Name______________________________

WEB INTERFACE FOR STATISTICS EDUCATION

SAMPLING DISTRIBUTION OF THE MEAN

EXERCISE #1

INTRODUCTION

Imagine you are a researcher investigating life satisfaction among college students. Like most researchers in behavioral sciences, you conduct your research project using a random sample of students drawn from the population of all students at your college. When the study was completed, you used statistics to make inferences about the population. Is the mean of the sample the same as the mean of the larger population? If another random sample of students were drawn from the population, would the results be the same? If the results vary between samples, how is it possible to make any conclusions about the population? How does sample size affect the accuracy of the sample mean?

To answer these questions we need to understand "the sampling distribution of the mean."

WHAT IS THE SAMPLING DISTRIBUTION OF THE MEAN?

The sampling distribution of the mean is a theoretical probability distribution of the values of the mean that would occur if an infinite number of same-sized random samples were drawn from a population. The sampling distribution tells us to what extent we can expect sample-to-sample variability in the sample mean due to chance or sampling error under a set of predefined conditions. It's one of the most important concepts in inferential statistics and provides a benchmark for comparing statistical results.

THE GOAL OF THE TUTORIAL EXERCISE

The goal of this exercise is to help you determine how accurate a sample mean is likely to be, and how this accuracy is related to the sample size. To achieve this goal, this tutorial exercise uses a small interactive animated program, called a JAVA applet, to pictorially demonstrate the properties of the sampling distribution of the mean. As you work through the interactive exercise, your interaction with the applet will allow you to *"see"* the impact of sample size on the sampling distribution of means. These demonstrations, along with working through the computational problems should enhance your understanding of this important statistical concept.

INSTRUCTIONS

It will take a few minutes for the applet to load on the computer. The applet will open in its own window. Read the attached "Introduction to the Sampling Distribution applet" while the applet is loading. Then when the applet is fully loaded, you will be ready to begin experimenting with the applet or navigating through the written exercise attached to this page. Complete the entire exercise.

How accurate is a sample mean?

WISE Computer Exercise 1: Life Satisfaction

Contents:

Overview

Goals of this exercise: This exercise will help you determine how accurate a sample mean is likely to be, and how this accuracy is related to the sample size.

What do I need to know? To make best use of this exercise, you should know how to use a z table to find probabilities on a normal distribution, and how to calculate the standard error of a mean. Relevant review materials are attached.

What do I need? You will need a table for the standardized normal distribution (z). For some of the exercises we will use the WISE Sampling Distribution applet (on the computer).

The Sampling Distribution applet is a small application program which loads into your computer. Because it is on your local computer, it responds very quickly to your input. It will offer you many choices, and you will see the results of your choices immediately. Feel free to experiment with the applet.

To access the applet: Start Netscape. Click file –- open page – type the following. http://wise.cgu.edu choose tutorials, then choose sampling distribution of the mean (note: the Central Limit Theorem tutorial found on the ‘tutorials’ page may be very useful for you as well).

Answers to questions: You will be asked to answer several questions. Approximate answers are provided so you can check your work. Detailed solutions are attached, but you should try to answer the questions on your own before consulting the detailed solutions. You will learn much more by doing the exercises yourself than if you merely read them and the answers.

WISE Exercise 1 (D. Berger, C. Aberson, M. Healy, V. Romero, D. Kyle 2/24/98)

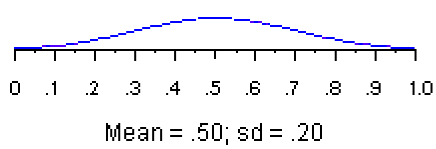
How accurate is a sample mean?
WISE Exercise 1

A friend of yours developed a scale to measure Life Satisfaction. For the population of American adults the scale has a normal distribution with a mean of .50 and a standard deviation of .20. She plans to measure Life Satisfaction for various groups of people, and she is very interested in knowing how accurate her sample means are likely to be. She is coming to you for advice. This exercise is designed to help you prepare for her questions.

Review - normal distributions: Life Satisfaction (n = 1)

We begin with a single score drawn from this normal distribution. (You don't need the applet yet.) A brief review of the normal distribution is attached. You may go there first if you wish.

Q1. What is the probability that a randomly selected American adult has a Life Satisfaction score within .05 of the population mean? (i.e., in the range .45 to .55)



Use the figure to make an approximation. A good way to make an approximation here is to guess what percent of the total area under the curve falls between .45 and .55. Remember, the total area is 1.0 (or 100%).

