ASUTOSH COLLEGE

Estd. 1916

92, Shyama Prasad Mukherjee Rd, Kolkata - 700026

(Affiliated to **University Of Calcutta**)

NAME: AHANA BOSE

STREAM: B.Sc. HONOURS

COLLEGE ROLL NO. : 85

UNIVERSITY ROLL NO: 193012-11-0286

UNIVERSITY REGISTRATION NO: 012-1211-0535-19

SUBJECT: STATISTICS

PAPER: DSE-B2 (Project work)

SEMESTER: VI YEAR: 2022

WHEAT PRODUCTION OF INDIA- A CASE STUDY AND DETAILED DATA ANALYSIS

CONTENTS

- Executive Summary of the Project 4
- Introduction 5-6
- Methodology 6-16
  - Graphical Representation of Data
    - Pie chart
    - Line diagram
  - Time Series Analysis Method
    - Simple Moving Average Method
    - Mathematical Curve Fitting
    - Significance Test
    - Descriptive Measures
    - MAD,MAPE,MSD
- Result & Analysis 17-41
  - Studying and comparing by graphical representation of data
  - Analysis Trend by Simple moving average method
  - Fitting a Mathematical Curve
  - Prediction/ Forecasting for future years
- Conclusion 42
- References 43
- Acknowledgement 44
- Appendix 45-50
- Collected Data 51
- Declaration 52

EXECUTIVE SUMMARY OF THE PROJECT:

Wheat is a grass widely cultivated for its seeds, a cereal grain which is a worldwide staple food. Wheat occupies, in both production and area' the second important position among food crops of India and is superseded only by rice. The total area under the crop is about 29.8 million hectares in the country. The production of wheat in India has increased significantly from 75.81 million metric tonnes in 2006-07 to an all time record high of 94.88 million metric tonnes in 2011-12. 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 wheat production. Numerous subsidies, ranging from fertilizer to irrigation, electricity, seeds, machinery are available and it has led to high productivity of wheat.

In this project it has been attempted to study and analyze the wheat production of the whole India and the production of the major wheat producing states of India. The data of wheat production of major wheat producing states for past 12years (2007-2018) and the total production of India from 1960-2020 have been collected.

At first we have compared the wheat production of the major states by explanatory 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 wheat production of the whole India over time and analysis the trend. Then we have used the time series modelling to fit a suitable trend line equation for our wheat production data and tried to predict the future production of upcoming years through the estimated trend line equation.

INTRODUCTION:

Wheat is the main cereal crop in India. The total area under the crop is about 29.8 million hectares in the country. The production of wheat in the country has increased significantly from 75.81 million MT in 2006-07 to an all time record high of 94.88 million MT in 2011-12. The productivity of wheat which was 2602 kg/hectare in 2004-05 has increased to 3140 kg/hectare in 2011-12. The major increase in the productivity of wheat has been observed in the states of Haryana, Punjab and Uttar Pradesh. Higher area coverage is reported from MP in recent years. Although there is huge variation in climatic conditions, Wheat is grown in all parts of India. Geographically India has been divided into six Wheat growing zones, namely:

1. Northern Hills Zone (NHZ)- The northern hills zone comprises of Western Himalayan regions of J&K (except Jammu and Kathua dist.), Himachal Pradesh (except Una and Paonta valley), Uttaranchal (except Tarai area), Sikkim and hills of West Bengal and N.E. States. This Zone has wheat growing area of 0.8 mha which is predominantly rainfed. Average productivity of wheat in this zone is 16.64 q/ha. The major constraints in wheat production in this zone is water stress, yellow and brown rusts, late sowing and low plant population.

2. North Western Plain Zone (NWPZ)- This zone comprises of Punjab, Haryana, Delhi, Rajasthan (except Kota and Udaipur divisions) and Western U.P. (except Jhansi division) Una district and Paonta valley of HP and Tarai region of Uttaranchal Pradesh. This zone has wheat growing area of about 9.5 million hectares. The average productivity of wheat in this zone is 39.4 q/ha. The major wheat production constraints of this zone are weed infestation particularly Phalaris minor and wild oat, Yellow and brown rusts, Karnal bunt, powdery mildew, foliar blight and termites.

3. North Eastern Plain Zone (NEPZ) - This zone comprises of Eastern Uttar Pradesh, Bihar, Jharkhand, West Bengal, Orissa, Assam, Sikkim and plains of far eastern states under irrigated conditions. This zone has wheat growing area of about 9.5 million hectares and average productivity of wheat in this zone is 25.1 q/ha. This zone has the second highest share in total wheat production of India. Major constraints of this zone are leaf blight & brown rust, delayed sowings, lack of seeds of improved varieties, weed like bathua; wild oat and phalaris minor, low plant population and zinc deficiency.

4. Central Zone (CZ) - This zone comprises of Gujarat, Madhya Pradesh, Chattisgarh, Jhansi division of UP and Kota and Udaipur division of Rajasthan. This zone has wheat growing area of about 4.5 million hectares and average productivity of wheat in this zone is 24.1 q/ha. Major constraints of this zone are leaf & stem rust, termites, rodents and pervalant drought conditions.

5. Peninsular Zone (PZ) - This zone comprises of Southern states of Maharashtra, Andhra Pradesh, Karnataka, Goa and plains of Tamil Nadu. This zone has wheat growing area of about 1.5 million hectares and average productivity of wheat in this zone is 29.8 q/ha. Besides bread and durum wheat, dicoccum wheat is also popularly grown in this zone. Major constraints of this zone are leaf and brown rust, attack of aphid, grain discoloration and water stress.

6. Southern Hills Zone (SHZ) - This zone comprises of hills of Tamil Nadu and Kerala comprising the Nilgiri and Palni hills of southern plateau. This zone has wheat growing area of about 0.2 million hectares and average productivity of wheat in this zone is 10 q/ha. Major constraints of this zone are attack of termites followed by lodging, attack of birds, delayed sowing and black rust.

❖ *METHODOLOGY:-*

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 tea of a company, productions of a crop) may be specified for individual points of time or for different periods 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 are 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 midpoints 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 more 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 different patterns 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 of countries last 15 years, say, we may draw multiple line diagrams.

- *PIE CHART-*

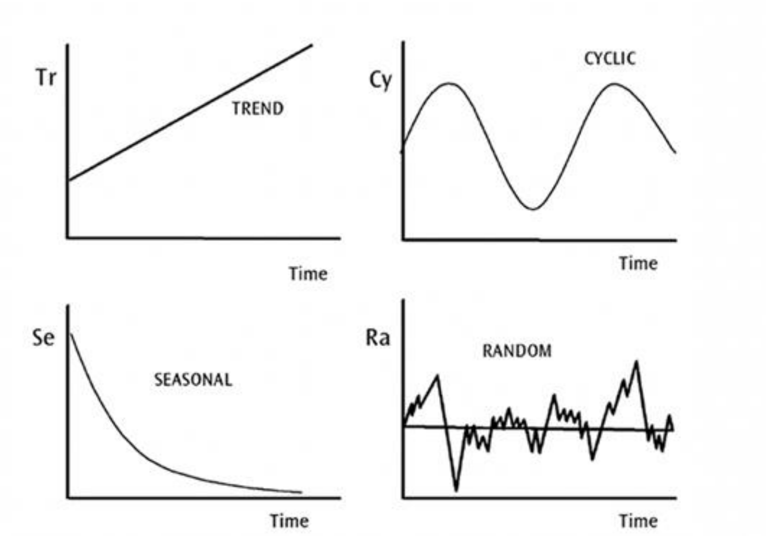
A pie chart is an appropriate diagram used for exhibiting the relative sizes of the different parts of a whole. In this case, a circle is partitioned into several sectors by drawing angles at the centre, the area of each sector indicating the corresponding percentage. In fact, the area enclosed by the circle is regarded as 100. Since, total area at the centre is 360 degrees, 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 pie chart, we can compare the relative contributions or importance of different categories to the study variable.

METHOD OF TIME SERIES ANALYSIS:

- *TIME SERIES ANALYSIS-*

Time series can be defined as an ordered sequence of data points spread across a period. Time, in this case, is usually an independent variable, whereas the other variables involved keep on changing the values. Time series data is monitored over a period of time with constant temporal intervals. Time Series Analysis is the process of identifying common patterns in the datasets over a specified time. We can classify these patterns as cyclic patterns, seasonal patterns, random patterns, and trends.



Time Series Models –

The various ways to model time series data can be classified as follows –

The first is the **Moving Average, MA model,** and it is the basic model used for univariate time series. The output variable in the MA model has a linear dependence on the current or the past values. A new series is developed based on the average of the current/past values. Trend spotting is easy using the moving average time series model.

Exponential Smoothing, ES is the model that also uses univariate time series. The values, however, in the case are determined from the weighted average of the past values. Trends and seasonality can be identified using the ES model.

Next is the **Auto-regressive model**. In an auto-regression model, we forecast the variable of interest using a linear combination of *past values of the variable*. The term *auto*-regression indicates that it is a regression of the variable against itself.

In this project we only work with the Moving Average model.

▪ *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 two-item 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

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

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,

$y_i = a_0 + a_1t_{1i} + a_2t_{2i} + a_3t_{3i} + ... + a_pt_{pi} + e_i$; $i = 1(1)n$ where e_i ~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,

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

vs.

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

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

$$SSE = \min_{a_i} \sum_{i=1}^{n} (y_i - a_0 - a_1t_{1i} - a_2t_{2i} - a_3t_{3i} + \cdots - a_pt_{pi})^2$$

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

$y_i = a_0 + a_1\acute{t}_{1i} + a_2\acute{t}_{2i} + a_3\acute{t}_{3i} + ... + a_p\acute{t}_{pi} + e_i$

where, $\acute{t}_{ji} = t_{ji} - \bar{t}_j \ \forall \ j = 1,2,…,p$

$$i.e., SSE = \sum_{i=1}^{n} (y_i - a_0 - a_1\acute{t}_{1i} - a_2\acute{t}_{2i} - a_3\acute{t}_{3i} - \cdots - a_p\acute{t}_{pi})^2$$

$$\therefore \frac{\partial}{\partial a_0} \sum_{i=1}^{n} (y_i - a_0 - a_1\acute{t}_{1i} - a_2\acute{t}_{2i} - a_3\acute{t}_{3i} - \cdots - a_p\acute{t}_{pi})^2 = 0$$

$$\sum_{i=1}^{n} (y_i - a_0 - a_1\acute{t}_{1i} - a_2\acute{t}_{2i} - a_3\acute{t}_{3i} - \cdots - a_p\acute{t}_{pi}) = 0$$

$$\sum_{i=1}^{n} y_i = na_0 + a_1 \sum_{i=1}^{n} \acute{t}_{1i} + a_2 \sum_{i=1}^{n} \acute{t}_{2i} + \cdots + a_p \sum_{i=1}^{n} \acute{t}_{pi}$$

$$Now, \sum_{i=1}^{n} \acute{t}_{ji} = \sum_{i=1}^{n} (t_{ji} - \bar{t}_j) = 0, \forall \ j = 1(1)p$$

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

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

$$\frac{\partial}{\partial a_j}\sum_{i=1}^{n}(\mathrm{y_i} - \mathrm{a_0} - \mathrm{a_1 t_{1i}} - \mathrm{a_2 t_{2i}} - \mathrm{a_3 t_{3i}} - \cdots - \mathrm{a_p t_{pi}})^2 = 0$$

$$\Rightarrow \sum_{i=1}^{n}(\mathrm{y_i} - \mathrm{a_0} - \mathrm{a_1 t_{1i}} + \mathrm{a_2 t_{2i}} - \mathrm{a_3 t_{3i}} - \cdots - \mathrm{a_p t_{pi}})t_{ji} = 0$$

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

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

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

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 < α 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.

- Mean- It is the most common measure of location (measure of the centre of a batch of a numbers). It is defined by-

$$\bar{x} = \frac{\sum x_i}{n}$$

Where
$\bar{x}$ is the mean
n is the number of observations and
$\sum x_i$ is the sum of all the observations

- Variance- It is the most common measure of dispersion (scatterness). It is defined as-

$$\sigma^2 = \sum (X_i - \bar{X})^2 / N$$

σ^2 = variance
X_i = the value of the ith element
$\bar{X}$ = the mean of X
N = the number of elements

- Median- Median is a statistical measure that determines the middle value of a dataset listed in ascending order (i.e., from smallest to largest value).

$$\mathrm{Med}(X) = \begin{cases} X[\frac{n}{2}] & \text{if n is even} \\ \frac{(X[\frac{n-1}{2}]+X[\frac{n+1}{2}])}{2} & \text{if n is odd} \end{cases}$$

- Mode- Mode or modal value is the value or number in a data set, which has a high frequency or appears more frequently.

- Skewness- Skewness is a measure of asymmetry or distortion of symmetric distribution. It measures the deviation of the given distribution of a random variable from a symmetric distribution, such as normal distribution.

$$\tilde{\mu}_3 = \frac{\sum_i^N \left(X_i - \bar{X}\right)^3}{(N-1) * \sigma^3}$$

$\tilde{\mu}_3$ = skewness
N = number of data points in the distribution
X_i = random variable
$\bar{X}$ = mean of the distribution
σ = standard deviation

- Kurtosis- Kurtosis is a measure of the combined weight of a distribution's tails relative to the centre of the distribution.

$$\text{kurtosis} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4/n}{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2/n\right]^2}$$

- Mean Absolute Deviation (MAD) - Mean absolute deviation (MAD) of a data set is the average distance between each data value and the mean. Mean absolute deviation is a way to describe variation in a data set. It helps us to get a sense of hoe 'spread out' the values in a data set are.

- Mean Absolute Percentage Error (MAPE) –The mean absolute percentage error (MAPE) is the mean or average of the absolute percentage errors of forecasts. Error is defined as actual or observed value minus the forecasted value. Percentage errors are summed up without regard to sign to compute MAPE.

- **Mean Square Deviation (MSD)** - The mean square deviation (MSD) measures the accuracy of the fitted time series values. Outliers have a greater effect on MSD than on MAD.

The measures *MSD*, *MAD* and *MAPE*:

$MSD = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2$	*Mean Squared Deviation* Comparable with *MSE* in regression models, but its value has another scale than the observations
$MAD = \frac{1}{n}\sum_{t=1}^{n}\lvert y_t - \hat{y}_t\rvert$	*Mean Absolute Deviation* Comparable with the square root of *MSE*, but less sensible to outliers. Has the same scale as the observations.
$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left\lvert\frac{y_t - \hat{y}_t}{y_t}\right\rvert \cdot 100$	*Mean Absolute Percentage Error* Expresses the mean absolute deviation in percentages of the level. Suitable for multiplicative models.

n is the number of time points, where both the *original* observation y_t and the *predicted* observation $\hat{y}_t$ exist

The smaller the MAD, MSD AND MAPE, the better the forecast. A model having smaller values of MAD, MSD and MAPE is a better fit to the data.

