International Journal of Mathematics and Computer Research

ISSN: 2320-7167

Volume 14 Special Issue 03 March 2026, Page no. – 180-186

Impact Factor: 8.751

DOI: 10.47191/ijmcr/v14iSPC3.35

Comparative Statistical Analysis of Major Crop Productivity in Maharashtra

Rupali Laxman Kamthe, Akshata Lembhe, Arati Dange, Deepali Akolkar, Seema Dokrimare

Department of Statistics, Dr. D. Y. Patil, Arts, Commerce & Science College, Pimpri, Pune, Maharashtra, India

ARTICLE INFO

Published Online:
17 March 2026

Corresponding Author:
Rupali Laxman Kamthe

ABSTRACT

This study presents a comprehensive statistical analysis of agricultural productivity in Maharashtra, focusing on five major crops Rice, Wheat, Soyabean, Sugarcane, and Cotton using district-level data from the ICRISAT dataset (1966–2015) . The objective is to evaluate the interrelationship between cultivated area, total production, and yield using descriptive statistics, correlation analysis, linear regression, ANOVA, and t-tests . The findings reveal that area and production are strongly correlated ($r \approx 0.9–1.0$), indicating that as cultivation area increases, production also rises proportionally, consistent with earlier findings . However, the correlation between area and yield is weak ($r \approx \pm 0.3$), suggesting that yield is influenced by non-land factors such as irrigation, technology, and rainfall . Regression results show that Wheat ($R^2 = 0.38$) has the strongest dependency between area and yield, while Rice and Cotton show weak relationships, aligning with prior econometric applications in agriculture . The t-test between Rice and Wheat yields ($p = 0.63$) indicates no significant difference in mean productivity. Time series analysis further demonstrates steady yield improvement in Wheat and Sugarcane over five decades . The study concludes that Maharashtra's agricultural growth depends more on improved resource efficiency and technology adoption than on area expansion.

KEYWORDS: Agricultural productivity, Maharashtra agriculture, Crop yield analysis, Correlation analysis, Linear regression, ANOVA, t-test, Time series analysis

I. INTRODUCTION

Agriculture is the backbone of Maharashtra's economy, employing nearly half of its population [4], [10]. Despite substantial technological progress, productivity gaps persist between districts and crops [11], [13]. Maharashtra exhibits varied climatic zones from humid coastal Konkan to semi-arid Vidarbha, making agricultural output highly sensitive to rainfall and irrigation [4], [12]. The state produces a variety of crops, with Rice and Wheat as major cereals, Soyabean and Cotton as key commercial crops, and Sugarcane as the dominant irrigated crop [3], [10]. However, while area under cultivation has expanded, yield growth has not always matched expectations [1], [11]. Therefore, it is essential to identify whether area expansion directly enhances yield or whether productivity depends more on other factors [9], [12], [13]. The present study applies statistical methods including regression, correlation, ANOVA, and time series analysis to measure relationships among area, production, and yield for key crops, providing insights into Maharashtra's long-term agricultural performance [6]–[8], [14], [15].

II. RELATED WORK

Several studies have examined agricultural productivity using statistical and econometric methods to understand the influence of cultivation practices and environmental factors on crop yield. Earlier research applied correlation and regression models to measure relationships among area, production, and yield [11], while others used ANOVA and t-tests to compare crop performances across regions [14]. Recent studies have emphasized time series analysis for identifying long-term yield trends and forecasting production patterns [15]. Reports from organizations such as FAO [5] and NITI Aayog [4] highlight that sustainable yield growth depends more on irrigation efficiency and modern technology than on land expansion. This study extends existing research by combining correlation, regression,

ANOVA, t-test, and time series analysis to statistically evaluate the productivity of major crops in Maharashtra.

III. METHODOLOGY

Dataset Description

The study uses secondary data from the ICRISAT District-Level Database (1966–2015) for Maharashtra [1]. It includes annual data for five major crops Rice, Wheat, Soyabean, Sugarcane, and Cotton representing both food and commercial crops.