Probability of a score from .45 to .55 is about ______.

Now, before going on, solve for the exact answer. (The answer is about 20%.) You will need to use a table of probabilities for the standardized normal distribution. The exact answer and detailed calculations are in the answer section. **Show your calculations in the space below.**

Sampling distribution of the mean: Life Satisfaction (n = 100)

For this exercise we will use the Sampling Distribution Applet. If you wish, you may consult the attached description of the various features in the applet now. If you haven't accessed the applet at this point, double click Netscape.

Our researcher friend plans to draw a sample of 100 people. To simulate this, select the following options: **normal, n=100, show sample data** (no other 'show' options for now). If you have questions about any of the options, please consult the attached description of the applet for clarification.

Click on **Draw a sample**, and write the value displayed for **Last mean =** in the first space below. Then click on **Draw a sample** nine more times and record each mean below.

____ ____ ____ ____ ____ ____ ____ ____ ____ ____

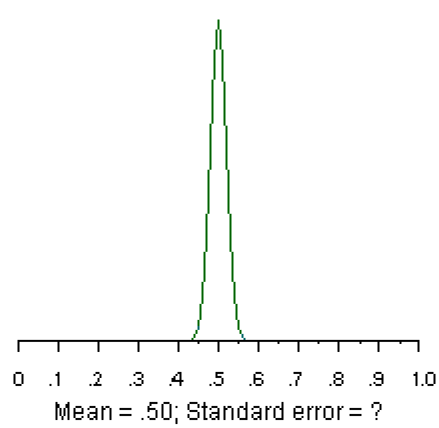
How many of your 10 sample means fell outside of the range .45 to .55? _______

Click on **Show obtained means** to see a display of all ten means**.**
Notice how tightly clustered the obtained means are compared to the individual scores.

Click **Draw 100 samples** to see the means for 100 different samples, each of size n=100.

Click **Show sampling distribution of the mean** to see how closely the observed sample means match the actual distribution of all possible means for samples of size n=100. If you wish, you may consult the attached review of the Central Limit Theorem now.

Q2. What is the probability that a randomly selected sample of n=100 American adults has a mean Life Satisfaction score within .05 of the population mean?



First, make an estimate of the answer. _______

You can also estimate the answer by counting the number of sample means out of 100 that fall within the range .45 to .55. It may be easier to see if you turn off **Show sampling distribution of the mean**. You don't need to get an exact count, but we see that when we draw a sample with n=100, it is unlikely that the sample mean is in error by more than .05 as an estimate of the mean Life Satisfaction for the population (i.e., it is unlikely to find a sample mean that deviates far from the mean). Were any of your 100 sample means .05 or more away from the population mean?

Now solve for the exact answer, using z-scores, before going on. The answer is over 98%. The exact answer and detailed calculations are in the answer section. **Show your calculations below.**

Hint: Consult the attached review of the Central Limit Theorem for more information, including the formula for the standard error of the mean (standard deviation of the sampling distribution).

Sampling distribution of the mean with samples of n=25

Our researcher wishes to know how accurate the sample mean is likely to be if she samples 25 people. To simulate this, select the following options: **normal, n=25, show sample data** (no other 'show' options).

Click on **Draw a sample**, and write the value shown for **Last mean** in the first space below. Then click on **Draw a sample** nine more times and record each mean below.

____ ____ ____ ____ ____ ____ ____ ____ ____ ____

How many of your 10 sample means fell outside of the range .45 to .55? _______

Click on **Show obtained means** to see all ten means displayed**.**
Notice how tightly clustered the obtained means are compared to the individual scores.

Click **Draw 100 samples** to see the means for 100 different samples, each of size n=25.

Click **Show sampling distribution of the mean** to see how closely the distribution of 100 observed sample means matches the actual distribution of possible means of size n=25.

Q3. What is the probability that a randomly selected sample of n=25 American adults has a mean Life Satisfaction score within .05 of the population mean?

First, estimate the answer by examining your ten sample means, the displays of 100 sample means with n=25 for each mean, and the sampling distribution of the mean. What is your best estimate based on these observations? _______

The exact answer is a little less than 80%. What is the exact answer? ________
A detailed solution is in the answer section, but try it on your own before consulting the solution. **Show your calculations below.**

Comparing sampling distribution of means for n=25 vs. n=100.

