STAT 441/541 Statistical Methods II

Homework Assignment 5 One-Way ANOVA

Submit a single pdf document to the Dropbox folder *Homework Assignment 5 One-Way ANOVA*

Start the solution for each dataset on a new page.

Dataset 1

Scenario: Etching Rate

An engineer is interested in investigating the relationship between the RF power setting and the etch rate of a tool for removing unwanted material from integrated circuits. Four treatments consisting of different power settings are used. Etch rate data (angstroms/minute) from the plasma etching experiment are given in the following data table:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Power (watts)** | **Etch Rate (angstroms/minute)** | | | | |
| **160** | 575 | 542 | 530 | 539 | 570 |
| **180** | 565 | 593 | 590 | 579 | 610 |
| **200** | 600 | 651 | 610 | 637 | 629 |
| **220** | 725 | 700 | 715 | 685 | 710 |

The R code file: Etching Rate R Code.R

NOTE: Convert the tabular data to a statistical dataset and save as Etching Rate.xlsx. Have two variables with names Power and EtchRate. Denote the four treatments for Power as: 160, 180, 200, and 220.

(a) Paste a screenshot of your dataset and explain how you converted the tabular data to a statistical dataset:

(b) Paste and comment on the boxplot of Observed Etching Rates:

(c) Paste and comment on the table of summary statistics for each RF Power Level:

(d) For this scenario, state the means model, describe all terms in the model, and give values for all subscripts:

(e) Do the power settings appear to yield different etching rates for removing unwanted material? Justify your answer using and use one-way ANOVA to test treatment effects. Paste the Analysis of Variance Table and perform the hypothesis test:

Hypotheses:

Test Statistic:

P-value:

Decision about the null hypothesis:

Conclusion:

(f) Check conditions for ANOVA based on the following assumptions:

Errors are normally distributed

There are no outliers

The variance of the errors, , is the same for all treatments (common variance)

(i) Paste and interpret the Residuals vs Fitted plot from R output:

(ii) Paste and interpret the Normal Q-Q plot from R output:

(iii) Paste and interpret the Scale-Location plot from R output:

(iv) Paste and interpret the Constant Leverage: Residuals vs Factor Levels plot from R output:

(v) Paste and interpret the histogram of residuals from R output:

(vi) Paste and interpret the boxplot of residuals from R output:

(vii) Paste the Shapiro-Wilk test from R output and perform a hypothesis test at a significance level of :

Hypotheses:

Test Statistic:

P-value:

Decision about the null hypothesis:

Conclusion:

(viii) Paste the Levene test from R output and perform a hypothesis test at a significance level of :

Hypotheses:

Test Statistic:

P-value:

Decision about the null hypothesis:

Conclusion:

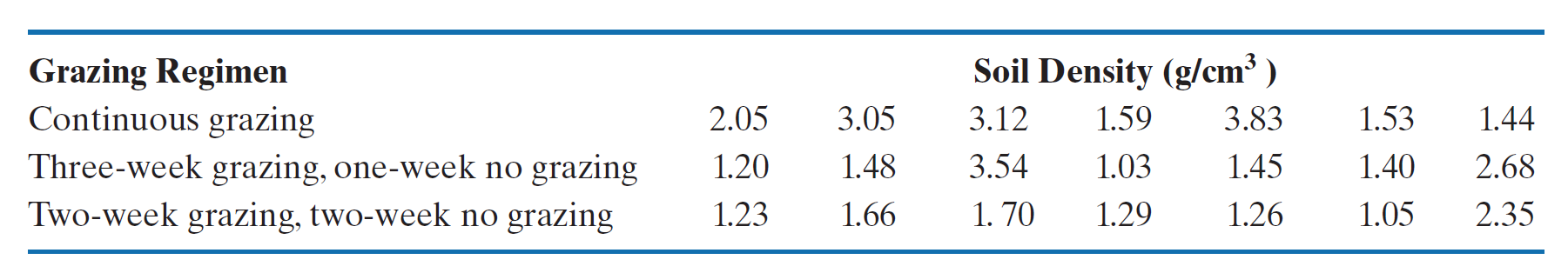
(ix) Have the assumptions been met? Justify your answer.

(g) What is your overall conclusion about this ANOVA analysis?

Dataset 2

Scenario: Soil Compaction

It is conjectured that when fields are overgrazed by cattle there will be a substantial reduction in the available grass during the subsequent grazing season due to the compaction of the soil. A horticulturist at the state agricultural experiment station designs a study to evaluate the conjecture. Twenty-one plots of land of nearly the same soil texture and suitable for grazing are selected for the study. Three grazing regimens selected for evaluation are randomly assigned to 7 plots each. The three grazing regimens are Continuous grazing, Three-week grazing and then one-week no grazing, and Two-week grazing and two-week no grazing. After the 21 plots are subjected to the grazing regimens for four months, the researcher randomly selects 10 soil cores from each plot and measures the bulk density (g/cm3) in each soil core. The mean soil density of the 10 cores from each plot is given in the following table:



The R code file: Soil Compaction R Code.R

NOTE: Convert the tabular data to a statistical dataset and save as Soil Compaction.xlsx. Have two variables with names Grazing and Density. Denote the three treatments for Grazing Regimen as: Continuous, 3wk\_1wk, and 2wk\_2wk.

(a) Paste a screenshot of your dataset and explain how you converted the tabular data to a statistical dataset:

(b) Paste and comment on the boxplot of Observed Soil Density:

(c) Paste and comment on the table of summary statistics for each Cattle Grazing Regimen:

(d) For this scenario, state the means model, describe all terms in the model, and give values for all subscripts:

(e) Do the grazing regimens appear to yield different degrees of effect on the amount of compacting in the soil? Justify your answer using and use one-way ANOVA to test treatment effects. Paste the Analysis of Variance Table and perform the hypothesis test:

Hypotheses:

Test Statistic:

P-value:

Decision about the null hypothesis:

Conclusion:

(f) Check conditions for ANOVA based on the following assumptions:

Errors are normally distributed

There are no outliers

The variance of the errors, , is the same for all treatments (common variance)

(i) Paste and interpret the Residuals vs Fitted plot from R output:

(ii) Paste and interpret the Normal Q-Q plot from R output:

(iii) Paste and interpret the Scale-Location plot from R output:

(iv) Paste and interpret the Constant Leverage: Residuals vs Factor Levels plot from R output:

(v) Paste and interpret the histogram of residuals from R output:

(vi) Paste and interpret the boxplot of residuals from R output:

(vii) Paste the Shapiro-Wilk test from R output and perform a hypothesis test at a significance level of :

Hypotheses:

Test Statistic:

P-value:

Decision about the null hypothesis:

Conclusion:

(viii) Paste the Levene test from R output and perform a hypothesis test at a significance level of :

Hypotheses:

Test Statistic:

P-value:

Decision about the null hypothesis:

Conclusion:

(ix) Have the assumptions been met? Justify your answer.

(g) What is your overall conclusion about this ANOVA analysis?