The Pharma Innovation Journal 2022; 11(12): 3062-3067

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277-7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2022; 11(12): 3062-3067

www.thepharmajournal.com
Received: 14-10-2022
Accepted: 17-11-2022

Varalakshmi A
Department of Agricultural Statistics, Applied Mathematics and Computer Application, GKVK, UAS, Bengaluru, Karnataka, India

Mohan Kumar TL
Department of Agricultural Statistics, Applied Mathematics and Computer Application, GKVK, UAS, Bengaluru, Karnataka, India

Sushma R
Department of Agricultural Statistics, Applied Mathematics and Computer Application, GKVK, UAS, Bengaluru, Karnataka, India

Vinay HT
Department of Agricultural Statistics, Uttar Banga Krishi Viswavidyalaya, Pundibari, Coochbehar, Karnataka, India

Harish Nayak GH
ICAR-IASRI, New Delhi, India

Corresponding Author:
Varalakshmi A
Department of Agricultural Statistics, Applied Mathematics and Computer Application, GKVK, UAS, Bengaluru, Karnataka, India

Statistical analysis of area, production and productivity of banana crop in Karnataka, India

Varalakshmi A, Mohan Kumar TL, Sushma R, Vinay HT and Harish Nayak GH

Abstract
Banana is a type of fruit from herbaceous plants of the genus *Musa*. *Musa* species grow in a wide range of environments and have varied human uses, ranging from the edible bananas and plantains of the tropics to cold-hardy fiber and ornamental plants. The present study was aimed to analyze the trend and forecast in area, production and productivity of banana crop of Karnataka by using the secondary data of area, production and productivity of banana crop for the period of 18 years (2000-01 to 2017-18) collected from Directorate of Economics and Statistics, Karnataka. For this purpose, linear, quadratic, exponential, logistic and Gompertz models were fitted and the best-fitted model was selected based on lowest MAPE and residual assumption tests such as Runs test and Shapiro-Wilks test were carried out. Result revealed that exponential model was best-fitted for area under banana and logistic model for production and productivity of banana crop. The results reveal that there is increasing trend on area, production and productivity of banana crop in Karnataka in above study period. Based on this trend, the banana crop area, production and productivity was forecasted for next five years.

Keywords: Linear and nonlinear models, MAPE, Shapiro-Wilks test, run test, banana

Introduction
Banana (*Musa paradisiaca* L.) are native to South-East Asia and one of the most important fruit crops of the world. The banana is grown in the tropics and, though it is most widely consumed in those regions, it is valued worldwide for its flavour, nutritional value, and availability throughout the year. A ripe fruit contains as much as 22 percent of carbohydrate and is high in dietary fibre, potassium, manganese, and vitamins B6 and C (Britannica, 2022) [2]. In India bananas are so predominant and popular among people that it is liked both by poor and rich alike. Considering the nutritive value and fruit value of bananas, it could be considered as "poor man's apple" and it is the cheapest among all other fruits in the country (Madhunaik and Yogish, 2020) [4]. In India, banana ranks first in production and third in area among fruit crops and is extensively grown in tropical, subtropical and coastal regions. It accounts for about 14 per cent of the total area and 32 per cent of the total production of fruit crops. The total area under cultivation of banana in India is 884.00 thousand hectares, production is around 30808.00 thousand metric tonnes whereas productivity is only 34.85 metric tonnes per hectare during 2017-18. Production is highest in Andhra Pradesh, followed by Gujarat. The other major banana producing states are Maharashtra, Tamil Nadu, Uttar Pradesh and Karnataka (Anon, 2018) [1].

Banana is a tropical fruit, which is grown in almost all the parts of the Karnataka districts like Chamarajanagar, Mysore, Chitradurga, Shimoga, Kolar, Haveri, Bellary, Gulbarga, Chikmagalur, Uttara Kannada, Belgaum and Mandya. The area, production and productivity of the crop during 2017-18 were 1.11 lakh hectare, 23.29 lakh metric tonnes and 21.07 metric tonnes per hectare respectively. Compared to last year an increase in area by 1.8 per cent and a drop in production and productivity by 1.2 and 15 per cent respectively was recorded (Anon, 2018) [1]. In this paper, an attempt is made to assess the trend in area, production and productivity of banana crop in Karnataka and also growth was forecasted up to 2022-23.

