CAT II PROJECT DOCUMENT

**17MDC36 – Business Statistics Lab using R**

Yield Production Analysis

**Team Members:**

* Saran Nithish T S - 2033031
* Dharun Prudhiv M - 2033010
* Dhiyanaesh P - 2033011

**Introduction:**

Yield prediction is an important agricultural problem. Every farmer is interested in knowing, how much yield he is about expect. In the past, yield prediction was performed by considering farmer's previous experience on a particular crop. The volume of data is enormous in Indian agriculture. The data when become information is highly useful for many purposes.

**Problem Statement:**

Forecasting future of commodity crops statistically and helping farmers cope up with their crop’s yield, production cost and cultivation cost.

**Objective:**

* To analyze and recommend the suitable commodity product for the farmers.
* To forecast the Cost of Production and Cost of Cultivation of the commodity over years.
* To visualize the abstracted information from the dataset

**Dataset Used:**

<https://www.kaggle.com/srinivas1/agricuture-crops-production-in-india?select=datafile+%281%29.csv>

- Agricultural Production in India (2013-2017)

- 43 columns

- Various Crops cultivation and Production

**Tool Description:**

**Libraries & Packages:**

library (ggplot2)

library (AICcmodavg)

library (qcc)

library (forecast)

**Software Used:**

R Studio

**Code:**

**One Way Anova and C- chart:**

library(ggplot2)

library (AICcmodavg)

library(qcc)

#Importing Data

c.dt <- read.csv ("D:/Remainder/Crops.csv", header = TRUE, colClasses = c("factor","factor","numeric","numeric","numeric"))

summary(c.dt)

#One-way ANOVA

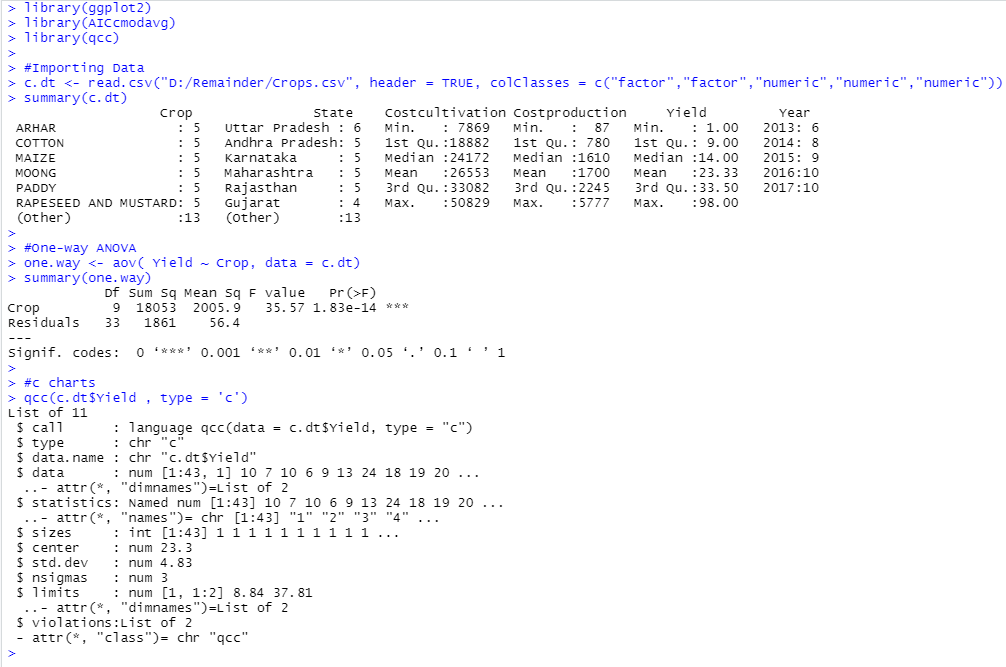
one.way <- aov(Yield ~ Crop, data = c.dt)

