CMDA-2006 Homework

Homework

Your name here

Due

```
setwd("C:/Users/jlutz/OneDrive/Documents/School/Spring Sem 2022/CMDA-STAT/Homework/HW_7")
```

Problem 1

Perform a one-way analysis of variance to test whether there is a significant difference in the average leaf sizes under the four growing conditions. Use alpha = 0.05. ## Part a

Example Chunk Below

```
Leaf_A <- c(27.7619, 27.8523, 21.3495, 31.9616, 19.4623, 12.2804, 21.0508, 19.5074, 26.2808, 26.1466)

Leaf_B <- c(4.2460, 14.1577, 7.0279, 7.0698, 0.8091, 13.9385, 11.0130, 10.9680, 6.9112, 9.6041)

Leaf_C <- c(15.5070, 5.0473, 18.3020, 16.0436, 10.2601, 19.0571, 17.1826, 16.6510, 18.8472, 12.4234)

Leaf_D <- c(33.0101, 44.9680, 34.2074, 28.9766, 42.9229, 36.6827, 32.7229, 34.5668, 28.7695, 36.6952)

Leaf_Data <- c(Leaf_A, Leaf_B, Leaf_C, Leaf_D)
leaves <- c(rep("Leaf A", 10), rep("Leaf B", 10), rep("Leaf C", 10), rep("Leaf D", 10))
anovafit0 <- aov(Leaf_Data ~ leaves)
print(anova0 <- anova(anovafit0))


Analysis of Variance Table

Response: Leaf_Data
          Df Sum Sq Mean Sq F value    Pr(>F)
leaves     3 4020.0  1340.0  54.699 1.736e-13 ***
Residuals 36  881.9    24.5
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Part b

What conclusions can you reach concerning the effect of growing conditions on the average leaf size?

Since our p-value, 1.736e-13 < alpha = 0.05, we reject the null hypothesis. We have enough statistical evidence to conclude that there is a significant difference concerning the effect of growing conditions on the average leaf size.

Part c

Perform a one-way analysis of variance to test whether there is a significant difference in the average nicotine contents under the four growing conditions. Use alpha = 0.05.

```
Nico_A <- c(10.0655, 9.4712, 9.1246, 11.3652, 11.3976, 11.2936, 10.6805, 8.1280, 10.5066, 10.6579)

Nico_B <- c(8.5977, 8.1299, 11.3401, 9.3470, 9.3049, 10.0193, 9.5843, 6.4603, 8.2589, 5.0106)

Nico_C <- c(6.7865, 10.9249, 11.3878, 9.7022, 8.0371, 10.7187, 11.2352, 7.7079, 7.5653, 9.0922)

Nico_D <- c(9.9553, 5.8495, 10.3005, 9.7140, 10.7543, 8.0262, 13.1326, 11.8559, 11.3345, 10.4763)

Nico_Data <- c(Nico_A, Nico_B, Nico_C, Nico_D)
titles <- c(rep("A Nico", 10), rep("B Nico", 10), rep("C Nico", 10), rep("D Nico", 10))

anovafit1 <- aov(Nico_Data ~ titles)
print(anova1 <- anova(anovafit1))
```

```
Analysis of Variance Table

Response: Nico_Data
         Df  Sum Sq Mean Sq F value Pr(>F)
titles    3  18.081  6.0271  2.1031  0.117
Residuals 36 103.170  2.8658
```

Part d

What conclusions can you reach concerning the effect of growing conditions on the average nicotine content?

Since our p-value, $0.117 >$ alpha $= 0.05$, we fail to reject the null hypothesis. We do not have enough statistical evidence to conclude that there is a significant difference concerning the effect of growing conditions on the average nicotine content.

Part e

Based on the conclusions you reached in parts (b) and (d), do you think the testimony of the tobacco companies scientists is supported by this experiment? Justify your conclusions.

I believe the testimony of the scientists is not supported by this experiment. From the one-way analysis of variance on the average leaf sizes from the four growing conditions, we determined that their is statistical evidence to conclude that their is a significant difference concerning the effect of growing conditions on the average leaf size. In the the one-way analysis on the growing conditions on the average nicotine content, we determined there is not sufficient statistical evidence to conclude significant difference concerning the growing conditions on the average nicotine content. Since there was evidence for significant difference concerning the effect of growing conditions on the average leaf size, but not on the average nicotine content, we have evidence to support that the companies scientists testimony was not supported by this experiment.

Part f

If you observe significant F -tests in either parts (b) or (d). Carry out the pair-wise tests to determine which mean(s) are different.