Materials and Methods
The present study is conducted with the overall objective of estimating suitable regression model that explains the trend of area, production and productivity of banana crop in Karnataka. For this study, the secondary data pertaining to area, production and productivity of banana for the period of 18 years from 2000-01 to 2017-18 was collected from Directorate of Economics

and Statistics, Government of Karnataka.
Trend analysis is the changes that have occurred because of general tendency of the data to increase or decrease over a long period of time. The trend of area, production and productivity of banana crops in Karnataka was analyzed using linear and Non-linear regression models.

Linear regression models
In this method, trend in area, production and productivity of banana crop is measured by establishing mathematical relation between time and the response variable, which is depending on time. The mathematical expression can be represented by:

1. Linear (Straight line)

$$Y_t = \alpha + \beta t + \varepsilon \quad (1)$$

2. Quadratic (Parabolic)

$$Y_t = \alpha + \beta t + kt^2 + \varepsilon \quad (2)$$

Where
α: Intercept or Average effect
β, k: Slope or Regression Coefficients (β= linear effect parameter and k: Quadratic effect parameter)
Y_t: Area, production or productivity of banana crop in time period t

ε: Error term or disturbance term
Coefficients α, β and k are constant parameters and need to be estimated. Here, the relation is so derived that the sum of the squared deviations (errors) of the observed values from the theoretical values is least. The process of minimization of the sum of the squared errors results in some equations called normal equations. The normal equations are the equations, which are used for finding the coefficients of the relation, which is fitted by the method of least square.

Non-linear regression models
The following are some of the important nonlinear growth curve, which are generally used to describe the growth of time-series.

Exponential $Y_t = \alpha\beta^t + \varepsilon \quad (3)$

Logistic $Y_t = \frac{\alpha}{1+\beta\ exp(-kt)} + \varepsilon;\ \beta = \frac{\alpha}{Y_0} - 1 \quad (4)$

Gompertz $Y_t = \alpha\, exp(-\beta\, exp(-kt)) + \varepsilon;\ \beta = ln\left(\frac{\alpha}{Y_0}\right) \quad (5)$

Where
Y_t Represents area, production or productivity of banana crop in time period t
α, β and k are parameters and
ε denotes the error term.

The parameter 'k' is the 'intrinsic growth rate', while the parameter 'α' represents the 'carrying capacity or yield ceiling'. For the third parameter, although the same symbol 'β' was used, yet this represented different functions of the initial value Y_0 for different models (Mohan Kumar *et al*, 2012 and Prathima, 2018) [5, 7]. The present statistical analysis was carried out by using the LM procedure available in PROC NLIN facility of SAS software package. The LM iterative method requires specification of the initial estimates of each parameter of the models to be estimated. Initial value specification is one of the most difficult problems encountered in estimating parameters of nonlinear models. Inappropriate initial values will result in longer iteration, greater execution time, non-convergence of the iteration and possibly convergence to unwanted local minimum sum of squares residual. To start the iterative procedure, many sets of initial values were tried to ensure global convergence. The iterative procedure was stopped when the reduction between successive residual sums of squares was found to be negligibly small. More details on methods of finding initial estimates of the parameters of models can be found in Draper and Smith (1998) [3].

Assumptions of error term
Once the parameters of the models were estimated, diagnostic check of residuals of the fitted models has to be analyzed to check any violations in the main assumptions of 'independence of residuals' and 'normality of residuals'. This assumption was verified using,

Shapiro-Wilk's (*W*) test
Is the standard test for normality. The test statistic *W* is the ratio of the best estimator of the variance (based on the square of a linear combination of the order statistics) to the usual corrected sum of squares estimation of the variance. The values of *W* ranges from 0 to 1. When *W*=1 the given data are perfectly normal in distribution. When *W* is significantly smaller than 1, the assumption of normality is not met.
Test statistic is given by:

$$W = \frac{\left[\sum_{i-1}^{n} a_i x_{(i)}\right]^2}{\sum_{i-1}^{n}(x-\bar{x})^2}$$

Where
$x_{(i)}$ is the i^{th}order statistic, *i.e.*,, the i^{th} smallest number in the sample

$\bar{x}$ is sample mean and the constants $a_i$is given by

$$(a_1, a_2, \cdots, a_n) = \frac{m^T V^{-1}}{\sqrt{(m^T V^{-1} V^{-1} m)}}$$

Where $m^T = (m_1, m_2, \cdots, m_n)^T$ and $m_1, m_2, \cdots, m_n$are the expected values of the order - statistics of independent and identically distributed random variables sampled from the standard normal distribution, and *V* is the covariance matrix of those order statistics (Shapiro *et al*., 1968) [8].
Run test can be used to test the randomness of residuals. A Run is defined as 'a succession of identical symbols in which are followed and preceded by different symbols or no symbols at all'. If very few runs occur, a time trend or some bunching owing to lack of independence is suggested and if many runs occur, systematic short period cyclical fluctuations seem to be influencing the scores.

