STAT 210
Applied Statistics and Data Analysis:
Homework 4

Due on Oct. 12/2025

You cannot use artificial intelligence tools to solve this homework.
Show complete solutions to get full credit. Writing code is not enough to answer a question. Your comments are more important than the code. Do not write comments in chunks. Label your graphs appropriately

Question 1

A car manufacturing company has developed a new exhaust system designed to reduce CO2 emissions. To evaluate its effectiveness, the company conducted an experiment using 25 identical cars, all driven under similar conditions in city traffic.

With the standard exhaust system, the average CO2 emissions were measured at 191.5 g/km. The emissions data collected using the new exhaust system are stored in the variable `emss1` within the `25Fhw4Q1` file.

Use a significance level of $\alpha = 0.01$ for all statistical tests in this question.

(a) Perform an exploratory analysis to assess whether the assumption of normality is reasonable for this dataset. You are required to produce two plots: First, combine a histogram, a graph of the estimated density and a curve for the normal density with parameters estimated from the sample. Use adequate names for the axes and add a legend. Second, do a quantile plot using the function `qqPlot` in the `car` library with the argument `line` set to `r`. What does this option do? Comment on what you observe in the plots. Do you think that the assumption of normality for the data is valid?

(b) Assuming that the variable `emss1` follows a Gaussian distribution, write down a formula for the lower one-sided confidence interval for the mean at level $(100 - \alpha)\%$ (This interval is bounded above but unbounded below). Calculate this confidence interval for the mean for the case $\alpha = 0.01$, check whether the reference value of 191.5 falls inside or outside and give an interpretation.

(c) What parametric test would be adequate for testing whether the new exhaust system decreases CO2 emissions for the car? State clearly what hypotheses you are testing and which assumptions are needed for the test. Explain why you think they are satisfied. Give a formula for the test statistic and calculate its value. Describe the sampling distribution and explicitly identify the type I and type II errors. Carry out this test and discuss the results.

(d) What non-parametric tests will be adequate for the problem in 1(c)? What assumptions are needed, and why do you think they are satisfied? Perform this test, discuss the results, and compare them with your previous results.

Question 2

A parmaceutical company wants to test a new weight reduction drug on rats. The experiment uses 15 rats which were weighted before and one week after receiving the drug. The results are stored in the file `25Fhw4Q2`. The column `wt1` holds the measurements for the weight at the beginning of the experiment and `wt2` has the values at the end.

(a) Use graphical tools to compare the weights before and after taking the drug and comment on what you observe.

(b) You want to test whether the drug effectively reduces weight using this data. What hypotheses do you want to test? What parametric test or tests could be appropriate here? What are the assumptions? Why do you think they are satisfied in this case? Identify the type I and type II errors. Carry out the test(s) and discuss your results.

(c) The pharmaceutical company is only interested in this drug if the reduction in weight for the rats is more than 2 g. To simplify the test, suppose you want to test that the reduction in weight is 2 g versus the alternative that it is more. How would you carry out this test with the data that you have? What assumptions are needed? Are they justified in this case? Carry out the test or tests and comment on your results.

(d) What non-parametric tests will be adequate for the problems in 2(b) and 2(c)? What assumptions are needed, and why do you think they are satisfied? Perform these tests, discuss the results, and compare them with your previous results.