Do not Write

INTRODUCTION

This handout is to be used with the Click & Learn "Sampling and Normal Distribution".

PROCEDURE

Follow the instructions and answer the questions in the spaces below as you read through the Click & Learn "Sampling and Normal Distribution".

Pre-assessment Question

1. A graph of a normally distributed data set is commonly called a bell curve. Describe what this type of curve indicates about a group of measurements (such as height or mass) of individuals in the population.

Read the text on the page titled "Sampling from a Normally Distributed Population" and answer the following items.

2. Distinguish between a sample and a population.

3. The size of the sample is represented by the letter _______.

4. On the graph, what is represented on the

   x-axis:

   y-axis:

5. The red line on the graph shows the distribution of masses in the population.
   a. Looking at this line, describe the measurements in a normally distributed population.

   b. Where on the curve is the mean value in a normal distribution?

PART 1 – SAMPLING FROM A NORMALLY-DISTRIBUTED POPULATION

First, let's take a look at the concept of standard deviation.

Set the following parameters:
Population Mean = 50 kg, Population Standard Deviation = 5kg, Sample Size = 4.
Click "Resample." Then, increase Population Standard Deviation in increments of 5 until you reach 50.

6. Describe what happens to the distribution of masses in the population (represented by the red line) as you increased the standard deviation from 5 kg to 50 kg.

7. In your own words, define "Population Standard Deviation."

Summary: In a population with a normal distribution, any measurements of a population, such as body size, are distributed symmetrically across a range, with most of the measurements occurring toward the middle of the distribution. For normally distributed data, the mean is a measure of the average of that distribution and the standard deviation (SD) is a measure of the variation, specifically how spread out individual measurements in sample are from the sample mean.

For normally distributed data, 68% of measurements in the population will fall within the range defined by one SD above and below the mean; 95% of the measurements will fall within the range defined by 2 SD above and below the mean.

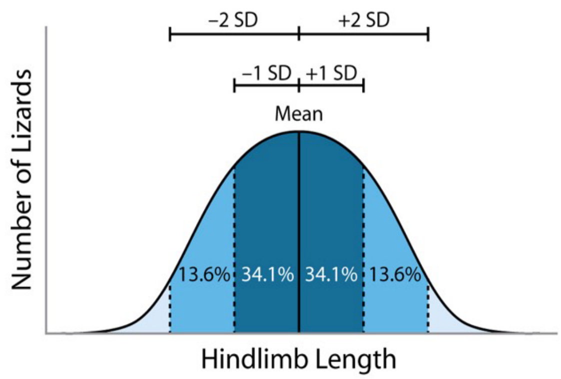


Figure 1. An Example of measurements with a normal distribution.

Now, let's look at mean and sample size

A sample is a randomly selected portion of a population. Each time you select a sample, you are randomly picking a number of individuals from the population to measure. Will the sample be an accurate representation of the whole population? Let's see.

Set the following parameters for the population:
Population Mean = 50 kg, Standard Deviation = 10 kg, Sample Size = 4.

8. Describe what a standard deviation of 10 kg means for this population.

9. Sample the population 10 times (by clicking "Resample") using a sample size of four. Then repeat the process for a sample size of 1,000. Record the means for each sample and other requested information in Table 1.

Table 1. Sample size and sample means – round the mean to the nearest tenth.

	Sample Size									
	4					1000				
Means of 10 samples										
Range between highest and lowest sample mean										
Describe the distribution of the data in the sample										

10. Which sample size results in the set of means that most consistently reflects the population's true mean? ______________

    a. Explain why this sample size results in a more accurate set of means.

    b. What might be a disadvantage of using this sample size?

Selecting the appropriate sample size

When designing an experiment and selecting an appropriate sample size from which to collect data, a balance must be found between obtaining an accurate representation of the whole population and the difficulties of obtaining data from a large number of individuals.

11. Your instructor will assign a sample size to the whole class. Using that sample size, resample the population 10 times.Record the means of the samples in Table 2.

Table 2. Sample size and sample means

Sample Size	Sample Means									

12. Plot the sample means on the grid provided by your instructor. You should plot one dot for each of your 10 samples.
13. For a population of infinite size, using the class-generated graphs, what sample size would be small enough to be feasible to collect data from while still giving you a good representation of the mean for the population? Explain why you selected this sample size.

