

Rice Yield Estimation in Nigeria: An Evaluation of Regression and Ratio Estimation Methods

UBA Tersoo; Ikughur, Atsua Jonathan; Onum Ene Sarah; Zaaya, Ipue Geoffrey

Department of Statistics, Joseph Sarwuan Tarka University Makurdi, Benue State, Nigeria.

DOI: <https://doi.org/10.51584/IJRIAS.2025.100700133>

Received: 12 July 2025; Accepted: 17 July 2025; Published: 21 August 2025

ABSTRACT

This study estimates the average rice yield in Nigeria using regression and ratio estimators. A sample of 15 states was selected out of the 36 states and Federal Capital Territory, Abuja, and data on land area (Hectares) and yield (tons) were collected from 2013 to 2021. The results show that both estimation methods produce similar estimates of the mean yield, with the regression estimator yielding 2.2745 tons/Ha and the ratio estimator yielding 2.2667 tons/Ha. The R-squared values (87.51% and 87.51%) indicate that both models explain a significant proportion of the variation in the data. The study highlights the importance of accurate estimation of crop yields for agricultural planning and policy-making.

Keywords: Rice Yield, Regression Estimator, Ratio Estimator, Nigeria, Agricultural Productivity.

INTRODUCTION

Rice yield refers to the amount of rice produced per unit area of cultivated land commonly measured in tons per hectare [8]. It is a critical indicator of agricultural productivity and efficiency in rice farming. In Nigeria, rice yield is influenced by several factors including land quality, access to irrigation, the use of modern agricultural inputs farm practices, etc. according to recent agricultural surveys, Nigeria's average rice yield varied across the country, reflecting regional differences in soil fertility, climate and farming technique ([10]. Understanding and accurately estimating rice yield is therefore essential for developing policies that can address these constraints and support sustainable rice production. In another study, [1] investigated the relationship between total agricultural land used and population growth rate in Nigeria from 1961 to 2018 using secondary data from Food and Agricultural Organization (FAO) of the United Nations and World Bank. They employed Descriptive, Trend equation and Correlation analysis to probe this association. Their findings showed that agricultural and arable land usage grew exponentially at 0.62% and 0.72% correspondingly, per annum, while, the total population growth rate stood at 2.5%. Meanwhile urban and rural populations grew at 4.75% and 1.6% respectively. Further findings revealed that agricultural and arable land utilization rate had significant positive correlation with the total population, urban and rural population. They however observed that most arable crops (such as yam, cassava, Cocoa bean, Maize, Sorghum, Groundnut and Rice) outputs increase mostly from land expansion rather than land productivity implying that the current situation cannot assure viable agricultural land use for food security in the near future.

In an attempt to investigate factors affecting yields of agricultural crops in Bahawalpur district in Pakistan, from 2008 to 2016, using multiple regression, [2] identified factors influencing higher yield/acre of crops as plough and rotator number, planking, irrigation, seed type, seed treatment DAP, Urea fertilizer and Farm yard manure. Others include latest crop varieties, certified seed, weed spray, disease and pests' spray, while soil type, excessive seed rate, home seed (local), weed disease and pest attacks were identified as negative factors affecting yields.

Relatedly, [3] explored factors crop production and those limiting yields. they identified technological factors such as genetic improvement of varieties, fertilizer technology, adaptive microbial technology, pesticides farm machinery and agronomic and management practices as those improving production whereas, late planting, wrong plant spacing, inaccurate planting method, insufficient sowing depth, delayed weeding, inaccurate pest and disease control, inappropriate fertilizer use and the use of low producing cultivars diminish yield expressively. This agrees with [4].

Apart from factors that affect productivity and yields, [5] noted in their work, the importance of crop residue as livestock feed. The review showed that when optimum management principles are followed, there is great probability for realizing the genetic potentials of the crop for residue yield and quality.

[6]) analyzed three interactive strategies that could probably raise future food production under global change. These includes Crop land expansion, Crop allocation and Agricultural intensification. They agreed that land expansion is less likely to increase food production and suggested proper allocation of crops in space and time, Climate Change, dietary shift and other socio-economic drivers which would shape the demand and supply side of food systems, crop-specific agricultural intensification for increasing yield per unit area of individual crop.

