5.5 Using Simulations and Samples to Estimate Probability in the Real World

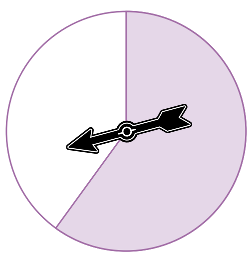
Project Connection

If your project uses data from a random sample of a population, you can construct a simulation based on your data. Several repetitions of your simulation will show you whether your particular sample results are unique or could simply have happened by chance.

? Think about The Need for a Simulation

Why is it necessary to do a simulation for this problem rather than use the strategies presented in earlier sections?

If you were to use a spinner, the sector angles should be 216° (for 60%) and 144° (for 40%).



SIMULATING A WORLD SERIES

Suppose that the Toronto Blue Jays and the New York Mets will play each other in the World Series. Imagine that the Blue Jays had a slightly better regular-season record and have a probability of 0.6 of winning any given game in the series. The first team to win four games in the series becomes world champion. What is the expected number of games required before a winner of the series is declared?

Analysis of the Problem

The World Series final can be thought of as a probability experiment in which each game is a trial. Each trial has one of two possible outcomes—a Blue Jays win (success with $p = 0.6$) or a Met win (failure with $q = 0.4$). Assume that the result of one game has no influence over the results of future games, so each trial is independent of the others. In this experiment, each trial is a Bernoulli trial. However, the series cannot be thought of as a binomial experiment because the random variable is not the number of successes in seven trials. This type of problem is called a *waiting-time* problem, and the random variable is the number of games that need to be completed before one team wins.

INVESTIGATION 1: SIMULATION DESIGN WITHOUT TECHNOLOGY

The mathematical tools used so far are not appropriate for a theoretical analysis of this waiting-time situation. We will estimate the probability with a simulation.

Purpose

To design a simulation of the World Series.

Procedure

A. The Blue Jays are expected to win 6 out of 10 games. Create a simulation using numbered slips of paper, playing cards (an ace and numbered cards 2 through 10), or a spinner. Randomly generate a number between 1 and 10. If its value is 1 through 6, record a Blue Jays win; otherwise, record a Mets win.

B. Repeat this process up to seven times, stopping as soon as one team has won four games.

C. Repeat the experiment several times (at least 100) and use the results to create a probability distribution. (The class could be divided into 10 groups and each group conducts the experiment 10 times. The results of each group are then combined to create 100 simulations.)

> **Think about The Frequency Distribution**
> - Should you be surprised if your simulation differs widely from this example?
> - How much of a difference would cause you to think that your simulation produced typical results or unexpected results?

The results of one such simulation are displayed below.

Number of Games	4	5	6	7
Frequency	16	28	30	26
Experimental Probability	0.16	0.28	0.30	0.26

Using this probability distribution, the expected number of games required before a winner is declared is

$$E(X) = 4(0.16) + 5(0.28) + 6(0.30) + 7(0.26) \\ = 5.66$$

INVESTIGATION 2: USING A TI-83 PLUS CALCULATOR

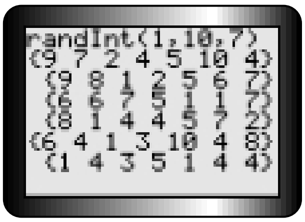
The simulation above generates 10 random integers between 1 and 10 for each trial and conducts up to 7 trials.

Purpose

To design a simulation of the World Series using a TI-83 Plus calculator.

Procedure

> **Technolink**
> For more information on using **randInt** with a TI-83 Plus calculator, see Appendix C.11 on page 408.



The **randInt** function on a TI-83 Plus calculator generates lists of seven random integers with values between 1 and 10. Use this to simulate one World Series. Pressing [ENTER] allows you to simulate as many series as you wish.

The sample screen in the margin represents the following win sequences.

Simulation	Winning Teams						
1	M	M	J	J	J	M	J
2	M	M	J	J	J	J	
3	J	J	M	J	J		
4	M	J	J	J	J		
5	J	J	J	J			
6	J	J	J	J			

INVESTIGATION 3: USING A SPREADSHEET

Purpose

To design a simulation of the World Series using a spreadsheet.

Procedure

Spreadsheet software was used to create the following 10-series simulation. Repeat this simulation to give at least 100 trials.

Technolink
The formula =INT(RAND()*10+1) was used to generate the random integers in this spreadsheet. See Appendix E.3 on page 426.

