ADVANCED MATHEMATICS

TEACHER'S EDITION

Advanced Modeling and Matrices

G. BURRILL, J. BURRILL, J. LANDWEHR, J. WITMER

DATA-DRIVEN MATHEMATICS

DALE SEYMOUR PUBLICATIONS®

Advanced Modeling and Matrices

TEACHER'S EDITION

DATA-DRIVEN MATHEMATICS

Gail F. Burrill, Jack Burrill, James M. Landwehr, and Jeffrey Witmer

Dale Seymour Publications®
Orangeburg, 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.

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

Dale Seymour Publications
125 Greenbush Road South
Orangeburg, NY 10962
Customer Service: 800–872–1100

Printed in the United States of America.

Order number DS21190

ISBN 1-57232-259-4

1 2 3 4 5 6 7 8 9 10-ML-01 00 99 98 97

This Book Is Printed On Recycled Paper

DALE SEYMOUR PUBLICATIONS®

Managing Editor: Cathy Anderson, Alan MacDonell

Senior Math Editor: Carol Zacny

Project Editor: John Sullivan

Production/Manufacturing Director: Janet Yearian

Senior Production Coordinator: Alan Noyes

Design Manager: Jeff Kelly

Text and Cover Design: Christy Butterfield

Composition: Joan Olson

Art: Carl Yoshihara

Cover Photo: David Madison/Bat Hitting Baseball

Authors

Gail F. Burrill
Whitnall High School
Greenfield, Wisconsin
University of Wisconsin-Madison
Madison, Wisconsin

Jack Burrill
Whitnall High School
Greenfield, Wisconsin
University of Wisconsin-Madison
Madison, Wisconsin

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

Jeffrey Witmer
Oberlin College
Oberlin, Ohio

Consultants

Emily Errthum
Homestead High School
Mequon, Wisconsin

Patrick Hopfensperger
Homestead High School
Mequon, Wisconsin

Henry Kranendonk
Rufus King High School
Milwaukee, Wisconsin

Vince O'Connor
Milwaukee Public Schools
Milwaukee, Wisconsin

Maria Mastromatteo
Brown Middle School
Ravenna, Ohio

Data-Driven Mathematics Leadership Team

Miriam Clifford
Nicolet High School
Glendale, Wisconsin

Kenneth Sherrick
Berlin High School
Berlin, Connecticut

Richard Scheaffer
University of Florida
Gainesville, Florida

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

Gail F. Burrill
Whitnall High School
Greenfield, Wisconsin
University of Wisconsin-Madison
Madison, Wisconsin

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 fields tests of the manuscripts.

The members of the *Data-Driven Mathematics* leadership team, the consultants and writers

Kathryn Rowe and Wayne Jones for their help in organizing the field-process and the Leadership Workshops

Barbara Shannon for many hours of word processing and secretarial services

Jean Moon for her advice on how to improve the field-test process

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 helps teachers motivate, develop, and reinforce concepts taught in current texts. The materials incorporate the 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 the statistical concepts most important for students to know and the relationship of these concepts to the secondary mathematics curriculum.

A diagram of the modules and possible relationships to the curriculum is on the back cover of each Teacher's Edition of the modules.

Using This Module

Why the Content Is Important

Many problems in the world require more than one variable to predict an outcome. Problems in medicine, science, engineering, and economics involve many variables, and a major task involves describing possible relationships among these variables. Matrices are an important mathematical tool in this process. They enable us to organize information and to efficiently apply different algorithms to data. As a set of mathematical elements, matrices also provide an example for investigating the field properties that are often taken for granted as we work with real numbers: the properties of closure, identity, inverse, associativity, commutativity, and distributivity. Finally, matrices have a very close link to vectors, which can be powerful tools in analysis. In each situation presented in the module, the geometry is brought out and its relation to an algebraic approach is explored.

The module is divided into two units. The first deals with a variety of variables and how these can be put together in a logical and mathematical way to provide a rating for a given subject. Based on the ratings, students investigate how to come up with a ranking order and a "best," such as the "best" company for working women or the "best" law school. Within this context, students use matrices to organize information, apply formulas by multiplying matrices, and investigate the effects of weighting variables. Students also explore the properties of multiplication with matrices, the transpose of a matrix, and how to solve systems of equations using matrices. This provides students with a different way to think about linear equations, concentrating on those of the form $ax + by = c$. As the number of variables increases, students move from exploring two-dimensional situations to three-dimensional and are introduced to an equation for a plane ($ax + by + cz = d$). Students make sketches in three dimensions and analyze the relationship between planes and lines using their knowledge of trigonometry. Finally, students generalize to an n-dimensional space and, in an extension, learn to describe the results using vectors.

The second unit involves using many variables to make a prediction. An outcome is often related to several different factors, and how to use these factors in a reasonable way is the focus of this unit. Using the fundamental notion of least squares linear regression and the sum of squared residuals, these same mathematical ideas can be used to deal with relationships among many variables. This replication of mathematical process, now in a broader context, provides students with a sense of the power of mathematics and introduces students to a valuable and widely used statistical method, multiple linear regression.

Students begin **this section** by stopping to reflect on the entire modeling process and the different components that must be considered to build a good mathematical model. They are introduced to problems with several possible explanatory variables through examples such as the relationship between standardized test scores and success in college. Matrices are used to express the problem in mathematical terms, and matrix operations are used as a tool in the analysis. From an understanding of the least squares algorithm for bivariate data, students generalize the procedure to situations where two variables are simultaneously used to predict a response. Through matrices, the analysis is extended to corresponding problems with multiple variables.

The emphasis is on students constructing their own mathematics in order to quantify the situation and to use information to solve a problem. Many of the problems are purposely open-ended; students have to make decisions about how to put data together as well as how to begin to analyze it. They need to recognize what assumptions they make when they arrive at formulas or rules or make extrapolations; and they must understand the consequences of changing these assumptions.

Content

Mathematics content: You will be able to:

- Use matrices to organize data.
- Multiply matrices and understand the limitations of matrix multiplication.
- Find the transpose of a matrix.
- Solve a system of equations using matrices.
- Graph equations of the form $ax + by = c$ given a, b.
- Plot points in three dimensions.
- Graph equations of the form $ax + by + cz = d$ given a, b, c.
- Identify parallel planes in space.
- Represent ordered n-tuples as vectors.

Statistics content: You will be able to

- Use least squares linear regression to find a model for paired data.
- Find and interpret the residuals for a given model.
- Calculate and interpret the mean and standard deviation for a set of data.
- Find and interpret the sum of squared errors for a model.
- Use multiple linear regression techniques to find a model for multivariate data.

Instructional Model

The instructional emphasis in *Advanced Modeling and Matrices,* as in all of the modules in *Data-Driven Mathematics,* is on discourse and student involvement. Each lesson is designed around a problem or mathematical situation and begins with a series of introductory questions or scenarios that can be used to prompt discussion and raise issues about that problem. These questions can provoke students' involvement in thinking about the problem and help them understand why such a problem might be of interest to someone in the world outside of the classroom. The questions can be used as part of whole-class discussions or by students working in groups. In some cases, the questions are appropriate to assign as homework to be done with input from families or from others not part of the school environment.

These opening questions are followed by the presentation of some discussion issues that clarify the initial questions and begin to shape the direction of the lesson that follows. Once the stage has been set for the problem, students begin to investigate the situation mathematically. As students work their way through the investigations, it is important that they have the opportunity to share their thinking with others and to discuss their solutions in small groups and with the whole class. Many of the exercises are designed for groups, where each member of the group does one part of the problem and the results are compiled for a final analysis and solution. Multiple solutions and solution strategies are also possible; it is important for students to recognize these situations and to discuss the reasoning that leads to different approaches. This will provide each student with a wide variety of approaches from which to build their own understanding of the mathematics.

In many cases, students are expected to construct their own understanding after being asked to think about the problem from several perspectives. They do need, however, validation of their thinking and confirmation that they are on the right track, which is why discourse among students and between students and teacher is critical. In addition, an important part of the teacher's role is to help students tie the ideas within an investigation together and to provide an overview of the "big picture" of the mathematics within the investigation. To facilitate this, a review and formalization of the mathematics is presented in a summary box following each investigation.

Each investigation is followed by a Practice and Applications section where students can revisit ideas presented within the lesson. These exercises may be assigned as homework, given as group work during class, or omitted altogether if students seem to be ready to move on.

At the end of each unit, assessment lessons are included within the student text. These lessons can be assigned as long-range take-home tasks, as group assessment activities, or as regular classwork. The ideas within the assessment provide a summary of the unit activities and can serve as a valuable way to enable students to demonstrate what they know and can do with the mathematics. It is helpful to pay attention to the strategies students use to solve a problem. This knowledge can be used as a way to help students grow in their ability to apply different strategies and learn to recognize those strategies that will enable them to find solutions efficiently.

Where to Use the Module in the Curriculum

This module is essentially on modeling and matrices. It should be used in a precalculus course, at the end of a second-year algebra course, or as a supplement to a calculus course.

- The first section of the module can be used in place of a standard chapter on matrices.
- Unit I can be used for students in an advanced geometry class to study applications in three dimensions.
- Unit I can be used in a study of analytic geometry.
- Unit II can be used in place of, or as an extension to, the section on fitting a curve in second-year algebra or precalculus.
- The entire module can be used in a calculus course after the Advanced Placement examinations have been given in the spring.
- The entire module can be used as a chapter in any upper-division mathematics course.
- The entire module can be used as a focal point in a mathematics topics course or a mathematics studies program.
- The module serves nicely as a chapter in a course that is designed around integrated mathematics.

Prerequisites

Students should have experience using variables to describe relationships, be able to graph linear equations, be able to simplify algebraic expressions, and understand something about the process of fitting a line to a set of data. The module follows three earlier modules from *Data-Driven Mathematics: Exploring Linear Relations, Exploring Regression,* and *Modeling with Logarithms.*

Pacing/Planning Guide

The table below provides a possible sequence and pacing for the lessons.

LESSON	OBJECTIVES	PACING
Unit I: Modeling, Matrices, and Ranks		
Introductory Activity: Where Do You Want to Live?	Practice using ratings to create ranks.	homework or 1/2 class period
Lesson 1: Representing Ratings with Matrices	Represent information in a matrix; multiply matrices and understand the constraints in the process; use matrices to apply a formula; solve systems of equations using a matrix.	1 to 3 class periods
Lesson 2: Ranking and Scatter Plots	Write a rating equation for two weighted variables; understand the relationships between the relative positions of points on a scatter plot; determine ranking using a scatter plot.	1 class period
Lesson 3: More on Ranking	Rank variables measured in different units; relate ranking with unequal scales to an equation and to a scatter plot.	1 to 2 class periods
Lesson 4: Ratings with Three or More Variables	Generalize the approach to the ratings problems using matrices and algebra; investigate algebraic representation of a plane.	3 to 4 class periods
Assessment: From Best Companies for Women to Cars	Use weights, a sweeping line, and matrices to create ratings and rankings.	1 class period
Unit II: Modeling, Matrices, and Multiple Regression		
Introductory Activity: What Affects Your Walking Speed?	Compare the strengths of different relationships.	homework or 1/2 class period
Lesson 5: Matrices and Linear Regression	Review the process of modeling a relationship between variables by fitting a line to the data; use matrices to express the least squares regression equation; review correlation, root mean squared error, and residuals.	1 to 3 class periods
Lesson 6: Multiple Variables and Modeling	Examine how two different factors can be used to predict an outcome; use recursion to generate a procedure.	1 to 2 class periods

LESSON	OBJECTIVES	PACING
Lesson 7: Comparing Order in Regression	Investigate the impact of order in the regression process; recognize what will occur as the process continues.	1 class period
Lesson 8: Matrices and Multiple Regression	Use matrices to represent the process of regression; justify the process mathematically; generalize the formula to many dimensions.	1 class period
Assessment: Rating Universities	Create ratings and ranks when high scores are good for one variable but bad for another variable.	1 class period
		3 weeks total time

Use of Data Sets and Teacher Resources

The data sets are on disks; student activity sheets, quizzes, and a test are available in the back of the teacher's guide.

LESSON	DATA SETS	RESOURCE MATERIALS
Unit I: Modeling, Matrices, and Ranks		
Introductory Activity: Where Do You Want to Live?		
Lesson 1: Representing Ratings with Matrices	Car Ratings Airline Complaints Ranking Metropolitan Areas	*Lesson 1 Quiz*
Lesson 2: Ranking and Scatter Plots	Dot-Matrix Printers Environmental Study for Selected Cities	*Activity Sheets 1, 2*
Lesson 3: More on Ranking	Hall of Fame Best Hitters Car Horsepower/Miles per Gallon Average Rainfall/Humidity	*Activity Sheets 3, 4* *Lesson 3 Quiz*
Lesson 4: Ratings with Three or More Variables	Baseball Hall of Fame Carl Yastrzemski Stats Rating Cities: Art, Education, Recreation Rating Cities: Cost of Living, Jobs, and Housing Best College Buys	*Test*
Assessment: From Best Companies for Women to Cars	Best Companies for Working Mothers Car Problems	
Unit II: Modeling, Matrices, and Multiple Regression		
Introductory Activity: What Affects Your Walking Speed?		
Lesson 5: Matrices and Linear Regression	Airplanes: Seats, Cost More Airplane Data Fast Food Calories/Fat	*Lesson 5 Quiz*
Lesson 6: Multiple Variables and Modeling	College Grade Point Average SAT Verbal, GPA Residuals SAT Verbal, GPA Residuals SAT Verbal, Math, GPA Residuals SAT Math, GPA Residuals SAT Math, Verbal GPA	

LESSON	DATA SETS	RESOURCE MATERIALS
Lesson 7: Comparing Order in Regression	GPA Verbal, GPA Math Residual Student 1 Comparing Sum of Squared Residuals Sum of Math, Verbal Best College Buys	
Lesson 8: Matrices and Multiple Regression	Cherry Wood Trees Cereal Data Airplane Data Cost of Stamps NFL Quarterback Rating	*Lesson 8 Quiz*
Assessment: Rating Universities	Data on Top Universities Rating Top Universities	

Technology

A graphing calculator is necessary for this module, preferably one that has the capability to interchange lists and matrices, although that is not absolutely necessary. (A graphing calculator resource section, entitled *Procedures for Using the TI-83 Graphing Calculator,* is included at the end of this module.) These manipulations might also be done with a computer and with spreadsheet and matrix manipulation software. Without technology, the procedures will become tedious and unmanageable, and the ideas will get lost in the manipulations.

Grade Level/Course

The module is appropriate for students in the latter part of second-year algebra, in precalculus, or in calculus.

UNIT I

Modeling, Matrices, and Ranks

INTRODUCTORY ACTIVITY

Where Do You Want to Live?

Materials: none
Technology: none
Pacing: 1/2 class period or as homework

Overview

"Where Do You Want to Live?" is designed to set the stage for this unit of the module. It can be used as an exercise for the entire class, as a homework assignment, or as an assignment students can use to share with their families.

Teaching Notes

Encourage students to think about things such as what kinds of jobs might be available, the climate, and whether they might want to live where there would be professional sports teams, theaters and cultural events, or lots of outdoor activities. They might also think about educational opportunities, health care, and many other variables. Students may have trouble rating the cities without any evidence, but at this point allow them to be purely subjective. If students are interested, you might return to the topic later in the unit, then have them collect some data about the cities they chose and apply the techniques they learned to that data. It is important to help them understand the difference between rating something (deciding whether something is good, bad, or indifferent; assigning a numerical value to something based on its perceived worth) and ranking (ordering a set according to some criteria from best to worst or highest to lowest).

Your students might not feel that choosing cities in which to live is an interesting topic. They will do some of this kind of rating in the lessons. You might instead allow them to choose another topic or suggest a different one, such as best quarterback in their school's conference, best basketball player, best rock group, best teacher, or best pizza. The focus should be on having students find ways to quantify their ratings and produce some rank order for their category. You could have different groups select different categories if that seems to be more appropriate.

STUDENT PAGE 3

INTRODUCTORY ACTIVITY

Where Do You Want to Live?

Which city in the United States has the best weather? the best schools? the best jobs?

Which factors are most important when choosing a place to live?

At some point in your life you may have the opportunity to move to a new location. It may be a different place within the state in which you already live, or it may be to a neighboring state, or to one across the continent. Your family may move, you may move to get a better job, or you may move for many other reasons.

OBJECTIVE

Practice using ratings to create ranks.

EXPLORE

Rate and Rank Cities

Assume you can move to any city in the United States. Consider how you might choose the city.

Data Collection and Analysis

1. List five factors about the city you would consider important in helping you make your choice.
2. Are all of these factors equally important?
3. Compare the factors you listed with those listed by others in your class. As a group decide on five factors you all agree are important.
4. Select five cities. Rate them on each of the factors and use your ratings to rank the cities. What was your top ranked city? How did your ranks compare to others in class?

LESSON 1

Representing Ratings with Matrices

Materials: *Lesson 1 Quiz*
Technology: graphing calculator with data sets stored
Pacing: 1 to 3 class periods

Overview

Students are introduced to matrices as a way to organize the data used by *Car and Driver* magazine to rate cars. They express a weighted rating system as a formula to find a rank for each car and use matrix multiplication as a way to apply the formula to the entire set of data. Some of the properties of matrix multiplication are explored, and the transpose is used in order to facilitate multiplication. Students learn to solve a system of equations using a matrix representation and the inverse of a matrix. The loss of some characteristics of the data when it is ranked is a consideration as students work with different data sets.

Teaching Notes

This lesson may be treated as an introductory lesson to matrices, or it may be covered quickly if your students already know how to multiply matrices and understand the constraints on the process. Be sure to have them work through at least one problem to see why multiplication makes sense. Students should understand the dimensions of a matrix, how the dimensions relate to the multiplication process, how to find and use the transpose of a matrix, and how to solve a system of equations using a matrix system.

The focus of the first section is to introduce matrices as an organizing tool and to use matrix multiplication as a way to facilitate using a formula. Students weight the variables involved in rating a car's quality and use the weights to produce a total rating for each car. By using matrices, they can experiment with several formulas at once and observe the impact on the results of changing the weights. Some students may have problems thinking about weighting and the final value if low numbers are desirable. They don't recognize that if you wish to double the weight of a variable you multiply the variable by 2; this seems to some as if you are giving that variable a lesser value because you are making it twice as much. Students do not see that you are giving every number in the category twice as much weight so the total number of possible points is doubled for that variable, making a larger overall total.

The set of airline data is the first students have worked with that involves rating with variables in different units. A common technique is to apply a ranking scheme to each category, but one disadvantage of this is that the ranks cover up the variability. A 0.3 difference separates first and second place in on-time arrival while 0.06 separates the two places for overbooking. Southwest is almost an outlier in overbooking, yet would receive the same tenth place rank as TWA for complaints. Students should recognize that you lose this information about the spread when you convert to ranks; but unless you do something, you will find it difficult to combine the variables in any sensible way. This idea of working with unlike variables in a methodical way will be revisited in Lesson 4.

The transpose is introduced in order to carry out the multiplication appropriate for a given context and is then used to produce the sum of squared residuals or differences from the mean. This notion of sum of squared residuals as a way to measure variability is an important idea and one that will recur in the second part of the module. You might point out to students that squaring a set of values will increase the impact of an outlier. (Two differences of 10 will have a sum of squares of 200, while one difference of 20 will contribute 400 to a sum of squares.) The sum of squares is useful because it lends itself to analysis by producing quadratics, curves that are easily understood with certain constant characteristics, one of which is a minimum value. (See Exploring Least Squares Regression from the *Data-Driven Mathematics* series if this is unfamiliar.)

The final part of the lesson deals with using a matrix representation for a system of equations. This again lays groundwork for ideas that will be developed in the second part of the module. Students should be cautioned that not every matrix has an inverse; only square matrices can even be considered. They might need to review the general fact that the product of an element and its inverse produces an identity. Students can experiment with several matrices to make sure they understand what an identity matrix is and how an inverse behaves. If a system consists of an infinite number of solutions or does not have a solution, the matrix system will fail because the matrix will not have an inverse. Students can inspect the equations to see whether one is a multiple of another or if there is some other relationship between the equations. They can also solve the system using row reduction with matrices or make use of other algebraic techniques, although using real data generates some very unwieldy numbers.

Technology

This lesson should be done with a graphing calculator or software to manipulate matrices. Initially, matrix multiplication should be done by hand so students can understand how the process works and why it makes sense. As this becomes routine, however, they should begin to use technology and to become comfortable with the matrix menu and how it operates. The use of technology allows them to consider "what if" questions and to explore the impact of different formulas on the data.

Follow-Up

Students might follow up this section by carrying out a rating scheme on a topic of their own or by finding instances of ratings and rankings in the newspapers, magazines, or other media. They should investigate the data used to make such claims as "best bad guy" in the movies and learn what questions to ask and what issues to consider as they think about rating something as "best."

STUDENT PAGE 4

LESSON 1

Representing Ratings with Matrices

If you were to buy a new car, which one would you buy?

What information would you want to know before you decided?

OBJECTIVES

Represent information in a matrix.

Multiply matrices and understand the constraints in the process.

Use matrices to apply a formula.

Solve systems of equations using a matrix.

"The best midsize sedan for 1994 was the Honda Accord," according to *Consumer Reports*. "Number one rating for the Intrepid." "Mustang leads the Pack." Minnesota is the healthiest state in Northwestern National Life Insurance Company's state health 1993, ranking. Rochester, Minnesota is one of the most "livable" cities according to *Money* magazine.

Cities, states, cars, and many other things are continually being rated to determine the "best." How is the "best" determined? In particular, what qualities do you think would give a car a top rating?

INVESTIGATE

Car Ratings

The staff of *Car and Driver* magazine selected six station wagons priced around $23,000, loaded them for a trip with fishing, boating and hiking gear and drove them through southern Ohio. The cars and the ratings given by the staff after the trips, as reported in the July 1994 issue of the magazine, are in the matrix below. A *matrix* is an ordered array of information. The dimensions of a matrix give the number of rows and columns in the matrix. The car rating matrix is a 6×10 matrix (read 6 by 10) because it has six rows and ten columns. The vehicles were rated from 1 (low) to 10 (high) in each category; the

Solution Key

Discussion and Practice

1. There are many different ways students may attempt to solve this problem. One possible way is to sum the ratings. For this method, the Camry will be the highest rated with a total of 87. Listed in order the cars would be

Toyota Camry	87
Volkswagen	85
Honda Accord	82
Mercury Sable	80
Subaru Legacy	79
Mitsubishi	75

2. a.

Honda Accord	54
Mercury Sable	46
Mitsubishi	38
Subaru Legacy	49
Toyota Camry	47
Volkswagen	51

The Honda is first according to this weighting scheme, followed by the Volkswagen. If students understand how to use matrices, they may already have solved the problem using matrices. If not, they would evaluate the formula using the data for each car.

$$\begin{bmatrix} 9 & 9 & 9 \\ 7 & 8 & 8 \\ 6 & 8 & 6 \\ 9 & 7 & 8 \\ 8 & 7 & 8 \\ 9 & 9 & 8 \end{bmatrix} \times \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 54 \\ 46 \\ 38 \\ 49 \\ 47 \\ 51 \end{bmatrix}$$

b.

Toyota Camry	54
Volkswagen	47
Honda Accord	40
Mercury Sable	47
Subaru Legacy	46
Mitsubishi	51

STUDENT PAGE 5

scores for each category were collected from each staff member and averaged; for example, the average score for the Honda Accord for handling was a 9.

Car Ratings

	Drive Train	Handling	Ride	Comfort (Drive)	Back Seat	Cargo Space	Cargo Utility	Style	Value	Fun to Drive
Honda Accord	8	9	8	9	8	7	7	9	9	8
Mercury Sable	8	7	8	7	7	10	10	8	8	7
Mitsubishi	7	6	8	8	9	10	7	8	6	6
Subaru Legacy	8	9	8	8	7	9	6	7	8	9
Toyota Camry	9	8	9	9	10	9	10	7	8	8
Volkswagen	9	9	8	9	9	8	7	9	8	9

Discussion and Practice

1. Use the ratings in the matrix to determine which wagon is the highest rated. Be ready to justify your method.

2. Chi decided that handling, styling, and value were the most important characteristics and reduced the table to a smaller matrix of just those ratings:

	Handling	Style	Value
Honda Accord	9	9	9
Mercury Sable	7	8	8
Mitsubishi	6	8	6
Subaru Legacy	9	7	8
Toyota Camry	8	7	8
Volkswagen	9	9	8

a. He felt that value (v) was the most important, handling (h) the next most important, and style (s) the least, so he weighted the categories to find the rating (r) as $r = 2h + 1s + 3v$. Find the weighted rating for each car using Chi's formula.

b. Tara felt that styling was more important than handling, so her formula was $r = h + 2s + 3v$. Compare Tara's weighted ratings to Chi's ratings.

A vector can be used to help organize the information. Chi's weighted ratings vector would be [2 1 3]. Written as a column vector, this would be: $\begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$

To find the weighted ratings for each car, you can use matrices to help organize your work.

The ranking would not be significantly affected by the different formula.

$$\begin{bmatrix} 9 & 9 & 9 \\ 7 & 8 & 8 \\ 6 & 8 & 6 \\ 9 & 7 & 8 \\ 8 & 7 & 8 \\ 9 & 9 & 8 \end{bmatrix} \times \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 54 \\ 47 \\ 40 \\ 47 \\ 46 \\ 51 \end{bmatrix}$$

STUDENT PAGE 6

Ratings Matrix · Weight Vector = Weighted Ratings

For the Honda Accord, the row vector for the Honda ratings can be multiplied by the column weight vector.

$$\begin{array}{c} \\ \text{Honda Accord} \end{array} \begin{array}{c} \text{Handling} \quad \text{Style} \quad \text{Value} \\ \begin{bmatrix} 9 & 9 & 9 \\ \underline{\quad} & \underline{\quad} & \underline{\quad} \\ \underline{\quad} & \underline{\quad} & \underline{\quad} \\ \underline{\quad} & \underline{\quad} & \underline{\quad} \\ \underline{\quad} & \underline{\quad} & \underline{\quad} \\ \underline{\quad} & \underline{\quad} & \underline{\quad} \end{bmatrix} \end{array} \cdot \begin{array}{c} \\ \text{Handling} \\ \text{Style} \\ \text{Value} \end{array} \begin{array}{c} \text{Chi} \\ \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} \end{array} = 9 \times 2 + 9 \times 1 + 9 \times 3 = \begin{bmatrix} 54 \\ \underline{\quad} \\ \underline{\quad} \\ \underline{\quad} \\ \underline{\quad} \\ \underline{\quad} \end{bmatrix}$$

For the Mercury Sable:

$$\begin{array}{c} \\ \text{Mercury Sable} \end{array} \begin{array}{c} \text{Handling} \quad \text{Style} \quad \text{Value} \\ \begin{bmatrix} \underline{\quad} & \underline{\quad} & \underline{\quad} \\ 7 & 8 & 8 \\ \underline{\quad} & \underline{\quad} & \underline{\quad} \\ \underline{\quad} & \underline{\quad} & \underline{\quad} \\ \underline{\quad} & \underline{\quad} & \underline{\quad} \\ \underline{\quad} & \underline{\quad} & \underline{\quad} \end{bmatrix} \end{array} \cdot \begin{array}{c} \\ \text{Handling} \\ \text{Style} \\ \text{Value} \end{array} \begin{array}{c} \text{Chi} \\ \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} \end{array} = 7 \times 2 + 8 \times 1 + 8 \times 3 = \begin{bmatrix} \underline{\quad} \\ 46 \\ \underline{\quad} \\ \underline{\quad} \\ \underline{\quad} \\ \underline{\quad} \end{bmatrix}$$

For the Mitsubishi:

$$\begin{array}{c} \\ \text{Mitsubishi} \end{array} \begin{array}{c} \text{Handling} \quad \text{Style} \quad \text{Value} \\ \begin{bmatrix} \underline{\quad} & \underline{\quad} & \underline{\quad} \\ \underline{\quad} & \underline{\quad} & \underline{\quad} \\ 6 & 8 & 6 \\ \underline{\quad} & \underline{\quad} & \underline{\quad} \\ \underline{\quad} & \underline{\quad} & \underline{\quad} \\ \underline{\quad} & \underline{\quad} & \underline{\quad} \end{bmatrix} \end{array} \cdot \begin{array}{c} \\ \text{Handling} \\ \text{Style} \\ \text{Value} \end{array} \begin{array}{c} \text{Chi} \\ \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} \end{array} = 6 \times 2 + 8 \times 1 + 6 \times 3 = \begin{bmatrix} \underline{\quad} \\ \underline{\quad} \\ 38 \\ \underline{\quad} \\ \underline{\quad} \\ \underline{\quad} \end{bmatrix}$$

Or, all at once:

Ratings Matrix · **Weight Vector** = **Weighted Ratings**

$$\begin{array}{c} \\ \text{Honda Accord} \\ \text{Mercury Sable} \\ \text{Mitsubishi} \\ \text{Subaru Legacy} \\ \text{Toyota Camry} \\ \text{Volkswagen} \end{array} \begin{array}{c} \text{Handling} \quad \text{Style} \quad \text{Value} \\ \begin{bmatrix} 9 & 9 & 9 \\ 7 & 8 & 8 \\ 6 & 8 & 6 \\ 9 & 7 & 8 \\ 8 & 7 & 8 \\ 9 & 9 & 8 \end{bmatrix} \end{array} \cdot \begin{array}{c} \\ \text{Handling} \\ \text{Style} \\ \text{Value} \end{array} \begin{array}{c} \text{Chi Weights} \\ \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} \end{array} = \begin{array}{c} \\ \text{Honda Accord} \\ \text{Mercury Sable} \\ \text{Mitsubishi} \\ \text{Subaru Legacy} \\ \text{Toyota Camry} \\ \text{Volkswagen} \end{array} \begin{array}{c} \text{Overall Rating} \\ \begin{bmatrix} 54 \\ 46 \\ 38 \\ 49 \\ 47 \\ 51 \end{bmatrix} \end{array}$$

3. **a.** The 47 represents the sum of $8 \times 2 + 7 \times 1 + 8 \times 3$ which is the product of the fifth row of the ratings matrix and the weight vector.

b.

$$\begin{bmatrix} 9 & 9 & 9 \\ 7 & 8 & 8 \\ 6 & 8 & 6 \\ 9 & 7 & 8 \\ 8 & 7 & 8 \\ 9 & 9 & 8 \end{bmatrix} \times \begin{bmatrix} 20 \\ 10 \\ 30 \end{bmatrix} = \begin{bmatrix} 540 \\ 460 \\ 380 \\ 490 \\ 470 \\ 510 \end{bmatrix}$$

c. The new results are shown in the weighted ratings matrix multiplied by 10. The ranks would not change.

4. **a.**

$$\begin{bmatrix} 9 & 9 & 9 \\ 7 & 8 & 8 \\ 6 & 8 & 6 \\ 9 & 7 & 8 \\ 8 & 7 & 8 \\ 9 & 9 & 8 \end{bmatrix} \times \begin{bmatrix} 5 & 3 \\ 2 & 3 \\ 8 & 5 \end{bmatrix} = \begin{bmatrix} 135 & 99 \\ 115 & 85 \\ 94 & 72 \\ 123 & 88 \\ 118 & 85 \\ 127 & 94 \end{bmatrix}$$

b. These two different weightings result in different ratings but almost the same ranking.

5. **a.** In the first matrix the columns represent the rating for Style, Value, and Fun to Drive for each car, and the rows indicate to which car those ratings correspond. In the second matrix the columns represent the weightings Pedro and Tina assigned to each of those features, and the rows represent the feature to which the weight is assigned.

b. $S + V + 5F = R$ for Pedro's ratings.

$3S + V + 3F = R$ for Tina's ratings.

STUDENT PAGE 7

3. Chi got 47 for the Toyota Camry.

a. Explain how he got that number.

b. Suppose Chi's weight vector was $\begin{bmatrix} 20 \\ 10 \\ 30 \end{bmatrix}$. Find the weighted ratings.

c. Compare the results Chi got using the two different weight vectors.

Paul and Kuo chose different weights for the three categories. The matrix representation below shows the ratings matrix and a matrix of the two weight vectors.

$$\begin{array}{l} \\ \text{Honda Accord} \\ \text{Mercury Sable} \\ \text{Mitsubishi} \\ \text{Subaru Legacy} \\ \text{Toyota Camry} \\ \text{Volkswagen} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Handling} & \text{Style} & \text{Value} \end{array} \\ \begin{bmatrix} 9 & 9 & 9 \\ 7 & 8 & 8 \\ 6 & 8 & 6 \\ 9 & 7 & 8 \\ 8 & 7 & 8 \\ 9 & 9 & 8 \end{bmatrix} \end{array} \qquad \begin{array}{l} \\ \text{Handling} \\ \text{Style} \\ \text{Value} \end{array} \begin{array}{c} \begin{array}{cc} \text{Paul} & \text{Kuo} \end{array} \\ \begin{bmatrix} 5 & 3 \\ 2 & 3 \\ 8 & 5 \end{bmatrix} \end{array}$$

4. To find Paul's weighted ratings for the Honda, multiply the first row, the ratings for each category, by the column with Paul's weight vector. To find Kuo's weighted ratings for the Honda, multiply the first row by Kuo's weight vector in the weighted ratings matrix.

a. Find the weighted ratings for Paul and Kuo.

b. How do the two compare?

5. Refer back to the original matrix on page 5.

a. Explain what each matrix below would represent.

Matrix R $\qquad\qquad$ Matrix W

$$\begin{array}{c} \\ H \\ MS \\ MI \\ S \\ T \\ V \end{array} \begin{array}{c} \begin{array}{ccc} S & V & F \end{array} \\ \begin{bmatrix} 9 & 9 & 8 \\ 8 & 8 & 7 \\ 8 & 6 & 6 \\ 7 & 8 & 9 \\ 7 & 8 & 8 \\ 9 & 8 & 9 \end{bmatrix} \end{array} \qquad \begin{array}{c} \\ S \\ V \\ F \end{array} \begin{array}{c} \begin{array}{cc} \text{Pedro} & \text{Tina} \end{array} \\ \begin{bmatrix} 1 & 3 \\ 1 & 1 \\ 5 & 3 \end{bmatrix} \end{array}$$

b. What formula would describe Pedro's weighted ratings? Tina's weighted ratings?

(5) c. $R_{42} = 8$. It is in the fourth row and the second column, and represents the value rating *Car and Driver* gave to the Subaru Legacy.

d.

$$\begin{bmatrix} 9 & 9 & 8 \\ 8 & 8 & 7 \\ 8 & 6 & 6 \\ 7 & 8 & 9 \\ 7 & 8 & 8 \\ 9 & 8 & 9 \end{bmatrix} \times \begin{bmatrix} 1 & 3 \\ 1 & 1 \\ 5 & 3 \end{bmatrix} = \begin{bmatrix} 58 & 60 \\ 51 & 53 \\ 44 & 48 \\ 60 & 56 \\ 55 & 53 \\ 62 & 62 \end{bmatrix}$$

The first column represents the weighted ratings of the cars according to Pedro's weighting scheme. The second column represents the weighted ratings for Tina.

6. a. If any one of the entries of the second matrix were zero, it would eliminate that category from consideration in the ranking scheme.

b. If Tina's weights were exactly double those of Pedro's, the ranking of the cars would be the same for both of students. However, Tina's ratings would be double those of Pedro.

7. a.

$$\begin{bmatrix} 9 & 9 & 7 & 9 \\ 7 & 7 & 10 & 8 \\ 8 & 6 & 10 & 6 \\ 8 & 9 & 9 & 8 \\ 9 & 8 & 9 & 8 \\ 9 & 9 & 8 & 8 \end{bmatrix} \times \begin{bmatrix} 3 \\ 5 \\ 2 \\ 10 \end{bmatrix} = \begin{bmatrix} 176 \\ 156 \\ 134 \\ 167 \\ 165 \\ 168 \end{bmatrix}$$

b. There is a change in ranking for each of the weighted values because the individuals did not weight the items the same and, in fact, were nearly of opposite opinions about the importance of a given item, with the exception of value.

STUDENT PAGE 8

c. Each entry in a matrix is sometimes called a *cell*. The 7 in the second row and third column (*MS, F*) of matrix *R* is in cell R_{23}. What entry is in cell R_{42}? What does the entry represent?

d. Find the product of the two matrices. Explain what the entries in the result represent.

6. Changing the entries in matrix *W* can affect the results.

a. Describe the effect of an entry in the second matrix being 0.

b. What would happen to the ratings and the ranking if Tina's weights for each category were exactly double Pedro's weights for those categories?

7. Use the ratings for driver comfort, handling, cargo space, and value with the following weights:

driver comfort 3, handling 5, cargo space 2, value 10.

a. Create the matrices and find the weighted ratings.

b. Explain why the following sets of weighted values don't give the same rankings:

Method 1: driver comfort 5, handling 2, cargo space 1, value 3
Method 2: driver comfort 2, handling 3, cargo space 5, value 3

8. The U.S. Department of Transportation periodically issues information about airlines, and on the basis of this information, airlines make claims about being the "Number 1" airline. The following table contains the information released for January to June 1995 for overbooking and complaints and for the first quarter of 1995 for mishandled baggage and on-time arrival.

Airline Complaints

Airline	Mishandled Baggage per 1,000 Passengers	Complaints per 100,000 Passengers	Percent On-Time Arrival	Overbooked per 10,000 Passengers
Southwest	4.14	0.25	80.5	3.04
US Air	4.81	0.48	80.2	1.67
American	4.99	0.59	75.2	0.44
America West	5.05	0.59	76.7	2.56
Continental	5.19	0.71	76.6	0.89
Delta	5.20	0.77	78.5	0.81
Alaska	5.25	0.85	77.7	1.65
United	5.28	1.09	79.0	0.32
Northwest	6.05	1.38	80.0	0.26
Trans World	6.33	1.51	72.0	0.68

Source: *Air Travel Consumer Report*, July, August, 1995

STUDENT PAGE 9

8. **a.** Student answers will vary. Some possible suggestions might be age of aircraft, safety record, age of their pilots, percent of on-time departures, and food service.

b. Student answers will vary. Look for student recognition of the fact that in some categories a high number is best and in others, low is best. Some values are percentages. Students should be aware of the fact that these values are not all equivalent.

c. Basaj recognized that for some of the items a high number is best and for others a low number is best. He set up an independent ranking from 1 to 10. The advantage to this process is that it eliminates the confusion caused by the fact that some of the categories are positive qualities and others are negative. Then the magnitude of one set of values does not have an overwhelming influence on his average. The disadvantage is that by simply averaging the ranks, Basaj gives equal value to each item.

9. In cell PI_{11}, this multiplication would combine the price of the three grades of oak board with the amounts of the three types of Grade I lumber and would therefore be nonsense.

a. What other factors might you consider important when rating an airline?

b. Based on the data given, how would you rank the airlines? Be ready to justify your method.

c. Basaj ranked the airlines from 1 to 10 in each category, then rated an airline by finding the average of the ranks. What are the advantages and disadvantages of his method?

Transpose of a Matrix

A lumber yard carried wood in different prices and grades. An inventory of the number of board feet of each kind they had is reported in matrix I. Matrix P is the price matrix.

Matrix I

$$\begin{array}{l} \\ \text{Oak bd. ft.} \\ \text{Cherry bd. ft} \\ \text{Pine bd. ft.} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Grade I} & \text{Grade II} & \text{Grade III} \end{array} \\ \begin{bmatrix} 750 & 526 & 300 \\ 200 & 127 & 300 \\ 250 & 750 & 750 \end{bmatrix} \end{array}$$

Matrix P

$$\begin{array}{l} \\ \text{Oak} \\ \text{Cherry} \\ \text{Pine} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Grade I} & \text{Grade II} & \text{Grade III} \end{array} \\ \begin{bmatrix} \$1.80 & \$1.95 & \$3.00 \\ \$1.42 & \$1.50 & \$2.50 \\ \$0.95 & \$1.10 & \$2.00 \end{bmatrix} \end{array}$$

9. Would it be useful to find the product PI? What would the entry in cell PI_{11} represent?

You can exchange the rows and columns of P and obtain a matrix called P^T, the *transpose* of P, as illustrated below.

Matrix P^T

$$\begin{array}{l} \\ \text{Grade I} \\ \text{Grade II} \\ \text{Grade III} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Oak} & \text{Cherry} & \text{Pine} \end{array} \\ \begin{bmatrix} \$1.80 & \$1.42 & \$0.95 \\ \$1.95 & \$1.50 & \$1.10 \\ \$3.00 & \$2.50 & \$2.00 \end{bmatrix} \end{array}$$

10. Find the product of IP^T and call it C.

a. What does the entry in C_{11} represent?

b. What does the entry in C_{12} represent?

c. How much money does the dealer have tied up in Cherry wood? Which cell in the matrix will answer the question?

10. **a.**

$$\overset{I}{\begin{bmatrix} 750 & 526 & 300 \\ 200 & 127 & 300 \\ 250 & 750 & 750 \end{bmatrix}} \times \overset{P^T}{\begin{bmatrix} \$1.80 & \$1.42 & \$0.95 \\ \$1.95 & \$1.50 & \$1.10 \\ \$3.00 & \$2.50 & \$2.00 \end{bmatrix}} = \overset{C}{\begin{bmatrix} \$3275.70 & \$2604.00 & \$1891.10 \\ \$1507.65 & \$1224.50 & \$929.70 \\ \$4162.50 & \$3355.00 & \$2562.50 \end{bmatrix}}$$

Entry C_{11} = \$3275.70. This represents the total value of all the oak the lumber yard has in stock.

b. Entry C_{12}= \$2604.00 This represents the sum of the products of the number of board feet of oak the lumber yard has and the price per board foot of the three grades of cherry wood and is, therefore, meaningless.

c. The lumber yard has \$1224.50 tied up in cherry wood, which is found in cell C_{22}.

d. The inventory of oak is worth \$3275.70, cherry is worth \$1224.50, and pine is worth \$2562.50. These answers are found on the main diagonal of the product matrix.

STUDENT PAGE 10

d. What is the value of the inventory in oak, cherry, and pine worth? How can you use a matrix to find the answer?

11. Use the same matrices I and P above.

a. Find I^TP.

b. What information can you obtain from the product?

12. Consider the following matrices:

$$A = \begin{bmatrix} 2 & 4 & 5 \\ 1 & 6 & 7 \end{bmatrix} \quad B = \begin{bmatrix} 3 \\ 2 \\ 8 \end{bmatrix} \quad C = \begin{bmatrix} 5 & 7 \\ 9 & 4 \end{bmatrix}$$

a. What are the dimensions of A^T? B^T? $(A^T)A$?

b. Find $(A^T)A$.

c. Find $(AB)^T$, $(A^T)(B^T)$, and $(B^T)(A^T)$. What conjecture can you make?

d. Check your conjecture using A and C.

13. A square matrix can have a multiplicative inverse. Remember that the product of an element and its inverse will yield an identity and that an identity will not change the value of the element when it is used as a factor. In the real numbers, 1 is the multiplicative identity, and $1/a$ is the multiplicative inverse of a, where a is not equal to 0.

a. Verify that $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ is the multiplicative identity matrix for all 2×2 matrices.

b. Why will only a square matrix have a multiplicative inverse?

c. Use the inverse key on your calculator to find the multiplicative inverse of the matrix $\begin{bmatrix} 2 & 4 \\ 5 & 1 \end{bmatrix}$ and verify that it is the inverse.

d. Find the inverse of $\begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix}$. What conclusion can you make?

Mean and Deviations

Another way to compare the car ratings is to consider the mean rating for each of the cars and how consistent those ratings are. You can use matrices to determine the deviation from the mean for the ratings in each of the categories and to find the sum of the squared deviations, an important statistic that is used in many different ways. Refer to the car rating matrix from the beginning of the lesson.

11. a.

$$\overset{I^T}{\begin{bmatrix} 750 & 200 & 250 \\ 526 & 127 & 750 \\ 300 & 300 & 750 \end{bmatrix}} \times \overset{P}{\begin{bmatrix} \$1.80 & \$1.95 & \$3.00 \\ \$1.42 & \$1.50 & \$2.50 \\ \$0.95 & \$1.10 & \$2.00 \end{bmatrix}} = \overset{V}{\begin{bmatrix} \$1871.50 & \$2037.50 & \$3250.00 \\ \$1839.64 & \$2041.20 & \$3395.50 \\ \$1678.50 & \$1860.00 & \$3150.00 \end{bmatrix}}$$

b. The values on the main diagonal represent the total value (money) tied up in the inventory of Grade I lumber, cell V_{11}, in Grade II lumber, cell V_{22}, and the value of Grade III lumber, cell V_{33}.

12. a. The dimension of A^T will be 3×2, the dimension of B^T will be 1×3, and the dimension of $(A^T)A$ will be 3×3.

b.

$$\begin{bmatrix} 2 & 1 \\ 4 & 6 \\ 5 & 7 \end{bmatrix} \times \begin{bmatrix} 2 & 4 & 5 \\ 1 & 6 & 7 \end{bmatrix} = \begin{bmatrix} 5 & 14 & 17 \\ 14 & 52 & 62 \\ 17 & 62 & 74 \end{bmatrix}$$

c. $(AB)^T = [\,54 \;\; 71\,]$; $(A^T)(B^T)$ is incompatible for multiplication because AT has dimensions 3×2 and B^T has dimensions 1×3; $(B^T)(A^T) = [\,54 \;\; 71\,]$.

A conjecture might be that $(AB)^T = (B^T)(A^T)$.

(12) d. $(CA)^T = \begin{bmatrix} 17 & 22 \\ 62 & 60 \\ 74 & 73 \end{bmatrix}$

$(A^T)(C^T) = \begin{bmatrix} 17 & 22 \\ 62 & 60 \\ 74 & 73 \end{bmatrix}$

13. a. The product

$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} =$

$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \times \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$

for any a, b, c, and d. This verifies that $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ is the multiplicative identity for a 2×2 matrix.

b. Suppose A is an $m \times n$ matrix and B, its inverse, has dimensions $p \times q$. So, $A \cdot B = B \cdot A = I$. But $A \cdot B$ cannot be performed unless $n = p$, and $B \cdot A$ cannot be performed unless $q = m$. So, B has dimensions $n \times m$. Now an $m \times n$ matrix times an $n \times m$ matrix produces an $m \times m$ matrix, and an $n \times m$ matrix times an $m \times n$ matrix produces an $n \times n$ matrix. If $A \cdot B = B \cdot A$, then n must be equal to m, and A and B are both square.

c. The inverse will be

$\begin{bmatrix} -1/18 & 2/9 \\ 5/18 & -1/9 \end{bmatrix}$ or $\begin{bmatrix} -0.055 & 0.222 \\ 0.277 & -0.111 \end{bmatrix}$.

$\begin{bmatrix} -1/18 & 2/9 \\ 5/18 & -1/9 \end{bmatrix} \times \begin{bmatrix} 2 & 4 \\ 5 & 1 \end{bmatrix} =$

$\begin{bmatrix} 2 & 4 \\ 5 & 1 \end{bmatrix} \times \begin{bmatrix} -1/18 & 2/9 \\ 5/18 & -1/9 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

d. The calculator has an error message: a singular matrix. This indicates there is no inverse for that matrix. An observant student might notice that the rows are multiples of each other. Not all matrices have inverses; if one row is a multiple of another or if one row is a linear combination of the others, the matrix will not have an inverse.

STUDENT PAGE 11

Car Ratings

	Drive Train	Handling	Ride	Comfort (Drive)	Back Seat	Cargo Space	Cargo Utility	Style	Value	Fun to Drive
Honda Accord	8	9	8	9	8	7	7	9	9	8
Mercury Sable	8	7	8	7	7	10	10	8	8	7
Mitsubishi	7	6	8	8	9	10	7	8	6	6
Subaru Legacy	8	9	8	8	7	9	6	7	8	9
Toyota Camry	9	8	9	9	10	9	10	7	8	8
Volkswagen	9	9	8	9	9	8	7	9	8	9

14. The mean rating over all the categories for the Mercury is 8. Make a 10×1 matrix R for the ratings for the Mercury and another 10×1 matrix M where each entry is the mean, 8.

a. What will $R - M$ give you?

b. Let $R - M = D$. What will you have if you find the sum of the entries in D?

15. Write the transpose of D.

a. What are the dimensions of D^T?

b. Find $(D^T)D$. What does this represent?

c. You can use $(D^T)D$ to find the standard deviation for the ratings. Divide $(D^T)D$ by the number of categories and take the square root. What is the standard deviation for the Mercury?

16. Divide the other cars among the members of your group.

a. For your car, find the mean, then use matrices to find the sum of the squared deviations from the mean.

b. Find the standard deviation of the ratings for your car.

c. Compare your results to those of others in your group. Which car has the smallest sum of squared deviations from the mean?

d. For which car are the ratings most consistent?

Solving Systems of Equations

Systems of equations are often used to solve problems involving data, and matrices can be very useful for solving systems of equations. For example, each year *Places Rated Almanac* rates the 343 metropolitan areas in the United States on living costs, job outlook, crime, health, transportation, education, the arts, recreation, and climate. The areas are rated on safety by using a linear combination of their violent crime rate (the sum of the

14. a. The matrix $R - M$ will represent a matrix for the differences between each of the rankings and the mean of those rankings.

$R - M = D = \begin{bmatrix} 0 \\ -1 \\ 0 \\ -1 \\ -1 \\ 2 \\ 2 \\ 0 \\ 0 \\ -1 \end{bmatrix}$

b. The sum of the entries in D will be zero.

15. $D^T = [0\ -1\ 0\ -1\ -1\ 2\ 2\ 0\ 0\ -1]$.

a. The dimensions of D^T will be 1×10.

b. $(D^T)D = [12]$. This is the matrix of the sum of the squares of the differences.

c. $\sqrt{12/10} = 1.095$. The standard deviation of the Mercury ratings is 1.095.

16. a.

Honda Accord	mean = 8.2 $(D^T)D = [5.6]$
Mitsubishi	mean = 7.5 $(D^T)D = [16.5]$
Subaru Legacy	mean = 7.9 $(D^T)D = [8.9]$
Toyota Camry	mean = 8.7 $(D^T)D = [8.1]$
Volkswagen	mean = 8.5 $(D^T)D = [4.5]$

b. Honda Accord standard deviation = $\sqrt{5.6/10} \approx 0.748$

Mitsubishi standard deviation = $\sqrt{16.5/10} \approx 1.285$

Subaru Legacy standard deviation = $\sqrt{8.9/10} \approx 0.943$

Toyota Camry standard deviation = $\sqrt{8.1/10} \approx 0.900$

Volkswagen standard deviation = $\sqrt{4.5/10} \approx 0.671$

Volkswagen has the smallest sum of the squared deviations from the mean.

c. The Volkswagen has the smallest standard deviation. Therefore, it has the most consistent ratings.

17. a. $1721a + 10934b = 2{,}281$
$549a + 4790b = 1{,}028$

STUDENT PAGE 12

murder, robbery, and aggravated assault rates) and their property crime rate (the sum of the burglary, larceny-theft, and motor vehicle theft rates). The rates are per 100,000 people. The formula for computing the crime rating, r, can be expressed as

$$av + bp = r$$

where v is the violent crime rate, p is the property crime rate, a and b are the coefficients for the formula.

17. The Miami rate for violent crime was 1,721 and for property 10,934. Miami's total crime rating was 2,821. Washington, DC had a 549 violent crime rate and a 4,790 property crime rate for a total rating of 1,028.

a. Write the system of equations you would need to solve to find the coefficients for the formula used to obtain the total crime rating.

The information can be written as a matrix system:

$$C \cdot A = R$$

where C is the matrix of the crime ratings, A is the coefficient matrix, and R is the rating matrix,

or $\begin{bmatrix}1721 & 10934\\549 & 4790\end{bmatrix} \times \begin{bmatrix}a\\b\end{bmatrix} = \begin{bmatrix}2821\\1028\end{bmatrix}$.

b. Verify that matrix multiplication will produce the system you wrote for part a.

c. To solve the system for the coefficients, use the inverse of C, C^{-1}. Write the formula to find the total crime rating. Do the coefficients make sense? Why or why not?

d. What is the total crime rating for Ottawa-Hull in Ontario, Canada, if the violent crime rate is 218 and the property crime rate is 5,732?

Summary

A matrix is an ordered array of elements. If the matrix has m rows and n columns, the dimensions of the matrix are $m \times n$. To multiply matrices, multiply each row of the first matrix by each column of the second. If the number of entries in each row of the first matrix does not match the number of entries in each column of the second, you cannot multiply the matrices. The entry in the i, j position of the product matrix is the result of multiplying the entries in the ith row by those in the jth column

b.

$$\begin{bmatrix}1721 & 10934\\549 & 4790\end{bmatrix} \times \begin{bmatrix}a\\b\end{bmatrix} = \begin{bmatrix}1721a & 10934b\\549a & 4790b\end{bmatrix}$$

c. $C^{-1}CA = C^{-1}R$

$$\begin{bmatrix}.0021376065 & -.0048794551\\-.000244999161 & .0007680210494\end{bmatrix} \times \begin{bmatrix}1721 & 10934\\549 & 4790\end{bmatrix} \times \begin{bmatrix}a\\b\end{bmatrix} =$$

$$\begin{bmatrix}.0021376065 & -.0048794551\\-.000244999161 & .0007680210494\end{bmatrix} \times \begin{bmatrix}2281\\1028\end{bmatrix}$$

Hence $\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} -.1401993195 \\ .2306825525 \end{bmatrix}$ or a is approximately $-.1402$, b is approximately .2307, and the formula is $-0.1402v + 0.2307p = r$.

Student answers will vary about the sensibility of the coefficients. They might question the negative sign for a.

(17) d.

$$[218\ 5732] \times \begin{bmatrix} -.1401993195 \\ .2306825525 \end{bmatrix} = [1291.71],$$

which is the rating for Ottawa-Hull, Canada, or $(218)(-.1401993195) + (5732)(.2306825525) = 1291.71$.

Practice and Applications

18. a. The best climate is in San Francisco, followed by San Diego.

b.

$$\begin{array}{cccc} \text{Cost Living} & \text{Crime} & \text{Health} & \text{Climate} \end{array}$$

$$\begin{bmatrix} 302 & 240 & 30 & 15 \\ 343 & 288 & 8 & 1 \\ 326 & 285 & 66 & 2 \\ 269 & 51 & 41 & 98 \\ 308 & 215 & 10 & 54 \end{bmatrix} \times \begin{bmatrix} 4 \\ 2 \\ 10 \\ 5 \end{bmatrix} = \begin{bmatrix} 2063 \\ 2033 \\ 2544 \\ 2078 \\ 2032 \end{bmatrix}$$

Washington, DC, has the best rating for Miguel's weights. (Recall that low scores are better.)

STUDENT PAGE 13

and summing the products. The transpose of a matrix is the matrix obtained by exchanging the rows and the columns of the original matrix such that the ith row becomes the ith column.

Practice and Applications

18. The top five metropolitan areas noted in *Places Rated Almanac* for 1993 are given in the matrix below with the ranks assigned by the Almanac to each of the categories. The rankings are based on scores that are a composite of data collected for that category. For example, recreation scores are based on assets such as good restaurants, public golf courses, bowling lanes, zoos, aquariums, NCAA Division I football and basketball teams, and miles of ocean or Great Lakes coastline. The metropolitan areas are then ranked according to those scores, where a 1 is the top ranked or best score for a given category.

Ranking Metropolitan Areas

	Cost of Living	Jobs	Housing	Transp.	Educ.	Health	Crime	Arts	Recreation	Climate
Seattle, WA	302	16	312	15	31	30	240	24	16	15
San Francisco, CA	343	71	543	21	23	8	288	6	5	1
San Diego, CA	326	7	328	27	12	66	285	33	3	2
Pittsburgh, PA	269	109	214	24	24	41	51	36	150	98
Washington, DC	308	5	314	10	6	10	215	3	105	54

Source: *Places Rated Almanac*, 1993

Note that in this problem, low numbers are better than high numbers.

Miguel Hernandez, who is planning to retire, weighted the categories as follows: cost of living 4, crime 2, health care 10, climate 5, and the rest 0.

a. Of the cities given, which have the best ratings for climate?

b. Find the weighted ratings for Miguel. Which town gets the best rating?

c. Chris Strom and Rex McNall rated the categories differently:

Chris: cost of living 8, crime 1, health care 5, climate 5

c.

$$\begin{array}{cccc} \text{Cost Living} & \text{Crime} & \text{Health} & \text{Climate} \end{array}$$

$$\begin{bmatrix} 302 & 240 & 30 & 15 \\ 343 & 288 & 8 & 1 \\ 326 & 285 & 66 & 2 \\ 269 & 51 & 41 & 98 \\ 308 & 215 & 10 & 54 \end{bmatrix} \times \begin{bmatrix} 8 & 5 \\ 1 & 0 \\ 5 & 8 \\ 5 & 10 \end{bmatrix} = \begin{bmatrix} 2881 & 1900 \\ 3077 & 1789 \\ 3233 & 2178 \\ 2898 & 2653 \\ 2999 & 2160 \end{bmatrix}$$

Choosing a different weighting scheme can cause a different ranking.

d. Student answers will vary. The purpose of this exercise is to give students practice in multiplication of matrices.

e.

$$\begin{bmatrix} \text{Arts} & \text{Health} & \text{Educ} \\ 24 & 30 & 31 \\ 6 & 8 & 23 \\ 33 & 66 & 12 \\ 36 & 41 & 24 \\ 3 & 10 & 6 \end{bmatrix} \times \begin{bmatrix} 1 & 10 & 3 & 5 \\ 2 & 20 & 2 & 25 \\ 3 & 30 & 1 & 45 \end{bmatrix}$$

$$= \begin{bmatrix} 177 & 1770 & 163 & 2265 \\ 91 & 910 & 57 & 1265 \\ 201 & 2010 & 243 & 2355 \\ 190 & 1900 & 214 & 2285 \\ 41 & 410 & 35 & 535 \end{bmatrix}$$

These weights produce different scores, but the same final rankings:

1. Washington, DC
2. San Francisco
3. Seattle
4. Pittsburgh
5. San Diego

19. a.

$$\begin{bmatrix} \text{Jobs} & \text{Trans} & \text{Educ} & \text{Rec} \\ 29 & 30 & 4 & 57 \\ 1 & 197 & 7 & 49 \\ 71 & 1 & 1 & 15 \end{bmatrix}$$

$$\times \begin{bmatrix} 10 & 10 \\ 5 & 1 \\ 10 & 5 \\ 2 & 8 \end{bmatrix} = \begin{bmatrix} 594 & 796 \\ 1163 & 634 \\ 755 & 836 \end{bmatrix}$$

The different schemes produced different rankings. Anaheim is last under the first weighting and first under the second weighting.

b. Cell P_{32} is the rating for New York according to the second weighting. The calculations were $71 \times 10 + 1 \times 1 + 1 \times 5 + 15 \times 8$.

STUDENT PAGE 14

Rex: cost of living 5, crime 0, health care 8, climate 10.

How do the three different weightings compare? Which weighting scheme do you think is the best? Why?

d. As a group, select three or four of the variables you all consider important. Each of you should independently assign weights to the variables you choose. Use matrix multiplication to compare the effects of the different weighting schemes.

e. Suppose the weights assigned to the categories are changed as given below. Create matrices and describe the effect on the rankings of the metropolitan areas.

Arts 1, health care 2, education 3

Arts 10, health care 20, education 30

Arts 3, health care 2, education 1

Arts 5, health care 25, education 45

19. The rankings for three other metropolitan areas are given in the following matrices, along with two different weighting schemes, I and II.

$$\begin{array}{l|cccc} & \text{Jobs} & \text{Transp} & \text{Educ} & \text{Rec} \\ \hline \text{Boston} & & & & \\ \text{Anaheim} & & & & \\ \text{New York} & & & & \end{array}$$

$$\begin{array}{lc} & \begin{array}{cccc} \text{Jobs} & \text{Transp} & \text{Educ} & \text{Rec} \end{array} \\ \begin{array}{l} \text{Boston} \\ \text{Anaheim} \\ \text{New York} \end{array} & \begin{bmatrix} 29 & 30 & 4 & 57 \\ 1 & 197 & 7 & 49 \\ 71 & 1 & 1 & 15 \end{bmatrix} \end{array} \qquad \begin{array}{lc} & \begin{array}{cc} \text{I} & \text{II} \end{array} \\ \begin{array}{l} \text{Jobs} \\ \text{Transp} \\ \text{Educ} \\ \text{Rec} \end{array} & \begin{bmatrix} 10 & 10 \\ 5 & 1 \\ 10 & 5 \\ 2 & 8 \end{bmatrix} \end{array}$$

a. Find the product P of the two matrices. How do the rating schemes compare?

b. What does the entry in cell P_{32} mean? What calculations created that entry?

20. Consider the matrices below:

$$\begin{bmatrix} 1 & 3 & 5 \\ 2 & 1 & 4 \\ 1 & 2 & 2 \end{bmatrix} = A \qquad \begin{bmatrix} 3 & 7 & 2 \\ 1 & 4 & -2 \\ 3 & 2 & -1 \end{bmatrix} = B \qquad \begin{bmatrix} 4 & -2 \\ 3 & 1 \\ 1 & 2 \end{bmatrix} = C$$

a. Find AB and BA and compare your results. What conclusion can you draw?

b. Find AC and CA. What has to be true about the dimensions of the matrices you want to multiply?

c. If you multiply a matrix that is 3×5 and one that is 5×2, what will the dimensions of the product be?

20. a.

$$\begin{bmatrix} 21 & 29 & -9 \\ 19 & 26 & -2 \\ 11 & 19 & -4 \end{bmatrix} = AB; \qquad \begin{bmatrix} 19 & 20 & 47 \\ 7 & 3 & 17 \\ 6 & 9 & 21 \end{bmatrix} = BA$$

$AB \neq BA$

Matrix multiplication is not commutative, and, therefore, order is very important.

b.

$$\begin{bmatrix} 18 & 11 \\ 15 & 5 \\ 12 & 4 \end{bmatrix} = AC$$

The number of columns in the first matrix must be equal to the number of rows in the second matrix to be compatible for multiplication. Since Matrix C is (3×2) and matrix A is (3×3), product CA is not possible.

c. When a (3×5) matrix is multiplied by a (5×2) matrix, the result is a (3×2) matrix.

STUDENT PAGE 15

21. More information about the five wagons reviewed by *Car and Driver* is given in the matrix below.

	0–30 mph	Braking Dist. 70 mph (ft)	City (mpg)	Highway (mpg)	700-Mile trip (mpg)	Price ($)
Honda	3.2	188	23	29	22	21,850
Mercury	3.0	197	19	28	20	21,645
Mitsubishi	3.8	193	18	24	18	26,790
Subaru	2.5	185	18	23	18	23,645
Toyota	2.9	174	18	24	21	23,303
Volkswagen	3.3	178	18	25	20	23,890

a. For which of these six categories are low numbers considered to be good?

b. Using the categories from part a, devise two different weight formulas to rank the cars. Use matrix multiplication to find the weighted ratings. Compare the results in ranking from these two weight formulas.

c. For which of these six categories are high numbers considered to be good?

d. Using the categories from part c, devise two different weight formulas to rank the cars. Use matrix multiplication to find the weighted ratings. Compare the results in ranking from these two weight formulas.

22. The National Association for Stock Car Auto Racing (NASCAR) sponsors the Winston Cup races, covering between 400 and 600 miles each on 17 designated speedways around the country. For each race, drivers are given a certain number of points for winning or for the place in which they finish, for the number of laps in which they lead, and five bonus points for the driver who leads the most laps. In the 1995 Martinsville, Virginia, Goody's 500, Dale Earnhardt won the race and earned 1440 points. He led in 252 of the 500 laps. In the Dover, Delaware, race held in late summer 1995, Jeff Gordon won with a total of 2,180 points and led in 400 of the 500 laps.

a. Set up a system of equations to determine how many points are given for first place and for the number of laps a driver leads.

b. Use matrices to solve the system.

c. How many points would a racer earn in a race if he led in 150 of the 500 laps and won the race?

---

21. a. 0–30 mph, Braking, and Price

b. Answers will vary; the sample here shows the ratings for the weights 1, 1, and 3 and the weights 3, 1, and 2.

$$\begin{bmatrix} 3.2 & 188 & 21850 \\ 3.0 & 197 & 21645 \\ 3.8 & 193 & 26790 \\ 2.5 & 185 & 23645 \\ 2.9 & 174 & 23303 \\ 3.3 & 178 & 23890 \end{bmatrix} \cdot \begin{bmatrix} 1 & 3 \\ 1 & 1 \\ 3 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 65741.2 & 43897.6 \\ 65135.0 & 43496.0 \\ 80566.8 & 53784.4 \\ 71122.5 & 47482.5 \\ 70085.9 & 46788.7 \\ 71851.3 & 47967.9 \end{bmatrix}$$

Although the scores are different, the rankings are virtually the same under both schemes, with Mercury ranked first. (Remember that low scores are better for these characteristics.)

c. City (mpg), Highway (mpg), and 700-mile trip (mpg)

d. Answers will vary; the sample here shows the ratings for the weights 3, 2, and 1 and the weights 1, 3, and 3.

$$\begin{bmatrix} 23 & 29 & 22 \\ 19 & 28 & 20 \\ 18 & 24 & 18 \\ 18 & 23 & 18 \\ 18 & 24 & 21 \\ 18 & 25 & 20 \end{bmatrix} \cdot \begin{bmatrix} 3 & 1 \\ 2 & 3 \\ 1 & 3 \end{bmatrix}$$

$$= \begin{bmatrix} 149 & 176 \\ 133 & 163 \\ 120 & 144 \\ 118 & 141 \\ 123 & 153 \\ 124 & 153 \end{bmatrix}$$

Although the scores are different, the rankings are virtually the same under both schemes, with Honda ranked first.

22. a. $F + 252L + 5 = 1440$

$F + 400L + 5 = 2180$

$$\begin{bmatrix} 1 & 252 \\ 1 & 400 \end{bmatrix} \cdot \begin{bmatrix} F \\ L \end{bmatrix} = \begin{bmatrix} 1435 \\ 2175 \end{bmatrix}$$

b.

$$\begin{bmatrix} F \\ L \end{bmatrix} = \begin{bmatrix} 1 & 252 \\ 1 & 400 \end{bmatrix}^{-1} \cdot \begin{bmatrix} 1435 \\ 2175 \end{bmatrix}$$

$$= \begin{bmatrix} 175 \\ 5 \end{bmatrix}$$

First place is worth 175 points, and each lap is worth 5 points.

c. 925 points

d. 170 points are awarded for second place. $(180 - 2 \cdot 5 = 170)$

23. a.

$$\begin{bmatrix} 120 & 120 & 200 & 250 \\ 100 & 90 & 150 & 150 \\ 150 & 175 & 175 & 200 \end{bmatrix} \cdot \begin{bmatrix} 6 & 12 \\ 7 & 12 \\ 7 & 15 \\ 8 & 18 \end{bmatrix} = \begin{bmatrix} 4960 & 10380 \\ 3480 & 7230 \\ 4950 & 10125 \end{bmatrix}$$

The dealer has \$4,960 invested in white T-shirts;

\$3,480 invested in red T-shirts;

\$4,950 invested in gray T-shirts.

b.

$$\begin{bmatrix} 25 & 50 & 75 & 70 \\ 50 & 78 & 123 & 100 \\ 15 & 82 & 105 & 84 \end{bmatrix} \times \begin{bmatrix} \$6 & \$12 \\ \$7 & \$12 \\ \$7 & \$15 \\ \$8 & \$18 \end{bmatrix} = \begin{bmatrix} \$1585 & \$3285 \\ \$2507 & \$5181 \\ \$2071 & \$4251 \end{bmatrix}$$

	Dealer's Cost	Sales Income	Profit
White	\$1585	\$3285	\$1700
Red	\$2507	\$5181	\$2674
Gray	\$2071	\$4251	\$2180

The dealer made the most money on red T-shirts.

c.

$$\begin{array}{c} \text{Stock} \\ \begin{array}{cccc} S & M & L & EL \end{array} \\ \begin{bmatrix} 120 & 120 & 200 & 250 \\ 100 & 90 & 150 & 150 \\ 150 & 175 & 175 & 200 \end{bmatrix} \end{array} - \begin{array}{c} \text{Sold} \\ \begin{array}{cccc} S & M & L & EL \end{array} \\ \begin{bmatrix} 25 & 50 & 75 & 70 \\ 50 & 78 & 123 & 100 \\ 15 & 82 & 105 & 84 \end{bmatrix} \end{array}$$

$$= \begin{array}{c} \text{Inventory} \\ \begin{array}{cccc} S & M & L & EL \end{array} \\ \begin{bmatrix} 95 & 70 & 125 & 180 \\ 50 & 12 & 27 & 50 \\ 135 & 93 & 70 & 116 \end{bmatrix} \end{array}$$

STUDENT PAGE 16

d. Terry Labonte finished second in the 1995 Richmond race and led in 2 laps. He earned a total of 180 points. How many points are awarded by NASCAR for a second-place finish?

23. After the Wisconsin Badgers won the Rose Bowl in 1994, Badger tee shirts were on sale everywhere in Wisconsin. At one store, they had three colors: red, gray, and white. They came in small, medium, large, and extra large. On the day of the Rose Bowl, the inventory was as given in I, the inventory matrix below.

Matrix I

$$\begin{array}{c|cccc} & \text{Small} & \text{Medium} & \text{Large} & \text{Extra Large} \\ \hline \text{White} & 120 & 120 & 200 & 250 \\ \text{Red} & 100 & 90 & 150 & 150 \\ \text{Gray} & 150 & 175 & 175 & 200 \end{array}$$

The dealer's cost and selling price for each size of shirt is given below.

$$\begin{array}{c|cccc} & \text{Small} & \text{Medium} & \text{Large} & \text{Extra Large} \\ \hline \text{Dealer cost} & \$6 & \$7 & \$7 & \$8 \\ \text{Selling Price} & \$12 & \$12 & \$15 & \$18 \end{array}$$

a. Create matrices and use matrix multiplication to determine how much money the dealer had tied up in each color of tee shirt.

b. On the day after the Rose Bowl, the store had sold tee shirts as recorded in matrix S.

Matrix S

$$\begin{array}{c|cccc} & \text{Small} & \text{Medium} & \text{Large} & \text{Extra Large} \\ \hline \text{White} & 25 & 50 & 75 & 70 \\ \text{Red} & 50 & 78 & 123 & 100 \\ \text{Gray} & 15 & 82 & 105 & 84 \end{array}$$

On which color tee shirt did the dealer make the most profit? Explain how you found your answers.

c. Make an inventory matrix for his remaining stock.

LESSON 2

Ranking and Scatter Plots

Materials: *Activity Sheets 1 and 2*
Technology: dynamic geometry software (optional)
Pacing: 1 class period

Overview

In Lesson 2, students consider how to produce a rating from data that are not already ranks. The scores given to printers for graphics and text are plotted in a scatter plot, and students use the plot to determine where a "best" printer would be. It seems reasonable to add the scores to determine a rating, but several different printers could have the same sum or rating ($G + T = R$). This produces a line, and as the rating varies so does the line. If you begin with a large intercept, the possible ratings will produce a sweeping line beginning in the upper right hand corner of the plot. The best printer will be the printer that the line touches first. If the weights are changed, the slope of the sweeping line is changed, but the way you find a "best" rating remains the same.

Teaching Notes

The mathematics in this section should be familiar to students, although the lesson presents straight lines in a slightly different way. The algebraic approach, to add the scores for the two variables, is a reasonable and simple technique to use. It is important, however, to focus on the geometry both to enhance student understanding of what is actually taking place and to provide students with an alternative way to think about the situation. Students inspect a scatter plot to determine if any printer "dominates" the others in terms of either of the variables. If so, there is clearly a best printer. This could be seen by looking at the individual scores. If the same printer has the highest score for both variables, it is clearly the best. Many students more easily understand what the numbers indicate, however, from looking at a graphical representation.

The sum of the scores yields an equation of the first degree, and for a given printer that equation will become a line with a specific rating R: $G + T = R$ with slope -1. Have students plot the series of lines for the given printers and observe that they are parallel. If you sweep a line with a very large intercept, you will find the best printer where the line meets one of the points representing a printer. Students should recognize that if low scores are valued, the sweeping line will begin with a small intercept near the origin and sweep upward.

If the variables are weighted, the equation of the line becomes $w_1G + w_2T = R$ for the weights $[w_1\ w_2]$. The slope of the line is now $-w_1/w_2$ if T is treated as the dependent variable. The sweeping line will behave the same way but sweep at a different angle. You might have students experiment with the weights to see if they can find a slope that will make a given printer the best.

Technology

Technology is not necessary for this lesson. You might use dynamic geometry software that works with coordinates (Cabri or a new version of Geometer's Sketchpad) so students can experiment easily with the sweeping line and with different weights (slopes).

Follow-Up

Students might consider how to deal with a situation where a high value is good for one variable and a low value is good for the other. What would the sweeping line look like? Will the same reasoning apply?

STUDENT PAGE 17

LESSON 2

Ranking and Scatter Plots

What is the best brand of tennis shoes?

Which kind of shampoo works the best?

How do you decide what brand to buy?

It is very important to make a good decision when you are about to make a major purchase such as a VCR or stereo or mountain bike. Many people look through consumer and trade magazines for ratings on the quality of the item they are considering.

In this lesson you will investigate how these comparisons can be better understood through the use of scatter plots, straight lines that are related to weighted values, and matrices that summarize the relationships.

OBJECTIVES

Write a rating equation for two weighted variables.

Understand the relationships between the relative positions of points on a scatter plot.

Determine ranking using a scatter plot.

INVESTIGATE

Selecting a Printer

Suppose you wanted to buy a dot-matrix printer to use with a personal computer. The table that follows contains quality rating scores for seven printers that were reviewed in the *1993 Buying Guide by Consumer Reports*. Each printer is rated in each of two areas: quality when printing graphics and quality when printing text.

Solution Key

Discussion and Practice

1. **a.** The highest quality in text is NEC. The highest quality in graphics is Panasonic KX-P1124;

b. The lowest quality rating in text is KX-P1124. The lowest quality rating in graphics is NEC.

STUDENT PAGE 18

Dot-Matrix Printers

Model	Graphics	Text
Star NX-2420 Rainbow	81	77
Epson LQ-510	78	78
Panasonic KX-P1124	84	71
Star NX-2420 Multipoint	82	73
Citizen GSX-140	77	76
NEC P3200	60	82
Panasonic KX-P1123	72	79

A scatter plot of Text score (T) versus Graphics score (G) is in Figure 2.1.

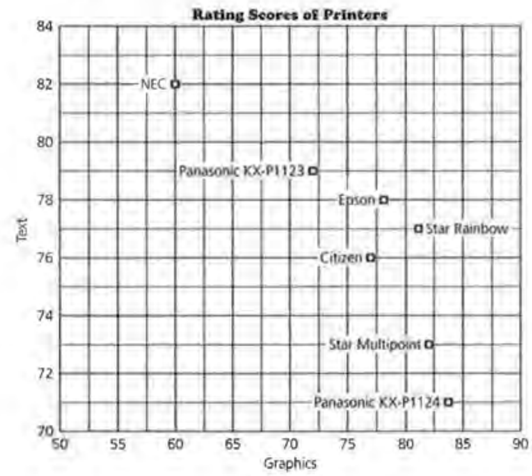


Figure 2.1

Discussion and Practice

1. Use the data in the table on Dot-Matrix Printers.

a. For each variable, which printer has the highest rating?

b. For each variable, which printer has the lowest rating?

To analyze the data more carefully, begin with only two printers, the Epson and the Citizen. The scatter plot in Figure 2.2 shows only the points for these two printers.

2. **a.** For the better quality rating in text, choose Epson with a quality rating of 78.

   b. For the better quality rating in graphics, choose Epson with quality rating of 78.

   c. Epson is better in both fields.

STUDENT PAGE 19

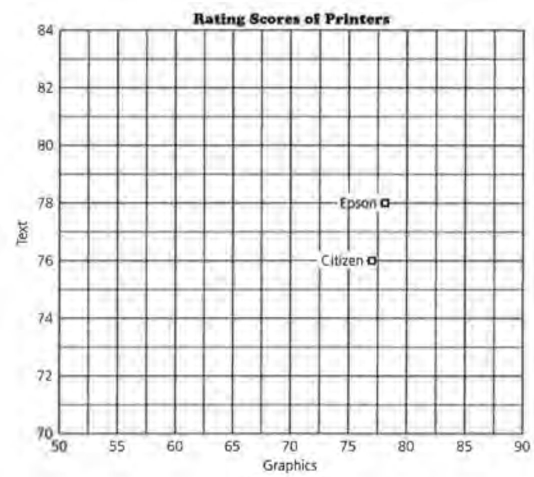
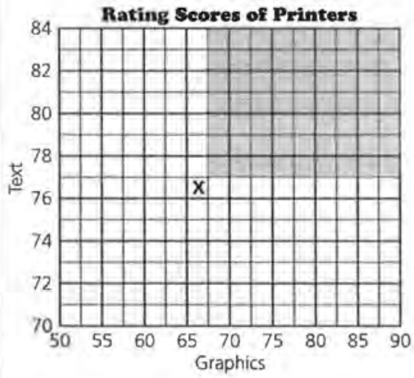


Figure 2.2

2. Use the data to compare the Epson and Citizen printers.
   a. If you only cared about printing text, which of these two printers would you choose? Why?
   b. If you only cared about printing graphics, which printer would you choose? Why?
   c. In terms of the two variables, which printer would you rank higher? Explain why.

In general for a scatter plot in which larger values are better for both variables, you can say that a given point *dominates* all those points having lower values on both variables. That is, if the first printer has larger values on both variables and larger indicates better performance, then the first printer was better than the second printer in terms of these two variables. The

3. a–b. Point P can be anywhere in the crosshatched region.



STUDENT PAGE 20

indicated point X in the scatter plot in Figure 2.3 dominates any point that is below and to the left, that is any point for another printer that lies in the shaded region.

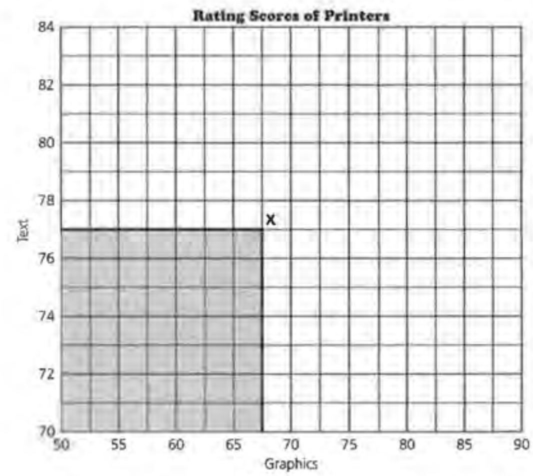


Figure 2.3

3. Suppose point P for another printer dominates the point labeled X in Figure 2.3.

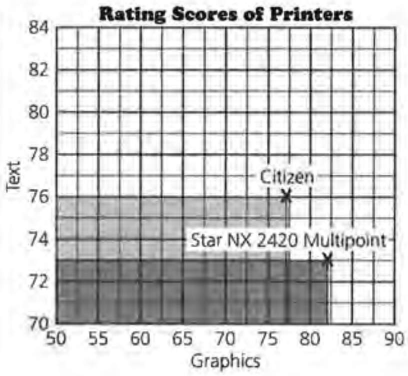
a. Make a sketch of the scatter plot and place point P so that it dominates point X.

b. Sketch the region in the scatter plot that contains all the points that dominate X, that is, all possible locations for P.

4. **a.** Star is superior in graphics, and Citizen is superior in text.

   b. There is no complete dominance by either printer.

5. **a–b.**



   c. The region in which the double crosshatch appears is the region dominated by both printers.

   d. The regions that are cross-hatched in only one direction are dominated by the individual printer.

STUDENT PAGE 21

4. Now consider the Citizen and Star Multipoint, shown in Figure 2.4.

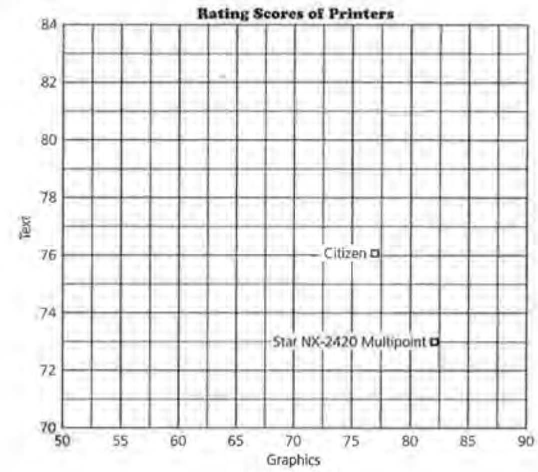


Figure 2.4

   a. How do the two printers compare on quality of printing graphics and text?

   b. Does one printer dominate the other?

5. Make a sketch of Figure 2.4.

   a. Shade the region containing points dominated by the Citizen.

   b. Use another color to shade the region containing the points dominated by the Star.

   c. Identify the region containing points dominated by both printers.

   d. Identify the region containing points dominated by one printer but not the other.

6. **a.** Panasonic KX-P1124 is Jose's choice.

   b. NEC should be Gina's choice.

7. The points are NEC, Panasonic KX-P1123, Epson, Star Rainbow, Star Multipoint, Panasonic KX-P1124.

   a. For each of them, there is no printer that is better for both text and graphics.

   b. The Citizen loses to the Star (and the Epson) on both variables.

STUDENT PAGE 22

6. Return to the original scatter plot in Figure 2.5 for the printers.

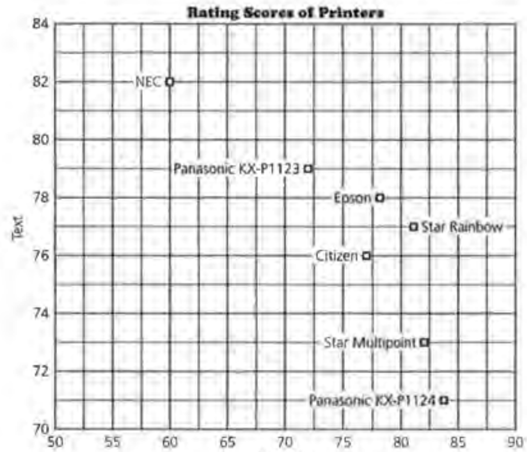


Figure 2.5

a. Jose thinks that graphics is the only thing that should matter when ranking printers. Which printer does he select as best?

b. Gina, on the other hand, thinks that text is all that should matter and that graphics should be ignored. Which printer does she select as best?

7. Which points in Figure 2.5 are not dominated by any other?

   a. Explain why an argument could be made for any of these printers as the best-quality printer.

   b. Explain why a reasonable argument cannot be made for any of the other printers as the best-quality printer.

In Lesson 1 you used weights to combine information on variables. There are two variables in the printer data, text quality and graphics quality—but when Jose ranks the printers, he gives all of the weight to graphics and no weight to text. His weights for (*graphics, text*) are [1 0]. Gina gives zero weight to graphics and all of the weight to text, so her weights are [0 1].

8. a. Answers will vary. Jeff's system tries to get the best of both qualities.

b. Any value as long as they are equal, probably 1 and 1.

c. The product matrix shown comes from adding the ratings for Text and for Graphics to get the overall rating according to Jeff's weights.

d.

$$\begin{bmatrix} 81 & 77 \\ 78 & 78 \\ 84 & 71 \\ 82 & 73 \\ 77 & 76 \\ 60 & 82 \\ 72 & 79 \end{bmatrix} \times \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 158 \\ 156 \\ 155 \\ 155 \\ 153 \\ 142 \\ 151 \end{bmatrix}$$

Star Rainbow is rated highest.

STUDENT PAGE 23

8. Jeff suggested the following compromise: use equal weights for the two variables.

a. Do you think this is a good compromise? Why or why not?

b. What weights will Jeff use?

c. Explain how the matrices below relate to Jeff's solution.

Data matrix · Weight matrix = Ratings matrix

$$\begin{bmatrix} 81 & 77 \\ 78 & 78 \\ 84 & 71 \\ 82 & 73 \\ 77 & 76 \\ 60 & 82 \\ 72 & 79 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \\ \\ \\ \\ \\ \\ \\ \end{bmatrix}$$

d. Compute the ratings for each printer in the matrix above. Which printer(s) has the highest rating?

When the weights are [1 1], the overall rating is Graphics score + Text score, or $G + T$. In matrix terms, this can be written as $[G\ T]\begin{bmatrix} 1 \\ 1 \end{bmatrix} = [R]$, or as the equation $G \cdot 1 + T \cdot 1 = R$.

To find the printer that is the best choice for Jeff, look at Figure 2.6.

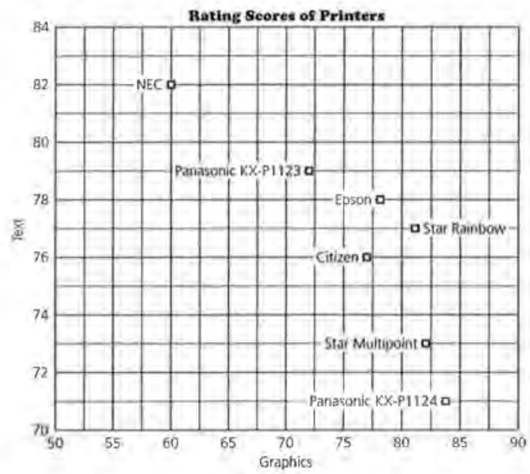
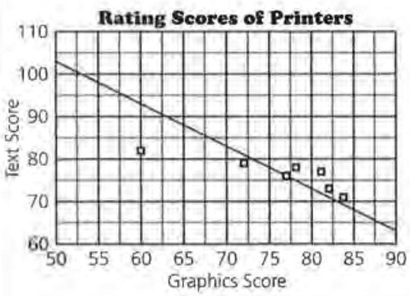


Figure 2.6

9.



a. The slope of the line is –1.

b. Only one of the printers, the Citizen GSX-140, is on the line.

10. a. $G + T = 156$. The slope is again –1.

b. Star Rainbow has a rating of 158.

c. The point for Star Rainbow is above the line.

d. The point will lie below the line.

11. a. All lines have the same slope because both the graphics and text carry the same weight.

b. The statement would be true because the value of the ratings represents the y-intercept, i.e., $T = -G + \text{Rating}$.

STUDENT PAGE 24

For the Citizen, the value of Text score + Graphics score is 153. Any other printer for which Text score + Graphics score is 153 would be just as appealing to Jeff as the Citizen printer—no more appealing and no less appealing. So all printers that fall on the line determined by $G + T = 153$ are equivalent from Jeff's point of view.

9. Graph the line $G + T = 153$ on *Activity Sheet 1.*

a. What is the slope of this line?

b. How many printers fall on that line?

10. Draw in the line determined by the rating for the Epson. Use *Activity Sheet 2.*

a. What is the equation? the slope?

b. What was the rating for the Star Rainbow?

c. Where is the point for the Star Rainbow in terms of the line for the Epson?

d. If a printer has a lower rating than 153, where will the point for that printer lie with respect to the line $G + T = 153$?

11. Figure 2.7 is the scatter plot with several lines added.

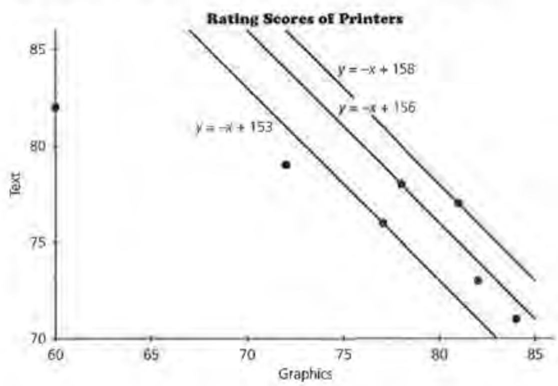


Figure 2.7

a. Why does each line have the same slope?

b. Comment on the statement:

STUDENT PAGE 25

c. Jeff will choose the Star Rainbow because it has the highest value for $G + T$, that is the point highest and to the right.

12. If the categories are equally rated, the slope will always be –1. As the line sweeps down, the first point it touches will have the largest y-intercept, and thus, the highest rating.

13. **a.** Using the rating $[\frac{1}{2}\ \frac{1}{2}]$ would reduce each rating to $\frac{1}{2}$ its original value.

b. The lines will not change because the slope remains the same.

c. The highest rating will remain Star Rainbow.

Since Jeff wants $G + T$ to be large, he wants the printer that falls on the line having the greatest y-intercept.

c. Which printer will Jeff choose and why?

You could find Jeff's ideal printer by taking a *sweeping line* having slope –1, starting on the top right-hand part of the scatter plot beyond any data points, and "sweeping" down and to the left ("to the southwest"), keeping the slope of the line equal to –1, until the moving line touches one of the points as in Figure 2.8.

Rating Scores of Printers

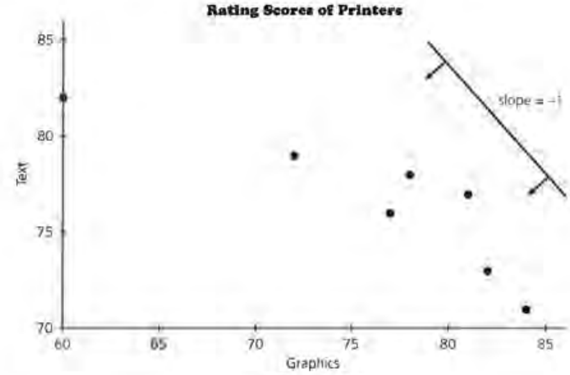


Figure 2.8

12. Why will "sweeping" down using a line with slope –1 find the best printer if the two variables are rated equally?

13. Suppose you had used weights $[\frac{1}{2}\ \frac{1}{2}]$ rather than [1 1].

a. How would these weights have changed the overall rating for each printer?

b. Would the lines on the graph change? Explain.

c. Which printer would now have the highest rating?

Other Slopes

14. Emily is more concerned with graphics than with text, but text quality does matter to her. She wants to take a weight-

14. a.

$$\begin{bmatrix} 81 & 77 \\ 78 & 78 \\ 84 & 71 \\ 82 & 73 \\ 77 & 76 \\ 60 & 82 \\ 72 & 79 \end{bmatrix} \times \begin{bmatrix} \frac{2}{3} \\ \frac{1}{3} \end{bmatrix} = \begin{bmatrix} 79.66 \\ 78 \\ 79.66 \\ 79 \\ 76.66 \\ 67.33 \\ 74.33 \end{bmatrix}$$

b. Star Rainbow and Panasonic KX-P1124 are now tied for the highest rating. The points representing those printers are on the same line, which students may observe.

c. $\frac{2}{3}G + \frac{1}{3}T = R$ implies $2G + T = 3R$, which implies $T = -2G + 3R$ and the slope is -2.

d. 2

STUDENT PAGE 26

ed average of text score and graphics score in which she gives twice as much weight to graphics as is given to text:

Emily's weighted average for a printer will be:

$R = \frac{2}{3} \cdot \text{graphics score} + \frac{1}{3} \cdot \text{text score}.$

a. Using the data matrix of $\begin{bmatrix} 81 & 77 \\ 78 & 78 \\ 84 & 71 \\ 82 & 73 \\ 77 & 76 \\ 60 & 82 \\ 72 & 79 \end{bmatrix}$ and a weight matrix of $\begin{bmatrix} \frac{2}{3} \\ \frac{1}{3} \end{bmatrix}$, compute the rating for each printer.

b. Which printer(s) has the highest rating? Could you have predicted this before you did the calculations? If so, how?

Emily can take a line having slope –2, start at the upper right-hand part of the scatter plot, and sweep down and to the left ("to the southwest"), keeping the slope of the line equal to –2, until the moving line touches one of the points as in Figure 2.9.

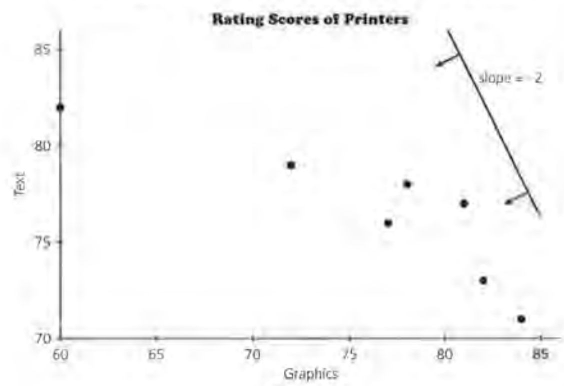


Figure 2.9

c. Verify that the slope of the equation for the sweeping line with these weights is –2.

d. Suppose Emily is considering a printer with ratings of 80 for graphics and 80 for text. What is the drop in text

15. a. The line simultaneously touches Star Rainbow and Panasonic KX-P1124. They have the highest weighted averages.

b. $T + 2G = 239$

STUDENT PAGE 27

score that Emily would give up in order to get a one-point increase in graphics score:

$\frac{\text{change in } T}{\text{change in } G} = ?$

Rating Scores of Printers

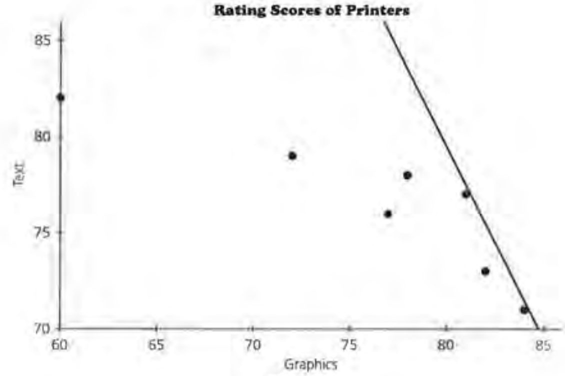


Figure 2.10

15. Consider the sweeping line in Figure 2.10.

a. What point(s) does the line touch first? What does this tell you?

b. What is the equation of this line?

Rating Scores of Printers

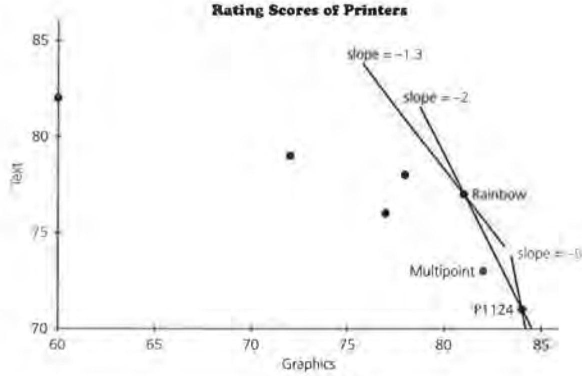


Figure 2.11

16. a. A slope of –5 indicates a preference toward graphics in a ratio of 5 to 1. A slope of –1.3 indicates a preference toward graphics in a ratio of 13 to 10. A slope of –.75 indicates a preference toward text in a ratio of 4 to 3.

b. A slope of zero would indicate that the only consideration is text.

c. The steeper the preference line, the more preference is given to graphics and vise versa.

d. A vertical line would indicate total preference for graphics. The slope would be undefined.

e. No. When the slope of the preference line is ≥ –2, the Rainbow Star beats Multipoint . When the slope is ≤ –2, the Panasonic KX-1124 beats the Multipoint.

Practice and Applications

17. Yes, because one can imagine a line passing through Epson only, above the other points and so indicating a domination of all other points.

STUDENT PAGE 28

16. Figure 2.11 shows lines with several different slopes.

a. What does it mean if a line has slope –5? –1.3? –0.75?

b. What does it tell you if the slope of the preference line is zero?

c. What does a steep "preference line" indicate?

d. What would a vertical line indicate? What would the slope of the vertical line be?

e. Is there a slope that will indicate a preference for the Multipoint? How can you tell?

Summary

If the weights on two variables, v_1 and v_2, are equal, the equation that can be used to generate the ratings is $1 \cdot v_1 + 1 \cdot v_2 = R$. For any value R, this equation will determine a straight line in terms of v_1, and v_2 with slope of –1. Using sets of parallel lines, all with slope –1, you can sweep down toward a plot of (v_1, v_2) from the upper right corner of your plot until you reach a point on the scatter plot. That point will have the maximum rating. If you change the weights, the slope of the line will also change, but the technique will still work.

Practice and Applications

17. Suppose you worked for the people who make the Epson printer. Then you might try to find a rating system combining text and graphics that would make the Epson appear to be the best printer. Think about the plot. Is there a pair of weights for text score and graphics score that make Epson the "winner"? How do you know?

18. **a.** Epson lies above that line.

b. NEC (60, 82) and Star Rainbow (81, 77). $\frac{82-77}{60-81} = \frac{5}{-21}$, which is the slope of the line passing through NEC and Star Rainbow.

c. The slope of $\frac{-5}{21}$ will establish Epson as preferred.

d. Five times the graphics and 21 times the text.

$\begin{bmatrix} 5 \\ 21 \end{bmatrix}$ as a matrix.

19. No. See the answer to 16e. The line connecting NEC and Epson passes above the Panasonic KX-1123.

STUDENT PAGE 29

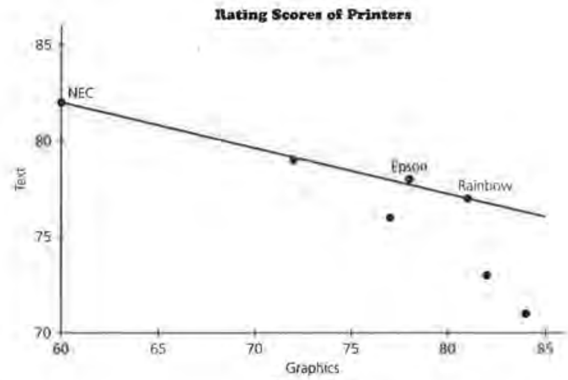


Figure 2.12

18. Look at the line in Figure 2.12 connecting the NEC and Rainbow points.

a. Where does the Epson fall in relation to the line through the NEC and the Rainbow points?

b. What is the slope of the line through the NEC and the Rainbow points?

c. Find a slope for which the Epson is the preferred printer.

d. What pair of weights correspond to the slope you found?

19. Is there a slope that will make the Panasonic KX-P1123 the preferred printer? Explain why or why not.

20. a. They would be the ones located farthest to the left and the farthest down. Jacksonville, San Antonio, San Francisco, Boston, Chicago, Philadelphia, and New York.

b. Houston, Atlanta, and Los Angeles.

c. If the ratings are equal, the slope of the sweeping line starting from the upper right with a large y-intercept (high rating number) will be -1. The first point such a line will pass through will be Houston (73, 65).

$$M - 65 = -1(A - 73)$$
$$M = -A + 138$$
$$M + A = 138$$

d. If you desire an equal rating, choose a sweeping line with slope -1. Sweep from the lower left and move up to the right maintaining the slope of -1. The first point the line touches will have the lowest and, in this case, the best rating. That point would be San Antonio, which would have the lowest rating for air quality and motor vehicle use.

STUDENT PAGE 30

20. Each year the World Resources Institute releases information about the environment for areas in the United States (as well as the rest of the world). The plot in Figure 2.13 below shows the ratings for 14 of the largest cities in the United States. The ratings are based on studies done by the World Resources Institute. For example, air quality is based on daily readings for five pollutants regulated by the Clean Air Act: sulfur dioxide, nitrogen oxides, particulate matter, carbon monoxide, and ozone. The lower the number the better the rating.

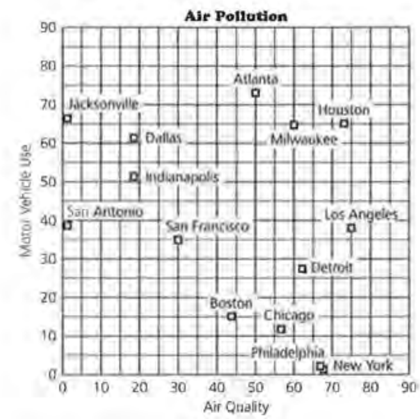


Figure 2.13

Source: *The 1994 Information Please Environment Almanac*

a. Smaller values are now better. How does that affect your interpretation of the plot? Which cities are not dominated by any others with respect to both variables?

b. Looking at the plot, which cities are likely candidates for the "worst" cities environmentally, if you judge only on the two variables in the plot?

c. If you weight air quality and motor vehicle use equally, find the line that will give you the worst (highest) rating. Explain how you found your answer.

d. Explain how you can use the sweeping line to determine the best (lowest) city.

STUDENT PAGE 31

21. The actual data from the environmental study are in the table below.

Environmental Study for Selected Cities

City	Unhealthy Air Quality	Toxic Releases	Energy Use	Motor Vehicle Use
Atlanta	50	31	20	73
Boston	44	17	49	15
Chicago	57	64	68	11
Dallas	19	33	29	61
Detroit	62	47	71	27
Houston	73	71	14	65
Indianapolis	19	46	56	52
Jacksonville	1	51	7	66
Milwaukee	60	52	74	64
Los Angeles	75	32	2	38
New York	68	27	42	1
Philadelphia	67	61	43	2
San Antonio	1	14	19	39
San Francisco	30	11	3	35

Source: *The 1994 Information Please Environment Almanac*

a. Make a scatter plot of (toxic releases, energy use). Which cities seem likely to be the worst cities based on these two variables if they are equally weighted?

b. Weight the toxic releases twice as much as the energy use, and then compute the ratings.

c. What is the slope of the line that will give you the worst city using the weights from part b? Draw in a series of lines with that slope and use them to verify your answer to b.

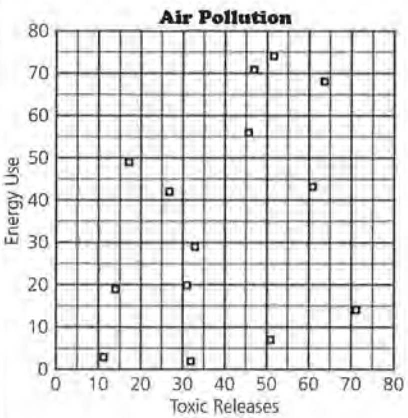
d. How can you tell from the plot if there are any cities that are tied in the ratings?

22. Examine the scatter plot.

a. Can you find a weighting that will make Houston the worst city? Justify your answer.

b. Can you find a weighting that will make Detroit the worst? Again, justify your answer.

21. a.



Milwaukee and Chicago seem to be the worst.

b. The ratings will be

$$\begin{bmatrix} 82 \\ 83 \\ 196 \\ 95 \\ 165 \\ 156 \\ 148 \\ 109 \\ 178 \\ 66 \\ 96 \\ 165 \\ 47 \\ 25 \end{bmatrix}.$$

The worst city is Chicago.

c. The slope would be –2.

$$[T\ E] \times \begin{bmatrix} 2 \\ 1 \end{bmatrix} = 2T + E = R.$$

Then $E = -2T + R$.

d. The line with slope –2 would pass through all points that are tied.

The points would all be on the same straight line.

22. a. Choose Chicago (64, 68) and Houston (71, 14) to determine the slope.

$\frac{68 - 14}{64 - 71} = \frac{54}{-7}$. This means 7 times the toxic releases and 54 times the energy. If the line is drawn any steeper, Houston is worse.

b. No, Milwaukee (52, 74) dominates Detroit (47, 71).

LESSON 3

More on Ranking

Materials: *Activity Sheets 3 and 4; Lesson 3 Quiz*
Technology: graphing calculator
Pacing: 1 to 2 class periods

Overview

In Lesson 1, students worked with airline data in different units, percents, and rates. In this lesson working within a baseball context, students will learn a procedure to equalize units and then build on the sweeping line process from Lesson 2. By specifying some way to make the scores on one variable equivalent to the scores on another, you can, in effect, equalize the contributions of both variables to the rating. Building on the notion of summing the variables from Lesson 2, the rating is now a linear combination or sum of the form $w_1x + w_2y$. The weighted ratings are determined by applying the weight matrix to the data matrix.

Teaching Notes

Be sure students recognize the problem caused by introducing variables measured in different units. In the case of *batting averages* and *runs batted in* used in the lesson, the units are extremely different, which should help make the point: batting averages are decimals between 0 and 1, while runs batted in are in thousands. Some students may feel other data such as the number of games played should be used as well in making any statement about ranking best baseball hitters. Suggest that this will be done later in the module, but right now the data is restricted to the two given variables or else the procedure will become too complicated. Other students may be unfamiliar with baseball. Assure them that such knowledge is not necessary to do the lesson.

Finding some way to equalize the variables is a critical point and one that is actually quite subjective. (This is usually the case in any statistical analysis. Some statistical summaries are often more logical than others for a given set of data, but there are usually several possible ways to approach a data set.)

You can state that two variables are equivalent by using a common measure of center for both variables, by taking minimum values, by using some established norm based on the context for the data, or by another method that seems reasonable. Once this baseline is established, the reciprocal of each measure can be used as the weights. This in effect will establish a sweeping line with a slope of $-w_1/w_2$.

Students may have trouble graphing the lines because the slope is the quotient of the weights and not a number with which they are used to working. Rather than use the slope to draw the line, students can find two points that are not close together that satisfy the equation and use them to sketch the line. This does provide students with a good lesson on scale and how difficult it can be to apply some of the techniques they learned using small integer values in actual situations.

The rest of the lesson is spent looking at the relationship between the weighted ratings, the equation, the plot, and the slope of the sweeping line. Students should be able to go from one to another and to understand how the four are connected and what they tell you about the total ratings. A major problem can occur if students concentrate on making the numbers work out exactly and forget about the context and the questions they are trying to resolve. Encourage students to refer back continually to the situations as they work through the problems.

Follow-Up

You might have students use one of the other methods of equalizing the variables such as using the minimum rather than the mean, and investigate how this approach compares to the one used in the lesson. You might also have students apply the formula to a set of baseball players from their favorite team or to the softball players from their school.

STUDENT PAGE 32

LESSON 3

More on Ranking

Who is the best quarterback? the leading tennis player?

What breed of dog is the most intelligent?

How do you combine successful operations with quality of staff and cost to produce a rating for a hospital or the number of touchdowns, interceptions, and completions to rate a quarterback?

OBJECTIVES

Rank variables measured in different units.

Relate ranking with unequal scales to an equation and to a scatter plot.

Sports figures are constantly being ranked as well as hospitals, movies, and schools: best quarterback, leading tennis player, top-ranked golfer. In each case there are many variables that could be considered important in determining ratings. The variables, however, are not always on the same scale. Some of them are percents, some might be in tens and some in thousands. Is there a way to determine weights that can be adjusted for these different scales? In this lesson you will learn some techniques to help you out when you work with two variables that are on different scales. You will also investigate how algebraic calculations using matrices relate to geometric interpretations of the ratings.

INVESTIGATE

Rating Baseball Players

Baseball players are elected to the Hall of Fame for a variety of reasons: pitching, hitting, home run hitting, and good all-around play. Which player in the Baseball Hall of Fame is the best career hitter? How can you rank some of the top hitters? The career batting average and runs batted in for several candidates are given in the table that follows.

Solution Key

Discussion and Practice

1. a. Those treated here are batting average and runs batted in. Student answers about which are more important will vary with each student.

b. Examples are home runs, triples, doubles; student answers will vary.

c. The batting average would not be affected by the number of seasons since it is a mean. The number of runs batted in would be directly affected by the number of years in a career since it is a total number.

d. Answers will vary. An advantage might be that using a ranking for each category will be an easy way to combine all of the variables even though they have different units. A disadvantage is that the variability in the data is lost; you cannot tell whether the difference between a 3 and 4 rank was 10 points or 100 points.

STUDENT PAGE 33

Hall of Fame Best Hitters

Player	Batting Average (*BA*)	Runs Batted In (*RBI*)
Hank Aaron	.305	2,297
Rod Carew	.328	1,015
Ty Cobb	.367	1,961
Lou Gehrig	.340	1,990
Rogers Hornsby	.358	1,584
Reggie Jackson	.262	1,702
Mickey Mantle	.298	1,509
Willie Mays	.302	1,903
Stan Musial	.331	1,951
Babe Ruth	.342	2,211
Ted Williams	.344	1,839

Source: *Universal Almanac*, 1994

Discussion and Practice

1. Consider the data in the table above.

a. Which aspects of a hitter's performance are treated as important here? Do you think these aspects are important? Why or why not?

b. What other variables might be considered important?

c. Does the number of seasons in the player's career directly affect batting average? Does it directly affect runs batted in? Explain.

d. Suppose you ranked the players from 1 to 11 in each category and used the ranking to obtain a combined rating. Name an advantage and a disadvantage of finding the ratings this way.

2. **a.** No, Musial is dominated by Gehrig, Cobb, and Ruth in both categories.

b. Cobb is dominant in batting average; Aaron is dominant in runs batted in.

c. Carew is weakest in runs batted in, and Jackson is weakest in batting average.

d. No one player completely dominates every player. No point on the graph is both above and to the right of every point simultaneously.

3. **a.** The matrix containing the weighting must have equal values so 1 and 1 are most reasonable choices. The multiplication of the matrices will result in adding the two factors together to get the ranking.

STUDENT PAGE 34

2. Figure 3.1 is a scatter plot of the two variables.

Figure 3.1

a. Is there a way to make Musial come out with the best rating using these variables? Explain.

b. For each of the variables, which of the eleven players seems to be the strongest?

c. For each of the variables, which of the eleven players seems to be the weakest?

d. Is there any player who completely dominates the others? Explain how you can tell.

3. Suppose you decide to use equal weights to ranking the players.

a. Explain how the matrices below relate to your work.

$$\begin{array}{ccccc} \text{Data Matrix} & & \text{Weight Matrix} & & \text{Ratings Matrix} \\ \begin{bmatrix} .305 & 2297 \\ .328 & 1015 \\ .367 & 1961 \\ .340 & 1990 \\ .358 & 1584 \\ .262 & 1702 \\ .298 & 1509 \\ .302 & 1903 \\ .331 & 1951 \\ .342 & 2211 \\ .344 & 1839 \end{bmatrix} & \cdot & \begin{bmatrix} 1 \\ 1 \end{bmatrix} & = & \begin{bmatrix} 2297.305 \\ 1015.328 \\ 1961.367 \\ 1990.340 \\ 1584.358 \\ 1702.262 \\ 1509.298 \\ 1903.302 \\ 1951.331 \\ 2211.342 \\ 1839.344 \end{bmatrix} \end{array}$$

(3) **b.** The ratings are completely dominated by the runs batted in.

4. **a.** They seem equivalent because each is the mean of its individual category.

b. Student answers will vary. They could take the player they feel is best or the one with the highest *RBI*s and *BA*. This could provide a lively discussion.

STUDENT PAGE 35

b. Why are the ratings in the Ratings Matrix problematic?

There is a way to use weights to combine information in a problem such as this. You need to find weights that take into consideration the fact that RBIs are large numbers while batting averages are all less than 1.

In general, with a weight w_1 for BA and a weight w_2 for $RBIs$, a player's overall rating will be

$$BA \cdot w_1 + RBI \cdot w_2 = R.$$

In matrix terms, this can be written as $[BA \;\; RBI]\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = [R]$.

Variables and Different Scales

Since BA and $RBIs$ are in different units, here is one way to get a reasonable pair of weights. First, decide what score on one variable you would consider to be equivalent to a given score on the other variable. That is, these two scores have the same contribution to the overall rating. For example, you could use the minimum batting average of .262 as equivalent to the minimum number of runs batted in, 1015, for the 11 players; or you could consider the maximums in both cases to be equivalent, .367 and 2297. Suppose you decide to use the mean of the batting averages for all 121 of the Hall of Fame players listed in the 1995 *Universal Almanac*, .306, and decide it should have the same contribution to the overall rating as the mean of the $RBIs$ earned by all of the Hall of Fame players, 1171. You can achieve this by choosing

$$w_1 = \frac{1}{.306} = 3.268, \text{ and } w_2 = \frac{1}{1171} = 0.000854.$$

Thus, a player with $BA = .306$ and $RBI = 1171$ would receive a rating of

$$\begin{aligned} w_1 \cdot BA + w_2 \cdot RBI &= 3.268 \cdot 0.306 + 0.000854 \cdot 1171 \\ &= 1 + 1 \\ &= 2. \end{aligned}$$

4. Consider the scores of .306 batting average and 1171 runs batted in.

a. Why might these scores be considered equivalent?

b. Think of other choices for batting average and the number runs batted in that you could consider to be equivalent.

5. **a.** Using the equation from the text above problem 4,

$w_1 \cdot BA + w_2 \cdot RBI = R$

$w_2 \cdot RBI = -w_1 \cdot BA + R$

$RBI = -\frac{w_1}{w_2} \cdot BA + \frac{R}{w_2}$

So the slope $= -\frac{w_1}{w_2} = -\frac{3.268}{0.000854} \approx -3826.7$.

b. The two variables will have equal contributions to the rating because each is the reciprocal of the weights; the product of a number and its reciprocal will be one.

6. **a.** Answers will vary. Two examples are:

$BA = .350$ and $RBI = 1002$;
$3.268(.350) + 0.000854(1002) = 2.0$

$BA = .280$ and $RBI = 1270$;
$3.268(.280) + 0.000854(1270) = 2.0$

b. All points on the line are ordered pairs that make the equation a true statement when substituted. Therefore, all points should generate 2.

c. The point would lie above the line. The value of $\frac{1}{.306}BA + \frac{1}{1171}RBI$ would be greater than 2; the *RBI*-intercept of the line containing that point would be above the *RBI*-intercept of the original line.

STUDENT PAGE 36

5. Consider the equation with weights $w_1 = \frac{1}{306}$ and $w_2 = \frac{1}{1171}$ that would give a player an overall rating of 2 if the player's batting average was .306 and had 1171 runs batted in.

a. What is the slope of the line represented by your equation?

b. Why will these two values for *BA* and *RBI* give equal contribution to the rating that uses the weights $\frac{1}{306}$ and $\frac{1}{1171}$?

6. Consider this relationship in terms of Figure 3.2. The plot contains a graph of the equation from Question 5. *X* represents a player with $(BA, RBI) = (.306, 1171)$.

Figure 3.2

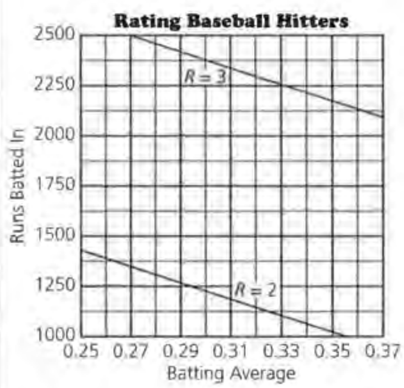
a. Find another point on the line and calculate the overall rating that a player would have with that combination of *BA* and *RBI*s.

b. Explain why any player with (*BA*, *RBI*) falling on the line would have the same rating.

c. If a player had a greater rating using the same weights, would the point (*BA*, *RBI*) lie above or below the line? Explain.

7. The equation using the same weights that would represent all of the players with a rating of 3 would be $3.268BA + 0.000854RBI = 3$. The graph of the line containing any player with a rating of 3 is below. One way to make the graph is to use the same slope and determine what the runs batted in would be for a player with a .250 batting average using the equation $RBI = 3513 - 3826.7BA$.

a.



b. The two equations have the same slope but different *y*-intercepts. The only difference is in the rating.

8. **a.** The matrix for the ratings would be:

$$\begin{bmatrix} 2.958378 \\ 1.938714 \\ 2.87405 \\ 2.81058 \\ 2.52268 \\ 2.309724 \\ 2.26255 \\ 2.612098 \\ 2.747862 \\ 3.00585 \\ 2.694698 \end{bmatrix}$$

STUDENT PAGE 37

7. Find an equation using the weights [3.268 0.000854] that would represent all of the players who have a rating of 3.
   a. Draw that equation in the plot.
   b. How does this equation compare to the equation in problem 6?
8. Return to the original question. How do the 11 baseball players compare, using the weights $[w_1\ w_2] = [3.268\ 0.000854]$?
   a. Using matrices, compute the rating of all 11 players using these weights. That is, calculate

$$\begin{bmatrix} .305 & 2297 \\ .328 & 1015 \\ .367 & 1961 \\ .340 & 1990 \\ .358 & 1584 \\ .262 & 1702 \\ .298 & 1509 \\ .302 & 1903 \\ .331 & 1951 \\ .342 & 2211 \\ .344 & 1839 \end{bmatrix} \cdot \begin{bmatrix} 3.268 \\ 0.000854 \end{bmatrix}.$$

   b. Which player has the highest rating? Does this make sense in terms of the lines your drew in the plot above?
9. Now, consider the original scatter plot redrawn in Figure 3.3, with Henry Aaron displayed as *A*.

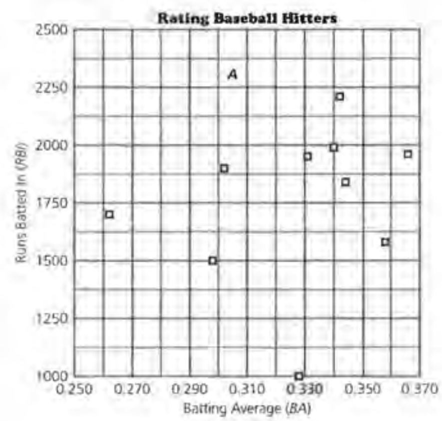


Figure 3.3

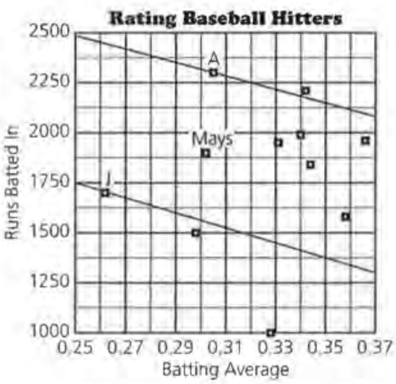
Player	Rating	Ranking
Aaron	2.96	2
Carew	1.94	11
Cobb	2.87	3
Gehrig	2.81	4
Hornsby	2.52	8
Jackson	2.31	9
Mantle	2.26	10
Mays	2.61	7
Musial	2.75	5
Ruth	3.01	1
Williams	2.69	6

b. Babe Ruth has the highest rating. Yes, since the line containing the point for Ruth will have the same slope as the others, but the higher rating (3.006) will create a *y*-intercept greater than the line containing any of the other players. Thus, the points representing any of the other players will be below the line containing Ruth.

9. **a.** Aaron's rating is 2.9583.

b. The equation is $3.268(BA) + 0.000854(RBI) = 2.9583$ or $RBI = -3826.7(BA) + 3464.1$.

Answers will vary about scores with the same rating; sample: a *BA* of .320 and an *RBI* of 2240 would have the same rating as that of Aaron.



c. No other player has the same rating, since no other player's point lies on the line.

10. **a.** Jackson's rating is 2.309724.

b. The equation is $3.268(BA) + 0.000854(RBI) = 2.309724$ or $RBI = -3826.7(BA) + 2704.6$.

Answers will vary about scores with the same rating; sample: a *BA* of .300 and an *RBI* of 1557 would have the same rating as that of Jackson. See the solution for 9b for the graph.

c. No other player has the same rating, since no other player's point lies on the line.

11. Willie Mays would lie between Jackson and Aaron, above the line for Jackson and below the line for Aaron. See the plot in the solution for Problem 9.

12. All points below Aaron's line will have a rating less than Aaron's 2.9583. This is because all equal ratings are points on the line, all

STUDENT PAGE 38

a. What is Henry Aaron's rating using these weights?

b. Create a *BA* and *RBI* for a player who would have the same rating as the rating Aaron got in part a but different scores for *BA* and *RBI*.

Find the equation for all *BA* and *RBI* that would tie Aaron. Graph the equation for this (*BA*, *RBI*) on *Activity Sheet 3*.

c. Does any other player have exactly the same rating as Aaron? How can you tell?

10. Repeat Problems 9a–9c for Reggie Jackson.

11. The rating for Willie Mays lies between the ratings for Reggie Jackson and Henry Aaron. Where does the point for Willie Mays lie in relation to the straight lines from Problems 9b and 10b?

12. Consider the straight line representing all points with rating equal to that of Aaron's. What property do all the points *below* this line have? Explain why.

13. Draw a "sweeping line" on *Activity Sheet 3*.

a. Using your scatter plot and a "sweeping line" such as that in Figure 3.4, explain how you could determine which player has the highest rating (using these weights) without actually doing the calculations.

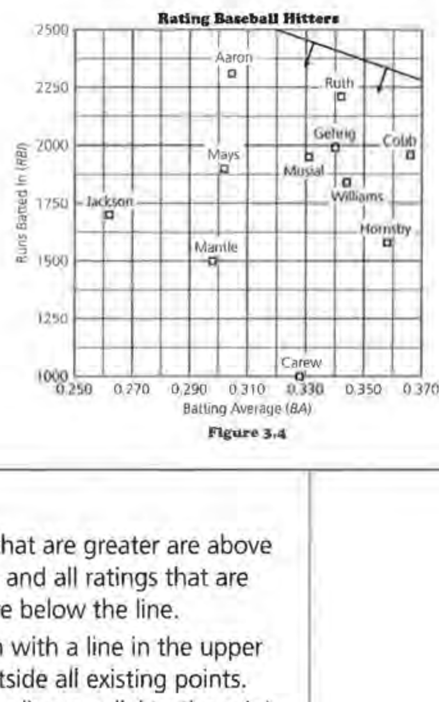


Figure 3.4

ratings that are greater are above the line, and all ratings that are lower are below the line.

13. **a.** Begin with a line in the upper right outside all existing points. Move the line parallel to the original line toward the points to determine which point it will intersect first. That point would belong to the player with the highest rating.

STUDENT PAGE 39

(13) **b.** The slope would change, and you could possibly get a different player with the highest rating.

14. No, there is no way to change the weights to make Williams the top ranked because he is dominated by Cobb.

15. **a.** There is no change in the ranking. The ratings are just larger.

b. The lines will be the same.

c. The answers will not change. The values of the rating will be higher; however the line will remain the same.

16. $RBI = \frac{2}{0.000854} - \frac{3.268BA}{0.000854} =$ $2341.92 - 3826.7(BA)$

a. The slope of the line is –3826.7.

b. The y-intercept is 2341.92.

b. What do you think will happen in terms of the sweeping line if you change your weights and make the batting average worth twice as much as the number of runs batted in?

14. Is there a way to choose a set of weights that will make Ted Williams the top-ranked hitter? Explain why or why not.

15. Earlier you used the weights $\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 3.268 \\ 0.000854 \end{bmatrix}$ to rate baseball hitters. Suppose you multiply both weights by the same positive number, say 50, so the weights become $\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = 50 \cdot \begin{bmatrix} 3.268 \\ 0.000854 \end{bmatrix} = \begin{bmatrix} 163.400 \\ 0.0427 \end{bmatrix}$.

a. How will the new ratings for these players compare with the ratings you found above in problem 8?

b. Consider the straight lines on a scatter plot that represent points with the same rating using the weights 3.268 and 0.000854. How will these straight lines change when the weights are multiplied by 50?

c. How will the answers to parts a and b change if the multiplier is 187.4 instead of 50?

For the weights $[w_1\ w_2] = [3.268\ 0.000854]$, you found that all ($BA$, RBI) combinations that result in a rating of 2 are the solutions of the equation

$$2 = 3.268 \cdot BA + 0.000854 \cdot RBI.$$

16. Rewrite the equation so the number of runs batted in (RBI) is a function of the batting average, BA.

a. What is the slope of the line?

b. What is the y-intercept?

Consider the situation in general for any positive weights $[w_1\ w_2]$. The (BA, RBI) pairs that satisfy a given rating R for those weights determine the equation

$$R = w_1 \cdot BA + w_2 \cdot RBI.$$

Thus,

$$w_2 \cdot RBI = R - w_1 \cdot BA$$

$$RBI = \frac{R}{w_2} - \frac{w_1}{w_2} \cdot BA.$$

17. In general, the slope is the negative of the ratio of the weight applied to *BA* to the weight applied to *RBI* or $-\frac{w_{BA}}{w_{RBI}}$.

a. The slopes are the same; they both have the same weights.

b. The lines will be parallel.

18. a. You have to do the calculations over again to get the new ratings.

b. The lines will have a different slopes.

The slope of the line with weights w_1, w_2 will be $\frac{-w_1}{w_2}$, while the slope of the line with weights w_1 and $3w_2$ will be $\frac{-w_1}{3w_2}$. The second line will be closer to horizontal.

c. Both values will increase, but the one with the higher *RBI* will increase more. In this case, it will be player *A* who will have the higher rating.

19. a. If $w_2 = 0$, the graph will be a vertical line.

If $w_1 = 0$, the graph will be a horizontal line.

b. If $w_2 = 0$, the players are rated solely on *BA*, which would make Cobb first.

If $w_1 = 0$, the players are rated solely on *RBI*, which would make Aaron first.

STUDENT PAGE 40

17. In general, what is the slope of the line for (*BA*, *RBI*)?

a. Explain why the line representing a rating of 3.2 will be parallel to the line representing a rating of 2.5.

b. What can you conclude about all lines determined by a weighting of $[w_1\ w_2]$?

18. Suppose you have ratings using $[w_1\ w_2]$ but think that *RBIs* deserve more weight in the overall rating. You decide to rate the hitters using the weights $[w_1\ 3w_2]$.

a. Can the new ratings be determined from the previous ratings, or do you have to use the original (*BA*, *RBI*) data?

b. How will the straight line for a constant rating of 2 from the new weights compare with the straight line for a constant rating of 2 from the previous weights?

c. Consider the plot in Figure 3.5 showing two players with equal rating according to the original weights.

Rating Baseball Hitters

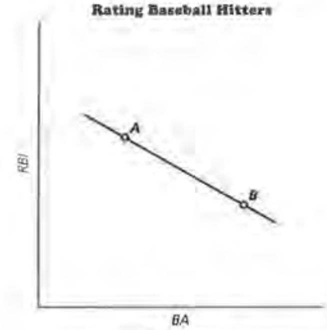


Figure 3.5

Which player has the higher rating using the new weights? Explain why.

19. Consider the lines $w_1 \cdot BA + w_2 \cdot RBI = R$ for a constant rating *R*.

a. What happens to these lines when $w_2 = 0$, and when $w_1 = 0$?

b. Describe how the players are rated and rank-ordered when $w_2 = 0$, and when $w_1 = 0$.

20. a. If $c = \frac{w_2}{w_1}$, the two sets of weights will be equivalent. From $[w_1\ w_2]$, the only way to get the 1 in the first position in the matrix is to divide by w_1.

So if $\begin{bmatrix} \frac{w_1}{w_1} & \frac{w_2}{w_1} \end{bmatrix} = [1\ c]$,

then $c = \frac{w_2}{w_1}$. You might also think about the slopes: $\frac{-w_1}{w_2} = \frac{-1}{c}$, so $c = \frac{w_2}{w_1}$.

b. The ratio of the weights or the slope of the line will have an effect on the ratings and rank orderings. As the ratio of the weights changes, so will the slope of the line. This determines the angle of the sweeping line and thus will affect the order in which the line touches the points, which really determines the ranking.

21. a. The line would also pass through Ruth's point.

b. $\frac{2297 - 2211}{.305 - .342} = \frac{86}{-.037} \approx -2324$

c. Sample: $w_1 = 86$ and $w_2 = .037$

d. $86(BA) + 0.037(RBI) = R$

$86(.305) + 0.037(2297) = 111.219$

$86(.342) + 0.037(2211) = 111.219$

STUDENT PAGE 41

20. Different weights sometimes result in equivalent rankings.

a. For what c will the weights $[w_1\ w_2]$ yield equivalent rankings to those from the weights $[1\ c]$? Explain your reasoning.

b. For any positive weights w_1 and w_2, what is the only property of these two numbers that has a real effect on the ratings and rank-ordering of the hitters? Why?

You can see by examining the graph that neither Henry Aaron nor Babe Ruth dominate each other's career-hitting performance in terms of *BA* and *RBI*. Given a pair of weights for these variables, you can rate these hitters and rank them. The next question to investigate is the following: Is there some pair of weights that will give Aaron and Ruth equal ratings?

21. Suppose some pair of weights gives Aaron and Ruth equal ratings.

a. If so, what can you say about the line of constant rating that passes through Aaron's point?

b. The (*BA*, *RBI*) data for Aaron is (.305, 2297) and for Ruth is (.342, 2211). What is the slope of the line through these two points?

c. Recall that, for weights $[w_1\ w_2]$, the slopes of the lines of constant rating are all equal to $-\frac{w_1}{w_2}$. Find a pair of weights $[w_1\ w_2]$ that should give Aaron and Ruth equal rating.

d. Verify that the two players have equal ratings using your weights $[w_1\ w_2]$ from part c.

Summary

If you have variables that are not in the same scale, determine what might be typical for each of the scales. Use the reciprocal of each of the typical values to define weights that would have equal value in your rating scheme. In general the equation you work with will be $w_1 \cdot y + w_2 \cdot x = R$.

You can find the ratings using either a system of matrices $[x\ y] \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = [R]$ or by generating a sweeping line. If you work with the sweeping line, the slope of the line will be determined by your weights. It is possible to change the weights for your variables and still use either the matrix system or the sweeping line

STUDENT PAGE 42

to find the rankings. If the slope of a line is $-\frac{a}{b}$, you can use the weights $[a \;\; b]$ to determine your ratings.

Practice and Applications

Cars

22. The table below shows horsepower and gas mileage (EPA highway miles per gallon) for each of seven cars. A scatter plot is given in Figure 3.6.

Cars HP and MPG

Car	HP	MPG
Toyota Camry	130	27
Nissan NX200	140	30
Saturn SC	124	33
Toyota Paseo	108	34
Mazda MX-3	130	28
Honda Accord	140	29
Honda Prelude	160	26

Source: *Consumer Reports*, March 1992, July 1992, January 1993.

Figure 3.6

a. Describe the trend in the scatter plot of (HP, MPG).

b. Find a set of weights that will give equal value to both the miles per gallon and the horsepower. Calculate the ratings of each of the cars using those weights.

c. Explain how you could use a sweeping line to select the highest rated car.

23. For some of the cars there is a set of weights for which that car has the highest rating.

Practice and Applications

22. a. The trend seems to be the higher the horsepower, the lower the miles per gallon.

b. Students might use a median, mean, or other value to equalize ratings. The example uses the mean. The mean horsepower = 133.14, and the mean miles per gallon = 29.57.

Let $w_1 = \frac{1}{133.14}$ and $w_2 = \frac{1}{29.57}$.

Then $w_1(hp) + w_2(mpg) = \frac{1}{133.14}$.

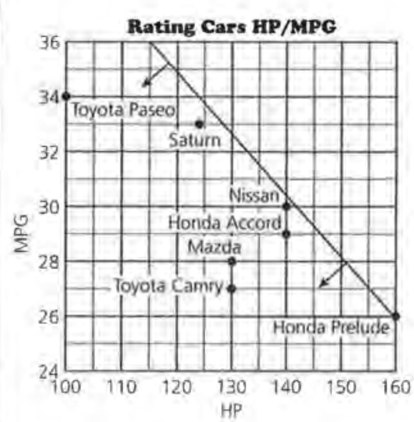
$(133.14) + \frac{1}{29.57}(29.57) = 2$

$0.0075(hp) + 0.0338(mpg) = R$

$$\begin{bmatrix} 130 & 27 \\ 140 & 30 \\ 124 & 33 \\ 108 & 34 \\ 130 & 28 \\ 140 & 29 \\ 160 & 26 \end{bmatrix} \times \begin{bmatrix} 0.0075 \\ 0.0338 \end{bmatrix} = \begin{bmatrix} 1.8876 \\ 2.0640 \\ 2.0454 \\ 1.9592 \\ 1.9214 \\ 2.0302 \\ 2.0788 \end{bmatrix}$$

Car	Rating	Rank
Toyota Camry	1.89	7
Nissan NX2000	2.06	2
Saturn	2.05	3
Toyota Paseo	1.96	5
Mazda MX-3	1.92	6
Honda Accord	2.03	4
Honda Prelude	2.08	1

c. Draw the line having slope –0.2219 and move the line from the upper right down to the left, maintaining the same slope, until it hits one or more points. (You may want to show this on an overhead projector.)



23. a. The Toyota Paseo, Saturn, Nissan, and Honda Prelude could be highest rated, since they form the "top-right boundary" of the points in the scatter plot.

b. Student responses will vary. For example,

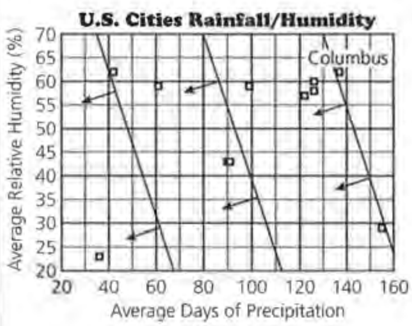
$\frac{34-30}{108-140} = \frac{4}{-32}$, implying $w_1 = 32$ and $w_2 = 4$. These weights make the Toyota Paseo and the Nissan NX2000 tie. Since the Saturn is above the line connecting these two, it will have a higher rating with these weights.

$$\begin{bmatrix}130 & 27\\140 & 30\\124 & 33\\108 & 34\\130 & 28\\140 & 29\\160 & 26\end{bmatrix} \times \begin{bmatrix}4\\32\end{bmatrix} = \begin{bmatrix}1384\\1520\\1552\\1520\\1416\\1488\\1472\end{bmatrix}$$

Saturn = 1552

24. a. Phoenix is the least in rainfall and least in relative humidity. Seattle has the most rain. San Diego and Columbus have the highest humidity.

b.



The city with the highest rating is Columbus, Ohio. The rating means that you favor rain over humidity.

STUDENT PAGE 43

a. Use the scatter plot to determine which cars those are.

b. Find a set of weights for which the Saturn SC has the greatest weighted average.

24. The average number of inches of rainfall and the average relative humidity in the afternoon (in percent) for selected cities in the United States are contained in the table below and graphed in Figure 3.7.

Average Rainfall/Humidity

City	Yearly Average Number of Precipitation Days	Yearly Average Relative P.M. Humidity (Percent)
Mobile, AL	122	57
Phoenix, AZ	36	23
San Diego, CA	42	62
Boise, ID	90	43
Chicago, IL	126	60
Boston, MA	126	58
Omaha, NB	99	59
Albuquerque, NM	61	59
Columbus, OH	137	62
Seattle, WA	155	29
Salt Lake City, UT	91	43

Source: *American Almanac*, 1994–95

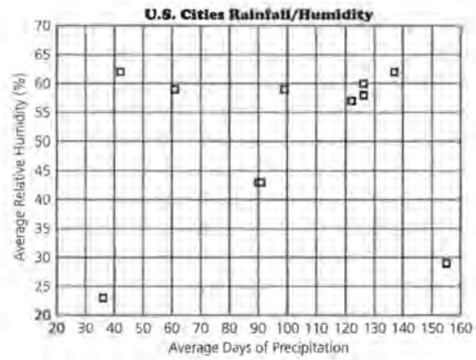


Figure 3.7

a. Which of the cities seems to have the least rain and the lowest humidity? the most?

STUDENT PAGE 44

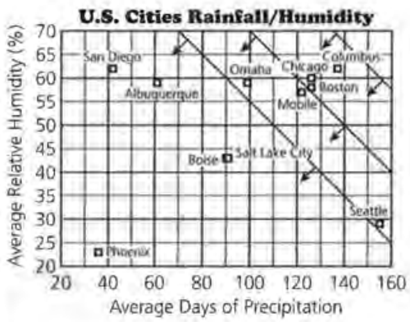
c. The ratio of the weights for precipitation to humidity is 3 to 2, three times the rainfall and two times the humidity. A slope of –3/2 would give you an equation of the form $3P + 2H = R$, for weights of [3 2].

25. a. The mean of the rainfall is 98.6, and the mean of the relative humidity is 50.5.

$w_1 = \frac{1}{98.6} = 0.0101$; $w_2 = \frac{1}{50.5}$ $= 0.0198$.

$0.0101(R) + 0.0198(H) = \text{rating}$

b.



Columbus has the highest rating. This would indicate that if you weighted the variables as .0101 and .0198, Columbus would have the highest ranking for those variables.

c.

$$\begin{bmatrix} 122 & 57 \\ 36 & 23 \\ 42 & 62 \\ 90 & 43 \\ 126 & 60 \\ 126 & 58 \\ 99 & 59 \\ 61 & 59 \\ 137 & 62 \\ 155 & 29 \\ 91 & 43 \end{bmatrix} \times \begin{bmatrix} 0.0101 \\ 0.0198 \end{bmatrix} = \begin{bmatrix} 2.3608 \\ 0.8190 \\ 1.6518 \\ 1.7604 \\ 2.4606 \\ 2.4210 \\ 2.1681 \\ 1.7843 \\ 2.6113 \\ 2.1397 \\ 1.7705 \end{bmatrix}$$

Columbus rating = 2.6113

b. Suppose the slope of the sweeping line was –1.5. Draw in some of the lines using *Activity Sheet 4*. Which of the cities would have the highest rating? What does this rating tell you?

c. Suppose the slope of the sweeping line was –1.5. What does this tell you about the weights for the two variables?

25. Find two weights that you think would make the variables have equal contribution.

a. Find the equation you could use to determine the ratings.

b. Draw in the sweeping lines on the second plot on *Activity Sheet 4* and determine which city would have the highest rating.

c. Calculate the ratings using matrices. How do your results compare to those you found with the sweeping line?

d. Are there any weights that would make the ratings tied for two of the cities? How can you tell?

Extension

26. Return to the top hitters in the Baseball Hall of Fame. Use your weights from problem 21, part c, to answer the following:

a. Suppose a player has a career *BA* of .305. What number of *RBIs* will result in *BA* and *RBI* having equal contribution to this player's overall rating?

b. If a second player has *BA* of .342, what number of *RBIs* will result in *BA* and *RBI* having equal contribution to this player's overall rating?

c. Now recall the weights you found in problem 21, part c. Would you conclude that these weights are rather reasonable, or that *BA* is weighted too heavily, or that *BA* is not weighted heavily enough? Explain your reasoning.

City	Rating	Rank
Mobile, AL	2.36	4
Phoenix, AZ	0.82	11
San Diego, CA	1.65	10
Boise, ID	1.76	9
Chicago, IL	2.46	2
Boston, MA	2.42	3
Omaha, NB	2.17	5
Albuquerque, NM	1.78	7
Columbus, OH	2.61	1
Seattle, WA	2.14	6
Salt Lake City, UT	1.77	8

Columbus is the highest ranked city using equal weights for the two variables, which agrees with the results suggested by the sweeping line.

d. Two groups of cities have ratings that are very close to tied using this weighting.

Salt Lake City	1.76
Albuquerque	1.78
Boise	1.77
Chicago	2.46
Boston	2.42

Two cities will have tied ratings if the points representing the two cities are on the same sweeping line.

Extension

26. a. $86(BA) + 0.037(RBI) = \text{Rating}$

$86(0.305) = 0.037(RBI)$

$708.91 = RBI$, implying $RBI = 709$

b. $86(0.342) = 0.037(RBI)$

$794.92 = RBI$ implying $RBI = 795$

c. Responses will vary. To some it may seem to place too little emphasis on batting average, since a *BA* of .342 should contribute more to the rating than 795 RBIs. Thinking about the actual data, we see that a .342 batting average is one of the highest batting averages, while 795 is a very low number for the number of runs batted in, lower than any of the group of Hall of Fame players.

LESSON 4

Ratings with Three or More Variables

Materials: three-dimensional models kit, clay, glue, pipe cleaners or other similar items, Test
Technology: graphing calculator, unit-to-unit line
Pacing: 3 to 4 class periods

Overview

In this lesson, students extend the procedure for finding a formula for weighted ratings from two variables to three, then to four variables or more. The algebraic extension is logical and follows quite directly from the work in Lesson 3. The geometric interpretation, however, introduces students to three-dimensional geometry. They investigate planes and equations of planes in terms of rating baseball players and cities.

Teaching Notes

Students begin by looking at a plot of data in three dimensions. They might need practice plotting the points if this idea is new. When possible, use a concrete model to enhance students' understanding. A model might be made by using Tinkertoy® or one of the common geometric construction kits. Be sure there is a way to mark the axes in units. The premise is that a sweeping plane will yield the *best* point for a given set of weights just as a sweeping line did in two dimensions. (A sheet of paper or light cardboard can be used as the plane.) A single equation in three dimensions with one, two, or three variables is the equation of a plane. (In this case, it is the equation that is generated by using the three weights and the data for one of the players.) Students consider the line that an equation in one of the coordinate planes (xy-plane) makes and investigate what will happen when a third variable, z or height, is added and that variable is varied. To grasp what a particular plane looks like, students are asked to make a trace of the plane by drawing the line segment the plane would make as it intersects each coordinate plane.

Once students understand that a given equation in three variables yields a plane, the next step is to consider when two planes are parallel. Using the intercepts to fix a given plane, students investigate the relationship between the intercepts and the angle that the trace of the plane makes with a coordinate plane. They discover that the angles are determined by the intercepts: $\tan A = w_1/w_2$, independent of the rating factor R. This means there is an infinite set of planes with traces that will have those same angles. If two planes are parallel, the angles made by each trace with a coordinate plane will be congruent. Essentially, this means that the ratio between any two of the intercepts must be constant for all parallel planes. Thus, $w_1x + w_2y + w_3z$ for a given set of weights w_1, w_2, w_3 will produce a series of parallel planes for the data points (x, y, z). The plane that touches the first data point as it sweeps toward the origin from the outside in will be the plane that determines the highest rating or the *best*.

The algebraic extension to four or more variables is straightforward and follows from the formulas for two and three variables. This is a good time to make the point that we need algebra to carry out work that involves more than two or three variables and cannot be done geometrically. Pictures help make things clear and provide a way to gain insights into a situation, but they are not always practical nor possible.

There is an extension in the practice exercises that allows students to generalize the geometric interpretation to n-space using vectors. The exposition is done in only two dimensions, and students may need to see an actual example worked through before the procedure is generalized. For example, you might plot four specific points such as (2, 2), (6, 1), (5, 3), and (4, 5). Then use simple weights such as [4 3] and let the point (x, y) in the extension be (12, 5).

Technology

Students should use the matrix menu of a graphing calculator to find the weighted ratings for the different data sets. If they are able to download the data and link with a primary source, it will enable them to concentrate on the mathematics instead of entering the data.

Materials

Materials for building three-dimensional models and to show the data points in the model would be useful. You might use Tinkertoys® or a commercial three-dimensional model kit. Try to show where a data point would be in the space by using a piece of clay, a gumdrop, or a ball of glue for hanging pictures attached to a pipe cleaner or straw or Tinkertoy® base.

Follow-Up

Encourage students to look for items being rated in the newspaper and discuss how these ratings might have been constructed. Students might even select some activity around the school or community, collect their own data, and create a rating formula.

STUDENT PAGE 45

LESSON 4

Ratings with Three or More Variables

Baseball statistics are kept on many variables: batting average, home runs, hits, number of games, runs batted in. How do you use all of the data to rate the players?

Just as cities, hospitals, and airlines are rated on many variables, so are colleges. How can all of the variables be used to find a rating?

When you rate something you usually have many variables to consider. In Lessons 2 and 3 you investigated rating formulas with two variables and the corresponding geometric representation in the two-dimensional coordinate plane. When you use three variables, you can find an algebraic formula much as you did before, but now the geometric representation is in three dimensions. In this lesson you will find weighted rating formulas for three or more variables, learn to plot points in 3-D, write the equation of a plane and of a line in 3-D, and think about the geometry of finding weighted ratings for three variables.

OBJECTIVES

Generalize the approach to the ratings problems using matrices and algebra.

Investigate algebraic representation of a plane.

INVESTIGATE

Baseball Statistics

Consider once again the Hall of Fame baseball players from Lesson 3. Suppose you knew the number of home runs hit by each of the players in addition to the information about batting average and runs batted in.

STUDENT PAGE 46

Batting Average, Runs Batted In, Home Runs

Player	Batting Average (*BA*)	Runs Batted In (*RBI*)	Home Runs (*HR*)
Hank Aaron	.305	2,297	755
Rod Carew	.328	1,015	92
Ty Cobb	.367	1,961	118
Lou Gehrig	.340	1,990	493
Rogers Hornsby	.358	1,584	301
Reggie Jackson	.262	1,702	563
Mickey Mantle	.298	1,509	536
Willie Mays	.302	1,903	660
Stan Musial	.331	1,951	475
Babe Ruth	.342	2,211	714
Ted Williams	.344	1,839	521

Source: *Universal Almanac*, 1994

A plot of the three variables for each player would be a three-dimensional plot such as the one shown below.

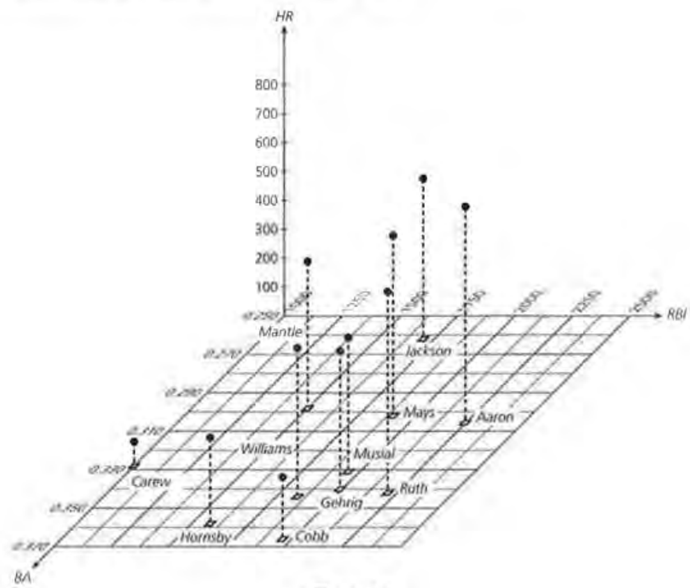


Figure 4.1

STUDENT PAGE 47

Solution Key

Discussion and Practice

1. **a.** Aaron dominates with respect to home runs. The point representing Aaron is the farthest out with respect to the home run axis.

b. Henry Aaron dominates with respect both to home runs and to runs batted in. The point representing Aaron would be located in the first octant and be farthest from the axes for those variables.

2. Yes, it is possible to have two or more players with the same rating but different statistics for batting average, runs batted in, and home runs. Consider the following two players, both with a rating of 5.63.

Player	Batting Average	Runs Batted In	Home Runs
1	.350	2115	509
2	.335	2250	496

3. 3.268 (.305) + 0.000854 (2297) + 0.00526 (755) = 6.93

Aaron would have a rating of 6.93.

Discussion and Practice

1. Study the plot carefully.

a. Is there one player who dominates with respect to home runs? Where is that player in the plot?

b. Is there one player who dominates with respect to two variables? If so, where is that player in the plot?

The algebraic approach for two variables began with identifying some sort of sensible weight to make the variables approximately equivalent or some sort of standard to use as a baseline weight for each variable. For the Hall of Fame players, you used the means of *BA* (.306) and *RBI* (1171) as weights to find a general equation:

$$R = w_1\ BA + w_2\ RBI$$
$$R = \frac{1}{.306}\ BA + \frac{1}{1171}\ RBI \text{ or}$$
$$R = 3.268\ BA + 0.000854\ RBI$$

If a third variable, home runs, is included in the analysis, it seems reasonable to use the mean of the home runs for all of the players in the Hall of Fame as the baseline for the weight for *HR*s. Using this mean of 190 and extending the equation for two variables, you get

$$R = w_1\ BA + w_2\ RBI + w_3\ HR$$
$$R = \frac{1}{.306}\ BA + \frac{1}{1171}\ RBI + \frac{1}{190}\ HR \text{ or}$$
$$R = 3.268\ BA + 0.000854\ RBI + 0.00526\ HR.$$

2. Is it possible to have more than one person with the same *R* but with different batting averages, number of runs batted in, and number of home runs? Explain your answer and give an example.

3. Consider the ordered triple (*BA*, *RBI*, *HR*) that belongs to Henry Aaron: (.305, 2297, 755). What rating would the formula give for Aaron?

Where on the plot would you find the players who have the same rating as Aaron? To answer, consider a simple equation: $y = 6$. In one dimension, this equation will generate a point. In two dimensions, it will generate a set of points where x is any value, and y is always 6. In three dimensions, it will generate a set of points where x is any value, z is any value, and y is always 6.

STUDENT PAGE 48

4. **a.** $\{(x, y) \mid y = 6\}$ represents a horizontal line passing through the point (0, 6) in a two-dimensional plot.

b. $\{(x, y, 0) \mid y = 6\}$ represents a line parallel to the x-axis, lying in the xy-plane and perpendicular to the y-axis in a three-dimensional coordinate system.

c. $\{(x, y, 3) \mid y = 6\}$ represents a line parallel to the x-axis and passing through the point (0, 6, 3) in a three-dimensional coordinate system.

4. Consider the equation $y = 6$ in two dimensions and in three dimensions.

a. What geometric figure is represented in a two-dimensional coordinate plane by $\{(x, y) \mid y = 6\}$?

b. What figure is represented in a three-dimensional coordinate system by $\{(x, y, 0) \mid y = 6\}$ pictured in the diagram below?

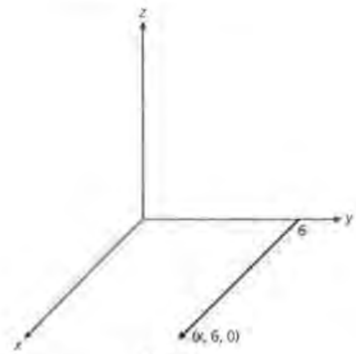


Figure 4.2

c. What figure is represented in a three-dimensional coordinate system by $\{(x, y, 3) \mid y = 6\}$ pictured in the diagram below?

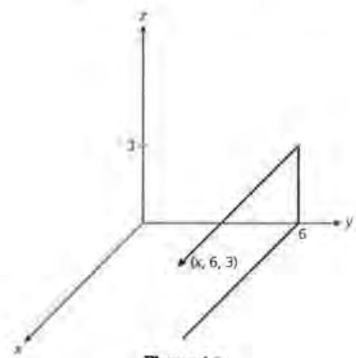
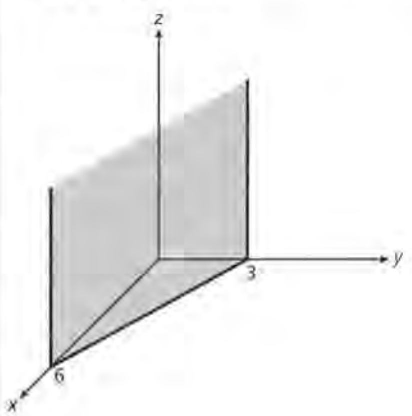


Figure 4.3

d. $\{(x, y, z) \mid y = 6\}$ represents a plane parallel to the xz-plane and perpendicular to the xy-plane creating the trace of $y = 6$ in the xy-plane.

5. The statement is true. Student justifications will vary. In the xy-plane, the graph will be a line. The coefficient of z is 0, so the line will have all possible heights (z-coordinates), thus creating a plane perpendicular to the xy-plane.



STUDENT PAGE 49

d. What figure is represented in a three-dimensional coordinate system by $\{(x, y, z) \mid y = 6\}$?

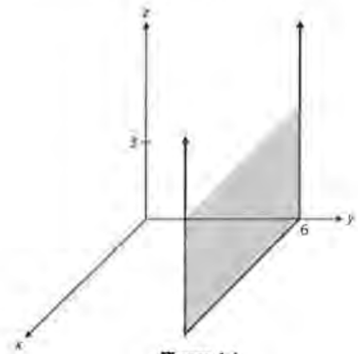


Figure 4.4

5. Consider the statement: For some a and b not both equal to zero, the graph of $ax + by = c$ in three dimensions will be a plane parallel to the z-axis. Use the equation $x + 3y = 6$ to help decide whether the statement is true or false.

6. A trace of a plane is the line in which the plane intersects one of the coordinate planes (the xy-plane, the yz-plane, or the xz-plane). To draw a picture of a plane, you can often draw the traces to show how the plane would look in the first octant.

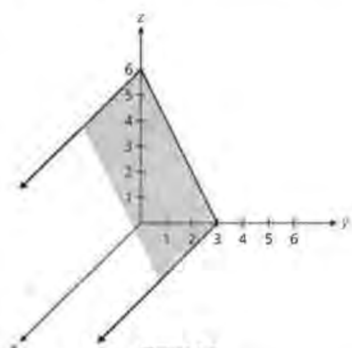
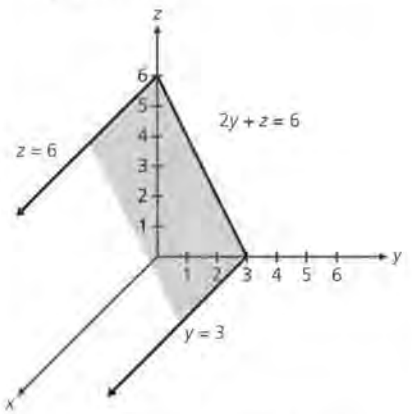


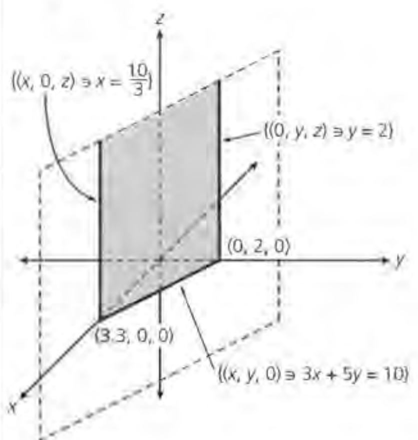
Figure 4.5

6. a. The three traces shown in the figure are given by the following equations.

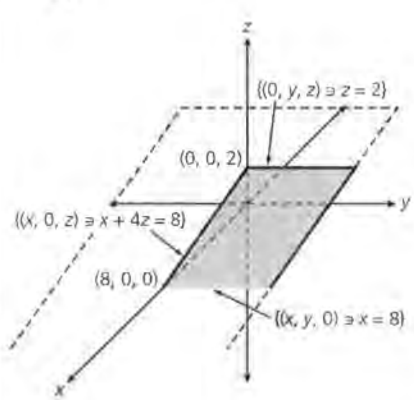


xy-plane $\{(x, y, 0) \mid y = 3\}$
yz-plane $\{(0, y, z) \mid 2y + z = 6\}$
xz-plane $\{(x, 0, z) \mid z = 6\}$

b.



c.



STUDENT PAGE 50

a. What is the equation for each of the three traces in the figure drawn above?

Sketch the traces for each of the equations to see how the plane would look in the first octant:

b. $3x + 5y = 10$

c. $x + 4z = 8$

d. $2y + 5z = 10$

7. The traces of the equation $x + 2y + 3z = 6$ are shown in the drawing below. Write an equation for each of the traces and explain their origin.

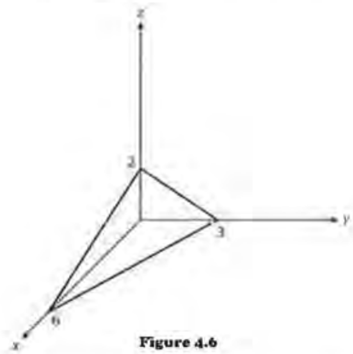


Figure 4.6

To answer the question posed earlier, all of the points that satisfy the equation determined by the weights [3.268 0.000854 0.00526] and Aaron's point (.305, 2297, 755) will be on a *plane* in three-dimensional space. This means that all players who have the same rating of 6.93 for the given weights will lie on that plane. The equation of that plane is

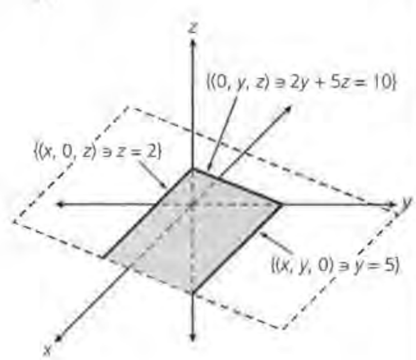
$$3.268\,BA + 0.000854\,RBI + 0.00526\,HR = 6.93.$$

8. Find a possible (*BA*, *RBI*, *HR*) for a player who had the same total rating as Aaron.

a. Explain how you know they have the same rating.

b. Write the equations of the traces of the plane representing all of the players with a rating of 6.93. Make a rough sketch of the plane using the traces.

d.



7. The three traces shown have the following equations.
xy-plane $\{(x, y, 0) \mid x + 2y = 6\}$
yz-plane $\{(0, y, z) \mid 2y + 3z = 6\}$
xz-plane $\{(x, 0, z) \mid x + 3z = 6\}$

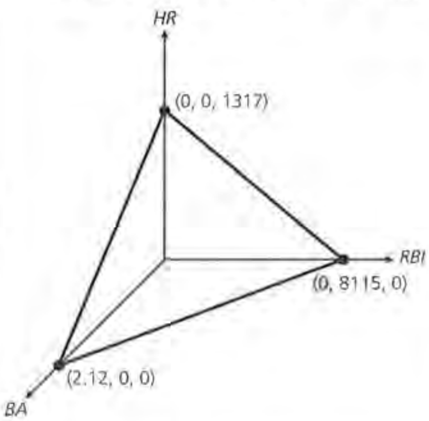
8. a. To determine the rating, you evaluate the equation:
$3.268BA + 0.000854\,RBI + 0.00526HR = R$
for a given *BA*, *RBI*, and *HR*. Aaron's rating was 6.93. One example that will give the same rating as Aaron is $BA = .342$, $RBI = 2211$, and $HR = 746$.

STUDENT PAGE 51

b. The three traces have the following equations: In the *BA-RBI*-plane (*BA*, *RBI*, 0), the equation will be $3.268BA + 0.000854RBI = 6.93$ with intercepts (2.12, 0, 0) and (0, 8115, 0).

In the *RBI-HR*-plane (0, *RBI*, *HR*), the equation will be $0.00854RBI + 0.00526HR = 6.93$ with intercepts (0, 8115, 0) and (0, 0, 1317).

In the *BA-HR* plane (*BA*, 0, *HR*), the equation will be $3.268BA + 0.00526HR = 6.93$ with intercepts (2.12, 0, 0) and (0, 0, 1317).



9. **a.** The equation for Willie Mays will be $3.268BA + 0.000854RBI + 0.00526HR = 6.08$.

b. All players that have the same rating as Willie Mays will be represented by points on the same plane as Mays, defined by that rating. The plane will have the following traces.

BA-RBI plane: (*BA*, *RBI*, 0) where $3.268BA + 0.000854RBI = 6.08$

RBI-HR plane: (0, *RBI*, *HR*) where $0.00854RBI + 0.00526HR = 6.08$

BA-HR plane: (*BA*, 0, *HR*) where $3.268BA + 0.00526HR = 6.08$

10. Any two planes will be parallel, intersect to form a line, or coincide.

11. **a.** *GF* is parallel to *DB*. One possible explanation follows.

In the *xy*-plane, the figure *OBD* is a triangle.

9. To see how the other players fit into the picture, use the same weights but a new player, say Willie Mays.

   a. Write the equation that will give the rating for Willie Mays and determine the rating.

   b. Where is the set of players who, with these weights, will have the same rating as Willie Mays?

The next question to consider is how the plane containing Willie Mays is related to the plane containing Henry Aaron.

10. Describe the possible geometric relationships between any two planes.

11. Consider the plane $x + 2y + 3z = 6$ and the plane $x + 2y + 3z = 12$. The trace for each of the planes is drawn in Figure 4.7.

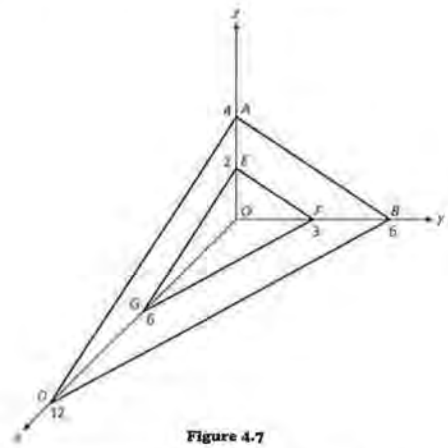


Figure 4.7

   a. Consider the *xy*-plane. How is the line containing *GF* related to the line containing *DB*? Explain how you know.

   b. Consider the *yz*-plane. How is the line containing *EF* related to the line containing *AB*? Explain how you made your conclusion.

   c. What conclusion can you make about the plane *ADB* and the plane *GEF*? Explain your reasoning.

OG = 6 and *OD* = 12; therefore, *G* is the midpoint of *OD*.

OF = 3 and *OB* = 6; therefore, *F* is the midpoint of *OB*.

The segment joining midpoints of two sides of a triangle is parallel to the third side and equal to one-half its length. Therefore, the line containing segment *GF* is parallel to the line containing the segment *DB*.

b. *EF* is parallel to *AB*. The reasoning is similar to that in 11a.

In the *yz*-plane, the figure *OAB* is a triangle.

OF = 3 and *OB* = 6; therefore, *F* is the midpoint of *OB*.

OE = 2 and *OA* = 4; therefore *E* is the midpoint of *OA*.

The segment joining midpoints of two sides of a triangle is parallel to the third side and equal to one-half its length. Therefore, the line containing segment *EF* is parallel to the line containing the segment *AB*.

(11) c. Since the lines containing *AB* and *DB* have a point in common, they are intersecting straight lines. Likewise, lines containing *EF* and *GF* are intersecting straight lines. Knowing that $AB \parallel EF$ and $DB \parallel GF$ and that two intersecting lines form a plane, you can conclude that the planes defined by the intersecting lines must be parallel.

d. Lines *AD* and *GE* are parallel because corresponding angles are equal in the *xz*-plane. Likewise, *EF* and *AB* are parallel. Since the two intersecting lines form a plane, the plane *ADB* is parallel to the plane *GEF*.

12. a. $\tan \angle EGO = \dfrac{\frac{R}{w_3}}{\frac{R}{w_1}} = \frac{w_1}{w_3}$

$\tan \angle GEO = \frac{w_3}{w_1}$

$\tan \angle EFO = \frac{w_2}{w_3}$

$\tan \angle FEO = \frac{w_3}{w_2}$

$\tan \angle GFO = \frac{w_2}{w_1}$

$\tan \angle FGO = \frac{w_1}{w_2}$

b. $\tan \angle EFO = \frac{5}{1}$

$\angle EFO = 78.7°$; so $\angle FEO = 11.3°$

$\tan \angle EGO = \frac{1}{1}$; $\angle EGO = 45°$;

so $\angle GEO = 45°$.

$\tan \angle GFO = \frac{5}{1}$; $\angle GFO = 78.7°$;

so $\angle FGO = 11.3°$.

c. $\tan \angle EFO = \frac{10}{2}$; $\angle EFO = 78.7°$;

so $\angle FEO = 11.3°$.

STUDENT PAGE 52

d. Suppose you know that $\angle EGO$ is congruent to $\angle ADO$ and that $\angle EFO$ is congruent to $\angle ABO$. What conclusions can you make?

12. Now consider the general plane $w_1 x + w_2 y + w_3 z = R$ for some given R. The intercepts can be written as $(\frac{R}{w_1}, 0, 0)$, $(0, \frac{R}{w_2}, 0)$, $(0, 0, \frac{R}{w_3})$.

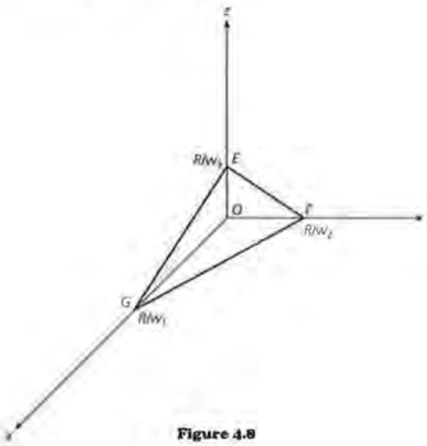


Figure 4.8

a. Find a trigonometric relationship between $\angle EGO$ and the intercepts. Do the same thing for $\angle GEO$, $\angle EFO$, $\angle FEO$, $\angle GFO$, and $\angle FGO$.

b. Find the angles that the traces of the plane make with each of the axes if the plane is given by $x + 5y + z = 8$.

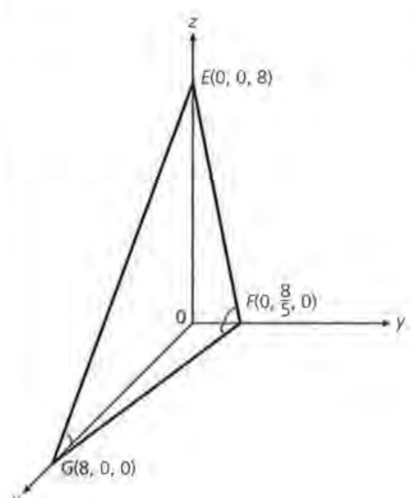
c. Find the angles for a plane given by $2x + 10y + 2z = 8$. What conclusion can you make about this plane and the plane from b?

13. Consider the planes $x + 2y + 4z = 8$ and $2x + y + 4z = 12$.

a. Make a trace of each plane.

b. Find the angle each trace makes with each coordinate axis.

c. Are the two planes parallel or not? Justify your answer.



$\tan \angle EGO = \frac{2}{2} = 1$; $\angle EGO = 45°$;

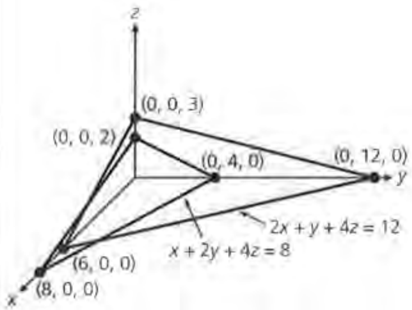
so $\angle GEO = 45°$.

$\tan \angle GFO = \frac{10}{2}$; $\angle GFO = 78.7°$;

so $\angle FGO = 11.3°$.

The two planes are parallel. The corresponding angles are congruent.

13. a.



b. For the plane $x + 2y + 4z = 8$;

The angle trace $x + 4z = 8$ makes in the xz-plane with the x-axis is $\tan \angle A = 1/4$; $\angle A = 14°$; with the z-axis; $\tan \angle B = 4$; $\angle B = 76°$.

The angle the trace $2y + 4z = 8$ makes in the yz-plane with the y-axis is $\tan \angle C = 1/2$; $\angle C = 26.6°$; with the z-axis; $\tan \angle D = 2$; $\angle D = 63.4°$.

The angle the trace $x + 2y = 8$ makes in the xy-plane with the x-axis is $\tan \angle E = 1/2$; $\angle E = 26.6°$; with the y-axis, $\tan \angle F = 2$; $\angle F = 63.4°$.

For the plane $2x + y + 4z = 12$;

The angle trace $2x + 4z = 12$ makes in the xz-plane with the x-axis is $\tan \angle A = 1/2$; $\angle A = 26.6°$; with the z-axis; $\tan \angle B = 2$; $\angle B = 63.4°$.

The angle the trace $y + 4z = 12$ makes in the yz-plane with the y-axis is $\tan \angle C = 1/4$; $\angle C = 14°$; with the z-axis; $\tan \angle D = 4$; $\angle D = 76°$.

The angle the trace $2x + y = 12$ makes in the xy-plane with the x-axis is $\tan \angle E = 2$; $\angle E = 63.4°$; with the y-axis, $\tan \angle F = 1/2$; $\angle F = 26.6°$.

c. The planes are not parallel because the traces are not parallel in any of the coordinate planes. The angles made by the traces with each plane to the x-axis are not congruent. This is enough to indicate that the planes are not parallel.

STUDENT PAGE 53

14. It seems reasonable to make the following conjecture: If the corresponding angles made by the traces of two different planes and the axes are congruent, the planes are parallel. Construct a proof of the conjecture.

15. Return to the planes containing Aaron and Mays. Sketch the three traces for each of the equations on the same axes.

a. How are these two planes related?

b. Describe the angles made by the traces of the planes for Mays and for Aaron with the *BA*, *RBI* coordinate plane.

c. Is the plane containing Mays above or below the plane for Aaron? How can you tell?

16. Suppose one player using the same weights had a rating of 10.

a. Write the equation of the plane.

b. Is the plane parallel to the plane for Aaron and Mays? How can you tell?

c. Is the plane for Aaron above or below this plane? How do you know?

17. Now consider a plane such that the traces form angles to each axis determined by the weights $[w_1 \;\; w_2 \;\; w_3]$. Begin sweeping the plane from afar toward the origin until you touch one of the points in the space.

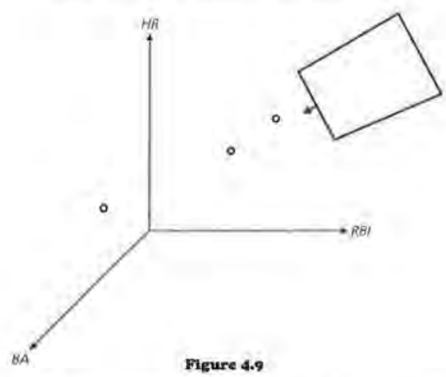


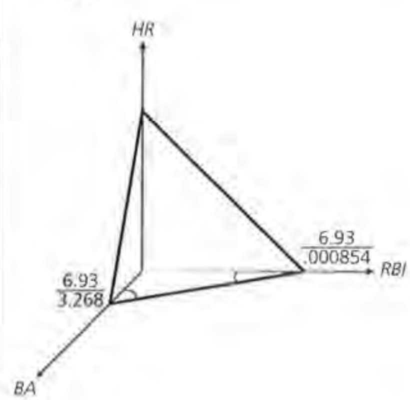
Figure 4.9

a. How will the sweeping plane determined by those weights determine the best player with respect to the three variables?

14. If the corresponding angles made by the traces are congruent, then the traces in each coordinate plane are parallel. The traces intersect in the plane, so the planes are parallel.

15. a. The planes for Mays and Aaron are parallel.

b.



tan −1((6.93/0.000854)/(6.93/3.268)) = 89.98°. The two angles are complementary because they are in a right triangle. Therefore, the two angles are 0.02° and 89.98°.

(15) c. The trace for Mays's plane is behind the trace for Aaron's plane in the first octant view because the constant term in Aaron's equation is larger than the constant term in Mays's equation. The intercepts in each case are R/W_i for any of the three weights, so the larger the R or constant term, the larger the intercept.

16. a. $3.268BA + 0.000854RBI + 0.00526HR = 10.$

b. Yes, the plane is parallel to Mays's and Aaron's planes. Since the coefficients are the same while the ratings are different, and the angles the trace lines make with the axes are congruent, the planes are parallel.

c. Aaron's plane will be below, closer to the origin of the coordinate system because his rating is 6.93, a number less than the 10 for the other player's equation.

17. a. As the plane sweeps closer and closer to the origin, the first point it touches will represent the player with the highest rank with respect to the three weights defining the sweeping plane.

STUDENT PAGE 54

b. Using the agreed-on weights from problem 1, who is the best player among the eleven listed at the beginning of the lesson?

c. How do the other players rank? Explain what this means in terms of the sweeping plane.

18. Suppose you felt that batting average should be three times as important as the other variables.

a. Write an equation of the possible planes for those weights.

b. What angle will the traces of those planes in the *BA*-*HR*-coordinate plane make with the *BA*-axis?

c. Find the equation of the plane that would contain Henry Aaron and sketch its traces.

d. How will the plane that contains Henry Aaron for these weights compare to the planes that used the original weights?

19. Using the new weights, find the ratings of each player.

a. Determine the ranking based on these ratings.

b. Is it possible to find a set of weights that will enable Willie Mays to be in first place? Explain.

Four or More Variables

A more complete set of data for the Hall of Fame players consists of the following.

Baseball Hall of Fame

Player	Games	At Bats	Hits	Batting Average (*BA*)	Runs Batted In (*RBI*)	Home Runs
Hank Aaron	3,298	12,364	3,771	.305	2,297	755
Rod Carew	2,469	9,315	3,053	.328	1,015	92
Ty Cobb	3,034	11,429	4,191	.367	1,961	118
Lou Gehrig	2,164	8,001	2,721	.340	1,990	493
Rogers Hornsby	2,259	8,173	2,930	.358	1,584	301
Reggie Jackson	2,820	9,864	2,584	.262	1,702	563
Mickey Mantle	2,401	8,102	2,415	.298	1,509	536
Willie Mays	2,992	10,881	3,283	.302	1,903	660
Stan Musial	3,026	10,972	3,630	.331	1,951	475
Babe Ruth	2,503	8,399	2,873	.342	2,211	714
Ted Williams	2,292	7,706	2,654	.344	1,839	521

Source: *Universal Almanac* 1994

b.

Player	Rating	Rank
Aaron	6.93	1
Carew	2.42	11
Cobb	3.49	10
Gehrig	5.40	5
Hornsby	4.11	9
Jackson	5.27	6
Mantle	5.08	8
Mays	6.08	3
Musial	5.25	7
Ruth	6.76	2
Williams	5.44	4

Aaron is best according to the ratings using these weights.

c. The plane will touch the point representing Aaron first, then Ruth, Mays, Williams, Gehrig, Jackson, Musial, Mantle, Hornsby, Cobb, and Carew. Musial and Jackson are almost on the same plane because their ratings are very close.

18. a. $9.804BA + 0.000854RBI + 0.00526HR = R.$

b. $\tan^{-1}\left(\frac{9.804}{0.00526}\right) = 89.97°.$

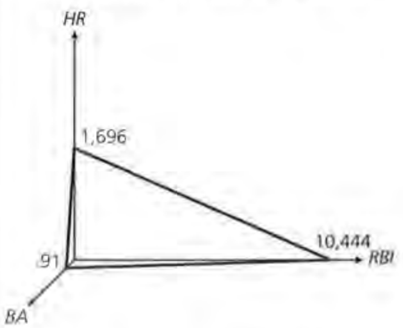
c. 9.804(.305) + 0.000854(2297) + 0.00526(755) = 8.92.

The equation of the plane for Aaron would be $9.804BA + 0.000854RBI + 0.00526HR = 8.92$.

BA, *RBI* trace $9.804BA + 0.000854RBI = 8.92$ with intercepts (0.91, 0, 0) and (0, 10445, 0);

BA, *HR* trace $9.804BA + 0.00526HR = 8.92$ with intercepts (0.91, 0, 0) and (0, 0, 1696);

RBI, *HR* trace $0.000854RBI + 0.00526HR = 8.92$ with intercepts (0, 10445, 0) and (0, 0, 1696).



d. The planes are not parallel. Student explanations will vary. Some may show that the angles made by the trace of each with the *BA* axis are not congruent. Others may find a point the two planes have in common.

19. a.

Aaron	8.92	2
Carew	4.57	11
Cobb	5.89	10
Gehrig	7.63	5
Hornsby	6.45	9
Jackson	6.98	8
Mantle	7.03	7
Mays	8.06	3
Musial	7.41	6
Ruth	8.99	1
Williams	7.68	4

STUDENT PAGE 55

20. The number of hits divided by the number of at bats will give the batting average. The only variable that is not really already involved in the formula is the number of games. Again, you want to have a standard for the number of games. Because all of the other data are based on means, it makes sense to continue to do so. The mean number of games for all the players in the Hall of Fame is 2030. Adding in this variable in the same way that the third variable was included would produce an equation for finding the total ranking of the form

$$3.268\,BA + 0.000854\,RBI + 0.00526\,HR + 0.000493G = R.$$

a. What total ranking would this produce for Aaron?

b. Use the new equation to find a ranking for each of the players.

c. How does the ranking with four variables compare to that with three?

d. Carl Yastrzemski, elected to the Baseball Hall of Fame in 1989, had the statistics shown below. How does Yastrzemski compare to the other players you have been ranking?

Carl Yastrzemski Stats.

Player	Games	At Bats	Hits	Batting Average (*BA*)	Runs Batted In (*RBI*)	Home Runs
Yastrzemski	3308	11,988	3,419	.285	1,844	452

21. What procedure could you use if you had five variables?

The geometry for four variables is impossible to sketch in a two-dimensional plane. The weights form a vector in four dimensions, and the point for each player is projected onto a scalar multiple of this vector. The player whose projection creates the longest vector is rated the "best" according to the given weights. An extension exercise in the Practice and Applications demonstrates how this works in two dimensions.

Summary

In this lesson you learned that an equation with three variables was the equation of a plane. The line in which a given plane intersects a coordinate plane is called a *trace of the plane*. If you considered equations of the form $w_1x + w_2y + w_3z = R$ for different values for *R*, you would produce a series of planes, parallel to each other. The intercepts of a plane are the points

b. Willie Mays is dominated by Henry Aaron in each category. Thus, it will never be possible to find a set of weights that would put him above Aaron; the weights would affect Aaron's rating the same way they would affect Mays's rating.

20. Some students may feel that bringing in the number of games played might be more effective if the number is used as a divisor for the other categories, creating ratios.

You might have students investigate their ideas as an extension. Be sure they can find a mathematical justification for their conclusions and that they compare what they are doing with the process followed in this problem.

a. 3.268(0.305) + 0.000854(2297) + 0.00526(755) + 0.000493(3298) = 8.56

(20) b.

Player	Rating	Rank
Aaron	8.56	1
Carew	3.64	11
Cobb	4.99	10
Gehrig	6.47	7
Hornsby	5.22	9
Jackson	6.66	5
Mantle	6.27	8
Mays	7.56	3
Musial	6.74	4
Ruth	8.00	2
Williams	6.57	6

c. There is not really much difference in the rankings. Aaron is first , Ruth is second, and Mays is third in both sets of rankings. Fourth, fifth, sixth, and seventh places are different. The rest are the same order.

d. Using the four variables, Yastrzemski has a rating of 6.51, which would place him in seventh position, replacing Lou Gehrig. Using only three variables from the earlier ratings, Yastrzemski has a rating of 4.88, which would place him ninth replacing Roger Hornsby.

21. You could find a mean for that variable, then use the reciprocal of that mean to determine another weight to be used in the formula.

Practice and Applications

22. Student explanations may vary.

a. The planes are parallel; the coefficients have the same ratio, which implies that the angles the traces make with the axes will be congruent.

b. The planes are the same; they coincide. The coefficients are multiples, so intercepts will be the same and the angles will be congruent.

STUDENT PAGE 56

$(\frac{r}{w_1}, 0, 0)$, $(0, \frac{r}{w_2}, 0)$, $(0, 0, \frac{r}{w_3})$, and the tangent of the angle a trace makes with an axis can be expressed in terms of the intercepts of the plane:

$$\tan A = \frac{\frac{R}{w_2}}{\frac{R}{w_1}} = \frac{w_1}{w_2}.$$

Two planes are parallel if the angles they make with the axes are congruent. You can use weights determined by some standard to establish a formula to rank something using as many variables as you choose, although you can only see the geometric interpretation in two and three dimensions.

Practice and Applications

22. Are the following planes parallel? Why or why not?

a. $x + y + z = 10$ and $x + y + z = 5$

b. $x + 2y + 4z = 8$ and $3x + 6y + 12z = 24$

c. $6x + 9y + 15z = 8$ and $4x + 6y + 10z = 10$

d. $x + 2z = 8$ and $3x - z = 8$

23. In *Places Rated Almanac*, the data for rating cities is collected for a variety of categories, then put into a formula to determine a score. The scores are then ranked in order from best to worst and the ranks are used to determine the composite rating. The table below contains the scores that cities earned in the Arts, Education, and Recreation. In each case, the higher the number the better the score.

Rating Cities: Arts, Education, Recreation

City	Arts	Education	Recreation
Ann Arbor, MI	944	1461	1480
Austin, TX	560	1730	1611
Jacksonville, FL	512	440	2642
Minneapolis, MN	1660	2212	2273
Portland, OR	597	944	2851
Richmond, VA	690	269	1130
San Jose, CA	1685	1854	1430
Santa Fe, NM	501	60	2294

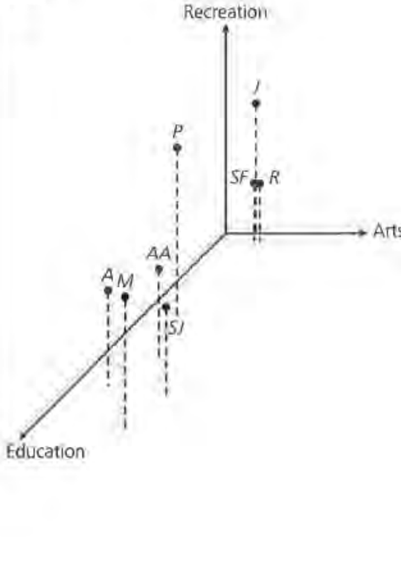
Source: *Places Rated Almanac*, 1993

a. Make a three-dimensional sketch of the scores. Does any city dominate the others? Explain.

c. The planes are parallel; the coefficients have the same ratio; trace angles will be congruent.

d. The planes are not parallel; they intersect in the *xz*-coordinate plane.

23.



a. No city dominates, since no city has the highest score in all categories.

b. Use the reciprocal of the medians for each category to determine a weight for a rating equation.

c. $R = 0.00333$ (Arts) + 0.00323 (Education) + 0.000685 (Recreation)

d. Minneapolis will have a rating of 14.23. The plane will have intercepts (4273, 0, 0), (0, 4406, 0), and (0, 0, 20774). Minneapolis dominates only in education but is second in the other two categories. You would have to see whether the points representing other cities are on the same plane or would satisfy the equation.

e.

	Rating	Rank
Ann Arbor	8.88	3
Austin	8.56	4
Jacksonville	4.94	6
Minneapolis	14.23	1
Portland	6.99	5
Richmond	3.94	7
San Jose	12.58	2
Santa Fe	3.43	8

24. a. Student responses will vary. Art, education, and recreation are more concerned with the quality of life, while cost of living, jobs, and housing with the practical matter of living. So some students would think the ratings might be very different.

b. Use the reciprocal of the median to determine weights.

STUDENT PAGE 57

b. Describe a method to make the scores equivalent. Following is some information you might find useful, based on all cities in the *Places Rated Almanac.*

For Arts, the scores ranged from 9681 to 46 with a median score of 300.

For Education, the scores ranged from 6728 to 0 with a median score of 309. For Recreation, the scores ranged from 3940 to 200 with a median score of 1460. You may choose to use some other measure if it seems reasonable.

c. Create a set of weights using the standards you chose in part b and write a formula to calculate a total rating based on those weights.

d. Find the total rating for Minneapolis using the formula and describe the plane this would generate. Will any of the other cities have the same rating as Minneapolis? How can you tell?

e. Find the total ratings for the cities using your weights and rank the eight cities accordingly. Would any of these cities be on the same plane?

24. The scores given in *Places Rated Almanac* for the eight cities on Cost of Living, Jobs, and Housing are in the table below. In the case of Cost of Living and Housing, a low score is better than a high score.

Rating Cities: Cost of Living, Jobs, Housing

City	Cost of Living	Jobs	Housing
Ann Arbor, MI	12,024	3,080	9,404
Austin, TX	10,158	5,448	8,277
Jacksonville, FL	9,598	3,275	7,535
Minneapolis, MN	11,099	6,242	9,316
Portland, OR	11,142	6,321	8,904
Richmond, VA	10,475	3,872	8,612
San Jose, CA	20,493	5,845	29,395
Santa Fe, NM	12,374	2,518	12,987

Source: *Places Rated Almanac,* 1993

a. Do you think these scores will yield a very different ranking from the one you found using the Arts, Education, and Recreation? Why or why not?

b. Describe a method to make the scores equivalent. (Remember that a low score is better for Housing and

(24) c. Because housing and cost of living are both better with a low score, some adjustment has to be made in finding weights. The example uses a negative for those two weights.

-0.000105 (cost of living) + 0.000411 (jobs) + -0.000144 (housing) = R

d. The rating for Minneapolis would be 0.059 using the equation from 24c. This would create the plane $-0.000105C + 0.000411J + -0.000144H = 0.059$.

e.

	Rating	Rank for *C, J, H*	Rank for *A, E, R*
Ann Arbor	−1.35	6	3
Austin	−.0194	3	4
Jacksonville	−.747	4	6
Minneapolis	.059	2	1
Portland	.146	1	5
Richmond	−.749	5	7
San Jose	−3.98	8	2
Santa Fe	−2.13	7	8

Minneapolis went from first to second. Portland moved from fifth to first, while San Jose went from second to eighth (housing was very costly in San Jose). Ann Arbor went from third to sixth. Jacksonville and Richmond are very close to being on the same plane.

25. a–b. Some students may wish to sum all of the ranks to find an overall rating as they did in Lesson 1. You might discuss what is lost in this method. (The great difference in San Jose's housing score and cost of living score is not apparent, nor is Santa Fe's low rank in education. These scores are last in the rank, but the ranks cover up the difference; they do not show the variability.)

STUDENT PAGE 58

Cost of Living.) Following is some information you might find useful.

For the Cost of Living, the scores ranged from 7,325 to 21,932 with a median of 9,514. For Jobs, the scores ranged from 23,028 to 1,623 with a median score of 2,434. For Housing, the scores ranged from 4,102 to 32,211 with a median of 6,944. You may choose to use some other measure if it seems reasonable.

c. Create a set of weights using the standards you choose in part b and write a formula to calculate a total rating based on those weights.

d. Find the total rating for Minneapolis using the formula and describe the plane this would generate.

e. Find the total ratings for the cities using your weights and rank the eight cities accordingly. How did the rankings compare with those using the Arts, Education, and Recreation?

25. Rank each of the cities in problems 23 and 24 from one to eight for each of the six categories, where one is the best rank.

a. Find the total rating for each city for Arts, Education, and Recreation by summing the individual ranks. Use the results to rank the cities for the three categories. How did the results compare to those you found in problem 23?

b. Find the total rating for each city for Cost of Living, Jobs, and Housing by summing the individual ranks. Use the results to rank the cities for the three categories. How did the results compare to those you found in problem 24?

26. Suppose the angle the trace of a plane made in the xz-coordinate plane with the x-axis was 78.69°, and the angle in the yz-coordinate plane with the z-axis was 26.56°.

a. Make a sketch of a plane that satisfies the two conditions.

b. What is an equation for the plane?

c. What is the angle made by the trace of the plane with the y-axis in the xy-coordinate plane?

27. *Money Guide* ranked the "100 best college buys," those schools they feel have the highest quality education for the tuition and fees. Their rankings are based on what the edi-

City	Rank in Arts	Rank in Ed	Rank in Rec	Rank in Cost of Living	Rank in Jobs	Rank in Housing	Sum of Ranks *A, E, R*	Sum of Ranks *C, J, H*
Ann Arbor	3	4	6	6	7	6	13	19
Austin	6	3	5	2	4	2	14	8
Jacksonville	7	6	2	1	6	1	15	8
Minneapolis	2	1	4	4	2	5	7	11
Portland	5	5	1	5	1	4	11	10
Richmond	4	7	8	3	5	3	19	11
San Jose	1	2	7	8	3	8	10	19
Santa Fe	8	8	3	7	8	7	19	22

STUDENT PAGE 59

(25. a–b. cont.)

tors consider will provide an excellent education at a much lower price than at schools of similar quality. The table below has some of the factors listed in the *Guide* for the top ten universities according to their ranking. The student academic level is based on class rank, test scores, and high school grades, where 1 represents those with students with the highest academic records.

Best College Buys

	Tuition and Fees	Room and Board	% Students Receiving Aid	% of Need Met	Student-Faculty Ratio	% Who Graduate in 6 yrs	Student Academic Level
California Institute of Technology	17,586	6,620	75	100	3:1	78	1
New College of University of South Florida	7,950	3,847	70	93	10:1	60	1
Northeast Missouri State	3,975	3,330	N.A.	80	22:1	49	4
Rice University	12,034	5,900	84	100	9:1	88	1
State University of New York at Binghamton	8,679	4,654	49	62	19:1	80	1
State University of New York at Albany	8,856	4,836	90	80	18:1	73	2
Spelman College	8,875	5,890	76	38	15:1	74	2
Trenton State College	6,658	5,650	48	90	15:1	78	2
University of Illinois at Urbana/Champaign	9,130	4,408	81	75	13:1	79	2
University of North Carolina-Chapel Hill	10,162	5,350	44	95	10:1	85	1

Source: *Money Guide*, 1996 Edition

In 1994–95, the average tuition at private universities was \$10,333 and at public universities, \$2,730. On average, 69% of the students in the schools surveyed received financial aid for about half of the costs. On the average, 55% of the students in those schools graduated in six years. The typical student-faculty ratio was 34 to 1. Ranking for student academic levels went from 1 to 5, with a 1 as best. According to the U.S. Census, the average cost of room and board at a private college was \$4,793 and at a public college \$3,680 in 1994.

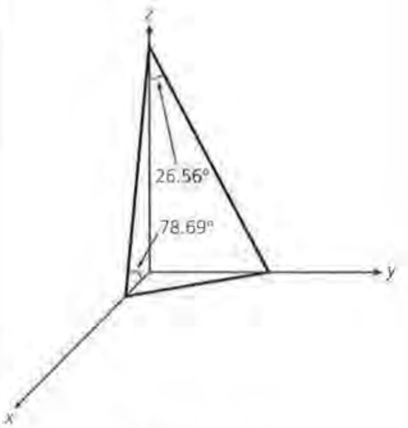
a. Using the information above, rate and then rank the ten universities in terms of percentage of students receiving aid, percentage of need met by aid, and the percentage

	Rank *A, E, R* (Prob. 23)	Ranks by Summing Ranks *A, E, R*	Rank *C, J, H* (Prob. 24)	Rank by Summing Ranks *C, J, H*
Ann Arbor	3	4	6	6
Austin	4	5	3	1
Jacksonville	6	6	4	1
Minneapolis	1	1	2	4
Portland	5	3	1	3
Richmond	7	7	5	4
San Jose	2	2	8	6
Santa Fe	8	7	7	8

The ranks obtained by summing the ranks for arts, education, and recreation are close to the ranks found by using the formula.

The ranks found by summing the ranks for cost of living, jobs, and housing were much different. Portland went from first to third, and Jacksonville from fourth to first.

26. a.



b. tan 26.56 = .4999, approximately .5, so $\frac{w_3}{w_2} = \frac{1}{2}$.

tan 78.69 = 4.9999, approximately 5, so $\frac{w_1}{w_3} = \frac{5}{1}$.

$2y + z = C$ and $5x + z = C$. Each of these two equations forms a plane, and the two planes intersect in a line that contains an infinite number of (x, y, z) that will satisfy the conditions. A particular solution might be when $z = 10$, $y = 5$ and $x = 2$. This indicates the intercepts could be (2, 0, 0), (0, 5, 0), and (0, 0, 10). An equation of a plane with those intercepts would be $5x + 2y + z = 10$. Any equation of the form $5x + 2y + z = R$ would also have the same angles.

c. $\tan^{-1}\left(\frac{w_2}{w_1}\right) = \tan^{-1}\left(\frac{2}{5}\right) = 21.8$

The angle made by the trace in the xy-coordinate plane will be 21.8°.

27. a. An equation would be $(1/69)FA + (1/50)NM + (1/55)G = R$ where FA is financial assistance, NM is financial needs meet, and G is the percent graduating in six years. Using these variables, Rice places first, followed by California Institute of Technology, and the State University of New York at Albany.

STUDENT PAGE 60

of students who graduate in 6 years. (You might use the typical 69% as the replacement for the NA category for Northeast Missouri State.)

b. Find a formula to rate and then rank the ten universities using all seven categories given. (You might use the average of the cost and tuition for public and private schools.)

c. Compare the rankings from parts a and b.

d. In their rating scheme, *Money Guide* actually used 16 variables, not all of which are given here. Using these variables, the universities ranked as follows.

Ranking

California Institute of Technology	5
New College of University of South Florida	1
Northeast Missouri State	3
Rice University	2
State University of New York at Binghamton	7
State University of New York at Albany	10
Spelman College	8
Trenton State College	4
University of Illinois at Urbana/Champagne	9
University of North Carolina at Chapel Hill	6

Compare your rankings to those from *Money Guide*.

Extension

28. Consider a pair of weights $[w_1 \ \ w_2]$ for two variables. These weights will determine a vector in a two-dimensional plane.

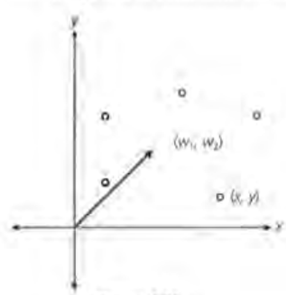


Figure 4.10

Ratings with Three Variables

	Rating	Ranking
California Institute of Technology	4.51	2
New College of University of South Florida	3.97	6
Northeast Missouri State	3.49	8
Rice University	4.82	1
State University of New York at Binghamton	3.40	9
State University of New York at Albany	4.23	3
Spelman College	3.21	10
Trenton State College	3.91	7
University of Illinois at Urbana/Champaign	4.11	4
University of North Carolina at Chapel Hill	4.08	5

b. Using the formula

$$\frac{-1}{6531.5}T + \frac{-1}{4236.5}RB + \frac{1}{69}FA + \frac{1}{50}NM + \frac{1}{55}G + \frac{-1}{34}\left(\frac{S}{F}\right) + \frac{-1}{3}SA = R,$$

you obtain the ratings in the table. The four categories where a low value is good were assigned a negative weight. Tuition and cost weights were obtained by taking the averages of the public and private schools. Students may suggest another way to find a weight for those two or for the Student Academic level. (The weight used here was the median value.)

Ratings with all Seven Variables

	Rating	Ranking
California Institute of Technology	–.172	9
New College of University of South Florida	1.213	1
Northeast Missouri State	.116	7
Rice University	.984	2
State University of New York at Binghamton	.085	8
State University of New York at Albany	.538	5
Spelman College	–.650	10
Trenton State College	.453	6
University of Illinois at Urbana/Champaign	.623	4
University of North Carolina at Chapel Hill	.637	3

c. The rankings from parts a and b are shown below.

	Ranking with Three Variables	Ranking with Seven Variables
California Institute of Technology	2	9
New College of University of South Florida	6	1
Northeast Missouri State	8	7
Rice University	1	2
State University of New York at Binghamton	9	8
State University of New York at Albany	3	5
Spelman College	10	10
Trenton State College	7	6
University of Illinois at Urbana/Champaign	4	4
University of North Carolina at Chapel Hill	5	3

d.

	Ranking with Three Variables	Ranking with Seven Variables	Ranking by Money Guide
California Institute of Technology	2	9	5
New College of University of South Florida	6	1	1
Northeast Missouri State	8	7	3
Rice University	1	2	2
State University of New York at Binghamton	9	8	7
State University of New York at Albany	3	5	10
Spelman College	10	10	8
Trenton State College	7	6	4
University of Illinois at Urbana/Champaign	4	4	9
University of North Carolina at Chapel Hill	5	3	6

The other variables used by *Money Guide* made a great difference in the ratings or the weights they used were very different from those used in 27a and b. However, Rice University was consistently in first or second place.

STUDENT PAGE 61

28. a. The equation of the line will be $w_1x + w_2y = R$. The slope of the line will be $-w_1/w_2$. The ray will pass through the origin (0, 0) and contain all points of the form (aw_1, aw_2). The slope of the ray will be w_2/w_1. Thus, the slopes of the two are negative reciprocals, and the ray and line are perpendicular.

b. The equation of the line that contains OV is $y = (w_2/w_1)x$ or $w_2x - w_1y = 0$. The distance from D to OV, $DV = \frac{|w_2x - w_1y|}{\sqrt{w_2^2 + w_1^2}}$

c. $OD = \sqrt{x^2 + y^2}$

Using those weights, each data point determines a rating and the equation for a line that will contain that data point.

a. Each of these lines will be perpendicular to the ray formed by taking all of the positive scalar multiples of $[w_1 \ \ w_2]$. Why is this so? (Recall how the slope of the line is related to the weights.)

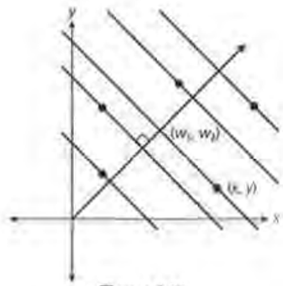


Figure 4.11

This means that each of the data points can be projected orthogonally (at right angles with the line) onto the vector $[aw_1 \ aw_2]$ for some a.

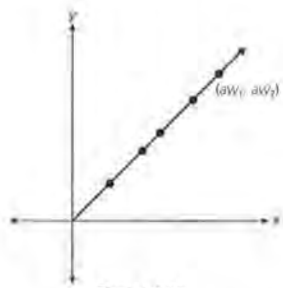


Figure 4.12

Another way to determine the data point that is the "best" according to a given set of weights is to use this projection onto a multiple of the weight vector. The data point for which the projection produces the vector of greatest magnitude will be the

STUDENT PAGE 62

(28)d. $OD^2 = OV^2 + DV^2$

$$OV^2 = OD^2 - DV^2$$

$$= x^2 + y^2 - \left(\frac{|w_2x - w_1y|}{\sqrt{w_2^2 + w_1^2}}\right)^2$$

$$= x^2 + y^2 - \frac{w_2xw_2x - 2w_1yw_2x + w_1yw_1y}{w_2^2 + w_1^2}$$

$$= \frac{x^2w_2^2 + x^2w_1^2 + y^2w_2^2 + y^2w_1^2 - w_2xw_2x + 2w_1yw_2x - w_1yw_1y}{w_2^2 + w_1^2}$$

$$= \frac{x^2w_1^2 + y^2w_2^2 + 2w_1yw_2x}{w_2^2 + w_1^2}$$

$$= \frac{(xw_1 + yw_2)^2}{w_2^2 + w_1^2}$$

$$OV = \sqrt{\frac{(xw_1 + yw_2)^2}{w_2^2 + w_1^2}} = \frac{xw_1 + yw_2}{\sqrt{w_2^2 + w_1^2}}$$

But, $R = w_1x + w_2y$ from the original weighted equation, so

$$OV = \frac{R}{\sqrt{w_2^2 + w_1^2}}.$$

e. The magnitude of any of these vectors will be $\frac{R}{\sqrt{w_2^2 + w_1^2}}$ for given weights w_1 and w_2. Any data point D will project onto the ray at V and form the vector $[aw_1\ aw_2]$ with that magnitude. For the largest R, OV will be the largest, and so for the greatest R you will have the vector of the greatest magnitude. This vector will be the *best* according to those weights.

data point that has the highest weighted rating and is the "best" according to those weights. To prove that statement, consider the following:

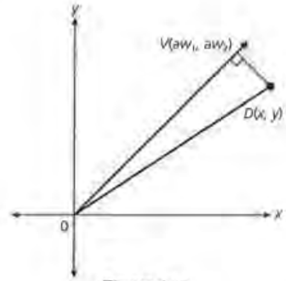


Figure 4.13

b. What is the equation of the line that contains OV? Use the equation of this line to find the distance from D to OV. (The distance from a point (x_1, y_1) to a line $Ax + By + C = 0$ is given by $\frac{|Ax_1 + By_1 + C|}{\sqrt{A^2 + B^2}}$.

c. Find length of OD.

d. Remember that the equation of the line containing DV is $w_1x + w_2y = R$.

Prove that the length of $OV = \frac{R}{\sqrt{w_1^2 + w_2^2}}$.

e. Show how this supports the statement made above: The data point for which the projection produces the vector of greatest magnitude will be the data point that has the highest weighted rating and is the "best" according to those weights.

The method above works for any number of variables. In each case, for n variables, there is an n-dimensional weight vector. Each data point is projected orthogonally onto that vector, and the point that has the highest rating will be the point for which $\frac{R}{\sqrt{w_1^2 + w_2^2 + w_3^2 + \ldots + w_n^2}}$ is the largest.

29. Demonstrate that the statement above works using the games played, home runs, batting average and runs batted in data.

29.

Player	Rating	Rank	$\frac{R}{\sqrt{w_1^2 + \ldots w_4^2}}$	Rank
Aaron	8.56	1	2.62	1
Carew	3.64	11	1.11	11
Cobb	4.99	10	1.53	10
Gehrig	6.47	7	1.98	7
Hornsby	5.22	9	1.60	9
Jackson	6.66	5	2.04	5
Mantle	6.27	8	1.92	8
Mays	7.56	3	2.31	3
Musial	6.74	4	2.06	4
Ruth	8.00	2	2.45	2
Williams	6.57	6	2.01	6

The ranking found by using the formula projecting each data point onto the ray determined by the vector (3.268, 0.000854, 0.00526, 0.000493) is the same as the ranking found by using the weighted formula $3.268BA + 0.000854RBI + 0.00526HR + 0.000493G = R$.

ASSESSMENT

From Best Companies for Working Women to Cars

Materials: none
Technology: graphing calculator
Pacing: 1 class period

Overview

The assessment begins by asking students to use what they have learned about weights and sweeping lines as they create ratings and ranks based on several variables. In the second part of the assessment, students must deal with a situation in which high numbers are good for one variable but bad for another variable.

Teaching Notes

This assessment might be best done by students working in pairs or in sets of three, so they can share their analysis and do some brainstorming. The questions asked are open-ended, but the data provide the opportunity for students to apply all of the rating and ranking techniques they have learned in Unit I if students so choose.

The two assessment problems raise different issues. In the first, the companies were selected because all of them employed women in the fields of engineering or accounting. If students are interested in other career options, they can find the article and select other companies to analyze. The first four columns of Table 1 have data that are similar to the data students used in Lesson 1. The other data are in different units and can be used in pairs or as ordered triples to produce rankings using the concept of the sweeping line or plane. Some students might even consider combining the data from the first four columns with the data from the last three to find some kind of composite rating. You might want to tell students that the median salary earned by an engineer according to *Jobs Rated* is \$64,000. Typical benchmarks for the other two lists might be 50%; you might expect half of the employees to be female and also half of the highest paid executives to be female if the company had total equality. Students have the opportunity here to show you what they know and how they can apply what they have learned without constraints imposed by your questions. Encourage them to demonstrate in their responses what kind of mathematics they have learned and how they can use mathematics (and matrices) as a tool to solve the problem.

Students may be interested in knowing something about the companies. The Calvert Group manages mutual funds; Corning, Incorporated manufactures glass products and runs clinical testing labs. The Dow Chemical Company and the Dupont Company are the nation's second largest and the largest chemical manufacturers, respectively, in the United States. Fannie Mae provides funds for home mortgages. Fel-Pro Incorporated makes gaskets, sealants, and lubricants for cars. General Motors manufactures cars and trucks, and Johnson & Johnson manufactures health care products. Mattel, Inc. is the world's largest toy maker. Merek & Co., Inc. makes prescription drugs. Motorola makes cellular phones, radios, and semi-conductors. Quad/Graphics is a printer, and Security Benefit Group of Companies sells mutual funds and life insurance.

If the questions seem too vague for your students, you might choose to focus them more directly by asking the following:

a. Use the rating given by *Woman's World* for money, opportunities to advance, child care support, and family-friendly benefits to find total ratings for each company, then rank the companies.

b. Use the salary and percent of professionals who are women to establish a rating for each company. Explain how a sweeping line can help you determine the rankings.

c. What would the equation of the plane that contained the ratings for Johnson & Johnson, Merek, and Fannie Mae based on the last three variables look like? Does any other company have a point on that same plane? How do you know?

d. Would any of the planes that contain the points for the remaining companies be parallel to the plane that contains Johnson & Johnson, Fannie Mae, and Merek? How do you know?

e. Based on all or some of the data given, find a way to rank all of the companies with respect to working mothers. Explain your method carefully and show how it uses the mathematics you have learned in this section.

In the second assessment problem, students have to deal with a situation where the "best" is a high value for one variable and a low value for the second. They must decide how to combine these values and decide what the graphical representation will be. Responses will enable you to see how well students can modify what they have learned to fit the situation. Some may choose to use negative values for the low numbers; others may choose to reverse them by finding a complement of some sort. Encourage students to be creative in their solutions.

STUDENT PAGE 63

ASSESSMENT

From Best Companies for Women to Cars

Each year *Woman's World* identifies the best 100 companies for working mothers. The following data came from the October 1995 issue.

OBJECTIVE

Use weights, a sweeping line, and matrices to create ratings and ranks.

Best Companies for Working Mothers

Companies	Pay Rating	Opportunity to Advance	Child Care Support	Family-Friendly Benefits	Median Salary	% of Women Professionals	% of Highest Paid Women
Calvert Group	4	4	4	4	39,000	53	43
Corning Incorporated	4	2	5	3	30,200	28	12
The Dow Chemical Company	4	2	3	4	44,400	25	7
Dupont Company	5	2	4	4	42,840	21	14
Fannie Mae	4	3	2	4	34,950	49	43
Fel-Pro Incorporated	4	2	5	5	37,999	40	19
General Motors	4	2	3	4	39,570	25	10
Johnson & Johnson	5	3	5	4	31,000	45	26
Mattel, Inc.	4	3	3	4	36,376	52	29
Merck & Co., Inc.	5	3	5	4	37,000	47	24
Motorola	3	2	4	3	38,000	23	6
Quad/Graphics, Inc.	3	2	5	3	30,000	50	9
Security Benefit Group of Companies	3	2	4	4	25,194	59	30

Source: *Woman's World*, October 1995

Woman's World staff rated each company in the first four categories from 1 to 5 based on certain criteria, where 5 is the best rating. A typical salary, the percent of women professionals employed in the company, and the percent of the highest paid employees who were women constituted part of the information given about each company and used to create the rankings.

Solution Key

1. a. Student responses will vary. Some may choose to use equal weights; others may think that pay and opportunity to advance are very important and use weights such as the following:

$3P + 3A + C + .5F = R,$

where P represents pay rating, A opportunity to advance, C child care support, and F family benefits. Students will find that with equal weights, there are many ties in the total ratings (five in eighth place) because the original rates are only from 1 to 5 and most are 4s and 5s. Their explanations should include setting up the rating matrix and the weight matrix and showing how the product of the two will produce the total ratings.

Best Companies for Working Mothers

Company	Rating Equal Weights	Rank	Rating from Sample Equation	Rank
Calvert Group	16	3	30	3
Corning Incorporated	14	6	24.5	8
The Dow Chemical Company	13	8	23	9
Dupont Company	15	5	27	4
Fannie Mae	13	8	25	7
Fel-Pro Incorporated	16	3	25.5	6
General Motors	13	8	23	9
Johnson & Johnson	17	1	31	1
Mattel, Inc.	14	6	26	5
Merck & Co., Inc.	17	1	31	1
Motorola	12	13	20.5	13
Quad/Graphics, Inc.	13	8	21.5	11
Security Benefit Group of Companies	13	8	21	12

Source: *Woman's World,* October 1995

STUDENT PAGE 64

The companies in the table were those whose salary was given for engineers or accountants.

1. Use the set of data given to answer the following:

a. How would you find the total ratings and ranks for the companies based on the first four numerical ratings, and how could you use matrices to do so?

b. How would you use two variables that are in different units to obtain a ranking of the companies? Explain in this context how to use a sweeping line.

c. How would you use three variables that are in different units to obtain a ranking of the companies? Explain in this context how to use a **sweeping plane.**

d. Based on your work for a to c and anything else that seems appropriate, how would you rank these thirteen companies in terms of the best companies for working mothers? Explain your reasoning.

2. The data in the table below are customer satisfaction ratings. The data are based on the number of problems per 100 cars during the first three months of ownership according to JD Power and Associates Initial Quality Survey and the amount of damage caused by four crashes (front and rear into flat barrier, rear into a pole, front into barrier at angle) at a speed of 5 mph. Note that high customer satisfaction is good as is low damage cost.

Car Problems

Car	Customer Satisfaction Rating	Total Cost of Damage ($)
Nissan Maxima	138	3,605
Honda Accord	149	1,433
Saab 900S	132	1,734
Subaru Legacy	155	1,966
Volvo 850	155	2,131
Toyota Camry	148	2,328
Ford Taurus	128	2,814
Lumina	125	2,629

With equal weights for the four categories, Johnson & Johnson and Merck & Co., Inc. tied for first, Calvert Group and Fel-Pro Incorporated tied for third. The following matrix multiplication could be used to obtain the above ratings:

$$\begin{bmatrix} 4&4&4&4 \\ 4&2&5&3 \\ 4&2&3&4 \\ 5&2&4&4 \\ 4&3&2&4 \\ 4&2&5&5 \\ 4&2&3&4 \\ 5&3&5&4 \\ 4&3&3&4 \\ 5&3&5&4 \\ 3&2&4&3 \\ 3&2&5&3 \\ 3&2&4&4 \end{bmatrix} \cdot \begin{bmatrix} 1&3 \\ 1&3 \\ 1&1 \\ 1&.5 \end{bmatrix} = \begin{bmatrix} 16&30 \\ 14&24.5 \\ 13&23 \\ 15&27 \\ 13&25 \\ 16&25.5 \\ 13&23 \\ 17&31 \\ 14&26 \\ 17&31 \\ 12&20.5 \\ 13&21.5 \\ 13&21 \end{bmatrix}$$

b. Student responses will vary. If they choose median salary S and women professionals P in a company, they might use the average of the salaries given, \$35,887, and 50% as a standard for the percent of women professionals in a company. The corresponding ratings equation would be:

$\frac{1}{35,887}S + \frac{1}{50}P = R$ or

$0.000028S + 0.02P = R$

This would produce ratings such as the following:

STUDENT PAGE 65

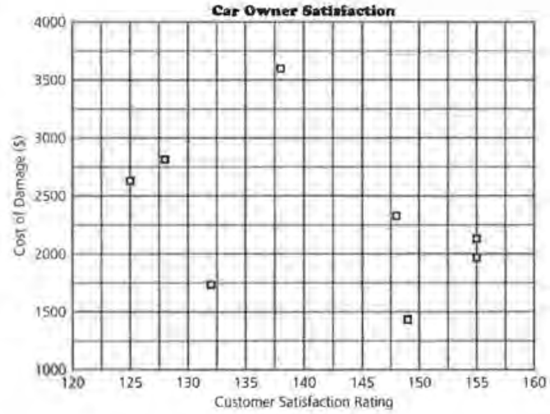


Figure A1.1

a. The average customer satisfaction rating for all cars was 100. Find the average cost for damage in the crashes for all of the cars tested. Use this information to find weights such that both variables contribute about equally to the rating.

b. What rating will the Volvo have using those weights?

c. Graph the equation for the line that shows cars with the same rating as the Volvo. How will the equation containing the rating for the other cars compare to the equation for the Volvo? Sketch at least one other equation on the graph.

d. Can a sweeping line help you identify the "best" car according to those weights for the variables? Explain why or why not.

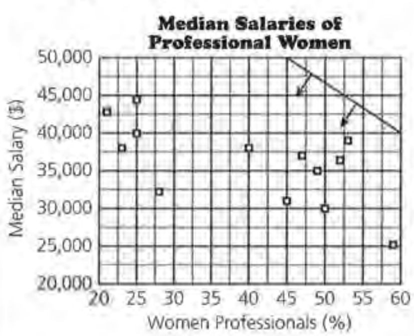
e. Use your equation to find a total rating for each of the cars. Rank each of the cars according to the ratings.

f. How reliable do you think the rating you produced in part e will be? Write a paragraph describing your conclusions.

Best Companies for Working Mothers

Company	Median Salary	% of Women Professionals	Rating	Rank
Calvert Group	39,000	53	2.15	1
Corning Incorporated	30,200	28	1.41	13
The Dow Chemical Company	44,400	25	1.74	9
Dupont Company	42,840	21	1.62	10
Fannie Mae	34,950	49	1.96	4
Fel-Pro Incorporated	37,999	40	1.86	6
General Motors	39,570	25	1.61	11
Johnson & Johnson	31,000	45	1.77	8
Mattel, Inc.	36,376	52	2.06	2
Merck & Co., Inc.	37,000	47	1.98	3
Motorola	38,000	23	1.52	12
Quad/Graphics, Inc.	30,000	50	1.84	7
Security Benefit Group of Companies	25,194	59	1.89	5

Calvert Group is first, Mattel, Inc. is second, and Merck & Co., Inc. is third. Note that three of these also ranked high using the first set of data.



The sweeping line would have a slope of about 718/1 or about \$3600 for every 5%. As the line sweeps down, it will encounter the first point at (53, 39,000), the Calvert Group, which would give them the top rating. They might point out that Fannie Mae and Merck & Co., Inc. are almost on the same line and thus have approximately the same rating. Students may choose to give more weight to one of the variables, which will produce a different equation and line.

c. If the variables selected are the median salary S, percentage of women professionals P, and percentage of highest paid women H, with standards the average salary listed and using 50% for the two percentage variables, the weighted equation would be $0.000028S + 0.02P + 0.02H = R$.

The ratings and ranks are given in the table. Students might comment that by including the percentage of highest paid women, Fannie Mae moved from fourth place to second place. Calvert Group is still first.

Best Companies for Working Mothers

Company	Median Salary	% of Women Professionals	% of Highest Paid Women	Rating	Rank
Calvert Group	39,000	53	43	3.01	1
Corning Incorporated	30,200	28	12	1.65	13
The Dow Chemical Company	44,400	25	7	1.88	10
Dupont Company	42,840	21	14	1.90	9
Fannie Mae	34,950	49	43	2.82	2
Fel-Pro Incorporated	37,999	40	19	2.24	7
General Motors	39,570	25	10	1.81	11
Johnson & Johnson	31,000	45	26	2.29	6
Mattel, Inc.	36,376	52	29	2.64	3
Merck & Co., Inc.	37,000	47	24	2.46	5
Motorola	38,000	23	6	1.64	12
Quad/Graphics, Inc.	30,000	50	9	2.02	8
Security Benefit Group of Companies	25,194	59	30	2.49	4

Source: data from *Woman's World*, October 1995

The sweeping plane will start from the outer right corner of the first quadrant and move in until it touches a point. The first point it will touch will be that for the Calvert Group if the equation used is the one with equal weights.

d. Students may choose to use all of the data or some subset of it. They might decide to weight some of the variables more than others. If they choose to use the ratings, they might select 3 as an average rating when given the option from 1 to 5 or they might use the average of the ratings given. A possible equation using all of the variables could be

$0.333PR + 0.333A + 0.333C + 0.333F + 0.000028S + 0.02P + 0.02H = R$.

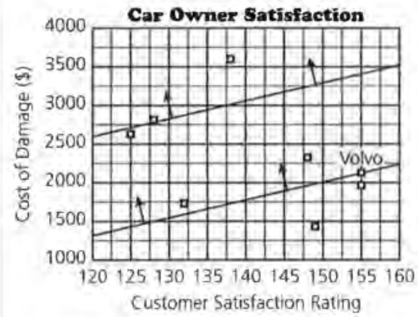
In this case, Calvert Group again has the best rating at 8.34, followed by Merck & Co, Inc. with 8.12.

2. As mentioned in the introduction, a small damage cost (C) is good while a high customer satisfaction rating (S) is good. One way students might account for this is to use a negative weight for a high damage rating. They may suggest other alternatives, however. The answers below are given using a negative weight.

a. The average for damage in the crashes for the cars tested was \$2,330. A weighted equation for finding ratings using the two variables is $0.01S + -0.00043C = R$.

b. The Volvo will have a rating of 0.63367.

c. Students may draw different examples for the sweeping line, but they all should have a slope of approximately $\frac{23}{1}$, or about a change of \$230 for every 10-unit increase in customer satisfactory rating.



d. A sweeping line can help, but because of the negative cost factor, you have to bring it from the lower right with a slope of about $\frac{23}{\$1}$ or $\frac{233}{\$10}$. The equations for the other cars will have a lower rating; the better the rating, the lower the y-intercept (because cost has a negative weight). According to the sweeping line, the best car will be the Honda, the point at the lower right.

e.

Car Problems

Car	Rating	Rank
Nissan Maxima	–0.1702	8
Honda Accord	0.87381	1
Saab 900S	0.57438	4
Subaru Legacy	0.70462	2
Volvo 850	0.63367	3
Toyota Camry	0.47896	5
Ford Taurus	0.06998	7
Lumina	0.11953	6

f. Student responses will vary. They may believe they should gather other data before they can make any kind of a statement about the cars.

UNIT II

Modeling, Matrices, and Multiple Regression

INTRODUCTORY ACTIVITY

What Affects Your Walking Speed?

Materials: measuring device; stopwatch
Technology: graphing calculator
Pacing: 1/2 class period or as homework

Overview

This activity is designed to set the stage for the second unit of the module, investigating data where there is a relationship between two or more variables and finding a mathematical model to describe that relationship. Students are to design an experiment to determine the length of a person's stride and the time it takes to walk a certain course to investigate the relationship between stride length and walking speed and the relationship between height and walking speed.

Teaching Notes

To measure stride length, students might moisten the bottoms of their shoes, walk on a dry pavement, and then and measure the distance between footprints. To control variability, they may choose to measure several strides for one person and find an average or a median stride.

Some students will have had experience fitting an equation to data and may fit a line or some other model to the data from the experiment. If the class as a whole has this expertise, you may want to treat Lesson 5 as a review and point out the critical aspects of fitting a curve to data as described in the lesson overview.

If they make both plots on the same axes, students will probably have trouble because the heights will be much larger than the stride length. If they use two different sets of axes, they will find it difficult to decide which of the two relationships is the stronger because scaling will affect their perception, unless one of the relationships is clearly linear and the other is not. This points out the need for a measure such as correlation to answer the question about which relationship is stronger.

Problem 3 challenges students to begin thinking about how to work with two variables predicting an outcome. They may decide to combine stride and height in some way, or they may not think it possible to use both at once. Have students share their strategies and try to decide as a class whether any of the suggestions might be reasonable.

STUDENT PAGE 69

INTRODUCTORY ACTIVITY

What Affects Your Walking Speed?

Do people who have long strides walk faster than those who do not?

Does the length of your stride affect how fast you can walk?

We are often interested in the relationship between two or more variables. We can use linear relationships to make predictions for one variable when we know the value of the other variable. In this unit you will explore ways to use several variables to predict the values of another variable.

OBJECTIVE

Compare the strengths of different relationships.

EXPLORE

Take a Walk

1. Design an experiment to determine the length of each person's stride and the time it takes the person to rapidly walk a given course. Have each person in class walk the course as fast as possible and record both the person's stride length and time.
   a. Plot (length of stride, time). Describe the plot.
   b. Does there appear to be a relationship between the variables?
2. Does your height affect your speed? Do taller people walk faster?
   a. Collect data on height from the class. Plot (height, time).
   b. Describe the plot. Does there appear to be a relationship between the variables?

STUDENT PAGE 70

c. Which appears to have the stronger relationship: (height, time) or (length of stride, time)? How did you decide?

3. How could you use both length of stride and height to determine time? Would a plot help? Explain your answer.

4. What other factors might affect your walking speed? How could these factors be incorporated into the process for predicting your walking speed?

LESSON 5

Matrices and Linear Regression

Materials: *Lesson 5 Quiz*
Technology: graphing calculator
Pacing: 1 to 3 class periods

Overview

This lesson provides a review of the basic concepts in fitting curves to data: least squares linear regression, correlation, residuals, sum of squared residuals, and root mean squared error. The focus is on using matrices as a tool in the process. Students write a prediction equation in matrix form, multiply matrices to make predictions, and subtract matrices to find residuals. The product of the transpose of the residual matrix and the residual matrix produces the sum of squared residuals. Students are asked to integrate all of these concepts to find a good fit or model that describes the relationship between two variables and to think carefully about making predictions or extrapolations from a model.

Teaching Notes

The lesson assumes students have had earlier experience with fitting a curve to data. Correlation as a measure of association between two variables is revisited with questions designed to expose how correlation can be misinterpreted. Examples of strong correlation on plots of data that are not really linear, such as a set of data with two clusters of points, and of the poor correlation for a perfect quadratic are designed to help students recognize that correlation does not describe how well a curve fits a set of data. Analyzing residuals, the differences between the actual values and the values predicted by an equation, is a critical component in the curve-fitting process. There should be neither pattern nor predictability in residuals. If plots of residuals against the independent variables reveal a pattern, the model chosen is not really satisfactory. The curve-fitting process usually involves finding a least squares regression line, the line that minimizes the sum of squared residuals. When data are not linear, a transformation can be used to "straighten" them and a least squares regression line calculated for the transformed data. Essentially, the root mean squared error is the square root of the average of the squared residuals and provides a measure of error on the same scale as the response variable. Students should recognize that because everything described above in the modeling process is a function of the mean, all of these tools are sensitive to outliers and can be greatly influenced by a few very large or very small values.

If your students have done curve fitting in great detail before, use this section for a light review. Students should recognize how to use matrices in the process and understand what the various components are about.

Technology

The use of a graphing calculator or curve-fitting software package and some way to manipulate matrices are essential for this lesson. Students should be able to fit curves to data, find residuals, and express their work in matrices quickly and efficiently.

Follow-Up

Students might reflect on what they have learned about fitting curves to data and how their understanding has grown. You might ask them to write about the process to describe their understanding, the various components, and the problems that can be encountered and to suggest examples or situations where curve fitting is actually used.

STUDENT PAGE 71

LESSON 5

Matrices and Linear Regression

How can you find a model for the relationship between height and weight? between the number of calories and the grams of fat in fast food? between time and the change in temperature in a physics lab?

How can matrices be useful in the process?

In earlier work you studied different methods to represent relationships: linear models and nonlinear models such as exponential, logarithmic, or power models. You considered questions relative to the ideas that are important in the modeling process. How do you know if a model is appropriate?

You can use matrices to help you in the modeling process. In the first section you studied how to use matrices to solve equations, to evaluate formulas, and to determine weighted values for what is "best." Matrices can be used in a variety of situations to help represent and manipulate information. An important area in mathematics and science is to find an algebraic model to describe the relationship in a set of data. In this lesson you will learn how to use matrices to help determine how good your model is for making predictions.

OBJECTIVES

Review the process of modeling a relationship between variables by fitting a line to the data.

Use matrices to express the least squares regression equation.

Review correlation, root mean squared error, and residuals.

INVESTIGATE

Airline Operating Costs

Is there any relationship between the number of seats of an aircraft and the operation costs? The matrix on the next page lists the operating costs and the number of seats on different airplanes.

Solution Key

Discussion and Practice

1. **a.** The points are nearly on a straight line with a positive slope. As one variable increases, so does the other.

STUDENT PAGE 72

Airplanes: Seats, Cost

Airplane	No. of Seats	Operating Cost per Hour
B747-100	405	$6132
L-1011-100/200	296	3885
DC-10-10	288	4236
A300 B4	258	3526
A310-300	240	3484
B767-300	230	3334
B767-200	193	2887
B757-200	188	2301
B727-200	148	2247
MD-80	142	1861
B737-300	131	1826
DC-9-50	122	1830
B727-100	115	2031
B737-100/200	112	1772
F-100	103	1456
DC-9-30	102	1778
DC-9-10	78	1588

Source: *World Almanac*, 1992

Discussion and Practice

1. Figure 5.1 is a scatter plot of (*number of seats, operating costs*).

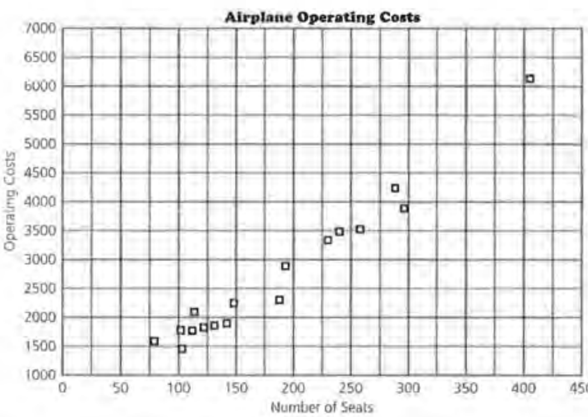


Figure 5.1

b. Student answers will vary. (150, 2250) (350, 5000) are two points on one line drawn through the data. The rate of change would be $\frac{5000-2250}{350-150} = \frac{2750}{200} = \frac{55}{4}$.

For every increase of four seats, the cost increases $55.00.

c. A sample answer for the line above is OC = $2250 + 13.75($N$ – 150), where N = the number of seats, or OC = $13.75 × N + $187.5.

d. OC = $13.75(240) + $187.50 = $3487.50, which is close to the actual value, $3484.

STUDENT PAGE 73

a. Describe the relationship between the variables in the scatter plot.

b. Draw a line that you think captures the trend in the relationship. Determine the rate of change for the line and explain what it represents in terms of the data.

c. Write the equation of the line you drew.

d. Using the equation of the line you created, predict the cost to run a plane that has 240 seats. Compare your value with the actual value given in the original data.

The process of fitting a model to data is called *regression*. One of the ways to determine whether you have a good model is to analyze the way the model behaves in relation to the data used to create the model. Finding the predicted values for a given model can be represented using matrices. Suppose the line you created for (number of seats, operating cost) has the equation: $C = 425 + 13.3S$. To find the predicted C for each value of S you will have to substitute for S and do the calculation. Just as you did in the earlier lessons, you can write this as a matrix problem.

$425 + 13.3S = C$ can be written

$1 \cdot 425 + S \cdot 13.53 = C$ and in matrices for some value S,

$[1\ S] \cdot \begin{bmatrix} 425 \\ 13.3 \end{bmatrix} = [C]$ or $X\,B = \hat{Y}$

where X is the matrix that contains the data for the number of seats along with a 1; B is the slope and intercept matrix, and $\hat{Y}$ is the predicted value matrix.

For $S = 405$ seats, the system would give you $[1\ 405] \cdot \begin{bmatrix} 425 \\ 13.3 \end{bmatrix} = [5811.5]$,

or an airplane with 405 seats will cost about $5811.50 per hour to operate.

Rather than doing just one x-value at a time, rewrite matrix X with a column vector of 1s as the first column and a column vector of all possible values for the number of seats as the second column.

STUDENT PAGE 74

$$X \cdot B = \begin{bmatrix} 1 & 405 \\ 1 & 296 \\ 1 & 288 \\ 1 & 258 \\ 1 & 240 \\ 1 & 230 \\ 1 & 193 \\ 1 & 188 \\ 1 & 148 \\ 1 & 142 \\ 1 & 131 \\ 1 & 122 \\ 1 & 115 \\ 1 & 112 \\ 1 & 103 \\ 1 & 102 \\ 1 & 78 \end{bmatrix} \cdot \begin{bmatrix} 425 \\ 13.3 \end{bmatrix}$$

2. Multiplying X and B will give you matrix $\hat{Y}$, the column vector of all predicted y-values.

a. Find XB and justify the dimensions of B.

b. Find the predicted operating cost for 296 seats using matrix multiplication.

c. $1,781 is one of the entries in XB. Explain how it was obtained.

Residuals

A *residual* is the difference between the predicted operating cost and the actual operating cost for a given number of seats. A careful study of the residuals will help you determine whether a model is a good fit for the data. If there is a pattern to the residuals, the error is predictable and another model might be better. The best models will have residuals that are not predictable and that seem to be randomly scattered above and below the line where the residuals equal 0.

2. a.

$$\begin{bmatrix} 5811.5 \\ 4361.8 \\ 4255.4 \\ 3856.4 \\ 3617.0 \\ 3484.0 \\ 2991.9 \\ 2925.4 \\ 2393.4 \\ 2313.6 \\ 2167.3 \\ 2047.6 \\ 1954.5 \\ 1914.6 \\ 1794.9 \\ 1781.6 \\ 1462.4 \end{bmatrix}$$

The dimensions of B are $(17 \times 2) \times (2 \times 1)$ which is (17×1).

b. The operating costs for 296 seats is predicted to be $4361.80 or about $4362.

c. If you have 102 seats, the operating cost would be $1781. It was obtained by multiplying 1×425 and 102×13.3 and adding the results.

STUDENT PAGE 75

The residuals are indicated on the plot in Figure 5.2.

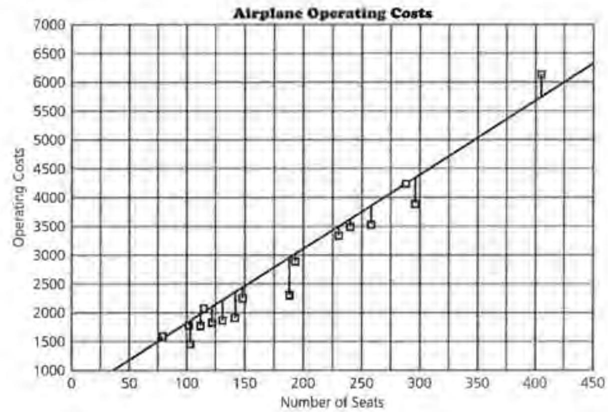


Figure 5.2

The residuals vector (or matrix) can be found by taking the difference between the matrix of the vector for the actual cost, Y, and the matrix of the vector for the predicted cost, $\hat{Y}$.

