International Journal of Statistics and Applied Mathematics 2023; 8(3): 48-53

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2023; 8(3): 48-53

https://www.mathsjournal.com
Received: 25-03-2023
Accepted: 27-04-2023

Rama Shankar Yadav
Department of Agricultural Statistics, College of Agriculture, Acharya Narendra Deva University of Agriculture and Technology (ANDUAT), Kumarganj, Ayodhya, Uttar Pradesh, India

Vishal Mehta
Department of Agricultural Statistics, College of Agriculture, Acharya Narendra Deva University of Agriculture and Technology (ANDUAT), Kumarganj, Ayodhya, Uttar Pradesh, India

Gokul Krishnan KB
Department of Agricultural Statistics, College of Agriculture, Acharya Narendra Deva University of Agriculture and Technology (ANDUAT), Kumarganj, Ayodhya, Uttar Pradesh, India

Corresponding Author:
Vishal Mehta
Department of Agricultural Statistics, College of Agriculture, Acharya Narendra Deva University of Agriculture and Technology (ANDUAT), Kumarganj, Ayodhya, Uttar Pradesh, India

Regression models for forecasting nutri cereals area and production in India

Rama Shankar Yadav, Vishal Mehta and Gokul Krishnan KB

DOI: https://doi.org/10.22271/maths.2023.v8.i3a.979

Abstract
Making of necessary steps for maintaining food security and overcome environmental problems is mainly achieved by undergoing estimation of crop production and prediction of crop area. The art of anticipating the future values of cultivated crop in advance before its harvest is called as forecasting of crop production. In this study, the main objective was to develop a clear portray about the area and production of nutria-cereals in India. India is the leading producer of nutri-cereals which is almost 33.3 percent of total nutri cereals produced all over the world. The largest nutri cereals-producing state in India is Rajasthan which almost contributes 41.03 percent of total nutri cereals produced with in the country. Regression modeling was applied for both area and production of nutri cereals in India for a period 70 years (1951-2020) using linear, logarithmic, quadratic, cubic, power and exponential models. The results concluded that cubic model shown better performance compared to other models for prediction of area and production of nutri cereals for India with maximum accuracy.

Keywords: Nutri Cereals, area, production, regression analysis, modeling, forecasting

1. Introduction

The class of cereals like Sorghum (Jowar), barnyard millet (Sawa/Jhangora), kodo millet (Kodo), Pearl Millet (Bajra), etc., which are made of small seed and are placed in the *Poaceae* family. It is an important fact that our country, India, is termed as the original home of nutri cereals. Nutri cereals are tolerant to drought and other extreme weather conditions and it require only low inputs of fertilizers and pesticides. Almost there are thirty-five species of grasses from which 20 genera falls in the group called as small millets (https://nutricereals.dac.gov.in/). The world's largest leading producer of nutri cereals is India which almost share34 percent of total production all over the world [2]. A Nutri cereal is covering almost about 60 percent of gross cropped area of India [3]. During 2016-17, 14.72 million ha of land area in India was cultivated under nutri-cereal crops. Rajasthan is the leading state in India for nutri-ceral production which is approximately 4.515 million tonnes of nutri cereals which is 28 percent of total nutri cereals production [4]. In nutri cereals, Ragi has the highest mineral content [5].

In data science, regression plays a pivotal role in predicting different continuous targets such that it is not overemphasized and considered as ubiquitous task in practical and perception terms. Even nowadays, the studies related to regression methods are thoroughly inspected by researchers since it get maximum recognition in important [6].

In the present study, the nutri cereals area and production was fitted and predicted by regression modelling approach. The main outlook of the studywas application of regression models for forecasting of area and production. The main outcome expected from the current study was to obtain complete sketch about future anticipation on nutri cereals which may help the farmer to make and adopt the plans and also for the government for making related policies.

