STAA 553: HW2

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 most questions is from Section 03 or earlier.
Unadjusted pairwise comparisons (Q6) is discussed in Section 04.

Weight Loss (Q1 - Q7)

Ott & Longnecker describe a weight loss study with g = 5 treatments (C, T1, T2, T3, T4). Trt C represents a "control" treatment. 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.

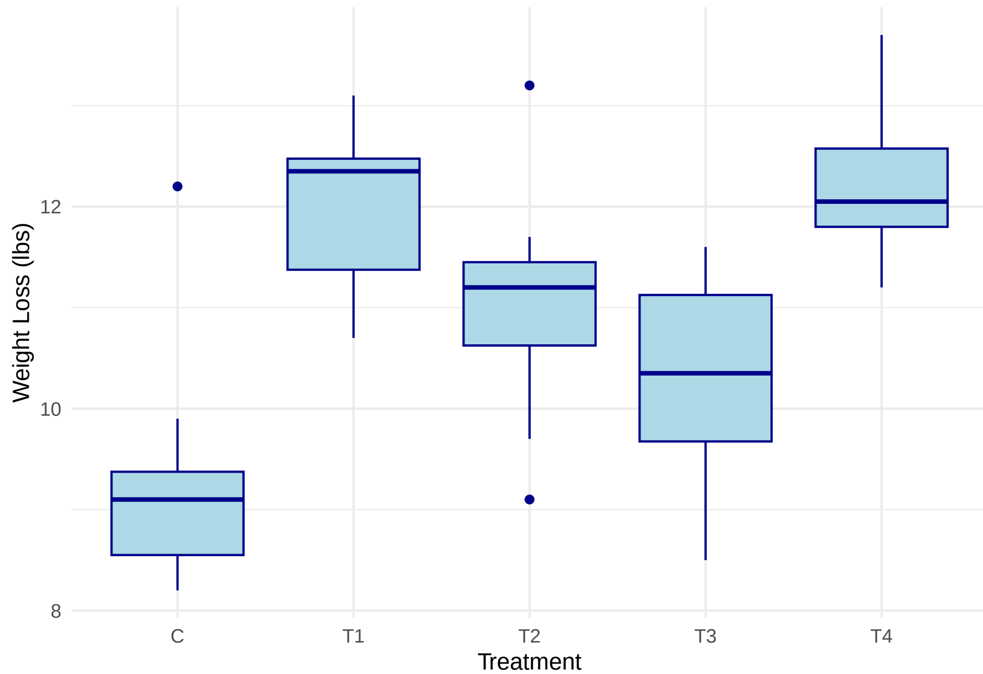
Q1 (3 pts)

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

```
## # A tibble: 5 x 4
##   Trt       n  mean    sd
##   <chr> <int> <dbl> <dbl>
## 1 C        10  9.27 1.16
## 2 T1       10 12.0  0.829
## 3 T2       10 11.0  1.12
## 4 T3       10 10.3  1.03
## 5 T4       10 12.2  0.756
```

Q2

Create an appropriate summary plot of the data.



Q3

Fit an appropriate one-way model with default contrasts.

Q3A

Show the design matrix.

```
##   (Intercept) TrtT1 TrtT2 TrtT3 TrtT4
## 1           1     1     0     0     0
## 2           1     1     0     0     0
## 3           1     1     0     0     0
## 4           1     1     0     0     0
## 5           1     1     0     0     0
## 6           1     1     0     0     0
```

Q3B

Show the coefficient (or parameter) estimates.

```
## (Intercept)       TrtT1       TrtT2       TrtT3       TrtT4
##        9.27        2.78        1.75        1.00        2.97
```

Q3C (5 pts)

Use the coefficient (or parameter) estimates to calculate the predicted mean for each of the treatments. Notes: You must show your work to get full credit for this question. Use echo = TRUE to show your work for this question.

```r
#Q3C

coefs <- coef(mod1)
#C
pred_C <- coefs["(Intercept)"]

#T1
pred_T1 <- coefs["(Intercept)"] + coefs["TrtT1"]

#T2
pred_T2 <- coefs["(Intercept)"] + coefs["TrtT2"]

#T3
pred_T3 <- coefs["(Intercept)"] + coefs["TrtT3"]

#T4
pred_T4 <- coefs["(Intercept)"] + coefs["TrtT4"]

# Combine predictions into a data frame
predicted_means <- data.frame(
  Treatment = c("C", "T1", "T2", "T3", "T4"),
  Predicted_Mean = c(pred_C, pred_T1, pred_T2, pred_T3, pred_T4)
)
predicted_means
```

