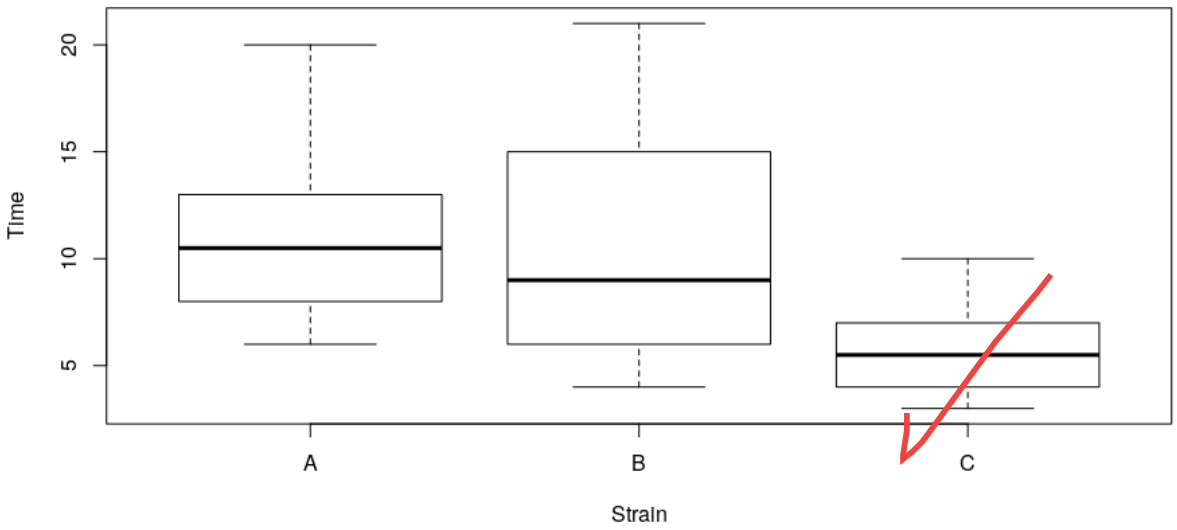
Colm Mooney – ST204 - 20325583

#3

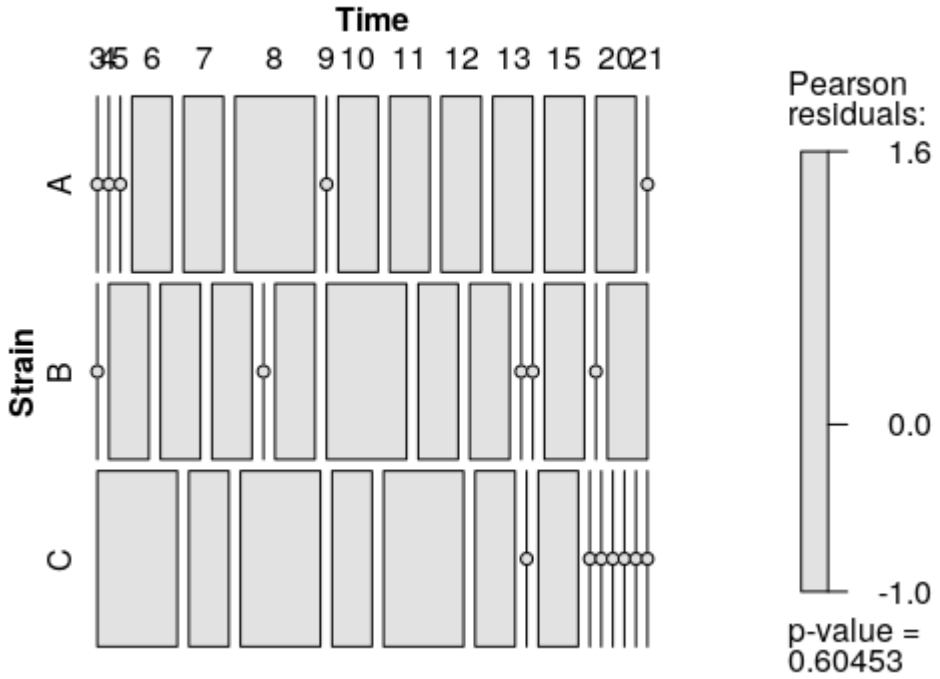
```
library(readr)

#(a)

CA3_FLU <- read_csv("CA3_Flu.csv")

CA3_FLU


#(a) Summarise the raw data using summary statistics and graphs as appropriate

tab_overall <- xtabs(~ Strain + Time, data = CA3_FLU)

addmargins(tab_overall)
```

```
> addmargins(tab_overall)
      Time
Strain  3  4  5  6  7  8  9 10 11 12 13 15 20 21 Sum
   A    0  0  0  1  1  2  0  1  1  1  1  1  1  0  10
   B    0  1  1  1  0  1  2  1  1  0  0  1  0  1  10
   C    2  1  2  1  2  1  0  1  0  0  0  0  0  0  10
   Sum  2  2  3  3  3  4  2  3  2  1  1  2  1  1  30
```


```
#From Tab overall, we can get the individual values for A,B and C.

A <- c(6,7,8,8,10,11,12,13,15,20)

B <- c(4,5,6,8,9,9,19,11,15,21)

C <- c(3,3,4,5,5,6,7,7,8,10)

y <- c(A, B, C)

x <- rep(c(1,2,3), c(10,10,10))

xf <- factor(x, labels=c("A", "B", "C"))

plot(y ~ xf, xlab = "Strain", ylab = "Time")
```



