Monte Carlo Explorations
Sampling Distribution

An Aside on Study Habits

I've setup this worksheet with places for you to stop and write out answers to check whether or not you understand the concept being introduced. For you to be an effective student, you must learn to do that for yourself. You have to check your comprehension. You must develop "metamemory savvy" to know how your memory system works and the special steps you have to take to cause learning to occur. Most often, you will be responsible for providing your own comprehension checks. I'll remind you every now and again. Remember: it's your learning, not mine.

What you will be doing

In theory you are going to be reaching into a big trash can with 5000 poker chips--each chip with a number written on it. You will be drawing a sample of chips and recording the values written on each poker chip. You will calculate the mean and standard deviation of that sample. Then you will draw additional samples, calculate the same statistics, and be guided to discover how the value of the mean of the samples changes in a lawful manner.

In practice, you'll be interacting with a computer program. The software will assume the burden of drawing the samples and making the calculations. You will be free to focus on the significance and meaning of the numbers that are calculated.

It is important that you stay awake. Don't go to sleep and fall into a pattern of trying to get done just to be done. (The terms "awake" and "asleep" are metaphorical, of course).

When you finish the tasks, form a small group with other students who have finished and review the tasks. Compare your results to theirs. Compare your answers to the questions to theirs.

To Begin

Launch the "**Statistics Tool**".
- Make sure you are using the version with the correct "Last Modified" data.
- Mouse down on "**Laboratory**" button in the "**Sampling Distributions**" block.
- Make sure the "**Task**" pull down menu is set for "***Student Draws Sample***".
- Click on "**Launch Histogramer**. This opens a second window
- Click the "**Plot Data in Histogram**" button to go to the histogram screen.
- By clicking on "**Next**" and "**Prev**" you can go to/from the Frequency Distribution Table.

First Task

The white field in the lower-left corner of the screen contains 5,000 numbers that form a $N(28,5)$ population.

Comprehension Check #1*: If you were to sort the 5,000 numbers in the enumerated population. How many of them would be between 18 and 28? How many of them would be between 20 and 28**? It is important to write out your answers to each comprehension check. Compare your answer with a colleague before continuing.*

To be sure the N(28,5) population hasn't been inadvertently altered, plot the 5,000 numbers in a Histogram.

     • Do this by copying the population of 5,000 numbers to the clipboard (click on the "paperclip" icon), click on the "Histogramer" application, Go to the "Plot Card" and paste the data by clicking on the "Paste" button. Then click the "Plot" button again.

     • Replot the data by changing the "**Number of bins**" until you get a plot that reveals the distribution most clearly.

     • You can also view the Frequency Distribution Table by clicking the "Next" button.

     • Click on the plot and abscissa labels and provide the appropriate label. Be sure to put your name in the "User Name" Box.

     • Print the histogram screen by clicking on the "Print" button. Click the button that prints the page sideways (landscape mode).

Question*: Is the population of 5,000 numbers reasonable "Normal"?*

Go back to the "Statistical Tool" application.

What you will do: You will draw a random sample of 10 observations from the population. You will calculate the mean and standard deviation. You will calculate the sampling error for your sample.

Comprehension Check #2*: You are going to draw a random sample from the enumerated population. Before you actually draw the sample, write down the value you expect the sample mean to be. Write down the value you expect the sample standard deviation to be.*

Draw a sample of size 10 by clicking on the population field and then dragging the observation to the "***Sample***" field. When you click, an observation is randomly selected from the population of 5,000 values. Where you click is NOT important!

> *NOTE: The actual mousing is to click the mouse-button to select the observation. Then release the mouse-button. Then click and drag the observation to the ' **Sample**" area.*

Repeat the "click & drag" steps until you have 10 observations.

Comprehension Check #3*: pretend you have a large trash can with 5,000 poker chips with numbers from the enumerated population written on them. How would you draw a sample of size 10 randomly?[You would get different results by taking a handful of chips out and selecting 10 versus taking 10 from the large container? Why?]*

Using your calculator, calculate the sample mean and the standard deviation of the sample using your calculator.

