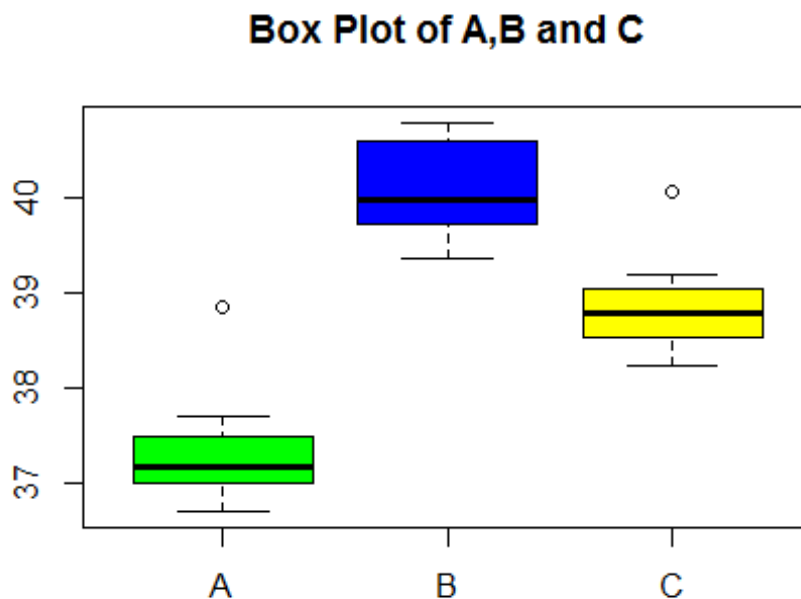# Assignment 10 (S-520)

FNU Anirudh

November 19, 2015

Solution 1

```
A<-
c(37.54,37.01,36.71,37.03,37.32,37.01,37.03,37.70,37.36,36.75,37.45,38.85)
B<- c(40.17,40.80,39.76,39.70,40.79,40.44,39.79,39.38)
C<- c(39.04,39.21,39.05,38.24,38.53,38.71,38.89,38.66,38.51,40.08)
ALL<-c(A,B,C)
boxplot(A,B,C,main="Box Plot of A,B and C",names=c('A','B','C'),col =
        c("Green","Blue","Yellow"))
```
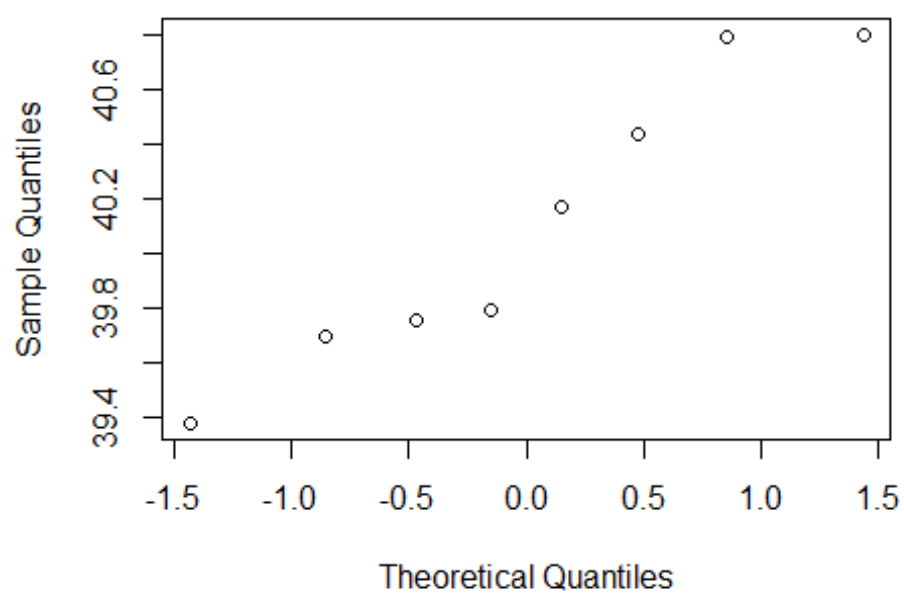

