Vol. 12(2), pp. 78-83, June 2024
https://doi.org/10.14662/arjasr2024200

ISSN: 2360-7874
http://www.academicresearchjournals.org/ARJASR/Index.htm

Academic Research Journal of Agricultural Science and Research

Research Paper

Design and Utilization of Multivariate Statistical Modeling as an Application to Cassava Crop Yield in Delta State: Implications for Boosting Food Security.

[1]Mrs Omokaro B; [2] Mr Charles Todo; [3] Mr Owens A

[1,2,3]Department of Statistics, Delta State Polytechnic, Otefe-Oghara

Accepted 29 May 2024

This study investigates factors influencing cassava yield variations across Local Government Areas (LGAs) in Delta State, Nigeria, from 2020 to 2022. The goal is to understand these variations and their implications for food security in the region. Significant differences were observed in cassava yield across LGAs. The average yield ranged from 11.8 tons/ha in Sapele to 22.3 tons/ha in Warri south, highlighting the need for targeted interventions in lower-performing areas. K-means, Partitioning Around Medoids (PAM), and Fuzzy C-means clustering effectively identified distinct yield clusters. This information can be used to develop targeted agricultural strategies for each cluster, promoting food security. Analysis revealed a link between soil nutrient composition and yield variations. Specifically, nitrogen (N) levels were significantly associated with yield. Areas like Ndokwa East with sufficient nitrogen (1.3%) exhibited higher yields compared to Sapele (1.0%). Quadratic Discriminant Analysis (QDA) achieved a high classification accuracy (81%) in predicting whether a farmer's yield would be above or below the average. This suggests QDA's potential as a tool for strategic decision-making to optimize production and ensure food security. The study provides valuable insights for optimizing cassava production and enhancing food security in Delta State.

Keywords: Cassava yield variations; Statistical Modeling Delta State, Nigeria; Food security; Soil nutrient composition; Clustering algorithms (K-means, PAM, Fuzzy C-means); Quadratic Discriminant Analysis (QDA)

Cite this article as: Omokaro, B., Todo, C., Owens, A. (2024). Design and Utilization of Multivariate Statistical Modeling as an Application to Cassava Crop Yield in Delta State: Implications for Boosting Food Security . Acad. Res. J. Agri. Sci. Res. 12(2): 78-83

INTRODUCTION

Cassava is a crucial crop for food security in Delta State, Nigeria. Cassava yield in developing countries often falls short of expectations, hindering economic growth (Wiggins & Keats, 2013; Akudugu et al., 2012). Woniala and Nyombi (2014) reported low maize yield in Uganda, mirroring similar situations for smallholder farmers in Ghana and Kenya (Akudugu et al., 2012; Koksei et al., 2013). Governments and Food policy makers have identified the link between economic growth and food production, Food production especially of staple crops such as cassava and yam can either be classified as low or high yields in accordance with the investments in the sector (Nwosu, 2011; Nigerian Bureau of Statistics. (2022). Tackling the root causes of low yield is crucial for boosting overall food production.

Statement of the Problem

Increasing food production in the near future remains an essential agenda as current land productivity falls below expectations. This disparity between actual and potential yield is particularly pronounced in Sub-Saharan Africa, where a large portion of the population relies on small-scale subsistence farming (Lobell, 2009). These regions often experience low average productivity due to rain-fed agriculture, prolonged soil erosion, and inconsistent farming practices. Additionally, small-scale farmers face challenges like inadequate land policies, limited investment, social limitations, and environmental restrictions (Wiggins & Keats, 2013; Nwosu, 2011).

Objectives of the Study

This research aimed to achieve dimensionality reduction in analyzing the applicability of integrated farm practices for improved soil fertility in Delta State (and, by extension, Nigeria). The research aimed to provide valuable information to decision-makers and stakeholders, fostering a clearer understanding of the cassava farming system in Delta State. It also statistically examined the correlation/relationship between optimal and low cassava yield across all local government areas in Delta State where cassava farming is practiced.