Null hypothesis
H_0: Sequence is random

Alternative Hypothesis

H_1: Sequence is not random

Let 'n_1', be the number of elements of one kind and 'n_2' be the number of elements of the other kind in a sequence of $N = n_1 + n_2$ binary events. For small samples *i.e.,*, both n_1 and n_2 are equal to or less than 20 if the number of runs r fall between the critical values, we accept the H_0 (null hypothesis) that the sequence of binary events is random otherwise, we reject the H_0.

For large samples *i.e.,*, if either n_1 or n_2 is larger than 15, a good approximation to the sampling distribution of r (runs) is the normal distribution, with

$$\text{Mean } (\mu_r) = \frac{2n_1n_2}{n_1+n_2}+1$$

$$\text{Variance } (\sigma_r{}^2) = \sqrt{\frac{2n_1n_2(2n_1n_2-n_1-n_2)}{(n_1n_2)^2(n_1+n_2-1)}}$$

Then H_0 can be tested using test statistic:

$$Z = \frac{r-\mu_r}{\sigma_r{}^2} \sim N(0,1)$$

The significance of any observed value of 'Z' computed using the equation may be determined from a normal distribution table.

Model adequacy checking

The goodness of fit of all the fitted models is assessed by computing Mean Absolute Percent Error (MAPE) which is given by:

Mean Average Percentage Error (MAPE)

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_t-\hat{Y}_t}{Y_t}\right| X\ 100$$

Where, Y_t= Actual values, $\hat{Y}_t$= Predicted values and n= number of observations

In the present study, to estimate trend area, production and productivity of banana crop, linear and non-linear models explained were fitted to find the best-fitted model. Only for those models, all the parameters are found to be significant at given level of significance, and assumptions of 'independence of residuals' and 'normality of residuals' are satisfied were considered as good fitted models. Among all the good fitted models, the best-fitted model was selected based on minimum MAPE values. This selected best-fitted model was used to of banana area, production and productivity in Karnataka for the period of 5 years from 2018-19 to 2022-23

Results and Discussion

The present study was undertaken with a view to analyze the trends in area, production and productivity of banana crop of Karnataka. The annual data pertaining to area, production and productivity of banana for the period of 18 years from 2000-01 to 2017-18 was used to build both linear model *viz.* linear, quadratic form of model and nonlinear growth model *viz.* exponential, logistic and Gompertz models. Such type of information generated proved to be very important and helpful for the policy makers in framing the strategies.

Model based trend analysis for area under banana in Karnataka

For the area under banana, the linear and nonlinear models were fitted. The results presented in the Table 1 reveals that, among the different fitted models, the parameters of Exponential model are found to be significant at 5 per cent level of significance with the minimum MAPE value of 9.49. Further, results from Table 1 also revealed that for the above fitted model, the number of runs and Shapiro-Wilk test statistic was found to be non-significant (p-value > 0.05) at 5 per cent level of significance indicating that assumptions of 'independence of residuals' and 'normality of residuals' were satisfied and considered as good fitted models. The observed and predicted area of banana by best fitted model for the study period is tabulated in Table 4 and the same is plotted in Fig. 1. The findings shown that the banana area has upward trend over the study period from 2000-01 to 2017-18.

Table 1: Parameter estimates and goodness of fit criteria by different models for area under banana (in thousand hectare) for the period of 2000-2001 to 2017-2018

Parameter	Models				
	Linear	Quadratic	Exponential	Logistic	Gompertz
α	19.61*	29.92*	30.41*	197.70*	8.77^{NS}
β	4.97*	1.87^{NS}	1.08*	6.64*	1.00^{NS}
k	-	0.16 ^{NS}	-	0.12*	2.92^{NS}
Test for randomness, normality of residuals and goodness of fit criteria					
Runs test (Z): (p –value)	2.08 * [0.04]	1.57 ^{NS} [0.12]	1.57^{NS} [0.12]	2.48* [0.01]	4.13* [0.0001]
Shapiro-Wilk(W): (p –value)	0.88* [0.03]	0.90* [0.05]	0.95^{NS} [0.39]	0.89* [0.04]	0.87* [0.02]
MAPE	13.21	9.42	9.49	10.34	100.00

