

Homework 5

STAT 462 (Fall 2020)

UNAUTHORIZED DISTRIBUTION AND/OR UPLOADING OF THIS DOCUMENT IS STRICTLY PROHIBITED.

Clearly label your answers to each question and each sub-question. Your answers MUST be uploaded to Canvas as a <HWx_Yourfirstname.nb.html> file by the deadline.

- (1) . Many different interest groups such as the lumber industry, ecologists, and foresters benefit from being able to predict the volume of a tree just by knowing its diameter. One classic data set (*shortleaf.txt*) reported by C. Bruce and F. X. Schumacher in 1935 concerned the diameter (in inches) and volume (in cubic feet) of 70 shortleaf pines.

A researcher is interested in learning about the relationship between the diameter and volume of shortleaf pines.

- (a). Identify the response variable and explanatory variable for the problem
- (b). Draw a scatter plot to show how volume of a tree and its diameter are associated. Comment on your observations.
- (c). Fit a regression line for the problem, write down the estimated equation (define any terms you might have used), and mark the estimated line on the scatter plot in part (b). Provide all outputs. Interpret the estimated parameters clearly in the context of the problem.
- (d). Obtain the diagnostics for the fitted model in part (c). Clearly state your observations. Provide all the outputs you used.
- (e). Identify (i) the point with highest residual (studentized residual), (ii) the point with highest leverage, and (iii) the point with highest Cook's distance. Suppose a friend of the researcher suggested that there is an influential point in the data, and should be investigated. Do you agree with this comment? Explain your reasoning.

```
#identifying the point with highest residual you can use a code line similar to below.  
#shortleafData is the name of my data object  
#I have previously saved the studentized residuals into an object named stud.residuals  
#abs() function gives the absolute value of argument inside the function and  
#max() function gives the maximum of the argument passed inside the function  
#== is a logical function that checks left-hand side is equal to right-hand side  
shortleafData[abs(stud.residuals)==max(abs(stud.residuals)),]
```

- (2). Manufacturer of a laundry detergent was interested in testing a new product prior to market release. One concern is the relationship between height of the detergent suds in a washing machine as a function of the amount of detergent added to the washing cycle. In a standard size washing machine with water at full level, the manufacturer makes random assignments of amounts of detergent tested them on the washing machine. Resulting data is in *SoapSuds.csv* file.

- (a). Recognize the response and the explanatory variables in this study.
- (b). Make a scatterplot and comment on your observations.
- (c). Obtain the least squares regression line and overlay the line on the scatterplot.

- (d). What is the equation of the least squares regression line.
- (e). Do a lack of fit test for the fitted model. Clearly state all the steps, outputs and conclusion.
- (f). Obtain the diagnostics for the fitted model. Clearly state your observations. Provide all the outputs you used.