ASUTOSH COLLEGE

Estd. 1961

92, Shyama Prasad Mukherjee Rd, Kolkata - 700026

(Affiliated o **University of Calcutta**)

Department of Statistics

NAME : ANIKET SANTRA

STREAM : B.Sc. HONOURS

COLLEGE ROLL NO. : 1195

UNIVERSITY ROLL NO. : 183012-21-0432

UNIVERSITY REGISTRATION NO. : 012-1111-1357-18

SUBJECT : STATISTICS

PAPER : DSE-B2 (PROJECT WORK)

SEMESTER : VI

YEAR : 2021

Rice Production of India – A Case Study and Detailed Data Analysis

Contents

- Executive Summary of the Project 1
- Introduction 2
- Collected Data 3-5
- Methodology 6-12
  - □ Graphical Representation of data *6*
    - Δ Line Diagram *6*
    - Δ Bar Diagram *6*
    - Δ Divide Bar Diagram and Pie Chart *7*
  - □ Time Series Analysis Method *7*
    - Δ Simple Moving Average Method *7*
    - Δ Mathematical Curve Fitting *8*
    - Δ Significance Test *9*
- Result and Analysis 13-36
  - □ Studying and Comparing by Graphical Representation of the data *13*
  - □ Analysis Trend by Simple Moving Average Method *28*
  - □ Fitting a Mathematical Curve *32*
- Conclusion 37
- References 38
- Acknowledgement 39
- Declaration 40

• Executive Summary of The Project:-

Rice is the staple food for more than half of the total population in India and hence a key pillar for food security. It comprised 42% of total food grain production in 2008-09 in India. The Green Revolution in the late 1960s in India made a huge impact in rice production. After that with the help of state government, the Indian Government implemented several policies to boost rice production. Numerous subsidies, ranging from fertilizer to irrigation, electricity, seeds, machinery are available and it has led to high productivity of rice.

In this project it has been attempted to study and analyse the rice production of the whole India and the production of the major rice producing states of India. The data of rice production of major rice producing states for past 10 years (2009-18) and the total production of India from-1950 to 2018 have been collected.

At first we have compared the rice production of the major states by exploratory data analysis with the help of some statistical diagrams like bar diagrams, pie charts and left our conclusion.

In the next part we have used time series analysis method to study the underlying pattern of the rice production of whole India over time and analysis the trend. Then we have used time series modelling to fit a suitable trend line equation for our rice production data and tried to predict the future production of upcoming years through the estimated trend line equation.

• Introduction:-

Rice is one of the chief grains of India. Moreover, this country has the largest area under cultivation, as it's one of the principal food crops. It is, in fact, the dominant crop of the country.

Rice is the basic food crop and being a tropical plant, it flourishes comfortably in hot and humid climate. Rice is mainly grown in rain-fed areas that receive heavy annual rainfall. That is why it is fundamentally a kharif crops in India. It demands temperature of around 25 degree Celsius and above, and rainfall of more than 100 cm. Rice is also grown through irrigation in those areas that receive comparatively less rainfall. Rice is the staple food of eastern and southern parts of India.

The rice growing areas in the country can be broadly grouped into five regions as given below:

i. North-Eastern Region: This region comprises of Assam and North eastern states. In Assam rice is grown in the Basin of Brahmaputra River. This region receives very heavy rainfall and rice is grown under rain fed condition.

ii. Eastern Region: It region comprises of Bihar, Chhattisgarh, Jharkhand, Madhya Pradesh, Odisha, Eastern Uttar Pradesh and West Bengal. In this region rice is grown in the basins of Ganga and Mahanadi rivers and has the highest intensity of rice cultivation in the country. This region receives heavy rainfall and rice is grown mainly under rainfed conditions.

iii. Northern Region: This region comprises of Haryana, Punjab, Western Uttar Pradesh, Uttarakhand, Himachal Pradesh and Jammu & Kashmir. The region experiences low winter temperature and single crop of rice from May-July to September-December is grown.

iv. Western Region: This region comprises of Gujarat, Maharashtra and Rajasthan. Rice is largely grown under rainfed condition during June-August to October - December.

v. Southern Region: This region comprises of Andhra Pradesh, Karnataka, Kerala and Tamil Nadu. Rice is mainly grown in deltaic tracts of Godavari, Krishna and Cauvery rivers and the non-deltaic rainfed area of Tamil Nadu and Andhra Pradesh. Rice is grown under irrigated condition in deltaic tracts.

India is the world's second largest producer of rice, and the largest exporter of rice in the world. Production increased from 42.22 million tons in financial year 1970 to 116.42 million tons in financial year 2018. In 2020-21 it's almost 120 million tons.

Collected Data On Annual Rice Production Of 15 Major Rice Producing States of India For Past Ten Years:-

Table 1

Production (in million tonnes)											
State	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	10 Years Total
West Bengal	14.34	13.05	14.61	15.02	15.37	14.68	15.75	15.30	14.97	16.05	149.14
Uttar Pradesh	10.81	11.99	14.02	14.42	14.64	12.17	12.51	13.75	13.27	15.54	133.12
Punjab	11.24	10.84	10.54	11.37	11.27	11.11	11.82	11.59	13.38	12.82	115.98
Andhra Pradesh	10.54	14.42	12.90	11.51	12.72	7.23	7.49	7.45	8.17	8.25	100.68
Odisha	6.92	6.83	5.81	7.30	7.61	8.30	5.88	8.33	6.55	7.31	70.84
Telangana	0.00	0.00	0.00	0.00	0.00	4.44	2.96	5.17	6.26	6.70	25.53
Chhattisgarh	4.11	6.16	6.03	6.61	6.72	6.32	6.09	8.05	4.93	6.53	61.55
Tamil Nadu	5.67	5.79	7.46	4.05	5.35	5.73	7.98	2.37	6.64	6.45	57.49
Bihar	3.60	3.10	7.16	7.53	5.50	6.36	6.49	8.24	8.09	6.04	62.11
Assam	4.34	4.74	4.52	5.13	4.93	5.22	5.14	4.73	5.28	5.14	49.17
Haryana	3.63	3.47	3.76	3.98	4.00	4.01	4.14	4.45	4.52	4.52	40.48
Jharkhand	1.54	1.11	3.13	3.16	2.81	3.36	2.88	3.84	4.08	2.89	28.8
Maharashtra	2.18	2.70	2.84	3.06	3.12	2.95	2.63	3.11	2.73	3.28	28.6
Karnataka	3.69	4.19	3.96	3.36	3.57	3.54	2.70	2.60	3.02	3.43	34.06
Madhya Pradesh	1.26	1.77	2.23	2.78	2.84	3.63	3.58	4.23	4.12	4.50	30.94
Others	5.22	5.82	6.33	5.95	6.20	6.43	6.37	6.49	6.75	6.97	62.53
Total	89.09	95.98	105.30	105.23	106.65	105.48	104.41	109.70	112.76	116.42	1051.02

Collected Data On Annual Rice Production Of Whole India From 1950 To 2018:-

Table 2

Year	Production (in million tonnes)
1950	20.58
1951	21.30
1952	22.90
1953	28.21
1954	25.22
1955	27.56
1956	29.04
1957	25.53
1958	30.85
1959	31.68
1960	34.58
1961	35.66
1962	33.21
1963	37.00
1964	39.31
1965	30.59
1966	30.44
1967	37.61
1968	39.76
1969	40.43
1970	42.22
1971	43.07
1972	39.24
1973	44.05
1974	39.58
1975	48.74
1976	41.92
1977	52.67
1978	53.77
1979	42.33
1980	53.63
1981	53.25
1982	47.12
1983	60.10
1984	58.34
1985	63.83

1986	60.56
1987	56.86
1988	70.49
1989	73.57
1990	74.29
1991	74.68
1992	72.86
1993	80.30
1994	81.81
1995	76.98
1996	81.73
1997	82.54
1998	86.08
1999	89.68
2000	84.98
2001	93.34
2002	71.82
2003	88.53
2004	83.13
2005	91.79
2006	93.36
2007	96.69
2008	99.18
2009	89.09
2010	95.98
2011	105.30
2012	105.23
2013	106.65
2014	105.48
2015	104.41
2016	109.70
2017	112.76
2018	116.42

***Footnotes:** Here we have collected the rice production data at the year time point and the production's unit is million tonnes.

**We will use this unit (million tonnes) of production further for studying or analysing in our entire project.

• Methodology

Δ Comparison by Diagrammatic Representation

Diagrams like graphs, charts, maps, pictures etc. are attractive and effective means for presentation of statistical data. It is more effective than tabular representation, being easily intelligible to a layman. Indeed, diagrams are almost essential whenever it is required to convey any statistical information to the general public. Diagrams are readily capable of revealing some features of the exhibited data.

Line Diagram

This diagram is meant for representing chronological data. In fact, it exhibits the relationship of the variable (e.g. sales of coffee of a company, productions of a crop) may be specified for individual points of time or for different period of time

In constructing a line diagram, two axis of co-ordinates are taken, the horizontal one for time and the vertical one for variable. The scale for each axis is then selected and the data are plotted as different points on the plane, the plotting of variable values being done against points of time or mid-points of the time interval (for time period). The successive points are now joined by straight line segments and the chart so obtained is called a line diagram for the given data.

Two or mutually related time series data having same unit of measurement can be represented using the same axis of co-ordinates, by drawing a number of line diagrams, one for each series. These different line diagrams are mutually distinguished by using distinct pattern of lines such as broken lines, dotted lines or multiple coloured lines. The resulting diagram is known as a **Multiple Line Diagram.** It is used for comparing two mutually related time series data e.g. if we want to compare the literacy rates for a number countries last 15 years, say, we may draw multiple line diagram.

Bar Diagram

Another mode of diagrammatic representation of data is the use of bar diagrams. These have more general applicability than line diagrams in the sense that they may be used for series varying either over time or over space. In this method bars of equal width are taken for the different items of the series, drawn over base line. The length or height of a bar representing the value of the variable concerned. It is preferable to take the bars horizontally for data varying over space and vertically in the case of a series varying over time. We can compare the different items of the series by visualizing bar diagram.

Divided Bar Diagram and Pie Chart

In some situations, the value of a variable are available for a number of components and comparison among different components or the relation between each part and the whole may be necessary. In this context, the proportions or percentages of the variable components are given more importance than the absolute values, proportions or percentages are expected to give a better idea of the relative importance of each components. For this purpose, one may draw either a divided bar diagram or pie chart

In the first case, a bar of suitable length and width is taken, its total area being regarded as 100. If a vertical bar is chosen, then this area is divided into a number of sections by drawing lines parallel to the base, in such a way that the area of each sections represent the percentage for the corresponding category.

A pie chart is another appropriate diagram used for exhibiting the relative sizes of the different parts of a whole. In this case, a circle is partitioned into several sectors by drawing angles at the centre, the area of each sector indicating the corresponding percentage. In fact, the area enclosed by the circle is regarded as 100. Since the total angle at the centre is 360^o, the desired angle for some particular category will be 3.6 times the relevant percentage. The diagram, thus constructed, is termed as pie diagram.

By using divided bar diagram and pie chart we can compare the relative contributions or importance of different categories to the study variable.

Δ Method of Time Series Analysis

◻ Simple Moving Average

A moving average is a calculation to analyse the data points by creating a series of averages of different subsets of the full dataset. The simple moving average of period k of a time series gives us a new series of arithmetic means, each of k successive observations of the time series. We start with the first k observations. At the next stage, we leave the first and include the $(k+1)^{st}$ observation. This process is repeated until we arrive at the last k observations. Each of these means is centred against the time which is the mid-point of the time interval included in the calculation of the moving average. Thus when k, the period of the moving average, is odd, the moving average values correspond to tabulated time values for which the time series is given. When the period is even, the moving average fails midway between two tabulated values. In this case, we calculate a subsequent twoitem moving average to make the resulting moving average values correspond to the tabulated time periods.

Moving averages are an important analytical tool used to identify current trends and the potential for a change in an established trend.

□ Mathematical Curve Fitting

It is an essential part of the concept of trend that the movement over fairly long periods is smooth. This is perhaps the best and most rational method of determining the trend. In this case, a suitable trend equation is selected and then the constants involved in the equation are estimated on the basis of the data in hand. After derived the estimated trend equation we can predict the future trend values.

Polynomial Trend Line

Here we assume that our suitable trend equation is a polynomial in time element 't'. So here we consider a mathematical model:-

$\mathbf{Y_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3 + \ldots.. + a_p t^p + e_t}$

where a_i's, $i = 1(1)p$ are constants and e_t denotes a random error with $\mathbf{E(e_t) = 0}$ and $\mathbf{V(e_t) = \sigma^2}$

Now we can estimate the constants a_i's by method of least squares and fit the polynomial model. In this method the constants are determined by minimizing,

$$S = \sum_t \left(y_t - a_0 - a_1 t - a_2 t^2 - \cdots - a_p t^p\right)^2$$

The normal equations are,

$$\frac{\partial S}{\partial a_j} = 0, \forall\, j = 0(1)p$$

$$\Rightarrow \sum_t t^j y_t = a_0 \sum_t t^j + a_1 \sum_t t^{j+1} + a_2 \sum_t t^{j+2} + \cdots + a_p \sum_t t^{j+p}, j = 0(1)p$$

By solving the normal equations we can obtain the estimates of the constants a_0, a_1, a_2,…, a_p, and fit a trend equation on the time element 't'.

Growth Curves

The family of curves (polynomials) described above represents a simple and very useful type but a curve of $Y_t = a_0 + a_1t$ or $Yt = a_0 + a_1t + a_2t^2$ etc may not be a satisfactory description of the trend of some time series for the period shown or for the prediction purpose also. Perhaps of even greater general utility, in the analysis of time series, are curves of a semi-logarithmic (exponential) type.

Exponential Curve

The simplest exponential curve may be written as $\mathbf{Y_t = ab^t}$, where $a > 0$, $b > 0$

Now taking logarithm both side, $\mathbf{\log Y_t = \log a + t \times (\log b)}$, which is a straight line in t

So, here we assume a mathematical model log $\mathbf{Y_t = \log a + t \times (\log b) + e_t}$, where et denotes a random error with $\mathbf{E(e_t)} = 0$ and $\mathbf{V(e_t) = \sigma^2}$

Now we can apply a least squares method to the logarithm of the original data (Y_t) to estimate a and b and fit an exponential trend equation.

