Measures of Center

Def: Descriptive measures that indicate where the center or "most typical" value of a data set lies are called measures of central tendency or measures of center.

These are also referred to as *averages*.

The Mean

Def: The mean of a data set is the sum of the observations divided by the number of observations.

The Mean

If we had a variable called `ages` with values $21, 22, 26, 18, 19$, and 21, the mean would be

$$\frac{\text{sum of values}}{\text{total \# of observations}} = \frac{21 + 22 + 26 + 18 + 19 + 21}{6}.$$

The mean or *average* age is 21.167.

SUMMATION NOTATION EXAMPLE

The Mean

For a variable x, we denote the sample mean by $\bar{x}$.

Symbolically, we write

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$$

where n is the number of observations.

The Median

- The **median** takes the data and splits it in half.
- To find the median, arrange our data in increasing order and then find the halfway point.

The Median

If we have an odd number of observations, say,

$$3, 1, 5, 4, 2$$

we sort and take the observation exactly in the middle.

In this case,

$$1, 2, \mathbf{3}, 4, 5$$

3 is the median.

The Median

If we have an even number of observations

$$3, 6, 1, 5, 4, 2$$

we sort and cut the data exactly in half

$$1, 2, 3 \quad | \quad 4, 5, 6.$$

The median is the average of the two observations closest to the halfway point

$$\frac{3+4}{2} = 3.5$$

The Mode(s)

We can also find modes numerically:

Find the frequency of each value in the data set. Any value that occurs with the greatest frequency is a mode of the data set.

If we had data $1, 2, 3, 4, 5$, there is no mode.

For data $3, 6, 1, 8, 3, 3, 6$,

Number	Frequency
1	1
3	3
6	2
8	1

The mode is 3.

Measures of Center

400	400	300	940	300
300	400	300	400	
450	800	450	1050	

Measure of Center	Result
Mean	474
Median	350
Mode	300

Which to use?

The mean is highly sensitive to outliers.

A resistant measure is one which is not sensitive to the influence of a few extreme observations.

A trimmed mean removes a percentage of the highest and lowest observations before calculating a mean. This can help improve resistance.

Which to use?

- We use the **mean** when the data are relatively **symmetric**.
- The **median** is used for data that are **skewed**.
- The **mode** is best for **qualitative data**.