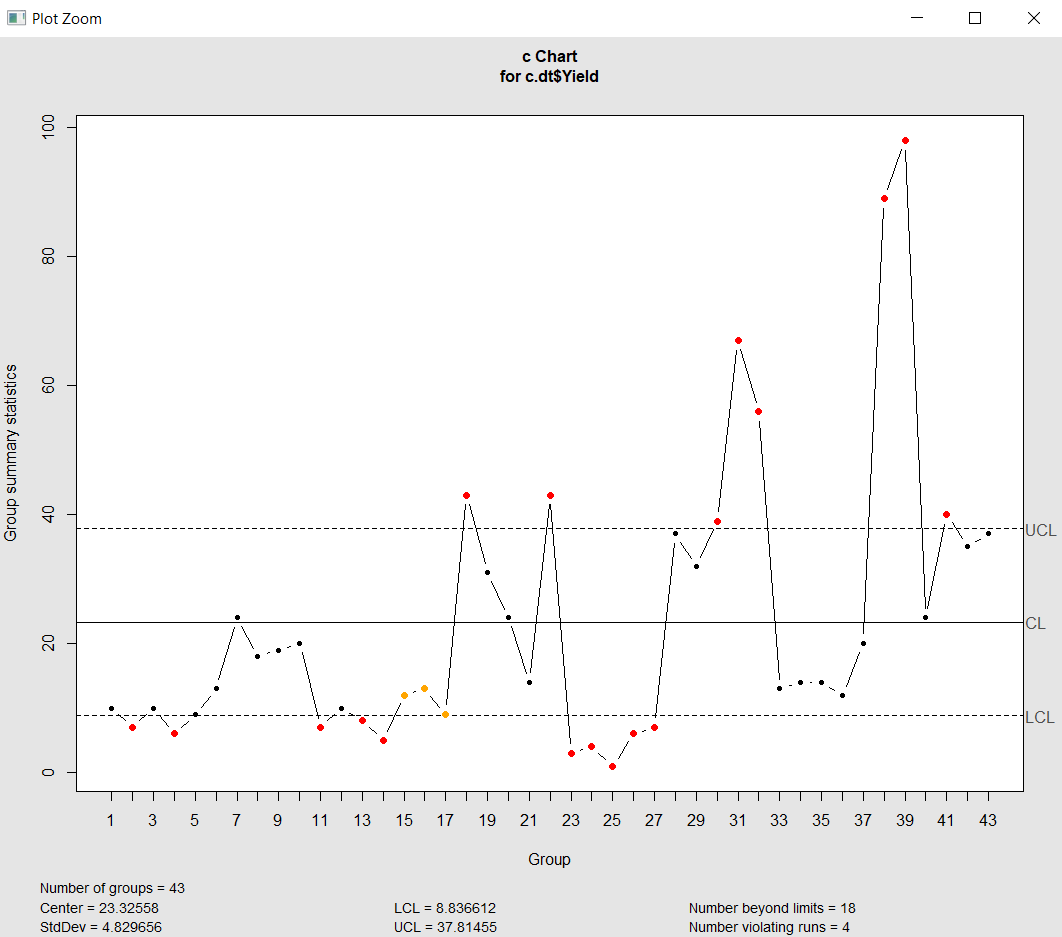
summary(one.way)

#c charts

qcc(c.dt$Yield , type = 'c')

**Output:**





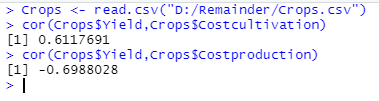
**Correlation:**

Crops <- read.csv("D:/Remainder/Crops.csv")

cor(Crops$Yield,Crops$Costcultivation)

cor(Crops$Yield,Crops$Costproduction)

**Output:**



**Linear Regression for Crop’s Yield vs Crop’s CostofCultivation:**

Crops <- read.csv("D:/Remainder/Crops.csv")

#Linear Regression

x <- c(Crops$Crop)

y <- c(Crops$Yield)

relation <- lm(y~x)

print(relation)

#print(summary(relation))

z <- c(Crops$Costcultivation)

y <- c(Crops$Yield)

relation1 <- lm(z~y)

print(relation1)

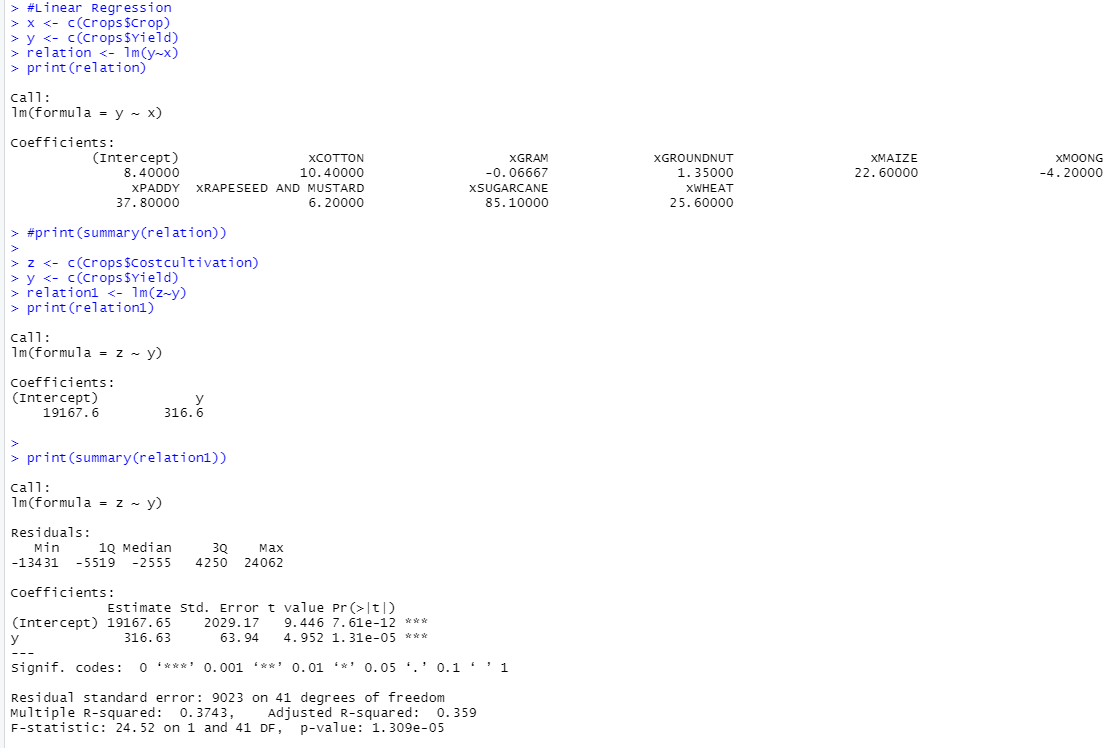
print(summary(relation1))

