1.

airline=read.table(file="airline\_cleaned.txt",header=T)

attach(airline)

airline$trans.TaxiIn=log(TaxiIn)

studynew <- subset(airline, UniqueCarrier =="UA" | UniqueCarrier =="AA"| UniqueCarrier =="NW")

# select=c(FlightNum,UniqueCarrier))

ua <- subset(airline, UniqueCarrier =="UA")

aa <- subset(airline, UniqueCarrier =="AA")

nw <- subset(airline, UniqueCarrier =="NW")

detach(airline)

attach(studynew)

library(lattice)

studynew.new <- droplevels(studynew)

# studynew.new <- droplevels(studynew)

Comp = studynew.new$trans.TaxiIn

Group = studynew.new$UniqueCarrier

# boxplot(log(studynew.new$TaxiIn) ~studynew.new$UniqueCarrier, studynew.new)

quartz()

trace<-rep(1, length(Group))

interaction.plot(Group,trace,Comp,fun=mean,legend=F)

quartz()

trace<-rep(1, length(Group))

interaction.plot(Group,trace,studynew.new$TaxiIn,fun=mean,legend=F)

print(tapply(Comp, Group, length))

print(tapply(Comp, Group, mean))

print(tapply(Comp, Group, sd))

histogram(~studynew.new$trans.TaxiIn | studynew.new$UniqueCarrier, layout=c(1,3),type="density",

 panel=function(x,...)

  {panel.histogram(x,...)

   panel.mathdensity(dmath=dnorm,col="blue",lwd=2,

      args=list(mean=mean(x, na.rm=T), sd = sd(x,na.rm=T)),...)

   panel.densityplot(x,col="red",lwd=2,...)

   })

bwplot(~studynew.new$trans.TaxiIn | studynew.new$UniqueCarrier, layout = c(1, 3), pch = "|") #Boxplots side-by-side

qqmath(~log(abs(studynew.new$TaxiIn)) | studynew.new$UniqueCarrier, data = studynew.new, panel = function(x){

panel.qqmath(x)

panel.qqmathline(x)

})

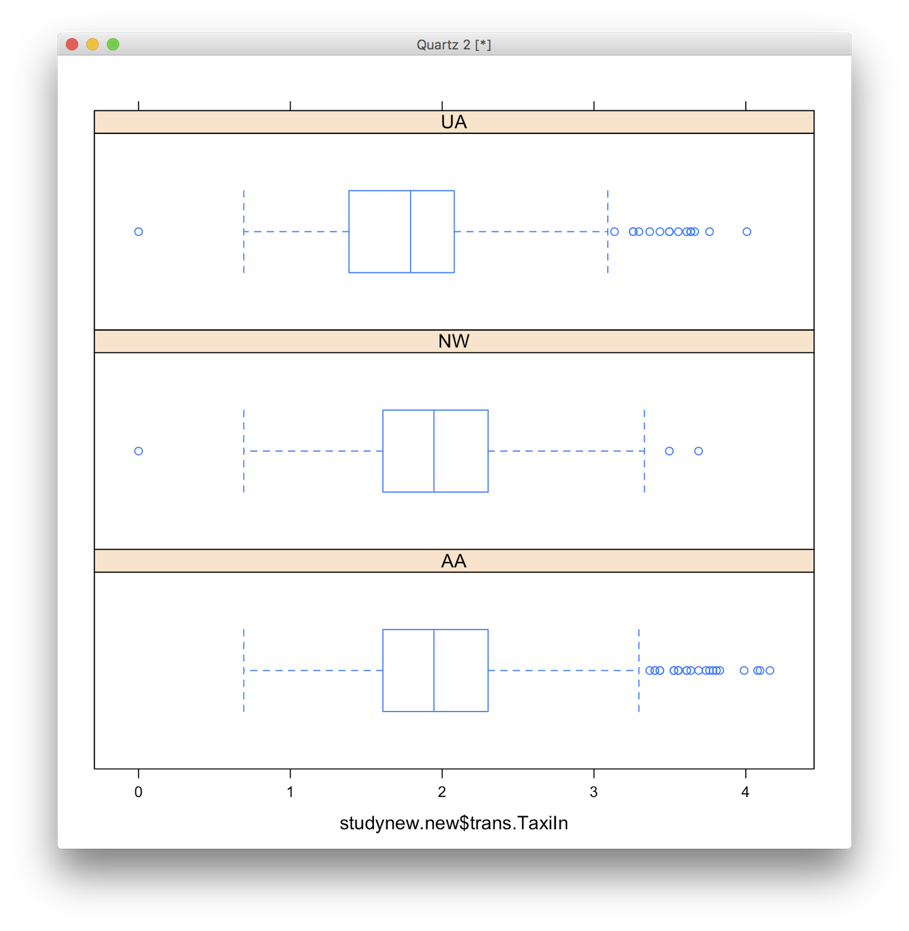
fit <- aov(Comp ~ Group, data=studynew.new)

