**PROBLEM STATEMENT**

To find the optimal conditions for the growth of *Streptomyces erythreus* in order to test the effectiveness of its antibiotic *Erythromycin.*

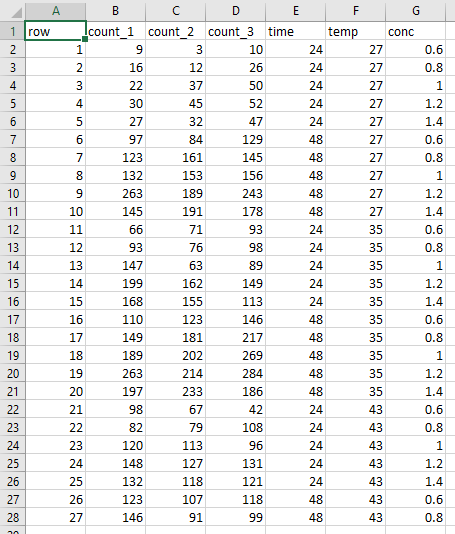
**ABSTRACT**

Bacteria are cultured in medical laboratories to identify them so patients can be treated correctly.

The dataset contains measurements of *Streptomyces erythreus* counts following the culturing of three strains of *Erythromycin - Resistant Streptomyces erythreus*, incubation time of bacteria, incubation temperature and concentration of T-broth to provide the basic nutrients and growth factors needed to support bacterial growth.

ANOVA is used with one-way, two-way, and factorial models with interactions, to identify the significant factors. Optimal factors are found for each strain from box plots.

**DATASET**

****

**CODE**

library(ggpubr)

library(DescTools)

library(ggplot2)

library(plotly)

data <- read.csv("C:/Users/ritik/OneDrive/Desktop/SEM 3/LAB - COMPUTING III/anova.csv", header = TRUE)

data

#Describing Dataset

summary(data)

# 2 sample t - test

t.test(data$count\_1, data$count\_2, var.equal = FALSE)

t.test(data$count\_2, data$count\_3, var.equal = FALSE)

t.test(data$count\_3, data$count\_1, var.equal = FALSE)

#p value is greater than 5%, data is significant to each other(accepts the null hypothesis)

#COUNT\_1 analysis

#checking if variable fits a normal distribution

h<-hist(data$count\_1,col="red",xlab="Count 1",main="Checking Normality")

x<-data$count\_1

xfit<-seq(min(x),max(x),length=40)

yfit<-dnorm(xfit,mean=mean(x),sd=sd(x))

yfit <- yfit\*diff(h$mids[1:2])\*length(x)

lines(xfit, yfit, col="blue", lwd=2)

#plot

plot\_ly(data=data, y = ~data$count\_1, x=~data$time, color = I("red"),type="box",

alpha = 0.5, boxpoints = "suspectedoutliers",name = "count1")

#show that 48 hours give higher growth than 24 hours

plot\_ly(data=data, y = ~data$count\_1, x=~data$conc, color = I("red"),type="box",

alpha = 0.5, boxpoints = "suspectedoutliers",name = "count1")

#optimal concentration is about 1.2%

plot\_ly(data=data, y = ~data$count\_1, x=~data$temp, color = I("red"),type="box",

alpha = 0.5, boxpoints = "suspectedoutliers",name = "count1")

#optimal temperature is about 35 degrees

#One way anova

aov.res1<- aov(count\_1 ~ time, data = data)

aov.resu1<- aov(count\_1 ~ temp, data = data)

aov.resul1<- aov(count\_1 ~ conc, data = data)

summary(aov.res1)

summary(aov.resul1)

summary(aov.resu1)

plot(aov.res1,1)#Homogenity(check variation among groups)

plot(aov.res1,2)#Normality

plot(aov.resu1,1)

plot(aov.resu1,2)

plot(aov.resul1,1)

plot(aov.resul1,2)

#Two way anova

aov2.resul1<- aov(count\_1 ~ conc + time, data = data)

summary(aov2.resul1)

aov2.resu1<- aov(count\_1 ~ temp + conc, data = data)

summary(aov2.resu1)

aov2.res1<- aov(count\_1 ~ temp + time, data = data)

summary(aov2.res1)

plot(aov2.resul1,1)

plot(aov2.resul1,2)

#2 way plot

mean.data1=data %>% group\_by(time,conc) %>% summarise(count\_1=median(count\_1))

two.way.plot1 <- ggplot(data,aes(x=conc, y=count\_1, group=time))+geom\_point(cex =1.5, pch=1.0,position =position\_jitter(w=0.1, h=0))

two.way.plot1 <-two.way.plot1+stat\_summary(fun.data = 'mean\_se', geom = 'errorbar', width = 0.2) +

stat\_summary(fun.data = 'mean\_se', geom = 'pointrange') +

geom\_point(data=mean.data1, aes(x=conc, y=count\_1)+

facet\_wrap(~ conc))

two.way.plot1 <- two.way.plot1 +

labs(title = "Count 1 in response to time and concentration",

x = "CONC of Tryptone by weight",

y = "COUNT1")

two.way.plot1

#Interactions(3 null hypotheses)

i.resul1<- aov(count\_1 ~ conc \* time, data = data)

i.resu1<- aov(count\_1 ~ temp \* time, data = data)

i.res1<- aov(count\_1 ~ temp \* conc, data = data)

summary(i.resul1)

summary(i.resu1)

summary(i.res1)

#COUNT\_2 analysis

#checking if variable fits a normal distribution

h<-hist(data$count\_2,col="magenta",xlab="Count 2",main="Checking Normality")

x<-data$count\_2

xfit<-seq(min(x),max(x),length=40)

yfit<-dnorm(xfit,mean=mean(x),sd=sd(x))

yfit <- yfit\*diff(h$mids[1:2])\*length(x)

lines(xfit, yfit, col="blue", lwd=2)

#plot

p1 <- plot\_ly(data=data, y = ~data$count\_2, x=~data$time, color = I("magenta"),type="box",

alpha = 0.5, boxpoints = "suspectedoutliers",name = "count2")

p1

c1 <- plot\_ly(data=data, y = ~data$count\_2, x=~data$conc, color = I("magenta"),type="box",

alpha = 0.5, boxpoints = "suspectedoutliers",name = "count2")

c1

T1 <- plot\_ly(data=data, y = ~data$count\_2, x=~data$temp, color = I("magenta"),type="box",

alpha = 0.5, boxpoints = "suspectedoutliers",name = "count2")

T1

#Oneway anova

aov.res2<- aov(count\_2 ~ time, data = data)

aov.resu2<- aov(count\_2 ~ temp, data = data)

aov.resul2<- aov(count\_2 ~ conc, data = data)

summary(aov.res2)

summary(aov.resu2)

summary(aov.resul2)

plot(aov.res2,1)

plot(aov.res2,2)

plot(aov.resu2,1)

plot(aov.resu2,2)

plot(aov.resul2,1)

plot(aov.resul2,2)

#Twoway anova

aov2.resul2<- aov(count\_2 ~ conc + time, data = data)

summary(aov2.resul2)

aov2.resu2<- aov(count\_2 ~ temp + conc, data = data)

summary(aov2.resu2)

aov2.res2<- aov(count\_2 ~ temp + time, data = data)

summary(aov2.res2)

plot(aov2.resul2,1)

plot(aov2.resul2,2)

mean.data2=data %>% group\_by(time,conc) %>% summarise(count\_2=median(count\_2))

two.way.plot2 <- ggplot(data,aes(x=conc, y=count\_2, group=time))+geom\_point(cex =1.5, pch=1.0,position =position\_jitter(w=0.1, h=0))

two.way.plot2 <-two.way.plot2+stat\_summary(fun.data = 'mean\_se', geom = 'errorbar', width = 0.2) +

stat\_summary(fun.data = 'mean\_se', geom = 'pointrange') +

geom\_point(data=mean.data2, aes(x=conc, y=count\_2)+

facet\_wrap(~ conc))

two.way.plot2 <- two.way.plot2 +

labs(title = "Count 2 in response to time and concentration",

x = "CONC of Tryptone by weight",

y = "COUNT2")

two.way.plot2

#Interactions

i.resul2<- aov(count\_2 ~ conc \* time, data = data)

i.resu2<- aov(count\_2 ~ temp \* time, data = data)

i.res2<- aov(count\_2 ~ temp \* conc, data = data)

summary(i.resul2)

summary(i.resu2)

summary(i.res2)