mosaic(~ Strain + Time, data = CA3_FLU, shade = TRUE, legend = TRUE)

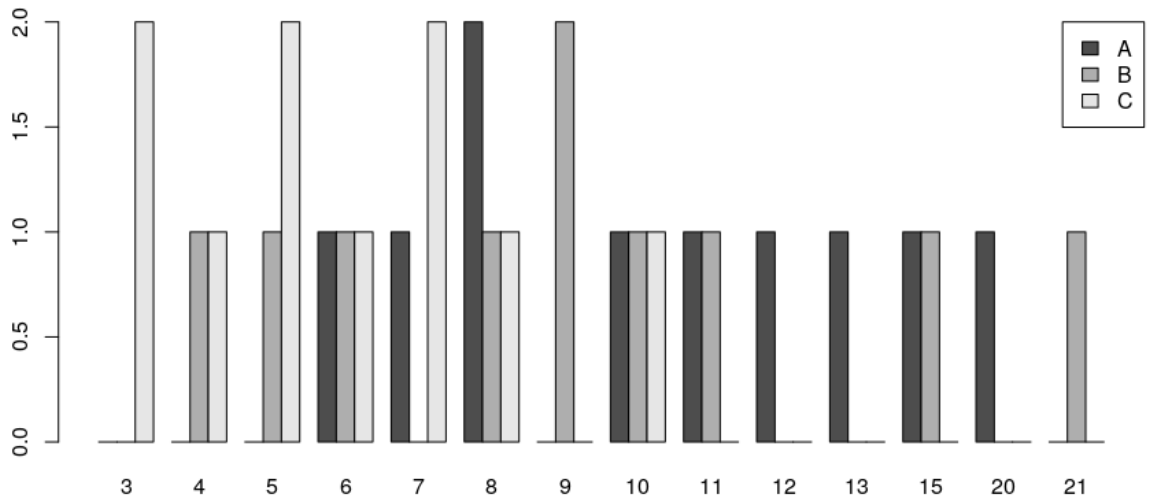


barplot(tab_overall,

legend = TRUE,

beside = TRUE,

args.legend = list(x = "topright"))



summary(tab_overall)

summary(CA3_FLU)

cdfA <- ecdf(A)

