STAT 231 — Linear Models — Fall 2021

Homework 8 (R) – Multiple Linear Regression Part II

NAME: _____

Directions: Create a R markdown file in order to write up your solutions. You must show all of your work to receive full credit. Put a box around your final answer, whenever possible. Put the following command between each problem's writeup:

```
\pagebreak
```

1. (10 pts) Researchers at National Semiconductor experimented with tin-lead solder bumps used to manufacture silicon wafer integrated circuit chips. The failure times of the microchips (in hours) were determined at different solder temperatures (°C). These data can be found in `WAFER.Rdata`. The researchers want to predict failure time (y) based on solder temperature (x).

   (a) Construct a scatterplot for the data. What type of relationship, linear or curvilinear, appears to exist between failure time and solder temperature?

   (b) Fit the model, $E(y) = \beta_0 + \beta_1 x + \beta_2 x^2$, to the data. Give the least squares prediction equation.

   (c) Conduct a test to determine if there is upward curvature in the relationship between failure time and solder temperature. (Use $\alpha = .05$.) Interpret the result.

2. (10 pts) In the Journal of Experimental Psychology: Learning, Memory, and Cognition (July 2005), University of Basel (Switzerland) psychologists tested the ability of people to judge risk of an infectious disease. The researchers asked German college students to estimate the number of people who are infected with a certain disease in a typical year. The median estimates as well as the actual incidence rate for each in a sample of 24 infections are provided in `INFECTION.Rdata`. Consider the quadratic model, $E(y) = \beta_0 + \beta_1 x + \beta_2 x^2$, where y = actual incidence rate and x = estimated rate.

   (a) Fit the quadratic model to the data, then conduct a test to determine if incidence rate is curvilinearly related to estimated rate. (Use $\alpha = .05$.)

   (b) Construct a scatterplot for the data. Locate the data point for Botulism on the graph. What do you observe?

   (c) Repeat part a, but omit the data point for Botulism from the analysis. Has the fit of the model improved? Explain.

3. (10 pts) The Journal of Accounting Education (Vol. 25, 2007) published the results of a study designed to gauge the best method of assisting accounting students with their homework. A total of 75 accounting students took a pretest on a topic not covered in class, then each was given a homework problem to solve on the same topic. The students were assigned to one of three homework assistance groups. Some students received the completed solution, some were given check figures at various steps of the solution, and some received no help at all. After finishing the homework, the students were all given a posttest on the subject. The dependent variable of interest was the knowledge gain (or test score improvement). These data are saved in the `ACCHW.Rdata` file.

   (a) Propose a model for the knowledge gain (y) as a function of the qualitative variable, homework assistance group.

(b) In terms of the β's in the model, give an expression for the difference between the mean knowledge gains of students in the "completed solution" and "no help groups."

(c) Fit the model to the data and give the least squares prediction equation.

(d) Conduct the global F-Test for model utility using $\alpha = .05$. Interpret the results, practically.

4. (10 pts) Which insect repellents protect best against mosquitoes? Consumer Reports (June 2000) tested 14 products that all claim to be an effective mosquito repellent. Each product was classified as either lotion/cream or aerosol/spray. The cost of the product (in dollars) was divided by the amount of the repellent needed to cover exposed areas of the skin (about 1/3 ounce) to obtain a cost-per-use value. Effectiveness was measured as the maximum number of hours of protection (in half-hour increments) provided when human testers exposed their arms to 200 mosquitoes. The data from the report are saved in the `REPELLANT.Rdata`.

   (a) Suppose you want to use repellent type to model the cost per use (y). Create the appropriate number of dummy variables for repellent type and write the model.

   (b) Fit the model, part a, to the data.

   (c) Give the null hypothesis for testing whether repellent type is a useful predictor of cost per use (y).

   (d) Conduct the test, part c, and give the appropriate conclusion. Use $\alpha = .10$.

   (e) Repeat parts a-d if the dependent variable is maximum number of hours of protection (y).

5. (10 pts) Refer to the Journal of Engineering for Gas Turbines and Power (January 2005) study of a high-pressure inlet fogging method for a gas turbine engine, Exercise 4.13 (p. 188). Consider a model for heat rate (kilojoules per kilowatt per hour) of a gas turbine as a function of cycle speed (revolutions per minute) and cycle pressure ratio. The data are saved in the `GASTURBINE.Rdata` file.

   (a) Write and fit a complete second-order model for heat rate (y). Summarize the results.

   (b) Give the null and alternative hypotheses for determining whether the curvature terms in the complete second-order model are statistically useful for predicting heat rate (y).

   (c) For the test in part b, identify the "complete" and "reduced" model.

   (d) Write and fit the reduced model for heat rate (y). Summarize the results.

   (e) Find the values of SSE_R, SSE_C, and MSE_C.

   (f) Compute the value of the test statistic for the test of part b. Find the rejection region for the test of part b using $\alpha = .10$. State the conclusion of the test in the words of the problem.