Variables

Dependent Variable: Crop yield (Kg per hectare),Independent Variables: Area (1000 ha), Production (1000 tons), and Year (1966–2015)

Statistical Methods

1. Descriptive Statistics– to summarize mean and variability of crop yields.
2. Correlation Analysis – to assess relationships between area, production, and yield[11].
3. Regression Analysis – to estimate the effect of area on yield.[6]-[8].
4. ANOVA and t-Test – to test yield differences among and between crops[14].
5. Time Series Analysis – to examine long-term yield trends using linear trend estimation[15].

All analyses were performed using Python (Pandas, Seaborn, and Statsmodels) at a 5% significance level.

IV. RESULTS AND DISCUSSION

Descriptive Statistics

Sugarcane has the highest mean yield (≈4200 Kg/ha) followed by Wheat and Rice (≈1500 Kg/ha), while Soyabean and Cotton show lower yields (<1000 Kg/ha). This suggests Sugarcane's dependence on irrigation and technological inputs[3],[4],[5].

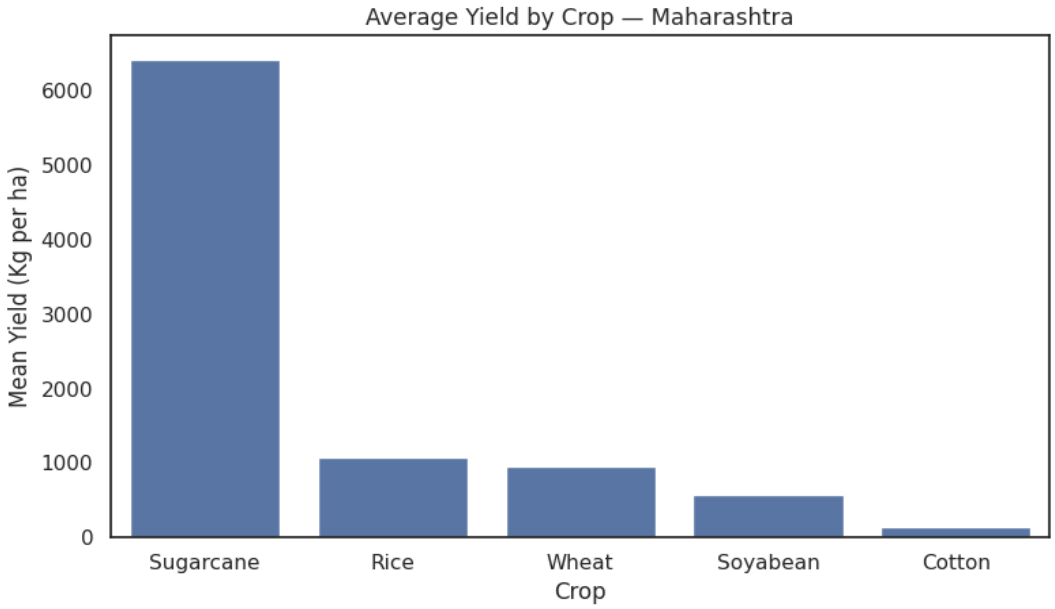

Fig 1: Average Yield by crop- Maharashtra

Correlation Analysis

The correlation heatmap shows that area and production are almost perfectly correlated (r ≈ 0.9–1.0), meaning larger cultivated area results in proportionally higher output. In contrast, area and yield show weak relationships (r ≈ ±0.3), confirming that yield variation is influenced more by external factors than land area[11].[12]

