Assignment 3
STAT 705, Spring 2020

This is a graded assignment, worth 25 points.
The data for this homework assignment is in the file "Homework 3 data.txt".

General information

Bacteria are cultured in medical laboratories to identify them so patients can be treated correctly. The dataset contains measurements of bacteria counts that were generated by culturing the bacterium Staphylococcus aureus. (This is the bacterium that can cause a persistent and hard-to-treat "staph" infection.) Each culture was incubated for 24 hours at 27°C, and contain varying percentages of tryptone, a nutrient designed to enhance bacteria growth. Each percentage of tryptone concentration was cultured (i.e., replicated) an equal number of times. The response variable is Count, which is the number of colony forming units in the millions.

As with the earlier homework assignments, you will write the SAS code to answer the following questions and then enter your answers directly into Canvas. In addition to answering these questions, you will need to to upload your SAS program file. This is the file with the "sas" extension, and it contains the statements that appear on the Code tab in SAS Studio. Please do <u>not</u> upload your SAS output file, the SAS log file or the HTML file. I need your original program file so that I can execute it and generate your output for myself.

You can use the Discussion area in Canvas to communicate with your classmates about this assignment, but you must write your own SAS code and answer the questions in your own words. Zero points are given for duplicate answers or SAS programs. Please email your instructor if you have any questions.

The questions are on the next page.

Perform analysis of variance on these data, and answer the following questions.

Question 1.   Does the assumption of normal errors appear to be violated? (yes or no) Describe what you see in the SAS output that enables you to answer this question.

Question 2.   Does the assumption of equal variances appear to be violated? (yes or no) Describe what you see in the SAS output that enables you to answer this question. If there is an official hypothesis test to help you answer this question, provide the p-value of the test.

Question 3.   How many different concentrations of tryptone are in the data?

Question 4.   Which concentration is the reference level?

Question 5.   How many indicator variables is SAS using in this analysis?

Question 6.   How many treatments are in this experiment?

Question 7.   What is the number of replications in this experiment?

Question 8.   Do all levels of concentration have the same mean count? Provide the p-value of the appropriate test.

Question 9.   Is there a linear trend? Provide the p-value of the appropriate test.

Question 10. Is there a quadratic trend? Provide the p-value of the appropriate test.

Question 11. Is the mean count for concentration 1.2 significantly different that the average count for all the other concentrations? (yes or no) Provide the p-value of the appropriate test.

Question 12. Suppose that the researchers ask you to determine which concentration they should use in order to produce the largest amount of bacteria. What do you recommend?

Question 13. Perform a regression analysis on these data, using Count as the response and Concentration as the predictor. According to the regression equation, how much of an increase (in millions of bacteria count) should we expect to see if the concentration increases by 1 percent?

Question 14. If you were going to summarize the results of this dataset, would you use the results from the analysis of variance or would you use the results from the regression?

Question 15. Briefly explain why you chose your answer in the previous question.