#COUNT\_3 analysis

#checking if variable fits a normal distribution

h<-hist(data$count\_3,col="green",xlab="Count 3",main="Checking Normality")

x<-data$count\_3

xfit<-seq(min(x),max(x),length=40)

yfit<-dnorm(xfit,mean=mean(x),sd=sd(x))

yfit <- yfit\*diff(h$mids[1:2])\*length(x)

lines(xfit, yfit, col="blue", lwd=2)

#plot

p3 <- plot\_ly(data=data, y = ~data$count\_3, x=~data$time, color = I("green"),type="box",

alpha = 0.5, boxpoints = "suspectedoutliers",name = "count3")

p3

c3 <- plot\_ly(data=data, y = ~data$count\_3, x=~data$conc, color = I("green"),type="box",

alpha = 0.5, boxpoints = "suspectedoutliers",name = "count3")

c3

T3 <- plot\_ly(data=data, y = ~data$count\_3, x=~data$temp, color = I("green"),type="box",

alpha = 0.5, boxpoints = "suspectedoutliers",name = "count3")

T3

#Oneway anova

aov.res3<- aov(count\_3 ~ time, data = data)

aov.resu3<- aov(count\_3 ~ temp, data = data)

aov.resul3<- aov(count\_3 ~ conc, data = data)

summary(aov.res3)

summary(aov.resu3)

summary(aov.resul3)

plot(aov.res3,1)

plot(aov.res3,2)

plot(aov.resu3,1)

plot(aov.resu3,2)

plot(aov.resul3,1)

plot(aov.resul3,2)

#Twoway anova

aov2.resul3<- aov(count\_3 ~ conc + time, data = data)

summary(aov2.resul3)

aov2.resu3<- aov(count\_3 ~ temp + conc, data = data)

summary(aov2.resu3)

aov2.res3<- aov(count\_3 ~ temp + time, data = data)

summary(aov2.res3)

plot(aov2.resul3,1)

plot(aov2.resul3,2)

mean.data3=data %>% group\_by(time,conc) %>% summarise(count\_3=median(count\_3))

two.way.plot3 <- ggplot(data,aes(x=conc, y=count\_3, group=time))+geom\_point(cex =1.5, pch=1.0,position =position\_jitter(w=0.1, h=0))

two.way.plot3 <-two.way.plot3+stat\_summary(fun.data = 'mean\_se', geom = 'errorbar', width = 0.2) +

stat\_summary(fun.data = 'mean\_se', geom = 'pointrange') +

geom\_point(data=mean.data3, aes(x=conc, y=count\_3)+

facet\_wrap(~ conc))

two.way.plot3 <- two.way.plot3 +

labs(title = "Count 3 in response to time and concentration",

x = "CONC of Tryptone by weight",

y = "COUNT3")

two.way.plot3

#Interactions

i.resul3<- aov(count\_3 ~ conc \* time, data = data)

i.resu3<- aov(count\_3 ~ temp \* time, data = data)

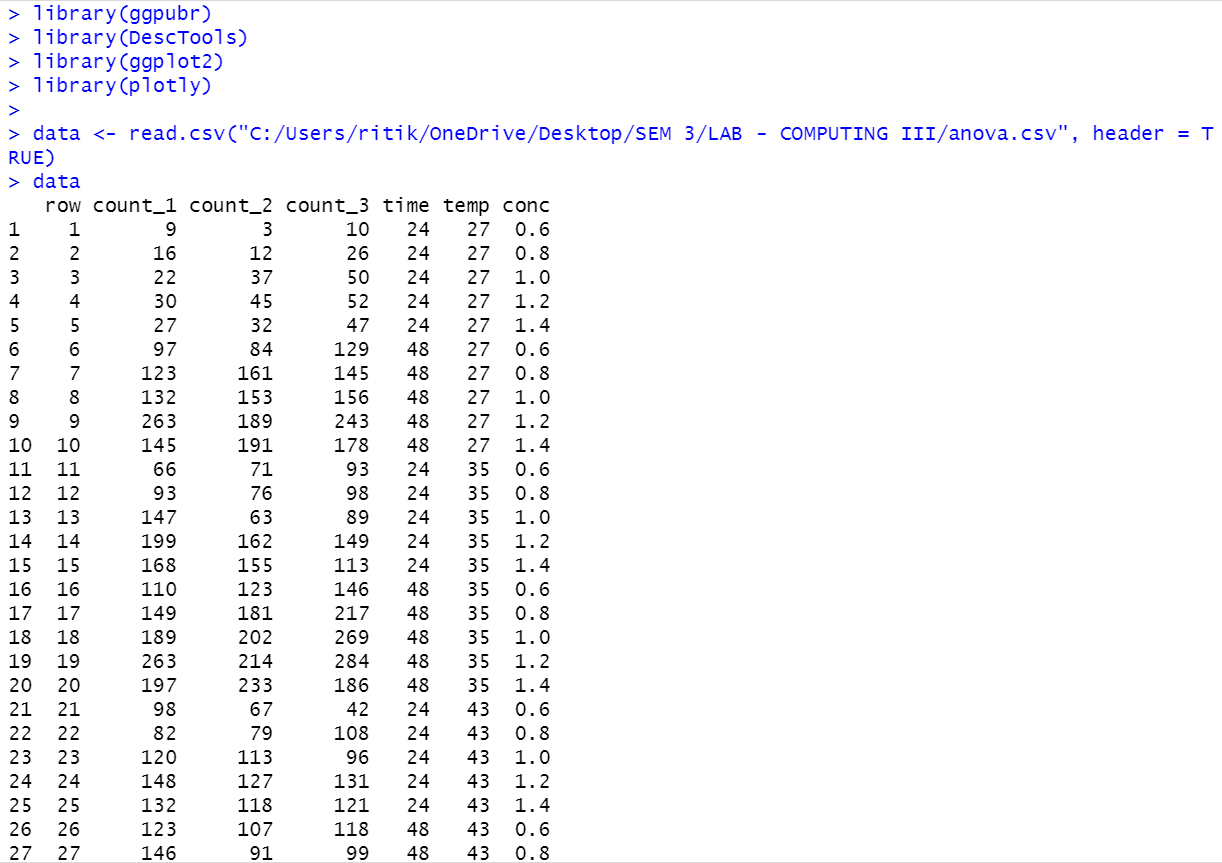
i.res3<- aov(count\_3 ~ temp \* conc, data = data)

summary(i.resul3)

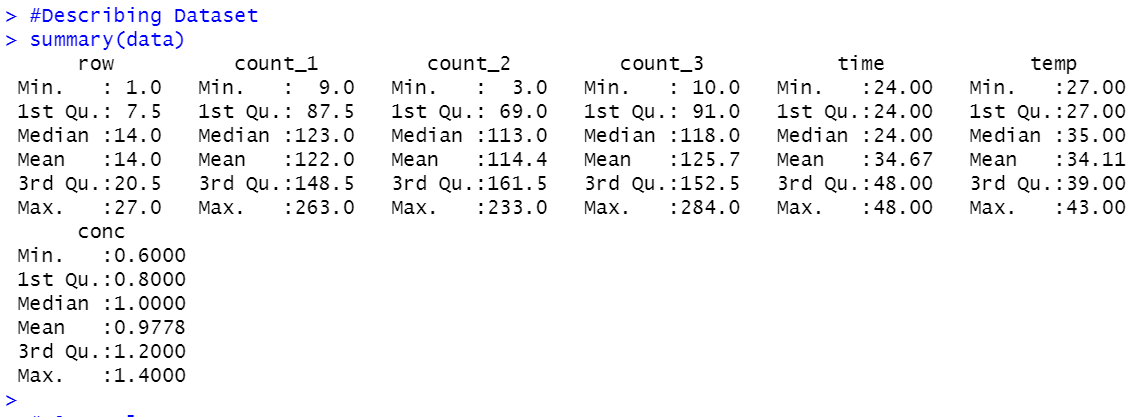
summary(i.resu3)

summary(i.res3)

