3.1: Measures of Center

Now, we will begin studying some numerical measures that describe data sets. There are two basic types:

- Measures of central tendency (this section)
- Measures of dispersion (next section)

Summation Notation:

Summation notation is a compact way to write "add up n numbers" of "do something to n numbers first, and then add them up." The numbers are represented as $x_1, x_2, \ldots, x_n$"

Σ Greek letter capital Sigma (stands for sum)

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \ldots + x_n$$

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \ldots + x_n$$

index variable - counts upward from 1 to n

Example 1: Consider the numbers 8, 2, 6, 10, 4, 9. Find $\sum_{i=1}^{6} x_i$ and $\sum_{i=1}^{6} x_i^2$.

$$\sum_{i=1}^{6} x_i = 8+2+6+10+4+9 = \boxed{39}$$

$$\sum_{i=1}^{6} x_i^2 = x_1^2 + x_2^2 + x_3^2 + \ldots + x_6^2$$

$$= 8^2 + 2^2 + 6^2 + 10^2 + 4^2 + 9^2 = 64 + 4 + 36 + 100 + 16 + 81 = \boxed{301}$$

The Mean: Ungrouped Data: (average)

The *mean* of a set of quantitative data is equal to the sum of all the measurements in the data set divided by the total number of measurements in the set.

If $x_1, x_2, \ldots, x_n$ is a set of n measurements, then the *mean*, or *average*, is given by

$$[\text{mean}] = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \ldots + x_n}{n} \quad \text{where}$$

$\bar{x}$ $\quad \bar{x} = [\text{mean}]$ if data set is a sample

μ $\quad \mu = [\text{mean}]$ if data set is the population

Greek letter mu

μ (used for population mean)

The median:

Sometimes the mean can be misleading for a data set. Suppose that a math class had 7 students with test scores (out of a possible 100) of 88, 99, 7, 78, 89, 94, and 75.

$$\text{mean} = \frac{88+99+7+78+89+94+75}{7} = 75.71$$

7 75 78 88 89 94 99

Median = 88

The *median* is unaffected by extreme values (outliers). Essentially it is the "middle" of the data set.

To find the median, you'll need to sort the data in numerical order.

<u>The Median (Ungrouped Data)</u>:

- If the number of measurements is odd, the median is the middle measurement when the measurements are arranged in descending or ascending order.
- If the number of measurements is even, the median is the mean of the two middle measurements when the measurements are arranged in descending or ascending order.

<u>Example 2:</u> Find the median of the test scores 88, 99, 7, 78, 89, 94, and 75.

<u>Example 3:</u> Find the median of the test scores 88, 85, 99, 7, 78, 89, 94, and 75.

<u>Example 4:</u> Provide some everyday examples in which the median is more useful than the mean.

The mode:

> <u>The Mode</u>:
>
> The *mode* is the most frequently occurring value in a data set, provided it occurs at least twice. There may be a unique mode, several modes, or no mode.
>
> A data set with two modes is called *bimodal*.

<u>Example 5:</u> Find the median and mode for the following data sets.

a. {4, 5, 5, 5, 5, 6, 7, 8, 12}

Median: 5

mode: 5

b. {1, 2, 3, 3, 3, 5, 7, 7, 7, 23}

Median: $\frac{3+5}{2} = \frac{8}{2} = \boxed{4}$ (find mean of the middle two points)

Modes: $\boxed{3, 7}$ (bimodal)

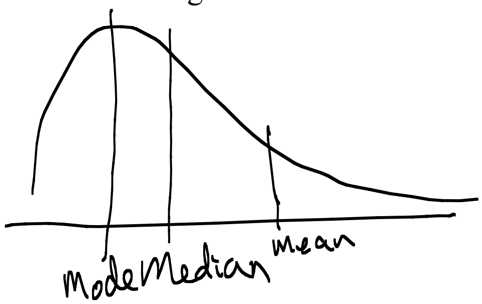
c. {1, 3, 5, 6, 7, 9, 11, 15}

<u>Example 6:</u>

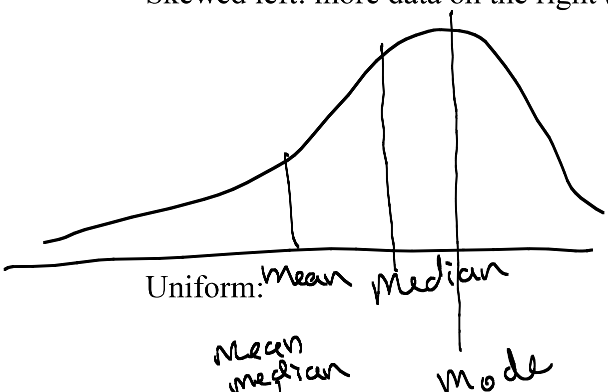
Mean, median, and mode for distributions of different shapes:

Median: half the area is on each side

Skewed right: more data on the left (lower) end.



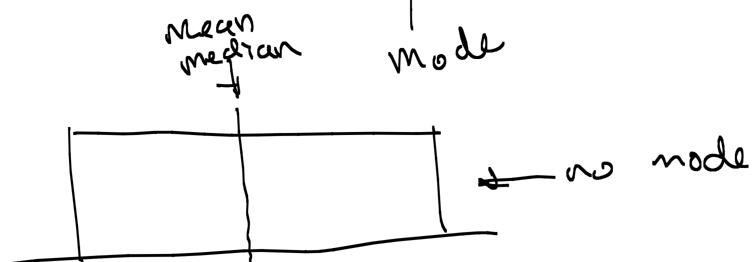
Skewed left: more data on the right (upper) end.



If data set is symmetric, the mean and median are equal.

The mode will be the peak

Uniform:



Symmetric: