

BIOST 509: Homework 3

Instructor: Amy Willis, Biostatistics, UW

Due date: 1:00pm on October 25, 2019 via Canvas

Instructions

All questions for this homework relate to the dataset “biomarkers.csv”. This dataset contains longitudinal data on 2 biomarkers for 4 subjects. **ID** gives the subject ID, **Date** gives the date that a sample was taken, and **biomarker1** and **biomarker2** contain the biomarker

The dataset is available from Canvas in **Pages/Week 3 materials** or **Files/datasets**.

Read in the dataset and answer the following questions. Submit your answers to the below questions in a R Script (.R), Word (.doc or .docx) or pdf file to Canvas. Don't forget to include both your code and answers in your response.

Questions

1. How many observations were there for each subject ID?
2. What is the earliest recorded observation for each of the 4 subjects?
3. One of the dates contains an error. Find it and create a new tibble/data frame with the observation with the erroneous date removed.
4. What are the different ways that missing data on Biomarker 1 were entered? Which was the most common way to code missing data?
5. Create a new data frame with missing observations coded in the most common way it was coded in the original data set, and convert **biomarker1** to numeric data.
6. Optional: Remove all rows from the dataset that contain missing data on *both* biomarkers 1 and 2. How many observations are left?