Lesson 11: Estimating Probability Distributions Empirically

Classwork

Exploratory Challenge 1/Exercise 1

In this lesson, you will use empirical data to estimate probabilities associated with a discrete random variable and interpret probabilities in context.

1. Collect the responses to the following questions from your class.

Question 1:   Estimate to the nearest whole number the number of hours per week you spend playing games on computers or game consoles.

Question 2:   If you rank each of the following subjects in terms of your favorite (number 1), where would you put mathematics: 1, 2, 3, 4, 5, or 6?
English, foreign languages, mathematics, music, science, and social studies

Exploratory Challenge 1/Exercises 2–5: Computer Games

2. Create a dot plot of the data from Question 1 in the poll: the number of hours per week students in class spend playing computer or video games.

3. Consider the chance experiment of selecting a student at random from the students at your school. You are interested in the number of hours per week a student spends playing games on computers or game consoles.

    a. Identify possible values for the random variable number of hours spent playing games on computers or game consoles.

    b. Which do you think will be more likely: a randomly chosen student at your school will play games for less than 9 hours per week or for more than 15 hours per week? Explain your thinking.

    c. Assume that your class is representative of students at your school. Create an estimated probability distribution for the random variable number of hours per week a randomly selected student at your school spends playing games on computers or game consoles.

    d. Use the estimated probability distribution to check your answer to part (b).

4. Use the data your class collected to answer the following questions.

   a. What is the expected value for the number of hours students at your school play video games on a computer or game console?

   b. Interpret the expected value you calculated in part (a).

5. Again, assuming that the data from your class is representative of students at your school, comment on each of the following statements.

   a. It would not be surprising to have 20 students in a random sample of 200 students from the school who do not play computer or console games.

   b. It would be surprising to have 60 students in a random sample of 200 students from the school spend more than 10 hours per week playing computer or console games.

   c. It would be surprising if more than half of the students in a random sample of 200 students from the school played less than 9 hours of games per week.

Exploratory Challenge 2/Exercises 6–10: Favorite Subject

6. Create a dot plot of your responses to Question 2 in the poll.

    a. Describe the distribution of rank assigned.

    b. Do you think it is more likely that a randomly selected student in your class would rank mathematics high (1 or 2) or that he or she would rank it low (5 or 6)? Explain your reasoning.

7. The graph displays the results of a 2013 poll taken by a polling company of a large random sample of 2,059 adults 18 and older responding to Question 2 about ranking mathematics.
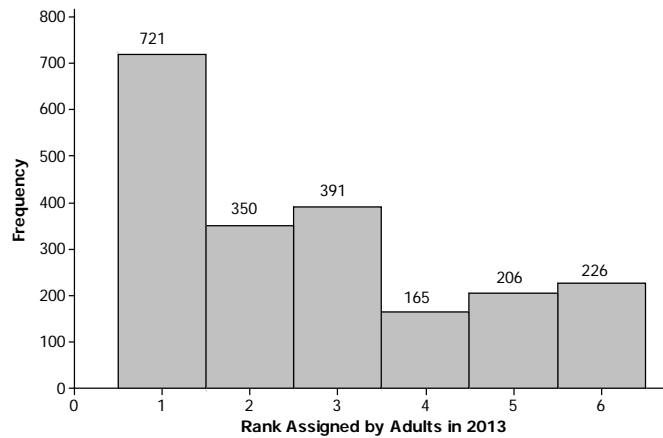
a. Describe the distribution of the rank assigned to mathematics for this poll.

b. Do you think the proportion of students who would rank mathematics 1 is greater than the proportion of adults who would rank mathematics 1? Explain your reasoning.

8. Consider the chance experiment of randomly selecting an adult and asking them what rank they would assign to mathematics. The variable of interest is the rank assigned to mathematics.

a. What are possible values of the random variable?

    b.   Using the data from the large random sample of adults, create an estimated probability distribution for the rank assigned to mathematics by adults in 2013.

    c.   Assuming that the students in your class are representative of students in general, use the data from your class to create an estimated probability distribution for the rank assigned to mathematics by students.

9.   Use the two estimated probability distributions from Exercise 2 to answer the following questions.

    a.   Do the results support your answer to Exercise 1, part (b)? Why or why not?

    b.   Compare the probability distributions for the rank assigned to mathematics for adults and students.

    c.   Do adults or students have a greater probability of ranking mathematics in the middle (either a 3 or 4)?

10. Use the probability distributions from Exercise 2 to answer the following questions.

    a. Find the expected value for the estimated probability distribution of rank assigned by adults in 2013.

        .

    b. Interpret the expected value calculated in part (a).

    c. How does the expected value for the rank students assign to mathematics compare to the expected value for the rank assigned by adults?

Lesson Summary

In this lesson you learned that

- You can estimate probability distributions for discrete random variables using data collected from polls or other sources.

- Probabilities from a probability distribution for a discrete random variable can be interpreted in terms of long-run behavior of the random variable.

- An expected value can be calculated from a probability distribution and interpreted as a long-run average.

