Seeing Red Part II—Building a Better Mousetrap

1. Planning for Data Collection

The first step in designing a statistical study is planning for data collection. This involves deciding what the population of interest is, deciding what characteristics of the population are of interest and how they will be measured, considering alternative methods of sample selection, and finally deciding on a sampling strategy.

Recall from the **Seeing Red I** activity that the school librarian is interested in estimating the proportion of red books in the library so that an informed decision about how many red and how many white security tags to purchase can be made. In that activity, estimates were produced without much thought being given to the actual process of data collection. Now that you know more about sampling methods, this activity will guide you through a planning process that should lead to a better data collection strategy.

Let's start with defining the population of interest and deciding how the characteristic of interest will be measured. After a class discussion, answer the following questions.

What is the population of interest?

What is red?

How will you decide whether a book is red?

To test your definitions, your teacher will show you some books. For each book, decide whether you would classify it as a "red book."

Now that you know what the population of interest is and what constitutes a "red book," it is time to think about a reasonable way to select a sample from the population. One possibility would be to select a simple random sample of books from the library. Unfortunately, assigning a number to each book in the library and then using a random number generator to select the books that will be in the sample is probably impractical. Discuss with your class other possible strategies that would produce a "near random" sample—that is, a sample that would come close to having the desirable properties of a random sample. You should also decide whether you will sample with or without replacement.

Once your class has come to a consensus on a sample selection strategy, record the steps that you will use to select books in the space below.

Considering Other Sampling Strategies

Suppose a decision to use stratified random sampling had been made. Discuss the following questions with your class:

1. How would a stratified random sampling plan be different from the “near random” sampling plan described in the steps above?
2. What is an example of a stratification scheme that would make sense in the context of this activity?

Now suppose that a cluster sampling plan is proposed. Discuss the following questions with your class:

1. How would a cluster sampling plan be different from the “near random” sampling plan?
2. What is a reasonable way to define clusters in the context of this activity?