PART 2 – STANDARD ERROR OF THE MEAN

Click the arrow in the upper right-hand corner to get to the next section of the Click & Learn titled "Standard Error of the Mean." Read the text in the left-hand panel and answer the following items:

1. What does the bottom graph show?

2. The 500 sample means in the bottom graph are normally distributed. The standard deviation of the means is often called the ________________________.

3. Based on your understanding of standard deviation, complete this sentence. ______% of the time you collect a sample of four individuals from this population the mean will be within the range of ± _______ from the population mean.

Now, you will investigate what happens to the standard error of the mean ($SE_{\bar{x}}$) as you increase the sample size.

4. Based on what you learned in Part 1, predict what will happen to the $SE_{\bar{x}}$ as you increase the size of the sample the mean is calculated for.

5. Why do you think this is the case?

6. Complete the following table by changing the sample size and clicking "Resample." Your values may not be the same as those of the person next to you because the program is generating 500 randomly selected samples from a normally distributed population. Don't worry, in the end you will come to the same conclusions.

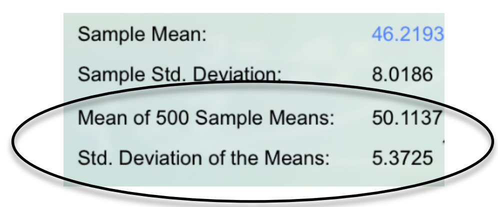


Table 3. Standard Deviation of 500 Sample Means

Sample Size	**Population Mean (kg)**	**Mean of 500 Sample Means (kg)** Round to the nearest tenth	**Standard Deviation of the Means** Round to the nearest tenth
4	50		
9	50		
16	50		
25	50		
100	50		
400	50		

7. Compare the mean and standard deviation of the means between six sample sizes.

   Means:

   Standard Deviations of the sample means:

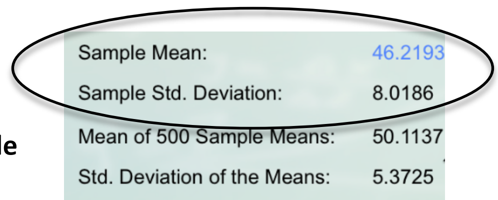
8. Provide the reasoning for the relationship you observed between distribution of the means and sample size.

We refer to the standard deviation of the sample means as "standard error of the mean ($SE_{\bar{x}}$)."

9. According to your data in Table 3, the standard error of the mean ________________ (increases/decreases) as sample size increases.

Fortunately, we don't need to sample a population 500 times to get the standard error of the mean. Statisticians have developed a formula that allows you to estimate the standard error of the mean based on a single sample's standard deviation and sample size.

$$SE_{\bar{x}} = \frac{SD}{\sqrt{n}}$$



You will now compare the standard deviation of 500 sample means to the standard error of the mean calculated using data from a single sample using the formula above.

Transfer the values for "Standard Deviation of the Means" from Table 3 into the appropriate column of Table 4. Resample the population a single time at each sample size in the table.

10. Record the sample mean and sample standard deviation in the table below. Leave the last two columns empty for now.

Table 4. Standard Error & 95% Confidence Intervals

Sample Size	Sample Mean	Sample Standard Deviation	Standard Error of the Mean ($SE_{\bar{x}}$)		68% of samples should have a mean between … ($\pm 1\ SE_{\bar{x}}$)	95% of samples should have a mean between … ($\pm 2\ SE_{\bar{x}}$)
			Using $SE_{\bar{x}}$ equation $\frac{SD}{\sqrt{n}}$	Standard Deviation of the Means		
4						
9						
16						
25						
100						
400						

11. Use the formula for standard error of the mean to calculate the $SE_{\bar{x}}$ for each sample based on the sample standard deviation and sample size.
    a. Compare and contrast the two standard errors for each sample size (determined by the standard deviation of 500 sample means and by using the formula).

b. Compare and contrast your standard error with those of a small group of other students. Describe what you observed.

c. For large sample sizes, the formula for determining the standard error of the mean is an effective way to estimate the range of means you would expect if you were to randomly select many samples from the same population. As you observed, with very small sample sizes the $SE_{\bar{x}}$ formula is a less reliable way to determine the range of means you would expect. Explain why this is the case.

