

Summary Samples from Tuesday class

A: We collected commute times from the class and recorded 17 times in minutes. While R provides functions for calculating all basic statistics, it's possible to calculate them from scratch. Example functions are mean, median, and sd. The calculation of standard deviation has two forms: one from a sample and one from a population.

--

B: We collected 17 commute times in minutes, the least was 5 minutes (two cases) and the longest was 120 minutes. We quickly created a stem and leaf display, but we didn't have enough space to report the longest commute times. I also noticed that the longest commute times always seem to rounded to the nearest 10 minutes. We used R to calculate basic statistics. The arrow in R is comparable to the single-equal-sign for variable assignment in some programming languages. If you have a vector of data values in R, it's easy to get basic statistics and produce plots and charts.

--

C: In this script, we started with a dataset of 17 commute times (in minutes) collected from the class, then explored several ways to describe the distribution of that data. First, we calculated the **mean** and **median** to compare measures of center — the median is more resistant to extreme values (like the 90- and 120-minute outliers in our data), which is why it differs from the mean. Next, we computed the **deviations** from the mean to see how far each observation sits from the average, then took the absolute values of those deviations and averaged them to get the **mean absolute deviation (MAD)** — a straightforward measure of spread. Finally, we squared the deviations instead and averaged those to get the **variance**, then took the square root to arrive at the **population standard deviation**, which brings the spread measure back into the original units (minutes) and is the most commonly used summary of variability.

--

D: We analyzed 17 commute times in R. I noticed that the mean was larger than the median. I think that's because some people have really long commute times. When R presented a histogram, I noticed that there were small frequency items to the right of the graph.

--

E: We analyzed a dataset of 17 commute times collected from the class, computing the mean and median to compare measures of center, then explored spread through the mean absolute deviation, variance, and standard deviation. We also visualized the data several ways: a stem and leaf display on the whiteboard, and a boxplot and histogram in R.

--

F: For this activity, we collected data values in class and did some analysis with it. Using R, we worked out several statements and functions. We also produced some graphs that revealed the properties of the data that we collected. The graphs were helpful in seeing the average values from the data that we collected.