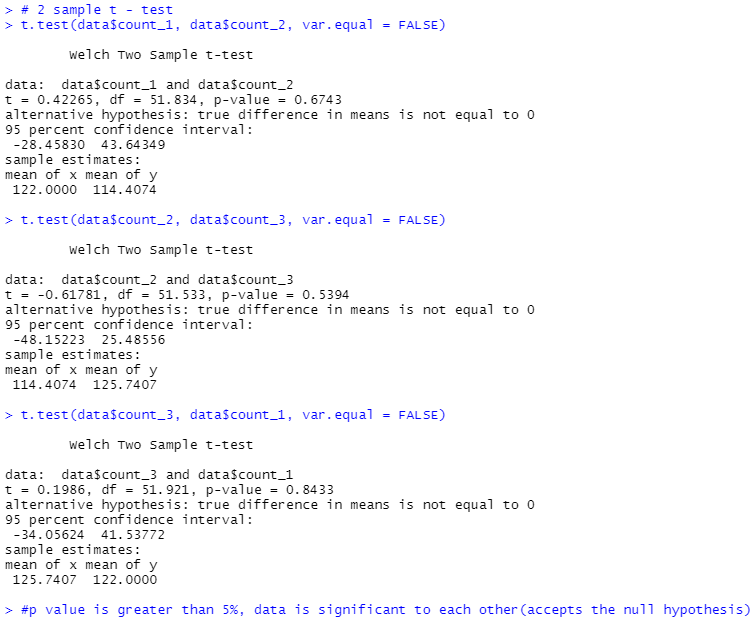
**OUTPUT**

****

* Summarising dataset

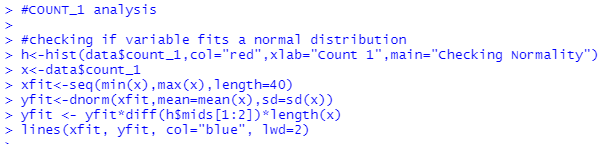
****

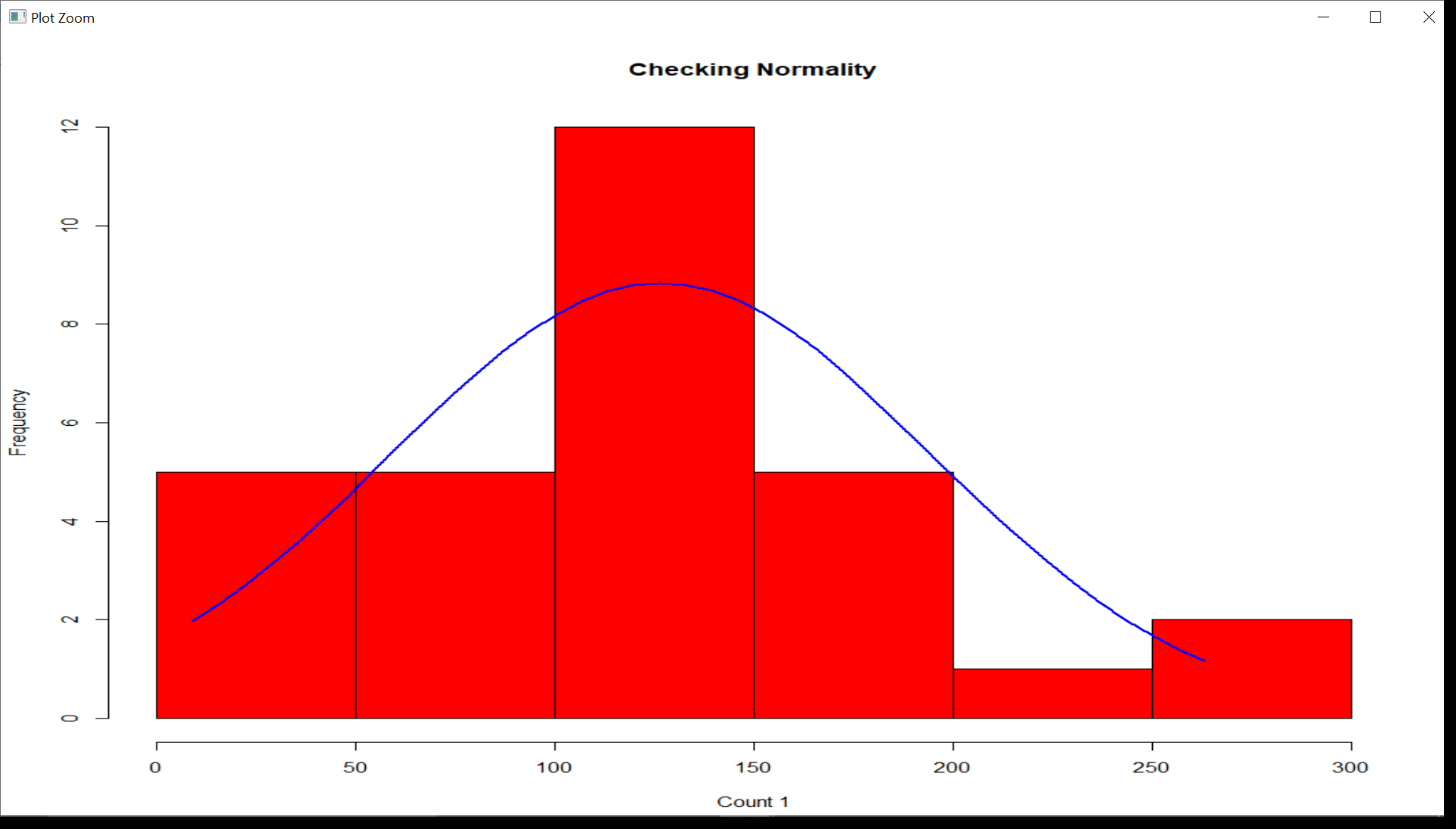
* Performing 2 sample t-test to show that each of the 3 strains are different from each other.

****

* COUNT 1 ANALYSIS

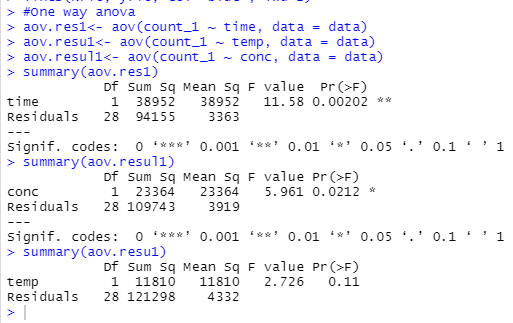
Normality of the count is checked.

****



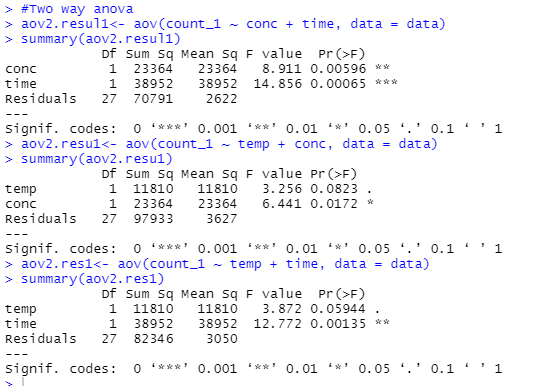
* One Way ANOVA is performed between Count 1 and the predictor variables.

Shows that time and concentration are significant factors.