```
pairwise.t.test(Leaf_Data, leaves, conf.level = 0.95)

    Pairwise comparisons using t tests with pooled SD

data:  Leaf_Data and leaves

       Leaf A  Leaf B  Leaf C
Leaf B 3.4e-07 -       -
Leaf C 0.0010  0.0068  -
Leaf D 1.3e-05 1.8e-13 2.6e-10

P value adjustment method: holm
```

Part g

Do the nicotine content data suggest violations of the AOV conditions? If you determine that the conditions are not met, perform an alternative analysis, and compare your results to the earlier analysis. Additionally if you observe a significant test, do the pair-wise comparisons using a non-parametric procedure with a Bonferroni correction.

```
#install.packages("car")
library(car)
bartlett.test(Nico_Data ~ titles)
```

```
        Bartlett test of homogeneity of variances

data:  Nico_Data by titles
Bartlett's K-squared = 3.3479, df = 3, p-value = 0.341


leveneTest(Nico_Data ~ titles)


Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  3  0.7624 0.5226
      36


fligner.test(Nico_Data ~ titles)


        Fligner-Killeen test of homogeneity of variances

data:  Nico_Data by titles
Fligner-Killeen:med chi-squared = 2.4446, df = 3, p-value = 0.4854
```

Since all three tests agree that we do not reject the null hypothesis, we will assume that all the variances are equal, we do not have suggestions of violations of the AOV conditions.

```
pairwise.t.test(Nico_Data, titles, p.adjust.method = "bonf")


        Pairwise comparisons using t tests with pooled SD

data:  Nico_Data and titles

       A Nico B Nico C Nico
B Nico 0.21   -      -
C Nico 1.00   1.00   -
D Nico 1.00   0.30   1.00

P value adjustment method: bonferroni
```

Problem 2

```
selenium <- read.csv("selenium.csv", header = T)
```

Part a

First, conduct the analysis of variance on this data set.

```
selenium$Meat <- factor(selenium$Meat)
anovafit2 <- aov(Selenium ~ Meat, data = selenium)
print(anova2 <- anova(anovafit2))


Analysis of Variance Table

Response: Selenium
           Df Sum Sq Mean Sq F value    Pr(>F)
Meat        3  21262  7087.3  27.001 7.684e-14 ***
Residuals 140  36747   262.5
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Part b

Specifically identify the null and alternative hypothesis statements for the contrasts that we are testing in the above questions.

For Question 1 : H0: VEN <= SQU HA: VEN > SQU For Question 2: H0: RRB == NRB HA: RRB != NRB For Question 3: H0: (VEN + SQU) == (RRB + NRB) HA: (VEN + SQU) != (RRB + NRB)