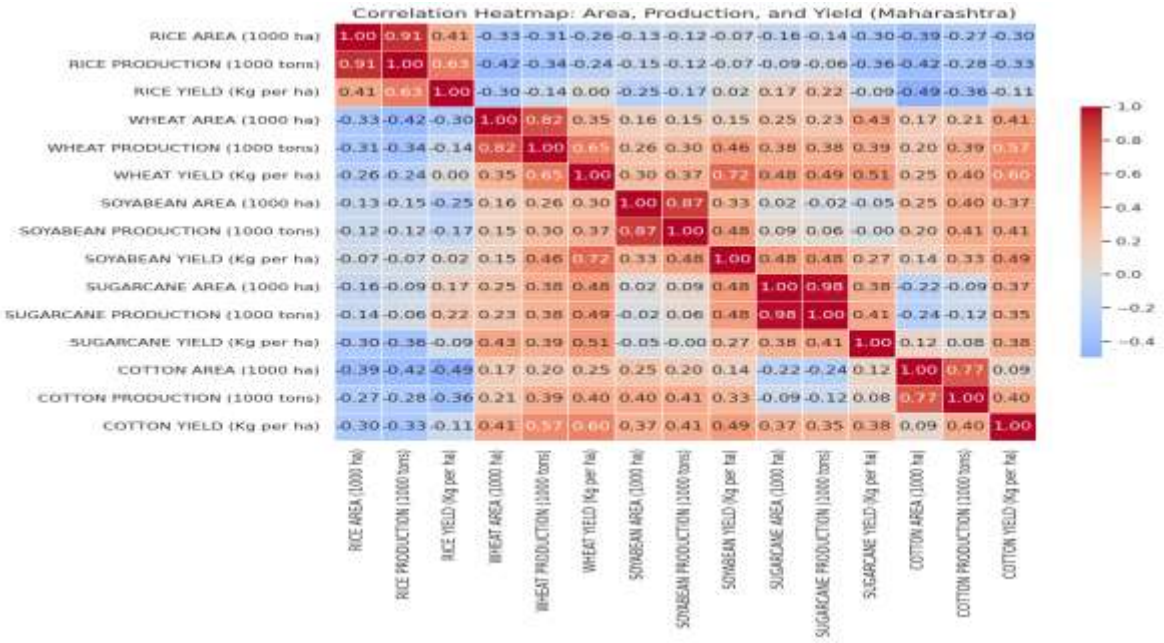

Fig 2: Correlation Heatmap:Area,Production ,and Yield (Maharashtra)

Regression Analysis

The regression analysis shows that the independent variable has a statistically significant effect on all five crops ($p < 0.001$). However, the strength of these relationships varies. Wheat exhibits the strongest association with an R^2 of 0.381, indicating that about 38% of its yield variation is explained by the model. Soybean shows a moderate relationship ($R^2 = 0.181$), while Rice, Sugarcane, and Cotton have very low R^2 values (below 0.06), meaning the predictor explains little of their yield variation. Although Sugarcane has the highest coefficient (25.06), suggesting a large effect size, its weak R^2 indicates high variability. Overall, while the predictor significantly influences all crops, it is most effective in explaining yield variation for Wheat and least for the others.[6]-[9]

Crop	Coefficient	Intercept	R^2	p-value
Rice	1.31	1317.88	0.048	<0.001
Wheat	6.65	980.06	0.381	<0.001
Soyabean	3.71	192.23	0.181	<0.001
Sugarcane	25.06	4201.78	0.055	<0.001
Cotton	0.68	105.65	0.059	<0.001