* Significant at 5% level of significance, NS: Not Significant; Values in [.] indicate Probability values

Model based trend analysis for production of banana in Karnataka

The parameter estimates of the all fitted models for banana production are presented in Table 2. Logistic model was found to be the best fit for production of banana as the parameters of this model is found to be significant at 5 per cent level of significance with the minimum value of MAPE (22.14). Further, results from Table 2 also revealed that for all the above fitted models, the number of runs and Shapiro-Wilk test statistic was found to be non-significant (p-value > 0.05) at 5 per cent level of significance indicating that assumptions of 'independence of residuals' and 'normality of residuals' were satisfied and considered as good fitted models. The observed and predicted production of banana by best fitted models for the study period is tabulated in Table 4 and the same is plotted in Fig. 2. The findings shown that the banana production has upward trend over the study period from 2000-01 to 2017-18.

Model based trend analysis for productivity of banana in Karnataka

Linear and nonlinear models were fitted to estimate the trend in productivity of banana. The results presented in the Table 3 reveals that, among the different fitted models, logistic model was found to be the best fit with the minimum MAPE value of 15.56 and also the parameters of this model is found to be significant at 5 per cent level of significance. Further, results from Table 3 revealed that for all the above fitted models, the number of runs and Shapiro-Wilk test statistic were found to be non-significant (p-value > 0.05) at 5 per cent level of significance indicating that assumptions of 'independence of residuals' and 'normality of residuals' were satisfied and considered as good fitted models. The observed and predicted productivity of banana by best fitted model for the study period is tabulated in Table 4 and the same is plotted in Fig. 3. The findings shown that the productivity of banana has upward trend over the study period from 2000-01 to 2017-18.

Table 2: Parameter estimates and goodness of fit criteria by different models for banana production (in thousand metric tonnes) for the period of 2000-01 to 2017-2018

Parameter	Models				
	Linear	**Quadratic**	**Exponential**	**Logistic**	**Gompertz**
α	82.13^{NS}	340.31^{NS}	524.40*	3011.00*	21950015.00^{NS}
β	144.53 *	67.08^{NS}	1.10*	14.08 *	1.00^{NS}
k	-	4.08^{NS}	-	0.26*	44118.20^{NS}
Test for randomness, normality of residuals and goodness of fit criteria					
Runs test Z: (p –value)	1.60^{NS} [0.08]	2.48* [0.01]	2.65 * [0.01]	1.70^{NS} [0.09]	4.13* [0.0001]
Shapiro-Wilk(W): (p –value)	0.87* [0.02]	0.94^{NS} [0.25]	0.94^{NS} [0.24]	0.94^{NS} [0.24]	0.87* [0.02]
MAPE	29.42	23.99	24.10	22.14	100.00

* Significant at 5% level of significance, NS: Not Significant; Values in [.] indicate Probability values

Table 3: Parameter estimates and goodness of fit criteria by different models for banana productivity (in metric tonnes per hectare) for the period of 2000-01 to 2017-2018

Parameter	Models				
	Linear	**Quadratic**	**Exponential**	**Logistic**	**Gompertz**
α	14.36*	14.77*	14.99 ^{NS}	37.24*	9.83^{NS}
β	0.63 *	0.51^{NS}	1.03^{NS}	1.62*	1.00^{NS}
k	-	0.01^{NS}	-	0.07*	9.83^{NS}
Test for randomness, normality of residuals and goodness of fit criteria					
Runs test (Z): (p –value)	1.57 ^{NS} [0.12]	1.67^{NS} [0.095]	1.57^{NS} [0.12]	1.57^{NS} [0.12]	4.13* [0.0001]
Shapiro-Wilk(W): (p –value)	0.93^{NS} [0.20]	0.92^{NS} [0.14]	0.92^{NS} [0.12]	0.94^{NS} [0.26]	0.92^{NS} [0.10]
MAPE	15.65	15.59	15.61	15.56	100.00

* Significant at 5% level of significance, NS: Not Significant; Values in [.] indicate Probability values

