International Journal of Statistics and Applied Mathematics 2024; SP-9(5): 67-80

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2024; SP-9(5): 67-80

www.mathsjournal.com
Received: 02-07-2024
Accepted: 06-08-2024

LK Kumawat
M.Sc. Scholar, Department of Agricultural Statistics, C. P. College of Agriculture, SDAU, Sardarkrushinagar, Gujarat, India

PB Marviya
Assistant Professor, Department of Agricultural Statistics, C. P. College of Agriculture, SDAU, Sardarkrushinagar, Gujarat, India

S Kumawat
Ph.D., Scholar, Department of Agricultural Economics, S. K. N. College of Agriculture, SKNAU, Jobner, Jaipur, Rajasthan, India

GK Chaudhary
Associate Professor, Department of Agricultural Statistics, C. P. College of Agriculture, SDAU, Sardarkrushinagar, Gujarat, India

DM Padalia
M.Sc., Department of Agricultural Statistics, C. P. College of Agriculture, SDAU, Sardarkrushinagar, Gujarat, India

Tejas S
M.Sc., Department of Agricultural Statistics, C. P. College of Agriculture, SDAU, Sardarkrushinagar, Gujarat, India

IB Savaliya
M.Sc., Department of Agricultural Statistics, C. P. College of Agriculture, SDAU, Sardarkrushinagar, Gujarat, India

Corresponding Author:
LK Kumawat
M.Sc. Scholar, Department of Agricultural Statistics, C. P. College of Agriculture, SDAU, Sardarkrushinagar, Gujarat, India

Statistical modeling for forecasting of area, production and productivity of cumin in Banaskantha district of Gujarat

LK Kumawat, PB Marviya, S Kumawat, GK Chaudhary, DM Padalia, Tejas S and IB Savaliya

DOI: https://doi.org/10.22271/maths.2024.v9.i5Sb.1810

Abstract
The present study was carried out to estimate the trends of area, production and productivity of cumin in Banaskantha district of Gujarat. The time series data on area, production and productivity of cumin of Banaskantha district for the period of 2000-01 to 2022-23 were collected from the published reports by Directorate of Agriculture, Gujarat state, Gandhinagar. The data from 2000-01 to 2018-19 for cumin were used for model building and remaining data 2019-20 to 2022-23 for validation of the forecast model. An attempt was made in present investigation to fit polynomial and ARIMA models to arrive at a methodology that can precisely explain the fluctuation in area, production and productivity of cumin in Banaskantha district of Gujarat and to compare different models. The first, second and third degree polynomial models were fitted on original data as well as three, four and five year moving average data approach for the area, production and productivity. The Autoregressive Integrated Moving Average (ARIMA) models were fitted to original time series data after checking the stationary condition of the data. Among the fitted polynomial models, the suitable model was identified on the basis of significance of regression coefficient, adjusted R^2, root mean square error (RMSE); mean absolute error (MAE), normality by the Shapiro-Wilk test and randomness of residuals by the Run test. Different orders of ARIMA models (p, d, q) were judged on the basis of autocorrelation function (ACF) and partial autocorrelation function (PACF) at various lags. Different possible ARIMA models were fitted and from these, the models were selected on the basis of significant autoregressive and moving average term, Akaike's Information Criteria (AIC), Schwartz-Bayesian Criteria (SBC) values, adjusted R^2, normality of residuals by Shapiro-Wilk (1965) test and Box-Ljung (1978) test. For the trend, cubic model with five year moving average data approach was best suitable polynomial model for area and production of cumin in Banaskantha district of Gujarat. For the trend, quadratic model with five year moving average data approach was best suitable polynomial models for productivity of cumin in Banaskantha district of Gujarat. Among the ARIMA models, ARIMA (2, 1, 2) model was found suitable to explain the pattern of cumin productivity. None of the ARIMA models were suitable for area and production of cumin due to lack of one or more criteria for selection of models. Thus, in general because of crucial requirements of model selection criteria in polynomial as well as ARIMA models, few models could get selected.

Keywords: ARIMA, polynomial, forecasting, ACF, PACF, RMSE, cumin, AIC, SBC, Box-Ljung and Shapiro-wilk test

Introduction
In Gujarat, cumin (*Cuminum cyminum* L.) is known as *Jiru* or *Jeera* and ranks second after pepper in perspective of its importance. It is dried from an herb (*Cuminum cyminum*) and was called as "*SUGANDHAN*" in Ancient India. Cumin is the dried white fruit with greyish brown colour of a small slender annual herb. Cumin can be used as whole seed or in powder form. Cumin is helpful in digestion, lactation, increasing appetite and increasing eye sight. In India, total sowing area and production was 9.34 lakh ha and 6.29 lakh tonnes respectively. Gujarat state is highest in the country with 3.22 lakh tonnes of total production with 4.35 lakh ha of total sowing area of the country. Banaskantha district ranks second position in Gujarat state with coverage of 0.78 lakh ha total sowing area and 0.66 lakh tonnes total production of cumin, which is 16.30 percent of total sowing area of cumin. (Anonymous, 2022) [1].

Time series analysis is very useful in our day-to-day life. It is mainly used for forecasting purpose that means to estimate future values of key variable by using past values of some core indicator variables. Time series analysis is concerned with the techniques to study the nature of this kind of dependency among the observations. The statistical procedures and techniques based on the independence assumption are not valid in this case, hence requires different approaches.

One of the important time series model is Box Jenkins' autoregressive integrated moving average (ARIMA) methodology (Box *et al*., 1976) [2], which is widely used for analysis of univariate time series data. Due to its statistical properties, the ARIMA model becomes so popular. ARIMA is a flexible class of models including autoregressive (AR) models, moving average (MA) models and combination of AR and MA models *i.e.*, ARMA models. In addition to ARIMA, various exponential models can also be used to forecast a linear time series process (Brokwell and Davis, 1991) [3].

In ARIMA model, it is assumed that the future value of a variable is linearly related to the past values of the variable itself and random errors also. It is a linear univariate time series model, which expresses a time series process. This model is an extrapolation method for forecasting like any other such methods, which requires only the historical time series data on the variables under forecasting. It incorporates the futures of all such methods; do not require the investigator to choose initial values of any variables and values of the various parameters a priori. It is robust to handle any data pattern. As one would expect, this is quite a difficult model to develop any apply as it involves transformation of the variables, identification of the model, estimation through nonlinear method, verification of the model and derivation of the forecasts.

Therefore, the present investigation was undertaken to study fluctuation in area, production and productivity for cumin in Banaskantha district of Gujarat with the objectives *viz.*, to fit Linear, Quadratic, Cubic and Autoregressive Integrated Moving Average (ARIMA) models on area, production and productivity of selected major seed spices; to compare the performance of different Linear, Quadratic, Cubic and Autoregressive Integrated Moving Average (ARIMA) models on area, production and productivity of selected major seed spices and to validate the best fitted models of area, production and productivity of selected major seed spices.

Methodology

The time series data on area, production and productivity of cumin crop for the period 2000-01 to 2022-23 were obtained from Directorate of Agriculture, Gujarat state, Gandhinagar. In present investigation, the polynomial (Linear, quadratic and third degree polynomial) techniques were employed to study the trend pattern of area, production and yield of mustard in North Gujarat. Among the fitted polynomial models, the suitable model identified on the basis of significance of regression coefficients, R^2, root mean square error, mean absolute error, normality (Shapiro-Wilk test) and randomness of residual's (Run test) distribution.

Linear Regression Approach (Rangaswamy, 2006) [20]:

$$Y = a + bt$$

Where, a and b are the regression constant and regression coefficient, respectively to be estimated.

Quadratic Regression Approach (Montgomery *et al*., 2003) [14]:

$$Y = a + bt + ct^2$$

The unknown parameter *viz*. a, b and c estimated by using 'Principle of least square'.

Third Degree Polynomial Approach (Montgomery *et al*., 2003) [14]:

$$Y = a + bt + ct^2 + dt^3$$

The constant a and coefficient b, c and d estimated using least square method.

Goodness of fit of the models (Montgomery *et al*., 2003) [14]

To test the goodness of fit of the fitted polynomial models, the coefficient of determination R^2 can be also calculated as under

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}$$

Where, R^2 indicates the amount of variation of dependent variable accounted by the explanatory variables.

To test the overall significance of the model, the F test was used.

$$F = \frac{R^2 / k}{(1 - R^2) / (n - k - 1)}$$

Which follows F distribution with k, (n-k-1) degrees of freedom.

The individual regression coefficients was tested using the t test under the null hypothesis

$$t = \frac{b_j}{S.E.(b_j)}$$

Which follows t – distribution with (n-k-1) degrees of freedom, bj = estimated j^{th} coefficient and S.E. (bj) is the standard error of bj.

In addition to the above criteria, two more reliability statistics Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) were computed to measure the adequacy of the fitted model. They can be computed as under (Liew *et al*., 2000),

$$RMSE = [\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 / n]^{\frac{1}{2}}$$

$$MAE = \sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right| / n$$

The fitted models, which had lower values of these estimates, were considered to be better.

ARIMA model

An autoregressive integrated moving average (ARIMA) model is characterized by the notation ARIMA (p, d, q) where p, d and q denotes orders of auto-regression, integration (differencing) and moving average respectively. ARIMA is a parsimonious approach which can represent both stationary and non-stationary process. An ARMA (p, q) process is defined by the equation

$$Y_{1t} = \mu + \phi_1 Y_{1t-1} + \phi_2 Y_{1t-2} + \cdots + \phi_p Y_{1t-p} + \theta_0 \varepsilon_{1t} + \theta_1 \varepsilon_{1t-1} + \cdots + \theta_q \varepsilon_{1t-q}$$

Where,

Y_{1t} = Value of series at time period

μ= constant term.