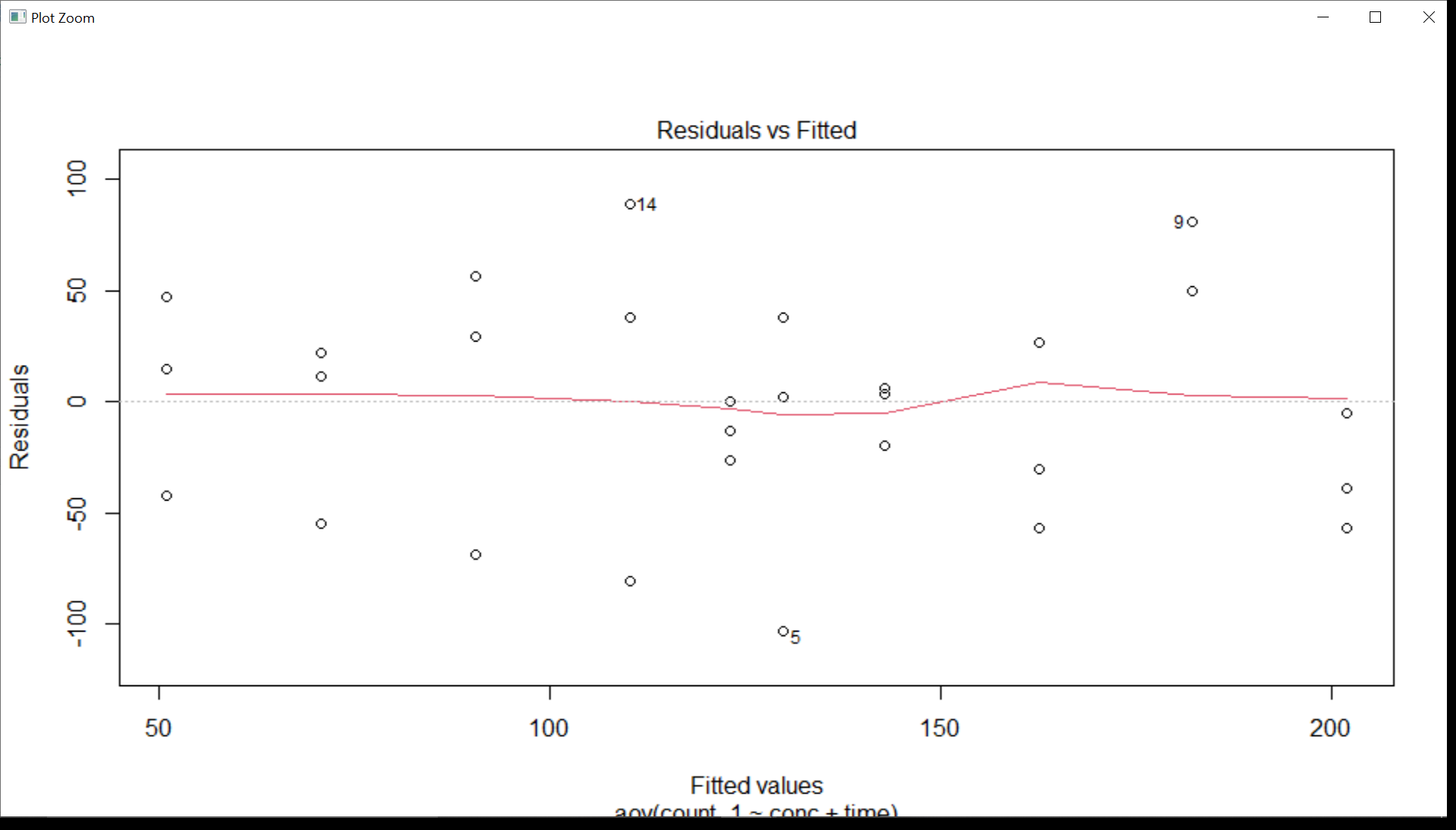
****

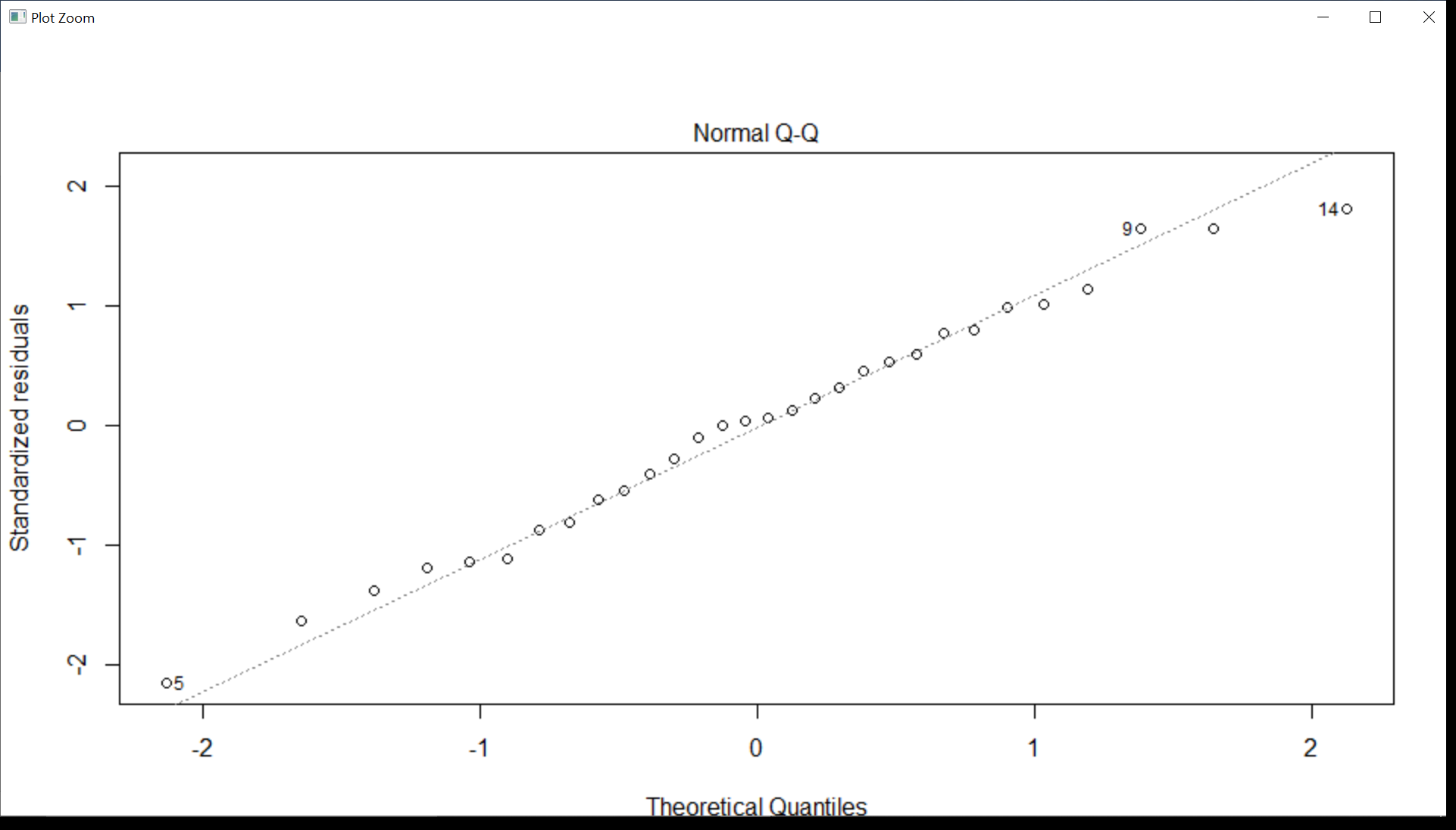
* Two Way ANOVA is performed between count 1 and two predictor variables.

Again we get that time and concentration are significant factors.

