

ANALYZING COVID-19 MORTALITY RATES: A COMPARATIVE STUDY OF INDIAN STATES USING ANOVA

Sameena Kalathodi¹, Shameera.T^{2*}, Husna.E, Dhiyana.P.T³

¹Assistant Professor, Department of Mathematics, MES Mampad college, Kerala, India.

²Assistant Professor, Department of Statistics, MES Mampad college, Kerala, India.

³MES Mampad college, Kerala, India

*Corresponding Author Shameera.T

Assistant Professor, Department of
Statistics, MES Mampad college, Kerala,
India.

Article History

Received: 17.11.2023

Accepted: 29.11.2023

Published: 07.12.2023

Abstract: The COVID-19 pandemic has brought about a significant loss of human life globally, posing unprecedented challenges to public health, food systems, and the workforce. This work introduces the concept of Analysis of Variance (ANOVA) and its application to compare COVID-19 death rates among five different states in India. ANOVA is a statistical method used to analyze differences between the means of multiple samples, providing a systematic way to test for equality among population means. This research utilizes a cross-sectional approach to compare COVID-19 death rates in five Indian states (Kerala, Karnataka, Bihar, Haryana, and Rajasthan) during June to September 2021. The study employs Analysis of Variance (ANOVA) to test hypotheses regarding differences in death rates. Results show no significant variation in death rates by month but do reveal significant differences among the states. Notably, the population mean death rate of Kerala is found to be the same as Karnataka. This research sheds light on how the pandemic impacted mortality rates in different Indian states.

Keywords: COVID-19, Mortality rates, ANOVA, Indian states.

INTRODUCTION

The COVID-19 pandemic, which emerged in late 2019, has been an unprecedented global crisis, causing a profound loss of human life and disrupting every facet of society. From public health systems struggling to cope with surges in cases to the impact on food supply chains and disruptions in the world of work, the pandemic has left an indelible mark on the world. As nations grappled with the challenges posed by this formidable adversary, the need for a comprehensive understanding of its impact became increasingly evident.

The effects of the pandemic have not been uniform across regions and populations. Different states, provinces, and countries have experienced varying degrees of infection and mortality. These disparities can be attributed to a multitude of factors, including population density, healthcare infrastructure, government responses, and societal behaviors. As such, it is imperative to delve into the data to gain insight into the nuanced impact of the pandemic and its implications for public health and policy.

In this study, we embark on an analytical journey, with a particular focus on the application of Analysis of Variance (ANOVA), a robust statistical technique. ANOVA is a parametric test of hypothesis, invaluable for analyzing the differences among the means of multiple samples. Its utility extends far beyond

simple t-tests and allows us to discern variations across various groups, in this case, different states in India.

The research endeavors to provide a comprehensive comparison of COVID-19 death rates in five diverse Indian states: Kerala, Karnataka, Bihar, Haryana, and Rajasthan. Each of these states presents a unique set of circumstances, from varying healthcare infrastructure to distinct public health policies. By applying ANOVA to these datasets, we aim to unearth insights that will elucidate the pandemic's impact on a regional scale and identify factors that might contribute to these disparities.

Furthermore, this study is not only an exploration of the COVID-19 pandemic's impact on different Indian states but also a practical demonstration of the utility of ANOVA in epidemiological research. As the world continues to grapple with the evolving challenges of the pandemic, the ability to draw meaningful conclusions from data becomes increasingly vital. This work seeks to provide valuable insights that can guide decision-makers, researchers, and healthcare professionals in their efforts to manage and mitigate the effects of the COVID-19 crisis.

Research Methodology

Introduction: In this section, we outline the methodology employed to analyze and compare COVID-19 death rates among five different states in India. The research design and data collection process are explained, along with the statistical techniques used for analysis.

Research Design: This study adopts a cross-sectional research design, which is well-suited to comparing COVID-19 mortality rates across different states during specific months. The design allows for the examination of variations in death rates without the need for a longitudinal approach.