Check the accuracy of your calculation by having "**Statistics Tool**" calculate the descriptive statistics for the sample.

• Select "***Calculate Button Functions***" from the "**Task**" pull down menu. Click the button labeled "Sample" (it has a "Summation" icon) to calculate the descriptive statistics.

Is the standard deviation of the sample calculated using the n or (n - 1) divisor?

Did your sample give you the exactly the values you expected for the sample mean and the sample standard deviation? What were the values you obtained? What values did you expect?

Note: Sampling error is a fundamental concept in inferential statistics

"Sampling Error" ($\mathcal{E}$) is the difference between a statistic ($\hat{\Phi}$) and the parameter (Φ) it estimates. That is,

"sampling error = the value of the statistic - the value of the parameter"

or

$$\mathcal{E} = \hat{\Phi} - \Phi$$

Comprehension Check #4*: Calculate the sampling error for your first sample for central tendency and dispersion.*

Standard error of the mean (SEM): For your sample divide the standard deviation by the square root of sample size. Write down the value of the SEM for your first sample.

$$SEM = \hat{\sigma}_{\bar{x}} = \frac{\hat{\sigma}}{\sqrt{n}} = \underline{\hspace{3cm}}$$

Comprehension Check # 5: *The standard deviation measures how scattered individual scores are. Based on your first sample, find the two numbers that are your best guess of the center 68% of the distribution of the population by adding and subtracting one sample standard deviation to/from the population mean.*

Comprehension Check #6: *The standard error of the mean measures how scattered means of samples will be. Based on your first sample, find the two numbers that are your best guess of the center 68% of the distribution of sample means (for samples of size 10) by adding and subtracting one SEM to/from the population mean.*

Critical Point. The concept "standard error of the mean [SEM]" (or standard deviation of the mean) describes a characteristic of reality--how scattered will the means of repeated samples be.

One way to estimate the SEM is to take the standard deviation of one sample and divide by the square root of sample size. Obviously you will get a different value for SEM each time you draw a sample because the sample standard deviation will be different for each sample.

A second way to put a number on how scattered means will be is to divide the population standard deviation by the square root of sample size. This number is not a sample value (that is, it is not a statistic since it is NOT a number derived from a sample). What is the value of the population standard error for the population from which we are sampling?

$$SEM = \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \underline{\hspace{2cm}}$$

The variability of scores in ONE sample tells us how scattered sample means from future samples (of that same size will be).

Divide the standard deviation of a sample by the square root of n to get an estimate of the standard error of the mean.

$$SEM = \hat{\sigma}_{\bar{x}} = \frac{\hat{\sigma}}{\sqrt{n}} = \underline{\hspace{2cm}}$$

• Choose "**Transfer Button Operates**" from the "**Task**" pull down menu.
• Click on the "**Transfer**" button near the bottom of the space between the "Sample" and the "Rep" columns. This transfers the statistics for your sample to a table on the screen where all the sample results can be displayed.

Second Task

What you will do: You will draw more samples. For each sample you will calculate the mean and the standard deviation. Then you will calculate the standard deviation of the means.

This process has been automated.
• Select the "**Computer Draws Sample**" from the "**Task**" pull down menu.
• Click the "**Clear Sample**" button -- It looks like a pencil erasing.
>>>>>>>>Do **NOT** click the "**Erase All**" near the "**Print**" button. <<<<<<<<<

Draw 14 more samples of size 10. For each sample:
• Click the "Generate" button to automatically draw a sample.
• Click the "Calculate" button to calculate the descriptive statistical values.
• Click the "Transfer" button to copy the results to display with the earlier samples.
• Click the "Clear Sample" button so you can collect another random sample.

When you have collected 15 samples each with 10 observations, <u>calculate the standard deviation of the 15 means</u>.

