Homework assignment 08

Use black text (if possible) for everything you include in this document. Keep both your answers and the original questions. Save this document in PDF format and submit it on Canvas. Please follow the general requirements described in the grading rubric.

1. Show a documentation header.

2. Download the file data-08-bacterial-soap.txt and review the documentation associated with this file. Prepare a data dictionary from the information provided in that documentation. If some of the documentation is incomplete (for example, no units of measurement specified), note this in your data dictionary.

3. Import the data into SPSS. Calculate the means and standard deviations of bacterial counts for each group. Draw boxplots for bacterial counts by group.

4. Do you think there are problems with meeting the equal variances assumption? Explain why or why not.

5. Run a oneway ANOVA model for this data. Display an analysis of variance table with only three rows properly labeled as between, within, and total. Use at least four significant digits for the sums of squares (SS) and mean squares (MS) but round the F-ratio and p-value to two significant digits. This will require you to re-type the analysis of variance table rather than cutting and pasting an image from the SPSS output.

6. Write the formal null and alternative hypotheses for this model. Would you accept or reject the null hypothesis? Explain.

7. Compute the Tukey post hoc comparison of pairwise differences. Display them below.

8. Interpret the results of the hypothesis test and Tukey post hoc comparison.

9. Compute the predicted values and residuals from the oneway ANOVA model. Display the first ten rows below.

10. Draw a QQ plot of the residuals and display it below.

11. Interpret the QQ plot.

6.

The following is the assignment from last year. I am keeping it around until I finalize this year’s assignment.

The data set used for this Homework assignment contains medical information from a study that assessed Cholesterol Level for persons obtaining labs associated with an annual screening visit to their primary care giver.

A medical student on rotation was interested in assessing the relationship between cholesterol level and weight. They wanted to know if those people in a heavier weight group have higher a mean total cholesterol level as compared to other weight groups.

Open the data set *HW - One-Way ANOVA – 2021.* Note that there are several variables in this data set. You must select the correct Independent (Factor) Variable and the correct Dependent Variable based upon the scenario and hints above.

Complete the following:

1. **Identify the Factor in this data set.** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
2. **How many levels of the factor are there?** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
3. **What are the levels (name them)** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
4. Run all tests to assess Assumptions.
   1. Complete the Assumptions table below. Include ALL assumptions (Add rows as needed)
   2. If assumptions are not met, state why and then proceed with the analysis regardless.
   3. **DO NOT CHANGE THE DATA BEFORE RUNNING THE ANALYSIS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Assumption** | **Method used to test assumption** | **Result of method** | **Decision (Met or did not meet assumption)** |

**Complete this Table after answering questions 5-9 below and if necessary, after the analysis.**

**(Questions 5 – 10 address the various methods for determining normality)**

1. **Calculate the Skewness and Kurtosis for each level of the factor.** Show the equation with the numbers that you used. Type your answers in the table below.
   1. **Is there evidence of Skewness?** Yes No
   2. **Is there evidence of Kurtosis?** Yes No
2. **Obtain the mean and median for each level of the factor. (put your values in the following table):** (Add rows as needed)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Weight Group | Mean | Median | Skewness | Kurtosis | Normally Distributed? |

1. Determine normality based on the mean, median, skewness and kurtosis.
   1. **What group(s) APPEAR to be normally distributed?** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
   2. **What group(s) APPEAR to not be normally distributed?**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
2. Determine Normality based on the Shapiro Wilks Test, Histograms and P-P plots. Type your answers in the table below. (Add rows)

|  |  |  |  |
| --- | --- | --- | --- |
| Weight Group | Shapiro-Wilkes test Significance (p value) | P-P Interpretation  (Normal or Not Normal) | Histogram Interpretation  (Normal or Not Normal) |

* 1. **What group(s) APPEAR to be normally distributed?** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
  2. **What group(s) APPEAR to be not normally distributed?**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Make a final decision about the Assumption of Normality using all evidence from above in Questions 5 – 8.
   1. Enter your decision in the table below for each group using all evidence. (Add rows)

|  |  |
| --- | --- |
| Weight Group | Decision – Normally Distributed or Not Normally Distributed |

1. In a “real-world analysis” Would you analyze this data with ANOVA? Yes No

**(Questions 11 – 15 address the analysis)**

1. Run the analysis:
   1. The Output that you will obtain includes: (There are other tables and graphs that we will not assess)
      1. Descriptives Table
      2. Test of Homogeneity of Variances Table
      3. ANOVA Table
      4. Robust Test of Equality of Means Table
      5. Multiple Comparisons Table
      6. Means Plot graph
2. Interpret the table for Test of Homogeneity of Variances.
   * 1. **What was the test statistic value?** \_\_\_\_\_\_\_\_\_\_
     2. **What was the p-value?** \_\_\_\_\_\_\_\_\_\_\_
     3. **Was the data homogenous in terms of variance?** YES NO
3. Interpret the ANOVA Table.
   1. **The Between groups df was** \_\_\_\_\_\_\_. **How was this calculated?** Write the equation.
   2. **The Within groups df was** \_\_\_\_\_\_\_. **How was this calculated?** Write the equation.
   3. **The Total df was** \_\_\_\_\_\_\_. **How was this calculated?** Write the equation
   4. **What was the F-ratio obtained in this analysis?** \_\_\_\_\_\_\_\_
      1. **Using information from the ANOVA table, how was this calculated?** (Write the equation with the correct numbers.)\_\_\_\_\_\_\_\_\_\_\_\_
   5. **Was the F-ratio significant?** YES NO
   6. **What was the exact p-value?** \_\_\_\_\_\_\_\_\_\_\_
   7. **Write the decision about the null hypothesis regarding this analysis**
4. Interpret the Multiple Comparisons Table(s).
   1. **Which groups were found to be different?**
      1. Using Tukey \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
      2. Using LSD \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
      3. Using Scheffe’\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
   2. **Which of these three tests is the least powerful test?** \_\_\_\_\_\_\_\_\_\_\_\_\_
   3. **Which of these three tests is the more powerful test? \_\_\_\_\_\_\_\_\_\_\_\_\_**
5. Hand Calculate the Effect Size for the main ANOVA analysis. Show your work.
6. **Create the final write up**. Be complete.

**\*\*\*All pertinent tables and graphs for testing Assumptions should be copied and pasted below here.**

**\*\*\*All pertinent tables and graphs for the ANOVA and Post-hoc analysis should be copied and pasted below here.**