$Y - \hat{Y} = R$ where R is the matrix of residuals.

$$\begin{bmatrix} 6132 \\ 3885 \\ 4236 \\ 3526 \\ 3484 \\ 3334 \\ 2887 \\ 2301 \\ 2247 \\ 1861 \\ 1826 \\ 1830 \\ 2031 \\ 1772 \\ 1456 \\ 1778 \\ 1588 \end{bmatrix} - \begin{bmatrix} 5811.5 \\ 4361.8 \\ 4255.4 \\ 3856.4 \\ 3617.0 \\ 3484.0 \\ 2991.9 \\ 2925.4 \\ 2393.4 \\ 2313.6 \\ 2167.3 \\ 2047.6 \\ 1954.5 \\ 1914.6 \\ 1794.9 \\ 1781.6 \\ 1462.4 \end{bmatrix} = \begin{bmatrix} 320.5 \\ -476.8 \\ -19.4 \\ -330.4 \\ -133.0 \\ -150.0 \\ -104.9 \\ -624.4 \\ -146.4 \\ -452.6 \\ -341.3 \\ -217.6 \\ 76.5 \\ -142.6 \\ -338.9 \\ -3.6 \\ -125.6 \end{bmatrix}$$

Root mean squared error

In earlier work on fitting lines to data, you probably studied different ways to determine how well a line fits data. One of the measures was the sum of the squared residuals, $\Sigma(y_i - \hat{y})^2$. The *root mean squared error* is the square root of the average

3. **a.** No, you cannot multiply two matrices in which the number of columns in the first matrix is different from the number of rows in the second.

b. The dimensions of R^T are (1×17).

c. The result of the product of the transpose of the residual matrix and the residual matrix would have dimensions (1×1) and therefore be a scalar. The value in the matrix would be the sum of the squares of the residuals.

d. $\Sigma(y_i - \hat{y}_i)^2$

4. $R^T \times R = [1427611.09]$

a. $\sqrt{\frac{1427611.09}{17}} \approx 289.79$, the root mean squared error.

b. $C = \$425 + \$13.3(400) = \$5745$

You can expect the prediction to be around ± \$290 of the actual operating costs.

STUDENT PAGE 76

of this sum, $\sqrt{\frac{(\Sigma(y_i - \hat{y}_i)^2)}{n}}$. This tells you how large in magnitude a typical residual is. To find the root mean squared error, take the residuals, square them, find the sum of the squares, divide by the number of residuals, and then take the square root. The smaller the root mean squared error (or, equivalently, the smaller sum of the squared residuals), the better the line will yield predicted y-values that are close to the actual y-values.

3. In order to get the sum of squared residuals, you have to square each individual residual.

a. Could you square the residual matrix to accomplish this? Why or why not?

b. Remember that the transpose of a matrix, A^T, is formed when the rows and columns of the matrix are interchanged. What are the dimensions of the transpose of the residual matrix R?

c. In the product AB, A is said to premultiply B. Describe the result if the transpose of the residual matrix premultiplies the residual matrix itself.

$$R^T \cdot R$$

$$[320.5\ -476.8\ -19.4\ -330.4\ 133.0\ -150.0\ -104.9\ -624.4\ -146.4\ -452.6\ -341.3\ -217.6\ 76.5\ -142.6\ -338.9\ -3.6\ -125.6] \cdot \begin{bmatrix} 320.5 \\ -476.8 \\ -19.4 \\ -330.4 \\ -133.0 \\ -150.0 \\ -104.9 \\ -624.4 \\ -146.4 \\ -452.6 \\ -341.3 \\ -217.6 \\ 76.5 \\ -142.6 \\ -338.9 \\ -3.6 \\ -125.6 \end{bmatrix}$$

d. Which of the following is equivalent to $(R^T)(R)$?

i. R^2 ii. $\Sigma(y_i - \hat{y}_i)^2$ iii. $(R)(R^T)$ iv. $\Sigma(y_i - \bar{y})^2$

4. Calculate the product $(R^T)(R)$.

a. Use it to find the root mean squared error.

b. Use the line $C = 425 + 13.3S$ to predict the operating costs for a plane that has 400 seats. What does the root mean squared error indicate about your prediction?

c. It appears that the predictions are off by around $290 on average.

5. **a.** A residual is the vertical distance between the actual value and the predicted value for a given explanatory or independent variable. If the residual values are small in absolute value and there is no pattern to them, the model is probably a good fit.

b. The plot looks rather random; however, a majority of the residuals are negative.

c. The values of the residuals are the same. The display is now the residual against the x-coordinate, and they are plotted on a graph containing the line residual = 0. A majority of the residuals are negative.

d. The equation does not seem to be a good fit because of the predominance of negative residuals.

STUDENT PAGE 77

c. In general, what conclusions can you make about the prediction model?

5. Inspecting a plot of the residuals can help you determine whether there is a pattern in them.

a. How does a residual relate to the graph of the data and to the equation? How can you tell from the plot when the fit is relatively good?

b. Figure 5.3 contains a plot of (*number of seats, residuals*). This is called a *residual plot*. Describe the plot.

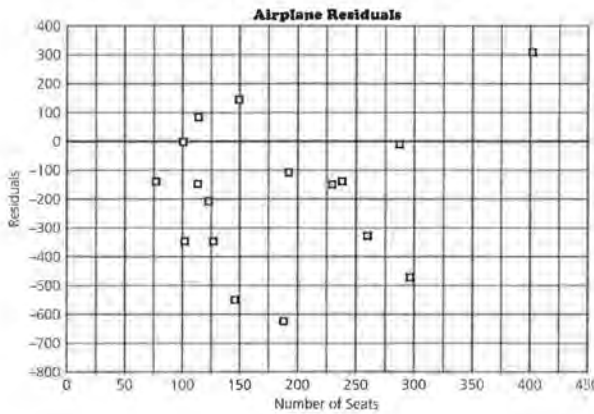


Figure 5.3

c. How does the plot of the residuals in Figure 3 compare to the residuals you observed in the plot in Figure 2?

d. Write an argument explaining why or why not the line $C = 425 + 13.3S$ is a *good* line to represent the relationship between the number of seats and the operating costs of the planes.

Linear regression

One of the lines you studied earlier is the *least squares linear regression* line. You may remember from your earlier work that the least squares linear regression equation has the smallest sum of squared errors of all lines that can be fit to a data set.

STUDENT PAGE 78

Most graphing calculators or computers can easily calculate the least squares linear regression line after you enter the data.

6. Calculate the least squares linear regression line for (*number of seats, operating costs*).

a. Use it to write a matrix representation for the prediction model.

b. Find $\hat{Y}$, the predicted y-values matrix, and R, the residuals matrix.

c. Use the matrix model to find the sum of the squared residuals. How does the sum compare to the one you found earlier?

7. Find the root mean squared error for the predicted operating costs based on the number of seats using the equation for the least squares regression line.

a. Use the least squares regression line to predict the operating costs for a plane with 240 seats. What does the root mean squared error indicate about your prediction?

b. How do you think outliers will affect the root mean squared error?

8. Plot (*number of seats, residuals*) using the least squares regression line for (*number of seats, operating costs*).

a. Describe the plot.

b. How well does the least squares linear regression model fit the data?

Correlation

Another tool in the modeling process is correlation. *Correlation* is a measure of the degree of linear association between two variables. The correlation coefficient, r, indicates how tightly packed the data are around a line; r^2 indicates the percent of change in the y-values that can be attributed to changes in the x-values. If the correlation is weak (close to zero), a linear model is not appropriate, although another model may be. It is important to look at a plot of the data before drawing any conclusions about the model. A strong correlation (close to 1 or −1) does not necessarily mean the model you are investigating is the best model. Correlation is very sensitive to outliers and clusters of data points.

6. The least squares linear regression line is $y = 13.71x + 174.47$.

a. The matrix equation is $\mathbf{XB} = \hat{\mathbf{Y}}$ when $\mathbf{B} = \begin{bmatrix} 174.47 \\ 13.71 \end{bmatrix}$.

b.

$$\hat{\mathbf{Y}} = \begin{bmatrix} 5727.02 \\ 4232.63 \\ 4122.95 \\ 3711.65 \\ 3464.87 \\ 3327.77 \\ 2820.50 \\ 2751.95 \\ 2203.55 \\ 2121.29 \\ 1970.48 \\ 1847.09 \\ 1751.12 \\ 1709.99 \\ 1586.60 \\ 1572.89 \\ 1243.85 \end{bmatrix} \qquad R = \begin{bmatrix} 404.98 \\ -347.63 \\ 113.05 \\ -185.65 \\ 19.13 \\ 6.23 \\ 66.50 \\ -450.95 \\ 43.45 \\ -260.29 \\ -144.48 \\ -17.09 \\ 279.88 \\ 62.01 \\ -130.60 \\ 205.11 \\ 344.15 \end{bmatrix}$$

c. [404.98 −347.63 113.05 −185.65 19.13 6.23 66.5 −450.95 43.45 −260.29 −144.48 −17.09 279.88 62.01 −130.6 205.11 344.15]

$$\cdot \begin{bmatrix} 404.98 \\ -347.63 \\ 113.05 \\ -185.65 \\ 19.13 \\ 6.23 \\ 66.50 \\ -450.95 \\ 43.45 \\ -260.29 \\ -144.48 \\ -17.09 \\ 279.88 \\ 62.01 \\ -130.60 \\ 205.11 \\ 344.15 \end{bmatrix} = R^2 = \Sigma(y_i - \hat{y}_i)^2$$

$R^T \times R = \Sigma(y_i - \hat{y}_i)^2 = 890833.64$ = the sum of the squared residuals. The sum of the squared residuals from the first model was larger, 1,427,611. This model is a definite improvement.

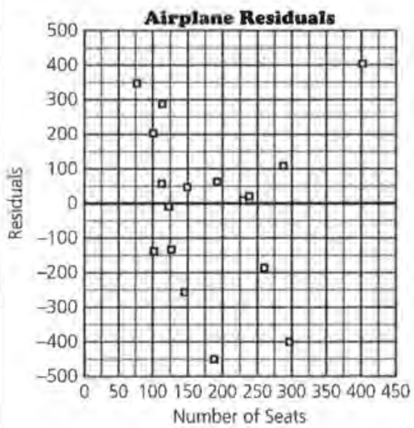
7. The root mean squared error = $\sqrt{\frac{890887.43}{17}} \approx 228.91$.

a. $\$13.71(240) + \$174.47 = \$3464.87$. We expect that the prediction is within ± \$228.91 or about \$230 of the actual cost.

b. If there are any outliers, the root mean squared error will be influenced by being made larger. The outlier will have a large residual, so it will contribute a very large value to $\Sigma(y_i - \hat{y}_i)^2$.

STUDENT PAGE 79

8.



a. Answers will vary. There does not seem to be a pattern to the plot. (There is a *slight* amount of curvature, but it is not pronounced.) The residuals are positive for both a large and a small number of seats.

b. The least squares line is a fairly good fit for the data.

Correlation

9. The correlation coefficient for (number of seats, operating costs) is 0.982.

a. The relationship is very nearly linear.

b. $r^2 = 0.9643$ meaning approximately 96% of the variability in operating costs can be attributed to variability in the number of seats.

c. The data does not contain any apparent outliers.

10. There seems to be a strong relationship between seats and operating costs. This is supported by a relatively small root mean squared error and a high correlation coefficient. (The residuals may cause one to look for an additional relation since there does appear to be *some* curvature in the residual plot.)

9. Using a calculator or computer, find the correlation for (*number of seats, operating costs*).

a. What does it tell you about the relation between the two variables?

b. What is the value of r^2? What does it tell you about the relationship?

c. Are there any outliers? If there are, how do they affect the correlation?

10. Summarize the work you have done in problems 6–9 to find a model for the relationship between the number of seats and the operating costs for the set of planes.

There are many other variables related to the operating costs of airplanes. Will a least squares linear regression model help you describe any of these relationships? Additional information about airplanes is included in the table below.

More Airplane Data

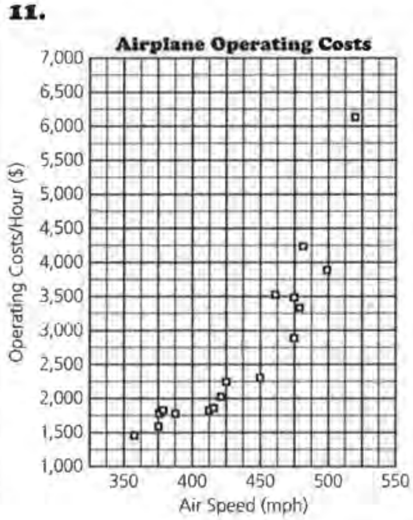
Aircraft	No. of Seats	Air Speed (mph)	Operating Cost/Hour ($)
B747-100	405	519	6132
L-1011-100/200	296	498	3885
DC-10-10	288	484	4236
A300 B4	258	460	3526
A310-300	240	473	3484
B767-300	230	478	3334
B767-200	193	475	2887
B757-200	188	449	2301
B727-200	148	427	2247
MD-80	142	415	1861
B737-300	131	413	1826
DC-9-50	122	378	1830
B727-100	115	422	2031
B737-100/200	112	388	1772
F-100	103	360	1456
DC-9-30	102	377	1778
DC-9-10	78	376	1588

Source: *World Almanac*, 1992

11. Consider the speed at which a plane flies and its relation to operating costs. Plot (*speed, operating costs*).

a. Describe the association you can see in the plot.

11.



a. The plot does not appear linear. It appears that the costs increase at a faster rate as speed increases.

(11) b. The correlation coefficient is .8958. However, since the plot is not linear we cannot use it as a reliable measure of the relationship between the speed and the cost per hour.

c. The regression line does not appear to be the best fitting model. It helps to see graphically that the plot is not linear. The least squares regression equation is $OC = -7211.14 + 22.8305S$.

12.

$$\begin{bmatrix} 1 & 519 \\ 1 & 498 \\ 1 & 484 \\ 1 & 460 \\ 1 & 473 \\ 1 & 478 \\ 1 & 475 \\ 1 & 449 \\ 1 & 427 \\ 1 & 415 \\ 1 & 413 \\ 1 & 378 \\ 1 & 422 \\ 1 & 388 \\ 1 & 360 \\ 1 & 377 \\ 1 & 376 \end{bmatrix} \times \begin{bmatrix} -7211.14 \\ 22.83 \end{bmatrix} = \begin{bmatrix} 4637.63 \\ 4158.2 \\ 3838.58 \\ 3290.66 \\ 3587.45 \\ 3701.6 \\ 3633.11 \\ 3039.53 \\ 2537.27 \\ 2263.31 \\ 2217.65 \\ 1418.6 \\ 2423.12 \\ 1646.9 \\ 1007.66 \\ 1395.77 \\ 1372.94 \end{bmatrix}$$

a.

$$\begin{bmatrix} 6132 \\ 3885 \\ 4236 \\ 3526 \\ 3484 \\ 3334 \\ 2887 \\ 2301 \\ 2247 \\ 1861 \\ 1826 \\ 1830 \\ 2031 \\ 1772 \\ 1456 \\ 1778 \\ 1588 \end{bmatrix} - \begin{bmatrix} 4637.63 \\ 4158.2 \\ 3838.58 \\ 3290.66 \\ 3587.45 \\ 3701.6 \\ 3633.11 \\ 3039.53 \\ 2537.27 \\ 2263.31 \\ 2217.65 \\ 1418.6 \\ 2423.12 \\ 1646.9 \\ 1007.66 \\ 1395.77 \\ 1372.94 \end{bmatrix} = \begin{bmatrix} 1494.37 \\ -273.2 \\ 397.42 \\ 235.34 \\ -103.45 \\ -367.6 \\ -746.11 \\ -738.53 \\ -290.27 \\ -402.31 \\ -391.65 \\ 411.4 \\ -392.12 \\ 125.1 \\ 448.34 \\ 382.23 \\ 215.06 \end{bmatrix}$$

STUDENT PAGE 80

b. Find the correlation for (*speed, operating costs*). What does this indicate about the data?

c. Find the least squares linear regression line. How well do you think it describes the relationship?

12. Create a matrix system to use the least squares linear regression line to predict operating costs using speed.

a. Use the matrix system to find the residuals.

b. What is the root mean squared error? Estimate the error if you use your linear model to predict the operating costs for a plane that travels 400 miles per hour.

c. Plot (*speed, residuals*). What does this plot indicate about using the model?

Transforming Data

When a linear regression model is not appropriate, but the data do seem to form a pattern, you can try transforming the data to determine a better model. Because logarithms use a different scale, they may help to find another model. In this case, to get a better fit it might help to try a transformation that would lower the points with large operating costs on the plot.

13. Divide the work among group members so that each member has a strategy to transform the speed and operating costs. Select your strategy from the list below.

(*speed, ln(operating costs)*)

(*speed, multiplicative inverse of operating costs*)

(*ln(speed), operating costs*)

(*ln(speed), ln(operating costs)*)

a. Plot the paired data after each transformation. Which of the transformations seems to produce the most linear relationship?

b. Find the least squares linear regression line for the transformed data you chose in part a.

14. Use the least squares regression line you found above to determine how well the line will fit the data by answering the following questions.

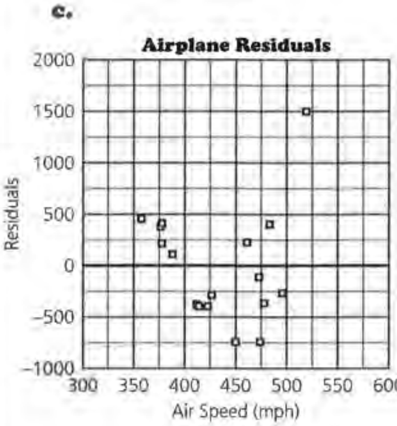
a. Write a matrix representation of the linear equation. Use it to create a set of predicted values.

b. $R^T \times R = 4900562.992$

The root mean squared error = 536.91.

$22.83(400) - 7211.14 = 1920.86$. You can expect an error in the prediction to be around \$537, or the cost to be between \$1384 and \$2458.

c.



There seems to be a pattern to the residual plot. The points in the middle are below the line, indicating that there is probably another model that will fit better.

13. a. The transformation of $\left(\text{speed}, \frac{1}{\text{operating costs}}\right)$ seems to be the most linear when graphed.

b. The least squares regression line is $y = -0.000003093S + 0.0017755$, where $y = \frac{1}{\text{cost}}$.

14. a.

$$\begin{bmatrix} 1 & 519 \\ 1 & 498 \\ 1 & 484 \\ 1 & 460 \\ 1 & 473 \\ 1 & 478 \\ 1 & 475 \\ 1 & 449 \\ 1 & 427 \\ 1 & 415 \\ 1 & 413 \\ 1 & 378 \\ 1 & 422 \\ 1 & 388 \\ 1 & 360 \\ 1 & 377 \\ 1 & 376 \end{bmatrix} \times \begin{bmatrix} .0017755132 \\ -.000003093 \end{bmatrix}$$

$$= \begin{bmatrix} .000170 \\ .000235 \\ .000279 \\ .000353 \\ .000313 \\ .000297 \\ .000306 \\ .000387 \\ .000455 \\ .000492 \\ .000498 \\ .000606 \\ .000470 \\ .000575 \\ .000662 \\ .000609 \\ .000613 \end{bmatrix}$$

= The prediction for $\frac{1}{\text{cost}}$.

The predicted cost would be

$$\begin{bmatrix} 5882.4 \\ 4255.3 \\ 3597.1 \\ 2832.9 \\ 3194.9 \\ 3367 \\ 3268 \\ 2584 \\ 2197.9 \\ 2032.5 \\ 2008 \\ 1650.2 \\ 2127.7 \\ 1739.9 \\ 1510.6 \\ 1642 \\ 1631.3 \end{bmatrix}.$$

Note: If you store the predictions for $\frac{1}{\text{cost}}$ in a calculator or computer and then take reciprocals, you may get slightly different answers due to round-off error.

b.

$$\begin{array}{ccccc} \text{Actual Cost} & & \text{Predicted Cost} & & \\ \begin{bmatrix} 6132 \\ 3885 \\ 4236 \\ 3526 \\ 3484 \\ 3334 \\ 2887 \\ 2301 \\ 2247 \\ 1861 \\ 1826 \\ 1830 \\ 2031 \\ 1772 \\ 1456 \\ 1778 \\ 1588 \end{bmatrix} & - & \begin{bmatrix} 5882.4 \\ 4255.3 \\ 3597.1 \\ 2832.9 \\ 3194.9 \\ 3367 \\ 3268 \\ 2584 \\ 2197.9 \\ 2032.5 \\ 2008 \\ 1650.2 \\ 2127.7 \\ 1739.9 \\ 1510.6 \\ 1642 \\ 1631.3 \end{bmatrix} & = & \begin{bmatrix} 249.65 \\ -370.3 \\ 638.88 \\ 693.14 \\ 289.11 \\ -33 \\ -381 \\ -283 \\ 49.198 \\ -171.50 \\ -182 \\ 179.83 \\ -96.66 \\ 32.87 \\ -54.57 \\ 135.96 \\ -43.32 \end{bmatrix} \end{array}$$

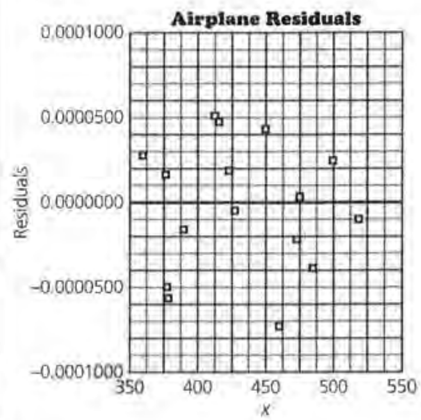
$R^T \times R = 1529040.26$

Root mean squared error

$= \sqrt{\frac{1529040.26}{17}} \approx 299.91$

If we use this model to predict operating costs from the speed, the predictors will tend to be off by \$300 or so.

c.



There is no pattern in the residual plot, which indicates that the relationship between $\frac{1}{\text{cost}}$ and speed is linear.

d. Using the model $\left(\text{speed}, \frac{1}{\text{cost}}\right)$, the prediction equation is

$\frac{1}{\text{cost}} = -0.000003093OS + 0.0017755.$

This will yield $0.0017755132 - 0.000003093(550) = 0.0000743632$ where $\frac{1}{\text{cost}} = 0.0000743632$, so cost = \$13,448.

The prediction will be quite unstable because 550 mph exceeds the largest speed used to make the model, and there is no guarantee that the same model will hold for a different domain. (In fact, this can be a very good investigation for students to pursue. Use a speed of 750 mph and ask them to predict the operating cost, and to explain the mathematics. The curve is a hyperbola with a vertical asymptote around 575 which means that after 575, the values are negative. The model is only appropriate for the domain from about 350 to 525.)

STUDENT PAGE 81

15. a. No. The process of matrix multiplication involves the multiplication of the row times the column and adds those values. If only numbers are involved you create constants; if variables are involved you create polynomial equations. In no case will it be possible to create an exponential equation using matrix multiplication in this manner. The system $[1\ x] \times \begin{bmatrix} 10 \\ 0.52 \end{bmatrix} = [y]$ yields $y = 10 + 0.52x$, not $y = 10x^{0.52}$.

b. $\ln(y) = 0.52\ln(x) + \ln(10)$

$\ln(y) = 0.52\ln(x) + 2.30258$

c. cost $= 10(7^{0.52}) = 27.51$; $\ln(\text{cost}) = .52(\ln(7)) + 2.30258 = 1.011873 + 2.30258$; $\text{antiln}(3.31445) = 27.51$

The answers are the same.

b. Use matrices to find the residuals and the root mean squared error for your model. What do these indicate about using your model for the relation between speed and operating costs?

c. Make a residual plot. What does this indicate about the relationship?

d. Consider using the model you found to be the best from working the first part of this problem. If you were to predict the operating costs for a speed of 550 miles per hour, how reliable do you think your prediction will be? Explain how you found your answer.

15. A model that seems to be a good fit for (age of plane, cost) is $y = 10x^{0.52}$ where y is in hundreds of dollars.

a. Does the matrix system below represent the model? Explain why or why not.

$[1\ x] \cdot \begin{bmatrix} 10 \\ 0.52 \end{bmatrix} = [y]$

b. Take the natural logarithm of both sides of the equation.

$\ln y = \ln 10x^{0.52}$

Use the properties of logarithms to express the relationship in the form of a sum.

c. Use both the equation $y = 10x^{0.52}$ and your answer for part b to predict the cost of operating a seven-year-old airplane. How do the two compare?

Summary

Matrices can be a useful tool in the modeling process. If you see a pattern in a set of data, you can search for a model to describe the pattern. If the pattern is linear, you can use the least squares linear regression line. If the data are not linear, you can try to straighten the data by transforming them, using what you know about the situation and about the plot. In either case, you can find the predicted values using your model by writing a matrix expression: $XB = \hat{Y}$, where X is the independent variable, Matrix B is the coefficient matrix for the model, and $\hat{Y}$ is the matrix of predicted values. The residuals are the difference between $\hat{Y}$ and Y, $Y - \hat{Y} = R$. If the residuals are small and seem to be randomly distributed, the model is a good model for the data. Correlation and root mean squared error are two other tools that can be an aid in the modeling

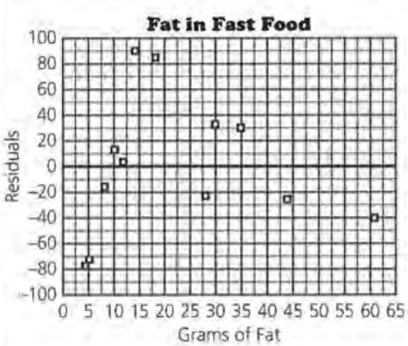
Practice and Applications

16. a.

$$\begin{bmatrix} 1 & 10 \\ 1 & 4.5 \\ 1 & 28 \\ 1 & 8 \\ 1 & 5 \\ 1 & 44 \\ 1 & 18 \\ 1 & 30 \\ 1 & 14 \\ 1 & 12 \\ 1 & 61 \\ 1 & 35 \end{bmatrix} \times \begin{bmatrix} 182 \\ 12.5 \end{bmatrix} = \begin{array}{c} \text{Predicted} \\ \text{Number of} \\ \text{Calories} \\ \begin{bmatrix} 307 \\ 238.25 \\ 532 \\ 282 \\ 244.5 \\ 732 \\ 407 \\ 557 \\ 357 \\ 332 \\ 944.5 \\ 619.5 \end{bmatrix} \end{array}$$

b.

$$\begin{array}{c} \text{Actual No.} \\ \text{Calories} \\ \begin{bmatrix} 320 \\ 162 \\ 510 \\ 267 \\ 172 \\ 706 \\ 492 \\ 589 \\ 447 \\ 334 \\ 905 \\ 649 \end{bmatrix} \end{array} - \begin{array}{c} \text{Predicted No.} \\ \text{Calories} \\ \begin{bmatrix} 307 \\ 238.25 \\ 532 \\ 282 \\ 244.5 \\ 732 \\ 407 \\ 557 \\ 357 \\ 332 \\ 944.5 \\ 619.5 \end{bmatrix} \end{array} = \begin{array}{c} \text{Residuals} \\ \\ \begin{bmatrix} 13 \\ -76.25 \\ -22 \\ -15 \\ -72.5 \\ -26 \\ 85 \\ 32 \\ 90 \\ 2 \\ -39.5 \\ 29.5 \end{bmatrix} \end{array}$$



Because there seems to be a pattern where the residuals get smaller as the number of grams of fat increases, you would suspect you did not have a very good model.

STUDENT PAGE 82

process. The sum of the squared errors is the product $R^T \cdot R$ where R is the residual matrix.

Practice and Applications

16. The least squares linear regression line for (*fat grams, calories*) for a variety of fast foods can be expressed in the following matrix system:

$[1\ x] \cdot \begin{bmatrix} 182 \\ 12.5 \end{bmatrix} = [y]$

where x = number of fat grams, y = calories.

a. Use the system to find the predicted number of calories for the following food items:

Fast Foods Calories/Fat

Place	Food	Calories	Fat (gms)
McDonald's	McLean Deluxe	320	10
	Chicken salad	162	4.5
	Quarter Pounder	510	28
Burger King	Broiler	267	8
	Chunky chicken salad	172	5
	Whopper with cheese	706	44
Pizza Hut	Cheese pizza	492	18
	Pan supreme pizza	589	30
Taco Bell	Bean burrito	447	14
	Chicken burrito	334	12
	Taco salad	905	61
	Nachos-Bell Grande	649	35

Source: *Eating on the Run*, Human Kinetics, Tribole, Evelyn. *USA Today*, Jan. 7, 1992

b. Find the residuals using matrices and make a residual plot. What does this tell you about your model?

17. An economics student claims the relationship between the per capita gross national product in 1992 (G) and the percent of students enrolled in secondary schools in the following countries (E) can be given by the equation $G = -8152 + 223E$.

a. Write the matrix system for the equation. Use the system to predict the gross national product for the given countries.

17. a.

$$\begin{bmatrix} 1 & 44 \\ 1 & 76 \\ 1 & 53 \\ 1 & 67 \\ 1 & 63 \\ 1 & 23 \\ 1 & 28 \\ 1 & 97 \end{bmatrix} \times \begin{bmatrix} -8152 \\ 223 \end{bmatrix} = \begin{bmatrix} 1660 \\ 8796 \\ 3667 \\ 6789 \\ 5897 \\ -3023 \\ -1908 \\ 13479 \end{bmatrix}$$

(17) **b.**

$$\begin{bmatrix} 380 \\ 3010 \\ 2190 \\ 1020 \\ 3470 \\ 340 \\ 1840 \\ 23030 \end{bmatrix} - \begin{bmatrix} 1660 \\ 8796 \\ 3667 \\ 6789 \\ 5897 \\ -3023 \\ -1908 \\ 13479 \end{bmatrix} = \begin{bmatrix} -1280 \\ -5786 \\ -1477 \\ -5769 \\ -2427 \\ 3363 \\ 3748 \\ 9551 \end{bmatrix}$$

$R^T \times R = [-1280\ -5786\ -1477\ -5769\ -2427\ 3363\ 3748\ 9551] \times$

$$\begin{bmatrix} -1280 \\ -5786 \\ -1477 \\ -5769 \\ -2427 \\ 3363 \\ 3748 \\ 9551 \end{bmatrix}$$

$= [193048289] = SSR$

c. The root mean squared error is $\sqrt{\frac{193048289}{8}} \approx 4912.335$.

Since the error is larger than all but one of the original values, the model does not seem appropriate.

d. The model does not fit well because Germany is an outlier here.

STUDENT PAGE 83

	% Enrolled Secondary School	Predicted GNP (in US $)
China	44	
Hungary	76	
Iran	53	
Peru	67	
Mexico	63	
Kenya	23	
Thailand	28	
Germany	97	

Source: *Universal Almanac*, 1995

b. If the actual GNPs were as follows, use matrices to find the residuals and the sum of squared residuals.

	% Enrolled Secondary School	Predicted GNP (in US $)	Actual GNP	Residuals
China	44		380	
Hungary	76		3010	
Iran	53		2190	
Peru	67		1020	
Mexico	63		3470	
Kenya	23		340	
Thailand	28		1840	
Germany	97		23030	

c. What is the root mean squared error and what does it tell you about your model?

d. Comment on the economic student's attempt to find a linear model to relate secondary education to gross national product.

18. a. The plot does not appear to be linear. The points on the ends of the data are above the straight line, and the points in the middle are below the line.

b. The removal of the outlier will change the correlation from the original

$r = 0.894$ to $r = 0.925$.

c. The residuals have a definite pattern that is visible when they are graphed. This supports the conclusion that the relationship is not linear.

STUDENT PAGE 84

18. A plot of (*speed, length of flights*) is given in Figure 5.4 with a regression line drawn through the data.

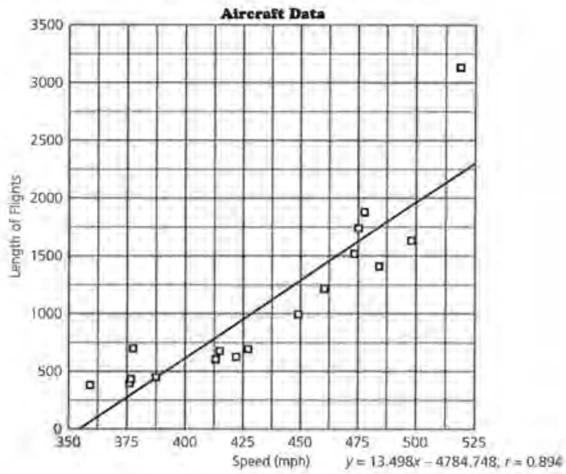


Figure 5.4

a. How well does the line seem to represent the relationship in the data?

b. The correlation for the line is 0.894. If the outlier around 520 mph were removed from the data, how do you think the correlation would change? Why?

c. Figure 5.5 contains a plot of the residuals. Describe what the residual plot indicates about using the line to predict the length of flights from the speed.

STUDENT PAGE 85

19. a. There appears to be a funnel-type pattern, very narrow (close to the line) on the left and widening as the value of length of flight increases. The error is increasing as the length of the flight increases.

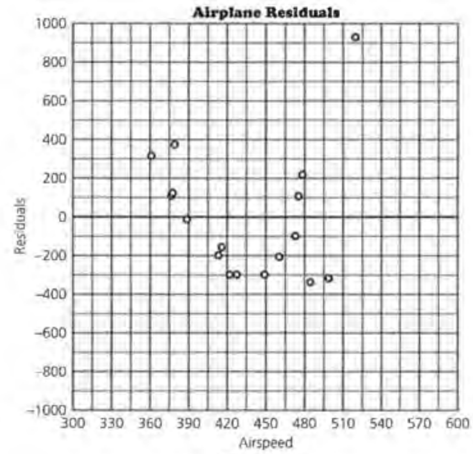


Figure 5.5

19. Figure 5.6 is a residual plot for the least squares linear regression line on (*flight length, fuel consumption*) for the set of planes.

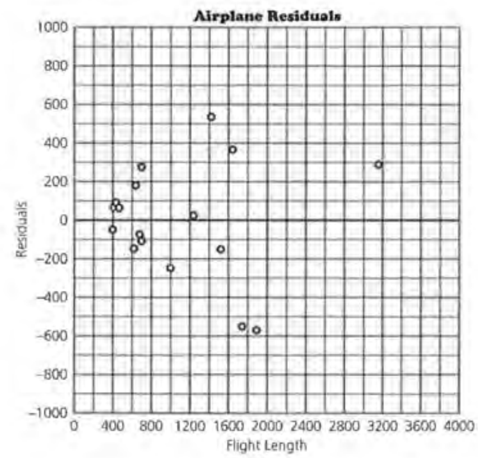


Figure 5.6

b. This model would become less accurate for longer flight lengths.

20. a. i. The value of the correlation coefficient implies that the data have some linear trend. This is supported by the graph.

ii. The value of the correlation coefficient implies that the data are very linear. The graph, however, does not support this but shows how an outlier can affect the correlation.

STUDENT PAGE 86

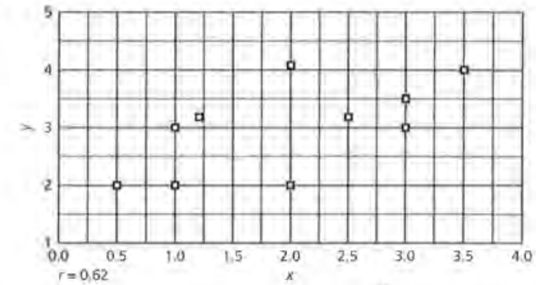
a. Describe any pattern you see in the residuals.

b. What do the residuals indicate about using the model that generated the residuals to predict the fuel consumption from the length of the flight?

20. The correlation coefficient is given for each of the plots i, ii, iii, and iv.

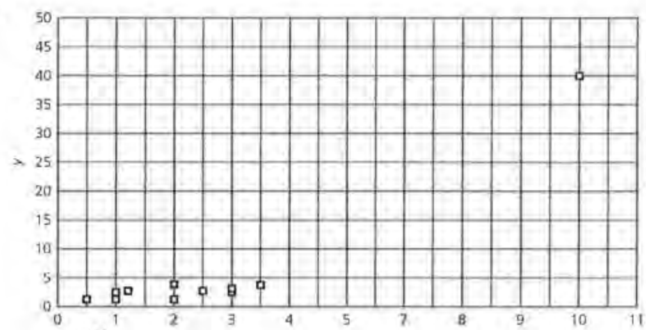
a. Comment on the relation between the plot and the numerical value of the correlation.

i.



$r = 0.62$

ii.



$r = 0.94$

(20. a.) iii. The value of the correlation coefficient implies that these data are not highly correlated. The visual appearance of the plot tends to dispute this with the exception of the one outlier. Again the outlier has a large effect on the correlation. (This time the regression line will act more like an angle bisector.)

iv. This graph does not appear to be linear, and the correlation coefficient implies the same conclusion. There is a definite pattern, however, but patterns that are not linear are not measured by the correlation coefficient.

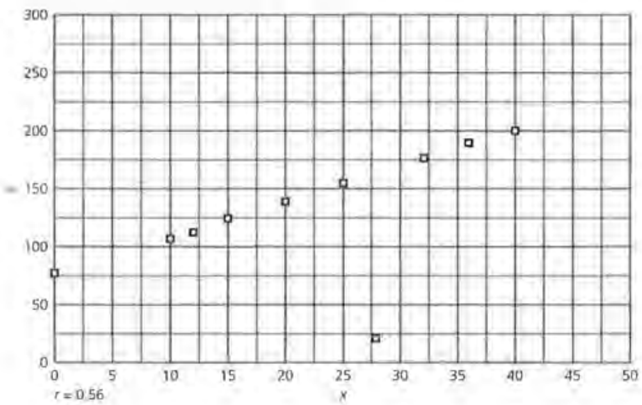
b. The inclusion of the outlier emphasizes that the correlation coefficient is sensitive to outliers and extreme values. The size of the coordinates of this point effectively causes the correlation coefficient to be calculated from two points, the outlier and one representing the cluster of points. The cluster is effectively one point because the cluster of points is so nearly equal in comparison with the outlier. If you determine the equation from two points, that equation should be linear and the correlation will be 1 or −1. So in this case the correlation is close to 1 since you nearly have 2 points.

c. Correlation can be used as one of several tools to determine the strength of the linear relationship. It should not be the only tool, however, since it is sensitive to outliers and can be misleading. It also does not reflect any relationship other than linearity.

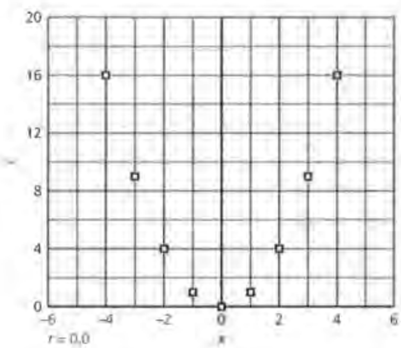
21. This problem deals with student data, and the answers will vary.

STUDENT PAGE 87

iii.



iv.



b. Plots i and ii are of the same data set with one exception. The point (10, 40) was added in plot ii. Comment on the effect of this one point on the correlation.

c. What observations can you make about using only correlation as a tool to help you find a good model for a data set?

21. Find the prices of 12 used cars of the same size or type. Plot (*age of car, price*) and look for a model that describes the price of used cars as a function of the age of the car. Explain how you chose your model and which of the tools you used to help make your choice.

LESSON 6

Multiple Variables and Modeling

Materials: none
Technology: graphing calculator
Pacing: 1 to 2 class periods

Overview

In most real situations more than one variable affects an outcome. In this lesson, two variables are used to predict a third. The mathematical process developed for such a situation builds on what students know about using a least squares regression line to explain the relationship between two variables. Students begin with one of the two independent variables and use it to predict the response variable. The new step is to explain the error in prediction (the residuals) by thinking of that error as a function of the second independent variable. Through a recursive process, students calculate a series of least squares regression lines alternating between the two independent variables with the last set of residuals as the response variable. The sum of squared residuals is used to determine whether each iteration is an improvement, and the process continues until students understand what is happening mathematically.

Teaching Notes

The recursion process is complicated and might almost be better done as a whole-class exercise where students create regression lines, calculate residuals, inspect graphs, and repeat the process under teacher direction. In order to make the investigation in the following lesson easier, have students store their results (residuals from each new regression) in a program or, if using technology with more than six lists, in another list. This will allow students to use these results as they continue their analysis later without having to redo the computations. *Be sure students label the stored data so they can recall what it represents.* A paper directory might be useful where they describe each program in more detail.

The notation can be confusing: The hat over a variable indicates it is a prediction; the difference between the variable with and without the hat is a residual. The notation used in the lesson is designed to keep as much contact with the context as possible so students can see how the pieces are being used. It helps to make a list of what each variable represents and where it is stored on the calculator or software package. It also helps if every student has the same set of data stored in the same list. For example:

List 1	2	3	4	5	6
SATV	SATM	GPA	Predicted GPA Using V	Actual GPA – Predicted GPA Using V	Actual Residuals – Predicted Residuals Using M
V	M	Y	$f(V_Y)$	$Y - f(V_Y)$	$y - f(V_Y) - g(M)$

On a separate chart, have students record the sum of squared residuals for each iteration. To find the sum of squared residuals on a graphing calculator, calculate the one-variable statistics for the list containing the residuals for that iteration and read Σy^2. Students should notice that for each iteration, the sum of squared residuals gets progressively smaller.

After the process has been carried out for two or three iterations, use overhead transparencies of the plots and review what has been happening. It may be useful to write the following mathematics from the lesson on the board or overhead.

$GPA = f(V) + e_1$ where e_1 is a function of M

$e_1 = g(M) + e_2$ where e_2 is a function of V

$e_2 = h(V) + e_3$ where e_3 is a function of M

A regression line was found for (V, GPA):

$GPA = f(V) + e_1$ where e_1 is a function of M, so a regression line was found for (M, e_1);

$e_1 = g(M) + e_2$ where e_2 is a function of V, so a regression line was found for (V, e_2);

$e_2 = h(V) + e_3$ where e_3 is a function of M, so a regression line was found for (M, e_3), and so on.

It is not important that students be able to replicate the process each time they have two independent variables, but rather that they understand the process and how it works. The practice exercises provide the opportunity for students to review the process by beginning with the second variable and allow students to try the process on their own. After a whole-class analysis of the investigation, you might have students work individually or in pairs on the practice exercises. If you have students with considerable mathematics backgrounds, you might discuss the process to begin with as a whole class, then assign half the class to do the investigation and the other half to do the practice exercises. It is not necessary for students to fill in all of the charts by hand if they can see the results on a graphing calculator or software. Be sure however, that someone stores their results in a program or another list so that the results are available for later work.

Note that round-off error is common, so the answers given here might differ from students' answers in the third or fourth decimal place.

Technology

Technology is essential to calculate the necessary least squares regressions and to keep track of the results.

Follow-Up

Ask students to explain the process used to fit a curve when you have two independent variables. This might even be done as a quiz the following day; the words students use will help you understand what they understood from the investigation done in class. As an extension, have students consider what they might do if they had three independent variables.

STUDENT PAGE 88

LESSON 6

Multiple Variables and Modeling

What factors might have an impact on your success in college?

Can you predict the grade point average a student will have in college?

What variables might be useful to help make a good prediction: high school grades, the courses the student took in high school, SAT scores, or ACT scores?

OBJECTIVES

Examine how two different factors can be used to predict an outcome.

Use recursion to generate a procedure.

In earlier lessons you learned how to find models to make predictions when you had only one independent variable. In this lesson you will learn how to use two independent variables to make a prediction. Students with high SAT verbal (*SATV*) and SAT math (*SATM*) scores tend to get better grades in college than do students who do not score well on the SAT exam. Suppose you only had those two variables available. The investigation in this lesson will show that it is possible to model the way in which college grade point average, *GPA*, depends on both *SATV* and *SATM*.

INVESTIGATE

College Entrance Data

The table that follows contains the data on *GPA*, *SATV*, and *SATM* for each of 15 college students.

STUDENT PAGE 89

College Grade Point Average

Student Number	*GPA*	*SATV*	*SATM*
1(X)	3.58	670	710
2	3.17	630	610
3	2.31	490	510
4	3.16	760	580
5	3.39	450	510
6	3.85	600	720
7	2.55	490	560
8	2.69	570	620
9	3.19	620	640
10	3.50	640	660
11	2.92	730	780
12	3.85	800	630
13	3.11	640	730
14	2.99	680	630
15	3.08	510	610

Source: Oberlin College, 1993

The scatter plot in Figure 6.1 shows the relationship between *GPA* and *SATV*. The fitted least squares linear regression line has been added to the plot. The first student on the list is shown with an *X* in the plot and on the table so you can keep track of this student as you progress through the lesson. ($\hat{GPA}$ is a notation for the predicted grade point average using the regression model.)

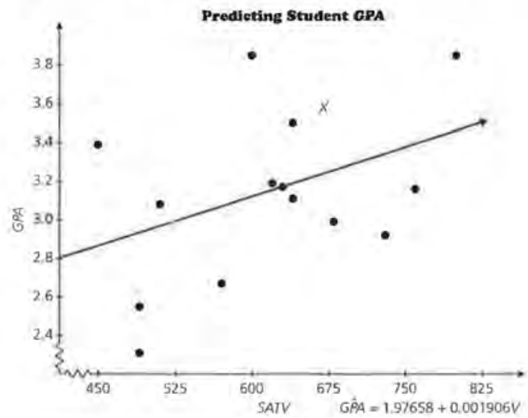


Figure 6.1

STUDENT PAGE 90

Solution Key

Discussion and Practice

1. **a.** The slope is 0.001906. It indicates that for every increase of one point in the *SAT* verbal score, the grade point of that person rises 0.001906 points on average.

b. If the *SAT* verbal would rise 100 points, the predicted grade point average should rise 0.1906 points.

c. The student with the higher *SATV* should have a higher *GPA*, and we would expect it to be about 0.19 points higher.

Figure 6.2 shows the relation between *SATM* and *GPA*.

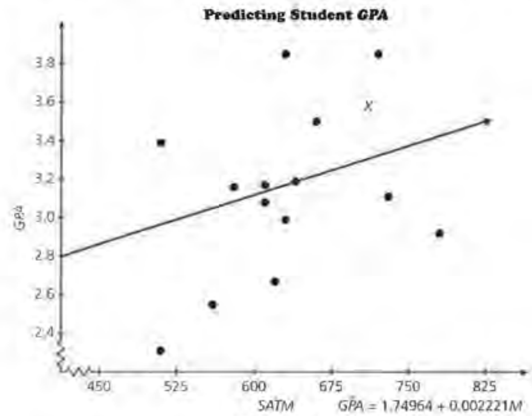


Figure 6.2

Clearly, you could use either *SATV* or *SATM* to predict *GPA*. However, it would seem even better to use both *SATV* and *SATM* as predictors.

SATV and the Regression Line

Consider again the plot of *GPA* versus *SATV*. The equation of the fitted linear regression model is $\hat{GPA} = 1.97658 + 0.001906V$. (Note that the coefficients are given to several decimal places; in doing the calculations on a computer or a calculator, carry as many digits of accuracy as possible.)

Discussion and Practice

1. Enter the data into your calculator or computer. Verify the equation for the least squares linear regression line.

  a. What is the slope of the line and what does it tell you about the relation between *SATV* and *GPA*?

  b. If *SATV* score goes up by 100, by how much on average will the *GPA* change?

  c. If one student has an *SAT* verbal score that is 100 points higher than that of a second student, what can you say about their *GPAs*?

(1) **d.** The *y*-intercept is (0, 1.97658) and implies if the *SATV* score were zero the predicted *GPA* would be 1.98, which is meaningless since *SATV* scores cannot be less than 200.

2. **a.** The first two values are the actual information concerning Student 1, the third value is the *GPA* predicted by the regression line:

$1.97658 + 670(0.001906) = 3.254.$

b.

Student Number	*GPA*	*SATV*	$\hat{GPA}_V$
1*x*	3.58	670	3.2536
2	3.17	630	3.1774
3	2.31	490	2.9105
4	3.16	760	3.4251
5	3.39	450	2.8343
6	3.85	600	3.1202
7	2.55	490	2.9105
8	2.69	570	3.0630
9	3.19	620	3.1583
10	3.50	640	3.1964
11	2.92	730	3.3680
12	3.85	800	3.5014
13	3.11	640	3.1964
14	2.99	680	3.2727
15	3.08	510	2.9486

STUDENT PAGE 91

d. What is the intercept on the vertical axis (*GPA*-intercept) and what could it mean in terms of the data?

As you may know, the lowest possible *SATV* score is 200 (and the highest possible score is 800), so zero is not a possibility. Because of this, the *GPA*-intercept only tells you where the regression line crosses the *GPA*-axis, but it does not give the predicted *GPA* for any student.

The table below contains a listing of the 15 students in the sample, the *GPA* and *SATV* for each, and the predicted *GPA* for the first student based on the equation:

$\hat{GPA}_V = 1.97658 + 0.001906\,V.$

We will call the predicted *GPA*s from this equation $\hat{GPA}_V$, since this model uses the SAT verbal score.

SAT Verbal, GPA

Student Number	*GPA*	*SATV*	$\hat{GPA}_V$
1(*X*)	3.58	670	3.254
2	3.17	630	
3	2.31	490	
4	3.16	760	
5	3.39	450	
6	3.85	600	
7	2.55	490	
8	2.69	570	
9	3.19	620	
10	3.50	640	
11	2.92	730	
12	3.85	800	
13	3.11	640	
14	2.99	680	
15	3.08	510	

2. Consider student number 1, marked with an X.

a. Explain the numbers in the first row for student 1.

b. Fill in the column on the table for the predicted *GPA* for the rest of the students.

3. **a.** The most accurate will be the students whose scores are very close to the regression line. The difference between the actual *GPA* and predicted *GPA* will be small.

b. $r_V = 0.3264$. The residual indicates the actual *GPA* is 0.3264 above the predicted value of that student.

STUDENT PAGE 92

Residuals

3. Look carefully at the plot of the data with the regression line in Figure 6.3 and at the values you now have in the table on the previous page. Remember that a residual is the difference of the predicted and the actual *GPA*.

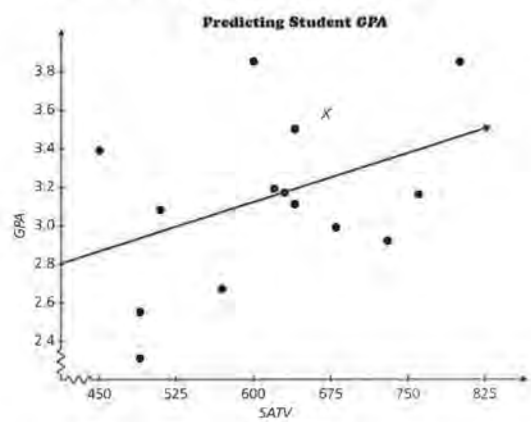


Figure 6.3

a. For which students will the prediction be the closest?

b. Let r_V represent the residuals. What is r_V for the first student? What does that tell you?

(3) c.

Student Number	*GPA*	$\hat{GPA}_V$	r_V
1*x*	3.58	3.2536	.3264
2	3.17	3.1774	–.0074
3	2.31	2.9105	–.6005
4	3.16	3.4251	–.2651
5	3.39	2.8343	.5557
6	3.85	3.1202	.7298
7	2.55	2.9105	–.3605
8	2.69	3.0630	–.3730
9	3.19	3.1583	.0317
10	3.50	3.1964	.3036
11	2.92	3.3680	–.4480
12	3.85	3.5014	.3486
13	3.11	3.1964	–.0864
14	2.99	3.2727	–.2827
15	3.08	2.9486	.1314

d. The negative sign indicates that the regression prediction is higher than the actual *GPA*. Example: Student 2 has an actual *GPA* of 3.17; the predicted value is 3.1774. $r_V = 3.17 - 3.1774 = -.0074$

STUDENT PAGE 93

Residuals *SAT Verbal, GPA*

Student Number	GPA	$\hat{GPA}_V$	r_V
1	3.58	3.254	
2	3.17	3.178	
3	2.31	2.911	
4	3.16	3.425	
5	3.39	2.834	
6	3.85	3.120	
7	2.55	2.911	
8	2.69	3.063	
9	3.19	3.159	
10	3.50	3.197	
11	2.92	3.368	
12	3.85	3.502	
13	3.11	3.197	
14	2.99	3.273	
15	3.08	2.949	

c. Complete the table above by finding the residuals for all of the students. The table contains values of

- the actual *GPA*;
- the predicted values found by $\hat{GPA}_V = 1.97658 + 0.001906V$;
- and r_V, the residuals obtained using the equation for $\hat{GPA}_V$ for each of the 15 students and calculated by $r_V = GPA - \hat{GPA}_V$.

d. What does a negative sign in a residual indicate? Use an actual data point in your explanation.

4. Because the plot looks random (that is there does not seem to be an apparent pattern) we infer that the model is a good fit, although the size of the residuals has to be considered also.

5. **a.** The smaller the sum of the squared errors, the closer the typical prediction would be to the actual *GPA*.

   b. The mean of the *GPA* is 3.156.

   c. $\Sigma (y-\bar{y})^2 = 2.70876$.

   d. The sum of the squared residuals using the equation $1.97658 + 0.001906V = \hat{GPA}$ is 2.16807, a better fit than using just the average, $\bar{y}$.

STUDENT PAGE 94

4. Figure 6.4 contains a residual plot, the plot of ($SATV$, r_V) obtained from predicting *GPA* using *SATV*. What does the residual plot tell you about how good this model is?

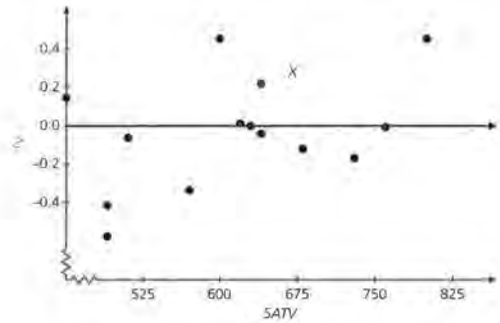


Figure 6.4

5. Remember from earlier work that you would like to have the smallest possible sum of squared residuals.

   a. Why do you want the smallest sum of squared residuals?

   b. If you did not have the *SATV* or *SATM* data and were trying to predict the *GPA* of one of the students, the best you could do would be to use the average, $\bar{y}$, as the predicted value. Find the mean *GPA*.

   c. Verify that using the mean $\bar{y}$ as the predicted *GPA* for each of the 15 students yields a sum of squared residuals of $\Sigma(y_i - \bar{y})^2 = 2.709$.

   d. What is the sum of squared residuals from the model that uses *SATV*? What conclusions can you make about the model?

The fitted model is $\hat{GPA}_V = 1.97658 + 0.001906V$. It is quite clear, however, that the model using *SATV* is not as good as it could be. There is an "error" in using the model. A more complete description of the relationship between *GPA* and *SATV* is $GPA = 1.97658 + 0.001906V + r_V = \hat{GPA}_V + r_V$. The residual term, r_V, represents the part of *GPA* that is not explained by *V*, the *SAT* verbal score, in the regression model.

6. a. This plot represents the numerical difference between the actual *GPA* and the predicted *GPA*, the residual, plotted against the student's *SAT* score in mathematics.

b. The plot looks quite random and no strong trend is obvious. It is different than the residuals versus *SATV*, however.

STUDENT PAGE 95

Adding SAT Mathematics Score

SATV is not the only thing that determines the *GPA* of a student. As you saw earlier, one of the other factors is *SATM*. You would like the residuals to be as small as possible, so possibly *SATM* can help explain r_V.

6. Figure 5 is a scatter plot of r_V versus *SATM*.

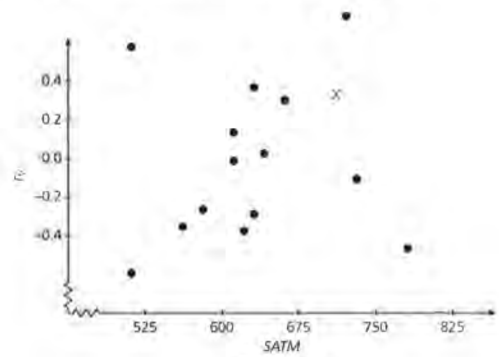


Figure 6.5

a. Explain in your own words what the plot represents.

b. What trend can you see in the plot?

7. It seems reasonable that *GPA* would depend on both *SATV* and *SATM*, probably with some error; in other words, $GPA = f(V) + h(M) + e$ for some functions f and h. So far you have $f(V) = 1.97658 + .001906V$ which predicts *GPA* with error r_V. Think of $h(M) + e$ as r_V. This indicates that r_V can be explained by M or $r_V = h(M) + e$. Thus, to find r_V, fit a regression line to $(SATM, r_V)$, and you will get the equation $\hat{r}V = -.52534 + 0.00083\ M$. This is the equation of the line shown in Figure 6.6.

7. **a.** This equation represents an attempt to use *SATM* to predict the residual from the regression of *GPA* against the *SATV*. Some of the residual error can be explained by the student's *SATM*.

b. $r_V = -0.52534 + 0.00083M$ Student answers will vary slightly due to round off.

STUDENT PAGE 96

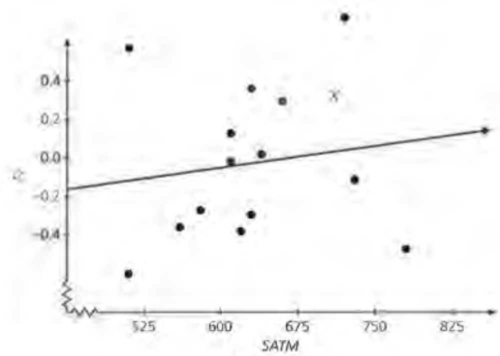


Figure 6.6

a. Explain what the equation above represents.

b. Use your own copy of the data to find the least squares regression equation for (M, r_V). How does it compare to the equation for the line in Figure 6.6?

Now, the equation for using only V to predict GPA is $\hat{GPA}_V = 1.97658 + 0.001906V$. Another statement of this relationship is that GPA depends on V and a residual amount r_V: $GPA = 1.97658 + 0.001906V + r_V$. If you replace r_V by the predicted value of r_V, from the equation $\hat{r}_V = -.52534 + 0.00083M$, you get a prediction equation for GPA in terms of V and M:

$$\hat{GPA} = 1.97658 + 0.001906V + (-.52534 + 0.00083M)$$

or

$$\hat{GPA} = 1.45124 + 0.001906V + 0.00083M$$

You found this equation by first using $SATV$ to predict GPA and then adding $SATM$ to the process, so the predicted values from this equation could be labeled $\hat{GPA}_{VM}$. The notation r_{VM} will denote the residuals from this model ($r_{VM} = GPA - \hat{GPA}_{VM}$).