□ Significance Test

Consider the regression model

$Y_t = a_0 + a_1t_i + a_2t_i^2 + a_3t_i^3 + \ldots.. + a_pt_i^p + e_{ti}$; $i = 1(1)n$

we assume that $e_{ti} \sim$ iid $N(0, \sigma^2)$

So here are p covariates t, t^2, t^3,…,t^p, the response variable Y_t and let there are n observations of pair (t, Y_t). Now for simplification we define the covariates as $t_1 = t$, $t_2 = t^2$, $t_3 = t^3$,…,$t_p = t^p$; response $Y_t = y$ and error $e_t = e$. Therefore the model becomes,

$y_i = a_0 + a_1t_{1i} + a_2t_{2i} + a_3t_{3i} + \ldots.. + a_pt_{pi} + e_i$; $i = 1(1)n$ where $e_i \sim$iid $N(0, \sigma^2)$

Now we want to test whether the covariate has any significant effect on response variable or not. So, our testing problem will be,

$H_0 : a_j = 0 \ \forall\ j = 1,2,\ldots,p$

vs.

$H_1 : a_j \neq 0$ for at least one j

Now, the unrestricted sum square error (SSE) will be,

$$SSE = \min_{a_i} \sum_{i=1}^{n} \left(\mathrm{y_i} - \mathrm{a_0} - \mathrm{a_1 t_{1i}} - \mathrm{a_2 t_{2i}} - \mathrm{a_3 t_{3i}} + \cdots - \mathrm{a_p t_{pi}}\right)^2$$

For the sake of simplicity, we rewrite the model as,

$y_i = a_0 + a_1 t'_{1i} + a_2 t'_{2i} + a_3 t'_{3i} + \ldots.. + a_p t'_{pi} + e_i$

where, $t'_{ji} = t_{ji} - \bar{t}_j \; \forall \; j = 1,2,\ldots,p$

$$i.e., SSE = \sum_{i=1}^{n} \left(\mathrm{y_i} - \mathrm{a_0} - \mathrm{a_1 t'_{1i}} - \mathrm{a_2 t'_{2i}} - \mathrm{a_3 t'_{3i}} - \cdots - \mathrm{a_p t'_{pi}}\right)^2$$

$$\therefore \frac{\partial}{\partial a_0} \sum_{i=1}^{n} \left(\mathrm{y_i} - \mathrm{a_0} - \mathrm{a_1 t'_{1i}} - \mathrm{a_2 t'_{2i}} - \mathrm{a_3 t'_{3i}} - \cdots - \mathrm{a_p t'_{pi}}\right)^2 = 0$$

$$\sum_{i=1}^{n} \left(\mathrm{y_i} - \mathrm{a_0} - \mathrm{a_1 t'_{1i}} - \mathrm{a_2 t'_{2i}} - \mathrm{a_3 t'_{3i}} - \cdots - \mathrm{a_p t'_{pi}}\right) = 0$$

$$\sum_{i=1}^{n} y_i = n a_0 + a_1 \sum_{i=1}^{n} t'_{1i} + a_2 \sum_{i=1}^{n} t'_{2i} + \cdots + a_p \sum_{i=1}^{n} t'_{pi}$$

$$Now, \sum_{i=1}^{n} t'_{ji} = \sum_{i=1}^{n} (t_{ji} - \bar{t}_j) = 0, \forall \; j = 1(1)p$$

$$\Rightarrow a_0 = \frac{1}{n} \sum_{i=1}^{n} y_i$$

$$\Rightarrow \hat{a}_0 = \bar{y}$$

$$\frac{\partial}{\partial a_j} \sum_{i=1}^{n} \left(\mathrm{y_i} - \mathrm{a_0} - \mathrm{a_1 t'_{1i}} - \mathrm{a_2 t'_{2i}} - \mathrm{a_3 t'_{3i}} - \cdots - \mathrm{a_p t'_{pi}}\right)^2 = 0$$

$$\Rightarrow \sum_{i=1}^{n} \left(\mathrm{y_i} - \mathrm{a_0} - \mathrm{a_1 t'_{1i}} + \mathrm{a_2 t'_{2i}} - \mathrm{a_3 t'_{3i}} - \cdots - \mathrm{a_p t'_{pi}}\right) t'_{ji} = 0$$

We can solve the above normal equations and obtain the value of $\hat{t}_1, \hat{t}_2, \hat{t}_3, \ldots, \hat{t}_p$

$$i.e. \; SSE = \sum_{i=1}^{n} \left(\mathrm{y_i} - \mathrm{\hat{a}_0} - \mathrm{\hat{a}_1 t_{1i}} - \mathrm{\hat{a}_2 t_{2i}} - \mathrm{\hat{a}_3 t_{3i}} - \cdots - \mathrm{\hat{a}_p t_{pi}}\right)^2 \; with \; df = n - k - 1$$

$$Now, under \; H_0, a_j = 0 \; \forall \; j \; and \; SSE_{H_0} = \min_{a_0} \sum_{i=1}^{n} (y_i - a_0)^2$$

Here, $\hat{a}_0^{H0} = \bar{y}$ [SD is the least RMSD]

$$i.e.\ SSE_{H_0} = \sum_{i=1}^{n}(y_i - \bar{y})^2\ with\ df = n - 1$$

$$\therefore \frac{SSE_{H_0} - SSE}{\sigma^2} \sim \chi^2_{n-1-(n-k-1)} \equiv \chi^2_k\ \ and\ \frac{SSE}{\sigma^2} \sim \chi^2_{n-k-1}$$

Also it can be shown that SSE_{H0} – SSE is independent of SSE

$$F = \frac{\frac{\frac{SSE_{H_0} - SSE}{\sigma^2}}{k}}{\frac{\frac{SSE}{\sigma^2}}{n-k-1}} \sim F_{k,n-k-1}$$

So, we reject H_0 at level α if $F_{obs} > F_{\alpha;k,n-k-1}$

In case of p-value if our p-value i.e. $P(F>F_{obs}) < \alpha$ we reject our null hypothesis i.e. $a_j = 0\ \forall\ j = 1,2,\ldots,p$

So if our p-value $< \alpha$ we can say the covariates have significant effect on response and in that case our model is statistically significant.

R^2 – Coefficient of Determination and Adjusted R Square

In statistics the coefficient of determination denoted by R^2 or r^2 and pronounced "R squared", is the proportion of the variation in the dependent variable that is predicted from the independent variable(s).

Definition:-
A data set has n values marked $y_1, y_2,\ldots y_n$(denoted as y_i) each associated with a fitted(or modelled, or predicted) value $f_1,f_2,\ldots f_n$(denoted as f_i or sometimes $\hat{y}_i$)

Define the residuals as $e_i = y_i - f_i$

If $\bar{y}$ is the mean of the observed data:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Then the variability of the dataset can be measured with two sum of squares formulas:

The total sum of squares (proportional to variance of the data):

$$SS_{tot} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

The sum of squares of residuals, also called residual sum of squares:

$$SS_{res} = \sum_{i=1}^{n} (y_i - f_i)^2 = \sum_{i=1}^{n} e_i^2$$

Then the most general definition of the coefficient of determination is

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

R^2 is a statistic that will give some information about the goodness of fit of a model. In regression it measures how well the regression predictions approximate the real data points. The value of R^2 lies between 0 and 1. More the value of R^2 close to 1 indicates a better fit of model. R^2 of 1 indicates that the regression predictions perfectly fit the data.

Adjusted R^2

However each time we add a new predictor variable to the model the R- squared is guarantee to increase even if the predictor variable isn't useful.

The adjusted R-squared or adjusted R^2 is a modified version of R^2 that adjusts for the number of predictors in a regression model. It is calculated as

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

R^2 : The coefficient of determination of the model
n : The number of observations
k : The number of predicted variables

Since R^2 always increases as you add more predictors to a model, adjusted R^2 can serve as a metric that tells you how useful a model is, adjusted for the number of predictors in a model. Therefore the adjusted R^2 tells us the percentage of variation explained by only the independent variables that actually affect the dependent variable. Same as R^2, the value of adjusted R^2 lies between 0 and 1. More the value of adjusted R^2 close to 1 indicates that all the predictor variables in model have better significant effects.

• Analysis and Results:-

First we intend to study the state wise rice production of India. So the major rice producing states are selected and their annual rice production for past ten years are collected. Then we will analyse the total annual production of rice in whole India by using appropriate time series analysis method.

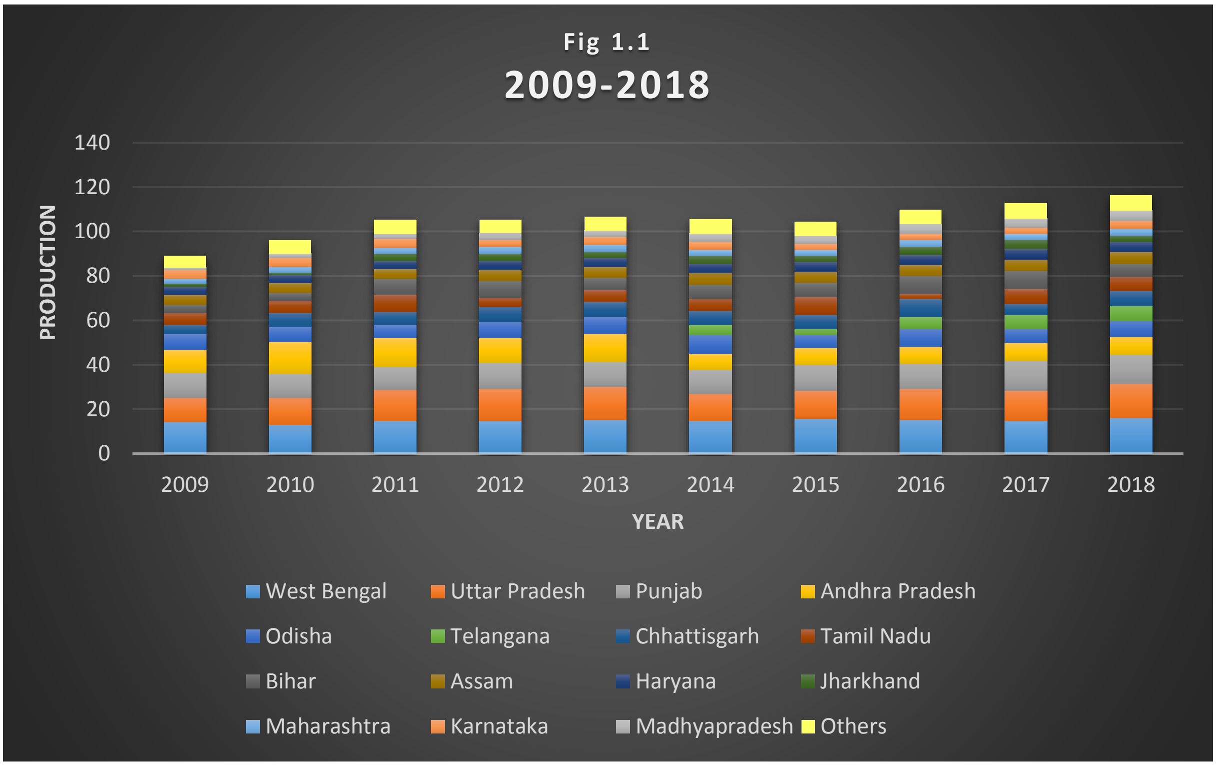
□ Studying and Comparing Data Using Their Graphical Representation:-

For studying and comparing the rice production of different state we use Line Diagrams, Bar Diagrams, Multiple Bar Diagrams, Divided Bar Diagrams and Pie Charts.

The major rice producing states of India are West Bengal, Uttar Pradesh, Punjab, Andhra Pradesh, Odisha, Telangana, Chhattisgarh, Tamil Nadu, Bihar, Assam, Haryana, Jharkhand, Maharashtra, Karnataka, Madhya Pradesh. Now, we want to compare their production and contribution in India's total rice production.

At first we draw a multiple bar diagram and a pie chart of rice production in ten years from 2009-2018 for 15 major rice producing states. Then for each year from 2009-2018 we draw bar diagrams and pie charts of rice production for those states to compare their production. Lastly we uses line diagram for each of the 15 states' production in the specified ten years period to study their production and production's movement over the time.

