INTRODUCTORY ALGEBRA ALGEBRA

Probability Through Data

P. HOPFENSPERGER, H. KRANENDONK, R. SCHEAFFER

DATA-DRIVEN MATHEMATICS

DALE SEYMOUR PUBLICATIONS®

Probability Through Data: Interpreting Results from Frequency Tables

DATA-DRIVEN MATHEMATICS

Patrick W. Hopfensperger, Henry Kranendonk, and Richard Scheaffer

Dale Seymour Publications®
White Plains, New York

This material was produced as a part of the American Statistical Association's Project "A Data-Driven Curriculum Strand for High School" with funding through the National Science Foundation, Grant #MDR-9054648. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Managing Editor: Alan MacDonell

Senior Mathematics Editor: Nancy R. Anderson

Consulting Editor: Maureen Laude

Production/Manufacturing Director: Janet Yearian

Production/Manufacturing Manager: Karen Edmonds

Production Coordinator: Roxanne Knoll

Design Manager: Jeff Kelly

Cover and Text Design: Christy Butterfield

Cover Photo: Steve Satushek, Image Bank

This book is published by Dale Seymour Publications®, an imprint of Addison Wesley Longman, Inc.

Dale Seymour Publications
10 Bank Street
White Plains, NY 10602
Customer Service: 800-872-1100

Printed in the United States of America.

Order number DS21170

ISBN 1-57232-225-X

1 2 3 4 5 6 7 8 9 10-ML-03 02 01 00 99 98

This Book Is Printed
On Recycled Paper

DALE
SEYMOUR
PUBLICATIONS®

Authors

Patrick W. Hopfensperger
Homestead High School
Mequon, Wisconsin

Henry Kranendonk
Rufus King High School
Milwaukee, Wisconsin

Richard Scheaffer
University of Florida
Gainesville, Florida

Consultants

Jack Burrill
National Center for Mathematics
Sciences Education
University of Wisconsin-Madison
Madison, Wisconsin

Emily Errthum
Homestead High School
Mequon, Wisconsin

Maria Mastromatteo
Brown Middle School
Ravenna, Ohio

Vince O'Connor
Milwaukee Public Schools
Milwaukee, Wisconsin

Jeffrey Witmer
Oberlin College
Oberlin, Ohio

Data-Driven Mathematics Leadership Team

Gail F. Burrill
National Center for Mathematics
Sciences Education
University of Wisconsin-Madison
Madison, Wisconsin

Miriam Clifford
Nicolet High School
Glendale, Wisconsin

James M. Landwehr
Bell Laboratories
Lucent Technologies
Murray Hill, New Jersey

Richard Scheaffer
University of Florida
Gainesville, Florida

Kenneth Sherrick
Berlin High School
Berlin, Connecticut

Acknowledgments

The authors thank the following people for their assistance during the preparation of this module:

- The many teachers who reviewed drafts and participated in the field tests of the manuscripts
- The members of the *Data-Driven Mathematics* leadership team, the consultants, and the writers
- Kathryn Rowe and Wayne Jones for their help in organizing the field-test process and the Leadership Workshops
- Jean Moon for her advice on how to improve the field-test process
- The many students and teachers from Rufus King High School and Homestead High School who helped shape the ideas as they were being developed

Table of Contents

About *Data-Driven Mathematics*

Historically, the purposes of secondary-school mathematics have been to provide students with opportunities to acquire the mathematical knowledge needed for daily life and effective citizenship, to prepare students for the workforce, and to prepare students for postsecondary education. In order to accomplish these purposes today, students must be able to analyze, interpret, and communicate information from data.

Data-Driven Mathematics is a series of modules meant to complement a mathematics curriculum in the process of reform. The modules offer materials that integrate data analysis with high-school mathematics courses. Using these materials will help teachers motivate, develop, and reinforce concepts taught in current texts. The materials incorporate major concepts from data analysis to provide realistic situations for the development of mathematical knowledge and realistic opportunities for practice. The extensive use of real data provides opportunities for students to engage in meaningful mathematics. The use of real-world examples increases student motivation and provides opportunities to apply the mathematics taught in secondary school.

The project, funded by the National Science Foundation, included writing and field testing the modules, and holding conferences for teachers to introduce them to the materials and to seek their input on the form and direction of the modules. The modules are the result of a collaboration between statisticians and teachers who have agreed on statistical concepts most important for students to know and the relationship of these concepts to the secondary mathematics curriculum.

Using This Module

Why the Content Is Important

A student's introduction to probability is typically connected to a study of percent or an overview of statistics. Although the general idea of probability is not difficult, students frequently view this special branch of mathematics as a complex study of game theory and counting techniques. An introduction to probability is rarely able to provide an adequate range of the questions investigated by this topic. Probability is certainly about counting. It is also, however, about ordering, organizing, and summarizing. This module develops this topic by involving you in an active process of forming a collection of data. The formal counting techniques are not important at this time. This module further develops examples in which probability is based on organizing outcomes. The range of questions presented to you in this module expands the type of applications traditionally studied in an introduction to probability.

The content of this module involves you in creating data sets and in observing collected data. In each case, you are required to organize the data to form frequency charts and relative frequency charts. This information is then generalized and used to develop the introductory topics of probability. Symbolic representation of the process is primarily an attempt to organize and summarize the main topics.

Similar to other introductory books related to probability, this module works with problems involving tossing a coin or landing in a certain area of a spinner. The connections to games, however, are minimal. The initial goal is to develop the observations from relatively simple simulations as a process to trace the beginning roots of what is probability and the types of questions associated with this topic. After the initial topics are introduced, you are directed at organizing collected data into charts and graphs. This information is used to introduce new questions and new problems. The initial foundation of probability is still there, but the range of problems and situations related to the topic is expanded.

This module develops data sets from a variety of applications. In the third unit of the module, *Data Tables and Probability,* the primary format for organizing the information is two-way tables. You are guided to organize frequency charts, relative frequency charts, and graphs from a two-way table. Questions introducing the topics of compound events, complementary events, and conditional probabilities are developed. This relatively simple method of organizing data provides an excellent opportunity to expand the questions involving probability. The topics developed in Unit III also guide you into an introduction of *association* in Unit IV. Association is not a simple idea. This module, however, provides an initial definition of association related to data collected from surveys and research experiments. Although you are not directed at measuring association through a formal statistic, this module provides you with a clearer idea of analyzing this complex topic by relating it to your work with probability. From a statistical perspective, you work with data to summarize the answer to a question of the solution to a problem. Unlike other problems, however, you are guided into realizing that the summary of your initial questions leads you into more complex questions. This is intentional, as this module attempts to use data from research studies as an example of real-world applications.

Mathematics Content

You will be able to

- Organize collected data into charts.
- Summarize data into percents.
- Identify and interpret relative frequencies as probability estimates.
- Graph frequency data into bar and line graphs.
- Create conditional relative frequency charts.
- Interpret conditional relative frequencies as an estimate of probability.
- Graph information from a chart and interpret the graph as a measure of association.

Statistics Content

You will be able to

- Determine relative frequencies from a simulated collection of data.
- Find expected values from relative frequencies.
- Find and interpret expected values as an estimate of an outcome from a specific problem.
- Interpret graphical summaries as an indication of association of two variables.
- Analyze survey data presented in two-way tables.

Unit I

Probability as Relative Frequency

LESSON 1

Relative Frequency

If the weather forecaster said that there is 60% chance of rain today, would you carry an umbrella?

What if the weather forecaster said 40%?

Weather forecasting depends on a forecaster's knowledge of probability and analysis of data collected over a period of time.

The theory of probability is a part of mathematics that deals with uncertainty. The foundations for the theory were laid in the 1500s and 1600s by mathematicians who were interested in questions about games of chance. Since that time, probability has been associated with many other fields, such as meteorology.

OBJECTIVES

Organize data collected from a probability experiment into a table.

Convert outcomes from a probability experiment into relative frequencies.

Recognize that increasing the number of trials will cause the experimental probability to approach the theoretical probability.

INVESTIGATE

Coin Tossing

If you toss a coin into the air, will it land heads or tails? Sometimes it lands heads, sometimes tails. You cannot say for sure what the next outcome will be. The tossing of a coin is an example of a ***random event.*** For a random event, individual outcomes are unpredictable; but after a great many trials, a long-term pattern may emerge. In this lesson, you will investigate the long-term patterns in the number of heads that occur after tossing a coin many times.

Discussion and Practice

1. Consider tosses of a coin.

a. Predict the number of heads that will occur when a coin is tossed 30 times. Explain your prediction.

b. Actually toss a coin 30 times and record the outcomes in a table similar to the one on page 4.

Toss Number	Outcome (H or T)	Toss Number	Outcome (H or T)
1	______	16	______
2	______	17	______
3	______	18	______
.	______	.	______
.	______	.	______
.	______	.	______
15	______	30	______

c. Record the total number of heads and tails.

Outcome	Number
H	______
T	______

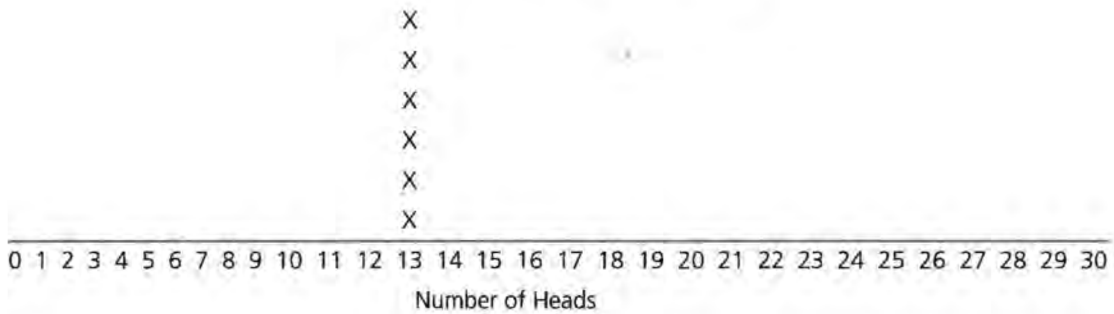
d. Compare the number of heads that you observed from the experiment with the number you predicted.

2. Ask the other students in class how many heads each one observed in his or her 30 tosses of the coin. In a table similar to the one shown below or *Activity Sheet 1*, record the results from all the students in your class.

Number of Heads	Tally	Frequency
0	______	______
1	______	______
2	______	______
3	______	______
4	______	______
5	______	______
6	______	______
.	______	______
.	______	______
.	______	______
30	______	______

3. Use the data collected from each class member to construct a graph of the frequencies of the number of heads from 30 tosses by writing an X above the number of heads for each time that number occurred in class. For example, if six students counted 13 heads when they tossed the coin 30 times, you would write six Xs above the number 13, as shown below.

Class Results for Number of Heads



a. Describe the shape of the graph of the class data.

b. Study the graph. Which number of heads occurred most often? What percent of all the tosses in class does this value represent?

c. If you were to toss another coin 30 times, would you be surprised if you observed 12 heads? 18 heads? 24 heads? Explain how you decided.

The ratio of heads to the total number of tosses is called the ***relative frequency*** of heads. This relative frequency can be expressed as a fraction, a decimal, or a percent.

$$\text{relative frequency of heads} = \frac{\text{number of heads tossed}}{\text{total number of tosses}}$$

4. Consider tosses of a coin.

a. What would you expect the relative frequency of heads to equal after 30 tosses of a coin? Compare this value with the results from your 30 tosses of the coin.

b. If a coin is tossed 1000 times, what value would you predict for the relative frequency of heads?

c. If the relative frequency of heads for 50 tosses of a coin is 0.54, how many heads were observed?

d. If the relative frequency of heads for 80 tosses of a coin is 0.55, what is the relative frequency of tails?

5. Calculate the relative frequency for both heads and tails for your 30 tosses. Express your answer as a fraction, a decimal, and a percent.

6. Use the data gathered by the class.

   a. What is the total number of heads and tails for your entire class?

   b. Calculate the relative frequency of heads and tails for the class data. Express your answer as a fraction, a decimal, and a percent.

   c. Compare the value of the relative frequency of heads from part b with the value of the relative frequency of heads calculated from 30 tosses of a coin in Problem 5.

The table below shows the results of 10 tosses of a coin. This table contains new columns, *Cumulative Number of Heads* and *Cumulative Number of Tails*. The values reflect the total number of heads or tails after each toss. For example, on the third toss a head was observed for the first time; so the cumulative number of heads is 1. The fourth toss was a tail; therefore, the cumulative number of heads remained 1.

Relative Frequency of Heads

Toss Number	Outcome	Cumulative Number of Heads	Relative Frequency of Heads	Cumulative Number of Tails	Relative Frequency of Tails
1	T	0	$\frac{0}{1} = 0$	1	$\frac{1}{1} = 1.00$
2	T	0	$\frac{0}{2} = 0$	2	$\frac{2}{2} = 1.00$
3	H	1	$\frac{1}{3} \approx 0.33$	2	$\frac{2}{3} \approx 0.66$
4	T	1	$\frac{1}{4} = 0.25$	3	$\frac{3}{4} = 0.75$
5	H	2	$\frac{2}{5} = 0.4$	3	$\frac{3}{5} = 0.60$
6	H	3	$\frac{3}{6} = 0.5$	3	$\frac{3}{6} = 0.5$
7	H	4	$\frac{4}{7} \approx 0.57$	3	$\frac{3}{7} \approx 0.43$
8	T	4	$\frac{4}{8} = 0.5$	4	$\frac{4}{8} = 0.5$
9	T	4	$\frac{4}{9} \approx 0.44$	5	$\frac{5}{9} \approx 0.56$
10	T	4	$\frac{4}{10} = 0.4$	6	$\frac{6}{10} = 0.6$

As the number of tosses increases, the relative frequency of heads changes. This value is calculated by dividing the *cumulative number of heads* by the *toss number*. For example, after the fifth toss there were 2 heads, so the relative frequency is $\frac{2}{5}$, or 0.4.

7. Use the results from your 30 tosses to complete a table similar to the one shown below, or use the table on *Activity Sheet 2*. Give the relative frequency of heads and tails as a decimal.

Relative Frequency of Heads

Toss Number	Outcome	Cumulative Number of Heads	Relative Frequency of Heads	Cumulative Number of Tails	Relative Frequency of Tails
1	______	______	______	______	______
2	______	______	______	______	______
3	______	______	______	______	______
.	______	______	______	______	______
.	______	______	______	______	______
.	______	______	______	______	______
30	______	______	______	______	______

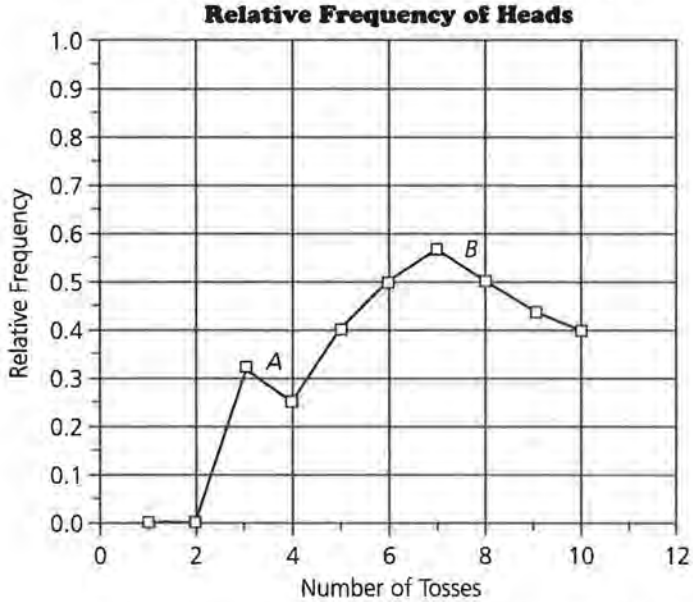
8. Study the table you just completed.

a. What is the sum of the cumulative numbers of heads and tails for any row?

b. What is the sum of the relative frequency of heads and the relative frequency of tails for any row of the table?

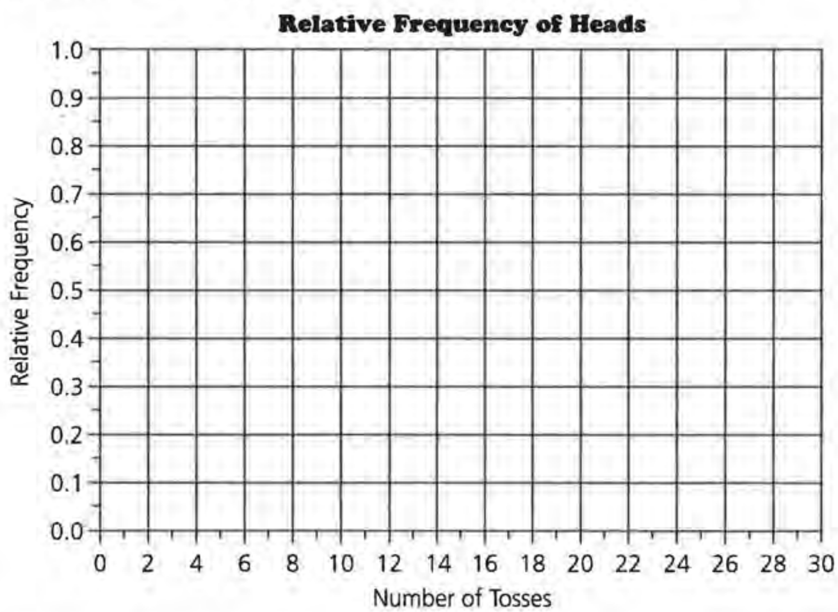
9. As the number of tosses increases, describe what happens to the relative frequency of heads.

10. The data you calculated in Problem 7 can also be displayed in a line graph of relative frequencies. The graph below is a line graph of the data in the table in Problem 6. Point *A* shows that after the third toss, the relative frequency of heads was approximately 0.33. What does point *B* show?



11. Use the data from your 30 tosses and a grid like the one below, also on *Activity Sheet 3*.

a. Construct a line graph of relative frequency of heads.



b. Describe any trends that you observe.

c. On the same graph, copy another student's line graph. Describe the similarities and differences in the graphs.