What have you done so far? You began by finding the regression equation that uses $SATV$ to predict GPA. To obtain a better model, you modified the original prediction, $\hat{GPA}_V$, by

STUDENT PAGE 97

8. a. The first row represents the student's actual *GPA*, the *SATV*, the *SATM*, the predicted *GPA* using the regression equation $G\hat{P}A_{VM}$, and the residual determined by *GPA* – predicted *GPA* using the regression equation.

b.

refining the residual term. That is, $GPA = G\hat{P}A_V + r_V$ and $r_V = \hat{r}_V + r_{VM}$. Your new equation is now $GPA = G\hat{P}A_V + r_{VM}$. College *GPA* is a function of *SAT* verbal, *SAT* math, and still another "error."

8. The table below summarizes what you should have so far. $G\hat{P}A_{VM} = 1.45124 + 0.001906V + 0.00083M$ represents the prediction model using both *SATM* and *SATV*, and r_{VM} represents the residuals found using this model.

Residuals SAT Verbal, Math, GPA

Student Number	*GPA*	*SATV*	*SATM*	$G\hat{P}A_{VM}$	Residual (r_{VM})
1x	3.58	670	710	3.318	0.262
2	3.17	630	610		
3	2.31	490	510		
4	3.16	760	580		
5	3.39	450	510		
6	3.85	600	720		
7	2.55	490	560		
8	2.69	570	620		
9	3.19	620	640		
10	3.50	640	660		
11	2.92	730	780		
12	3.85	800	630		
13	3.11	640	730		
14	2.99	680	630		
15	3.08	510	610		

a. Explain what the values in the first row represent for student *X*.

b. Complete the table.

c. Find the sum of the squared residuals r_{VM}. How does it compare to the sum of the squared residuals r_V?

Can you do even better than this? That is, can you improve the model

$$G\hat{P}A_{VM} = 1.45124 + 0.001906V + 0.00083M$$

to get even better predictions?

Iterating the Process

You created a model, which will be called model *VM*, by first regressing *GPA* on *SATV*. You then computed residuals, which

Student Number	*GPA*	*SATV*	*SATM*	$G\hat{P}A_{VM}$	r_{VM}
1x	3.58	670	710	3.3176	0.2624
2	3.17	630	610	3.1583	0.0116
3	2.31	490	510	2.8085	–0.4985
4	3.16	760	580	3.3812	–0.2212
5	3.39	450	510	2.7322	0.6578
6	3.85	600	720	3.1924	0.6576
7	2.55	490	560	2.8500	–0.3000
8	2.69	570	620	3.0523	–0.3623
9	3.19	620	640	3.1642	0.0258
10	3.50	640	660	3.2189	0.2811
11	2.92	730	780	3.4900	–0.5700
12	3.85	800	630	3.4989	0.3511
13	3.11	640	730	3.2770	–0.1670
14	2.99	680	630	3.2702	–0.2802
15	3.08	510	610	2.9296	0.1504

c. $\Sigma(r_{VM})^2 = 2.10972 < \Sigma(r_V)^2$

STUDENT PAGE 98

you labeled r_V, and regressed the residuals, r_V, on *SATM*. You combined the two equations

$$GPA = 1.97658 + 0.001906V + r_V$$

and

$$\hat{r}_V = -.52534 + 0.00083M$$

to get

$$\hat{GPA}_{VM} = 1.45124 + 0.001906V + 0.00083M.$$

Another way to express the relationship is to say that *GPA* depends on *V* and *M* plus a residual, labeled r_{VM}: $GPA = 1.45124 + 0.001906V + 0.00083M + r_{VM}$. You would like to improve the model and get smaller residuals, but *V* and *M* are the only predictor variables available.

The last step was to add *SATM* to the model. How are the residuals r_{VM} related to *SATV*? Figure 6.7 is a scatter plot of $(SATV, r_{VM})$.

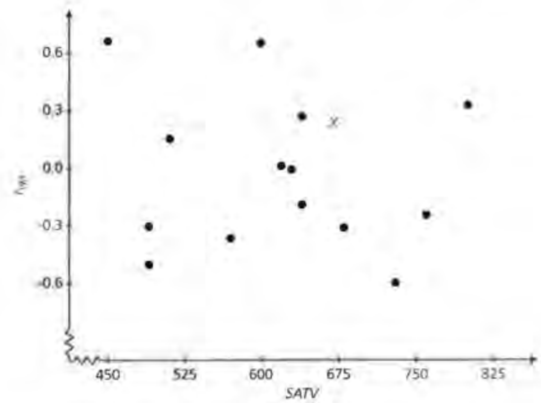


Figure 6.7

If *SATV* contributed a great deal to the determination of r_{VM}, then the pattern in this plot would be closer to linear. Apparently there is not a very strong relationship here, but you can capture the trend that exists by fitting a regression line (Figure 6.8).

STUDENT PAGE 99

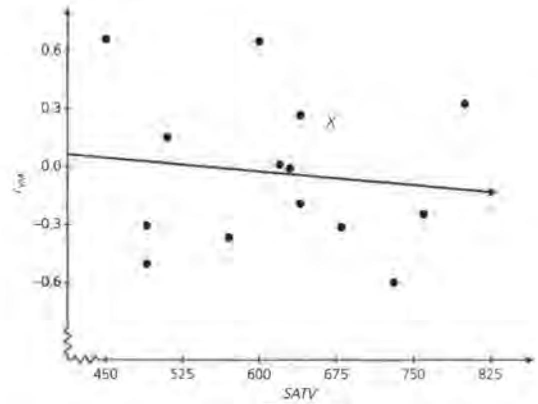


Figure 6.8

9. Find the prediction equation for this regression line, $\hat{r}_{VM}$. (Note: you are plotting *SATV* and the residuals from the previous model.)

   a. The equation to predict *GPA* can be written

   $GPA = 1.45124 + 0.001906V + 0.00083M + \hat{r}_{VM}$.

   This prediction equation will be labeled $G\hat{P}A_{VMV}$. Explain why.

   b. Write the equation for $G\hat{P}A_{VMV}$ substituting for $\hat{r}_{VM}$.

   c. Find the sum of squared residuals for model *VMV*. Compare model *VMV* to model *VM*.

You can continue the process of adding *V*, followed by *M*, followed by *V*, and so on. The best model you have so far is model *VMV*, which yields predicted values $G\hat{P}A_{VMV}$ from an expression of the form $b_0 + b_1V + b_2M$. Another statement of the situation is that *GPA* depends on *V* and *M* plus a residual, which can be labeled r_{VMV}:

$GPA = b_0 + b_1V + b_2M + r_{VMV}$

9. $\hat{r}_{VM} = 0.2135131314 + -0.000345256V$

   a. It is called $G\hat{P}A_{VMV}$ because first to get $G\hat{P}A_V$ you regressed the *GPA* against *V*. Next, the residuals of this regression were regressed against *M* forming $G\hat{P}A_{VM}$ and finally, the residuals from this regression were regressed against *V* forming $G\hat{P}A_{VMV}$. The subscripts help track the order of the regressions.

   b. $G\hat{P}A_{VMV} = G\hat{P}A_{VM} + r_{VM}$

   $G\hat{P}A_{VMV} = 1.45124 + 0.001906V + 0.00083M + 0.2135131314 - 0.000345256V$

   $G\hat{P}A_{VMV} = 1.664753 + 0.0015707V + 0.00083M$

   c. $\Sigma(r_{VMV})^2 = 2.091825767$

   $\Sigma(r_{VM})^2 = 2.10972$

   Since $\Sigma(r_{VMV})^2 < \Sigma(r_{VM})^2$, model "*VMV*" is considered better than model "*VM*."

STUDENT PAGE 100

Following the pattern of adding V, then M, then V, consider how r_{VMV} is related to M. Figure 6.9 is a scatter plot of r_{VMV} versus $SATM$.

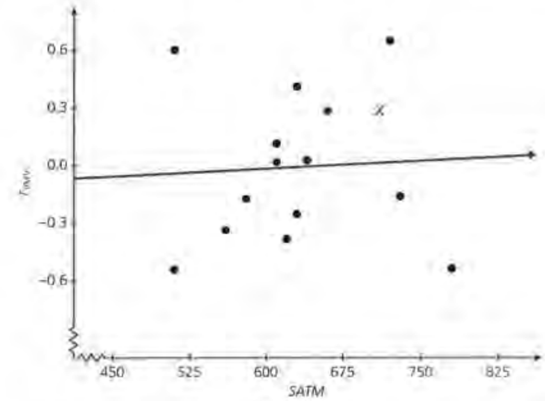


Figure 6.9

10. There is only a very weak association between r_{VMV} and $SATM$. The equation of the regression line is $\hat{r}_{VMV} = -0.15916 + 0.000252M$.

a. Write the equation for $G\hat{P}A_{VMVM}$.

b. The sum of squared residuals from this model is 2.086. Compare model $VMVM$ to model VMV.

11. Reflect back over the lesson so far.

a. Fill in the table with the corresponding equations and squared sum of residuals.

Sum of Squared Residuals Verbal, Math

	Equation	Sum of Squared Residuals
GPA, SATV	$G\hat{P}A_V =$	
GPA, SATV, SATM	$G\hat{P}A_{VM} =$	
GPA, SATV, SATM, SATV	$G\hat{P}A_{VMV} =$	
GPA, SATV, SATM, SATV, SATM	$G\hat{P}A_{VMVM} =$	

b. If you wrote down the equations for $G\hat{P}A_V$, $G\hat{P}A_{VM}$, $G\hat{P}A_{VMV}$, and $G\hat{P}A_{VMVM}$ and then continued the

10. a. $G\hat{P}A_{VMVM} = 1.664753 + 0.0015707V + 0.00083M + 0.000252M - 0.15916$

$G\hat{P}A_{VMVM} = 1.505593 + 0.0015707V + 0.001082M$

b. $\Sigma(r_{VMV})^2 = 2.091825767$

$\Sigma(r_{VMVM})^2 = 2.086$

Since $\Sigma(r_{VMVM})^2 < \Sigma(r_{VMV})^2$, $G\hat{P}A_{VMVM}$ is a better prediction equation.

11. a.

	Equation	Sum of Squared Residuals
GPA, SATV	$G\hat{P}A_V = 1.97658 + 0.001906V$	2.16807
GPA, SATV, SATM	$G\hat{P}A_{VM} = 1.45124 + 0.001906V + 0.00083M$	2.109572
GPA, SATV, SATM, SATV	$G\hat{P}A_{VMV} = 1.664753 + 0.0015707V + 0.00083M$	2.091826
GPA, SATV, SATM, SATV, SATM	$G\hat{P}A_{VMVM} = 1.505593 + 0.0015707V + 0.001082M$	2.086

b. The sum of the squared residuals would reach a limit or minimum value, and the prediction equation would change less and less until it also reaches a limit.

Practice and Applications

12. a. $G\hat{P}A_M = 1.749642072 + 0.0022205651M$

STUDENT PAGE 101

iterative process to add $G\hat{P}A_{VMVMV}$, $G\hat{P}A_{VMVMVM}$, and so on, what do you think you would see?

Summary

One process of finding a model when two variables, x_A and x_B, are used to predict a third, y, is an iterative one. You begin by finding a regression line for (x_A, y). The residuals from this model can be written as a function of the second variable, x_B. The regression line for the residuals from the plot of (x_B, r_A) can be used to adjust the original model. The residuals from (x_B, r_A) can be explained by using (x_A, r_{AB}) and this can be used to obtain a new model. The process continues, alternating between the variables and the residuals, and each time you adjust the model you had before.

Practice and Applications

Using *M* followed by *V*

12. Rather than using *SATV* first as a predictor of *GPA* and then adding *SATM*, you could use these two variables in the reverse order. First, plot (*M*, *GPA*) and fit a regression line (Figure 6.10). (Because this model uses *M*, and only *M*, to predict *GPA*, the notation $G\hat{P}A_M$ will denote the predicted values. In the same way, the residuals can be called r_M.)

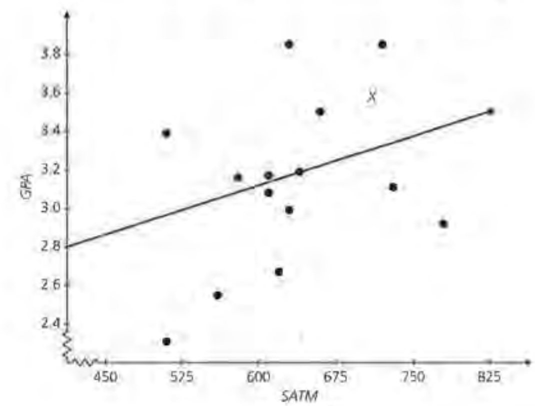


Figure 6.10

b.

STUDENT PAGE 102

a. What is the equation of the least squares regression line?

b. Use your equation to complete the table below.

Residuals *SAT* Math, *GPA*

Student Number	*GPA*	*SATM*	Predicted *GPA* $\hat{GPA}_M$	Residuals r_M
1	3.58	710		
2	3.17	610		
3	2.31	510		
4	3.16	580		
5	3.39	510		
6	3.85	720		
7	2.55	560		
8	2.69	620		
9	3.19	640		
10	3.50	660		
11	2.92	780		
12	3.85	630		
13	3.11	730		
14	2.99	630		
15	3.08	610		

A plot of r_M versus *SATM* shows no alarming pattern (Figure 6.11).

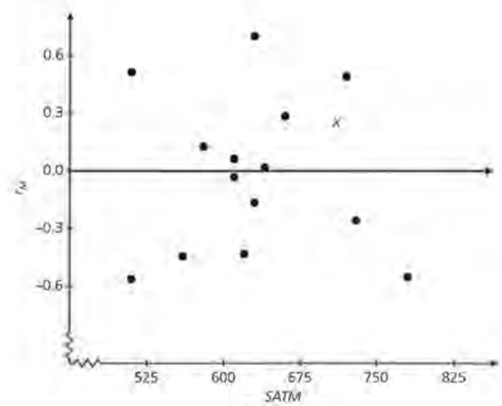


Figure 6.11

Student Number	*GPA*	*SATM*	$G\hat{P}A_M$	Residuals r_M
1*x*	3.58	710	3.3262	0.2538
2	3.17	610	3.1042	0.0658
3	2.31	510	2.8821	−0.5721
4	3.16	580	3.0376	0.1224
5	3.39	510	2.8821	0.5079
6	3.85	720	3.3484	0.5016
7	2.55	560	2.9932	−0.4432
8	2.69	620	3.1264	−0.4364

Student Number	*GPA*	*SATM*	$G\hat{P}A_M$	Residuals r_M
9	3.19	640	3.1708	0.0192
10	3.50	660	3.2152	0.2848
11	2.92	780	3.4817	−0.5617
12	3.85	630	3.1486	0.7014
13	3.11	730	3.3707	−0.2607
14	2.99	630	3.1486	−0.1586
15	3.08	610	3.1042	−0.0242

(12) c. $\Sigma(r_M)^2 = 2.2899614$
$\Sigma(r_V)^2 < \Sigma(r_M)^2$, which is evidence that *SATV* is a better prediction of *GPA* than is *SATM*.

13. a. The association is weak.

STUDENT PAGE 103

c. Find the sum of squared residuals from the model that uses *SATM*. How does it compare to the sum of squared residuals for the model that uses *SATV*?

13. You can improve the model that uses only *SATM* by adding *SATV* to it. A scatter plot of r_M versus *SATV* is in Figure 6.12.

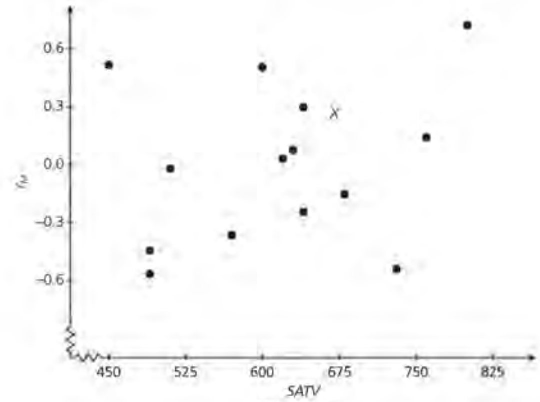


Figure 6.12

a. Describe the association in the plot.

b. $\hat{r}_M = -0.607231744 + 0.0009815276V$

c. $G\hat{P}A_{MV} = 1.142410328 + 0.0009815276V + 0.0022205651M$

STUDENT PAGE 104

b. The least squares regression line has been plotted in Figure 6.13. Find the equation.

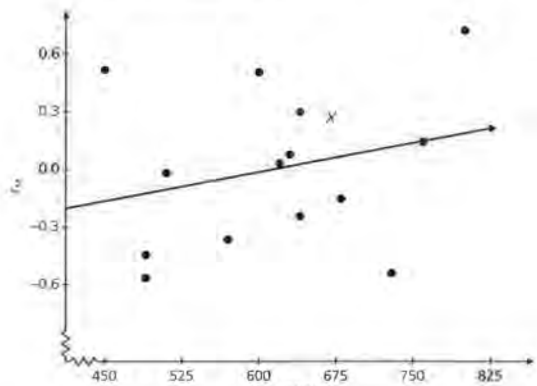


Figure 6.13

c. Another statement of this relationship is that *GPA* depends on *M* and a residual amount r_M:

$GPA = a_0 + a_1M + r_M$.

If you replace r_M by the predicted value from the equation $\hat{r}_M$, you get a prediction equation for *GPA* in terms of both *V* and *M*:

$$G\hat{P}A = a_0 + a_1M + (c_1 + a_2V)$$

or

$$G\hat{P}A = a_3 + a_1M + a_2V$$

Write the new prediction equation for $G\hat{P}A$ and use your equation to fill in the $G\hat{P}A_{MV}$ values and the corresponding residuals for each of the 15 students in the table that follows. (You found this equation by first using *M* to predict *GPA* and then adding *V* to the process, so call the predicted values from this equation $G\hat{P}A_{MV}$ and the resulting residuals r_{MV}.)

Student Number	*GPA*	*SATV*	*SATM*	$G\hat{P}A_{MV}$	r_{MV}
1*x*	3.58	670	710	3.3766	0.2034
2	3.17	630	610	3.1153	0.0547
3	2.31	490	510	2.7558	–0.4458
4	3.16	760	580	3.1763	–0.0163
5	3.39	450	510	2.7166	0.6734
6	3.85	600	720	3.3301	0.5199
7	2.55	490	560	2.8669	–0.3169
8	2.69	570	620	3.0786	–0.3886

Student Number	*GPA*	*SATV*	*SATM*	$G\hat{P}A_{MV}$	r_{MV}
9	3.19	620	640	3.1721	0.0179
10	3.50	640	660	3.2362	0.2638
11	2.92	730	780	3.5910	–0.6710
12	3.85	800	630	3.3266	0.5234
13	3.11	640	730	3.3916	–0.2816
14	2.99	680	630	3.2088	–0.2188
15	3.08	510	610	2.9975	0.0825

(13) d. $\Sigma(r_{MV})^2 = 2.1466458$

$\Sigma(r_{VM})^2 < \Sigma(r_{MV})^2$

STUDENT PAGE 105

Residuals SAT Math, Verbal, GPA

Student Number	GPA	SATV	SATM	$\hat{GPA}_{MV}$	Residual (r_{MV})
1	3.58	670	710		
2	3.17	630	610		
3	2.31	490	510		
4	3.16	760	580		
5	3.39	450	510		
6	3.85	600	720		
7	2.55	490	560		
8	2.69	570	620		
9	3.19	620	640		
10	3.50	640	660		
11	2.92	730	780		
12	3.85	800	630		
13	3.11	640	730		
14	2.99	680	630		
15	3.08	510	610		

d. What is the sum of the squared residuals for r_{MV}? How does it compare to the sum of squared residuals for r_{VM}?

14. You used *M* and then added *V* to the model. Now use *M* again by plotting r_{MV} against *M* and fitting a regression line.

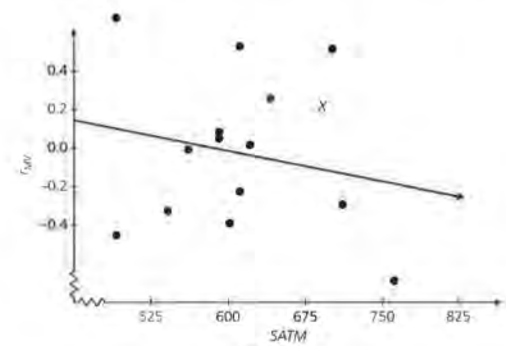


Figure 6.14

STUDENT PAGE 106

a. The equation of the regression line is $\hat{r}_{MV} = 0.453533 - 0.000716M$. Find the refined prediction equation for *GPA* by substituting $0.453533 - 0.000716M$ for r_{MV} in the equation

$$GPA = 1.142410328 + 0.0009815276V + 0.0022205651M + r_{MV}.$$

b. Why does it make sense to call the predicted values from this equation $G\hat{P}A_{MVM}$?

c. The sum of squared residuals for model *MVM* is 2.103. How do these compare to the sum of squared residuals for model *MV*?

15. Fill in the chart using the results of problems 12, 13, and 14.

Sum of Squared Residuals Math Verbal

	Equation	Sum of Squared Residuals
GPA, SATM	$G\hat{P}A_M =$	
GPA, SATM, SATV	$G\hat{P}A_{MV} =$	
GPA, SATM, SATV, SATM	$G\hat{P}A_{MVM} =$	

What observations can you make?

16. The plots below are of the residuals for each successive iteration. What observation can you make? Does the development of the model support your observation? If so, how?

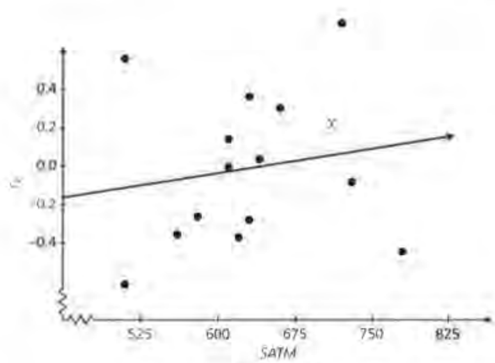


Figure 6.15

14. a. $G\hat{P}A_{MVM} = 1.142410328 + 0.0009815276V + 0.0022205651M + 0.453533 - 0.000716M$

$G\hat{P}A_{MVM} = 1.59594333 + 0.0009815276V + 0.0015045651M$

b. The order of the regression is first on *M* then on *V* followed by a regression on *M*.

c. $\Sigma(r_{MV})^2 = 2.1466458$ and $\Sigma(r_{MVM})^2 = 2.10309$.
Since $\Sigma(r_{MVM})^2 < \Sigma(r_{MV})^2$, model "*MVM*" gives better predictions, overall, than does model "*MV*".

15. (See chart below).
The sum of the squared residuals is becoming smaller and approaching a limit, implying the predictor equation is getting better.

	Equation	Sum of Squared Residuals
GPA, SATM	$G\hat{P}A_M = 1.749642072 + 0.0022205651M$	2.2899614
GPA, SATM, SATV	$G\hat{P}A_{MV} = 1.142410328 + 0.0009815276V + 0.0022205651M$	2.1466458
GPA, SATM, SATV, SATM	$G\hat{P}A_{MVM} = 1.59594333 + 0.0009815276V + 0.0015045651M$	2.10309

16. The regression line appears to become more and more horizontal, implying that the influence of that variable is becoming less.

17. If there are two variables involved in making a prediction, create a "best" prediction equation by doing the following.

1. Arbitrarily choose one variable to begin the prediction.
2. Regress against that item and determine the residuals.
3. Regress the residuals from the first regression against the second variable and determine the new set of residuals.
4. Regress these new residuals against the first variable and determine the residuals.
5. Iteration of this process will result in a single "best" prediction equation, where "best" minimizes the sum of squared residuals.

STUDENT PAGE 107

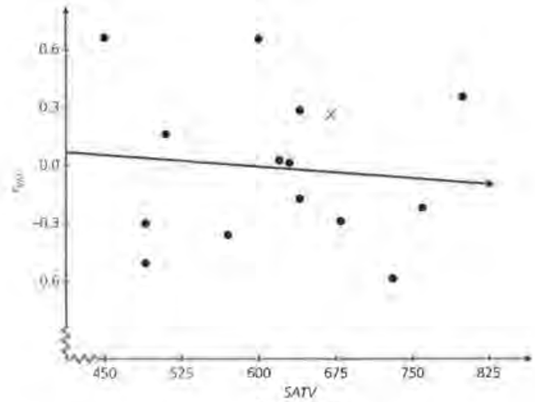


Figure 6.16

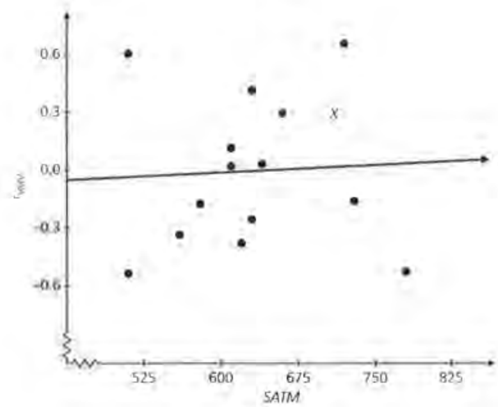


Figure 6.17

17. Write a summary of the process that can be used to find a model if two variables are involved in making a prediction.

LESSON 7

Comparing Order in Regression

Materials: none
Technology: graphing calculator
Pacing: 1 class period

Overview

In this lesson, students investigate the effect of order in the regression process when there is more than one independent variable. The question is whether beginning with one variable or the with the other will eventually yield the same model. The answer is found by studying the residuals. A scatter plot of the residuals after the first regression (*SATV*, *GPA*) and (*SATM*, *GPA*) graphically illustrates how the two models differ for each data point. A plot of the residuals from the second regression in the iterative process shows that overall, the differences are becoming smaller, and the scatter plot is becoming more linear. A plot of the residuals after the third regression shows even more clearly how the two sets of residuals are approaching the line $y = x$, which indicates that the two models are producing the same results. Using a table to record the sum of squared residuals for each regression shows the sums are approaching each other; after two iterations the sum of squared residuals are the same to the tenths place, and the argument continues to improve. As a final point, the sum of the squared residuals from (*SATV* + *SATM*, *GPA*) is compared to the table and found to be larger than those from the iterative process. The conclusion is that the model obtained using either order will produce the same result, the least squares multiple regression model.

Teaching Notes

It will be helpful to have students recall the data they have stored from their work in Lesson 6 to make the plots and investigate the sum of squared residuals in this lesson. Even if only one student in a group has access to the data, it will allow students to focus on the mathematics instead of trying to recreate the data. You may want to put the plots of the two sets of residuals on an overhead to facilitate class discussion.

The fact that both orders converge to the same model is quite impressive and should be clearly demonstrated for students. It illustrates that the mathematics underlying the process is valid and powerful.

Technology

Technology is essential for this lesson to retrieve stored information, to organize sum of squared residuals, and to make plots from those data.

Follow-Up

You might ask students to think of examples where an iterative process converges to a result. Ask students who have taken calculus to see if they can find a parallel between the process in the lesson and the process of finding area using Riemann sums.

STUDENT PAGE 108

LESSON 7

Comparing Order in Regression

Does it make a difference if you start finding a model using either *SATM* or *SATV*?

Which variable should you use first, *SATM* or *SATV*? How would you choose?

What would you do if you had more variables to use in predicting *GPA*?

OBJECTIVES

Investigate the impact of order in the regression process.

Recognize what will occur as the process continues.

You have used *SATV* and *SATM* to make predictions of *GPA* in two different ways. One approach was to start with *SATV* as a predictor and then to take *SATM* into consideration; this led to the list of predictions labeled $\hat{GPA}_V$, $\hat{GPA}_{VM}$, and so on. The second approach was to reverse the order, first using *SATM* as a predictor and then adding *SATV* to the model; you labeled the predicted values from this model $\hat{GPA}_M$, $\hat{GPA}_{MV}$, and so on. In this lesson you will compare the two different approaches by looking at the coefficients, the graphs, and the sums of the squared residuals for each iteration.

INVESTIGATE

Comparing Two Models

The equations to predict *GPA* if you began with *SATV* and if you began with *SATM* were:

$\hat{GPA}_V = 1.97658 + 0.001906V$ and
$\hat{GPA}_M = 1.74964 + 0.002221M$.

The coefficients in the two models are different, but how different are the results the models produce? One way to compare the two is to consider the predictions they give for a single

STUDENT PAGE 109

student, such as student number 1. Student number 1 was plotted with an X in all of the scatter plots. Notice that in both of the scatter plots in Figure 7.1, student number 1 is above most of the other points. In particular, student number 1 has a positive residual in each case.

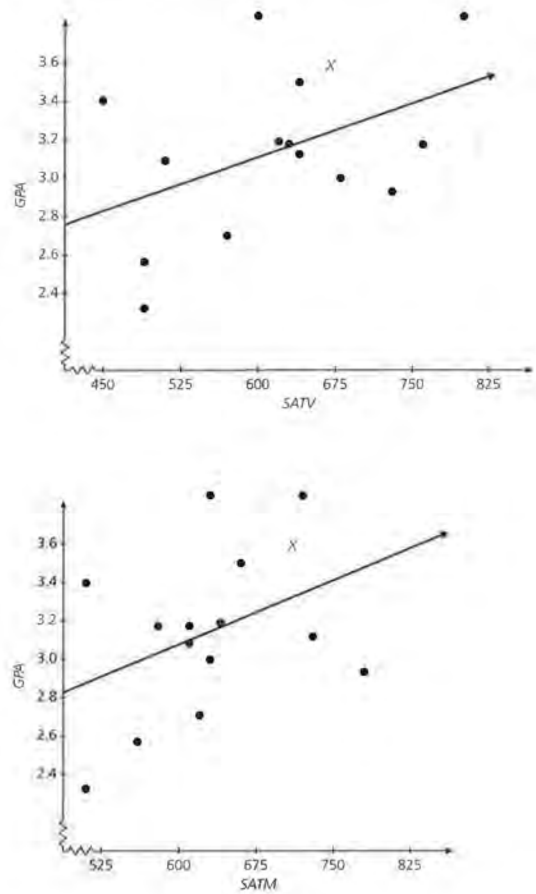


Figure 7.1

STUDENT PAGE 110

Solutions

1. The predicted values are: $\hat{GPA}_V = 3.2536$ and $\hat{GPA}_M = 3.32655$.

a.

Discussion and Practice

1. What are the predicted *GPA* values from the two models for student number 1?

a. Complete the table with the predicted values for each student.

GPA Verbal, GPA Math

Student Number	*GPA*	*SATV*	*SATM*	$\hat{GPA}_V$	$\hat{GPA}_M$
1x	3.58	670	710		
2	3.17	630	610		
3	2.31	490	510		
4	3.16	760	580		
5	3.39	450	510		
6	3.85	600	720		
7	2.55	490	560		
8	2.69	570	620		
9	3.19	620	640		
10	3.50	640	660		
11	2.92	730	780		
12	3.85	800	630		
13	3.11	640	730		
14	2.99	680	630		
15	3.08	510	610		

b. If GPA_V and GPA_M were equally good at predicting *GPA*, describe the plot of $(\hat{GPA}_V, \hat{GPA}_M)$.

c. Make a plot of $(\hat{GPA}_V, \hat{GPA}_M)$. How are the predictions from the two models related to each other?

Consider models *MV* and *VM*. You constructed model *VM* by first using *SATV* to predict *GPA*, computing the residuals r_V, and then regressing r_V on *M*. You constructed model *MV* by first using *SATM* to predict *GPA*, computing the residuals, r_M, and then regressing r_M on *V*. The two equations that resulted were

$$\hat{GPA}_{VM} = 1.45124 + 0.001906V + 0.00083M$$

and

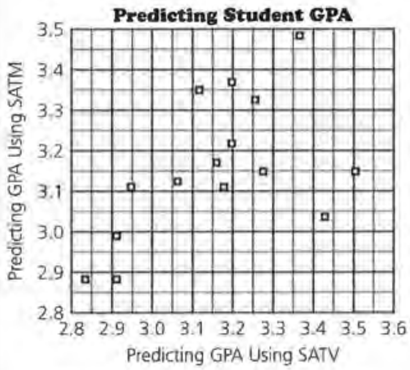
$$\hat{GPA}_{MV} = 1.14241 + 0.0009815276V + 0.0022205651M$$

The order in which you add *V* and *M* to the model affects the coefficients in the model. That is, the coefficients in the model

Student Number	*GPA*	*SATV*	*SATM*	$\hat{GPA}_V$	$\hat{GPA}_M$
1x	3.58	670	710	3.2536	3.3262
2	3.17	630	610	3.1774	3.1042
3	2.31	490	510	2.9105	2.8821
4	3.16	760	580	3.4251	3.0376
5	3.39	450	510	2.8343	2.8821
6	3.85	600	720	3.1202	3.3484
7	2.55	490	560	2.9105	2.9932
8	2.69	570	620	3.0630	3.1264
9	3.19	620	640	3.1583	3.1708
10	3.50	640	660	3.1964	3.2152
11	2.92	730	780	3.3680	3.4817
12	3.85	800	630	3.5014	3.1486
13	3.11	640	730	3.1964	3.3707
14	2.99	680	630	3.2727	3.1486
15	3.08	510	610	2.9486	3.1042

b. The plot would be a straight line because both equations would produce the same estimates of *GPA*.

c.



There is a positive association although not a perfectly linear relationship.

2. The predicted values are: $G\hat{P}A_{VM} = 3.3176$ and $G\hat{P}A_{MV} = 3.3766$.

a. $G\hat{P}A_M - G\hat{P}A_V = 0.07295$ and $G\hat{P}A_{MV} - G\hat{P}A_{VM} = 0.0590$

The two predictions for $G\hat{P}A_{MV}$ and $G\hat{P}A_{VM}$ are closer to the same value or the same prediction.

b. If the models were equivalent, then the two predictions would be the same.

c. The two planes are not parallel. Students justifications will vary. Some may refer to the ratio between the coefficients (much like that of the weights from the first lessons) as different; others may use the intercepts and traces to discuss why the planes are not parallel. Others may find a point on both planes.

3. **a.** The line is where the predicted values from using model "*VM*" would equal the predicted values using the model "*MV*."

b. There is a clear linear trend. The points are close to the line $G\hat{P}A_{MV} = G\hat{P}A_{VM}$. This tells you the models give predictions that are very close.

STUDENT PAGE 111

VM are not exactly the same as the coefficients in the model *MV*.

2. What are the predicted *GPA*s for student 1 using models *MV* and *VM*?

- **a.** How do these results compare to those for model *V* and model *M*?
- **b.** If the models were equivalent, what do you know about the two predictions?
- **c.** You learned in the first part of this module that the equations for $G\hat{P}A_{VM}$ and $G\hat{P}A_{MV}$ would represent planes. Decide whether the two planes are parallel or intersect and justify your decision.

Graphs

It is not sufficient to compare the models on only one student. Just as you did in problem 1, consider the ordered pair $(G\hat{P}A_{VM}, G\hat{P}A_{MV})$ for each of the 15 students.

3. Figure 7.2 is a plot of those pairs of predictions, with student number 1 again plotted using an *X*.

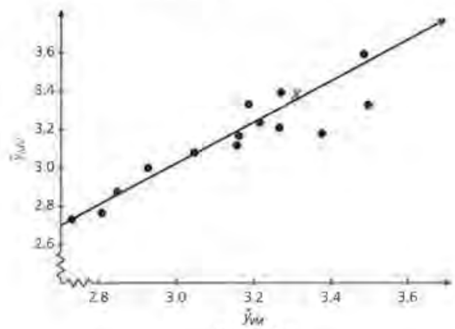


Figure 7.2

- **a.** What does the line in Figure 7.2 represent?
- **b.** Describe the trend in the plot. What does this tell you about the relationship between the *GPA*s predicted by the two models?

(3) **c.** The plot would be the straight line $G\hat{P}A_{MV} = G\hat{P}A_{VM}$.

4. **a.** $G\hat{P}A_{VMV} = 3.3064$ for Student 1.

$G\hat{P}A_{MVM} = 3.32181$ for Student 1.

$G\hat{P}A_{MVM} - G\hat{P}A_{VMV} = 0.01541$

b. $G\hat{P}A_{MV} - G\hat{P}A_{VM} = 0.0590$; therefore, the predictions using "*VMV*" and "*MVM*" are more similar.

c. The trend is more linear than $(G\hat{P}A_{VM}, G\hat{P}A_{MV})$. The points are closer to the line $G\hat{P}A_{MVM} = G\hat{P}A_{VMV}$.

d. The points in the upper right for higher *GPA's* are becoming closer and approaching the line of equality.

STUDENT PAGE 112

c. If the two models made exactly the same prediction for each of the 15 *GPA*s, describe what the plot would look like.

4. The prediction equations for $G\hat{P}A_{VMV}$ and $G\hat{P}A_{MVM}$ are similar but have different coefficients:

$$G\hat{P}A_{VMV} = 1.664753 + 0.0015707V + 0.00083M$$

and

$$G\hat{P}A_{MVM} = 1.59594333 + 0.0009815276V + 0.0015045651M$$

a. Compare the predictions that the models *VMV* and *MVM* give for the *GPA* of student number 1.

b. How do these predictions compare to those using model *VM* and model *MV*?

c. Figure 3 is a scatter plot of $(G\hat{P}A_{VMV}, G\hat{P}A_{MVM})$ for each of the 15 students. Describe the trend in the plot.

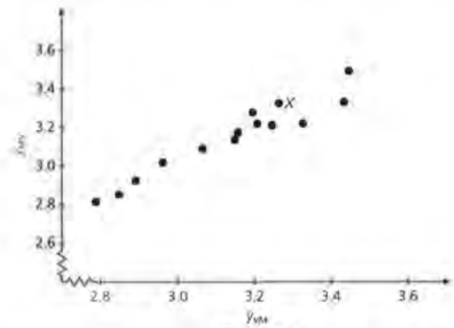


Figure 7.3

d. Look carefully at the plot you made for problem 1 and at Figure 7.2 and Figure 7.3. What observation can you make?

STUDENT PAGE 113

Residuals

Another way to compare the two models is to consider the residuals from each.

5. This table has the predicted *GPA* for student 1 for each of the models.

Residuals Student 1

Student	GPA	$\hat{GPA}_V$	$\hat{GPA}_M$	$\hat{GPA}_{VM}$	$\hat{GPA}_{MV}$	$\hat{GPA}_{VMV}$	$\hat{GPA}_{MVM}$
1	3.58	3.2536	3.3262	3.3176	3.3766	3.3064	3.3218

a. Find and compare the residuals for each case:

r_V, r_M
r_{VM}, r_{MV}
r_{VMV}, r_{MVM}

b. What conclusions can you draw?

c. Why is investigating the sum of squared residuals a better way to compare the models than using an individual student?

6. The sum of squared residuals for each of the models, except for the model *MVMV*, is given in the following table.

Comparing Sum of Squared Residuals

Model	Sum of Squared Residuals for Model	Model	Sum of Squared Residuals for Model
V	2.168	*M*	2.290
VM	2.110	*MV*	2.147
VMV	2.092	*MVM*	2.103
VMVM	2.086	*MVMV*	

a. Back in Lesson 6, you learned that if you had no information at all about SAT scores and were trying to predict GPA using only the mean GPA, the sum of the squared residuals for the mean would be 2.709. What does this say about using either of the two models to begin with?

b. Describe the trend in the sum of squared residuals for each group of models in the table.

c. Use the information above to comment on the statement: The predictions from the model that uses only *M* are about as accurate, overall, as the predictions generated by the model that uses only *V*. The equations for

5. a.

r_V	r_M	r_{VM}	r_{MV}	r_{VMV}	r_{MVM}
0.3264	0.2538	0.2624	0.2034	0.2736	0.2582

b. The difference in the residuals goes from 0.0726 to 0.0590 to 0.0154. It is hard to make a conclusion based on those three numbers, although the difference is least for the last model.

c. Comparing the sum of the squared residuals will use the entire sample and include all outliers or extreme cases. You cannot generalize based on one example. You care about how a model works overall, not just for one person.

6. a. Either model "*V*" or model "*M*" is superior to simply using the mean *GPA* to make the prediction. The sum of squared residuals in both models is smaller than 2.709, the sum of squared residuals using the mean.

b. Within each group, the sum of the squared residuals is getting smaller.

c. This is a reasonable statement. The sums of the squared residuals are similar.

STUDENT PAGE 114

(6) d. $\Sigma(r_M)^2 - \Sigma(r_V)^2 = 0.122$; $\Sigma(r_{MV})^2 - \Sigma(r_{VM})^2 = 0.037$; $\Sigma(r_{MVM})^2 - \Sigma(r_{VMV})^2 = 0.011$.

e. You might expect it to be close to the value of $\Sigma(r_{VMVM})^2$, which is 2.086.

7. a.

the two models are rather different, but overall, each model makes predictions that are reasonable.

d. Compare the difference between the sum of squared residuals for model *V* and model *M*; model *VM* and model *MV*; model *VMV* and model *MVM*.

e. What do you think the sum of squared residuals is for model *MVMV*?

Conclusion

If you continue to repeat the process outlined in Lesson 6 using either order, you end up with the following prediction equation:

$$\hat{GPA} = 1.52885 + 0.00141V + 0.001192M$$

This is the *least-squares multiple regression model* for predicting *GPA* using *SATV* and *SATM*.

7. Use the multiple regression model to predict the *GPA* for each of the students.

a. What is the sum of squared residuals for this model? How does it compare to the sum of squared residuals from the earlier models?

b. If you were a college entrance administrator at the school from which these data came and a student applied for entrance with an *SATV* of 450 and an *SATM* of 480, what would you tell that student?

8. Think back to the original problem. How can you use both *SATV* and *SATM* to predict college *GPA*? You might have wondered what would happen if you had added the two values and worked with the total *SAT* score rather than going through this entire process. That is, why not add *SATM* to *SATV* to get *SAT TOTAL* and use *SAT TOTAL* to predict *GPA*?

a. Do you think that this will give better predictions then you found with the models above? Use the information in the table that follows to find the linear regression model for (*SAT TOTAL*, *GPA*) and calculate the sum of squared residuals to help you answer the question.

b. Think about different variables that could be used to predict *GPA*. Would it always make sense to add two predictor variables together?

Student Number	*GPA*	*SATV*	*SATM*	$\hat{GPA}$	Residuals
1*x*	3.58	670	710	3.3199	0.26013
2	3.17	630	610	3.1443	0.02573
3	2.31	490	510	2.8277	–0.5177
4	3.16	760	580	3.2918	–0.1318
5	3.39	450	510	2.7713	0.61873
6	3.85	600	720	3.2331	0.61691
7	2.55	490	560	2.8873	–0.3373
8	2.69	570	620	3.0716	–0.3816

Student Number	*GPA*	*SATV*	*SATM*	$\hat{GPA}$	Residuals
9	3.19	620	640	3.1659	0.02407
10	3.50	640	660	3.2180	0.28203
11	2.92	730	780	3.4879	–0.5679
12	3.85	800	630	3.4078	0.44219
13	3.11	640	730	3.3014	–0.1914
14	2.99	680	630	3.2386	–0.2486
15	3.08	510	610	2.9751	0.10493

$\Sigma r^2 = 2.0841$

The sum of the squared residuals is smaller than any of the previous sums.

b. The student's predicted *GPA* would be 2.736, but there are other factors such as work habits, motivation, and so on, that are not used in the model that could make a difference.

8. **a.** $G\hat{P}A = 1.502886702 + .001320378(SAT)$; $r = .4797732941$

Student Number	*GPA*	*SATV*	*SATM*	Total *SAT*	Predicted *GPA*	Residuals
1*x*	3.58	670	710	1380	3.325	0.25499
2	3.17	630	610	1240	3.1402	0.02984
3	2.31	490	510	1000	2.8233	–0.5133
4	3.16	760	580	1340	3.2722	–0.1122
5	3.39	450	510	960	2.7704	0.61955
6	3.85	600	720	1320	3.2458	0.60421
7	2.55	490	560	1050	2.8893	–0.3393
8	2.69	570	620	1190	3.0741	–0.3841
9	3.19	620	640	1260	3.1666	0.02344
10	3.50	640	660	1300	3.2194	0.28062
11	2.92	730	780	1510	3.4967	–0.5767
12	3.85	800	630	1430	3.3910	0.45897
13	3.11	640	730	1370	3.3118	–0.2018
14	2.99	680	630	1310	3.2326	–0.2426
15	3.08	510	610	1120	2.9817	0.09829

Σ(squared residuals) = 2.0853. This does not give better predictions than the equation in problem 7.

b. The variables might be in different units and not compatible for adding.

(8) c. Answers will vary; sample: *SATV* and *SATM* deal with different aspects of a student's ability.

Practice and Applications

9. Model beginning with ROOM:

$\hat{T}_R = -4026.25 + 2.65757R$

$\Sigma(r_R)^2 = 48613824$

$\hat{T}_{RN} = -4026.25 + 2.65757R - 4060.7 + 49.9471N = -8086.95 + 2.65757R + 49.9471N$

$\Sigma(r_{RN})^2 = 40156500$

$\hat{T}_{RNR} = -8086.95 + 2.6575R + 49.9471N + 493.861 - 0.097823R = -7593.089 + 2.559747R + 49.9471N$

$\Sigma(r_{RNR})^2 = 40066899$

$\hat{T}_{RNRN} = -7593.089 + 2.559747R + 49.9471N - 43.0213 + 0.529168N = -7636.1103 + 2.559747R + 50.476268N$

$\Sigma(r_{RNRN})^2 = 40065949$

Model beginning with NEED MET:

$\hat{T}_N = 4161.04 + 64.323N$

$\Sigma(r_N)^2 = 100717908$

$\hat{T}_{NR} = 4161.04 + 64.323N - 12780.7 + 2.53159R = -8619.66 + 64.323N + 2.53159R$

$\Sigma(r_{NR})^2 = 40708521$

$\hat{T}_{NRN} = -8619.66 + 64.323N + 2.53159R + 1113.36 - 13.6945N = -7506.3 + 50.6285N + 2.53159R$

$\Sigma(r_{NRN})^2 = 40072747$

$\hat{T}_{NRNR} = -7506.3 + 50.6285N + 2.53159R - 135.406 + 0.026821R = -7641.706 + 50.6285N + 2.558411R$

$\Sigma(r_{NRNR})^2 = 40066011$

The results are converging.

STUDENT PAGE 115

Sum of Math, Verbal

Student Number	*GPA*	*SATV*	*SATM*	*SAT TOTAL*	Predicted *GPA*	Residuals
1	3.58	670	710			
2	3.17	630	610			
3	2.31	490	510			
4	3.16	760	580			
5	3.39	450	510			
6	3.85	600	720			
7	2.55	490	560			
8	2.69	570	620			
9	3.19	620	640			
10	3.50	640	660			
11	2.92	730	780			
12	3.85	800	630			
13	3.11	640	730			
14	2.99	680	630			
15	3.08	510	610			

c. Why is it better to use *SATM* and *SATV* as two separate predictors rather than combining them into *SAT TOTAL*?

Summary

You can find a prediction equation in two variables by beginning with one variable, taking the residuals from the model, regressing them on the other variable, and continuing to repeat the process. The equations are slightly different, depending on which variable you choose to begin the process. As you compare the *GPA* predictions for each step, however, the differences begin to lessen. The plots of predicted values from the iterations using different orders of the variables $\hat{GPA}_{VM}$ and $\hat{GPA}_{MV}$ converge to the straight line $y = x$. The sums of the squared residuals for each case converge toward a common sum. The equation you approach is called the *least squares multiple regression model.*

Practice and Applications

9. Work with a partner on the following data about colleges. One of you should begin with *room and board cost* (R) and use it to predict *tuition and fees*. The other should begin with *% of need met* (N) and use it to predict *tuition and fees*. Use the process of iteration to generate $\hat{T}_{RNRN}$ and $\hat{T}_{NRNR}$ and compare your results. (Note that *% of need met* is a measure of financial aid at the college.)

10. The model beginning with W is $\hat{P}_W = 4.01471 + 2.69118W$; $\Sigma(r_W)^2 = 50.1471$.

$\hat{P}_{WL} = 4.01471 + 2.69118W - 5.30402 + 1.68764L = -1.28931 + 2.69118W + 1.68764L$; $\Sigma(r_{WL})^2 = 7.83193$.

$\hat{P}_{WLW} = 1.28931 + 2.69118W + 1.68764L + 1.91632 - 0.583229W = 0.62701 + 2.107951W + 1.68764L$; $\Sigma(r_{WLW})^2 = 1.22319$.

$\hat{P}_{WLWL} = 0.62701 + 2.107951W + 1.68764L - 0.828377 + -0.263575L = -0.201367 + 2.107951W + 1.951215L$; $\Sigma(r_{WLWL})^2 = 0.191036$.

$\hat{P}_{WLWLW} = -0.201367 + 2.107951W + 1.951215L + 0.29929 + -0.091088W = 0.097923 + 2.016863W + 1.951215L$; $\Sigma(r_{WLWLW})^2 = 0.029836$.

$\hat{P}_{WLWLWL} = -0.097923 + 2.016863W + 1.951215L + 0.129375 + -0.041165L = -0.031452 + 2.016863W + 1.99238L$; $\Sigma(r_{WLWLWL})^2 = 0.00466$.

STUDENT PAGE 116

After 4 iterations the model is close to $\hat{P} = 2W + 2L$, and the sum of squared residuals is close to zero. After 6 iterations the model is very close to the limiting model, and the sum of squared residuals is very close to zero.

11. There must be a linear relationship between the three variables, and one of the variables must be influenced by the other two. It is called the *least squares* regression model because the model minimizes the sum of squared residuals.

12. The residuals are considered to be the part of y, the response variable, not explained by the immediate regression. You regress those residuals against the other variable, and the resulting residuals are now considered to be the amount not influenced by this variable. Since you are considering only two variables, you regress the new residuals against the first variable, and so on.

13. You would begin by regressing the predicted value against any of the three variables. The residuals obtained from this first regression are now regressed against one of the two remaining variables. The residuals obtained from this second regression are now regressed against the third variable. The residuals obtained from this third regression are now regressed against the first variable once more. This iteration continues.

Best College Buys

School	Tuition and Fees ($)	Room and Board ($)	% of Need Met
California Inst. of Tech.	17,586	6,620	100
New College of U. of South Florida	7,950	3,847	93
Northwest Missouri State	3,975	3,330	80
Rice University	12,034	5,900	100
State U. of NY at Binghamton	8,679	4,654	62
State U. of NY at Albany	8,856	4,836	80
Spelman College	8,875	5,890	38
Trenton State College	6,658	5,650	90
U. of Illinois at Urbana/ Champaign	9,130	4,408	75
U. of North Carolina at Chapel Hill	10,162	5,350	95

Source: *Money Guide*, 1996 Edition

10. The following table gives the width, length, and perimeter of each of seven rectangles. Suppose you did not know that $P = 2W + 2L$, and you wanted to determine the relationship between P, W, and L by using regression. Use the iterative process to find $\hat{P}_{WLWL}$, etc. How many iterations does it take until the sum of squared residuals gets close to zero and the fitted model gets close to $P = 2W + 2L$?

W = width	L = length	P = perimeter
2	2	8
3	5	16
4	4	16
1	1	4
5	5	20
2	3	10
6	2	16

11. What are some underlying assumptions that are necessary if a least squares multiple regression model is to be a good model? Why do you think the model is called the least squares multiple regression model?

12. Explain the role of residuals in the process of finding the least squares multiple regression model.

13. Make a conjecture about how you might find a model if you had three variables available to determine a prediction equation.

LESSON 8

Matrices and Multiple Regression

Materials: *Lesson 8 Quiz*
Technology: graphing calculator
Pacing: 1 class period

Overview

The process used in Lessons 6 and 7 to find a multiple regression model is long and complicated. Using matrices and writing the regression problem in matrix form leads to a matrix solution that quickly and easily reveals the multiple regression equation. In addition, the process is general and can be used to find a regression equation for any number of independent variables. The matrix method can also be extended to fitting a quadratic equation with both a quadratic and linear term to a set of data or fitting a polynomial of any degree to the data.

The justification that this model does produce the least sum of squared residuals is also a function of operating with matrices and is demonstrated as the final part of the investigation. The exercises provide practice using the matrix formula.

Teaching Notes

Students have to recall the procedures they learned in Lesson 5 for writing a regression equation in the form of a matrix system and the facts about matrices and matrix operations from Lesson 1. Using a regression equation yields a matrix equation that can be "solved" for the missing coefficients of the regression equation. The solution, however, depends on finding an inverse for the matrix that contains the independent variables, a matrix that is not square. An "inverse" is created by using the product of a matrix and its transpose to produce a square matrix that can have an inverse. It is important to recognize that this solution does not actually "solve" the system but rather provides as close a solution as possible, in fact the estimate that minimizes the sum of squared residuals.

Technology

Technology that allows students to manipulate matrices is essential for this lesson.

Follow-Up

Students should have some experience deciding whether variables actually contribute to the regression equation. You might provide them with several choices for independent variables and ask them to decide which, if any, seems to be a factor in predicting some outcome. Potential areas for investigation are sports: points scored, rebounds, experience, predicting salary, or factors involved in cost of something such as size, quality, number of competitors.

STUDENT PAGE 117

LESSON 8

Matrices and Multiple Regression

How can you find a model more quickly? Is there a method that would be more efficient?

Is there any way that another mathematical approach for finding a model can help?

The iterative process outlined in Lessons 6 and 7 seems to be a long, involved process. Is there an easier way to find a good model using both variables? Is there a mathematical technique that can help? What happens if there are more than two variables that can be used to find a prediction equation? In this lesson you will investigate a shortcut formula using matrices that will enable you to find a model with any number of variables.

OBJECTIVES

Use matrices to represent the process of regression.

Justify the process mathematically.

Generalize the formula to many dimensions.

INVESTIGATE

Using Matrices in Regression

In Lesson 5 you learned to write a system of matrices to represent a regression rule. Equations of the form $aw + bx = c$ were written as

$$[w\ x] \cdot \begin{bmatrix} a \\ b \end{bmatrix} = [c]$$

When used with a least squares regression equation, the matrix $[w\ x]$ became an $n \times 2$ matrix that contained a column vector of 1s and a column of the x-values; [c] became an $n \times 1$ column vector with the predicted y-values. $\begin{bmatrix} a \\ b \end{bmatrix}$ was the coefficient matrix determined by the least squares regression equation. So the matrix equation could be written:

STUDENT PAGE 118

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 & y_1 \\ 1 & y_2 \\ \vdots & \vdots \\ 1 & y_n \end{bmatrix}$$

You have been considering prediction equations of the form

$$\hat{y} = b_0 + b_1x_1 + b_2x_2.$$

For example, the model labeled *VM* is

$$G\hat{P}A_{VM} = 1.45124 + 0.001906V + 0.00083M.$$

In this model, $b_0 = 1.45124$, $b_1 = 0.001906$, and $b_2 = 0.00083$.

You can rewrite the model as

$$b_0(1) + b_1V + b_2M = G\hat{P}A.$$

to make explicit the fact that b_1 is multiplied by V, b_2 is multiplied by M, and b_0 is multiplied by 1. In matrix form, this is

$$[1\ V\ M] \cdot \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = [G\hat{P}A].$$

A matrix formulation of this model for predicting *GPA*s is $Xb = \hat{Y}$.

$$X = \begin{bmatrix} 1 & 670 & 710 \\ 1 & 630 & 610 \\ 1 & 490 & 510 \\ 1 & 760 & 580 \\ 1 & 450 & 510 \\ 1 & 600 & 720 \\ 1 & 490 & 560 \\ 1 & 570 & 620 \\ 1 & 620 & 640 \\ 1 & 640 & 660 \\ 1 & 730 & 780 \\ 1 & 800 & 630 \\ 1 & 640 & 730 \\ 1 & 680 & 630 \\ 1 & 510 & 610 \end{bmatrix}$$

The first column in X is a column of 1s, the second column in X holds the values of *SATV*, and the third column holds the values of *SATM*.

The matrix b is a 3×1 matrix that holds the parameters b_0, b_1, and b_2 for the equation $G\hat{P}A = b_0 + b_1V + b_2M$. For model *VM*,

$$b = \begin{bmatrix} 1.45124 \\ 0.001906 \\ 0.00083 \end{bmatrix}$$

Solution Key

Discussion and Practice

1. a.

$$Xb = \begin{bmatrix} 3.31756 \\ 3.15832 \\ 2.80848 \\ 3.3812 \\ 2.73224 \\ 3.19244 \\ 2.84998 \\ 3.05226 \\ 3.16416 \\ 3.21888 \\ 3.49002 \\ 3.49894 \\ 3.27698 \\ 3.27022 \\ 2.9296 \end{bmatrix}$$

b.

$$\begin{bmatrix} 3.58 \\ 3.17 \\ 2.31 \\ 3.16 \\ 3.39 \\ 3.85 \\ 2.55 \\ 2.69 \\ 3.19 \\ 3.50 \\ 2.92 \\ 3.85 \\ 3.11 \\ 2.99 \\ 3.08 \end{bmatrix} - \begin{bmatrix} 3.31756 \\ 3.15832 \\ 2.80848 \\ 3.3812 \\ 2.73224 \\ 3.19244 \\ 2.84998 \\ 3.05226 \\ 3.16416 \\ 3.21888 \\ 3.49002 \\ 3.49894 \\ 3.27698 \\ 3.27022 \\ 2.9296 \end{bmatrix} = \begin{bmatrix} 0.26244 \\ 0.01168 \\ -0.49848 \\ -0.2212 \\ 0.65776 \\ 0.65756 \\ -0.29998 \\ -0.36226 \\ 0.02584 \\ 0.28112 \\ -0.57002 \\ 0.35106 \\ -0.16698 \\ -0.28022 \\ 0.1504 \end{bmatrix}$$

c. $\Sigma(r)^2 = 2.109564418$. To determine the sum of the squared residuals using matrices, multiply the transpose of the matrix of residuals times the matrix itself.

