Charles Hesketh

Exam 2

Problem 1)

a)

Source	**Degrees of Freedom**	**Sum of Squares**	**Mean Squares**
Block	2	520	260
Treatment	4	498	124.5
Residual	8	40	5
Total	14	1058	

b) Using the qtukey function in R with an alpha value of 0.01 (level requested), 5 means, and 3 blocks then, using the Tukey test, we get a critical value of 4.684 to compare with our differences in treatments. It was assumed that there was an equal number of samples, since there was no evidence of the contrary, and thus each difference of means was divided by sqrt(5*2/3). This yielded the following combinations that were greater than the previously calculated 4.684, meaning they are statistically significant:

Combination	**Test Statistic**
A vs B	7.120
A vs E	6.025
B vs C	6.573
B vs D	7.120
C vs E	5.477
D vs E	6.025

c) The Tukey method of comparing treatments test whether there is a significant difference between two specific treatments. If this method found only one pair of significant differences, let alone six, then we expect the F test at the same level to have a p-value less then 0.01 since it is testing if there is any difference between any of the combinations. Thus, we would reject the null hypothesis that the treatments are the same based on the above six differences Tukey's method discovered.

Problem 2)

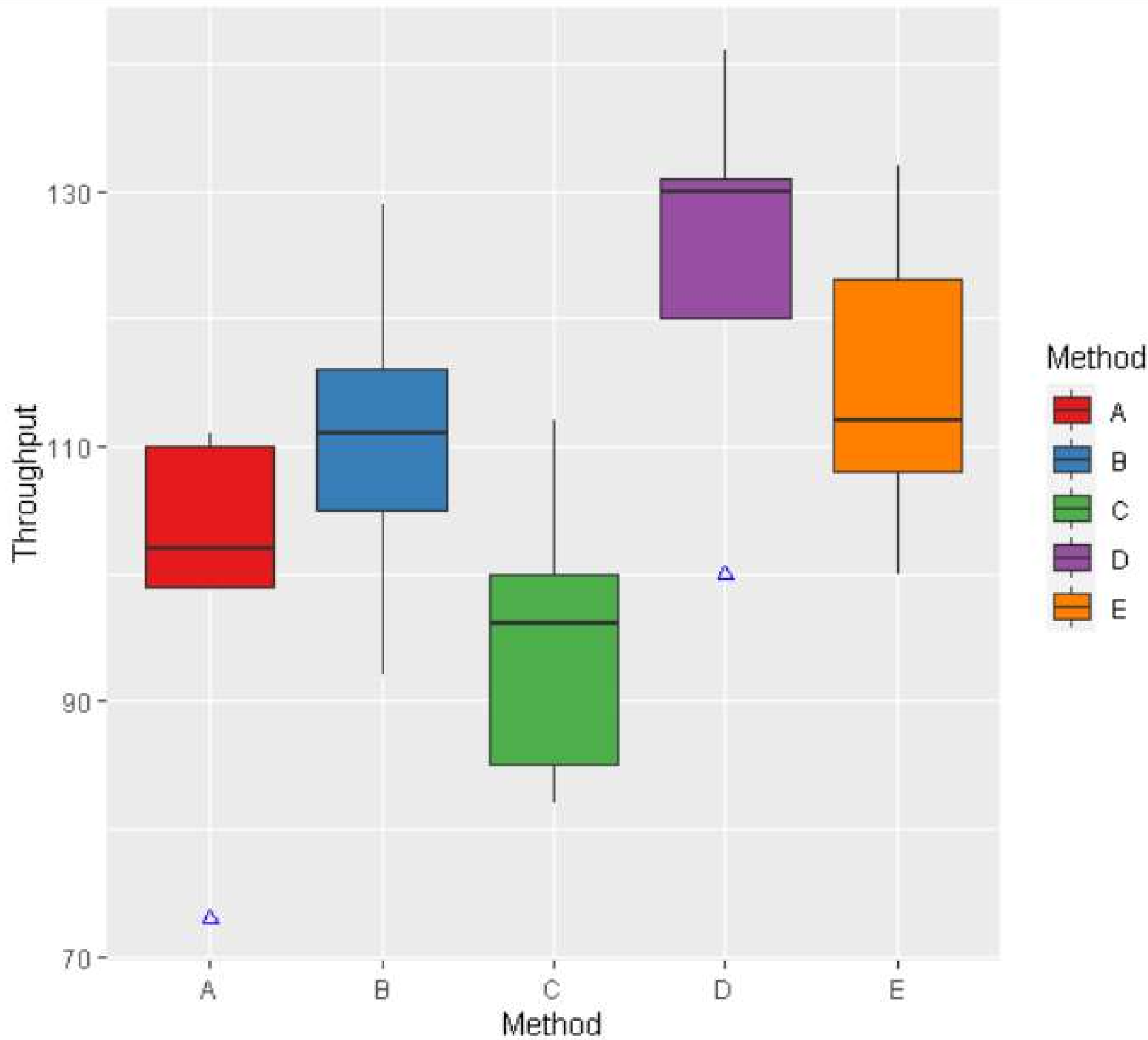
a) The experiment used was a paired comparison test using 1 block with 6 levels and 1 factor with 2 levels (A and B). This was used because we want to determine which catalyst would provide a better yield between two different catalysts, leading to a paired comparison as our initial analysis (which provided acceptable results and other tests were not needed to be attempted).
b) Since we wanted to know if B was better than A we used a one tailed t-test to obtain the p-value. The paired t-test produced a p-value of 0.02283 suggesting B has a greater yield then A at an alpha of 0.05.