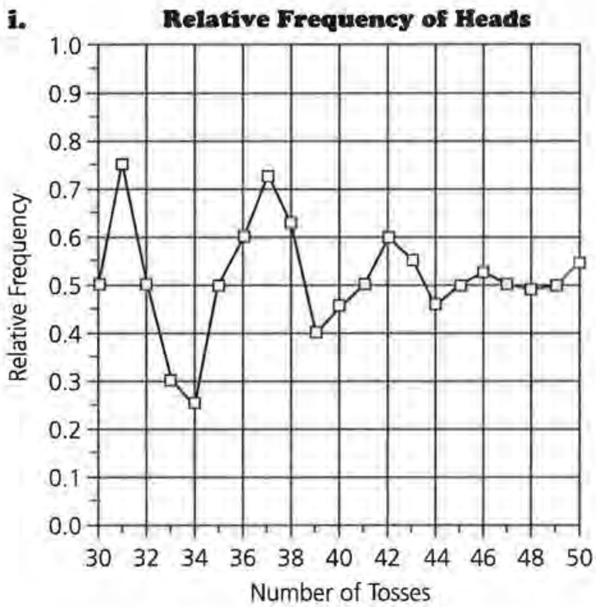
d. Draw a horizontal line across the graph at a relative frequency of 0.50. What does this line represent?

e. As the number of tosses increases, describe the changes in the line graph.

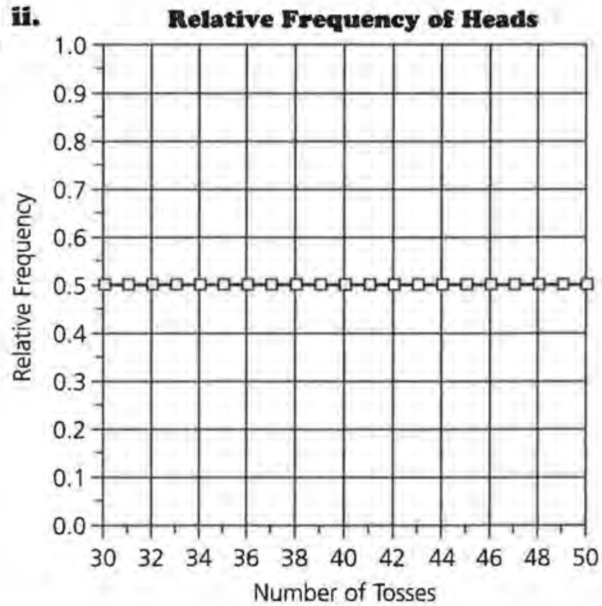
f. As the number of tosses increases, what relative frequency does the line graph seem to approach?

12. If you toss a coin 20 more times and combine these tosses with your previous 30 tosses, what do you think would happen to the line graph? Which of these three graphs comes closest to your expectation? Explain your answer.

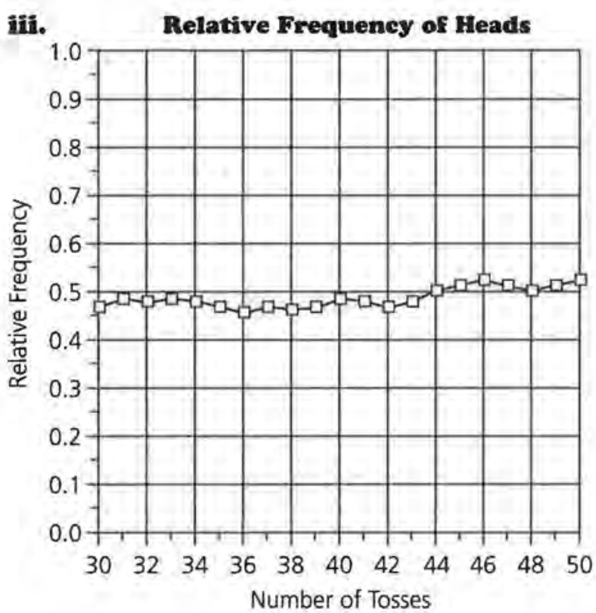
i.



ii.



iii.



The value that the line graph approaches can be thought of as the probability of observing a head in the toss of a coin. The ***probability*** of a head is the relative frequency of heads in a very great number of tosses.

13. What does a line graph of relative frequencies tell you about the probability of a head when tossing a coin?

Summary

This lesson was based on the results of 30 tosses of a coin. You observed that the relative frequency of heads varied with each toss, but this variability decreased as the number of tosses increased. After a great number of tosses, the relative frequency of heads approached the value 0.5, which is the probability of getting a head when tossing a coin.

Practice and Applications

14. A fair number cube has six faces that are numbered from 1 to 6. A student rolls the cube 60 times and records whether the face that turns up is an even or an odd number.

a. Make a sketch of a line graph that you think would show the results of the relative frequencies of getting an even number as the number of rolls of the die increases.

b. What does the line graph of relative frequencies tell you about the probability of getting an even number when rolling a fair die?

c. Work with your group, and roll a die 60 times. On *Activity Sheet 3*, construct a line graph of the relative frequencies of getting an even number. Compare your graph to the one in part a.

15. During World War II in a German prison, John Kerrich, an English mathematician, tossed a coin 10,000 times. After 10 tosses, he observed 4 heads; after 20 tosses, 10 heads; after 30 tosses, 17 heads; after 100 tosses, 44 heads; after 1000 tosses, 490 heads; and after 10,000 tosses, 5067 heads.

a. Create a table showing the toss number, the number of heads and tails, and the relative frequency of heads and tails for the data above.

b. Construct a line graph showing the number of heads and the relative frequency of heads. When making the line graph, the horizontal axis (toss number) should start at 0 and end at 10,000, counting by 500.

c. Discuss how the line graph helps to illustrate why the probability of getting a head when tossing a coin is 0.50.

Applying Relative Frequency

Did you ever watch the TV show *Wheel of Fortune*? A contestant spins the wheel that has numerous pie-shaped sections labeled with amounts of money.

If you watch the show long enough and tally which amounts of money come up, do you think that a pattern will emerge?

OBJECTIVE

Estimate the theoretical probability of an event based on the convergence of the relative frequencies.

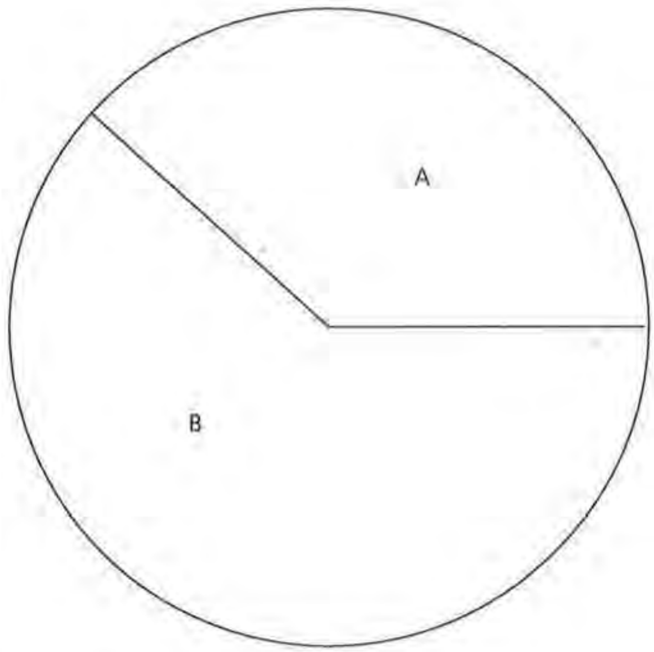
The experiment that you performed in Lesson 1 required you to toss a coin 30 times and record the outcomes. Based on the results, you found the relative frequency of heads and constructed a line graph of these frequencies. From this line graph, you observed that the relative frequency of heads varied with each toss of the coin, but the relative frequency converged on the value 0.50.

Probability describes the uncertainty of any single event by summarizing what happens after many trials of an experiment. In this lesson, you will observe the outcomes of some experiments and, based on your observations, estimate the probability of an event.

INVESTIGATE

Spinning Around

You will need a paper clip and a copy of the figure on page 13, reproduced on *Activity Sheet 4*.



Discussion and Practice

1. Where is the spinner more likely to land, in the area marked A or the area marked B?
   a. Estimate the probability that on a random spin the spinner will land in area A.
   b. Use a paper clip on the tip of a pencil and the circle with areas A and B marked, as on *Activity Sheet 4*, to assemble a spinner. Then spin the spinner 30 times and record your results in a table similar to the one that follows. You many record your results in the first two columns of *Activity Sheet 5*.

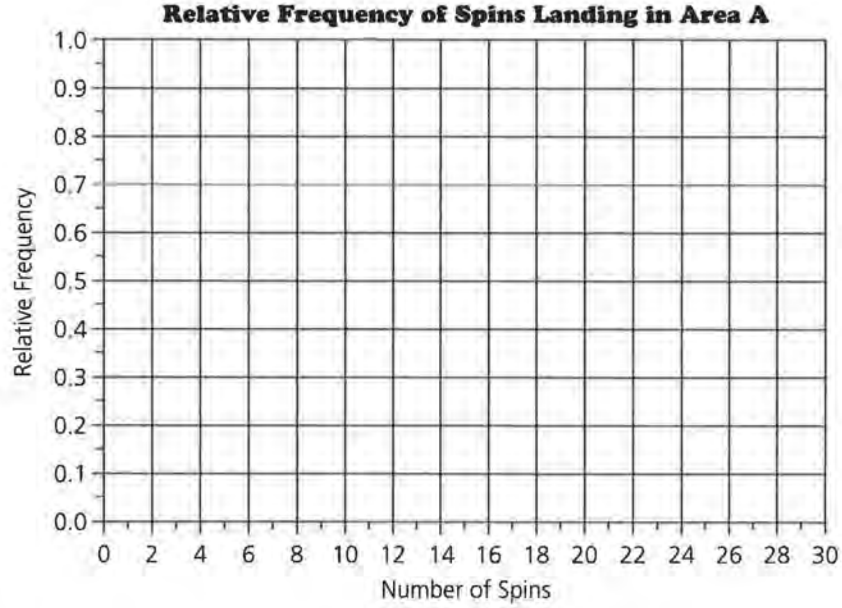
Spin Number	Outcome (Area A or Area B)
1	______
2	______
3	______
.	______
.	______
.	______
30	______

c. Complete a table similar to the one shown below or complete the remaining columns in the table on *Activity Sheet 5*.

Spin Number	Outcome (Area A or Area B)	Cumulative Number in Area A	Cumulative Number in Area B	Relative Frequency of Spins Landing in Area A
1				
2				
3				
.				
.				
.				
30				

d. Use the data in the relative frequency column from the table you completed in part c to construct a line graph of the relative frequency of spins landing in area A for each spin. You may use the grid on *Activity Sheet 6*.

Relative Frequency of Spins Landing in Area A



e. What happens to the line graph as the number of spins increases?

f. As the number of spins increases, what relative frequency does the line graph appear to approach?

g. On the same graph, copy a classmate's graph. Describe the similarities and differences in the two graphs.

The relative frequency value the line graph approaches can be thought of as an estimate of the ***probability*** that the spinner will land in area A. Probability, as the value that the relative frequency approaches, can be generalized by the following rule:

As an experiment is repeated again and again, the relative frequency of occurrences of outcomes defining an event approaches the probability of the event. The value that the line graph of relative frequencies approaches can be used as an estimate for the ***probability of the event.***

2. Use your line graph from Problem 1d.

   a. What does the line graph of relative frequency tell you about the probability of landing in area A on any spin?

   b. If you were to spin 1000 times, what value would you predict for the relative frequency of the number of times landing in area A?

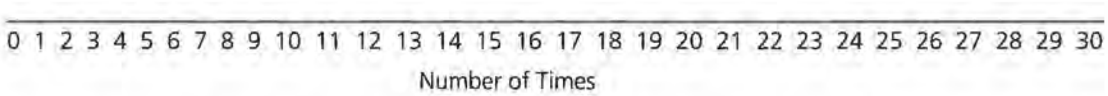
3. Ask other students in class for the number of times their spinner landed in area A.

   a. Record the results in a table similar to the one shown below. You may use the table on *Activity Sheet 7.*

Number of Times Spinner Landed in Area A	Tally	Frequency

   b. Use the data collected from each class member to construct a graph using a number line similar to the one that follows. Record the frequencies of the number of times the spinner landed in area A out of 30 spins by writing an X above the number of times for each time that number occurred in class.

Class Results for Number of Times Spinner Landed in Area A



c. Read the graph to determine the most likely number of times the spinner would land in area A after 30 spins.

d. Do you think it would be likely for the spinner to stop from 18 to 22 times out of 30 spins in area A? Explain your answer.

e. If you were to spin the spinner 1000 times, how many times would you expect to land in the area marked A? How does this answer relate to the answer to Problem 2b?

4. Use the relative frequency that your line graph approached.

a. Predict the measure of the central angle for area A.

b. Use a protractor to find the measure of the angle.

c. Compare the results from parts a and b.

Skittles

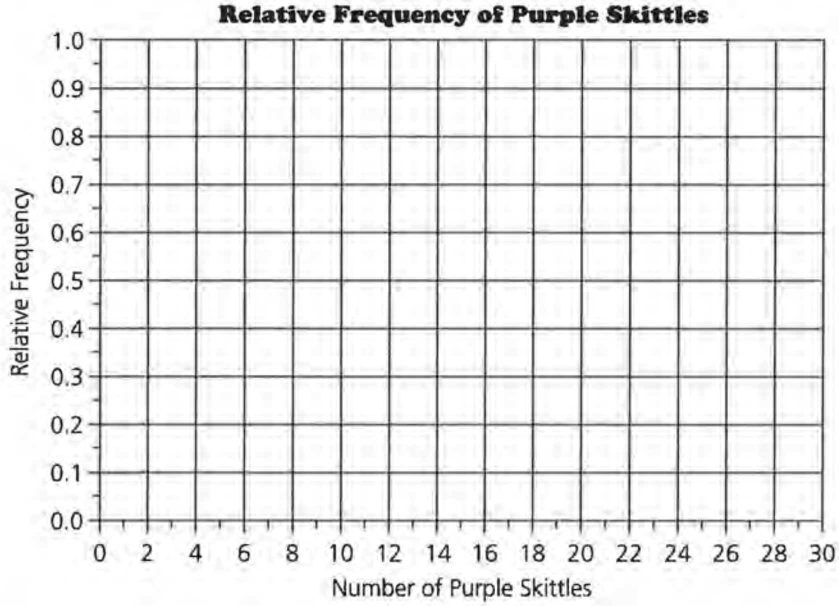
You will need a small bag of Skittles® for this problem.

5. Consider the question "What is the probability that a randomly chosen Skittle will be purple?"

a. Open a small hole in the bag of Skittles. Allow only one Skittle at a time to fall out of the bag and note the color. Record the results in a table similar to the one that follows. As you collect the data, calculate the relative frequency of purple Skittles, and construct a line graph of the relative frequency. Use a grid like the one shown, reproduced on *Activity Sheet 6*.

Skittle Number	Outcome	Cumulative Number of Purple Skittles	Relative Frequency of Purple Skittles

Relative Frequency of Purple Skittles



b. How many Skittles did you sample until you were confident that you knew the probability of randomly selecting a purple Skittle?

i. Why did you stop at this number of Skittles?

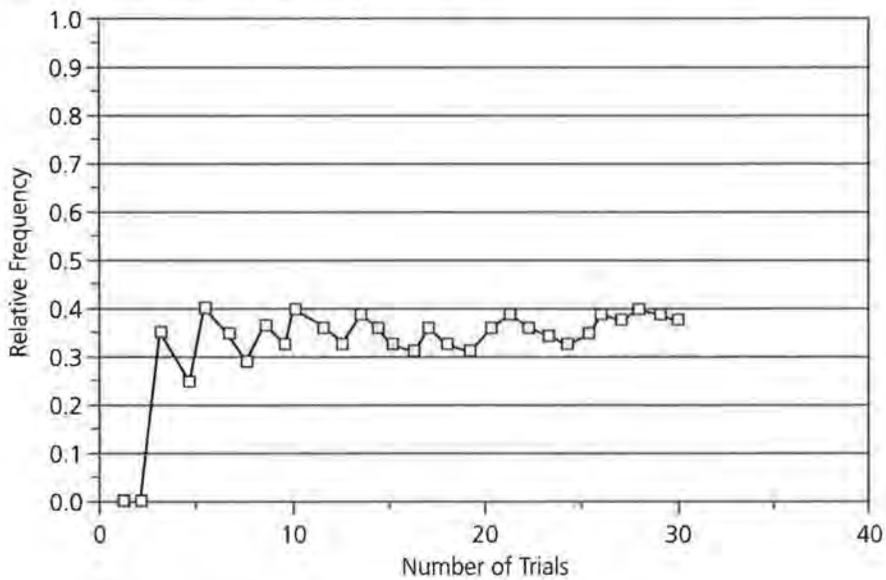
ii. What is your estimate for the probability of randomly selecting a purple Skittle?

c. Compare your estimate with those of other students in your class.

d. If there are 65 Skittles in a bag, how many purple Skittles do you expect to be in the bag? Explain your answer.

Practice and Applications

6. The line graph below shows the relative frequency of a success for 30 trials of an experiment.



a. As the number of trials increased, what relative frequency did the line graph approach?

b. If this experiment were repeated for 500 trials, how many successes would you expect to see?

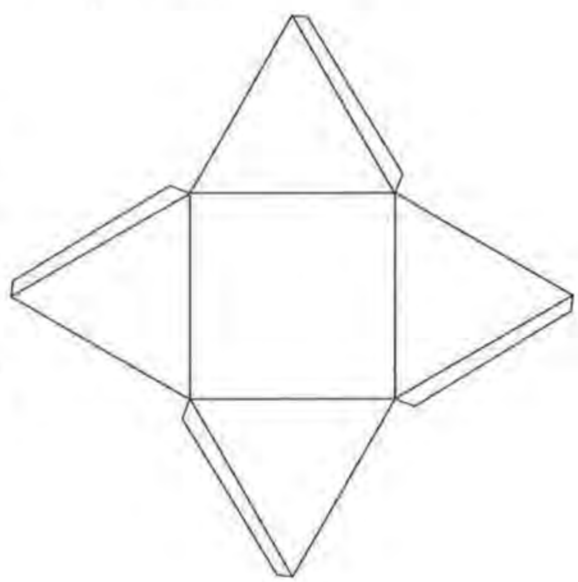
c. What is your estimate for the probability of a success?

d. Think of a situation that might be described by the graph.

