Assignment 1 – Central Limit Theorem

Introduction:

The central limit theorem (CLT) has been described as “the most important theorem in statistics”. While this is an overstatement, it is true that much of statistical inference relies upon the central limit theorem. The CLT in a nutshell states that if random samples are drawn from a large population with mean μ and variance σ^2, then the sampling distribution of the mean is approximately normally distributed with mean $\mu_{\bar{X}} = \mu$ and standard deviation $\sigma_{\bar{X}} = \sigma / \sqrt{n}$; hence,

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

is the value of a standard normal variable Z. What is equally important is that this is true regardless of the distribution of the population sampled. We usually don't know the population variance, so we estimate it with s^2; for $n \geq 30$ this is usually sufficient, but for $n < 30$ we use the *Student-t* distribution.

Objectives:

Using simulation, we will demonstrate some properties and implications of the CLT.

Instructions:

Using the software of your choice, generate 1000 random samples of size 50 from a normal distribution with mean 30 and standard deviation 9. Plot a histogram of the sampling distribution of the means. Calculate a 95% confidence interval for the mean for each sample. What proportion of the confidence intervals include the true mean?

What is the 95^{th} percentile for the sample means? Transform this value to a z-score. Using a Z table or a built-in function for the normal distribution in your software application, find P(Z>z) for the z-score you calculated. Does the probability meet your expectation?

Repeat all of the steps above, except generate samples from something other than a normal distribution. Try a continuous uniform on the interval {0,1}, for example.

Product:

Submit your answers in professional memo format. This assignment is due in two weeks, at the beginning of class on Wednesday January 26, 2005.