Your researcher friend says "I'd really rather use samples of n=25 than n=100, because the smaller sample size is less expensive. Why should I use a sample of n=100 rather than n=25?"

How would you respond? Be specific, using your findings.

Sampling distribution of the mean with n = 5

Your researcher friend has considered using sample sizes of only five people. She will ask you to explain the advantages and disadvantages of this plan.

To begin, select the following options: **normal, n=5, show sample data** (no other 'show' options).

Click on **Draw a sample**, and record the value shown for **Last mean** in the first space below. Then click on **Draw a sample** nine more times and record each mean below.

____ ____ ____ ____ ____ ____ ____ ____ ____ ____

How many of your 10 sample means fell outside of the range .45 to .55? _______

Click on **Show obtained means** to see all ten means displayed**.**
Notice how tightly clustered the obtained means are compared to the individual scores, and to the distributions of means when n=100.

Click **Draw 100 samples** to see the means for 100 different samples, each of size n=5.

Click **Show sampling distribution of the mean** to see how closely the observed sample means match the actual distribution of possible means of size n=5.

How does this sampling distribution of the mean (for n=5) compare to the sampling distribution of the mean for n=100? **Indicate your answer below.**

Q4. What is the probability that a randomly selected sample of n=5 American adults has a mean Life Satisfaction score within .05 of the population mean?

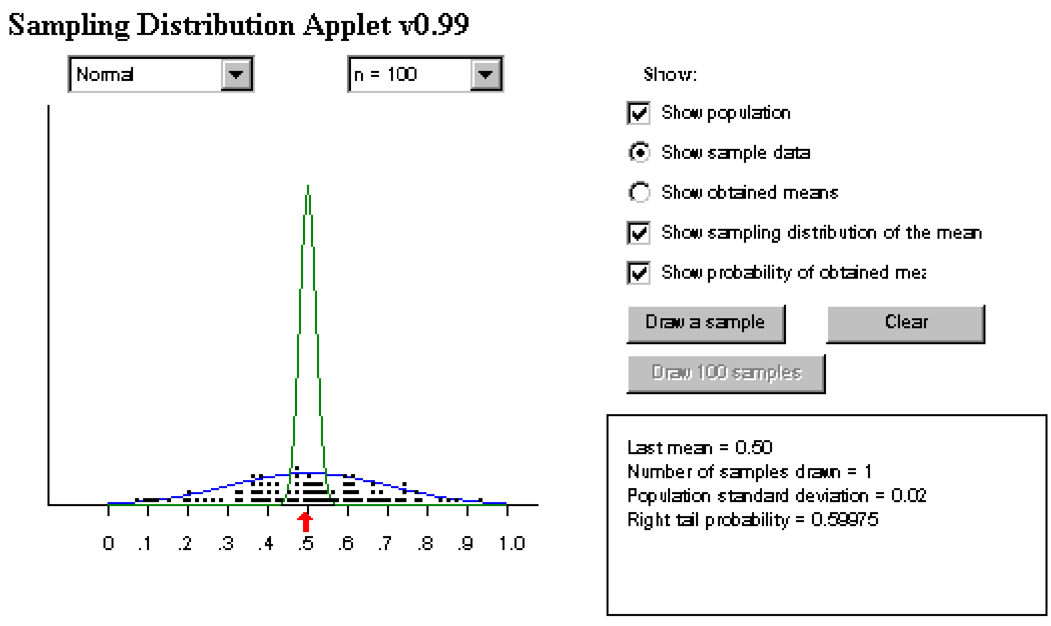
You can estimate the answer by examining your ten sample means and the displays of 100 sample means with n=5 for each mean. What is your estimate based on these observations?

____

The answer is a little more than 40%. What is the exact answer? ________
A detailed solution is in the answer section, but try it on your own before consulting the answer.
Show your calculations below.

Your researcher friend says "we know that for any population, the best estimate of the mean is the sample mean -- therefore, it shouldn't matter what size sample I use, right? Since that is the case, I'll use a sample of n=5 as this will save a good deal of time and money." What do you tell your friend? In your answer, include information from the exercises you have just completed.

Introduction to the Sampling Distribution applet



In the upper left corner, there is a box that will display either Normal, Binomial, or Uniform. This box allows you to select the type of population distribution.

The next box to the right shows n=2 (or 5, 25, or 100), and it allows you to select the size of your sample.

