Discussion Sheet 13
The Central Limit Theorem

We now study the most important Theorem that lets us make inferences about populations from samples based on the normal distribution and distributions related to it.

Random Sample
If a sample of \( n \) things is selected from a population in such a way that every set of \( n \) things in the population has an equal probability of being selected, then that sample is called a random sample.

1. Suppose that all of a doctor's patient records are stored in alphabetical order by patient's last name. If a sample of 200 by is selected by taking first 200 records, would that be likely to be a random sample with respect to whether the patient had arthritis?

2. Suppose instead that the doctor's records above were stored by age and then alphabetically by patient last name. If a sample of 200 is made by by taking first 200 records, would that be likely to be a random sample with respect to whether the patient had arthritis?

Sampling Distribution
Suppose that a large number of random samples in a certain situation is taken and if a particular statistic (e.g., the mean or the variance) is calculated for each sample taken. The set of statistics calculated in this way would have its own distribution, not necessarily that of the original population. That distribution would be called the sampling distribution of that statistic.

3. Suppose that a sample of 100 values is drawn from a population with a uniform distribution and the mean of that sample calculated. Suppose we did this for 5000 different samples of 100 values from that population. Would the sampling distribution represented by those 5000 different means have a uniform distribution? What can we know about this sampling distribution?

Standard Error and Mean of a Sampling Distribution
If we have the sampling distribution of a particular statistic, the mean of that sampling distribution is simply called the mean of the sampling distribution. The standard deviation of that sampling distribution is traditionally called the standard error of the sampling distribution.

4. When polls or surveys are reported on television they are usually labeled as accurate "plus or minus \( n\% \)." That percentage is the standard error for the result reported which is usually a proportion of the voters or a certain group that believe or will do a certain thing. Can we tell anything about this standard error if we know the mean and standard deviation of the original population (not of any sample drawn from it) but don't know what kind of distribution it has?
5. Look Figure 7.9 on pages 304 and 305 of your textbook. This shows the relative frequencies of the mean of 1000 samples of size 5, 15, 25, 50, and 100 drawn from a population with a uniform distribution from 0 to 1 (see Figure 7.8 a at the top of page 304). Relative frequency is an approximation to the sampling distribution of a statistic, in this case the mean. What can you say about the mean of the sampling distribution for the samples of different sizes (5, 15, etc.)? What can you say about the standard error of the sampling distribution for the samples of different sizes? What can you say about shape of the sampling distribution for the samples of different sizes? How do these three things appear to change as the size of a sample gets larger?

6. For the uniform distribution shown at the top of page 304, what is its mean? What is its standard deviation? How do these relate to the means and standard errors of the sampling distributions approximated by the relative frequency histograms shown for samples of different sizes?

7. Look Figure 7.11 on pages 306 to 308 of your textbook. This shows the relative frequencies of the mean of 1000 samples of size 5, 15, 25, 50, and 100 drawn from a population with a standard normal distribution (see Figure 7.10a at the top of page 306). What can you say about the mean of the sampling distribution for the samples of different sizes (5, 15, etc.)? What can you say about the standard error of the sampling distribution for the samples of different sizes? What can you say about shape of the sampling distribution for the samples of different sizes? How do these three things appear to change as the size of a sample gets larger?

8. For the normal distribution shown in Figure 7.10a, what is its mean? What is its standard deviation? How do these relate to the means and standard errors of the sampling distributions approximated by the relative frequency histograms shown for samples of different sizes?
9. Look at Figure 7.12 on pages 308 to 310 of your textbook. This shows the relative frequencies of the mean of 1000 samples of size 5, 15, 25, 50, and 100 drawn from a population with an exponential distribution (see Figure 7.10b at the top of page 306). What can you say about the mean of the sampling distribution for the samples of different sizes (5, 15, etc.)? What can you say about the standard error of the sampling distribution for the samples of different sizes? What can you say about shape of the sampling distribution for the samples of different sizes? How do these three things appear to change as the size of a sample gets larger?

10. Based on your answers to Questions #5 through #9, what would you say is true in general about the means of a sampling distribution of means regardless of the probability distribution of the original population? Of the standard error of the sampling distribution? Of the shape or type of the sampling distribution? How does this change in general as the samples involved in calculating the statistic get larger?

### Central Limit Theorem

If a random sample of $n$ observations, $y_1, y_2, \ldots, y_n$ is drawn from a population for a random variable $Y$ with a finite mean $\mu$ and variance $\sigma^2$ then $\bar{y}$ is normally distributed with mean $\mu$ and variance $\frac{\sigma^2}{n}$ for sufficiently large $n$ regardless of the distribution of the original random variable $Y$.

Note: "Sufficiently large" turns out to be about $n \geq 30$.

11. Based on the Central Limit Theorem, how would you answer the basic questions posed in Question 10?