Write the value off the standard deviation of 15 means here: _____

The standard deviation of a statistic is called a standard error. **The term "standard error" means "the standard deviation of a statistic."** You have calculated the standard error of mean when you calculated the standard deviation of the 15 means because the mean of a sample is a statistic and you have calculated the standard deviation of a statistic.

Write the value of the standard error of your 15 means here: _____

Write the value of the standard deviation of your 15 means here: _____

<u>You should have written the same value in both of these spaces.</u> Be sure you understand why the same number is called both a "standard deviation" and a "standard error"! Ask your proctor to explain if you are unsure.

• Click the "Calculate History" button to have the software calculate the mean and standard deviation of the 15 means to check your calculations.

Very Important Note: When you have calculated the standard deviation of the 15 means (each from a sample of 10 observations), you have the standard error the mean. You do **NOT** divide by the square root of n. The standard deviation of a set of means is, by definition, an estimate of standard error of the mean.

If you have the standard deviation of **one** sample of 10 observations, you can estimate how dispersed the means of future samples (of size 10) will be by dividing the standard deviation by the square root of sample size.

Comprehension Check #7: *Why is the value of the standard deviation of sample means smaller than the standard deviation of individual observations?*

Comprehension Check #8: *What value of SEM do you expect for samples of size 10 from the enumerated population?*

There are three ways to get an idea of how much sample means will vary from one sample (of size n) to the next sample (of size n).

1. Find the standard deviation of means of multiple samples. Enter that value here: _____

2. Find the SEM based on one sample of observations. Enter that value here: _____

3. Find the population SEM by dividing the Population Standard Deviation by the square root of sample size. Enter that value here: _____

Compare the size of the SEM of your 15 sample means ("1" above) with the value you predicted from knowing the population standard deviation ("3" above). Compare these with the estimate of standard error from your first sample ("2" above). Are all three numbers about the same magnitude? Why (or why not)?

• Click the "**Print**" button to print this screen for later reference.

• Choose the "**Computer Does Lab**" from the "**Task**" pull down menu.
• Click the **"Clear All"** button.
• Click the "**Generate**" button to have the software repeat everything we have done to now.

Third Task

• Click the "**First**" button to go to the title screen. Then click the "**Std Dev'n & Std Error**"

This time the software will draw **r** samples each of size n from the N(28,5) population. "**r**" stands for replications and indicates how many samples were drawn.

• Check to be sure you are drawing 15 samples each with 10 observations. Click "**Generate**"

The mean, standard deviation, and standard error of the first six samples are displayed.
The means of the r samples are displayed in a scrolling field at the right of the screen
The mean and standard deviation of the r means are displayed in the lower box on the right
The population value of the Standard Error of the Mean is displayed (in Red).

• Record the three values for the standard error of the mean.

When you clicked "Generate" you repeated the step-by-step process you have followed in this worksheet

Repeat the process 6 more times each time writing down the three values for the standard error of the mean.

Simulation	1	2	3	4	5	6	7
SD of 15 means; each mean based on 10 observations:							
The True SEM:							
The Estimate of SEM from the **first** sample:							

Notice how the 3 numbers in a column are all "approximately" the same value. Also note how the numbers in the 7 columns "replicate" from one run of the simulation to the next.

Fourth Task

What you will do: Draw **10,000** samples and inspect the values of means, and variances.

Go to the title screen of the "**Statistics Tool**" application and click on the "**Generate Example Distributions**" button

This card automates the process of drawing repeated samples.
> • Click the "**Clear**" button to ensure the card has been reset.
> • Click on the "**Mean**" radio button
> • Click on the "**Reality Hidden**" cover to choose the parent population.
> • Choose the "**N(28,5)**" population.
> • Click on the "**n1**" value and set it to 10
> • Click on the "**reps**" value and set it to 10000.
> • Click the "**Generate**" button to start the process of calculating the mean of 10,000 samples.

