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 $x_1, x_2, \ldots, x_n$"

index variable

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

Σ: Greek letter sigma stands for sum

x_i = the i^{th} x-value

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

$x_1 = 8$, $x_2 = 2$, $x_3 = 6$, etc

$$\sum_{i=1}^{6} x_i = x_1 + x_2 + \ldots + x_6 = 8+2+6+10+4+9 = 39$$

$$\sum_{i=1}^{6} x_i^2 = x_1^2 + x_2^2 + \ldots + x_6^2 = 8^2+2^2+6^2+10^2+4^2+9^2$$

$$= 64+4+36+100+16+81 = 301$$

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

μ = [mean] if data set is the population

x-bar: $\bar{x}$ = sample mean

mu: μ = population mean

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.

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

All but one of the students scored at least 75. The 7 drags the mean way 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.

The Median (Ungrouped Data): (data that's the list of data points, rather than a frequency table. Frequency table is a summary of the data)

- 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.

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

~~7~~, ~~75~~, ~~78~~, 88, ~~89~~, ~~94~~, ~~99~~

$\boxed{\text{Median} = 88}$

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

~~7~~, ~~75~~, ~~78~~, $\underbrace{85, 88}$, ~~89~~, ~~94~~, ~~99~~

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

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

home prices,

salary etc

The mode:

> The Mode:
>
> 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*.

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

Modes: 3, 7 (bimodal)

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

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

No mode

Example 6: In a frequency table, the modal class is the most frequently occurring category

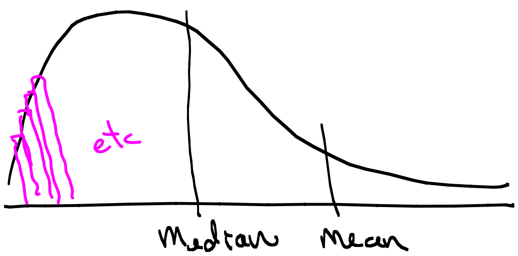
Age	
<20	5
21-29	7
30-39	22
40-49	40
50-59	36

Modal class: 40 – 49

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

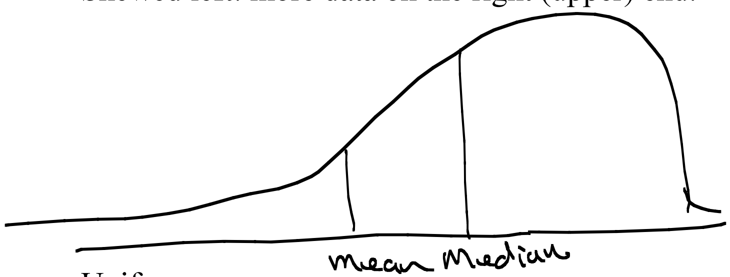
Median: Half the area is left of the median, half the area to the right

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



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

The outliers in the tail will pull the mean in the direction of the tail.



Uniform:
(symmetric)

Symmetric:
(bell-shaped)