ϕ_i (i = 1,2,...,p) and θ_1 (j=0,l,2,...,q) are model parameters.

ε_{1t} = Random Error at time period t.

ε_{1t} ~ IID (0, σ^2)

However, in practical applications, residuals obtained after fitting of appropriate ARIMA model may have non-constant error variance. Engle (1982) proposed to model time-varying conditional variance with auto-regressive conditional heteroscedasticity (ARCH) process using lagged disturbance.

Diagnostic checking

The ARIMA (p,d,q) model is assumed to be efficient when the residuals estimated from it are of white noise which can be ensured only when the residuals of the fitted model are used for diagnostic checking (Abdulla and Hossain, 2015). In this way, the estimated model is checked to verify if it adequately represents the series. In this study, diagnostic checks were performed on the residuals to see if they are randomly and normally distributed by using Jarque – Bera (JB) test for normality. In addition, the adequacy of the selected model was checked using Box-Ljung test.

The Jarque – Bera (JB) test of the following specification (Jarque and Bera, 1987) was used in the present study:

$$JB = \frac{n}{6}\left(s^2 + \frac{(K-3)^2}{4}\right) \sim \chi^2_{(2)}$$

Where,

$$\text{Skewness (s)} = \frac{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^3}{\left[\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2\right]^{\frac{3}{2}}}$$

$$\text{Kurtosis (k)} = \frac{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^4}{\left[\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2\right]^{2}}$$

Where, n is the number of observations, $\bar{Y}$ is the mean of the underlying variable series under study and y_i refers to the individual values of the variable under study. The statistic JB has an asymptotic chi-square distribution with 2 degrees of freedom and can be used to test the hypothesis of skewness being zero and excess kurtosis being zero. If JB >χ^2 (α, 2), then the null hypothesis will be rejected and it will be concluded that the data do not follow normal distribution.

In addition, an overall check of the model adequacy is to be made using Box-Ljung test (Ljung and Box, 1978) [11]. The test statistics is given by:

$$Q = n(n+2)\sum_{k=1}^{m}\frac{r_k^2}{n-k}$$

Where, n is the number of observations, r_k is the estimated autocorrelation of the series at lag k =1, 2,…, m and m is the number of lags being considered, $\chi^2_{(1-\alpha,h)}$is the chi-square distribution table value with ‘h’ degrees of freedom and level of significance ‘α’ such that P ($\chi^2_{(h)}$>$\chi^2_{(1-\alpha,h)}$) = 1- α and degrees of freedom, h = (m-p-q); p and q are the numbers of AR and MA terms, respectively.

In the present study, a formal test regarding the overall fitness of the model was done using Box-Ljung test of the residuals in the following manner: (i) Null hypothesis (H_0): The errors are distributed randomly, and (ii) Alternate hypothesis (H_1): The errors are non-random. Accordingly, the null hypothesis will be rejected if Q >$\chi^2_{(1-\alpha,h)}$ and the errors are not considered to be independent. On the other hand, the null hypothesis will be accepted, *i.e.* the errors are independent if Q <$\chi^2_{(1-\alpha,h)}$. Thereby, if the Q values happens to be significantly large than zero exceeding the table $\chi^2_{(1-\alpha,h)}$value then it is to be concluded that the residuals of the estimated model are probably auto-correlated and the entire model has to be reformulated.

Forecasting

Once the three previous steps of ARIMA model are over, then we are able to obtain the forecasted values by estimating appropriate model. ARIMA models was used to forecast the corresponding variable. For that, the entire data were segregated into two parts: one for sample period forecasts and the other for post-sample period forecasts. The former was used to develop confidence in the model and the latter was used to generate genuine forecasts for use in future planning.

In this regard, the actual value of the left out period and the forecasted value of the left out period from the selected model are used for cross-validation. For this, the percentage error is calculated such as:

$$\text{\% of Forecasting Error} = \left(\frac{Y - \hat{Y}}{Y}\right) \times 100$$

Where, Y is the observed value of remaining period and $\hat{Y}$ is the forecast values of given period under consideration.

Lower the value of forecasting error percentage, better is the prediction by the selected model. Besides, the accuracy of the forecasts for both ex-ante and ex-post were tested for the minimum values of Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and maximum value of coefficient of determination (R^2).

Goodness of fit for ARIMA model (Sarda and Prajneshu, 2002).

The suitability of ARIMA model was examined by

1. Akaike's Information Criterion (AIC).
2. Schwartz-Bayesian Information Criterion (SBC).
3. Akaike's Information Criterion (AIC).
4. The AIC value for ARIMA model was calculated as under.

$$\text{AIC} = \text{n} \log\hat{\sigma}^2 + 2\text{n}$$

Where, n denotes number of effective observation and $\hat{\sigma}^2$ is the white noise variance.

Schwartz-Bayesian Criterion (SBC)

SBC is computed as $\text{SBC} = \text{n}\log\hat{\sigma}^2 + \text{m}\log\text{n}$

Where, m is the sum of number of AR and MA parameters i.e. m = p + q.
The model with lower values of AIC and SBC were considered as better model.

Results and Discussion

Fitting of trends on cumin area in Banaskantha district

Fitting of polynomial models for area of cumin

The results of fitted polynomial models on cumin area in Banaskantha district of Gujarat are given in Table 1. The results indicated that in all polynomial models, the coefficient of determination (R^2) was found to be highly significant. All regression constants in linear, quadratic and cubic models were found highly significant. Linear regression coefficients were found to be highly significant except all quadratic models and original data approach of cubic model. Quadratic regression coefficients were found to be significant for quadratic model with 5 year moving average data approach and highly significant for cubic model with 3, 4 and 5years moving average data approach. All cubic regression coefficients were found to be highly significant except original data approach. In original data approach, the value of R^2 was improved by 0.019 percent in quadratic regression as compared to linear regression while, in case of cubic regression, it was improved by 0.032 percent in case of first, second and third degree polynomial models, respectively over original data. Thus, higher improvement in coefficient of determination (R^2) was observed due to moving average approach.

Table 1: Fitted polynomial models for cumin area in Banaskantha district of Gujarat

Model	Moving Average	Regression Constant	Regression Coefficients			R^2	Adj R^2	RMSE	MAE	S-W Statistics	Run Test |Z|
		a	b	c	d						
Linear	Original	177.494**	29.797**	-	-	0.732**	0.716	98.586	9773.298	0.942	1.408
	3 Years	204.595**	29.966**	-	-	0.768**	0.753	80.586	6494.047	0.927	1.997*
	4 Years	199.990**	31.853**	-	-	0.817**	0.804	69.581	4841.499	0.936	2.329*
	5 Years	199.109**	33.727**	-	-	0.860**	0.849	58.904	3469.625	0.942	2.136*
Quadratic	Original	247.852**	9.695	1.005	-	0.751**	0.720	95.142	9051.987	0.882	1.882
	3 Years	295.906**	1.131	1.602	-	0.810**	0.783	72.965	5323.878	0.930	2.499*
	4 Years	284.362**	3.729	1.654	-	0.854**	0.831	62.165	3864.427	0.946	2.329*
	5 Years	285.691**	3.168	1.910*	-	0.900**	0.884	49.663	2466.412	0.938	2.136*
Cubic	Original	370.613**	-55.725	8.977	-0.266	0.783**	0.739	88.948	7911.774	0.940	1.882*
	3 Years	491.671**	-113.352**	17.057**	-0.572**	0.906**	0.884	51.334	2635.184	0.963	1.494
	4 Years	468.482**	-109.454**	17.805**	-0.633**	0.940**	0.925	39.887	1590.958	0.954	1.294
	5 Years	446.507**	-101.021**	17.676**	-0.657**	0.967**	0.958	28.607	818.388	0.963	1.597

*, ** indicates significant at 5% and 1% level of significance, respectively

The cubic model with five year moving average data approach showed comparatively lower values of RMSE (28.607), MAE (818.388) and highest R^2 (0.967).
The criteria for testing normality (Shapiro-Wilk test) of residuals indicated that all polynomial models had normally distributed residuals. The test of randomness of residuals (Run test) indicated that original data approaches of linear and quadratic models and all moving averages of cubic model except original data approach were fulfilled the assumption of randomness of residuals. Thus, cubic model with five year moving average data approach was found to be best fitted.
Post Sampled period forecasts and validation.
It is observed that the percentage difference between predicted and actual production based on cubic model with five year moving average data approach was only 12.23 percent (Table 2). Hence, it was found the best fitted model for forecasting the area of cumin in Banaskantha district during the period under study *i.e.* 2019-20 to 2022-23.

Table 2: Post Sampled period forecasts using cubic model with five year moving average data approach for the cumin area in Banaskantha district

Year	Actual value (00'ha)	Cubic model with five year moving average		
		Predicted value	Deviation	Percent Deviation
2019-20	851.15	664.155	186.995	21.96
2020-21	772.08	609.673	162.407	21.03
2021-22	508.67	523.529	14.859	2.86
2022-23	373.52	401.781	28.261	7.56
Average	626.355	549.784	76.571	12.23

Fitting of ARIMA models

Fitting ARIMA model for area of cumin in Banaskantha:

To build ARIMA model, the preliminary step is to identify the parameter of the model and then estimation is done by least square method. At last diagnostic checking of the model is done.

