STAA 553: HW4

YOUR NAME HERE

See Canvas Calendar for due date.
44 points total, 2 points per problem unless otherwise noted.
Add or delete code chunks as needed.
Content for Q1-Q6 is from section 06.
Content for Q7-Q15 is from section 07.

Tomatoes (Q1 - Q5)

A study is being planned to compare a new vs standard fertilizer for tomatoes. Investigators plan to test for a difference in the mean yield (lb/plant) comparing the two treatments. For analysis, they will use a two-sample (and two-sided) t-test assuming equal variance with alpha = 0.05. The experimental units will be individual tomato plants of the same variety and age (randomly assigned to either new or standard fertilizer). Based on a previous study, they conjecture that the within group standard deviation is 1.2. They want to be able to detect a meaningful difference (between means) of 1.5 (lb/plant).

Q1

Find the power with n = 8 per treatment (n = 16 total) for the scenario above.

Power =

Q2

If the sample size was larger (more than 8), would the power be higher or lower than that calculated in Q1? Just answer higher or lower; no need to justify.

Response

Q3

If we use alpha = 0.01 (instead of 0.05), would the power be higher or lower than that calculated in Q1? Just answer higher or lower; no need to justify.

Response

Q4

If the standard deviation was larger (more than 1.2), would the power be higher or lower than that calculated in in Q1? Just answer higher or lower; no need to justify.

Response

Q5

Return to the original scenario and find the sample size required (per treatment) to achieve 80% power. Remember to "round" up to an integer value.

n = XX per treatment.

Apples (Q6)

A study is being planned to compare g = 4 different pesticides for apple trees. The response variable is yield per tree (lb/tree). For analysis, they will use a one-way ANOVA F-test with alpha = 0.05. They conjecture that the within group standard deviation is 10 and μ_1=50, μ_2=50, μ_3=40, μ_4=35.

Q6 (4 pts)

Find the sample size required (number of trees per treatment) to achieve 90% power. Remember to "round" up to an integer value. Hint: Watch out for sd versus variance!

n = XX per treatment

Peppers (Q7 - Q15)

A researcher is interested in how water level (applied during growing season) might affect the "heat" level of jalapeno peppers (measured in SHU = Scoville Heat Units). It is well known that varieties differ in heat levels. They conduct a study using four common jalapeno Varieties (A, B, C, D) and two Water levels (low or high). There are n = 4 individual plants per treatment combination. Hence there are a total of 4 x 2 x 4 = 32 observations. The response variable is SHU (in thousands). The data is available from Canvas as Peppers.csv.

Note: The variable Group is only used in Q15 below.

Q7 (4 pts)

Fit an appropriate model (including Variety and Water main effects plus interaction) and provide the Type 3 ANOVA table. **Important note:** Be sure to use contr.sum to get meaningful type 3 tests.

Q8

Calculate MSResid for this model.

Q9

Create a summary graph (of emmeans) using code similar to what is provided.

Q10 (4 pts)

Use residual diagnostic plots to discuss whether model assumptions are satisfied. You do NOT need to include the plots in your assignment. But for full credit it should be clear which plot is being used to check which assumption.

Response

Q11

Using the ANOVA table from above and considering the "Factorial Principle", should we focus on interaction or main effects? Reference a specific test (and p-value) from the ANOVA table.

Response

Q12

Use emmeans to estimate and test a comparison of high vs low Water, *averaging over Varieties*.

Q13

Considering your answer to the previous question, do we find evidence of a difference in mean response for high vs low Water? Be sure to discuss whether high or low water has the higher mean response.

Response

Q14 Peppers No Interaction Model

We now consider an alternate analysis. A colleague considers your ANOVA table from Q7 and suggests that you drop the interaction from the model.

Q14A

Fit the model with Variety and Water main effects only (no interaction) and provide the Type 3 ANOVA table.

Q14B

Calculate MSResid for this model.

Q14C

Use emmeans to estimate and test a comparison of high vs low Water, averaging over Varieties.

Q14D

By comparing the emmeans output (SE, t.ratio and p-value) for Q12 vs Q14C, we can see that power dropped (slightly) when the interaction term was dropped from the model. Briefly explain why this happened.

Response

Q15 Peppers One-way Model

We consider another alternate analysis. Fit a one-way model using Group as the predictor.

Q15A

Provide the ANOVA table.

Q15B

Use emmeans contrast() to estimate and test a comparison of high vs low Water, *averaging over Varieties*. Notes: You will need to set up emmeans(Model3, ~ Group) before using contrast(). This question is "self-checking" in the sense that these results should exactly match the results from Q12 above.

Q15C

The models used in Q7 and this question are equivalent (in the sense that they are different parameterizations of the same model). State at least one piece of output that confirms the models are equivalent. Note: This can be taken from summary() output or ANOVA table.

Response

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

#Q5

#Q6

#Q7

#Q8

#Q9
library(emmeans)
#emmip(Model1, Water ~ Variety)
#Q10

#Q12

#Q14A

#Q14B

#Q14C

#Q15A

#Q15B
```