Data Analysis

by Sophia

1. Data Analysis

Data analysis is what we do once we've collected our data. In this lesson, we will look at data analysis to identify the trends or key features of a data set. There are four components of data analysis that are key:

- The shape that a distribution will have
- The center of that distribution
- The spread of the data
- Any outliers in the data

TERM TO KNOW

Data Analysis

The understanding of the key features of a set of data--shape, center, spread, and outliers.
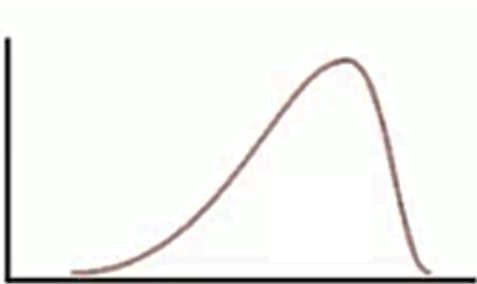
2. Shape

Shape is a qualitative notion telling us where most of the points lie in the distribution.

↪ EXAMPLE For instance, in this shape below, you would say that most of the data points are in the hump, where the line is highest on the y-axis. There are not a lot of data points on the far right side, in what we'd call the tail of the graph.
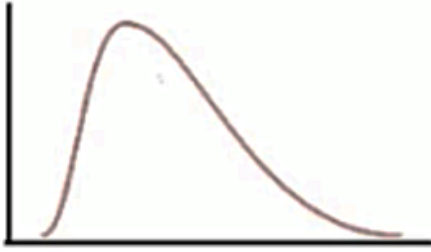
Shapes can be either skewed to the left or the right: The distribution in the example above is called skewed to the right because it has a hump on the left and a tail on the right.

In contrast, the distribution below is called skewed to the left. It has a tail on the left and a hump on the right.


🗎	TERM TO KNOW

Shape
> The qualitative description of the clustering of data points in a certain location when the data are graphed.

3. Center

The term **"center"** is essentially what it sounds like: it's wherever the middle is. There are a couple of different ways to measure the center.

In the graph below, a few arrows are pointing to the different measurements of the middle.

All different measures of centers
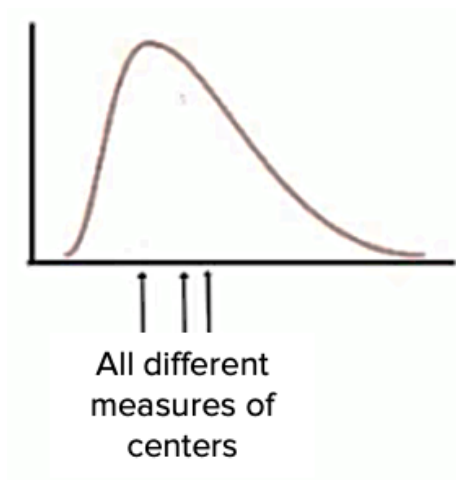
- The first arrow (the arrow furthest to the left on the x-axis) falls directly below the peak.
- The second arrow (the one in the middle) is a little further off to the right. It appears that if you drew a line directly through this arrow, about half the area of the graph would be to the left of it and about half the area would be to the right of it.
- The third arrow is farthest to the right of the x-axis.

Which one is the correct measure of center? They're all different measures, and they can all be correct in different situations.
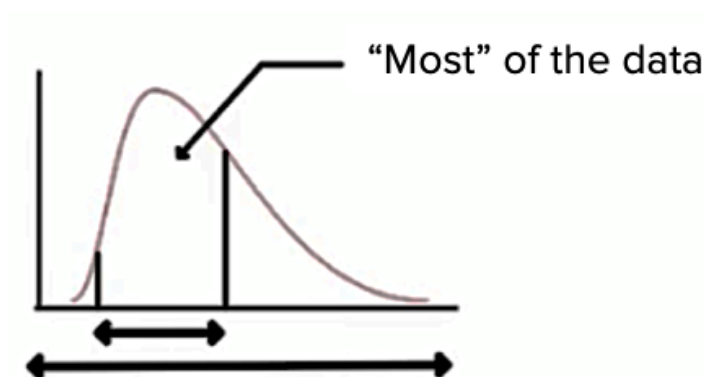
> 📄 **TERM TO KNOW**
>
> **Center**
> The "middle" of the data set. There are many measures of center.

---

4. Spread

Spread gives a numerical value relating how spread out the data points are. Just as with center, there are several different measures for spread.

Perhaps you are interested in where most of the data points lie, which would be below the hump:

"Most" of the data

Or, perhaps you are interested in the full range of data points from the lowest all the way the highest:

All the data
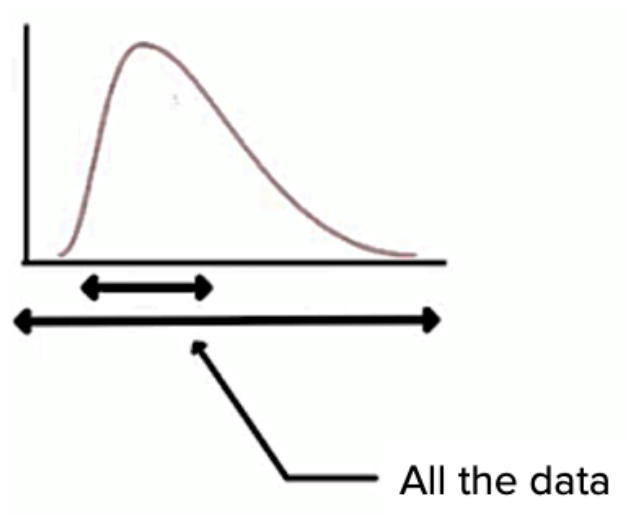
These would both be different, and correct, measurements of the spread.

📄 TERM TO KNOW

Spread
   The numerical description of how close the numbers are to the center.

5. Outliers

When analyzing a data set, it is important to look for **outliers**, which are not just the highest or lowest numbers, but are numbers that are very far above the next highest number in the data set or very far below the next lowest number in the data set.

⇗ EXAMPLE Suppose that a small class took an exam, and the scores were as follows:

90, 98, 89, 88, 46, 90, 91, 84, 94

Some students did very well on this test. In fact, most students scored in the 80's or the 90's. However, one person scored only 46. That 46 would be considered an outlier because it's so much lower than the rest of the pack.

⭐ BIG IDEA

Outliers are important data points because they are so high or low that they would be considered unusual.

📄 TERM TO KNOW

Outliers
   Points in a data set that are so high or so low as to be unusual, given the rest of the values.

✔ SUMMARY

Data analysis consists of clearly describing the four key elements of the data set: shape, center, spread, and outliers (if there are any). Some standard descriptions are used to describe the shape, such as skewed to the left and skewed to the right, and there are also several different measures for the center and spread, which are typically numbers. Outliers are values that are so high above or so far below the rest of the data set that they would be considered unusual.

Good luck!

Source: THIS TUTORIAL WAS AUTHORED BY JONATHAN OSTERS FOR SOPHIA LEARNING. PLEASE SEE OUR **TERMS OF USE**.

📄 TERMS TO KNOW

Center
The "middle" of the data set. There are many measures of center.

Data Analysis
The understanding of the key features of a set of data - shape, center, spread, and outliers.

Outliers
Points in a data set that are so high or so low as to be unusual, given the rest of the values.

Shape
The qualitative description of the clustering of data points in a certain location when the data are graphed.

Spread
The numerical description of how close the numbers are to the center.