Identification of parameter

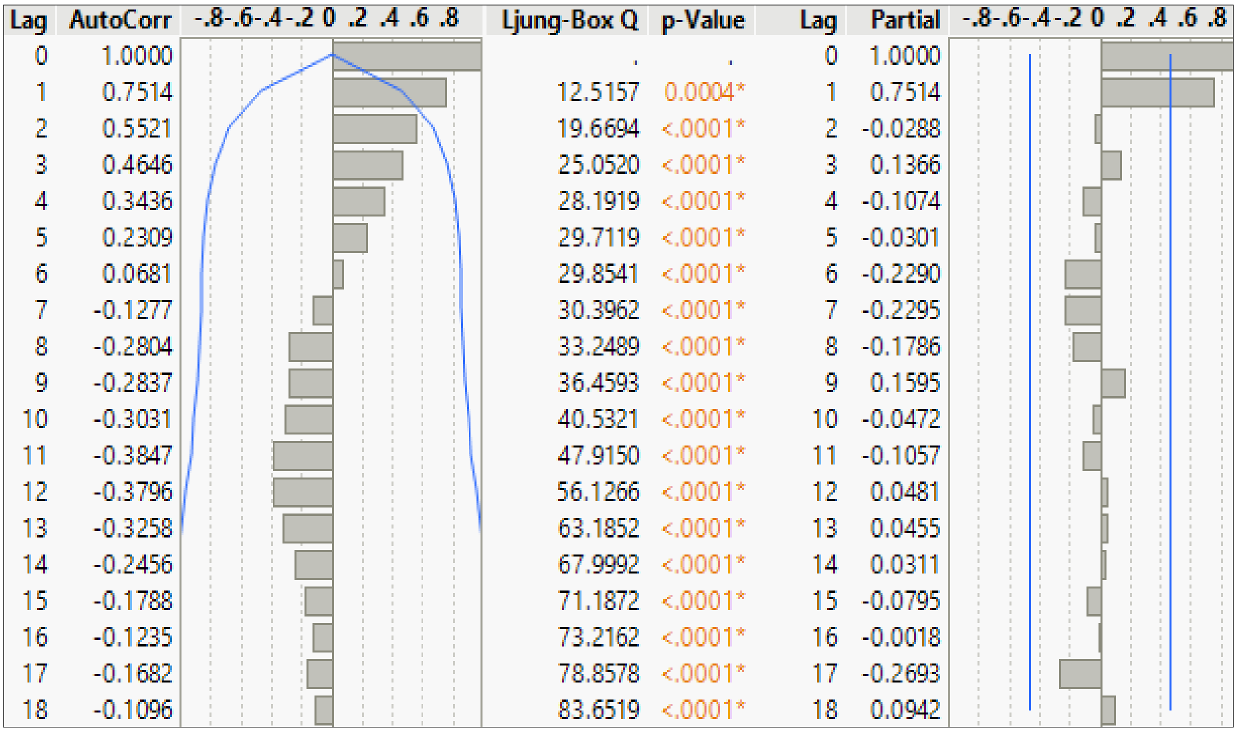
The ARIMA model includes three parameters *viz.*, p, d and q. Where 'p' is the number of Autoregressive (henceforth, AR) terms, 'd' is the number of times the series is to be differenced in order to make it stationary and 'q' is the number of moving average (henceforth, MA) terms. The parameter 'd' was found by performing unit root test, where area of cumin in Banaskantha district, the order of integration was found to be one, *i.e.*, I (1) or 'd = 1' as shown in table 3.

Table 3: Stationary test for area series in Banaskantha district

Data	Series	Statistic	Probability
Cumin Area	Original	-1.067	0.301
	1st difference	-4.192**	0.0008

Augmented Dickey Fuller (ADF) test was used to test the stationarity of time series. Results presented in Table 3 revealed the test result of cumin area, which showed that 1st differenced series was found stationary.

The parameters ‘p’ and ‘q’ are identified with the help of ACF and PACF of the differenced series because the differenced series is stationary. The ACF and PACF for area of cumin is given in Figure 4.1. As it could be seen from the figure, both ACF and PACF values were obtained for 18 lags length. In ACF and PACF, first lag length was found to cross the standard error limits. Thereby, both the parameters ‘p’ and ‘q’ were traced of first sacred orders. Accordingly, for area of cumin in Banaskantha district, the parameters ‘p’, ‘d’ and ‘q’ were identified as ARIMA (1, 1, 1) model, based on information given by ACF and PACF of level of differenced series.



Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	0.7514	12.5157	0.0004*	1	0.7514
2	0.5521	19.6694	<.0001*	2	-0.0288
3	0.4646	25.0520	<.0001*	3	0.1366
4	0.3436	28.1919	<.0001*	4	-0.1074
5	0.2309	29.7119	<.0001*	5	-0.0301
6	0.0681	29.8541	<.0001*	6	-0.2290
7	-0.1277	30.3962	<.0001*	7	-0.2295
8	-0.2804	33.2489	<.0001*	8	-0.1786
9	-0.2837	36.4593	<.0001*	9	0.1595
10	-0.3031	40.5321	<.0001*	10	-0.0472
11	-0.3847	47.9150	<.0001*	11	-0.1057
12	-0.3796	56.1266	<.0001*	12	0.0481
13	-0.3258	63.1852	<.0001*	13	0.0455
14	-0.2456	67.9992	<.0001*	14	0.0311
15	-0.1788	71.1872	<.0001*	15	-0.0795
16	-0.1235	73.2162	<.0001*	16	-0.0018
17	-0.1682	78.8578	<.0001*	17	-0.2693
18	-0.1096	83.6519	<.0001*	18	0.0942

Fig 1: ACF and PACF of cumin area in Banaskantha district

Estimation of parameter

Based upon the conditions, we have following expected ARIMA (p, d, q) models given in the Table 4. To select the best suitable model for forecast, Akaike Information Criteria (AIC), Schwarz Bayesian Criteria (SBC), Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE) and Adjusted R^2 were employed. The summary of each of the fitted ARIMA model in the time series for data of area of cumin area can be seen for further inference (Table 4). The estimation of parameter of various ARIMA models was carried out using JMP18 statistical software package. The model with least AIC, SBC, MAPE, and MAE with higher value of adjusted R^2 was proposed to be selected.

Table 4: Summary of the ARIMA model for cumin area in Banaskantha district

Model	AIC	SBC	MAPE	MAE	Adj. R^2
ARIMA (0, 1, 0)	218.389	219.279	15.751	67.826	0.715
ARIMA (1, 1, 1)	219.371	222.043	16.600	67.816	0.719
ARIMA (0, 1, 1)	220.275	222.055	15.729	68.848	0.699
ARIMA (1, 1, 0)	220.295	222.075	15.652	68.453	0.698
ARIMA (0, 1, 2)	220.566	223.237	19.523	77.439	0.692
ARIMA (1, 1, 2)	221.100	224.662	16.836	67.006	0.700
ARIMA (2, 1, 0)	222.193	224.864	15.889	68.555	0.680
ARIMA (2, 1, 1)	224.294	227.855	15.647	68.450	0.655
ARIMA (2, 1, 2)	224.844	229.296	17.045	70.419	0.641
ARIMA (1, 0, 0)	234.903	236.791	20.940	77.620	0.613
ARIMA (2, 0, 0)	236.825	239.659	21.402	78.521	0.591
ARIMA (1, 0, 1)	236.828	239.661	21.400	78.509	0.590
ARIMA (2, 0, 1)	237.873	241.651	22.147	80.426	0.580
ARIMA (1, 0, 2)	238.809	242.586	21.321	78.381	0.565
ARIMA (0, 0, 2)	239.193	242.026	28.105	98.901	0.542
ARIMA (2, 0, 2)	239.870	244.592	22.088	80.382	0.550
ARIMA (0, 0, 1)	244.255	246.144	30.595	115.926	0.444

It can be clearly observed from Table 4 that, lower AIC, SBC, MAPE and MAE with highest value of adjusted R^2 are for ARIMA (1, 1, 1).

Table 5: Estimates of the fitted ARIMA (1, 1, 1) model for area of cumin

Variables	Estimates	Standard Error	t ratio	Probability
AR1	0.57*	0.20	2.78	0.0140
MA1	0.99**	0.14	6.88	0.0001
Intercept	30.06**	7.02	4.28	0.0007

*, ** indicates significant at 5% and 1% level of significance, respectively

The coefficients of MA1 and intercept under this model were found to be highly significant and AR1 was found significant as shown in Table 5. Hence, the ARIMA (1, 1, 1) model was found to be best fitted for area of cumin in Banaskantha district.This model was yet to be confirmed by diagnostic checking of residuals.

Diagnostic checking

The residual correlogram of the fitted ARIMA (1, 1, 1) model for cumin area in Banaskantha district is depicted in the Figure 4.2. It resembles like a white noise series, which indicate that there was no serial correlation in the residuals because all the lags were fitted within the standard error limits. The Q-statistic accepts the null hypothesis of no serial correlation in the residuals which satisfy one of the assumptions of ARIMA model (independence of errors). Thus, the fitted ARIMA (1,1, 1) model can adequately describe the dynamic linear dependence of area of cumin in Banaskantha district.