In another investigation, [7] examined the characterizing of the effect that environmental and induced factors have on crop yields. He outlined a framework which can be used to estimate the probability distribution function (pdf) of crop yield, which he argued may not be normal or log-normal as usually assumed. The study suggested the use of the Weibull distribution function.

Furthermore, [9] employed polynomial regression to estimate the effect of farm size on rice yield across various states in Nigeria, using cluster sampling to account for geographical variation. They found out that a second-degree polynomial model provided a better fit than a linear regression model for predicting rice yield based on farm size. They also noted that farms located in regions with access to irrigation and modern farming techniques had significant higher yields. Their study concluded that polynomial regression under cluster sampling is an effective for capturing the non-linear relationship between farm size and yield.

In another study [11] applied linear regression and ridge regression to estimate maize yield in clustered farms across northern region of Nigeria. Cluster sampling was used to group farms based on soil type and climate conditions. Their study found out that while linear regression provided reasonable estimates, ridge regression significantly improved model accuracy by addressing multicollinearity among the predictor variables such as soil quality and fertilizer use.

The review provides an overview of the importance of estimating crop yields, including rice, and highlights various factors that can impact yields. It also discusses different statistical models used to estimate crop yields. However, there is limited focus on land area and yield data: While the review mentions various factors affecting crop yields, it does not specifically focus on the relationship between land area and yield, which is the primary focus of this study. So, in this study we use the classical regression and ratio estimators of finite population mean under simple random sampling to estimate mean yield of rice (tons/ha) in Nigeria from 2013 to 2021

METHODOLOGY

Source of Data

The data used in the work is annual data on yield of rice (tons) and land area (ha) in the 36 states of Nigeria, including the Federal Capital City (FCT), Abuja from 2013 to 2021. It is available in Annual Abstracts of the National Bureau of Statistics (NBS). The Python programming language was used for data analysis.

Description of Simple Random Sampling Scheme

Consider a finite population, $U = \{U_1, U_2, \dots, U_N\}$. We draw a sample of size n from this population using Simple Random Sampling without replacement (SRSWOR) scheme. Let y be the study variable of interest and x the auxiliary variables and y_i, x_i be the observations in the i^{th} unit of the study variable and the auxiliary variable under consideration.

Preliminary Definitions

In this study, $y = \text{Yield of rice (tons)}$, and $x = \text{Land Area (Ha)}$. $n = \text{sample size}$, $N = \text{population size}$, $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i : \text{sample mean of the auxiliary variable } x$;

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i : \text{sample mean of the study variable } y,$$

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i : \text{population mean of the auxiliary variable } x;$$

$$f = \frac{n}{N} : \text{sampling fraction}; \quad \hat{R} = \frac{\bar{y}}{\bar{x}} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i} : \text{ratio estimator};$$

$$S_y^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \bar{Y})^2 : \text{Population variance of study variable};$$

$$S_x^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2 : \text{population variance of the auxiliary variable } x;$$

$$S_{xy} = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y}) : \text{population covariance of } x \text{ and } y;$$

$$s_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 : \text{sample variance of study variable},$$

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 : \text{sample variance of auxiliary variable } x;$$

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) : \text{sample covariance of } x \text{ and } y.,$$

The Classical Ratio Estimator of Finite population mean

The classical ratio estimator of finite population mean is given as:

$$\bar{y}_r = \bar{y} \frac{\bar{X}}{\bar{x}} = \hat{R} \bar{X} \quad (1)$$

The sample estimator of the bias is given as

$$\hat{B}(\bar{y}_r) = \frac{1-f}{n\bar{X}} (\hat{R}s_x^2 - s_{xy}) \quad (2)$$

The variance is given as

$$V(\bar{y}_r) = \frac{1-f}{n} (S_y^2 + R^2 S_x^2 - 2RS_{xy}) \quad (3)$$

Its sample estimator of the variance is

$$\hat{V}(\bar{y}_r) = \frac{1-f}{n} (s_y^2 + \hat{R}^2 s_x^2 - 2\hat{R}s_{xy}) \quad (4)$$

Note that the classical ratio estimator is preferable when there exists a strong positive relationship between the study variable y and the auxiliary variable, x and the regression line passes through the origin (or near the origin).