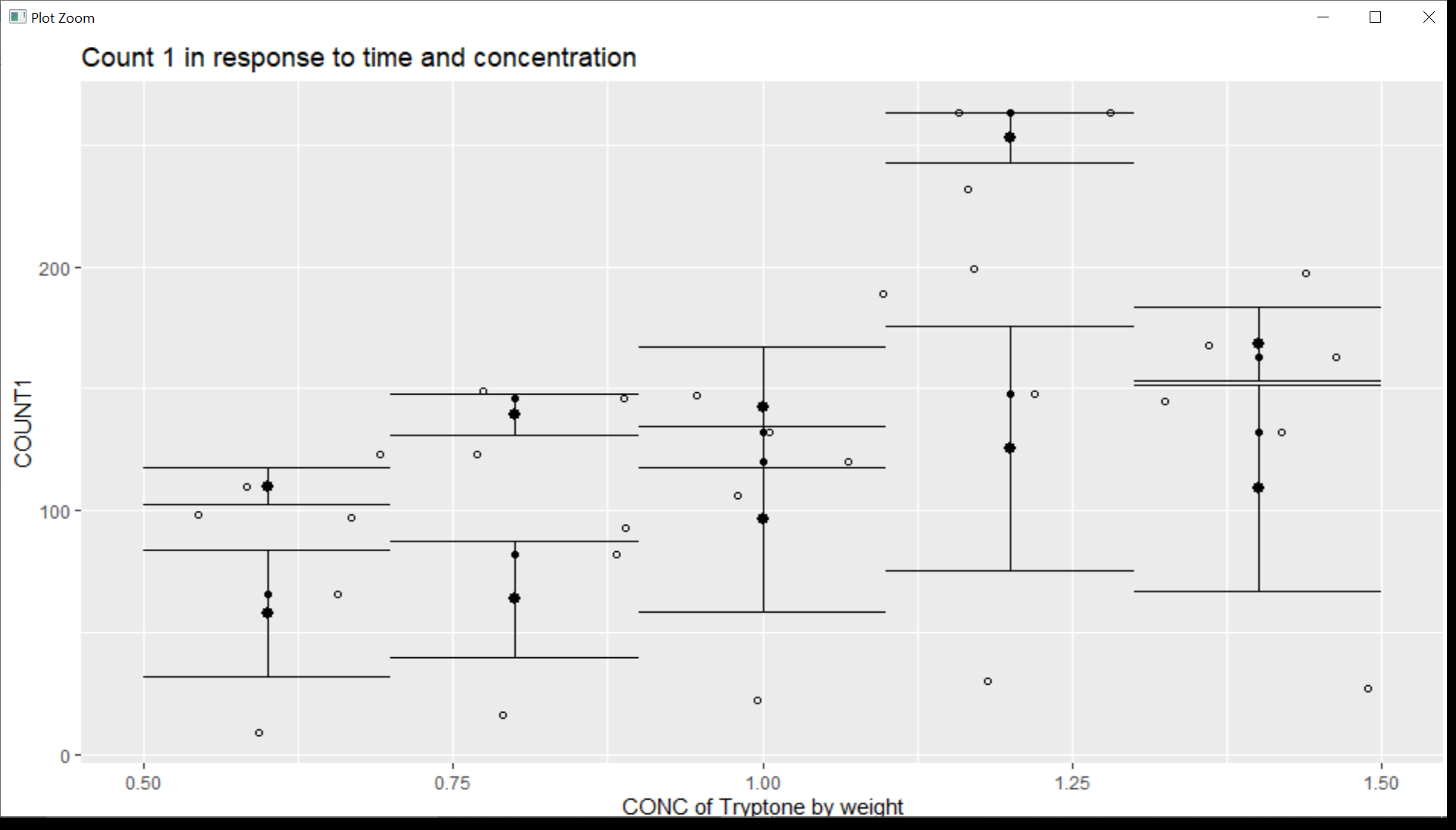
****

****





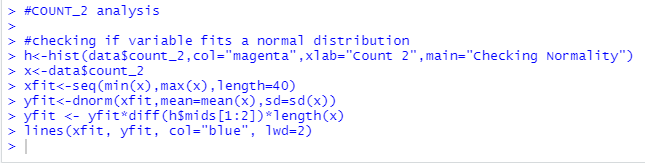
* A two way BOX PLOT is visualised from the results.

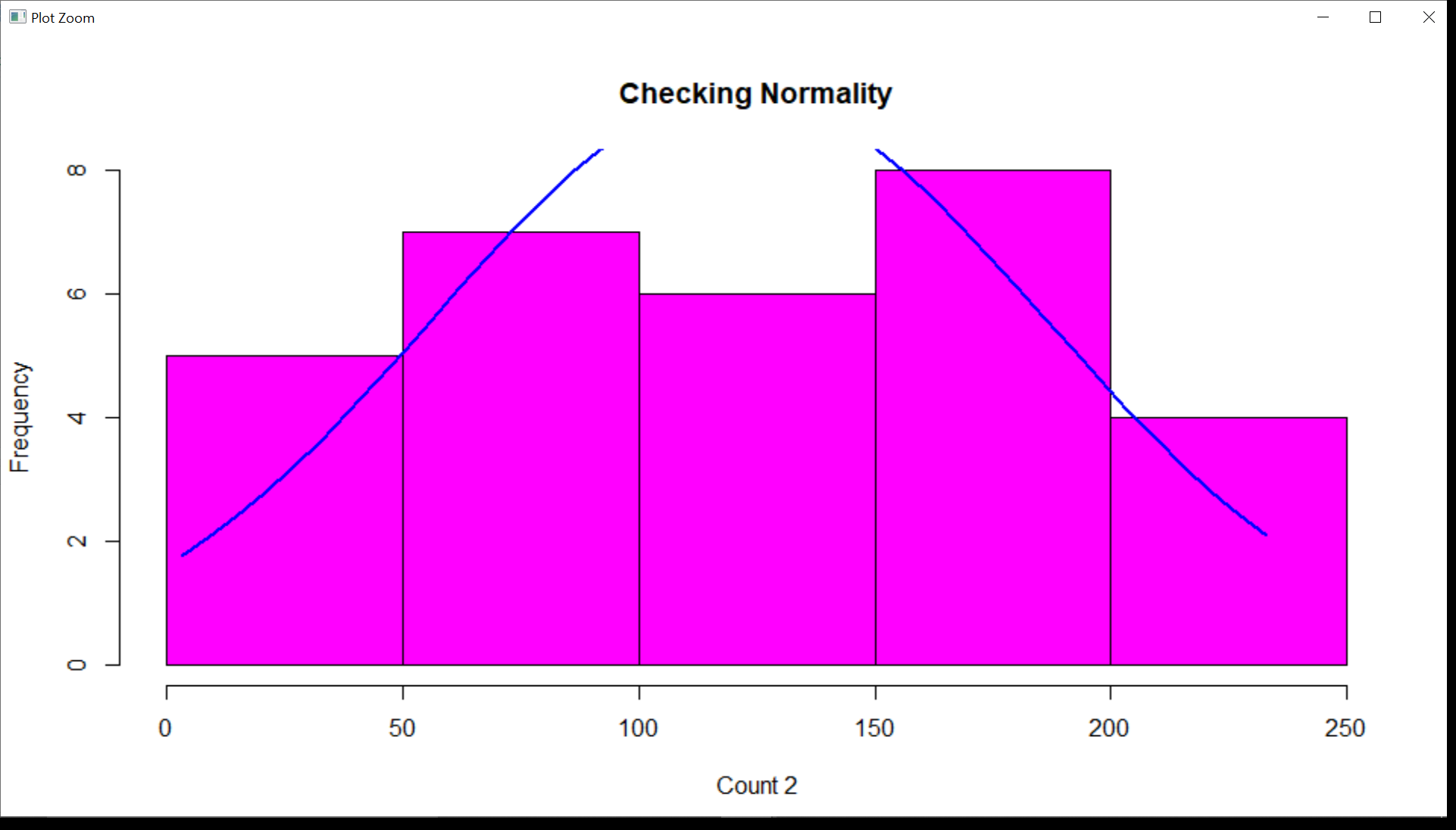


-COUNT 2 Analysis

Similar results are obtained for the other two strain counts as well.

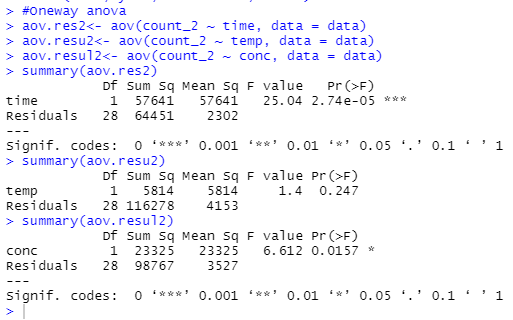
Normality of the count is checked.