7. For this problem, you will need scissors, tape, and a copy of the figure shown below, also on *Activity Sheet 8*.

a. Cut out the figure. Then fold and tape it to form a square pyramid.

b. Conduct an experiment to determine the probability that the pyramid will land on the square base when it is randomly tossed. Your work should include tables, graphs, and an explanation of how you arrived at an answer.



Unit II

Simulation

LESSON 3

Designing a Simulation

How likely is each situation described below?

A basketball player went to the free-throw line 10 times in a game and made all 10 shots.

A student guessed all the answers on a 20-question true-false test and got 18 of them wrong.

All the Skittles in a bag of Skittles were purple.

INVESTIGATE

Unusual? Some people would say that the chances that these events would happen are very small. These three situations have certain common characteristics. They all involve repetition of the same event, and they all have as a goal counting the number of "successes" in a fixed number of repetitions.

OBJECTIVES

Recognize probability problems that are of the form "How many successes occur in *n* repetitions of an event?"

Simulate the distribution of the number of successes in *n* repetitions of an event.

Use simulated distributions to make decisions.

Discussion and Practice

One Outcome per Trial

In this lesson, you will learn how to construct a simulation model for events of the type described above so that you can approximate their probabilities and decide for yourself whether or not the events are unusual.

1. Read the article on page 22 entitled *Non-Cents: Laws of Probability Could End Need for Change.*

NON-CENTS: LAWS OF PROBABILITY COULD END NEED FOR CHANGE

(*Milwaukee Journal,* May, 1992)

Chicago, Ill. [AP] Michael Rossides has a simple goal: to get rid of that change weighing down pockets and cluttering up purses.

And, he says, his scheme could help the economy.

"The change thing is the cutest aspect of it, but it's not the whole enchilada by any means," Rossides said.

His system, tested Thursday and Friday at Northwestern University in the north Chicago suburb of Evanston, uses the law of probability to round purchase amounts to the nearest dollar.

"I think it's rather ingenious," said John Deighton, an associate professor of marketing at the University of Chicago.

"It certainly simplifies the life of a businessperson and as long as there's no perceived cost to the consumer it's going to be adopted with relish," Deighton said.

Rossides' basic concept works like this:

A customer plunks down a jug of milk at the register and agrees to gamble on having the \$1.89 be rounded down to \$1 or up to \$2.

Rossides' system weighs the odds so that over all transactions, the customer would end up paying an average \$1.89 for the jug of milk but would not be inconvenienced by change.

That's where a random number generator comes in. With 89 cents the amount to be rounded, the amount is rounded up if the computerized generator produces a number from 1 to 89; from 90 to 100 the amount is rounded down.

Rossides, 29, says his system would cut out small transactions, reducing the cost of individual goods and using resources more efficiently.

The real question is whether people will accept it.

Rossides was delighted when more than 60% of the customers at a Northwestern business school coffee shop tried it Thursday.

Leo Hermacinski, a graduate student at Northwestern's Kellogg School of Management, gambled and won. He paid \$1 for a cup of coffee and muffin that normally would have cost \$1.30.

Rossides is seeking financial backing wants to test his patented system in convenience stores.

But a coffee shop manager said the system might not fare as well there.

"Virtually all of the clientele at Kellogg are educated in statistics, so the theories are readily grasped," said Craig Witt, also a graduate student. "If it were just to be applied cold to average convenience store customers, I don't know how it would be received."

a. Would you shop at a store that used the system described by Mr. Rossides?

b. Do you think that you would pay too much money under his proposal? Explain your reasoning.

Suppose the soft-drink machine you use charges \$0.75 per can. The scheme proposed by Mr. Rossides requires you to pay either \$0 or \$1, depending upon your selection of a random number. You select a random number between 1 and 100. If the number you select is 75 or less, you pay \$1. If the number you select is greater than 75, then you pay nothing.

2. Simulate the buying of a can of soda by randomly selecting a number between 1 and 100 using a graphing calculator or a similar process. Did you have to pay \$0 or \$1 for the can of soda? Explain your answer.

3. The article suggests that things will even out in the long run. Suppose that over a period time, you purchase 60 drinks from this machine and use a random mechanism for payment each time. This can be simulated by choosing 60 random numbers between 1 and 100. Make such a selection of 60 random numbers using a calculator or a similar process and keep track of how much you paid for each soda.

   a. How many times did you pay \$1? What is the total amount you paid for 60 drinks?

   b. If you had paid \$0.75 for each drink, how much would you have paid for 60 drinks? Does the system proposed by Mr. Rossides seem to balance out in the long run?

4. Now suppose you are buying a box of cookies that cost \$3.23. You pay either \$3 or \$4, depending on the outcome of a random-number selection.

   a. For what values of the random numbers should you pay \$3? For what values of the random numbers should you pay \$4?

   b. Use the rule you determined in part a to simulate what will happen if you and your group buy 100 boxes of these cookies. How many times did you have to pay \$3? \$4?

   c. How much did you and your group pay in all for the 100 boxes of cookies from the simulation? How much would you have paid for the 100 boxes if you had paid \$3.23 per box? Does Mr. Rossides' proposal seem fair? Explain your answer.

Many Outcomes per Trial

You decide to guess at all of the answers on a true-false test. To investigate your chances of correctly answering more than half of the questions, use a simulation.

5. There are five questions on the test, so the simulation must be designed to simulate guessing the answers to the five questions and keep track of the number of correct answers.

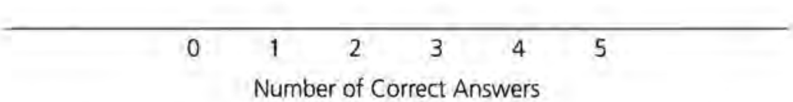
   a. What is the probability of guessing the correct answer for any one question? What could you use to simulate this probability?

   b. One way to simulate guessing the correct answer to a true-false question is to randomly choose a 0 or a 1, with 0 representing a correct answer. Since there are five questions, you will need to select five random numbers. Use a graphing calculator or some other process to select five random numbers from 0 and 1 and count the 0s, or correct answers. Record this number of 0s.

   c. Repeat the procedure for a total of 15 trials. These trials represent taking the test 15 times. Record the number of correct answers for each trial.

   d. Combine the number of correct answers from all members of your group. Using a number line similar to the one below, construct a graph of your group's results by writing an X above the number of correct answers. This graph is the distribution of the number of correct answers.

Distribution of Correct Answers



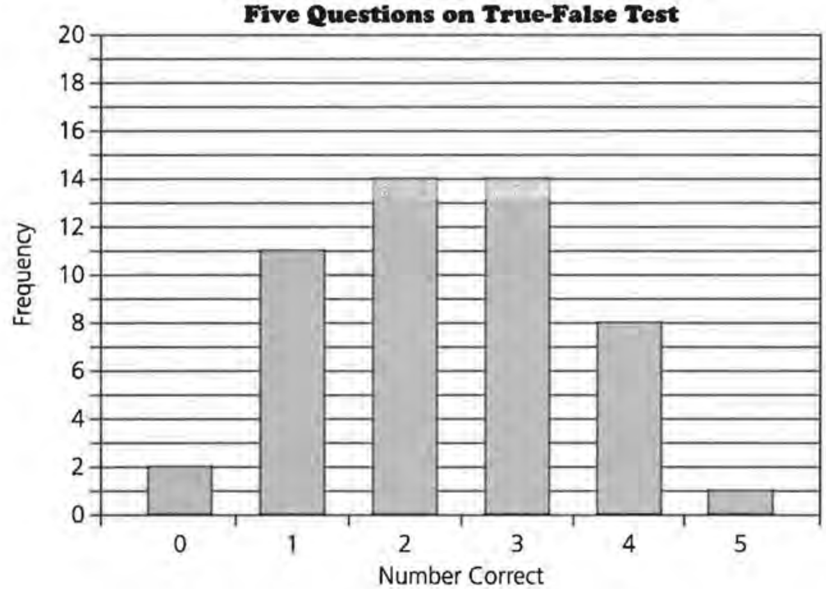
Your distribution of numbers of correct answers may look something like the one shown. In this example, three or more correct answers were obtained 23 times out of 50, or 46% of the time. The average number of correct answers per trial is 2.36. The average, also known as the ***expected value,*** was found by multiplying the number of correct answers by the frequency, finding the sum of these products, and then dividing by the total number of trials.

Distribution of Correct Answers

Number of Correct Answers	Frequency	Product
0	2	0
1	11	11
2	14	28
3	14	42
4	8	32
5	1	5
5	1	5
Totals	50	118

$$\text{Average} = \frac{118}{50} = 2.36$$

Number of Correct Answers Out of Five Questions on True-False Test



e. Approximate the probability that you would correctly answer three or more questions by guessing.

f. What is the average number of questions correctly answered per trial? Show your work.

The average in Problem 5f is an approximation of your expected number of correct answers when you take the test by guessing.

Summary

Many probability questions ask how many times a specific outcome—that is, a success—will occur in a fixed number of opportunities. An example of this situation includes how many questions a student got correct on an exam, how many hits a baseball batter got in his or her last 10 trips at-bats, or how many babies among the 20 born in a local hospital were girls. The distributions of the number of successful outcomes can be simulated by the following steps:

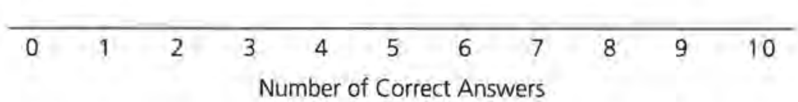
- **i.** Determine the probability for each random selection.
- **ii.** Determine how to construct an event having this probability from random numbers.
- **iii.** Determine how many selections are in a trial of the simulation.
- **iv.** Run many trials of the simulation, recording the number of successes for each.
- **v.** Make an appropriate graph of the results for all trials.
- **vi.** Use the simulated distribution of outcomes to answer questions about probabilities and averages or expected values.

Practice and Applications

6. Suppose you are now guessing your way through a ten-question true-false test. Conduct a simulation for approximating the distribution of the number of correct answers.

- **a.** Does the number of random selections per trial change? If yes, what is it now?
- **b.** With your group, conduct at least 50 trials and record the number of correct answers for each trial. Remember, one trial means you have simulated taking the ten-question test one time.
- **c.** Construct a graph of the results using a number line similar to the one below.

Distribution of Correct Answers



d. Approximate the probability of getting more than half the questions correct. How does this compare with the answer to Problem 5e? If you had to guess at the answers on a true-false test, would you want to take a long test or a short one? Explain your answer.

e. What is your average or expected number of correct answers when guessing on this test? Explain how you found your answer.

7. Suppose you are now guessing your way through a ten-question multiple-choice test in which each question has five possible choices, only one of which is correct. Conduct a simulation for approximating the distribution of the number of correct answers.

a. You must choose a random number to correspond to guessing on this multiple-choice test. What is the probability of guessing the correct answer to any one question?

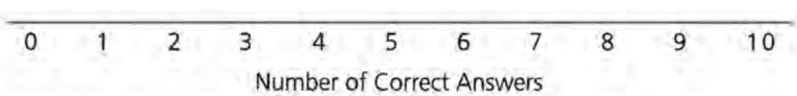
Since there are five choices, you could randomly select a 1, 2, 3, 4, or 5, with a 1 representing the correct answer.

b. For each trial, how many numbers from 1 to 5 must you randomly select?

c. With your group, conduct at least 50 trials and record the number of correct answers for each trial.

d. Make a graph of the results, using a number line similar to the one below.

Distribution of Correct Answers

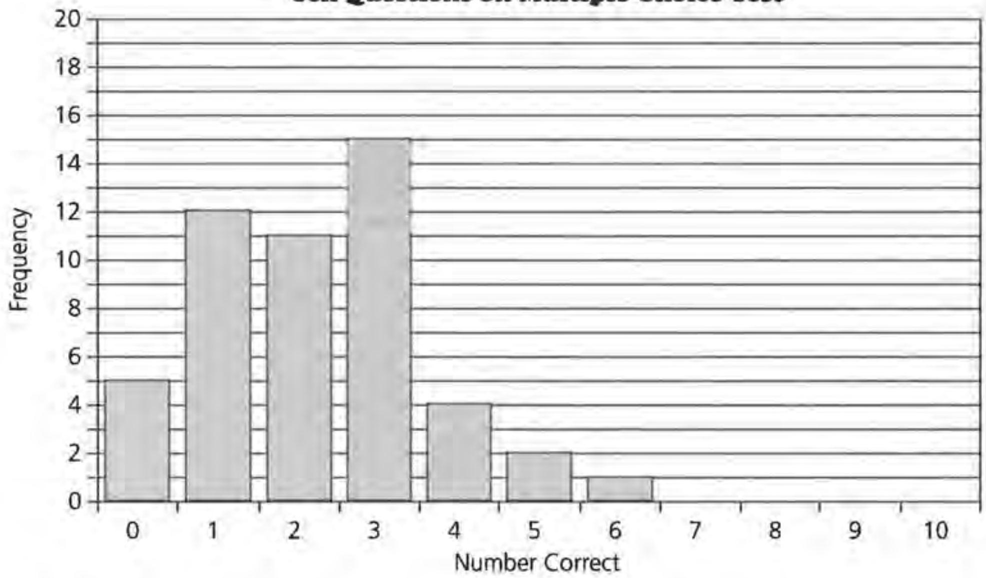


e. What is the approximate probability of getting more than half of the answers correct?

f. What is the average number of correct answers per trial? Show your work.

g. How many answers would you expect to get correct when guessing your way through this test?

Number of Correct Answers out of Ten Questions on Multiple-Choice Test



Your simulation of the multiple-choice test may look something like the one below. Here the chance of getting six or more answers correct is only $\frac{1}{50}$, or 0.02. The average number of correct answers per trial is 2.2.

Extension

8. For each of the simulations in this lesson, identify the following variables:

n = the number of selections per trial

p = the probability of obtaining a specific outcome—that is, a success—on any one selection

m = the average, or ***mean,*** number of successes per trial in the simulation

	n	p	m
Five-Question True-False Test	______	______	______
Ten-Question True-False Test	______	______	______
Ten-Question Multiple-Choice Test	______	______	______

Do you see any relationship between m, calculated from the simulation data, and n and p? Write a formula for m, using the variables n and p.

LESSON 4

Waiting for Success

How many children will a couple have before the first girl arrives?

How many times will I have to play a game with a friend before I finally win? How often will I have to take my driver's test before I pass?

How many medications must a physician try with a sick patient until she finds one that works?

How many times have you considered questions like these? Waiting-time problems such as those above are common in probability and statistics. The problems associated with these questions all have common characteristics that will be invested in this lesson.

OBJECTIVES

Recognize probability problems that are of the form "How long will I have to wait until the first success occurs?"

Simulate the distribution of the number of trials until a success is achieved.

Use simulated distributions to make decisions.

INVESTIGATE

Waiting for a Customer

The president of a local bank knows that 50% of the bank's account customers each have more than two accounts with the bank. The president would like to discuss the bank's services with one customer who has more than two accounts.

Discussion and Practice

1. If the president randomly selected customers, how many customers would he or she have to contact before finding one that has more than two accounts?

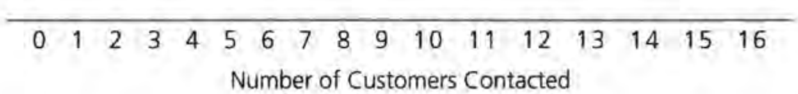
A simulation can be designed to help find the average number of customers that the president would have to contact. Since 50% of the customers have more than two accounts each, you could flip a coin and designate a head as a customer with more than two accounts and a tail as a customer with one or two accounts.

In Lesson 3, you knew the number of trials; in this problem, however, the number of trials is unknown. It is, in fact, the answer to the question "How many randomly selected customers would the bank president have to contact before he or she finds a customer with more than two accounts?"

2. Conduct 30 trials of the simulation. One trial is flipping a coin until you observe the first head and recording the number of times you flipped the coin, representing the number of customers who had to be contacted.

  a. Plot the results of the 30 trials.

Waiting for a Customer



  b. Use the results of your simulation to determine the greatest number of customers that the bank president had to contact.

  c. Find the average number of customers who had to be contacted. Show how you found this average.

Waiting for Blood

The manager of a blood bank needs a donor with blood type B+ (B positive) and knows that about 10% of the donors have type B+ blood. Simulating the distribution of the waiting time until the first B+ donor shows up can help the manager see how long she might have to wait. The simulation can help her decide whether or not to issue a special call for more donors.

3. Set up the simulation for this waiting-time problem.

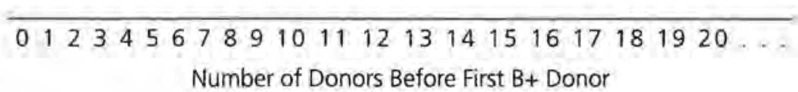
  a. Assume you will be selecting a random number from 1 to 100. Which numbers will represent a donor with B+ blood?

b. Describe how to conduct one trial.

c. Conduct at least 30 trials of this simulation. Record the number of donors up to and including the first B+ donor for each simulated trial.

d. Plot the outcomes of the 30 trials.

Waiting for Blood



e. Describe the shape of this distribution.

f. On the average, how many donors must show up before the first donor with B+ blood shows up?

g. What is the approximate probability that it will require 15 or more donors to find the first donor with B+ blood?

Summary

Many probability problems are of the form "How many times must this event be repeated before success is achieved?" The distribution of the number of repetitions until the first success can be simulated by following these steps:

i. Determine the probability for the basic random event.

ii. Determine how to construct an event with this probability from random numbers or some other randomization device.

iii. Determine how many repetitions make up one trial; this is a key step.

iv. Run many trials of the simulation, recording the number of repetitions in each.

v. Make an appropriate plot of the results for all trials.

vi. Use the simulated distribution of outcomes to answer questions about probabilities and expected values.

Practice and Applications

Many board games that involve the rolling of a number cube provide an advantage to the player who can roll a 6 fairly often. Suppose you are playing a game that requires rolling a 6 before you can take your first turn.