2. The matrix method will not work because you would have to solve the system by finding the inverse of X. In this case X is a 15×3 matrix and does not have an inverse. Only square matrices have inverses.

STUDENT PAGE 119

When you multiply X by b you have

$$\begin{bmatrix} 1 & 670 & 710 \\ 1 & 630 & 610 \\ 1 & 490 & 510 \\ 1 & 760 & 580 \\ 1 & 450 & 510 \\ 1 & 600 & 720 \\ 1 & 490 & 560 \\ 1 & 570 & 620 \\ 1 & 620 & 640 \\ 1 & 640 & 660 \\ 1 & 730 & 780 \\ 1 & 800 & 630 \\ 1 & 640 & 730 \\ 1 & 680 & 630 \\ 1 & 510 & 610 \end{bmatrix} \cdot \begin{bmatrix} 1.45124 \\ 0.001906 \\ 0.00083 \end{bmatrix}$$

Discussion and Practice

1. Enter the data for X and b into two matrices on your calculator.

a. Find the predicted *GPAs*, Xb.

$$\text{Let } [Y] = \begin{bmatrix} 3.58 \\ \\ 3.08 \end{bmatrix}.$$

Enter the data for Y into a matrix on your calculator.

b. Find the residuals using matrices.

c. What is the sum of squared residuals? Explain how to find this using matrices.

In the example above, you used the coefficients from the model *VM*. In general, how can you find good choices for b_0, b_1, and b_2? You want to find a matrix b such that Xb is a good approximation of Y, where Y is an $n \times 1$ matrix that contains the known values of the variable y. To find the matrix formula for the least-squares solution (to find the best values of b_0, b_1, and b_2) you have to find the best b to use in the equation Y = Xb. You determined a way to find b by the iterative process you used in Lessons 6 and 7.

2. You might also think about using the matrix method we used in Lesson 1 to find b_0, b_1, and b_2 that in general will give the best prediction for *GPA* based on *SATV* and *SATM*. If Xb = Y, then $b = X^{-1}Y$. Will this method work? Why or why not?

STUDENT PAGE 120

A Matrix Solution

Fortunately, there is another method, and it uses matrices! If you premultiply both sides of the equation Xb = Y by the transpose of X, you get $X^TXb = X^TY$. X^TX has an inverse as long as no column in X is a linear combination of any other column. You can find the inverse of X^TX and use that to find b.

$$Xb = Y$$
$$X^TXb = X^TY$$
$$((X^TX)^{-1})(X^TXb) = (X^TX)^{-1}(X^TY)$$
$$\text{or } ((X^TX)^{-1})(X^TX)b = (X^TX)^{-1}X^TY$$
$$\text{so } b = (X^TX)^{-1}X^TY$$

Therefore, if you premultiply Xb = Y by $(X^TX)^{-1}X^T$, you get $b = (X^TX)^{-1}X^TY$. Note that b is not the exact solution to the equation Xb = Y because you did not directly "undo" multiplication by X. Instead, you found a way to create the possibility of an inverse, so what you have found is the "best" approximation you can for the solution. In the last part of this section, you will see why it is the "best."

3. Consider the points (2, 5) and (4, 7) and the graph of the line below.

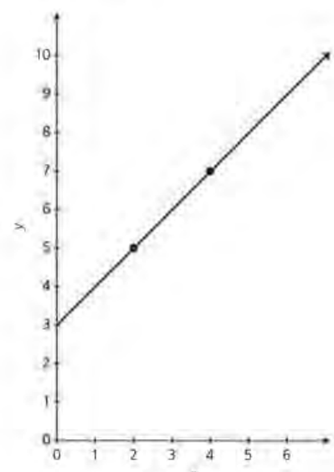


Figure 8.1

3. **a.** $y - 5 = \frac{7-5}{4-2}(x-2)$

$y - 5 = \frac{2}{2}(x-2)$

$y = x + 3$

$\Sigma(r)^2 = 0$

b. The least squares linear regression line obtained from a calculator will also be $Y = X + 3$.

c.

$$X = \begin{bmatrix} 1 & 2 \\ 1 & 4 \end{bmatrix}; \quad X^T = \begin{bmatrix} 1 & 1 \\ 2 & 4 \end{bmatrix}$$

$Xb = Y$

$b = (X^TX)^{-1} X^T Y$

$$\underset{X}{\begin{bmatrix} 1 & 2 \\ 1 & 4 \end{bmatrix}} \cdot \underset{b}{\begin{bmatrix} b_0 \\ b_1 \end{bmatrix}} = \underset{Y}{\begin{bmatrix} 5 \\ 7 \end{bmatrix}}$$

$$\mathbf{b} = \left(\begin{bmatrix} 1 & 1 \\ 2 & 4 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 1 & 4 \end{bmatrix}\right)^{-1} \cdot \begin{bmatrix} 1 & 1 \\ 2 & 4 \end{bmatrix} \cdot \begin{bmatrix} 5 \\ 7 \end{bmatrix}$$

$$= \begin{bmatrix} 5 & -1.5 \\ -1.5 & 0.5 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 \\ 2 & 4 \end{bmatrix} \cdot \begin{bmatrix} 5 \\ 7 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & -1 \\ -.5 & .5 \end{bmatrix} \cdot \begin{bmatrix} 5 \\ 7 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} \rightarrow y = 3 + x$$

4. The matrix formula works. To show that it works, students might take an example with three points and use least squares linear regression to find a fit, then compare that line to the one obtained by using the matrix formula. For the three points shown in Figure 2, the equations will be:

by regression on the calculator, $y = 3.66667 + x$;

by the matrix method, $Xb = Y$

$$\text{or} \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} 5 \\ 8 \\ 7 \end{bmatrix};$$

then $b = (X^TX)^{-1}X^TY$, and

$$\left(\begin{bmatrix} 1 & 1 & 1 \\ 2 & 3 & 4 \end{bmatrix} \cdot \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}\right)^{-1} \cdot$$

$$\begin{bmatrix} 1 & 1 & 1 \\ 2 & 3 & 4 \end{bmatrix} \cdot \begin{bmatrix} 5 \\ 8 \\ 7 \end{bmatrix} = \begin{bmatrix} 3.66667 \\ 1 \end{bmatrix}$$

implying $y = 3.66667 + x$.

STUDENT PAGE 121

a. Write the equation of the line by using the two points. What is the sum of the squared residuals?

b. Write the equation of the least squares linear regression line that would fit two points.

c. Write the equation of the line you would find by using the matrix formula to determine the coefficients as described above.

All three of these methods give the same answer, that is, the same values for the slope and the intercept of the least squares regression line when there are only two points.

4. Hector claims the matrix method using the formula works in the general case in which there are three or more data points. Is he correct? Justify your reasoning. Use the points (2, 5), (3, 8), and (4, 7) shown in Figure 8.2.

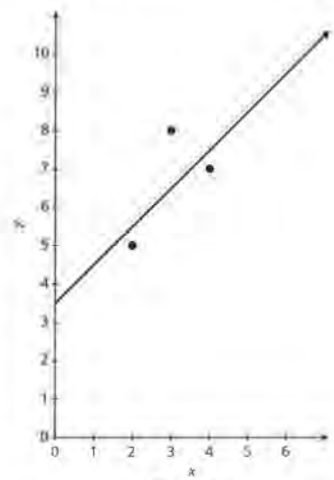


Figure 8.2

5. It is impossible to use either of the two methods described in problems 3a and b above to find the equation for the SAT verbal and mathematics scores to predict college grade point.

a. Explain why.

5. **a.** You cannot find the slope and intercept of the equation because it is not a linear equation in two variables. You cannot find a least squares linear regression line as you usually do because there are two independent variables.

(5)b. The equation using the matrix formula is
$\hat{y} = 1.528854829$
$+ 0.0014099537V$
$+ 0.0011918744M$.

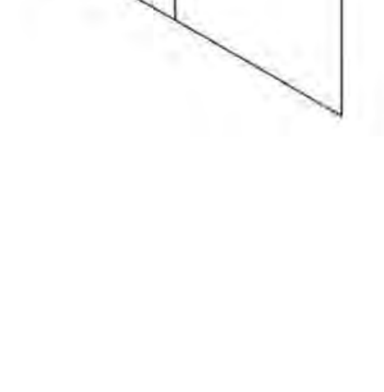
The iteration approach was converging to the model above.

$\hat{y}_{MVM} = 1.59594333$
$+ 0.0009815276V$
$+ 0.0015045651M$

$\hat{y}_{VMVM} = 1.505247875$
$+ 0.001560712V + 0.001082M$

6. a. The foresters could measure the length of the shadow of the tree and measure the angle of elevation from the end of the shadow to the top of the tree. Then they could determine the height of the tree by using the tangent relationship.

Or, the foresters could use similar triangles to determine the height of the tree. They would measure the distances from the eye to the pencil and from the eye to the tree; then knowing the length of the pencil, they could determine the height of the tree. (See diagram)



STUDENT PAGE 122

b. Apply the formula $b = (X^TX)^{-1}X^TY$ to the original SAT and GPA example. How does your solution compare to the one you found using the iterative process?

6. Hardwood trees are used to make furniture. The amount of lumber that can be realized from a tree is critical to determine how much the tree is worth when it is sold. To measure volume before the tree is cut, two measurements are taken: the diameter of the tree 4.5 feet from the ground and the height of the tree measured with sighting instruments. After the tree is cut, the actual volume can be measured. The better the estimate, the more likely the seller will get the correct amount for a crop of trees. Therefore, a good regression model will be useful. The data in the table below are on black cherry trees from the Allegheny National Forest, Pennsylvania.

Cherry Wood Trees

Diameter (ft)	Height (ft)	Volume (ft³)	Diameter (ft)	Height (ft)	Volume (ft³)
0.691	70	10.3	1.075	74	19.1
0.716	65	10.3	1.075	74	22.2
0.733	63	10.2	1.108	85	33.8
0.875	72	16.4	1.142	86	27.4
0.892	81	18.8	1.150	71	25.7
0.900	83	19.7	1.183	78	34.5
0.917	66	15.6	1.208	80	31.7
0.917	75	18.2	1.333	74	36.3
0.925	80	22.6	1.358	72	38.3
0.933	75	19.9	1.442	77	42.6
0.942	79	24.2	1.458	81	55.4
0.950	76	21.0	1.492	82	55.7
0.950	76	21.4	1.500	80	58.3
0.975	69	21.3	1.500	80	51.5
1.00	75	19.1	1.717	87	77.0

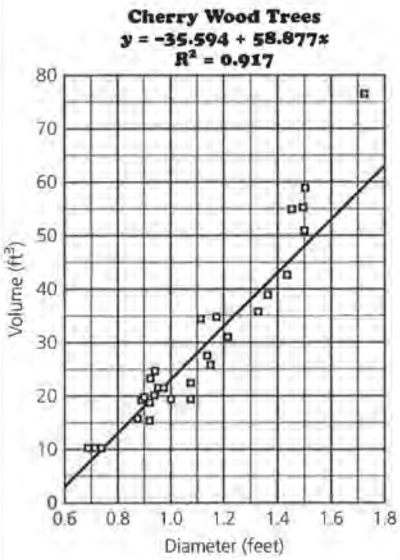
Source: B. F. Ryan, B. L. Joiner, and T. A. Ryan, *Minitab Handbook* 2d ed., PWS-Kent, Boston 1986

a. How do you think the foresters measure the height of a tree? Use an example in your answer.

b. Make a plot of (*diameter, volume*). Find the least squares linear regression model for the data by using the matrix formula and by using the linear regression on your calculator. How do the two models compare?

c. Find the sum of squared residuals for the least squares model.

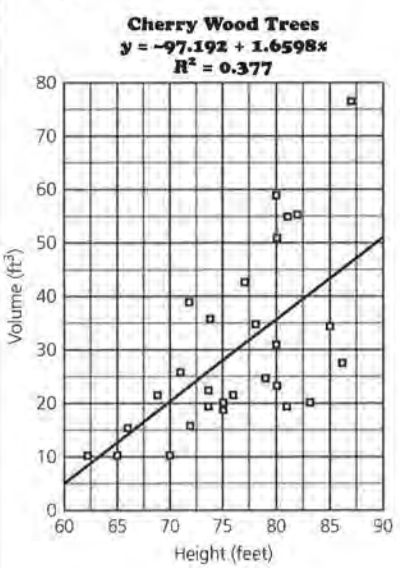
b.



The two models should be the same as the equation recorded above the graph.

c. $\Sigma(r_d)^2 = 641.28$

d.



$\Sigma(r_h)^2 = 4821$ (Notice that the sum of squared residuals is much larger for this model. The correlation coefficient was not as strong for this model either. There is a much better relationship between diameter and volume than between

STUDENT PAGE 123

d. Make a plot of (*height, volume*). Find the least squares linear regression model for the plot. Find the sum of squared residuals.

e. Find a regression model to predict volume from both diameter and height. Compare the three models and describe how well each fits the data.

7. Think carefully about the procedure for finding the coefficients for a model that can serve as a predictor equation. How would the procedure have to be modified if you had three independent variables? Consider using height, shoulder width, and age to predict the weight of a person.

a. Write the general form of the appropriate model and set up a matrix system using the three variables.

b. Suppose you have the measurements for 12 people. Describe the system you would be using.

c. Explain how you could use the procedure described above to find the coefficients of the model that will give you the least sum of squared residuals.

8. Obtain from your teacher the grades for a quiz, homework or a project, a test, and a final or semester exam for at least ten students. Find a model that could be used to predict the final exam grade based on the set of other grades.

Why the Formula Works (Optional)

The formula $b = (X^TX)^{-1}X^TY$ will always give you the least squares regression model for any number of variables. You must be careful, however, because, as in any curve-fitting process, a variable may have no relation to the response variable and so will not improve the model. The idea behind least squares regression is to find values of the parameters b_0, b_1, and b_2 that make the sum of squared residuals as small as possible. (Hence the term "*least squares.*") A residual is the difference between an actual value of y and the corresponding predicted value $\hat{y}$, where you subtract $\hat{y}$ from y: $r = y - \hat{y}$.

In matrix terms, the predicted values (i.e., the $\hat{y}$s) are given by $Xb = \hat{Y}$. Thus, you can use matrices to find the residuals by subtracting Xb from the matrix Y. This gives an $n \times 1$ matrix of residuals: $Y - \hat{Y}$. If you want to calculate the sum of squared residuals, take this $n \times 1$ matrix and premultiply it by its transpose (which is a $1 \times n$ matrix). The product is a 1×1 matrix

height and volume, although height is definitely a factor.)

e. $\hat{y} = -56.098 + 54.736d + 0.329h$

$\Sigma(r_{dh})^2 = 561.62$

$\Sigma(r_d)^2 = 641.28$

$\Sigma(r_h)^2 = 4821$

The sum of the squared residuals is smaller in the multiple regression model, implying it is best. Since the sum of the squared residuals for the regression of the diameter against the volume is smaller than the sum for the regression of the height against the volume, the model using the diameter is preferred over the model using the height.

7. a.

$X = [1\ h\ s\ a]$ $\qquad$ $Xb = Y$

$$b = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} \qquad [1\ h\ s\ a] \times \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = [w]$$

$Y = [\text{weight}]$

(7) b. Matrix X would have dimensions (12×4).

Matrix b would have dimensions (4×1).

Matrix Y would be (12×1).

c. If you multiply the multiplicative inverse of $[X^TX]$ by the product of X^T and Y, the result will be a matrix of the predicted coefficients (including the constant) for the prediction equation.

8. Student answers are dependent on their own data.

STUDENT PAGE 124

(i.e., a scalar) that equals the first residual squared + the second residual squared + ...

$$\text{Sum of squared residuals} = SSR = [Y - \hat{Y}]^T[Y - \hat{Y}] = [Y - Xb]^T[Y - Xb]$$

To find the least squares values for the parameters b_0, b_1, and b_2, you want to minimize $[Y - Xb]^T[Y - Xb]$. First, multiply out this expression, using the rules identified in Lesson 1. This will involve four terms:

$$\begin{aligned} SSR &= [Y - Xb]^T[Y - Xb] \\ &= [Y^T - (Xb)^T]\,[Y - Xb] \\ &= Y^TY - Y^T(Xb) - (Xb)^TY + (Xb)^T(Xb) \\ &= Y^TY - Y^TXb - b^TX^TY + b^TX^TXb \text{ since } (Xb)^T = b^TX^T \end{aligned}$$

Note that Y^TXb is the product of a $1 \times n$ matrix (Y^T) with an $n \times 1$ matrix (Xb). Thus, it is a 1×1 matrix. Likewise, b^TX^TY is a 1×1 matrix. Since the transpose of a 1×1 matrix (i.e., a scalar) is equal to itself and since the transpose of b^TX^TY is Y^TXb, it follows that b^TX^TY equals Y^TXb. Thus,

$$SSR = Y^TY - 2Y^TXb + b^TX^TXb.$$

You wish to minimize this expression by finding good values for the parameters b_0, b_1, and b_2. That is, you want to find a good matrix b. The first term in *SSR*, Y^TY, does not involve b, so you can ignore it and concentrate on minimizing $-2Y^TXb + b^TX^TXb$.

Earlier you found that a good choice of b is $b = (X^TX)^{-1}X^TY$. You will see that this is the best choice of b. You will see that if you change $(X^TX)^{-1}X^TY$ by adding a vector to it, you make matters worse, unless the vector is the zero vector (that is, a vector whose elements are all equal to zero).

Suppose $b = (X^TX)^{-1}X^TY + v$, where v is an $n \times 1$ vector. Then the sum of squared residuals

$$SSR = Y^TY - 2Y^TXb + b^TX^TXb$$

becomes

$$= Y^TY - 2Y^TX[(X^TX)^{-1}X^TY + v] + [(X^TX)^{-1}X^TY + v]^TX^TX[(X^TX)^{-1}X^TY + v].$$

If you expand the product of these terms you get

$$\begin{aligned} SSR = Y^TY &- 2Y^TX[(X^TX)^{-1}X^TY - 2Y^TX[v] \\ &+ [(X^TX)^{-1}X^TY]^TX^TX[(X^TX)^{-1}X^TY] \\ &+ [(X^TX)^{-1}X^TY]^TX^TX[v] \\ &+ [v]^TX^TX[(X^TX)^{-1}X^TY] \\ &+ [v]^TX^TX[v] \end{aligned}$$

STUDENT PAGE 125

This expression contains seven terms, but the first, second, and fourth terms do not involve v. Combine those terms and call them c (for "constant"). Then you have

$$\begin{aligned} SSR = c &- 2Y^TX[v] + [(X^TX)^{-1}X^TY]^TX^TX[v] \\ &+ [v]^TX^TX[(X^TX)^{-1}X^TY] \\ &+ [v]^TX^TX[v] \end{aligned}$$

The next thing to notice is that the second to last term, $[v]^TX^TX[(X^TX)^{-1}X^TY]$, can be simplified. The X^TX part cancels with the $(X^TX)^{-1}$ part, leaving $[v]^T[X^TY]$. Moreover, v is a $p \times 1$ vector (matrix), so $[v]^T$ has dimension $1 \times p$. X^TY is the product of a $p \times n$ matrix with an $n \times 1$ matrix, so it has dimension $p \times 1$. Thus, $[v]^T[X^TY]$ has dimension 1×1, which means that $[v]^T[X^TY]$ is equal to its transpose $Y^TX[v]$.

Likewise, the term $[(X^TX)^{-1}X^TY]^TX^TX[v]$ can be simplified. First, take the transpose of the leading part, $[(X^TX)^{-1}X^TY]$, to get $Y^TX[(X^TX)^{-T}]$, so that

$$[(X^TX)^{-1}X^TY]^TX^TX[v] = Y^TX[(X^TX)^{-T}]X^TX[v].$$

Now, (X^TX) is a symmetric matrix, which means that $(X^TX)^{-1}$ is also symmetric. Thus, $(X^TX)^{-T} = (X^TX)^{-1}$. Hence, $(X^TX)^{-T}$ cancels with (X^TX), giving us the result that $Y^TX[(X^TX)^{-T}]X^TX[v] = Y^TX[v]$.

Thus, you have simplified the expression for the sum of squared residuals to the form

$$\begin{aligned} SSR = c &- 2Y^TX[v] + Y^TX[v] \\ &+ Y^TX[v] \\ &+ [v]^TX^TX[v]. \end{aligned}$$

Notice that the middle part of this expression cancels, which means that

$$SSR = c + [v]^TX^TX[v].$$

Write the last term a bit differently to get

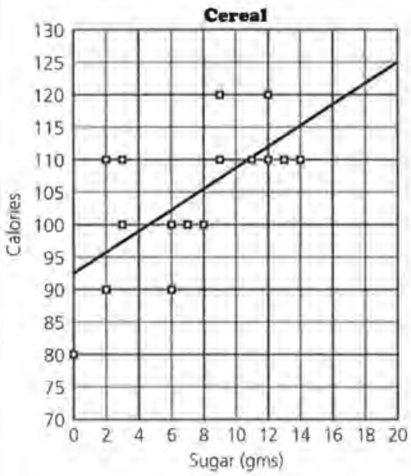
$$SSR = c + v^TX^TXv = c + [Xv]^T[Xv].$$

In summary, the sum of squared residuals equals a constant plus $[Xv]^T[Xv]$.

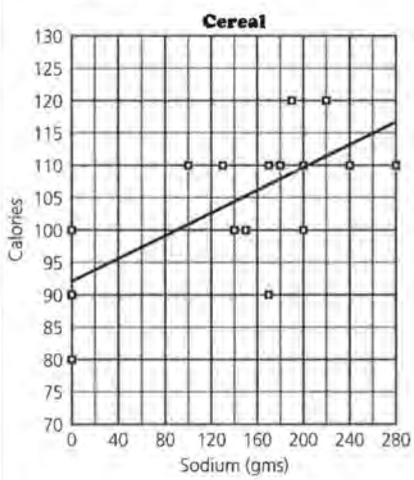
Now, Xv is the product of an $n \times p$ matrix and a $p \times 1$ matrix, so it has dimension $n \times 1$. That is, Xv is an $n \times 1$ vector. The term $[Xv]^T[Xv]$ is the sum of the squared values of the elements of Xv. That is, $[Xv]^T[Xv]$ equals the first element squared + the second element squared + ... You wish to minimize SSR, which means that you want $[Xv]^T[Xv]$ to be as

Practice and Applications

9. a.



Regression equation:
$C = 92.629 + 1.6063 \times \text{sugar}$
$\Sigma(r)^2 = 1041.81857$



Regression equation:
$C = 92.2585 + 0.0823 \times \text{sodium}$
$\Sigma(r)^2 = 1049.00832$

The two predictors work about equally well.

STUDENT PAGE 126

small as possible. Clearly, the smallest possible value for the sum of squared elements is 0, which is what you get if *each* of the elements equals 0.

You can make each of the elements of Xv equal to 0 by making v the zero vector. Thus, you have shown that the sum of squared residuals is minimized when v is the zero vector, which means that the least squares solution is

$$b = (X^TX)^{-1}X^TY + v = (X^TX)^{-1}X^TY.$$

This is what you wanted to show.

Summary

When you have more than one variable to use to predict an outcome, it is possible to find a good model by using an iterative regression process. There is a matrix formula, however, that will produce the model much faster. It can be shown that using this formula produces the model with the least sum of squared residuals. If you write the problem as a matrix system, you can find the elements in the coefficient matrix b by using the formula $b = (X^TX)^{-1}X^TY$.

Practice and Applications

9. What cereal do you eat for breakfast? How high is your cereal in calories? What variables affect the number of calories? The table shown on the next page contains the amount of sugar, the amount of sodium, and the number of calories for a one-ounce serving of 16 different cereals.

b. Regression equation:
$\hat{y} = 84.00166475 + 1.372788111$ sugar $+ 0.0702153524$ sodium.
$\Sigma(r)^2 = 515.884049$
This sum of the squared residuals is about half the size of either of the other two sums.

c. The prediction is 118.25 calories.

d. A change in the amount of sugar will increase the amount of calories faster because the coefficient of sugar is 1.37, while the coefficient of sodium is only 0.07. (Of course, sugar is measured in grams and sodium is measured in milligrams.)

e. Some other factors could be fat and carbohydrates. To include them you would add columns to matrix X and add the same number of constants to matrix b.

STUDENT PAGE 127

Cereal

Cereal	Sugar (gms)	Sodium (mg)	Calories
Cap'n Crunch	12	220	120
Apple Jacks	14	125	110
Corn Pops	12	90	110
Frosted Flakes	11	200	110
Frosted Mini Wheats	7	0	100
Life	6	150	100
Nut & Honey Crunch	9	190	120
Raisin Nut Bran	8	140	100
Raisin Squares	6	0	90
Rice Chex	2	240	110
Shredded Wheat	0	0	80
Wheaties	3	200	100
Golden Graham	9	280	110
Nutri Grain Wheat	2	170	90
Cocoa Puffs	13	180	110
Grape Nuts	3	170	110

Source: American Statistical Association, 1993 statistical graphics exposition

a. Plot the data for each of the following. Find a regression model and decide how well the model seems to fit the data.

i. (*sugar, calories*)

ii. (*sodium, calories*)

b. Find a multiple regression model using both sugar and sodium to predict the number of calories. How does this model compare to either of the other two?

c. Use your model from part b to predict the number of calories in a serving of Wheat Chex that has 5 grams of sugar and 390 milligrams of sodium.

d. Look carefully at your model from part b. Which variable has the greater effect on the prediction? Explain how you can tell.

e. Do you think other variables might also have an influence on the number of calories? If so, what might they be? Describe how would you adjust your model to accommodate them.

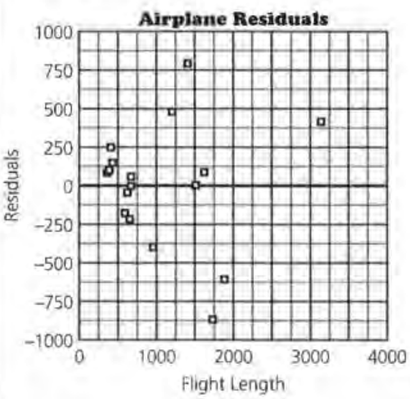
10. In Lesson 5 you studied the relation between two of the variables that might be factors in finding the operating cost of an airplane. Other variables are given in the table on the next page.

10. a. $OC = -1619.68 + 7.05637$ speed $+ 1.1686$ length

b. $\Sigma(r)^2 = 2{,}521{,}538$

c. The predicted cost would be $2375.95.

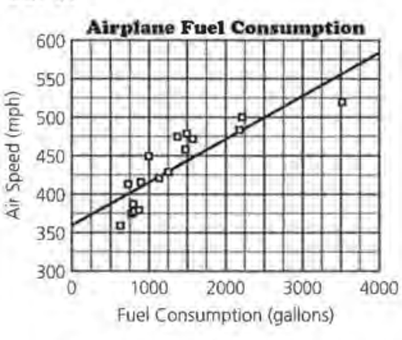
d. You could plot the residuals against the flight length and see if the residuals are larger for the longer flights. It appears in general that as the length of the flight increases, the residuals do get larger.



11. $OC = 673.44 + 7.1245$ seats $- 1.0060$ speed $+ 0.2006$ length $+ 0.7067$ fuel

$\Sigma(r_{s,l})^2 = 2{,}521{,}538$ and $\Sigma(r_{s,l,s,f})^2 = 282{,}310.79$. The multiple regression model results in a large reduction in the sum of squared residuals and so is preferred.

12. a.



The trend is that this plot is increasing. It also appears that the plot is curved.

STUDENT PAGE 128

Airplane Data

Aircraft	No. of Seats	Air Speed (mph)	Flight Length (miles)	Fuel Consumption (gal/hr)	Operating Cost/hour ($)
B747-100	405	519	3149	3529	6132
L-1011-100/200	296	498	1631	2215	3885
DC-10-10	288	484	1410	2174	4236
A300 B4	258	460	1221	1482	3526
A310-300	240	473	1512	1574	3484
B767-300	230	478	1886	1503	3334
B767-200	193	475	1736	1377	2887
B757-200	188	449	984	985	2301
B727-200	148	427	688	1249	2247
MD-80	142	415	667	882	1861
B737-300	131	413	605	732	1826
DC-9-50	122	378	685	848	1830
B727-100	115	422	626	1104	2031
B737-100/200	112	388	440	806	1772
F-100	103	360	384	631	1456
DC-9-30	102	377	421	804	1778
DC-9-10	78	376	394	764	1588

Source: *The World Almanac and Book of Facts*, 1993

a. Find a model for using speed and flight length to predict operating costs.

b. What is the sum of squared residuals?

c. Use your model to predict the operating cost for an airplane that travels at 500 mph and usually makes flights of around 400 miles.

d. Is your model more effective for short trips or for long ones or is there no difference? How can you tell?

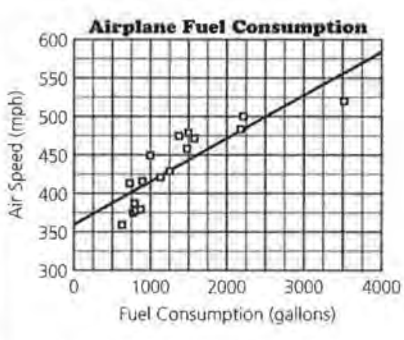
11. Use all of the factors to determine a model to predict the cost of operating an airplane. Compare your model using all of the factors to the one you found using only speed and flight length.

12. The matrix formula can be used to generate a prediction equation in one variable that is quadratic rather than linear. Suppose you had just speed to predict fuel consumption.

a. Plot (fuel consumption, speed). Describe the trend you see in the plot.

b. Find a least squares linear regression line to describe the plot. How well does your line fit the data?

b.

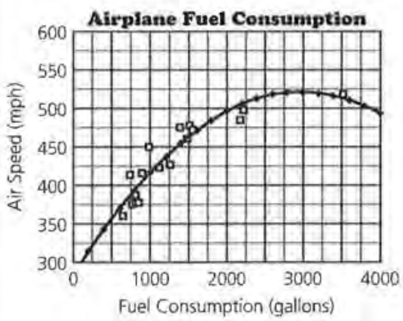


Speed = 359.844427 + 0.0562533537 × fuel. This line does not appear to fit very well. The points on the ends are below the line and the points in the middle are above the line.

c. $S = 282.0856255 + 0.1611949466$ (fuel) $- 0.00002709351325$ (fuel $)^2$

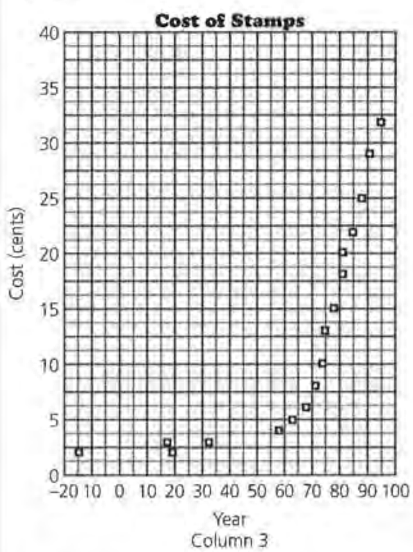
The matrix formula will determine the values of coefficients $b_0, b_1, b_2, \cdots$ for any number of factors you wish to consider. Here you are using x_1 = fuel and x_2 = (fuel)^2.

d.



This curve seems to be a much better fit of the data set we have. There is a concern about extrapolation as the equation appears to be dropping in value too rapidly beyond 3500 gallons per hour.

13. a.



$C = 1.197847031 - 0.267810276\,y + 0.0059007021y^2$,
where y = year – 1900.

b. The feature that has the most effect is that the rate of change between years is not constant or linear.

c. In the year 2000 the predicted cost of a stamp according to the quadratic model is 33.4¢ or simply 33¢. Note: $y = 100$ when the year is 2000.

STUDENT PAGE 129

c. Use the matrix formula to find an equation of the form

$b_0 + b_1x + b_2x^2 = y$

where $[1 \;\; x \;\; x^2]$ is the matrix just as $[1 \;\; V \;\; M]$ was in the lesson. Explain why the matrix formula can be used to find the model.

d. Graph your equation with the data. How does the new model compare to the linear model?

13. The cost of stamps has consistently risen over the years.

Cost of Stamps

Year	Cost of Stamps (Cents)
1885	2
1917	3
1919	2
1932	3
1958	4
1963	5
1968	6
1971	8
1974	10
1975	13
1978	15
1981	18
1981	20
1985	22
1988	25
1991	29
1995	32

Source: *The Milwaukee Journal*, December 1994

a. Plot (*year*, *cost*). Find a model to predict the cost of stamps.

b. What feature of the data has a strong impact on the model?

c. Predict the cost of stamps in the year 2000. How reliable do you think your prediction will be?

14. **a.** $R = -5.36250967 + 67.20470557\frac{C}{A} + 139.3422339\frac{T}{A} - 237.0252249\frac{I}{A} + 6.875124272\frac{Y}{A}$

b. NFL rating for Anderson is 81.85800745.

The rating for Anderson determined by the formula derived for problem part a is 82.60.

New formula rating for Montana is 94.12549703.

New formula rating for Young is 92.36213346.

The coefficients in the models differ, but they give similar ratings.

15. Students find data and create an original problem. You might use Problem 15 as an assessment of students' understanding of the process, or you might have students report their results to the class using a poster presentation that includes the data, the results, and any other interesting features. You could have the class grade each report based upon organization, quality of data, explanation of the model and how it works, ability to answer questions, and usefulness or interesting information provided as part of the solution. Presentations give students the opportunity to verbalize their knowledge using data they have selected.

STUDENT PAGE 130

14. The table below shows the all-time leading passers in the National Football League (NFL) as of 1993, along with the variables used by the NFL to rate its quarterbacks on their passing ability.

NFL Quarterback Rating

Player	Att	Comp	TD	Int	Yards	Rating
Ken Anderson	4475	2654	197	160	32838	81.9
Len Dawson	3741	2136	239	183	28711	82.6
Sonny Jurgensen	4262	2433	255	189	32224	82.6
Jim Kelly	3494	2112	179	126	26413	86.0
David Kreig	4178	2431	217	163	30485	82.0
Neil Lomax	3153	1817	136	90	22771	82.7
Dan Marino	5434	3219	298	168	40720	88.1
Joe Montana	4898	3110	257	58	37268	93.1
Roger Staubach	2958	1685	153	109	22700	83.4
Steve Young	1968	1222	105	58	15900	93.0

Source: *The Universal Almanac*, 1995

a. Using the variables $\frac{C}{A}$, $\frac{T}{A}$, $\frac{I}{A}$, and $\frac{Y}{A}$, where

R = the rating
A = attempts
C = number of completions
T = number of touchdowns
I = number of interceptions
Y = yards gained by passing,

find a model to predict the ratings in general.

b. The formula used by the NFL is

$$R = \frac{50 + 2000\left(\frac{C}{A}\right) + 8000\left(\frac{T}{A}\right) - 10000\left(\frac{I}{A}\right) + 100\left(\frac{Y}{A}\right)}{24}$$

How does your formula compare?

15. Choose a situation that has at least two variables involved in the outcome. Collect data and find a regression model. In a brief report, describe each of the following:

- the problem
- the data source
- the model and how you selected it
- an example of how the model works
- your opinion about how effective your model will be to make predictions

ASSESSMENT

Rating Universities

Materials: none
Technology: graphing calculator
Pacing: 1 class period unless given as a take-home test or long-range assignment

Overview

Students may work in pairs or individually. Having groups of more than two might not be useful in this case. If you give the assessment as an in-class test, students will not be able to truly explore the data in meaningful ways, although they should be able to demonstrate their understanding by finding immediate and obvious answers to the questions. If they are given the task as a long-term assignment, students will have the opportunity to display their understanding and knowledge of the mathematics to a much greater degree. A curious student might seek out the original article and obtain the rest of the data describing the universities and see how the prediction model can be refined based on further information.

Teaching Notes

Using data describing universities and admission characteristics, students are asked to find a model to predict the actual attendance rate of students who apply and are accepted by the university. In the first problem, students apply what they know about least squares linear regression and find a model to predict a response variable given one independent variable. They are expected to begin by looking at a graphic display of the data and then to demonstrate knowledge of correlation, sum of squared residuals, and residual plots as they look for a good model. Based on these results, students are then asked to select a second independent variable and produce a multivariate regression model by using the techniques they learned in Lesson 8.

These variables were some of those used by *U.S. News & World Report* to rate the universities. Students are given the final rating produced by the magazine and are asked to find a model that would predict that final rating based on the four variables they are given.

The problems lend themselves to an open investigation. A student can go directly to the solutions or can show an understanding of the process by investigating many different options. Some might look to see whether using four variables predicts better than using any other combination. Others might experiment to see if a quadratic or a cubic model might be more effective. Don't expect all students to exhaust the problem; however, the opportunity is there for the student who is interested and capable.

The process can be greatly simplified if the student knows how to make a calculator or computer software do the work. By entering data into matrices and using judicious editing or cutting and pasting, much of the repetition and tedious data entering can be avoided. This means that careful planning on the part of students before they begin punching keys is important. You might either have some of the students who are very good at this explain their procedures to the class before giving the assessment or select someone who did exceptionally well on the assessment to give an explanation after the task has been completed. One final caution, rounding the decimal places can result in extremely large errors. Students should keep every decimal place in their calculations as they proceed, although they may choose to round to four or five decimal places when recording results.

Observe whether students are using the linear regression key on their calculator or are actually finding the regression line by using the matrix formula, which works for situations in which there is only one independent variable as well.

STUDENT PAGE 131

ASSESSMENT

Rating Universities

The following data were taken from some of the information given in the *U.S. News & World Report*, September 18, 1995 issue on America's best colleges. The universities were the top fourteen in *U.S. News & World Report* list, given here alphabetically.

OBJECTIVE

Create ratings and ranks when high scores are good for one variable but bad for another variable.

Data on Top Universities

University	SAT/ACT 25th Percentile	Acceptance Rate %	% Attend After Accepted	Cost per Student ($)
Brown University (RI)	1210	22	49	22,704
California Institute of Technology	1350	25	46	63,575
Cornell University (NY)	1180	33	47	21,864
Dartmouth (NH)	1250	23	48	32,162
Duke University (NC)	1220	30	40	31,585
Harvard University (MA)	1320	14	75	39,525
Johns Hopkins University (MD)	1210	44	28	58,691
Massachusetts Institute of Technology	1290	30	51	34,870
Northwestern University (IL)	1160	39	36	28,052
Princeton University (NJ)	1280	14	56	30,220
Stanford University (CA)	1270	20	54	36,450
University of Chicago (IL)	1180	50	29	38,380
University of Pennsylvania	1190	36	47	27,553
Yale University (CT)	1290	19	53	43,514

1. The percentage of students who actually attend a university after having been offered an invitation to attend is given in the third column.

   a. Which factors do you think might influence the attendance rate for the universities?

   b. Analyze the data given above, looking for a relationship between two of the variables. Describe what you find and indicate how the elements of the curve-fitting process are part of the discussion.

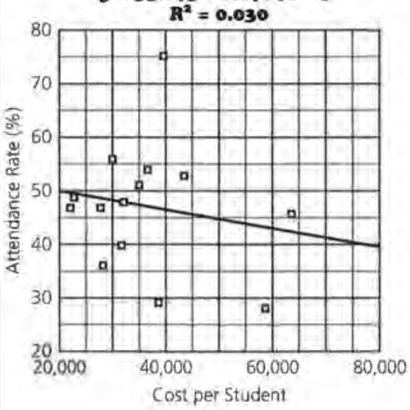
1. a. Factors such as financial aid, distance from home, or acceptance at other universities may influence the actual attendance rate.

   b. Student responses will vary. Some possible avenues for exploration are given in the example, but students may approach the problem in very different ways. Their discussion should include any patterns in the plot, the prediction equation (if appropriate), the sum of the squared residuals, residual plots, and correlation.

   Students should begin by looking at the plot. Each of the three possible independent variables and their value in predicting attendance rate can be investigated and compared. There is virtually no relationship between cost and attendance rate, and some relationship between the SAT scores and the attendance rate. The strongest relationship is between acceptance rate and attendance rate: the larger the percentage acceptance, the smaller the percentage attendance. The schools that are more selective seem to have a greater degree of acceptance for the students they invite.

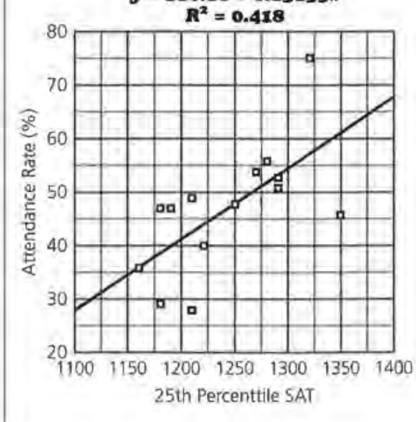
   Observations might include:

University Attendance/Cost
$y = 53.175 - 1.6784e - 4x$
$R^2 = 0.030$



For (*cost per student, attendance rate*), the plot shows there is no pronounced pattern of any kind. The correlation coefficient is 0.17, which reinforces the fact that the association is weak.

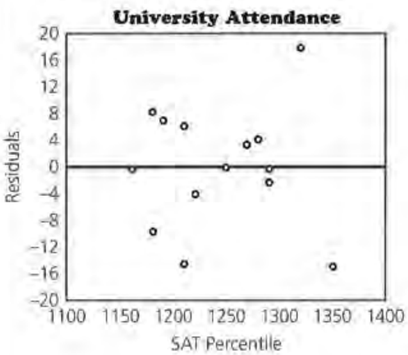
University Attendance
$y = 116.18 + 0.13135x$
$R^2 = 0.418$

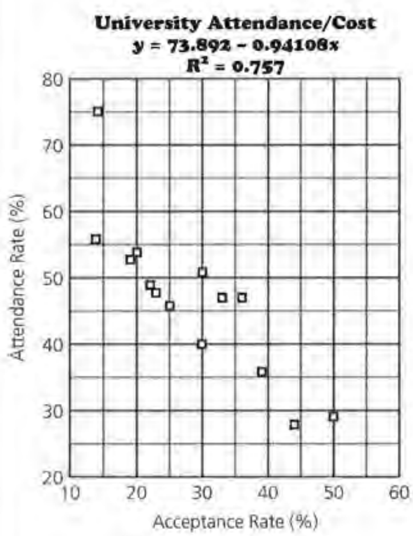


For (*twenty-fifth percentile on SAT, attendance rate*), the plot shows a possible linear trend; the equation would be

$AT = -116.18 + 0.13135SAT$.

The correlation coefficient is .65 (r^2 is 0.418), which indicates that 42% of the variability in attendance can be accounted for by knowing something about the twenty-fifth SAT percentiles of those attending the schools. The sum of the squared residuals is 1063. The residual plot is shown below.

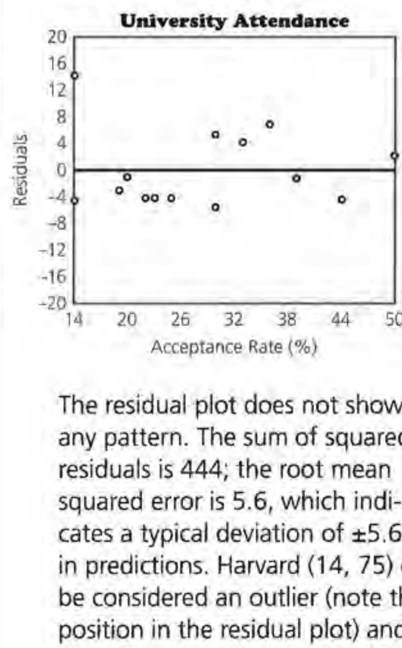




For (*acceptance rate, attendance rate*), the data seem to be quite linear; the higher the acceptance rate, the lower the actual attendance rate. The least squares linear regression would be

$AT = 73.892 - 0.94108AC$.

The slope of –0.94 indicates that for every 1% increase in the percent accepted, there is a corresponding decrease of 0.94 in the percent who actually attend. The correlation coefficient is –0.87 and r^2 is 0.757, which indicates that 76% of the variability in attendance can be accounted for by knowing the attendance rates. (You might observe whether students think that the 42% and the 76% should then account for more than 100% of the variability. The fact that they both explain that much about attendance indicates that these two variables also have some correlation with each other.)



The residual plot does not show any pattern. The sum of squared residuals is 444; the root mean squared error is 5.6, which indicates a typical deviation of ±5.6% in predictions. Harvard (14, 75) can be considered an outlier (note the position in the residual plot) and alters the sum of squared residuals and the correlation considerably. Harvard contributes 204 to sum of squared residuals for the linear model. Without Harvard, the correlation is 0.9, and the root mean squared error is now 3.8%.

(1) **c.** Based on the work for part b, the two variables that seem to have an impact on attendance rate would be the percentile on the SAT test and the acceptance rate.

A multivariate model that would predict the attendance rate (*AT*) using these two variables would be $AT = 58.1558 + 0.01165SAT - .89699AC$. The sum of squared residuals using this model is 441, which is an improvement over any of the fits using only one independent or explanatory variable.

d. Matrices allow us to find the coefficients for the multiple regression model easily.

2. **a.** Student responses will vary. Using twenty-fifth percentile on *SAT*, attendance rate (*AT*), acceptance rate (*AC*), and cost (*C*), the multiple regression model for the overall rating (*Ra*) would be:

$$Ra = 73.79087569 + 0.0170126216SAT + 0.0568506922AT - 0.0265693752AC - 0.00001523200893C$$

It can be found by entering a column vector of 1s and the four variables into a data matrix (*D*) and the actual ratings into a Ratings matrix (*Ra*).

From the matrix system, D · b = **Ra,** you get

$$[1\ SAT\ AT\ AC\ C] \cdot \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = [\mathbf{Ra}]$$

and can find **b** by the matrix formula $\mathbf{b} = (\mathbf{D}^T\mathbf{D})^{-1}\,\mathbf{D}^T\,(\mathbf{Ra})$. Store **b** in a new matrix, then use the product of **D** · **b** to predict the ratings and store the predictions into another matrix, Predicted Ratings (**PR**). **Ra** – **PR** = **R**, residuals which can be stored into yet another matrix; $\mathbf{R}^T\mathbf{R}$ = sum of squared residuals.

The sum of squared residuals using all four variables is 16.106.

b. If cost is not included, the multiple regression equation is

$Ra = 77.96 + .0131SAT + 0.0653AT - 0.0350AC$.

The sum of squared residuals for this model will be 16.198, not as small as the sum using four variables. Involving three variables can be done easily by eliminating the cost column on the matrix using four variables.

The model using only *SAT* percentile is $Ra = 63.58 + 0.026SAT$, and Σr^2 is 24.84.

As you saw in part a, $\Sigma r^2 = 16.106$ using all four variables. The smaller Σr^2 indicates a better model.

STUDENT PAGE 132

c. Look for two variables that may have an impact on the attendance rate. Find a multivariate regression equation that would predict the attendance rate.

d. Explain how using matrices simplifies the process.

2. The final rating for each of the universities as given by *U.S. News and World Report* is shown below.

Rating Top Universities

University	Overall Scores
Brown University (RI)	95.3
California Institute of Technology	95.5
Cornell University (NY)	94.0
Dartmouth (NH)	95.5
Duke University (NC)	96.8
Harvard University (MA)	100.0
Johns Hopkins University (MD)	94.6
Massachusetts Institute of Technology	98.0
Northwestern University (IL)	94.0
Princeton University (NJ)	98.8
Stanford University (CA)	98.1
University of Chicago (IL)	94.4
University of Pennsylvania	94.4
Yale University (CT)	98.8

a. Use appropriate variables and find a multivariable regression equation to predict the final rating based on those variables.

b. Show how you can verify that multiple regression produces a better model than any using just one independent variable.

Teacher Resources

LESSON I QUIZ

Representing Ratings with Matrices

NAME ______________________________

1. Researchers from California State University at Fresno studied 36 United States cities to measure the willingness of people to help strangers. They performed the following tests: If a researcher dropped a pen, would someone walking toward them pick it up and return it? Would someone help a researcher with a leg brace pick up dropped magazines? Would someone check for change for a quarter when asked? Would someone tell a researcher, who was wearing dark glasses and carrying a white cane, when a light turned green? Would someone mail a dropped stamped and addressed letter? They also looked at *density* and the *annual per capita contribution* to United Way. The table below contains the ranks in each category for the ten cities that were eventually rated "the friendliest" cities. (Note that these ranks come from all 36 cities, so are not from 1 to 10.)

City	Dropped Pen (*P*)	Hurt Leg (*H*)	Change (*C*)	Blind Person (*B*)	Lost Letter (*L*)	United Way (*U*)	Density (*D*)
Chattanooga, TN	27	1	15	28	12	2	2
Detroit, MI	5	22	4	24	3	18	24
East Lansing, MI	14	22	9	7	3	22	17
Houston, TX	3	12	2	13	20	19	10
Knoxville, TN	13	29	3	2	6	20	8
Louisville, KY	2	32	1	18	17	15	18
Memphis, TN	4	6	15	12	17	16	9
Nashville, TN	8	2	6	4	24	9	1
Rochester, NY	18	16	19	3	12	1	23
St. Louis, MO	18	15	5	10	12	5	22

Boston Globe, July 7, 1994

a. Which city seems to have the largest number of high individual rankings?

b. Suppose you use the following formula (Formula 1) to calculate the total rating for each city:

Formula 1: $P + 2H + C + 3B + 2L + U + 5D = R$

What weight was assigned to each of the variables?

(1) c. Write the formula you would use if each category was given an equal weight. Call this Formula 2.

d. Use matrix multiplication to find the total ratings for each city for both formulas. Record your results in the table below and rank the cities in each case. How do the ranks compare for the two different formulas?

City	Ratings from Formula 1	Rank	Ratings from Formula 2	Rank
Chattanooga, TN				
Detroit, MI				
East Lansing, MI				
Houston, TX				
Knoxville, TN				
Louisville, KY				
Memphis, TN				
Nashville, TN				
Rochester, NY				
St. Louis, MO				

2. Let Matrix $A = \begin{bmatrix} 12 & 22 & 18 \\ 2 & 41 & 25 \\ 8 & 14 & 31 \end{bmatrix}$; Matrix $B = \begin{bmatrix} 1 & 0 \\ 5 & 3 \\ 2 & 1 \end{bmatrix}$;

and Matrix $C = \begin{bmatrix} 1 & 3 & 5 \\ 0 & 1 & 0 \\ 2 & 4 & 1 \end{bmatrix}$.

a. Can we multiply $A \cdot B$? Can we multiply $B \cdot A$?

b. Is $(AC)^T = A^T C^T$? Explain your thinking.

c. Is $A^T A$ necessarily the same as $A\,(A^T)$ for any matrix A? Why or why not?

3. Scoring for the Indianapolis Car Racing competition is based only on finishing in the first twelve places and on the number of laps the driver led. One driver won 3 races, finished second twice, and led 48 laps. Another driver won 1 race, finished second 3 times, and led 14 laps. A third driver won 1 race and led 12 laps. The first driver earned 140 points. The second driver earned 82 points. The third driver earned 32 points.

Source: *USA Today*, October 21, 1994

a. Set up a system of equations to determine how many points the IndyCar system awards for first place, for second place, and for leading a lap.

b. Use matrices to find the solution.

c. How many points did Rusty Wallace earn in the IndyCar system if he placed first in the Hanes 500 at Martinsville, Virginia, in 1994 and led the race for two laps?

4. Can you solve every system of equations using a matrix system? Explain why or why not.

Bonus

Rochester, New York, was the "friendliest" city among the 36. (Patterson, New Jersey, was the least!) What weights could you use that would ensure that, at least for the data above, Rochester would rank first?

Jobs and Basketball

NAME ______________________________

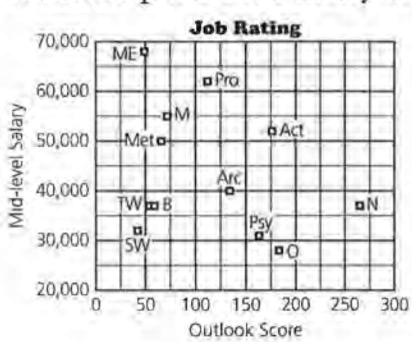
Jobs

Jobs Rated Almanac ranks 250 jobs by a variety of variables including salary, stress, outlook, and environment. In each case, a score for a category is created by evaluating factors in that category. Outlook is defined as the quality of a job's future and involves unemployment rates, expected government growth, potential salary growth, and potential for promotion. Points are assigned to each of these areas and compiled for a total score, where a high score is good. The table contains scores for outlook and the mid-level incomes for twelve jobs.

Job	Mid-level Income ($) (*I*)	Outlook Score (*L*)
Architect (ARC)	40,000	134
Mechanical Engineer (ME)	68,000	49
Accountant (ACT)	52,000	176
Nurse (N)	37,000	265
Mathematician (M)	55,000	72
Meteorologist (MET)	50,000	66
Professor (PRO)	62,000	112
Optician (O)	28,000	183
Biologist (B)	37,000	59
Psychologist (PSY)	31,000	163
Social Worker (SW)	32,000	42
Technical Writer (TW)	37,000	54

Source: *Jobs Rated Almanac*, 3d Edition, 1995.

1. A scatter plot of the salary and outlook score is below.



a. Is there any one occupation that clearly dominates the others with respect to both factors (salary and outlook)? Explain why or why not.

b. In 1993 the median income (I) earned by a worker over the age of 25 in the United States was \$25,636. The median outlook score (L) was 23. Explain how these two numbers can be used to determine the weights in the equation

$$0.000039I + 0.043L = R$$

for a rating (R) based on both salary and outlook score.

c. Use the equation to determine the rating for both variables for a mathematician. With those weights, does any other career have the same total rating?

2. Write the equation of the line that would describe all of the jobs that would have the same rating as a mathematician.

a. How is the slope of the line related to the weights for salary and outlook? Graph that line.

b. Airline pilots have a total rating of 3.93. Will the line containing the point representing pilots be above or below the line for mathematicians? Explain how you know.

c. Describe how a sweeping line can help you determine the *best* job with respect to salary and outlook.

3. Find the total ratings for each of the jobs.

a. Use the ratings to rank the jobs.

Job	Mid-level Income ($)	Outlook Score	Ratings	Ranking
Architect	40,000	134		
Mechanical Engineer	68,000	49		
Accountant	52,000	176		
Nurse	37,000	265		
Mathematician	55,000	72		
Meteorologist	50,000	66		
Professor	62,000	112		
Optician	28,000	183		
Biologist	37,000	59		
Psychologist	31,000	163		
Social Worker	32,000	42		
Technical Writer	37,000	54		

(3) b. Saundra calculated her total ratings and found that nursing rated first, followed by psychologist. She studied her graph and knew she had an incorrect answer. How could she tell from the plot that she had done something incorrectly?

c. If you weighted salary three times as much as outlook, how would the ratings change?

d. What weights would make the occupations of accountant and mathematician have approximately the same rating?

Basketball

4. Who had the greatest shooting season in the NBA? The field goals and free-throw statistics for their best seasons are given in the table below for ten of the all-time top shooters.

Player	Number of Field Goals	Number of Free-Throws
Wilt Chamberlain	1597	835
Rick Barry	1011	753
Elvin Hayes	930	467
Kareem Abdull-Jabbar	1159	504
Michael Jordan	1098	833
Dominque Wilkins	888	527
George Gervin	1024	505
Alex English	959	406
Adrian Dantley	909	632
Nate Archibald	1028	663

Source: *World Almanac and Book of Facts,* 1995

a. Use what you have learned to rate and then rank the players. Be sure to explain your reasoning.

b. Are there any weights that would make Jordan first? Explain.

Rating Educational Institutions

NAME ______________________________

1. The September 18, 1995, *US News and World Report* contained an article ranking the best colleges in the United States. The table below contains the ranks given to the top ten universities for six categories: academic reputation, student selectivity, faculty resources, financial resources, retention rank, and alumni satisfaction.

College	Alumni Satisfaction (*AS*)	Retention Rank (*RE*)	Financial Resources (*FI*)	Faculty Resources (*FA*)	Student Selectivity (*S*)	Academic Reputation (*AR*)
Brown University	13	6	29	12	7	14
California Institute of Technology	24	28	1	2	5	8
Dartmouth College	7	4	12	18	7	17
Duke University	9	5	11	13	9	8
Harvard University	3	2	5	1	1	1
Johns Hopkins University	10	22	2	15	21	4
Massachusetts Institute of Technology	16	10	7	6	4	1
Princeton University	1	2	14	5	3	4
Stanford University	33	7	6	3	6	1
Yale University	6	1	4	10	2	4

a. To determine the total rating, the magazine used the following weights for each category: alumni satisfaction 5%, retention rank 25%, financial resources 10%, faculty resources 20%, student selectivity 15%, and academic reputation 25%. Write a formula to determine the total rating for a university.

b. Find a rating for each university using the following: the formula you wrote in part a, a formula that would give equal weights to the categories, and a formula that gives 0 weights to faculty resources.

c. Rank the universities according to the three sets of ratings. How do they compare?

d. Using the weights given, explain how matrix multiplication can help you find the ratings.