Game 1	Game 2	Game 3	Game 4	Game 5	Game 6	Game 7	Games Required
8	6	10	9	4	9	6	6
8	8	8	8	1	5	3	4
6	9	2	5	6	3	4	5
4	2	2	3	3	4	10	4
10	4	1	6	2	8	7	5
7	6	5	6	8	1	9	6
3	2	9	5	8	9	8	7
8	3	6	1	1	6	10	5
7	6	1	8	2	6	3	6
9	8	8	8	9	2	5	4
2	2	2	1	4	10	8	4
10	2	3	2	1	5	2	5
8	5	4	8	10	9	4	6
10	9	4	2	8	10	4	6
3	2	6	7	8	3	3	6
1	10	3	4	10	10	9	7
8	7	10	6	9	6	1	5
5	1	8	8	5	5	7	6
7	8	8	2	2	7	9	6
9	8	10	6	1	5	8	7

The probability distribution resulting from this simulation appears below.

Number of Games	4	5	6	7
Frequency	4	5	8	3
Experimental Probability	0.20	0.25	0.40	0.15

Using this probability distribution, the expected number of games required before a winner is declared is

$$E(X) = 4(0.20) + 5(0.25) + 6(0.40) + 7(0.15)$$
$$= 5.5$$

Discussion Questions

1. How does the number of repetitions affect the probability distribution?
2. How does the number of repetitions affect the expected value?

3. Repeat your simulation several times. (If other students have carried out the same simulation, you could combine the results.) Describe the distribution of the expected values. Do they appear to be normally distributed? Explain your answer.

Think about The Effectiveness of Two Drugs

Why do the sample results not provide absolute evidence that one drug is actually more effective than the other?

SIMULATIONS BASED ON SAMPLE RESULTS

Drug A has been in use for a number of years and has an observed success rate of 80% when used to treat a particular illness. In an experimental trial group of 100 patients suffering from the same medical condition, 90 showed improvement using drug B. How likely is it that drug B is really more effective than drug A?

Analysis of the Problem and Simulation Design

No theoretical probability is associated with the success rate of either drug. If the entire population of people with that particular medical condition were given drug A, a certain proportion would show improvement. Assume the same for drug B. Those proportions are unknown.

Suppose that the effectiveness of drug B were the same as that for drug A. It is possible that a larger group of people who improved were selected by chance in the sample. If the probability of this happening is low, then drug B can be said to be more effective than drug A. If the probability of this happening is high, then it cannot be said with certainty that drug B is more effective.

In statistics, if the probability of an event happening by chance is small (the cutoff of 5% is commonly used), it is considered that the event does not happen by chance. Thus, there is a high level of confidence that when the event occurs, it does not occur by chance and is, therefore, noteworthy.

INVESTIGATION 4: SAMPLE RESULTS

Design a simulation of the sampling process for drug B assuming that it has the same effectiveness as drug A, namely 80%.

Purpose

To design a simulation to compare the effectiveness of drugs A and B without the use of technology.

Procedure

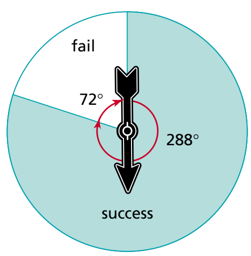
A. One way to carry out this simulation is to use a large number of pieces of paper, 100 for example. Label 80% of them with the letter B to indicate that drug B caused significant improvement.

B. Randomly draw a sample of 10 and record the number of papers that have the letter B.

C. Repeat the experiment at least 200 times.

D. Create a probability distribution for the number of B's in a sample of 10.

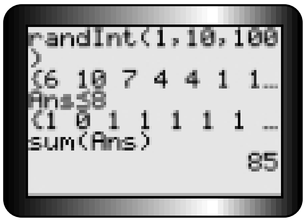
A sample spinner for drug A is shown below.



While it would work, this simulation is very impractical to carry out. It would be simpler to use a spinner set to reflect the 80% success rate, and spin it 100 times repeatedly.

If P is small, there is a small chance that this happened as a result of a random selection of patients. Therefore, the difference in the results of the two drugs is significant and not due to chance. The assumption that drug B was only as effective as drug A is rejected.

Technolink
The screen image below shows the simulation steps as they are carried out one step at a time using a TI-83 Plus calculator.



INVESTIGATION 5: USING A TI-83 PLUS CALCULATOR

It may not be practical to approach this problem using labelled pieces of paper. Simulate the sampling process by generating random numbers as described below.

