Central Limit Theorem

The Central Limit Theorem (CLT) is perhaps the most important theorem in statistics. It allows you to apply a common approach to problems that otherwise appear to be very different. In words, what it says is that even though you are measuring a variable that follows a probability distribution that may be quite unusual, the sample average does not follow the same distribution. In fact, in most situations, repeated samples will yield a collection of averages that, when plotted, will follow a normal distribution.

The sampling distribution of any quantity, say the sample average, is that conceptual distribution that would arise if one actually took all possible samples from a population, computed the average of each sample and then plotted the results. In practice, this does not happen since you usually only take one sample, but as stated above, the CLT allows you to understand what would happen without having to do it.

The use of the Central Limit Theorem requires some assumptions, most of which are technical and often accepted quickly. The one that is not, however, is that the CLT requires the sample to be large enough for the averaging to counteract the shape of the original distribution. How large that needs to be depends upon what is being measured, but in many cases it is felt that a sample of size 30 or more is sufficient.

Under the assumptions mentioned above, the CLT states that if you take a sufficiently large <u>sample of size n</u> from a population with a <u>mean of μ</u> and a <u>variance of σ^2</u> then the sampling distribution of the sample average, $\bar{\mathbf{X}}$, follows a normal distribution, where the mean of the sampling distribution equals the mean of the original observations and the variance of the sampling distribution equals the variance of the original observations divided by n.

This is often abbreviated as $\bar{\mathbf{X}} \sim \mathbf{N}(\mu, \sigma^2/n)$.