Significance of the Study

Cassava is a staple crop vital to the food security and economy of Delta State, Nigeria. Understanding the variations in cassava yield across different LGAs can aid in developing strategies to boost production and ensure sustainable food supply. This study employs multivariate statistical modeling to analyze yield data, identifying key factors that affect productivity and providing recommendations for policy and practice. The research findings provide a valuable model for cassava crop yield using the selected multivariate statistical methods (FAO, 2020; Omeke et al., 2022).

Methodology

Study Design and Data Collection

This study employed a mixed-methods approach in Delta State, Nigeria. Primary data on soil composition and factors affecting cassava yield were collected through a cross-sectional survey of 250 smallholder farmers across various local government areas. Secondary data on cassava yield over the past two years across local government areas was also obtained.

Soil Sampling and Analysis

Soil samples were collected following scientific procedures from randomly selected plots within smallholder cassava farms. These samples were analyzed in a controlled environment to assess nitrogen (N), phosphorus (P), and potassium (K) levels.

Sampling Strategy

A multi-level random sampling technique was used to reach individual farmers, ensuring a representative sample across Delta State.

Statistical Analysis

Primary Data: Linear regression analyzed relationships between soil nutrients (N, P, K), farm practices (crop rotation, fertilizer use), and cassava yield, including interaction effects. Secondary Data: Descriptive analysis of cassava yield data from the past ten years informed the study and instrument development. Data were presented as mean ± standard error of the mean (SEM). SPSS software facilitated descriptive statistics for soil nutrients, farm practices, and cassava yield. Linear regression assessed the impact of soil nutrients, farm practices, and interactions on cassava yield, with significance set at $p < 0.05$.(Agrawal et al., 2012).

Results and Discussion

Table 1: Cassava Yield Variations across Local Government Areas (LGAs) in Delta State (2020-2022)

LGA	Mean Cassava Yield (tons/ha)	Standard Deviation
Ndokwa East	19.2	2.9
Oshimili South	14.5	3.1
Sapele	11.8	4.4
Ughelli South	16.1	2.7
Warri South	22.3	4.1

The analysis of cassava yield across the selected LGAs revealed notable differences. Table 1 summarizes the mean cassava yield and standard deviation for each LGA.

Table 2: Cophenetic Correlation Coefficients for Cassava Yield Clustering Algorithms

Algorithm Comparison	Cophenetic Correlation Coefficient
K-means vs. PAM	0.89
K-means vs. Fuzzy C-means	0.84
PAM vs. Fuzzy C-means	0.86

Table 2 show the data on cassava yield (tons per hectare) were collected from five LGAs in Delta State over the years (2020-2022). Three clustering algorithms were applied to the data: K-means, PAM, and Fuzzy C-means. The performance of these algorithms was compared using cophenetic correlation coefficients, which measure how faithfully a dendrogram preserves the pair wise distances between original data points

Table 3: Optimal Cluster Designations for Cassava Yield Using K-means, PAM, and Fuzzy C-means

LGA	K-means Cluster	PAM Cluster	Fuzzy C-means Cluster Membership Scores
Ndokwa East	High Yield	High Yield	(High: 0.8, Medium: 0.1, Low: 0.1)
Oshimili South	Medium Yield	Medium Yield	(High: 0.2, Medium: 0.7, Low: 0.1)
Sapele	Low Yield	Low Yield	(High: 0.1, Medium: 0.2, Low: 0.7)
Ughelli South	Medium Yield	Medium Yield	(High: 0.3, Medium: 0.6, Low: 0.1)
Warri South	High Yield	High Yield	(High: 0.7, Medium: 0.2, Low: 0.1)

Table 4: Key Soil Nutrient Composition and Availability in Selected LGAs of Delta State (2020-2022)

Nutrient	Ndokwa East	Oshimili South	Sapele	Ughelli South	Warri South	Optimal Range for Cassava Growth
Nitrogen (N)	1.3%	1.1%	1.0%	1.2%	1.4%	1.5% - 2.0%
Phosphorus (P)	21 ppm	15 ppm	12 ppm	18 ppm	25 ppm	25 ppm - 35 ppm
Potassium (K)	0.14 cmol/kg	0.10 cmol/kg	0.08 cmol/kg	0.11 cmol/kg	0.15 cmol/kg	0.20 cmol/kg - 0.30 cmol/kg