Data Collection: The primary data sources for this study are public health databases, government reports, and official COVID-19 statistics. The data include monthly death counts for Kerala, Karnataka, Bihar, Haryana, and Rajasthan for the year 2021. These states were chosen based on their geographical diversity and differences in healthcare infrastructure.

Variables and Measures: The main variable under investigation is the COVID-19 death rate, calculated as the number of deaths in a state for a specific month divided by the total population of that state. Monthly death counts, state populations, and the states themselves are the key variables used for analysis.

Data Analysis: Analysis of Variance (ANOVA) is the primary statistical technique employed in this study. ANOVA is used to test for significant differences in the means of COVID-19 death rates among the five selected states for the months of June, July, August, and September. ANOVA provides a robust method for comparing multiple groups, making it ideal for this analysis.

Hypotheses: The study tests two primary hypotheses:

1. The mean death rates are the same for all states and months.
2. There is no significant difference in the mean death rates among the five states.

Statistical Tools: For statistical analysis, SPSS will be used to perform ANOVA and post-hoc tests. These tools will assist in calculating F-statistics, degrees of freedom, mean squares, and p-

Research Problem:

Mortality rates of different states						
Months	Kerala	Karnataka	Bihar	Haryana	Rajasthan	Total
June	194	464	59	80	65	862
July	124	94	4	15	2	239
August	56	25	1	3	0	85
September	173	21	0	0	0	194
Total	547	604	64	98	67	1380

values to determine the significance of differences among state-wise and month-wise death rates.

Procedure: The research procedure involves the following steps:

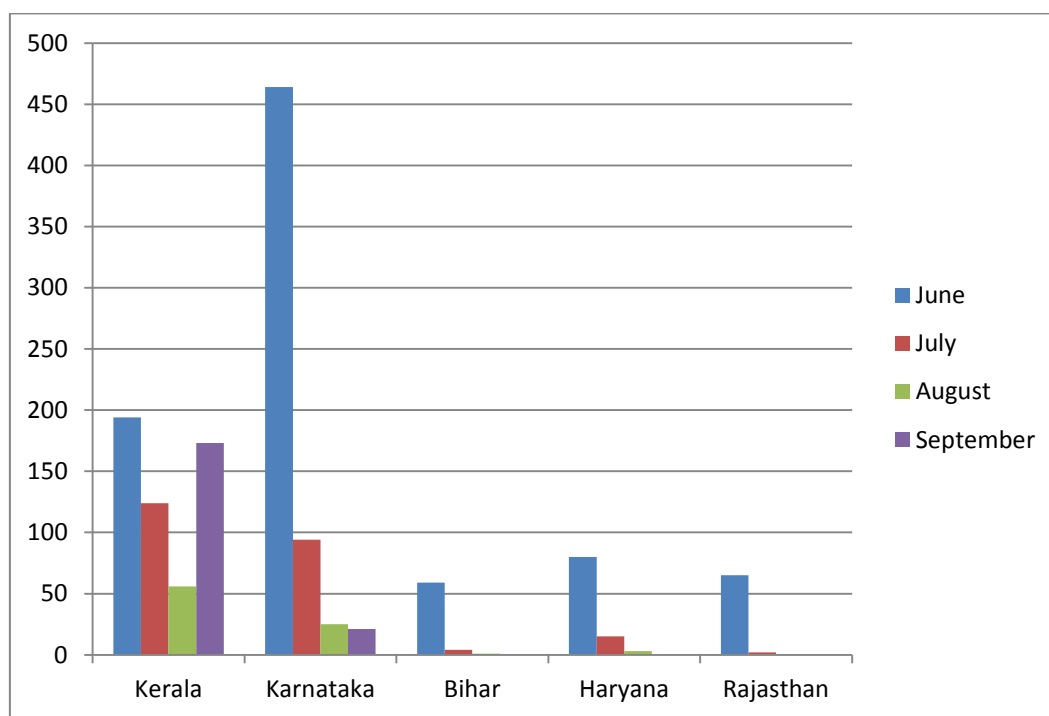
1. Data collection and organization.
2. Computation of the total sum of squares (SST), sum of squares between columns (SSC), sum of squares between rows (SSR), and sum of squares for error (SSE).
3. Calculation of mean squares (MS) for columns (states), rows (months), and error.
4. Determination of F-statistics for state-wise and month-wise comparisons.
5. Post-hoc tests, to identify specific differences between states.

