

## BS805 Fall 2021 Week 5

Be sure to follow the *Assessment Guideline 1: Writing up Homework* at the end of the syllabus in preparing the homework for submission.

Homework assignments need to be uploaded to the blackboard website by 2 PM on the due date.

In each homework report, be sure to include an introductory and a summary paragraph. Also, include the relevant parts of your SAS code where appropriate in your answer for each question.

A study was conducted to determine the degree of recovery that takes place 90 minutes following treatment in 36 children diagnosed at a clinic with moderate-to-severe dehydration. Patients were treated with an electrolyte solution in popsicle-form at different doses (measured in mEq/L). A rehydration scale score was subjectively assigned based on examination and parent report. The scores were scaled to potentially range from 0 to 100 (as a percent of recovery). In addition to dose, the age and weight of each child were recorded. The data on this study are available for you in the Excel file, *dehydration\_f22.xlsx*.

- 1) There could be problem values in this data set. If present, they could be influence points, outliers, or both. Run a simple linear regression with rehydration score as the outcome and dose as the predictor variable. Plot the residuals and predicted values from this regression, and then print them out. On the basis of this plot and listing, give the ID of the problem value(s) if present, and report if it is an outlier, an influence point, or both. If there are problem values, create a saved SAS data set that excludes them. If not, create a permanent SAS data set including all data.
- 2) Make another program that reads in the saved SAS data set. Perform the remainder of the analysis with this data set.
  - a) Find the correlation values between the four numeric variables (rehydration score; dose; age; weight) using the CORR procedure. Do any of the predictor variables have a significant correlation with another predictor at the .05 level?
  - b) Run a simple linear regression predicting rehydration score from dose. Determine if the relationship between these two variables is significant at the 0.05 level. What number from the output quantifies this relationship?
  - c) Run a multiple linear regression predicting rehydration score from dose, age, and weight. How good is the overall model in predicting rehydration score? Is the overall model statistically significant at the 0.05 level? What is the most predictive factor? Present specific numeric evidence from the output that answers these questions.

d) Using the model from parts b) and c) as a base, is there a statistically significant effect modification (interaction) at the 0.05 level between:

- i) Dose and age;
- ii) Dose and weight?

(Use two separate models to individually assess each interaction.) What would the presence of a significant interaction for either of these mean here with respect to how children with dehydration should be treated?