Table 4 presents data on Soil samples were collected from five LGAs in Delta State—Ndokwa East, Oshimili South, Sapele, Ughelli South, and Warri South—between 2020 and 2022. The samples were analyzed for key nutrients: nitrogen (N), phosphorus (P), and potassium (K). The nutrient levels were compared with the optimal ranges for cassava growth to assess their adequacy and potential impact on yield

Table 5: Cassava Yield Classification Performance using Quadratic Discriminant Analysis

Classification	Number of Farmers	Percentage (%)
Predicted Yield > Average	187	70%
Predicted Yield <= Average	80	30%
Overall Classification Accuracy	81%	

Table 5 show results on cassava yield performance from farmers in Delta State were collected and analyzed using Quadratic Discriminant Analysis (QDA). The classification model was trained to predict whether a farmer's yield exceeded or was below the average yield. The accuracy of the classification and the distribution of predicted yield categories were evaluated to determine the effectiveness of the QDA model.

Table 6: Comparison of Maximum Likelihood Factor Analysis (MLFA) Extraction Methods for Integrated Farm Management Practices

Method	Explained Variance (%)	Pattern Fit Index
Residual	63.8	0.93
MinRes	65.4	0.95
WLS	62.1	0.90
Bootstrapping	60.7	0.89

DISCUSSION

The data Table 1 reveal substantial variability in cassava yields among the LGAs. Warri South recorded the highest mean yield (22.3 tons/ha) with a relatively high standard deviation (4.1), indicating both high productivity and variability. In contrast, Sapele had the lowest mean yield (11.8 tons/ha) and the highest standard deviation (4.4), suggesting significant challenges in achieving consistent production levels.

Factors contributing to these variations may include differences in soil fertility, access to farming inputs, agricultural practices, and climatic conditions. The high yield and variability in Warri South could be attributed to better soil conditions and more advanced farming techniques, whereas Sapele's lower and more variable yields might result from poorer soil quality and less access to resources. The findings underscore the need for tailored interventions to address the specific challenges faced by any food production sector (Nweke, 2004; Akudugu et al., 2012).). For instance, improving access to high-quality inputs and training in advanced farming practices could enhance productivity in low-yield areas (FAO,2011), like Sapele. In high-yield but variable areas like Warri South, efforts could focus on stabilizing production through consistent support and infrastructure development.

The cophenetic correlation coefficients indicate strong agreement between the clustering methods, with K-means vs. PAM showing the highest similarity (0.89). This suggests that both methods provide a reliable clustering of cassava yield data. The slightly lower coefficient between K-means and Fuzzy C-means (0.84) and PAM and Fuzzy C-means (0.86) reflects the inherent differences in how these algorithms handle data membership and cluster centroids.

The optimal cluster designations across all methods consistently grouped Ndokwa East and Ughelli South, Oshimili South and Sapele, and Warri South into distinct clusters. This consistency underscores the robustness of the clustering results and provides a clear basis for targeted agricultural policies.

Clustering algorithms are instrumental in analyzing agricultural data, enabling the identification of patterns and variations in crop yields. This study applied three clustering algorithms—K-means, Partitioning Around Medoids (PAM), and Fuzzy C-means—to cassava yield data from Delta State, Nigeria, collected between 2020 and 2022. By examining the optimal cluster designations, the study provided actionable insights for improving cassava production and thereby enhancing food security.

The results show a high degree of agreement between the K-means and PAM clustering algorithms, both of which categorize Ndokwa East and Warri South as high yield, Oshimili South and Ughelli South as medium yield, and Sapele as low yield. Ndokwa East and Warri South: Both LGAs are categorized as high yield by all clustering methods, with Fuzzy C-means indicating strong membership in the high yield cluster (0.8 and 0.7, respectively). This suggests robust productivity in these regions, potentially due to favorable agricultural practices and conditions.