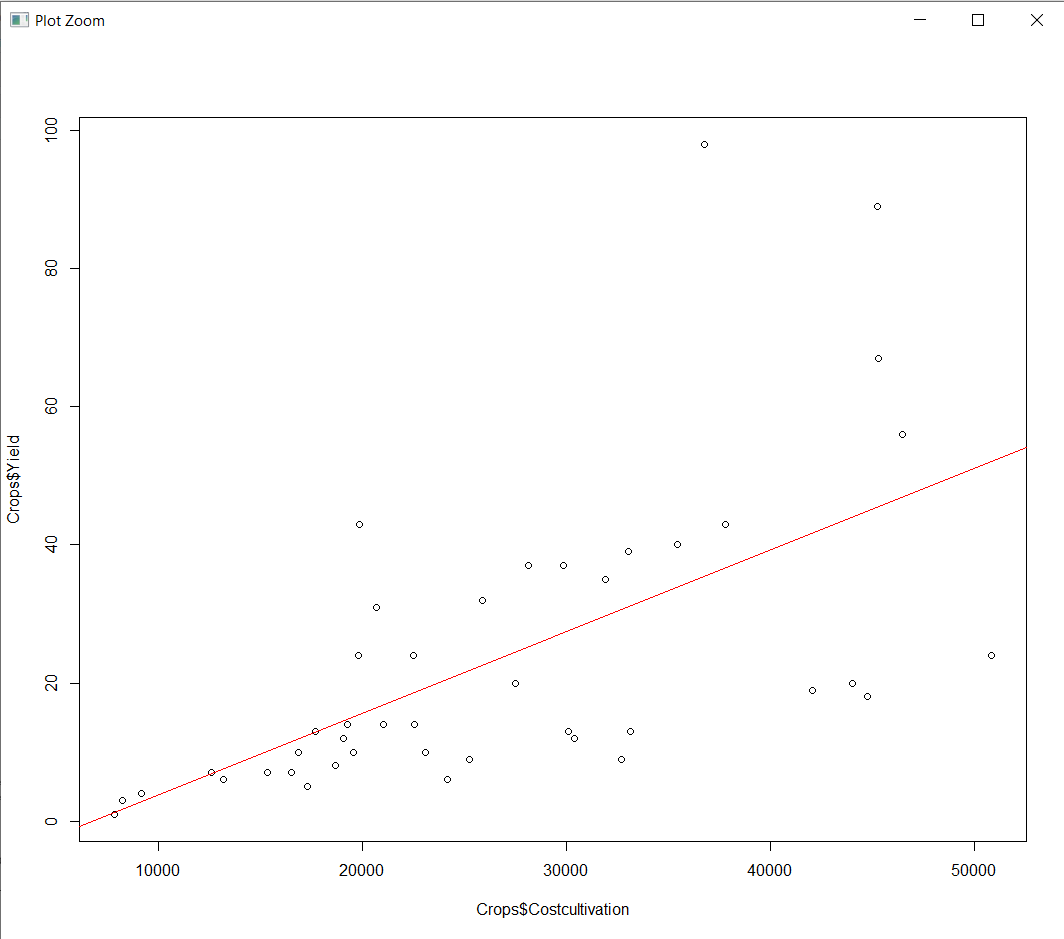
plot(Crops$Costcultivation,Crops$Yield)

model2=lm(Yield~Costcultivation,data = Crops)

abline(model2, col="red")

**Output:**

****

**  
Linear Regression for Crop’s Yield vs Crop’s CostofProduction:**

Crops <- read.csv("D:/Remainder/Crops.csv")

#Linear Regression

x <- c(Crops$Crop)

y <- c(Crops$Yield)

relation <- lm(y~x)

print(relation)

#print(summary(relation))

z <- c(Crops$Costproduction)

y <- c(Crops$Yield)

relation1 <- lm(z~y)

print(relation1)