❖ANALYSIS AND RESULTS :-

First we intend to study the state wise wheat production of India. So the major wheat producing states are selected and their annual wheat production for past 12years are colleced. Then we analyse the total annual production of wheat in whole India by using appropriate time series analysis method.

- *Studying and comparing data using their graphical representation:-*

For studying and comparing the wheat productions of different states we use Line Diagrams and Pie Charts.

The major wheat producing states of India are Punjab, Uttar Pradesh, Haryana, Madhya Pradesh, Maharashtra, Himachal Pradesh, Gujarat, Rajasthan, Assam, Bihar, Meghalaya, Uttarakhand, Delhi, West Bengal, Jharkhand, Jammu & Kashmir, Odisha, Karnataka, Chattisgarh. Now we want to compare their production and contribution in India's total wheat production.

At first we draw multiple bar diagram and a pie chart of wheat production in 12 years from 2006-2017 for 19 major wheat producing states. Then for each we draw pie charts of wheat production for those states to compare their production. Lastly we use the line diagram for each of the 19 states' production in the specified 12 years time period to study their production and productions' movement over the time.

- ***PIE CHARTS OF WHEAT PRODUCTION IN THE YEAR 2006 & 2007 FOR MAJOR 19 WHEAT PRODUCING STATES IN INDIA:***

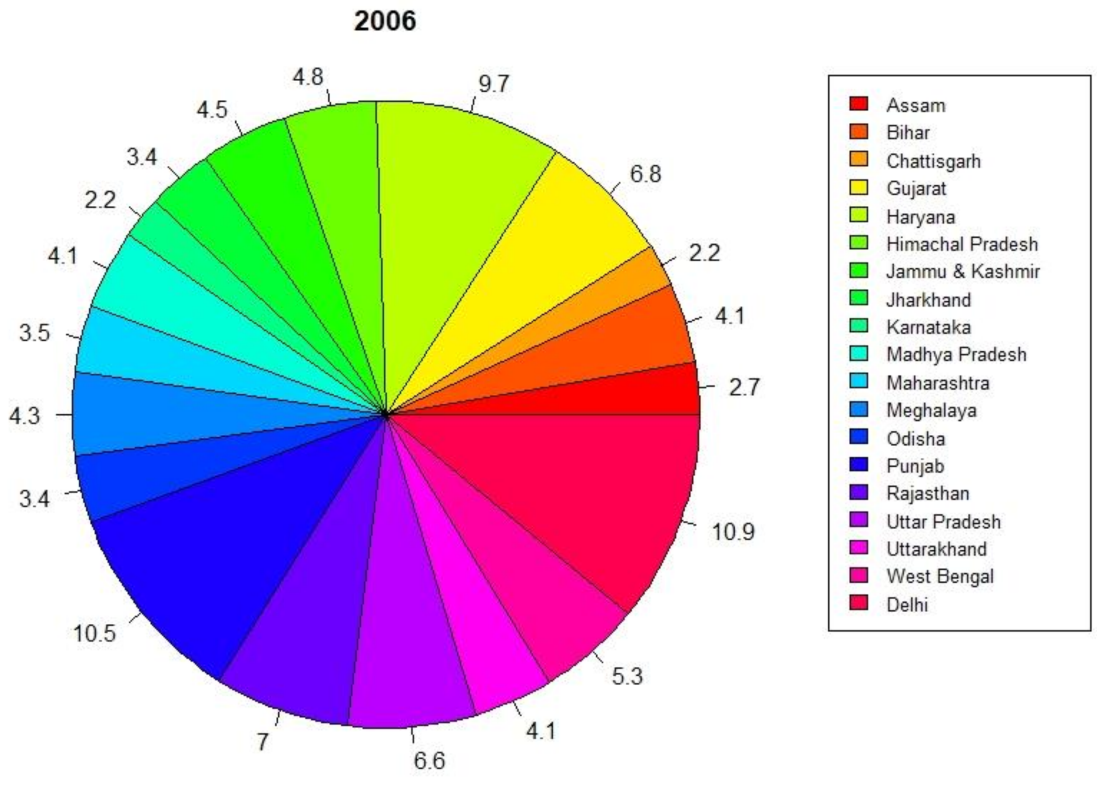


Figure 1.1

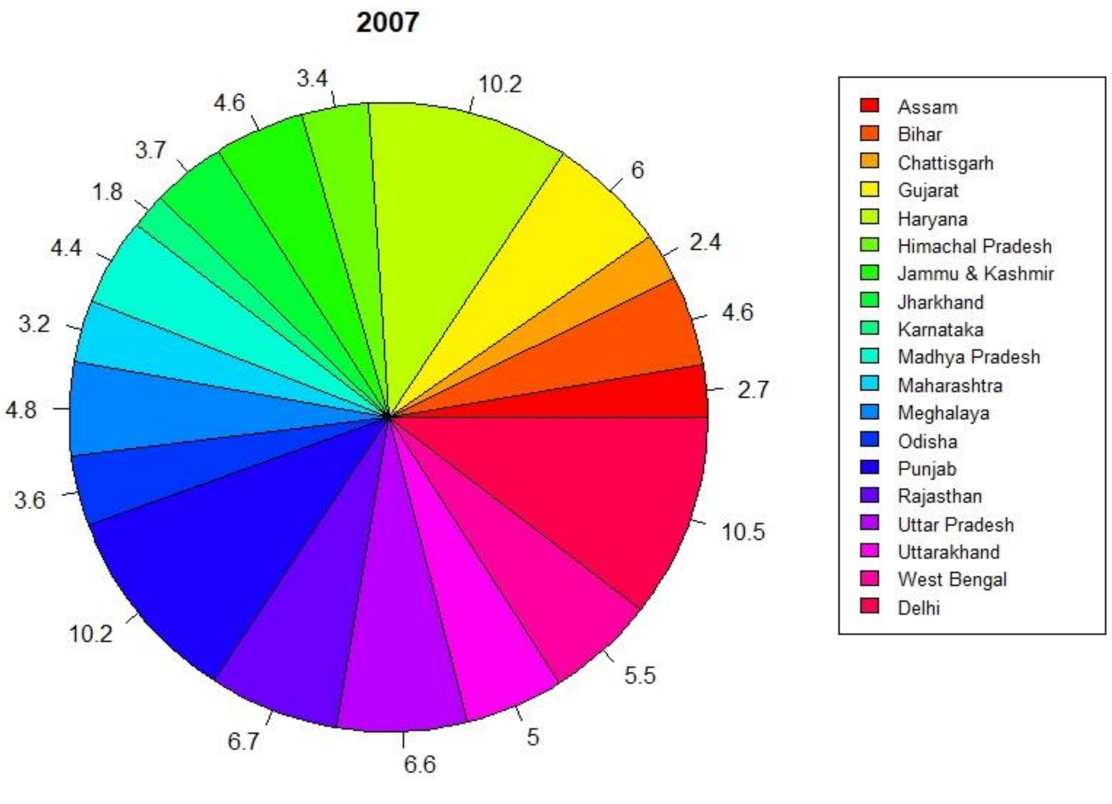


Figure 1.2

- ***PIE CHARTS OF WHEAT PRODUCTION IN THE YEAR 2008 & 2009 FOR MAJOR 19 WHEAT PRODUCING STATES IN INDIA:***

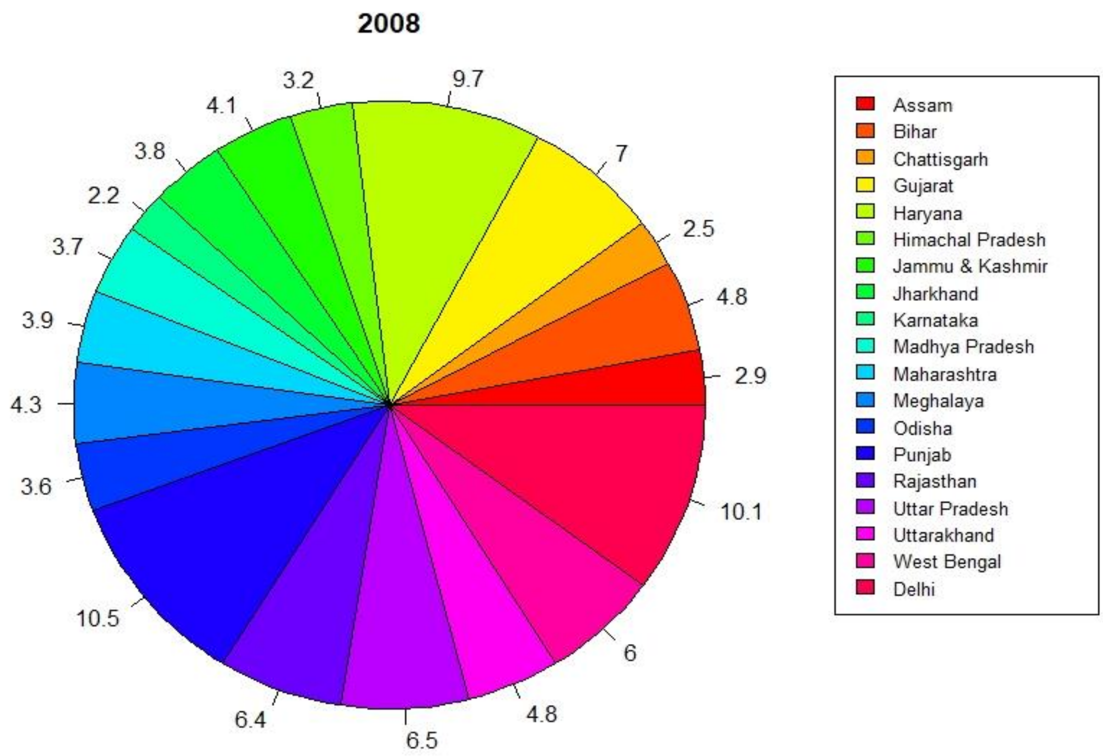


Figure 1.3

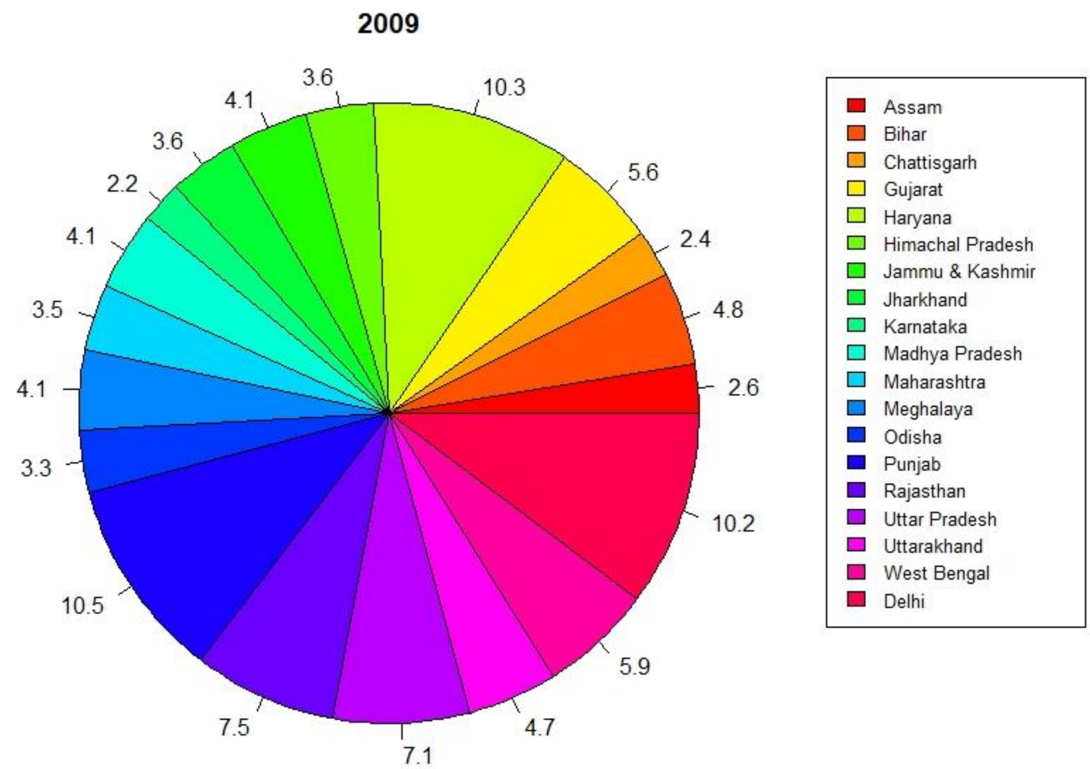


Figure 1.4

- ***PIE CHARTS OF WHEAT PRODUCTION IN THE YEAR 2010 & 2011 FOR MAJOR 19 WHEAT PRODUCING STATES IN INDIA:***

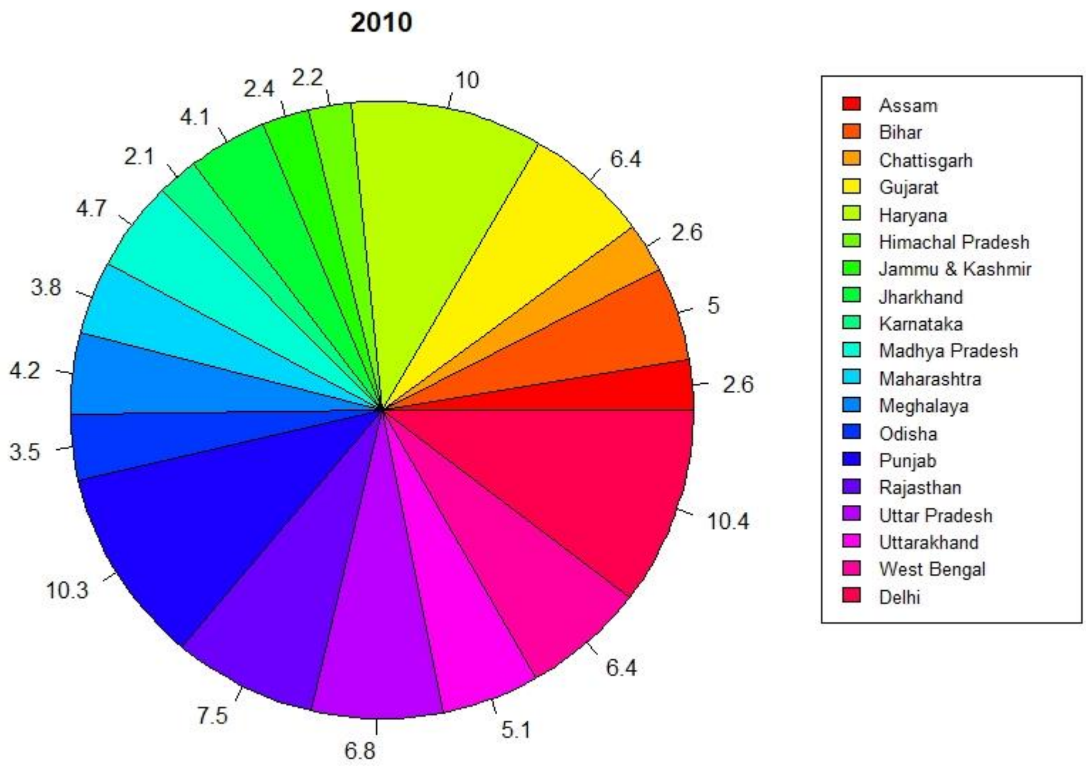


figure 1.5

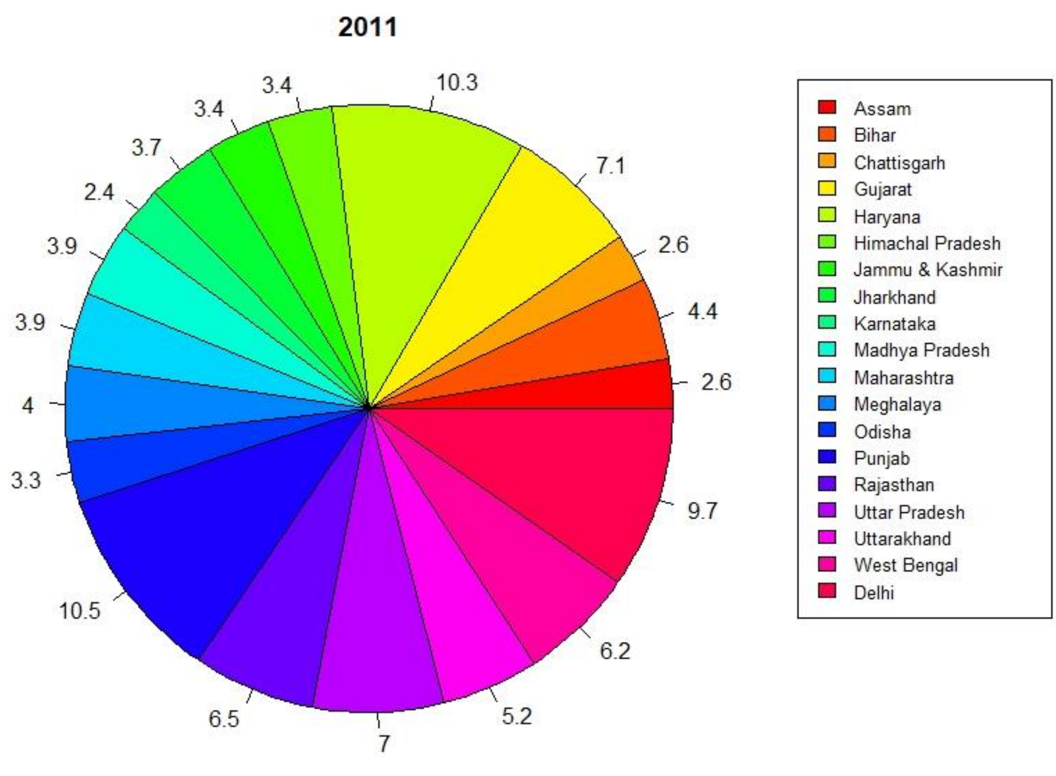


Figure 1.6

- ***PIE CHARTS OF WHEAT PRODUCTION IN THE YEAR 2012 & 2013 FOR MAJOR 19 WHEAT PRODUCING STATES IN INDIA:***

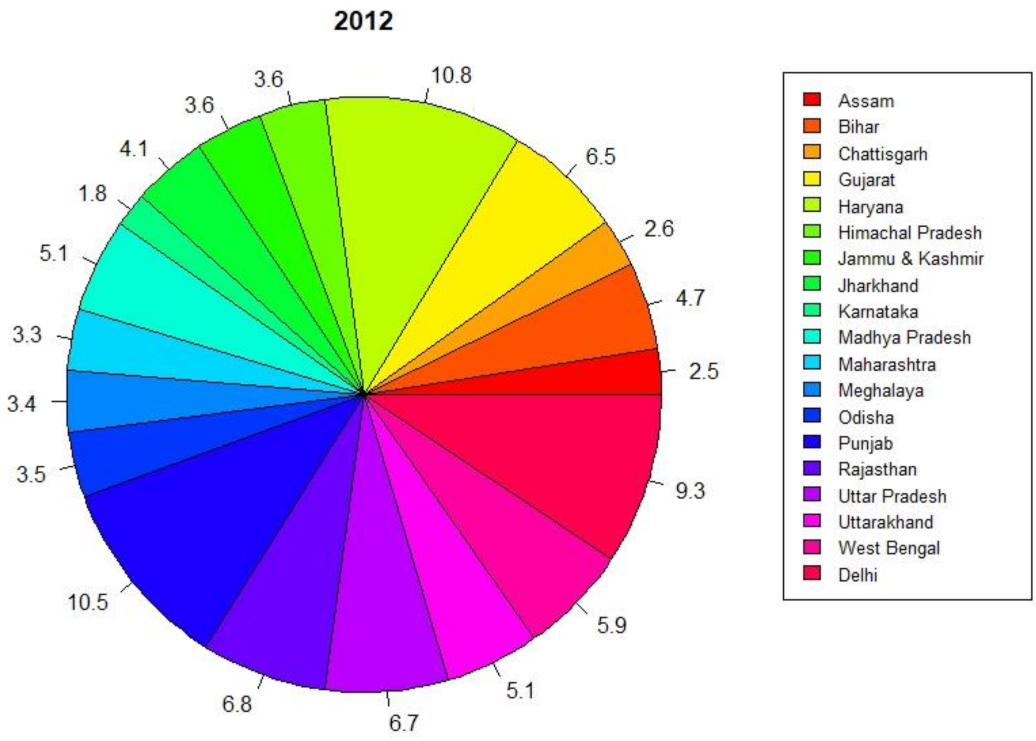


Figure 1.7

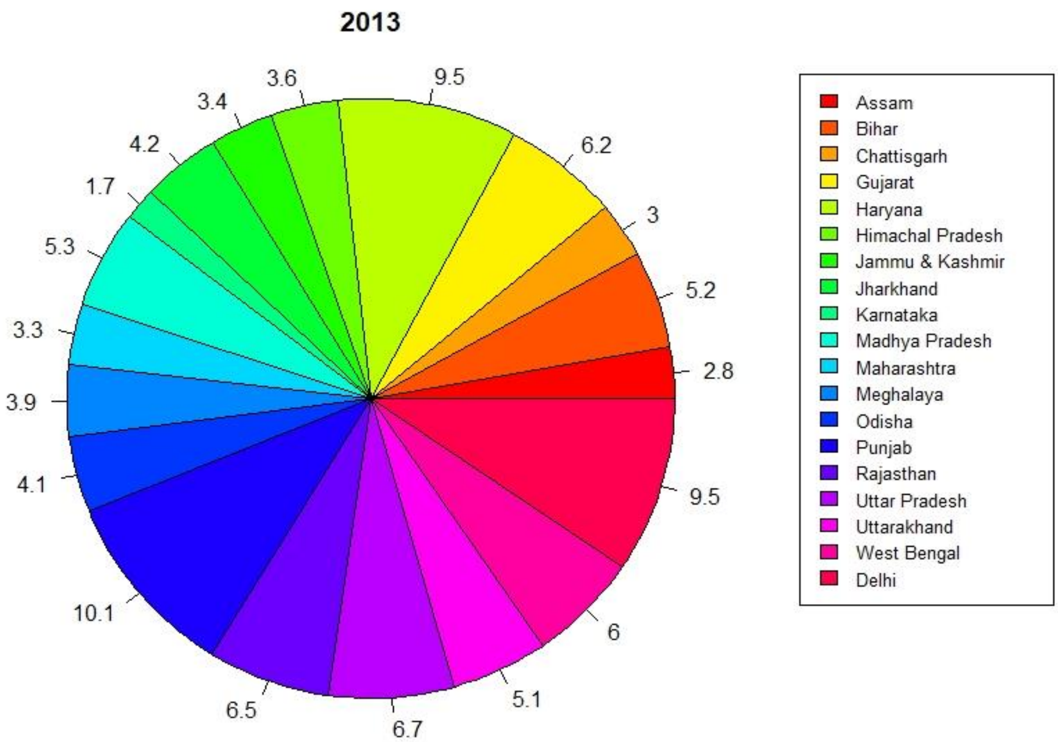


Figure 1.8

- ***PIE CHARTS OF WHEAT PRODUCTION IN THE YEAR 2014 & 2015 FOR MAJOR 19 WHEAT PRODUCING STATES IN INDIA:***

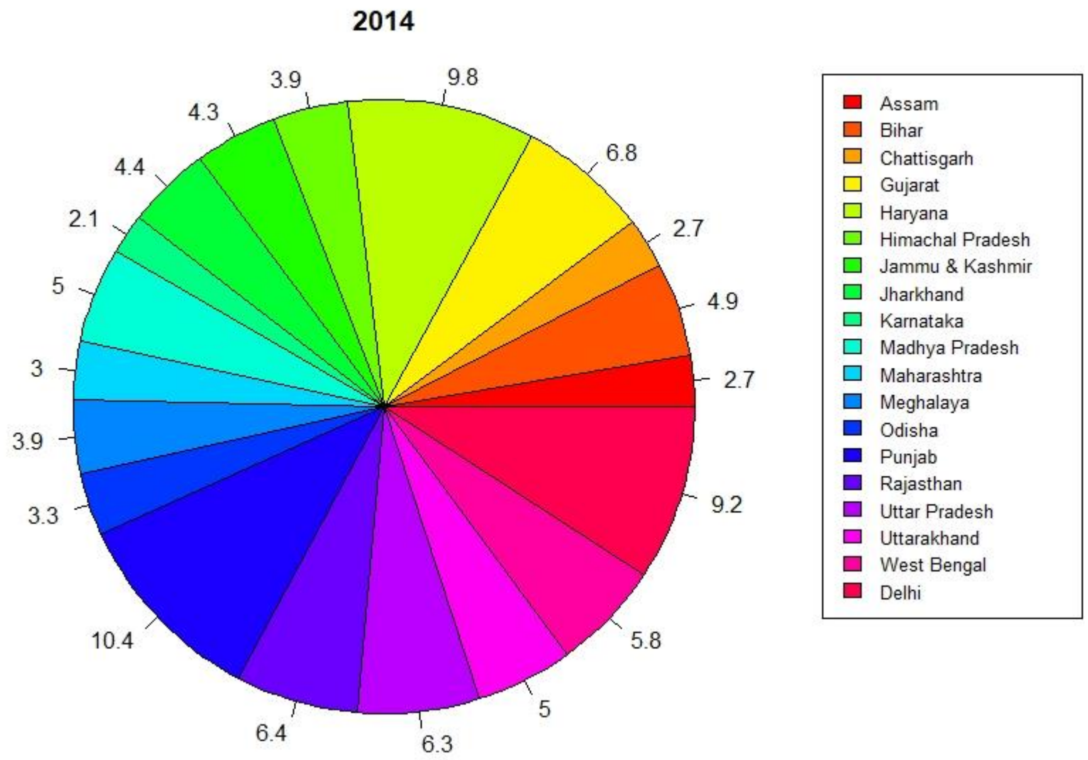


Figure 1.9

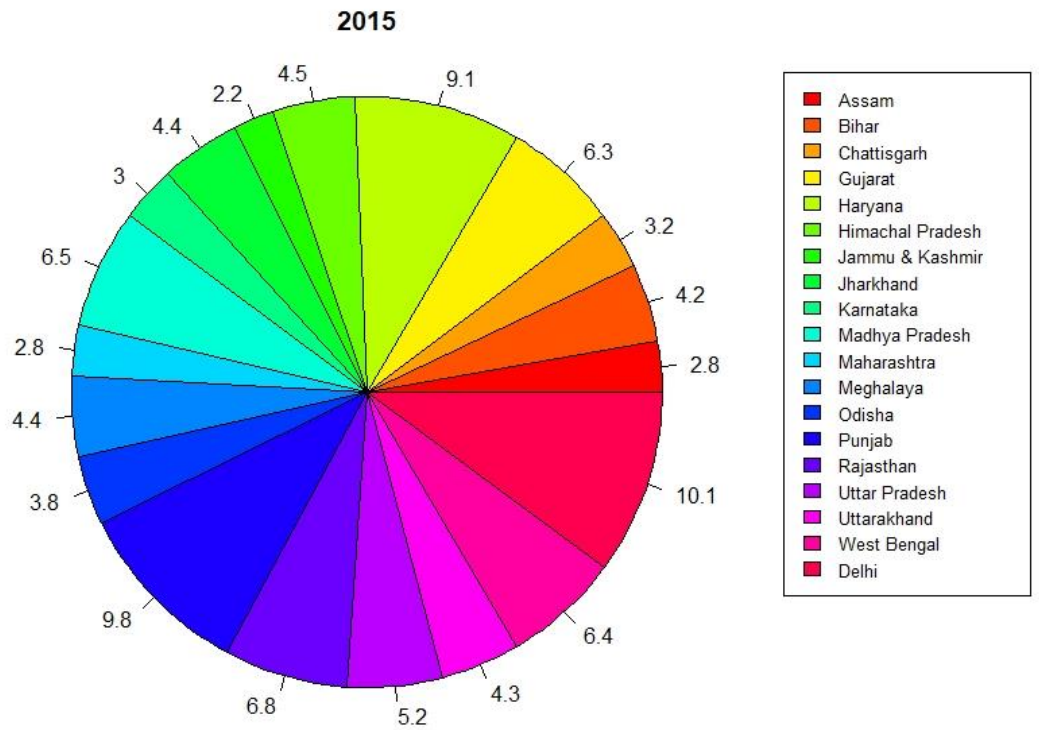


Figure 1.10

- ***PIE CHARTS OF WHEAT PRODUCTION IN THE YEAR 2016 & 2017 FOR MAJOR 19 WHEAT PRODUCING STATES IN INDIA:***

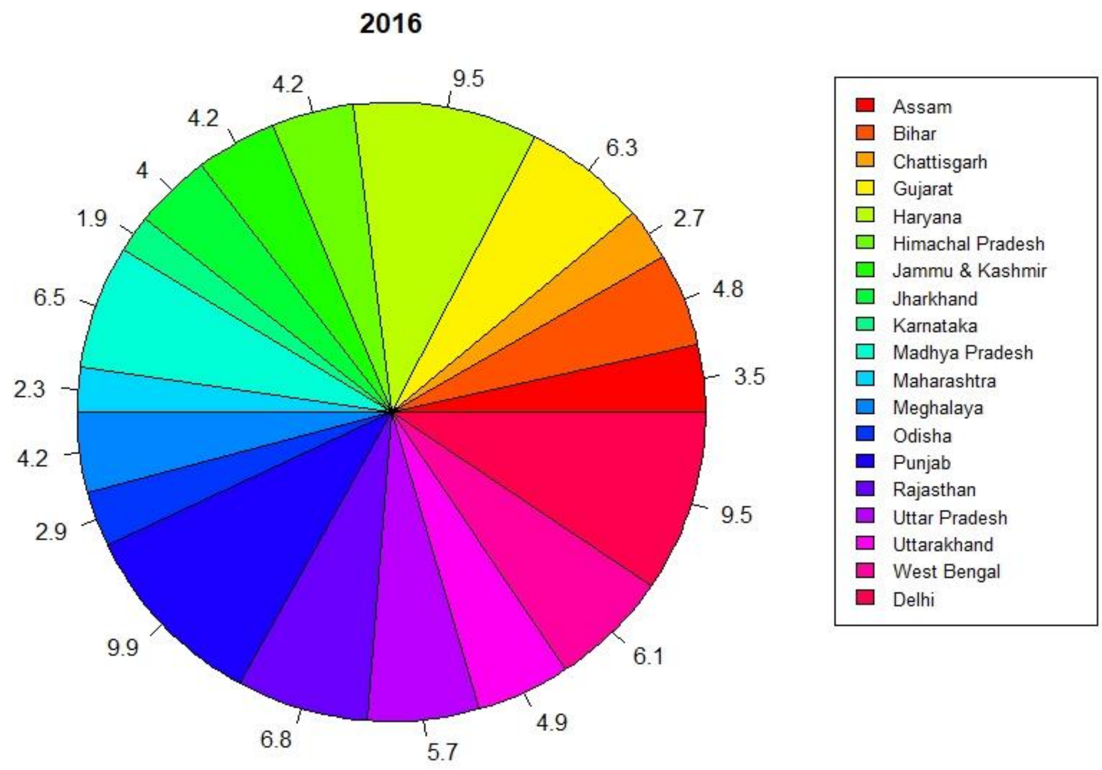


Figure 1.11

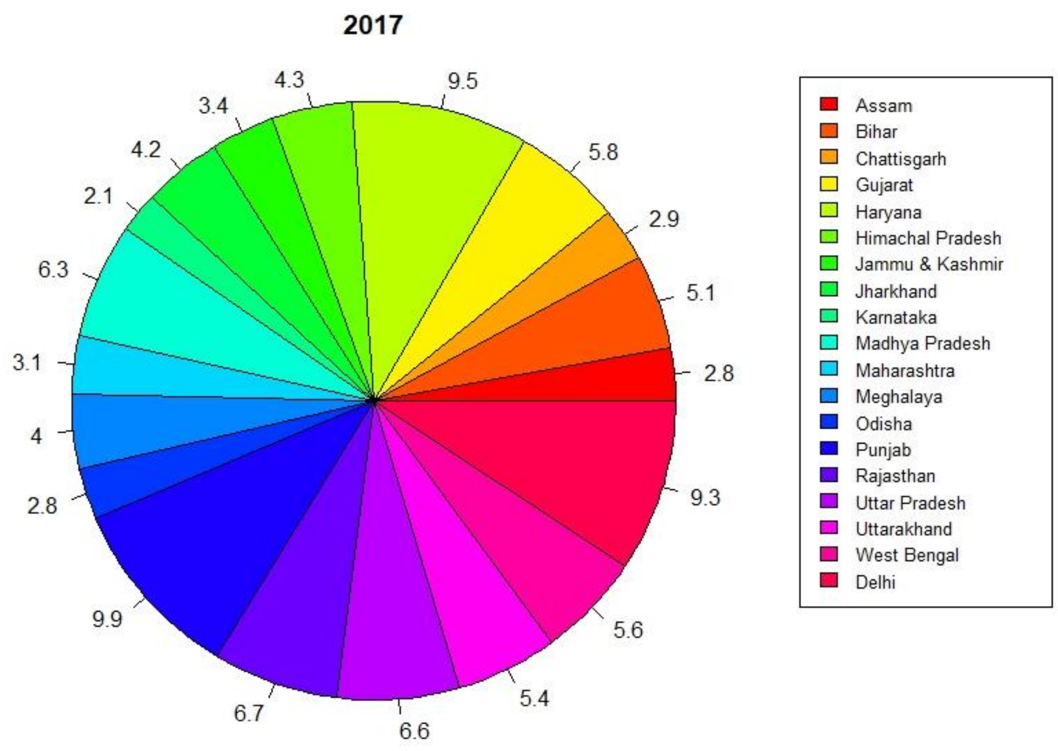


Figure 1.12

MULTIPLE LINE DIAGRAMS OF WHEAT PRODUCTION FOR 12 YEARS 2006-2017 OF MAJOR 19 WHEAT PRODUCING STATES OF INDIA:-

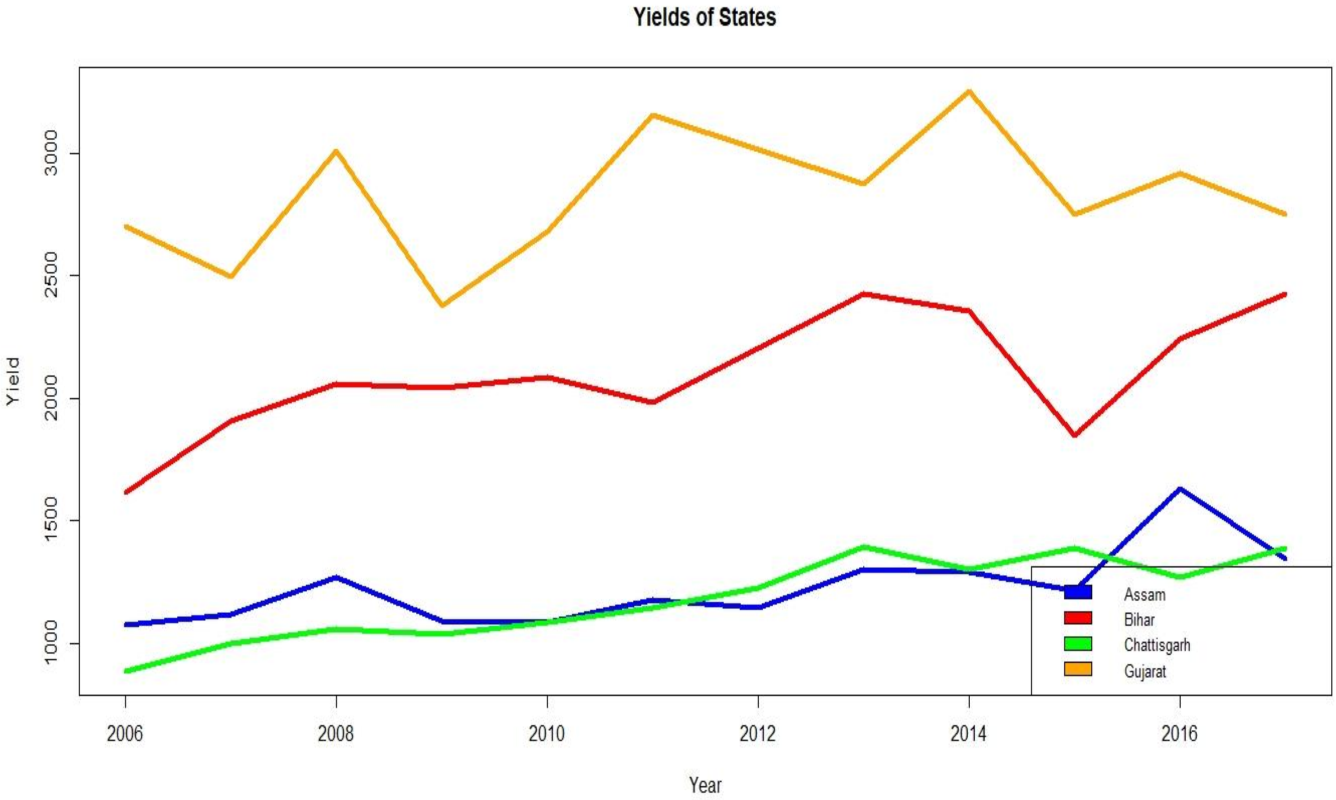


Figure 2.1

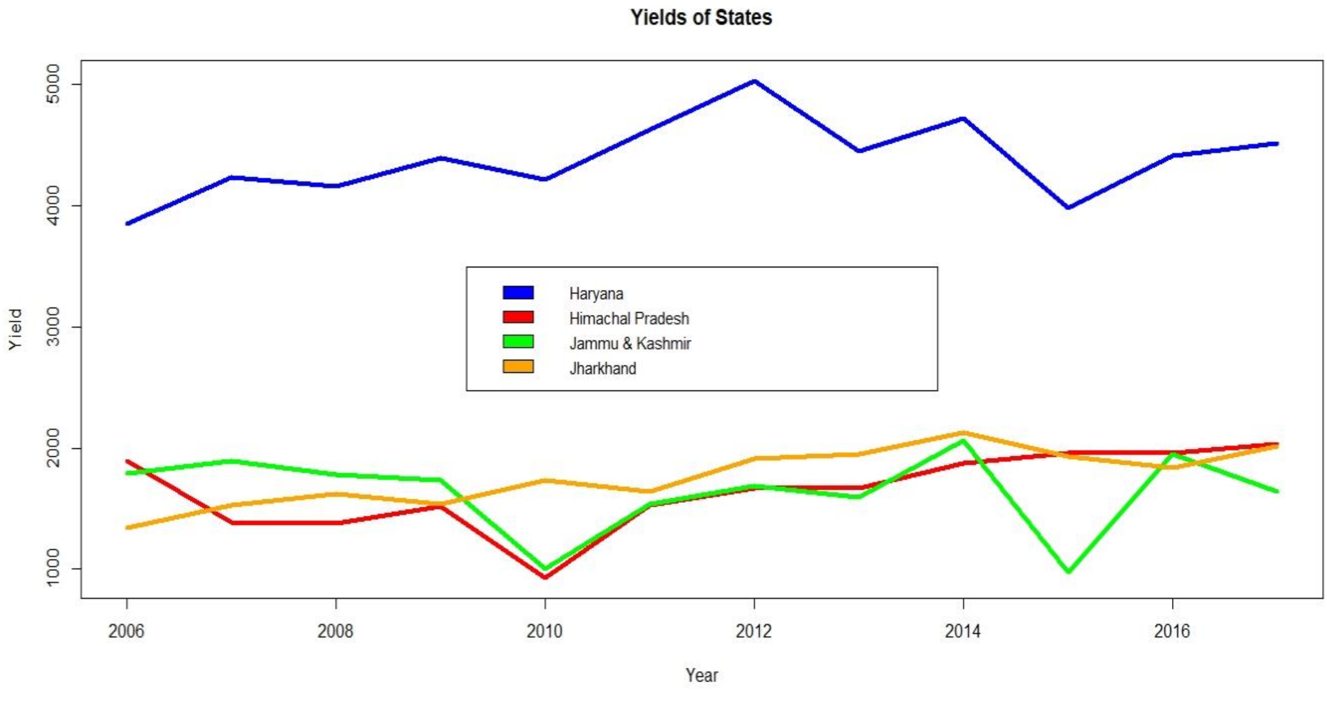


Figure 2.2

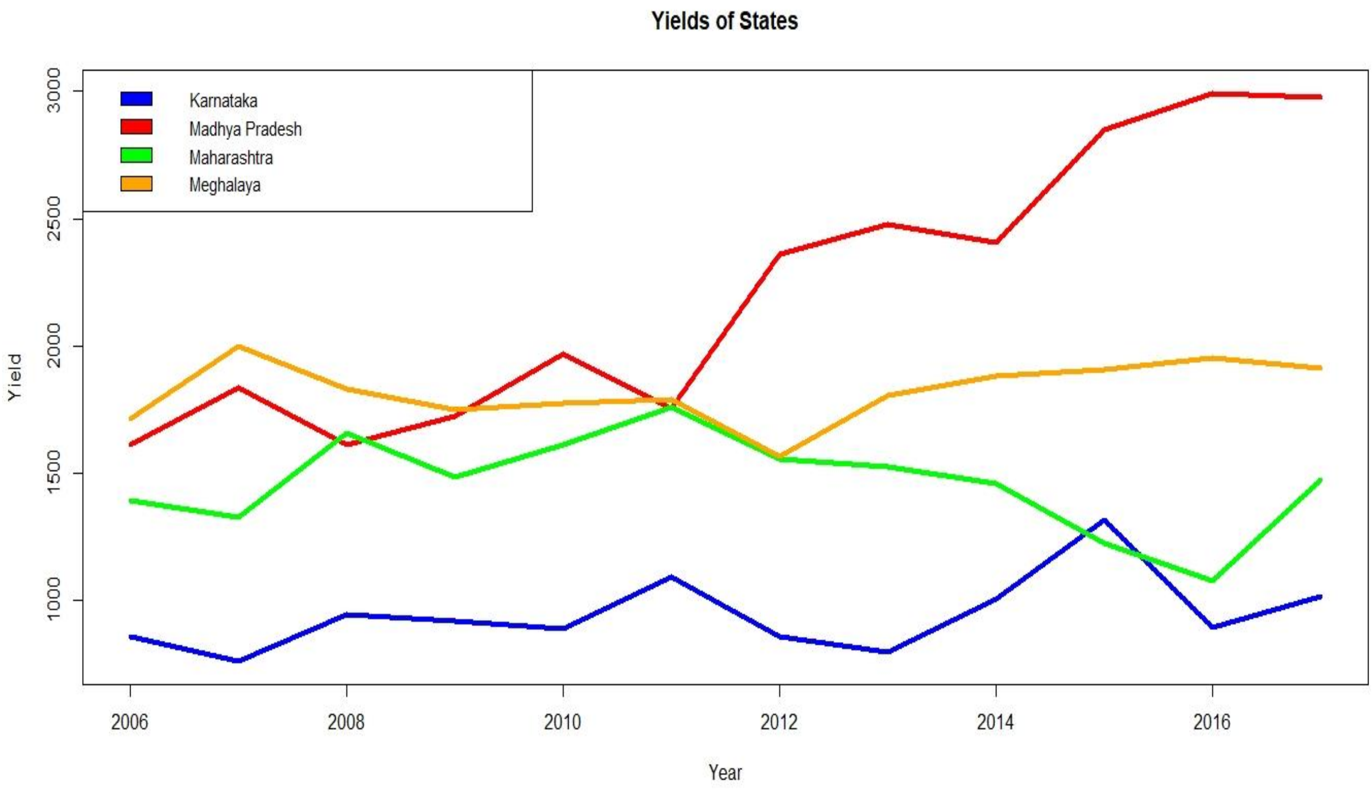


Figure 2.3

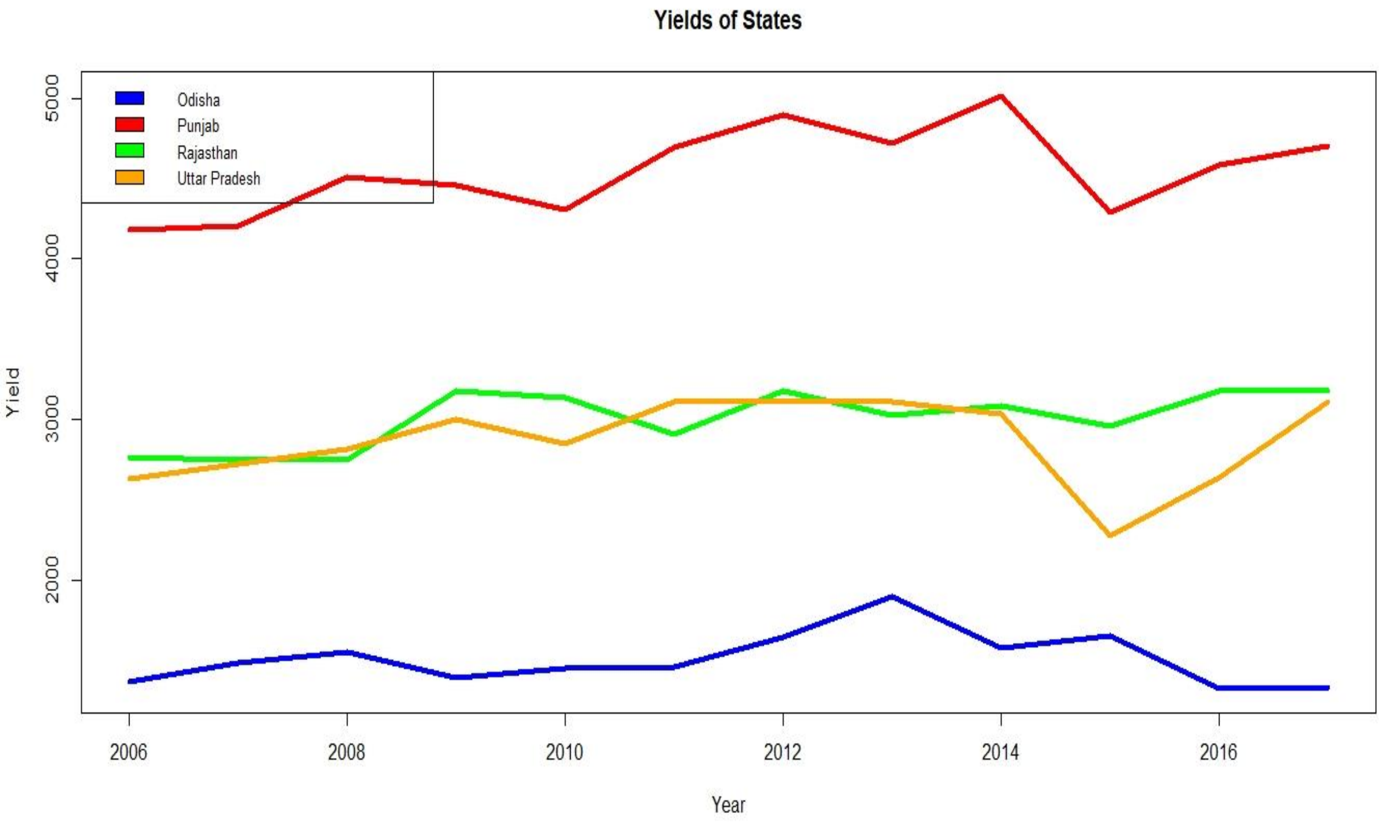


Figure 2.4

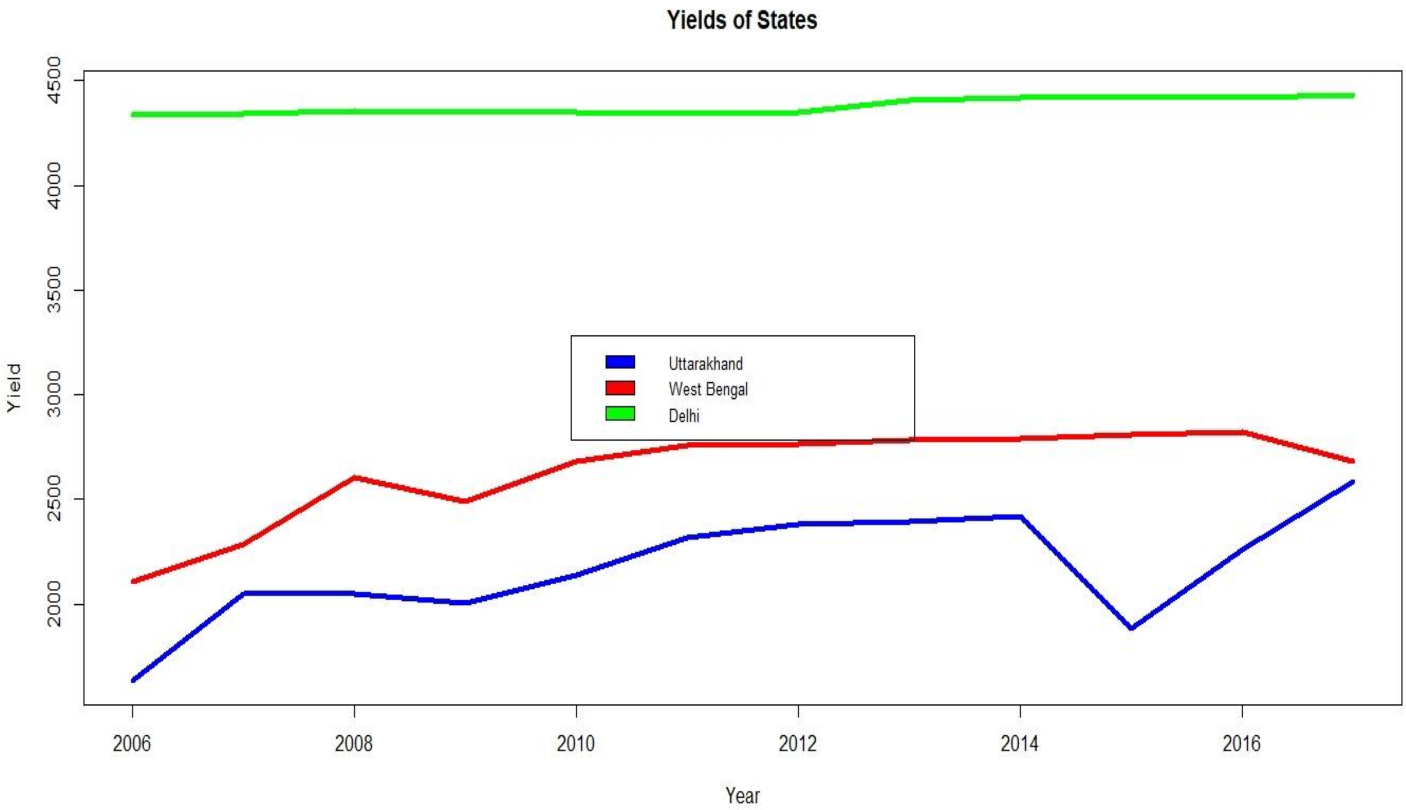


Figure 2.5

INTERPRETATION- From the pie charts we can see that in the last 12 years (2006-2017) total Punjab has been the most wheat producing state in India and the states Delhi, Haryana securing the next positions. After these states comes Rajasthan, Uttar Pradesh, Gujarat, West Bengal and Uttarakhand.Then the other major states have almost the same amount of contribution in total wheat production in that time period but less than the previous eight states.

Now we are studying the annual wheat production of past 12 years for the major states individually and trying to compare them :-

1.PUNJAB- From the pie charts we can see that except for the years 2012, Punjab has been the leading wheat producing state in the 12 years. From the line diagram for Punjab in Figure 2.4 the production of wheat in Punjab is mostly increasing but has a significant declination in the year 2015 and slight declination in 2010.