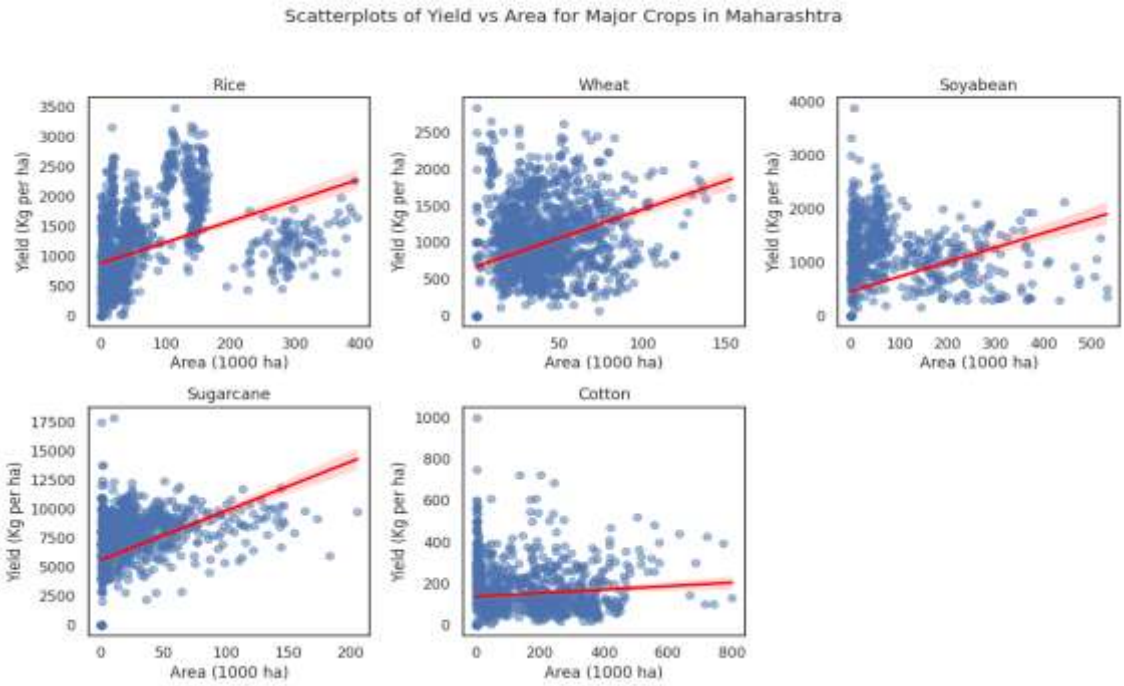

Fig 3: Scatterplot of Yield Vs Area for Major Crops in Maharashtra

ANOVA Analysis

ANOVA tests were conducted to examine yield differences among crops. Results show a significant F-statistic ($F = 1674.6$, $p < 0.001$), confirming that average yields differ across crops[14]. Sugarcane yields are statistically higher, while Soyabean and Cotton yields are significantly lower.

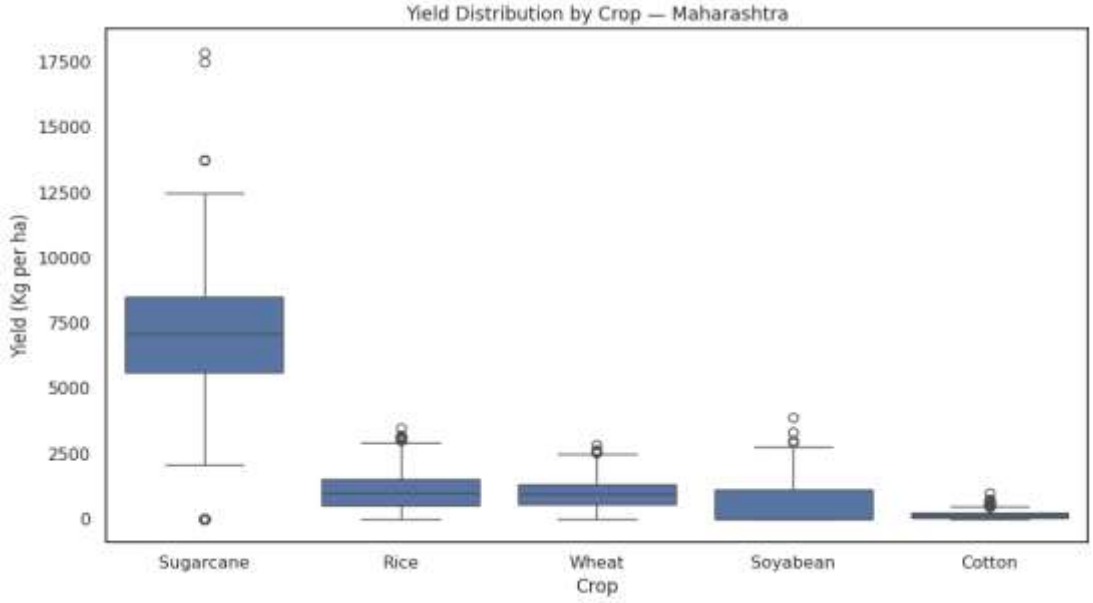

Fig 4: Yield Distribution by Crops- Maharashtra

t-Test: Rice vs Wheat Yields

The independent t-test compared the mean yields of Rice and Wheat. The t-statistic = -0.4837, p-value = 0.6286, indicating no significant difference in mean yields between Rice and Wheat. Both crops exhibit comparable productivity.[15]

