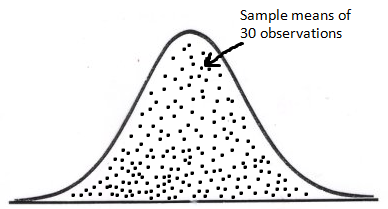
Reason for using a Sample Size of 30

*By: Wen Cai*

Statistical inference allows the researcher uses a sample to estimate a given statistic of a population. With a “large” sample size, the researcher can make inferences about the population regardless the size and shape of population distribution. The reason: a sample of 30 is considered a “large” based on the criterion called the central limit theorem.

The central limit theorem says that as the sample size increases, the frequency of the sample means, when plotted, will be normally distributed. Therefore, the sample statistics can be reflective of the population statistics. To illustrate this, if researcher takes infinitely repeated samples of 30 observations and plotted their means, the resulting distribution will be a normal curve. The center of this curve will have a mean that is equal to the population mean. Therefore, when the researcher takes one sample of 30 observations, the mean of that sample will fall somewhere in the center, very close to the population mean.



Infinitely repeated sampling of 30 observations, the mean of the sample distribution equals to the population mean.

µ

Now the question is: why are observations 30 or greater considered “large”? The reason is the resulting distribution approaches a normal curve at about 30. With infinitely repeated samples of observations less than 30, the resulting curve will still remain bell shaped, but the distribution will be wider, with a larger standard deviation. As the size of the sample increases, the curve will become skinner and look closer to a normal distribution. By about 30, the distribution will approach the normal curve; which is why a sample of 30 is generally considered a sufficiently large size for statistical inference.