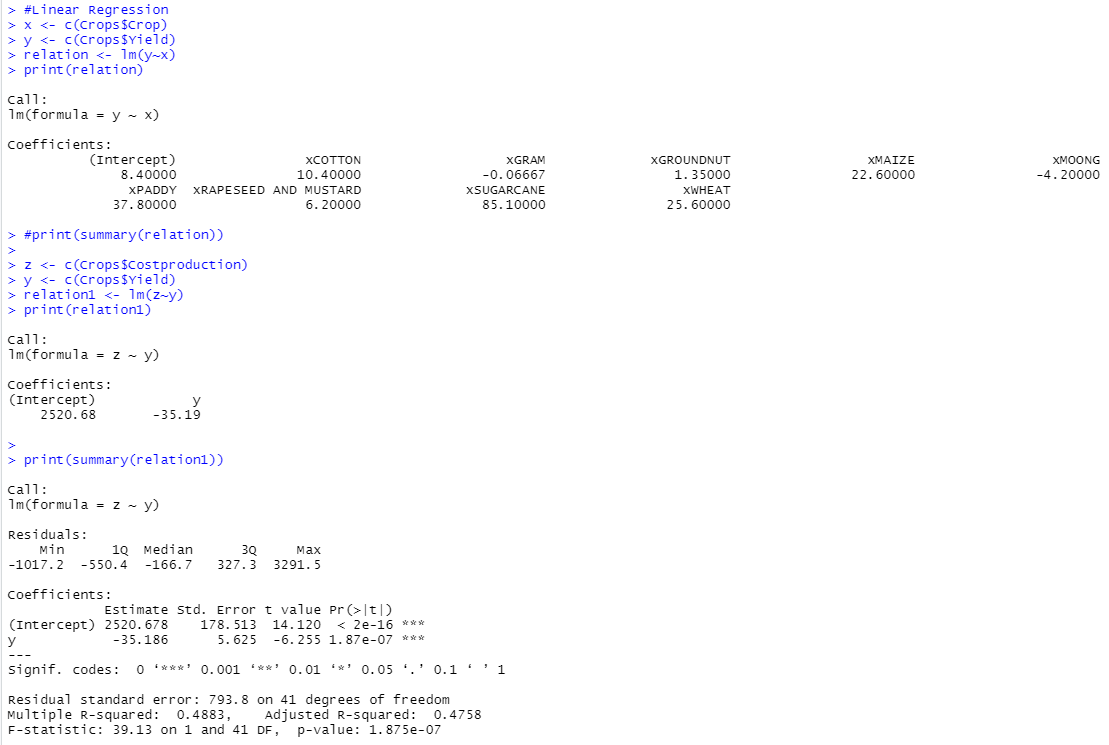
print(summary(relation1))

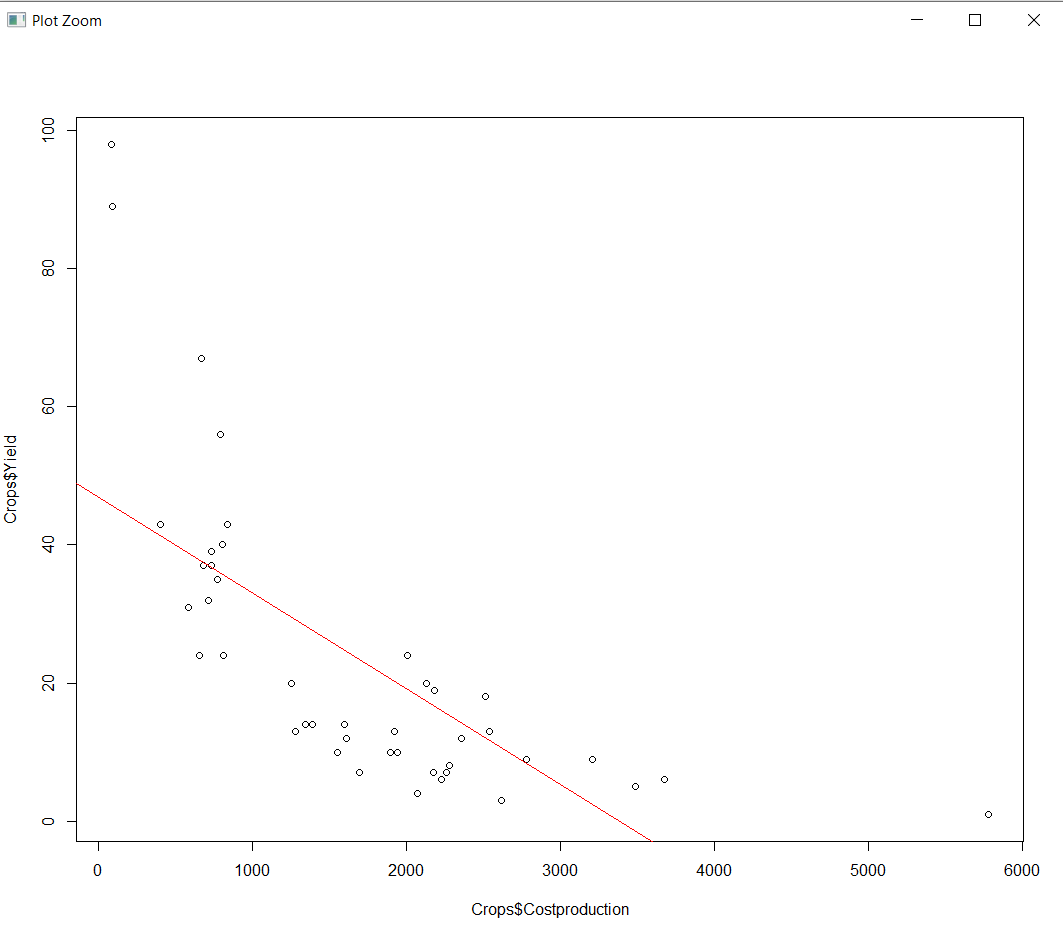
plot(Crops$Costproduction,Crops$Yield)

model2=lm(Yield~Costproduction,data = Crops)

abline(model2, col="red")

**Output:**

****

****

**Time series Graph:**

Crops <- read.csv("D:/Remainder/Crops3.csv")

# Convert it to a time series object.

Crops.timeseries <- ts(Crops,start = 2013,end=2017,frequency = 1)

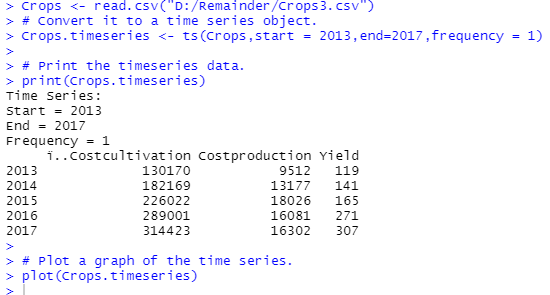
# Print the timeseries data.

