JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 3512-3523
https://publishoa.com
ISSN: 1309-3452

Design and Development of ARIMA Model for Bajra (*Pennisetum glaucum*) Production in India

T. Jai Sankar *and* **P. Pushpa**
Department of Statistics, Bharathidasan University, Tiruchirappalli, Tamilnadu, INDIA
tjaisankar@gmail.com *and* pushbkr@gmail.com

Received: 2022 March 15; **Revised:** 2022 April 20; **Accepted:** 2022 May 10.

Abstract

This study deals with design and development of autoregressive integrated moving average (ARIMA) model for Bajra (*Pennisetum glaucum*) production in India based on Bajra (*P. glaucum*) production during the years from 1951 to 2018. The study considers Autoregressive (AR), Moving Average (MA) and Autoregressive Integrated Moving Average (ARIMA) processes to select the appropriate ARIMA model for Bajra (*P. glaucum*) production in India. Based on ARIMA (p, d, q) and its components Autocorrelation Function (ACF), Partial Autocorrelation Function (PACF), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Normalized BIC and Box-Ljung Q statistics estimated, ARIMA (0,1,1) was selected. Based on the chosen model, it could be predicted that Bajra (*P. glaucum*) production would increase from 9.21 million tons in 2018 to 10.52 million tons in 2025 in India.

Key Words: ARIMA, BIC, Forecasting, Bajra Production.

Introduction

Bajra is a widely grown kind of millet. The botanical name of Bajra is *Pennisetum glaucum*. This crop has been grown in the Indian subcontinent and Africa since prehistoric times.

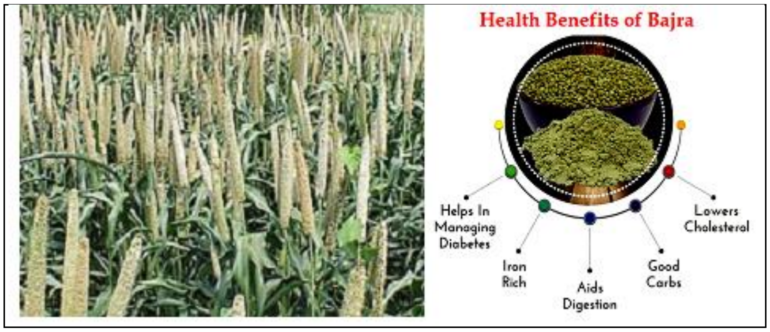

Figure 1 - Health Benefits of Bajra

India is the largest producer of Bajra. India began growing bajra between 1500 and 1100 BCE. Kambu is the Tamil name of Bajra and is an important food across the Indian state of Tamil Nadu. It is the second important food for Tamil people consumed predominantly in the hot humid summer months from February through May every year. It is made into a gruel and consumed along with buttermilk or consumed as dosa or idly. The potential health benefits associated with regularly eating bajra are weight loss, improved diabetes management, and a higher intake of nutrients that support healthy hair, nails, and skin (Figure 1). However, Bajra is a good source of many nutrients known for contributing to healthy hair, skin, and nails, including protein, vitamin B6, niacin folate, iron and zinc.

Material and Methods

As the aim of the study was to implement of ARIMA model for Baajra (*P. glaucum*) production in India, various forecasting techniques were considered for use. ARIMA model, introduced by Box and Jenkins (1976), was frequently applied for discovering the pattern and predicting the future values of the time series data. Box and Pierce (1970) measured the distribution of residual autocorrelations in ARIMA. Akaike (1970) found the stationary time series by an AR (p), where p is finite and bounded by the same integer. Moving Average (MA) models were applied by Slutzky (1973). Md. Moyazzem Hossain and Faruq Abdulla (2016) applied ARIMA (0,2,1) model for yearly potato production in Bangladesh for the period from 1971 to 2013. Borkar et al. (2016) found that ARIMA (2,1,1) is the appropriate model for forecasting the production of cotton in India. Kour et al., (2017) applied ARIMA model for forecasting of productivity of pearl millet of Gujarat for the period from 1960-61 to 2011-12 and validated ARIMA (0,1,1) model performs quite satisfactorily as the RMAPE value is less than 6 percent. Hemavathi and Prabakaran (2018) found rice production data during 1990-2015 and applied ARIMA (0,1,1) model up to 2020. Bhola Nath et al. (2019) discovered ARIMA (1,1,0) model for wheat production in India for the period from 1949-50 to 2016-17 and forecasted up to 2026-27.