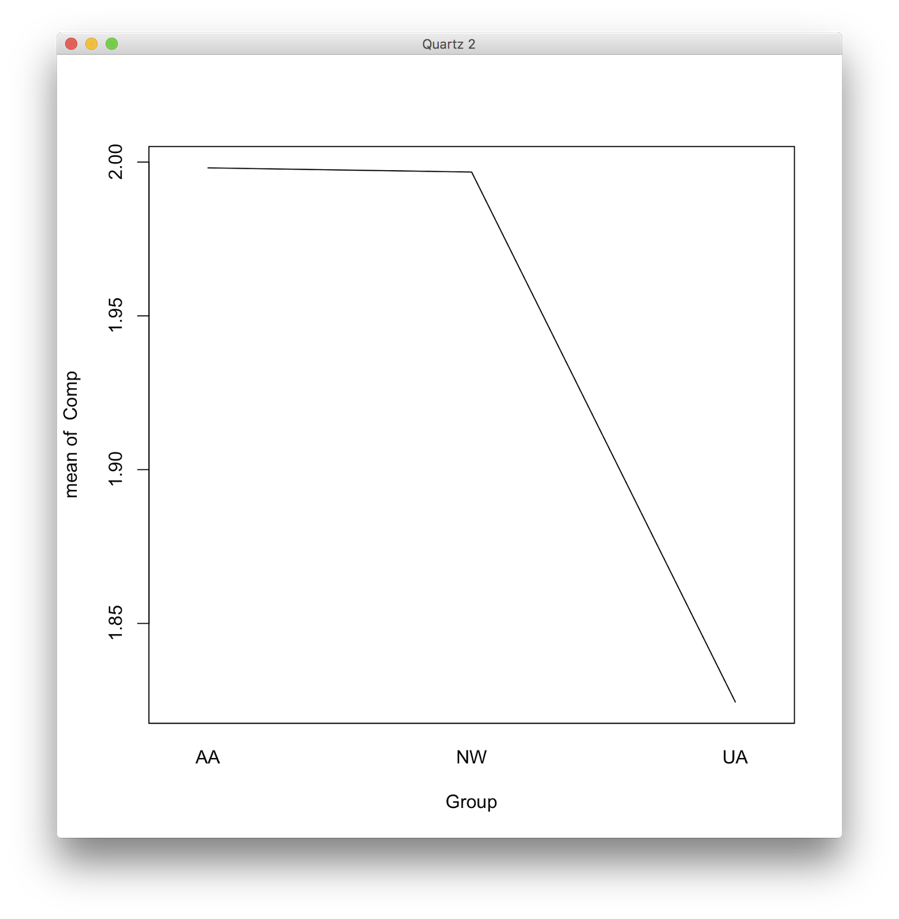
summary(fit)

