Fall 2005: Statistics for Biology 120 and 150 Labs

We use statistics to summarize data. It is more informative to look at an average than at the list of 50 values which was used to generate it.

We take **samples** (selected subsets of a population) in order to generate statistics which we can use to estimate the true values for the population. For example, you want to know the average height of American men. If you could measure every American man both here and overseas, you could determine their true average height. American men are your **population** of interest. However, measuring every American man is nearly impossible. So, you measure as many American guys as you can, you determine their average height, and if your data was pretty good, then you can safely say that the average height for your sample closely approximates the true average height of American men.

But...how do you tell if your data were good?

1) <u>Bigger sample sizes result in better data</u>. For example, say the average height of American men is 6 feet. You measure two guys, one who is 5'8" and one who is 5'. The average height for that sample would be 5'4", well under the true average height for the population. Because the sample was so small, the 5' measurement had a large effect on your average height for the sample. If you measured 10 guys, 9 who were 5'8" and one who was 5', now the average height for your sample would be about 5'7", much closer to the population average of 6'. <u>With large sample sizes, the occasional unusual measurement has less of an effect on your sample average</u>.

2) <u>The less variable your sample measurements are, the more confident you can be that your sample probably reflects your population</u>. For example, you have a can of mixed marbles and you want to know which color is most common. If you randomly remove 10 marbles, and 9 are white, you'll probably feel safe saying that most of the marbles are white. If you take out 10 marbles and each is a different color, you'll probably want to look at more before you decide on the most common color.

So, one way to tell if your sample is representative of the population it came from is to get a feel for how variable your data is. If your data are highly variable, you may need to sample more.

Variance is one way to look at this. For example, if the average height for a sample of 50 men is 5'9", it would be informative to know the average amount by which an individual measurement differed from the average. (So, if you randomly grab one of your 50 guys and measure him, how close to being 5'9" tall is he likely to be?) To determine the variance, you find out how far off average height each of your 50 guys is (guy 1 is 5'11" or 2" taller; guy 2 is 5'7" or 2" shorter; etc.), you square these numbers and add them up, and then divide them by one less than the sample size (49 in this case). Note: you square the differences in height so that when you add the 2" taller and 2" shorter measurements they don't cancel each other out, because [2 plus -2 is 0] while [2 squared plus -2 squared is 4 plus 4 or 8]. (When you square a negative number the result is always a positive number.)

Unfortunately, while you wanted to know how much taller or shorter than average any one of your 50 guys was likely to be, because you squared the numbers you added, your variance is now in units of inches squared, which is hard to interpret when you're looking at height, which is measured in inches.

To make the units easier to interpret, take the square root of the variance. This is the **standard deviation** of your sample, and is now back in units of inches, which makes sense for height measurements.

Now, you've figured out how similar any single measurement is likely to be to your sample's average value. In this case, how close any of your guys' heights are likely to be to 5'9". But remember, the real point sampling was to try to estimate the average height of all American men. So what you really want to know is if your sample average is likely to be similar to the true average height of all American men. So, how variable is your sample average?

If you measure a different sample of 50 men and determine their average height, you are **replicating** your first sample. Say you do 50 replicates of sampling. Now, instead of asking how close is an individual guy likely to be to being average height, ask how close the average height from one sample is likely to be to the average of all fifty of your sample means. The average height calculated from 50 different samples is more likely to accurately reflect the true average height of American men than is the average height of just one sample of 50 guys (because, as we discussed above, 50 replicates is a bigger sample size than 1 replicate).

As we did before, you can calculate a variance and a standard deviation for your set of mean heights. However, since we are now looking at variation between average heights for certain samples and the overall average, rather than between individual heights and the average height, the standard deviation of these averages is called the **standard error of the mean (SE)**.

Finally, if your data have a **normal distribution** (most measurements are close to being average, and the number of measurements far above average or far below average are about the same) then you can say the following about your data...

<u>68% of all sample means will fall within plus or minus 1 SE of the overall mean value</u>. (For example, if your overall average height was 5'9", and your SE was 2", 68 out of 100 means generated by randomly measuring 50 guys will be between 5'7" and 5'11).
<u>95% of all sample means will fall within plus or minus 2 SE of the overall mean value.</u> (For example, with your SE of 2", if your overall average height was 5'9", and your SE was 2", 95 out of 100 times, when you measure 50 guys and average their heights, your average will be between 5'5" and 6'1" tall).
<u>99.7% of all sample means will fall within plus or minus 3 SE of the overall mean value</u>. This means you can confidently say that based on this data, if you sample 50 guys and average their height, the average will fall between 5'3" and 6'3".

Standard errors of the mean are also useful if you want to know how likely two samples are to have different (or similar) means. Say you want to know if women really are on average shorter than men. If you generate a SE for a set of average women's heights just like you did for men, you can then compare average women's height to average men's height.

For example, say your average height for men was 5'9" with a SE of 2", and your average height for women was 5'7" with a SE of 2". Assuming a normal distribution of your data, this means that if you measure 50 women and average their heights, 68% of the time, the average height should fall between 5'5" and 5'9". If you do the same with men, the average should fall between 5'7" and 5'11". Notice that this means you can take samples of men or women where the average height will be around 5'7" to 5'8". Because the standard errors for your two sets of samples (men and women) overlap, you can't say that men are generally taller than women, because even though your 5'9' average for men seemed taller than 5'7' for women, there is a good chance that if you sample again, you may get a sample of men with an average height of 5'7", or a sample of women with an average height of 5'9".

A useful rule of thumb: If the SE is greater than 10% of your mean, your sample size is too small. This makes your mean unreliable, because 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.