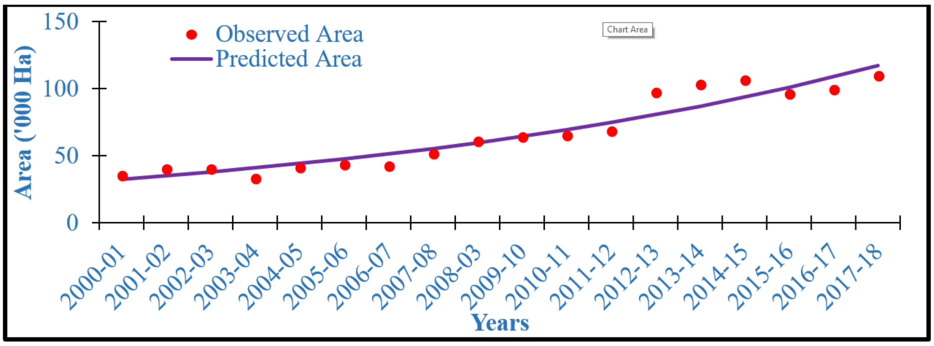

Fig 1: Observed and predicted values of area under banana by exponential model for the period of 2000-01 to 2017-18

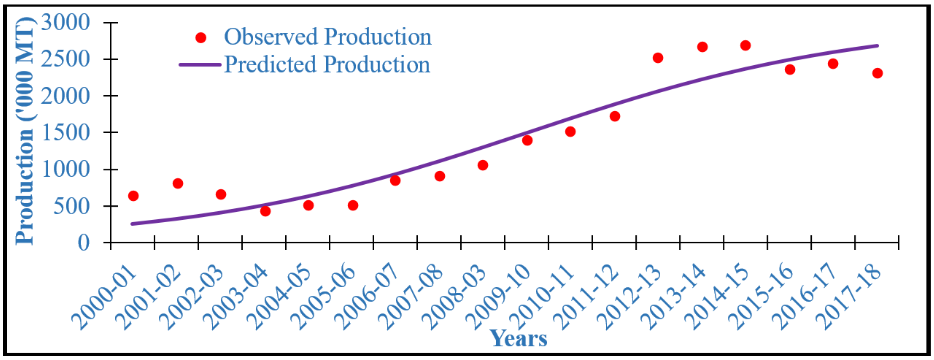

Fig 2: Observed and predicted values of banana production by logistic model for the period of 2000-01 to 2017-18

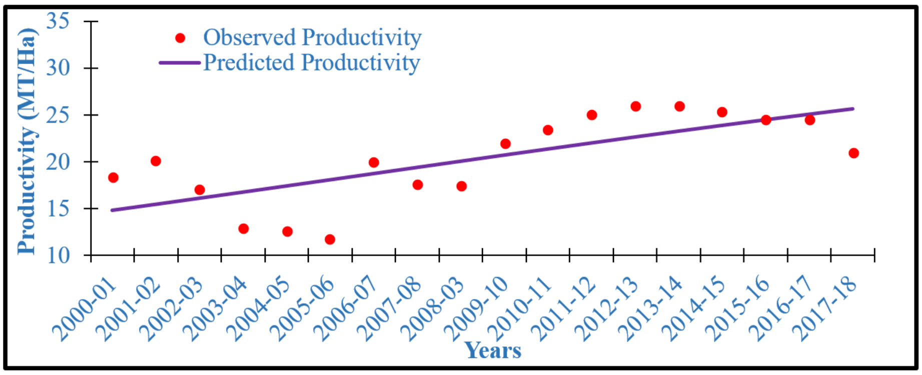

Fig 3: Observed and predicted values of banana productivity by logistic model for the period of 2000-01 to 2017-18

Table 4: Trend values of area (in thousand hectare), production (in thousand metric tonnes) and productivity (in metric tonnes per hectare) of banana by best fitted models