****



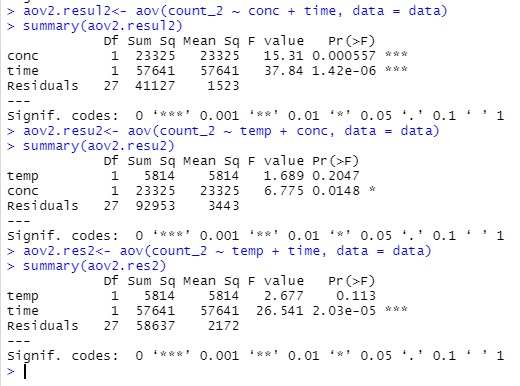
* One Way ANOVA is performed between Count 2 and the predictor variables.

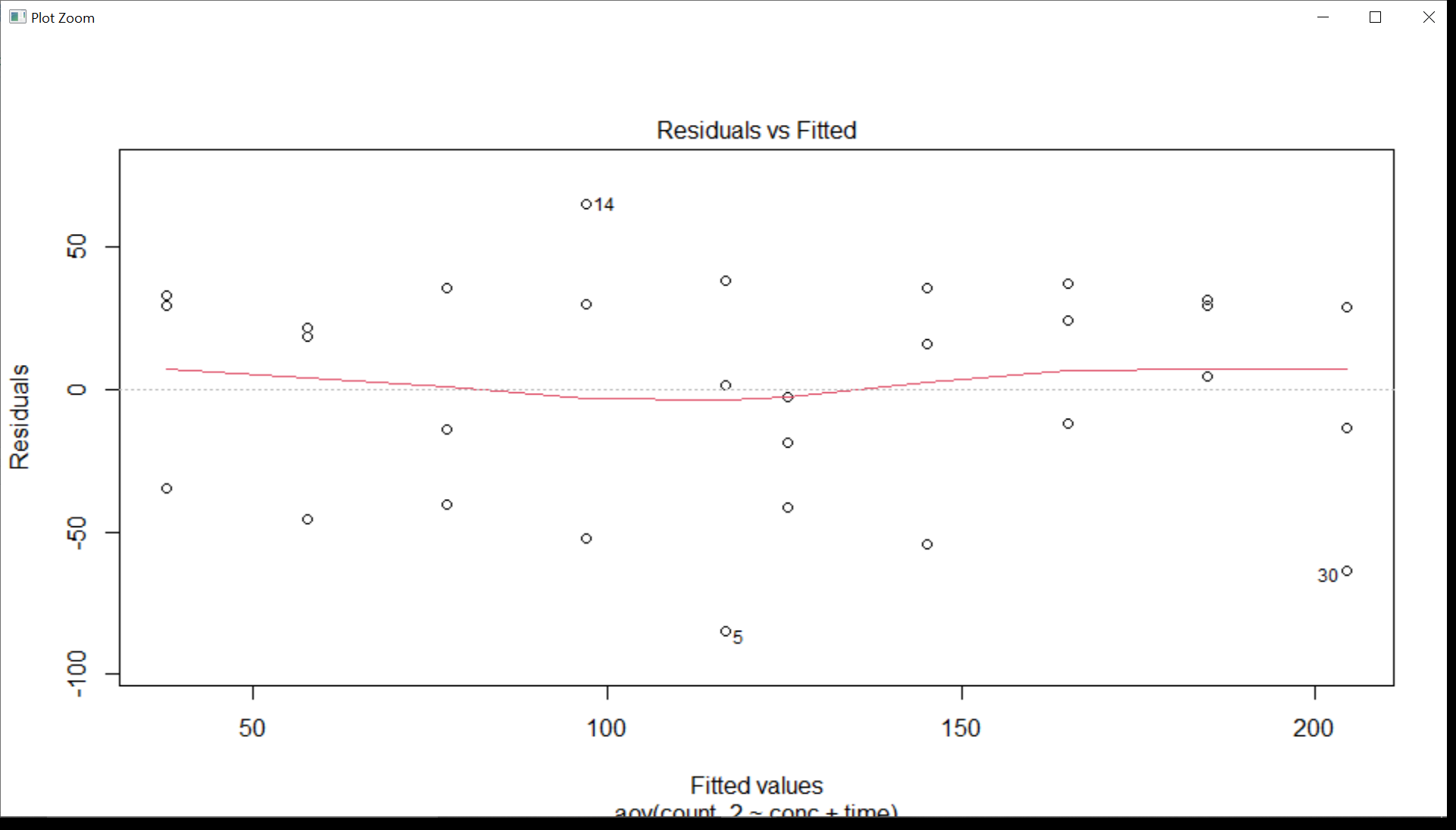
Shows that time and concentration are significant factors.

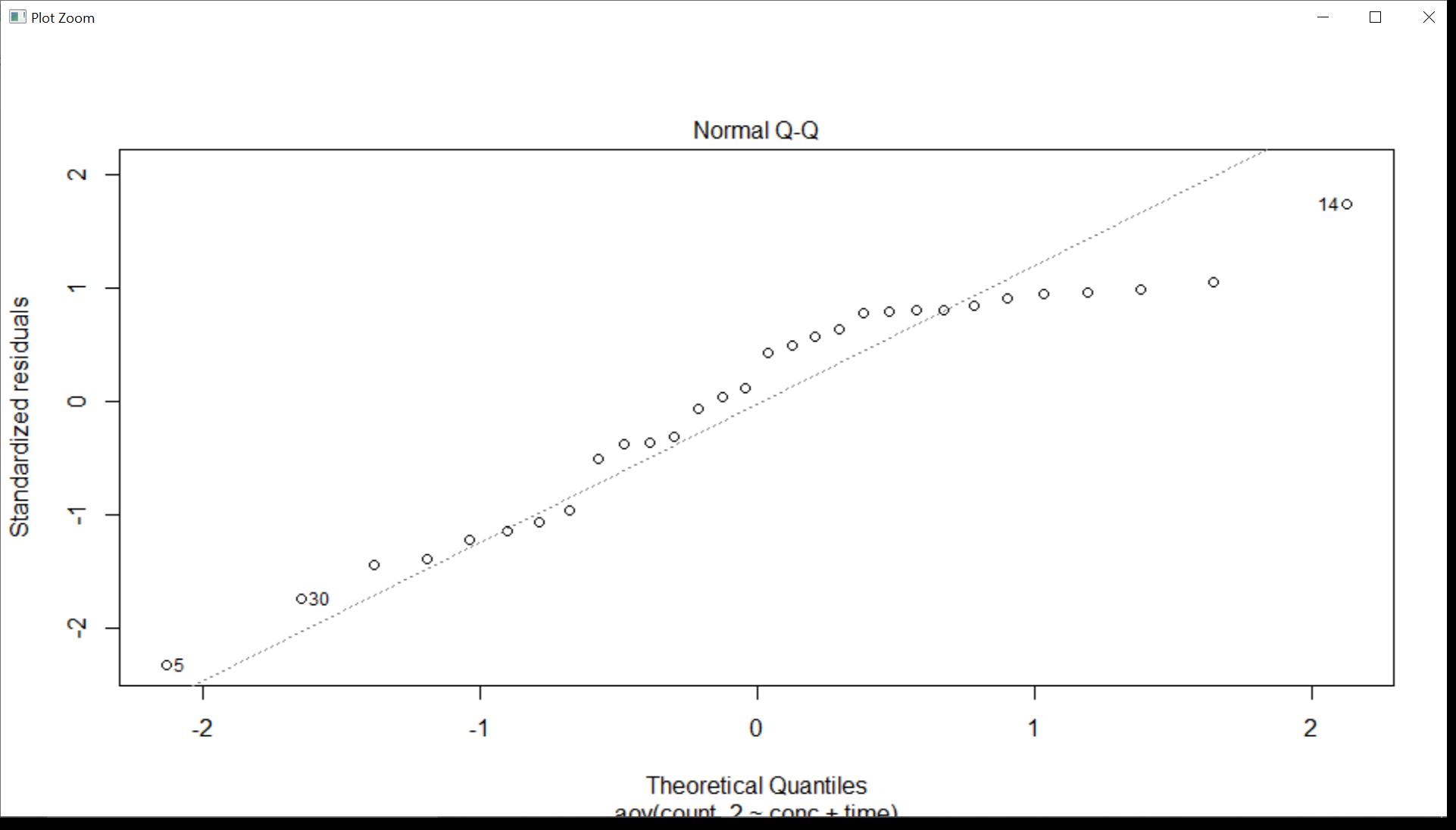
****

* Two Way ANOVA is performed between count 2 and two predictor variables.

