Faux Fish Figuring
Day 25

OVERVIEW

The purpose of this lesson is for the students to learn about normal curves, sampling, and variability. The advantage of using faux fish over real data is that students will know everything about the population of fish from which they are sampling. They can compare the information they get from different sizes of samples to the truth. Sampling theory has been developed by testing techniques in fake situations where the truth is known. Then, scientists and statisticians can apply what they have learned about sampling to situations where the truth is unknown. When actually conducting research, scientists never have the luxury of knowing about the whole population of what they are sampling. This lesson will be referred to over and over in the following lessons on statistical testing.

Focus Questions: *What can you learn from a sample? How does sample size affect the accuracy of estimates about the population?*

SCIENCE SKILLS

- ✔ Students will be able to distinguish between samples and populations.
- ✔ Students will be able to state the basic properties of a normal distribution.
- ✔ Students will be able to state that a sample can estimate the true average of a population but that it does not estimate this perfectly.
- ✔ Students will be able to explain why samples with more observations are better estimates of the true population.

BACKGROUND

Background information describing the concepts in this lesson is presented in the following sections:

Section I: Research Q & H – *Hypothesis Versus Null Hypothesis*
Section II: Experimental Design – *Treatment Types*
Section III: Analyzing Results – *Sampling from a Population*
Section III: Analyzing Results – *The Normal Distribution*
Section III: Analyzing Results – *Sampling Variability*
Section III: Analyzing Results – *Statistical Testing*

MATERIALS

- ✔ 1 envelopes per student pair
- ✔ 1 pair of scissors per student pair
- ✔ butcher block paper or overhead transparencies to display class graphs.
- ✔ 1 histogram of wild chinook weight (g) per student pair (provided)
- ✔ student worksheet (provided)

DEVELOPMENT OF LESSON

1. Before the class begins, prepare two graphs at the front of the room. Label one graph "3-Observation Averages" and the other graph "10-Observation Averages." You can use large pieces of butcher block paper or overhead transparencies.

2. Discuss normal curves as a class (refer to the Background Materials for this section). The basic points that should be covered include the following:

- ✔ The normal curve can describe many things found in nature (heights, weights etc.).
- ✔ The center or balance point of the normal curve is the average.
- ✔ The spread of the curve describes the variance.

Compare two normal curves – one with a small variance and one with a large variance. How are they similar? Different? Would they have the same average? How might data from a population with very little variability differ from data from a population with high variability.

3. Explain to the class the difference between a population and a sample. Solicit examples of the population versus the sample from a variety of LTRP projects. Tell the students that in today's class, they will discover what samples can tell you about a population. Understanding samples will help students use their data to answer their LTRP research question. It will help the students to imagine sampling if they remember that their LTRP data probably come from a normal distribution – most things in their population have a value near the average but some are particularly large or small.

4. Hand out the fish histogram. These fish come from a normal distribution but with only 97 fish, it doesn't look smooth. Discuss as a class what information is contained in the histogram (numbers of different weights of fish in the population, the mean of the population, the spread or variance of the population). Are there many extreme data values - really small or large fish? The average weight of fish in this population is 29 grams. Are most fish close to the average? Tell the class to imagine that this is the population of all wild chinook in the ocean. They know the weights of every single fish.

5. Have the students work in pairs to cut out all the fish and put them in an envelope. It will make the process go much faster if they cut the fish in rows or columns first and then snip off the fish. Of course, the fish won't be cut out along the lines, they'll just be on squares of paper.

6. Hand out the worksheet. Tell the students that they are going to test how well sampling works. They know the real average of the population (29). Have each student take a sample of three fish from the envelope and calculate the average value of their 3 observations. (Each student should calculate his or her own average from his or her own 3-observation sample. There should be one 3-observation average from each student on the class graph. Two students can sample from the same ocean).

7. As each student finishes, have them graph the 3-observation average on the 3-observation class graph at the front of the room.

8. Repeat the exercise with a sample of 10 fish. Have the students graph their 10-observation average on the 10-observation graph at the front of the room.

9. Discuss the results of the activity as a class. What is the difference between the two graphs? Point out that, in both cases, the average of several observations was different from sample to sample – even though the population was exactly the same. Point out that the averages form a normal curve – most averages fall toward the center of the graph (near the true average) but a few are in the tails. What does that tell you about a sample? What does a sample tell you about the population? Is the average of a sample always a perfect estimate of the average of the population? Is it usually pretty good? Can it be off sometimes? What is the difference between a sample of 3 observations and a sample of 10 observations? Which is more variable? What does that tell you about the LTRP? What would happen if you could collect more data? What would a graph of a 25-observation sample look like.

10. The most important concept for the students to understand is that you can take two samples from the same population and get two different averages. This is a critical concept. When flipped around, it means that if you have different averages for the two treatments of the LTRP project, it doesn't necessarily mean that that the populations are different. Even if the null hypothesis is true, you will get slightly different values for the average of all observations in each treatment. Students will need some sort of procedure or decision rule to determine if (1) the two averages are too close together to be able to determine if the populations are the same or different or if (2) the two averages are so far apart that the populations must really be different. The statistical test that students learn in Practicing T-Tests (Day 26) will help them decide how sure they are that the populations are different. Save lots of time for the discussion questions!

DISCUSSION QUESTIONS

1. What happens to your estimate of the population average as you make more and more observations?

2. What would happen to your estimate of the population average if you were taking observations from a population that was really variable? A population that had almost no variability?

3. Why might it be helpful to know that the averages have a normal distribution? *You know that the averages of most samples fall toward the center (the true population average) and only a few are too small or too large.*

4. If you could take a sample of 100 observations of student height at your school, would it be a perfect estimate of the true average student height? *No, but it would be better than a sample of 50 observations.*

5. Since two samples that look different can come from the same population, how can you tell if the samples come from different populations? What information would go into your decision rule? *The difference between the two averages, variability of replicate observations, the number of observations in your sample.*

6. There are 97 fish in the ocean that you sampled in today's lesson. How many do you think you would have to sample to get a really good estimate of the true population average?

7. How many replicates did you have for each treatment in your LTRP? How reliable do you think your sample average is as an estimate of the truth? Would your answer change if your replicate observations were less variable? More variable?