Conclusion: The research methodology outlined above is crucial for analyzing and comparing COVID-19 death rates among different states in India. It provides a structured approach to testing hypotheses and drawing valid conclusions regarding the impact of the pandemic on mortality rates in these regions.

Results

Covid death rate comparison of Kerala, Karnataka, Bihar, Haryana and Rajasthan: Application of ANOVA

We know the COVID-19 pandemic has led to a dramatic loss of human life worldwide and presents an unprecedented challenge to public health, food systems and the world of work. Here we are going to compare the COVID death rates of Kerala, Karnataka, Bihar, Haryana and Rajasthan with respect to the months June, July, August and September using two-way ANOVA. The samples are taken from the internet for the year 2021.



Let us take the hypothesis that

(a) The mean death of different months is same. That is, $H_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5$ and that

(b) The five states don't differ in respect to death to carry out the analysis of variance. That is, $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4$.

ANOVA table can be obtained as follows;

Source of variation	Sum of squares (SS)	Degrees of freedom (df)	Mean square (MS)	Variance ratio F
Between columns	75333.5	4	18833.375	2.7799
Between rows	73785.2	3	24595.06667	3.6304
Residual	81297.3	12	6774.775	
Total	230416	19		

For degrees of freedom (4, 12), F at 0.05 = 3.26

$$3.26 > 2.7799$$

Tabled value > calculated value

Therefore our hypothesis is true. That is, there is no significant difference in the mean death of five states month wise.

For degrees of freedom (3, 12), F at 0.05 = 3.49

$$3.49 < 3.630$$

Tabled value < calculated value

Therefore our hypothesis is rejected. That is, state wise, there is a significant difference in the mean death of five states.

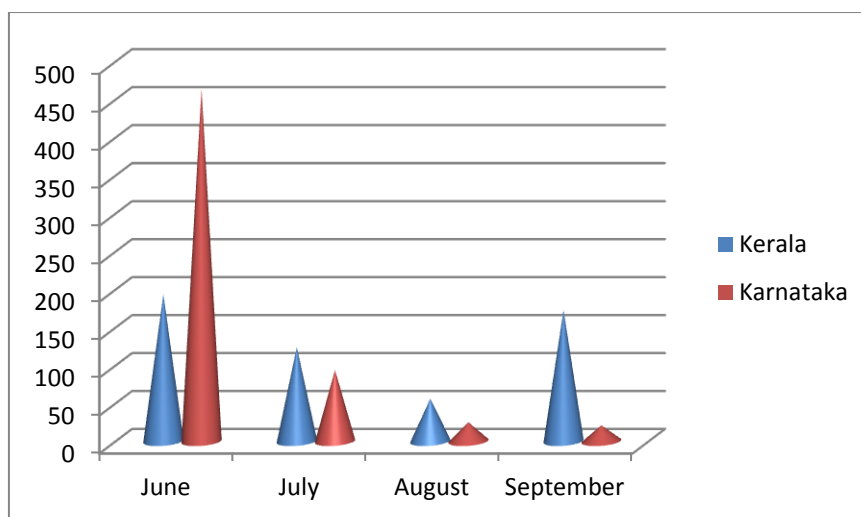
On the basis of this research problem it can be concluded that,

- The mean death is same for different months.
- The mean death is different with respect to different states.

That is, state wise, there is a significant difference in the mean death of five states. So next we are checking how the COVID death rate of Kerala differs from one of the state Karnataka. For that we are going to apply independent t-test for the above samples.