Years	By Exponential model		By Logistic model		By Logistic model	
	Actual Area	Area Trend values	Actual Production	Production Trend values	Actual Productivity	Productivity Trend values
2000-01	35.34	32.79	649.57	254.69	18.38	14.84
2001-02	40.25	35.34	815.57	323.03	20.26	15.48
2002-03	39.81	38.10	681.66	407.00	17.12	16.12
2003-04	33.22	41.07	431.46	508.67	12.99	16.77
2004-05	41.98	44.28	526.90	629.59	12.55	17.43
2005-06	43.56	47.73	517.26	770.40	11.87	18.09
2006-07	42.65	51.45	851.44	930.41	19.97	18.75
2007-08	52.19	55.46	923.00	1107.22	17.69	19.41
2008-09	61.00	59.79	1059.93	1296.70	17.38	20.07
2009-10	63.98	64.45	1407.16	1493.17	21.99	20.72
2010-11	65.76	69.48	1536.23	1690.06	23.36	21.37
2011-12	69.26	74.90	1730.52	1880.74	24.99	22.02
2012-13	97.40	80.74	2529.60	2059.39	25.97	22.65
2013-14	102.71	87.04	2675.63	2221.66	26.05	23.28
2014-15	106.61	93.83	2710.84	2364.94	25.43	23.89
2015-16	96.63	101.15	2370.95	2488.33	24.54	24.49
2016-17	99.46	109.04	2446.03	2592.32	24.59	25.08
2017-18	110.55	117.55	2328.98	2678.39	21.07	25.65

Forecasting of area, production and productivity of banana crops

Based on minimum MAPE, the best-fitted model was selected from among all the fitted Linear and nonlinear models to forecast the area, production and productivity of banana crop for the period of next 5 years from 2018-19 to 2022-23.

The forecasted area, production and productivity of banana in Karnataka are presented in the Table 5. The result in Table 5 shown that, the forecasted area under banana will be increased from 126.72 to 171.12 thousand hectares during the period from 2018-19 to 2022-23 by indicating an upward trend for the next five years. Forecasted banana production will also have increased trend from 2748.56 to 2913.76 thousand metric tonnes during the period from 2018-19 to 2022-23 by indicating an upward trend for the next five years. Further, an increasing trend can also be observed in productivity of banana from 26.21 to 28.28 metric tonne per hectare for the period from 2018-19 to 2022-23. Similarly, Okeke and Akarue (2018) [6] also reported that there is an increasing in the forecasted values for area, production and yield of cashew in Nigeria.

Table 5: Forecasted area, production and productivity of banana in Karnataka

Year	Area ('000 Ha)	Production ('000 MT)	Productivity (MT/Ha)
2018-19	126.72	2748.56	26.21
2019-20	136.60	2805.06	26.75
2020-21	147.25	2850.11	27.28
2021-22	158.74	2885.75	27.79
2022-23	171.12	2913.76	28.28

Conclusion

The current study attempted to study the trend in area, production and productivity of banana crop in Karnataka, for this purpose two linear and three non-linear models were fitted. Results of the present study revealed that exponential model was the best-fit for area under banana, and logistic model was best fit for production and productivity of banana. It was observed that the area, production and productivity of banana have shown an upward trend for the study period i.e., from 2000-01 to 2017-18 and forecasted period i.e., from 2018-19 to 2022-23.

Further, it can be concluded from the study that the area, production and productivity of banana crop has been increased since 2000, because Karnataka agriculture has undergone a major shift in the recent past, moving away from the cereal to non-cereal crops cultivation, especially toward the horticulture crops. Therefore, it can be concluded that the area production and productivity of banana crop has been increasing by replacing cereal crops.

References

1. Anonymous. Horticulture statistics at a glance. Horticulture Statistics Division, Department of Agriculture, Cooperation and Farmer's Welfare Ministry of Agriculture and Farmer's Welfare, Government of India; c2018.
2. Britannica. Banana. Encyclopaedia Britannica. Inc. 2022;32:1-4.
3. Draper NR, Smith H. Applied Regression Analysis. 3rd edition. John Wiley and Sons. U.S.A; c1998.
4. Madhunaik GG, Yogish SN. An economic analysis of banana cultivation in Chikmagalur district of Karnataka state. International Research Journal on Advanced Science Hub. 2020;2(9):105-107.
5. Mohan Kumar TL, Darshan MB, Sathishgowda CS, Sheelarani S. Comparison of nonlinear statistical growth models for describing coffee export trend in India. An Asian Journal of Horticulture. 2012;7(1):31-35.
6. Okeke DC, Akarue BO. Forecasting of cashew area harvested, yield and production using trend analysis. International Journal of Agricultural Economics. 2018;3(4):65-71.
7. Prathima CM. Statistical modelling of arrivals and prices of tender coconut in Maddur market. M.Sc. Thesis (UnPub.), University of Agricultural Sciences, Bengaluru; c2018.
8. Shapiro SS, Wilk MB, Chen HJ. A comparative study of various tests for normality. Journal of the American statistical association. 1968 Dec 1;63(324):1343-1372.