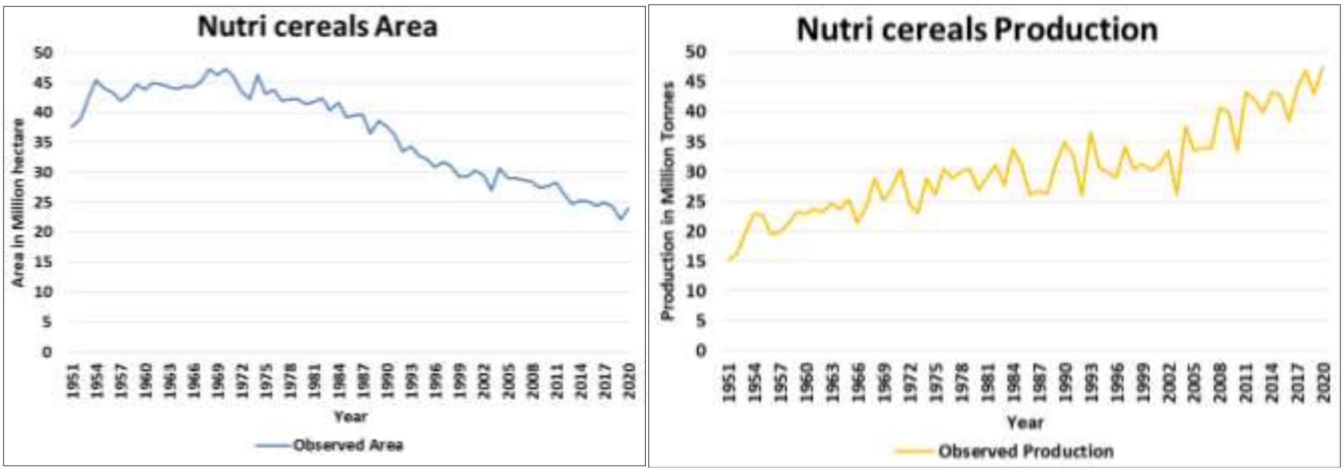

Fig 1: Observed area and production

2. Data and Methodology

Collection of data

The secondary data required for the study pertaining to area, production and yield of nutri cereals in India of 70 years (1951-2020) was collected from the Agricultural Statistics at a Glance 2020, Government of India (GOI) [7].

Table 1: Yearly data of nutri cereals area and Production

Area - Million Hectares
Production - Million Tonnes

Year	Area	Production	Year	Area	Production	Year	Area	Production
1951	37.67	15.38	1975	43.15	26.13	1999	29.34	31.34
1952	38.88	16.09	1976	43.80	30.41	2000	29.34	30.33
1953	42.45	19.61	1977	41.94	28.88	2001	30.26	31.08
1954	45.37	22.97	1978	42.28	30.02	2002	29.52	33.38
1955	43.92	22.82	1979	42.23	30.44	2003	26.99	26.07
1956	43.45	19.49	1980	41.36	26.97	2004	30.80	37.60
1957	42.02	19.87	1981	41.78	29.02	2005	29.03	33.46
1958	42.91	21.23	1982	42.45	31.09	2006	29.06	34.07
1959	44.66	23.18	1983	40.43	27.75	2007	28.71	33.02
1960	43.79	22.87	1984	41.71	33.90	2008	28.48	40.75
1961	44.96	23.74	1985	39.21	31.17	2009	27.45	40.04
1962	44.73	23.22	1986	39.47	26.20	2010	27.68	33.55
1963	44.29	24.63	1987	39.74	26.83	2011	28.34	43.40
1964	43.93	23.72	1988	36.55	26.36	2012	26.42	42.01
1965	44.35	25.37	1989	38.68	31.47	2013	24.76	40.04
1966	44.34	21.42	1990	37.69	34.76	2014	25.22	43.29
1967	45.09	24.05	1991	36.32	32.70	2015	25.17	42.86
1968	47.34	28.80	1992	33.42	25.99	2016	24.39	38.52
1969	46.24	25.18	1993	34.42	36.59	2017	25.01	43.77
1970	47.24	27.29	1994	32.82	30.82	2018	24.29	46.97
1971	45.95	30.55	1995	32.17	29.88	2019	22.15	43.06
1972	43.57	24.60	1996	30.88	29.03	2020	24.02*	47.48*
1973	42.21	23.14	1997	31.81	34.11			
1974	46.24	28.83	1998	31.05	30.40			

***Source* of Data:** Directorate of Economics & Statistics, DAC&FW, New Delhi.

***4th Advance Estimates**

Regression Analysis

In regression analysis, relationship associated between response variable and explanatory variables are obtained after undergoing analysis of the data. In this study, the variables selected for undergoing analysis are year, area and production. Area and production of nutri cereals are considered as response variables and year is recognized as explanatory variable for this study.

The different steps undertaken for conducting this study is illustrated asa flow chart in Fig. 2. The collection of data for area and production of nutri-cereals was done for a period of 70 years (1951-2020) from Agricultural Statistics at a Glance 2020, Government of India. The modeling of area and production of nutri-cereals was undergone by using simple regression analysis. The future observation of response variable was predicted using the relationship existed in among an independent and a dependent variable. Therefore, the analysis was conducted with an assumption that the future values predicted for area and production of nutri-cereals would follow the past trend. The easiness and ability to predict any data accessible was told as the main asset of this approach [8, 9, 11].