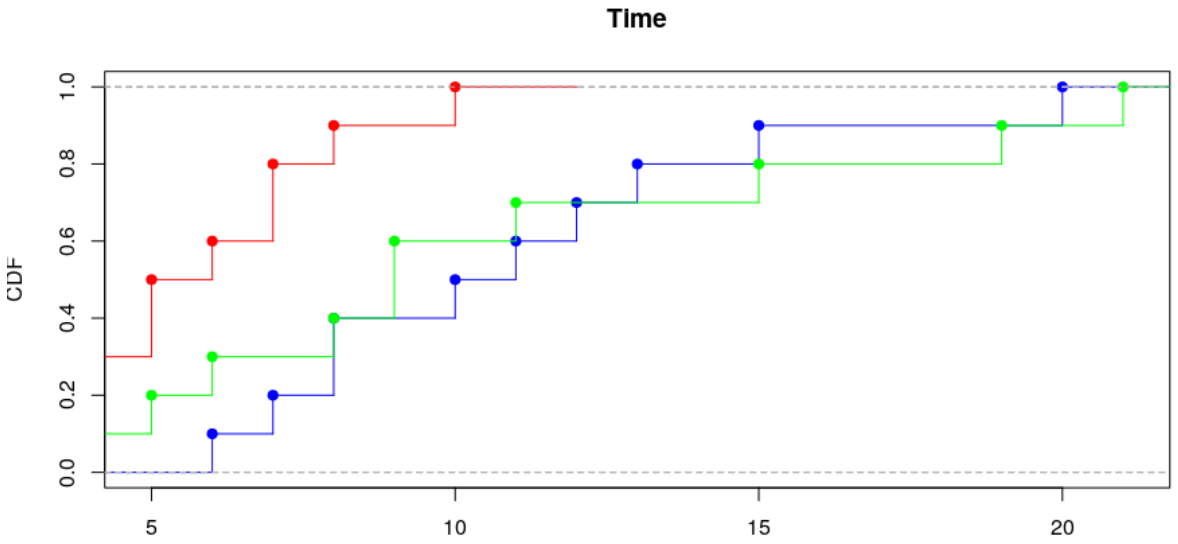
cdfB <- ecdf(B)

cdfC <- ecdf(C)

plot(cdfA, verticals=TRUE, col.points="blue", col="blue", main="Time", ylab="CDF", xlab="t")

lines(ecdf(B), verticals=TRUE, col.points="green", col="green")

lines(ecdf(C), verticals=TRUE, col.points="red", col="red")



#(b) Use a parametric test to compare the groups. State your conclusion.

#Find confidence intervals for the pairwise differences using Tukey's method.

#Assess the model fit via residual diagnostics.

#Part 1

```
A <- c(6,7,8,8,10,11,12,13,15,20)

B <- c(4,5,6,8,9,9,19,11,15,21)

C <- c(3,3,4,5,5,6,7,7,8,10)

y <- c(A, B, C)

x <- rep(c(1,2,3), c(10,10,10))

xf <- factor(x, labels=c("A", "B", "C"))

plot(y ~ xf)
```

####FIT ONE-WAY ANOVA MODEL (i.e. APPLY F TEST)

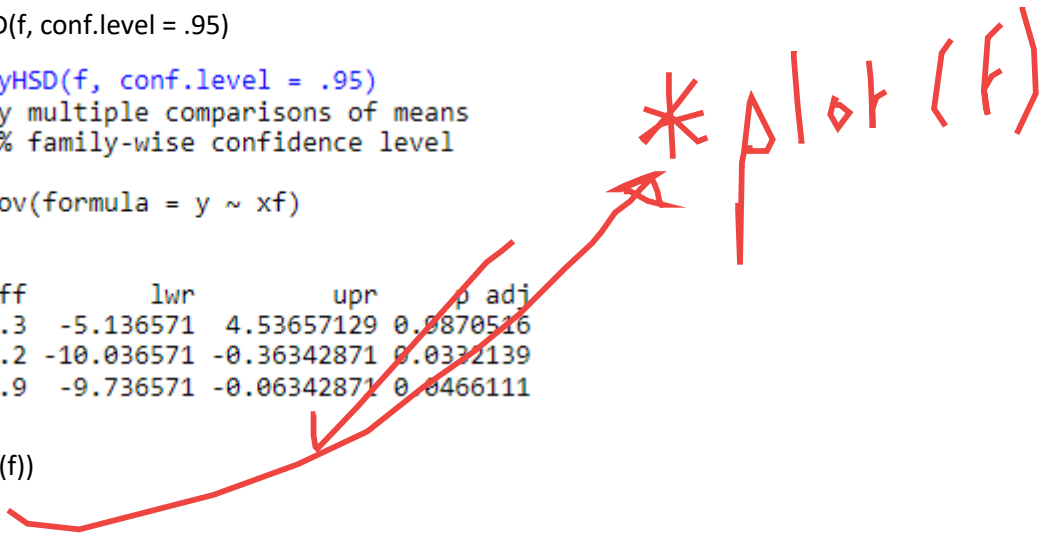
```
f <- aov(y ~ xf) # aov fits the model

anova(f)        # anova extracts the ANOVA table

summary(f)      # since the model *is* an ANOVA model, the summary is the same as an ANOVA table from anova()
```