The Classical Regression Estimator of Population Mean

The classical regression population mean is given as

$$\bar{y}_{lr} = \bar{y} - \hat{\beta}(\bar{x} - \bar{X}) \quad (5)$$

where: $\hat{\beta}$ is the sample estimator of β , the regression coefficient. $\hat{\beta} = \frac{s_{xy}}{s_x^2}$.

The large sample approximation to the mean square error of \bar{y}_{lr} is

$$V(\bar{y}_{lr}) = \frac{1-f}{n} (S_y^2 + \beta^2 S_x^2 - 2\beta S_{xy}) \quad (6)$$

In large sample, the sample estimator of the variance of \bar{y}_{lr} is given as

$$\hat{V}(\bar{y}_r) = \frac{1-f}{n} (s_y^2 + \hat{\beta}^2 s_x^2 - 2\hat{\beta} s_{xy}) \quad (7)$$

Or

$$\hat{V}(\bar{y}_r) = \frac{1-f}{n} s_y^2 (1 - \hat{\rho}^2) \quad (8)$$

where: $\hat{\rho}^2 = \frac{s_{xy}^2}{s_y^2 s_x^2}$.

The classical regression estimator is preferable when there exists a strong positive relationship between the study variable y and the auxiliary variable, x and the regression line has an intercept.

Coefficient of Variation (CV)

The coefficient of variation is a key metric used to compare the efficiency of different sampling methods or estimators. It is defined as the ratio of the standard deviation (σ) to the mean (μ) and expressed as a percentage.

$$CV = \frac{\sigma}{\mu} \times 100\% \quad (9)$$

Coefficient of Determination (R^2)

The coefficient of determination, denoted as (R^2), measures the proportion of the variance in the dependent variable that is predictable from the independent variable(s) in a regression model. It can be represented as mathematically as

$$R^2 = 1 - \frac{[\sum_{i=1}^n (y_i - \hat{y}_i)]^2}{[\sum_{i=1}^n (y_i - \bar{y}_i)]^2} \quad (10)$$

where:

y_i = Actual value, \hat{y}_i = predicted value and \bar{y}_i = mean of actual values. R^2 is a useful metric for evaluating the goodness of fit of a regression model.

PRESENTATION OF RESULTS

In this Section, we present results of the analyses on data in this investigation.

Table 1: Population Descriptive Statistics

Parameter	LAND_AREA (X)	YIELD (Y)
Value 1	333	333
Mean	103426.3	232995.3
Value 2	72184.69	186893.7
Value 3	2144.2	1700

Value 4	319324	843741
Correlation Coefficient	—	0.9122

Table 2: Total Area and Yield of Rice per State (2013-2021)

State	Land Area (Ha)	Yield (Tons)
Abia	2.592876e+05	4.309974e+05
Adamawa	1.231550e+06	2.918076e+06
Akwa Ibom	1.114172e+05	1.030414e+05
Bauchi	1.090099e+06	2.614342e+06
Bayelsa	3.407158e+05	5.618891e+05
Benue	1.791962e+06	4.563573e+06
Borno	9.545778e+05	1.226434e+06
Cross River	6.281870e+05	9.510781e+05
Delta	2.061061e+05	4.507953e+05
Ebonyi	7.390615e+05	1.344277e+06
Edo	5.724819e+05	9.401793e+05
Ekiti	6.503946e+05	8.541610e+05
Enugu	4.381915e+05	9.629072e+05
F.C.T(Abuja)	1.425293e+06	3.048970e+06
Gombe	1.115220e+06	2.706444e+06
Imo	4.670812e+05	8.569590e+05
Jigawa	8.473128e+05	1.986174e+06
Kaduna	1.612117e+06	4.106520e+06
Kano	1.424441e+06	3.951385e+06
Katsina	9.276795e+05	2.466850e+06
Kebbi	1.605867e+06	3.625189e+06
Kogi	2.151527e+06	5.187612e+06
Kwara	2.022607e+06	4.681823e+06
Lagos	6.203670e+05	1.259656e+06
Nassarawa	1.291818e+06	3.620975e+06
Niger	1.684375e+06	4.027459e+06
Ogun	7.622282e+05	1.764014e+06
Ondo	4.325120e+05	9.411906e+05
Osun	3.366895e+05	8.022730e+05
Oyo	5.349243e+05	8.893802e+05
Plateau	9.829618e+05	2.720125e+06
Rivers	5.575564e+05	9.914723e+05