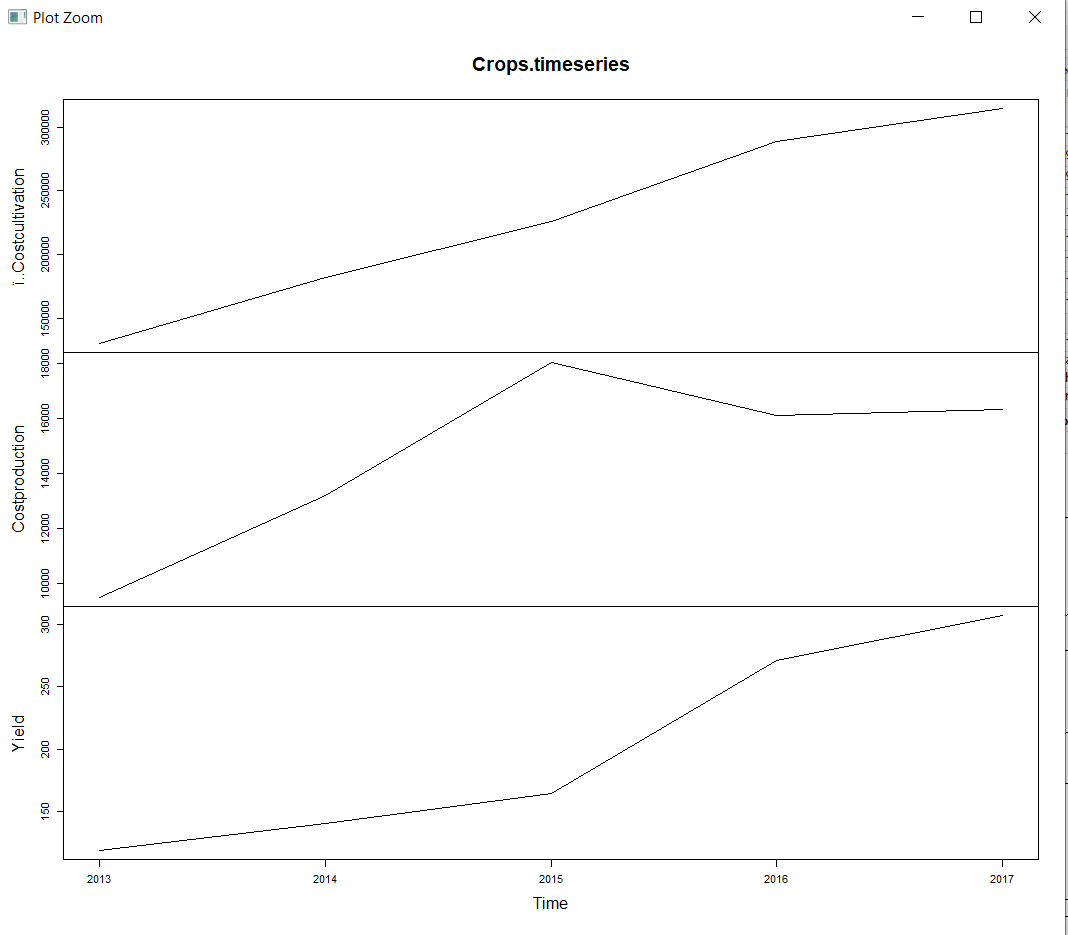
print(Crops.timeseries)

# Plot a graph of the time series.

plot(Crops.timeseries)

**Output:**

****

****

**Time Series forecasting for Cost of Production:**

library(forecast)

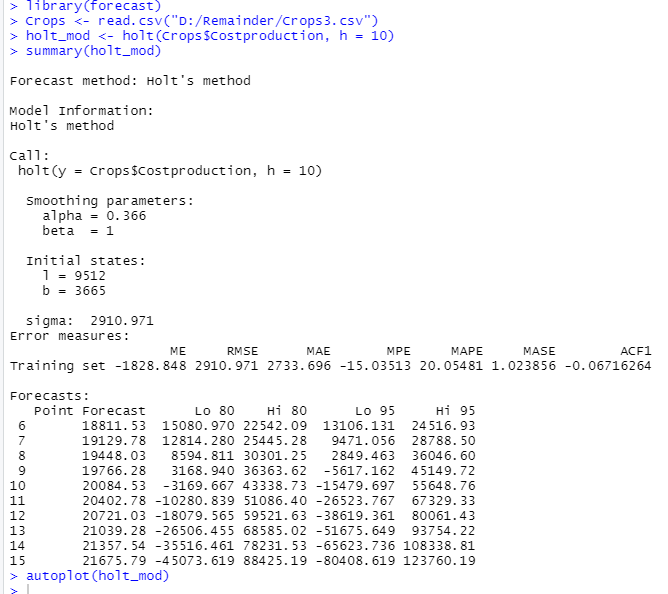
Crops <- read.csv("D:/Remainder/Crops3.csv")

