Normal Distribution Notes

Essential Understanding Standard deviation is a measure of how far the numbers in a data set deviate from the mean.

In the previous lesson you studied range and interquartile range. Each of these is a **measure of variation**. A measure of variation describes how the data in a data set are spread out.

Variance and **standard deviation** are measures showing how much data values deviate from the mean. The Greek letter σ (sigma) represents standard deviation. σ^2 (sigma squared) is the variance.

take note

Key Concepts Finding Variance and Standard Deviation

- Find the mean, $\bar{x}$, of the n values in a data set.
- Find the difference, $x - \bar{x}$, between each value x and the mean.
- Square each difference, $(x - \bar{x})^2$.
- Find the average (mean) of these squares. This is the variance. $\sigma^2 = \frac{\sum (x - \bar{x})^2}{n}$
- Take the square root of the variance. This is the standard deviation. $\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$

A **discrete probability distribution** has a finite number of possible events, or values.

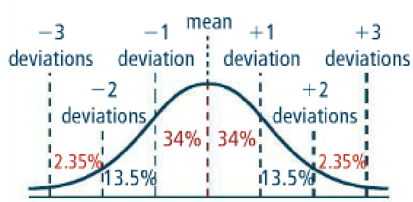
The events for a **continuous probability distribution** can be any value in an interval of real numbers. If a data set is large, the distribution of its discrete values approximates a continuous distribution.

Essential Understanding Many common statistics (such as human height, weight, or blood pressure) gathered from samples in the natural world tend to have a *normal distribution* about their mean.

A **normal distribution** has data that vary randomly from the mean. The graph of a normal distribution is a normal curve.

take note

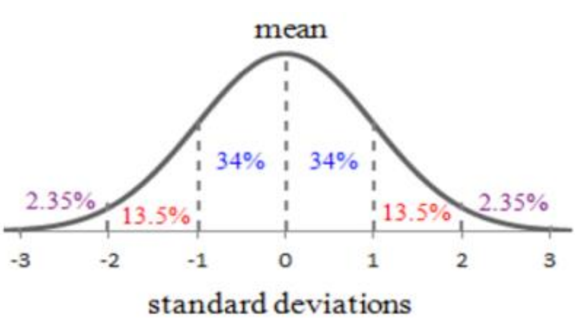
Key Concept Normal Distribution



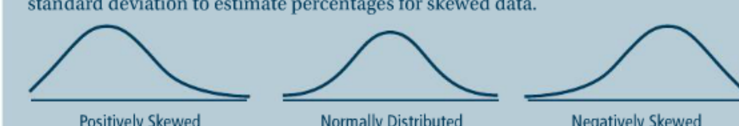
In a normal distribution,
- 68% of data fall within one standard deviation of the mean
- 95% of data fall within two standard deviations of the mean
- 99.7% of data fall within three standard deviations of the mean

A normal distribution has a symmetric bell shape centered on the mean.

Nearly all data lie within 3 standard deviations of the mean, μ. This is called the **empiricle** rule. The area under a normal curve is always 1. When calculating population percentages, the value will be less than 1, and needs to be converted to a percent.



Sometimes data are not normally distributed. A data set could have a distribution that is *skewed*, an asymmetric curve in which one end stretches out further than the other end. When a data set is skewed, the data do not vary predictably from the mean. This means that the data do not fall within the standard deviations of the mean like normally distributed data, and so it is inappropriate to use mean and standard deviation to estimate percentages for skewed data.



Example 1:

ACT test scores are approximately normally distributed. One year the scores had a mean of 21 and a standard deviation of 5.2.

a. Draw the normal distribution curve. Be sure to label the mean, standard deviations, and scores.

b. What is the interval that contains 95% of scores?

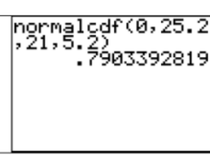
Not all scores are going to be exact standard deviations of the mean score, so we need to use a calculator. If you do not have a graphing calculator, you can go to an online normal distribution calculator.

http://www.mathcracker.com/normal_probability.php works great!

Use a calculator to calculate the following:

c. What percentage of ACT scores is less than 25.2?

Using a TI-83 or TI-84 Graphing Calculator

The distribution features are found by pushing 2ND DISTR VARS. A menu like the one at the right should appear. Option 2, normalcdf(is the feature that you want to use. This feature is the normal cumulative distribution function. It will calculate the percentage of data that fall between two numbers. Select option 2 by pushing L2 2 Z or by using your arrow keys to arrow down to 2 and pushing ENTRY SOLVE ENTER.	
The syntax required for this feature is normalcdf(lower bound, upper bound, mean, standard deviation). In the case of our example, it would be normalcdf(0, 25.2, 21, 5.2). Approximately 79% of the scores are below 25.2.	

d. What percentage of ACT scores is between 28 and 36?

Use a calculator to find the percentage. Use normalcdf(.

```
normalcdf(28,36,
21,5.2)
        .0871669871
```

e. What percentage of scores are greater than 27?