Stochastic time-series ARIMA models were widely used in time series data which are having the characteristics (Alan Pankratz, 1983) of parsimonious, stationary, invertible, significant estimated coefficients and statistically independent and normally distributed residuals. ARIMA model was used in this study, which required a sufficiently large data set and involved four steps: identification, estimation, diagnostic checking and forecasting. Model parameters were estimated to fit the ARIMA models.

Autoregressive process of order (p) is, $Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \varepsilon_t$;

Moving Average process of order (q) is, $Y_t = \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - .... - \theta_q \varepsilon_{t-q} + \varepsilon_t$; and

The general form of ARIMA model of order (p, d, q) is

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - .... -$$

where Y_t is Bajra (*P. glaucum*) production, ε_t's are independently and normally distributed with zero mean and constant variance σ^2 for t = 1,2,..., n; d is the fraction differenced while interpreting AR and MA and ϕs and θs are coefficients to be estimated.

Trend Fitting **:** The Box-Ljung Q statistics was used to transform the non-stationary data into stationarity data and also to check the adequacy for the residuals. For evaluating the adequacy of AR, MA and ARIMA processes, various reliability statistics like R^2, Stationary R^2, Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and BIC were used. The reliability statistics viz. RMSE, MAPE, BIC and Q statistics were computed as below:

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

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{(Y_i - \hat{Y}_i)}{Y_i} \right|$$ *and* BIC(p,q) = ln

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 3512-3523
https://publishoa.com
ISSN: 1309-3452

v*(p,q)+(p+q) [ln(n) / n]

where p and q are the order of AR and MA processes respectively and n is the number of observations in the time series and v* is the estimate of white noise variance σ^2.

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

where n is the number of residuals and rk is the residuals autocorrelation at lag k.

In this study, the data on Bajra (*P. glaucum*) production in India were collected from the Annual Report (2018), Directorate of Economics and Statistics, Department of Agriculture, Cooperation and Farmers Welfare, Ministry of Agriculture and Farmers Welfare, Government of India for the period from 1951 to 2018 (Table 1) and were used to fit the ARIMA model to predict the future production.

Results and Discussion

In this study, the data for Bajra production in India is collected from the period 1951 to 2018 is given in Table 1. To fit an Autoregressive model, Autoregressive process for any variable involves four steps: identification, estimation, diagnostic and forecasting. ARIMA (p,d,q) model is fitted to check stationarity through examining the graph or time plot of the data. Figure 2 reveals that the data is non-stationary.

Table 1. Actual Bajra (*P. glaucum*) production (million tons) in India

Year	**Production**	*Year*	**Production**	*Year*	**Production**	*Year*	**Production**
1951	2.60	*1969*	3.80	*1987*	4.51	*2005*	7.93
1952	2.35	*1970*	5.33	*1988*	3.30	*2006*	7.68
1953	3.19	*1971*	8.03	*1989*	7.78	*2007*	8.42
1954	4.55	*1972*	5.32	*1990*	6.65	*2008*	9.97
1955	3.52	*1973*	3.93	*1991*	6.89	*2009*	8.89
1956	3.43	*1974*	7.52	*1992*	4.67	*2010*	6.51
1957	2.87	*1975*	3.27	*1993*	8.88	*2011*	10.37
1958	3.62	*1976*	5.74	*1994*	4.97	*2012*	10.28
1959	3.87	*1977*	5.85	*1995*	7.16	*2013*	8.74

