Homework 5

John Kaspers

10/12/2020

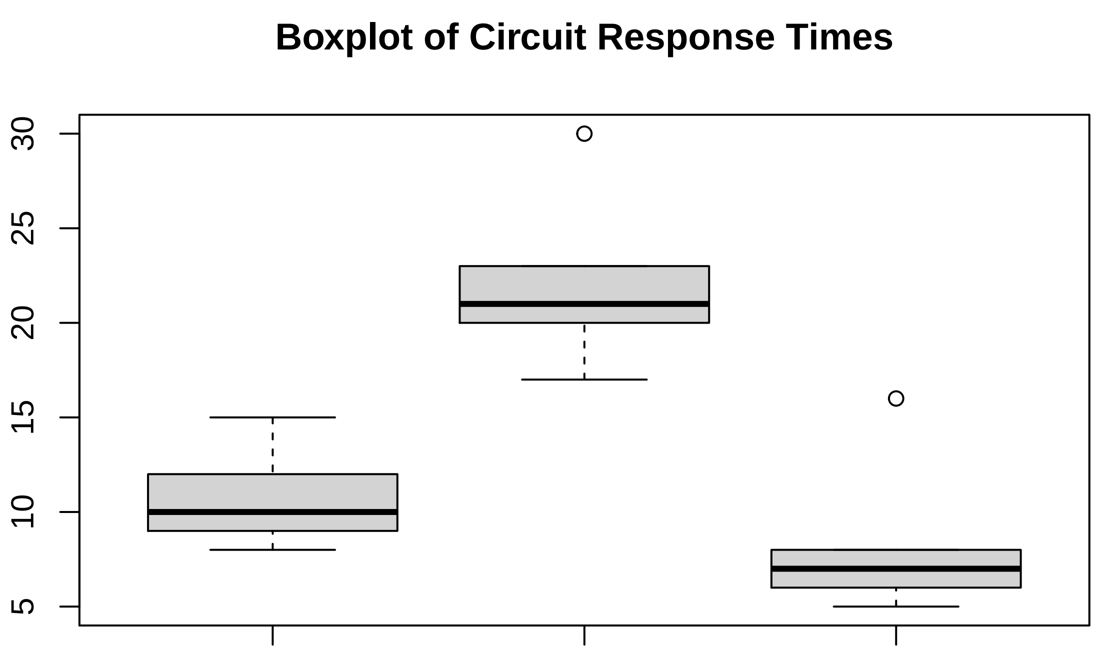
Questions for practice / review

Problem 1

#(i) # Draw comparative boxplot

```
Circuit1 <- c(9, 12, 10, 8, 15)
Circuit2 <- c(20, 21, 23, 17, 30)
Circuit3 <- c(6, 5, 8, 16, 7)

boxplot(Circuit1, Circuit2, Circuit3, main = "Boxplot of Circuit Response Times")
```




```
#Circuit2 has a higher mean and median response time than the other two, and Circuit3 has the lowest me
```

(ii)

```r
type = sort(rep(1:3,5))
response_time = c(9, 12, 10, 8, 15, 20, 21, 23, 17, 30, 6, 5, 8, 16, 7)
Type = as.factor(type)
model <- lm(response_time~Type)
summary.aov(model)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)    
## Type         2  543.6   271.8   16.08 0.000402 ***
## Residuals   12  202.8    16.9                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

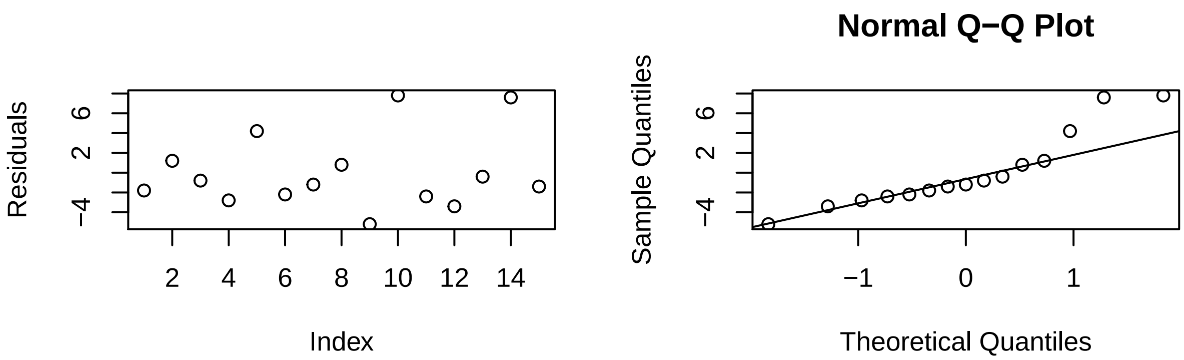
H_0: Response time Circuit 1 = Response time Circuit 2 = Response time Circuit 3

H_A: At least one response time is different

The appropriate test statistic is F = 16.08 on 2 and 12 degrees of freedom. The critical value (p-value) is 0.000402, which is less than alpha = 0.01. Therefore, we can reject the null hypothesis that there is no difference between mean response time across the three different circuit types and we can say there is evidence to prove that at least one response time is different.

(iii) Comment on model adequacy by examining the standardized residual plots. Test for equality of variances using Bartlett's test