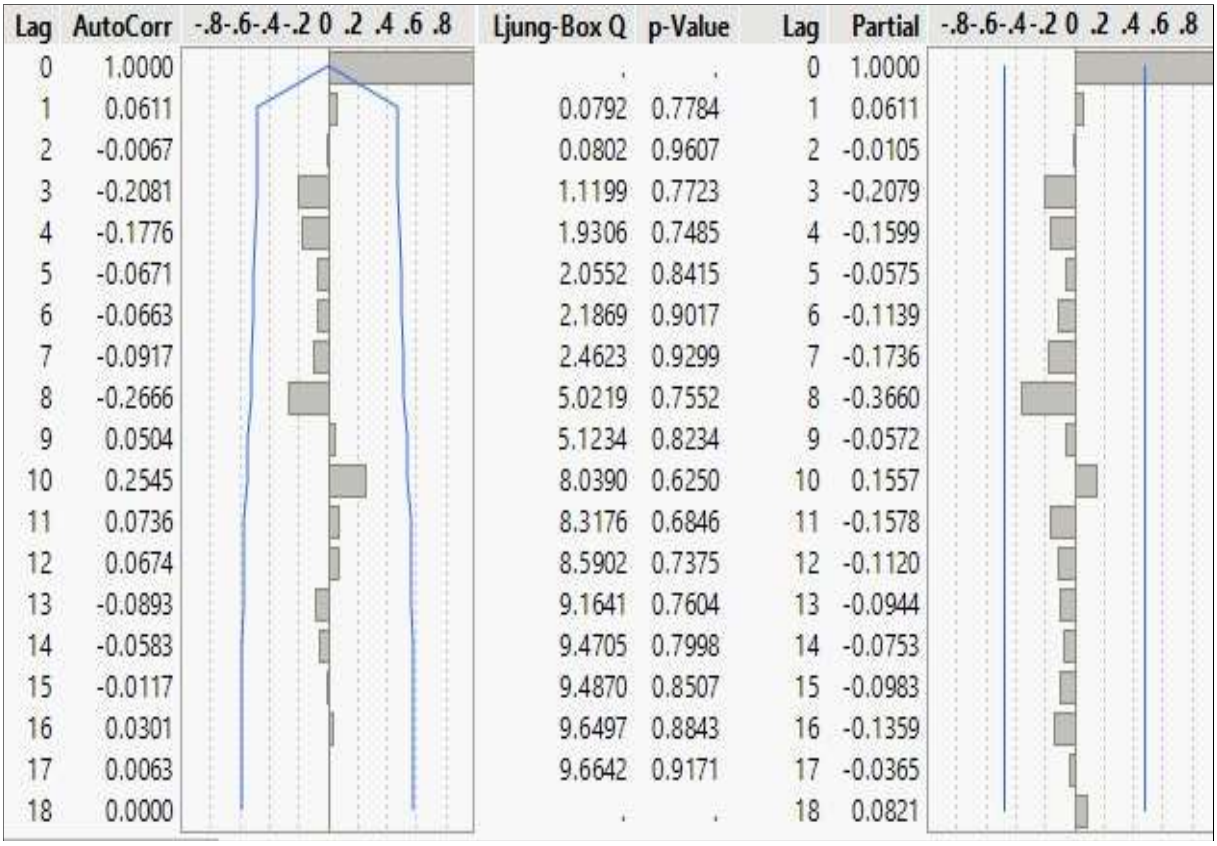


Fig2: Residual correlogram of cumin area in Banaskantha district

The Ljung-Box Q test: The adequacy of the model was checked by using Ljung-Box test. It was done to check the independence of the residuals *i.e.,* whether they are distributed randomly or not. The hypothesis was stated as below:

H_0: The errors are distributed randomly

H_a: The errors are non-random

The obtained Q value was 9.66 (p=0.917), which is non-significant. So, the errors were found independent. In other words, the estimated Q-statistics value was found to be insignificant revealing the overall fitness of the model for future forecasts.

Table 6: Post sampled period forecasts using ARIMA (1, 1, 1) model for the cumin area in Banaskantha district

Year	Actual value (00'ha)	ARIMA (1, 1, 1)		
		Predicted value	Deviation	Percent Deviation
2019-20	851.15	782.681	68.469	8.04
2020-21	772.08	810.286	38.206	4.94
2021-22	508.67	838.927	330.257	64.92
2022-23	373.52	868.169	494.649	132.43
Average	626.355	825.015	232.895	37.18

Post Sampled period forecasts and validation

In order to check the validity of these predicted values, they were compared with the actual values of area of cumin during post sampled period forecast *i.e.* for the period of 2019-20 to 2022-23.

It is observed that the average percentage difference between predicted and actual values based on ARIMA (1, 1, 1) model was 37.18 percent. Due to higher percent deviation, ARIMA (1, 1, 1) model was not selected to forecast the future trends of cumin area in Banaskantha district.

Fitting of trends on cumin production in Banaskantha district

Fitted polynomial models for production of cumin

The results of fitted polynomial models for production of cumin in Banaskantha district are given in Table 7. The results indicated that in all polynomial models, the coefficient of determination (R^2) was found to be highly significant. All regression constants in linear, quadratic and cubic models were found to be non-

Table 7: Fitted polynomial models for production of cumin in Banaskantha district of Gujarat

Model	Moving Average	Regression Constant	Regression Coefficients			R^2	Adj R^2	RMSE	MAE	S-W Statistics	Run Test |Z|
		a	b	c	d						
Linear	Original	-29.296	43.614**	-	-	0.909**	0.904	75.525	5704.084	0.959	2.829**
	3 Years	-6.252	45.753**	-	-	0.929**	0.924	61.894	3830.868	0.962	2.499*
	4 Years	5.364	47.084**	-	-	0.944**	0.940	52.905	2798.955	0.966	2.329*
	5 Years	16.897	46.608**	-	-	0.960**	0.957	45.596	2171.168	0.963	2.136*
Quadratic	Original	-5.901	36.929*	0.334	-	0.910**	0.899	74.996	5624.327	0.949	2.829**
	3 Years	16.286	38.636*	0.395	-	0.930**	0.921	61.316	3759.593	0.947	2.499**
	4 Years	32.587	38.010*	0.534	-	0.946**	0.938	51.935	2697.212	0.950	2.329*
	5 Years	51.450	36.413*	0.762	-	0.963**	0.957	40.995	1680.584	0.970	2.136*
Cubic	Original	155.889	-49.289	10.840*	-350*	0.942**	0.930	60.359	3643.208	0.965	0.935*
	3 Years	182.855**	-58.773	13.546**	-0.487**	0.966**	0.959	42.574	1812.584	0.969	1.494*
	4 Years	191.074**	-59.416*	14.436**	-0.545**	0.980**	0.975	31.817	1012.340	0.957	1.811
	5 Years	188.749**	-52.539*	14.223**	-0.561**	0.990**	0.987	21.893	479.305	0.943	1.597

*, ** indicates significant at 5% and 1% level of significance, respectively

significant except cubic model with 3, 4 and 5 years moving average data approaches.

Linear regression coefficients were found to be highly significant in linear model and significant in quadratic model and cubic model with 4 and 5years moving average data approach. Cubic regression coefficients were found to be significant for original data approach and highly significant for all other moving average data approaches. In original data approach, the value of R^2 was slightly improved by 0.001 percent in quadratic regression as compared to linear regression while, in case of cubic regression, it was improved by only 0.032 percent over quadratic model.

By taking moving averages of five years, the improvement in coefficient of determination (R^2) was observed very low *viz.* 0.051, 0.053 and 0.048 percent in case of first, second and third degree polynomial models, respectively.

The cubic model with five year moving average data approach showed comparatively lower values of RMSE (21.893), MAE (479.305) highest coefficient of determination (0.990).The criteria for testing normality (Shapiro-Wilk test) of residuals indicated that all polynomial models had normally distributed residuals. The test of randomness of residuals (Run test) indicated that residuals of all polynomial models were not fulfilled the assumption of randomness of residuals except cubic model with 4 and 5 years moving average data approaches. Thus, cubic model with five year moving average data approach was found to be best fitted.

Post sampled period forecasts and validation

Table 8: Post sampled period forecasts using cubic model with five year moving average data approach for production of in Banaskantha district

Year	Actual value (00' MT)	Cubic model with five year moving average		
		Predicted value	Deviation	Percent deviation
2019-20	902.22	691.35	210.87	23.37
2020-21	779.80	649.84	129.96	16.66
2021-22	513.76	579.54	65.78	12.80
2022-23	380.99	477.11	96.12	25.22
Average	644.19	599.46	44.73	6.94

It is observed that the percentage difference between predicted and actual production based on cubic model with five year moving average data approach was 6.94 percent (Table 8). Hence, it was found the best fit model for forecasting the area of cumin in Banaskantha district during the period under study *i.e.* 2019-20 to 2022-23.

Fitting of ARIMA models

Fitting ARIMA model for production of cumin:

To build ARIMA model, the preliminary step is to identify the parameter of the model and then estimation is doneby least square method or conditional or exact likelihood method. At last diagnostic checking is done.