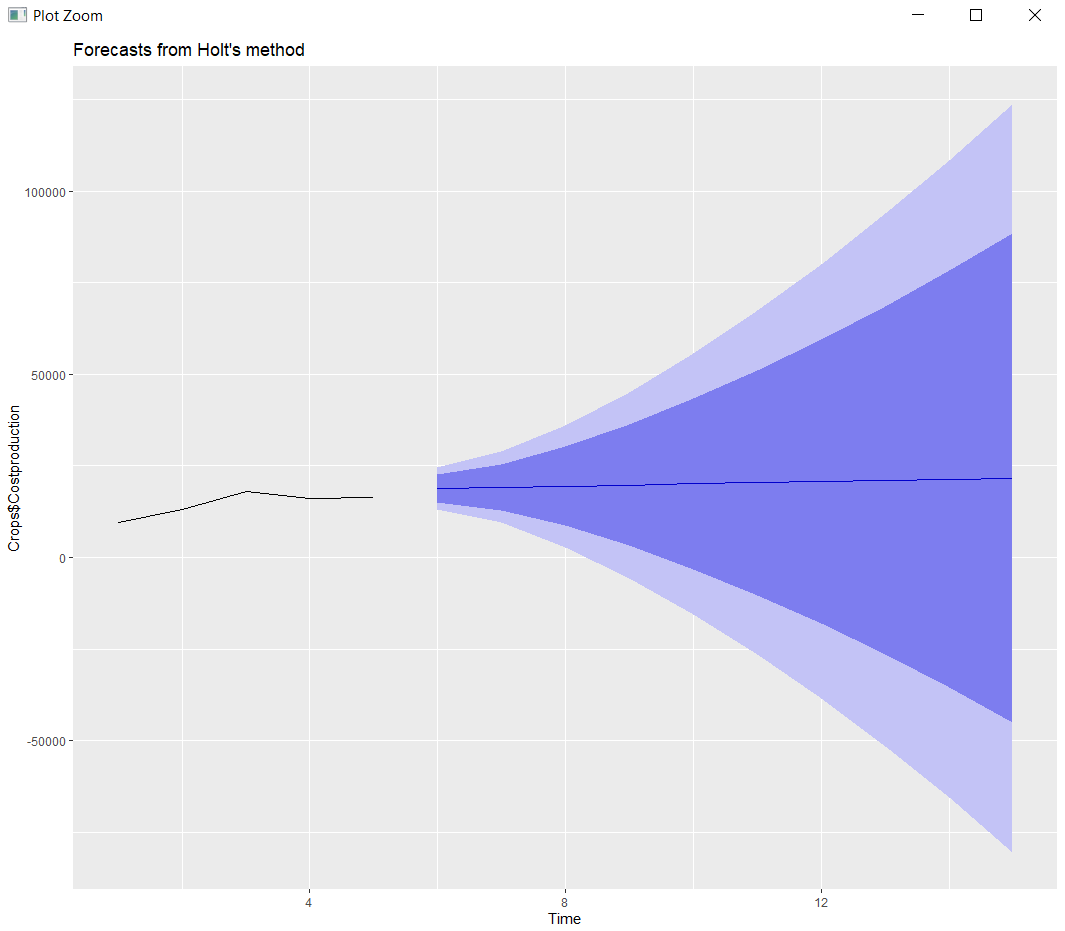
holt\_mod <- holt(Crops$Costproduction, h = 10)

summary(holt\_mod)

autoplot(holt\_mod)

**Output:**





**Time Series forecasting for Cost of Cultivation:**

library(forecast)

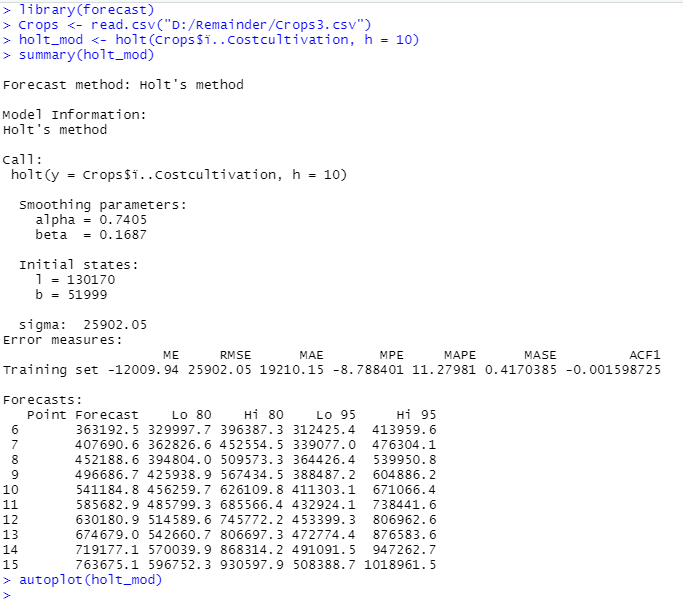
Crops <- read.csv("D:/Remainder/Crops3.csv")

