STAA 553: HW3

YOUR NAME HERE

See Canvas Calendar for due date.
40 points total, 2 points per problem unless otherwise noted.
Add or delete code chunks as needed.
Content for all questions is from Section 05 or earlier.

Weight Loss (Q1 - Q5)

We return to the weight loss study from HW2. Ott & Longnecker describe a weight loss study with g = 5 treatments (C, T1, T2, T3, T4). The response variable is weight loss (in pounds). A total of 50 (human) subjects were randomly assigned to treatments such that there are n = 10 subjects per treatment. The data is available from Canvas as WtLoss.csv.

Q1 (0 pts)

Fit an appropriate model (with default contrasts) and show the ANOVA table. Because we already did this for HW2, this question is worth 0 pts.

```
##   Trt Loss
## 1  T1 12.4
## 2  T1 10.7
## 3  T1 11.9
## 4  T1 11.0
## 5  T1 12.4
## 6  T1 12.3
```

```
## Analysis of Variance Table
##
## Response: Loss
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Trt        4 61.618 15.4045  15.681 4.164e-08 ***
## Residuals 45 44.207  0.9824
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Q2

Use the emmeans package to calculate Tukey adjusted pairwise comparisons.

Q2A

Show the Tukey adjusted comparisons (including estimates and p-values).

```
##  contrast estimate    SE df t.ratio p.value
##  C - T1      -2.78 0.443 45  -6.272  <.0001
##  C - T2      -1.75 0.443 45  -3.948  0.0024
##  C - T3      -1.00 0.443 45  -2.256  0.1784
##  C - T4      -2.97 0.443 45  -6.700  <.0001
##  T1 - T2      1.03 0.443 45   2.324  0.1563
##  T1 - T3      1.78 0.443 45   4.016  0.0020
##  T1 - T4     -0.19 0.443 45  -0.429  0.9927
##  T2 - T3      0.75 0.443 45   1.692  0.4490
##  T2 - T4     -1.22 0.443 45  -2.752  0.0618
##  T3 - T4     -1.97 0.443 45  -4.444  0.0005
##
## P value adjustment: tukey method for comparing a family of 5 estimates
```

Q2B

Comparing Tukey comparisons to unadjusted comparisons (from previous assignment), do we find evidence of more or fewer differences? Use alpha = 0.05.

Response Yes there are fewer differences. In the last assignment, we saw all but 1 were signifficant. here it is all but 4. *****

Q2C

From the previous question, we can see that Tukey has lower power to detect differences, as compared to the unadjusted method. So what is the benefit of using Tukey's method?

Response Tukey helps controll our false discovery rate. This is importnat because at a .05 alpha level, 1/20 experiments is excpected to have a false positive. This becomes more likely as we move from doing one experiment to many. By adjusting with tukey, we keep the p value where it should be accross many experiments. *****

Q2D

Construct a Tukey adjusted CLD display.

```
##   Trt emmean    SE df lower.CL upper.CL .group
##   C     9.27 0.313 45     8.43     10.1  a
##   T3   10.27 0.313 45     9.43     11.1  ab
##   T2   11.02 0.313 45    10.18     11.9   bc
##   T1   12.05 0.313 45    11.21     12.9    c
##   T4   12.24 0.313 45    11.40     13.1    c
##
## Confidence level used: 0.95
## Conf-level adjustment: sidak method for 5 estimates
## P value adjustment: tukey method for comparing a family of 5 estimates
## significance level used: alpha = 0.05
## NOTE: If two or more means share the same grouping symbol,
##       then we cannot show them to be different.
##       But we also did not show them to be the same.
```

Q2E (4 pts)

Calculate the Tukey adjusted 95% ME for pairwise comparisons (ex: $\mu_i - \mu_j$). This is the HSD value. Notes: You must show your work to get full credit for this question. Use echo = TRUE to show your work for this question.