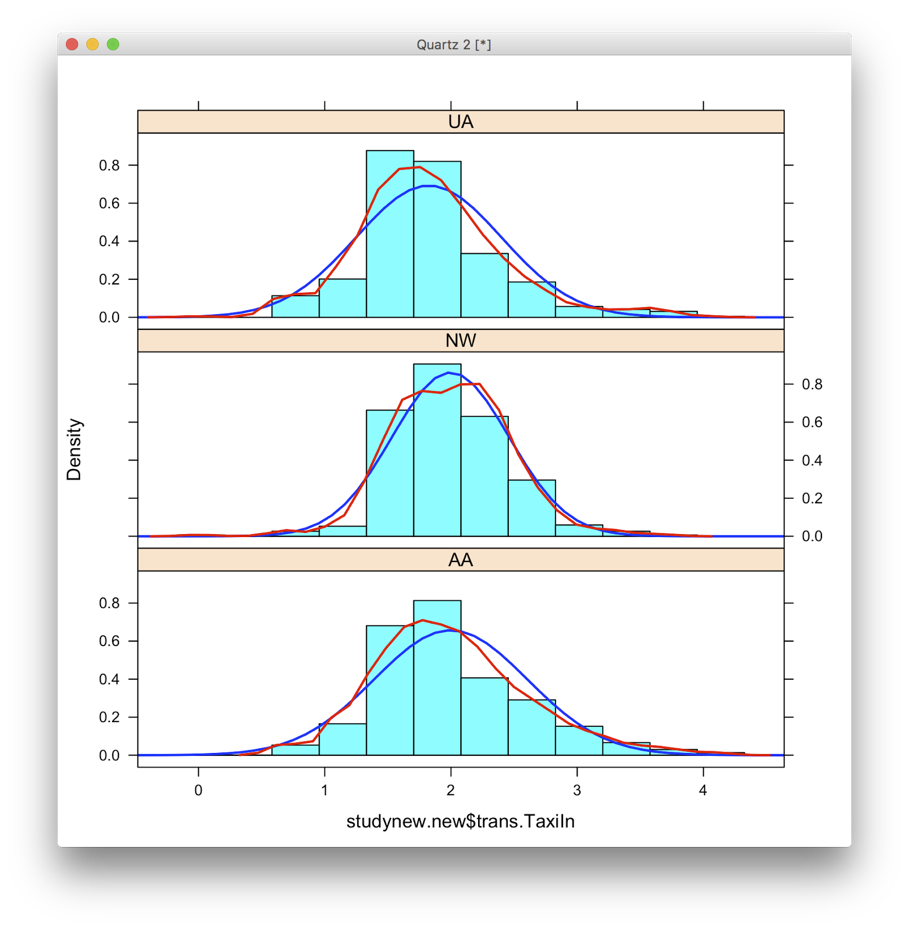
test.Tukey<-TukeyHSD(fit,conf.level=0.95)