```r
model_fit = model$fitted
residuals = model$residuals
par(mfrow=c(2,2))
stripchart(residuals~Type,vertical=T,xlab="Circuit Type",ylab="Residuals") # Testing constant variance
plot(model_fit,residuals,xlab="Y-hat",ylab="Residuals")#Testing if there are outliers
plot(residuals,ylab="Residuals") #against index; this could be run order
qqnorm(residuals)#is sample drawn from a normal distribution?
qqline(residuals)
```




```r
bartlett.test(response_time~Type)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  response_time by Type
## Bartlett's K-squared = 1.1345, df = 2, p-value = 0.5671
```

#(iv) If appropriate, compare pairs of treatment means using the following methods. In using each of the methods, specify clearly in what context the interval is appropriate

```r
# Tukey HSD (alpha = 0.05)

# Compare using Bonferroni's and Scheffe's procedures using the same significance level

# Would Dunnet's test be appropriate here (no calculations required)? Why?
```

(v) If you were a design engineer and you wished to minimize the response time, which circuit type would you select?

Answer

Questions for completion

#Problem 2

```r
#(i) Construct a set of orthogonal contrasts in treatment means

#Orthogonal contrasts
c1=c(1, -2, 1) #comparing Circuit 2 with other two
c2=c(1, 0, -1) #additional orthogonal contrast

#(ii) What are the estimates of these contrasts? Confirm sums of squares of these contrasts sum to sums
Circuit1 <- c(9, 12, 10, 8, 15)
Circuit2 <- c(20, 21, 23, 17, 30)
Circuit3 <- c(6, 5, 8, 16, 7)
Circuit = c(Circuit1, Circuit2, Circuit3)

#estimating the contrasts
contrast_estimate_1 <- mean(Circuit1)
contrast_estimate_2 <- mean(Circuit2)
contrast_estimate_3 <- mean(Circuit3)

#First contrast estimate:
contrast_estimate_1 - 2*contrast_estimate_2 + contrast_estimate_3
```

```
## [1] -25.2
```

```r
#Second contrast estimate
contrast_estimate_1 - contrast_estimate_3
```

```
## [1] 2.4
```

```r
#aov
a=3
n = 5
grp=as.factor(c(rep(1,5), rep(2,5), rep(3,5)))
gmeans=tapply(Circuit, grp, mean)
m=aov(Circuit~grp)

mse=summary(m)[[1]]$"Mean Sq"[2]
mse
```

```
## [1] 16.9
```

```r
dfe=m$df.residual
alpha=0.05

ss1=((sum(c1*gmeans))^2)/(sum(c1^2)/n);ss1
```

```
## [1] 529.2
```

```r
ss1
```

```
## [1] 529.2
```

```r
ss2=((sum(c2*gmeans))^2)/(sum(c2^2)/n);ss2
```

```
## [1] 14.4
```

```r
ss2
```

```
## [1] 14.4
```

```r
ss1+ss2 #Check that it adds up to SSTr
```

```
## [1] 543.6
```

```r
summary(m)[[1]]$"Sum Sq"[1]#SSTr
```

```
## [1] 543.6
```

```r
# (iii) Get a 95% C.I. for the contrasts using Scheffe's procedure. What do these tell us about these c

c1m=sum(c1*gmeans); c1m# mean contrasts
```

```
## [1] -25.2
```

```r
c2m=sum(c2*gmeans); c2m# mean contrasts
```

```
## [1] 2.4
```

```r
sc1=sqrt(mse*sum(c1^2)/n);sc1#std error
```

```
## [1] 4.503332
```

```r
sc2=sqrt(mse*sum(c2^2)/n);sc2#std error
```

```
## [1] 2.6
```

```r
scrit1=sqrt((a-1)*qf(1-alpha,2,dfe))*sc1 #quantile*se
scrit2=sqrt((a-1)*qf(1-alpha,2,dfe))*sc2 #quantile*se