On the right are several 'show' boxes which you can select to show any of the following:

Show population (blue) - this is the distribution of all scores in the population

Show sample data (black) - the data from every individual case in a sample of size n

Show obtained means (red) - the obtained means from several of your samples of size n. You can choose to display the data from a single sample or the means from several samples, but not both at the same time

Show sampling distribution of the mean (green) - the actual distribution of all possible means for random samples of size n

Show probability of obtained means - how likely is it that a sample mean will be at or below the observed value for your sample mean

Below the ‘show’ buttons are three gray rectangular ‘action’ buttons. The actions available will depend on which ‘show’ buttons have been selected.

Draw a sample - draws a single random sample of size n, and shows the sample mean or the n individual data points, depending on which ‘show’ buttons you have selected.

Clear - clears data from the display.

Draw 100 samples - draws 100 samples of size n and displays the 100 sample means.

On the bottom right is a box with a display of several statistics:

Last mean - the mean for the last single sample of size n

Number of samples drawn - the number of samples displayed with Show obtained means

Standard error of the mean - the standard deviation of the sampling distribution of the mean

Left tail probability - the probability of obtaining a sample mean at or below the observed value of the last sample mean.

A good way to use the applet is to explore the features, and develop explanations for your observations.

Review - Probability of scores in a normal distribution: SAT scores

A normal distribution is a specific mathematical distribution that is rarely observed in real data, but yet it has tremendously important statistical applications. As we will see, even when a real population of data does not have a normal distribution, the sampling distribution of means may be very close to a normal distribution.

When we have a normal distribution and we know the mean and the standard deviation of that distribution, we know everything there is to know about the shape of that distribution. We can calculate the probability of observing a score within any specific range.

Exercise: SAT scores among U.S. college students normally distributed with mean of 500 and a standard deviation of 100. What is the probability that a randomly selected individual from this population has an SAT score at or below 600?

Solve this yourself now. **_____** The solution follows.

A score of 600 is one standard deviation above the mean. Using a formula to calculate the z value, we find $z = (x - \mu)/\sigma = (600 - 500)/100 = +1.00$; z is the number of standard deviations that the score of interest differs from the mean. Now we can use a z table of probabilities (areas under the standard normal distribution) to find the desired probability. From the z table (available in your introductory statistics book) we find that the probability that a randomly selected z score in a normal distribution will exceed z=1.00 is .1587 or about 16%. This is equivalent to the probability that a randomly selected individual from this population will have an SAT score over 600 is about 16%. The probability that a randomly selected individual from this population will have an SAT score at or below 600 is 100% - 16% = 84%.

Exercise: What is the percentile score for a person in this population who has an SAT score of 650? _____ Solve this yourself now. The solution follows.

A score of 650 corresponds to a z score of (650 - 500)/100 = 150/100 = 1.50. An SAT score at or above 650 in this normally distributed population is as likely as a z score at or above 1.50 in a standardized normal distribution. We can consult a z table to find this probability, which is .0668. This tells us that 93% of the distribution is at or below this z value. Thus, an SAT score of 650 is at the 93rd percentile in this population.

Caution: It is not appropriate to use the z table to find probabilities unless you are confident that the shape of your distribution is very close to the normal distribution.

Review: Central Limit Theorem

Important facts about the distribution of possible sample means are summarized in the Central Limit Theorem, which can be stated as follows:

If a random sample of **n** cases is drawn from a population with mean μ and standard deviation σ, then the sampling distribution of the mean (the distribution of all possible means for samples of size **n**)

1) has a mean equal to the population mean μ;

2) has a standard deviation equal to the population standard deviation divided by the square root of **n**;

3) and the shape of the sampling distribution of the mean approaches normal as the sample size **n** increases.

The Central Limit Theorem states what the mean and the standard deviation of the sampling distribution of the mean will be for any given sample size, and it also states that the shape of this sampling distribution approaches normal as the sample size increases, *whatever the shape of the population distribution*.

This information allows us to determine the likely accuracy of a sample mean, especially if the sampling distribution of the mean is approximately normal.

If the population distribution is normal, then the sampling distribution of the mean will be normal for any sample size n (even n=1). If a population distribution is not normal, but it has a bump in the middle and no extreme scores and no strong skew, then a sample of even modest size (e.g., n=30) will have a sampling distribution of the mean that is very close to normal. However, if the population distribution is far from normal (e.g., extreme outliers or strong skew), then to produce a sampling distribution of the mean that is close to normal it may be necessary to draw a very large sample (e.g., n=500 or more).

