# LAB 3 INSTRUCTIONS

# SAMPLING DISTRIBUTIONS

In the lab instructions we will review the basic properties of the sampling distributions of a sample mean. In particular, we will use Excel to demonstrate the Central Limit Theorem. Note that some techniques required in lab 3 assignment are discussed in *Lab 2 Instructions*, in particular, random number generation, and Poisson and normal distribution probabilities.

## 1. Sampling Distribution of a Sample Mean

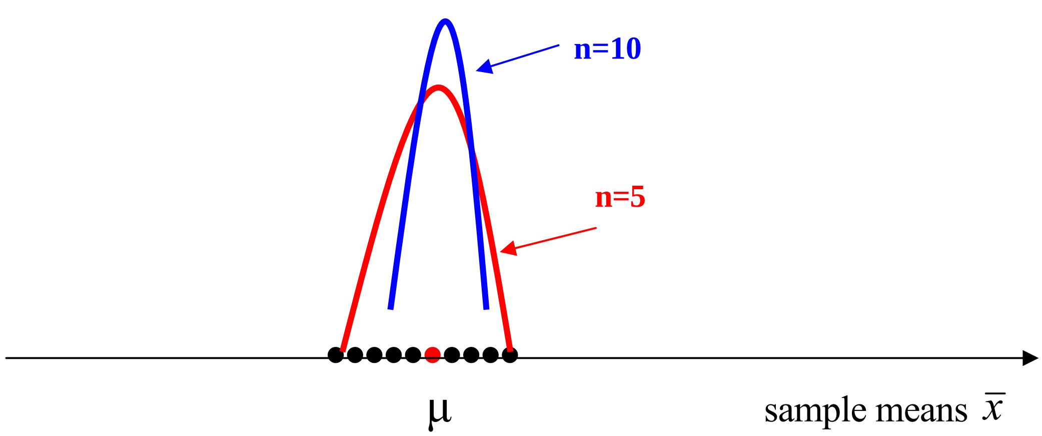
A statistic is any function of the observations in a random sample. In particular, sample mean or sample proportion are examples of statistics. A sampling distribution is the probability distribution of a given statistic based on a random sample of size $n$. It may be considered as the distribution of the statistic for *all possible samples* of a given size. The sampling distribution depends on the underlying distribution of the population (also called the parent population), the statistic being considered, and the sample size used.

For example, consider a normal population with mean $\mu$ and variance $\sigma^2$. Assume we repeatedly take random samples of a given size from this population and obtain the average of observations for each sample. The distribution of these averages is called the sampling distribution of the sample mean.

If a population has the mean $\mu$ and the standard deviation $\sigma$, then the sampling distribution of the sample mean $\bar{x}$ has the mean $\mu_{\bar{x}}$ and the standard deviation $\sigma_{\bar{x}}$ defined by the formulas:

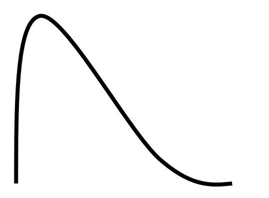
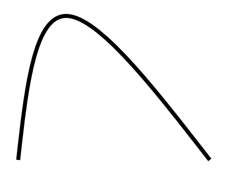
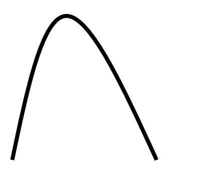
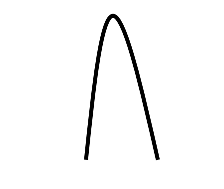
$$\mu_{\bar{x}} = \mu \qquad \text{and} \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}.$$

If the population from which the samples are drawn (parent population) is normal, then the distribution of the sample means will be normal regardless of the sample size. Based on the above equations, the sample means are centered at the population mean for any sample size and the spread of the sampling distribution of the sample mean decreases as the sample size increases.

