Homework No. 2 – Basic Statistics by Simulation

Objectives:

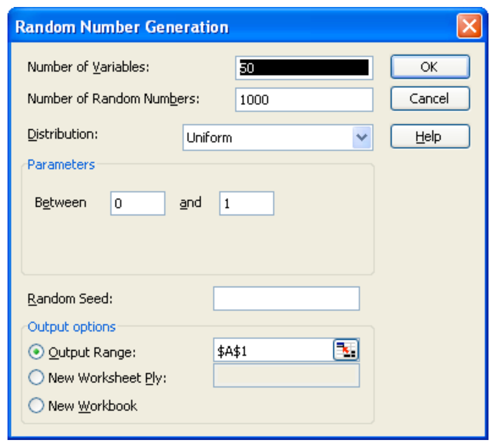
In this assignment we will work with a population that follows the continuous **uniform** distribution on the interval {0,1}. The objective is to prove empirically some statements about the Central Limit Theorem. We'll create a set of 1000 hypothetical samples, and look at the sampling distribution of the mean.

Instructions:

The Central Limit Theorem (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 normal with mean $\mu_{\overline{X}} = \mu$ and standard deviation $\sigma_{\overline{X}} = \sigma/\sqrt{n}$, with μ and σ from the original population. What is equally important is that this is true regardless of the distribution of the population sampled. We usually don't know σ, so we estimate it with s.

Creating the hypothetical set of samples

In Excel, use the random number generation tool to generate 1000 pseudo-random samples of size 50 from this population. The Analysis ToolPak will try to generate the samples in columns, but there are not 1000 columns in a spreadsheet, so if you try to do this it will fail. It is actually easier to generate "50 samples" of "size 1000" and just look at them the other way around. See the example at right.



You'll get 1000 **rows** that span 100 **columns**. Treat each row as if it is a sample. Be careful.

Analysis procedure

1. Construct a confidence interval around each of the 1000 sample means as if they were independent samples. What proportion of the confidence intervals contain the population mean?
2. Calculate the mean for each of the 1000 samples. Plot a histogram showing the distribution of the means. This is the sampling distribution of the mean. Does it look like it follows a normal distribution?
3. Find the mean of the 1000 means. How does this compare to the population mean for a continuous uniform?
4. Find $s_{\overline{X}}$, the sample standard deviation of the 1000 means. Recall this is called the "standard error of the mean". Using $s_{\overline{X}}$ as our best estimate of $\sigma_{\overline{X}}$, solve for the population standard deviation. How does this compare to the population standard deviation for the continuous uniform?

Note that a continuous uniform distribution has:

$$\mu = \frac{b-a}{2}, \text{ and } \sigma^2 = \frac{(b-a)^2}{12}.$$

Product:

Summarize your results, including any tables or figures, in a professional memo no longer than one double-sided page. Include a very concise description of your procedure.

You must put your Lab Section on your memo and address the memo to your TA.

Due Date:

This assignment is due at the beginning of class on Monday February 2, 2009.