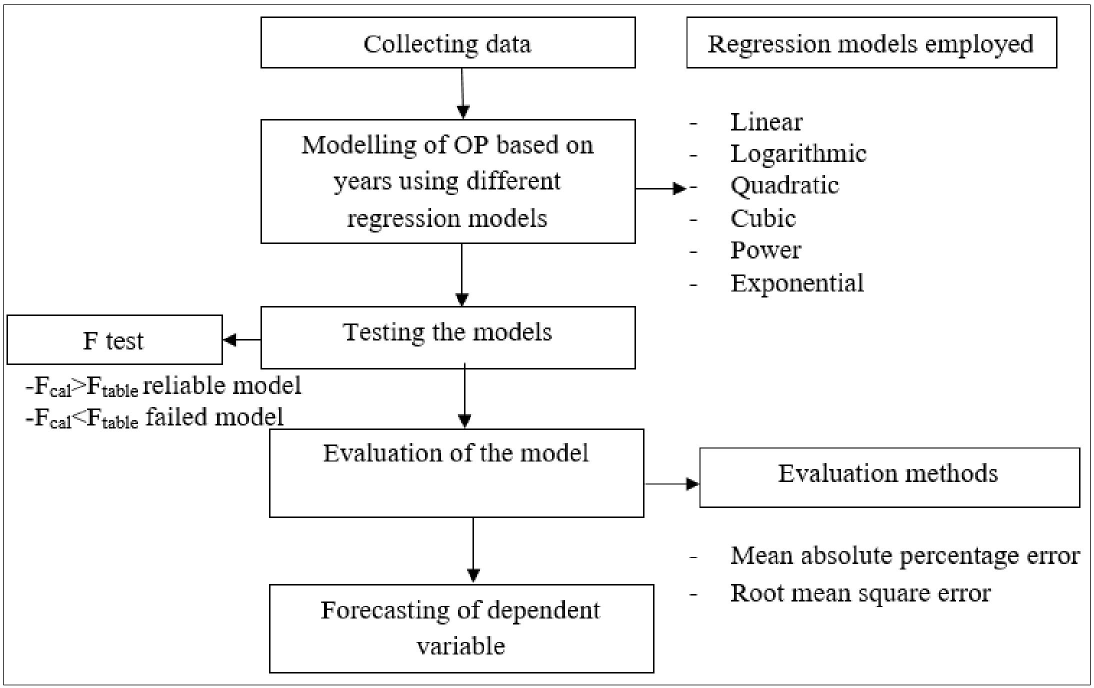

Fig 2: Flow chart of the study

For this study, regression analysis methods including both linear and non-linear methods like linear, logarithmic, power, exponential, quadratic and cubic regressions were utilized for fitting suitable models for area and production. It is noted that straight-line relationships exist for a linear model whereas it is curvilinear for a non-linear model. The data was split into two different datasets, training and testing datasets. The training dataset includes data for a period of 65 years (1951-2015) which was used for fitting the model whereas testing dataset consists of data for a period of last 5 years (2016-2020) used for evaluation of fitted regression model.

The regression models used in this study including both linear and non-linear models are mathematically expressed as:

Linear regression model:

$$Y_t = a + bt + \varepsilon_t \quad \text{------ (1)}$$

Quadratic regression model:

$$Y_t = a + bt + ct^2 + \varepsilon_t \quad \text{------ (2)}$$

Cubic regression model:

$$Y_t = a + bt + ct^2 + dt^3 + \varepsilon_t \quad \text{------ (3)}$$

Logarithmic regression model:

$$Y_t = a + b\ln(t) + \varepsilon_t \quad \text{------ (4)}$$

Power regression model:

$$Y_t = at^b + \varepsilon_t \quad \text{------ (5)}$$

Exponential regression model:

$$Y_t = ae^{bt} + \varepsilon_t \quad \text{------ (6)}$$

Where 'Y_t' indicates the dependent variable i.e., area or production, 't' denotes the independent variable, Time in years, 'a' is the intercept, 'b', 'c' and 'd' are representing the regression coefficients and 'ε_t' is the error term, $\varepsilon \sim N(0, \sigma^2)$.

