19, measured per the total number of positive COVID-19 cases

α = The regression constant

β_1 = The coefficient of regression

V_{it} = The error term

As previously mentioned, epidemiological data concerning COVID-19 cases were acquired from the Ministry of Health and Family Welfare (Lee and Lee, 2020). The simple regression model estimates the recovery rates and the Case Fatality Rate (CFR) as the output for India and its different states in 2020. The cumulative number of confirmed cases was used as the predictor and the number of recovered cases was used as the outcome variable of the regression analysis. The coefficient of determination R squared was used to determine the best fit. In contrast, the 95% Confidence Interval (CI) was used to calculate the standard error of the slope (Haciimamoğlu, 2021). The polynomial regression was applied in forecasting the number of expected patients during the next period.

There are different regression dimensions for analysis of the spread and effects of COVID-19, based on the data obtained at the end of 2020 when the virus had spread significantly (Sarafidis and Wansbeek, 2021). The unit root, as one of the angles of panel data analysis, can be tested using the Levin-Lin-Chu 12 and Hadri LM stationary tests. In this type of analysis, the null hypothesis will state that the panels contain the unit roots and its alternative hypothesis will state that the panels are stationary. If the p-value is less than six, the null hypothesis is rejected and the alternative hypothesis is accepted. In addition, the Constant Coefficient Model (CCM) from the regression analysis assumes that the coefficients included in the model remain the same across the cross-sectional units for the entire period specified. Thus, the CCM ignores the time and space aspects of the panel data and assumes that the coefficients are the same for all tested parameters (Das *et al.*, 2021). A straightforward application method uses the pooled panel data set and applies the Ordinary Least Squares (OLS) method to estimate the model's unknown parameters.

Another dimension involves using a specific-effect model that assumes the unobserved heterogeneity across the selected Indian district captured by the statistically insignificant parameter of the model (Sharma and Nigam, 2020). Thus, the main questions are often whether there exist individual-specific effects on how the virus impacts people, whether the dependent variable contributes to the spread of the virus and if fixed effects exist within the model. The absence of these effects indicates a random

impact on the model. The final angle is the fixed-effect, also known as the least-square dummy variable regression model. It considers the individuality of the different districts and cross-sectional units by allowing the intercept to vary for each of India's selected and high-risk districts (Gupta *et al.*, 2021). However, as a dimension, it still assumes that the slope coefficients are always constant across the districts that have recorded increased COVID-19 cases and that most of the people have perished from the virus's effects.

The differences in the intercept may result from preventive measures undertaken by the different states in the four districts chosen for analysis in the study (Rath *et al.*, 2020). In the fixed effects model, even though intercepts may differ across districts, each intercept does not exhibit changes over time and is thus considered time-invariant. This qualifies the model for implementation using the dummy variable technique. The time-invariant nature of the model is also attributed to the normalcy exhibited in the year in which the effects of the virus impacted the district at low but significant rates. Extremely low or high rates lead to substantial variations of the model in estimating and predicting results for the entire sector.

The variable D21 represents one observation from the Tenkasi district and is zero; D31 represents one observation from the Thoothukudi district and is zero; D41 represents one observation from the Tirunelveli district and is zero. a_1 is the intercept for the Kanyakumari district and a_2 , a_3 and a_4 are the different intercept coefficients representing the remaining three districts based on the linearity of the model. Given that dummies are used to estimate the fixed effects, the model is known as the Least-Squares Dummy Variable (LSDV) model. One of the major conclusions drawn is that the restricted panel regression model is invalid and thus the LSDV is approved as a valid option (Malavika *et al.*, 2021).

Other tests can supplement and confirm the results postulated by the standard regression model. The Random-Effect Model (REM) makes assumptions about individual-specific effects such as randomness of variables, which are thus included in the error term. This instrument is crucial in testing the effectiveness of the data collected as well as the reliability of the model for each state since all four states chosen for the study exhibit differences in population, the effectiveness of health measures for the prevention of the spread of COVID-19 and the characteristics of their people. In addition, each cross-section possesses some slope parameters and a composite error term (Kotwal *et al.*, 2020.) Furthermore, the Rho exists as the interclass correlation of the error and the fraction of the variance within the error term, due to the effects on individual members within the Indian districts. Another test is the Hausman test, which has a null hypothesis that states that the preferred model possesses random effects instead of fixed effects.

