Lab Activity: Discrete Random Variable – Binomial

In this lab activity, you will simulate a binomial experiment and compute binomial probabilities using Statcato.

Student Learning Outcomes

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

- Simulate a binomial experiment using Statcato
- Compute binomial probabilities using Statcato
- Demonstrate an understanding of long-term and cumulative probabilities

Preliminary

Read Chapter 4 Discrete Random Variables 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**):

  random variable, probability distribution function, expected value, Law of Large Numbers, binomial probability distribution, Bernoulli Trial

Simulating a Binomial Experiment

Suppose you want to know the grade you would get if you randomly guess the answers for a five-question quiz in which each question has five choices. You will estimate the probabilities for getting a specific number of questions correct by repeatedly generating a sequence of five guesses and counting how many answers are correct. The simulation procedure is as follows:

- Generate **30** sequences of **five** random guesses out of the **five** possible choices (a, b, c, d, e).
- Suppose that the correct answer sequence happens to be **a, a, a, a, a**. For each randomly generated sequence, count the number of correct answers (i.e. count the number of choice **a**)
- Tally the frequency for each possible number of correct answers (0, 1, 2, 3, 4, 5).

You will use Statcato to generate the random sequences.

 TATCATO **Generating Sequences of Random Samples with Replacement**

Enter the five possible choices **a, b, c, d, e** in column **C1**, starting from row 1 up to row 5.

Go to `Data > Generate Random Data > Samples from Column`.

- **Sample from Column**: C1
- **Store Samples in**: C2-C31

- Check the box **Sample with Replacement**
- **Number of Samples to Generate**: 5

There should now be 30 sequences of five choices in column C2 to C31.

Each of the 30 sequences represents the five random answers in a quiz. Now for each of the 30 sequences, count the number of choice **a** (which happens to be the correct answer for each of the five questions). The number should be between 0 and 5. Tally each number of correct answers in a quiz and sum to the tally to find the frequencies below:

Number of correct answers in a quiz	Tally (卌 卌 ‖)	Frequency (0, 1, …)
0		
1		
2		
3		
4		
5		

The sum of the frequencies should be 30. Find the relative frequency of each possible number of correct answers (divide frequency by 30) and record in **LR: Binomial Simulation**. The relative frequency is the empirical probability of each possible number of correct answers. Calculate the sample mean and standard deviation (**LR: Binomial Simulation**).

Computing Theoretical Probabilities

Now you will compute the theoretical probabilities of each possible number of correct answers if one makes random guesses in a five-question multiple-choice quiz. A question in the quiz is considered a trial, so there are five trials ($n = 5$). A correct answer is considered a success. Each question has five possible answers of equal chance, so the probability of success is $p = 1 / 5 = 0.2$.

Generating Binomial Probabilities

Go to `File > New Datasheet` to open a new datasheet.

Enter the numbers 0, 1, 2, 3, 4, 5 in column **C1**, starting from row 1 up to row 6.

Go to `Calculate > Binomial Probability Distributions > Binomial`.

- In the **Distribution** panel:
  - o **Number of trials**: 5

- o **Event probability**: 0.2
- o **Compute**: Probability density
- In the **Inputs and Outputs** panel:
  - o **Column**: C1
- Click **Compute**

Copy and paste the generated table of binomial probabilities in the Log window to **LR: Binomial Probabilities**. Calculate the mean and standard deviation of the distribution and record in **LR: Binomial Probabilities**.

Discussion

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

1. Compute the following probabilities based on the empirical and theoretical probability distribution:

Probabilities	Empirical	Theoretical
P(no questions right)		
P(at least one question right)		
P(no more than two questions right)		
P(two to four questions right)		
P(all questions right)		

2. Is the empirical distribution similar to the theoretical distribution? Explain.

3. How would you change the simulation experiment so that the empirical distribution is very close to the theoretical distribution?

4. What is the expected value for the number of correct answers? Given the expected value, do you think it is wise to complete the quiz by random guessing?