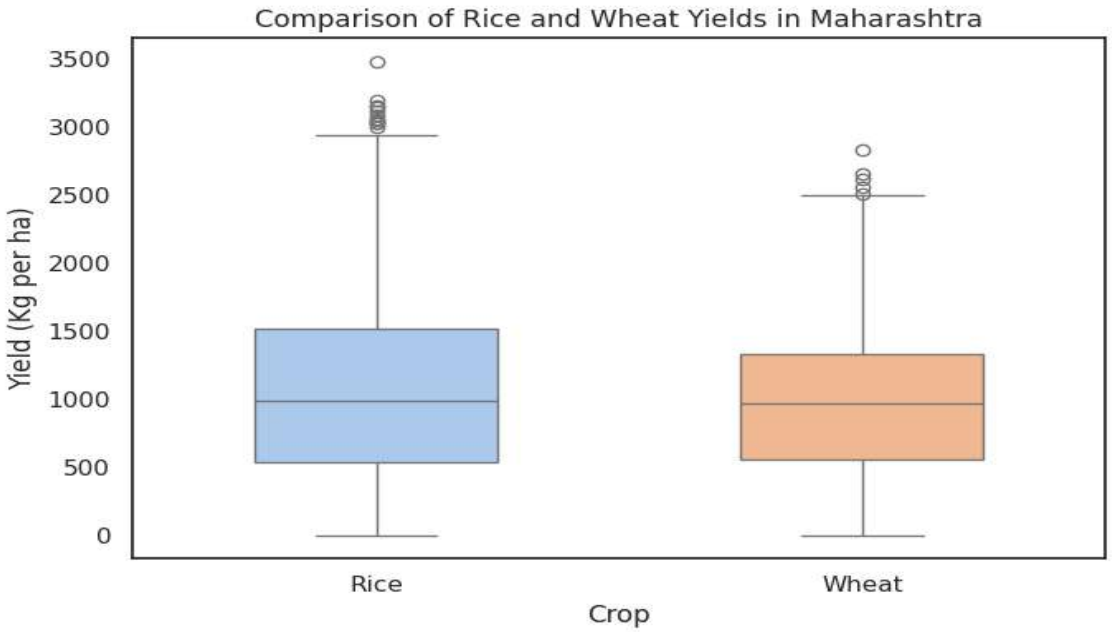

Fig 5: Comparison of Rice and Wheat Yield in Maharashtra

Time Series Analysis

Time series trends (1966–2015) show a steady increase in yields for all crops. Wheat and Sugarcane exhibit the most consistent upward trends. Rice shows gradual improvement, while Soyabean and Cotton display fluctuations due to rainfall variability.