1960	3.49	*1978*	4.73	*1996*	5.38	*2014*	9.25
1961	3.28	*1979*	5.57	*1997*	7.87	*2015*	9.18
1962	3.65	*1980*	3.95	*1998*	7.64	*2016*	8.07
1963	3.96	*1981*	5.34	*1999*	6.96	*2017*	9.73
1964	3.88	*1982*	5.54	*2000*	5.78	*2018*	9.21
1965	4.52	*1983*	5.13	*2001*	6.76		
1966	3.75	*1984*	7.72	*2002*	8.28		
1967	4.47	*1985*	6.05	*2003*	4.72		
1968	5.19	*1986*	3.66	*2004*	12.11		

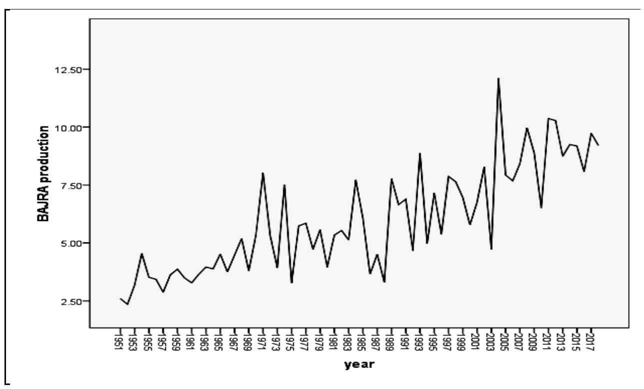

Figure 2 - Time plot of Bajra (*P. glaucum*) Production

The autocorrelation and partial autocorrelation coefficients of various orders of Y_t are computed Table 2. The graphs of ACF and PACF are given in Figure 3.