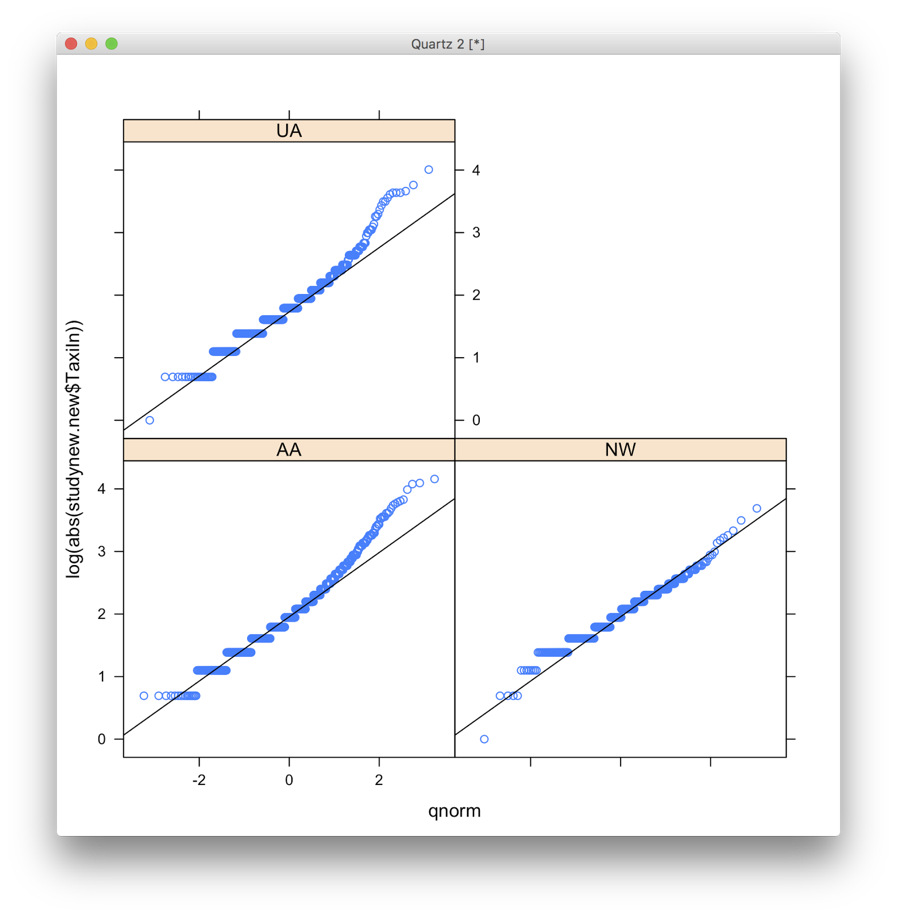
test.Tukey

2.