Identification of parameter

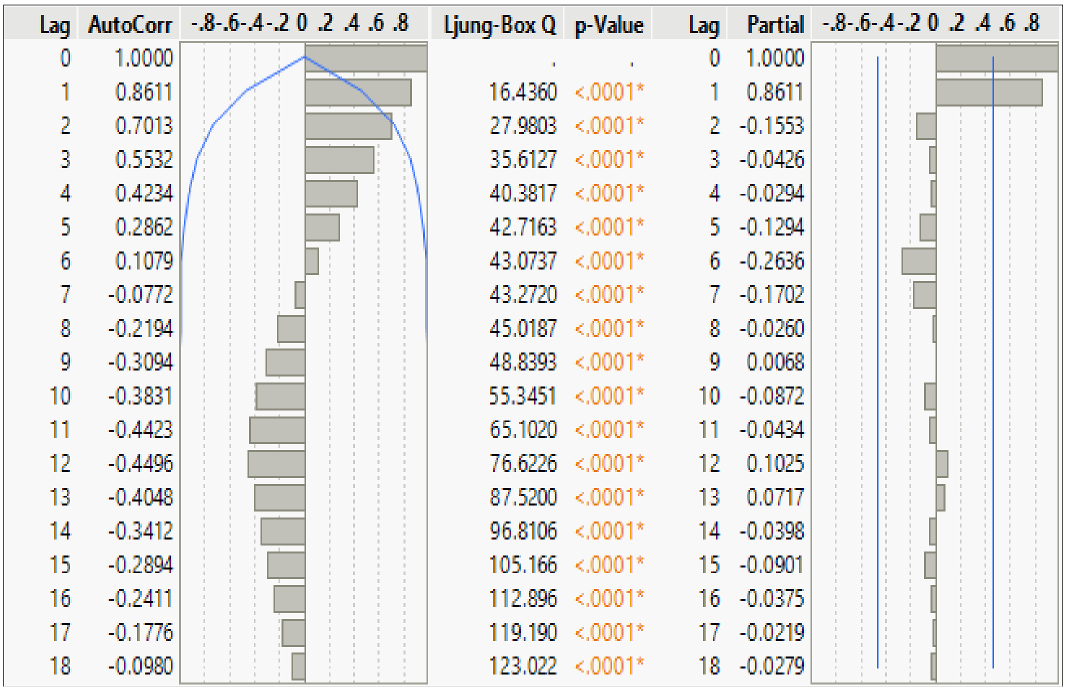
The ARIMA model includes three parameters *viz.*, p, d and q. Where 'p' is the number of Autoregressive (henceforth, AR) terms, 'd' is the number of times the series is to be differenced in order to make it stationary and 'q' is the number of moving average (henceforth, MA) terms. The parameter 'd' was found by performingunit root test, where for production of cumin in Banaskantha district, the order of integration was found to be two, *i.e.* I (2) or 'd = 2' as shown in Table9.

Table 9: Stationary test for production of cumin in Banaskantha district

Data	Series	Statistic	Probability
Cumin Production	Original	-0.35	0.89
	1st differenced	-2.78	0.08
	2nd differenced	-4.09**	0.007

Augmented Dickey Fuller (ADF) test was used to test the stationarity of time series. Results presented in Table 9 revealed the test result of production of cumin in Banaskantha district, which showed that 2nd differenced series was found stationary.

The parameter 'p' and 'q' is identified with the help of ACF and PACF of the differenced series because the differenced series is stationary. The ACF and PACF for production of cumin is given in Figure 4.3.As it could be seen from the figure, both ACF and PACF values were obtained for 18 lags length. In ACF and PACF, second and first leg length coefficients were found to cross the standard error limits, respectively. Thereby, both the parameters 'p' and 'q' were traced of second and first sacred orders, respectively. Accordingly, for production of cumin in Banaskantha district, the parameters 'p', 'd' and 'q' were identified as ARIMA (2, 2, 1) model, based on information given by ACF and PACF of the level differenced series.



Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	0.8611	16.4360	<.0001*	1	0.8611
2	0.7013	27.9803	<.0001*	2	-0.1553
3	0.5532	35.6127	<.0001*	3	-0.0426
4	0.4234	40.3817	<.0001*	4	-0.0294
5	0.2862	42.7163	<.0001*	5	-0.1294
6	0.1079	43.0737	<.0001*	6	-0.2636
7	-0.0772	43.2720	<.0001*	7	-0.1702
8	-0.2194	45.0187	<.0001*	8	-0.0260
9	-0.3094	48.8393	<.0001*	9	0.0068
10	-0.3831	55.3451	<.0001*	10	-0.0872
11	-0.4423	65.1020	<.0001*	11	-0.0434
12	-0.4496	76.6226	<.0001*	12	0.1025
13	-0.4048	87.5200	<.0001*	13	0.0717
14	-0.3412	96.8106	<.0001*	14	-0.0398
15	-0.2894	105.166	<.0001*	15	-0.0901
16	-0.2411	112.896	<.0001*	16	-0.0375
17	-0.1776	119.190	<.0001*	17	-0.0219
18	-0.0980	123.022	<.0001*	18	-0.0279

Fig3: ACF and PACF of production of cumin in Banaskantha district

Estimation of parameter

Table 10: Summary of the ARIMA model for production of cumin in Banaskantha district

Model	AIC	SBC	MAPE	MAE	Adj. R^2
ARIMA (0, 2, 2)	193.367	195.867	11.127	40.959	0.931
ARIMA (1, 2, 2)	194.300	197.633	9.934	36.713	0.931
ARIMA (1, 2, 1)	196.225	198.725	10.945	41.601	0.917
ARIMA (2, 2, 1)	196.342	199.675	10.454	37.340	0.922
ARIMA (0, 2, 1)	196.731	198.398	11.083	45.721	0.914
ARIMA (0, 2, 0)	197.045	197.878	13.811	52.623	0.901
ARIMA (2, 2, 2)	197.503	201.669	10.906	39.638	0.919
ARIMA (1, 2,0)	198.916	200.583	13.470	51.417	0.895
ARIMA (1, 1, 2)	199.035	202.597	11.761	33.243	0.963
ARIMA (2, 2, 0)	199.178	201.678	12.701	50.924	0.899
ARIMA (0, 1, 1)	199.621	201.402	11.125	37.056	0.948
ARIMA (0, 1, 2)	199.721	202.392	10.589	33.562	0.949
ARIMA (1, 1, 1)	200.032	202.703	10.347	33.588	0.948
ARIMA (2, 1, 1)	200.564	204.126	12.855	37.797	0.949
ARIMA (2, 1, 2)	201.029	205.480	11.716	33.177	0.950
ARIMA (2, 1, 0)	201.609	204.280	11.103	35.536	0.943
ARIMA (0, 1, 0)	202.024	202.915	11.722	42.083	0.934
ARIMA (1, 1, 0)	202.175	203.956	11.195	39.129	0.937
ARIMA (1, 0, 1)	217.947	220.781	32.639	51.511	0.830
ARIMA (2, 0, 1)	219.729	223.506	32.051	49.823	0.820
ARIMA (1, 0, 2)	219.789	223.567	32.210	50.228	0.820

Based upon the conditions, we have above expected ARIMA (p, d, q) models given in the Table 10. To select the best suitable model for forecast Akaike Information Criteria (AIC), Schwarz Bayesian Criteria (SBC), Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE) and Adjusted R^2 were employed. The summary of each of the fitted ARIMA model using timeseries data of production of cumin can be seen for further inference (Table 10). The estimation of parameter of various ARIMA models were carried out using JMP18 statistical software package. The model with least AIC, SBC, MAPE, and MAE with higher value of adjusted R^2 was proposed to be selected.

It can be clearly observed from Table 10 that, the lower AIC, SBC, MAPE and MAE and highest value of adjusted R^2 are for ARIMA (1, 1, 2).

Table 11: Estimates of the fitted ARIMA (1, 1, 2) model for production of cumin

Variables	Estimates	Standard Error	t ratio	Probability
AR1	0.43	0.23	1.87	0.08
MA1	0.00	-	-	-
MA2	1.00**	0.002	420.08	<0.0001
Intercept	41.92	-	-	-

*, ** indicates significant at 5% and 1% level of significance, respectively

The parameters for this model were estimated and depicted in Table 11. The coefficients of MA2 and intercept under this model were found to be highly significant at 1 percent level. Hence, the ARIMA (1, 1, 2) model was found to be best fitted for production of cumin in Banaskantha district. This model was yet to be confirmed by diagnostic checking of residuals.

Diagnostic checking

The residual correlogram of the fitted ARIMA (1, 1, 2) model for cumin production in Banaskantha district is depicted in the Figure 4. It resembles like a white noise series, which indicates that there was no serial correlation in the residuals because all the lags were fitted within the standard error limits. The Q-statistic accepts the null hypothesis of no serial correlation in the residuals which satisfy one of the assumptions of ARIMA model (independence of errors). Thus, the fitted ARIMA (1, 1, 2) model can adequately describe the dynamic linear dependence of production of cumin in Banaskantha district.