The verification of fitted model is completed by determining coefficient of determination (R^2) and an F test. R^2 was used to obtain total changes in a response variable due to regression model. The efficacy of model fitted was indicated by a higher R^2 value. In order to undergo acceptance of fitted model and independent variable sophisticated in the model, F test was undergone. F test was used to identify whether fitted model was significant or not. If the F value obtained after undergoing F test is higher than the F table value (from F distribution table), the model is accepted or otherwise rejected.

The best forecasting model was identified by checking the work of regression models established for the data. Least value for RMSE and MAPE was the selected criteria for checking the work of fitted models (Eq. 7 and 8). Smaller values of RMSE and MAPE indicate the better prediction.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_t - \hat{y}_t)^2}{n}} \quad \text{----- (7)}$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_t - \bar{y}_t}{y_t}\right| \quad \text{----- (8)}$$

Where

y_t is the observed value and $\hat{y}_t$ is the predicted value, n is the total number of measurements.

Table 2: Linear and Non-linear models of Nutri cereals Area in India

Area Model	Parameter				Criteria			
	a	b	c	d	R^2	Adj. R^2	RMSE	MAPE
Linear	716.19*	-0.34*			0.835*	0.833*	3.066	6.36
Logarithmic	5188.87*	-678.49*			0.833*	0.831*	3.088	6.43
Quadratic	-22506.50*	23.05	-0.01		0.917*	0.914*	2.185	4.87
Cubic	-2133901*	3213.67*	-1.61*	0.00*	0.969*	0.968*	1.314	3.08
Power	153.14*	-19.69*			0.834*	0.831*	3.620	7.16
Exponential	23.31*	-0.01*			0.836*	0.834*	3.593	7.09

Table 3: Linear and Non-linear models of Nutri cereals Production in India

Production Model	Parameter				Criteria			
	a	b	C	d	R^2	Adj. R^2	RMSE	MAPE
Linear	-616.071*	0.326*			0.809*	0.806*	3.199	8.93
Logarithmic	-4874.36*	645.887*			0.808*	0.805*	3.208	8.93
Quadratic	8057.86*	-8.413	0.002		0.821*	0.815*	3.098	8.93
Cubic	-1956537*	2960.369*	-1.493*	0.0002*	0.870*	0.864*	2.637	7.24
Power	-160.995*	21.646*			0.814*	0.814*	3.089	8.89
Exponential	-18.266*	0.011*			0.814*	0.811*	3.084	8.89

1. Source: Secondary data collected from Directorate of Economics & Statistics, DAC&FW, New Delhi and https://eands.dacnet.nic.in/
2. A model's criteria value in bold numbers indicates that it is superior to other models in that criterion.
3. * indicate significant at 5% level of probability.

Table 4: Results of the Validation Tests for the Developed Models

Validation test	Regression type					
	Linear	Logarithmic	Quadratic	Cubic	Power	Exponential
t (area)	345.90	339.96	367.90	707.54	340.42	347.25
t (production)	287.25	285.58	153.28	147.24	297.54	297.19

Table 5: the Criteria of MAPE for Model Evaluation Based on Lewis (1982) [10]

MAPE	Evaluation
MAPE ≤ 10%	High accuracy prediction
10% < MAPE ≤ 20%	Good prediction
20% < MAPE ≤ 50%	Reasonable prediction
MAPE > 50%	Inaccurate prediction

Modeling of nutri cereals area and production

The results of the regression analysis using different methods are described in Table 2. The R^2 values lie above 0.969 and 0.870 for fitted regression models, suggested an immense degree of ling among area and years, production and year respectively. The determination coefficients also revealed that only limited percent (0.011–0.010%) of the variation in the area and production are due other unknown factors rather than the predictors (see Eqs. 1to 6). According to model evaluation criterion, cubic model is best fit for forecasting for area and production of nutri-cereals (Eqs. 9 and 10).

Forecasted Cubic model is Area

$$\hat{Y}_{Area} = -2133901 + 3213.67t - 1.61t^2 + 0.00t^3 \quad \text{-- (9)}$$

Forecasted Cubic model is Production

$$\hat{Y}_{Production} = -1956537 + 2960.37t - 1.49t^2 + 0.00t^3 \quad \text{-(10)}$$