When the means of 10,000 samples have been drawn and displayed, click the "**Calculate** button and record the descriptive statistics for the data. Record the values in the table below. Copy and paste the values into the **Histogramer** and plot the 10,000 means. Enter information for the labels. Print the histogramer screen. Repeat the process of drawing 10,000 samples **variances**. Record the values; plot the 10,000 variances using the Histogramer. Print the plot.

| Statistic/Values of 10,000 samples	Reps	Mean	Median	est. Sigma
Mean				
Variance				

Comprehension Check
> • *Do the histograms of both Means and Variances show regularity with a particular Central Tendency or are the values randomly scattered across the histogram plot?*
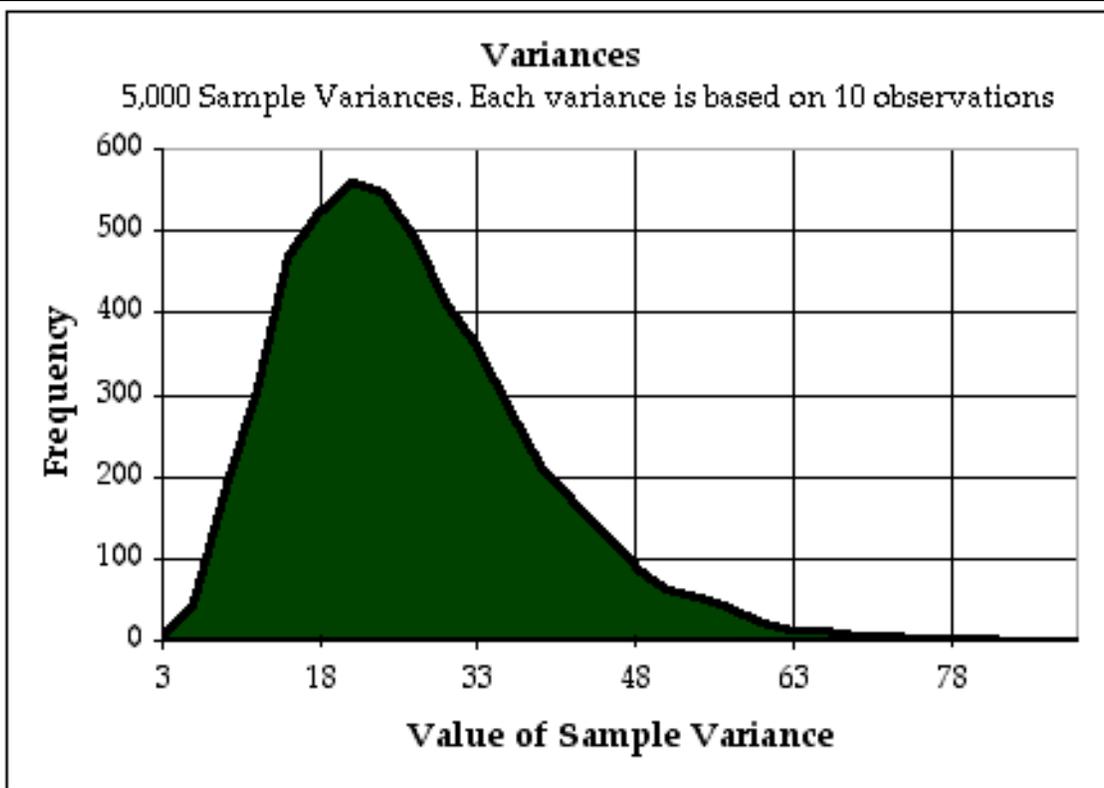
> • *How is the sampling distribution of the Mean different from the sampling distribution of the variance. They should look different. Is the sampling distribution of the variance normal?*

Final Step. Very important. Make an appointment with a colleague or two to compare the results you got with the results they got. Discuss how the results are similar and how they differ.

# Sampling Distributions for the Sample Mean and the Sample Variance (n = 10, r = 5,000)

Means
5,000 Means. Each mean is based on 10 observations.