Sokoto	6.587009e+05	1.330355e+06
Taraba	1.449142e+06	3.051659e+06
Yobe	1.058780e+06	2.573643e+06
Zamfara	1.118645e+06	2.286642e+06

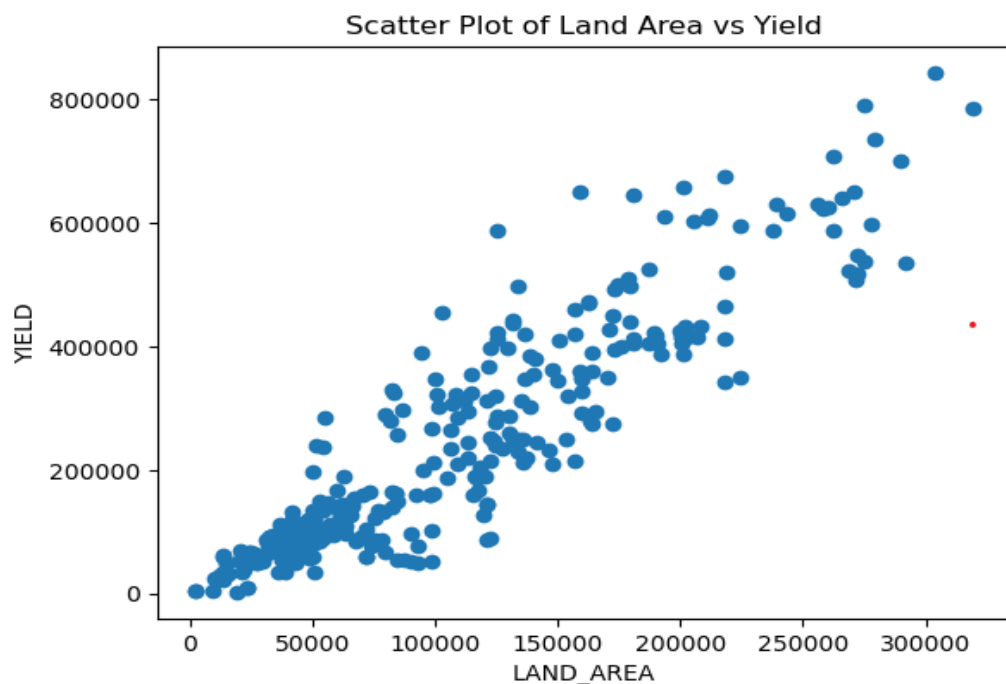


Figure 1: Showing Scatter Plot of Land Area (Ha) against Yield (tons)