Part c

Using the Bonferroni method, find the contrast estimates, along with their 95% simultaneous confidence intervals and the adjusted p-values for the above research questions. This can be done using a single contrast matrix in R. Display the results, and provide plots of the confidence intervals.

```
library(DescTools)
pairwise.t.test(selenium$Selenium, selenium$Meat, conf.level = 0.95, p.adjust.method = "bonf")


    Pairwise comparisons using t tests with pooled SD

data:  selenium$Selenium and selenium$Meat

    NRB      RRB     SQU
RRB 2.2e-11 -       -
SQU 0.00072 0.00117 -
VEN 1.7e-12 1.00000 9.0e-05

P value adjustment method: bonferroni


t <- 4
k <- t * (t - 1)/2
df.within <- anova2[2,"Df"]
library(multcomp)
comparisons <- glht(anovafit2, linfct = mcp(Meat = "Tukey"))
plot(confint(comparisons, calpha=qt(1-0.05/2/k, df.within)),
     main = "Bonferroni 95% family-wise confidence level")
```
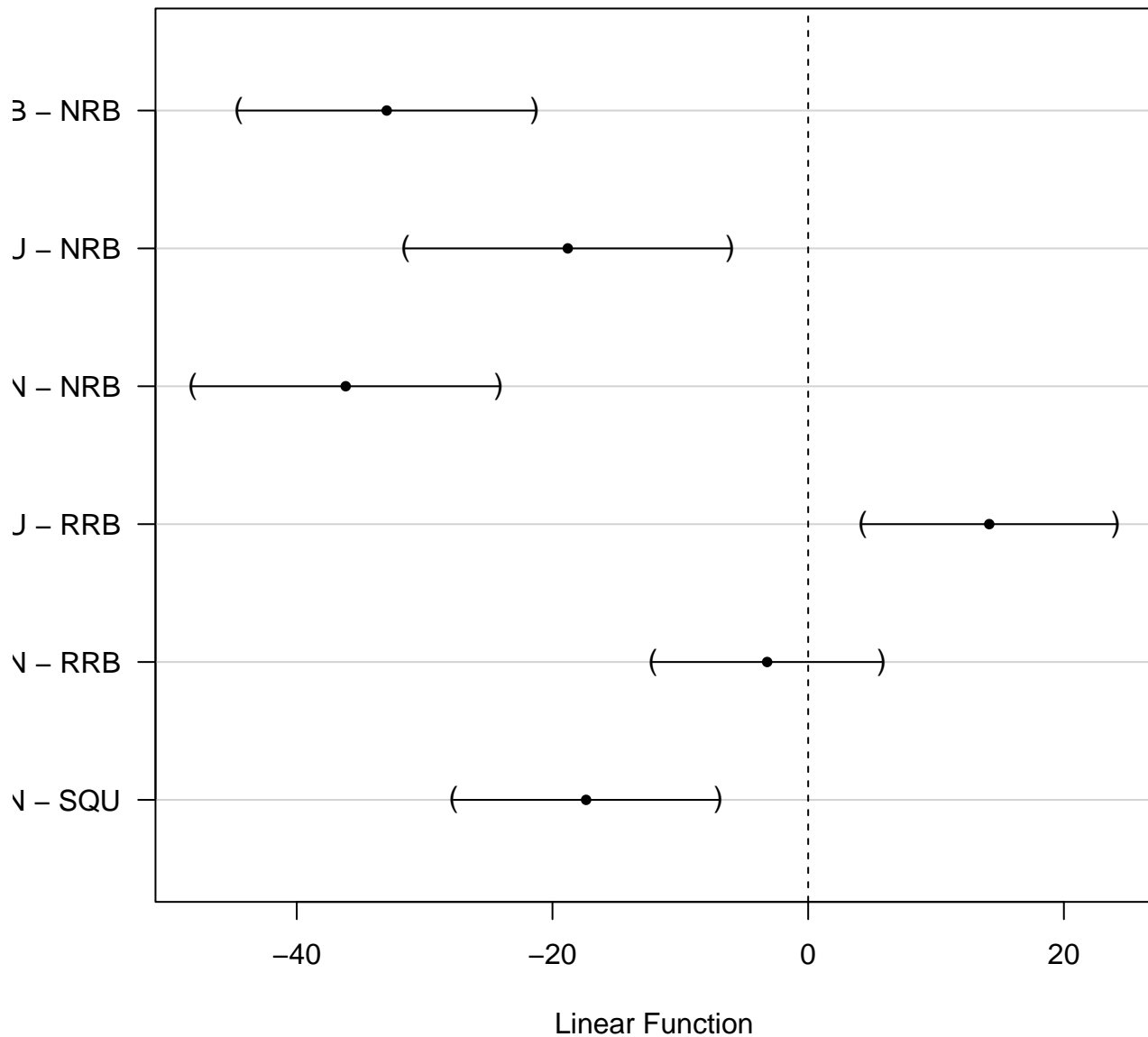
Bonferroni 95% family–wise confidence level

Part d

Don't forget to answer the original research questions using part (c).

Since our p-value 7.684e-14 < alpha = 0.05, we reject h). We have sufficient statistical evidence to conclude that there is a significant difference.

Part e

Finally, use Tukeys method to conduct a post-hoc analysis of all of the pairwise contrast estimates, their 95% confidence intervals. Display the results, and provide a plot of the confidence intervals. Comment on any interesting results that you see.

```
PostHocTest(anovafit2, method = "bonferroni")
```

```
 Posthoc multiple comparisons of means : Bonferroni
   95% family-wise confidence level

$Meat
              diff     lwr.ci     upr.ci    pval
RRB-NRB -32.963297 -44.557180 -21.369413 2.2e-11 ***
SQU-NRB -18.800649 -31.513356  -6.087942 0.00072 ***
VEN-NRB -36.170840 -48.158678 -24.183001 1.7e-12 ***
SQU-RRB  14.162648   4.256201  24.069094 0.00117 **
VEN-RRB  -3.207543 -12.164833   5.749748 1.00000
VEN-SQU -17.370190 -27.734930  -7.005451 9.0e-05 ***


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

plot(confint(comparisons, calpha=qt(1-0.05/2/k, df.within)),
     main = "Bonferroni 95% family-wise confidence level")
```


Bonferroni 95% family–wise confidence level