

## **Seeing Red**

### **Illustrating the Data Analysis Process**

**These activities are an extension of an activity used by Andre Lubecke at Lander University. We would like to recognize her for the idea for the first activity in this investigation and thank her for allowing us to run with that idea in developing this series of activities.**

This investigation consists of a collection of six activities suitable for use in an introductory statistics course. The six activities taken together guide the student through the data analysis process, from planning a statistical study to using the resulting data to make a recommendation. The first of the activities can be carried out as early as the first day of class and the other activities can be introduced at the appropriate times during the course. For each activity, a student handout and detailed teacher notes are provided. The activities that make up this collection are described below.

**Seeing Red I:** This activity works well as a first day of class activity. Students are asked to produce an estimate of the proportion of red books in the school library so that the school librarian can determine how many of each of two colors of security tags should be ordered. On the first day, students generally use less than optimal procedures for collecting data to produce this estimate. They also discover first hand the need to think carefully about the objective of a study and about the need to define what will be measured and how it will be measured prior to collecting data. This activity can lead to a

productive discussion of many issues related to sampling and designing a statistical study.

**Seeing Red II:** This activity builds on the previous activity and can be done just after the class has covered the topics of sampling methods and sources and types of bias. The activity engages students in the process of designing a reasonable data collection plan, asking them to consider several different types of sampling methods. They must also carefully define the population of interest, what the appropriate response variable is and how the response variable will be measured. The product of this activity is a study design that can be carried into the next activity, where students will actually implement the plan.

**Seeing Red III:** In this activity, students actually implement the data collection plan from the previous activity, with some students taking samples of size 50 and others working in pairs to select samples of size 100. Students then construct graphical displays of the sample proportions and begin to investigate sampling variability. This activity can be done at any point after **Seeing Red II**, but it might make most sense to do this activity just prior to beginning the study of sampling variability and sampling distributions.

**Seeing Red IV:** In this activity, students use the plots of the sample proportions observed in the previous activity to informally assess the validity of two claims about the proportion of red books in the library. Next, students construct a single estimate of the proportion of red books in the library by combining the data from all of the student

samples. Students then carry out a simulation of the sampling distribution of the sample proportion to get a sense of what values of the sample proportion would be likely to occur when the librarian's claim about the population proportion is true. Students then use the simulated distribution to decide if the difference between the class estimate of the proportion and the librarian's claimed proportions can be attributed solely to sampling variability. The notion of a P-value is also informally introduced here. This activity can be carried out prior to the formal development of hypothesis tests for proportions.

**Seeing Red V:** This activity revisits the idea of testing the librarian's claim about the proportion of red books in the library, and should be done after students are familiar with the definitions of Type I and Type II errors and know how to carry out a one sample hypothesis test for a population proportion. Students are asked to identify appropriate hypotheses, describe Type I and Type II errors in context, and to discuss the potential consequences of making each type of error. Students then must make an appropriate choice of significance level and carry out the test to reach a conclusion about the validity of the librarian's claim.

**Seeing Red VI:** This activity completes this investigation. Students are asked to make a recommendation to the librarian regarding how many of each color security tag to purchase. While students are not explicitly told how to arrive at a recommendation, they are prompted to explain how margin of error figured in to arriving at the recommendation. The relationship between margin of error and sample size is also briefly addressed.

If you have questions or comments about these activities, feel free to contact any of the authors listed below.

Roxy Peck	Cal Poly, San Luis Obispo	<a href="mailto:rpeck@calpoly.edu">rpeck@calpoly.edu</a>
Celia Rowland	Enloe High School	<a href="mailto:crowland@wcpss.net">crowland@wcpss.net</a>
Daren Starnes	The Lawrenceville School	<a href="mailto:dstarnes@lawrenceville.org">dstarnes@lawrenceville.org</a>