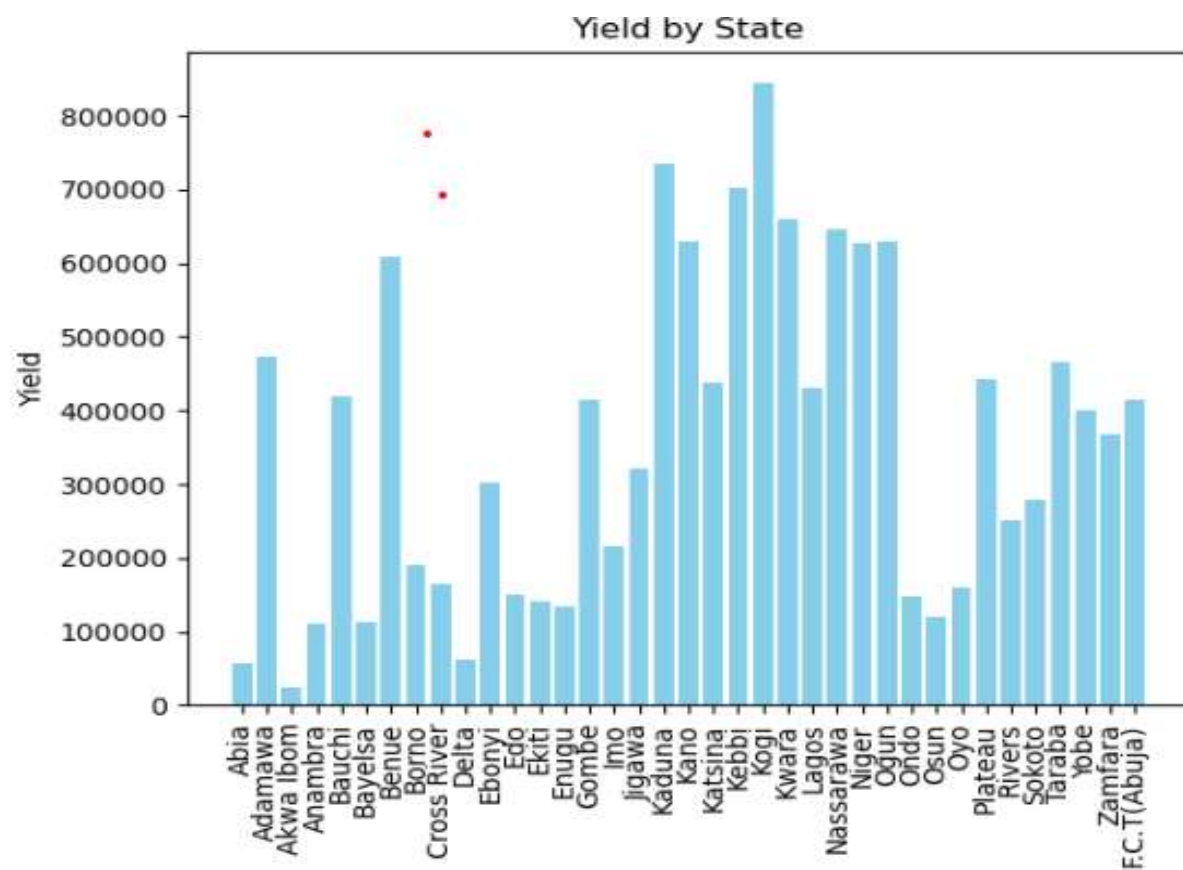


Figure 2: Bar Plot of Total Yield (tons) per State

Table 3: Sample Statistics of States, Land Area, Yield and Mean Yield per State

S/N	State	Land Area (Ha)	Yield (Tons)	Mean Yield/Hectare
1	Bayelsa	3.407158e+05	5.618891e+05	1.649143
2	Benue	1.791962e+06	4.563573e+06	2.546690
3	Edo	5.724819e+05	9.401793e+05	1.642286
4	Enugu	4.381915e+05	9.629072e+05	2.197458
5	Imo	4.670812e+05	8.569590e+05	1.834711
6	Jigawa	8.473128e+05	1.986174e+06	2.344086
7	Kebbi	1.605867e+06	3.625189e+06	2.257466
8	Lagos	6.203670e+05	1.259656e+06	2.030501
9	Niger	1.684375e+06	4.027459e+06	2.391070
10	Ogun	7.622282e+05	1.764014e+06	2.314286
11	Osun	3.366895e+05	8.022730e+05	2.382828
12	Oyo	5.349243e+05	8.893802e+05	1.662628
13	Plateau	9.829618e+05	2.720125e+06	2.767274
14	Taraba	1.449142e+06	3.051659e+06	2.105838
15	Yobe	1.058780e+06	2.573643e+06	2.430763

Table 4: Sample Estimates

Statistic	Estimator	
	Regression	Ratio
Mean Yield (tons/Ha)	2.2745	2.2667
R-squared	0.8752	0.8751
MSE	4083689397.3147	4084027987.2683
CV	28.21%	28.21%

DISCUSSION OF RESULTS

Table 1 provided statistics that offer insights into the distribution of land area and yield in the dataset. The average land area is approximately 103,426 hectares, with a standard deviation of 72,184 hectares. The land area ranges from 2,144 hectares to 319,323 hectares. The average yield is approximately 232,995 tons, with a standard deviation of 186,893 tons. The yield ranges from 1,700 tons to 843,741 tons. Variability: The standard deviations for land area and yield are substantial, indicating significant variability in the data.