Purpose

To design a simulation to compare the effectiveness of drugs A and B using a TI-83 Plus calculator.

Procedure

A. Generate single-digit random integers between 1 and 10. The numbers 1 through 8 will represent a patient for whom drug B had a strong effect. The numbers 9 and 10 represent a patient for whom the drug had no significant effect.

B. Generate 100 of these random numbers. This will correspond to one trial in the simulation.

C. The trial is considered an indication that drug B has a stronger effect if 90 or more of the numbers are between 1 and 8. For the manufacturer of drug B, this would be a success.

D. Perform many trials and use the results to construct a probability distribution for the number of patients for whom drug B has been effective.

E. Use the probability distribution to determine *P*(number of patients for whom drug B is effective).

Technolink
Sample results of combining these into a single command are shown below.

The following steps can be used with a TI-83 Plus calculator to carry out one trial of this simulation.

1. **randInt(1,10,100)** will generate a list of 100 random integers between 1 and 10.

2. **randInt(1,10,100)≤8** will create a list of 1s and 0s. A 1 will appear in each position in which the original random number was less than 8. A 0 will appear in each position in which this was not the case. The ≤ test can be found using [2nd] [MATH] **6:≤**.

3. **sum(randInt(1,10,100)≤8)** will add up the entries in the list in Step 2. This will indicate how many 1s were in this list and will correspond to the number of random numbers in Step 1 that were less than 8. This will count the number of patients for whom drug B was successful. The **sum** function can be found using [2nd] [STAT] [MATH] **5:sum(**.

INVESTIGATION 6: USING A SPREADSHEET PROGRAM

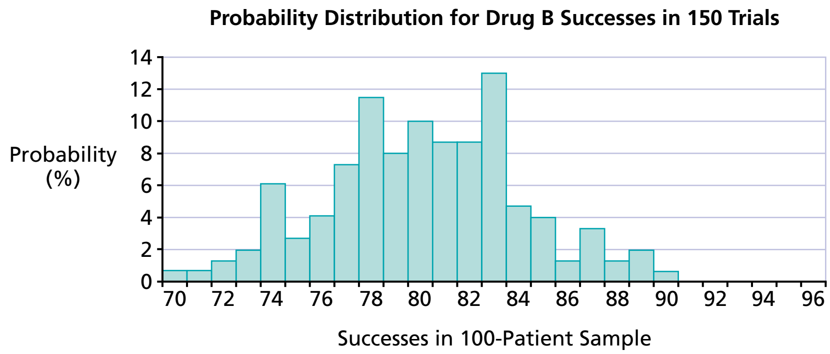
Spreadsheet software can also be used to generate several simulations. Below is part of a spreadsheet that shows 150 trials with 100 random integers each. The command required to generate each random integer is the same as that used in the World Series simulation.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
82	7	6	5	1	5	2	4	8	7	8	7	2	6	3	10	3	10	4	7	1	4	1	9	9	9	1
83	4	3	6	7	9	9	6	8	8	9	2	8	3	4	6	7	6	4	2	10	7	2	3	5	2	8
84	6	8	1	3	4	3	5	1	9	2	3	5	4	9	7	9	8	2	4	6	4	8	3	6	8	2
85	6	3	3	10	5	1	5	4	9	6	4	2	7	8	8	9	9	1	1	5	2	6	1	7	2	3
86	10	3	7	8	2	4	9	5	2	13	7	8	4	9	10	10	6	1	5	4	6	7	7	2	6	3
87	4	3	9	1	3	7	3	3	3	5	4	5	9	7	6	5	4	10	7	4	10	3	2	7	7	6
88	8	2	2	4	4	8	2	3	2	5	6	10	5	9	9	3	3	7	7	6	4	8	1	9	2	8
89	8	8	7	9	7	8	3	3	3	1	1	8	3	8	5	8	7	4	8	2	3	6	10	7	10	5
90	5	1	7	7	10	4	7	6	10	1	7	9	9	9	7	9	1	2	6	3	7	5	5	8	2	1
91	7	6	2	5	2	6	5	9	7	6	1	3	6	1	1	1	8	2	1	1	9	6	6	4	8	8
92	6	8	10	1	1	1	9	5	3	4	10	7	3	6	6	9	2	8	8	4	2	9	6	2	2	5
93	1	5	10	7	2	1	9	7	9	7	3	3	8	7	6	7	5	3	10	2	4	8	5	1	8	8
94	7	2	9	9	8	2	3	1	4	7	7	5	9	1	4	4	4	8	8	10	3	8	6	3	3	9
95	2	1	3	6	10	5	10	8	8	1	5	3	1	6	8	4	9	2	9	4	3	9	2	1	9	7
96	10	2	7	4	9	5	5	2	7	4	5	3	8	8	5	4	6	4	6	6	10	4	7	3	7	5
97	2	6	3	3	3	4	2	6	1	9	9	9	8	8	5	9	2	1	8	2	2	9	8	2	5	6
98	3	5	4	5	3	5	8	4	6	1	5	1	1	2	8	7	10	2	9	1	5	5	6	3	6	5
99	4	10	6	7	6	4	3	7	4	9	4	3	9	1	3	7	6	2	8	9	7	3	4	7	4	8
100	3	2	1	3	2	9	2	7	8	4	2	6	1	9	5	9	8	10	9	8	8	9	2	3	6	7
101	86	84	82	85	82	85	86	87	77	81	88	82	86	75	84	74	83	79	77	79	72	75	82	82	75	76