Again we get that time and concentration are significant factors.

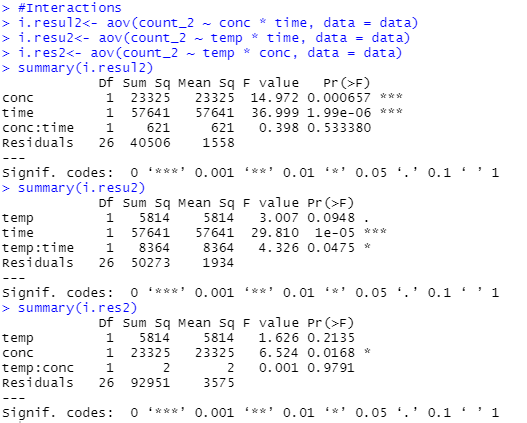
****





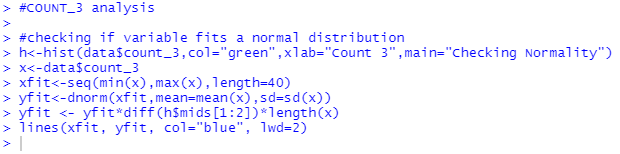
* Interactions are checked to see the significance of the predictor variables with count 2.

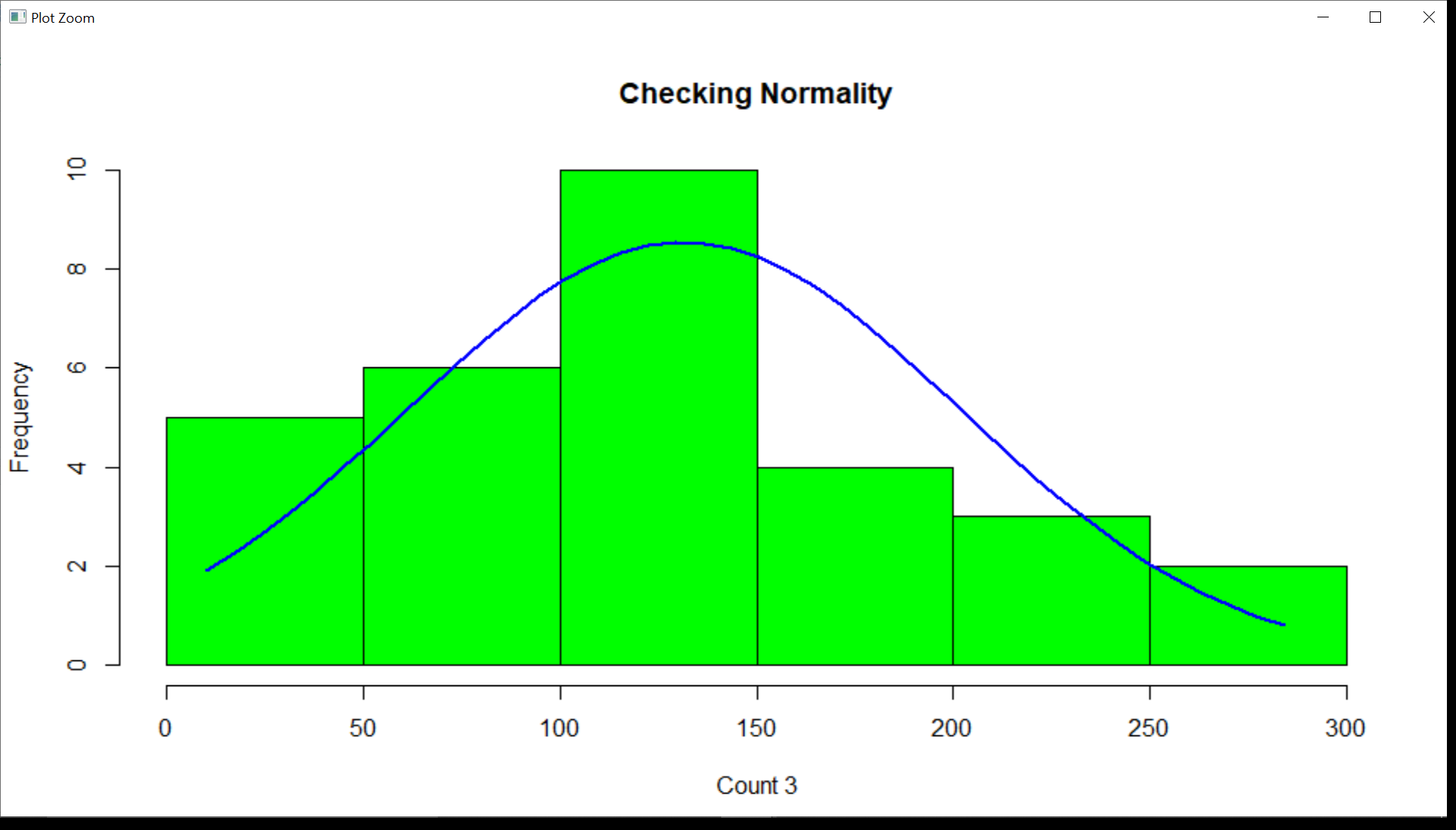
Clearly time and temperature together are significant factors.

****

* COUNT 3 Analysis

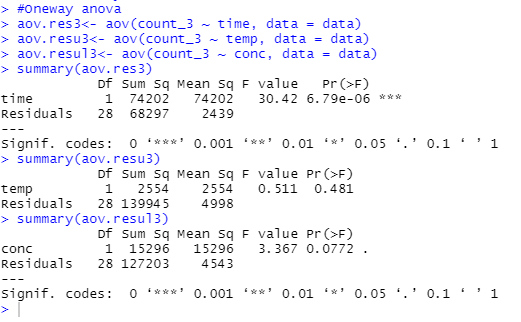
Similar results are obtained for the other two strain counts as well. Normality of the count is checked.

****



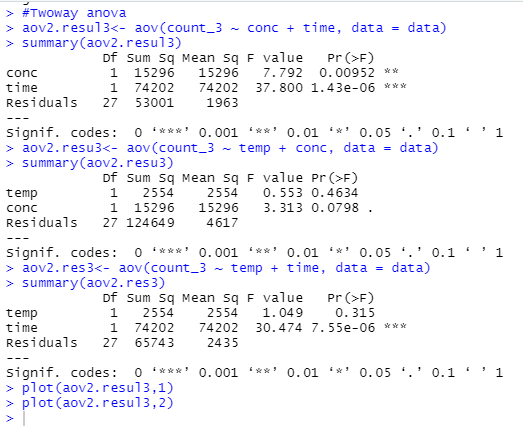
* One Way ANOVA is performed between Count 3 and the predictor variables.