Results and Discussion

Based on the results from the application of the model and its associated statistical tools, it is clear that COVID-19 has significantly affected the lives of people within the selected districts (Lewnard *et al.*, 2022). The number of registered COVID-19 by October 2020 was highest in Kanyakumari (118), followed by Tirunelveli (86), Thoothukudi (77), and Tenkasi (49). At the same time, the overall average number of COVID-19 cases was lowest in Tenkasi (3), followed by Tirunelveli (15), Thoothukudi (17), and Kanyakumari (25), as of October 2020. The number of infections followed a normal distribution in all four districts. Based on the Jargue-Bera tests, the obtained values were non-insignificant at 0.05. The normality in the distribution justifies the government's use of random testing for contact tracing during the initial stages of the pandemic in India (Goswami *et al.*, 2020).

In Table 1, the number of COVID-19 cases in the first column (0-50) is 7 for the Kanyakumari district, 31 for Tenkasi, and 15 for the Thoothukudi and Tirunelveli districts. The following column of 50-100 shows Kanyakumari with 22 and both Tirunelveli and Thoothukudi with 16. The 100-150 column shows the Kanyakumari district with only 2. Therefore, based on Fig. 1, it is clear that the total number of COVID-19 cases reported in October 2020 was highest in the Kanyakumari district, with a total of 2199, followed by Thoothudi with 1583, Tirunelveli with 1523 and Tenkasi with 484. The total number of cases recorded in Tenkasi was low due to awareness campaigns and the administration's effectiveness in ensuring all the health protocols were followed and those who violated them were punished (Gopal *et al.*, 2020).

Additionally, in the Kanyakumari district, the trend in infections was characterized by high rates on the first to the fourth day and a decline in the subsequent days. The declining trend indicates the effectiveness of the measures taken by the government concerning education and enforcement of preventative measures (Sulis *et al.*, 2021). Holt's linear smoothing indicates $a = 0.1413$ and $b = 0.1071$, representing a declining trend within the Kanyakumari district in the third quarter of December 2020. This aligns with the current state of the Indian districts. According to Pinchoff and colleagues (2020), based on data-driven modeling, COVID-19 rates decreased due to the effectiveness of stay-at-home orders, contact tracing, and quarantine interventions. Gopal *et al.* (2020) support this, postulating that the imposition of various measures such as lockdowns, quarantine, and enforcement of all healthcare measures was vital in reducing and maintaining the number of COVID-19 infections (Senapati *et al.*, 2021). The Tenkasi district indicated the highest number of infections in October, which later decreased to 4 cases towards the end of the month. The cases were reduced to the extent that only one digit number was registered.

After analyzing the regression model, it is clear that COVID-19 infections in both Tirunelveli and Kanyakumari exhibited an upward trend (Thayer *et al.*, 2021). The highest number of daily infections was registered in the Kanyakumari District, while the lowest number was registered in the Tenkasi District. In Table 2, the results from the analysis of variance of 44.29316 indicated by F-test and 64.31809 indicated by Welch F-test value with a p-value of 0.0000 suggest that both tests are highly significant, at a 99% significance level. This implies that the number of COVID-19 infections differs from one district to another. The uncorrelated nature of the total number of tests per district is supported by Sulis *et al.* (2021). The ANOVA tests show that the Kanyakumari results obtained concur with those produced by Holt's linear exponential smoothing. The results produced by ARIMA for Thoothukudi, Tenkasi, and Tirunelveli -- (0, 1, 1), (2, 0, 0,) and (2, 2, 2) respectively --- indicate that there is no correlation among infection rates for the four districts.