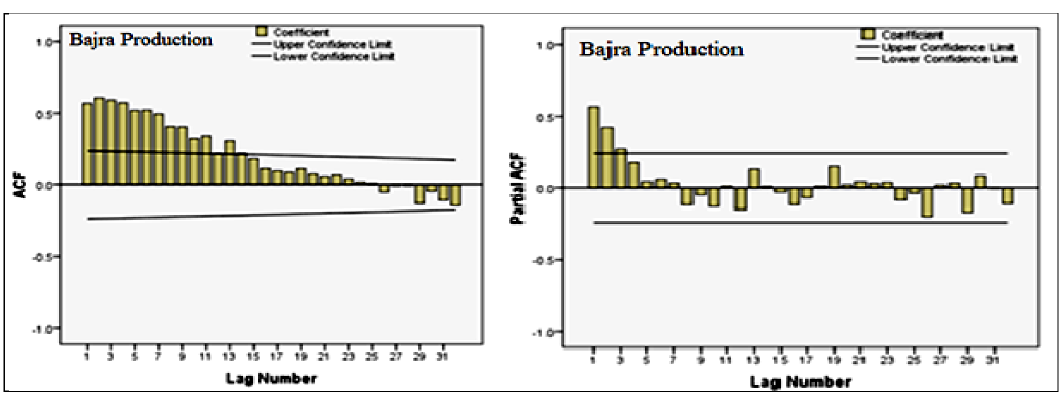

Figure 3 ACF and PACF of differenced data

Table 2 - ACF and PACF of Bajra (*P. glaucum*) Production

Lag	AC	Std. Error [a]	Box-Ljung Statistic	PAC	Std. Error	Lag	AC	Std. Error [a]	Box-Ljung Statistic	PAC	Std. Error
	Value	Df	Sig.[b]	Value	Df		Value	Df	Sig.[b]	Value	Df
1	0.569	0.119	23.008	0.569	0.121	***17***	0.098	0.104	226.404	-0.070	0.121
2	0.607	0.118	49.554	0.418	0.121	***18***	0.088	0.102	227.135	0.012	0.121
3	0.591	0.117	75.155	0.273	0.121	***19***	0.114	0.101	228.394	0.146	0.121
4	0.573	0.116	99.580	0.175	0.121	***20***	0.076	0.100	228.970	0.022	0.121
5	0.521	0.115	120.065	0.043	0.121	***21***	0.056	0.099	229.293	0.042	0.121
6	0.522	0.114	140.998	0.061	0.121	***22***	0.068	0.098	229.775	0.031	0.121
7	0.495	0.113	160.152	0.035	0.121	***23***	0.037	0.097	229.922	0.038	0.121
8	0.407	0.112	173.266	-0.118	0.121	***24***	0.015	0.096	229.947	-0.086	0.121
9	0.404	0.111	186.412	-0.049	0.121	***25***	0.005	0.095	229.950	-0.036	0.121
10	0.323	0.110	194.987	-0.129	0.121	***26***	-0.051	0.094	230.242	-0.199	0.121
11	0.340	0.109	204.654	0.012	0.121	***27***	-0.008	0.093	230.250	0.019	0.121
12	0.221	0.108	208.790	-0.149	0.121	***28***	-0.008	0.092	230.256	0.034	0.121
13	0.308	0.107	216.979	0.127	0.121	***29***	-0.128	0.091	232.250	-0.167	0.121
14	0.221	0.107	221.270	0.010	0.121	***30***	-	0.089	232.509	0.085	0.121

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 3512-3523
https://publishoa.com
ISSN: 1309-3452

							0.045				
15	0.184	0.106	224.298	-0.030	0.121	***31***	-0.106	0.088	233.968	-0.003	0.121
16	0.115	0.105	225.501	-0.116	0.121	***32***	-0.141	0.087	236.583	-0.111	0.121

[a]The underlying process assumed is independence (white noise).

[b]Based on the asymptotic chi-square approximation.

The models and corresponding BIC values are given in Table 3. The value of normalized BIC is 0.807 and R Squared value is 0.654. So the most suitable model for Bajra Production is ARIMA(0,1,1) as this model has the lowest BIC value.

Model Estimation: Model parameters were estimated and reported in Table 3 and Table 4. The model verification is concerned with checking the residuals of the model to improve on the chosen ARIMA (p,d,q). This is done through examining the autocorrelations and partial autocorrelations of the residuals of various orders, up to 32 lags were computed and the same along with their significance which is tested by Box-Ljung test are provided in Table 5. This proves that the selected ARIMA model is an appropriate model.

Table - 2 BIC values of ARIMA(p,d,q)

ARIMA (p,d,q)	**BIC Values**
0,1,0	1.599
0,1,1	**0.807**
0,1,2	0.828
1,1,0	1.233
1,1,1	0.829
1,1,2	0.908
2,1,0	1.127

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 3512-3523
https://publishoa.com
ISSN: 1309-3452

2,1,1	0.905
2,1,2	0.985
3,1,0	1.109
3,1,1	0.985
3,1,2	1.091

Table 3 - Estimated AR Model of Bajra (*P. glaucum*) Production

	Estimate	**SE**	**t**	**Sig.**
Constant	-2.118	1.873	-1.131	0.262
MA 1	0.994	0.480	2.072	0.042

Table 4 - Estimated AR Model Fit Statistics

ARIMA (p,d,q)	**Stationary R^2**	**R^2**	**RMSE**	**MAPE**	**MaxAPE**	**MAE**	**MaxAE**	**Normalized BIC**
0,1,0	0.000	0.174	2.089	26.125	133.056	1.519	7.296	1.599
0,1,1	**0.581**	**0.654**	**1.362**	**17.051**	**85.595**	**0.953**	**4.398**	**0.807**
0,1,2	0.604	0.673	1.334	16.936	99.218	0.937	3.794	0.828
1,1,0	0.358	0.47	1.686	21.434	96.773	1.234	5.127	1.233
1,1,1	0.604	0.673	1.335	16.974	96.51	0.94	3.742	0.829
1,1,2	0.604	0.673	1.346	17.132	101.383	0.944	3.738	0.908
2,1,0	0.467	0.56	1.549	19.619	81.019	1.106	4.819	1.127
2,1,1	0.605	0.674	1.344	16.973	100.966	0.936	3.716	0.905
2,1,2	0.605	0.674	1.356	16.974	99.402	0.941	3.736	0.985
3,1,0	0.516	0.6	1.488	19.256	85.948	1.051	4.986	1.109
3,1,1	0.605	0.674	1.356	16.99	98.818	0.939	3.779	0.985
3,1,2	0.594	0.665	1.385	18.163	93.688	0.997	3.915	1.091

