Measures of Center

Reading

Sections 1.6.2, 1.6.5, 1.6.6

Practice Problems

1.9.6 (page 65) 1.42, 1.44 a,b, 1.45, 1.47 (for "mean" part of the questions), 1.49b, 1.55, 1.56, 1.62

Notes

- There are two main measures of center, and another less used:

  Median The "middle" value. Half the values are below it, half above.

  Denoted by M.

  The median is stable to the effects of outliers.

  Mean Numerical average of all values.

  Denoted by $\bar{x}$.

  The mean has nice numerical properties, but it is affected by outliers.

  Trimmed-mean A mean computed after a percent of data from each end is removed.

  Resembles the mean, but is not much affected by outliers.

- NOTE: The book, and other resources, talk about the **mode**, referring to the "most frequent value". We will not use that term for this purpose in this class, but you should be aware of it.

- One important consideration is how these measures behave to changes in the data. This is the notion of "resistance" or "robustness":

  Resistant A measure is called *resistant*, or *robust*, if its value is not considerably affected by extreme changes to a few data values.

  In particular, outliers can be considered as values that were near others but their value somehow changed radically. So resistant measures are not affected much by outliers.

- Based on this description, the median is a resistant measure, but the mean is not. (THINK ABOUT IT!)

- Measures of center and skeweness:

  Symmetric Mean and Median about the same

Skewed right The top half of the numbers pulled up, they take the mean with them.

Mean bigger than Median

Skewed left The bottom half of the numbers pulled down, they take the mean with them.

Mean smaller than Median

- Example to consider:

  If we imagine the populations of each of the largest 150 countries as our data points, we will find a mean / average population of 4.6 million and a median population of 1.07 million. Think about these numbers and whether/how they make sense.