Exploring Sampling Distributions

A **sampling distribution** of a statistic is the distribution of values taken by the statistic in all possible samples of the **same size** from the **same population**. The reason we look at a sampling distribution is because we want to answer the question, "What would happen if a sample of this size was taken many, many times?" Would we be able to reach any conclusions about the population? As long as data is produced by random sampling or randomized experimentation, we can comfortably use inference to make conclusions about the population.

1. Open the .tns document, **Sampling Distributions** from the Statistics folder.

2. Go to the third page by pressing ctrl ▶ to get to the spreadsheet. The first column, named **popdata**, contains data of a population. Insert a Data & Statistics page and graph this variable. Insert a Calculator page on, *Calculator* to find the mean and standard deviation.

Describe this distribution. Calculate the mean and standard deviation of <u>popdata</u>.

3. Using a sample size of *n*=4, let's examine the distribution of the $\bar{x}$'s from 50 random samples of size, 4. To do this we will randomly sample 4 values with replacement from **popdata** and calculate the mean.

4. In the *Lists & Spreadsheet*, arrow to cell **f1** define a variable **formean** by typing the following→ **formean**:=**mean**(randSamp(**popdata**,4,0)). In the spreadsheet, there is a column **fourmeans** that has been set up to "capture" **formean** every time you press ctrl R. Press ctrl R about 50 times.

NOTE: When you have put 50 values in the list, be sure to delete the "0" in the first cell. To do this, arrow to the first cell that contains **0** and press ctrl menu. Select *Delete Cell*. When you are prompted to "Do you want to overwrite…" select *Yes*. Only the **0** should be deleted.

5. Examine the values in **fourmeans** and insert another Data & Statistics page. Graph the data in a histogram.

Describe this histogram and provide a labeled sketch in the space below.

6. Insert another Calculator page, ctrl I, *Add Calculator*.

Calculate the mean and standard deviation of <u>fourmeans</u>.

How do these values compare with the mean and standard deviation of <u>popdata</u>?

7. Now choose a sample size of 9. In the *Lists & Spreadsheet* define a second variable in cell **f2** as **ninmean**:= **mean(**randSamp(**popdata**,9,0)**)**. Press ctrl R 50 times and these sample means will be stored in another column named **ninemeans**. Be sure to delete the *0* (see note on previous page).

8. Insert a Data & Statistics page and select **ninemeans** as the horizontal axis.

Describe this histogram and provide a labeled sketch of it in the space below.

Calculate the mean and standard deviation.

How do these values compare with the mean and standard deviation of the population, <u>popdata</u>?

How do they compare with the mean and standard deviation of <u>ninemeans</u>?

9. Let's examine another value of *n*. For *n*=16, repeat the above procedure in #7 and #8, using the variable **sxtnmean.** The command to type in on the calculator page is **sxtnmean**:=**mean(**randSamp(**popdata**,16,0)**)**. These values will be stored in **sixtnmeans.**

Compare this histogram to the previous two histograms and provide a labeled sketch of it in the space below.

Calculate the mean and standard deviation.

How do these values compare with the mean and standard deviation of the population, <u>popdata</u>?

How do they compare with the mean and standard deviation of <u>fourmeans</u> and <u>ninemeans</u>?

10. Let's examine one more value of *n*. For *n*=25, repeat the procedure in #7 and #8, using the variable **twtfvmean.** The command to type in cell **f4** is **twtfvmean**:=**mean(**randSamp(**popdata**,25,0)**)**. These values will be stored in **twtyfivemeans.**

Compare this histogram to the previous 3 graphs and provide a labeled sketch of it in the space below.

Calculate the mean and standard deviation.

How do these values compare with the mean and standard deviation of the population, popdata?

Looking at the means and standard deviations for *n*=25, *n* =16, *n*=9, and *n*=4. Do you notice any patterns or trends?

For the following sample sizes, *n*, list what you believe would be the mean and standard deviation of the $\bar{x}$'s from popdata, IF you were to generate them.(Do not use the calculator for this!).

1. ***n*=50**

2. ***n*=90**

3. ***n*=100**

For ANY sample size, *n*, what would be a general formula for calculating the mean and standard deviation of $\bar{x}$'s for popdata?

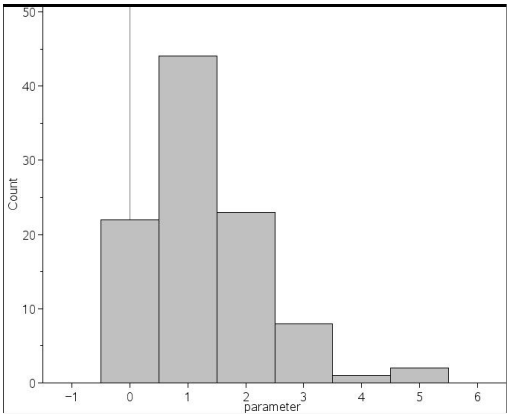
Using a couple of sentences, describe what you have learned in this activity about the sampling distribution of a mean.

