STAT 5000 LAB #8
FALL 2024 DUE TUE, OCT 29TH NAME: _____

Directions: Complete the exercises below. When you are finished, turn in any required files online in Canvas, then check-in with the Lab TA for dismissal.

Two-Factor ANOVA in R

To demonstrate the two-factor study analysis in R, we will explore the minnow larvae example from lecture where the goal is to examine the effects of different concentrations of copper and zinc in water on the ability of minnow larvae to produce protein. The variables for the experiment were:

Factor A: Concentration of copper (0 or 150 ppm)

Factor B: Concentration of zinc (0, 750 or 1500 ppm)

Response Variable: protein content (μg/tank)

This is a full factorial treatment design including six treatments. The six treatments were randomly applied to two water tanks containing minnow larvae each, giving a total of 12 tanks (experimental units). Hence, the experimental design is CRD. The data are given in the table below and in the `minnow.txt` file posted in Canvas.

Copper Conc.	Zinc Concentration		
	0 ppm	750 ppm	1500 ppm
0 ppm	201	173	115
	186	162	124
150 ppm	163	184	114
	182	157	108

The R code for analyzing this data is shown below:

- First, load in the data using the *Import Dataset* tool in R Studio. Be sure to include the column (variable) names and to change the variable type on the copper and zinc columns to "factor" and enter "0, 150" and "0, 750, 1500" as the levels, respectively:

```
library(readr)
minnows <- read_table("minnows.txt",
                      col_names = c("copper", "zinc", "protein"), col_types =
                      cols(copper = col_factor(levels = c("0","150")),
                      zinc = col_factor(levels = c("0", "750", "1500"))))
```

- Next, compute the ANOVA table using the `aov()` function, specifying the model with interactions using the `*` operator, indicating that you want to use the sum-to-zero constraints by setting the `"contr.sum"` option beforehand (you can change this to the baseline constraints using `"contr.treatment"` option instead), and outputting the design matrix using the `model.matrix()` function:

```
options(contrasts = c("contr.sum", "contr.sum"))
sumto0.anova <- aov(protein ~ copper + zinc + copper*zinc, data = minnows)
summary(sumto0.anova)
sumto0.anova$coefficients
model.matrix(sumto0.anova)
```

```
options(contrasts = c("contr.treatment", "contr.treatment"))
baseline.anova <- aov(protein ~ copper + zinc + copper*zinc, data = minnows)
summary(baseline.anova)
baseline.anova$coefficients
model.matrix(baseline.anova)
```

- Alternatively, you can use the `lm()` function to achieve similar results:

```
options(contrasts = c("contr.sum", "contr.sum"))
summary(lm(protein ~ copper + zinc + copper*zinc, data = minnows))
```

- Then, you can use the `interaction.plot()` function to create interaction plots:

```
interaction.plot(minnows$copper, minnows$zinc, minnows$protein,
                 main="Minnows Experiment",
                 xlab="Zinc Concentration",
                 ylab="Mean Protein Level",
                 trace.label="Copper\n Conc.")
interaction.plot(minnows$zinc, minnows$copper, minnows$protein,
                 main="Minnows Experiment",
                 xlab="Copper Concentration",
                 ylab="Mean Protein Level",
                 trace.label="Zinc\n Conc.")
```

- You can use the `emmeans()` function to estimate the simple effects (cell means), and then use the `pairs()` function to test all the pairwise comparisons:

```
library(emmeans)
simple.effects <- emmeans(sumto0.anova, c("copper", "zinc"))
pairs(simple.effects, adjust=NULL)
```

- Similarly, these functions can be used to estimate and test for main effects (marginal means):

```
copper.effects <- emmeans(sumto0.anova, "copper")
pairs(copper.effects, adjust="tukey")

zinc.effects <- emmeans(sumto0.anova, "zinc")
pairs(zinc.effects, adjust="tukey")
```

- Finally, you can diagnose the assumptions by making residual plots and checking for normality:

```
plot(sumto0.anova$fitted.values, sumto0.anova$residuals,
      main="Diagnostic Plot for Minnows Experiment",
      xlab="Fitted Values", ylab="Residuals")
plot(as.numeric(minnows$copper), sumto0.anova$residuals,
     main="Diagnostic Plot for Minnows Experiment",
     xlab="Copper Level", ylab="Residuals")
plot(as.numeric(minnows$zinc), sumto0.anova$residuals,
     main="Diagnostic Plot for Minnows Experiment",
     xlab="Zinc Level", ylab="Residuals")

qqnorm(sumto0.anova$residuals)
qqline(sumto0.anova$residuals, col="red")
```

```
shapiro.test(sumto0.anova$residuals)

library(moments)
mean(sumto0.anova$residuals)
median(sumto0.anova$residuals)
skewness(sumto0.anova$residuals)
kurtosis(sumto0.anova$residuals)-3
```

Assignment

Run the code you created in R Studio for the minnow protein example of a two-factor experiment to complete the following exercises:

1. Using baseline constraints, report the six unique rows in the design matrix $\mathbf{X}$ and the parameter vector β for this two-factor experiment.
2. Using sum-to-zero constraints, report the six unique rows in the design matrix $\mathbf{X}$ and the parameter vector β for this two-factor experiment.
3. Use the estimated parameters from the R output to calculate the six sample treatment means.
4. Use the ANOVA table from the R output to conduct the overall F-test for the effects of the six treatments on the minnow protein production. Interpret the results of the test in the context of the study.
5. Use the ANOVA table from the R output to test for the main effect of copper concentration on the minnow protein production. Assuming there is no significant interaction, interpret the results of the test in the context of the study.
6. Use the ANOVA table from the R output to test for the main effect of zinc concentration on the minnow protein production. Assuming there is no significant interaction, interpret the results of the test in the context of the study.
7. Use the ANOVA table from the R output to test for the interaction effect between copper and zinc concentrations on the minnow protein production. Interpret the results of the test in the context of the study.
8. Study the interaction plot in the R output. Explain why, based on this plot, the interaction is not significant in the model.
9. Use the LSD method to perform all pairwise comparisons of the cell means (simple effects) for the six treatments. Summarize any significant findings in the context of the study.
10. Use Tukey's HSD method to perform pairwise comparisons of the marginal means for the two copper concentrations (main effect of copper) and the marginal means for the three zinc concentrations (main effect of zinc). Summarize any significant findings in the context of the study.
11. Study the plots of the residuals versus: a) predicted value, b) the copper concentrations, and c) the zinc concentrations. Do these plots show any cause for concern?
12. Study the normal probability plot for the residuals. Is there anything of concern in this plot?

Total: 50 points **# correct:** _____ **%:** _____