Using the constant coefficient model (Panel OLS), the number of new cases was captured and a justified prediction was made (Senapati *et al.*, 2021). The number of new cases is the dependent variable, while time is the explanatory variable. The results indicate that the constant and slope are highly significant, at a 1% level of significance. The line of a good fit, represented by R squared, is 29%, suggesting that the model can explain 29% of variations in the dependent variable. In addition, the Durbin-Watson value of 0.242 shown in Table 3 indicates the presence of autocorrelation aspects within the data. Identifying the existence of autocorrelation is instrumental in ensuring that the results obtained from each district under study portray the actual individual-specific state of the country. Since the model assumes that the coefficient slope and the time-explanatory variable are the same for all four districts, the model fails to provide significant reliability for making appropriate conclusions on the state of the existing healthcare issues within India (Sulis *et al.*, 2021).

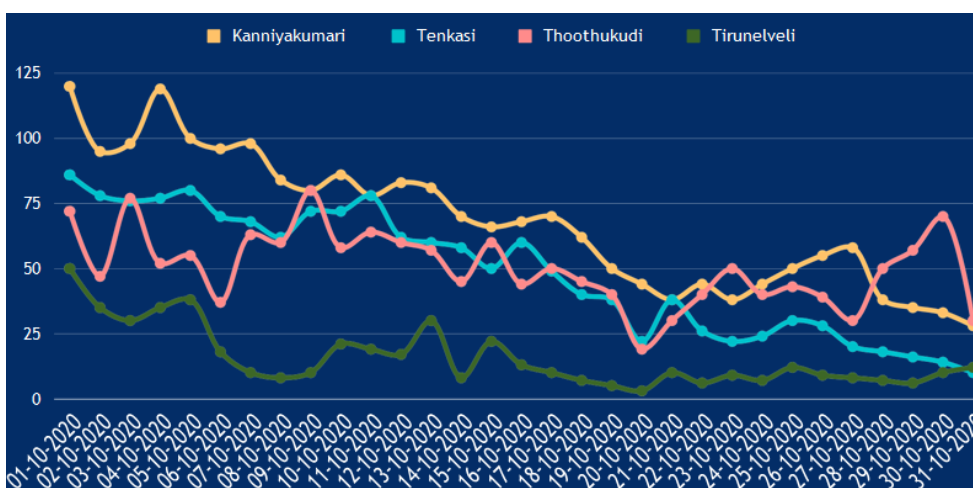


Fig. 1: Illustrating the number of infected cases in October 2020 in four Indian districts

Table 1: Descriptive statistics of covid-19 infection status per district

Statistics	Kanya Kumari	Tenkasi	Thoothukudi	Tirunelveli	
Mean	70.935480	15.612900	51.064520	49.129030	
Median	70.000000	10.000000	51.000000	50.000000	
Maximum	118.000000	49.000000	77.000000	86.000000	
Minimum	25.000000	3.000000	17.000000	15.000000	
Std. dev	24.611560	11.976860	14.193740	22.860070	
Skewness	-0.019244	1.131092	-0.001396	0.023114	
Kurtosis	2.154139	3.372580	2.676387	1.508788	
Jarque-bera	0.926077	6.789374	0.135281	2.875059	
Probability	0.629368	0.033551	0.934597	0.237514	
Sum	2199.000000	484.000000	1583.000000	1523.000000	
Sum sq.dev	18171.870000	4303.355000	6043.871000	15677.480000	
Category	No. of cases				
0-50	7.000000	31.000000	15.000000	15.000000	68
50-100	22.000000	0.000000	16.000000	16.000000	58
100-150	2.000000	0.000000	0.000000	0.000000	2

Table 2: Result of variance analysis

Method	df	Value	Probability
ANOVA F-test	-3,120.0000	44.29316	0
Welch F-test*	(3,64.3377)	64.31809	0

Table 3: Results of autocorrelation analysis

Variable	Coefficient	Std. Error	t-Statistics	Probability
C	73.490000	4.258	17.260	0
X	-1.680000	0.232	-7.210	0
Root MSE	22.950000	R-Squared	0.299	
Mean dependent error	46.690000	Adjusted R-Squared	0.293	
SD dependent var	27.520000	SE of Regression	23.140	
Akaike info criterion	9.137000	Sum squared residual	65302.000	
Schwarz criterion	9.182000	Log-Likelihood	-564.000	
Hannan-quinn criterion	9.155000	F-Static	52.000	
Durbin-watson stat	0.242000	Prob. (F static)	0.000	
Crossid	Effect			
1	24.045620			
2	-30.810700			
3	4.342125			
4	2.422954			