Box Plot of A,B and C

```
# Check Normality
qqnorm(A,main="QQ Plot for Sample A")
```
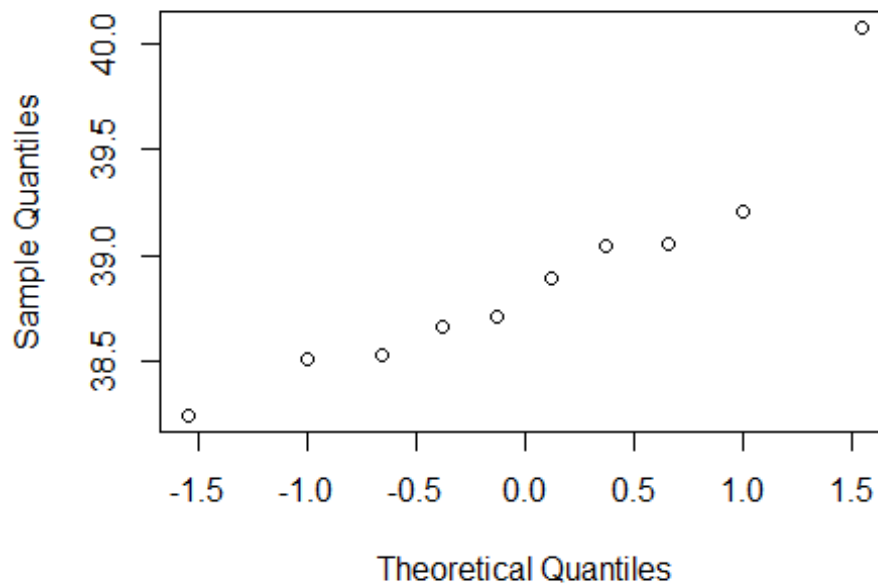
QQ Plot for Sample A
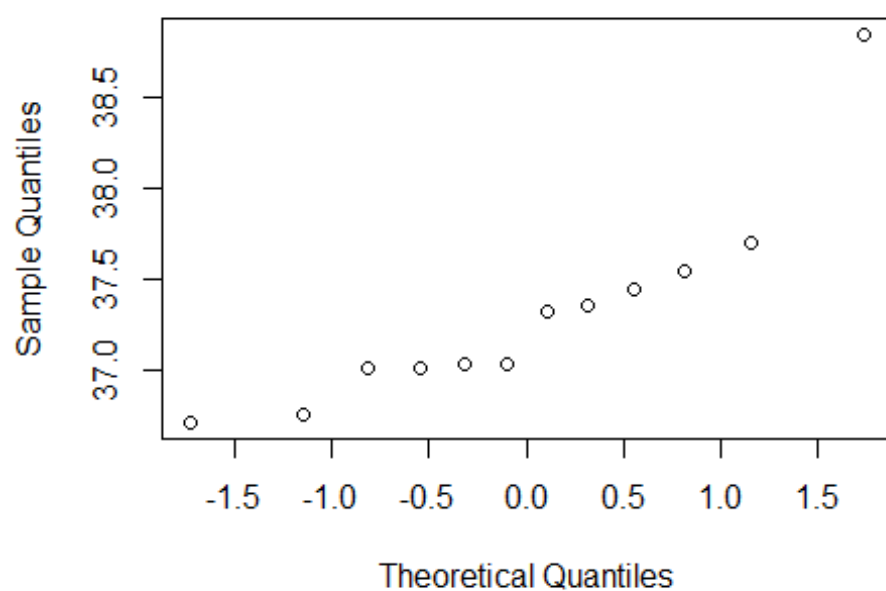

```
qqnorm(B,main="QQ Plot for Sample B")
```

QQ Plot for Sample B


```
qqnorm(C,main="QQ Plot for Sample C")
```

QQ Plot for Sample C


```
n1=length(A)
n2=length(B)
n3=length(C)
N=n1+n2+n3
# Check homoscedasticity
sda=sd(A)
sdb=sd(B)
sdc=sd(C)
sda
```

```
## [1] 0.5727975
```

```
sdb
```

```
## [1] 0.5313846
```

```
sdc
```

```
## [1] 0.510812
```

```
# Assumptions of Normality:-
# The distribution resembles closer to normality for sample B though
# location of median may question it's symmetry and normal distribution.
# We cannot conclude that data is normally distributed. Sample A and C are
pretty much symmetric and have outliers that questions normal distribution.
# Assumption of Homoscedasticity:-
# Standard Deviation of all the samples are more or less equal
# Hence Homoscedasticity assumption is quite plausible.
```

```r
meana=mean(A)
meanb=mean(B)
meanc=mean(C)
grand.mean=mean(ALL)

# Total Sum of Squares
SST = sum( (ALL-grand.mean)^2 )
total.df = N - 1

# Between sum-of-squares and mean-square
SSB = n1*(meana-grand.mean)^2 +
  n2*(meanb-grand.mean)^2 +
  n3*(meanc-grand.mean)^2
between.df = 2
between.meansquare = SSB/2

# Within sum-of-squares and mean-square
SSW = sum( (A-meana)^2 ) +
  sum( (B-meanb)^2 ) +
  sum( (C-meanc)^2 )
# Alternative formula
SSW = (n1-1)*var(A) +
  (n2-1)*var(B) +
  (n3-1)*var(C)
within.df = N - 3
within.meansquare = SSW/within.df

# Check these are equal
SST

## [1] 46.7349

SSB + SSW

## [1] 46.7349

# The two are equal.

# Are these close?
between.meansquare

## [1] 19.40044

within.meansquare

## [1] 0.2938524

# F-test
F = between.meansquare/within.meansquare
# P-value
P=1 - pf(F, df1=between.df, df2=within.df)
```

```
sites<-
c(between.df,round(SSB,4),round(between.meansquare,4),round(F,4),round(P,4))
residuals<-c(within.df,round(SSW,4),round(within.meansquare,4),"-","-")
ANOVA<-as.data.frame(rbind(sites,residuals))
rownames(ANOVA)<-c("SITES","RESIDUALS")
colnames(ANOVA)<-c("Df","Sum.sq","Mean.sq","F:Value","Pr(>F)")
ANOVA

##            Df  Sum.sq Mean.sq F:Value Pr(>F)
## SITES       2 38.8009 19.4004  66.021      0
## RESIDUALS  27   7.934  0.2939       -      -
```

$p = 4.008649e-11 < ALPHA = 0.05$ ---> reject H0

Since p value is so small hence we reject the Null Hypothesis.

Solution 2

```
x1<- c(124,42,25,45,412,51,1112,46,103,876,146,340,396)
x2<- c(81,461,20,450,246,166,63,64,155,859,151,166,37,223,138,72,245)
x3<- c(248,377,189,1843,180,537,519,455,406,365,942,776,372,163,101,20,
       283)
x4<- c(1234,89,201,356,2970,456)
x5<- c(1235,24,1581,1166,40,727,3808,791,1804,3460,719)
X<- c(x1,x2,x3,x4,x5)
y1<- log(x1)
y2<- log(x2)
y3<- log(x3)
y4<- log(x4)
y5<- log(x5)
Y<- c(y1,y2,y3,y4,y5)
n1=length(x1)
n2=length(x2)
n3=length(x3)
n4=length(x4)
n5=length(x5)
N=n1+n2+n3+n4+n5
par(mfrow=c(1,2))
boxplot(x1,x2,x3,x4,x5,main="Box Plot of X",names =
c("X1","X2","X3","X4","X5"),col=c("Red","Blue","Green","Yellow","Orange"))
boxplot(y1,y2,y3,y4,y5,main="Box Plot of Log(X)",
        names = c("Y1","Y2","Y3","Y4","Y5"),col=c("Red","Blue","Green",
                                                  "Yellow","Orange"))
```