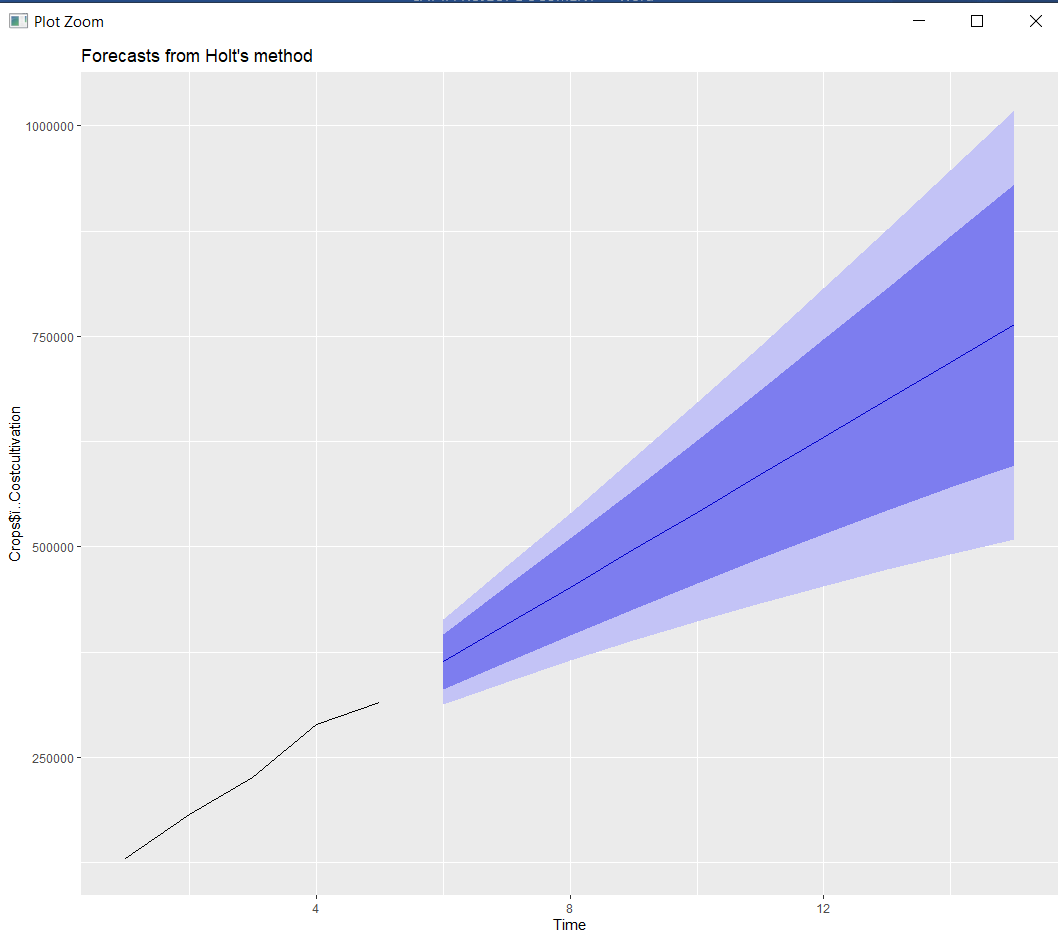
holt\_mod <- holt(Crops$ï..Costcultivation, h = 10)

summary(holt\_mod)

autoplot(holt\_mod)

**Output:**

****

****

**Time Series forecasting for Crop’s Yield:**

library(forecast)

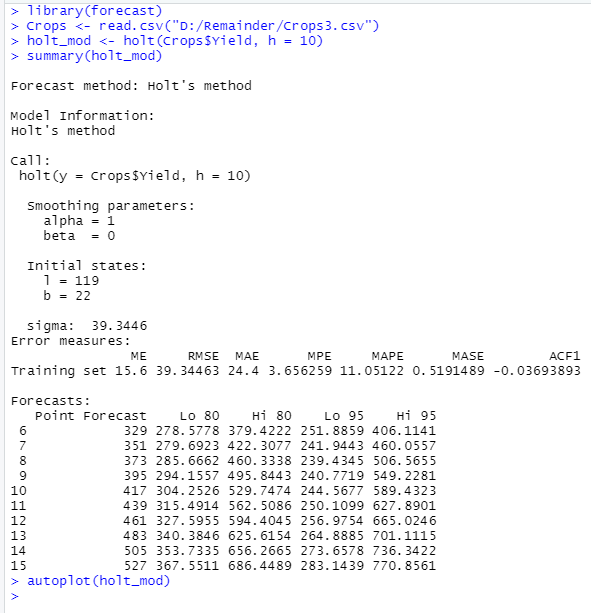
Crops <- read.csv("D:/Remainder/Crops3.csv")

