The Central Limit Theorem and Estimation

If we select many, many samples (each of size $n$) from a parent population with mean $\mu$ and standard deviation $\sigma$ and compute each of the sample means, this collection of sample means (called the sampling distribution of sample means – the SDSM) will be:

1. **normally distributed**, if the parent population is normally distributed. The mean of the SDSM is given by $\mu_\tau = \mu$ and the standard deviation of the SDSM is given by $\sigma_\tau = \frac{\sigma}{\sqrt{n}}$.

   OR

2. **approximately normally distributed**, if the parent population is not normally distributed (provided $n \geq 30$). The mean of the SDSM is given by $\mu_\tau = \mu$ and the standard deviation of the SDSM is given by $\sigma_\tau = \frac{\sigma}{\sqrt{n}}$. As $n$ increases beyond 30, the approximation improves.

**Note:** in each of the possible scenarios above, the parameter $\sigma_\tau$ is known as the *standard error of the mean*, or the S.E. mean (in class we called $\sigma_\tau$ the standard deviation of the SDSM).

So, assuming that we know both $\mu$ and $\sigma$ (actually somewhat unrealistic in real life) we can calculate probabilities of sample means having certain values. For example, see Class Practices on Sampling Distributions #1 and #2.

However, what is really *more useful* to us is the technique of interval estimation which becomes possible due to the Central Limit Theorem.

Now let us *assume that we do not know* the mean $\mu$ of a parent population (finally a realistic assumption) and that we *do know* the standard deviation $\sigma$ of that same population (unrealistic, but let’s go with it for now). In order to determine an interval estimate for the population mean $\mu$, we select a sample of size $n$ and compute its mean $\bar{x}$. This $\bar{x}$ becomes a point estimate for $\mu$. We then build the interval estimate by first computing the error of the estimate $E$ which is given by the formula $E = \frac{z\sigma}{\sqrt{n}}$. The $z$ value, also denoted by the notation $z(\alpha/2)$, in this formula is fixed or determined by knowing the level of confidence that is specified in the problem. The level of confidence is then considered to be a probability, or equivalently, an area that is centered under the normal distribution. This allows us to find $z$ by using the table on page 810. Once $E$ has been determined we then compute both the

$LCL = \text{lower confidence limit} = \bar{x} - E$ and
UCL = upper confidence limit $= \bar{x} + E$. 

Summary of significant vocabulary associated with Interval Estimation of $\mu$:

**parent population** – the population that we wish to know something about; also the population from which we select a random sample of size $n$

**SDSM** or sampling distribution of sample means – this is the collection of sample means that would be obtained if we were to select *many* samples from a parent population and compute the mean ($\bar{x}$) of each sample

$\mu_{\bar{x}} = \mu$ is the mean of the SDSM

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ is the standard deviation of the SDSM

**confidence coefficient** – this is the $z$ value used in the formula $E = \frac{z\sigma}{\sqrt{n}}$; when the level of confidence is $1 - \alpha$, there will be $\alpha/2$ area in each tail of the SDSM and because of that, we sometimes use the symbol $z(\alpha/2)$

**lower confidence limit (LCL)** – this is a lower bound on the population mean $\mu$

**upper confidence limit (UCL)** – this is an upper bound on the population mean $\mu$

**level of confidence** – a probability that is selected before the two confidence limits are computed; typical levels of confidence are 90%, 95%, and 99%