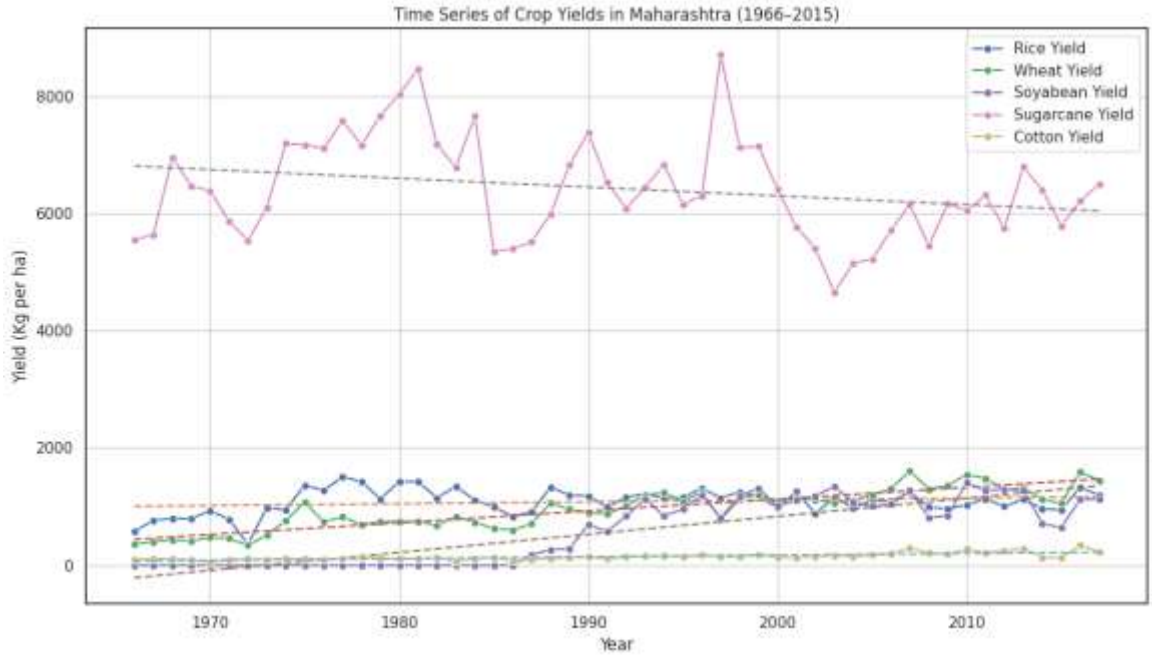

Fig 6: Time Series of Crop Yield in Maharashtra

Discussion

The integrated results of correlation, regression, ANOVA, t-test, and time series analyses collectively demonstrate that crop productivity in Maharashtra is shaped by both land-use dynamics and technological factors [3], [4], [5], [13]. While cultivated area expansion continues to play a role in production [1], [11], yield improvement through technology adoption and irrigation remains the most significant contributor to agricultural growth [4], [12], [13]. The statistical tests confirm that yield variability across crops is significant, yet certain crops (like Rice and Wheat) maintain stable productivity levels. The time series trends further highlight the gradual improvement in overall productivity but also reveal the persistent challenges faced by rained crops [15]. These findings underscore the need for targeted policy interventions focusing on water management, crop diversification, and input optimization to sustain long-term productivity growth.

V. CONCLUSION AND FUTURE SCOPE

Summary of Findings

This study statistically examined the productivity of major crops in Maharashtra — Rice, Wheat, Soyabean, Sugarcane, and Cotton — using correlation, regression, ANOVA, t-test, and time series analysis.

The results revealed significant variations in yield patterns among crops. Wheat and Sugarcane showed strong positive trends over time, while Soyabean and Cotton displayed high variability due to rainfall dependence. Correlation analysis confirmed a close relationship between production and cultivated area, but yield improvement was influenced more by technological and irrigation factors than by land expansion.

ANOVA results demonstrated significant differences in mean yields among crops, and the t-test indicated no meaningful difference between Rice and Wheat yields. Time series analysis showed gradual improvement in overall productivity from 1966–2015, highlighting the role of modernization in agriculture.

Recommendations

Based on the findings, the following recommendations are suggested:

1. **Focus on technology-driven productivity** — Adoption of high-yielding seed varieties and mechanization should be expanded.
2. **Enhance irrigation infrastructure** — Targeted water resource management can reduce yield variability.
3. **Promote balanced crop diversification** — Encourage both food and cash crops to maintain economic stability.
4. **Use statistical monitoring systems** — Regular regression and time series evaluations can help identify yield fluctuations and policy impacts.

Future Research Directions

Future studies can extend this work by:

1. Incorporating climatic and soil quality data to improve model accuracy.
2. Using non-linear and multivariate models such as ARIMA, VAR, or machine learning for yield forecasting.
3. Expanding the analysis to include district-level comparisons and post-2015 data to observe the impact of new agricultural policies.
4. Examining price–productivity relationships to link statistical yield performance with economic outcomes.

