

Measuring the Centre of Data

Today we are going to talk about mean, median, and mode.

The mean is simply the arithmetic average.

The median is the middle value of an ordered data set.

The mode is the most frequently occurring value –aka “the most popular” value.

Some cool notation:

x A data value

n The number of data values in either the sample or the population

Σ A sum

\bar{x} The mean of a **sample** [sometimes called “x-bar”]

μ The mean of a **population** [pronounced “mu”]

$$\mu = \frac{\Sigma x}{n} \quad \text{OR} \quad \bar{x} = \frac{\Sigma x}{n}$$

To find the median, order your data set, and find the middle number. If you have an odd number of data, then the median will be one of the data values. If you have an even number of data, then the median is the average of the two middle values and might not be one of the data values.

If there are n data values, find $\frac{n+1}{2}$. The median is the $\left(\frac{n+1}{2}\right)$ th data value.

When $n = 23$, the median = $\frac{23+1}{2} = 12$. Which means the median is the 12th ordered data value.

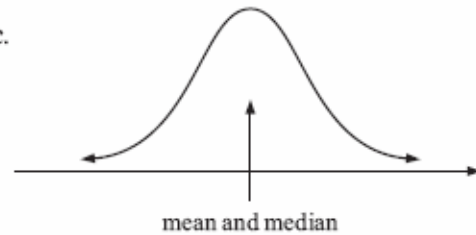
When $n = 22$, the median = $\frac{22+1}{2} = 11.5$. Which means the median is the average of the 11th and 12th data values.

The mean and the median are often different values. Based on my own empirical data from years of teaching, whenever I calculate the statistics from any test that I have given, the mean and the median are often different due to “outliers”.

See the graphs on page 427

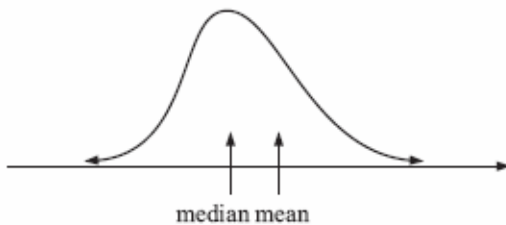
THE RELATIONSHIP BETWEEN THE MEAN AND THE MEDIAN FOR DIFFERENT DISTRIBUTIONS

First of all consider a distribution that is **symmetric**.

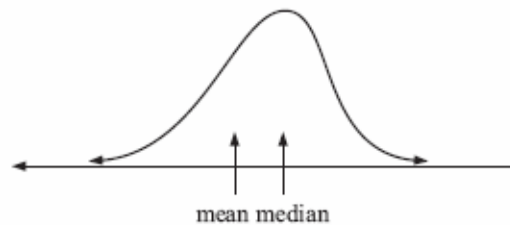


Note: Other symmetrical shapes exist other than this 'bell' type shape. Any symmetrical distribution will have a mean and median that are approximately equal.

positively skewed distribution



negatively skewed distribution

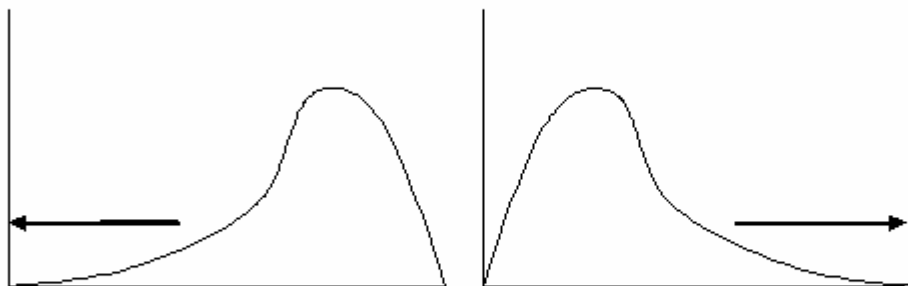


Hence if the data set has symmetry, both the mean and the median should accurately measure the centre of the distribution.

Note: The mean is influenced by all data values in the data set whereas the median is not.

From Wikipedia: [that fountain of knowledge!]

1. **positive skew:** The right tail is longer; the *mass* of the distribution is concentrated on the left of the figure. The distribution is said to be *right-skewed*.
2. **negative skew:** The left tail is longer; the mass of the distribution is concentrated on the right of the figure. The distribution is said to be *left-skewed*.



Negative Skew

Elongated tail at the **left**

More data in the left tail than would be expected in a normal distribution

Positive Skew

Elongated tail at the **right**

More data in the right tail than would be expected in a normal distribution

Let's do the investigation on page 427. If you need TI instructions, you can use your book's cd and click on the "TI" icon on page 427.

Be careful when entering the data. It is very easy to make a typo.

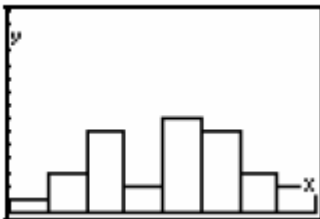
Here's what our window should be:

Press **WINDOW** to display the window editor.
 The X values should take into account the minimum and maximum values in the dataset.
 Set the values as shown in the screen dump at right.

```

WINDOW
Xmin=240
Xmax=280
Xscl=5
Ymin=0
Ymax=15
Yscl=1
Xres=1
    
```

And here is what the graph should look like:



Now let's measure the centre of data using our TI!

<pre> 1-Var Stats L1 </pre>	<pre> 1-Var Stats x̄=260.7633333 Σx=7822.9 Σx²=2042308.47 Sx=9.064880444 σx=8.912518661 ↓n=30 </pre>	<pre> 1-Var Stats ↑n=30 minX=244.6 Q1=253.9 Med=263 Q3=266.5 maxX=277.5 </pre>
---------------------------------	--	--

$\bar{x} = 260.763$

Median = 263

Is the mean accurate? $MEAN < median$
negatively skewed

Let's do #3 in the investigation to see what happens.

median same
mean smaller

Now we know that we can easily find the mean and the median using our TI. The mode – the most popular data value – is also easy to find. It is the value that occurs the most often. Data sets can be bimodal which means that two data values have the same high frequency. It is possible to have a data set with no mode.

Let's try #1a on page 429.

Our data set:

2, 3, 3, 3, 4, 4, 4, 5, 5, 5, 5, 6, 6, 6, 6, 6, 7, 7, 8, 8, 8, 9, 9

Homework: page 429 and 430 #1d, 2, 4, and 5