Homework 7

STAT 462 (Fall 2020)

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

The cotton aphid is pale to dark green in cool seasons and yellow in hot, dry summers. Generally distributed throughout temperate, subtropic, and tropic zones, the cotton aphid occurs in all cotton-producing areas of the world. These insects congregate on lower leaf surfaces and on terminal buds, extracting plant sap.

If weather is cool during the spring, populations of natural enemies will be slow in building up, and heavy infestations of aphids may result. When this occurs, leaves begin to curl and pucker; seedling plants become stunted and may die. Most aphid damage is of this type. If honeydew resulting from late-season aphid infestations falls onto open cotton, it can act as a growing medium for sooty mold. Cotton stained by this black fungus is reduced in quality and brings a low price for the grower.

Entomologists studied the aphids to determine weather conditions that may result in increased aphid density on cotton plants. The data were reported in Statistics and Data Analysis (*Peck, Olson, and Devore, 2005*) and come from an extensive study as reported in the article "*Estimation of the Economic Threshold of Infestation for Cotton Aphid*" [Mesopotamia Journal of Agriculture (1982): 10, 71–75].

The data are given in the file *aphideData.csv*. In the data file,

y = infestation rate (aphids/100 leaves)

x_1 = mean temperature (°C)

x_2 = mean relative humidity

---

(1). Draw a scatter plot matrix and comment on your observations

(2). Obtain a correlation plot for your data and comment on your observations.

(3). Fit the first order full model for this data (here after will be referred to as model 1). Discuss the validity of the models assumptions. Provide R outputs for all the tools you may have used.

(4). Write down the estimated first order full model, and interpret the parameters in the context of the problem. Provide any R outputs you might have used.

(5). Ignore the validity of the model assumptions, and test the overall fit of model at 5% level. Clearly write down the all the steps. Provide any R outputs you might have used.

(6). What is the proportion of variability explained by model 1.

(7). Test the partial slope of x_2 at 5% level. Clearly write down the all the steps. Provide any R outputs you might have used.

(8). Test whether x_1 has any additional predictive power above contributed x_2. Clearly write down the all the steps. Provide any R outputs you might have used.

(9). Based on your conclusion for part (8), what would be your next step.

(10). Run a SLR model with x_2 as the predictor. Hereafter refer to this model, as model 2. Write down the estimated SLR model.

(11). Compare model 1 with model 2 at 5% significant level. Clearly write down the all the steps. Provide any R outputs you might have used.

(12). Based on your comparison at part (11), which model (model 1 or model 2) would you choose for this data. Explain reasons for your choice.

(13). Run a SLR model with x_1 as the predictor. Hereafter refer to this model, as model 3. Based on the R^2 values for models 2 & 3, which model would you choose? Explain your reasons.

(14). Would you recommend using model 2 to predict the average infestation rate when $x_2 = 115$? Explain your answer.

(15). Using model 2, find an interval estimate of the infestation rate for a cotton plant with mean relative humidity of 65.