

Homework 1: Describing the Data

(Due: Wednesday Sep 02, before class)

Please submit a hard copy of your homework before class and email your R code, if applicable, to Alex Goebel at AJGOEBEL@email.sc.edu

1. Transit survey components [15 points]

The transportation office wants to estimate the proportion of all currently enrolled students who would use a late-night shuttle. It samples university email addresses and obtains 620 completed questionnaires.

- Identify the target population, sampled population, sampling unit, observation unit, and response.
- Explain why the target and sampled populations may differ.

2. Commute-time distribution [20 points]

The following commute times in minutes were recorded for 24 students.

- Construct a frequency table using classes 10-14, 15-19, ..., 45-49.
- Draw a histogram and describe shape, center, spread, and unusual values.

12	18	21	16	24	19	15	27
22	18	35	20	17	23	25	19
14	31	18	26	20	22	45	16

3. Center, spread, and boxplot [30 points]

Using the commute data, compute the mean, median, mode, range, sample variance, sample standard deviation, Q1, Q3, IQR, and the book's scaled MAD. Use the median-of-halves quartile convention. Apply the 1.5 IQR rule and sketch a boxplot.

4. Effect of the outlier [15 points]

Remove the 45-minute value and recompute the mean, median, sample SD. Which measures change most, and why?