Box Plot of X   **Box Plot of Log(X)**
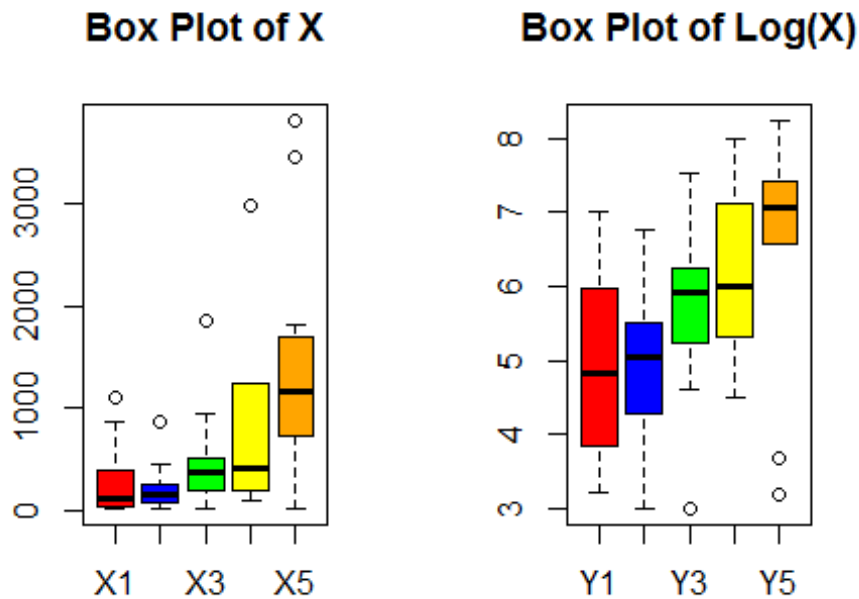
```
# Assumptions of Normality:-
# As seen from Box Plot, All X- Values have outliers and doesn't look
# normally distributed looking at median values. Box plot of Log values
# look little closer to normal distribution specially Y1. (Y2 AND Y4 being
close) hence Y1,Y2 and Y4 can be considered normal. Log values
# can be assumped to satisfy ANOVA assumptions.

sdx1=sd(x1)
sdx2=sd(x2)
sdx3=sd(x3)
sdx4=sd(x4)
sdx5=sd(x5)
sdy1=sd(y1)
sdy2=sd(y2)
sdy3=sd(y3)
sdy4=sd(y4)
sdy5=sd(y5)
# Assumptions Homoscediasticity
# Looking at the values of Standard Deviation there is a some difference
# in values for X whereas there is not much difference in values for Y
# Hence Homoscediasticity is plausible for Y.

#Let Ho be that mean survival time be same for different organ affected #Let
Ha be that mean survival time be different for at least one organ affected

y1m = mean(y1)
```

```
y2m = mean(y2)
y3m = mean(y3)
y4m = mean(y4)
y5m = mean(y5)
grand.mean = mean(Y)
k=5

SST = sum( (Y-grand.mean)^2 )
total.df = N - 1

SSB = n1*(y1m-grand.mean)^2 +
    n2*(y2m-grand.mean)^2 +
    n3*(y3m-grand.mean)^2 +
    n4*(y4m-grand.mean)^2 +
    n5*(y5m-grand.mean)^2
between.df = k-1
between.meansquare = SSB/between.df
SSW = sum( (y1-y1m)^2 ) +
    sum( (y2-y2m)^2 ) +
    sum( (y3-y3m)^2 )+ sum((y4-y4m)^2)+sum((y5-y5m)^2)
within.df = N - k
within.meansquare = SSW/within.df

F = between.meansquare/within.meansquare
# P-value
P=1 - pf(F, df1=between.df, df2=within.df)

organs<-
c(between.df,round(SSB,4),round(between.meansquare,4),round(F,4),round(P,4))
residuals<-c(within.df,round(SSW,4),round(within.meansquare,4),"-","-")
ANOVA<-as.data.frame(rbind(organs,residuals))
rownames(ANOVA)<-c("Organs","Residuals")
colnames(ANOVA)<-c("Df","Sum.sq","Mean.sq","F:Value","Pr(>F)")
ANOVA

##            Df  Sum.sq Mean.sq F:Value Pr(>F)
## Organs      4 24.4866  6.1216   4.286 0.0041
## Residuals  59 84.2696  1.4283       -      -
```

P = 0.004121 < ALPHA = 0.05---> reject H0

Since P-value is so small, we reject Null Hypothesis.