4. Design a simulation to find the number of rolls you might expect to make before you can take your first turn.

   a. Describe what random numbers you used, and describe how to conduct one trial.

   b. Record the results for the 30 trials, and construct a plot of the results.

   c. Describe the shape of this distribution.

   d. Over many plays of the game, what is the average number of rolls needed to begin play?

   e. What is the approximate probability that you will need to roll the die more than 10 times before you can begin play?

5. According to the United States Bureau of the Census, about 12% of American families have 3 or more children. The Gallup Organization wants to find a randomly selected family with 3 or more children in order to test a questionnaire on family entertainment.

   a. Design a simulation that will help the Gallup Organization determine the number of families needed in order to find one with 3 or more children. Write a detailed description of how you conducted your simulation. Your description should include a graph and any calculations.

   b. Do you think that the Gallup Organization will need to select more than 20 families to find one with 3 or more children? Explain your answer.

Extension

6. Problem 3 asks about the waiting time until the first donor with B+ blood arrives at the blood bank. Suppose the blood bank needs two B+ donors.

   a. Design a simulation that will help determine the number of donors that must arrive before two B+ donors arrive at the blood bank. Write a detailed description of how you conducted your simulation. Your description should include a graph and any calculations.

b. How many donors should the blood bank have to check before finding the two B+ donors?

c. How does this waiting-time distribution compare to the one in Problem 3?

7. For each simulation in this lesson, identify the following variables:

p = the probability of obtaining the outcome of interest on any one selection

m = the mean number of repetitions up to and including the first success

Do you see any relationship between m, calculated from the simulation data, and p? That is, can you write the expected value of the number of repetitions to achieve success as a function of p?

Assessment for Units I and II

OBJECTIVE

Apply the concepts of simulation to answer probability questions.

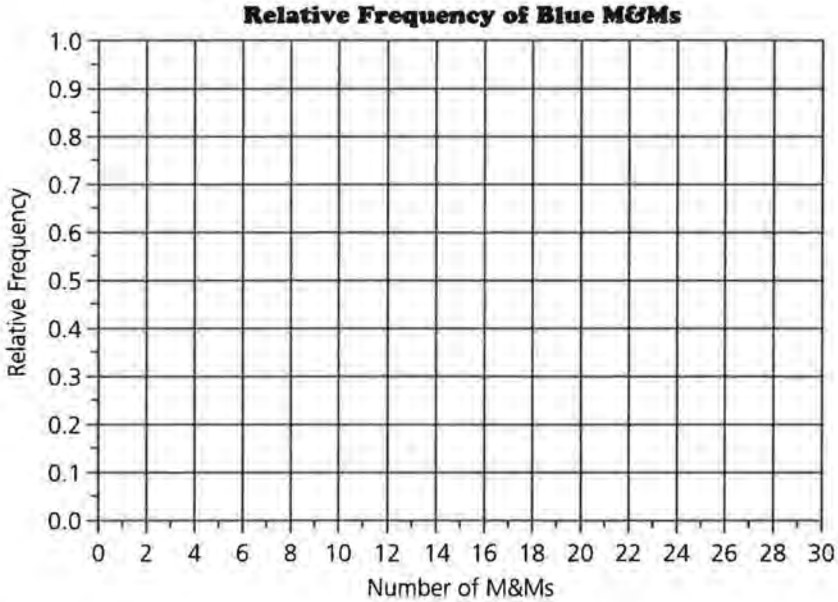
1. A student was interested in the percent of M&Ms® produced that are blue. The student conducted an experiment and collected the following data.

M&M Number	Outcome (Color)	M&M Number	Outcome (Color)
1	Red	9	Brown
2	Brown	10	Green
3	Red	11	Yellow
4	Blue	12	Blue
5	Green	13	Brown
6	Brown	14	Brown
7	Blue	15	Orange
8	Red		

a. Complete the following table, or use the table on *Activity Sheet 9.*

M&M Number	Outcome	Cumulative Number of Blue M&Ms	Relative Frequency of Blue M&Ms
1	______	______	______
2	______	______	______
3	______	______	______
.	______	______	______
.	______	______	______
.	______	______	______
15	______	______	______

b. Use the data from the 15 M&Ms to construct a line graph of the relative frequency of blue M&Ms. Use a grid like the following or the grid on *Activity Sheet 10.*



c. What is your estimate for the probability of randomly selecting a blue M&M?

d. If 15 more M&Ms were sampled, how would this change the line graph? Extend your line graph to show your answer.

e. Predict the number of blue M&Ms in a bag containing 75 M&Ms.

2. The star free-throw shooter on the girls' basketball team makes 80% of her free throws. She takes about 10 such shots per game.

a. Design and conduct a simulation that shows the approximate distribution of the number of successful free throws per game for this player. Show all the steps of the simulation, including what random numbers you used, what constituted one trial, and a plot of the resulting data.

b. What is the approximate probability that she will make more than 80% of her free throws in any one game?

c. How many free throws should she expect to make in a typical game?

d. Over the course of a 15-game season, how many free-throw points should this player expect to score?

e. Are there any assumptions built into the simulation that might not be realistic? Explain your answer.

3. About 33% of the people who come into a blood bank to donate blood have type A+ blood. The blood bank gets about 20 donors per day.

   a. Set up and conduct a simulation that shows the approximate distribution of the number of A+ donors coming into the blood bank per day. Be sure to show all your work for each step of the simulation.

   b. If the blood bank needs ten A+ donors tomorrow, is the bank likely to get them? Explain your answer.

   c. How many A+ donors can the blood bank expect to see each day?

4. Refer to the A+ blood donor in Problem 3. How would the setup of the simulation change if the blood bank were interested in the number of donors who must be seen until the first A+ donor shows up? Outline the simulation; however, you need not carry it out.

Unit III

Data Tables and Probability

LESSON 5

Probability and Survey Results

Do you eat breakfast on a regular basis?

Do other members of your family ever skip meals?

OBJECTIVE

Find an estimate of the probability of an event given the results of a survey.

Students at Rufus King High School in Milwaukee, Wisconsin, as part of a research project, conducted a survey asking the question "Do you eat breakfast at least 3 times a week?" Instead of contacting every student in the school, the students took a ***random sample*** of all the students.

After the survey was conducted, the students used the survey results to draw conclusions about the entire student body. As they organized and analyzed the data from the survey, they noticed similarities between their survey results and the experiments conducted in Lessons 1 and 2. In this lesson you will investigate these similarities and use results from a survey to draw conclusions about the probability of an event.

INVESTIGATE

Breakfast Survey

A ***simple random sample*** is a sample chosen in such a way that every possible sample of a given size has an equal chance of being selected.

To select 50 students from the student body, the students could put the name of every student on a card, put the cards in a box, mix them up, and draw 50 cards. Since every possible sample of 50 students has the same chance of being chosen, this would be one way to take a random sample of the entire student body.

Can you think of some other methods that the students might have used to take a random sample of the entire student body? Do you think your class is a random sample of the entire student body in your school?

Discussion and Practice

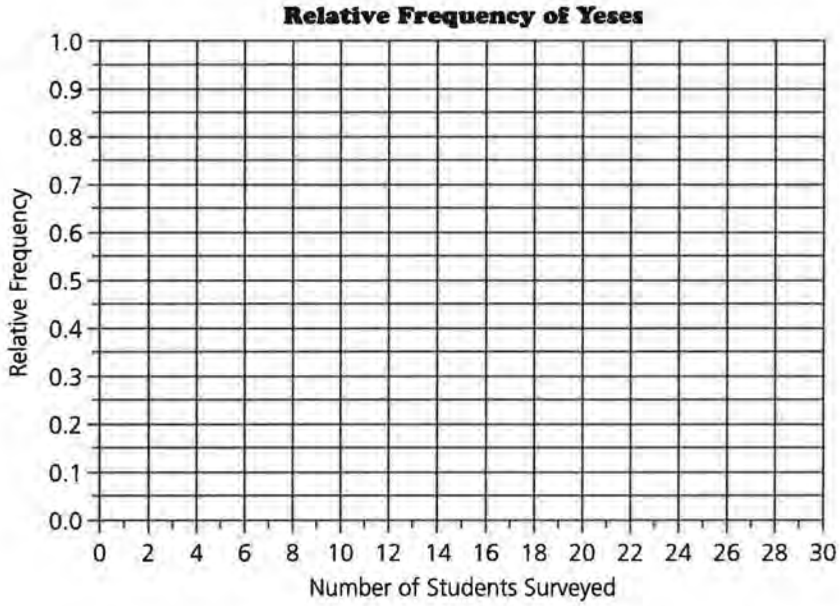
In this survey, your class will take a random sample of 30 students in your school. To ensure that you are taking a random sample, your class might give each student in school a number and then select 30 numbers at random.

1. Ask each student in the survey this question: "Do you eat breakfast at least 3 times a week?" Collect the results of the survey in a table similar to the following, also available on *Activity Sheet 11*.

Do You Eat Breakfast?

Student	Outcome (Yes or No)	Cumulative Number of Yeses	Relative Frequency of Yeses
1	______	______	______
2	______	______	______
3	______	______	______
.	______	______	______
.	______	______	______
.	______	______	______
30	______	______	______

2. Use the results of your survey to construct a line graph of the relative frequency of yeses. Use a grid like the one on page 41, also available on *Activity Sheet 12*.



a. As you tallied more and more surveys, describe what happened to the line graph.

b. As the number of students surveyed increased, what relative frequency does the line graph approach? Draw a horizontal line across your graph at this point.

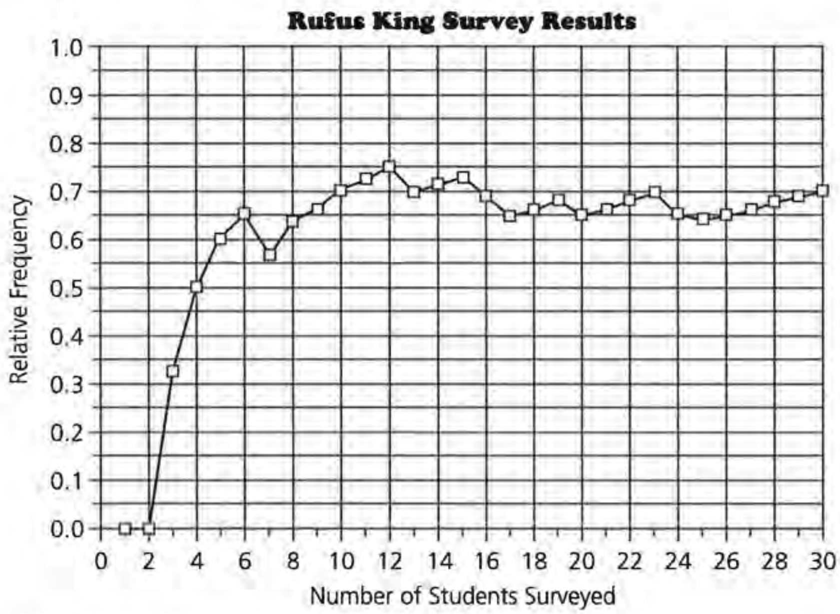
c. What does the line graph of relative frequency tell you about the percent of students at your school who eat breakfast at least 3 times a week?

3. Use the results from the survey in Problem 2.

a. Suppose you randomly selected one student. What do you think the probability is that the student will answer "yes" to the question "Do you eat breakfast at least 3 times a week?"

b. If you randomly sampled 100 students, how many students would you expect to say that they eat breakfast at least 3 times a week?

4. The line graph below shows the results of the survey conducted by the Rufus King students. The graph shows the variation in the relative frequency of "yes" responses to the question "Do you eat breakfast at least 3 times a week?" Compare the results of your class survey to those of the King students.



The Rufus King students sampled more than 30 students. The following table summarizes the survey results after the students tallied the remaining surveys collected at the school.

Number of Students	Number of Yeses	Relative Frequency of Yeses
40	30	0.75
80	59	0.74
100	74	0.74
150	112	0.75

a. What do the data tell you about the percent of Rufus King students who eat breakfast at least 3 times a week?

b. If one King student were randomly selected, what do you think the probability is that the student will say "yes" to the question "Do you eat breakfast at least 3 times a week?"

c. If 850 students attend Rufus King High School, how many students would you expect to eat breakfast at least 3 times a week? Explain how you arrived at your answer.

Different Survey Results and Probability

5. A survey with a large sample size can be used to answer questions concerning the entire population under study. A poll asked a random sample of 2500 high-school students the question "When you go out on a date, who pays for the date?" The results are shown below.

Responses	Frequency
Boy	1200
Split Costs	800
Girl	75
Girls' Parents	25
Boys' Parents	25
Don't Date	375
Total	2500

The results of this survey can be used to draw some conclusions about the population of high-school students.

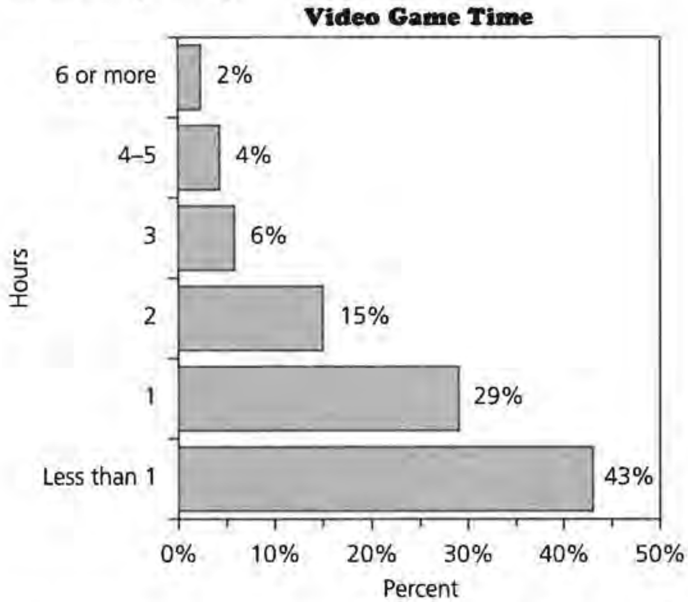
a. Make a table of the data with a column added for relative frequency of responses. Calculate the relative frequency for each response as a fraction, a decimal, and a percent. Find the sum of the relative frequencies.

b. Suppose that only one high-school student were randomly selected. What is an estimate of the probability that the student would say the costs are split when they date?

c. What is an estimate of the probability that a randomly selected student would say that he or she does not date?

In Problem 5, the results of the survey were given in terms of the number of people in each category. Many times, survey results are reported using the percent of outcomes in each category. These percents can also be thought of as relative frequencies. Since the survey had a very large sample, these percents can be used as an estimate for the probability of individual outcomes.

6. A study reported that nearly 90% of children aged 9–13 play video games an average of 1.4 hours a day. The graph below shows the percent of children playing a given number of hours per day.



a. If the survey consisted of 750 children aged 9–13, how many of these children would say they play video games 3 or more hours a day?

b. If a child aged 9–13 is randomly selected, what is an estimate of the probability that the child plays video games less than 1 hour a day?

c. If a child aged 9–13 is randomly selected, what is an estimate of the probability that the child plays video games more than 1 hour a day?

d. What is the sum of the percents given in the graph? Why don't they add to 100%?

7. A survey reported the age distribution and the percent of people who purchased running shoes during the last year. The results shown below describe the distribution.

Age of Purchasers	Percent
Under 14 Years Old	8.3%
14 to 17 Years Old	10.5%
18 to 24 Years Old	10.0%
25 to 34 Years Old	28.5%
35 to 44 years Old	24.0%
45 to 64 Years Old	15.2%
65 Years Old and Over	3.5%

a. Construct a bar graph with the ages of the purchasers as individual categories and the percent as the height or length of each bar. This graph represents the distribution of running-shoe buyers. What do you notice about the age ranges in each category?

b. Describe the age of the "typical" running-shoe buyer. Explain your answer.

c. What is an estimate of the probability that a randomly selected purchaser of running shoes is

i. 14 to 17 years old?

ii. 45 years old or older?

iii. 18 to 34 years old?

iv. 17 years and under?

v. at least 18 years old?

The table of the data contains all the possibilities of ages of purchases of running shoes. If you think of the percents as probabilities, then the sum of the percents gives the total of the probabilities. The sum of the probabilities of all the outcomes is exactly 100%, or 1. (Note: Due to rounding, there are times when the sum of the decimals or percents used is not exactly 1.)

8. The data below show how long cattle ranches have been in one family, as reported by the *USA Today* on July 8, 1994. The first column gives the number of years the ranch has been in one family. The second column gives the cumulative percent of ranches having been in one family fewer than the given number of years. For example, 59% of ranches have been held in one family for fewer than 50 years. This figure also includes the 21% that have been in one family for fewer than 25 years.

Number of Years Owned by One Family	Percent
Fewer than 25 years	21%
Fewer than 50 years	59%
Fewer than 100 years	89%

a. How do the percents given in this chart differ from the percents given in previous problems?

b. If a cattle ranch is randomly selected, what is the probability that the ranch has been held in the family

   i. for fewer than 50 years?

   ii. for 100 or more years?

   iii. for 25 to 49 years?

Summary

The results of surveys can be given in many different forms. The data can be presented in terms of items as in the "Who pays on a date?" survey. To find the probabilities of these results, the data should be converted to a relative frequency given as a fraction, a decimal, or a percent.

Sometimes the results of a survey are given as a percent. The percent could represent the information about a single item as in the video-game-time data or it could be a cumulative percent as in the cattle-ranch problem. Careful attention must be given to what percents represent.

Practice and Applications

9. The table below shows the distribution of methods of generating electricity in the United States as reported by the Energy Information Administration.

Method	Percent
Coal	56.9%
Nuclear	21.2%
Gas	9.0%
Hydro Power	9.3%
Oil	3.5%
Other	

a. What is the percent for the *Other* category?

b. List some other ways by which electricity is generated in the United States.

10. According to the United States Bureau of the Census, the number of cars available to American households is given by the following cumulative percents.

Number of Cars	Cumulative Percent of Households
0	15%
1	62%
2	91%

If a household is randomly selected, what is the probability that the household will have

a. one car available?

b. two or fewer cars available?

c. three or more cars available?

11. Find a survey from a newspaper or magazine. Summarize the results of the survey and make a list of four probability questions that could be asked using the results of the survey.

LESSON 6

Compound Events

Do you know what a census is?

How often is a census of the United States conducted?