Problem Set

1. The results of a 1989 poll in which each person in a random sample of adults ranked mathematics as a favorite subject are in the table below. The poll was given in the same city as the poll in Exercise 6.

   a. Create an estimated probability distribution for the random variable that is the rank assigned to mathematics.

   Table: Rank assigned to mathematics by adults in 1989

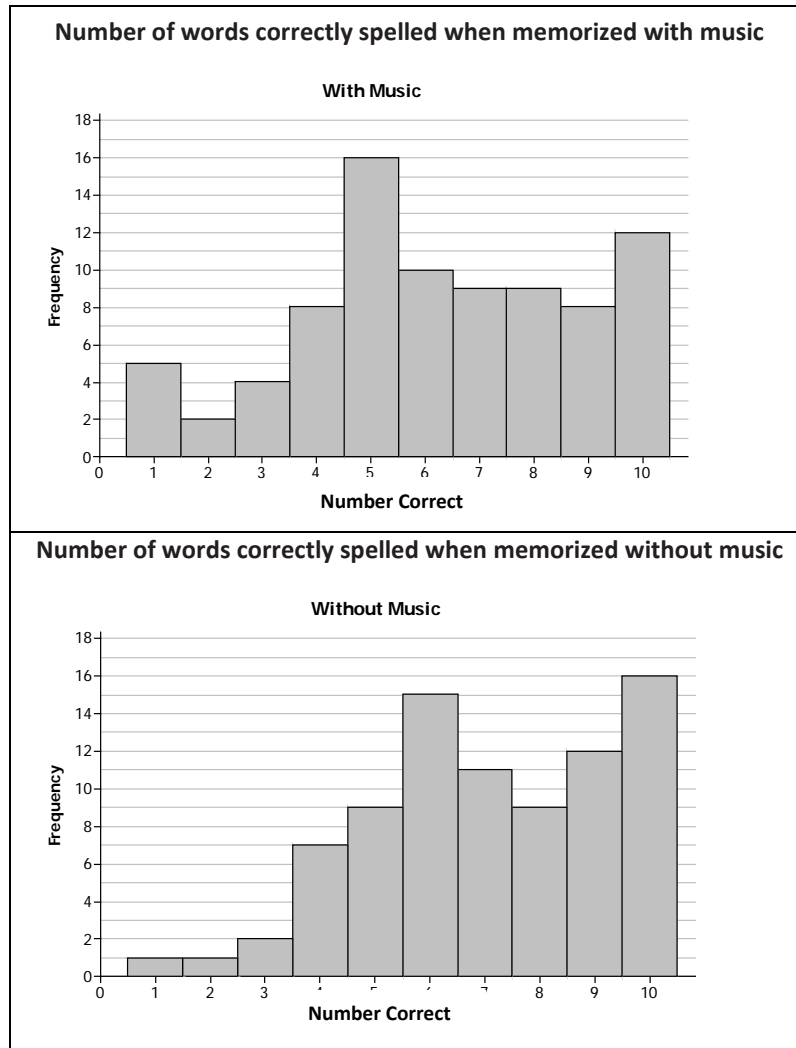 | Rank | 1 | 2 | 3 | 4 | 5 | 6 |
   |-----------|----|----|----|----|----|----|
   | Frequency | 56 | 43 | 12 | 19 | 39 | 61 |

   b. An article about the poll reported, "Americans have a bit of a love-hate relationship with mathematics." Do the results support this statement? Why or why not?

   c. How is the estimated probability distribution of the rank assigned to mathematics by adults in 1989 different from the estimated probability distribution for adults in 2013?

   Table: Rank assigned to mathematics by adults in 2013

Rank	1	2	3	4	5	6
Frequency	0.35	0.17	0.19	0.08	0.1	0.11

---

2. A researcher investigated whether listening to music made a difference in people's ability to memorize the spelling of words. A random sample of 83 people memorized the spelling of 10 words with music playing, and then they were tested to see how many of the words they could spell. These people then memorized 10 different words without music playing and were tested again. The results are given in the two displays below.

Number of words correctly spelled when memorized with music

Number of words correctly spelled when memorized without music

a. What do you observe from comparing the two distributions?

b. Identify the variable of interest. What are possible values it could take on?

c. Assume that the group of people that participated in this study is representative of adults in general. Create both a table and a graph of the estimated probability distributions for number of words spelled correctly when memorized with music and number of words spelled correctly when memorized without music. What are the advantages and disadvantages of using a table? A graph?

d. Compare the probability that a randomly chosen person who memorized words with music will be able to correctly spell at least eight of the words to the probability for a randomly chosen person who memorized words without music.

e. Make a conjecture about which of the two estimated probability distributions will have the largest expected value. Check your conjecture by finding the expected values. Explain what each expected value means in terms of memorizing with and without music.

3. A random variable takes on the values $0, 2, 5,$ and 10. The table below shows a frequency distribution based on observing values of the random variable and the estimated probability distribution for the random variable based on the observed values. Fill in the missing cells in the table.

Table: Distribution of observed values of a random variable

Variable	0	2	5	10
Frequency	18	12	2	?
Probability	0.3	0.2	??	0.47