STAR 513: HW 3

YOUR NAME HERE

Total points: 40
Questions are worth **2 pts** each, except where noted.
See Canvas calendar for due date.

Homework should be submitted as a pdf, doc or docx file via Canvas.
Use of R markdown HW template is strongly encouraged.
Add or delete code chunks as needed.
Knit frequently to avoid last minute problems!
Your submitted assignment should be neatly formatted and organized.

Ott & Longnecker Example 8.7: It is conjectured that if fields are overgrazed by cattle there will be soil compaction (which could lead to reduced grass). A horticulturist at the agriculture experiment station designed a study to evaluate the conjecture. Three grazing regimens are considered:
- Continuous: continuous grazing
- Rest1week: three-week grazing then one-week no grazing and
- Rest2weeks: two-week grazing then two-weeks no grazing.

A total of 21 similar plots of land are selected for the study. Each of the three grazing regimens are randomly assigned to 7 plots per regimen. After the plots are subject to the grazing regimens for 4 mounts, the the soil density (g/cm3) is measured for each plot.

The data GrazeData.csv is available from Canvas.

Prior to starting the statistical analysis, you will first need to transpose the data to long format using code something like the following. Modify the code as needed. For consistency, the levels of graze should match the bullet list above. Be sure to check the modified data!

Q1 (4 pts)

What is the predictor variable (x)? Is this categorical or numeric?
What is the response variable (y)? Is this categorical or numeric?

The predictor variable (x) is
The response variable (y) is

Q2 (4 pts)

Create a boxplot of the data. Your plot should include axis labels that include the units where appropriate. Briefly comment on at least one thing you learn from this plot.

Response

Q3 (4 pts)

Create a table of summary statistics of SoilDensity by Graze including n, mean and standard deviation. Hint: Use tidyverse group_by() and summarise().

Q4

Fit an appropriate model and include the detailed "coefficients table" in your assignment. This table includes estimates, standard errors, test statistics and p-values. This can be done using tidy() or summary().

Q5 (3 pts)

Calculate the estimated average soil density for each of grazing regimens, using the coefficient (or parameter) estimates from the previous question.
Notes:
(1) You must show your work to get full credit for this question.
(2) Use echo = TRUE to show your work for this question.
(3) Check your own work using simple means from Q3.

```
#Q5
#GrazeContinuous

#Rest1week

#Rest2weeks
```

Q6 (3 pts)

The estimate labeled "Intercept" is $\hat{\beta}_0$. Provide a detailed one-sentence interpretation of what is being estimated in context of this study.

Response

Q7 (4 pts)

The estimate labeled "GrazeRest1week" is $\hat{\beta}_1$. Provide a detailed one-sentence interpretation of what is being estimated in context of this study.

Response

Q8

Use model.matrix() to examine the design or model matrix (but you do not need to include it in your assignment). Provide a brief description of the second column labeled "GrazeRest1week".

Response

Q9

Which grazing regimen does R treat as the reference group? Why is this the reference group?

Response

Q10

Provide an ANOVA table corresponding to the model.

Q11

The ANOVA F-test from the previous question corresponds to a null hypothesis of $H_0 : \mu_1 = \mu_2 = \mu_3$. Provide a brief, statistical conclusion using $\alpha = 0.05$ (ex: Reject H0 or Fail to Reject H0.)

Response

Q12

Now provide a conclusion in context of the research study, limiting statistical jargon. Remember we don't accept H0!

Response

Q13

Use emmeans() to provide the emmeans (estimated marginal means).

Q14

Regardless of any previous results, use pairs(, adjust = "none") to provide the unadjusted pairwise comparisons. Note: I ask for unadjusted pairwise comparisons here for learning purposes. In most cases, we prefer Tukey adjusted pairwise comparisons.

Q15

Which of the pairwise comparisons from Q14 already appeared in the default output (from Q4)? Why do we get one additional comparison in Q14?

Response

Appendix

```
#Retain this code chunk!!!
knitr::opts_chunk$set(echo = FALSE)
knitr::opts_chunk$set(message = FALSE)
library(knitr)
#Import and transpose
library(tidyverse)
GrazeData <- read.csv("GrazeData.csv")
GrazeData <- GrazeData %>%
  pivot_longer(cols = everything(), names_to = "Graze", values_to = "SoilDensity") %>%
  mutate(Graze = as_factor(Graze)) %>%
  arrange(Graze)
#Q2

#Q3

#Q4

#Q5
#GrazeContinuous

#Rest1week

#Rest2weeks


#Q8

#Q10

#Q13

#Q14
```