What kind of information could be summarized from the data gathered in a census?

OBJECTIVES

Determine the probability that both event A and event B will occur.

Determine the probability that either event A or event B will occur.

Organize data on two variables in a two-way table.

In the preceding lessons, we introduced the basic concept of probability by examining the relative frequency of an outcome as the number of trials increased. In some investigations, *every* member of the population is measured or counted. This is called a ***census***. The U.S. Bureau of the Census is involved in counting the entire population of the United States. Why is a census of the United States important?

For census data, the exact probability of an event can be found. School records contain the age of every student in your school; therefore, you could find the exact probability of randomly selecting a 14-year-old student.

INVESTIGATE

Movie Survey

As a class, identify a movie and encourage your classmates to answer the question "Did you like the movie?" with "yes," "no," or "didn't see." Record the data in a table similar to the one below, using Y for "yes," N for "no," and DS for "didn't see." Save your table for future work.

Student Name	Response (Y, N, DS)
________	________
________	________
________	________
________	________

Discussion and Practice

1. Combine the students' responses from your table like the one below to show the number of each response.

Movie Responses

Response	Number
Yes	________
No	________
Didn't See	________
Total	________

2. Use the results from your survey in Problem 1.

a. Find the relative frequency of the "yes" responses.

b. Find the relative frequency of the "no" responses.

3. Use your data from Problem 1.

a. If a student from your class is randomly selected, what is the probability that he or she answered "yes"?

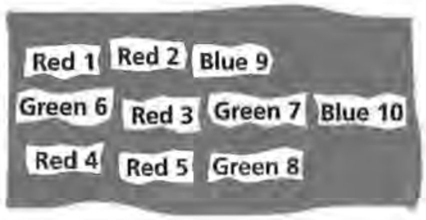
b. Is your answer to part a an estimate of the probability or the exact probability? Explain your answer.

c. If one student from your class is randomly selected, what is the probability that the student answered "no" or "didn't see" the movie?

d. What is the sum of the relative frequencies for the three responses? Explain why the relative frequencies must add up to this number.

The *AND* Rule

4. A brown paper bag contains ten slips of paper. As shown below, five of the slips are red labeled with the numbers 1 to 5; three of the slips are green labeled with the numbers 6 to 8; and two of the slips are blue labeled with the numbers 9 and 10.



a. If one slip of paper is randomly selected, what is the probability

i. that an even number will be drawn?

ii. that a blue slip will be drawn?

iii. that an even-numbered blue slip will be drawn?

b. If one slip of paper is randomly selected, what is the probability

i. that an odd number will be drawn?

ii. that a green slip will be drawn?

iii. that an odd-numbered green slip will be drawn?

c. One student found the answer to the last question using the following method:

probability of green slip = $\frac{3}{10}$

probability of odd-numbered slip = $\frac{5}{10}$

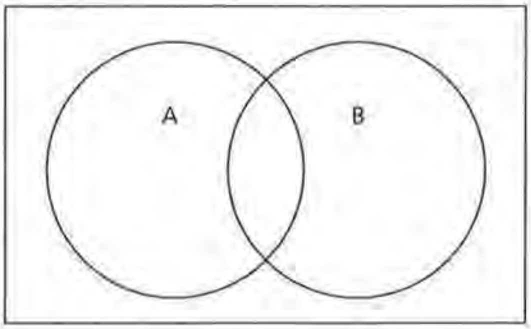
probability of odd-numbered green slip = $\frac{3}{10} + \frac{5}{10} = \frac{8}{10}$

Another student disagreed and said the answer was $\frac{1}{10}$.

Which student do you think is correct? Explain why.

Venn diagrams can be used to help find the probability that an odd-numbered green slip will be drawn. A Venn diagram is a picture showing how two or more events are related.

5. In the first Venn diagram below, let circle A represent the event that a randomly drawn slip is green, and let circle B represent the event that a randomly drawn slip contains an odd number.



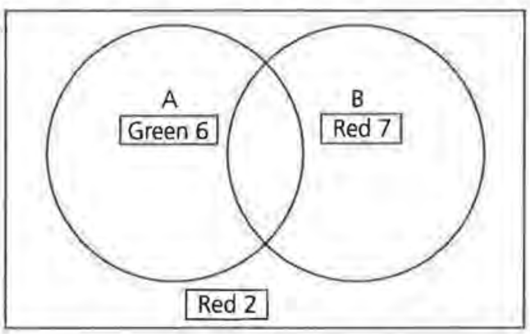
a. How many slips of paper would be represented by circle A?

b. How many slips of paper would be represented by circle B?

c. What is represented by the area where the two circles overlap? How many slips of paper are represented by this area?

d. What is represented by the area outside the two circles? How many slips of paper are represented by this area?

e. Three slips of paper are shown in the Venn diagram below and on *Activity Sheet 13*. Give the proper location for the other seven slips.



The overlap of the two circles is called the ***intersection*** of event *A* and event *B*. This can be represented by ***A and B***. ***P*(*A and B*)** represents the probability that both event *A* and event *B* will occur.

6. Let event A represent the drawing of a red slip and event B represent the drawing of an odd slip.

a. Draw a Venn diagram showing the location of each of the ten slips of paper. You can use *Activity Sheet 13.*

b. Find $P(A \text{ and } B)$.

Summary

P*(*A and B*)** is the probability that event A and event B occur together. The word *and* indicates the ***intersection, or overlap, of the two events.

7. A dealership had two colors of cars in stock. When he compared the compact cars and full-sized sedans, the dealer recorded the information in the table. How many of each type of car does he have in stock?

		Color	
		Dark Green	**Red**
Type of Car	**Compact**	5	12
	Full-Sized	9	6

a. Dark-green cars

b. Compact cars

c. Dark-green compact cars

d. Draw a Venn diagram showing the relationship between dark-green cars and compact cars. Label each section of the Venn diagram and indicate how many cars are in each section. You can use *Activity Sheet 13.*

e. If a car is selected at random, what is the probability that the car is a dark-green compact car?

The *OR* Rule

As a class, identify another movie and have your classmates answer the question "Did you like the movie?" Again, students may answer "yes" (Y), "no" (N), or "didn't see" (DS). Add the data to the table you made in Problem 1, as shown below.

	Movie 1	Movie 2
Student Name	**Responses (Y, N, DS)**	**Responses (Y, N, DS)**
________	________	________
________	________	________

8. Combine the responses for the second movie in a table similar to the one below.

Student	Response (Y, N, DS)
1	______
2	______
3	______
4	______
.	______
.	______
.	______

9. If one student from your class is randomly selected, what is the probability that

   a. he or she answered "yes" to the Movie 2 question?

   b. he or she answered "no" or "didn't see" to the Movie 2 question?

   c. he or she answered "yes" or "didn't see" to the Movie 2 question?

   d. he or she answered "yes," "no," or "didn't see" to the Movie 2 question?

The table below shows a way to organize the results of the two survey questions. This ***two-way table*** can be very useful when investigating whether or not there is a relationship between two variables.

		Movie 2 Responses			
		Y	N	DS	Totals
Movie 1 Responses	Y	a	b	c	
	N	d	e	f	
	DS	g	h	i	
	Totals				

The cell labeled "a" will contain the number of students who answered "yes" to the Movie 1 question and "yes" to the Movie 2 question.

10. Write a description of what each cell represents.

   a. cell b b. cell f

   c. cell g d. cell i

11. Use the class data compiled in Problems 1 and 8. Record the results of responses in a table similar to the one below.

		Movie 2 Responses			
		Y	N	DS	Totals
Movie 1 Responses	Y				
	N				
	DS				
	Totals				

12. Use your table to give the number of students in each of the following categories.

a. Didn't see either movie

b. Didn't see Movie 1 and liked Movie 2

c. Didn't see Movie 1 or didn't see Movie 2

When finding the probability that event A or event B will occur, you can use the ***Addition Rule.***

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B),$$

where $P(A \text{ and } B)$ is the probability that event A and event B will occur together. You can think of the word *or* as the ***union*** of the two events.

13. One student found the answer to Problem 12c by adding the total number of students who said they didn't see the first movie to the total number who said they didn't see the second movie. Another student disagreed with this method because she said some students were counted twice. Do you think the second student is correct? Explain.

14. If one student from your class is randomly selected, what is the probability that the student

a. said he or she liked the first movie?

b. said he or she did not see the second movie?

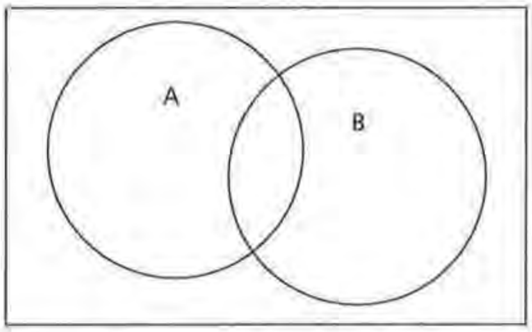
c. said he or she liked the first movie and did not see the second movie?

d. said he or she liked the first movie or did not see the second movie?

e. said he or she did not like the first movie and did not see the second movie?

f. said he or she did not see the first movie or did not like the second movie?

15. In the Venn diagram below, circle A represents students who did not see the first movie and circle B represents students who did not see the second movie.



a. Copy the Venn diagram above or use *Activity Sheet 13*. Label each distinct section with the correct number of students.

b. Shade the part that represents those students who did not see either movie.

c. On another copy of the Venn diagram, shade the part that represents those students who did not see the first movie or did not see the second movie.

d. On a third copy of the Venn diagram, shade the part that represents those students who saw the first movie.

Summary

***P*(*A or B*)** is the probability that event *A* or event *B* will occur. You find the probability by adding the probability that event *A* will occur to the probability that event *B* will occur and then subtract the probability that event *A* and event *B* will occur together. You can think of the word *or* as the *union* of the two events.

Practice and Applications

16. The two-way table below shows data collected from a group of students in a science class who had been studying the relationship between eye dominance and hand dominance.

		Eye Dominance		
		Left	**Right**	**Totals**
Hand Dominance	**Left**	3	2	5
	Right	5	14	19
	Totals	8	16	24

Let A represent the set of students who are left-hand dominant.

Let B represent the set of students who are right-eye dominant.

Let C represent the set of students who are right-hand dominant.

a. Draw a Venn diagram that shows the relationship between left-hand and right-eye dominance. You can use *Activity Sheet 13*. On the diagram label all the sections and show the number of students in each circle, in the intersection of the two circles, and in the exterior of both circles.

Since A represents the set of students who are left-hand dominant, then the symbol $P(A)$ represents the probability of selecting a student who is left-hand dominant.

b. Use the descriptions for A, B, and C to describe each of the following and find the indicated probability.

i. $P(C)$

ii. $P(B)$

iii. $P(A \text{ and } B)$

iv. $P(A \text{ or } B)$

v. $P(B \text{ and } C)$

vi. $P(B \text{ or } C)$

17. A government agency was interested in how the descriptions of workers in the United States changed from 1985 to 1995. The table below shows the relative frequencies of new workers from 1985 to 1995.

Description of Workers

	White	Nonwhite	Immigrant	Totals
Men	0.15	0.07	0.13	0.35
Women	0.42	0.14	0.09	0.65
Totals	0.57	0.21	0.22	1.00

a. What is the probability that a randomly selected new worker is

i. an immigrant?

ii. a woman?

iii. an immigrant and a woman?

iv. an immigrant or a woman?

v. white or nonwhite?

vi. a man or nonwhite?

vii. nonwhite or an immigrant?

b. For which of the questions in part a can the Addition Rule be used?

Complementary Events

If 3 students in a class of 28 are absent, how many students are not absent?

If you toss a number cube and it does not land with a 5 up, what numbers might be showing?

If every member of your family went to a movie, how many stayed home?

OBJECTIVE

Find the probability of the complement of an event.

Each of the situations above asks you to find the *complement* of an event. In your class, the female students are the complement of the male students. In the alphabet, the vowels are the complement of the consonants.

INVESTIGATE

In this lesson, you will study a number of situations which involve events and their complements.

Discussion and Practice

1. A school administrator, interested in how many students were absent from an 8th-hour study hall, kept track of absences over the last 40 days of school. The table shows that there were 7 days when no students were absent, 8 days when 1 student was absent, and 4 days when 2 students were absent.

Number of Students Absent	Days
0	7
1	8
2	4
3	6
4	6
5	3
6	2
7	2
8	1
9	0
10	1
Total	40

a. For how many days were 5 students absent from study hall?

b. For how many days was at least 1 student absent from study hall? Explain how you found your answer.

c. For how many days were at least 5 students absent from study hall? Explain how you found your answer.

d. For how many days were fewer than 10 students absent from study hall?

When you found the number of days with fewer than 10 students absent from study hall, you were finding the complement of the number of days with 10 students absent. The ***complement*** of event A is the event that A does *not* occur, written "not A."

2. Consider the data given in Problem 1.

a. If A represents the number of days that no students were absent from study hall, what is the complement of A?

b. If B represents the number of days that at least 1 student was absent from study hall, what is the complement of B?

3. If the school administrator randomly selects one of the 40 days, what is the probability

a. that no students were absent from study hall?

b. that at least 1 student was absent from study hall?

4. What is the sum of the probabilities from Problems 3a and 3b?

Rule of Complementary Events

$P(A) + P(\text{not } A) = 1$

$P(\text{not } A) = 1 - P(A)$

$P(A) = 1 - P(\text{not } A)$

The rule above can be used to find the probability that 1 or more students were absent, or the complement of no students absent.

$$P(\text{one or more students absent}) = 1 - P(\text{no students absent})$$
$$= 1 - \frac{7}{40}$$
$$= \frac{33}{40}$$

5. The following table gives the approximate relative frequency of Americans that have certain blood types.

Type	Relative Frequency
O	0.45
A	0.40
B	0.10
AB	0.05

Let O represent a person who has type O blood.

Let A represent a person who has type A blood.

Let B represent a person who has type B blood.

Let AB represent a person who has type AB blood.

a. What is the complement of A?

b. If a random American is chosen, find each of the following:

i. $P(A)$

ii. $P(O)$

iii. $P(A \text{ or } O)$

iv. $P(\text{not } B \text{ and not } A)$

6. Below is the tally of the number of wrong answers that a class of 30 students received on a mathematics test.

Number Wrong	Number of Students
1	/
2	//
3	/
4	////
5	~~////~~
6	////
7	///
8	/
9	/
10	
11	//
12	/

The total number of students is 30.

a. If A represents the event that a student got 10 answers wrong, what is the complement of A?

b. If B represents the event that a student got 2 or more answers wrong, what is the complement of B?

c. Use the Rule of Complementary Events to find P(2 or more answers wrong).

d. What is the probability that a randomly selected student

   i. got 10 wrong?

   ii. did not get any answers wrong?

   iii. got at least 1 answer wrong?

   iv. got 4, 5, or 6 answers wrong?

   v. got at most 1 answer wrong?

   vi. did not get exactly 4 answers wrong?

e. Explain how the Rule of Complementary Events could be used for part d iii.

Practice and Applications

7. In the late 1960s and the early 1970s, men were drafted into the U.S. Army by random selections of birthdays. Assume that each of the 366 birthdays has an equal chance of being chosen.

   a. If B represents the event that a person was born in June, what is the complement of B?

   b. Find the probability that a randomly selected birthday is not in June.

   c. Explain how the Rule of Complementary Events helps to answer part b.

   d. What is the probability that a randomly selected birthday is

      i. in June or July?

      ii. the 31st of a month or in October?

      iii. not in September or November?

      iv. not June 20th?

8. The table below contains Census data on the education attainment for adults more than 65 years old.

Education	Number (thousands of persons)
Did Not Complete High School	13,183
Completed High School	9,412
College, 1–3 Years	2,915
College, 4 or More Years	3,018
Total	28,528

   a. If E is the event that a person did not complete high school, what is the complement of E?

   b. Find the probability that a randomly selected person did not complete high school.

   c. Find the probability that a randomly selected adult more than 65 years old

      i. completed high school.

      ii. completed some college but less than 4 years of college.

      iii. completed high school or college.

9. During the 1992 Winter Olympics, a random survey asked 2600 American adults these questions.

Question 1: What city hosted the 1992 winter Olympics?

Question 2: What city will host the 1996 Summer Olympics?

The table below shows the relative frequencies for the results of the answers to these two questions.

		Question 1	
		Knew	**Did Not Know**
Question 2	**Knew**	0.54	0.30
	Did Not Know	0.01	0.15

If a randomly selected American adult had been selected, what is an estimate of the probability that

a. he or she knew the answer to Question 1?

b. he or she knew the answer to Question 2?

c. he or she knew the answer to both questions?

d. he or she knew the answer to Question 1 or Question 2?

e. he or she did not know the answer to either question?

f. If A represents the event that a person did not know what city would host the 1996 Summer Olympics, what is the complement of A?

Conditional Probability

When you survey people with two questions, does a "yes" response to one question help you to predict the answer to the other question?

How do basic rules of probability help you to make predictions?

OBJECTIVES

Construct and interpret relative frequencies from columns or from rows of a table.

Interpret column or row relative frequencies as conditional probabilities.

In the preceding lessons, you have investigated several of the basic rules of probability. Many of the examples involved the examination of data, summarized in a two-way table, that resulted from asking people two questions. Is a person who has seen one popular movie likely to have seen another? Is a person who has right-hand dominance likely to have right-eye dominance?

INVESTIGATE

In this lesson, you will use tables to help you understand probability and make predictions.

Discussion and Practice

According to the United States Bureau of the Census, the number of family households in the country classified according to household head and status of children for 1993 is as shown in the following table.

	Married-Couple Household Head (millions)	Male Household Head (millions)	Female Household Head (millions)	Totals
No Children Under 18	28.4	1.7	4.7	[illegible]
Children Under 18	24.7	1.3	7.2	[illegible]
Totals	53.1	3.0	[illegible]	[illegible]

Source: United States Bureau of the Census

1. Use the data from the table above for these problems.

a. What does the number 28.4 million represent?

b. Copy and complete the table.

c. How many households headed by a married couple have children under 18?

d. How many households headed by a male have no children under 18?

e. What is the total number of family households in the United States?

2. Use the data from the table above for these problems. If the A.C. Nielsen company, which conducts TV ratings surveys, randomly selected a household, what is the probability that the household

a. is headed by a female?

b. has children under 18?

c. is headed by a male and has children under 18?