Figure 1: Multiple Divided Bar Diagram And Pie Chart Of Rice Production In 10 Years 2009-2018 For Major 15 Rice Producing States In India:-



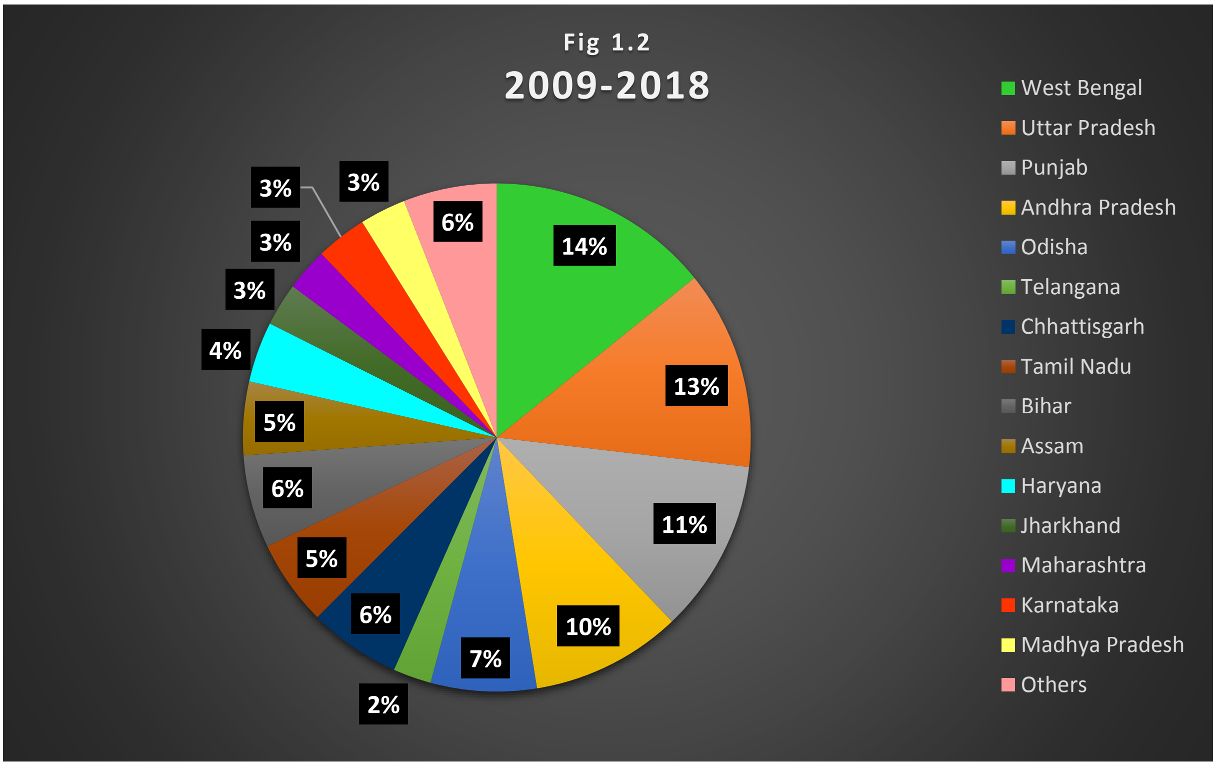
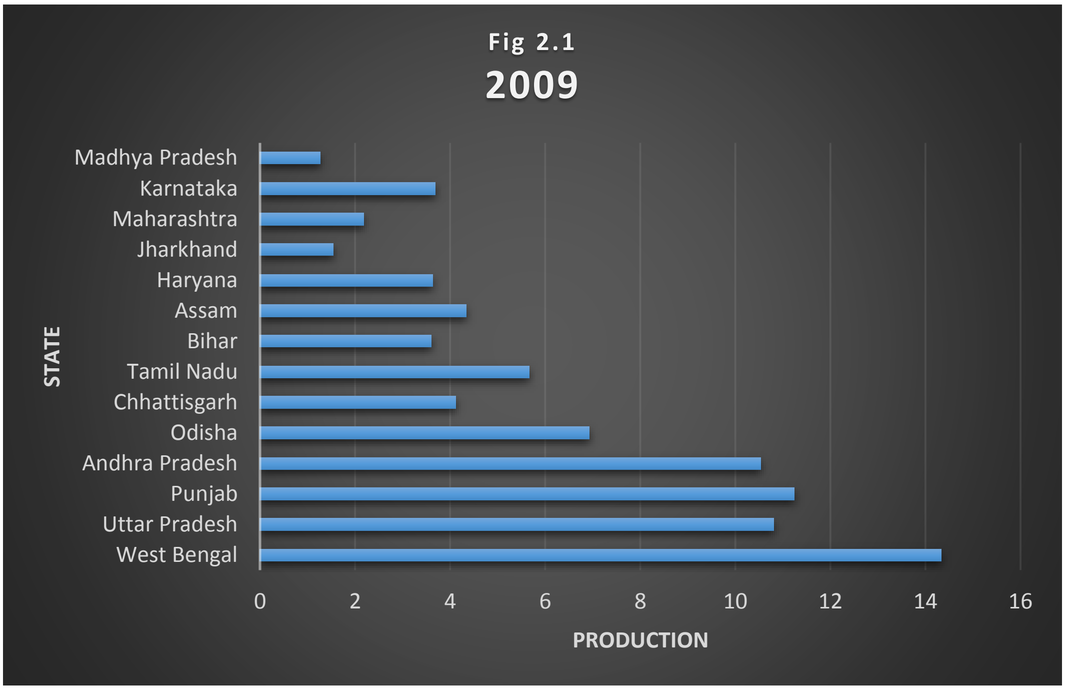


Figure 2: Bar Diagram And Pie Chart Of Rice Production In The Year 2009 For Major 15 Rice Producing States In India:-



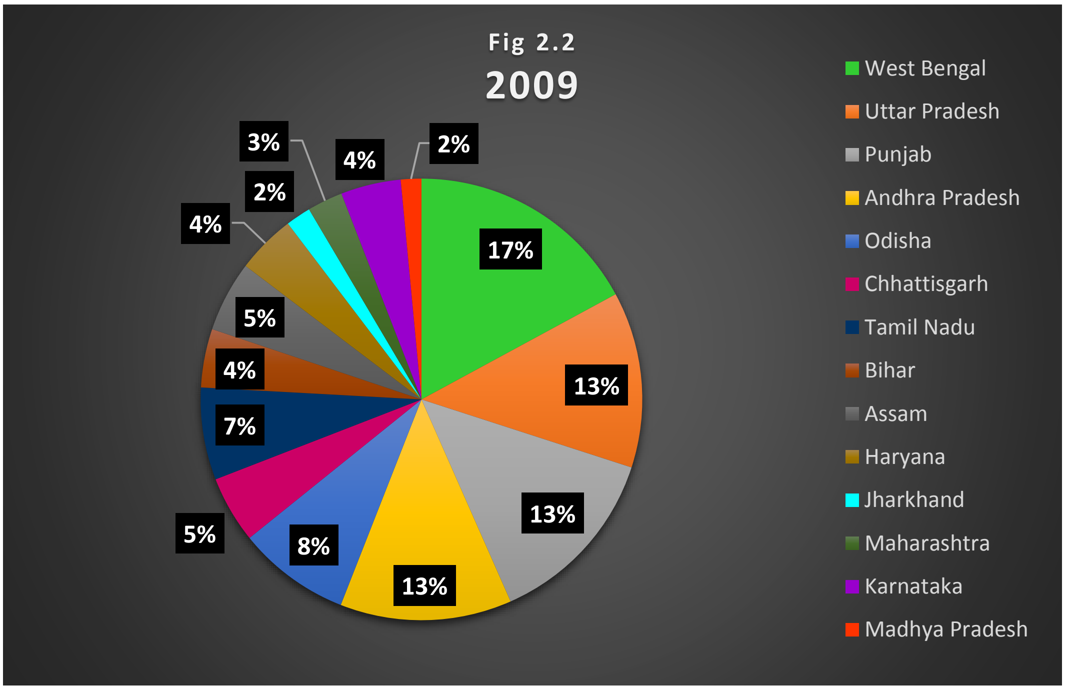
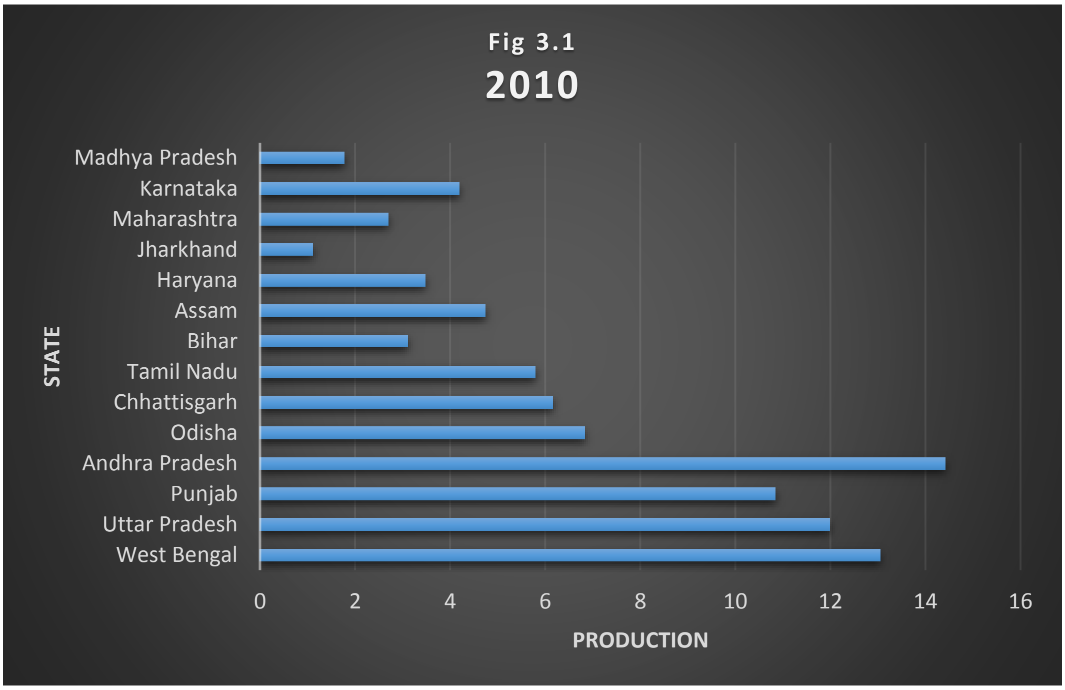


Figure 3: Bar Diagram And Pie Chart Of Rice Production In The Year 2010 For Major 15 Rice Producing States In India:-



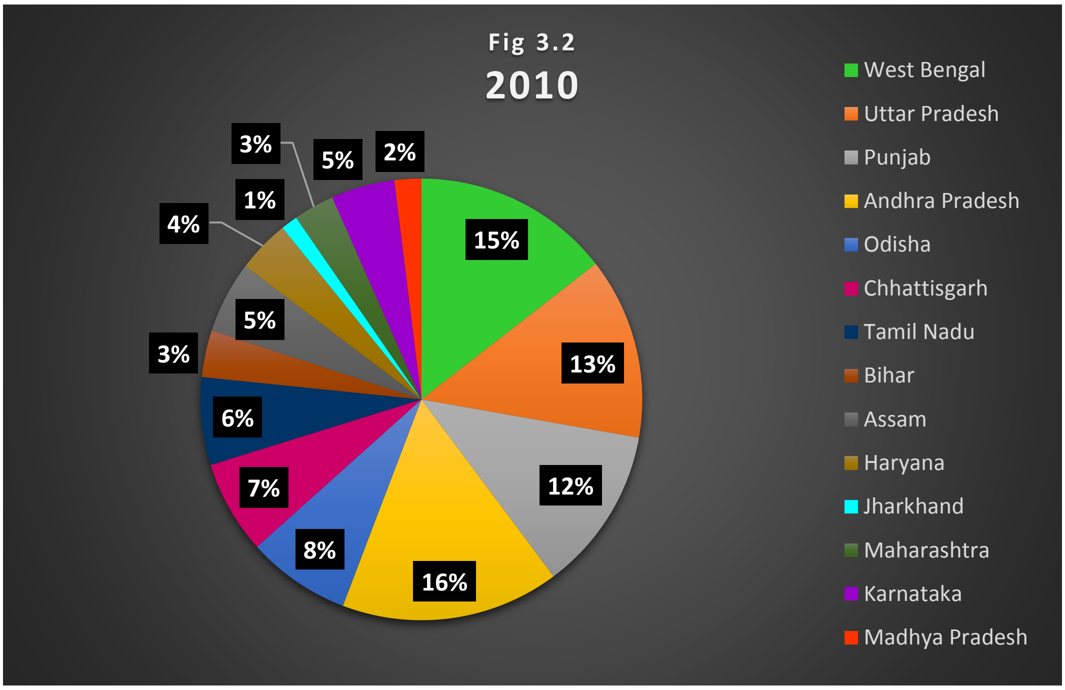
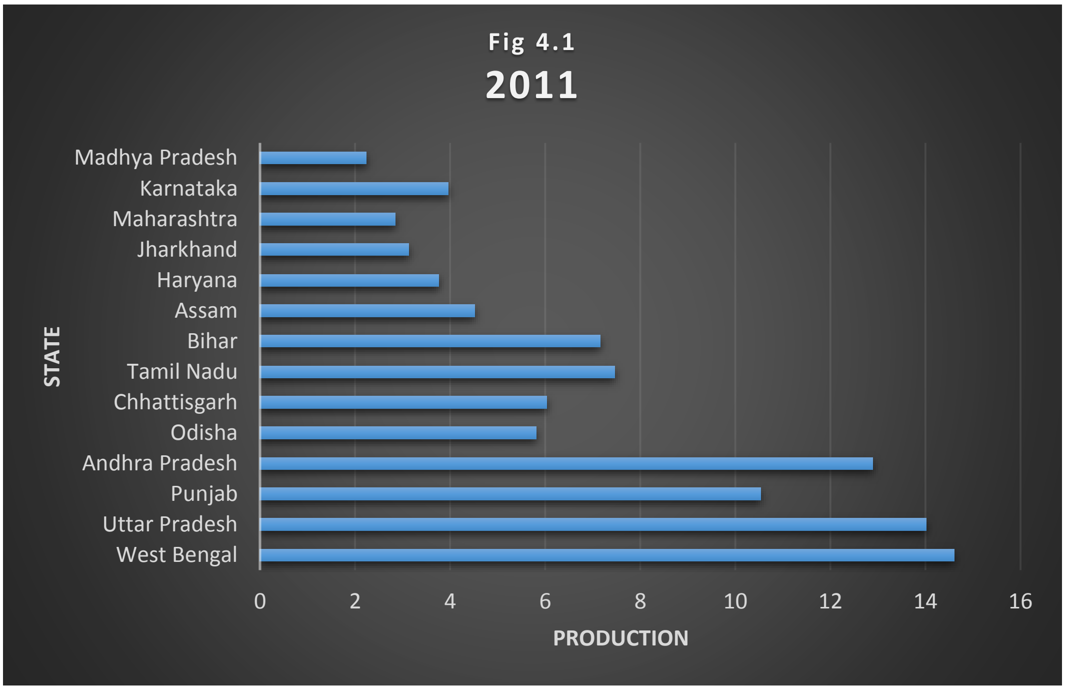


Figure 4: Bar Diagram And Pie Chart Of Rice Production In The Year 2011 For Major 15 Rice Producing States In India:-



