Exploring the Central Limit Theorem

You will now take a closer look at what happens when we graph the means of sampling distributions of size *n*, and what happens to the *mean of the means*. By the end of this activity, you should be able to make a general statement about the *mean of the means* as *n* becomes large.

1. You will be working with a partner for this activity. Each partner should count out 50 pennies from the "population" of pennies in the bucket. Be sure you EACH have n=50 pennies.

2. Open a New Document in your handheld and add a Lists & Spreadsheet page. Record the age of each penny (this year **minus** the year on the penny) in Column A. One partner will need to name their column A, **fiftyage** and the other partner will to name their column A, **fiftypenny**.

3. Look at the distribution of *your* 50 pennies using a histogram in the Data & Statistics page.

Describe the distribution.

Make a guess for the following (for *your* 50 pennies):

Mean age:_____ **Median:**_____ **Standard Deviation:**_____

Mean Age of *all the pennies* in the bucket: _____

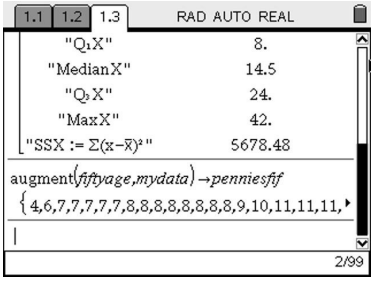
4. Insert a calculator page a produce the one-variable statistics for **fiftyage**.

State the following:

Mean age: _____ **Standard deviation:** _____ **Median Age:** _____

5. To send your **fiftyage** data to your partner you will need to do the following:
   a. Save your document as **CLT1** and have your partner save it as **CLT2**.
   b. Get a link cable and link your handheld to your partner's. You will need to send your document to your partner and vice versa. (doc▾), *File*, *Send*. This will send the current document to your partner's handheld. Repeat to get your partner's file.
   c. Within your open document arrow up until your **fiftyage,** or **fiftypenny** is highlighted. Press (ctrl) [C] to copy this column.

d. Open your partner's document, either **CLT1** or **CLT2.** (Choose to save and press (enter).
e. Go to the Spreadsheet and in Column B, named **mydata**, press (ctrl) [V] and your data should appear.
f. Now, press (ctrl) ▶ until you reach the Calculator page. Type **augment(<partner's column>, mydata)** (var) **penniesfif.**
g. In your spreadsheet page, name your column C, **penniesfif** and the one hundred penny ages will fill the column.



6. Create a histogram of your data (n=100) in a separate Data & Statistics page.

Describe the distribution, using any calculations that are appropriate. Note any interesting observations. Be thorough but concise.

We will put all of our data together to examine the group of *all pennies*. Physically, that means we must put all the pennies into a single bucket. Graphically that means we need to find out how many pennies there are of each age, record this data, and display it in a single histogram. Once we draw this histogram, we can compare its center, shape, and spread to that of the means of the samples we will draw. Describe the shape of the histogram of all the pennies.

7. Save your document (you only need to do this for one handheld within each partnership) as **CLT<your intials>**.

8. Go to the teacher's computer and link your handheld. I will drag-drop your document to create the class histogram.

9. Next, you will draw samples of varying sizes (n=9, n=16, n=25, n=50) and compute the mean age of the pennies in each sample. For the n=50 pennies sample, use your original data. You will be working with your partner to find these sample means. Record the following information.

n=9 mean:_____ n=16 mean: _____ n=25 mean: _____ n=50 mean: _____

We will all list and plot the results of our samples on the corresponding histogram of the total collection of pennies on the board. Be sure to plot your sample results on the histogram by using 4 different colors/symbols depending on the sample size: for n=9 use

a light color and shade the box; for n=16 use a bright color and draw a circle; for n=25 use a contrasting color and mark a dot in each box; for n=50 use your final color and draw a diamond in the box. You should see all 4 distributions superimposed on the original.

10. At this point you should set up 4 lists that will contain the means of everyone's sampling results. Each list will be for a given sample size (n=9, 16, 25, 50 respectively). After you enter all the means, you can run 1-VAR STATS on each list. This will give you the mean and standard deviation of each of the sampling distributions. Be sure you can identify the statistics for each sampling distribution. Compare your results with hose of your partner to be sure you have analyzed the data correctly. These 8 values are very important. Use these values to answer the following questions.

11. Once the results of the samples are listed, compute the ***mean of the means*** for each sample size.

    (n=9) $\mu_{\bar{x}}$ = _____ (n=16) $\mu_{\bar{x}}$ = _____ (n=25) $\mu_{\bar{x}}$ = _____ (n=50) $\mu_{\bar{x}}$ = _____

12. Now calculate the ***standard deviation of the means*** we have just collected. We call this value the standard error of the mean, $\sigma_{\bar{x}}$, for it describes the spread or variability of the sampling distribution of the sample means. Since we gathered samples in four sizes, we need to calculate the standard errors of each of those four sampling distributions.

    (n=9) $\sigma_{\bar{x}}$ = _____ (n=16) $\sigma_{\bar{x}}$ = _____ (n=25) $\sigma_{\bar{x}}$ = _____ (n=50) $\sigma_{\bar{x}}$ = _____

What we are seeing is the ***Central Limit Theorem*** revealing itself. Pierre Simone LaPlace formalized The Central Limit Theorem. You need to state it in easy to understand English (informal yet accurately and clearly stated, ***based on our simulation***. Use another sheet of paper. You may look up the CLT in the text if you feel that it is necessary. Most importantly, explain what the CLT means in the context of our data gathering exercise and simulation. Be sure to discuss the center, shape, and spread of the 4 sampling distributions, the role of sample size, and the relationship of the sampling distribution statistics ($\mu_{\bar{x}}$ and $\sigma_{\bar{x}}$) to the parameters (μ_x and σ_x). Think of the CLT with the *n* data you gathered. We usually gather just one sample – which one do you think you will get??