# Week 5: Discussion

## David Diez

**Central Limit Theorem ::** This Theorem says the following:

> If we collect a sample from a distribution with mean $\mu$ and standard deviation $\sigma$ (which need not be normal), then if the sample is sufficiently large, the *sample mean* $\bar{x}$ will be approximately normally distributed with mean $\mu$ and standard deviation $\sigma/\sqrt{n}$, where $n$ is the sample size.

There are a few very important aspects of this Theorem:

- We expect the sample mean, $\bar{x}$, to be near the true mean, $\mu$.

- If $n$ is sufficiently large, then $\bar{x}$ is about normally distributed.

- The variation of $\bar{x}$ is given by $\sigma/\sqrt{n}$.

Let's assume for a moment that $n$ is "sufficiently large". Then $\bar{x}$ is just some normally distributed value. We expect it to be within the distance $\sigma/\sqrt{n}$ about 68% of the time, within $2\sigma/\sqrt{n}$ about 95% of the time, and within $3\sigma/\sqrt{n}$ about 99.7% of the time. We will begin to exploit these nice properties.

This of course all assumes we know $\sigma$. Later on we will take on the case where we don't know $\sigma$ (ie, we estimate it using our sample).

**Sample proportion ::** The sample proportion, $\hat{p}$, is similar to a sample mean and, if $n$ is sufficiently large, then we will say $\hat{p}$ is about normally distributed. It will have mean $p$ (the true proportion) and standard deviation $\sqrt{\frac{p(1-p)}{n}}$. So we can have the same properties with $\hat{p}$ as we do with $\bar{x}$ but we have a few conditions:

- $pn$ and $(1-p)n$ are each at least 10. If we are uncertain of the precise value of $p$, then if $\hat{p}n$ (the number of "successes" in the sample) and $(1-\hat{p})n$ (the number of "failures" in the sample) may be used as substitutes.

- The sample is a "simple random sample" (ie, completely random) from the population.

- The sample is no more than 10% of the population. (Otherwise it is no longer like the individuals in our sample are really independent.)

Can we ever continue if we don't have conditions? Well, yes, but it doesn't mean it will be accurate! If you have a problem with one of the 3 conditions above for a problem, make a note that the results may not be accurate but you may continue anyways.

**Quiz 5, attempt 1 review ::** Problems 3 and 4 will be omitted since these were discussed in lab.