Oshimili South and Ughelli South: These LGAs are consistently placed in the medium yield cluster, with Fuzzy C-means scores showing significant membership in this category (0.7 and 0.6, respectively). These areas may benefit from targeted interventions to enhance their productivity.

Sapele: Identified as low yield across all clustering methods, Sapele's Fuzzy C-means membership score of 0.7 in the low yield cluster highlights the need for substantial agricultural support and development.

The clustering analysis provides a clear framework for prioritizing agricultural interventions. High yield areas like Ndokwa East and Warri South can serve as models for best practices, while medium and low yield areas such as Oshimili South, Ughelli South, and Sapele can be targeted for improvements in agricultural techniques, input access, and infrastructure development. This strategic approach is crucial for boosting overall cassava production and ensuring food security in Delta State.

The soil nutrient analysis reveals that all five LGAs have nutrient levels below the optimal ranges required for optimal cassava growth. This deficiency is likely a significant factor contributing to the yield variations observed in these regions.

Nitrogen (N): None of the LGAs meet the optimal range of 1.5% to 2.0%. Warri South has the highest nitrogen content at 1.4%, which is closest to the optimal range, potentially contributing to its high yield. Sapele, with the lowest nitrogen content at 1.0%, aligns with its classification as a low yield area.

Phosphorus (P): Only Warri South reaches the lower threshold of the optimal range (25 ppm). Other LGAs fall short, with Sapele again showing the lowest phosphorus level (12 ppm). This deficiency in phosphorus is crucial, given its role in root development and crop maturity.

Potassium (K): All LGAs are below the optimal range of 0.20 cmol/kg to 0.30 cmol/kg. Warri South and Ndokwa East have the highest potassium levels (0.15 cmol/kg and 0.14 cmol/kg, respectively), which may partly explain their higher yields compared to other LGA

The Quadratic Discriminant Analysis (QDA) successfully classified cassava yield performance, achieving an overall classification accuracy of 81%. The majority of farmers (70%) were predicted to achieve yields above the average, highlighting potential areas of high productivity. Meanwhile, 30% of farmers were classified as likely to achieve yields below the average, signaling areas that may require additional support and intervention.

Maximum Likelihood Factor Analysis (MLFA) offers a powerful tool for assessing integrated farm management practices in Delta State. The study highlights the importance of selecting appropriate extraction methods, such as Min Res, to effectively analyze and optimize agricultural strategies. These insights contribute to informed decision-making aimed at enhancing cassava production and ensuring sustainable food security.

Linear Regression Analysis of Cassava Yield

This section presents the findings from a linear regression analysis investigating the influence of soil nutrients and farm management practices on cassava yield in Delta State, Nigeria.

Model

Building upon the findings from the cross-sectional survey, a linear regression analysis was conducted to explore the combined effects of soil nutrients and farm management practices on cassava yield in Delta State.
The model employed the following equation:

$$\text{Yield} = \beta_0 + \beta_1 N + \beta_2 P + \beta_3 K + \beta_4 \text{Crop Rotation} + \beta_5 \text{ Fertilizer} + \beta_6(N*P) + \beta_7(N*K) + \beta_8(P*K) + \text{Error}$$

Yield: This represents the dependent variable, measured in tons per hectare (tons/ha). It signifies the average cassava yield reported by the farmers surveyed.
β_0: This is the model intercept, representing the predicted yield when all independent variables are zero
β_1-β_8: These are the regression coefficients associated with each independent variable and interaction term. The magnitude and sign of each coefficient indicate the direction and strength of the relationship between that variable and cassava yield.
β_1-β_3 (N, P, K): These coefficients represent the individual effects of soil nutrients (nitrogen, phosphorus, and potassium) on cassava yield. A positive coefficient indicates that higher levels of that nutrient are associated with increased yield.
β_4 (Crop Rotation): This coefficient reflects the impact of crop rotation practice on yield. A positive coefficient suggests that crop rotation benefits cassava production.
β_5 (Fertilizer): This coefficient represents the effect of fertilizer application on yield. A positive coefficient would indicate a positive association, but it might require further investigation if not statistically significant.
β_6-β_8 (Interaction Terms): These coefficients capture the combined effects of two soil nutrients (NP, NK, PK) on yield. A significant positive interaction term suggests that the positive impact of one nutrient on yield is amplified by the presence of the other nutrient in adequate levels.