AA  NW  UA

808 407 518

     AA       NW       UA

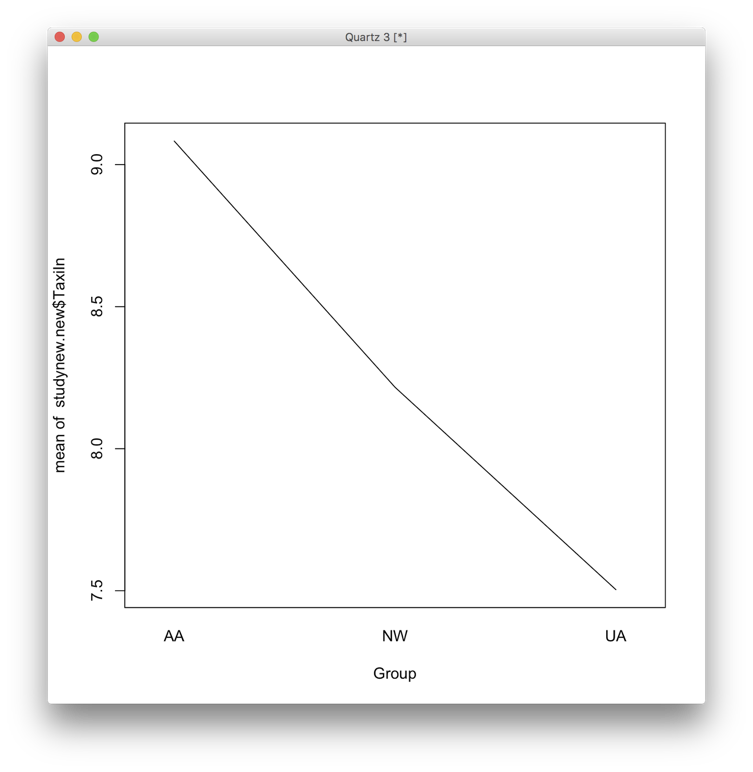
1.998115 1.996744 1.824460

      AA        NW        UA

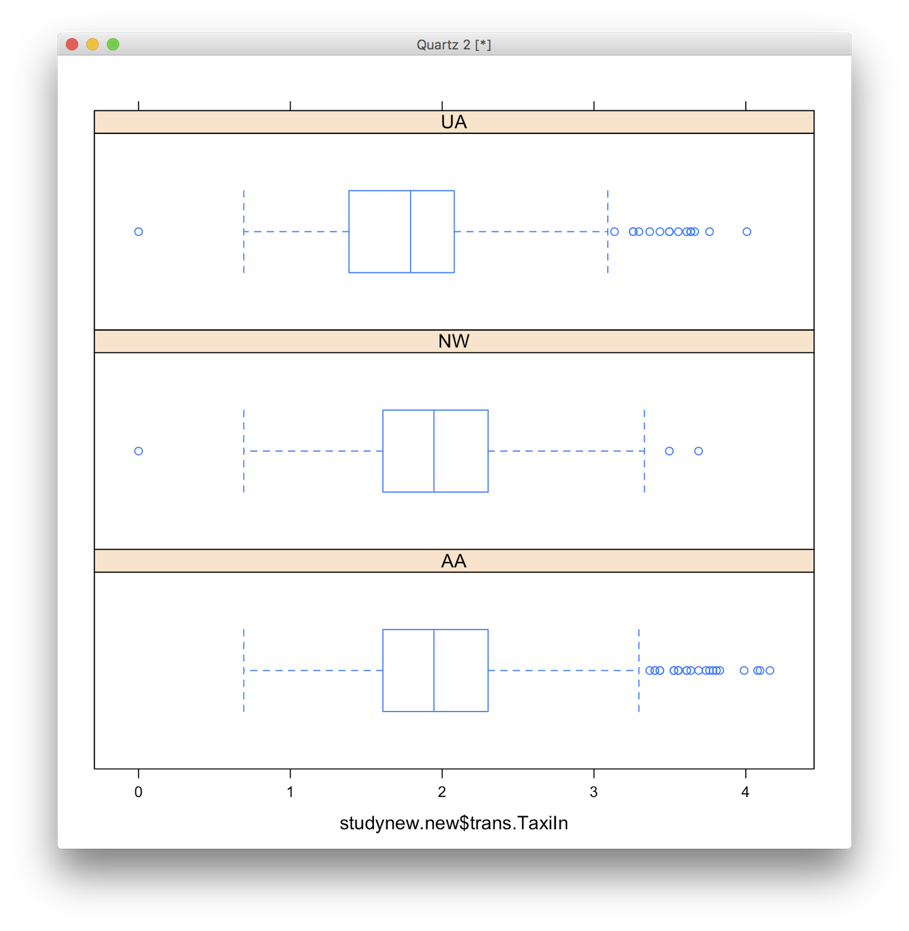
0.6071220 0.4632140 0.5759199

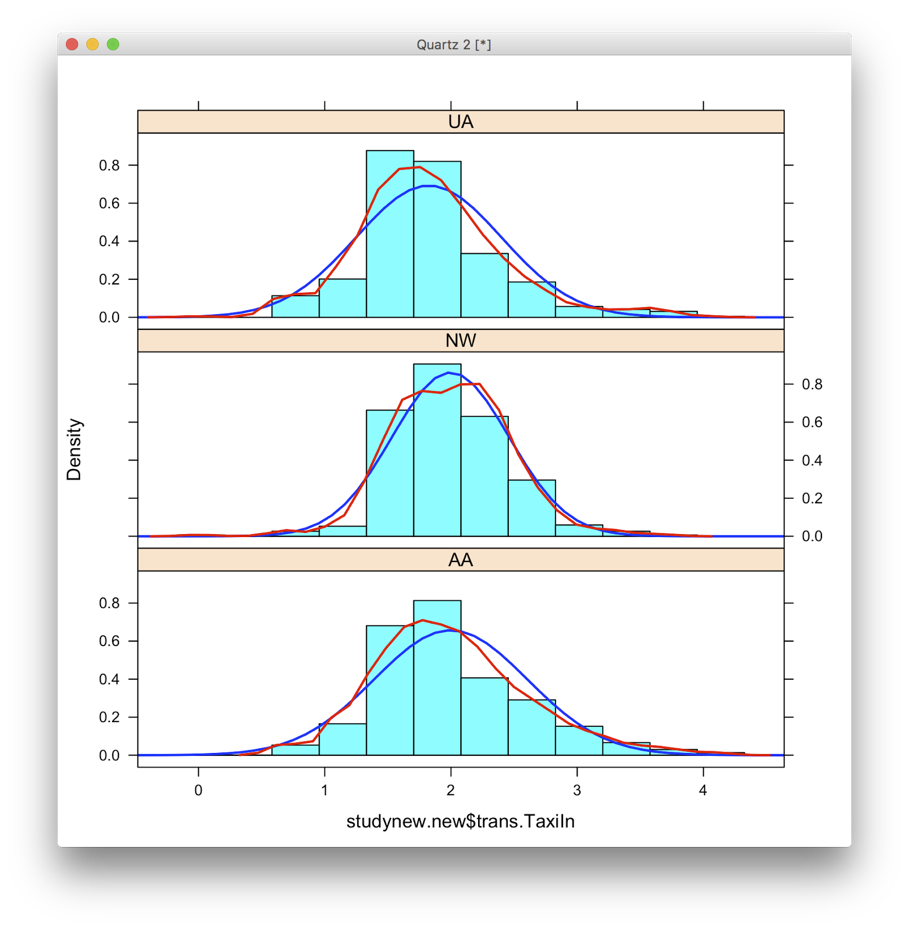
|  |  |  |  |
| --- | --- | --- | --- |
| Group | n | sample mean | sample standard deviation |
| AA | 808 | 1.998115 | 0.6071220 |
| NW | 407 | 1.996744 | 0.4632140 |
| UA | 518 | 1.824460 | 0.5759199 |