3. In the Rufus King High School survey introduced in Lesson 5, two of the many questions that were asked were:

Do you consider your diet to be healthy?

Do you eat breakfast at least 3 times a week?

Among the 50 students in a random sample from the school, 20 answered "yes" to the first question and 36 answered "yes" to the second question.

Several questions are listed below. Which of the questions can be answered from the survey data given and which cannot? If a question cannot be answered, explain why not.

a. How many students do not eat breakfast at least 3 times a week?

b. How many students eat breakfast at least 3 times a week and consider their diet healthy?

c. How many students do not consider their diet healthy?

d. If the total enrollment in the school is 1200 students, approximately how many students in the school would you expect to eat breakfast at least 3 times a week?

Two-Way Tables

When analyzing the Rufus King data, it was not possible to answer the question "How many surveyed students eat breakfast at least 3 times a week and consider their diet to be healthy?" To answer a question like this implies that there are responses for each student on both parts of the question. If you are interested in investigating a relationship between the two variables, eating breakfast and perceptions of healthy diet, you must keep track of how each individual responded to both parts. As you saw in Lesson 6, one way to do this is to record the data in a two-way table.

4. The two-way table below shows the results of the Rufus King survey. The 17 represents the number of students who answered "yes" to both questions.

		Do You Think Your Diet Is Healthy?		
		Yes	**No**	**Totals**
Do You Eat Breakfast at Least 3 Times a Week?	**Yes**	17		36
	No			
	Totals	20		50

a. Copy and complete the table above.

b. Compare your results to those of others in your group.

c. How many students answered "no" to the breakfast question?

d. How many students answered "no" to the healthy-diet question?

e. How many students answered "no" to both questions?

Totals for each row and column are called ***marginal totals.*** The 36 in the first row of the table above is the marginal total of the students who said "yes" to the breakfast question; 20 is the marginal total of the number of students who answered "yes" to the diet question. To help in an analysis of the data, it is sometimes helpful to convert the marginal totals to marginal relative frequencies or percents.

5. Use the table from the Rufus King survey.

   a. Copy and complete the table below by converting the marginal totals to marginal relative frequencies.

		Do You Think Your Diet Is Healthy?		
		Yes	No	Totals
Do You Eat Breakfast at Least 3 Times a Week?	Yes			$\frac{36}{50} = 0.72$
	No			
	Totals	$\frac{20}{50} = 0.40$		$\frac{50}{50} = 1.00$

   b. What does the marginal relative frequency 0.72 tell you?

   c. What is the probability that a student randomly selected from the sample of 50 students will answer "yes" to the breakfast question?

   d. What does the marginal relative frequency under the *No* column for the diet question tell us?

   e. Of the 1200 students in the school, approximately how many would you expect to consider their diet healthy?

The values in the table provide information on the ***joint*** behavior of students in response to the two questions. The value 17, which is the number of students who answered "yes" to both questions, is called a ***joint frequency.***

6. Consider the marginal relative frequencies from the table above.

   a. What is the joint frequency of the students who answered "yes" to the diet question and "no" to the breakfast question?

   b. What is the joint frequency of the students who answered "no" to the diet question and "yes" to the breakfast question?

To convert the data to ***joint relative frequencies,*** divide each joint frequency by the total number. The joint relative frequency of the students who answered "yes" to both questions is $\frac{17}{50} = 0.34$.

7. Use the table in Problem 4.

a. Convert the joint frequencies to joint relative frequencies.

b. Compare your results to those of others in your group.

c. What does the joint relative frequency of $\frac{17}{50}$, or 0.34, represent?

d. What does the joint relative frequency in the cell represented by "yes" on the diet question and "no" on the breakfast question represent?

e. What is the probability that a randomly selected student from among the 50 students selected in the sample will answer the diet question "yes" and the breakfast question "no"?

If a student said that he or she ate breakfast at least 3 times a week, is this student more likely to say that he or she has a healthy diet than an unhealthy diet? To help answer this question, you will need to study the first row of the table—that is, the row that represents the students that said "yes" to the breakfast question.

		Do You Think Your Diet Is Healthy?		
		Yes	**No**	**Totals**
Do You Eat Breakfast at Least 3 Times a Week?	**Yes**	17	19	36

From this total of 36 students, 17, or 47%, of the students said "yes" to the diet question. The 47% is called the ***conditional relative frequency.***

8. Use the data above to answer the following questions.

a. Among the students known to eat breakfast at least 3 times a week, what is the conditional relative frequency of students who think their diet is healthy?

b. Among the students known to eat breakfast at least 3 times a week, what is the conditional relative frequency of students who think their diet is not healthy?

c. If a student said that she ate breakfast at least 3 times a week, is she more likely to say that she has a healthy diet than an unhealthy diet?

d. Of the 1200 students in the school, approximately how many would you expect to eat breakfast at least 3 times a week? Among these breakfast-eaters, approximately how many would you expect to think their diet is healthy?

The conditional relative frequency can also be thought of as a ***conditional probability.***

9. Consider those students who said "no" to the breakfast question and how they answered the diet question.

		Do You Think Your Diet Is Healthy?		
		Yes	**No**	**Totals**
Do You Eat Breakfast at Least 3 Times a Week?	**No**	3	11	14

a. If a student is randomly selected from those who are not regular breakfast-eaters, what is the conditional probability that this student will say that his or her diet is healthy?

b. If a student is randomly selected from those who are not regular breakfast-eaters, what is the conditional probability that this student will say that his or her diet is not healthy?

c. Of the 1200 students in the school, approximately how many would you expect to not eat breakfast at least 3 times a week? Among those who don't eat breakfast, approximately how many would you expect to think their diet is healthy?

Summary

Data from variables that classify items into categories can be summarized by recording the frequencies for each category. If two categorical variables are to be compared, the frequencies can be arranged in a two-way table. A ***joint relative frequency*** or probability is the ratio of a cell frequency to the overall number of times classified in the table. A ***marginal relative frequency*** or probability is the ratio of a row or column total to the overall frequency of items for the whole table. Marginal data provide information on either the row or column variable by itself. A ***conditional relative frequency*** or probability is the ratio of a cell frequency to either its row total or its column total.

Practice and Applications

10. The Rufus King High School survey also included the following two questions:

Do you like school?

Do you participate in a sport at school?

a. Copy and complete the table below.

		Do You Like School?		
		Yes	No	Totals
Do You Participate in a Sport?	Yes		14	32
	No	12		
	Totals			50

b. In your table, circle the values that represent the marginal totals.

c. Construct a table showing the data converted to marginal and joint relative frequencies.

d. What is the approximate probability that a randomly selected student will answer "yes" to the school question?

e. What is the approximate probability that a randomly selected student will answer "yes" to both questions?

f. Consider only the students who said that they have participated in a sport. What is the conditional relative frequency of those students who like school within this group?

g. Consider only the students who said that they have participated in a sport. What is the conditional relative frequency of those students who do not like school within this group?

h. Of 1200 students in the school, approximately how many would you expect to have participated in a sport? Among those who participated, approximately how many would you expect to like school?

i. Consider only the students who said that they do not like school. What is the conditional relative frequency of the students who do not participate in sports within this group?

j. Consider only the students who said that they like school. What is the conditional relative frequency of those students who do not participate in sports within this group?

11. The following data were collected from a high-school mathematics class.

Number of Letters in First Name

	3 or Fewer	4 to 6	More than 6	Totals
Male	1	6	3	
Female	3	8	5	
Totals				

a. Copy the table and fill in the marginal totals for each row and column.

b. What percent of students have 4 to 6 letters in their first names?

c. What percent of females have 4 to 6 letters in their names?

d. If you were to choose a student from this class at random, what is the probability that the student would have 3 or fewer letters in his or her name?

e. If you were to choose a student from this class at random, what is the probability that the student would be male and have more than 6 letters in his name?

f. If the person randomly chosen from this class is a male, what is the probability that this male will have more than 6 letters in his name?

g. If a female is chosen from this class, what is the probability that she has 3 or fewer letters in her name?

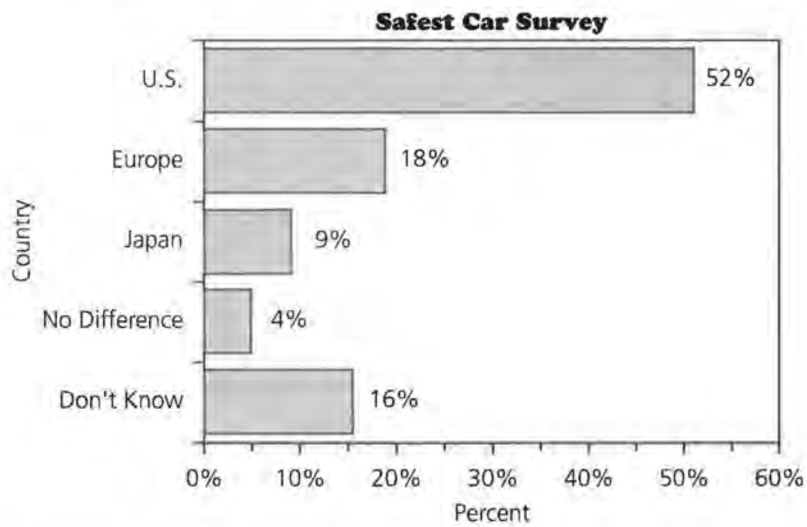
h. If the name that is chosen has 6 or more letters, what is the probability that the person is female?

Assessment for Unit III

OBJECTIVE

Apply the concepts of compound events, complementary events, and conditional probability.

1. In a nationwide survey, 1250 adults were asked this question: "Where in the world are the safest cars produced?" The following graph summaries their responses.



What is an estimate for the probability that a randomly chosen adult will answer that the safest auto is produced

a. in the U.S.?

b. not in the U.S.?

c. in Europe or in Japan?

d. not in Europe or not in Japan?

e. in the U.S. or in Europe or in Japan?

2. The following table shows the results of a Channel One News survey. The table shows the number of middle- and high-school students who say guilty teens deserve corporal punishment for certain crimes.

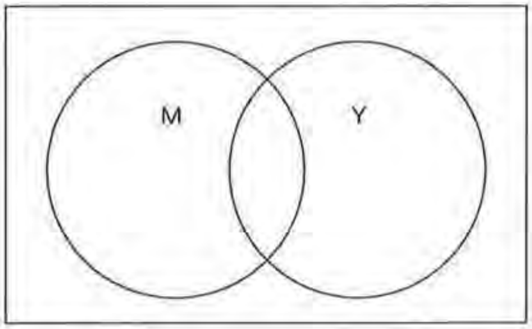
Deserve Corporal Punishment

	Yes	No	Totals
Middle/Junior High School	401	381	
High School	464	304	
Totals			

a. How many high-school students took part in the survey?

b. How many students answered "yes" to the question?

c. In the Venn diagram below, circle M represents students in middle school and circle Y represents students who answered "yes" to the question on corporal punishment.



i. Copy the Venn diagram and shade the part of the diagram that represents the students who are in middle school and answered "yes" to the question.

ii. Make another copy of the Venn diagram and shade the part of the diagram that represents the students who are in middle school or said "yes" to the question.

iii. Make a third copy of the Venn diagram and shade the part of the diagram that represents the students who are in high school.

3. Use the data from Problem 2 for the following.

If a middle/junior high-school or high-school student is randomly selected, what is the probability that

a. the student is in high school?

b. the student answered "yes" to the survey question?

c. the student is in high school and answered "no" to the question?

d. the student is in middle/junior high or answered "yes" to the question?

e. the student is not in high school?

4. Use the data from Problem 2 to answer the following questions.

a. Consider only the high-school students. What is the probability that a randomly selected high-school student answered "yes" to using corporal punishment?

b. What is the probability that a randomly selected student is in high school if you know that the person selected the answer "yes" to using corporal punishment?

5. The National Computing Survey of 2800 adults and children reported that 63% of the children aged 11–17 would rather use a computer than read a book and 59% would rather use a computer than watch TV.

a. If a person aged 11–17 is randomly selected, what is the probability that the person would say that he or she would rather read a book than use a computer?

b. Can you determine the probability of a person aged 11–17 who would rather use a computer than read a book and rather use a computer than watch TV? Explain your answer.

6. The table below contains information on the educational background for adults aged 25 to 34 in 1993. This data is part of the Census information gathered by the U.S. government.

Education	Percent of Population
Did Not Complete High School	13.1%
Completed High School	35.9%
Some College, No Degree	19.2%
Associate or Vocational Degree	7.9%
Bachelor's Degree	18.6%
Advanced Degree	5.2%

Source: *Statistical Abstract of the United States,* 1994

a. Find the probability that a randomly selected adult aged 25 to 34

i. did not complete high school by 1993.

ii. received an associate, vocational, or bachelor's degree by 1993.

b. Find the sum of the relative frequencies for each of the five categories. What is the significance of this value?

c. Suppose the Gallup Organization polled 1000 randomly selected adults aged 25 to 34. How many would you expect to say that they completed high school?

7. Prior to the 1996 Summer Olympic Games in Atlanta, the Atlanta Committee for the Olympic Games surveyed 606 Americans and asked how much interest they had in going to Atlanta for the Olympic games. The results are shown below.

	Amount of Interest		
Age Group	**Very Interested**	**Somewhat Interested**	**Not Interested**
18–34	160	96	64
35–54	109	87	90

a. Find the marginal totals and the marginal relative frequencies.

b. What is the joint relative frequency that a person from this survey is 18–34 years old and says that he or she is very interested in going to Atlanta for the Olympics?

c. What is the probability that a person from this survey is 35–54 years old or says that he or she is somewhat interested in going to Atlanta for the Olympics?

d. If it is known that the person surveyed is aged 18–34, what is the probability that the person has no interest in going to Atlanta for the Olympics?

e. If it is known that the person surveyed is very interested in going to Atlanta for the Olympics, what is the probability that the person is aged 35–54?

f. If it known that the person surveyed is aged 18–34, what is the probability that he or she is aged 35–54?

8. A random survey of 1000 adults found that 48% said that they do have enough leisure time. Only 14% of the 1000 adults said they would work fewer hours for less pay. The survey also found that only 5% said that they have enough leisure time and would be willing to work fewer hours for less pay.

   a. Copy and complete the two-way table with the joint frequencies.

Would You Work Fewer Hours for Less Pay?	Do You Have Enough Leisure Time?		
	Yes	No	Totals
Yes			
No			
Totals			1000

   b. Write four probability questions and their answers involving the information in your two-way table. At least one question should involve the joint frequencies and at least one question should involve conditional relative frequency.

Unit IV

Understanding Association

LESSON 9

Association

If you are right-handed, are you also right-eyed?

Do people who never smoked cigarettes think that smoking reduces stress?

In this lesson you will investigate how to answer questions that involve relationships, or ***association,*** between two variables.

OBJECTIVE

Understand how conditional probability can be used to measure association between two variables.

INVESTIGATE

Violence on TV and in Movies

Have you ever walked out of a movie or turned off a TV show because of the violence in the show? Is there a relationship between how men and women answer this question?

Discussion and Practice

In September, 1996, the American Medical Association released the results of a survey concerning violence in TV shows, movies, music lyrics, and computer games. The survey included a randomly selected nationwide sample of 800 registered voters. The results of one question on the survey are shown below.

Question: Have you ever walked out of a movie or turned off a TV show because of the violence in the show?

	Yes	No
Men	176	224
Women	320	80

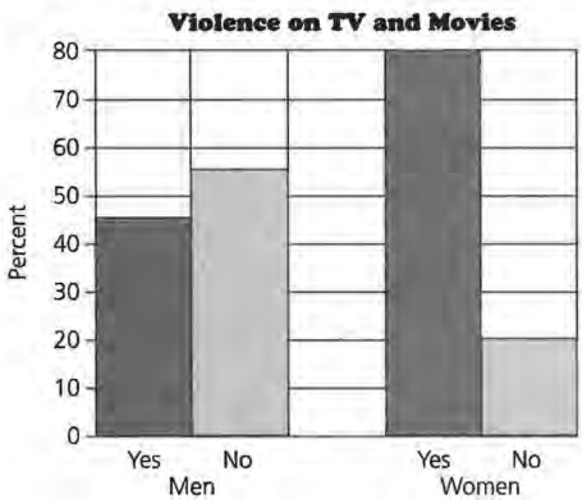
1. Make a copy of the table above and include the marginal totals.
2. What is the relative frequency of people in the survey that answered "yes" to the question?

3. Consider only the men in the survey. What is the conditional relative frequency of men who answered "yes"? What is the conditional relative frequency of men who answered "no"?

4. Consider only the women in the survey. What is the conditional relative frequency of women who answered "yes"? What is the conditional relative frequency of women who answered "no"?

A bar graph of the conditional relative frequencies can be used to decide if there is a relationship between the sex of the person and how the person answered the question about leaving a show because of violent content.

The bar graph below shows the distribution of the conditional relative frequencies.



5. Use the conditional relative frequencies and the bar graph to comment on whether there seems to be a relationship between the sex of a person and whether or not he or she would leave a show because of violent content. In other words, do the conditional relative frequencies seem to differ between men and women?

Left-Handed, Left-Eyed?

Is there a relationship between hand dominance and eye dominance? Check your eye dominance by making a circle with your thumb and forefinger and focusing, with both eyes open, on an object on the wall of the classroom. Close one eye and see if the object appears to move. If it does not move, the open eye is dominant.

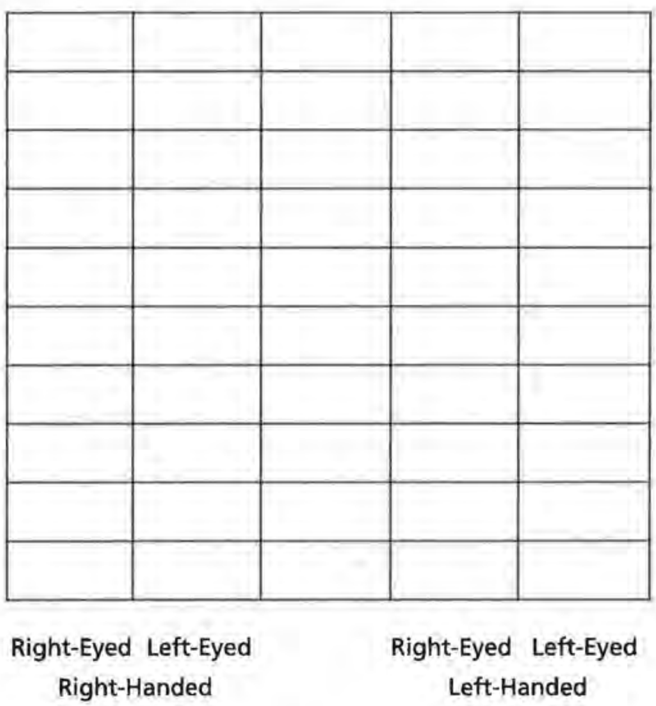
6. Collect the class data on eye dominance in a two-way table similar to the one below. In your table, fill in the values for cells a, b, c, and d.