#Part 2

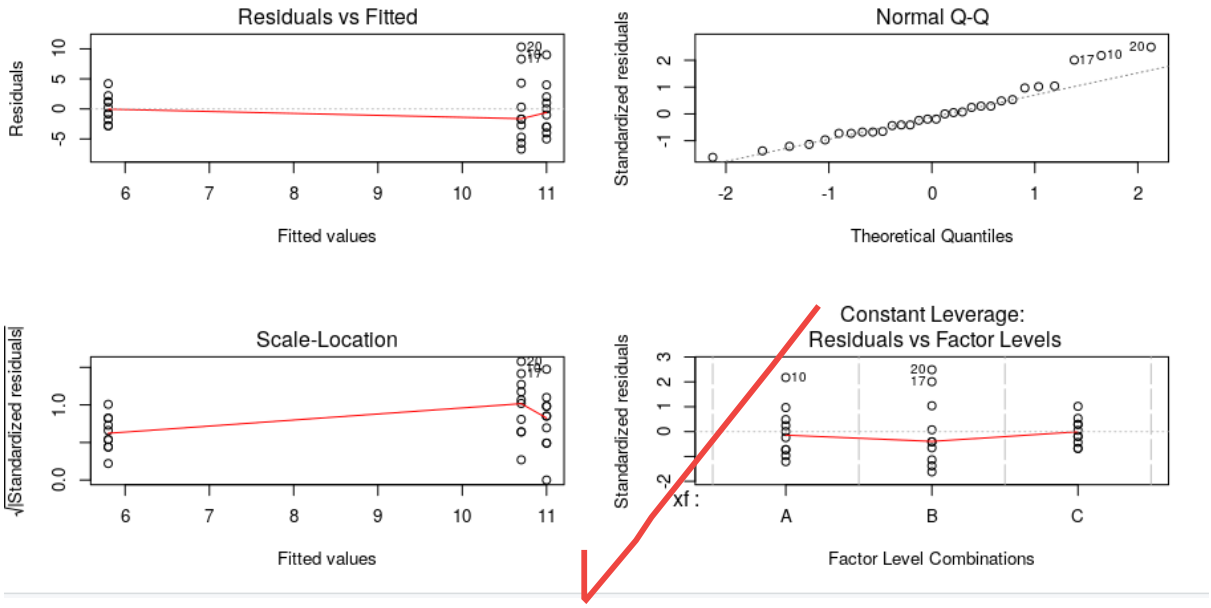
```
library(DescTools)

#Confidence Intervals are shown here:

TukeyHSD(f, conf.level = .95)
```

```
> TukeyHSD(f, conf.level = .95)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = y ~ xf)

$xf
     diff        lwr         upr     p adj
B-A -0.3  -5.136571  4.53657129 0.9870516
C-A -5.2 -10.036571 -0.36342871 0.0332139
C-B -4.9  -9.736571 -0.06342871 0.0466111
```

```
hist(resid(f))
```



```
par(mfrow=c(2,2))
plot(f, conf.level = .95)
par(mfrow=c(1,1))
```




```
# not normal according to QQ plot 3rd plot
# non-constant variance according to 1st and

#(c)Use nonparametric tests to compare the groups. Look for both overall and pairwise differences.
#State all appropriate hypotheses and conclusions
#do a Kruskal-Wallis test followed by pairwise Bonferroni adjusted WMW tests.

#Kruskal Wallis.
ks.test(A,B, alternative = "greater") #p-value: .90
ks.test(A,C, alternative = "greater")#P-value: 1
ks.test(B,C, alternative = "greater")#P-value' 1
#These are for Correlation
cor.test(A, B, alternative="greater")
cor.test(A, C, alternative="greater")
cor.test(B, C, alternative="greater")
```

#Bonferroni Adjustments

sigma <- 0.05

bottom<- 3 #3(3-1)/2

sigma/bottom #Bonferroni Level of significane is 0.01666667

wilcox.test(A,B) #p = 0.7

wilcox.test(A,C) #Significance #p = 0.003461

wilcox.test(B,C) #p = 0.03362

#A is significantly different from C, but the other comparisons are not significant.

#Conclusion: There is no correlation between each of the samples.

#(d) Which approach is more appropriate here?

#I believe that using the nonparametric tests are more appropriate to use here.

#The reasons for using nonparametric tests are:

#1: The population sample size is small enough.

#2: The underlying data do not meet the assumptions about the population sample.

#3: We are dealing with Ordinal data.

Index of comments