Value of Sample Mean
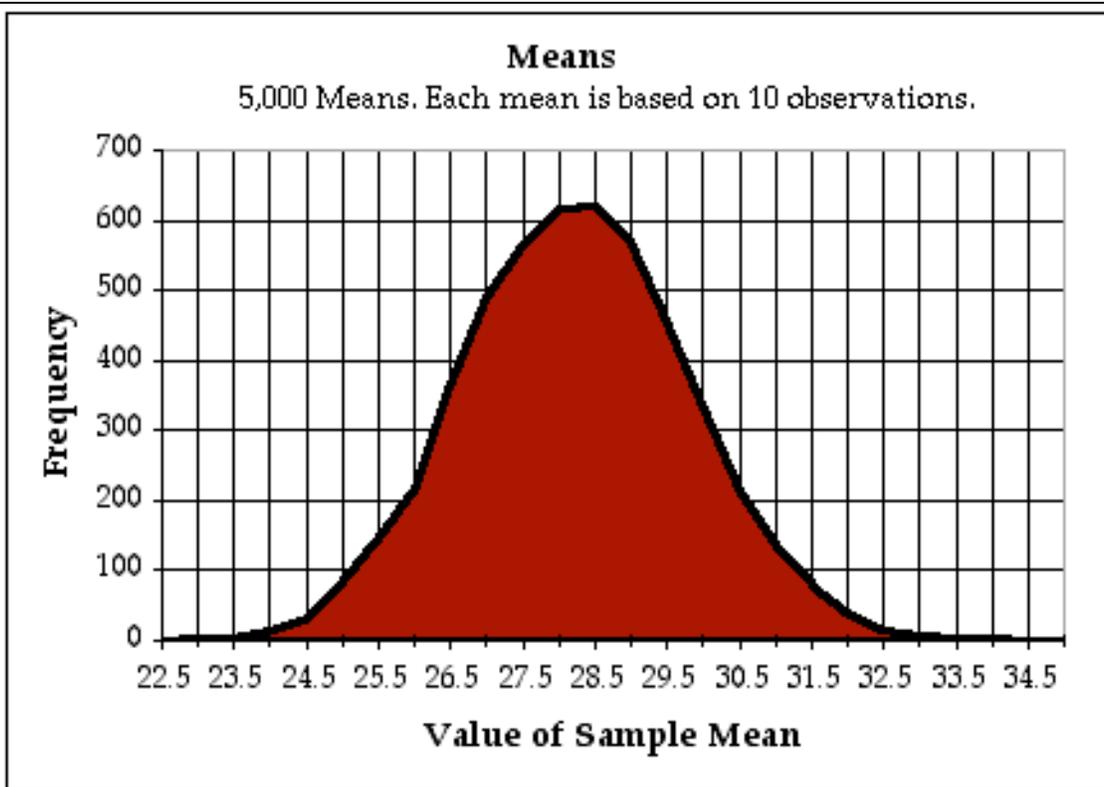
The sampling distribution of the sample mean is Normal . Among other things this means it is symmetrical.

Variances
5,000 Sample Variances. Each variance is based on 10 observations

Value of Sample Variance

The sampling distribution of the sample variance is NOT normal. It is not symmetrical. It is skewed to the right.

Shown below is the histogram for the enumerated population.