Table 2 and Figure 2 show the total land area utilized for rice production and yield per state in Nigeria from 2013-2021. Some key observations:

The top 5 states with the highest total yield are Kogi (5.18 million tons), Kwara (4.68 million tons), Benue (4.56 million tons), Kaduna (4.11 million tons), and Niger (4.03 million tons). These states are likely to be major contributors to the country's rice production. The top 5 states with the highest total land area are Kogi (2.15 million hectares), Kwara (2.02 million hectares), Benue (1.79 million hectares), Kaduna (1.61 million hectares), and Kebbi (1.60 million hectares). These states have the largest areas dedicated to rice cultivation.

Anambra (0.34 million hectares) and Osun (0.34 million hectares) have relatively smaller land areas dedicated to rice production. There is significant variation in rice yield across states, with some states like Kogi and Benue having much higher yields than others like Anambra and Osun. The data suggests regional differences in rice production, with some states in the north and middle belt regions having larger land areas and higher yields. Given that the total area and yield refer specifically to rice production within the period of study, let's calculate the average yield per hectare.

Relationship between Land Area and Yield:

The scatter plot of land area against yield (Figure 1) suggests a positive relationship between the two variables. States with larger land areas tend to have higher yields, indicating that land area is an important factor in determining rice production. The results suggest that policymakers and stakeholders should focus on supporting rice farmers in the top-producing states, particularly Kogi, Kwara, Benue, Kaduna, and Niger. Strategies to improve rice production could include increasing access to inputs, improving irrigation systems, and providing training and extension services to farmers. Additionally, policymakers could consider implementing policies to support sustainable agriculture practices and reduce the environmental impact of rice production.

The results for the estimates of the mean yield per hectare are presented in Table 4. The study uses two estimation methods: regression and ratio estimators. Here's a comparison of the results:

Mean Yield Estimates

The mean yield estimates from both methods are similar, with the regression estimator yielding 2.2745 tons/Ha and the ratio estimator yielding 2.2667 tons/Ha. This suggests that both methods produce comparable estimates of the mean yield.

R-Squared Values

The R-squared values for both methods are identical (0.8751 and 0.8752), indicating that both models explain a significant proportion of the variation in the data.

Mean Squared Error (MSE)

The MSE values for both methods are also similar, with the regression estimator having an MSE of 4083689397.3147 and the ratio estimator having an MSE of 4084027987.2683. This suggests that both methods have similar levels of accuracy.

Coefficient of Variation (CV)

The CV values for both methods are identical (28.21%), indicating that both methods have similar levels of variability.

Comparison of Methods

Based on the results, it appears that both regression and ratio estimators produce similar estimates of the mean yield and have similar levels of accuracy. This suggests that both methods can be used to estimate rice yield in Nigeria. However, the choice of method may depend on the specific research question and data characteristics.

Implications

The similarity in results between the two methods has implications for researchers and policymakers. It suggests that both regression and ratio estimators can be used to estimate rice yield in Nigeria, and the choice of method may depend on the specific context and research question.

Overall, the results provide valuable insights into the performance of regression and ratio estimators in estimating rice yield in Nigeria. Further research can build on these findings to explore the use of these methods in other

contexts and to identify the most effective approach for estimating crop yields.

CONCLUSION

The study demonstrates the effectiveness of regression and ratio estimators in estimating rice yield in Nigeria. The findings have implications for policy-making, research, and agricultural development. Further research can build on these findings to explore the use of these methods in other contexts and to identify the most effective approach for estimating crop yields. The study's results can also inform the development of more effective agricultural practices and policies that support sustainable rice production in Nigeria. For instance, policymakers could use the study's results to identify areas with high potential for rice production and target interventions to support farmers in those regions. Moreover, the study's findings on the effectiveness of regression and ratio estimators can inform the development of more accurate and reliable crop yield estimation methods, which can in turn support more informed decision-making in agricultural policy and planning.

RECOMMENDATIONS

Some potential policy recommendations based on the study's findings include:

- i. Targeted interventions: Policymakers could target interventions to support farmers in regions with high potential for rice production, such as providing access to improved seeds, fertilizers, and irrigation systems.
- ii. Crop yield estimation: The study's findings on the effectiveness of regression and ratio estimators can inform the development of more accurate and reliable crop yield estimation methods, which can support more informed decision-making in agricultural policy and planning.
- iii. Agricultural development programs: Policymakers could use the study's results to inform the design and implementation of agricultural development programs, such as programs to promote sustainable rice production practices or improve farmer access to markets and credit.