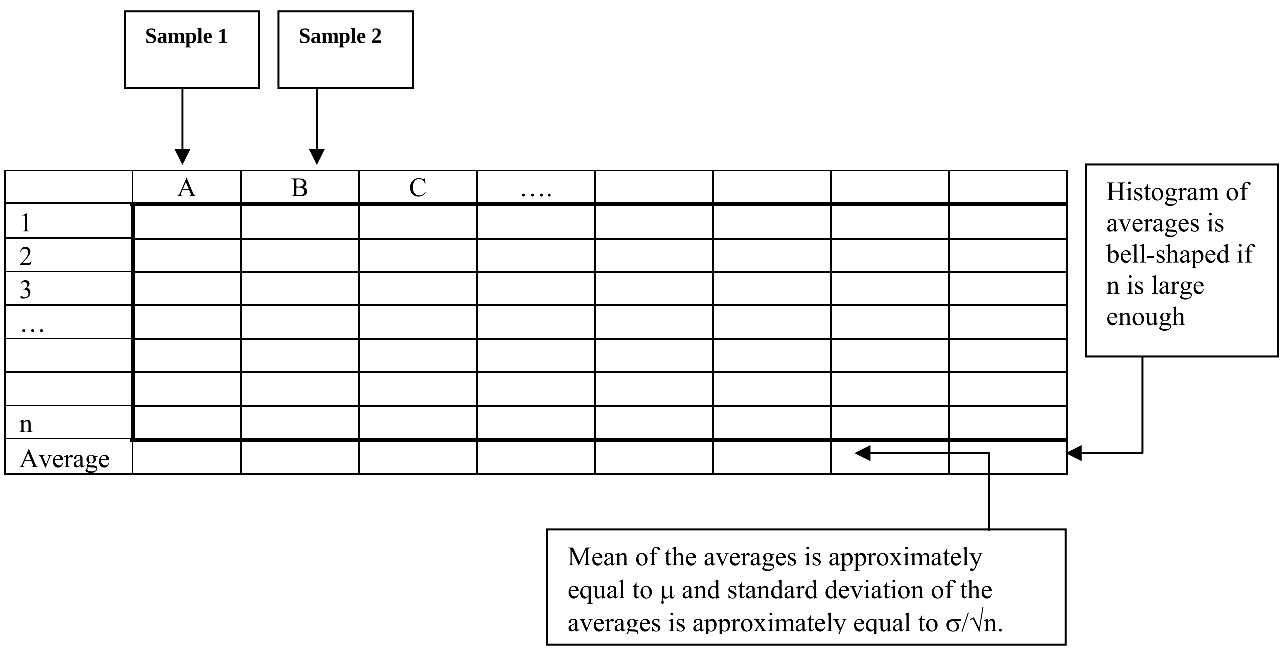


**2. The Central Limit Theorem**

The Central Limit Theorem states that if the sample size is large enough ($n \geq 30$), then the distribution of the means of those samples (the sampling distribution of the sample mean) is approximately normal regardless of the population distribution. The larger the sample size n, the better the normal approximation.

| **Parent Population** | **Sampling Distribution of $\bar{x}$** | | |
|---|---|---|---|
| | **n=5** | **n=10** | **n=30** |
|  |  |  |  |

The Central Limit Theorem can be demonstrated in Excel by obtaining first a large number of samples of size n from the population, where n is large enough. Each sample corresponds to one column in an Excel worksheet and consists of n observations (n rows in the worksheet). The samples can be easily obtained by using the *Random Number Generation* feature (see *Lab 2 Instructions*). Then the averages for each column (sample) are calculated using the *Insert Function*. This is demonstrated in the drawing on the next page.




If the sample size n is large enough, the histogram of the averages should resemble a normal curve. The *Insert Function* applied to the last row in the table allows you to calculate the mean and standard deviation of the averages. The law of large numbers states that the actually observed sample mean of a large number of observations must approach the mean of the population.