The least-square dummy variable regression model, also known as the fixed effects regression model, shows 82% variation in the dependent variable. Thus, the dependent variable represented by the new cases can be explained by the variations in the explanatory variables defined by time, a measure put in place by the national government, and other factors captured in the error term (Gopal *et al.*, 2020). The model is highly significant, at a 1% level of significance. At both the 1% significance level and 82% R squared, it can be concluded that all the fixed effects models are the mode efficient model compared to the panel least-squares regression model. Looking at the fixed effects among the four districts, the 24.25 positive value for Kanyakumari is high compared to the other three districts. One of the main reasons for this is the extremely high infection rates recorded in this district compared to the other three. For instance, Tenkasi recorded a -31.0725, which indicates extremely low rates of new infections. In the Thoothukudi district, the cross-sectional fixed effects were 4.3790, which is also exceptionally low compared to the Kanyakumari district. The Tirunelveli district's 2.4435 value was extremely low compared to both the Kanyakumari and Thoothukudi districts.

A. Random Effects Model

The random effects model was used to analyze COVID-19 infection rates in the four major districts in India at the end of 2020 when the spread and effects of the virus were enormous. The results were highly significant, at a 1% level of significance, indicating a high R-squared value of 62% with an SE of regression of 1.7257 and a root MSE of 11.6307. Thus, both the slope and the intercept indicate an increased significance of 1%, like that exhibited by the fixed effect model. The slope was

recorded at -1.675101, which is highly significant, indicating that new COVID-19 infections are decreasing at the rate of approximately 1.68%. The reduction in the infection rate suggests the effectiveness of measures intended to curb the spread of the virus, such as sessional movement of people, quarantine measures, sanitation, mask-wearing, and the opening of more testing centers. Additionally, the rho value of 0.7915 implies a value of 0.8 of personal effects for the cross-sections. Thus, there is no correlation among the four districts' selections as far as infection and testing are concerned (Gopal *et al.*, 2020).

The results show the normally distributed residuals with a 3.1389 Jarque-Bera test value, which is insignificant at a 5% significance level. The model thus allows for the future by considering numerous other parameters related to the inclusion of populations that have issues and ailments associated with respiration, etc. Age and associated chronic illness increase patients' risk of COVID-19 infection, due to the reduced capacity of the immune system. This explains why most of both the infection rates and the devastating effects of the virus are witnessed in the older generations. Since a significant percentage of India's population is elderly, the impact of the virus is felt strongly, as these older people are pillars of the existing economy of the country.

Thus, based on the different regression models used to analyze the COVID-19 situation in India, it can be observed that the emergence and spread of the virus extensively relied on the number of people within a given district, the measures undertaken by local and national governments and the infection rate (Sulis *et al.*, 2021). Higher populations have an increased risk of exposure to the virus, as shown by the data for Kanyakumari compared to other districts. Panel data effectively assessed the trends in COVID-19 cases in the four

districts. They thus predicted future trends given the specific characteristics of the population, healthcare measures, and infection rates. The random effects model exhibits clear and reliable COVID-19 case data for the different parts of the four districts (Kotwal *et al.*, 2020).

The registered number of new COVID-19 cases was highest in Kanyakumari, followed by Changotra *et al.* (2021). As indicated by the R-squared, the explanatory variables could explain approximately 90% of the dependent variable of new COVID-19 cases. In addition, the 0.8% of personal effects for the cross-sections indicate a late correlation among the four districts, implying that each of the districts was exposed to similar risks and conditions that facilitated the spread of the virus, as opposed to being influenced by individual actions. One limitation of the models is their failure to take into account existing waves of COVID-19, which have had a significant impact on the infection rates. The ever-evolving and mutational nature of the virus make it unpredictable, especially recently, as the second and third waves have had devastating effects on the entire population. Furthermore, the lessening of government restrictions in India offered contradictory results, since this action increases people's risk of exposure, which has to be accounted for in the model's specifications (Kotwal *et al.*, 2020). The COVID-19 infection rate had decreased by 1.68% as of October 2020.