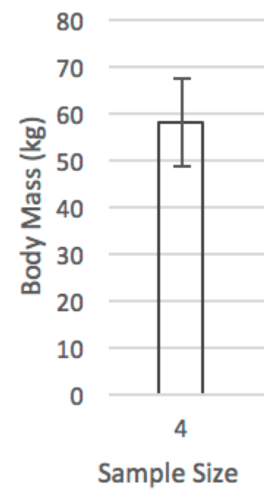
Summary: What does standard error of the mean ($SE_{\bar{x}}$) tell you about the reliability of the mean for your sample? The larger the $SE_{\bar{x}}$, the less likely your sample mean will be a true reflection of the mean of the whole population. While that could be concerning if you are comparing two different populations, $SE_{\bar{x}}$ can help us estimate the range of sample means that could come from randomly selected samples from the two populations. Remember that 68% of the means of samples from the population should fall between one $SE_{\bar{x}}$ above and below the mean (± 1 $SE_{\bar{x}}$) and 95% of the means of the samples from the population should fall between two $SE_{\bar{x}}$ above and below the mean (± 2 $SE_{\bar{x}}$). Therefore, we can be 95% confident that the true mean of the population is somewhere within the range of 2 standard errors of the mean above and below the mean.

This statistic is called the 95% confidence interval (CI).

12. In Table 4, fill in the range of means for one and two standard errors of the mean. For the 68% range, subtract one $SE_{\bar{x}}$ from the mean for the low end of the range and add one $SE_{\bar{x}}$ to the mean for the high end of the range. For the 95% range, perform the same procedure but use 2 $SE_{\bar{x}}$ in your calculations.

How can we show a 95% confidence interval on a graph? The first step is to make a bar graph.
In this exercise, we will graph the sample means of the different sized samples you collected in the last part of the activity. Use the grid below. (*Note – it is important that bars are two or three squares wide and that you DO NOT shade in bars)*

Now it is time to illustrate the range of means within which we are 95% confident the true mean should be found. This is done by including what are called 95% confidence interval (CI) error bars to the bars representing your sample means. One error bar should extend 2 $SE_{\bar{x}}$ above the sample mean and the other error bar should extend 2 $SE_{\bar{x}}$ below the sample mean. See the example to the right.



Sample mean: 57.8 kg
Standard Error of the Mean: ± 4.7 kg
95% Confidence Interval: ± 9.4 (2 $SE_{\bar{x}}$)

13. Describe the amount of overlap you see in the 95% confidence interval errors bars from the different sample sizes.

Summary: When 95% CI error bars from two samples overlap, we cannot claim that they have different means. This would make sense in this case because all of the samples were taken from the same population. When the error bars between two samples do not overlap we can say that the means of the samples appear to be different from each other.

PART 3 – APPLY WHAT YOU HAVE LEARNED

A group of students has conducted an experiment to investigate which enzyme (cellulase or pectinase) or combination of the two enzymes would produce the most apple juice from small chunks of apples. They mixed the specified amount of each enzyme with 50 grams of apple sauce and then filtered the apple juice for 15 minutes into a graduated cylinder. They repeated the control and experimental groups for a total of ten trials each. The students' results are given below.

Table 5. Apple Juice Production

	Amount of Apple Juice Produced from Apple Sauce (mL)			
	10 drops of water	**10 drops of cellulase**	**10 drops of pectinase**	**5 drops of pectinase + 5 drops of cellulase**
Mean of 10 trials	**4**	**3.5**	**12.5**	**10**
St. Dev.	**0.61**	**0.50**	**0.71**	**0.82**
$SE_{\bar{x}}$				
95% CI				

1. Calculate the standard error of the mean ($SE_{\bar{x}}$) and 95% confidence interval for the control and experimental groups.
2. Graph the control and experimental group means and include 95% CI error bars.

3. Describe in a sentence or two what the 95% CI error bars tell you about the students' data.

4. Answer the following questions. Include how the 95% CI error bars help you in making your conclusion.
   a. What can you conclude regarding the effectiveness of cellulase in making apple juice from apple sauce compared to the control?

   b. What can you conclude regarding the effectiveness of pectinase in making apple juice from apple sauce compared to using cellulase?

   c. What can you conclude regarding the effectiveness of the combination of pectinase and cellulase in making apple juice from apple sauce compared to using pectinase alone?