

# ST 517: Data Analytics I

## Module 2 Homework

1. (3 points) (Adapted from Exercise 21, Section 6.7, *The Statistical Sleuth*, 2nd Ed.) The dataset `bearings.csv` contains failure times (measured in millions of cycles) of engine bearings made from five different compounds.
  - (a) (1 point) Read in the data. What type of data object is `bearings`? With `ggplot2`, create side-by-side boxplots of the failure times by compound.
  - (b) (2 points) Determine the pairs of engine ball bearing compounds for which there is a significant difference in mean failure times. Present your findings in a **short** statistical report ( $\approx 4$  sentences).
2. (3 points) Using the data from the last weeks lab and homework, `case0501` in the `Sleuth3`, answer the question “Which diets differ in their mean lifetime?”
3. (2 points) A soda company is developing a new soda. They are trying to determine how much sugar to put in it to give it the best taste. In order to evaluate this, they have made samples with ten different sugar levels. Each level of sugar is assigned to a random sample of seven people, and each person rates the soda on a scale from 0 to 10. The company would like to make inference on the difference between mean ratings between each pair of sugar level.
  - (a) How many pairwise comparisons are there?
  - (b) Name 2 procedures that the company could use to control the familywise Type I error rate on the differences of means? Explain.
4. (2 points) A consumer research group names seven types of department stores. They take a random sample of six department stores for each type and record their yearly sales. They wish to find significant pairwise differences in mean yearly sales for seven types of stores.
  - (a) How many pairwise comparisons are there?
  - (b) The group wishes to control the familywise Type I error rate at 1% using Bonferroni methods. What should be the Type I error rate of each pairwise comparison?