Conclusion

The analysis shows that the outbreak of the novel coronavirus in 2019 in Wuhan, China has led to devastating effects on the economies of more than 200 countries across the globe. More importantly, the number of patients and casualties increased significantly in 2020, as most countries recorded persistent increases in their numbers of confirmed cases. Due to the devastating effects of the virus, the World Health Organization declared COVID-19 a worldwide pandemic in March 2020. In India, the first case of the virus was recorded on 30th January 2020 in Kerala, in a student who had returned from Wuhan University in China. Soon, the pandemic spread to other parts of the country, due to massive importation of the virus from students who had returned from other places. The outbreak of COVID-19 (SARS-CoV-2) within the different states of India has increased the overall death rate by 69% as of December 2021, with the highest rates recorded at 10,000 per day. Vaccines have recently been emphasized by practitioners and health experts as complete prevention mechanisms against spreading the virus. With approximately 1.34 billion people, making it the second most populated country in the world, India has had difficulty treating its severe COVID-19 cases.

As a result, a capacity of roughly 50,000 ventilators has been insufficient in serving the existing population. If the number of new infections exceeded India's ability to accommodate the conditions, the nation would experience catastrophic effects; therefore, there would be a need for sound decision-making using the available tools for the necessary preventive and recovery solutions. Another challenge associated with the increased infection rate is the difficulty in identifying the channels of infection and the people who come into contact with them. This often demands the identification of multiple strategies for handling the outbreak, including statistical and quantitative analyses and computational modeling for enhancing the development of vaccines and drug treatments.

Therefore, the impact of COVID-19 has crippled systems within the political, social, and economic landscapes of many countries. More importantly, the health-associated complications that patients suffer from infection have led to significantly high death rates. Using the appropriate statistical tools, various predictive information has been derived from the existing features of the virus and the characteristics surrounding the infections. Thus, the regression model is a crucial tool to provide predictive results that are instrumental for both national and local governments' planning, to ensure that future infections and their effects on the people are reduced and prevented. The analysis shows that infection rates have been declining and that results from the four districts sampled do not relate to each other. Thus, the existing measures from the Ministry of Health and the government, in general, are instrumental in the management of future cases.

However, the data used was based on the year 2020, when infection rates were relatively low and generally distributed to provide reliable information from the data obtained. Additionally, the outbreak of COVID-19 in China was catastrophic, leading to multiple changes within the healthcare sector and medical system to counter the pandemic. Compared to the systems and processes of China, India was less capable of managing the pandemic, hence the high death tolls recorded in early 2021, especially in the elderly population. It is the role of the Ministry of Health and Family Welfare to ensure that effective strategies and measures are put in place for the prevention of future waves and effects of the virus. Given that the elderly and a subset of the adult population have higher risks of suffering severe consequences from the virus, priority has to be placed on interventions that directly target them. The existence of the high-risk population justifies the existence of strict measures that aim to ensure that the virus is effectively contained. One challenge associated with the regression model is that while the COVID-19 infections continue to manifest in different waves, the model does not allow adjustment to account for any deviations, making the existing data unreliable. Accounting for these variations is thus instrumental in making rational and realistic predictions

regarding the future state of the pandemic. Further studies should focus on incorporating the various infection rates caused by the constantly evolving waves of the pandemic, which place the majority of Indians at risk of infection.

Acknowledgment

The author would like to acknowledge the support provided by Al Maarefa University while conducting this research work.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

