Various questions of the type we will learn to answer in this course (Chs. 14 – 22):

- What does this sample data tell us about the mean level of pesticides in whale blubber from whales in the Pacific Ocean off the coast of South America?
- Does the sample data provide significant evidence that the mean tip (as a percentage of the bill) is lower in The Crab Shack than at Red Lobster?

Terminology:

A parameter is a number that describes the population. A statistic is a number that can be computed from the sample data without making any use of any unknown parameters. We use a statistic to estimate an unknown parameter.

For example: We have 20 samples of whale blubber from a sample of whales in the appropriate area. We can measure the level of pesticides in each piece of whale blubber. We compute the sample mean (statistic) of those 20 values to estimate the population mean level (parameter) of pesticides in all the blubber in whales in the area.

Preliminary question to investigate (Ch. 11):

Suppose we have access to full information about the entire population, and we take many random samples of size __. (Fill in a size that is interesting to you.)

1. How variable can the sample means be?
2. Estimate what would be unusual scores for those sample means.
3. Estimate what would be typical scores for those sample means.

Activity 1. Ages of coins.

I have a box of about 500 pennies that I had collected from my pocket change in 2009-2010. I consider this to be representative of the population of pennies that was in circulation in Austin at that time. I want to consider the mean age of a sample of five pennies from this population at the time they were collected in 2010. What would be unusual scores for that mean age? What would be typical scores for that mean age?

From the pennies in the container, pick five and write their five ages here. _____
(Example: The date on a penny is 2002. Age = 2010 – 2002 = 11.)

Write the mean of their five ages here: _____

As a class, build the (popn) histogram of the ages of the pennies on one of the graphs on the table at the front and build the (sample mean) histogram of the means of the ages on the other graph on the table at the front.

(For the results from a much larger class, see the link on the website to “Coin example”.)

What do you see?

- Compare the center (mean) of histogram (popn) and histogram (sample mean). _____
- Compare the spread of histogram (popn) and histogram (sample mean). _____
- Compare the shapes of histogram (popn) and histogram (sample mean). _____

Activity 2: Investigate this type of question further using the Statkey software. On our detailed course calendar, follow the link to Central Limit Theorem discussion and then to the Statkey software.

Discussion: Look at the illustrations of a process like this from "See these pictures."

- If the sample size for each mean was considerably larger than 5, what would you expect to be different **about the distribution of sample means from what you saw with the pennies**?_____

Fill in the following: For any population distribution,

- The distribution of sample means has mean _____
- If the sample size is large enough, the distribution of sample means has shape like a _____ distribution.
- The distribution of sample means has standard deviation which is (choose one: smaller or larger) than that of the population distribution.

Look at the text "The Central Limit Theorem." See that it says all of this and more – it says what the standard deviation of the distribution of sample means is. Write that formula here: _____

Activity 3: For each of the following, sketch both the population distribution and the distribution of sample means for the given sample size.

a. Sample size 36.
Population: Shape is strongly right-skewed, center (mean) is 55, standard deviation is 18.

b. Sample size 9.
Population: Shape is normal, center (mean) is 20, standard deviation is 6.

Activity 4: For the distribution in part b, start with the pictures you drew and modify them to show how to compute each of the following:

b.1. The probability that a score in the population is greater than 26.

b.2. The probability that a sample mean from a sample of size 9 is greater than 26.

For more examples, see the link from our web page to "CLT computation examples."

Discussion / Lecture:

The distribution of the sample means is called the "sampling distribution of the sample mean for samples of size n." Throughout the rest of the course, it will be very important that you be able to distinguish between these:

- The distribution of the original population. (dist'n of the ages of all the pennies in the box)
- The distribution of the values in one particular sample. (dist'n of the five values from one of your samples.)
- The sampling distribution of the sample mean for samples of a given size. (dist'n of all the possible means for all the possible samples of size 5. Our histogram of nickels was a small part of this.)

Activity 5. The National Council for Education Statistics reported 4th grade reading scores in 2003 to have a mean of 217 and a standard deviation of 120. (The scale for these is 0 to 500.) The distribution of these scores was approximately normal.

A particular subgroup of the 4th graders had 3600 students had standard deviation about the same as that of the overall population of 120, but their sample mean was 212.

1. Use the 68-95-99.7 rule to estimate the probability of getting a random sample with a mean as low as this, or lower, from this population.
2. Then you are told that this subgroup of size 3600 was all the 4th graders in charter schools in this state. Which of the following statements are supported by these data?
    a. The mean score for these charter school 4th graders is unusually high compared to the mean score for all 4th graders.
    b. The mean score for these charter school 4th graders is fairly close to the mean score for all 4th graders.
    c. The mean score for these charter school 4th graders is unusually low compared to the mean score for all 4th graders.
3. Suppose the subgroup of all charter-school 4th graders had only 400 students with sample mean 212 and standard deviation 120. Choose the best of a, b, or c for this situation.
    a. The mean score for these 400 charter school 4th graders is unusually high compared to the mean score for all 4th graders.
    b. The mean score for these 400 charter school 4th graders is fairly close to the mean score for all 4th graders.
    c. The mean score for these 400 charter school 4th graders is unusually low compared to the mean score for all 4th graders.

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
Quiz 7: Due Wed. Mar. 6 at the beginning of class:

11.26, 11.34, 14.20, 14.22, 14.28

On each of 11.26b and 11.34, give an appropriate two-part picture as shown in the examples in class and on the website.