#Confidence interval for C1
c1m-scrit1;c1m+scrit1
```

```
## [1] -37.75339
```

```
## [1] -12.64661
```

```r
#Confidence interval for C2
c2m-scrit2;c2m+scrit2
```

```
## [1] -4.847701
```

```
## [1] 9.647701
```

#Problem 3 *Answer* is directly on Gradescope Does type of stove, gas or electric, lead to a different mean time to boil 2 cups of water, controlling for volume and temperature of water.

Questions graded for correctness

Problem 4. Q3.50

```r
Brand1 = c(100, 96, 92, 96, 92)
Brand2 = c(76, 80, 75, 84, 82)
Brand3 = c(108, 100, 96, 98, 100)

Weeks_of_Life = data.frame(Brand1, Brand2, Brand3)
Weeks_of_Life
```

```
##   Brand1 Brand2 Brand3
## 1    100     76    108
## 2     96     80    100
## 3     92     75     96
## 4     96     84     98
## 5     92     82    100
```

```r
#In order to get an estimate of sigma^2, we must find an average of the sd across the three brands of b
sigma_hat <- (sd(Brand1)+sd(Brand2)+sd(Brand3))/3

#Referring to lab06,
alpha=.01
a=3 #number of treatments
D=10 #desired diff in means to detect with prob 1beta
sigsq=sigma_hat^2 #our estimate of sigma^2
ni=2:10 #the common sample size for each treatment (try several)
Fcrit=qf(1-alpha,a-1,a*(ni-1))#value at which we reject H0
lam=ni*(D^2)/(2*sigsq) #noncentrality paramter (ncp)
beta=pf(Fcrit,a-1,a*(ni-1),ncp=lam)
power=1-beta
cbind(ni,Fcrit,beta,power) #display results by combining in columns
```

```
##       ni     Fcrit       beta      power
##  [1,]  2 30.816520 0.93257412 0.06742588
##  [2,]  3 10.924767 0.76251099 0.23748901
##  [3,]  4  8.021517 0.53787107 0.46212893
##  [4,]  5  6.926608 0.33648189 0.66351811
##  [5,]  6  6.358873 0.19144974 0.80855026
```

```
## [6,]  7  6.012905 0.10095960 0.89904040
## [7,]  8  5.780416 0.05003070 0.94996930
## [8,]  9  5.613591 0.02353852 0.97646148
## [9,] 10  5.488118 0.01059653 0.98940347
```

```
#Thus, we should use a sample size of 8 or more.
```

Answer Thus, a sample size of 8 should be used. To get a preliminary estimate of sigma^2, I first calculate the sd of each brand of battery, then I summed those up and divided by 3.

(b)

```
alpha=.01
a=3 #number of treatments
P=25 #desired percentage increase in sd to detect with prob 1beta
ni=10:20 #the common sample size for each treatment (try several)
Fcrit=qf(1-alpha,a-1,a*(ni-1))#value at which we reject H0
lam=a*ni*((1+.01*P)^2-1) #noncentrality paramter (ncp)
beta=pf(Fcrit,a-1,a*(ni-1),ncp=lam)
power=1-beta
cbind(ni,Fcrit,beta,power) #display results by combining in columns
```

```
##       ni    Fcrit        beta     power
##  [1,] 10 5.488118 0.190325663 0.8096743
##  [2,] 11 5.390346 0.137717655 0.8622823
##  [3,] 12 5.312029 0.097880152 0.9021198
##  [4,] 13 5.247894 0.068444533 0.9315555
##  [5,] 14 5.194413 0.047157162 0.9528428
##  [6,] 15 5.149139 0.032052624 0.9679474
##  [7,] 16 5.110318 0.021515894 0.9784841
##  [8,] 17 5.076664 0.014277425 0.9857226
##  [9,] 18 5.047210 0.009373481 0.9906265
## [10,] 19 5.021217 0.006093079 0.9939069
## [11,] 20 4.998110 0.003924159 0.9960758
```

```
#power.t.test(n = NULL, delta = 10, sd = 1.25*(sigma_hat), sig.level = 0.05, power =
#0.90, type = c("two.sample"), alternative = c("two.sided"))
```

Answer Given an increase of 25% in Standard Deviation, we would need a sample size of at least 12, based on the above output.

#(c) If we wish to construct a 95% confidence interval on the difference in two mean battery lives that has an accuracy of 2 weeks, how many batteries of each brand must be tested?