References

- Anwar, S., Nasrullah, M., & Hosen, M. J. (2020). COVID-19 and Bangladesh: Challenges and how to address them. *Frontiers in Public Health*, 8. <https://doi.org/10.3389/fpubh.2020.00154>
- Bandekar, S. R., & Ghosh, M. (2022). Mathematical modeling of COVID-19 in India and its states with optimal control. *Modeling Earth Systems and Environment*, 8(2), 2019-2034. <https://link.springer.com/article/10.1007/s40808-021-01202-8>
- Changotra, R., Rajput, H., Rajput, P., Gautam, S., & Arora, A. S. (2021). The largest democracy in the world crippled by COVID-19: Current perspective and experience from India. *Environment, Development, and Sustainability*, 23(5), 6623-6641. <https://link.springer.com/article/10.1007/s10668-020-00963-z>
- Das, A., Ghosh, S., Das, K., Basu, T., Dutta, I., & Das, M. (2021). Living environment matters: Unravelling the spatial clustering of COVID-19 hotspots in Kolkata megacity, India. *Sustainable Cities and Society*, 65, 102577. <https://doi.org/10.1016/j.scs.2020.102577>
- Folliot, P. (2020) The COVID-19 pandemic and the middle east and north africa region. *Frontiersin*. (2021). COVID-19: Mitigation strategies and their implications for the global environment.
- Gopal, A., Sharma, A. J., & Subramanyam, M. A. (2020). Dynamics of psychological responses to COVID-19 in India: A longitudinal study. *PloS one*, 15(10), e0240650. <https://doi.org/10.1371/journal.pone.0240650>
- Goswami, K., Bharali, S., & Hazarika, J. (2020). Projections for COVID-19 Pandemic in India and effect of temperature and humidity. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 14(5), 801-805. <https://doi.org/10.1016/j.dsx.2020.05.045>
- Guleria, P., Ahmed, S., Alhumam, A., & Srinivasu, P. N. (2022, January). Empirical Study on Classifiers for Earlier Prediction of COVID-19 Infection Cure and Death Rate in the Indian States. In *Healthcare* (Vol. 10, No. 1, p. 85). MDPI. <https://doi.org/10.3390/healthcare10010085>
- Gupta, A. K., Singh, V., Mathur, P., & Travieso-Gonzalez, C. M. (2021). 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>
- Hacıımamoğlu, T. (2021). Is Per Capita Real Gdp Stationary In G-7 Countries? New Evidence from A Fourier Panel Unit Root Test. *Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 22(2), 123-144. <https://doi.org/10.37880/cumuiibf.948351>
- Jaipuria, S., Parida, R., & Ray, P. (2021). The impact of COVID-19 on tourism sector in India. *Tourism Recreation Research*, 46(2), 245-260. <https://doi.org/10.1080/02508281.2020.1846971>
- Ketu, S., & Mishra, P. K. (2022). India perspective: CNN-LSTM hybrid deep learning model-based COVID-19 prediction and current status of medical resource availability. *Soft Computing*, 26(2), 645-664. <https://doi.org/10.1007/s00500-021-06490-x>
- Kotwal, A., Yadav, A. K., Yadav, J., Kotwal, J., & Khune, S. (2020). Predictive models of COVID-19 in India: A rapid review. *Medical Journal Armed Forces India*, 76(4), 377-386. <https://doi.org/10.1016/j.mjafi.2020.06.001>
- Kumar, N., & Kumar, H. (2022). A novel hybrid fuzzy time series model for prediction of COVID-19 infected cases and deaths in India. *ISA transactions*, 124, 69-81. <https://doi.org/10.1016/j.isatra.2021.07.003>
- Kumar, P., Singh, R. K., Nanda, C., Kalita, H., Patairiya, S., Sharma, Y. D., ... & Bhagavathula, A. S. (2020). Forecasting COVID-19 impact in India using pandemic waves Nonlinear Growth Models. *MedRxiv*. <https://doi.org/10.1101/2020.03.30.20047803>
- Lee, C. F., & Lee, J. C. (Eds.). (2020). *Handbook Of Financial Econometrics, Mathematics, Statistics and Machine Learning* (In 4 Volumes). World Scientific. <https://doi.org/10.1142/11335>
- Lewnard, J. A., Mahmud, A., Narayan, T., Wahl, B., Selvavinayagam, T. S., & Laxminarayan, R. (2022). All-cause mortality during the COVID-19 pandemic in Chennai, India: an observational study. *The Lancet Infectious Diseases*, 22(4), 463-472. [https://doi.org/10.1016/S1473-3099\(21\)00746-5](https://doi.org/10.1016/S1473-3099(21)00746-5)
- Maddock, J. (2020). Has pandemic fatigue set in? Here's why you might have it. <https://edition.cnn.com/2020/11/12/health/pandemic-fatigue-wellness-partner/index.html>

