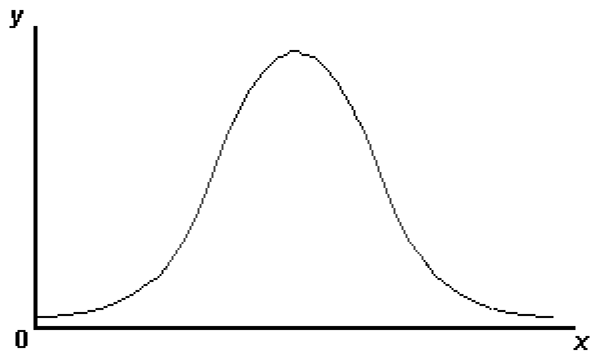
Statistics that Might be Helpful

The **mean** of data is commonly called the average and is calculated with the formula:

$$\overline{x} = \frac{\sum_{i=1}^{N} x}{N}$$

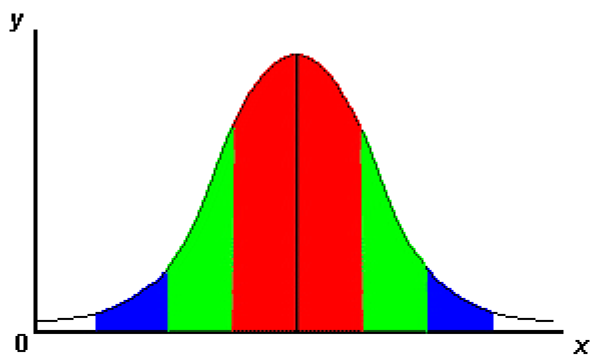
where x is an individual measurement, $\overline{x}$ is the mean, and N is the total number of measurements beginning with the i=1 measurement.

The **standard deviation** (σ) is a measurement that tells you how tightly all of the data are around the mean in a normal distribution (you may have heard of this as a bell curve).



Normal distribution

One standard deviation away from the mean (red area) accounts for ~68% of the data while 95% of the data can be found within two standard deviations (red and green) from the mean. The red, green, and blue areas account for 99% of the data.



The standard deviation can be calculated by the formula:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \overline{x})^2}{N-1}}$$

where σ^2 is referred to as the **variance**. So, for example, suppose in a class of 10 students the following exam scores were obtained:58, 75, 74, 62, 76, 93, 48, 63, 74, and 78. The mean is calculated as follows:

$$\bar{x} = \frac{58+75+74+62+76+93+48+63+74+78}{10} = 70.1$$

and the standard deviation calculation is shown below:

x	x - $\bar{x}$	$(x$ - $\bar{x})^2$
58	-12.1	146.4
75	4.9	24.0
74	3.9	15.2
62	-8.1	65.6
76	5.9	34.8
93	22.9	524.4
48	-22.1	488.4
63	-7.1	50.4
74	3.9	15.2
78	7.9	62.4

$$s = \sqrt{\frac{146.4+24.0+15.2+65.6+34.8+524.4+488.4+50.4+15.2+62.4}{9}} = 12.6$$

Now, this suggests that 68% of all students should be ± 12.6 units from the mean of 70.1 or should have scored between 57.5 and 82.7. Scores between 44.9 and 95.3 should account for 95% of all scores. Typically, those students scoring within one standard deviation within the mean would obtain a C on the exam while those within two standard deviations would get either a B or a D. Those within 3 standard deviations would get A's and F's. With a larger sample size, the 68%, 95%, and 99% data typically become more accurate i.e., the data fit a bell curve better. Also, with a sample size of typically 30 or more measurements, the standard deviation equation changes to reflect this

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - m)^2}{N}}$$

where m is the arithmetic mean for an infinite number of such measurements, i.e., the mean.

The last piece that may be of interest to you is the **median**. This is the central piece of data, or if there are an even number of measurements, it would be the average of the middle two pieces of measurements. The median for the exam above is:

$$\frac{74+74}{2} = 74$$