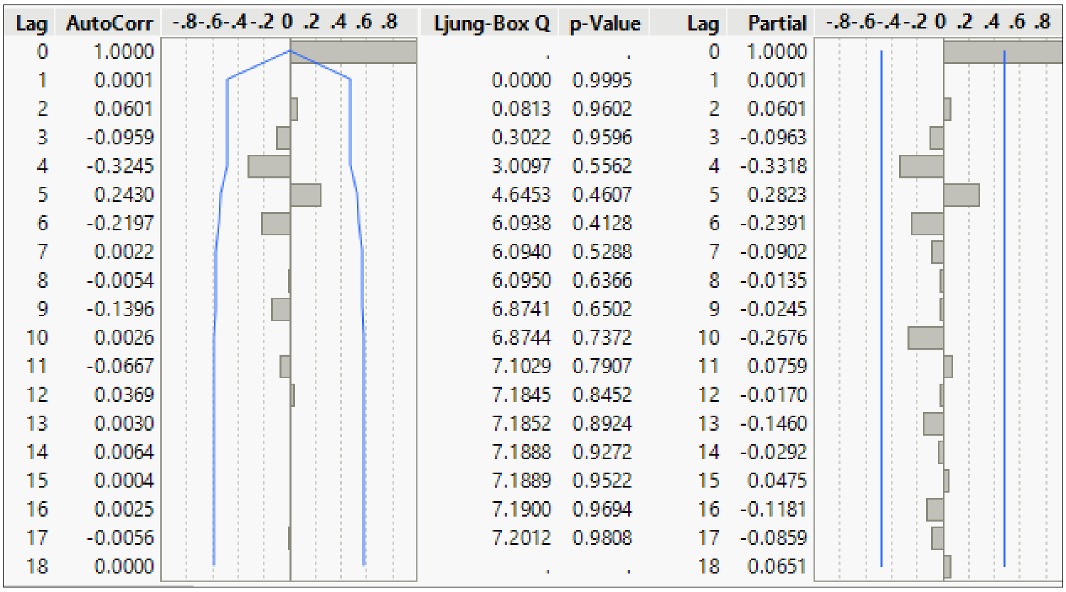


Fig 4: Residual correlogram of cumin production in Banaskantha district

The Ljung-Box Q test:

The adequacy of the model was checked by using Ljung-Box test. It was done to check the independence of the residuals *i.e.*, whether they are distributed randomly or not. The hypothesis was stated as below:

H_0: The errors are distributed randomly

H_a: The errors are non-random

The obtained Q value was 7.20(p=0.9808) which is non-significant. So, the errors were found independent. In other words, the estimated Q-statistics value was found to be insignificant revealing the overall fitness of the model for future forecasts.

Post sampled period forecasts and validation

Table 12: Post sampled period forecasts using ARIMA (1, 1, 2) for production of cumin in Banaskantha district

Year	Actual value (00'MT)	ARIMA (1, 1, 2)		
		Predicted value	Deviation	Percent Deviation
2019-20	902.22	827.69	74.53	8.26
2020-21	779.80	871.11	91.31	11.71
2021-22	513.76	913.67	399.91	77.84
2022-23	380.99	955.87	574.88	150.89
Average	644.19	892.09	246.90	38.33

In order to check the validity of these predicted values, they were compared with the actual values of production of cumin during post sampled period forecast *i.e.* for the period of 2019-20 to 2022-23.

It is observed that the average percentage difference between predicted and actual values based on ARIMA (1, 1, 2) model was 38.33 percent. Due to higher percent deviation, ARIMA (1, 1, 2) model was not selected to study the future trends on production of cumin in Banaskantha district.

Fitting of trends on cumin productivity in Banaskantha district

Fitting of polynomial models for productivity of cumin

The results of fitted polynomial models on cumin productivity in Banaskantha district of Gujarat are given in Table 13. The results indicated that in all polynomial models, the coefficient of determination (R^2) was found to be highly significant. Regression constants in linear, quadratic and cubic models were found to be highly significant and for original data approach of quadratic and cubic models it was found to be significant.

Linear regression coefficients in all data approaches were found to be highly significant in all polynomial models except original data approach of cubic model. Quadratic regression coefficients were found to be highly significant for all data approaches of quadratic model and cubic model with 3, 4 and 5 years moving average data approaches was found to be significant. Cubic regression coefficients were found to be non-significant for cubic model with all data approaches.

In original data approach, the value of R^2 improved by 0.094 percent in quadratic regression as compared to linear regression while, in case of cubic regression it was improved by only 0.003 percent over quadratic model. By taking moving average of five years, the improvement in coefficient of determination (R^2) was observed to be 0.062, 0.066 and 0.064 percent in case of first, second and third degree models, respectively over original data approach.

The quadratic with five year moving average data approach showed comparatively lower values of RMSE (0.027), MAE (0.001) and higher value of R^2 (0.989).

The criteria for testing normality (Shapiro-Wilk test) of residuals indicated that all models were normally distributed residuals except 5 years moving average data approach of cubic model. The test of randomness of residuals (Run test) indicated that residuals of all polynomial models were fulfilled the assumption of randomness of residuals. Thus, quadratic model with five year moving average data approach was found to be best fitted.

Table 13: Fitted polynomial models for productivity of cumin in Banaskantha district of Gujarat.

Model	Moving Average	Regression Constant	Regression Coefficients			R^2	Adj R^2	RMSE	MAE	S-W Statistics	Run Test |Z|
		a	b	c	d						
Linear	Original	0.361**	0.043**	-	-	0.829**	0.818	0.106	0.011	0.951	1.408
	3 Years	0.395**	0.045**	-	-	0.880**	0.872	0.082	0.007	0.953	3.002
	4 Years	0.429**	0.044**	-	-	0.887**	0.879	0.073	0.005	0.933	2.847
	5 Years	0.455**	0.045**	-	-	0.891**	0.883	0.066	0.004	0.936	2.674
Quadratic	Original	0.154*	0.102**	-0.003**	-	0.923**	0.913	0.071	0.005	0.909	0.000
	3 Years	0.192**	0.109**	-0.004**	-	0.987**	0.985	0.066	0.004	0.962	2.499
	4 Years	0.245**	0.106**	-0.004**	-	0.985**	0.983	0.053	0.003	0.944	1.811
	5 Years	0.281**	0.106**	-0.004**	-	0.989**	0.987	0.027	0.001	0.953	0.521
Cubic	Original	0.210*	0.072	0.001	0.000	0.926**	0.911	0.404	0.164	0.946	3.776
	3 Years	0.171**	0.122**	-0.005*	0.000062	0.987**	0.984	0.217	0.047	0.914	2.499
	4 Years	0.219**	0.122**	-0.006*	0.000089	0.986**	0.983	0.319	0.102	0.898	3.364
	5 Years	0.257**	0.121**	-0.006*	0.000097	0.990**	0.987	0.263	0.069	0.867*	3.213

*, ** indicates significant at 5% and 1% level of significance, respectively

Post sampled period forecasts and validation

In order to check the validity of these predicted values, they were compared with the actual values of productivity of cumin during post sampled period forecast *i.e.* for the period of 2019-20 to 2022-23.

Table 14: Post sampled period forecasts using quadratic model with five year moving average data approach for productivity of cumin in Banaskantha district

Year	Actual value (00' MT/ha)	Quadratic model		
		Predicted value	Deviation	Percent deviation
2019-20	1.06	0.95	0.11	10.37
2020-21	1.01	0.92	0.09	8.91
2021-22	1.01	0.89	0.12	11.88
2022-23	1.02	0.85	0.17	16.66
Average	1.02	0.90	0.12	11.76

It is observed that the percentage difference between predicted and actual production based on quadratic model with five year moving average data approach was 11.76 percent (Table 14). Hence, it was found the best fit model for forecasting the productivity of cumin in Banaskantha district during the period under study.

Fitting of ARIMA models

Fitting ARIMA model for productivity of cumin

To build ARIMA model, the preliminary step is to identify the parameter of the model and then estimation is done by least squares method or conditional or exact-likelihood method. At last diagnostic checking is done.

Identification of parameter

The ARIMA model includes three parameters *viz.*, p, d and q. Where 'p' is the number of Autoregressive (henceforth, AR) terms, 'd' is the number of times the series is to be differenced in order to make it stationary and 'q' is the number of moving average (henceforth, MA) terms. The parameter 'd' was found by performing unit root test. For productivity of cumin in Banaskantha district, the order of integration was found to be two, *i.e.* I (2) or 'd = 2' as shown in Table 15.

Table 15: Stationary test for productivity of cumin in Banaskantha district

Data	Series	Statistic	Probability
Cumin Productivity	Original	-2.23	0.20
	1st differenced	-0.72	0.81
	2nd differenced	-13.79**	0.00

Augmented Dickey Fuller (ADF) test was used to test the stationarity of time series. Results presented in Table 15 revealed the test result of cumin productivity, which showed that 2nd differenced series was found stationary.

The parameter 'p' and 'q' is identified with the help of ACF and PACF of the 2nd differenced series because the 2nd differenced series is stationary. The ACF and PACF for productivity of cumin is given in Figure 4.5.