You should not assume that the sampling distribution of the mean is normal without considering the shape of the population distribution and the size of your sample.

Answer Section
WISE Exercise 1

(Note: Selected Answers Only)

Q1. What is the probability that a randomly selected American adult has a Life Satisfaction score within .05 of the population mean? _______

A review of the normal distribution is attached.

The population distribution is normal with a mean of .50 and a standard deviation of .20. First we calculate the standardized z score corresponding to a deviation of .05 from the mean to be $z = (x - \mu)/\sigma = (.55 - .50)/.20 = .05/.20 = .25$. Then we use the z table to find that the probability that a randomly selected score has a z value between zero and .25 is .0987. Next, to find the probability of observing a z score between -.25 and +.25, we double .0987 to get .1974, or about 20%.

Q2. What is the probability that a randomly selected sample of n=100 American adults has a mean Life Satisfaction score within .05 of the population mean?

If we knew the standard deviation of the sampling distribution of the mean, we could calculate the z score for any observed mean and then use z tables to find the probability of observing a mean within any specific range. Fortunately, from the Central Limit Theorem we know that the standard error of the mean (i.e., the standard deviation of the sampling distribution of the mean) is equal to the standard deviation of the population divided by the square root of the sample size. In our example, the standard error of the mean is .20 / (square root of 100) = .20/10 = .020. Thus, a sample mean that is .02 greater than the population mean has a z score of 1.00.

In our example, we would like to know how likely it is that a sample mean differs from the population mean by .05. This corresponds to a z score of .05/.02 = 2.50. The probability of observing a sample mean between .45 and .55 in our example is the same as the probability of observing a z score between -2.50 and +2.50 on a standardized normal distribution. From the z table we find the probability of observing a z score greater than 2.50 is only .0062; the probability of observing a z score less than -2.50 is also .0062. Thus, the probability of observing a z score between -2.50 and +2.50 is 1.0000 - .0062 - .0062 = .9876.

We conclude that the probability is over 98% that a randomly selected sample of n=100 American adults will have a mean Life Satisfaction score within .05 of the actual population mean.

Q4. What is the probability that a randomly selected sample of n=5 American adults has a mean Life Satisfaction score within .05 of the population mean?

The standard error of the mean is .20 / (square root of 5) = .20/2.236 = .08944. A sample mean that is .05 greater than the population mean has a z score of .05/.08944 = .56. From the z table we find the probability of observing a z score greater than .56 is .2877. Thus, the probability of observing a z score between -.56 and +.56 is 1.0000 - .2877 - .2877 = .4246.

We conclude that the probability is about 42% that a randomly selected sample of n=5 American adults will have a mean Life Satisfaction score within .05 of the actual population mean, and a 58% chance that the error in estimate is greater than .05.

Follow Up Questions

NAME ___________________________

<u>Note: These are graded</u>

1. As you increase the sample size (N), the dispersion of the sample means (i.e., the variance of the possible values of a sample mean)
   a. becomes less.
   b. becomes greater.
   c. remains the same.
   d. varies.
   e. cannot be determined from the above information.

2. True or False? The standard deviation of a sampling distribution of means is smaller than the standard deviation of the population when the sample size is 2. Why?

3. Imagine you are working as a research consultant for a large organization. The organization is interested in examining employee satisfaction (essentially, how happy workers are with their jobs). The organization employs over 10,000 employees so having everyone complete the scale is not practical. The organization decides to take a random sample of employees and only administer the scale to these employees. The organization is assuming that the sample of employees will represent the attitudes of the entire population of employees. The organization is unsure as to how many people to administer the survey. One researcher at the organization wants to use a sample of 20 employees whereas the other researcher wants to use a sample of 100 employees. Which sample size should they choose? Why? Present statistical evidence for your answer.

4. Imagine we measured the height of all the male students at a particular college. We found that the average height of men at this school was 70 inches (5'10") with a standard deviation of 2 inches. If we were to randomly select one male student from this college, what is the probability that this student was 73 inches (6'1") or taller?

5. Still think about the same college in question 4. Imagine we selected 16 students at random. What is the probability that the average height of these 16 students is 73 inches or taller?

6. Imagine that you took a sample of students from the college in the prior questions and found a mean height of 68 inches. For which of the sample sizes below would you find this most surprising. (note: n = size of sample)
    a. n = 10
    b. n = 50
    c. n = 100
    d. n does not matter

7. For problem 5 above, what is the population mean?

8. For problem 5 above, what is the sample mean?