The N(28,5) [N=5000] Popula...
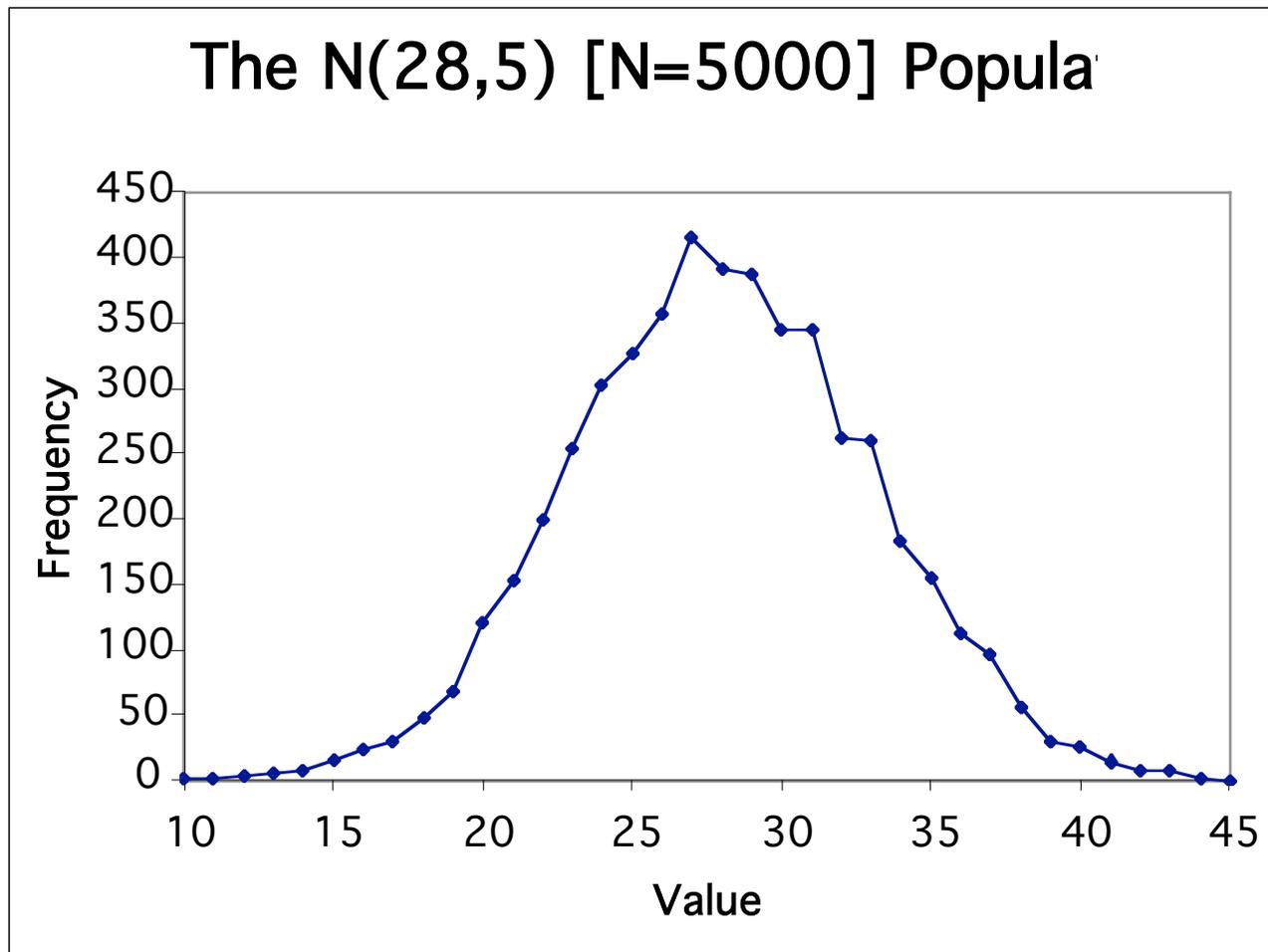
Frequency Distribution Table of the Enumerated Population (N = 5000)

Value	Frequency	Value	Frequency	Value	Frequency
10	1	22	199	34	182
11	1	23	253	35	154
12	3	24	303	36	112
13	6	25	327	37	97
14	8	26	356	38	55
15	15	27	415	39	29
16	23	28	392	40	25
17	30	29	387	41	14
18	47	30	344	42	7
19	67	31	345	43	8
20	121	32	261	44	1
21	153	33	259		