Math Lesson 2: How are Data Collected?

Hawaii DOE Content Standards:
Math standard: [Data Analysis, Statistics, and Probability]-Pose questions and collect, organize, and represent data to answer those questions.

Key concept:
Randomness depends on chance outcomes.

Performance indicator:
After completing this lesson, students will collect data so that a sample represents the population.

Vocabulary:
Sample population, subjects, bias, randomness

Time:
One to two class periods

Materials:
1. *DDM-Exploring Projects*
   Lesson 5, "Selecting a Sample," pp. 25-29.
2. Worksheet 1 (See Appendix A.)

Discussion

Someone once said, "You don't have to eat a whole ox to know that the meat is tough." That statement gives us the idea of what "sampling" means. Suppose we wondered what the average weight of mullet in He'eia fishpond was. Should we weigh every mullet in the pond? Of course not – it is impractical and almost impossible to accomplish without doing considerable harm to the pond as a whole. But we could take a **sample** of the **population** (the entire group of fish or things about which we want information). From this **sample** (or subset of the population) we could compute an average weight, and then draw conclusions about the average weight of all the mullet in the pond, **as long as the data in our sample were representative of the entire population**. In other words, should we limit our sample subjects (individual members of the population) to a collection from one area of the pond? Would the pua (fingerlings) congregate more in an area where the salinity was lower? Would larger animals seek deeper areas of the pond? We would want to be sure that our sample is representative of the entire population, so sampling different areas of the pond would better represent the total mullet population within the loko 'ia. The more areas sampled, the more representative the data are of the population.

What if we wanted to get an idea, for instance, of the size of the average Hawaiian household. Should we go to the homes of Native

sample population

Hawaiians in our neighborhood only and ask how many people lived in each home? Could we assume that our data would reflect the average size of households across Hawai'i and on the mainland?

It is very important to collect sample data correctly, and that often is not an easy thing to do. If our data are bad, (that is, the data entries do not represent the entire population), then the conclusions we draw from our data will not be sound. We cannot assume, for example, that the average weight of mullet sampled from only a single area of He'eia pond would represent the average weight of the total mullet population.

When you set out to collect data for the first time, you might, for example in the "Akaka Bill" opinion poll, be tempted to sample only your friends because they are easy to contact and you enjoy talking with them. Do you think this is a good idea? Why or why not? This type of sample is called a *convenience sample* (for obvious reasons). Why would this **not** be a good way to get data that represent the entire population?

Another not so good way to collect data way is to ask for information voluntarily. For example, fisheries researchers from the state frequently ask recreational fishermen to report their catch. What type of response do you think they receive? Would a fisherman only report his larger catch and fail to note the smaller and/or undersized fish? Perhaps only fishermen who are proud of their catch would file a report. Or perhaps some would fail to mention the capture of a less desirable kind of fish. This type of sample is called a *voluntary response sample*, because responders provide the information voluntarily, and the person collecting the data can't be sure that the responses are accurate. A way to avoid such **bias** (when data in a sample favor a certain outcome that is not present in the population) is to select the subjects for the sample **randomly**, that is, by chance. (See Appendix C.)

Activities

1. Use *Data-Driven Mathematics (DDM) – Exploring Projects*
    Lesson 5: "Selecting a Sample", pp. 25-29
    (This lesson will take one or two class periods.)

2. *(Note: assign for overnight)* Each student will ask ten people what they think is the importance of mathematics in society, on a scale of 1 to 9, with 1 being completely useless or not important at all, and 9 being of highest importance. Students will also record the gender of the responder. (See Appendix A: Worksheet 1.) Before students begin to collect their data they should determine a plan for how they will select their sample subjects randomly. When they return to class with the data from their samples, they must describe in a written paragraph what kind of sampling strategy they used and how they collected the data.

convenience sample

voluntary response sample

bias

random sample