3.1: Measures of Center (measures of central tendency)

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 $x_1, x_2, \ldots, x_n$"

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

i = index variable. It counts from 1 to n.

$\sum$ is the capital Greek letter Sigma (stands for "sum")

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

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

$$\sum_{i=1}^{6} x_i^2 = 8^2+2^2+6^2+10^2+4^2+9^2$$
$$= 64+4+36+100+16+81 = \boxed{301}$$

$$\sum_{i=1}^{6} (x_i-1)^2$$
$$= (8-1)^2+(2-1)^2+(6-1)^2+(10-1)^2+(4-1)^2+(9-1)^2$$
$$= 49+1+25+81+9+64$$
$$= \boxed{229}$$

(check it)

The Mean: Ungrouped Data:

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

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

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

$\mu =$ [mean] if data set is the population

$\bar{x}$ = sample mean

μ = population mean

μ: Greek letter mu

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.

Population mean: $\mu = \dfrac{88+99+7+78+89+94+75}{7} = \boxed{75.71}$

Note: All but one student scored at least 75. The low score of 7 pulled the mean down.

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.

$\cancel{7}$ $\cancel{75}$ $\cancel{78}$ 88 $\cancel{89}$ $\cancel{94}$ $\cancel{99}$

median is 88.

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

$\cancel{7}$ $\cancel{75}$ $\cancel{78}$ $\underbrace{85 \quad 88}$ $\cancel{89}$ $\cancel{94}$ $\cancel{99}$

$\text{Median} = \dfrac{85+88}{2} = \boxed{86.5}$

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

Home prices, Salaries

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. {$\cancel{4}$, $\cancel{5}$, $\cancel{5}$, $\cancel{5}$, 5, $\cancel{6}$, $\cancel{7}$, $\cancel{8}$, $\cancel{12}$}

Median: 5

Mode: 5

b. {$\cancel{1}$, $\cancel{2}$, $\cancel{3}$, $\cancel{3}$, 3, 5, $\cancel{7}$, $\cancel{7}$, $\cancel{7}$, $\cancel{23}$}

Median: $\frac{3+5}{2} = 4$

Mode: 3, 7

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

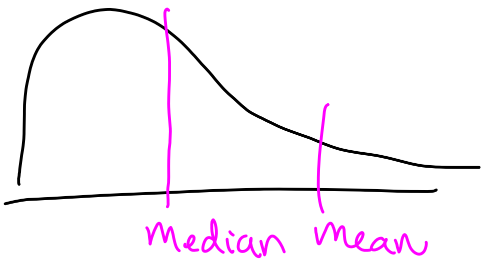
Median: $\frac{6+7}{2} = \frac{13}{2} = 6.5$

No mode

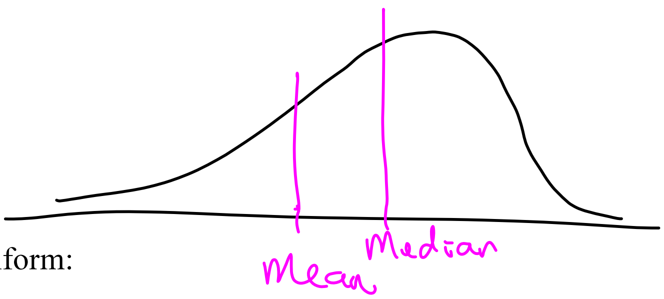
<u>**Example 6:**</u>

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

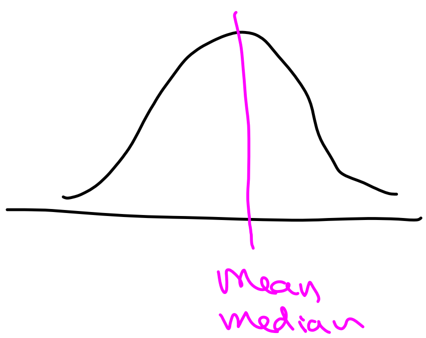
Skewed right: more data on the left (lower) end.

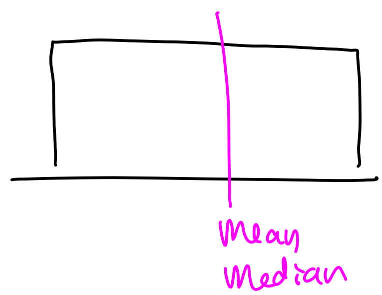


Skewed left: more data on the right (upper) end.



Uniform:





Symmetric: