Statistics Summary for Biology 120 Lab

We use **statistics** to summarize data.

Example: It's easier to interpret an average height than it is to interpret the list of heights for 50 people that you measured.

We use sample statistics to estimate values for the population of interest.

Example: You can get the average height for a subset (a **sample**) of 1,000 American men relatively easily, but it would be impossible to go out and measure every American man on the planet (the **population**) and get the true average height of American men.

Highly variable data is less likely to result in sample statistics that accurately reflect your population. However, variability has less of an impact on your data set if you increase your sample size.

Example: If you have a sample of one 6'4" guy and one 7' guy, their average height (6'8") doesn't really reflect the typical height of most American guys. However, if you have a sample of 50 guys who are about 5'8", the one 6'4" guy and the one 7' guy, the average height of all these guys is now just over 5'8", which is more realistic.

The less variable you measurements, the more repeatable your sampling will be, and the more reliable your results will be. Also, the bigger your sample size, the better it will reflect the overall population. This is especially important in correcting for variable data. Thus, <u>bigger sample sizes = more reliable data</u>.

If you do replicate samples (measure the heights of 50 men at the mall, and then 50 more men on the NKU campus), their averages will be slightly different, because you're sampling different individuals. If you get a completely different average every time you sample, then you can't really tell which (if any) of those averages accurately reflect the average value for your population.

The standard error of the mean (SE) gives you a feel for how different your averages will be if you resample, and thus for how reliable your numbers are. The calculation of the SE takes into account both your sample size and the amount of variability in your data set.

You can also use SE to help you decide if two samples are likely to have different average values, or whether they're really similar.

A useful rule of thumb: If the SE is greater than 10% of your mean, your sample size is too small to give you reliable results. Chances are high that your sample mean is not really representative of the population's mean, and that you might get a very different value if you took another sample.