```r
#Q2E

anova_out <- anova(mod1)
MSE <- anova_out$"Mean Sq"[2]
df_error <- anova_out$"Df"[2]
g <- length(levels(dat$Trt))
n <- 10

alpha <- 0.05
q_crit <- qtukey(p = 1 - alpha, nmeans = g, df = df_error)

HSD <- q_crit * sqrt(MSE / n)

HSD
```

```
## [1] 1.259489
```

Q3

Use the emmeans package to calculate Dunnett adjusted pairwise comparisons vs control (Trt = C).

Q3A

Show the Dunnett adjusted comparisons (including estimates and p-values)

```
##  contrast estimate    SE df t.ratio p.value
##  T1 - C       2.78 0.443 45   6.272  <.0001
##  T2 - C       1.75 0.443 45   3.948  0.0010
##  T3 - C       1.00 0.443 45   2.256  0.0930
##  T4 - C       2.97 0.443 45   6.700  <.0001
##
## P value adjustment: mvt method for 4 tests
```

Q3B (4 pts)

Using the result from the previous question, briefly summarize your conclusions (in context) using alpha = 0.05.

Response Treatments 1,2 and 4 have a signifficant difference from controll at an alpha = .05 level

---

Q4 (2 pts per contrast)

Use the emmeans package to estimate and test the following contrasts. For this question, additional information about the treatments is needed:
C = Standard
T1 = Drug therapy with exercise and with counseling
T2 = Drug therapy with exercise but no counseling
T3 = Drug therapy no exercise but with counseling
T4 = Drug therapy no exercise and no counseling

A. Compare the mean for control versus the average of (the means for) the four other treatments.
B. Compare the averages of (the means for) the treatments with exercise versus those without exercise. (Ignore the control.)
C. Compare the averages of (the means for) the treatments with counseling versus the control. (Ignore treatments without counseling.)

```
##  contrast estimate    SE df t.ratio p.value
##  A           -2.12 0.350 45  -6.064  <.0001
##  B            0.28 0.313 45   0.893  0.3764
##  C            1.89 0.384 45   4.924  <.0001
```

Q5

Use the car package to perform a simultaneous test of the provided orthogonal contrasts. Recall that when testing contrasts with the car package, it is easiest to use the no intercept (or cell means) model. Just run the code here.

Q5A

Show the result of the simultaneous test. Note: This should exactly match the one-way ANOVA F-test from Q1.

```
##
## Linear hypothesis test:
## 4 TrtC - TrtT1 - TrtT2 - TrtT3 - TrtT4 = 0
## 3 TrtT1 - TrtT2 - TrtT3 - TrtT4 = 0
## 2 TrtT2 - TrtT3 - TrtT4 = 0
## TrtT3 - TrtT4 = 0
##
## Model 1: restricted model
## Model 2: Loss ~ 0 + Trt
##
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1     49 105.825
## 2     45  44.207  4    61.618 15.681 4.164e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Q5B

Check one pair of contrasts (given in rows) for orthogonality.

```r
#Q5B
c1 <- OrtConMat[1, ]
c2 <- OrtConMat[2, ]
sum(c1 * c2)
```

```
## [1] 0
```

Breakfast Study (Q6 - Q9)

A study was done to examine whether breakfast choice was associated with cholesterol levels in children. A total of n=35 fourth and fifth graders were included in the study. Based on survey response, children were identified as one of (g = 4) four (BKFST) breakfast types: Cereal_F (cereal with fiber), Cereal_O (other cereal), Other_Br (other breakfast) or Skip (no breakfast). The response variable was plasma total cholesterol (TC). The data is available from Canvas as Breakfast.csv.

Notes: - The BMI variable is not used in this analysis. - The sample sizes are unequal because this is an observational study.

Q6

Calculate a table of summary statistics including sample size, mean, sd by BKFST group.

```
## # A tibble: 4 x 4
##   BKFST        n Mean_TC SD_TC
##   <chr>    <int>   <dbl> <dbl>
## 1 Cereal_F    10    163.  8.19
## 2 Cereal_O    12    171.  6.54
## 3 Other_Br     8    172. 10.2
## 4 Skip         5    172.  7.75
```

Q7

Fit an appropriate model (with default contrasts) and show the ANOVA table.

```
## Analysis of Variance Table
##
## Response: TC
##           Df  Sum Sq Mean Sq F value  Pr(>F)
## BKFST      3  531.38 177.128  2.6891 0.06343 .
## Residuals 31 2041.92  65.869
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Q8

Use the emmeans package to calculate the emmeans (estimated marginal means).

Q8A

Show the emmeans table.