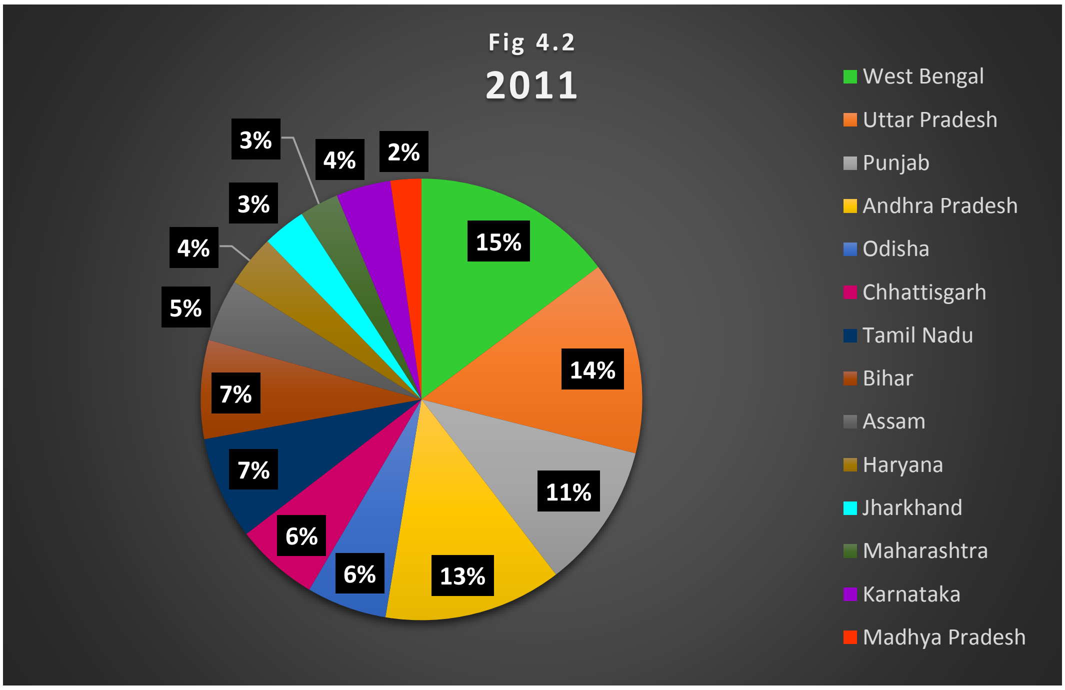
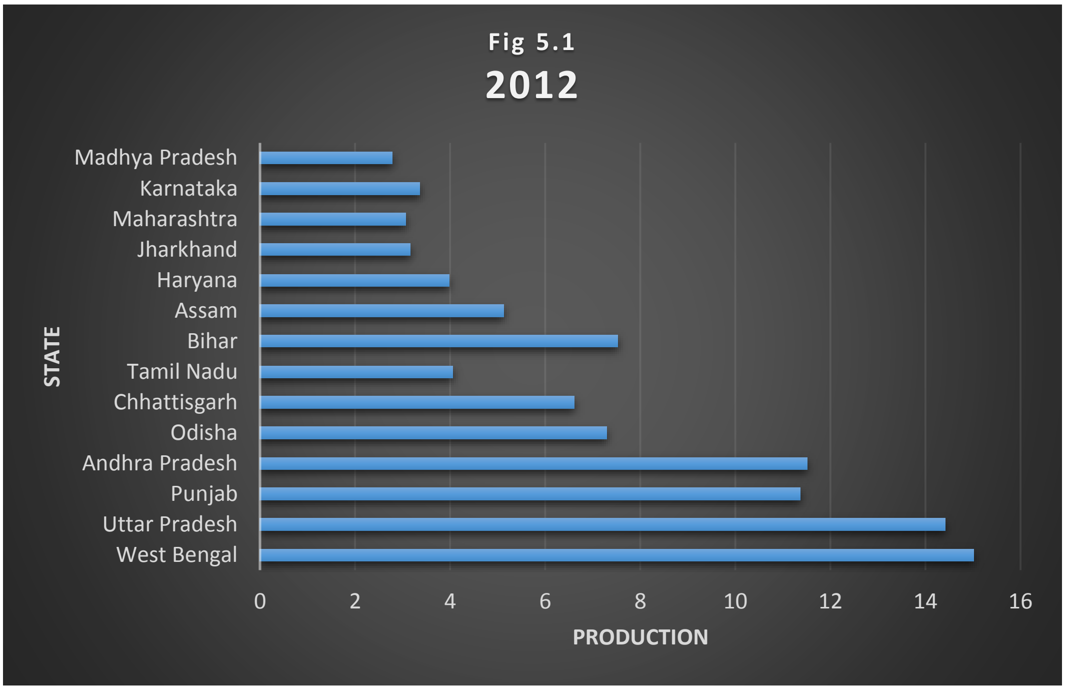


Figure 5: Bar Diagram And Pie Chart Of Rice Production In The Year 2012 For Major 15 Rice Producing States In India:-



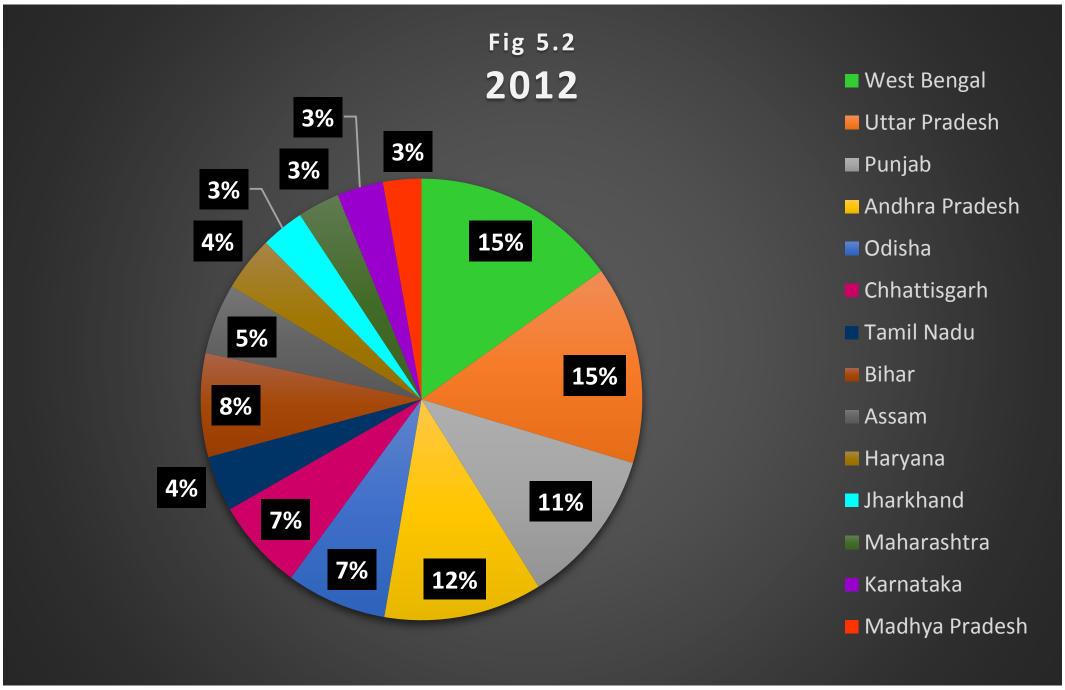
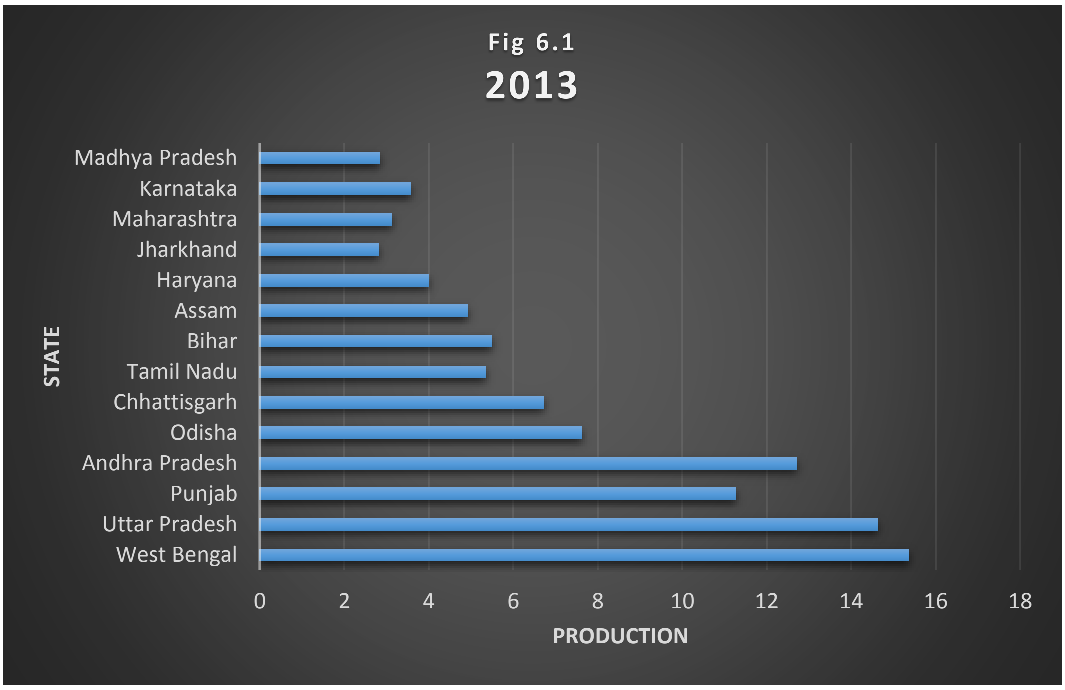


Figure 6: Bar Diagram And Pie Chart Of Rice Production In The Year 2013 For Major 15 Rice Producing States In India:-



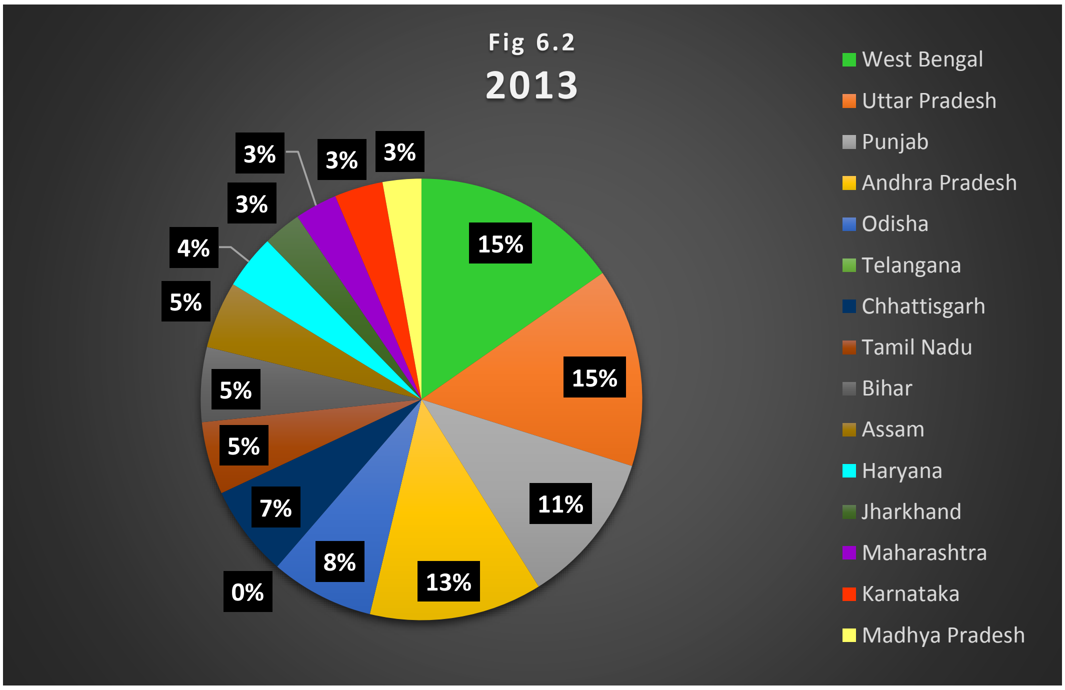
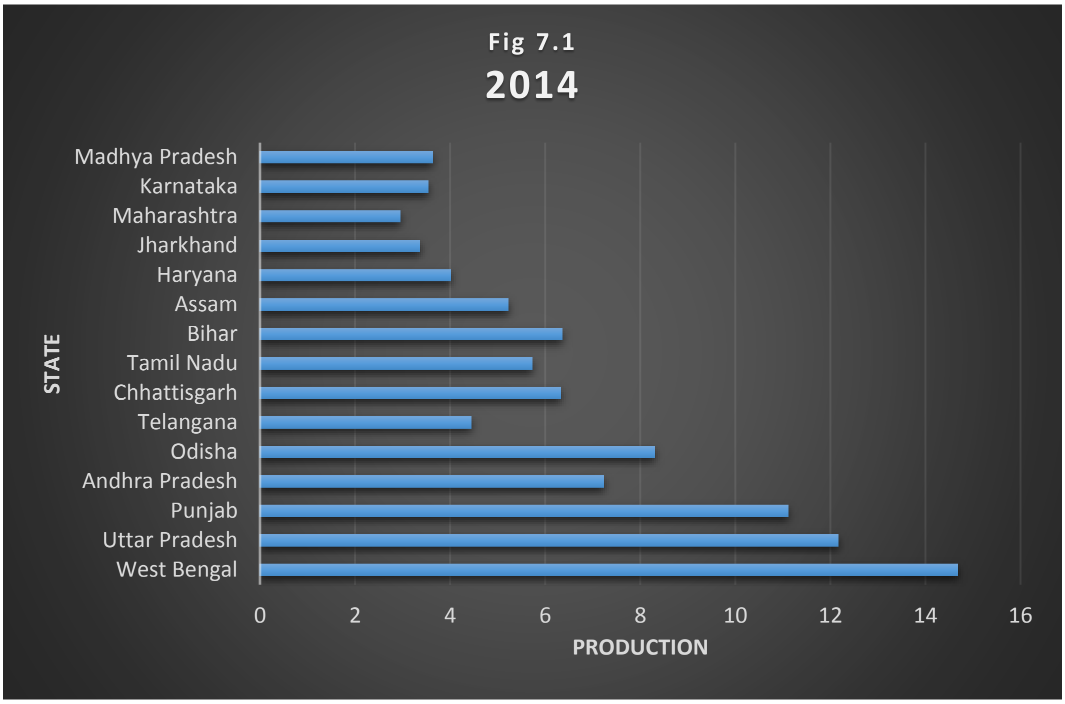