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 3512-3523
https://publishoa.com
ISSN: 1309-3452

The ACF and PACF of the residuals are given in Figure 4. It also indicates 'good fit' of the model. So the fitted ARIMA model for the Bajra Production data is

$$Y_t = \mu - \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

$$Y_t = -2.118 - 0.994\varepsilon_{t-1} + \varepsilon_t$$

Forecasting: Forecasted value of Bajra Production (Quantity in numbers) for the year 2019 through 2025 respectively given by 9.70, 9.83, 9.97, 10.11, 10.24, 10.38 and 10.52 are given in Table 6. To assess the forecasting ability of the fitted ARIMA (p,d,q) model, important measures of the sample period forecasts' accuracy were computed. This measure indicates that the forecasting inaccuracy is low. Figure 5 shows that the actual and forecasted value of Bajra Production data with 95% confidence limits.

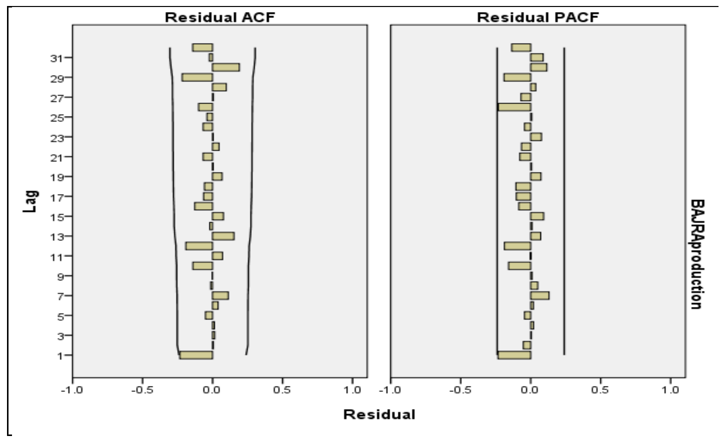

Figure 4 - Residuals of ACF and PACF

Table - 5 Residual of ACF and PACF of Bajra (*P. glaucum*) Production

Lag	**ACF**		**PACF**		***Lag***	**ACF**		**PACF**	
	Mean	**SE**	**Mean**	**SE**		**Mean**	**SE**	**Mean**	**SE**
1	-0.234	0.122	-0.234	0.122	***17***	-0.064	0.142	-0.103	0.122
2	0.006	0.129	-0.052	0.122	***18***	-0.057	0.143	-0.105	0.122
3	0.016	0.129	0.005	0.122	***19***	0.068	0.143	0.073	0.122
4	0.013	0.129	0.020	0.122	***20***	0.004	0.144	0.006	0.122
5	-0.050	0.129	-0.044	0.122	***21***	-0.067	0.144	-0.077	0.122
6	0.040	0.129	0.019	0.122	***22***	0.046	0.144	-0.066	0.122

7	0.112	0.129	0.131	0.122	***23***	0.005	0.144	0.076	0.122
8	-0.014	0.131	0.050	0.122	***24***	-0.067	0.144	-0.044	0.122
9	-0.004	0.131	0.009	0.122	***25***	-0.040	0.145	0.007	0.122
10	-0.140	0.131	-0.159	0.122	***26***	-0.100	0.145	-0.231	0.122
11	0.071	0.133	-0.001	0.122	***27***	0.006	0.146	-0.069	0.122
12	-0.191	0.133	-0.187	0.122	***28***	0.098	0.146	0.035	0.122
13	0.153	0.137	0.072	0.122	***29***	-0.217	0.147	-0.190	0.122
14	-0.020	0.140	0.009	0.122	***30***	0.192	0.152	0.115	0.122
15	0.078	0.140	0.093	0.122	***31***	-0.023	0.155	0.089	0.122
16	-0.127	0.141	-0.085	0.122	***32***	-0.141	0.155	-0.136	0.122