2. HARYANA- Comparing the pie charts it can be seen that Haryana was the 3rd most wheat producing state in India during 2006-2008 and again in 2010 and 2015(in these years Delhi's wheat production was more than Haryana) but during the other time periods it was the 2nd most wheat producing state of India.

The line diagram for Haryana in Figure 2. 2 says that Haryana is increasing from 2006-2017 with significant declination in the years 2010 and 2015 and has a maximum in the year 2012.

3. DELHI- The pie charts interpret that Delhi was the 2nd most wheat producing state in India during 2006-2008 and again in the years 2010 and 2015. For the rest of the years Haryana was in the 2nd place, making Delhi the 3rd most wheat producing state of India.

The line diagram of Delhi in figure 2.5 shows mostly stable trend over the years.

4. RAJASTHAN- Observing the pie charts it can be seen that Rajasthan was the 4th most wheat producing state in India, except for the years 2008, 2011 ans 2013(during these years Uttar Pradesh became the 4th most wheat producing state).

From figure 2.4 we can say that the wheat production of Rajasthan has been almost stable increasing throughout the years 2006-2017 with some slight declinations for two-three years.

5. UTTAR PRADESH- From the pie charts, it is observable that Uttar Pradesh has been the 5th most wheat producing state of India. It took 4th position during 2008, 2011 and 2013 by replacing Rajasthan.

The line diagram of Uttar Pradesh in Figure 2.4 shows an increasing pattern with a huge declination in 2015, after which it started increasing again.

6. GUJARAT, UTTARAKHAND AND WEST BENGAL – In the state wise wheat production of India,the positions of 6th, 7th and 8th most wheat producing states are permutated between Gujarat, west Bengal and Uttarakhand. But in most of the cases Gujarat's wheat production is more than the other two states.

