Chapter 6: The Standard Deviation as a Ruler and the Normal Model

The Standard Deviation is used to compare two very different-looking values and tells you how the whole collection of values vary. The standard deviation is a natural ruler in determining how far values are from the mean. In order to calculate distance from the mean we must take a data value and subtract it by the mean of all values. And then divide this difference by the standard deviation. The result is: z or the z-score.
The Z-score is a standardized value. Standardized values have no units because z-scores measure the distance of each data value from the mean in standard deviations. A z-score of 3 tells you that a data value is 3 standard deviations away.

$$Z = \frac{(y - (\text{mean of } y))}{s}$$

You must consider the three aspects of distribution when you standardize data:

- Standardizing your data to z-scores does not change the **shape** of a variable's distribution
- Standardizing your data to z-scores makes the mean 0, so it changes the **center**.
- Standardizing your data to z-scores makes the standard deviation 1, so it changes the **spread**.

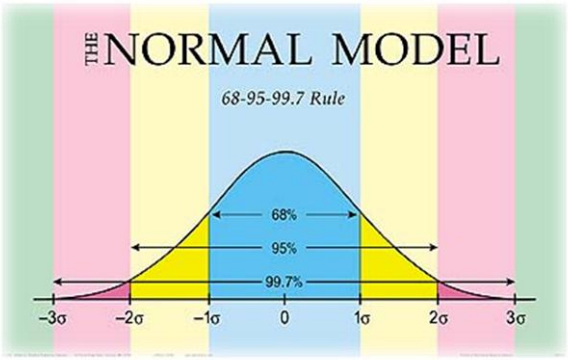
In order to determine if data values are unusual we use the **Normal model**.
If we used data with the normal model and standardize them, we get:

$$Z = \frac{y - \mu}{\sigma}$$

However, using the mean 0 and standard deviation 1 you should only use this if the distribution of the data values are unimodal and symmetric. If they are not, standardizing the data values will be useless.
This is called the Nearly Normal Condition. You must satisfy this condition in order to see if it's appropriate to use the normal model.

Normal models give us an idea of how extreme a value is by telling you how likely it is to find a particular data value far from the mean. In the normal model, about 68% of data values fall within 1 standard deviation, about 95% of the values are within 2 standard deviations away, and 99.7% of data values are within 3 standard deviations. Anything more than 3 standard deviations is very unlikely. This is called the 68-95-99.7 rule.

Oh! Don't forget the most important rule of all.
Make a Picture
Make a Picture
Make a Picture



David Jimenez