Introduction to Statistics

November 11, 2019

1 Describing Data

- **Statistics** are metrics which are used to summarize data.
- Statistics can also be used to make predictions about data.
- When describing data, what we are interested in is:
  - How are data distributed?
  - What is the center of the distribution?
  - What does a typical example look like?
  - How spread out are the data?
  - How do different distributions of data compare to each other?
- For example, let's suppose you want to explore the Exam 2 scores of the Spring 2019 and Fall 2019 sections of Quantitative Literacy.
  - **Spring 2019 Scores**

    18 27 27 34 36 36 43 50 57 59
    59 64 64 68 73 73 73 75 80 82
    84 100
  - **Fall 2019 Scores**

    57 58 62 68 69 74 83 85 88 89
    93 93 94 97 97 99 100 102 103 106
    106
- Which section did better?
- How can we quantify who did better?

2 Describing Data Graphically

- One easy way to summarize data is graphically.
- **Stem and Leaf Plots** provide a quick way to summarize data.
- To construct a stem and leaf plot:
  1. Look at the range of values.
  2. Write down "stems" this is the set of all but the last digits of the data. (for example, in the above data, the stems are the numbers 1-10)

3. Beside each stem, write down the leaves (the last digits) of the numbers.
4. Sort the leaves from smallest to largest.

- Let's make a stem and leaf plot of the Spring and Fall scores.
- What does this tell us?
- If we put the stem and leaf plots back to back, we can compare the distributions of the two.
- A **histogram** is a bar chart of frequencies.
- To construct a histogram:
  1. Determine the range of "bins" we are going to have.
  2. Count the number of data elements that fall within each bin.
  3. Draw a bar chart of these frequencies.
- Let's construct a histogram of the Spring and Fall exam data.
- We can compare histograms by placing them back to back.
- What does this tell us? Where is the center? What is the distribution of grades like?