The line diagram of Gujarat (Figure 2.1) show a huge declination between 2008 and 2009. The line diagram of West Bengal is increasing over the years 2006-2017. Line diagram of Uttarakhand shows less overall production of wheat than West Bengal and has noticeable declination in 2015.

The other major states have almost stable and increasing production of wheat with some inclinations and declinations in between. Only we see for

Maharashtra a decreasing trend till 2016 and a very abrupt trend for Jammu & Kashmir.

Now we will use methods of time series analysis to analyse the trend of annual wheat production of India. From the annual wheat production of India from the year 1960 to 2020,we visualize the time series data graphically below:

GRAPHICAL REPRESENTATION OF ANNUAL WHEAT PRODUCTION OF INDIA FROM 1960 TO 2020:

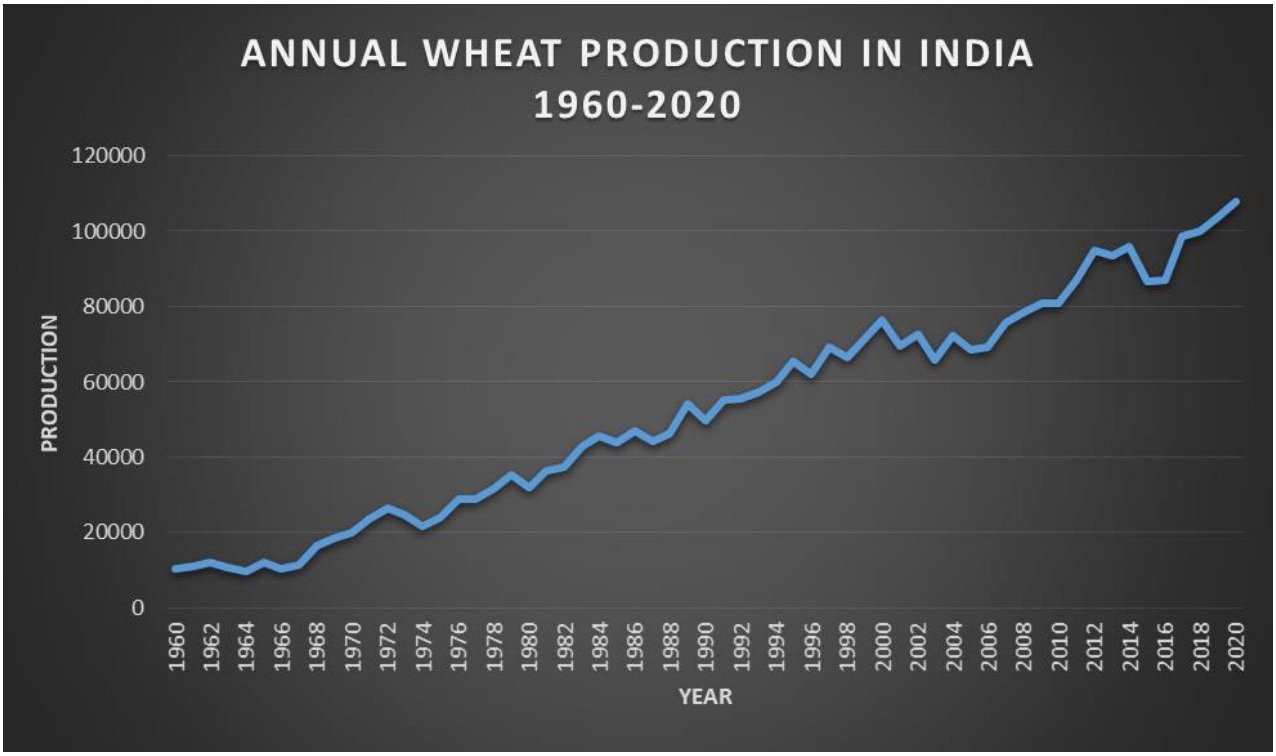


Figure 3

COLLECTED DATA ON ANNUAL WHEAT PRODUCTION OF WHOLE INDIA FROM 1960-2020:

Given below is the data on annual wheat production from 1960-2020 which we will use in our project for furthrt analysis:

YEAR	PRODUCTION (Yt)
1960	10320
1961	10995
1962	12076
1963	10779
1964	9854
1965	12258
1966	10394
1967	11393
1968	16540
1969	18651
1970	20093
1971	23832
1972	26410
1973	24735
1974	21778
1975	24104
1976	28846
1977	29010
1978	31749
1979	35508
1980	31830
1981	36313
1982	37452
1983	42794
1984	45476
1985	44069
1986	47052
1987	44323
1988	46169
1989	54110
1990	49850
1991	55134
1992	55690
1993	57210
1994	59840
1995	65470
1996	62097
1997	69350
1998	66350
1999	71288
2000	76369
2001	69681

YEAR	PRODUCTION(Yt)
2002	72766
2003	65761
2004	72156
2005	68637
2006	69355
2007	75807
2008	78570
2009	80679
2010	80804
2011	86874
2012	94882
2013	93506
2014	95850
2015	86527
2016	87000
2017	98510
2018	99870
2019	103600
2020	107860