Solution 3

```
stereograms = read.table(file.choose(), header=TRUE)
treatment = stereograms$time[stereograms$group==2]
control = stereograms$time[stereograms$group==1]
all.diffs<-stereograms$time
t.test(treatment, control)
```

```
##
##  Welch Two Sample t-test
##
## data:  treatment and control
## t = -2.0384, df = 70.039, p-value = 0.04529
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -5.95314090 -0.06493219
## sample estimates:
## mean of x mean of y
##  5.551429  8.560465

n1 = length(treatment)
n2 = length(control)
N=n1+n2
k=2
boxplot(treatment, control, range=0,
    names=c("Treatment","control"),
    ylab="Time",col=c("Red","Green"))
```
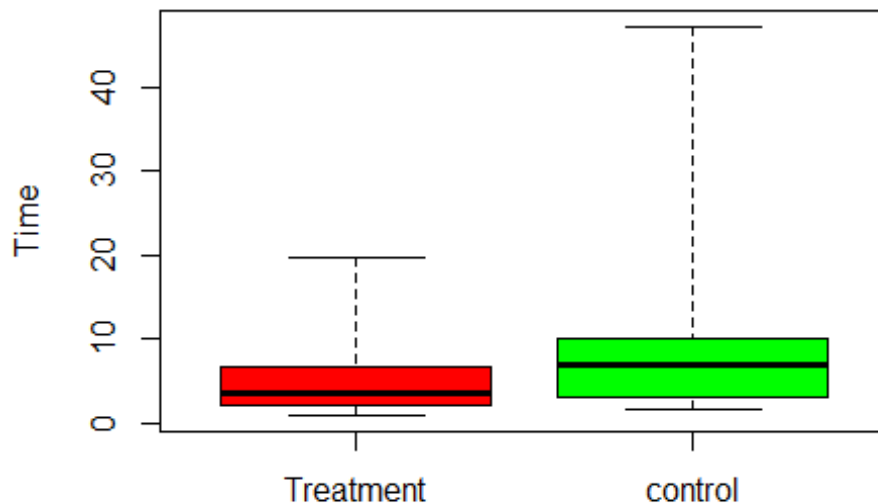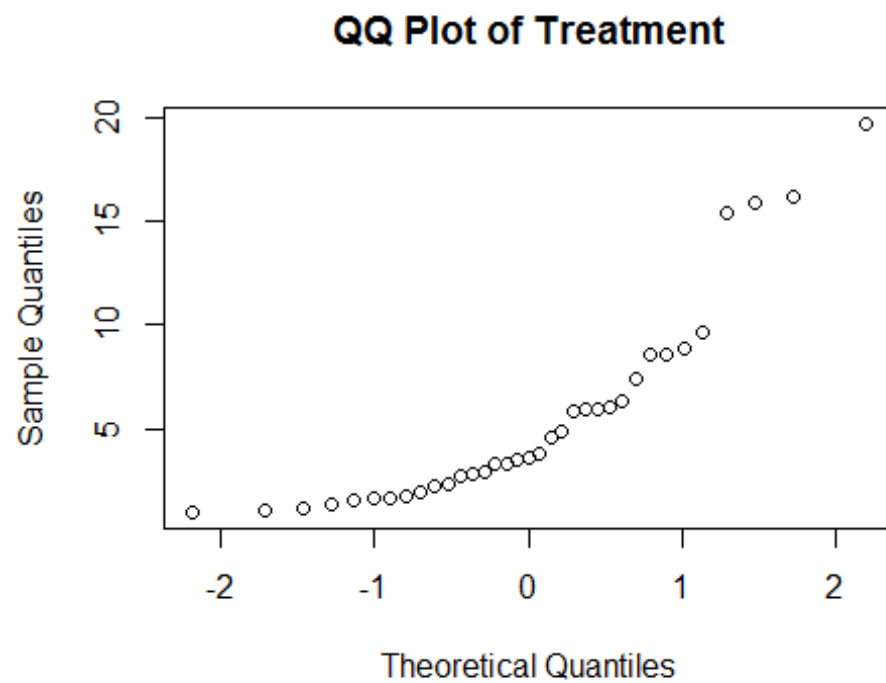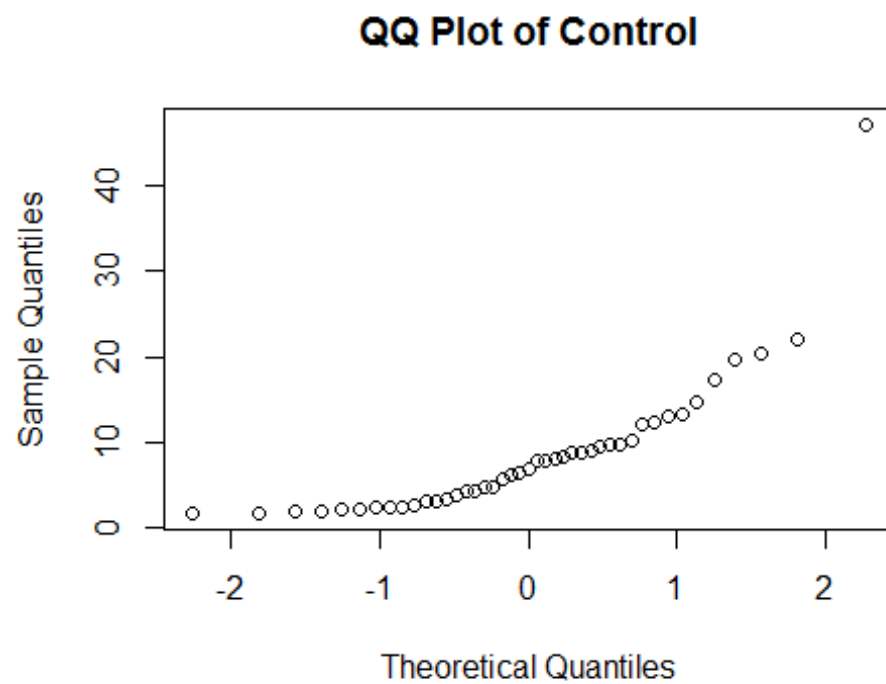


```
sd(treatment)

## [1] 4.801738

sd(control)

## [1] 8.085411
```

```
qqnorm(treatment,main="QQ Plot of Treatment")
```