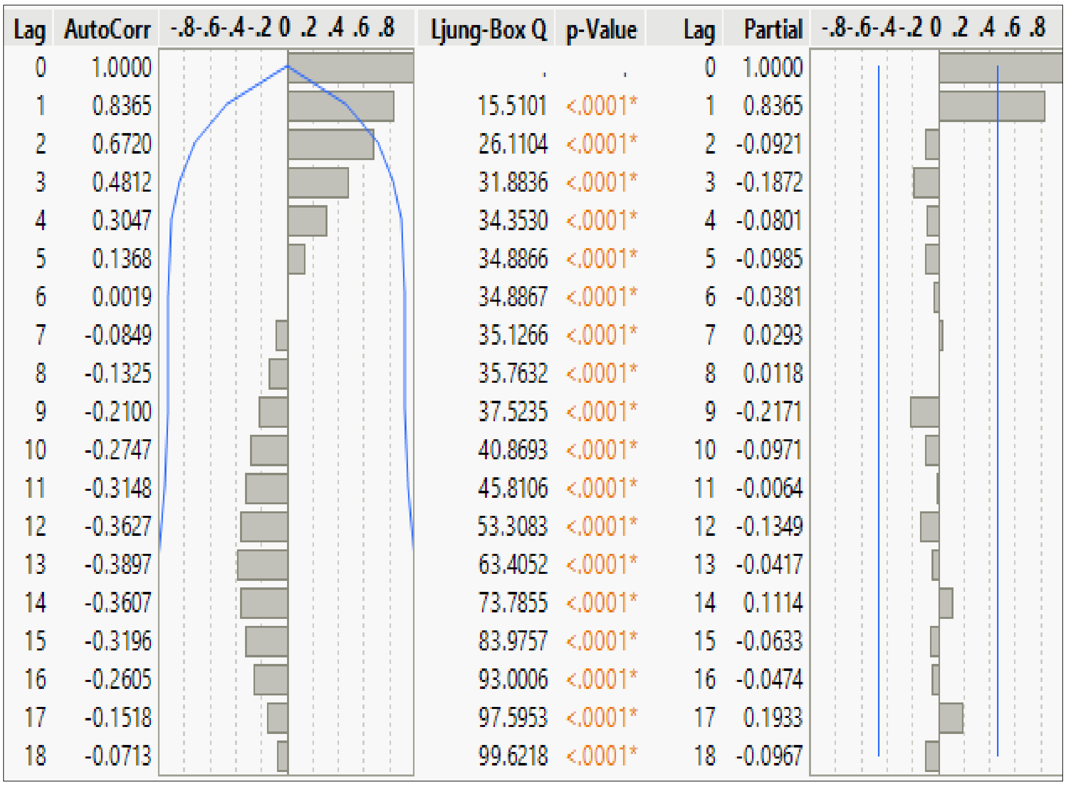


Fig5: ACF and PACF of productivity of cumin in Banaskantha district

As it could be seen from the figure, both ACF and PACF values were obtained for 18 lags length. In ACF and PACF, second and first leg length were found to cross the standard error limits, respectively.
Thereby, both the parameters 'p' and 'q' were traced of second and first sacred orders, respectively. Accordingly, for productivity of cumin in Banaskantha district, the parameters 'p', 'd' and 'q' were identified as ARIMA (2, 2, 1) model, based on information given by ACF and PACF of the 2nd differenced series.

Estimation of parameter

Based upon the conditions, we have above expected ARIMA (p, d, q) models given in the Table 16. To select the best suitable model for forecast Akaike Information Criteria (AIC), Schwarz Bayesian Criteria (SBC), Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE) and Adjusted R^2 were employed. The summary of each of the fitted ARIMA model using timeseries data of production of cumin can be seen for further inference (Table 16). The estimation of parameter of various ARIMA models were carried out using JMP18 statistical software package.

Table 16: Summary of the ARIMA model for productivity of cumin in Banaskantha district

Model	AIC	SBC	MAPE	MAE	Adj. R^2
ARIMA (2, 1, 2)	-31.495	-27.043	11.945	0.063	0.852
ARIMA (2, 1, 1)	-31.128	-27.567	12.334	0.067	0.853
ARIMA (2, 2, 0)	-31.067	-28.567	11.742	0.071	0.705
ARIMA (2, 1, 0)	-30.185	-27.514	12.469	0.069	0.847
ARIMA (0, 1, 0)	-30.061	-29.171	12.357	0.067	0.835
ARIMA (0, 1, 1)	-29.589	-27.808	12.700	0.069	0.839
ARIMA (2, 2, 2)	-29.391	-25.225	11.170	0.066	0.656
ARIMA (2, 2, 1)	-29.131	-25.798	11.577	0.070	0.682
ARIMA (1, 1, 0)	-28.927	-27.146	12.669	0.069	0.833
ARIMA (0, 1, 2)	-28.307	-25.636	12.775	0.070	0.832
ARIMA (1, 1, 1)	-27.63	-24.962	12.594	0.069	0.829
ARIMA (1, 1, 2)	-26.935	-23.374	13.716	0.074	0.805
ARIMA (1, 0, 0)	-26.211	-24.322	16.157	0.083	0.737
ARIMA (2, 0, 2)	-26.039	-21.317	17.954	0.088	0.681

ARIMA (2, 0, 1)	-24.683	-20.905	16.605	0.080	0.709
ARIMA (0, 2, 2)	-24.616	-22.116	11.453	0.069	0.677
ARIMA (1, 0, 1)	-24.224	-21.390	16.133	0.083	0.722
ARIMA (2, 0, 0)	-24.223	-21.390	16.136	0.083	0.722
ARIMA (0, 2, 1)	-22.981	-21.315	11.206	0.067	0.705
ARIMA (1, 2, 2)	-22.813	-19.480	11.073	0.067	0.662
ARIMA (1, 2, 1)	-21.851	-19.352	11.268	0.068	0.693
ARIMA (1, 2, 0)	-15.091	-13.425	14.596	0.089	0.602
ARIMA (0, 0, 2)	-13.655	-10.822	25.182	0.124	0.564
ARIMA (0, 2, 0)	-12.436	-11.603	16.627	0.104	0.504
ARIMA (1, 0, 2)	-11.384	-7.607	25.855	0.125	0.502
ARIMA (0, 0, 1)	-9.958	-8.069	27.898	0.140	0.487

The model with least AIC, SBC, MAPE, and MAE with largest value of adjusted R^2 was proposed to be selected.
It can be clearly observed from Table 16 that, the lowest AIC, SBC, MAPE and MAE and higher value of adjusted R^2 are for ARIMA (2, 1, 2).

Table 17: Estimates of the fitted ARIMA (2, 1, 2) model for productivity of cumin in Banaskantha district

Variables	Estimates	Standard Error	t ratio	Probability
AR1	-0.95**	0.17	-5.33	0.0001
AR2	-0.90**	0.10	-8.34	0.0001
MA1	-0.91*	0.40	-2.24	0.0428
MA2	-0.43	0.27	-1.59	0.1364
Intercept	0.03*	0.01	2.82	0.0144

*, ** indicates significant at 5% and 1% level of significance, respectively

The parameters for this model were estimated and depicted in Table 17. The coefficients of AR1, AR2, MA1 and intercept under this model were found to be statistically significant. Hence, the ARIMA (2, 1, 2) model was found to be best fitted for productivity of cumin in Banaskantha district.
This model was yet to be confirmed by diagnostic checking of residuals.

Diagnostic checking

The residual correlogram of the fitted ARIMA (2, 1, 2) model for cumin productivity in Banaskantha district is depicted in the Figure 4.6. It resembles like a white noise series, which indicates that there was no serial correlation in the residuals because all the lags were within the standard error limits. The Q-statistic accepts the null hypothesis of no serial correlation in the residuals, which satisfy one of the assumptions of ARIMA model (independence of errors). Thus, the fitted ARIMA (2, 1, 2) model can adequately describe the dynamic linear dependence of productivity of cumin series in Banaskantha district.

The Ljung-Box Q test

The adequacy of the model was checked by using Ljung-Box test. It was done to check the independence of the residuals *i.e.*, whether they are distributed randomly or not. The hypothesis was stated as below:
H_0: The errors are distributed randomly
H_a: The errors are non-random

The obtained Q value was 11.77(p=0.813) which is non-significant. So, the errors were found independent. In other words, the estimated Q-statistics value was found to be revealing the overall fitness of the model for future forecasts.

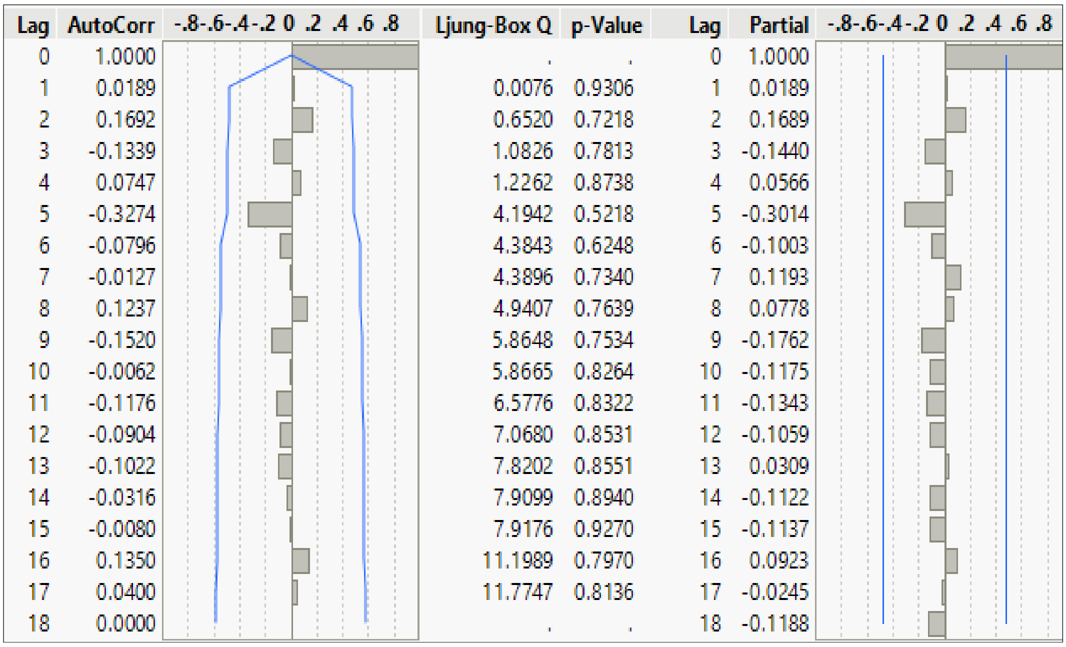


Fig6: Residual correlogram of cumin productivity in Banaskantha district

Post sampled period forecasts and validation

In order to check the validity of these predicted values, they were compared with the actual values of productivity of cumin during post sampled period forecast *i.e.* for the period of 2019-20 to 2022-23. It is observed that the average percentage difference between predicted and actual values based on ARIMA (2, 1, 2) model was 8.82 percent. Due to lower percent deviation, ARIMA (2, 1, 2) model was selected to study the future trends on productivity of cumin in Banaskantha district.