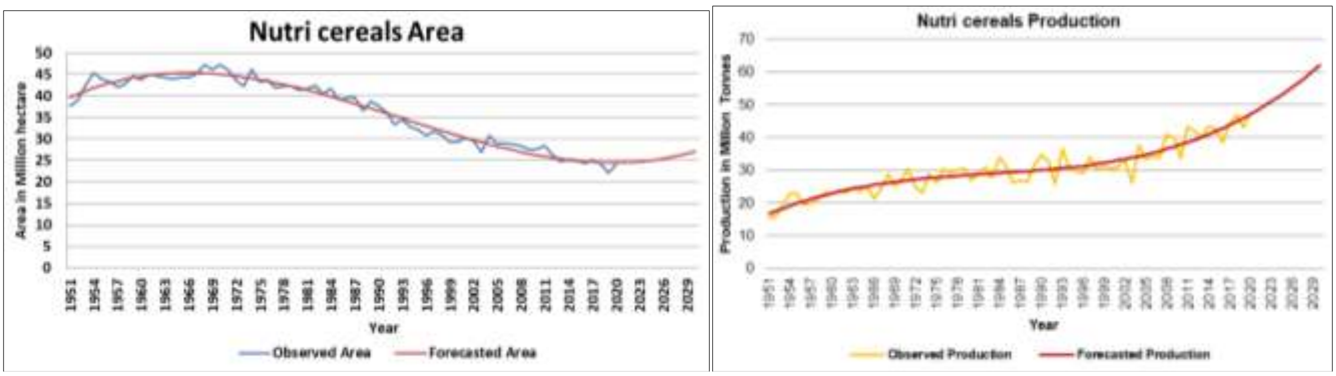

Fig 3: actual and forecasted value of area and production

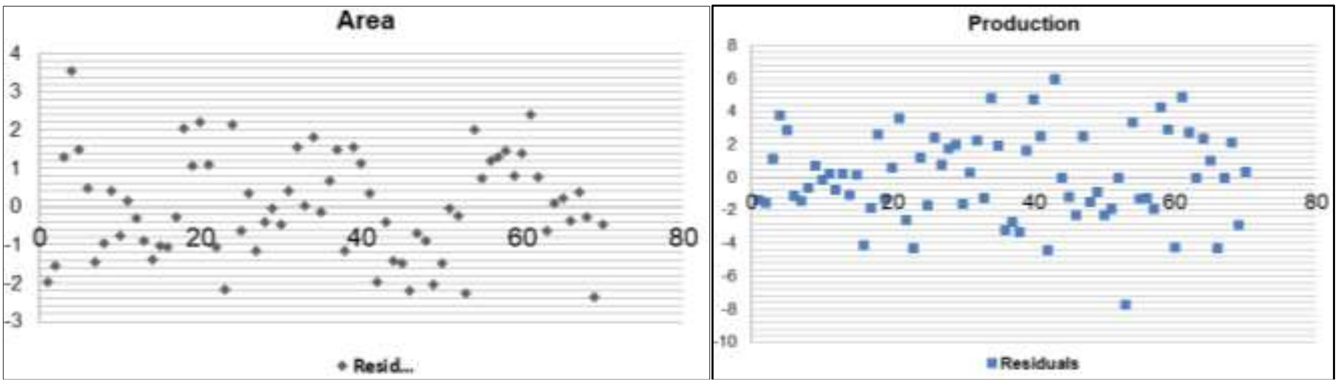

Fig 4: Forecasted residuals of area and production

The validation results of fitted models are described above. The calculated value of F is higher than the F table value which confirmed that Eqs. (7) to (8) are significant at 95% confidence level. Thus, it was concluded that explanatory variables employed in this study for fitting the regression models for area and production of nutri-cereals are significant.

3. Forecasting Nutri cereals

The MAPE values calculated for cubic model developed for area and production were determined as 3.08 and 7.24 respectively (see Table 2, 3). These results suggested that cubic model which is selected as best forming model can give decent prediction of the area and production of nutri-cereals. The cubic models have MAPE values which indicated that both models developed for area and production of nutri-cereals are high accuracy prediction models (see Table 2,3). It is also clearly visible in from Fig.3 that the actual and predicted values for the area and production of nutri-cereals lying very close to each other.

