Stat 311 R Assignment 5

Use `rmarkdown` to create HTML file output that you then convert to pdf. Your final product will be a pdf file uploaded to Canvas.

The R5 Tutorial introduces all the functions you will need to complete this assignment. Create a new .Rmd file for this assignment. To get you started, copy and paste the title section and setup chunk from the tutorial. Then copy, paste and edit code from the tutorial as needed for the assignment. Remember to include section and subsection headers as needed to make your assignment readable. Only include the output needed to answer the questions.

This assignment requires the patient diet data set. A data description document is included.

1. Do a quick EDA of the adherence time and weight loss variables. Describe the distributions of each variable, including the existence of outliers, if any.
2. Create a scatterplot of weight loss (y) on adherence time (x). Describe the relationship. Identify any potential outliers that could be influential in a linear regression, if any.
3. Run a linear regression of weight loss on adherence time. Do not use any transformations or subsets of data.
    a. Write out the regression equation.
    b. Interpret the slope parameter in the context of the problem.
    c. What is the standard error for regression? Interpret this value.
    d. How much variability in weight loss can be explained by adherence time?
4. Conduct linear regression model diagnostics. What are your thoughts on meeting the assumptions for inference for linear regression?
5. Assuming assumptions are met (regardless of what you find in (4) above),
    a. Conduct a hypothesis test for the claim that the population slope of weight loss on adherence is greater than zero. Make sure to write out the hypotheses and all pertinent numbers in your summary, including your decision and interpretation in the context of the problem.
    b. Find the 99% confidence interval for mean weight loss when adherence is 5.75 (median adherence). Interpret the interval in the context of the problem.
    c. Find the 99% prediction interval for the weight of a new, randomly chosen person when adherence is 5.75. Interpret the interval in the context of the problem.
6. Create a subset of the data that removes observation 37 (adherence level 9.25). Rerun the regression with this subset. Do you think observation 37 is influential?