```
n= seq(20, 40, 1)
alpha=0.05
dfe = 3*(n-1)
MSE = 15.6
qt <- qt(1-(alpha/2), dfe)
se = sqrt((2*MSE)/n)
num = qt*se
cbind(n, num)
```

```
##         n      num
##  [1,] 20 2.501079
##  [2,] 21 2.438161
##  [3,] 22 2.379773
##  [4,] 23 2.325395
##  [5,] 24 2.274588
##  [6,] 25 2.226976
##  [7,] 26 2.182239
##  [8,] 27 2.140096
##  [9,] 28 2.100308
## [10,] 29 2.062661
## [11,] 30 2.026971
## [12,] 31 1.993073
## [13,] 32 1.960822
## [14,] 33 1.930089
## [15,] 34 1.900759
## [16,] 35 1.872727
## [17,] 36 1.845901
## [18,] 37 1.820196
## [19,] 38 1.795536
## [20,] 39 1.771853
## [21,] 40 1.749084
```

Answer So we would need to test 31 or more batteries to get an accuracy of 2 weeks.

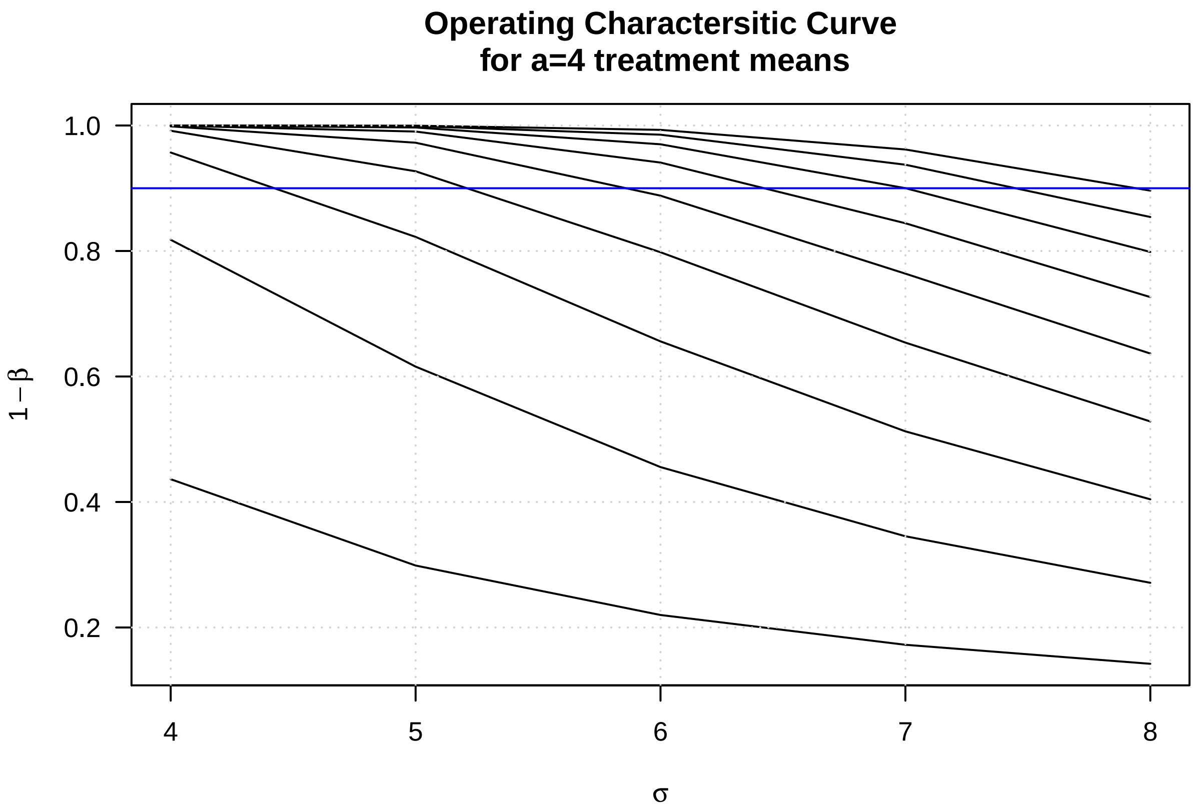
Problem 5

```r
alpha = 0.05
sd <- c(4, 5, 6, 7, 8)
nn <- seq(2, 10, 1)
group.means <- c(50, 60, 50, 60)
power <- matrix(NA, nr = length(sd), nc = length(nn))

for(i in 1:length(sd))
  power[i,] <- power.anova.test(groups = 4, n = nn, between.var = var(group.means), within.var = sd[i]^2
  colnames(power) <- nn; rownames(power) <- sd

  opar <- par(las=1, cex = 0.8)
  matplot(sd, power, type = "l", xlab = expression(sigma), ylab = expression(1-beta), col = 1, lty = 1)
grid()
text(rep(10, 2), power[length(sd)], as.character(nn), pos = 3)
title("Operating Charactersitic Curve\n for a=4 treatment means")
par(opar)

abline(h = 0.9, col = "Blue")
```



Operating Charactersitic Curve
for a=4 treatment means
1.0
0.8
0.6
0.4
0.2
1 − β
4
5
6
7
8
σ