Conflict of Interest: None.

REFERENCES

1. Akpan, S. B., & Ebong, V. O. (2021). Agricultural land use and population growth in Nigeria: The need for synergy for sustainable agricultural production. *Journal of Agribusiness and Rural Development*, 3(61), 269-278.
2. Shah, M. A. A., Mobsun, M., Chesneau, C., Zulfiqar, A., Jamal, F., Nadeem, K., & Sherwan, R. A. K. (2020). Analysis of factors affecting yield of agricultural crops in Bahawalpur district. *Proceedings of the Pakistan Academy of Science. A Physical and Computational Sciences*, 57(4), 99-112.
3. Yali, W., & Benga, T. (2022). Crops production and factors limiting yields. *Advances in Crop Science and Technology*, 10(10), 537-541.
4. Tadzi, N. L., & Mutengwa, S. C. (2020). Factors affecting yield of crops. *IntechOpen* <http://dx.doi.org/10.5772/intechopen.90672>.
5. S., Reddy, P. S., Bidinger, F., & Blumed, M. (2003). Crop management factors influencing yields and quality of crop residues. *Field Crops Research*, 64, 57-77.
6. Wu, W., Yu, Q., Verburg, P. H., You, L., Yang, P., & Tang, H. (2014). How could agricultural land systems contribute to raise food production under global change? *Journal of Integrative Agriculture*, 13(7). Available Online at www.sciencedirect.com.
7. Glover, T. F. (1985). Measuring factors affecting crop production (Economic Research Institute Study Paper No. 416). Utah State University. <https://digitalcommms.usu.edu/eri/416>.
8. [8] Food and Agriculture Organization of the United Nations. (2020). Rice production in Africa: Challenges and potential. <https://www.fao.org/publications/en>
9. Udoh, I., & Akpan, T. (2019). Polynomial regression in estimating farm yield: A case study of rice farms in Nigeria. *Journal of African Agricultural Research*, 16(6), 321-323.

10. Udoh, I., & Ogbonna, I. (2021). Efficiency of sampling methods in agricultural surveys. *Journal of African Agricultural Research*, 16(4), 231-245.
11. Ojo, A., & Afolabi, F. (2021). Polynomial and ridge regression models for estimating rice yield in Nigeria. *Journal of Agricultural Statistics*, 22(4), 135-148.