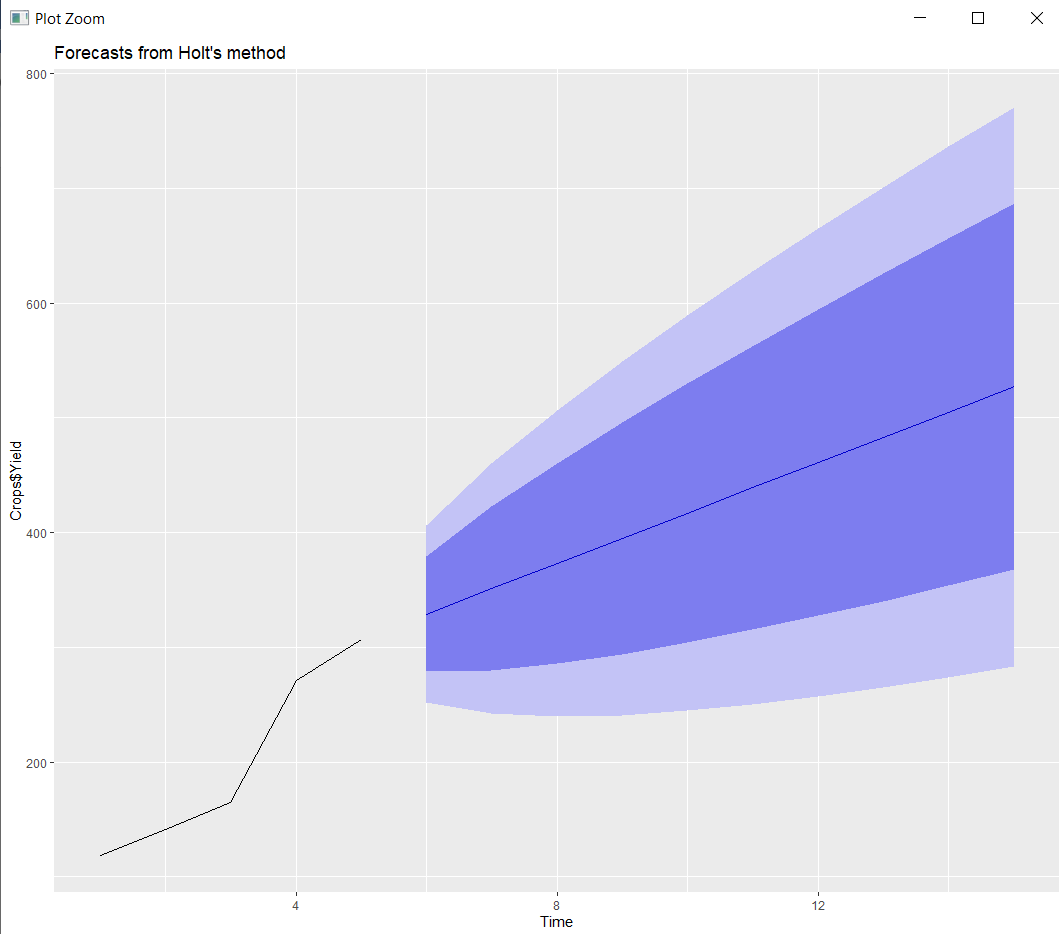
holt\_mod <- holt(Crops$Yield, h = 10)

summary(holt\_mod)

autoplot(holt\_mod)

**Output:**

****

****

**Performance Measure:**

1) One – Way Anova:

Df Sum Sq Mean Sq F value Pr(>F)

Crop 9 18053 2005.9 35.57 1.83e-14 \*\*\*

Residuals 33 1861 56.4

2) Correlation:

cor(Crops$Yield,Crops$Costcultivation)

0.6117691 (Positive value)

Here the value is positive which shows that Yield and Cost of cultivation are weakly dependent to each other.

cor(Crops$Yield,Crops$Costproduction)

-0.6988028 (Negative value)

Here the value is negative which shows that Yield and Cost of production are independent.

3) Regression:

Crop Yield vs Cost of Cultivation

Coefficients:

(Intercept) y

19167.6 316.6

When the correlation is positive, the regression slope will be positive. Cost of crop cultivation and Crop yield have strong effects on each other.

Crop Yield vs Cost of Production

Coefficients:

(Intercept) y

2520.68 -35.19

When the correlation is negative, the regression slope will be negative.

**Analysis report and inference:**

1) Correlation:

cor(Crops$Yield,Crops$Costcultivation) (0.6117691 (Positive value))

Here the value is positive which shows that Yield and Cost of cultivation are weakly dependent to each other.

cor(Crops$Yield,Crops$Costproduction) (-0.6988028 (Negative value))

Here the value is negative which shows that Yield and Cost of production are independent.

2) Linear Regression:

Crop Yield vs Cost of Cultivation

When the correlation is positive, the regression slope will be positive. Cost of crop cultivation and Crop yield have strong effects on each other.

Crop Yield vs Cost of Production

When the correlation is negative, the regression slope will be negative.

3) Time Series Forecasting:

* Yield:

Using Holt’s method, we have forecasted the next 10 years yield which shows that there will be no decrease in Yield of the chosen crops.

* Cost of Cultivation:

Using Holt’s method, we have forecasted the next 10 years yield which shows that there will be no decrease in Cost of Cultivation. We can also verify by the correlation value, which is positive so when the Yield increases Cost of Cultivation also increases.

* Cost of Production:

Using Holt’s method, we have forecasted the next 10 years yield which shows that there may be an increase or decrease in the Cost of Production. We can also verify by the correlation value, which is negative so when the Yield increases Cost of Production may increase or decrease.

**Conclusion:**

Using the Above methods, we have forecasted the future of commodity crops statistically and helping farmers cope up with their crop’s yield, production cost and cultivation cost.