		Eye Dominance	
		Left	**Right**
Hand Dominance	**Left**	a	b
	Right	c	d

a. Among the right-handed students, what is the conditional relative frequency that a person selected at random is right-eyed?

b. If a randomly selected student from your class is right-handed, what is the probability that this person has left-eye dominance?

c. Use a grid like the one below to make a bar graph showing the distribution of the conditional relative frequencies.



d. If a person is right-handed, do you think he or she is more likely to be left-eyed or right-eyed? Explain your answer.

7. The two-way table below shows a different distribution of hand and eye dominance.

		Eye Dominance		
		Left	Right	Totals
Hand Dominance	Left	7	1	8
	Right	2	18	20
	Totals	9	19	28

a. Among the right-handed students in this distribution, what is the conditional relative frequency that this person is right-eyed?

b. Among the right-handed students in this distribution, what is the conditional relative frequency that this person is left-eyed?

c. Use a grid like the one in Problem 6 to make a bar graph showing the distribution of the conditional relative frequencies.

d. If a person from this distribution is right-handed, do you think he or she is more likely to be left-eyed or right-eyed? Explain your answer.

8. The two-way table below shows another distribution of hand and eye dominance.

		Eye Dominance		
		Left	Right	Totals
Hand Dominance	Left	4	4	8
	Right	9	11	20
	Totals	13	15	28

a. Among the right-handed students in this distribution, what is the conditional relative frequency that this person is right-eyed?

b. Among the right-handed students in this distribution, what is the conditional relative frequency that this person is left-eyed?

c. Use a grid like the one in Problem 6 to make a bar graph showing the distribution of the conditional relative frequencies.

d. If a person from this distribution is left-handed, do you think he or she is more likely to be left-eyed or right-eyed? Explain your answer.

Problems 7 and 8 involve different distributions. The cells were set in Problem 7 to show that if a person is right-handed, then he or she is almost certain to be right-eyed. The cells in Problem 8 were set to show that if a person is right-handed, then there is no tendency toward being left-eyed or right-eyed.

9. Does it appear that your class distribution from Problem 6 is closer to the distribution shown in Problem 7, Problem 8, or neither problem? Explain your answer.

A relationship, or an association, exists if knowing the response to one of the variables helps to predict what the response might be to the other variable.

10. Study the tables in Problem 7 and Problem 8.

a. Which table shows strong association? Which table shows weak or no association?

b. Use your class distribution to determine whether there appears to be a strong relationship or association between hand dominance and eye dominance. Explain.

Summary

An association exists if knowing the response to one of the variables helps to predict what the response might be to the other variable. Knowing that person was right-handed helped predict his or her eye dominance. Knowing whether or not a person would walk out of a movie because of its violent content helped predict whether the person was a male or a female. If conditional relative frequencies for two groups are considerably different, then an association may exist.

Practice and Applications

11. On election day, exit polls are conducted by the media to determine who voted and why voters voted for a particular candidate. One such poll was conducted in 1996 by Voter News Service in Wisconsin. The results below are based on individual voter questionnaires after voters left polling places.

	Voted for		
	Clinton	**Dole**	**Totals**
Men	255	269	524
Women	382	238	620
Totals	637	507	1144

a. What percent of the voters voted for Clinton?

b. What percent of the voters were female?

c. Of the male voters, what percent voted for Clinton?

d. Of the female voters, what percent voted for Clinton?

e. Use a grid like that in Problem 6 to make a bar graph showing the distribution of conditional relative frequencies.

f. Use the conditional relative frequencies and the bar graph to comment on whether there seems to be an association between the sex of the Wisconsin voter and the candidate for whom the person voted.

12. The national Teenage Attitudes and Practices Survey obtained completed questionnaires from around the country by phone and mail from a randomly selected group of people 12–18 years old. The sample involved 9965 teenagers. Many of the questions asked the teenagers about their perception of the behavior of their peers. Why do you think the questions asked the students about their perceptions of peer behavior rather than ask questions about their own behavior?

Some of the questions on the Teenage Attitudes and Practices Survey asked the teenagers about their own behavior rather than their perceptions of their peers. The data shown are the results collected from the question "Do you believe cigarette smoking helps reduce stress?"

	Never Smoked	Experimented with Smoking	Former Smoker	Current Smoker
Yes	12.0%	18.7%	29.8%	46.5%
No	84.9%	78.5%	68.9%	51.7%
Don't Know	3.0%	2.5%	1.6%	1.6%

13. Use the data in the table on page 84 to answer the following questions.

a. How were these percents calculated?

b. Explain what the 12.0% represents.

c. Among students who have never smoked, what percent believe that cigarette smoking helps reduce stress?

d. Among students who are current smokers, what percent believe that cigarette smoking helps reduce stress?

e. If you knew that a teenager never smoked, what do you think his response to the question "Do you believe cigarette smoking helps reduce stress?" would be?

f. Are opinions on whether or not cigarette smoking helps reduce stress associated with the smoking status of the person responding? Write a short paragraph justifying your answer. Your paragraph should include appropriate data, calculations, and graphs.

LESSON 10

Constructing Tables from Conditional Probabilities

How many hours of TV do you watch each day?

Do your friends watch TV more or fewer hours each day than you do?

OBJECTIVES

Determine expected frequencies from conditional probabilities.

Interpret probability statements by constructing an appropriate table of expected frequencies.

Determine unconditional probabilities from a table of expected frequencies.

INVESTIGATE

Information often comes to you as conditional relative frequencies or conditional probabilities, even though it is not usually labeled as such. A recent poll in *Time* magazine states that 20% of U.S. students watch more than 5 hours of TV per day, but the figure is only 14% for France and 5% for Canada. Is this conditional information? Yes, it is conditional on the countries involved. There are about 61 million students in U.S. schools, and so this conditional percent can be translated in the fact that about $61(0.20) = 12.2$ million students watch more than 5 hours of TV per day in this country. What would you need to know in order to find out how many students in France watch more than 5 hours of TV per day? Do you think it could be greater than the number for the United States?

Discussion and Practice

Constructing a Table from Poll Results

1. A poll in *Life* magazine entitled "If Women Ran America" reports that two thirds of the women interviewed say that the problem of unequal pay for equal work is a serious one, while only half of the men interviewed have this opinion.

   a. Are the given poll results conditional information? Explain your answer.

b. About how many women would you expect to be in a group of 2000 Americans? Explain.

c. For a typical group of 2000 Americans, how many would you expect to be women who think that unequal pay for equal work is a serious problem?

d. For a typical group of 2000 Americans, how many would you expect to be men who think that unequal pay for equal work is a serious problem?

e. Copy the table below. Then use your results from parts b, c, and d to complete your table.

Is Unequal Pay for Equal Work a Serious Problem?

	Yes	No	Totals
Men			
Women			
Totals			2000

f. Find the marginal relative frequencies for the men and women rows and explain what each frequency represents.

2. In the same poll, one half of the women and one third of the men thought that discrimination in promotions was a very serious problem.

a. Copy and complete the following table to show how a typical group of 1000 Americans would be divided on this issue. Remember, the total numbers of men and women are approximately equal.

Is Discrimination in Promotions a Serious Problem?

	Yes	No	Totals
Men			
Women			
Totals			1000

b. Approximately what fraction of Americans think that discrimination in promotions is a very serious problem?

Blood Tests and Conditional Probability

ELISA is a popular test for screening blood samples for the presence of HIV. For blood samples known to contain HIV, ELISA shows a positive result 99% of the time. This means that 99% of the time the test will correctly identify blood that contains HIV. For blood samples known to be free of HIV, ELISA still reports a positive result about 2% of the time. This means that 2% of the time the test will incorrectly report that a blood sample contains HIV. This is called a *false positive.*

3. Is the 99% stated above conditional or unconditional information? Explain.

4. Assume that a medical lab tests 1000 blood samples using ELISA. Also assume that 50% of the samples contain HIV.

   a. How many blood samples known to contain HIV would you expect to test positive?

   b. How many false positives would you expect to see?

   c. Copy and complete the following table by filling in the remaining expected frequencies.

	Contain HIV	No HIV	Totals
Tested HIV Positive			
Tested HIV Negative			
Totals	500	500	1000

   d. Which cell contains the false positives?

   e. Which other cells show tests that are in error?

   f. One measure of the accuracy of the screening procedure is to examine the probability that an HIV sample will test positive. This is called the ***predictive value*** of the procedure. This means that the predictive value from the data in the table above would be found by dividing the number of samples that test positive and contain HIV by the total number of positive tests. Find this value. Would you say the test is doing a good job of screening? Why or why not?

Summary

Information is often presented in terms of conditional relative frequencies, which can be interpreted as conditional probabilities. To understand these conditional statements and to obtain unconditional information from them, it is helpful to think in terms of expected frequencies for a typical sample from the population under investigation. These expected frequencies are most easily understood when displayed in a two-way table.

Practice and Applications

5. A *Sports Illustrated for Kids* Omnibus Survey reported that about 44% of children aged 9–13 collect sports trading cards. Of this group who collected cards, about 1 in 4 of them are girls.

a. Are the given survey results conditional information? Explain.

b. About how many girls aged 9–13 would you expect to be in a random group of 800 children aged 9–13? Explain.

c. Of the 800 children surveyed, how many said that they collect sports trading cards?

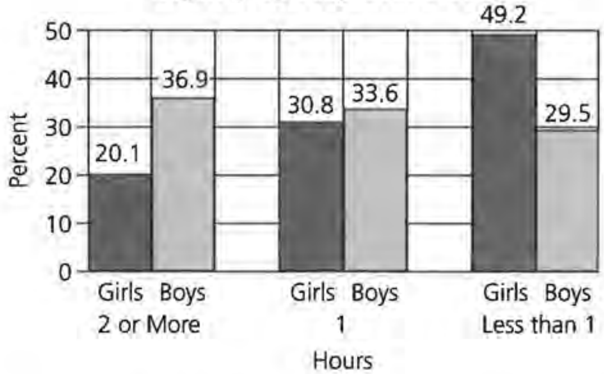
d. Of the 800 children surveyed, how many girls said that they collect trading cards?

e. Copy the table below. Then write the expected frequencies in the appropriate cells of your table.

	Do You Collect Sports Trading Cards?		
	Yes	No	Totals
Boys			
Girls			
Totals			800

6. The results of a *Sports Illustrated For Kids* poll, reported in *USA Today*, October 4, 1995, are shown in the graph below.

Number of Hours per Day Boys and Girls Aged 9–13 Play Video Games



a. Are the given poll results conditional information? Explain.

b. What does the 49.2% represent?

c. A student wanted to organize the results of the poll in a two-way table similar to the one that follows. Is it possible to complete the table from the data given? Can you fill in any of the marginal or joint relative frequencies? If so, which ones?

Time Playing Video Games

	2 or More Hours	1 Hour	Less than 1 Hour	Totals
Boys				
Girls				
Totals				

7. In all states, statewide exit polls are conducted on election day to help the news media predict which candidate will win the election. The results shown below are based on 1300 voters in Wisconsin on Election Day, 1996.

Percent of Total	Voted for Clinton	Dole	Perot	Other
47% Men	41%	44%	11%	4%
53% Women	55%	34%	8%	3%
36% Democrat	86%	6%	6%	2%
34% Republican	13%	80%	6%	1%
30% Independent/Other	42%	31%	20%	7%

a. Copy the following table. Then use the data on page 90 to complete your table.

	Voted for				
	Clinton	Dole	Perot	Other	Totals
Democrat					
Republican					
Independent/Other					
Totals					1300

b. If a Wisconsin voter is chosen at random, what is the probability that the person voted for Clinton?

c. If a Wisconsin voter is chosen at random from among those who said they are Democrats, what is the probability that the person voted for Clinton?

d. If a Wisconsin voter is chosen at random from among those who voted for Clinton, what is the probability that the person is a Democrat?

e. Use the data from part a to complete, as far as possible, the following table of frequencies of voters calculated from the exit-poll results.

	Voted for				
	Clinton	Dole	Perot	Other	Totals
Men					
Women					
Totals					1300

f. If a Wisconsin voter is chosen at random from among the women voters, what is the probability that the woman voted for Clinton?

g. If a Wisconsin voter is chosen at random from among those who voted for Clinton, what is the probability that the person is a woman?

LESSON 11

Comparing Observed and Expected Values

If you tossed a coin 100 times and you counted 35 heads, would you be surprised by the results?

If you took a sample of 500 people living in your community and 280 men and 220 women were chosen for your survey, would you think that your survey was not random?

OBJECTIVES

Simulate the variability that may be attached to a table of expected values.

Make decisions about the presence of association based on this variability.

In this lesson, you will compare data collected by a survey or an experiment to what you would have expected if the results happened purely by chance.

INVESTIGATE

Love Is Not Blind

"Love is not blind, and study finds it touching" is the headline of an article in the *Gainesville Sun*, June 22, 1992. Seventy-two blindfolded people each tried to distinguish his or her partner from a group of three people, one who was actually the partner, by touching the forehead. The blindfolded people making the selections were correct 58% of the time.

Discussion and Practice

1. Consider the data in the paragraph above.
   a. If a person picked his or her partner by chance, what percent of the time would you expect the person to be correct?
   b. Does the observed 58% success rate seem to be far from what you would expect due to chance?

Many statistical problems have the same structure as this experiment. Data are observed under certain experimental conditions. The question is "What is the probability that the observed data could have happened by chance?" To answer this question, you will design and run a simulation. The simulation model provides probabilities based on the occurrences that happen by chance. After the simulation is run, the observed data are compared to the results produced by the simulation. If they agree, then you could conclude that the observed data might have happened by chance. If the observed data and the simulated results do not agree, then you might conclude that the observed data represent a significant result.

Assume that the blindfolded participants in the study are merely guessing about their partners. With three possible choices for each blindfolded participant, the chance of choosing the correct partner by guessing is one third, or about 33%. How does the observed 58% compare with the 33% by chance? Is it a significant result?

It will be easier to work with frequencies rather than with percents. The number of correct decisions observed was 58% of 72, or 42. The number expected due to chance would be $\frac{1}{3}$ of 72, or 24. However, even if the participants were merely guessing, they might have had more or fewer than exactly 24 correct decisions.

2. If all 72 participants were guessing, what do you think is a reasonable interval of values that might contain the number of correct decisions?

One way to answer Problem 2 is to design and run a ***simulation.*** Simulation involves using a random number to model the behavior of the real event under investigation. The real event of choosing the correct partner purely by chance has a probability of $\frac{1}{3}$. To model this probability, random numbers 1, 2, and 3 can be used, designating 1 to represent "success" (choosing the right partner) and 2 and 3 to represent "failure" (*not* choosing the right partner).

3. Use a simulation to answer Problem 2.

a. Randomly select 1, 2, or 3 and record the outcome as "S" for success (1—picked the correct partner) or "F" for failure (2 or 3—did not pick the correct partner). Since the real experiment contacted 72 people picking their partners, you will need to choose a random number 72 times. Work with your group to randomly select a 1, 2, or 3 a total of 72 times. When finished, count the successes. Was your value close to 24, the expected number?

b. Repeat the 72 selections of random numbers 4 more times so you have 5 values for the number of success out of 72 decisions. Combine these 5 values with those from other groups in your class. Construct a graph of the class data. Use a number line like the one below.

15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35

c. Describe the shape of the graph you developed in part b. Mark the expected number of 24 on the number line. Do the points center around this value?

d. Compare the data from the experiment to the 42 successes reported in the research. Did anyone in class achieve 42 successes? What fraction of the data points on the graph are greater than or equal to 42?

e. Do you think people can choose their partners by touching them on the forehead; or do you think the results of the experiment happened by chance? Explain.

Independence Between Two Variables

4. The question of healthy diet versus eating breakfast was introduced in Lesson 5. The two-way table below shows the marginal totals.

Do You Eat Breakfast at Least 3 Times a Week?	Do You Think Your Diet Is Healthy?			
		Yes	No	Totals
	Yes			36
	No			14
	Totals	20	30	50

a. What fraction of the 50 polled students eat breakfast at least 3 times a week?

b. What fraction of the polled students think their diet is healthy?

c. If there are 1200 students in the school for which the poll was conducted, how many would you expect to say that they have a healthy diet?

5. Suppose that the two questions are not related. The statistical term we use to describe this is ***independence***. If the assumption that the questions are not related or independent for the students surveyed, then answering "yes" to the breakfast question will not affect the response to the diet question. In other words, the percent answering "yes" to the diet question should be about the same for each group (each row of the table) in the breakfast question. Since $\frac{20}{50}$, or 0.40, answered "yes" to the diet question, then among the students who answered "yes" to the breakfast question, you should expect 40% of 36, or 14.4, to have answered "yes" to the diet question.

   a. In a table like the one above, place 14.4 in the upper left-hand corner of the table (the "yes-yes" cell). Complete your table and describe your strategy.

   b. Describe what the value in the lower right-hand corner of your table (the "no-no" cell) represents.

6. The table below shows the original data table as presented in Lesson 5. Copy and complete the rest of the table using only the information presented.

		Do You Think Your Diet Is Healthy?		
		Yes	No	Totals
Do You Eat Breakfast at Least 3 Times a Week?	Yes	17		36
	No			
	Totals	20		50

The goal now is to see if the two questions are independent, or not related, by comparing the observed results (the table of original data) to the table of expected values. Since the joint information inside the table can be determined by knowing the marginal totals and one cell, the problem of determining whether or not the two variables are independent can focus on one cell. The problem can be restated as follows:

Is the 17 so far from the expected frequency of 14.4 for that cell that its occurrence probably did not happen by pure chance?

