STAT 3010: Assignment 5

Spring 2024: Due Mar. 31, 2024

Please answer each question as completely as possible. Solutions must be typed and submitted to Canvas by 11:59pm CST on the due date.

1. Efficient design of certain types of municipal waste incinerators requires that information about energy content of the waste be available. Data are given on y = energy content (kcal/kg); the three physical composition variables x1 = % plastics by weight, x2 = % paper by weight, and x3 = % garbage by weight; and the proximate analysis variable x4 = % moisture by weight for waste specimens obtained from a certain region.

   ```
   energy <- read.csv("hw5q1.csv", header = TRUE)
   ```

   (a) Fit a regression function with the four aforementioned variables as predictors of energy. Provide a summary of the result and comment on the direction of the relationships of the predictors and energy.

(b) Predict the value of energy content when plastics is 17.03, paper is 23.46, garbage is 32.45, and water is 53.23. Also determine the corresponding residual. (Hint: This is observation # 11 in the dataset)

(c) What proportion of observed variation in energy content can be explained by the approximate relationship between energy content and the four predictors?

(d) Perform stepwise regression to select variables. Summarize the selected model.

2. Historically, reinforced concrete structures used externally bonded steel plates to add strength and support. Recently, fiber reinforced polymer (FRP) plates have been used instead of steel plates because of their superior properties. Investigators developed a method to mathematically model bond strength between a carbon FRP and a concrete substrate. For each of 15 carbon FRP–concrete samples, they reported the maximum transferable load (kN) calculated by the model (Calc) and compared this with the corresponding maximum transferable load (kN) as measured in the laboratory (Meas). The data are given here:

```
conc <- read.csv("hw5q2.csv", header = TRUE)
```

(a) Construct a scatterplot of the data. Does it seem to be the case that, in general, when the measured load is low, the calculated load is also low? For each sample, are the two variables relatively close in value?

(b) Calculate the value of the sample correlation coefficient. Does it confirm your impression from the scatterplot?

3. The collapse of reinforced concrete buildings during earthquakes can result in significant loss of property and life. Often such collapses are caused by concrete column axial failure. A study investigated how y = maximum sustained shear (Vmax, in kN) is influenced by x_1= transverse reinforcement yield stress (MPa) and x_2 = concrete cylinder compressive strength (Mpa).

```
shear <- read.csv("hw5q3.csv", header = TRUE)
```

Use R to fit and summarize

(a) $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$

(b) $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$

(c) $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_1^2 + \beta_5 x_2^2 + \varepsilon$

(d) Which one is the best model? Justify.