Successes	Frequency	Probability
70	1	0.67%
71	1	0.67%
72	2	1.33%
73	3	2.00%
74	9	6.00%
75	4	2.67%
76	6	4.00%
77	11	7.33%
78	17	11.33%
79	12	8.00%
80	15	10.00%
81	13	8.67%
82	13	8.67%
83	19	12.67%

Successes	Frequency	Probability
84	7	4.67%
85	5	3.33%
86	2	1.33%
87	4	2.67%
88	2	1.33%
89	3	2.00%
90	1	0.67%
91	0	0.00%
92	0	0.00%
93	0	0.00%
94	0	0.00%
95	0	0.00%
96	0	0.00%

Each column represents one trial with 100 random numbers. The last row of the spreadsheet shows the count of the numbers less than 8. The simulation's probability distribution appears below.



There was one trial in which the number of successes was 90 or more. The probability of this happening was less than 0.67%—very small. As a result, it is appropriate to reject the assumption that the 90% success rate for drug B reported in the sample was only a chance result.

The conclusion is that there is strong evidence that drug B is more effective than drug A. You can be confident that this is a correct conclusion at least 95% of the time.

KEY IDEAS

using simulations to estimate probability distributions—We have dealt exclusively with the binomial distribution in this text. Many probability experiments can be best described using other types of distributions that you have not yet encountered. Other situations are so complex that it may be practical to determine a theoretical probability distribution for them.

A simulation can be used to estimate the probability distribution when it is not possible or it is impractical to determine the theoretical probability distribution.

using samples to estimate the probability distribution for a population—Often, the results of an experiment or a survey provide evidence that a particular characteristic occurs within a sample of the overall population with a certain relative frequency. The relative frequency value can be used as an estimate for the proportion of the entire population that possesses the characteristic. This estimate can be used to design a simulation and to estimate the probability distribution of the characteristic for the entire population.

5.5 Exercises

A

1. Calculate the experimental probability for each outcome in the following frequency tables.

(a)

X	7	8	9	10	11	12	13
Frequency	15	19	12	7	5	2	1

(b)

X	1	2	3	4	5	6	7	8
Frequency	1	7	11	13	9	5	2	0

(c)

X	3	4	5	6	7	8
Frequency	1	1	5	8	15	12

(d)

X	10	11	12	13	14	15
Frequency	2	7	10	6	8	5

2. Use technology to generate 50 random numbers between 1 and 10, and record the results in a frequency table.

3. Calculate the experimental probability for each outcome in the frequency table you created in Question 2.

4. **Knowledge and Understanding** Design a simulation that will give you an estimate of the number of people in a group of 20 who have the same birthday. For the purposes of this problem, ignore the year of birth and assume that there are only 365 days per year.

5. Suppose you have a container in which there are 13 balls numbered 1 through 13. Design a simulation to determine the probability that no ball is drawn in numerical order (ball 1 is not drawn first, ball 2 is not drawn second, and so on).

B

6. Raymond recorded the number of people at his office building that were sick each day for two months (61 days) and summarized his findings in the following frequency table, where the random variable, X, represents the number of sick people each day.

X	0	1	2	3	4	5	6	7
Frequency	44	8	4	2	1	0	1	1

(a) Calculate the experimental probability for each of the outcomes.
(b) Using the experimental results, calculate the probability that no one will be sick during the next 5 work days.
(c) Calculate the expected value of the discrete random variable, X.