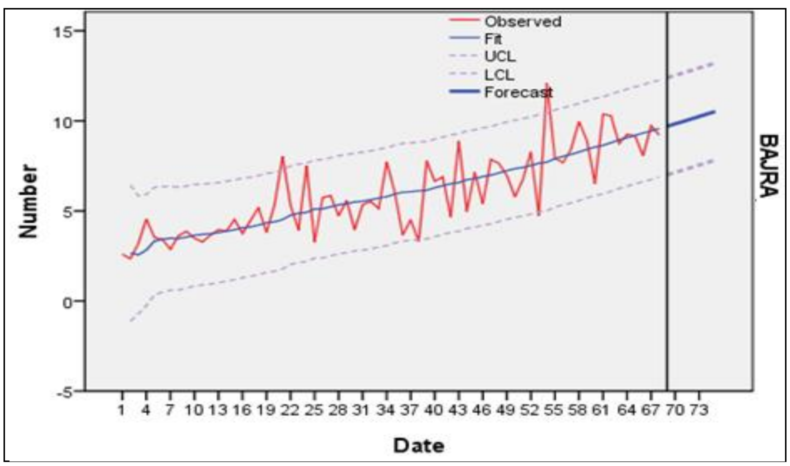

Figure 5 - Actual and Estimate of Bajra (*P. glaucum*) Production

Table 6 - Forecast of Bajra (*P. glaucum*) Production

Year	Predicted	LCL	UCL
2019	9.70	7.01	12.39

2020	9.83	7.14	12.53
2021	9.97	7.28	12.66
2022	10.11	7.41	12.80
2023	10.24	7.55	12.94
2024	10.38	7.69	13.07
2025	10.52	7.83	13.21

Conclusion

The most appropriate ARIMA model for Bajra (*P. glaucum*) production forecasting of data was found to be ARIMA (0,1,1). From the time series data, it can be found that forecasted production would increase to 10.52 million tons in 2025 from 9.1 million tons in 2018 in India for using time series data from 1951 to 2018 on Bajra (*P. glaucum*) production, this study provides an evidence on future Bajra (*P. glaucum*) production in the country, which can be considered for future policy making and formulating strategies for augmenting and sustaining Bajra (*P. glaucum*) production in India.

References

1. Alan Pankratz, Forecasting with Univariate Box - Jenkins Models: Concepts and Cases, John Wiley & Sons, New York, 1983.
2. Bhola Nath, DS Dhakre and Debasis Bhattacharya, Forecasting wheat production in India: An ARIMA modelling approach, *Journal of Pharmacognosy and Phytochemistry*, (2019) no. 8(1), 2158-2165.
3. Borkar, Prema and P.M. Tayade, Forecasting of Cotton Production in India Using ARIMA Model, *International Journal of Research in Economics and Social Sciences*, (2016), no. **6(5),** 1-7.
4. E. Slutzky, The summation of random causes as the source of cyclic processes, *Econometrica*, (1973), no. **5**, 105-146.
5. G.E.P. Box and D.A. Pierce, Distribution of Residual Autocorrelations in ARIMA Models, *J. American Stat. Assoc.,* (1970), no. **65**, 1509-1526.

6. G.E.P. Box and G.M. Jenkins, Time Series Analysis: Forecasting and Control, San Francisco, Holden-Day, California, USA, 1976.
7. H. Akaike, Statistical Predictor Identification, *Annals of Institute of Statistical Mathematics.* (1983) no. **22,** 203-270.
8. Kour Satvinder, U.K. Pradhan, Ranjit Kumar Paul and P.R. Vaishnav, Forecasting of Pearl Millet Productivity in Gujarat Under Time Series Framework, *Economic Affairs*, (2017), no. **62(1)**, 121-127.
9. M. Hemavathi and K. Prabakaran, 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), no. **7(2)**, 149-156.
10. Md. Moyazzem Hossain and Faruq Abdulla, Forecasting Potato Production in Bangladesh by ARIMA Model, *Journal of Advanced Statistics*, (2016), no. **1(4)**, 191-198.