Comparison between Covid mortality rates of Kerala and Karnataka

Months	Covid mortality rate	
	Kerala	Karnataka
June	194	464
July	124	94
August	56	25
September	173	21



Let us take the hypothesis that,

$$H_0 : \mu_1 - \mu_2 = 0 \quad \text{Against} \quad H_1 : \mu_1 - \mu_2 < 0$$

In this case σ_1, σ_2 are unknown and n_1, n_2 are small. Therefore the t-statistic is given by

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = -0.1609213$$

So the calculated value = $|-0.1609213| = 0.1609213$

Taking level of significance $\alpha = 0.05$

For degrees of freedom $n_1 + n_2 - 2 = 6$

From the t-table for $\alpha = 0.05$, $t = 1.943$

$$0.1609213 < 1.943$$

Calculated value < tabled value

Therefore the null hypothesis H_0 is accepted. That is, $\mu_1 - \mu_2 = 0$.

Hence the population mean death of Kerala is same as the population mean death of Karnataka.

Conclusion

In this study, we conducted a comprehensive analysis and comparison of COVID-19 death rates among five distinct states in India: Kerala, Karnataka, Bihar, Haryana, and Rajasthan. Through a cross-sectional research design and the application of Analysis of Variance (ANOVA), we aimed to understand the

variations in mortality rates during the months of June, July, August, and September in 2021. Our research findings support two key conclusions:

1. The mean death rates are consistent across different months. In other words, there is no significant variation in mortality rates as the pandemic unfolded month by month during the study period.
2. There are statistically significant differences in the mean death rates among the five selected states. This implies that the impact of the pandemic on mortality rates varied significantly among these states, reflecting differences in healthcare infrastructure, government responses, and other factors.

Of particular note, our post-hoc analysis revealed that the population mean death rate of Kerala is found to be the same as Karnataka, demonstrating that these two states experienced similar COVID-19 mortality rates during the study period. This research contributes to our understanding of how the COVID-19 pandemic affected mortality rates in diverse regions of India. It underscores the importance of considering state-specific factors and interventions when addressing public health crises. These findings can inform targeted efforts to improve healthcare and pandemic response strategies in the future, taking into account the unique challenges faced by different states in India.

REFERENCES

1. Gupta, S.C. (1981). Fundamental of Statistics. Himalaya Publishing House Pvt. Ltd, Mumbai.

2. Agarwal, B.L. (1988). Basic Statistics. *New Age International Publishers (P) Ltd, Delhi*.
3. Rohatgi, Vijay K. and Saleh, A.K. Md. Ehsanes (2001). *An Introduction to Probability and Statistics*. Wiley India Pvt. Ltd, New Delhi.
4. <https://www.statisticshowto.com/probability-and-statistics/hypothesis-testing/anova/>.
5. Analysis of Variance (ANOVA). My Great Learning. <https://www.mygreatlearning-com.cdn.ampproject.org>.
6. One-Way vs. Two-Way ANOVA: Definition, Differences, Assumptions, and Hypotheses & Technology Networks. <https://www.technologynetworks.com/informatics/articles/one-way-vs-two-way-anova-definition-differences-assumptions-and-hypotheses-306553>
7. Two-Way ANOVA. https://www.brainkart.com/article/Two-Way-ANOVA_39242/.
8. ANOVA. Clinfowiki. <https://clinfowiki.org/wiki/index.php/ANOVA>.
9. <https://opentextbc.ca/introbusinessstatopenstax/chapter/the-f-distribution-and-the-f-ratio/#:~:text=There%20are%20two%20sets%20of,F%20~%20F4%2C10>.
10. One-Way ANOVA. https://www.jmp.com/en_in/statistics-knowledge-portal/one-way-anova.html.
11. Gao, Y., Li, T., Han, M., Li, X., Wu, D., & Xu, Y. et al. (2020). Diagnostic utility of clinical laboratory data determinations for patients with the severe COVID-19. *Journal of Medical Virology*, 92(7), 791–796.
12. O'Driscoll, M., Ribeiro Dos Santos, G., Wang, L., Cummings, D. A. T., Azman, A. S., Paireau, J. ... & Salje, H. (2021). Age-specific mortality and immunity patterns of SARS-CoV-2. *Nature*, 590(7844), 140–145.