Table 18: Post sampled period forecasts using ARIMA (2, 1, 2) for productivity of cuminin Banaskantha district

Year	Actual value (00'MT/ha)	ARIMA (2, 1, 2)		
		Predicted value	Deviation	Percent Deviation
2019-20	1.06	1.05	0.01	0.94
2020-21	1.01	1.10	0.09	8.91
2021-22	1.01	1.13	0.12	11.88
2022-23	1.02	1.17	0.15	14.70
Average	1.02	1.11	0.09	8.82

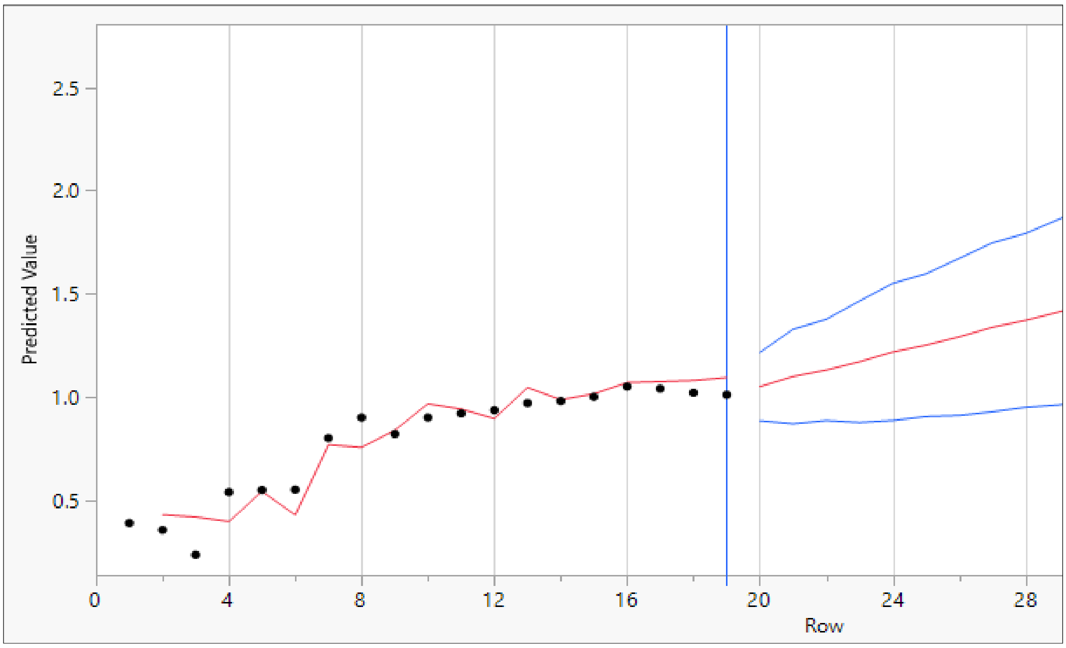


Fig 7: Forecasted productivity of cumin in Banaskantha district of Gujarat.

Conclusion

For area of cumin

The cubic model with five year moving average approach was satisfied all the criteria for selection of model. The selected model is,

$\hat{Y} = 446.507^{**} - 101.021^{**}t + 17.676^{**}t^2 - 0.657^{*}t^3$ (Adj. R^2= 95.80%)

The selected ARIMA (1, 1, 1) model was not found to be suitable to explain the trends on area of cumin in Banaskantha district due to higher average percent deviation in sampled forecasts validation.

For Production of cumin

The cubic model with five year moving average approach was satisfied all the criteria for selection of model. The selected model is,

$\hat{Y} = 188.749^{**} - 52.539^{*}t + 14.223^{**}t^2 - 0.561^{**}t^3$ (Adj. R^2= 98.70%)

The selected ARIMA (2, 1, 2) model was not found to be suitable to explain the trends on production of cumin in Banaskantha district due to higher average percent deviation in sampled forecasts validation.

For Productivity of cumin: The quadratic model with five year moving average approach was satisfied all the criteria for selection of model. The selected model is:

$\hat{Y} = 0.281^{**} + 0.106^{**}t - 0.004^{**}t^2$ (Adj. R^2= 98.70%)

The coefficients (AR1, AR2 and MA1) and intercept under ARIMA (2, 1, 2) model were found to be statistically significant. Hence, the ARIMA (2, 1, 2) model was found to be best fitted. The percentage difference between forecasted and actual productivity based on ARIMA (2, 1, 2) model was very low (8.32). Hence, it was found the best fitted for forecasting the productivity of cumin in Banaskantha district during the period under study.

References

1. Anonymous. Agricultural statistics at a glance-2022. New Delhi: Ministry of Agriculture and Farmers Welfare, Directorate of Economics & Statistics, Krishi Bhawan, Government of India; 2022.
2. Box GE, Jenkins GM, Reinsel GC. Autocorrelation function and spectrum of stationary processes and analysis of seasonal time series. In: Time series analysis: forecasting and control. 2nd ed. San Francisco: Holden-Day; 1976. p. 21-43.
3. Brockwell PJ, Davis RA. Stationary time series. In: Time series: theory and methods. 1st ed. 1991. p. 1-41.
4. Dhekale BS, Mahdi SS, Garde YA, Sahu PK. Parametric and nonparametric regression models for area, production and productivity trends of rice crop. Int J Agric Stat Sci. 2014;10(1):211-216.
5. Guha B, Bandyopadhyay G. Gold price forecasting using ARIMA model. J Adv Manag Sci. 2016;4(2):117-121.
6. Gupta GS. ARIMA model for forecasting tea production in India. Indian Econ J. 1993;41(2):88-110.

7. Harris RI. Testing for unit roots using the augmented Dickey-Fuller test: some issues relating to the size, power and the lag structure of the test. Econ Lett. 1992;38(4):381-386.
8. Hemavathi M, Prabakaran K. ARIMA model for forecasting of area, production and productivity of rice and its growth status in Thanjavur district of Tamil Nadu, India. Int J Curr Microbiol App Sci. 2018;7(2):149-156.
9. Kumari P, Parmar DJ, Kumar MS, Mahera AB, Lad YA. Prediction of area, production and productivity of total fruit crops in Gujarat. The Pharma Innovation J. 2022;11(5):750-754.
10. Kumari P, Parmar DJ, Kumar MS, Mahera A, Lad Y. Comparison of statistical models for prediction area, production and yield of Citrus in Gujarat. Biol Forum An Int J. 2022;14(2):540-545.
11. Ljung GM, Box GE. On a measure of lack of fit in time series models. Biometrika. 1978;65(2):297-303.
12. Makridakis S, Wheelwright SC, Hyndman RJ. Forecasting methods and applications. 3rd ed. New York: John Wiley & Sons; 2003. p. 312-372.
13. Mishra P, Padmanaban K, Dhekale BS, Tailor AK. Statistical investigation of production performance of cumin in India. Econ Affairs. 2018;63(2):547-555.
14. Montgomery DC, Peck EA, Vining GG. Introduction to linear regression analysis. New York: John Wiley & Sons; c2003. p. 221-258.
15. Nochai R, Nochai T. ARIMA model for forecasting oil palm price. In: Proceedings of the 2nd IMT GT regional conference on Mathematics, Statistics and Applications; 2006; Penang, Malaysia.
16. Pal S, Mazumdar D. Forecasting monthly rainfall using artificial neural network. RASHI J Soc. Appl. Stat Agric. Allied Sci. 2018;3(2):65-73.
17. Pankratz A. Detecting and treating outliers in dynamic regression models. Biometrika. 1993;80(4):847-854.
18. Parmar RS, Bhojani SH, Chaudhari GB. Application of regression models for area, production and productivity trend of maize crop for Panchmahal region of Gujarat state, India. Int J Trend Sci Res Dev. 2017;1(3):17-22.
19. Rajarathinam A, Parmar RS, Vaishnav PR. Estimating models for area, production and productivity trends of tobacco (*Nicotiana tabacum*) crop for Anand region of Gujarat state. Indian J Appl. Sci. 2010;10:2419-2425.
20. Rangaswamy R. A textbook of agricultural statistics. New Age International (P) Limited Publishers; c2006. p. 158-210.
21. Rao SG, Naidu GM. Statistical modeling on area, production and productivity of wheat in India. Int J Agric Stat Sci. 2021;17(2):539-543.
22. Ravita, Verma U. Application of ARIMA modeling for mustard yield prediction in Haryana. Int J Appl Math Stat Sci. 2016;5(4):23-28.
23. Sahu PK, Mishra P. Modelling and forecasting production behaviour and import-export of total spices in two most populous countries of the world. J Agric Res. 2013;51(1):81-97.
24. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika. 1965;52(3-4):591-611.
25. Varadharaj S, Prakash AR. An analytical study on the production and the growth trends of spices in India. Int J Res Anal Rev. 2018;5(3):227-281.