```
##  BKFST    emmean   SE df lower.CL upper.CL
##  Cereal_F    163 2.57 31      158      168
##  Cereal_O    171 2.34 31      166      176
##  Other_Br    172 2.87 31      166      178
##  Skip        172 3.63 31      164      179
##
## Confidence level used: 0.95
```

Q8B

Considering the results from the previous question, discuss how sample size effects SE.

Response Increased sample size leads to smaller standard errors

---

Q9

Use the emmeans package to calculate Tukey adjusted pairwise comparisons.

Q9A

Show the Tukey adjusted comparisons (including estimates and p-values).

```
##  contrast            estimate   SE df t.ratio p.value
##  Cereal_F - Cereal_O   -7.892 3.48 31  -2.271  0.1270
##  Cereal_F - Other_Br   -9.287 3.85 31  -2.413  0.0956
##  Cereal_F - Skip       -8.910 4.45 31  -2.004  0.2082
##  Cereal_O - Other_Br   -1.396 3.70 31  -0.377  0.9814
##  Cereal_O - Skip       -1.018 4.32 31  -0.236  0.9953
##  Other_Br - Skip        0.378 4.63 31   0.082  0.9998
##
## P value adjustment: tukey method for comparing a family of 4 estimates
```

Q9B

Considering the results from the previous question, identify the comparison with either the largest OR smallest SE and discuss the sample sizes corresponding to this comparison.

Response Other to skip has teh greatest stander error this makes sense because they have 8 and 5 samples respecitvely which is the samllest two accross all groups *****

Appendix

```
#Retain this code chunk!!!
library(knitr)
knitr::opts_chunk$set(echo = FALSE)
knitr::opts_chunk$set(message = FALSE)
knitr::opts_chunk$set(warning = FALSE)
#Q1

dat <- read.csv("WtLoss.csv")
dat$Trt <- factor(dat$Trt, levels = c("C","T1","T2","T3","T4"))

head(dat)

mod1 <- lm(Loss ~ Trt, data=dat)
anova(mod1)


#
library(emmeans)

emm <- emmeans(mod1, specs = ~ Trt)
tukey_pairs <- pairs(emm, adjust = "tukey")
tukey_pairs


#Q2D
library(emmeans)
library(multcomp)
library(multcompView)

mod1 <- lm(Loss ~ Trt, data = dat)
emm <- emmeans(mod1, specs = ~ Trt)

cld_tukey <- cld(
  emm,
  alpha   = 0.05,
  Letters = "abcdef",
  adjust  = "tukey"
)
```

```
cld_tukey

#Q2E

anova_out <- anova(mod1)
MSE <- anova_out$"Mean Sq"[2]
df_error <- anova_out$"Df"[2]
g <- length(levels(dat$Trt))
n <- 10

alpha <- 0.05
q_crit <- qtukey(p = 1 - alpha, nmeans = g, df = df_error)

HSD <- q_crit * sqrt(MSE / n)

HSD


#Q3A
dunnett_contr <- contrast(emm, method = "trt.vs.ctrl", ref = "C", adjust="mvt")
dunnett_contr

#Q4
my_contrasts <- list(
  "A" = c( 1, -0.25, -0.25, -0.25, -0.25 ),
  "B" = c(0, 0.5, 0.5, -0.5, -0.5),
  "C" = c(-1, 0.5, 0, 0.5, 0)
)

contrast(emm, my_contrasts, adjust="none")


#Q5
OrtConMat <- matrix(
             c(4, -1, -1, -1, -1,
               0, 3, -1, -1, -1,
               0, 0, 2, -1, -1,
               0, 0, 0, 1, -1),
              nrow = 4, byrow = TRUE)
#Q5A
library(car)

mod_cm <- lm(Loss ~ 0 + Trt, data = dat)
linearHypothesis(mod_cm, hypothesis.matrix = OrtConMat)

#Q5B
c1 <- OrtConMat[1, ]
c2 <- OrtConMat[2, ]
sum(c1 * c2)

#Q6
bdat = read.csv("Breakfast.csv")
library(dplyr)
```

```r
bdat %>%
  group_by(BKFST) %>%
  summarise(
    n = n(),
    Mean_TC = mean(TC, na.rm = TRUE),
    SD_TC   = sd(TC, na.rm = TRUE)
  )
#Q7
mod_bkf <- lm(TC ~ BKFST, data = bdat)
anova(mod_bkf)
#Q8A

emm_bkf <- emmeans(mod_bkf, specs = ~ BKFST)
emm_bkf
#Q9A
pairs(emm_bkf, adjust = "tukey")
```