c) In R we had to change the t.test call to allow for a two-sided test in order to get the appropriate confidence level. We have a mean difference of 2.333, a theoretical t-statistic of 2.5706, and a 95% CI of (0.0663,4.6004).

Problem3)

For analyzing the data I first plotted data using a box plot to show how the data is dispersed according to the different methods (A,B,C,D,E).

0



From this graph we would expect D to have the best results given its median Throughput is the greatest and C to have the worst results given its median is the least of the group. We will analyze the Graeco-Latin square's resutls using an anova, to determine if there is indeed a difference between methods, and a pair wise comparison, via the Tukey Method, to isolate the differences in methods. The data columns labaled Day, Operator, Machine, and Method were converted into factors and fed through the aov function in R with Throughput being the response variable. This led to a p-value of .00004 in respect to the Method variable. This needs further analysis to explain which pairs of Methods are signifcantly different. Using the Tukey method again with 5 means, 8 degrees of freedom, and an alpha of 0.01 yields a test statistic of 3.4547. The comparisons between methods yields:

Comparison	Statistic	Greater than 3.45475
A-B	**4.0608**	*
D-A	**8.8918**	*
E-A	**5.6011**	*
C-B	**5.4611**	*
D-B	**4.8310**	*
D-C	**10.2921**	*
E-C	**7.0014**	*
C-A	1.4003	
E-B	1.5403	
E-D	3.2907	

This implies that there are 7 different comparisons where a difference is significant between the methods (namely the first 7 in the above table).

Appendix Code:

```
library(reshape2)
library(DescTools)
library(ggplot2)

setwd("C:\\Users\\User\\Desktop\\School\\Math_531T\\Exam2")



####################################
#1a)




#b)

alpha=.01

qtukey(1-alpha,5,8)/sqrt(2)

yA=45
yB=58
yC=46
yD=45
yE=56

AB=abs(yA-yB)/sqrt(5*2/3)
AC=abs(yA-yC)/sqrt(5*2/3)
AD=abs(yA-yD)/sqrt(5*2/3)
```

AE=abs(yA-yE)/sqrt(5*2/3)

BC=abs(yB-yC)/sqrt(5*2/3)

BD=abs(yB-yD)/sqrt(5*2/3)

BE=abs(yB-yE)/sqrt(5*2/3)

CD=abs(yC-yD)/sqrt(5*2/3)

CE=abs(yC-yE)/sqrt(5*2/3)

DE=abs(yD-yE)/sqrt(5*2/3)

AB

AE

BC

BD

CE

DE

#c)

#Since these are above the calculated Tukey, then we would expect the F test to fail at a level of 0.01.

######################################

#2a)

#Paired comparison test with 6 blocks and 2 factors (A and B).

#b)

data=c(9,19,28,22,18,8,10,22,30,21,23,12)

```
data

qt(0.975,5)

#p-value:
t.test(data[7:12],data[1:6],paired = T,alternative="greater")

#c)
#for confidence interval:
t.test(data[7:12],data[1:6],paired = T,alternative="two.sided")


######################################
#3)

data=read.table("https://www2.isye.gatech.edu/~jeffwu/book/data/throughput.dat", h=T)
data

ggplot(data=data,aes(x=Method,y=Throughput,fill=Method))+geom_boxplot(outlier.shape=2,outlier.col
our="blue")+scale_fill_brewer(palette="Set1")

results=aov(formula = Throughput ~ as.factor(Day) + as.factor(Operator)+ as.factor(Machine)+
as.factor(Method), data = data)
summary(results)
anova(results)

analysis=PostHocTest(results,method="hsd")$'as.factor(Method)'

qtukey(.95,5,8)/sqrt(2) #2*4=8
```

```
abs(analysis[,1])/sqrt(20.4*(2/5))
```