The study concludes that while area expansion contributes significantly to total production, it has little impact on yield improvement. Yield depends mainly on technological and environmental factors. Wheat and Sugarcane show stronger correlations between area and yield, while Rice, Soyabean, and Cotton are more influenced by rainfall and soil fertility. The similarity in Rice and Wheat mean yields (confirmed by t-test) suggests balanced productivity between these cereals. Policymakers should prioritize resource-efficient farming, irrigation coverage, and modern agronomic practices rather than increasing cultivated area.

ACKNOWLEDGEMENT

The author gratefully acknowledges the contributions of the ICRISAT data repository for providing reliable district-level agricultural data essential for this study. Sincere thanks are extended to Dr. D. Y. Patil Arts, Commerce and Science College and Savitribai Phule Pune University (SPPU) for their academic support and research facilities. The author also appreciates the guidance of faculty members and colleagues whose insights on statistical analysis and agricultural research greatly enhanced the quality of this work.

REFERENCES

1. ICRISAT. (2020). *District-Level Database for India (1966–2015).* International Crops Research Institute for the Semi-Arid Tropics. Hyderabad, India.
2. Directorate of Economics and Statistics, Government of Maharashtra. (2024). *Agricultural Statistical Information System (MAHADES Portal).* Retrieved from https://mahades.maharashtra.gov.in
3. Ministry of Agriculture and Farmers Welfare. (2024). *Agricultural Statistics at a Glance 2024.* Government of India, New Delhi.
4. NITI Aayog. (2022). *Agriculture and Farmer Productivity Report.* Government of India, New Delhi.
5. Food and Agriculture Organization of the United Nations (FAO). (2023). *FAOSTAT: Global Agricultural Database.* Retrieved from https://www.fao.org/faostat
6. Gujarati, D. N., & Porter, D. C. (2020). *Basic Econometrics* (6th ed.). New York, NY: McGraw-Hill Education.
7. Draper, N. R., & Smith, H. (2014). *Applied Regression Analysis* (3rd ed.). Hoboken, NJ: John Wiley & Sons.
8. Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to Linear Regression Analysis* (5th ed.). Hoboken, NJ: Wiley.
9. Gujrati, N. D. (2013). *Econometric Applications in Agriculture.* New Delhi, India: Oxford University Press.
10. Government of India. (2023). *Handbook of Agricultural Statistics 2023.* Directorate of Economics and Statistics, Ministry of Agriculture.
11. Deshmukh, A., & Kale, S. (2021). Statistical analysis of crop productivity trends in Maharashtra using regression and correlation models. *Indian Journal of Agricultural Economics, 76*(2), 245–258.

12. Patil, R., & Kharat, S. (2020). Impact of technological interventions on crop yield in semi-arid regions of India. *Journal of Statistics and Agricultural Research, 8*(3), 67–75.
13. Rao, C. H. (2019). *Agricultural Growth and Productivity in India: Policy Perspectives.* New Delhi, India: Academic Foundation.
14. Panse, V. G., & Sukhatme, P. V. (1985). *Statistical Methods for Agricultural Workers* (Revised ed.). New Delhi, India: Indian Council of Agricultural Research (ICAR).
15. Kumar, S., & Sharma, A. (2022). Time series analysis of major crop yields in India using ARIMA models. *International Journal of Applied Statistics and Data Science, 5*(1), 31–44.

AUTHOR BIOGRAPHY

Rupali Laxman Kamthe holds academic qualifications in Statistics and is currently affiliated with Dr. D. Y. Patil Arts, Commerce and Science College**,** under Savitribai Phule Pune University (SPPU)**.** Her research interests include real estate analytics, statistical modeling, agricultural statistics, and urban economic analysis**.** She is actively engaged in applying quantitative methods to study developmental trends, resource utilization, and socioeconomic patterns across sectors. Her work reflects a strong emphasis on data-driven insights and interdisciplinary approaches in applied statistics.