STAT 6021: Linear Models for Data Science

Homework problem set for learning unit 2.

Textbook problems: 2.1 (a,b,c), 2.4 (a, b), 2.7 (a,b,d), 2.12 (a,b,c), 2.19 (see below). Additional problems are listed below.

Comments and additional instructions:

General comment: On this assignment, use R to help you as much as you can. Turn in your a paper copy of your solutions in class, which is to include the key pieces of your work that led to the answer, any explanations, and clearly labeled final answers. Submit electronic copies of the R code that led to your solution to the course web page in Collab, using the "Assignments" tool. Carefully annotate your R code to help the grader connect the sections of your code to the paper copy of your solutions. Do not expect that the grader will open your R code to find your answer or any key pieces of your work; the grader would only look at the R code to assign partial credit, if necessary.

Problems 2.1a, 2.4a, 2.7a, 2.12a: When reporting the results of fitting a model, write down the equation of the fitted regression model, filling in the numerical values of the least-squares estimates of the regression coefficients.

Problem 2.1b, 2.4b: When reporting an ANOVA table, put the table in the format of Table 2.4 in the textbook, filling in all of the quantities listed in that table.

Problem 2.1b, 2.4b, 2.7b, 2.12b: To test for significance of regression, use the test based on the ANOVA F statistic, F_0.

Problem 2.1b, 2.4b, 2.7b, 2.12b, 2.12c: When reporting the results of a hypothesis testing procedure, state the null and alternative hypotheses, the observed value of the test statistic, the decision rule in terms of critical values for the test statistic, and your decision regarding the null hypothesis; supplement this information by additionally reporting a P-value.

Problem 2.1b, 2.7d: When reporting an interval estimate, state the confidence level, the point estimate of the quantity being estimated, and the upper and lower bounds of the interval. If it is appropriate, also state the margin of error associated with the interval in the form "pt. est" $\pm$ "margin of error."

Problem 2.19: To complete this problem fit both models indicated in the instructions, one of satisfaction by severity and the other of satisfaction by age. **Part A:** Write down the equation of each fitted regression model. **Part B:** For each model create a scatter diagram and draw in the fitted regression line. **Part C:** For each model calculate the sum-of-squares statistics SS_R, SS_{Res}, and SS_T. **Part D:** Under which model does it appear that the residuals fall more tightly around the regression line. **Part E:** In which of the statistics that you calculated in **Part C** is your observation in **Part D** reflected. Explain.

Additional problems:

Problem 6: In the context of **Textbook Problem 2.1**, find a 95% confidence interval for the error variance, σ^2. State the confidence level, the point estimate of σ^2, and the upper and lower bounds of the interval.