Lab Activity: Data Collection and Sampling

In this lab activity, you will collect data samples by hand as well as generate random samples using Statcato. You will input the data samples you collected into Statcato, which you will use to create frequency tables of the data.

Student Learning Outcomes

By the end of this chapter, you should be able to

- Demonstrate and explain the use of a random sampling method

- Enter sample data into Statcato

- Generate random samples in Statcato

- Generate frequency table for data using Statcato

Preliminary

Read Chapter 1 Sampling and Data in:

Illowsky, Barbara, and Susan Dean. <u>Collaborative Statistics</u>. Connexions. 2 Mar. 2010 <http://cnx.org/content/col10522/1.37/>.

Make sure you understand the following **key terms** (**LR:Key Terms**):

population, sample, parameter, statistic, data, qualitative, quantitative, discrete, continuous, sampling, frequency

Data Collection

Eye Color Study

Suppose you are involved in a study of the distribution of eye colors. Collect 50 samples for the study using a random sampling method of your choice. Discuss the details and results of the data collection process in **LR: Data Collection**.

Here you will input the samples into Statcato for further processing later in the lab.

- Go to `File > Save Project` in order to save the project.

- Enter the 50 samples in column **C1**. Enter all letters in lowercase and spell all words correctly (this is needed for the frequency table construction to work correctly).

Dice Simulation

Suppose you want to examine whether a six-sided die is fair (each side of the die is equally likely to be landed). Such a study would involve tossing the die many times and counting the number of occurrences of each possible outcome. Now you will simulate such a study using Statcato by generating 100 integers between 1 and 6.

Go to `Data > Generate Random Data > Integer`.

- **Store Samples in**: C2

- **Number of Samples to Generate**: 100

- **Minimum**: 1

- **Maximum**: 6

Click **OK**.

Now you should have 100 random integers between 1 and 6 in C2.

Frequency Tables

For each of the sets of data that you collected, you will construct a frequency table to see how often a categorical or numerical datum occurs.

Eye Color Study

Go to `Calculate > Frequency Table`.

- For **Source Data**, choose **C1** in the drop-down menu.

- **Store Frequency Table** in

  o **Category column**: c4

  o **Frequency column**: c5

- For **Method of Computing Frequencies**, choose "Treat source data as categories. Count the frequency of each category."

- Click **OK**.

Now C4 contains the color categories, and C5 the corresponding frequencies.

Go to `Calculate > Calculator`.

- **Enter expression**: c5/50

- Check the box **Store in Worksheet**. For **Column**, enter c6.

- Click **OK**. Then click the close button in the upper right corner.

Now C6 contains the relative frequencies of the colors.

Given C4 the label **color** (enter color in the **Var** cell under C4). Label C5 **color-freq**. Label C6 **color-rel-freq**.

Dice Simulation

Click **Frequency Table** in the **Dialog History**. This brings up the Frequency Table dialog again.

- For **Source Data**, choose **C2** in the drop-down menu.
- **Store Frequency Table** in
    - **Category column**: c7
    - **Frequency column**: c8
- For **Method of Computing Frequencies**, choose "Treat source data as categories. Count the frequency of each category."

Click **OK**.

Now C7 contains six possible dice outcomes, and C8 the corresponding frequencies.

Go to `Calculate > Calculator`.

- **Enter expression**: c8/100
- Check the box **Store in Worksheet**. For **Column**, enter c9.
- Click **OK**. Then click the close button in the upper right corner.

Now C9 contains the relative frequencies of the dice rolls.

Label C7 **dice**. Label C8 **dice-freq**. Label C9 **dice-rel-freq**.

Copy the two frequency tables in **LR: Frequency Tables**.

Discussion

Answer the following questions in **LR: Discussion**. Use the knowledge you have so far to answer these questions—don't worry about getting the right answers.

1. The prevalence of eye colors is as follows: brown, hazel, blue, green, violet.
    a. How does the eye color distribution of your sample differ from the prevalence order listed, if any?
    b. Does your sample appear to be representative of the population?
2. Based on the frequencies of the die rolls, do you think the simulated die is fair?
3. Do you think the sample sizes chosen are sufficient (50 for eye colors and 100 for die rolls)? If not, what do you think the sample sizes should be?