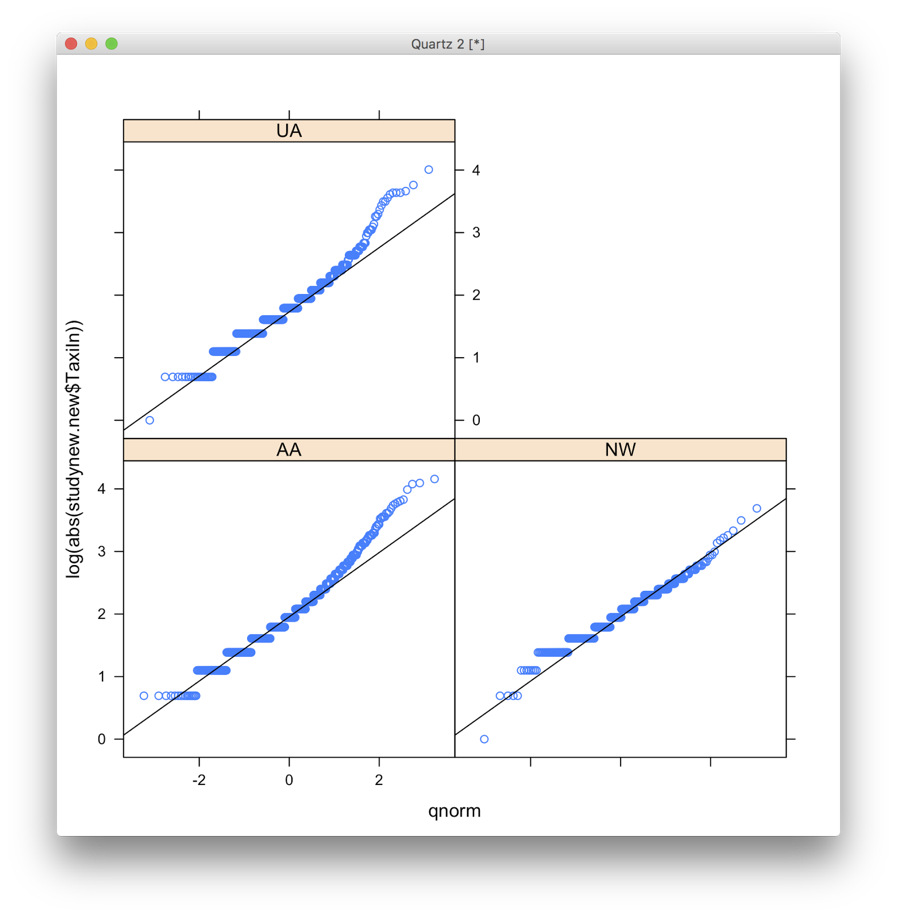
3.



4.







From the plot above, we can see that they are normal distributed. So it is appropriate  to use the ANOVA test to analysis it.

5.

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

Group          2   10.9   5.447   16.95 5.14e-08 \*\*\*

Residuals   1730  556.1   0.321

---

Signif. codes:  0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

**Step 1: Definition of the terms**

AA is the population mean TaxiIn score for the AA method.

NW is the population mean TaxiIn score for the NW method.

UA is the population mean TaxiIn score for the UA method.

**Step 2: State the hypotheses**

H0: AA = NW = UA

Ha: at least two i's are different.