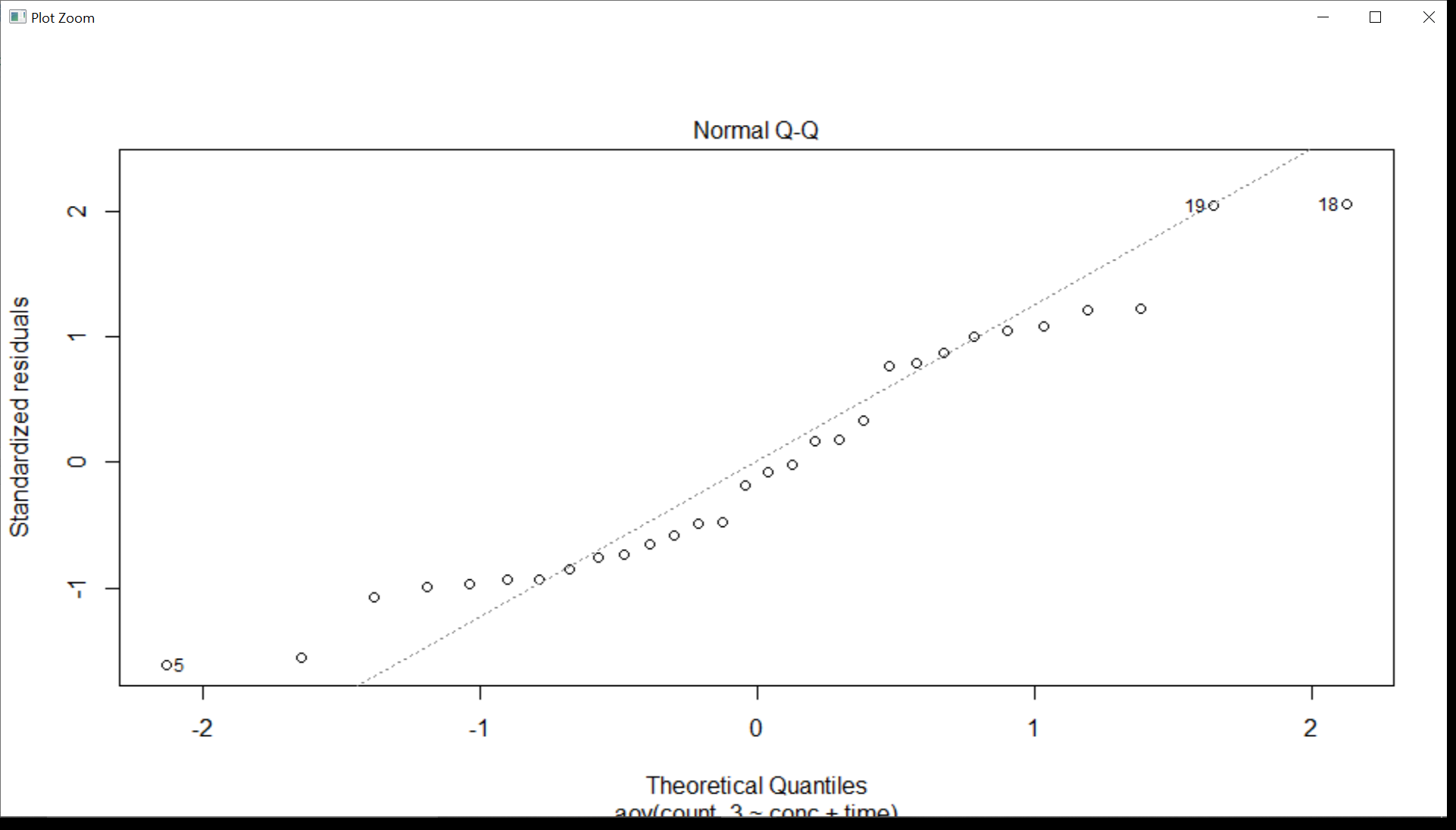
Shows that time and concentration are significant factors.

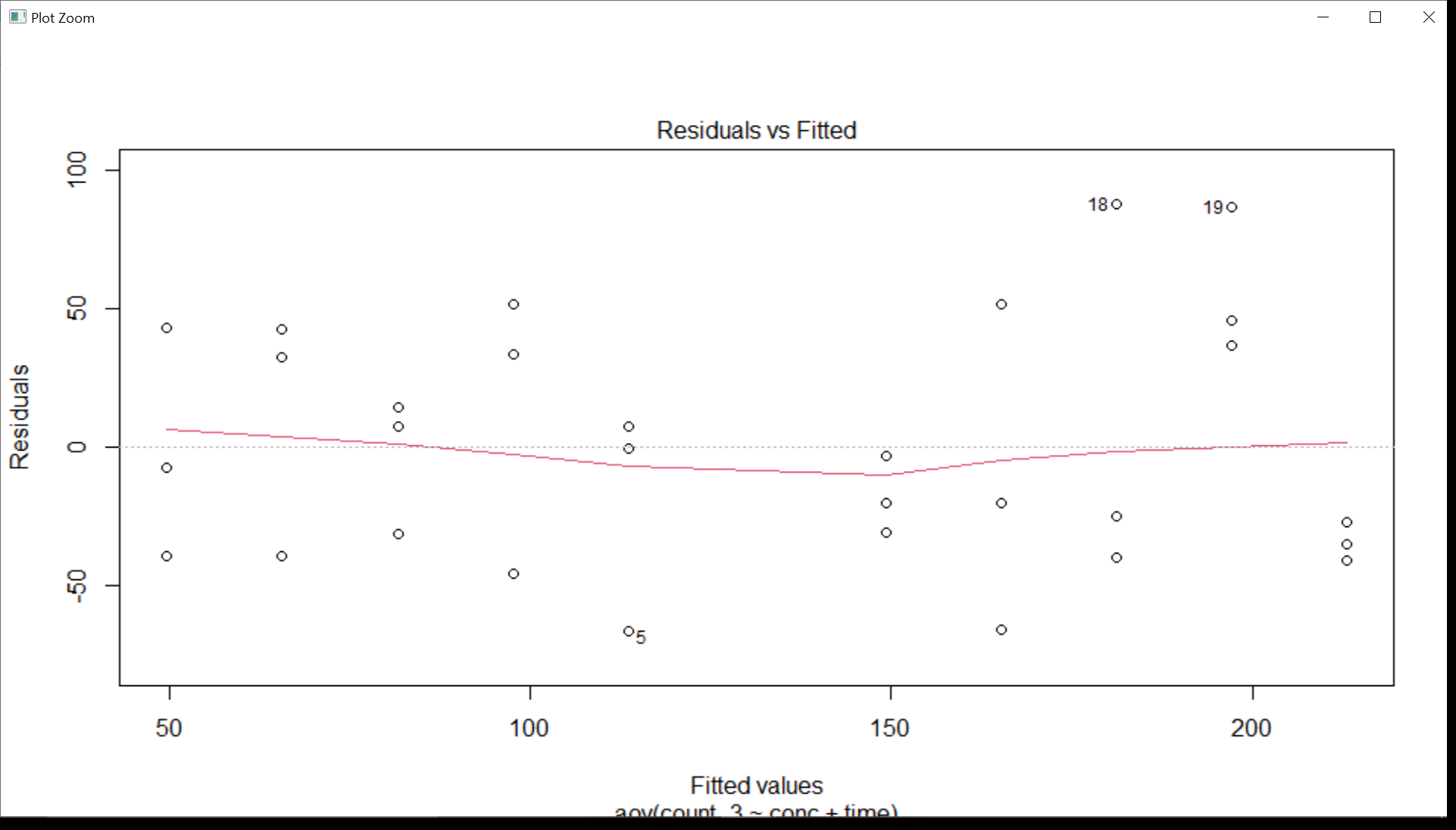
****

* Two Way ANOVA is performed between count 3 and two predictor variables.

Again we get that time and concentration are significant factors.

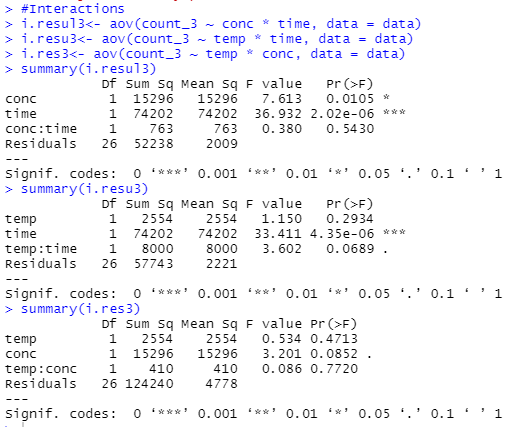
****





* Interactions are checked to see the significance of the predictor variables with count 3.

Clearly time and temperature together are significant factors.

****

**CONCLUSION**

For each of the three strain counts, out of incubation time of bacteria, incubation temperature and concentration of T-broth, significant factors have been found using One way ANOVA, two way ANOVA and Interactions and boxplots.