Figure 7: Bar Diagram And Pie Chart Of Rice Production In The Year 2014 For Major 15 Rice Producing States In India:-



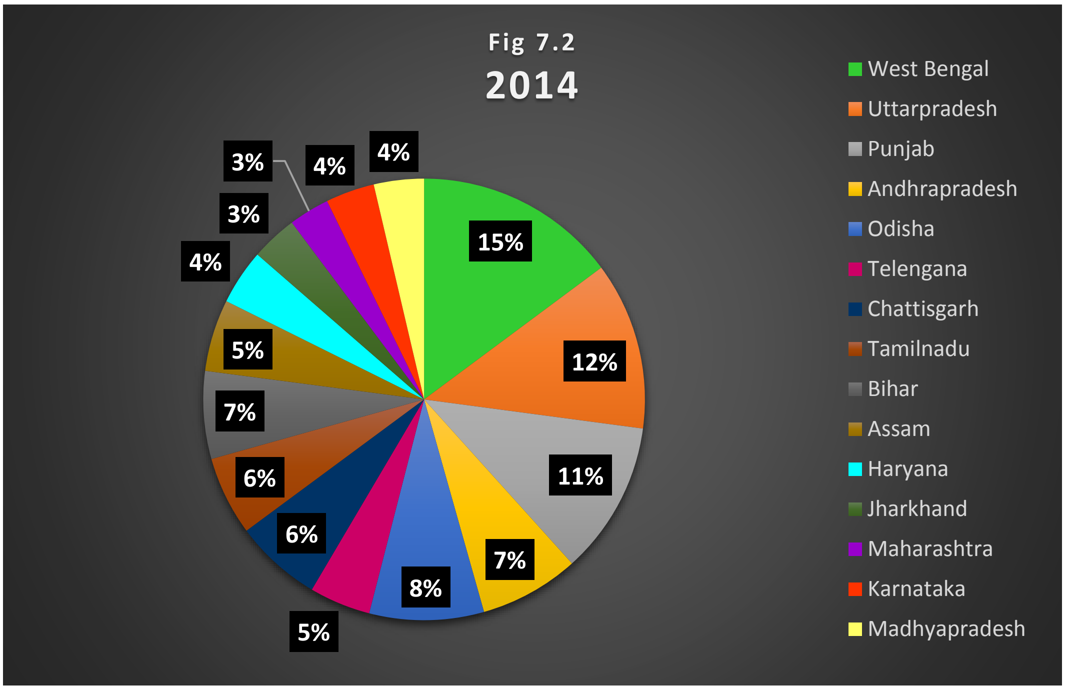
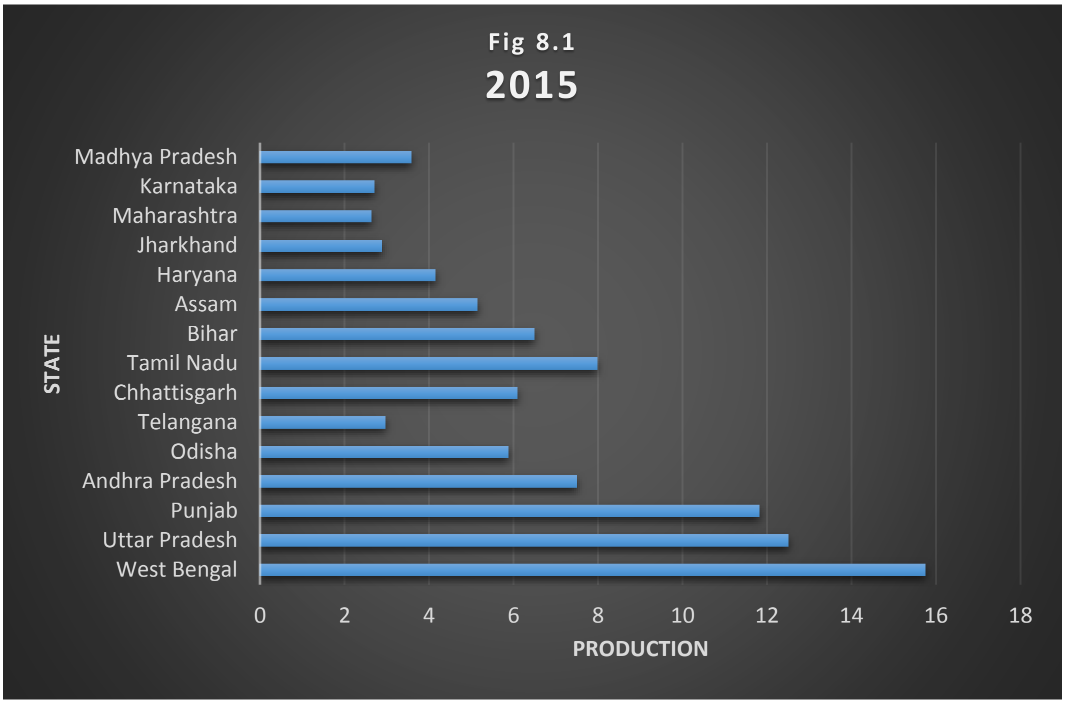


Figure 8: Bar Diagram And Pie Chart Of Rice Production In The Year 2015 For Major 15 Rice Producing States In India:-



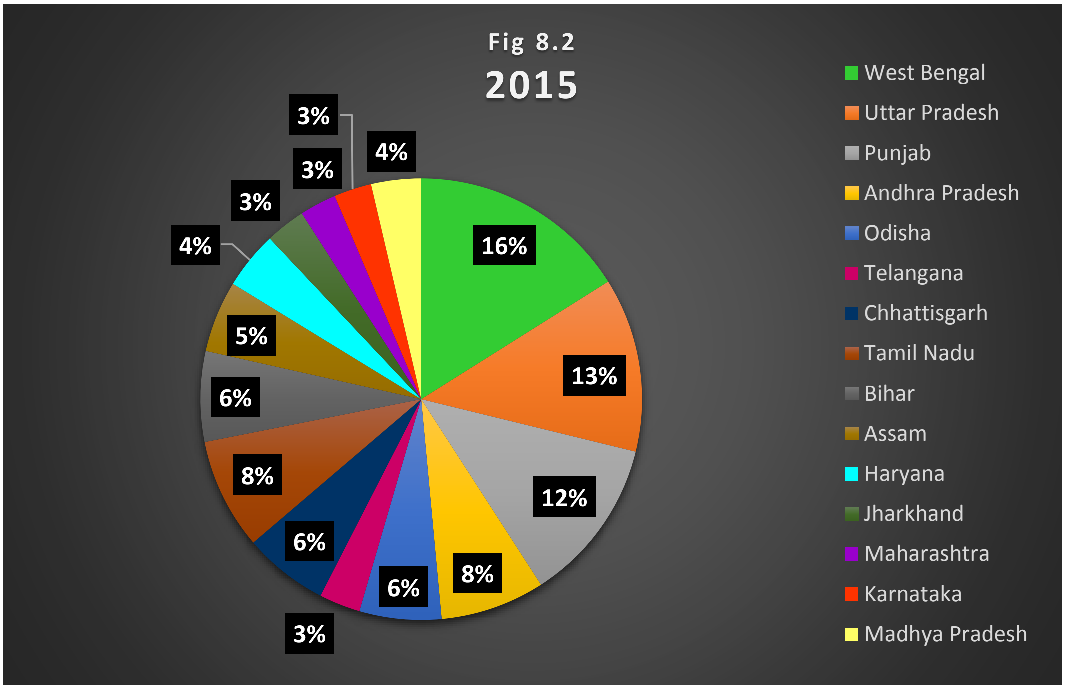
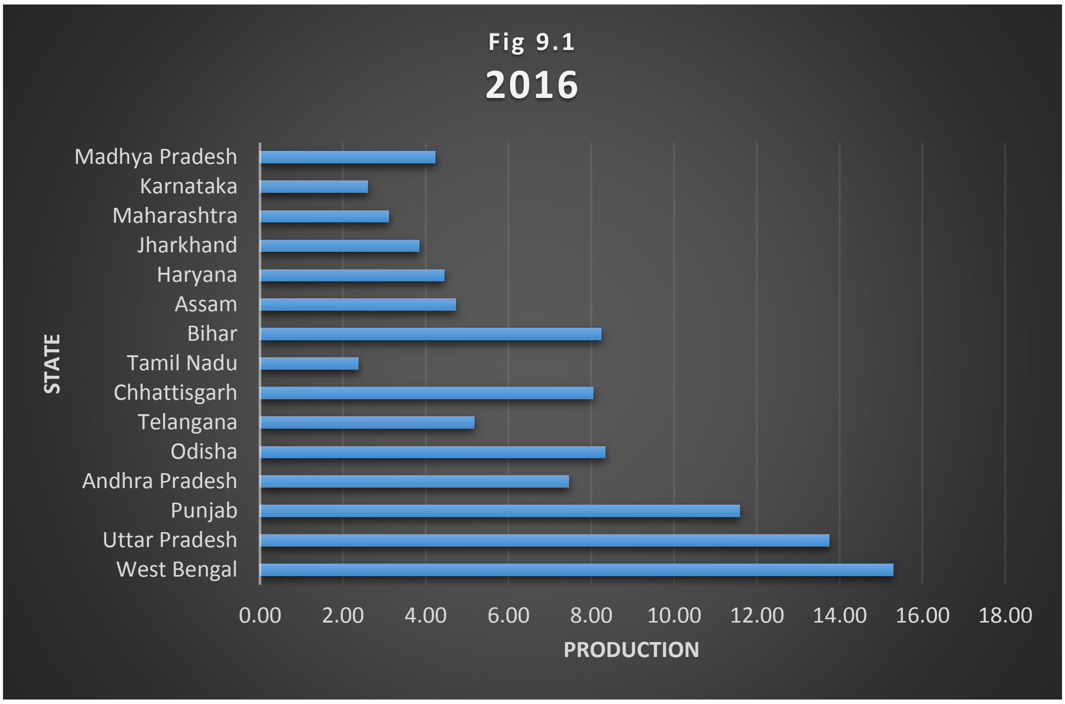


Figure 9: Bar Diagram And Pie Chart Of Rice Production In The Year 2016 For Major 15 Rice Producing States In India:-



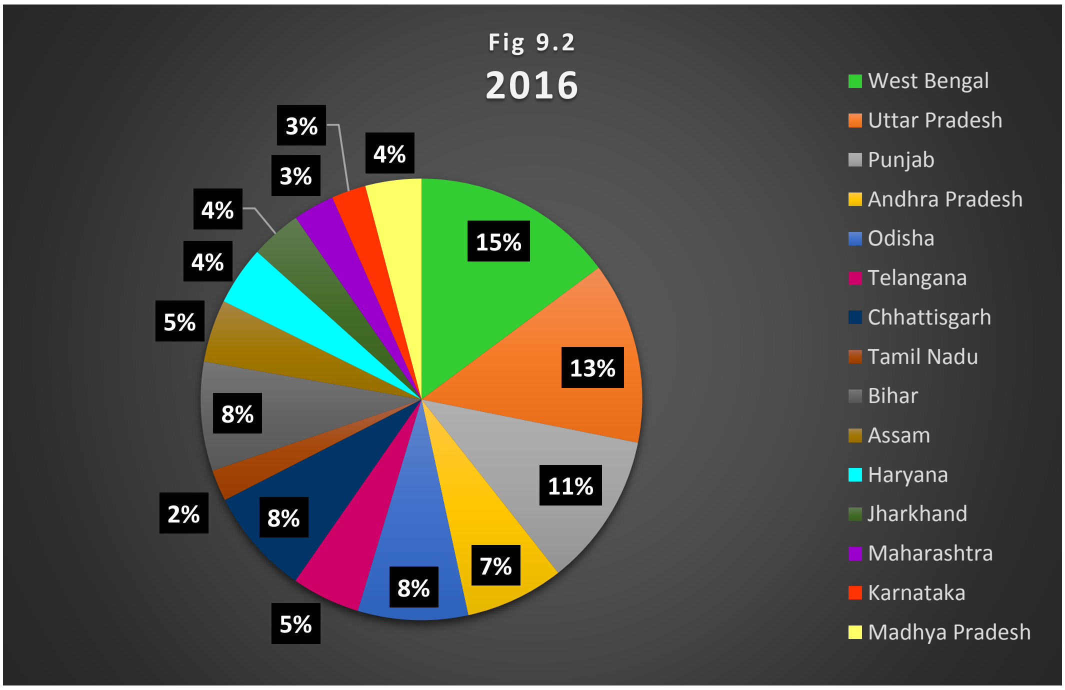
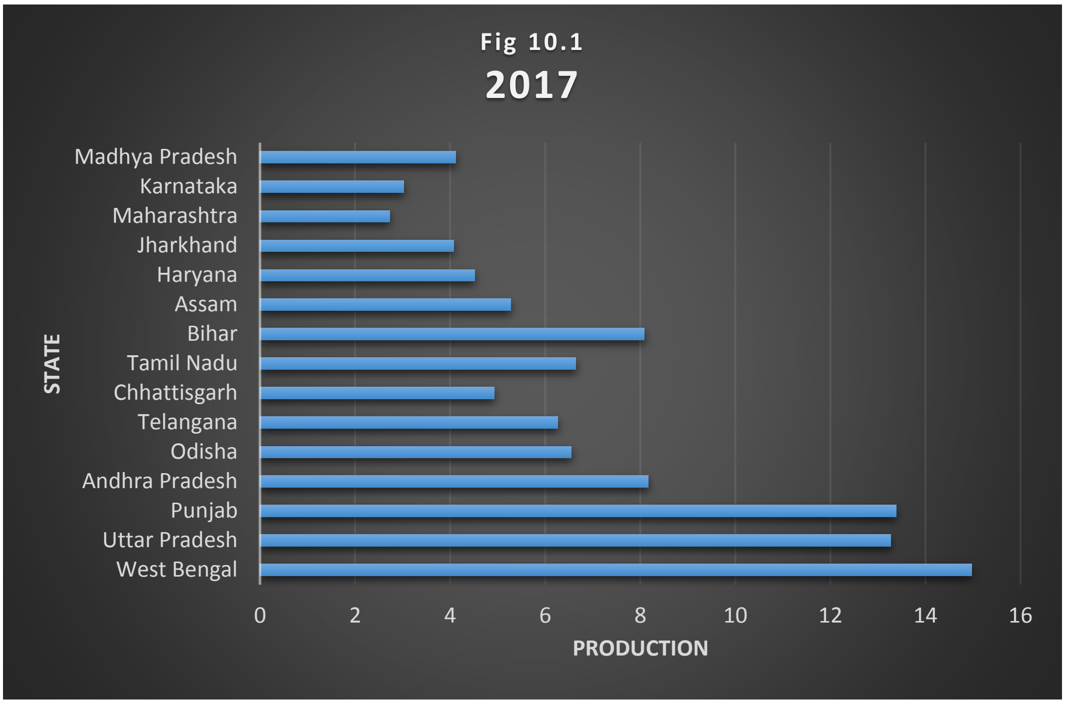