Table 6: Forecasted values of Nutri cereals Area and Production by Cubic Model

Year	Forecasted Area	Forecasted Production
2021	24.54217	48.34
2022	24.62659	49.62
2023	24.75832	50.95
2024	24.93896	52.35
2025	25.17015	53.81
2026	25.45349	55.33
2027	25.79062	56.92
2028	26.18314	58.57
2029	26.63268	60.29
2030	27.14085	62.09

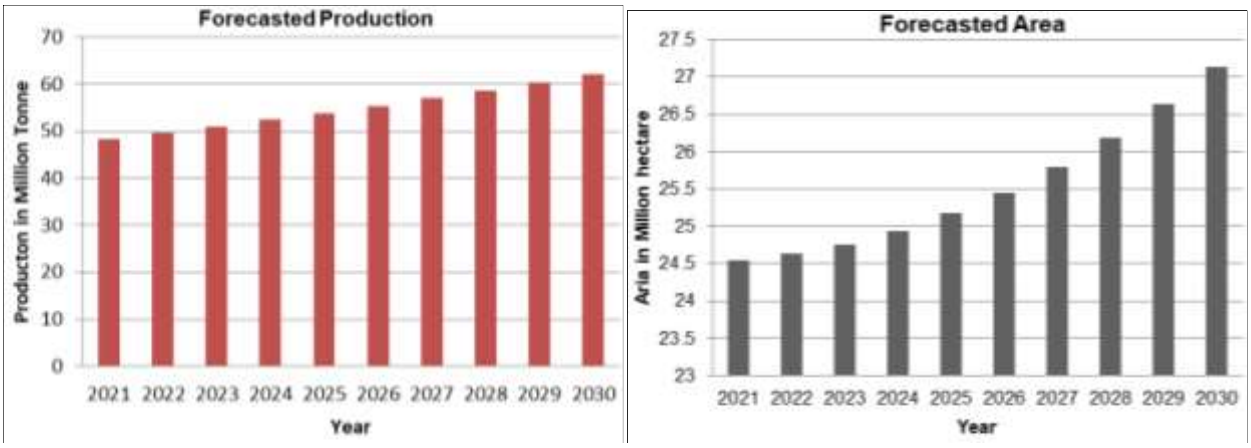

Fig 5: Bar graph of 10 years forecasted area and production

The area and production of nutri-cereals predicted using cubic model for the next 10 years are printed in Table 6 and illustrated in Fig. 3 and Fig. 5. The results suggested that the area and production of nutri-cereals may reach up to 27.14 million hectares and 62.09 million tonnes in 2030.

4. Conclusions

In this study, the regression modelling approach was used for predicting the area and production of nutri-cereals, which are defined for the prosperity of farmers in decision making and also helps the government in making of policy related to agriculture. The following are the main outcomes of the study:

- The study concluded that the cubic regression model outperformed all other models used in this study.
- The determination coefficients for the training data sets of area and production of nutri-cereals are 0.969 and 0.870 respectively, which indicated that fitted models are highly reliable.
- The F test showed a statistically significant result for the cubic model fitted for the area and production of nutri-cereals. The small values for MAPE and RMSE for the fitted models also concluded that best selected cubic model have high accuracy.

In future, the researchers must undergo the application of other models like artificial intelligence, grey theory forecasting, hybrid approach, multiple linear regression and other econometric methods for modelling the area and production of nutri-cereals. The results derived from the future studies can be compared with the present study for evaluation.

5. References

1. Website: https://nutricereals.dac.gov.in/
2. Website: https://ipad.fas.usda.gov/cropexplorer/cropview/commodityView.aspx?cropid=0459100&sel_year=2018&rankby=Production
3. Anbukkani P, Balaji SJ, Nithyashree ML. Production and consumption of minor millets in India- A structural break analysis. Annals of Agricultural Research. New Series. 2017;38(4):1-8.
4. Website: https://nutricereals.dac.gov.in/areaproduction.aspx
5. Website: https://vikaspedia.in/health/nutrition/nutritive-value-of-foods/nutritive-value-of-cereals-and-millets/millets-the-nutri-cereals#:~:text=Finger%20Millet%20(Ragi),-Finger%20millet%20is&text=Ragi%20has%20the%20highest%20mineral%20content.
6. Altman N, Krzywinski M. Simple linear regression. Nature Methods. Energy Sources. 2015;12(11):999-1000.
7. India. Ministry of Agriculture and Farmers Welfare, Department of Agriculture. Agricultural Statistics at a Glance. New Delhi, Government of India; c2021.
8. Kone CA, Buke T. Forecasting of CO_2 emissions from fuel combustion using trend analysis. Renew. Sustain. Energy Rev. 2010;14:2906-2915.
9. Aydin G. The application of trend analysis for coal demand modeling. Energy Sources Part B. 2015;10:183–191.
10. Lewis CD. International and business forecasting methods. London: Butterworths; c1982.
11. Yadav RS, Mehta V, Tiwari A. An application of time series ARIMA forecasting model for predicting nutri cereals area in India. The Pharma Innovation Journal. 2022;11(3):1260-1267.