- Malavika, B., Marimuthu, S., Joy, M., Nadaraj, A., Asirvatham, E. S., & Jeyaseelan, L. (2021). Forecasting COVID-19 epidemic in India and high incidence states using SIR and logistic growth models. *Clinical Epidemiology and Global Health*, 9, 26-33. <https://doi.org/10.1016/j.cegh.2020.06.006>
- Mohan, S., Solanki, A. K., Taluja, H. K., & Singh, A. (2022). Predicting the impact of the third wave of COVID-19 in India using hybrid statistical machine learning models: A time series forecasting and sentiment analysis approach. *Computers in Biology and Medicine*, 144, 105354. <https://doi.org/10.1016/j.compbiomed.2022.105354>
- Pinchoff, J., Santhya, K. G., White, C., Rampal, S., Acharya, R., & Ngo, T. D. (2020). Gender-specific differences in COVID-19 knowledge, behavior and health effects among adolescents and young adults in Uttar Pradesh and Bihar, India. *PloS one*, 15(12), e0244053. <https://doi.org/10.1371/journal.pone.0244053>
- 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(5), 1467-1474. <https://doi.org/10.1016/j.dsx.2020.07.045>
- Rowan, N. J., & Laffey, J. G. (2020). Challenges and solutions for addressing the critical shortage of supply chain for personal and protective equipment (PPE) arising from Coronavirus disease (COVID19) pandemic-A case study from the Republic of Ireland. *Science of The Total Environment*, 138532. <https://doi.org/10.1016/j.scitotenv.2020.138532>
- Sarafidis, V., & Wansbeek, T. (2021). Celebrating 40 years of panel data analysis: Past, present and future. *Journal of Econometrics*, 220(2), 215-226. <https://doi.org/10.1016/j.jeconom.2020.06.001>
- Senapati, A., Nag, A., Mondal, A., & Maji, S. (2021). A novel framework for COVID-19 case prediction through piecewise regression in India. *International Journal of Information Technology*, 13(1), 41-48. <https://doi.org/10.1007/s41870-020-00552-3>
- Sharma, V. K., & Nigam, U. (2020). Modeling and Forecasting of Covid-19 growth curve in India. *Transactions of the Indian National Academy of Engineering*, 5(4), 697-710. <https://doi.org/10.1007/s41403-020-00165-z>
- Sulis, G., Batomen, B., Kotwani, A., Pai, M., & Gandra, S. (2021). Sales of antibiotics and hydroxychloroquine in India during the COVID-19 epidemic: An interrupted time series analysis. *PLoS Medicine*, 18(7), e1003682. <https://doi.org/10.1371/journal.pmed.1003682>
- Takele, R. (2020). Stochastic modelling for predicting COVID-19 prevalence in East Africa Countries. *Infectious Disease Modelling*, 5, 598-607. <https://doi.org/10.1016/j.idm.2020.08.005>
- Thayer, W. M., Hasan, M. Z., Sankhla, P., & Gupta, S. (2021). An interrupted time-series analysis of the lockdown policies in India: A national-level analysis of COVID-19 incidence. *Health Policy And Planning*, 36(5), 620-629. <https://doi.org/10.1093/heapol/czab027>
- Wang, Y., Wang, Y., Chen, Y., & Qin, Q. (2020). Unique epidemiological and clinical features of the emerging 2019 novel coronavirus pneumonia (COVID- 19) implicate special control measures. *Journal of Medical Virology*, 92(6), 568-576. <https://doi.org/10.1002/jmv.25748>
- Worldometer. (2021). COVID-19 coronavirus pandemic. <https://www.worldometers.info/coronavirus>
- Young, E. (2020). N.J. expects second wave to peak in 2021's first quarter.
- Zhang, D., Ling, H., Huang, X., Li, J., Li, W., Yi, C., ... & Zhang, X. (2020). Potential spreading risks and disinfection challenges of medical wastewater by the presence of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)g viral RNA in septic tanks of fangcang hospital. *Science of the Total Environment*, 140445. <https://doi.org/10.1016/j.scitotenv.2020.140445>