Lab Activity: Probability

In this lab activity, you will collect data of two factors (or categories) and build a contingency table for the data in Statcato. You will compute various probabilities based on the contingency table.

Student Learning Outcomes

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

- Construct a contingency table using Statcato
- Compute probabilities using a contingency table

Preliminary

Read Chapter 3 Probability Topics in:

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

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

  experiment, outcome, sample space, event, probability, conditional probability, independent, mutually exclusive, multiplication rule, addition rule, contingency table

Sampling Data

Suppose you are involved in a study of gender vs. credit status of the students in your school. Survey 30 subjects for the study using a random sampling method of your choice. Ask each subject how many credit units he or she is enrolled in. Record the gender and the number of units of each subject. Discuss the details and results of the data collection process in **LR: Data**.

Building a Contingency Table

Before constructing a contingency table, you will enter the sample data into Statcato. The gender factor has two levels: M (male) or F (female). The credit status factor has three levels: NC (non-credit, 0 credit units), PT (part-time, 0-6 credit units--excluding 0 units), and FT (full-time, more than 6 credit units). (You may refine the credit status according to the unit scale of your school).

STATCATO

Enter the label "gender" in the **Var** cell under **C1**.

Enter the label "status" in the **Var** cell under **C2**.

For each student that you surveyed, enter the gender (M or F) in a cell in C1 and enter the credit status (NC, PT, or FT) in the corresponding cell in C2.

You will now construct a contingency table for the sample data in Statcato.

TATCATO

Go to `Statistics > Multinomial Experiments > Cross Tabulation and Chi-Square`.

- In the Inputs panel, choose **C1 gender** in the **Row** drop-down menu, and choose **C2 status** in the **Column** drop-down menu.
- In the **Chi-Square Test** panel, deselect **Perform Chi-square test**.
- Click **OK**.

The contingency table should now be shown in the Log. Highlight the table in the Log, and then copy (`Edit > Copy`, `Ctrl+C`, or click). Paste the results to the provided space in **LR: Contingency Table**.

Computing Probabilities

Let symbols F, M, FT, NC, and PT denote the following events:

Symbol	Event
F	gender = female
M	gender = male
FT	status = full-time
NC	status = non-credit
PT	Status = part-time

Now compute the following probabilities (**LR: Probabilities**):

Probability	Value	Hint *
P(F)		# females (F,All) / sample size (All,All)
P(M)		# males (M,All) / sample size (All,All)
P(FT)		# full-time (All,FT) / sample size (All,All)
P(NC)		# non-credit (All,NC) / sample size (All, All)
P(PT)		# part-time (All,PT) / sample size (All, All)
P(NC')		1 – P(NC)
P(FT and F)		# full-time female (F,FT) / sample size (All, All)
P(FT and M)		# full-time male (M,FT) / sample size (All, All)

P(FT | F)		P(FT and F) / P(F), OR
		# full-time female (F,FT) / # females (F,All)
P(FT | M)		P(FT and M) / P(M)
P(FT and PT)		Are the events FT and PT mutually exclusive?
P(FT | PT)		P(FT and PT)/P(PT)
P(F or M)		P(F) + P(M) – P(F and M)
PT(F or FT)		P(F) + P(FT) – P(F and FT)

* # stands for *number of*. (row, column) refers to the cell in the specified row and column in the contingency table.

Discussion

Answer the following questions in **LR: Discussion**.

1. What is the sample space of this experiment?

2. Give an example of a mutually exclusive event using the data collected for this lab. Show that the event is mutually exclusive using probabilities.

3. Are the events F (gender = female) and NC (status = non-credit) independent? Explain using probabilities.

4. What is the probability that a person in the study is a credit student (full-time or part-time)?

5. Assuming that the sample you have chosen is representative of the entire school of 10,000 students, what is the expected number of full-time female students in your entire school? What is the expected number of non-credit students?

6. Are there any relationships between the credit status and the gender of a student? Explain using any patterns you notice in the data.