Figure 10: Bar Diagram And Pie Chart Of Rice Production In The Year 2017 For Major 15 Rice Producing States In India:-



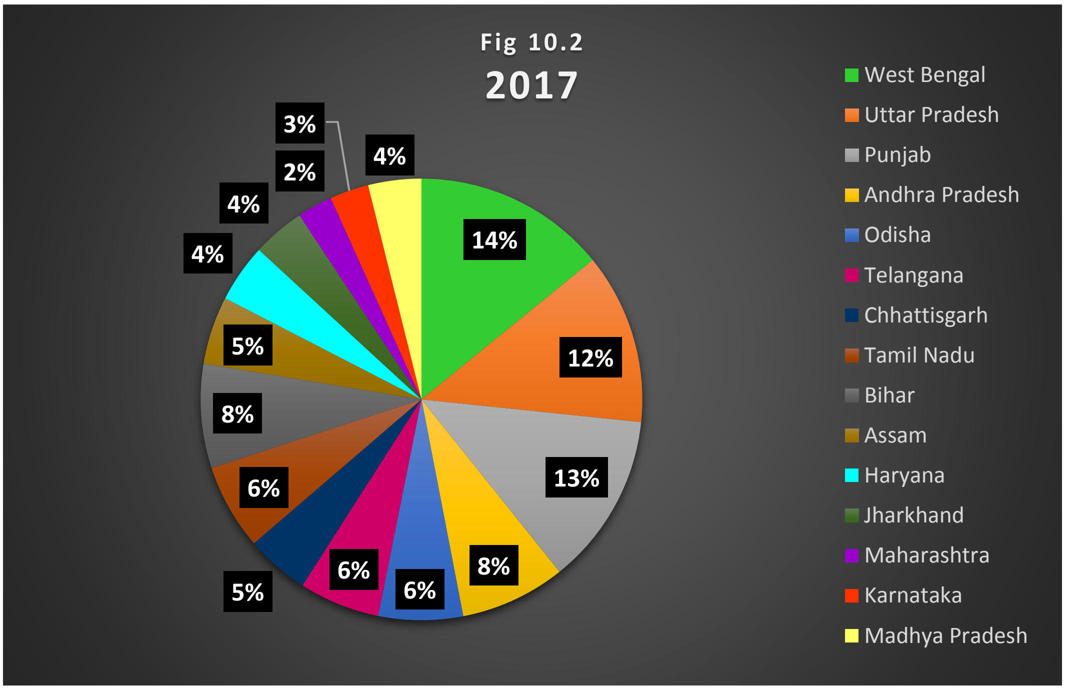
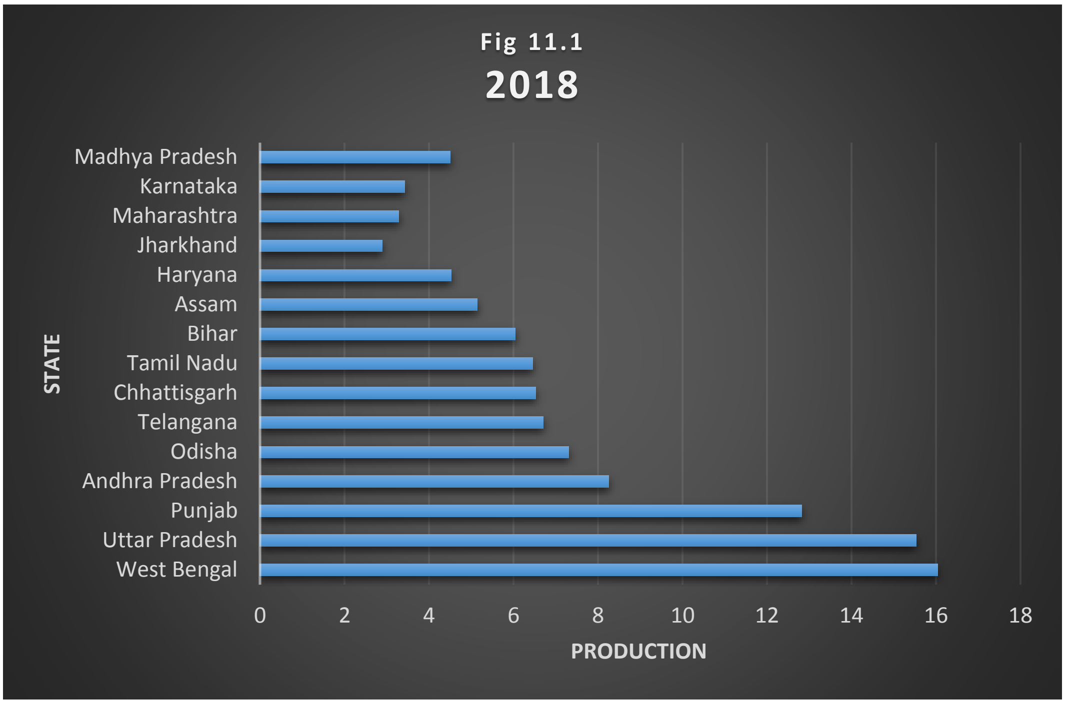


Figure 11: Bar Diagram And Pie Chart Of Rice Production In The Year 2018 For Major 15 Rice Producing States In India:-



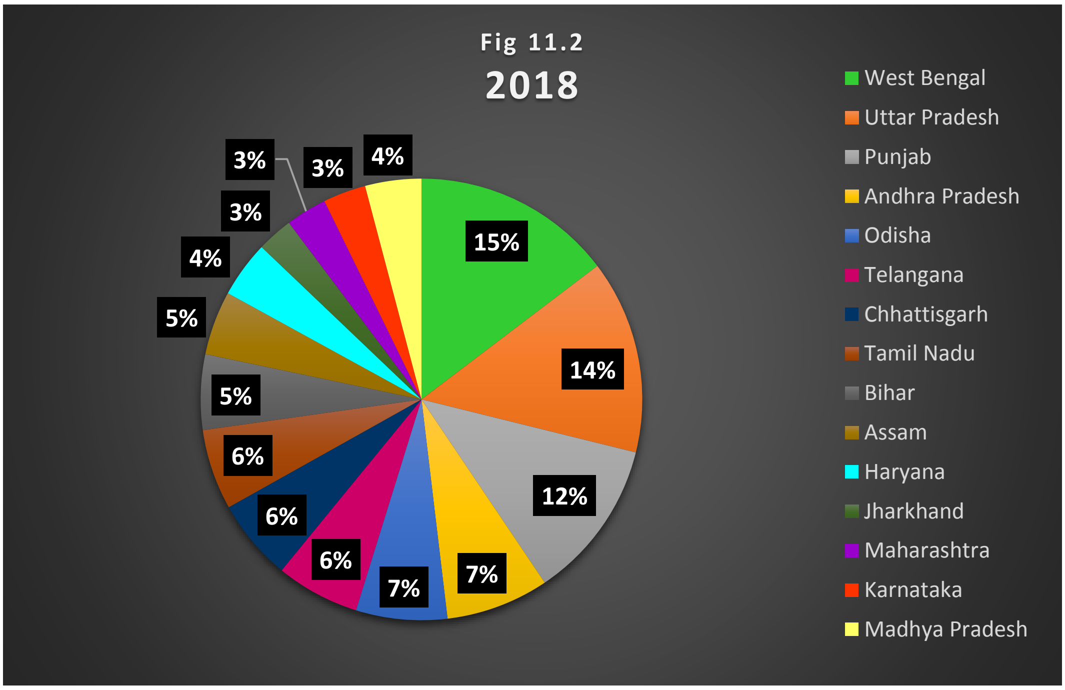
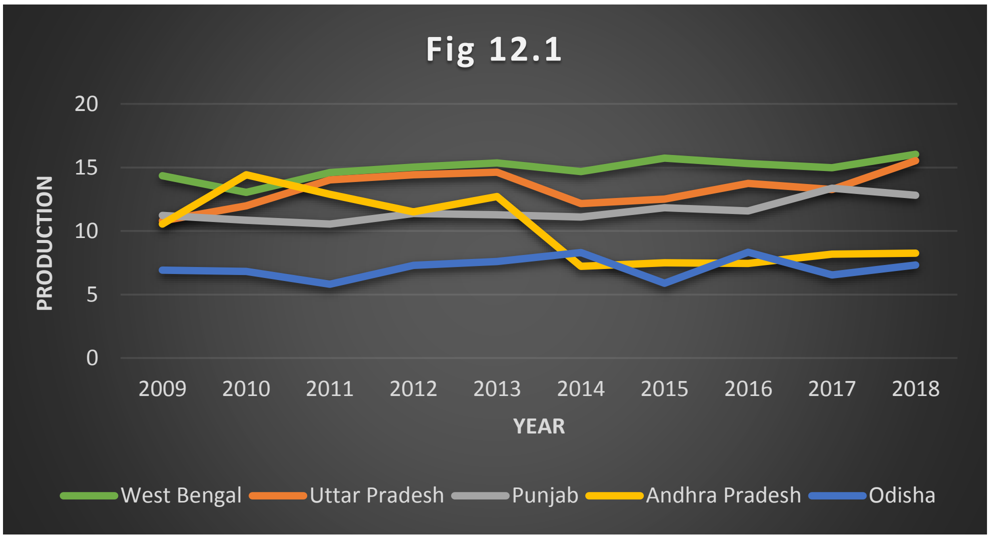
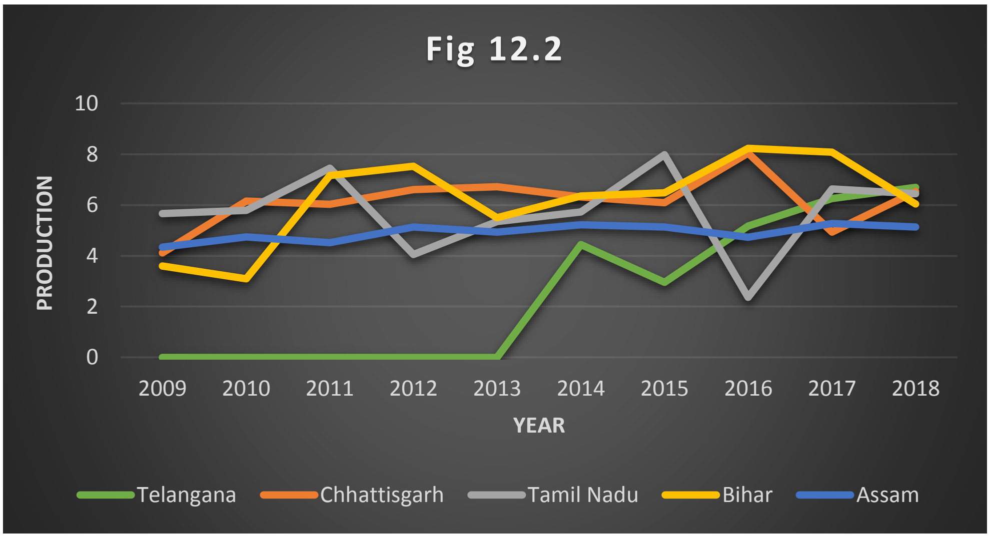
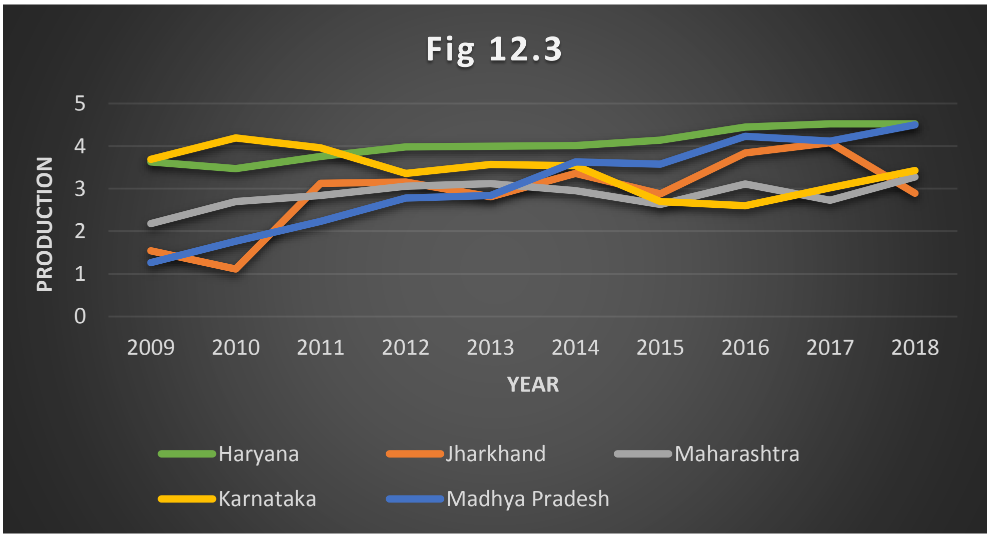


Figure 12: Multiple Line Diagrams of Rice Production For 10 Years 2009-2018 Of Major 15 Rice Producing States In India:-







In Fig 1.1 we can see that throughout the ten years period 2009-2018 the relative contribution of the major 15 states in total rice production of India has remained almost constant has not been changed significantly except for the state Andhra Pradesh. From 2014 the relative contribution of Andhra Pradesh has been decreased, it is because in 2014 the state Andhra Pradesh was divided into two states Andhra Pradesh and Telangana. So from 2014 the total rice production of Andhra Pradesh has been divided into for Andhra Pradesh and Telangana.