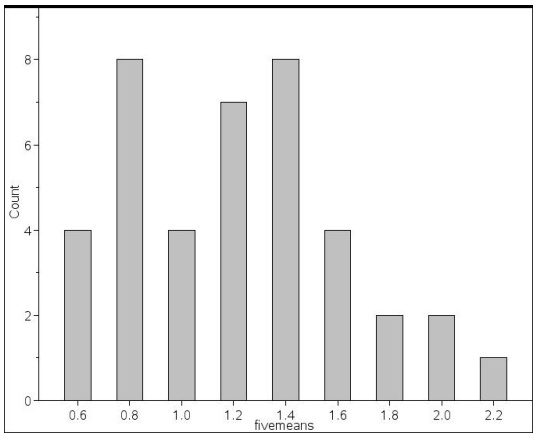
Here are the screens for this activity. NOTE: The students can use the counter at the bottom of the Calculator page to determine when they have reached 50 values.

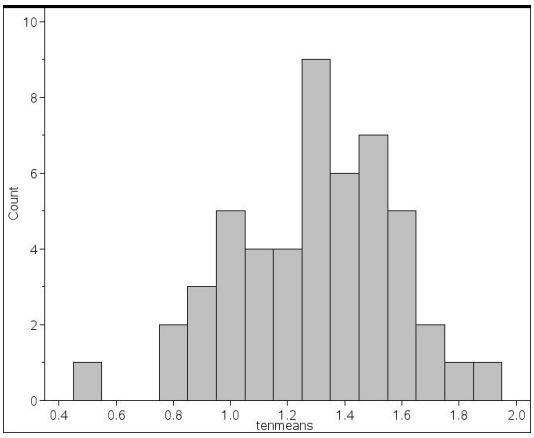
Sampling Distributions

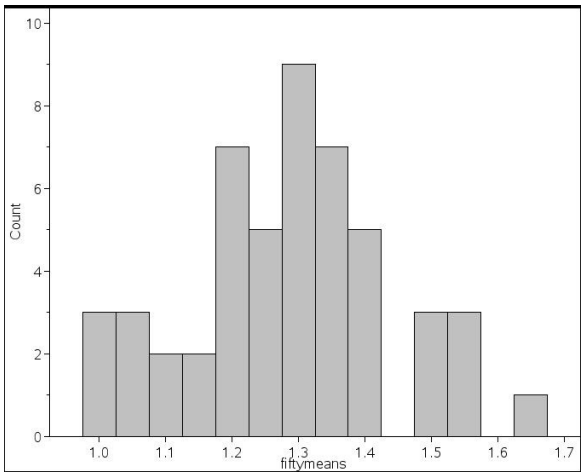
The **sampling distribution** of a statistic is the distribution of values taken by the statistic in all possible samples of the same size from the same population. On the page, a "population" of data has been created. We will sample this data using different values of *n* and look at what happens when we graph the means.

	A parameter	B fivemeans	C tenmeans	D fiftymeans	E
◆		=capture('smeans,1)	=capture(tenmean,1)	=capture(fiftymean,1)	
1	2	6/5	13/10	3/2	
2	1	1	3/2	32/25	
3	2	4/5	9/10	39/25	
4	2	3/5	17/10	32/25	
5	1	7/5	3/2	69/50	
6	1	4/5	11/10	51/50	
7	1	6/5	19/10	59/50	
8	3	7/5	7/5	77/50	
9	1	8/5	3/2	33/25	
10	0	7/5	8/5	37/25	
11	1	4/5	3/2	32/25	
12	2	1	7/5	63/50	
13	1	7/5	13/10	71/50	
14	0	11/5	7/5	26/25	
15	3	6/5	6/5	34/25	
16	1	8/5	13/10	59/50	
17	1	2	6/5	7/5	
A	parameter				









stDevPop(*parameter*)	1.04957
mean(*parameter*)	1.28
stDevSamp(*fivemeans*)	0.418514
stDevSamp(*tenmeans*)	0.285293
stDevSamp(*fiftymeans*)	0.15194
stDevPop(*parameter*)	1.04957
$\frac{1.0495713410722}{\sqrt{5}}$	0.469383
$\frac{1.0495713410722}{\sqrt{10}}$	0.331904
$\frac{1.0495713410722}{\sqrt{50}}$	0.148432
mean(*fivemeans*)	1.215
mean(*tenmeans*)	1.294
mean(*fiftymeans*)	1.2828

12/99