STAT 50 Group Project 2, Spring 2025

Instructions. This project assignment consists of two problems and a bonus problem based on the Excel data set posted on Canvas (AmericanRiver-Sp2024.xlsx) which involve temperature, pressure and oxygen saturation measurements from American River and tap water samples (courtesy of Dr. Deboleena Roy of American River College. The assignment is due by Wednesday night on December 7th. You are allowed to work in pairs or groups of size three, and submit the assignment jointly. In that case, one group member can submit the completed document with the names of the collaborators on the first page (or in the comments section on Canvas). You may also choose to submit the relevant hand-written or typed project documents in person or upload them on Canvas individually by stating names of your collaborators, if any. Please don't upload Excel or redundant data files via Canvas Assignments platform as the graphs on Excel files don't show properly when uploaded via Canvas Assignments.

1. Combine the dissolved oxygen (DO) measurements (in mg/L) from American River samples (column D) in the Excel sheet (so that there are 42 observations in total). Then, do the following:

   (a) (1.5 pts) Obtain a histogram, boxplot and normality plot of the measurements. Describe the shape of the data distribution and potential outliers. Does it appear to come from an approximate normal population? Why or why not? You may also consider applying *kstest* function of Matlab (to a standardized version of the sample data).

   (b) (1.5 pts) Determine the sample mean, median and standard deviation of the data in part (a). Does a z-interval procedure apply to this data set to estimate the population mean for dissolved oxygen amount in the river? Does a t-interval method apply? Explain.

2. Combine the dissolved oxygen measurements (in mg/L) from tap water samples (column I) in the Excel sheet. Then, do the following:

   (a) (1.5 pts) Repeat 1(a) for this data set.

   (b) (1.5 pts) Explain why an approximate t-interval procedure applies to this data set. Then, determine a lower bound confidence interval for the mean dissolved oxygen in tap water at 95% level of confidence.

3. (1 pt) Do one of the following for extra credit:

   (a) In order to explain the relationship between oxygen saturation (OD) and temperature, construct scatterplots of these data sets in the Excel file (by ignoring missing data pairs). You may also check sample correlation between the two data sets. Then, briefly state how this information can be used to explain (or adjust) some extreme OD measurements from the river.

   (b) Interpret the confidence interval you obtained in part 2 (b) in context. Based on the interval, can you reasonably conclude that the average dissolved oxygen concentration in tap water is above 7 mg/L? Justify your answer.

   (c) Are the pressure measurements related to the temperature or OD measurements in the Excel file (river or tap water)? Use scatterplots, correlations etc. to support your argument.