From the pie chart in Fig 1.2 we can see that in the last ten years (2009-2018) total West Bengal has been the most rice producing state in India and the state Uttar Pradesh, Punjab has secured 2nd and 3rd place respectively in this case. After these states it comes Odisha and Andhra Pradesh and they have almost same percentages of rice production. Then the other major states have almost same amount of contribution in total rice production in that period but less than the previous five states.

Now we are studying the annual rice production of past 10 years for the major states individually and trying to compare them:-

i. West Bengal:- From the bar diagrams and pie charts in Fig 2-11 we can see that except the year 2010 West Bengal has been the leading rice producing country in the ten years 2009-2018. From the line diagram for West Bengal in Fig 12.1 the production of rice in West Bengal in the ten years (2009-2018) mostly increasing only it has a declination which is because of the Cyclone Aila in year 2009 and slight declination in the years 2014 and 2016-17.

ii. Uttar Pradesh:- Comparing the bar diagrams and pie charts in Fig 2-11 it can be seen that 2009-10 it was the 3rd most rice producing state in India but the production was increasing and from 2011 it became the 2nd most rice producing state of India throughout the year 2018 (except the year 2017, in this year Punjab's rice production was more than UP).

The line diagram for UP in Fig 12.1 says that the rice production in UP is increasing from2009 to 2013, then there was a decline in the year 2014 but thereafter it was again increasing only had a slight declination in 2017

iii. Andhra Pradesh:- The bar diagrams and pie charts in Fig 2-11 interpret that in year 2010 Andhra Pradesh produced most rice in India among the states. After that it became 3rd most rice producing state in India up to 2013 but in 2014 Andhra Pradesh divided into two states (Andhra Pradesh & Telangana). From 2014 its position in rice production among the other states oscillates between position 4th and 5th.

The line diagram for Andhra Pradesh in Fig 12.1 says its rice production after 2010 has been declined up to 2012 but there was a major fall in 2014 for the previous reason of being divided into two states.

iv. Punjab:- Observing the bar diagrams and pie charts in Fig 2-11 it can be seen that Punjab was 4th most rice producing state in India up to 2014, after Andhra Pradesh being divided it has been 3rd most rice producing state.

From Fig 12.1 we can say the rice production of Punjab has been almost stable increasing throughout the year 2009-2018 with some slight declination for two-three years.

v. Bihar, Tamil Nadu and **Odisha:-** In the state wise rice production in India the positions of 5th, 6th and 7th most rice producing state is permutated between Bihar, Tamil Nadu and Odisha. But in 2012 & 2015 Tamil Nadu's rice production was unnaturally less for the Cyclone Nilam in 2012 and disaster of heavy rainfall and flood in South India in 2015.

The line diagram of Bihar says after 2010 Bihar has increased their rice production significantly by making progress in their agriculture though it had a declination in 2013 for flood. From line diagram of Tamil Nadu we can see there were two major fall in their rice production in 2012 and 2015, its reason has been mentioned earlier. Odisha has almost a constant amount of rice production throughout the years except of some declination in two years for natural disaster or phenomena.

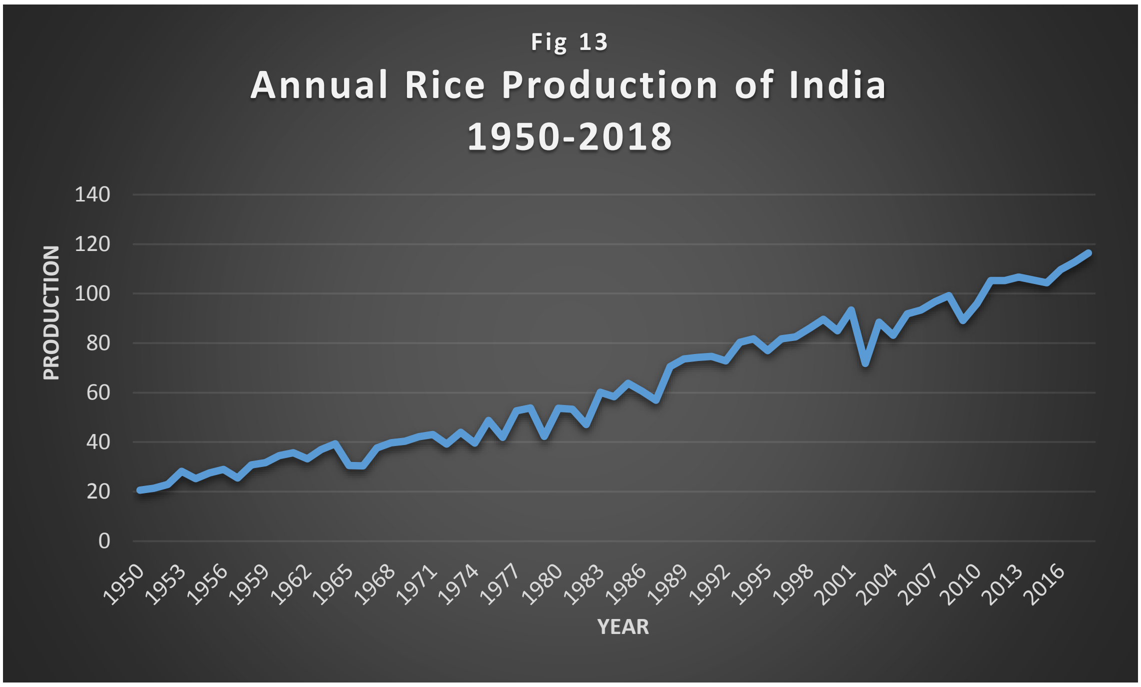
For rest of the major states **Assam's** rice production is almost stable and slightly increasing. It was the 8th most rice producing state up to the year 2015 but after **Telangana** state was created in 2014 it crossed Assam's production in 2016.

The other major states have also almost stable and increasing production of rice.

Only we see for **Chhattisgarh** in Fig 12.2 there was a major fall in production in 2017 because of the effect of drought and from the line diagram in Fig 12.3 for **Karnataka** we can see its production was almost decreasing till 2016, after that the production has been increasing.

Now we will use methods of time series analysis to analyse the trend of annual rice production of India. From the annual rice production data of India from the year 1950 to 2018 in Table 2, we visualize the time series data graphically below -

Graphical Representation of Annual Rice Production Data of India From 1950 To 2018:-



◻ Analysis Trend by Simple Moving Average Method:-

At first we will use Simple Moving Average Method for analysing the trend of the annual rice production of India. We obtain simple moving average for 3-Year, 5-Year and 10-Year period.

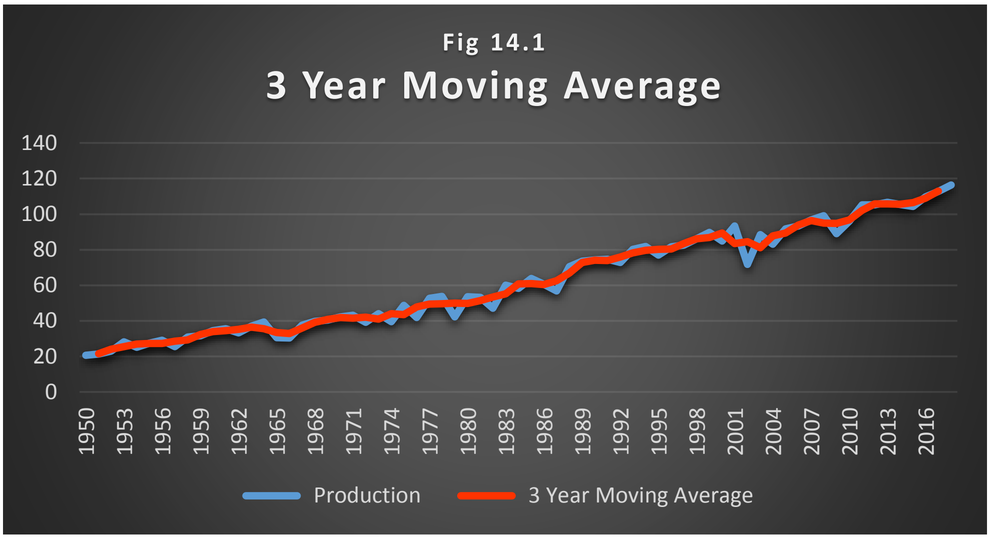
Yearly Moving Averages For Annual Rice Production Of India From 1950 To 2018:-

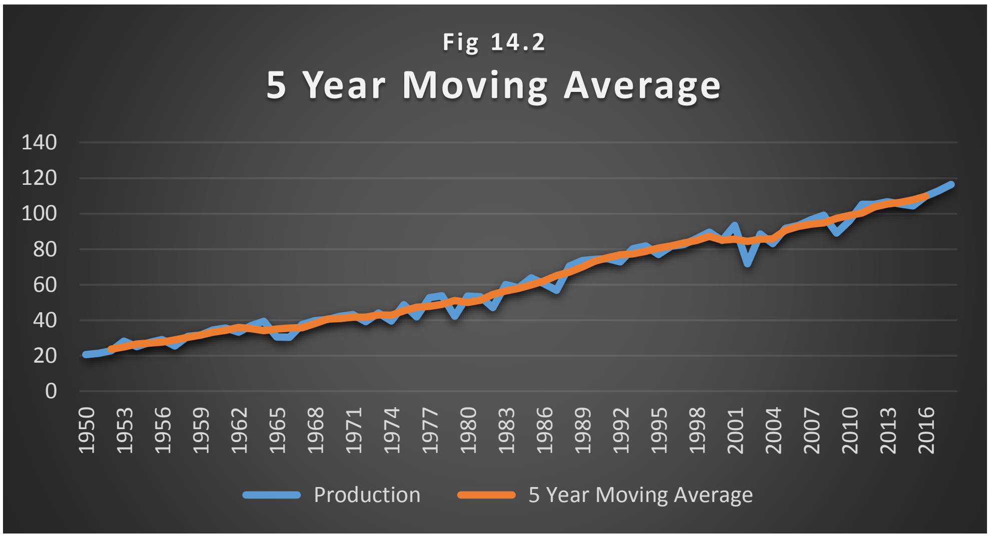
Table 3

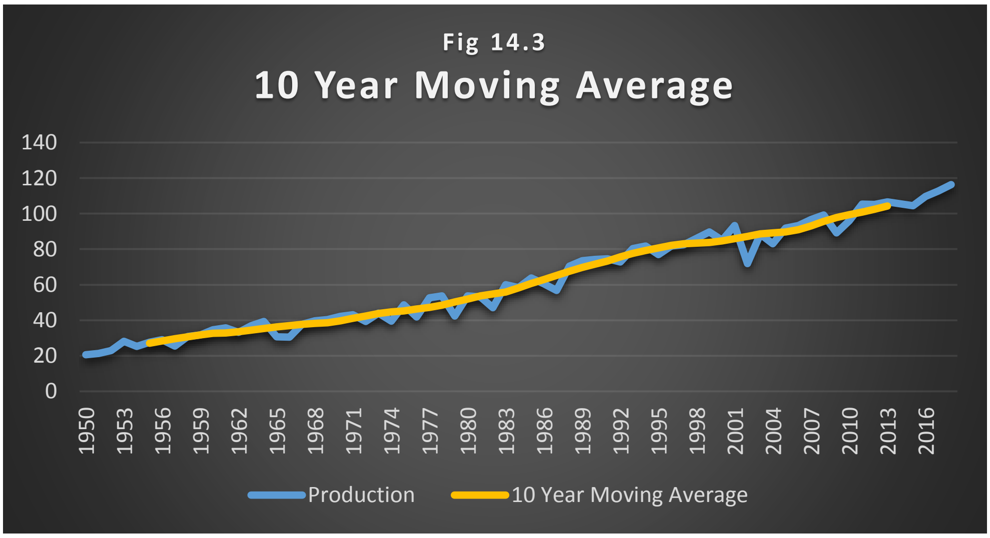
Year	Production	3 Year Moving Average	5 Year Moving Average	10 Year Moving Average
1950	20.58			
1951	21.30	21.593		
1952	22.90	24.137	23.642	
1953	28.21	25.443	25.038	
1954	25.22	26.997	26.586	
1955	27.56	27.273	27.112	26.987
1956	29.04	27.377	27.640	28.405
1957	25.53	28.473	28.932	29.639
1958	30.85	29.353	30.336	30.594
1959	31.68	32.370	31.660	31.738
1960	34.58	33.973	33.196	32.594
1961	35.66	34.483	34.426	32.815
1962	33.21	35.290	35.952	33.489
1963	37.00	36.507	35.154	34.539
1964	39.31	35.633	34.110	35.421
1965	30.59	33.447	34.990	36.241
1966	30.44	32.880	35.542	36.993
1967	37.61	35.937	35.766	37.665
1968	39.76	39.267	38.092	38.319
1969	40.43	40.803	40.618	38.685
1970	42.22	41.907	40.944	39.606
1971	43.07	41.510	41.802	41.088
1972	39.24	42.120	41.632	42.415
1973	44.05	40.957	42.936	43.868
1974	39.58	44.123	42.706	44.664
1975	48.74	43.413	45.392	45.330
1976	41.92	47.777	47.336	46.409
1977	52.67	49.453	47.886	47.312
1978	53.77	49.590	48.864	48.508
1979	42.33	49.910	51.130	50.249
1980	53.63	49.737	50.020	51.941
1981	53.25	51.333	51.286	53.628

1982	47.12	53.490	54.488	54.769
1983	60.10	55.187	56.528	55.815
1984	58.34	60.757	57.990	58.213
1985	63.83	60.910	59.938	60.808
1986	60.56	60.417	62.016	62.913
1987	56.86	62.637	65.062	65.271
1988	70.49	66.973	67.154	67.568
1989	73.57	72.783	69.978	69.752
1990	74.29	74.180	73.178	71.583
1991	74.68	73.943	75.140	73.299
1992	72.86	75.947	76.788	75.641
1993	80.30	78.323	77.326	77.704
1994	81.81	79.697	78.736	79.290
1995	76.98	80.173	80.672	80.630
1996	81.73	80.417	81.828	82.097
1997	82.54	83.450	83.402	82.978
1998	86.08	86.100	85.002	83.338
1999	89.68	86.913	87.324	83.815
2000	84.98	89.333	85.180	84.621
2001	93.34	83.380	85.670	85.944
2002	71.82	84.563	84.360	87.232
2003	88.53	81.160	85.722	88.595
2004	83.13	87.817	85.726	89.221
2005	91.79	89.427	90.700	89.741
2006	93.36	93.947	92.830	90.889
2007	96.69	96.410	94.022	93.157
2008	99.18	94.987	94.860	95.734
2009	89.09	94.750	97.248	97.758
2010	95.98	96.790	98.956	99.506
2011	105.30	102.170	100.45	100.954
2012	105.23	105.727	103.728	102.575
2013	106.65	105.787	105.414	104.240
2014	105.48	105.513	106.294	
2015	104.41	106.530	107.800	
2016	109.70	108.957	109.754	
2017	112.76	112.960		
2018	116.42			

Figure 14: Simple Moving Averages of 3 Year, 5 Year and 10 Year With Actual Production:-







The moving averages of 3 years, 5 years and 10 years have smoothed our data and make the trend clearer. The longer the period is chosen the more smooth the data. So, here in Fig 14.3, 10 year moving average makes the data smoothest as it takes more data points in calculation. Fig14.1, 14.2, 14.3 capture the natural underlying trend of our data and are giving us the general idea of the movement. The figures show us that our time series data has overall an upward trend.