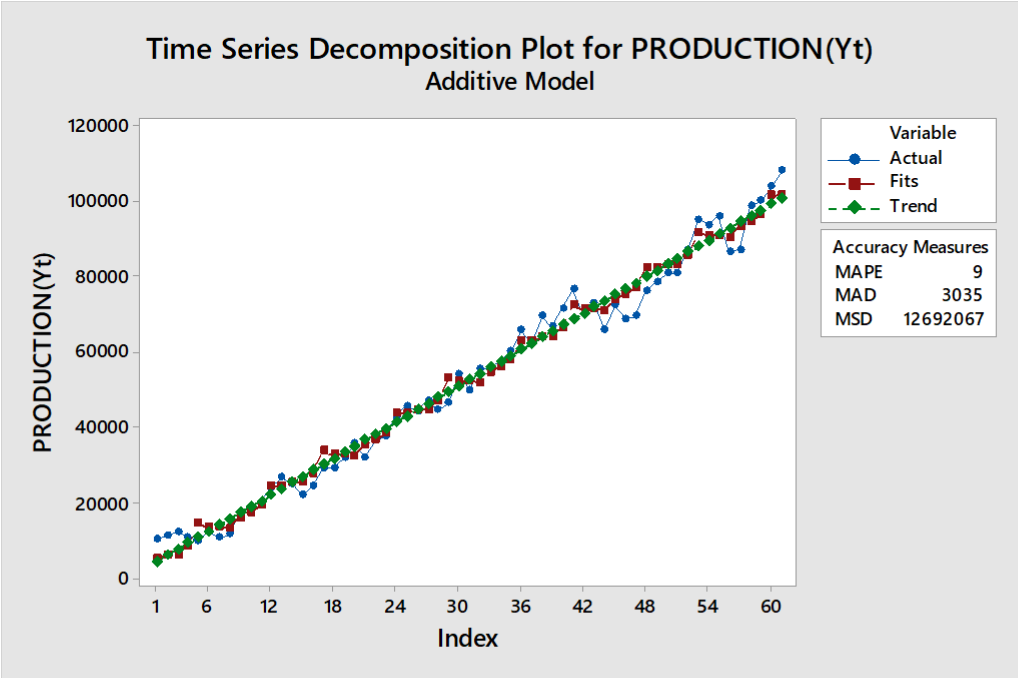
DESCRIPTIVE MEASURES:

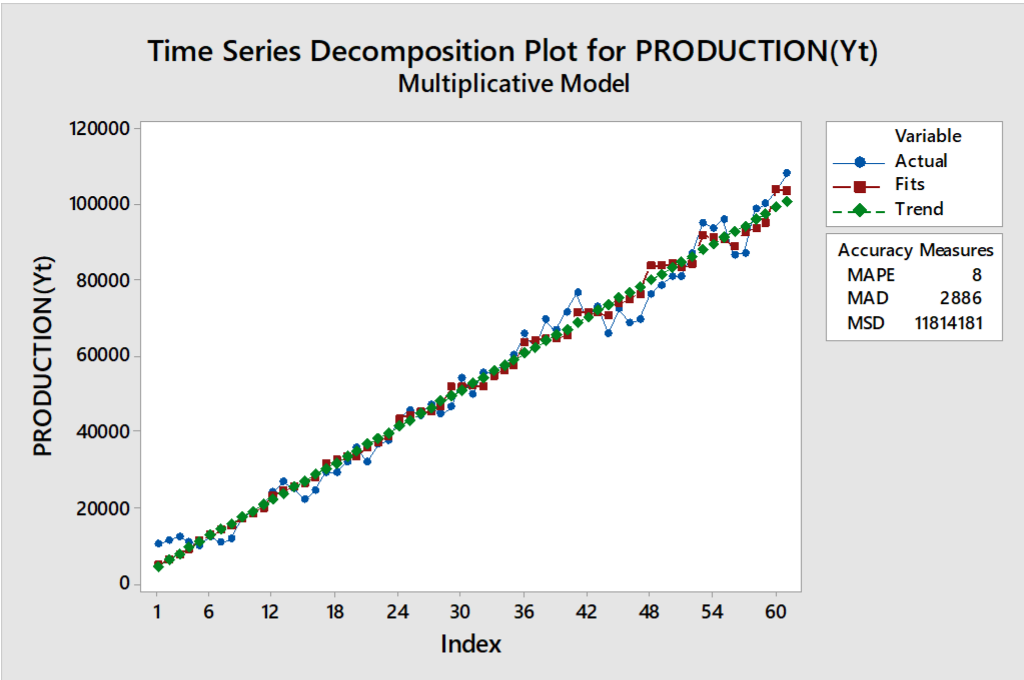
SUMMARY	
Mean	52397.64
Median	54110
Mode	#N/A
Standard Deviation	28781.53
Sample Variance	8.28E+08
Kurtosis	-1.1407
Skewness	0.103606
Minimum	9854
Maximum	107860
Sum	3196256
Count	61
Confidence Level (95.0%)	7371.293

The production data of wheat of whole India from 1960-2020, has mean production = 52397.64 and median production=54110. There's no mode present in the dataset. The minimum production is 9854 in the year 1964 and maximum production is 107860 in 2020. The standard deviation of the dataset is 28781.53. Total wheat production over the years 1960-2020= 3196256.

FOOTNOTES: Here we have collected the wheat production data at the year time point and the production's unit is 1000 mega tonnes (MT). We will use this unit (1000MT) of production further for studying or analysing in our entire project.

DECOMPOSITION OF TIME SERIES:-





Here MAD, MAPE and MSD of Multiplicative model is less than that of the additive model. Hence multiplicative model is a better fit to the Production of wheat over whole India from 1960 to 2020 data.

- ***ANALYSIS TREND BY SIMPLE MOVING AVERAGE METHOD:***

At first we will use Simple Moving Average Method for analysing the trend of the annual wheat production of India. We obtain simplemoving averrage for 3-Year, 5-Year and 10-year period.

Yearly Moving Average For Annual Wheat Production Of India From 1960-2020:

YEAR	PRODUCTION	3YEAR MOVING AVERAGE	5YEAR MOVING AVERAGE	10YEAR MOVING AVERAGE
1960	10320			
1961	10995	11130.33333		
1962	12076	11283.33333	10804.8	
1963	10779	10903	11192.4	
1964	9854	10963.66667	11072.2	
1965	12258	10835.33333	10935.6	
1966	10394	11348.33333	12087.8	12326
1967	11393	12775.66667	13847.2	13303.3
1968	16540	15528	15414.2	14587
1969	18651	18428	18101.8	16020.4
1970	20093	20858.66667	21105.2	17416
1971	23832	23445	22744.2	18608.4
1972	26410	24992.33333	23369.6	19793
1973	24735	24307.66667	24171.8	21638.2
1974	21778	23539	25174.6	23399.9
1975	24104	24909.33333	25694.6	24920.8
1976	28846	27320	27097.4	26606.5
1977	29010	29868.33333	29843.4	27780.2
1978	31749	32089	31388.6	29028.3
1979	35508	33029	32882	30132.5
1980	31830	34550.33333	34570.4	31938.4
1981	36313	35198.33333	36779.4	34308.2
1982	37452	38853	38773	36304.7
1983	42794	41907.33333	41220.8	38125.3
1984	45476	44113	43368.6	39656.6
1985	44069	45532.33333	44742.8	41098.6
1986	47052	45148	45417.8	42958.8
1987	44323	45848	47144.6	44760.8
1988	46169	48200.66667	48300.8	46642.9

1989	54110	50043	49917.2	48466.7
1990	49850	53031.33333	52190.6	49908.3
1991	55134	53558	54398.8	51344.7
1992	55690	56011.33333	55544.8	53484.8
1993	57210	57580	58668.8	54989.3
1994	59840	60840	60061.4	57492
1995	65470	62469	62793.4	59510.1
1996	62097	65639	64621.4	61227.9
1997	69350	65932.33333	66911	63879.8
1998	66350	68996	69090.8	65334.5
1999	71288	71335.66667	70607.6	67042.1
2000	76369	72446	71290.8	67897.2
2001	69681	72938.66667	71173	69128.8
2002	72766	69402.66667	71346.6	69445.5
2003	65761	70227.66667	69800.2	70171.3
2004	72156	68851.33333	69735	70817
2005	68637	70049.33333	70343.2	72039
2006	69355	71266.33333	72905	72978.1
2007	75807	74577.33333	74609.6	73421.6
2008	78570	78352	77043	75140.9
2009	80679	80017.66667	80546.8	77352.5
2010	80804	82785.66667	84361.8	80127
2011	86874	87520	87349	82496.4
2012	94882	91754	90383.2	84285.4
2013	93506	94746	91527.8	86049.9
2014	95850	91961	91553	88320.2
2015	86527	89792.33333	92278.6	90450.2
2016	87000	90679	93551.4	
2017	98510	95126.66667	95101.4	
2018	99870	100660	99368	
2019	103600	103776.6667		
2020	107860			

<u>Interpretation-</u> The moving averages of 3-years, 5-years and 10-years have smoothed our data and made the trend clearer.The longer the period is chosen, the more smoothed the data. So here 10 year moving average makes the trend smoothest as it takes more data points in calculation. The figures given below capture the 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.

SIMPLE MOVING AVERAGES OF 3YEAR, 5YEAR AND 10YEAR WITH ACTUAL PRODUCTION:

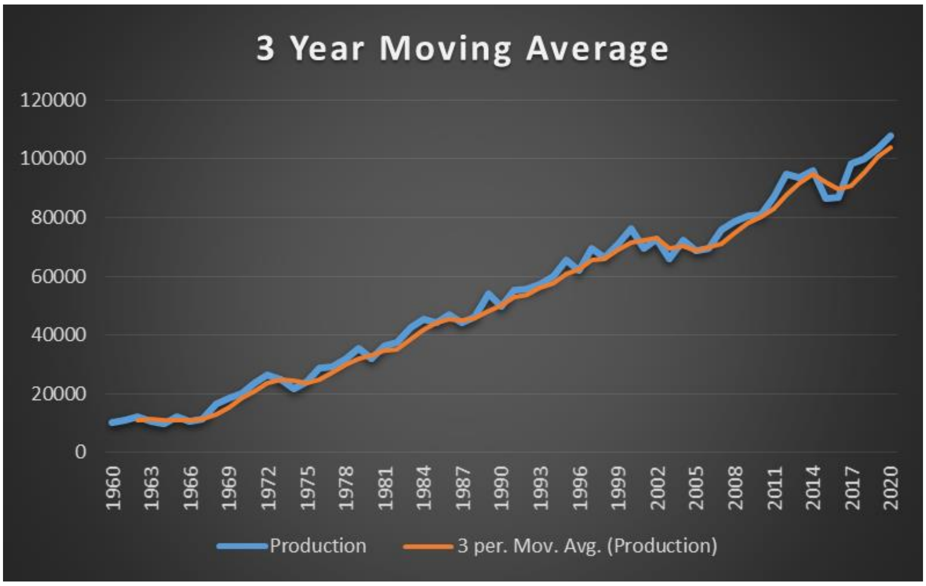


Figure 4.1

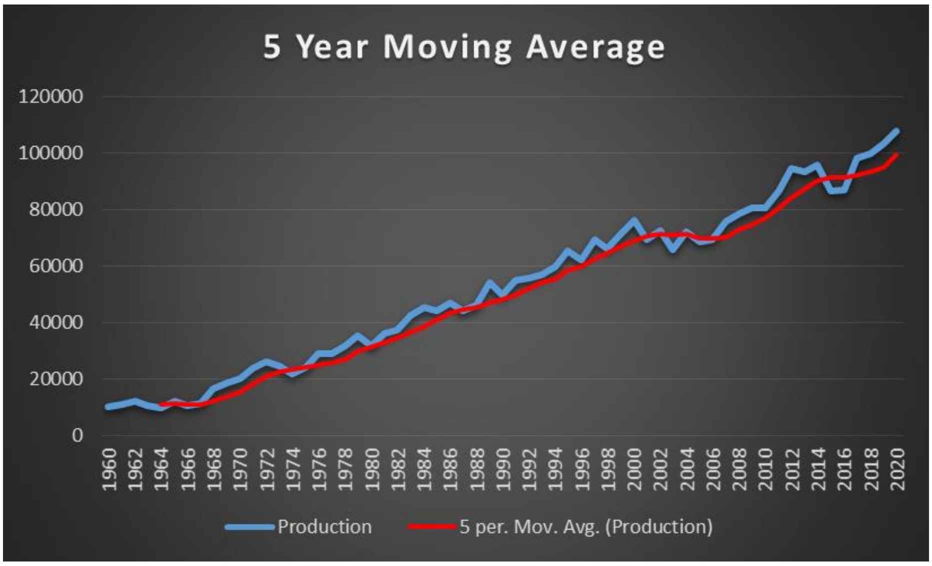


Figure 4.2

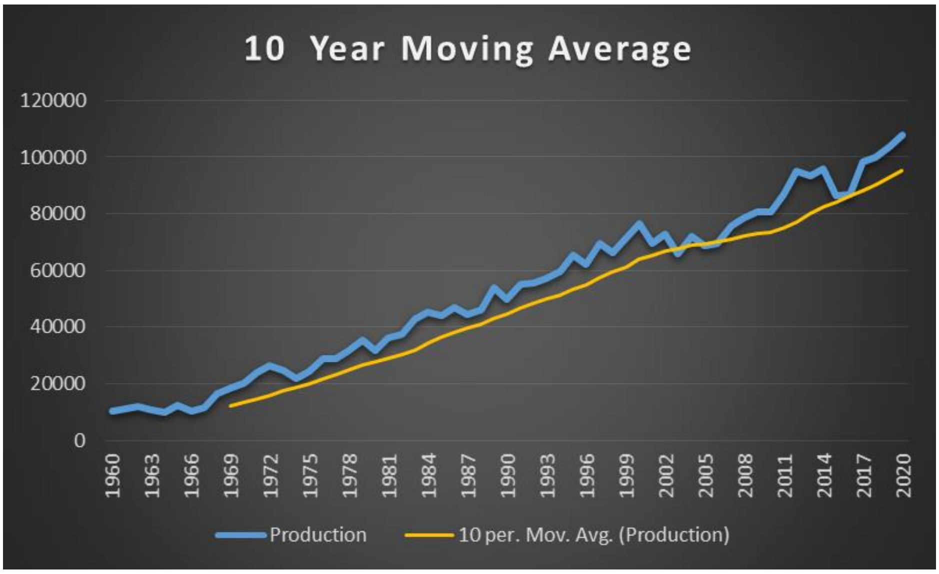


Figure 4.3

- ***FITTING A MATHEMATICAL CURVE:***

In Figure 3 the graphical visualization of our tine series data and in Figure 4 after smoothing the data by moving average we observed that the trend of our time series data i.e., the long term movement is increasing. So for polynomial trend line, we have successively fitted Linear and Quadratic models and not any polynomial higher degree than 2 for my data. We have performed their ANOVA testing and made decisions of accepting or rejecting on the basis of p-values provided by the test. Here for the ANOVA testing, 5% level of significance is chosen. We have also observed the value of coefficient of determination i.e., R^2 values and adjusted R^2 values in each case.

LINEAR MODEL

The fitted linear model is-

Yt= 1606.2t + 2604.7 , with origin at **1960** and **1 year** unit of time.

ANOVA

Table 5.1

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	48786940466	4.88E+10	3143.55	7.02394E-53
Residual	59	915661929.9	15519694		
Total	60	49702602396			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	4210.928609	996.5256782	4.22561	8.38E-05
X Variable 1	1606.223691	28.648099	56.06737	7.02E-53

Since the p-value < 0.05, we accept the linear model fitting.