7. The chance that a frequency of 17 or more could occur in that cell under the assumption of independence can be approximated by a simulation. The following steps lead through a simulation to help find this probability.

Simulation Steps

i. Use 50 slips of paper of equal size. Mark 20 of the slips "Y" to represent those that answered "yes" to the diet question and mark 30 "N" for those that said "no" to the diet question.

ii. Mix the 50 slips of paper in a box and randomly select 36 of the slips. These represent the 36 students who said "yes" to the breakfast question. Under the assumption that the questions are independent, these 36 should behave like a random sample from the whole group of 50. That is, the proportion of the 36 who say "yes" should be about the same as the proportion of the 50 who said "yes."

iii. Count the slips among the sample of 36 that have a "Y." This represents the number of students who also said "yes" to the diet question from those who said "yes" to the breakfast question. In other words, this represents the number who said "yes" to both questions.

iv. Repeat this procedure 5 to 10 times to obtain a set of possible values for the "yes-yes" cell.

a. Combine your data with the data collected by other groups in your class. Use the number line like the one below to show the distribution of frequencies for the "yes-yes" cell.

7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22

b. Describe the shape of the graph produced. Make the expected value of 14.4 on your graph. Do the points center around this value?

c. Compare the observed value of 17 to the class data from the simulation. Does the value of 17 occur very often? What is your estimate of the probability of seeing 17 or more in the "yes-yes" cell?

d. Does it seem reasonable to say that the result of 17 or more is quite likely to occur by chance? Explain.

If you answered "yes" to part d, then you are suggesting that the two questions are independent; that is, the independent model agrees with the data, and there is probably no association. Generally, if the probability that an event will occur is less than 10%, the investigator may decide that this is a rare event and, therefore, there may be an association between the two variables in question.

e. What do you conclude about the association between questions dealing with diet and breakfast at Rufus King High School?

Summary

One of the goals of a statistical investigation is to compare data with theoretical models to see if the data support the model. For example, to check a coin for balance, one could assume that the coin is balanced (the model), toss it many times to observe the fraction of heads (the data), and compare the observed fraction of heads to $\frac{1}{2}$, the ratio assumed for this model. If the observed fraction of heads differs considerably from $\frac{1}{2}$, then the model of a balanced coin is questioned.

To decide if two questions or variables are associated, you should construct a two-way table of expected frequencies under the assumption that the two variables are independent. Design and run a simulation based on a model of independence. Compare the results of the simulation with the observed data and find an estimate for the probability that the observed data could have occurred by chance. If this probability is small, usually less than 10%, you may conclude that there is an association between the two variables.

Simulating the distribution of results that could be obtained from the model allows the investigator to see if the original outcome is likely or unlikely under the assumptions of the model. A simulated distribution of possible outcomes under the model is a powerful tool for decision making.

Practice and Applications

8. A sample of 40 students were asked two questions:

Do you have a job outside of school?

Do you have a car or truck of your own?

Half of the students had a car or truck of their own. Twenty-six of the 40 students had jobs. Twelve of the students did not have a job and did not own a vehicle.

a. Copy and complete the two-way table of observed frequencies.

Observed Frequency Table

		Job Outside of School		
		Yes	No	Totals
Do You Own a Car or Truck?	Yes			
	No			
	Totals			40

b. To decide if these two questions are associated, we need to construct a two-way table of expected frequencies. Assume that the two questions are independent of each other and copy and complete the following table. Remember that the marginal totals are the same as found in the observed frequency table. Also, since you are assuming independence, the ratio of the number in the "yes-yes" cell to the total of the number of students who own a car or truck should be the same as the ratio of the number of students who have a job to 40, the total number of students in the sample.

Expected Frequency Table

		Job Outside of School		
		Yes	No	Totals
Do You Own a Car or Truck?	Yes			
	No			
	Totals			40

c. Do the values in the observed-frequency table seem to be far from the corresponding values in the expected-frequency table?

d. Based on the tables constructed above, does it appear that the two questions are associated?

9. To help answer Problem 8d, you need to find the probability that the observed value in the "yes-yes" cell could happen by chance under the assumption that the two questions are independent. To do this, we will design and run a simulation.

Simulation Steps

i. Prepare 40 slips of paper of equal size. Mark the appropriate number "Y." These should represent the number of students who answered "yes" they have a job. The rest of the slips should be marked "N."

ii. Randomly select the number of slips equal to the number of students who said that they own a car or truck. Record the number of "Y"s.

iii. Repeat this procedure several times. Combine your data with the data from the rest of your class. Construct a graph of the frequencies of "Y"s.

a. Determine the probability that the observed value in the "yes-yes" cell could happen by chance under the assumption that the two questions are independent.

b. Do the observed results in the original survey appear to be unusual under the independence assumption? What does that tell you about the association between the two questions? Does information on job status provide any information on whether or not a student might own a car or truck?

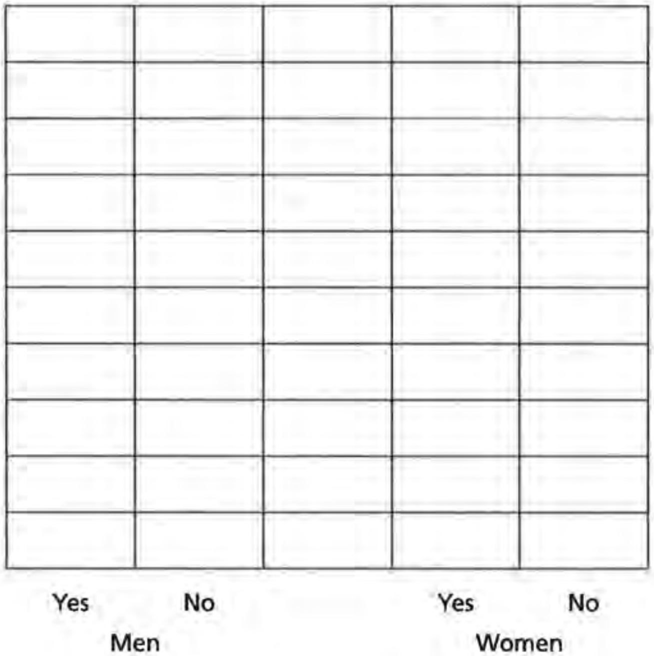
10. In March of 1995, a poll taken by the Gallup Organization for *U.S. News and World Report* reported that 43% of women and 34% percent of men said that they have been accused of being a "back-seat driver."

a. Assume that there were 50 men and 50 women in the survey. Copy and complete the following two-way table of observed frequencies.

Accused of Being a "Back-Seat Driver"

	Yes	No	Totals
Men			50
Women			50
Totals			100

b. Make a bar graph showing the distribution of the conditional relative frequencies. Uses a grid similar to the following.



c. Does it appear from the table of observed frequencies that there is an association between gender and whether or not the person was accused of being a back-seat driver? Explain.

d. What does it mean for the two variables to be independent?

e. Copy and complete the following two-way table of expected frequencies. What assumptions did you use to make your calculations?

	Accused of Being a "Back-Seat Driver"		
	Yes	**No**	**Totals**
Men			50
Women			50
Totals			100

f. After completing the table of expected frequencies, would you change your answer to part c?

g. Design and run a simulation to confirm your opinion of whether or not there is an association between the two variables. Write out the steps of your simulation and show the results of the simulation in graphical form. You should also state the probability of the observed frequency occurring by chance.

h. Write a paragraph explaining your conclusion on the association between gender and whether or not the person was accused of being a back-seat driver.

Assessment for Unit IV

OBJECTIVE

Apply the concepts of conditional probability and simulation to decide if there is an association between two variables.

1. A poll, taken by the Gallup Organization for *U.S. News and World Report* in March, 1995, reported that 43% of women surveyed and 34% of men surveyed said that they have been accused of being a "back-seat driver."

   a. Are the given poll results conditional information? Explain.

   b. The poll consisted of 806 randomly selected adults. How many women would you expect to be in this group of 806?

   c. How many women would you expect to have said that they have been accused of being a back-seat driver?

   d. Complete the two-way table below.

Accused of Being a "Back-Seat Driver"

	Yes	No	Totals
Men			
Women			
Totals			

2. A question on the Teenage Attitudes and Practices Survey was "Do you believe almost all doctors are strongly against smoking?" Responses to the question are recorded in percents.

Do You Believe Almost All Doctors Are Strongly Against Smoking?

	Never Smoked	Experimented with Smoking	Former Smoker	Current Smoker
Yes	80.1%	78.8%	80.1%	80.5%
No	17.3%	18.8%	17.3%	16.7%
Don't Know	2.5%	2.3%	2.6%	2.6%

   a. Are the percents joint, marginal, or conditional? Explain.

b. From the table on page 102, can you approximate the percent of teenagers who think that almost all doctors are strongly against smoking? Explain your reasoning.

c. Are the opinions that doctors are against smoking associated with the smoking status of the person responding? Justify your answer with appropriate references to conditional relative frequencies. Show your work.

3. The following summary of a poll reported in the January 30, 1995, issue of *Time* magazine is given in percent form.

POLL RESULTS

How Much Effort Are You Making to Eat a Healthy and Nutritionally Balanced Diet?

	Men	Women	Totals
Very Serious Effort	31%	43%	37%
Somewhat Serious Effort	47%	43%	45%
Not Very Serious Effort	12%	9%	10%
Don't Really Try	10%	5%	8%

a. Are these joint, marginal, or conditional percents? How can you tell?

b. Why are the percents for totals midway between the percents for men and for women? Explain your reasoning.

c. From a group of people who say they are very serious about eating a balanced diet, one person is randomly selected. What is the approximate probability that this person will be female?

d. Is there a strong association between gender and efforts at eating a balanced diet? Justify your answer with appropriate calculations of conditional relative frequencies.

4. The Koop Foundation conducted a survey of 1600 urban residents. The survey, reported in the *Milwaukee Journal Sentinel* on December 2, 1995, found that 42% of Americans watch 3 or more hours of television a day. Of those watching this amount of television, 62% were from families with an income of $25,000 or less, compared with 38% from families earning more than $25,000.

a. Are the results given from the Koop Foundation survey conditional information? Explain your answer.

b. Consider a sample of 1000 Americans. Is it possible to complete the following table from the data given? Can you fill in any of the marginal or joint totals? If so, which ones? If not, explain what other information you would need in order to complete the table.

		Income Level		
		$25,000 or Less	More than $25,000	Totals
Hours Watching TV	3 or More			
	Fewer than 3			
	Totals			1000

5. In September, 1996, researchers reported at a meeting of the American Society for Microbiology that a new oral vaccine that could protect babies against a sometimes fatal diarrheal disease might be available in 1998. The disease, *rotavirus,* is the most common cause of diarrhea in the world. The vaccine was given to 1190 babies, and 1208 babies received a placebo. Within two years, researchers counted 57 episodes of rotavirus in the vaccine group and 184 in the placebo group.

a. Copy and complete the following two-way table.

	Vaccine Group	Placebo Group	Totals
Rotavirus			
No Rotavirus			
Totals			

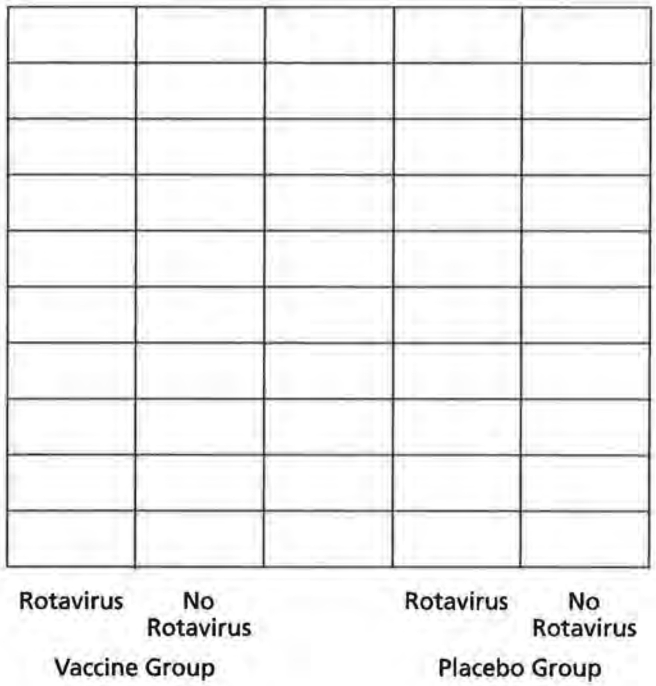
b. What proportion of the babies in the study got rotavirus within the two years?

c. What proportion of the vaccine group got rotavirus within the two years?

d. What proportion of the placebo group got rotavirus within the two years?

e. What does it mean to say that there is no association between what group the baby was in and whether or not the baby got rotavirus?

f. Make a bar graph showing the distribution of the conditional relative frequencies. Use a grid like the following.



g. Construct a two-way table of expected values, under the assumption that there is no association between the groups and whether or not the babies got rotavirus.

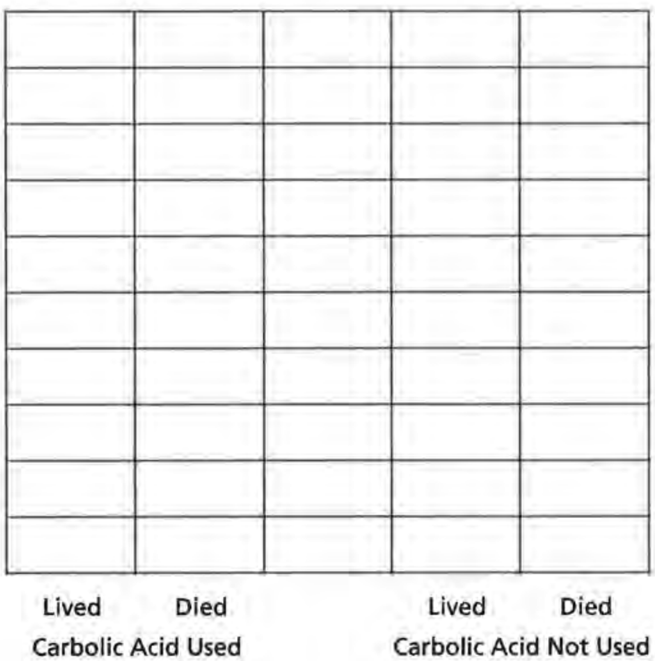
	Vaccine Group	**Placebo Group**	**Totals**
Rotavirus			
No Rotavirus			
Totals			

h. Describe how you would design and conduct a simulation to determine if there is an association between what group the baby was in and whether or not the baby got rotavirus.

6. Joseph Lister (1827–1912) was a surgeon at the Glasgow Royal Infirmary and one of the first to believe in the germ theory of infection. In one of his experiments, he used carbolic acid to disinfect operating rooms and compared the results from surgeries in these rooms to the results from rooms that had not been disinfected. Of the 40 patients in rooms where carbolic acid had been used, 34 lived. Of the 35 patients in rooms where carbolic acid had not been used, only 19 lived.

   a. Construct a two-way table of observed values.

   b. Make a bar graph using a grid like the following to show the distribution of the conditional relative frequencies.



   c. Construct a two-way table of expected values.

   d. Does it appear that the results of the surgeries are associated with the use of carbolic acid in the operating rooms? Explain.

   e. Design and run a simulation to confirm your opinion in part c. Record the steps that you followed and the simulation results.

   f. Write a paragraph explaining your conclusions on the association between the use of disinfectant and the success of surgeries.

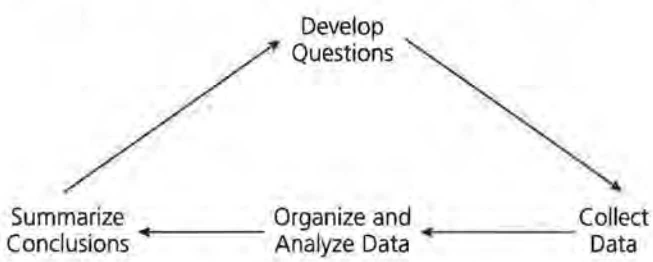
PROJECT

Analyzing Survey Results

OBJECTIVE

Analyze results of your own survey.

In lesson 5, the research project conducted by students at Rufus King High School was briefly discussed. In Lessons 9 through 11, the questions from the student survey concerning diet and breakfast were analyzed. The active research process conducted by the students consisted of asking questions, collecting data, analyzing the data, summarizing conclusions, and then refining and developing more questions. Frequently, successful research raises more questions than it answers. A diagram of the research process looks like this:



Project Analyzing Survey Results

In order to gain an insight into statistical methods, you are encouraged to carry out a random survey on a topic of interest to you.

Steps

i. Identify the population from which you will collect the information. This may be the entire student body or a particular class.

ii. Write a questionnaire that contains at least two questions. Carefully word your questions to avoid any misinterpretations or misunderstandings.

iii. Give your questionnaire to students in your group and your teacher, and have them give you feedback. Make any revisions necessary.

iv. Select a random sample of at least 30 people from the population you have identified. Recall from Lesson 3 the characteristics of random samples.

Issues that you need to address:

a. How will you choose your random sample?

b. How will you contact each person randomly selected?

c. How will the surveys be returned to you?

d. What will you do if someone does not respond to your questions?

v. Collect your results. Analyze and summarize the results both numerically by finding percents and graphically by constructing a bar graph or a similar summary.

vi. Decide on two questions from your survey that you think might be associated.

a. Place the results from these two questions in a two-way table of observed frequencies.

b. Complete a two-way table of expected frequencies based on the assumption that the two questions are independent.

c. Does there appear to be an association between the two questions?

d. Design and run a simulation to confirm your opinion about association.

e. What are your conclusions concerning whether or not there is an association between the two questions?

vii. Write an article for your school newspaper or Web page describing your results and the limitations of your survey.

Data-Driven Mathematics is a series of modules written by teachers and statisticians that focuses on the use of real data and statistics to motivate traditional mathematics topics. This chart suggests which modules could be used to supplement specific middle-school and high-school mathematics courses.

Dale Seymour Publications® is a leading publisher of K–12 educational materials in mathematics, thinking skills, science, language arts, and art education.

DALE SEYMOUR PUBLICATIONS®

90000

9 781572 322257

ISBN 1-57232-225-X

21170