□ Fitting a Mathematical Curve:-

In Fig 13 the graphical visualization of our time series data and in Fig 14 after smoothing the data by moving average we observed that the trend of our time series data i.e. is the long-term movement is entirely increasing and there is not any sign of changing its direction in downward. So for polynomial trend line, we have successively fitted Linear and Quadratic model and not any polynomial higher degree than 2 for my data. We have performed their ANOVA testing and made decision of accepting or rejecting basis on p values provided by the test. Here for the ANOVA testing is 1% level of significance is chosen. We have also observed the value of coefficient of determination i.e. the R^2 values and adjusted R^2 value in each case.

Linear Model

The fitted linear model is:-

$\mathbf{Y_t = 1.3734t + 15.9375}$ **,** with origin at **1950** and **1 year** unit of t

ANOVA

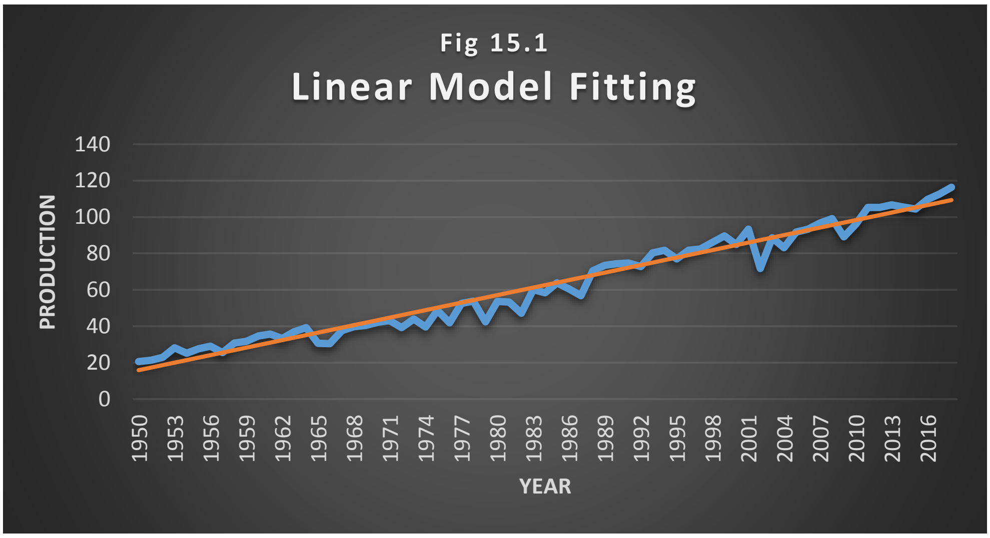
Table 4.1

ANOVA Table					
	df	*SS*	*MS*	*F*	*P-value*
Regression	1	51622.97744	51622.98	1815.063	2.94E-50
Residual	67	1905.575151	28.44142		
Total	68	53528.55259			

Since the p-value = 2.94E-50 < 0.01, we accept the linear model fitting.

Table 4.2

Regression Statistics	
Multiple R	0.982039089
R Square	0.964400772
Adjusted R Square	0.96386944
Standard Error	5.333049799
Observations	69



Quadratic Model

The fitted quadratic model is:-

$\mathbf{Y_t = 0.0069t^2 + 0.89815t + 21.244}$ **,** with origin at **1950** and **1 year** unit of t

ANOVA

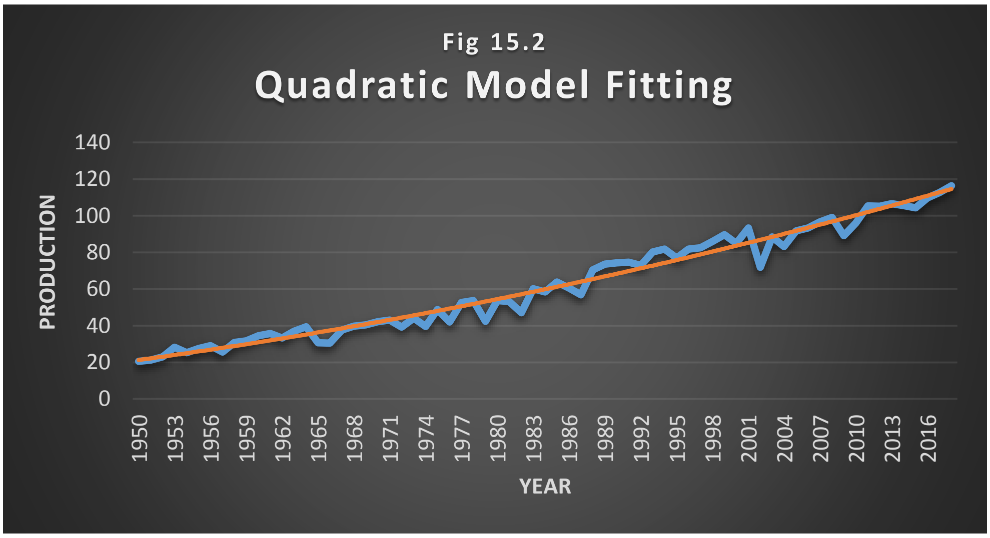
Table 5.1

ANOVA Table					
	df	*SS*	*MS*	*F*	*P-value*
Regression	2	52046.88422	26023.44	1159.198	3.9E-52
Residual	66	1481.668367	22.44952		
Total	68	53528.55259			

Since the p-value = 3.9E-52 < 0.01, we accept the quadratic model fitting.

Table 5.2

Regression Statistics	
Multiple R	0.986062897
R Square	0.972320037
Adjusted R Square	0.97148125
Standard Error	4.738092519
Observations	69



For growth curve now we are fitting exponential model and performing ANOVA testing for the linear model $\log Y_t = \log a + t \times (\log b)$ and made decision of accepting or rejecting basis on p values provided by the test. Here for the ANOVA testing is 1% level of significance is chosen.

The fitted exponential model is:-

$Y_t = 24.8458 \times 1.02423^t$, with origin at **1950** and **1 year** unit of t

<u>ANOVA:-</u>

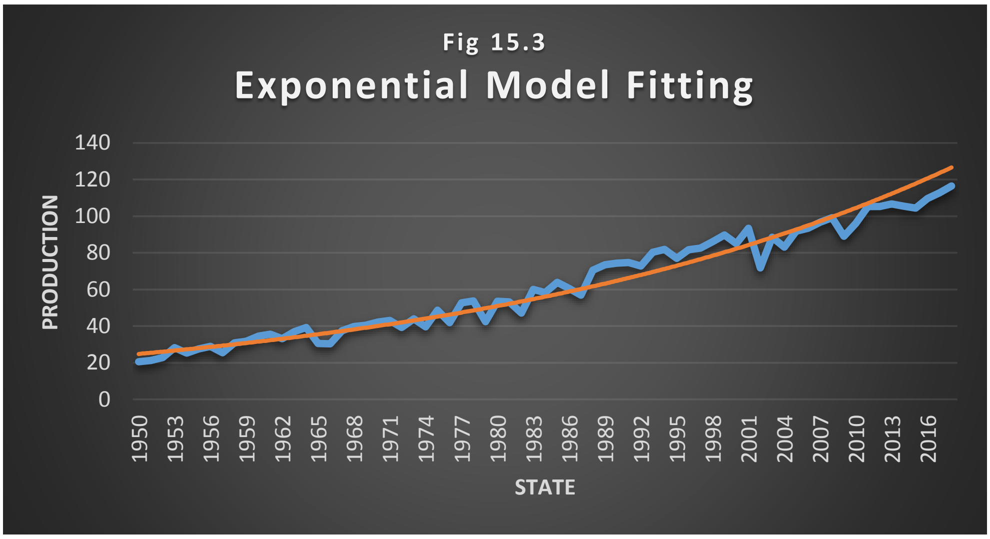
Table 6.1

ANOVA Table					
	df	*SS*	*MS*	*F*	*P-value*
Regression	1	2.958528	2.958528	1671.596	4.19E-49
Residual	67	0.118582	0.00177		
Total	68	3.07711			

Since the p-value = 4.1919E-49 < 0.01, we accept the exponential model fitting.

Table 6.2

Regression Statistics	
Multiple R	0.980542
R Square	0.961463
Adjusted R Square	0.960888
Standard Error	0.04207
Observations	69



Now our time series data the amount of production has no upper limit. From the time series plot graph in Fig 13 and after smoothing the data in Fig 14 we can see the series is increasing and the amount of growth doesn't decline by a constant percentages and approaches any upper limit. That is there is not any decrease in growth rate or increasing rate. So we need not to check the fitting of Modified Exponential model and Gompertz Model.

Now from Table 4.2, Table 5.2, Table 6.2 we can see that among linear, quadratic, exponential model; quadratic model has the largest R^2 value i.e. 0.972320037 and largest adjusted R^2 value which is 0.97148125. So we conclude that the fitted quadratic trend line equation is best for our time series data.

Now we can predict the production of rice for future years through our fitted quadratic trend model.

Prediction Of Annual Rice Production Of India For Future Years:-

Table 7

Year	t	Production (in million tonnes)
2019	69	116.0673
2020	70	117.9245
2021	71	119.7956
2022	72	121.6804
2023	73	123.5791
2024	74	125.4915
2025	75	127.4178
2026	76	129.3578
2027	77	131.3117
2028	78	133.2793
2029	79	135.2608
2030	80	137.256

• Conclusion:-

We have analysed the rice production data of major rice producing states in India for past 10 years in an explorative way. After analysing it can be concluded that the West Bengal is the most rice producing state in India followed by Uttar Pradesh, Punjab. Then the major rice producing states come are Odisha, Andhra Pradesh, Chhattisgarh, Tamil Nadu. West Bengal, Uttar Pradesh and Punjab contributes the most in the rice production in India.

From the time series trend analysis on our rice production data of whole India we can see that the production of rice in India has an upward trend.

In the trend equation part of the analysis for our data we can see that a quadratic trend is appropriate for the rice production of whole India. By the fitted quadratic trend equation we have predicted the production of rice of India for future 12 years. Observing the predicted value we conclude that the rice production will be increased in upcoming years i.e. there will be a rise in rice production for future years.

This study and analysis represents an overview of India's rice production scenario and can help in development and progression of India's rice production in future.

• Reference:-

◻ Sources of Theory:-

Fundamental of Statistics (Volume One & Two) – A.M. Goon, M.K. Gupta, B. Dasgupta

Introduction to Statistics – Prasanta Kumar Giri, Jiban Banerjee

https://www.statisticshowto.com/adjusted-r2/

https://www.statology.org/multiple-r-vs-r-squared/

https://outlier.ai/trendlines-moving-averages/

◻ Software Used:-

Microsoft Excel 2013

Minitab 18 Statistical Software

Microsoft Word 2013

• Appendix:-

◻ Sources of Data:-

Agricultural Statistics at a Glance 2018 – agricoop.gov.in

Statistical Year Book India 2018 – mospi.nic.in

https://agriexchange.apeda.gov.in/state wise rice production of India/

www.statista.com/production volume of rice India/

◻ Necessary Calculation and Analysis

https://drive.google.com/calculaion and analysis/

• Acknowledgement:-

I am indebted to so many people for helping me in the preparation of this project.

Firstly, Dr. Apurba Roy, Vice- Principal , Asutosh College, University of Calcutta, without whose help I couldn't have been a part of this prestigious college.

I owe a deep debt of gratitude to my supervisor Dr. Shirsendu Mukherjee for necessary guidance, for this presentation of this dissertation, valuable comments and suggestions. I am extremely grateful to him for the necessary stimulus, support and valuable time.

Special thanks to Smt. Ena Chatterjee, Head of the Department of Statistics, Asutosh College. I am greatly indebted to Dr. Dhiman Dutta, Dr. Sankha Bhattacharya, Dr. Parthasarathi Bera and Oindrila Bose. Faculty members often took pain and stood by me in adverse circumstances. Without their encouragement and inspiration it was not possible for me to complete this project.

Finally my earnest thanks go to my friends who were always beside me when I needed them without any excuses and made these three years worthwhile. This project is not only a mere project. It is the memories spend with the whole department which has created a mutual understanding among us. There are many emotions related to this piece of work, especially respect and duty towards teachers and vice versa, educational attachment with my friends and social attachment with my college.

• Declaration:-

I, Aniket Santra, a student of B.Sc. Sem-VI, Statistics Honours, of University of Calcutta, Registration No. 012-1111-1357-18; Roll No. 183012-21-0432 hereby declare that I have done this piece of project work entitled as ”Rice Production Of India – A Case Study and Detailed Data Analysis” under the supervision of Dr. Shirsendu Mukherjee, Assistant Professor, Department of Statistics, Asutosh College, as a part of B.Sc. Sem-VI examination according to the syllabus paper DSE-B2.

I further declare that the piece of project has not been published elsewhere for any degree or diploma or taken from any published project.