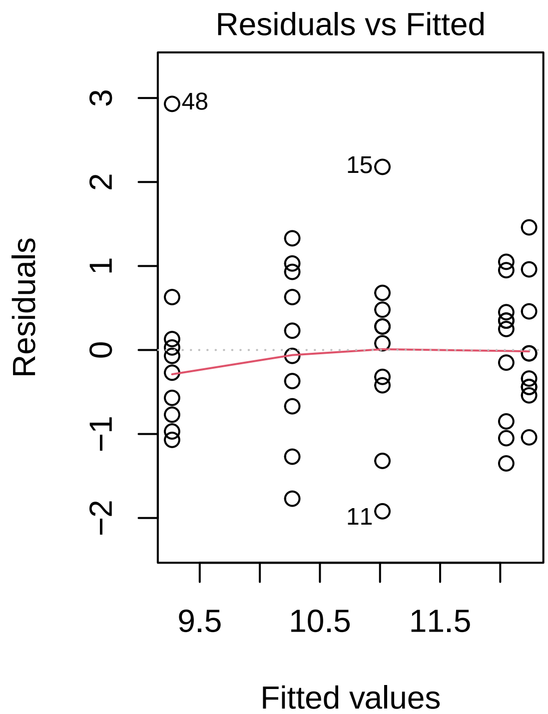
```
##   Treatment Predicted_Mean
## 1         C           9.27
## 2        T1          12.05
## 3        T2          11.02
## 4        T3          10.27
## 5        T4          12.24
```

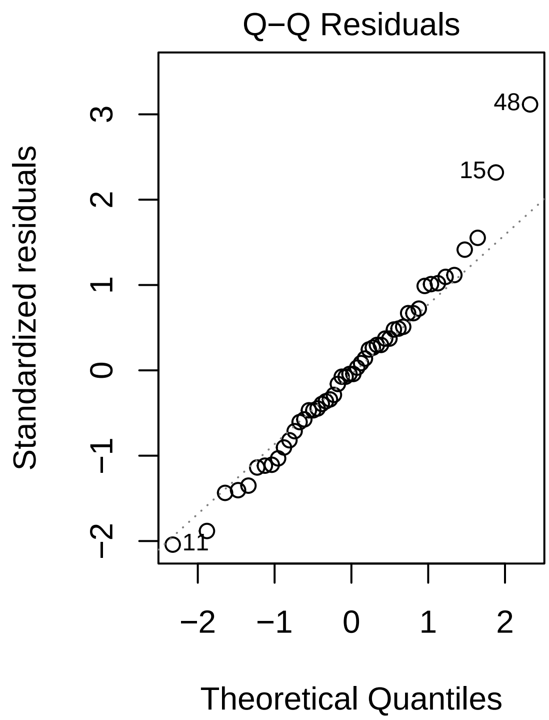
Q4

Now consider the diagnostic plots.

Q4A

Show the plots of Residuals vs Fitted values and QQplot of residuals





Q4B

Using a plot from above, briefly discuss whether the assumption of **equal variance** is satisfied. Note: In your discussion, make it clear what plot you are using to evaluate this assumption.

Response In the Residuals vs Fitted values, I see no clear funnel shape, and a generally random scatter. this makes me comfortable with our equality of variance assumption.

Q4C

Using a plot from above, briefly discuss whether the assumption of **normality** is satisfied. Note: In your discussion, make it clear what plot you are using to evaluate this assumption.

Response Looking at the Q-Q plot, we see geneneral normality. there are deviations at the tails which do raise question, but it is only a few data poitns so I believe we can continue with the normal assumption.

Q5

Provide an appropriate one-way ANOVA table.

Q5A

Show the ANOVA table.

```
## 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
```

Q5B

State the null hypothesis corresponding to the F-test. Be specific.

Response H0: All treatment means are equal.

Q5C

Make a conclusion for the F-test **in context**.

Response We can reject the null hypothesis and conclude that there is evidence that at least one treatment group has a mean weight loss different from the others.

Q6

Use the emmeans package to calculate the following.