PYTHON CODES

```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

from sklearn.linear_model import LinearRegression

from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score#

*Load data

df = pd.read_excel('C:/Users/UBA TERSOO\Desktop/RICE YIELD.xlsx')

df

df.head()

df.info()

df.describe()

# Calculate total land area and yield per state

state_totals = df.groupby('STATE').agg({'LAND_AREA': 'sum', 'YIELD': 'sum'})

print("Total Land Area and Yield per State:")

print(state_totals)

# Randomly select 15 states

sample_states = np.random.choice(df['STATE'].unique(), size=17, replace=False)

sample_df = df[df['STATE'].isin(sample_states)]

# Assuming 'sample_df' is the DataFrame containing the sample data

# Group the sample data by state and calculate total land area and yield

sample_table = sample_df.groupby('STATE').agg({'LAND_AREA': 'sum', 'YIELD': 'sum'}).reset_index()

# Calculate mean yield per hectare for each state

sample_table['Mean_Yield_per_Hectare'] = sample_table['YIELD'] / sample_table['LAND_AREA']

# Print the sample table

print(sample_table)
```

```
# Regression Estimator
```

```
X = sample_df[['LAND_AREA']]
```

```
y = sample_df['YIELD']
```

```
model = LinearRegression()
```

```
model.fit(X, y)
```

```
beta0 = model.intercept_
```

```
beta1 = model.coef_[0]
```

```
mean_yield_per_ha_reg = beta1
```

```
y_pred = model.predict(X)
```

```
r2_reg = r2_score(y, y_pred)
```

```
mse_reg = np.mean((y - y_pred) ** 2)
```

```
cv_reg = np.sqrt(mse_reg) / np.mean(y) * 100
```

```
# Ratio Estimator
```

```
mean_yield_per_ha_ratio = sample_df['YIELD'].sum() / sample_df['LAND_AREA'].sum()
```

```
y_pred_ratio = sample_df['LAND_AREA'] * mean_yield_per_ha_ratio
```

```
mse_ratio = np.mean((sample_df['YIELD'] - y_pred_ratio) ** 2)
```

```
r2_ratio = r2_score(sample_df['YIELD'], y_pred_ratio)
```

```
cv_ratio = np.sqrt(mse_ratio) / np.mean(sample_df['YIELD']) * 100
```

```
# Print results
```

```
print("Sample Estimates:")
```

```
print(f"Total Land Area: {total_land_area_sample} hectares")
```

```
print(f"Total Yield: {total_yield_sample} tons")
```

```
print(f"Mean Yield per Hectare (Regression): {mean_yield_per_ha_reg:.4f} tons/ha")
```

```
print(f"R-squared (Regression): {r2_reg:.4f}")
```

```
print(f"MSE (Regression): {mse_reg:.4f}")
```

```
print(f"CV (Regression): {cv_reg:.2f}%")
```

```
print(f"Regression Coefficient (beta1): {beta1:.4f}")
```

```
print(f"Mean Yield per Hectare (Ratio): {mean_yield_per_ha_ratio:.4f} tons/ha")
```

```
print(f"R-squared (Ratio): {r2_ratio:.4f}")
```

```
print(f'MSE (Ratio): {mse_ratio:.4f}')
```

```
print(f'CV (Ratio): {cv_ratio:.2f}%')
```

```
# Coefficient of Determination (R-squared)
```

```
X = df['LAND_AREA']
```

```
y = df['YIELD']
```

```
mean_y = y.mean()
```

```
ss_tot = np.sum((y - mean_y) ** 2)
```

```
model = np.polyfit(X, y, 1)
```

```
y_pred = np.polyval(model, X)
```

```
ss_res = np.sum((y - y_pred) ** 2)
```

```
r2 = 1 - (ss_res / ss_tot)
```

```
print(f'\nR-squared: {r2:.4f}')
```

```
# df is my DataFrame with LAND_AREA and YIELD columns
```

```
plt.figure(figsize=(8,6))
```

```
plt.scatter(df['LAND_AREA'], df['YIELD'], color='blue')
```

```
plt.title('Land Area vs Yield')
```

```
plt.xlabel('Land Area')
```

```
plt.ylabel('Yield')
```

```
plt.show()
```

```
# Create a figure with multiple subplots
```

```
fig, axs = plt.subplots(2, 2, figsize=(12,10))
```

```
# Plot 1: Yield by State
```

```
axs[0, 0].bar(df['STATE'], df['YIELD'], color='skyblue')
```

```
axs[0, 0].set_title('Yield by State')
```

```
axs[0, 0].set_xlabel('State')
```

```
axs[0, 0].set_ylabel('Yield')
```

```
axs[0, 0].tick_params(axis='x', rotation=90)
```

```
# Plot 2: Land Area vs Yield
```

```
axs[0, 1].scatter(df['LAND_AREA'], df['YIELD'], color='blue')
```

```
axs[0, 1].set_title('Land Area vs Yield')
```

```
axs[0, 1].set_xlabel('Land Area')
```

```
axs[0, 1].set_ylabel('Yield')
```

Plot 3: Yield over Year (if YEAR column exists)

```
if 'YEAR' in df.columns:
```

```
    axs[1, 0].plot(df['YEAR'], df['YIELD'], color='green')
```

```
    axs[1, 0].set_title('Yield over Year')
```

```
    axs[1, 0].set_xlabel('Year')
```

```
    axs[1, 0].set_ylabel('Yield')
```

Plot 4: Land Area distribution

```
axs[1, 1].hist(df['LAND_AREA'], color='orange', bins=10)
```

```
axs[1, 1].set_title('Land Area Distribution')
```

```
axs[1, 1].set_xlabel('Land Area')
```

```
axs[1, 1].set_ylabel('Frequency')
```

```
plt.tight_layout()
```

```
plt.show()
```