7. Twenty-five trials of an experiment produce the following values for the discrete random variable, X: 4, 2, 4, 3, 4, 2, 2, 3, 3, 4, 2, 4, 3, 2, 4, 1, 2, 3, 3, 4, 3, 3, 3, 1, 1.

(a) Create a frequency table for these experimental results.
(b) Calculate the experimental probability for each outcome.
(c) Calculate the expected value for the discrete random variable, X.

8. Suppose the Blue Jays and the Mets each has a probability of 0.5 of winning any game of the World Series. Design a simulation that will allow you to determine the probability that the Blue Jays win the series given that the Mets win the first game. (**Remember**: To win the World Series, you need 4 wins.)

9. A baseball player has a batting average of 0.300. Consider his next 20 at-bats. Design a simulation that will allow you to estimate the probability that he will have a run of 7 at-bats in a row without a hit.

10. You toss a coin 20 times and it comes up tails 18 times. Use a simulation that assumes the coin is unbiased to find evidence that it is a biased coin.

11. **Communication** A well-known basketball player has a career success rate of 80% when shooting free throws. He has missed four out of his last five attempts. Write a report, based on the results of a simulation, that discusses whether or not the coach should be concerned about this.

12. **Application** An experiment is conducted to determine whether a subject has extrasensory perception (ESP). The experimenter tosses a coin and the subject, seated in another room, states what he or she thinks the result is. Should a subject who correctly guesses 7 out of 10 tosses be considered to have ESP?

13. In a jar of 100 jellybeans, 90 are green and 10 are yellow. Five jellybeans are selected at random.

(a) Determine the probability that there are exactly 2 yellow jellybeans in the sample.
(b) Explain why this situation cannot be represented by a binomial distribution.
(c) Determine the probability distribution for the discrete random variable, X, the number of yellow jellybeans in a sample of 5 jellybeans.
(d) Design a simulation to verify the probability distribution you determined in part (c).
(e) Determine the expected number of yellow jellybeans in a sample of 5 jellybeans.

14. A survey was taken of 100 potential voters from a population of 10 000 voters. If 60% of the actual population will vote in the election, determine the probability of the following.

(a) All 100 in the sample will vote.
(b) None of the 100 in the sample will vote.
(c) At least 10 in the sample will vote.

? Think about Question 15

Rolling a die provides a good simulation for Question 15. Why?

15. A cereal company puts coloured pens in its cereal boxes as prizes. There are six different pens available. What is the expected number of boxes of cereal you would have to buy before you could collect all six colours?

C

16. **Thinking, Inquiry, Problem Solving** Recall the birthday problem in Question 4. Use a simulation to determine the number of people needed so that the probability that at least two of them have the same birthday is more than 0.5.

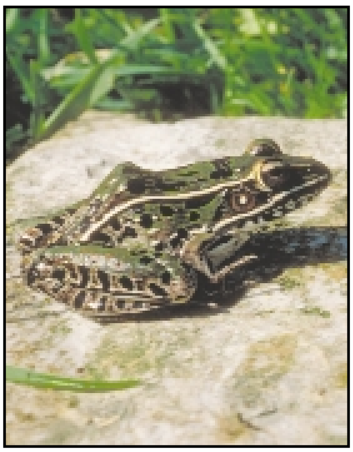
ADDITIONAL ACHIEVEMENT CHART QUESTIONS

17. **Knowledge and Understanding** Design a simulation that will give you an estimate of the number of people in a group of 30 who were born in June.

18. **Application** A small convenience store throws away milk that has expired. At this store, 75% of the 1-L milk cartons are sold before they expire. The store receives 12 new cartons every week. Design and conduct a simulation to determine the probability of throwing out three or more of these cartons.

19. **Thinking, Inquiry, Problem Solving** A newspaper poll of 50 voters showed that 30 approved a proposal for constructing a new expressway around the city. Mayor Haviva Lieberthal, however, says his mail is running against this proposal. Can the mayor be correct? What is the probability that a sample of 50 voters from a population that is equally split could produce 30 *yes* votes?

20. **Communication** Describe a real-world situation for which determining the theoretical probability distribution is impractical or impossible. Describe a simulation that you could use to provide an estimate for the probability distribution of this situation.



Chapter Problem

Are Frog Populations Declining?

CP4. In phase 2 of your study, suppose you capture 100 frogs:

Species	Number in Sample
Bullfrog	35
Spring peeper	50
Mink frog	15

Design a simulation to determine whether or not the distribution in your sample indicates that the number of mink frogs is seriously reduced relative to the number of spring peepers.