Q6A

Show the emmeans (estimated marginal means).

```
##  Trt emmean    SE df lower.CL upper.CL
##  C     9.27 0.313 45     8.64      9.9
##  T1   12.05 0.313 45    11.42     12.7
##  T2   11.02 0.313 45    10.39     11.7
##  T3   10.27 0.313 45     9.64     10.9
##  T4   12.24 0.313 45    11.61     12.9
##
## Confidence level used: 0.95
```

Q6B

Show the unadjusted pairwise comparisons. Hint: Use adjust = "none".

```
##  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  0.0003
##  C - T3      -1.00 0.443 45  -2.256  0.0290
##  C - T4      -2.97 0.443 45  -6.700  <.0001
##  T1 - T2      1.03 0.443 45   2.324  0.0247
##  T1 - T3      1.78 0.443 45   4.016  0.0002
##  T1 - T4     -0.19 0.443 45  -0.429  0.6702
##  T2 - T3      0.75 0.443 45   1.692  0.0976
##  T2 - T4     -1.22 0.443 45  -2.752  0.0085
##  T3 - T4     -1.97 0.443 45  -4.444  0.0001
```

Q6C (4 pts)

Using the result from the previous question, briefly summarize your conclusions **in context** using alpha = 0.05. Note: it may be easier to discuss which comparisons do NOT show evidence of differences.

Response All Comparisons are significant except T1-T4 and T2-T3. we also see that all treatments are signifficantly differen tthan the controll. This indicates that treatment 1 and treatment 4 are signifficantly better than the controll, but about the same as eachother. The same goes for T2 and T3. Additionally, the T1, T4 group appears to outperform the T2 and T3 group.

Q7

Now refit the one-way model using one "alternate" parameterization from Section 03 notes or example. Use echo = TRUE to show your work for this question.
Using the Sum To Zero param ## Q7A

Show the coefficient (or parameter) estimates.

```
#Q7A
options(contrasts = c("contr.sum", "contr.poly"))
mod2 <- aov(Loss ~ Trt, data = wtloss)

coef(mod2)
```

```
## (Intercept)        Trt1        Trt2        Trt3        Trt4
##       10.97       -1.70        1.08        0.05       -0.70
```

Q7B

Use the predict() function to calculate the model based predicted means. Note: These predicted means should (exactly) match the "simple" means from Q1 and the emmeans from Q6A.

```
## # A tibble: 5 x 3
##   Trt       n predicted_mean
##   <chr> <int>          <dbl>
## 1 C        10           9.27
## 2 T1       10          12.0
## 3 T2       10          11.0
## 4 T3       10          10.3
## 5 T4       10          12.2
```

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
library(dplyr)
library(readr)
```

```r
wtloss <- read.csv("WtLoss.csv")

summary_stats <- wtloss %>%
  group_by(Trt) %>%
  summarise(
    n = n(),
    mean = mean(Loss, na.rm = TRUE),
    sd = sd(Loss, na.rm = TRUE)
  )

# Display the summary table
summary_stats


#Q2
library(ggplot2)

ggplot(wtloss, aes(x = Trt, y = Loss)) +
  geom_boxplot(fill = "lightblue", color = "darkblue") +
  labs(x = "Treatment",y = "Weight Loss (lbs)") +
  theme_minimal()

#Q3A
mod1 <- aov(Loss ~ Trt, data = wtloss)

design_matrix <- model.matrix(~ Trt, data = wtloss)
head(design_matrix)

#Q3B
coef(mod1)

#Q3C

coefs <- coef(mod1)
#C
pred_C <- coefs["(Intercept)"]

#T1
pred_T1 <- coefs["(Intercept)"] + coefs["TrtT1"]

#T2
pred_T2 <- coefs["(Intercept)"] + coefs["TrtT2"]

#T3
pred_T3 <- coefs["(Intercept)"] + coefs["TrtT3"]

#T4
pred_T4 <- coefs["(Intercept)"] + coefs["TrtT4"]

# Combine predictions into a data frame
predicted_means <- data.frame(
  Treatment = c("C", "T1", "T2", "T3", "T4"),
  Predicted_Mean = c(pred_C, pred_T1, pred_T2, pred_T3, pred_T4)
```

```r
)
predicted_means
#Q4
par(mfrow = c(1,2))
plot(mod1, which = 1)
plot(mod1, which = 2)

#Q5
anova(mod1)

#Q6A
#install.packages("emmeans")
library(emmeans)
emm <- emmeans(mod1, ~ Trt)
emm

#Q6B
pairs(emm, adjust = "none")

#Q7A
options(contrasts = c("contr.sum", "contr.poly"))
mod2 <- aov(Loss ~ Trt, data = wtloss)

coef(mod2)

#Q7B
newdata <- wtloss
distinct_predictions <- predict(mod2, newdata = newdata)

prediction_df <- data.frame(Trt = newdata$Trt, predicted = distinct_predictions)

summary_stats <- prediction_df %>%
  group_by(Trt) %>%
  summarise(
    n = n(),
    predicted_mean = mean(predicted, na.rm = TRUE),
  )

summary_stats
```