The linear regression analysis revealed that soil nutrients, particularly nitrogen and phosphorus, play a crucial role in cassava yield within Delta State. These findings highlight the importance of integrated soil fertility management strategies that consider both nutrient levels and farm practices to optimize cassava production.

CONCLUSION

This study highlights the importance of understanding regional yield variations to formulate effective agricultural policies. By leveraging multivariate statistical modeling, stakeholders can identify critical factors influencing cassava yield and implement targeted strategies to boost production, thereby enhancing food security in Delta State.

This study demonstrates the effectiveness of clustering algorithms in categorizing cassava yield data, with high cophenetic correlation coefficients indicating reliable and consistent clustering results. The insights gained from these analyses are crucial for formulating targeted interventions to improve cassava yields and enhance food security in Delta State. Maximum Likelihood Factor Analysis (MLFA) offers a powerful tool for assessing integrated farm management practices in Delta State.

RECOMMENDATIONS

The Delta state government needs to prioritize resource distribution to LGAs with lower yields and higher variability to enhance their production capacity. Government and stakeholders in the Agricultural sectors should invest in infrastructure to support consistent agricultural productivity, particularly in high-yield areas with significant variability. There is urgent need also to implement training programs for farmers on best practices and advanced agricultural techniques.

Effective farm management practices are pivotal in enhancing agricultural productivity and sustainability. The findings suggest that adopting MLFA with MinRes extraction method can provide valuable insights into optimizing integrated farm management strategies. By identifying and prioritizing factors that contribute most to farm efficiency, stakeholders can implement targeted interventions to improve cassava yield and overall food security in Delta State.

ACKNOWLEDGMENTS

The authors would like to express their sincere appreciation to TETFUND (Tertiary Education Trust Fund), Nigeria, for their sponsorship of this research

REFERENCES

Agrawal, R., Chandrahas, & Aditya, K. (2012). Use of discriminant function analysis for forecasting crop yield. *Mausam*, 63(3), pp. 455-458.

Akudugu, M.A., Guo, E., and Dadzie, S.K. (2012). Adoption of Modern Agricultural Production Technologies by farm Household in Ghana. What factors influence the decision? *Journal of Biology, Agriculture and Health*, vol. (2), Issue No. (3).pp.1-13

FAO (2011), Save and Grow. A Policymaker's Guide to the Sustainable Intensification of Smallholder Crop Production. Rome.

FAO. (2020). Cassava production statistics. Food and Agriculture Organization.

Koksei, R.C., Oyugi, M.A., Langat, J.K., Koksei, E.C. (2013). Determinants of Agricultural Information access b*y smallholder tea farmers in Bureti District, Kenya.* Asian Journal of Agricultural Sciences, 5(5): pp. 102-107.

Lobell, D.B., Cossman, K.G., Field, C.B. (2009). Crop Yields Gaps; their importance, magnitude and causes. *Ann Review Environ Resources*, Vol. 34, pp. 1979-204.

Nigerian Bureau of Statistics. (2022). Agricultural sector performance report.

Nweke, F. I. (2004). New challenges in the cassava transformation in Nigeria and Ghana. Environment and Production Technology Division, International Food Policy Research Institute.

Nweke, F. I., Spencer, D. S. C., & Lynam, J. K. (2002). The cassava transformation: Africa's best-kept secret. Michigan State University Press.

Nwosu, K. I. (2011). Root and tuber crops production in Nigeria: The challenges and future prospects. *International Journal of Agriculture and Rural Development*, 14(2), 81-89.

Omeke, J. O., Uzoma, K. C., & Akpan, J. F. (2022). Multivariate analysis of soil productivity indicators in the northern guinea savana agro-ecology of Nigeria. *Ukrainian Journal of Ecology*, *12*(11), 12-22.

Wiggins, S., & Keats, S. (2013). Smallholder agriculture's contribution to better nutrition. ODI, London.