Table 5.2

Regression Statistics	
Multiple R	0.990745771
R Square	0.981577183
Adjusted R Square	0.981264932
Standard Error	3939.504249
Observations	61

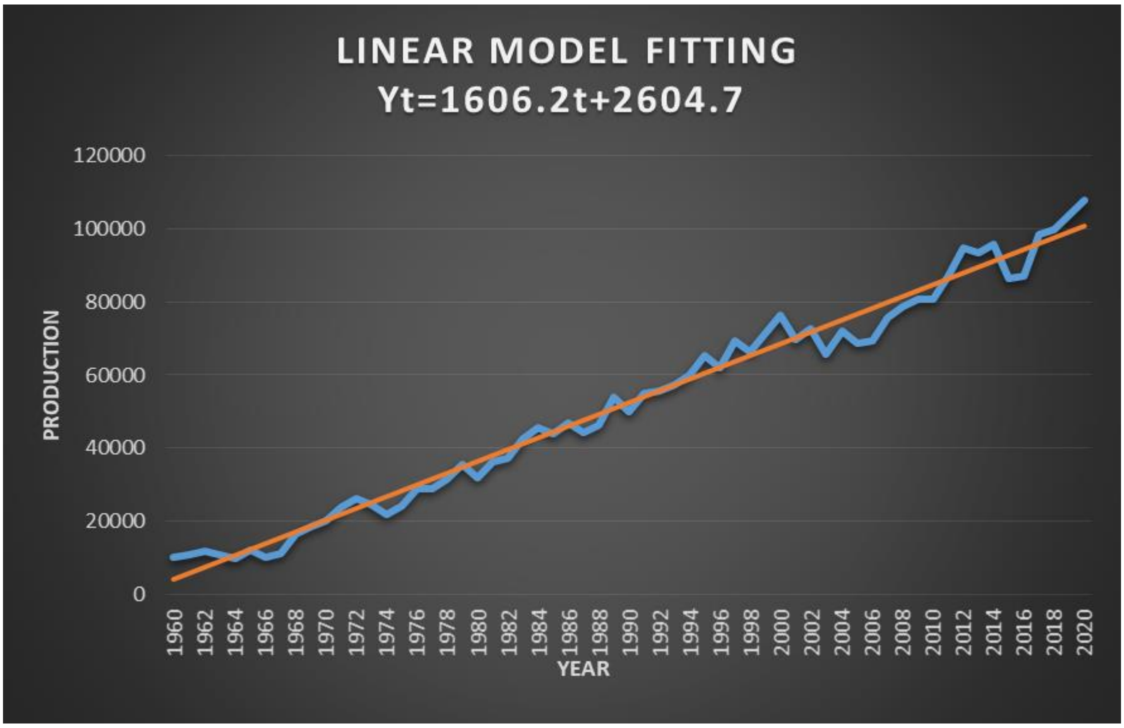


Figure 5

QUADRATIC MODEL

The fitted quadratic model is-

Yt= 3.003t^2+ 1420t+ 4559.7, with origin at **1960** and **1 year** unit of t.

ANOVA

Table 6.1

	df	SS	MS	F	Significance F
Regression	2	48829199043	2.44E+10	1621.298	1.25961E-51
Residual	58	873403352.6	15058678		
Total	60	49702602396			

	Coefficients	Standard Error	t Stat	P-value
Intercept	5982.721815	1442.991393	4.146055	0.000112
X Variable 1	1426.041331	111.199609	12.82416	1.41E-18
X Variable 2	3.003039332	1.792656231	1.67519	0.099281

Since again the p-value < 0.05, we accept the quadratic model fitting.

Table 6.2

Regression Statistics	
Multiple R	0.991174764
R Square	0.982427412
Adjusted R Square	0.981821461
Standard Error	3880.551313
Observations	61

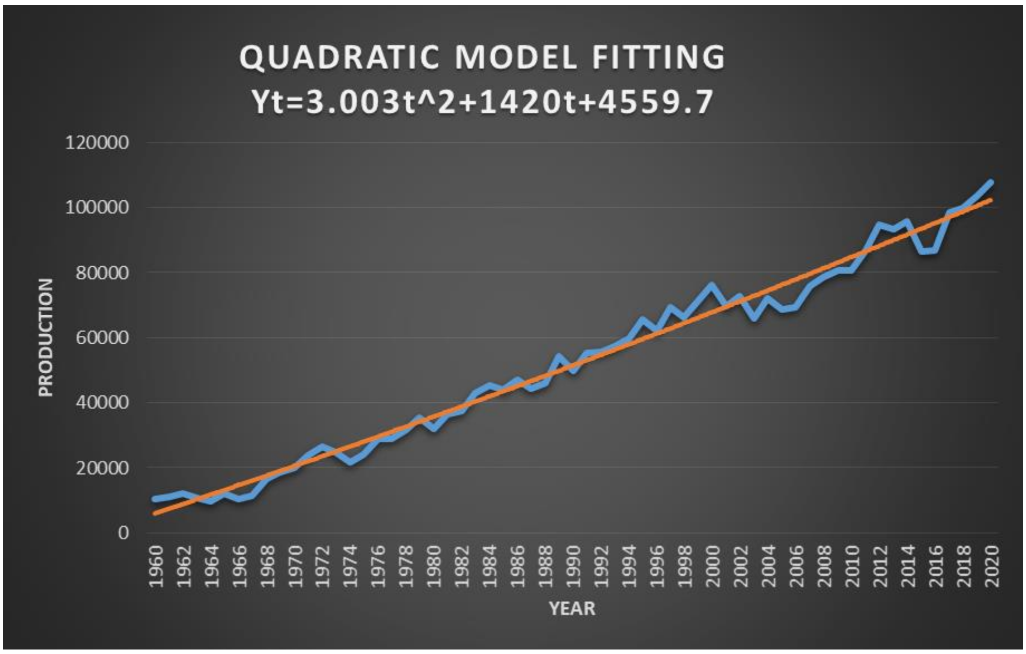


Figure 6

For growth curve now we are fitting exponential model and performing ANOVA testing for the linear model log Yt= log a + t* (log b) and made decision of accepting or rejecting on the basis of p-values provided by the test. Here for the ANOVA testing 5% level of significance is chosen.

The fitted exponential model is- **Yt= 12991*(1.03913^t)**, with origin at **1960** and **1 year** unit of time.

ANOVA

Table 7.1

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	5.254131362	5.254131	649.9984	1.51007E-33
Residual	59	0.476914641	0.008083		
Total	60	5.731046003			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	4.13030127	0.022742659	181.6103	8.77E-83
X Variable 1	0.016668814	0.000653805	25.49507	1.51E-33

Since p-values are < 0.05, we accept the exponential model fitting.

Table 7.2

Regression Statistics	
Multiple R	0.957488389
R Square	0.916784014
Adjusted R Square	0.915373574
Standard Error	0.089907169
Observations	61

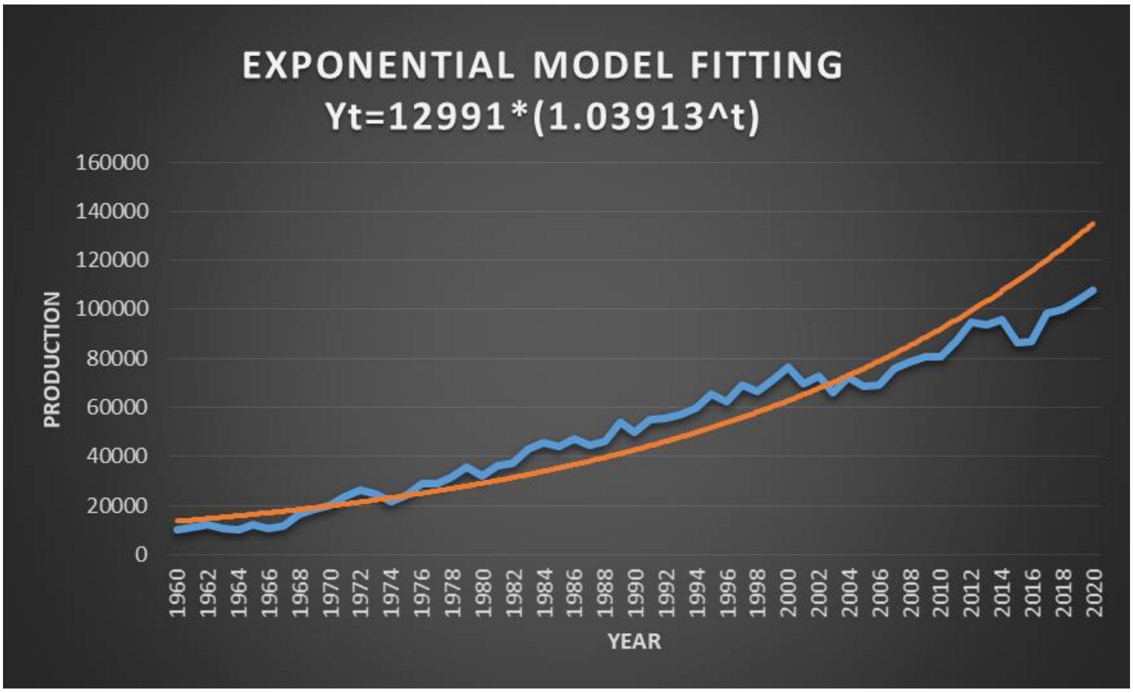


Figure 7

Now in our time series data the amount of production has no upper limit. From the time series plot graph in figure 3 and after smoothing the data in fig 4 we can see that the series is increasing and the amount of growth doesn't decline by 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 and Gompertz Model.

From the Tables 5.2, 6.2 and 7.2 we can see that among the linear, quadratic and exponential model, quadratic model has the largest R square value i.e., 0.982427412 and largest adjusted R square value which is 0.981821461. So we conclude that the fitted quadratic trend line equation is best for our time series data.

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

Comparison between Linear, Quadratic and Exponential Models using MAD, MAPE & MSD-

The fitted linear model is- **Yt= 1606.2t + 2604.7**

Method

Model type	Linear Trend Model
Data	PRODUCTION(Yt)
Length	61
NMissing	0

Accuracy Measures

MAPE	9
MAD	3075
MSD	15010851

The fitted quadratic model is- **Yt= 3.003t^2+ 1420t+ 4559.7**

Method

Model type	Quadratic Trend Model
Data	PRODUCTION(Yt)
Length	61
NMissing	0

Accuracy Measures

MAPE	8
MAD	3032
MSD	14318088

The fitted exponential model is - **Yt = 12991*(1.03913^t)**

Method

Model type	Growth Curve Model
Data	PRODUCTION(Yt)
Length	61
NMissing	0

Accuracy Measures

MAPE	18
MAD	8517
MSD	117202946

The MAPE, MAD & MSD of of the quadratic model is less than both the linear and exponential growth curve models. So, the quadratic model is a better fit to the data. So we use this model to find out future production of wheat in India.

Prediction Of Annual Wheat Production Of India For Future Years (Using Quadratic Model):

Table 8

YEAR	t	PRODUCTION(1000MT)
2021	61	102353.863
2022	62	104143.232
2023	63	105938.607
2024	64	107739.988
2025	65	109547.375
2026	66	111360.768
2027	67	113180.167
2028	68	115005.572
2029	69	116836.983
2030	70	118674.4

CONCLUSION:

We have analysed the wheat production data of major wheat producing states in India for the past 12 years in an explorative way. After analysing it can be concluded that Punjab is the most wheat producing state in India followed by Haryana, Delhi. Then the major wheat producing states come are Rajasthan, Uttar Pradesh, Gujarat, West Bengal and Uttarakhand contributes the most in the wheat production in India.

From the time series trend analysis on our wheat production data of whole India we can see that the production of wheat 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 wheat production of whole India. By the fitted quadratic trend equation we have predicted the production of wheat in India for future 10 years. We also used exponential smoothing to predict the future wheat production in India for the year 2021. Observing the predicted values we conclude that the wheat production will be increased in the upcoming years i.e. there will be a rise in wheat production for the future years.

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

REFERENCES:

- ***Sources of Theory –***

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

Introduction To Statistics – Prasanta Kumar Giri, Jiban Banerjee

The Analysis Of Time Series – Chatfield C.

Introduction to Time Series and Forecasting, 2nd Edition – Brockwell and Davis

https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/adjusted-r2/

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

https://www.tableau.com/learn/articles/time-series-analysis/

- ***Software used-***

Microsoft Excel 2013

Microsoft Word 2013

R Software (R x64 4.0.2)

Minitab 18

ACKNOWLEDGEMENT:

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

Firstly, I owe a deep debt of gratitude to my supervisor Ms. Oindrila Bose, for necessary guidance, for this presentation of this dissertation, valuable suggestions and comments. I am extremely grateful to her for the necessary stimulus, support and valuable time.

Special thanks to Dr. Dhiman Dutta, Head of the Department of Statistics, Asutosh College. I am greatly indebted to Dr. Shirsendu Mukherjee, Dr. Sankha Bhattacharya and Dr. Parthasarathi Bera. 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.

APPENDIX:

- ***Necessary Calculation & Analysis –***

- ***R CODES FOR PIE CHARTS-***

FOR 2006-

```
>
x=c(1074,1617,886,2700,3844,1894,1790,1340,858,1613,1393,1714,1364,4179,2762,2627,1633,210
9,4339)

> piepercent=round(100*x/sum(x),1)

> pie(x,radius=1,labels=piepercent,main="2006",col=rainbow(length(x)))

> legend("topright",c("Assam","Bihar","Chattisgarh","Gujarat","Haryana","Himachal
Pradesh","Jammu & Kashmir","Jharkhand","Karnataka","Madhya
Pradesh","Maharashtra","Meghalaya","Odisha","Punjab","Rajasthan","Uttar
Pradesh","Uttarakhand","West Bengal","Delhi"),fill=rainbow(length(x)),cex=0.8,inset=c(-0.4,0))
```

FOR 2007 -

```
>
x=c(1117,1908,1002,2498,4232,1385,1893,1529,762,1835,1325,2000,1487,4210,2751,2721,2049,22
82,4341)

> piepercent=round(100*x/sum(x),1)

> pie(x,radius=1,labels=piepercent,main="2007",col=rainbow(length(x)))

> legend("topright",c("Assam","Bihar","Chattisgarh","Gujarat","Haryana","Himachal
Pradesh","Jammu & Kashmir","Jharkhand","Karnataka","Madhya
Pradesh","Maharashtra","Meghalaya","Odisha","Punjab","Rajasthan","Uttar
Pradesh","Uttarakhand","West Bengal","Delhi"),fill=rainbow(length(x)),cex=0.8,inset=c(-0.4,0))
```

FOR 2008-

```
>
x=c(1268,2058,1059,3013,4158,1376,1782,1621,946,1612,1659,1833,1554,4507,2749,2817,2050,26
02,4354)

> piepercent=round(100*x/sum(x),1)

> pie(x,radius=1,labels=piepercent,main="2008",col=rainbow(length(x)))

> legend("topright",c("Assam","Bihar","Chattisgarh","Gujarat","Haryana","Himachal
Pradesh","Jammu & Kashmir","Jharkhand","Karnataka","Madhya
Pradesh","Maharashtra","Meghalaya","Odisha","Punjab","Rajasthan","Uttar
Pradesh","Uttarakhand","West Bengal","Delhi"),fill=rainbow(length(x)),cex=0.8,inset=c(-0.4,0))
```

