

Activity: Sampling Distributions and Introduction to the Central Limit Theorem

Concepts: Random samples from populations, parameters, statistics, sampling distributions, empirical sampling distributions, the Central Limit Theorem.

Prerequisites: The student should be familiar with random variables, distributions (probability and empirical probability), expected value, and statistics such as the sample mean and sample variance.

Recap: In our last class we used simulation (the software package *Sampling Sim*) to examine the sampling distribution for the sample mean statistic \bar{X} . First we saw that the sample mean \bar{X} is a random variable. Our investigation of the empirical probability distribution of \bar{X} by taking many samples of the same size, n , from the same population resulted in the following observations about the sampling distribution of \bar{X} :

Population Parameters: mean = μ , standard deviation = σ ;

Sample Statistics: mean = \bar{x} , standard deviation = s

Observations about the sampling distribution of \bar{X} :

- shape: Bell shaped (i.e. normal shaped) distribution for “large enough” sample sizes, n .
- center: Distribution of \bar{X} centered at the population mean μ .
- spread: Spread of \bar{X} depends on sample size, n . Spread decreases as n increases (actually spread is σ/\sqrt{n})

Our simulation results to compute the sampling distributions for the sample mean statistic \bar{X} illustrated the **Central Limit Theorem**. This theorem says the following about the sampling distribution of the sample mean \bar{X} :

- The mean of the sampling distribution of \bar{X} equals the population mean μ , regardless of the sample size or the population distribution, i.e. $E[\bar{X}] = \mu$.
- The standard deviation of the sampling distribution of \bar{X} equals the population standard deviation σ divided by the square root of the sample size, regardless of the population distribution, i.e. $Var[\bar{X}] = \sigma^2 / n$.
- As the sample size gets larger, the shape of the sampling distribution of \bar{X} approaches a normal distribution (i.e. it is approximately normal for “large” sample sizes), regardless of the population distribution, and it IS normal for ANY sample size when the population distribution is normal.

In this take-home activity, you will run the Sampling Sim program to investigate the sampling distribution for \bar{X} and thus see the Central Limit Theorem in action. The first part of this activity takes you through how to use the Sampling Sim program and reviews some basic concepts along the way (parts (a)-(i)). Once you know how to use the simulation and understand what information it is providing, please hand in parts (j)-(q).

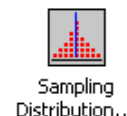
To download *Sampling Sim*, go to the website: http://www.gen.umn.edu/research/stat_tools/ then click on the Software button and download the proper compressed file for your machine (zip for windows machines). I will also have this software installed on the computer labs in Madison Hall and the MLC.

Scenario: Professor Lectures Overtime

Let X = amount of time a professor lectures after class should have ended. Suppose these times follow a Normal distribution with mean $\mu = 5$ min and standard dev $\sigma = 1.804$ min.

- Draw a rough sketch (and label) this distribution.
- Is μ a parameter or a statistic?
- Suppose you record these times for 5 days x_1, x_2, \dots, x_5 and calculate the sample mean \bar{x} . Is \bar{x} a parameter or a statistic?

To investigate the sampling distribution of these \bar{x} values, we will take many samples from this population and calculate the \bar{x} value for each sample. Open the program *Sampling SIM* by double clicking on its icon.



- Click the DISTRIBUTION button and select “Normal” from the list. You should see a sketch similar to what you drew in (a).
 - From the Window menu, select “Samples.”
 - Click Draw Samples and one observation from the population is selected at random (Note: the program may be very slow the first couple of times you click this button). This is one realization of the random variable X .
- How long did the professor run over this time?
 - Click Draw Samples again, did you observe the same time?
 - Change the value in the Sample Size box from 1 to 5 and click Draw Samples. How does this distribution compare (roughly) to the population distribution?
 - Click Draw Samples again. Did the distribution of your 5 sample values change?
 - Change the sample size from 5 to 25 and click Draw Samples. Describe how this distribution differs from the ones in (f) and (g). How does the shape, center, and spread of this distribution compare to that of the population (roughly)? (The mean of this distribution is represented by \bar{x} , the standard deviation of this distribution is represented by s . Compare these values to μ and σ .)
 - Click Draw Samples again. Did you get the same distribution? The same \bar{x} and s values?

The main point here is that results vary from sample to sample. In particular, statistics such as \bar{x} and s change from sample to sample. You will now look at the distribution of these statistics.

From the Windows menu, select Sampling Distribution. Move this window to the right so you can see all three windows at once. You should see one green dot in this window (it will be small and on the x-axis). This is the \bar{x} value from the sample you generated in (i). In the Sampling Distribution window, click on “New Series” so it reads “Add More.” Click the Draw Samples button. A new sample appears in the Sample Window and a second green dot appears in the Sampling Distribution window for this new sample mean. Click the Draw Samples button until you have 10 sample means displayed in the Sampling Distribution window. Note: You can click the F button in the Samples Window to speed up the animation. Record the values displayed in the “Mean of Sample Means” box and in the “Standard Dev. of Sample Means” box. These values are empirical. Compare these to the theoretical values predicted by the Central Limit Theorem.

	Mean of	Standard Dev. of
Sample Means	<input type="text"/>	<input type="text"/>

Mean of Sample Means _____

Standard Dev. of Sample Means _____

Be very clear you understand what these numbers represent. If not, ask your instructor!

(j) In the Population window, click on NORMAL to change the population to one of the following: Bimodal, Skew-, Skew+, Trimodal, U-Shaped, Uniform. Note, this changes the population mean μ and standard deviation σ as well. Please indicate which population distribution you are using and also the population mean and standard deviation.

(k) Now change Sample Size to 1 and number of samples to 500 (you definitely want to make sure you have the F button pressed in your Samples window to speed up the animation!). Click the Draw Samples button. Record the following information:

- Describe the shape, center (the mean of the sample means), and spread (the standard deviation of the sample means) of the Sampling Distribution of the \bar{x} values. In particular, how do the shape, center, and spread compare to the population distribution? You can click the purple population outline (upper left corner of Sampling Distribution window) for easier visual comparison.
- Now click the blue normal outline in the Sampling Distribution window. Which outline (population or normal) appears to be a better description of the sampling distribution of the sample mean \bar{x} values?

(l) Change the sample size to 5 (keep number of samples at 500) and click the Draw Samples button. Give the information asked for in part (k).

(m) Change the sample size to 25, click the Draw Samples button, and answer the same questions. Give the information asked for in part (k).

(n) Change the sample size to 50, click the Draw Samples button, and answer the same questions. Give the information asked for in part (k).

(o) Complete the table below. Are the theoretical values predicted by the Central Limit Theorem (CLT) close to the empirical values you got when you ran the simulations above?

Sample Size (n)	Population Mean	Empirical Mean of Sample Means	Theoretical Mean of Sample Means (via the CLT)	Population Standard Deviation	Empirical Standard Deviation of Sample Means	Theoretical Standard Dev. Of Sample Means (via the CLT)
1						
5						
25						
50						

(p) Repeat parts (j)-(p) for another non-normal population. Clearly indicate which population you use!

(q) Briefly summarize your results in terms of the Central Limit Theorem.