**Step 2: Find the *Test Statistic, p-value, report DF***

Fts = 16.95 5

DF1 = 2, DF2 = 1730

P-value = 5.14e-08

**Step 4: Conclusion:**

= 0.05

Since 5.14e-08 < 0.05, we should reject H0

The data provides sufficiently strong evidence (P-value = 5.14e-08) to the claim that the population mean values of at least one of the time that it takes to taxi in is different from the rest.

6.

> test.Tukey<-TukeyHSD(fit,conf.level=0.95)

> test.Tukey

 Tukey multiple comparisons of means

   95% family-wise confidence level

Fit: aov(formula = Comp ~ Group, data = studynew.new)

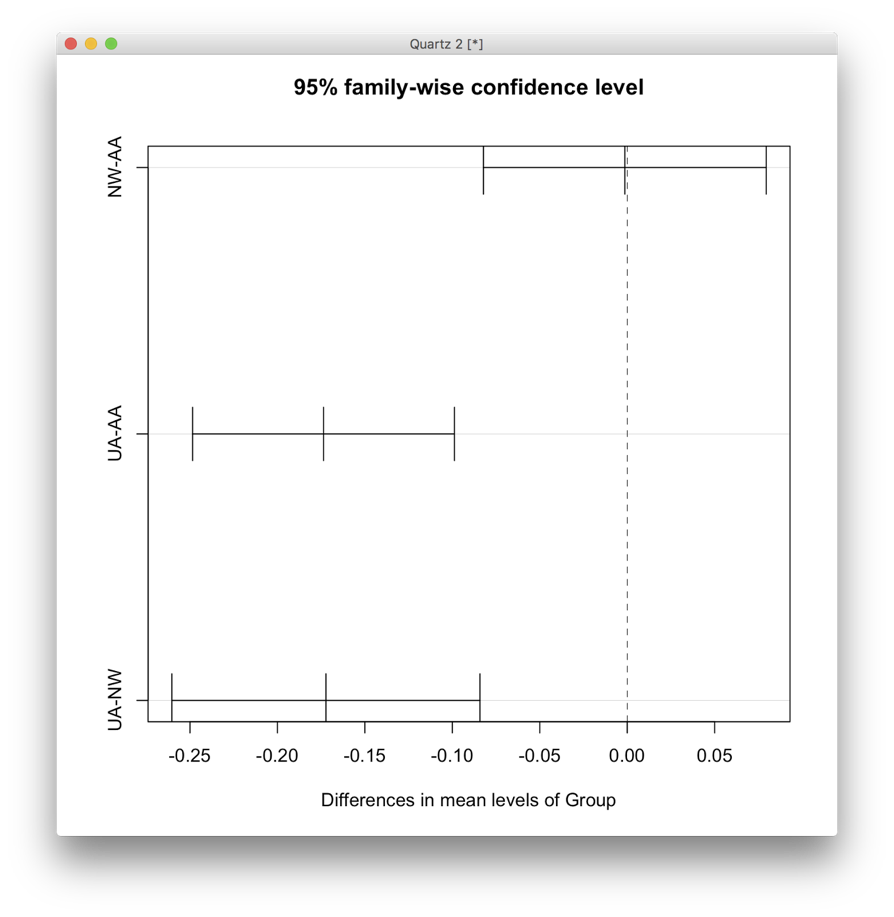
$Group

             diff         lwr         upr     p adj

NW-AA -0.001371685 -0.08220622  0.07946285 0.9991269

UA-AA -0.173655076 -0.24850871 -0.09880144 0.0000002

UA-NW -0.172283391 -0.26037215 -0.08419463 0.0000143



In this case, UA and AA are different and UA and NW are different. NW and AA are same.

This is easily seen using the following procedure:

1) Order the means in descending (or ascending order)

2) Draw a line when the groups are the same:

AA(1.998115) NW(1.996744) UA(1.824460)

------------------------------------------------------------------

Therefore the best company would be AA.

7.

From the ANOVA confidence interval, we can see there is no significant different between AA and NW. Because both of their ANOVA confidence interval contains 0.

But we can find UA is significantly less than AA and NW.