QQ Plot of Treatment


```
qqnorm(control,main="QQ Plot of Control")
```

QQ Plot of Control

```r
tmean = mean(treatment)
cmean = mean(control)
grand.mean = mean(all.diffs)
SST = sum( (all.diffs-grand.mean)^2 )
total.df = N - 1
SSB = n1*(tmean-grand.mean)^2 +
    n2*(cmean-grand.mean)^2

between.df = k-1
between.meansquare = SSB/between.df
SSW = sum( (treatment-tmean)^2 ) +
    sum( (control-cmean)^2 )


within.df = N - k
within.meansquare = SSW/within.df

F = between.meansquare/within.meansquare
# P-value
p=1 - pf(F, df1=between.df, df2=within.df)

group<-
c(between.df,round(SSB,3),round(between.meansquare,3),round(F,3),round(p,5))
residuals<-c(within.df,round(SSW,3),round(within.meansquare,3),"-","-")
ANOVA<-as.data.frame(rbind(group,residuals))
rownames(ANOVA)<-c("Group","residuals")
colnames(ANOVA)<-c("Df","Sum.sq","Mean.sq","F:Value","Pr(>F)")
ANOVA

##            Df  Sum.sq Mean.sq F:Value  Pr(>F)
## Group       1 174.702 174.702   3.762 0.05615
## residuals 76 3529.63  46.443       -       -
```

P- Value comes close to Student t-test value.