FOR 2009-

```
>
x=c(1090,2043,1040,2377,4390,1520,1735,1541,918,1723,1483,1750,1396,4462,3175,3002,2003,24
90,4351)

> piepercent=round(100*x/sum(x),1)

> pie(x,radius=1,labels=piepercent,main="2009",col=rainbow(length(x)))

> legend("topright",c("Assam","Bihar","Chattisgarh","Gujarat","Haryana","Himachal
Pradesh","Jammu & Kashmir","Jharkhand","Karnataka","Madhya
Pradesh","Maharashtra","Meghalaya","Odisha","Punjab","Rajasthan","Uttar
Pradesh","Uttarakhand","West Bengal","Delhi"),fill=rainbow(length(x)),cex=0.8,inset=c(-0.4,0))
```

FOR 2010-

```
>
x=c(1087,2084,1086,2697,4213,928,1003,1738,887,1967,1610,1773,1450,4307,3133,2846,2139,268
0,4352)

> piepercent=round(100*x/sum(x),1)

> pie(x,radius=1,labels=piepercent,main="2010",col=rainbow(length(x)))

> legend("topright",c("Assam","Bihar","Chattisgarh","Gujarat","Haryana","Himachal
Pradesh","Jammu & Kashmir","Jharkhand","Karnataka","Madhya
Pradesh","Maharashtra","Meghalaya","Odisha","Punjab","Rajasthan","Uttar
Pradesh","Uttarakhand","West Bengal","Delhi"),fill=rainbow(length(x)),cex=0.8,inset=c(-0.4,0))
```

FOR 2011-

```
>
x=c(1179,1948,1144,3155,4624,1530,1535,1642,1094,1757,1761,1791,1458,4693,2910,3113,2316,2
760,4340)

> piepercent=round(100*x/sum(x),1)

> pie(x,radius=1,labels=piepercent,main="2011",col=rainbow(length(x)))

> legend("topright",c("Assam","Bihar","Chattisgarh","Gujarat","Haryana","Himachal
Pradesh","Jammu & Kashmir","Jharkhand","Karnataka","Madhya
Pradesh","Maharashtra","Meghalaya","Odisha","Punjab","Rajasthan","Uttar
Pradesh","Uttarakhand","West Bengal","Delhi"),fill=rainbow(length(x)),cex=0.8,inset=c(-0.4,0))
```

FOR 2012-

```
>
x=c(1147,2206,1227,3014,5030,1671,1689,1908,858,2360,1558,1564,1644,4898,3175,3113,2379,27
65,4349)

> piepercent=round(100*x/sum(x),1)
```

```
> pie(x,radius=1,labels=piepercent,main="2012",col=rainbow(length(x)))

> legend("topright",c("Assam","Bihar","Chattisgarh","Gujarat","Haryana","Himachal Pradesh","Jammu & Kashmir","Jharkhand","Karnataka","Madhya Pradesh","Maharashtra","Meghalaya","Odisha","Punjab","Rajasthan","Uttar Pradesh","Uttarakhand","West Bengal","Delhi"),fill=rainbow(length(x)),cex=0.8,inset=c(-0.4,0))
```

FOR 2013-

```
> x=c(1304,2427,1396,2875,4452,1671,1595,1944,796,2478,1528,1806,1894,4724,3028,3113,2396,2786,4406)

> piepercent=round(100*x/sum(x),1)

> pie(x,radius=1,labels=piepercent,main="2013",col=rainbow(length(x)))

> legend("topright",c("Assam","Bihar","Chattisgarh","Gujarat","Haryana","Himachal Pradesh","Jammu & Kashmir","Jharkhand","Karnataka","Madhya Pradesh","Maharashtra","Meghalaya","Odisha","Punjab","Rajasthan","Uttar Pradesh","Uttarakhand","West Bengal","Delhi"),fill=rainbow(length(x)),cex=0.8,inset=c(-0.4,0))
```

FOR 2014-

```
> x=c(1292,2358,1304,3255,4722,1873,2061,2123,1005,2405,1460,1881,1574,5017,3083,3038,2422,2791,4418)

> piepercent=round(100*x/sum(x),1)

> pie(x,radius=1,labels=piepercent,main="2014",col=rainbow(length(x)))

> legend("topright",c("Assam","Bihar","Chattisgarh","Gujarat","Haryana","Himachal Pradesh","Jammu & Kashmir","Jharkhand","Karnataka","Madhya Pradesh","Maharashtra","Meghalaya","Odisha","Punjab","Rajasthan","Uttar Pradesh","Uttarakhand","West Bengal","Delhi"),fill=rainbow(length(x)),cex=0.8,inset=c(-0.4,0))
```

FOR 2015-

```
> x=c(1216,1851,1388,2751,3981,1957,979,1931,1318,2850,1226,1909,1650,4294,2961,2277,1881,2807,4419)

> piepercent=round(100*x/sum(x),1)

> pie(x,radius=1,labels=piepercent,main="2015",col=rainbow(length(x)))

> legend("topright",c("Assam","Bihar","Chattisgarh","Gujarat","Haryana","Himachal Pradesh","Jammu & Kashmir","Jharkhand","Karnataka","Madhya Pradesh","Maharashtra","Meghalaya","Odisha","Punjab","Rajasthan","Uttar Pradesh","Uttarakhand","West Bengal","Delhi"),fill=rainbow(length(x)),cex=0.8,inset=c(-0.4,0))
```

FOR 2016-

```
> x=c(1634,2244,1270,2919,4407,1958,1947,1835,897,2993,1077,1953,1324,4583,3175,2636,2258,2825,4419)
> piepercent=round(100*x/sum(x),1)
> pie(x,radius=1,labels=piepercent,main="2016",col=rainbow(length(x)))
> legend("topright",c("Assam","Bihar","Chattisgarh","Gujarat","Haryana","Himachal Pradesh","Jammu & Kashmir","Jharkhand","Karnataka","Madhya Pradesh","Maharashtra","Meghalaya","Odisha","Punjab","Rajasthan","Uttar Pradesh","Uttarakhand","West Bengal","Delhi"),fill=rainbow(length(x)),cex=0.8,inset=c(-0.4,0))
```

FOR 2017-

```
> x=c(1344,2427,1391,2751,4514,2033,1638,2011,1018,2976,1474,1913,1333,4704,3175,3113,2587,2682,4434)
> piepercent=round(100*x/sum(x),1)
> pie(x,radius=1,labels=piepercent,main="2017",col=rainbow(length(x)))
> legend("topright",c("Assam","Bihar","Chattisgarh","Gujarat","Haryana","Himachal Pradesh","Jammu & Kashmir","Jharkhand","Karnataka","Madhya Pradesh","Maharashtra","Meghalaya","Odisha","Punjab","Rajasthan","Uttar Pradesh","Uttarakhand","West Bengal","Delhi"),fill=rainbow(length(x)),cex=0.8,inset=c(-0.4,0))
```

- ***R CODES FOR LINE DIAGRAMS-***

FIG 2.1 :

```
> x=c(1074,1117,1268,1090,1087,1179,1147,1304,1292,1216,1634,1344)
> y=c(2006,2007,2008,2009,2010,2011,2012,2013,2014,2015,2016,2017)
> plot(y,x,main="Yields of States",xlab="Year",ylab="Yield",type="l",lwd=4.0,col="blue",ylim=c(886,3255))
> z=c(1617,1908,2058,2043,2084,1984,2206,2427,2358,1851,2244,2427)
> lines(y,z,col="red",type="l",lwd=4.0)
> t=c(886,1002,1059,1040,1086,1144,1227,1396,1304,1388,1270,1391)
> lines(y,t,col="green",type="l",lwd=4.0)
> u=c(2700,2498,3013,2377,2679,3155,3014,2875,3255,2751,2919,2751)
> lines(y,u,col="orange",type="l",lwd=4.0)
```

```
>
legend("bottomright",c("Assam","Bihar","Chattisgarh","Gujarat"),fill=c("blue","red","green","orange
"),cex=0.8)
```

FIG 2.2 :

```
> x=c(3844,4232,4158,4390,4213,4624,5030,4452,4722,3981,4407,4514)

> y=c(2006,2007,2008,2009,2010,2011,2012,2013,2014,2015,2016,2017)

> plot(y,x,main="Yields of
States",xlab="Year",ylab="Yield",type="l",lwd=4.0,col="blue",ylim=c(928,5030))

> z=c(1894,1385,1376,1520,928,1530,1671,1671,1873,1957,1958,2033)

> lines(y,z,col="red",type="l",lwd=4.0)

> t=c(1790,1893,1782,1735,1003,1535,1689,1595,2061,979,1947,1638)

> lines(y,t,col="green",type="l",lwd=4.0)

> u=c(1340,1529,1621,1541,1738,1642,1908,1944,2123,1931,1835,2011)

> lines(y,u,col="orange",type="l”,lwd=4.0)

> legend("center",c("Haryana","Himachal Pradesh","Jammu &
Kashmir","Jharkhand"),fill=c("blue","red","green","orange"),cex=0.9)
```

FIG 2.3 :

```
> x=c(858,762,946,918,887,1094,858,796,1005,1318,897,1018)

> y=c(2006,2007,2008,2009,2010,2011,2012,2013,2014,2015,2016,2017)

> plot(y,x,main="Yields of
States",xlab="Year",ylab="Yield",type="l",lwd=4.0,col="blue",ylim=c(762,2993))

> z=c(1613,1835,1612,1723,1967,1757,2360,2478,2405,2850,2993,2976)

> lines(y,z,col="red",type="l",lwd=4.0)

> t=c(1393,1325,1659,1483,1610,1761,1558,1528,1460,1226,1077,1474)

> lines(y,t,col="green",type="l",lwd=4.0)

> u=c(1714,2000,1833,1750,1773,1791,1564,1806,1881,1909,1953,1913)

> lines(y,u,col="orange",type="l",lwd=4.0)

> legend("topleft",c("Karnataka","Madhya
Pradesh","Maharashtra","Meghalaya"),fill=c("blue","red","green","orange"),cex=0.9)
```

FIG 2.4 :

```
> x=c(1364,1487,1554,1396,1450,1458,1644,1894,1574,1650,1324,1333)

> z=c(4179,4210,4507,4462,4307,4693,4898,4724,5017,4294,4583,4704)

> t=c(2762,2751,2749,3175,3133,2910,3175,3028,3083,2961,3175,3175)
```

```
> u=c(2627,2721,2817,3002,2846,3113,3113,3113,3038,2277,2636,3113)

> y=c(2006,2007,2008,2009,2010,2011,2012,2013,2014,2015,2016,2017)

> plot(y,x,main="Yields of
States",xlab="Year",ylab="Yield",type="l",lwd=4.0,col="blue",ylim=c(1324,5017))

> lines(y,z,col="red",type="l",lwd=4.0)

> lines(y,t,col="green",type="l",lwd=4.0)

> lines(y,u,col="orange",type="l",lwd=4.0)

> legend("topleft",c("Odisha","Punjab","Rajasthan","Uttar
Pradesh"),fill=c("blue","red","green","orange"),cex=0.8)
```

FIG 2.5 :

```
> x=c(1633,2049,2050,2003,2139,2316,2379,2396,2422,1881,2258,2587)

> y=c(2006,2007,2008,2009,2010,2011,2012,2013,2014,2015,2016,2017)

> plot(y,x,main="Yields of
States",xlab="Year",ylab="Yield",type="l",col="blue",lwd=4.0,ylim=c(1633,4419))

> z=c(2109,2282,2602,2490,2680,2760,2765,2786,2791,2807,2825,2682)

> lines(y,z,col="red",type="l",lwd=4.0)

> t=c(4339,4341,4354,4351,4351,4340,4349,4406,4418,4419,4419,4434)

> plot(y,x,main="Yields of
States",xlab="Year",ylab="Yield",type="l",col="blue",lwd=4.0,ylim=c(1633,4434))

> z=c(2109,2282,2602,2490,2680,2760,2765,2786,2791,2807,2825,2682)

> lines(y,z,col="red",type="l",lwd=4.0)

> t=c(4339,4341,4354,4351,4351,4340,4349,4406,4418,4419,4419,4434)

> lines(y,t,col="green",type="l",lwd=4.0)

> legend("center",c("Uttarakhand","West Bengal","Delhi"),fill=c("blue","red","green"),cex=0.8)
```

Collected data on State-Wise Wheat Yield(in Kg/hectare) Of 19 Major States in India from 2006-2017:

State	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
ASSAM	1074	1117	1268	1090	1087	1179	1147	1304	1292	1216	1634	1344
BIHAR	1617	1908	2058	2043	2084	1948	2206	2427	2358	1851	2244	2427
CHATTISGARH	886	1002	1059	1040	1086	1144	1221	1396	1304	1388	1270	1391
GUJARAT	2700	2498	3013	2377	2679	3155	3014	2875	3255	2751	2919	2751
HARYANA	3844	4232	4158	4390	4213	4624	5030	4452	4722	3981	4407	4514
HIMACHAL PRADESH	1894	1385	1376	1520	928	1530	1671	1671	1873	1957	1958	2033
JAMMU & KASHMIR	1790	1893	1782	1735	1003	1535	1689	1595	2061	979	1947	1638
JHARKHAND	1340	1529	1621	1541	1738	1642	1908	1944	2123	1931	1835	2011
KARNATAKA	858	762	946	918	887	1094	858	796	1005	1318	897	1018
MADHYA PRADESH	1613	1835	1612	1723	1967	1757	2360	2478	2405	2850	2993	2976
MAHARASHTRA	1393	1325	1659	1483	1610	1761	1558	1528	1460	1226	1077	1474
MEGHALAYA	1714	2000	1833	1750	1773	1791	1564	1806	1881	1909	1953	1913
ODISHA	1364	1487	1554	1396	1450	1458	1644	1894	1574	1650	1324	1333
PUNJAB	4179	4210	4507	4462	4307	4693	4898	4724	5017	4294	4583	4704
RAJASTHAN	2762	2751	2749	3175	3133	2910	3175	3028	3083	2961	3175	3175
UTTAR PRADESH	2627	2721	2817	3002	2846	3113	3113	3113	3038	2277	2636	3113
UTTARAKHAND	1633	2049	2050	2003	2139	2116	2379	2396	2422	1881	2258	2587
WEST BENGAL	2109	2282	2602	2490	2680	2760	2765	2786	2791	2807	2825	2682
DELHI	4339	4341	4354	4351	4352	4340	4349	4406	4418	4419	4419	4434

- ***DECLARATION:***

I, Ahana Bose, a student of B.Sc. Sem-VI, Statistics Honours, of University of Calcutta, Registration No: 012-1211-0535-19; Roll No: 193012-11-0286, hereby declare that I have done this piece of project work entitled as "WHEAT PRODUCTION OF INDIA- A CASE STUDY & DETAILED DATA ANALYSIS" under the supervision of Ms. Oindrila Bose, Professor of 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 this piece of project has not been published elsewhere for any degree or diploma or taken from any published project.