2. The following data were collected by staff at *Milwaukee Magazine* for the purpose of rating high schools in the Milwaukee, Wisconsin, area. The following ten schools were among those rated.

School	National Percentile 10th Grade Test	Grads to Four-year College %	% Enrolled in AP Took Test	% Passed AP test	Class Size Estimated	Student Per Computer	Teachers with Adv. Degrees %
Arrowhead-Hartland	82	70	11	62	25	10	65
Brookfield East	88	77	4	73	24	9	57
Cedarburg	82	68	7	74	26	7	76
Kettle Moraine	81	65	5	75	25	8	65
Mukwonago	80	60	5	72	22	6	51
New Berlin Eisenhower	81	63	17	42	25	5	61
Nicolet	85	85	9	88	22	9	83
Shorewood	87	89	6	96	23	7	66
West Allis Central	73	50	1	95	22.5	7	77
Whitnall	78	75	6	76	23.5	5	63

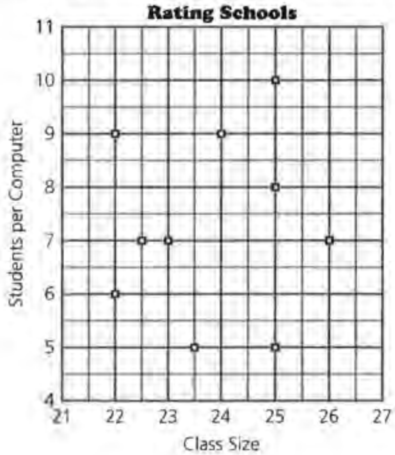
Source: *Milwaukee Magazine*, September 1995

a. Find the average of the students per computer values for the ten schools.

b. Use matrices to find the difference between the average number of students per computer and the number of students per computer at each school.

c. Show how you can use the transpose to find the sum of the squared differences. What does the sum of squared differences indicate?

3. A scatter plot of the class size and the number of students per computer for each of the ten schools is given below. Nationally, according to *Market Data Retrieval* data published by the U.S. Census, the number of students per computer was 9.7 in 1993. The average class size in 1994 was 17.3, according to the National Education Association.

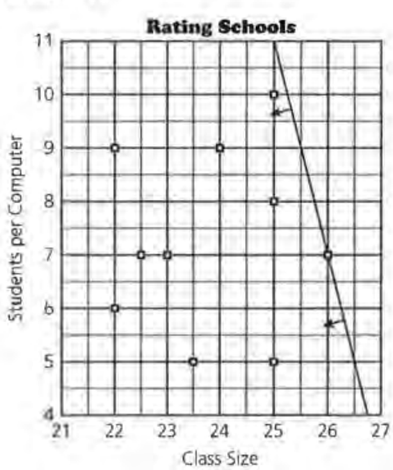


a. Create weights that could be used to make those variables equivalent, then write an equation for finding a total rating using the two variables.

b. Draw a sweeping line that could be used to determine the ranking for the ten schools and explain how the line would work.

c. Rank the ten schools according to the weights you found for part a. Does the ranking seem reasonable? Why or why not?

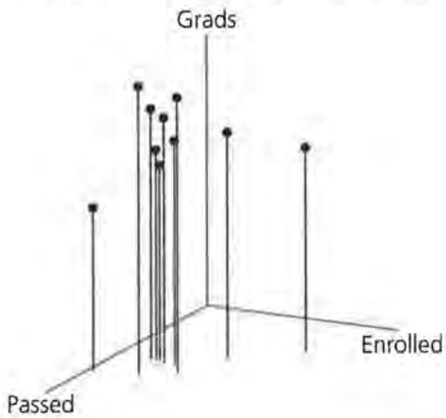
d. The plot below contains a sweeping line that was created by weighting one of the variables more than the other. Which of the two, class size or the number of students per computer, was weighted more heavily and how do you know?



(3.)e. Is it possible to find a set of weights that would make Brookfield East, Cedarburg, and Kettle Moraine have equal ratings? If so, find them. If not, explain why not.

4. The plot below is of

- the percent of students who attend a four-year college,
- the percent of those enrolled in Advanced Placement courses who took an Advanced Placement test, and
- the percent of those who took the Advanced Placement test who passed with a score of three or better.



Information you may find useful: According to 1992 data from the Census Bureau, 65.6% of high school graduates enrolled in college. 64% of those who took the Advanced Placement test passed the test in 1995. (You will have to ignore the difference in years to use the data; over several years, the percentages will not vary too significantly.)

a. Why is it necessary to use information on both who took an AP course and who passed? What do you think is the optimum percent for those enrolled who took the test and why?

b. Write an equation that you could use to find a total rating for each school and describe the graphical representation of the equation.

c. Use your equation to find a rating for the schools, then use those ratings to rank the schools.

d. Based on the data you have, how adequately do you think the results describe the school? How do they compare to your work from problem 3?

5. One student used weights for the three variables in problem 4 and found the following total ratings: Arrowhead 745; West Allis Central 1005; and Shorewood 1079. What were the weights the students used? Show how matrices can help you find your answer.

6. Add a fourth variable to the data you used in problem 4, the percentile score on a tenth grade national test given by the state.

a. Use the four variables to rank the schools and describe the process you used. (A reasonable standard for achievement on a national test might be to score in the 50th percentile.)

b. You were given certain information to determine weights for each variable used in the formula. Why was this necessary and what other information do you think would have been useful?

7. **a.** Describe the geometric representation in space of each of the following.
Make a sketch that illustrates your thinking.

$y = 9$

$3x + y = 9$

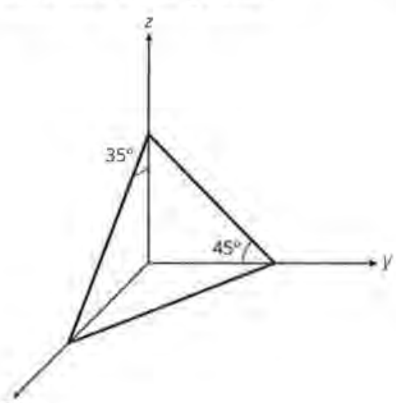
$3x + y + z = 9$

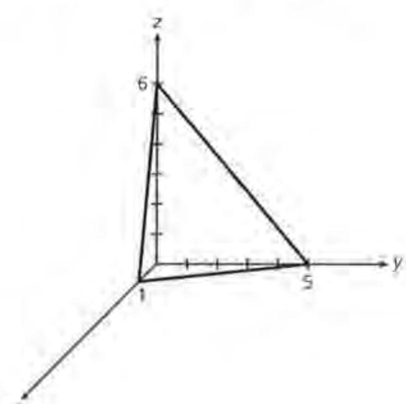
b. Decide whether the two planes are parallel and explain how you know.

$x + y + 3z = 7$

$2x + 2y + 3z = 14$

c. Given the information in the plots below, write an equation for each plane.



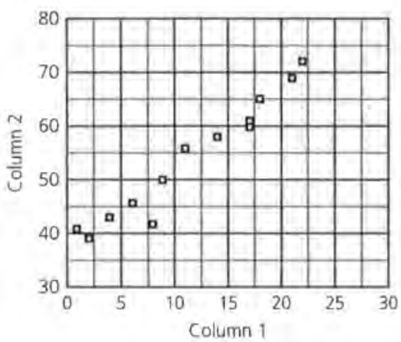


Modeling

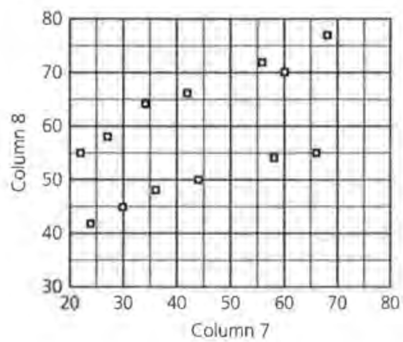
NAME ______________________________

1. For which of the following plots would you expect the data to have a correlation coefficient of 0.9 or greater? Explain why in each case.

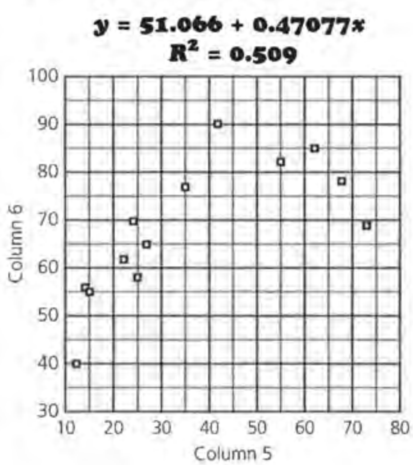
a.



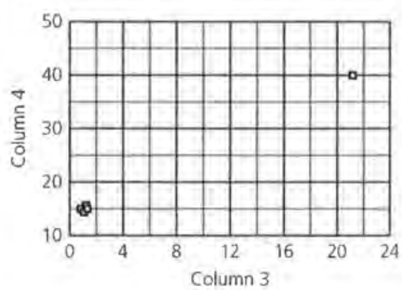
b.



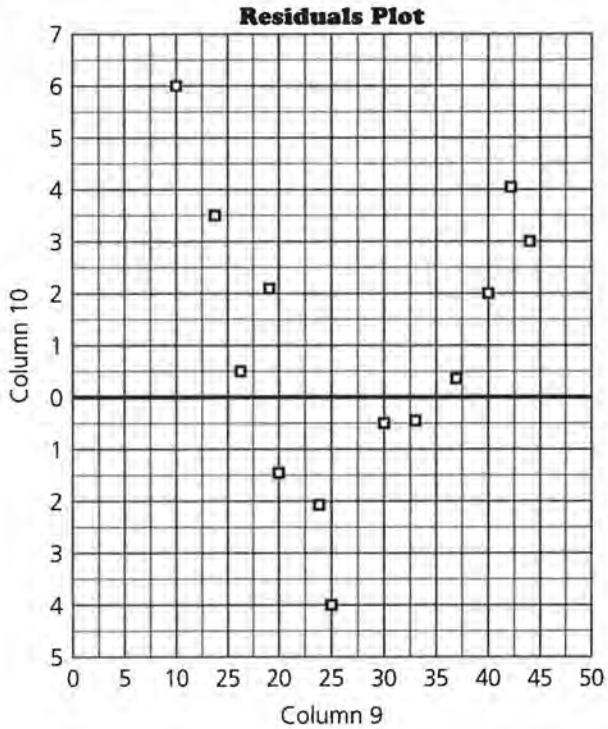
c.



d.



2. A residual plot for an equation that was fit to a set of data is below.



Make a rough sketch of data and a model that would produce residuals that were similar to those in the given plot. Explain your reasoning.

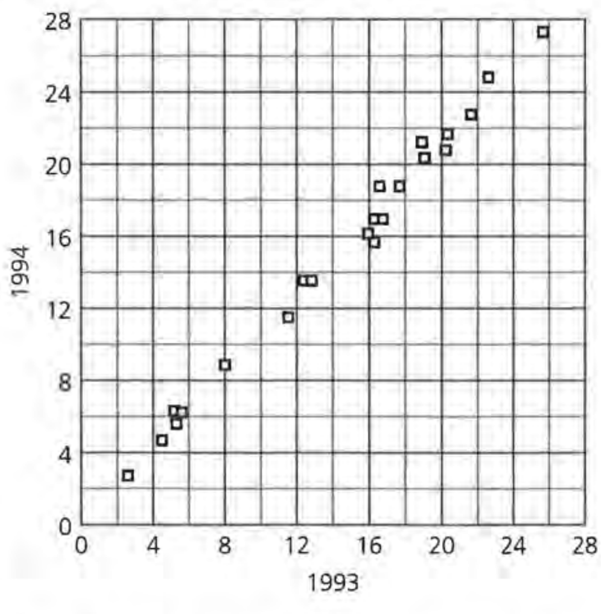
3. The table contains the hourly compensation in United States dollars for manufacturing jobs in various countries throughout the world for 1993 and for 1994.

Country	1993	1994	Country	1993	1994
Mexico	2.56	2.61	Denmark	19.11	20.44
United States	16.73	17.10	Finland	16.56	18.89
Canada	16.33	15.68	France	16.23	17.04
Australia	12.49	13.66	Germany	25.70	27.31
Hong Kong	4.29	4.80	Italy	16.00	16.16
Japan	19.01	21.42	Netherlands	20.21	20.91
Korea	5.51	6.25	Norway	20.21	20.91
New Zealand	8.01	8.93	Portugal	4.50	4.57
Singapore	5.25	6.29	Spain	11.50	11.45
Taiwan	5.22	5.55	Sweden	17.70	18.81
Austria	20.37	21.73	Switzerland	22.63	24.83
Belgium	21.62	22.97	United Kingdom	12.76	13.62

Source: U.S. Department of Labor, World Trade, September 1995

a. In which country was the compensation the highest? Explain what might cause the difference in wages between that country and the wages in the United States.

b. A plot of the data is below.



A model that seems to be a good fit to predict the 1994 compensation from the 1993 data can be given as the matrix system

$$[1\ W] \times \begin{bmatrix} 0.104 \\ 1.056 \end{bmatrix} = [PW].$$

Explain how this system works.

c. Use the system to predict the hourly compensation in Ireland for 1994 if you knew the hourly compensation in 1993 was $12.16.

d. Find the residual for the United States. Explain what this represents.

Bonus

Suppose that the model were constant for each year. In other words, the same model could be used to predict the 1995 wages from the 1994 wages. What would you predict for the wages for 1997 and how would you proceed?

Buying a Home

NAME ______________________________

1. Suppose you were moving into an area of the city and wanted to be sure you lived in an area that had good schools, but you were also concerned about the amount of taxes you would pay.

a. In some of the earlier lessons you studied the way schools were ranked. Describe the connection you think might exist between the rating of a school system, the amount of taxes paid for a home, and the cost of buying a home in that district. (The cost is the median price of the homes listed for sale in one weekend period; the tax is the equalized tax rate for the region of the state.)

b. What other variables might influence the price of a house?

2. Think back to the iterative process for finding a regression model when you have more than one independent or explanatory variable.

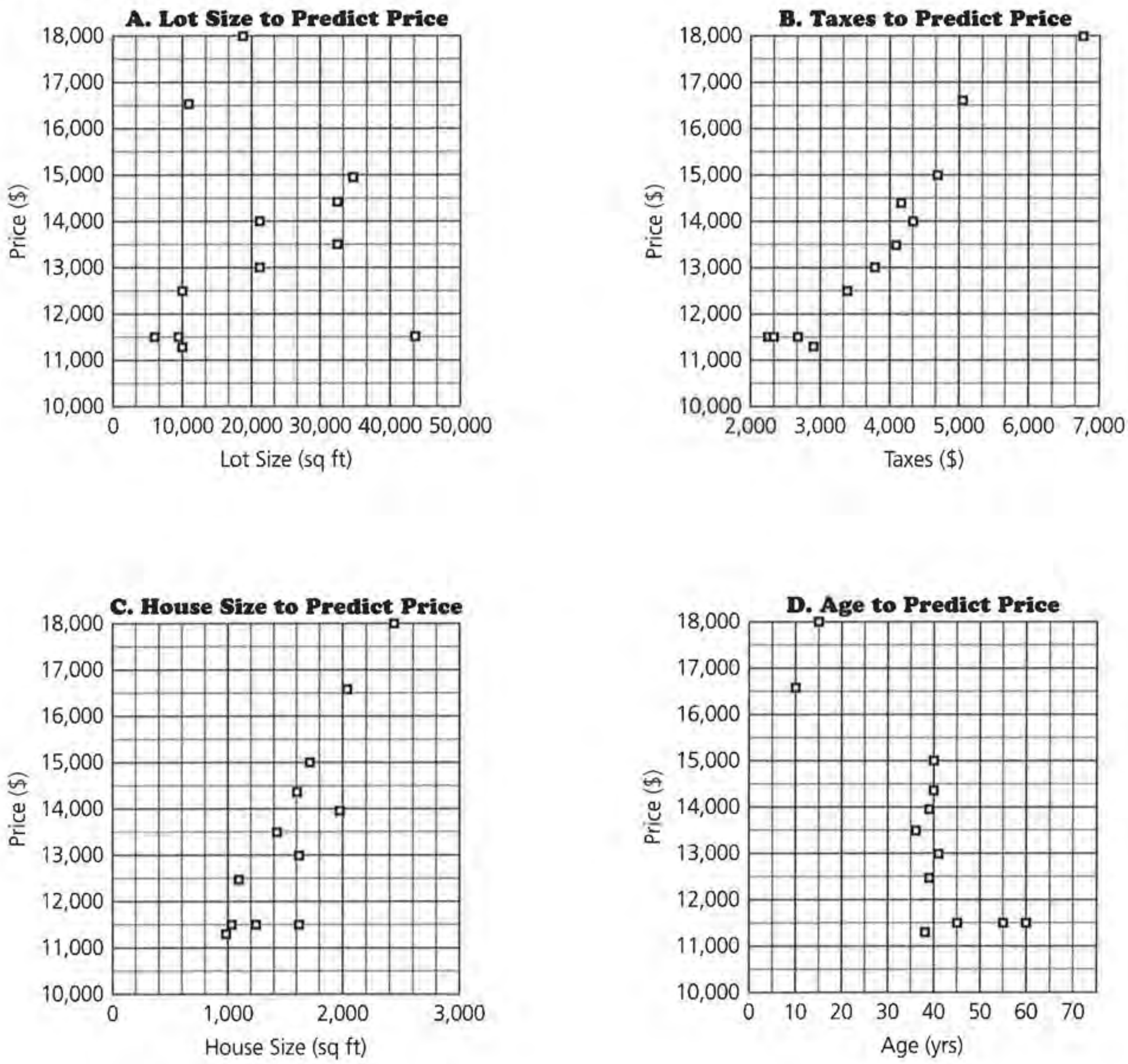
a. A regression for (rating of school system, cost) for 28 homes produced the following equation:
$C = 67642\ R - 80849$. Describe what you should do next in the iterative process in order to include the amount of taxes paid for a home in your regression model.

b. After two steps in the iteration process, suppose you had a model that looked like the following:
$C = 67642\ R + 19055\ T + 12351 + e$. Explain what e represents.

c. The sum of squared residuals for the model in part b was 6,354,291,529. Describe how this sum of squared residuals would compare to that for the model in part a and to the sum of squared residuals for the next model in the iteration process.

d. Suppose you had instead first chosen to find the regression equation for (*taxes, cost*). How would your results have been different?

3. a. Suppose that based on your visits and discussion you have selected a certain area of the city. Data concerning house prices were collected from a realtor for that area. Plots of (housing variables, cost) are below. Which of the following do you think would have a strong effect on the price of a house in the area in which you are interested? Explain your reasoning.



The data that were collected are in the following table.

Housing Variables

Lot Size (sq ft) (*L*)	Taxes (*T*)	House Size (sq ft) (*H*)	Age (*A*)	Price (*P*)
10000	2904	980	38	112900
43560	2689	1244	60	114900
9400	2337	1627	55	114900
6000	2242	1024	45	114900
10005	3400	1100	39	124900
21056	3784	1629	41	129900
32670	4091	1431	36	134900
21285	4323	1992	39	139900
32670	4179	1608	40	143900
34848	4696	1723	40	149900
10890	5062	2045	10	165900
18560	6775	2450	15	179900

Source: Wauwatosa Realty

b. Write down the matrix system to predict the price using the four variables given in the table. Give the dimensions of each matrix and explain what the matrix represents.

4. Enter the data into your calculator and find the coefficients of the multiple regression model.

a. Select an example to show how you can use the coefficients to predict the price of a house.

b. Predict the prices for the homes given and find the residuals. Which home had the smallest residual? the largest? What does this mean?

c. What is the sum of squared residuals? Explain how you found it. What will this give for the root mean squared error?

5. a. Use your system to determine your offer on the price to buy a 2,000 square foot home in which you were interested that was twenty years old, was on a lot about 100 ft × 115 ft, with taxes of about $4,500.

b. The house was listed as $165,000. Do you think the price was fair or not? Explain why.

Representing Ratings with Matrices

Teaching notes

This quiz essentially covers all of the major points in the first lesson on matrices. Students should be able to use matrices to evaluate a formula, to find the transpose of a matrix, to understand the process of matrix multiplication, and to set up and solve a system of equations using matrices. You may find some students using lists on their calculators or other methods to find some of the solutions. Encourage them to try to use matrices as a very efficient way to achieve their goal. It will facilitate their work if you have the city data on a calculator or disk so they can quickly download before they begin the quiz.

At some point you might have a discussion on why anyone would actually carry out such research. There seems to be a correlation between stress, anger and hostility, and heart disease which might have been a factor in the study. The results seemed to be also closely connected to the density of a city, which might have an effect on urban planning and on city management. The study was done by social psychologists who were looking at what they termed the psychological overload from living in such cities.

In problem 1, students should be able to use both formulas in one matrix with the city data in a second and multiply the two matrices to find the solutions for d.

Problem 2 asks students to think about multiplication and the transpose without actually asking for the product. Students may find the products as a low-level response, but be sure they understand that to fully respond to the question, they need to go beyond the numerical result. For 2b, they may find the two products (which will take some time), discuss it from the perspective of the dimensions of each matrix, or might even think about the elements in the products and how they will be different for the two cases.

Problem 3 gives students the opportunity to show how they can use matrices to set up and solve a system of equations. To answer problem 4, they have to consider what the mathematics of the procedure is and understand that some matrices might not have an inverse to use in the solution. They might also begin from the base that not every system has a unique solution and give examples of matrices that would represent such systems. (As an example, consider linear equations that are multiples of one another or where the slopes are the same but the intercepts differ.)

Scoring can be done by assigning points to each response or by giving a each question a rating, say from 1 to 5. Students might enjoy taking the ratings you assigned each of the four problems and deciding what an appropriate weighting would be to use in determining their scores.

Solution Key

1. **a.** Nashville has six top ten ratings which is more than any of the other cities.

b. Picking up and returning a dropped pen, checking for change, and contributing to United Way were all given weights of one; helping someone with a hurt leg and mailing a lost letter were considered twice as important as those three, helping a blind person three times as important, and the density of the area five times as important in calculating a total rating.

c. $P + H + C + B + L + U + D = R$

(1) d. Remember that low ratings mean high rankings.

City	Ratings from Formula 1	Rank	Ratings from Formula 2	Rank
Chattanooga, TN	164	4	87	5
Detroit, MI	269	10	100	9
East Lansing, MI	201	6	94	8
Houston, TX	177	5	79	2
Knoxville, TN	152	2	81	4
Louisville, KY	260	9	103	10
Memphis, TN	162	3	79	2
Nashville, TN	92	1	54	1
Rochester, NY	218	7	92	7
St. Louis, MO	222	8	87	5

Comparisons will vary. Nashville is ranked number 1 in both formulas.

2. a. Matrix A and B can be multiplied only in the order $A \cdot B$. In order to multiply two matrices, the number of columns in the left matrix must equal the number of rows in the right matrix. This is true for these two matrices in the product $A \cdot B$ but not in the product $B \cdot A$.

b. No. In general, $(AC)^T = C^T A^T$.

c. Responses will vary. Students might use an example, finding the product of A^TA and $A\,(A^T)$. They might also use the dimensions to build an argument: if A is a 2×3 matrix, A^T would be a 3×2 matrix. The product A^TA will have dimensions 3×3, but the product $A\,(A^T)$ will have dimensions 2×2. The matrices cannot be equal because they have different dimensions.

3. a. The system of equations would be

$3F + 2S + 48L = 140$

$F + 3S + 14L = 82$

$F + 12L = 32.$

In matrix notation, this would be

$$\begin{bmatrix} 3 & 2 & 48 \\ 1 & 3 & 14 \\ 1 & 0 & 12 \end{bmatrix} \times \begin{bmatrix} F \\ S \\ L \end{bmatrix} = \begin{bmatrix} 140 \\ 82 \\ 32 \end{bmatrix}.$$

b. The solution would be the matrix $\begin{bmatrix} 20 \\ 16 \\ 1 \end{bmatrix}$ or

20 points for first place, 16 for second place, and 1 point per lap.

c. He earned 22 points.

4. Not every system of equations can be solved using a matrix system because the coefficient matrix has to have an inverse. Matrices for dependent and inconsistent systems will not have inverses.

Bonus

You can make Rochester the friendliest city among the 36 if you decrease D, H, C, L, and D where it ranked quite low and increase B and U where it ranked very high. One example is

$$D + H + C + 10B + L + 15U + 0.5D = R.$$

This formula produces a rating of 121.5 for Rochester; the next best rating for a city is 215.5. Other formulas are possible although not all combinations will work!

Jobs and Basketball

Teaching Notes

This quiz can be given in class or as an assignment. Problems 1, 2 and 3 are quite directive, and student responses should enable you to see whether they understand how to find and use an equation to determine ratings and whether they understand the geometrical interpretation of the equation. Be sure students use a calculator or software to carry out the computations, or they will not finish their work in a timely fashion. You might want to have students explain the procedure they used to calculate the total ratings using the given formula and how matrices might help them find the solution to 3c quickly after having done 3**a**. You may wish to hand out the pages unstapled so students can use the first page with the plot and data to help them with the questions.

Initially, the four points on the upper right look as if they may be on the same line and have approximately the same ratings. As students work with the equation, they should realize that using the weights dramatically changes the line they would get if they just added the two variables. (Adding is not appropriate for variables that have such different units.) This again reinforces the notion that both a mathematical approach and a geometric approach are important.

Part II may be best given as a long range assessment because it is open ended and allows students to demonstrate what they have learned. There are no suggestions made for standard values to use to determine the initial weights. Students may choose to use the average of the points given in each category, or they may wish to do some research to find a more meaningful number. Check to see whether students plot the data and if they take time to think about how a plot might help them analyze the situation. You may have students work in pairs.

Scoring will depend on how you choose to approach the problems. The first three problems can be scored by assigning points for each solution, but be sure to value explanations and work as well as final solutions. Part II could be scored holistically or points could be assigned for certain aspects of the work: graphical representation and interpretation; determining a set of weights in an appropriate way; relating the weights, the equation, and the rolling lines to the graph; finding the total ratings and ranking the players.

Solution Key

1. **a.** No one occupation dominates with respect to both variables. The nursing profession has a much larger job outlook score than any of the others, while the mid-level income for a mechanical engineer is much higher than for any of the other occupations listed.

b. 1/(25,636) is approximately 0.000039 and 1/23 is approximately 0.043. Thus, the product of the weight for mid-level salary and the mid-level salary is 1; and the product of the weight for job outlook and the job outlook score is 1. With these weights, both of these variables contribute the same value (1) to the total rating.

c. An occupation as a mathematician has a rating of 5.241. No other occupation has the same rating.

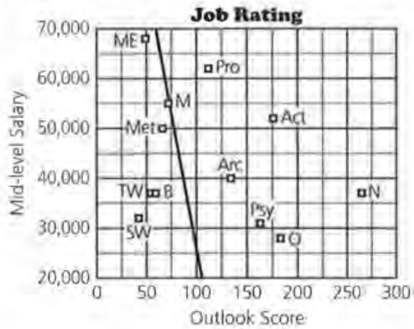
2. The line containing the rating for the mathematician would be determined by the equation: $0.000039I + 0.043L = 5.241$. If this is rewritten, the equation will be

$I = (-0.043/0.000039)L + 5.241/0.000039$

$= -1103L + 134385$

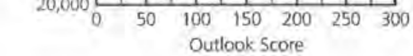
$= \left(\frac{-\text{weight for outlook}}{\text{weight for salary}}\right)L + 134{,}385$

(2) a. The slope of the line is the negative of the weight for outlook divided by the weight for salary. When the outlook score would be about 58, the income is about 70,000. Connecting that point with the point for the mathematician would give the line in the plot below.



b. The line containing the point for airline pilots will be below the line for containing mathematicians. The line will have the same slope as the mathematician line but a smaller y-intercept (3.93/0.000039 is about 100,769).

c. All of the y-intercepts will be in terms of weight for salary (W_I) and the rating. The first point the line touches as it sweeps down from the upper right will have the largest y-intercept (R/W_I). Because all of the y-intercepts are in terms of W_I, the line with the largest y-intercept will have the largest R or rating. The sweeping line using the given weights will touch the point representing nursing first.



3. a. Ratings will vary slightly due to round off.

Job	Mid-level income ($)	Out-look Score	Rat-ings	Rank-ing
Architect	40,000	134	7.32	5
Mechanical Engineer	68,000	49	4.76	9
Accountant	52,000	176	9.60	2
Nurse	37,000	265	12.84	1
Mathematician	55,000	72	5.24	7
Meteorologist	50,000	66	4.79	8
Professor	62,000	112	7.23	6
Optician	28,000	183	8.96	3
Biologist	37,000	59	3.98	10
Psychologist	31,000	163	8.22	4
Social Worker	32,000	42	3.05	12
Technical Writer	37,000	54	3.77	11

The nurse is the first occupation using those weights, followed by the accountant and then the optician.

b. Answers will vary, but there is no way the sweeping line could touch psychology before it touched accounting, optician, or professor. Look at the relationship of the points to each other in the plot.

c.

Job	Ratings Equal Weights	Rank-ing	Rating Income $\times 3$	Rank ing
Architect	7.32	5	10.44	6
Mechanical Engineer	4.76	9	10.06	7
Accountant	9.60	2	13.65	2
Nurse	12.84	1	15.72	1
Mathematician	5.24	7	9.53	8
Meteorologist	4.79	8	8.69	9
Professor	7.23	6	12.07	3
Optician	8.96	3	11.15	4
Biologist	3.98	10	6.87	10
Psychologist	8.22	4	10.64	5
Social Worker	3.05	12	5.55	12
Technical Writer	3.77	11	6.65	11

Weighting income three times as much as outlook score will essentially leave first and second places the same, but move the occupation of professor from 6 to 3. The other occupations change place only slightly.

d. Any weights that would determine the equation of the line through the points (176, 52,000) and (72, 55,000) would make the two occupations have the same rating. This would create the line $I = 52000 + (3000/-104)(L - 176)$ or approximately

$I = 57{,}077 + -28.85L.$

In the form you have been investigating, that would be $28.85L + I = 57{,}077$, where the weights would be 28.85 for outlook score and 1 for income. To put it into the same range of numbers involved in the problem, you might divide by 11,415 to produce the equation that has both occupations with a weight of 5:

$0.0025L + 0.000088I = 5$

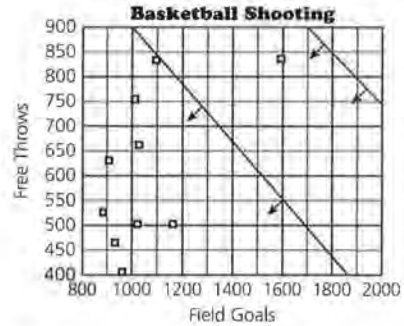
Students should recognize that this will be possible for a wide variety of weights as long as the equation is satisfied. (There is an infinite number of possible combinations of weights that will make the two occupations have the same rating.)

4. Answers will vary. Students have to select a standard for each factor. They might simply use the average of the data given, the median, or create a standard of their own (for example, 1,000 points for field goals, 500 points for free-throws). They should point out that Chamberlain clearly dominates with respect to both kinds of scoring so no matter what weights are used (if positive), he should be first. They should show a plot of the data as well as a numerical analysis.

a. If they use the average number of points from field goals (G), it is 1060 and the average number of free-throw points (F) will be 613. This will produce an equation with weights

$(1/1060)G + (1/613)F = R.$

The sweeping line for the plot (G, F) will have slope $-613/1060$ or approximately -0.58.



Player	Rating	Rank
Wilt Chamberlain	2.87	1
Rick Barry	2.18	3
Elvin Hayes	1.64	9
Kareem Abdull-Jabbar	1.92	5
Michael Jordan	2.39	2
Dominque Wilkins	1.70	8
George Gervin	1.79	7
Alex English	1.57	10
Adrian Dantley	1.89	6
Nate Archibald	2.05	4

b. There are no weights that will make Jordan first because Chamberlain dominates with respect to both field goals and free-throws. Any increase or decrease in weights will not change Jordan's position with respect to Chamberlain.

Rating Educational Institutions

Teaching Notes

This test is designed to be given after Lesson 4. It covers material from all four preceding lessons and can be given in class or as a take-home test. It would greatly facilitate student work if they were allowed to download the data into their calculator, and avoid input errors. You may chose to see how students approach the problem by providing them with the data both in matrices and in lists or spreadsheets. Using lists or spreadsheets rather than matrices will provide many of the solutions, but will take more time. Providing the data stored in programs as either lists or in matrices will enable students to make choices and to delete certain sets of the data as they use it without destroying the total set.

Students apply a formula generated by a magazine to data that are already in rating form, create a formula that will give equal weights and one that is quite open-ended, and use these to rank universities. The data on ranking high schools are in different units, and students have to equalize each factor using some standardization process. A base is suggested in each case, but you may have students select their own standardizing process which will create a whole new set of solutions. The initial problem concentrates on just two of the variables (working in two dimensions) and uses the sweeping line studied in Lesson 2. The problem that some students may not recognize is that in this case, a low rating is better than a high rating: you would like to have low class sizes and a low number of students per computer. This means the sweeping line should either begin from near the origin or if it begins from the upper right, the last point it touches will be the best according to the weights. Student understanding of the relation between the slope of the sweeping lines and the weights used in the formula will be demonstrated by their answer to problem d and e.

Students use three variables to rank the schools, this time three that are better for large values, and relate the data to a plane. In problem 7 they demonstrate their understanding of how to sketch the equation of a plane in the first octant. They also rank the schools using four variables and compare the results to those using just three. Students are not asked to use all of the variables to find a ranking because they might need more time to consider how to combine data with both large and small values; this can be assigned as a take-home follow up to the test or could be given as extra credit. The number of teachers with an advanced degree is also given, but without any standardizing information.

Do not be too concerned about answers exact to each decimal point; the answers for some of the ratings may vary because of the rounding factor. If students round 1/65.6 to .015, their total ratings will be slightly different than if they used 1/65.6 as the factor. In problem 5 where they are given the three ratings and asked to find the weights, the number of decimal places is necessary in order to have the weights obviously approach integral values.

Solution Key

1. a. $0.05AS + 0.25RE + 0.10FI + 0.20\ FA + 0.15S + 0.25AR = R$

b.–c. Student responses will vary. A formula that would give equal weights would be

$AS + RE + FI + FA + S + AR = R.$

This will vary. One formula that would give 0 weights to faculty resources could be $AS + RE + FI + S + AR = R.$

School	Formula a	Rank	Equal Weights	Rank	0-Weight Faculty Resources	Rank
Brown University	12	9	81	10	69	10
California Institute of Technology	11.45	7	68	8	66	9
Dartmouth College	11.45	7	65	7	47	6
Duke University	8.75	6	55	5	42	5
Harvard University	1.75	1	13	1	12	1
Johns Hopkins University	13.35	10	74	9	59	8
Massachusetts Institute of Technology	6.05	5	44	4	38	4
Princeton University	4.40	3	29	3	24	3
Stanford University	5.75	4	56	6	53	7
Yale University	4.25	2	27	2	17	2

Responses will vary. The top three rankings are the same in all of these rating schemes.

d. Each element of the first row, the ratings for each category, is multiplied by the corresponding element in the first column, the weight for that category, and the products are added to produce a total rating for a given school.

2. a. The average number of students per computer is 7.3.

b.

$$\begin{bmatrix} 10 \\ 9 \\ 7 \\ 8 \\ 6 \\ 5 \\ 9 \\ 7 \\ 7 \\ 5 \end{bmatrix} - \begin{bmatrix} 7.3 \\ 7.3 \\ 7.3 \\ 7.3 \\ 7.3 \\ 7.3 \\ 7.3 \\ 7.3 \\ 7.3 \\ 7.3 \end{bmatrix} = \begin{bmatrix} 2.7 \\ 1.7 \\ -.3 \\ .7 \\ -1.3 \\ -2.3 \\ 1.7 \\ -.3 \\ -.3 \\ -2.3 \end{bmatrix}$$

c. Let D be the difference matrix in part b. $D^T \cdot D$ will yield the sum of the squared differences.

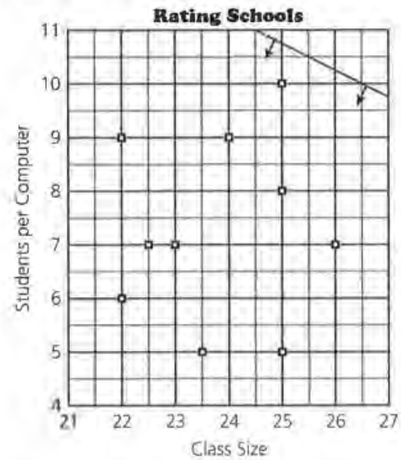
$$[2.7\ 1.7\ -.3\ .7\ -1.3\ -2.3\ 1.7\ -.3\ -.3\ -2.3] \cdot \begin{bmatrix} 2.7 \\ 1.7 \\ -.3 \\ .7 \\ -1.3 \\ -2.3 \\ 1.7 \\ -.3 \\ -.3 \\ -2.3 \end{bmatrix} = [26.1]$$

So the sum of the squared differences is 26.1. This is a measure of variability in the data.

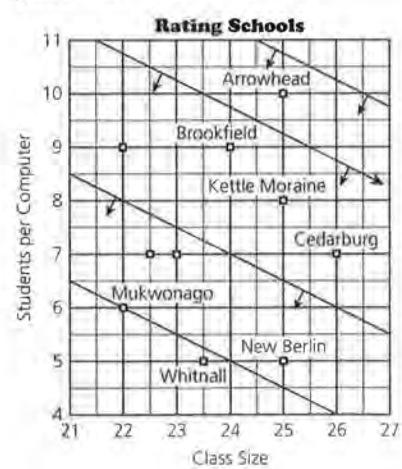
3. a. An equation with weights based on 9.7 students per computer (S/C) and an average class size (CS) of 17.3 could be written as

$\frac{1}{9.7}S/C + \frac{1}{17.3}CS = R$ or approximately $0.103S/C + 0.058CS = R$.

b. The slope of the sweeping line would be $-9.7/17.3$ or about -0.56. The line can begin at any point on the graph, either in the upper right corner or in the lower left corner near the origin. (In this case, a low rating is preferable; you want low class size and a low number of students per computer.)



Student explanations of the sweeping line will vary. The first point the line touches will have the lowest rating. In this case that will be (25, 10), which belongs to Arrowhead-Hartland High School. This indicates that with equal weights for the two variables, Arrowhead ranks lowest of the ten schools for the number of students per computer and class size.



As the line continues to sweep in, the second to the last point it will touch is Mukwonago at (22, 6) and the last will be Whitnall at (23.5, 5) indicating the schools in second and first places.

c.

School	Rating Equal Weights	Rank
Arrowhead-Hartland	2.48	10
Brookfield East	2.319	9
Cedarburg	2.229	7
Kettle Moraine	2.274	8
Mukwonago	1.894	2
New Berlin Eisenhower	1.965	3
Nicolet	2.203	6
Shorewood	2.055	5
West Allis Central	2.026	4
Whitnall	1.878	1

Students will agree or disagree depending on how they feel about the procedure.

d. Class size was weighted more heavily. The absolute value of the slope increased from about .56 to 4. To have a fraction increase in value means the numerator has to increase or the denominator decrease. The slope will be the weights: $\frac{W_{cs}}{W_{s/c}}$. This means that the coefficient of class size (the independent variable in this case) was increased or the coefficient of the response variable, students per computer, was decreased.

e. Any set of weights that satisfy the equation $CS + \frac{S}{C} = 33$ would make all three of the schools have equal ratings because all three of the points lie on the same straight line. Any variation of the equation is acceptable:

$$\frac{1}{33}CS + \frac{1}{33}\left(\frac{S}{C}\right) = 1$$

or any other equivalent equation.

4. a. It would be possible to have many students enrolled in Advanced Placement courses but none qualified to take the test, or to have only one student take and pass the test. The optimum would be to have 100% of those enrolled take and pass the test with a score of three or more.

b. $\frac{1}{64}P + \frac{1}{100}AP + \frac{1}{65.6}C = R$, where P is the percentage of those who pass the AP tests, AP is the percentage of those enrolled who took the tests, and C is the percentage of those who continue to a four-year school. This equation would represent a series of parallel planes for a set of possible ratings, R.

c.

School	Rating	Rank
Arrowhead-Hartland	2.14	8
Brookfield East	2.35	4
Cedarburg	2.26	5
Kettle Moraine	2.21	7
Mukwonago	2.09	9
New Berlin Eisenhower	1.79	10
Nicolet	2.76	2
Shorewood	2.92	1
West Allis Central	2.26	5
Whitnall	2.39	3

In this ranking, Shorewood is first, followed by Nicolet and Whitnall.

d. Student responses will vary. They should point out that when using computers and class size, the rankings yielded the top three as Whitnall, Mukwonago, and New Berlin Eisenhower. When Advanced Placement courses and college attendance are used as factors, a different set of three schools emerges: Shorewood, Nicolet, and Whitnall. The variables are measuring two very different qualities of schools, which may explain why the rankings are different.

5. This problem may be difficult for students.

For each school the ordered triple (enrolled AP, passed AP, college) would be: Arrowhead (11, 62, 70); West Allis (1, 95, 50); Shorewood (6, 96, 89). If a is the weight factor for percentage enrolled who took the AP test, b the weight factor for the percentage passing AP test, and c the percentage going on to a four-year college, the new system will be

$$11a + 62b + 70c = 745$$
$$a + 95b + 50c = 1005$$
$$6a + 96b + 89c = 1079.$$

The values for a, b, and c can be found by solving the matrix system. The weights would be $a = 5$, $b = 10$, and $c = 1$ to produce the equation $5AP + 10P + 1C = R$.

6. a. The equation used to find the total rating with the percentile score on a national test (T), the percentage of graduates continuing to a four-year college (C), the percentage enrolled in AP courses who took the AP test (AP), and the percentage who passed the AP test with a 3 or higher (P) is

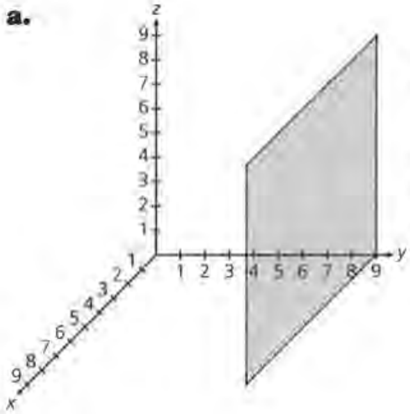
$$02T + .0152C + .01AP + .0156P = R.$$

The rankings using all four variables are in the table. Shorewood is first, followed by Nicolet. Using the four variables changed Whitnall from third to fourth, Brookfield from fourth to third, and changed the others only slightly.

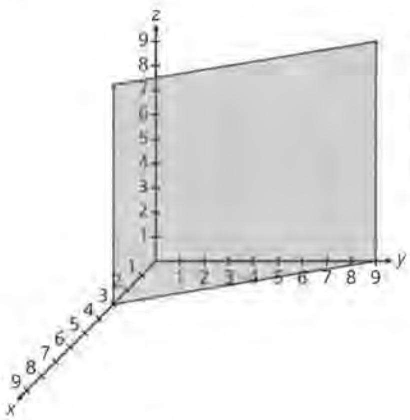
School	Rating	Rank
Arrowhead-Hartland	3.78	7
Brookfield East	4.11	3
Cedarburg	3.90	5
Kettle Moraine	3.83	6
Mukwonago	3.69	9
New Berlin Eisenhower	3.40	10
Nicolet	4.45	2
Shorewood	4.65	1
West Allis Central	3.71	8
Whitnall	3.95	4

b. Student responses will vary. They should mention that without some standardizing factor, mixing percentiles on tests with percentage taking a course involves combining numbers with different units. (In this case, they may argue that they are all based on 100, so it is not impossible to find another way to combine the data. The problem arises to a greater degree when class size and so on are included.) They may suggest other factors as important, such as the dropout rate or the average grade point of the students in the school.

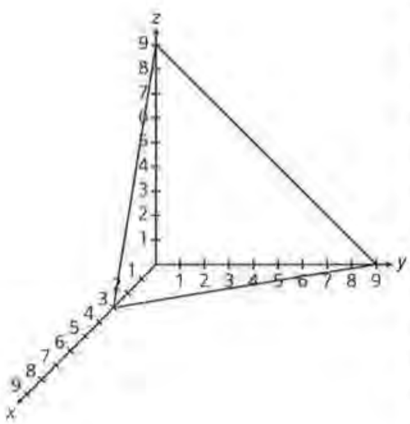
7. a.



$y = 9$ is a plane perpendicular to the y-axis at $y = 9$ and parallel to both the z-axis and the x-axis.



$3x + y = 9$ is a plane perpendicular to the xy-plane and parallel to the z-axis.



$3x + y + z = 9$ is a plane whose representation in the first octant can be observed from the traces connecting the x-intercept 3, the y-intercept 9, and the z-intercept 9.

b. The two planes are not parallel. Student justifications may differ. Some might suggest that they share the points (7, 0, 0) and (0, 7, 0) and thus, could not be parallel. Others may point out that the x- and y-intercepts are the same but the z-intercepts are different.

c. For the first plane, any equation that is a multiple of $10x + 7y + 7z = 70$ would have angles made by the traces with the coordinate planes as indicated.

Let:
a = the x-intercept
b = the y-intercept
c = the z-intercept

Then:
$\tan 35° = \frac{a}{c} = .7$

$\tan 45° = \frac{c}{b} = 1$

and
$a = .7c$
$c = b$

Thus, if $c = 10$, then $b = 10$ and $a = 7$. This would produce the equation $10x + 7y + 7z = 70$.

For the second plane, the equation that would have the given intercepts would be any multiple of the equation $30x + 6y + 5z = 30$.

Modeling

Teaching Notes

This quiz assesses what students know about the process of fitting a curve to a set of data. Some students may have strong backgrounds in this area and will only need to have some of the major points reviewed. If you suspect this might be the case, the quiz can be given as an assessment of how much students remember about the modeling process before you study the lesson. A rather different notion in the lesson is using a matrix system to represent the least squares linear regression equation. This may be very unfamiliar to students, despite their work in Lesson 1, because it is a different way to think about using regression. The quiz does not have students fit a model, but you can change the structure of the questions if you would like them to do so. The quiz is intended to probe their understanding of the process and the concepts involved rather than test their skills using a calculator to find a regression.

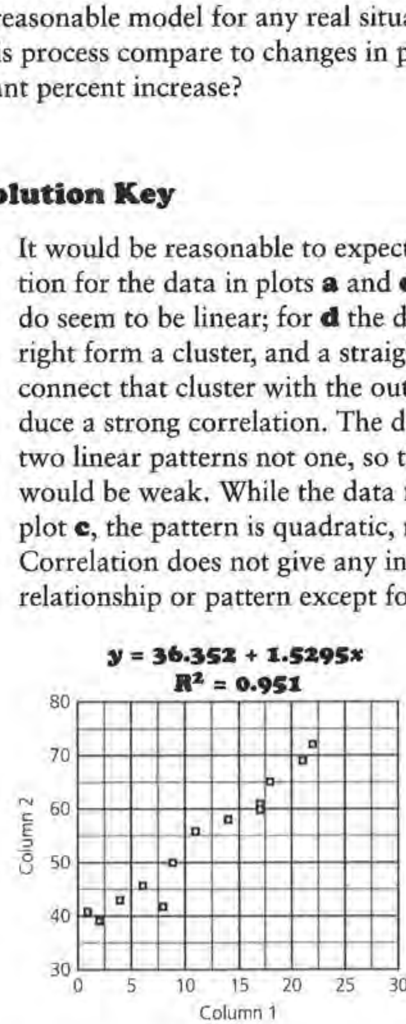
The first question focuses on correlation and how it can be misleading. Students study four plots to determine which might indicate a strong correlation coefficient relating the variables. Only one of the cases is really a linear relationship, even though several of the others might be interpreted that way. One of the plots is a quadratic, which will have a low correlation coefficient despite the pattern, because correlation measures only the strength of the linear relationship. The second question deals with residuals and how a pattern in the residuals relates to the original model. They also have to explain how the sum of squared residuals relates to linear regression (the least squares regression equation will minimize the sum of squared residuals) and to the plot of the data. Finally, they are given a set of data and the model written in matrix form and asked to explain how it will work.

The bonus question may take some time. You might have students download the data into their calculators and take the question home as an option. This is one way to see how well they can use their calculators, and you could raise questions about the recursive nature of the process. Is the process an example of recursion? What was the starting point? (the first data set) What is the rule? (the least squares regression equation for (1993, 1994)) What will happen to the model over time? Do you think this would ever be a reasonable model for any real situation? How does this process compare to changes in price due to a constant percent increase?

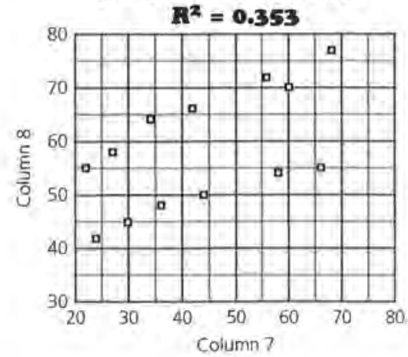
Solution Key

1. It would be reasonable to expect a strong correlation for the data in plots **a** and **d**. For **a**, the data do seem to be linear; for **d** the data in the lower right form a cluster, and a straight line would connect that cluster with the outlier point to produce a strong correlation. The data in plot **b** form two linear patterns not one, so the correlation would be weak. While the data form a pattern for plot **c**, the pattern is quadratic, not linear. Correlation does not give any indication about a relationship or pattern except for linearity.

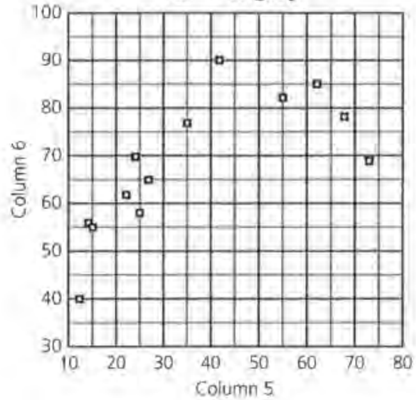
$y = 36.352 + 1.5295x$
$R^2 = 0.951$



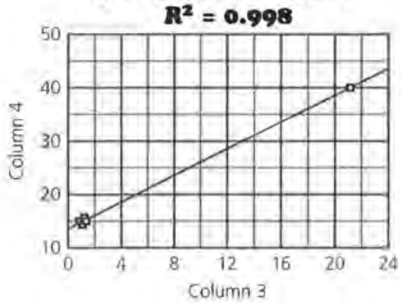
$y = 40.840 + 0.39697x$
$R^2 = 0.353$



$y = 51.066 + 0.47077x$
$R^2 = 0.509$



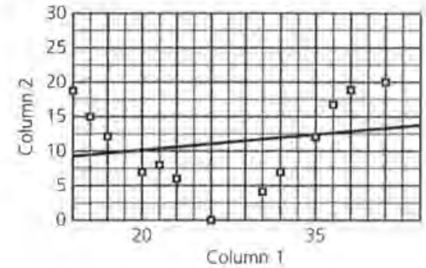
$y = 13.442 + 1.2638x$
$R^2 = 0.998$



2. The residual plot shows that the actual data points are above those predicted for domain elements that are less than 20 and for those that are greater than 35. For the data points between 20 and 35, the predicted values exceeded the actual values. Whenever there is a clear pattern in the residuals, when you can predict the error, the model is not appropriate.

Student plots will vary, but in each case, the points should reflect the statements above (data points above the model for less than 20 and greater than 35, below the model for those in between 20 and 35). Do not mark off for the incorrect number of data points or for the wrong placement relative to each other, but rather check to see if their graph has the proper arrangement of points with respect to the model students used.

Sample:



3. **a.** The wages were highest in Germany both years. Student responses will vary for the explanation. They might include points such as: high hourly wages do not necessarily mean German workers are better off financially than those in other countries, but could mean that cost of living is higher, the number of hours they work on a regular basis is lower, or a variety of other factors could be involved.

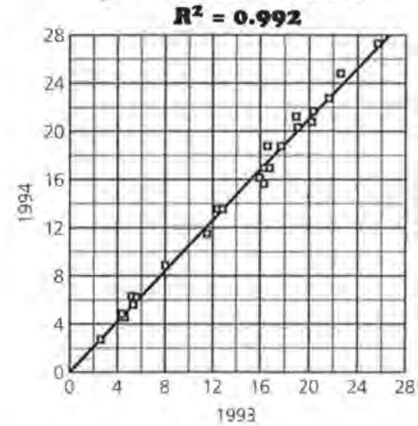
b. The first matrix contains the wages in 1993; the second the slope and intercept of the linear model used to predict the wages for 1994. $P\hat{W}$ is the set of predicted wages. The actual equation would be: $PW = 1.056W + 0.104$.

For example, in Mexico = \$2.56 in 1993, the system would give you

$$[1\ 2.56] \times \begin{bmatrix} .104 \\ 1.056 \end{bmatrix} = [2.81] .$$

In 1994 the wages for manufacturing jobs predicted for Mexico were \$2.81 per hour.

$y = 0.10386 + 1.0556x$
$R^2 = 0.992$



(3) c. The system would predict \$12.94 for hourly wages in 1994 in Ireland.

d. The residual for the United States would be the difference between the actual wage, \$17.10, and the predicted wage, \$17.77, \$ −0.67. The model would predict 67 cents too much for the wages in the United States.

Bonus

Let Y1 = 1.056X + .104.
Then,
Y1(1993) = Wages in 1994
Y1(1994) = Wages in 1995
Y1(1995) = Wages in 1996
Y1(1996) = Wages in 1997

Country	1993	1997	Country	1993	1997
Mexico	2.56	3.64	Denmark	19.11	24.22
United States	16.73	21.26	Finland	16.56	21.05
Canada	16.33	20.76	France	16.23	20.64
Australia	12.49	15.98	Germany	25.70	32.41
Hong Kong	4.29	5.79	Italy	16.00	20.35
Japan	19.01	24.09	Netherlands	20.21	25.58
Korea	5.51	7.30	Norway	20.21	25.58
New Zealand	8.01	10.41	Portugal	4.50	6.05
Singapore	5.25	6.98	Spain	11.50	14.75
Taiwan	5.22	6.94	Sweden	17.70	22.46
Austria	20.37	25.78	Switzerland	22.63	28.59
Belgium	21.62	27.34	United Kingdom	12.76	16.32

Source: 1993 U.S. Department of Labor, World Trade. September 1995

Buying a Home

Teaching notes

This quiz can be given after teaching Lesson 8. It can be given in class and will probably take most of a class period, or as a take-home assignment that will be graded. The questions are worded to find out what students understand about the iterative process of finding a multiple regression model without having them complete the actual process. The data provide a link to the work students did in the first unit on rating schools.

Students are expected to find a multiple regression equation with four variables using the matrix formula, so you may want to provide them with the data to download into their calculators or computers. Caution them that they should have some way to protect the original set of data as they work, or they will find themselves reentering data and wasting time. Some careful thought about how they will approach the problems and what they will do will help students complete the work in a timely and efficient way.

The last problem poses a real question for students. Is the seller actually asking a reasonable price for a house given the information about the surrounding neighborhood? These are decisions that people have to make in their lives as they confront many different situations, not only those involved in purchasing a home. The problem does not address the challenge of obtaining money for the purchase, which could be another lesson in mathematics and might make a nice extension for the quiz. Students could investigate the actual loan costs in their area and decide, based on that data, how they might go about financing the house and what it would actually cost them to do so.

Solution Key

1. **a.** Student responses will vary. They will probably suggest the higher the rating for the school system and the greater the amount of taxes paid, the more the cost of the house. They should recognize that there will probably be an association and be able to describe it using appropriate vocabulary.

b. Student responses will vary. Some of their suggestions might include condition of house, size of house, size of lot, location, schools, number of bedrooms, age.

2. **a.** The next step would be to find the residuals for (rating, cost) and use the taxes to try to explain part of the residuals or error in the model. To do this, you would find a regression equation for (taxes, first residuals). The process would continue by using rating once again to try to explain the new residuals produced by the model (taxes, first residuals), and so on.

b. The *e* represents the difference between the actual prices of the houses and the prices predicted by using the regression model. This is what you are trying to explain by involving another variable. As you continue the process of iteration, *e* should tend to become smaller and smaller for each house, eventually approaching a limit.

c. The sum of squared residuals for the model in part a (C_R) would be larger than the sum of squared residuals after the second regression (C_{RT}), and that would in turn have a larger sum of squares than the next regression model (C_{RTR}). As the iterative process continues, the sum of squares decreases and gradually approaches a limiting value. (Note that the sum of squares seems quite large but actually is a root mean squared error of around ±$15,000 housing prices.)

(2) **d.** The final results would be the same, but the order of regression would be reversed. Eventually both orders will produce the same regression model.

3. **a.** The strongest relation appears to be in plots B and C between taxes and price and between square feet of house and price (although the two newer houses make plot D appear to have a linear pattern). Lot size and age of the house do not really seem to contribute much to the price. The correlation coefficients are as follows:

(lot size, price): 0.11; (taxes, price): 0.97; (house size, price): 0.87; (age, price): –0.84.

Note that this last correlation is due to those two relatively new houses that seem to be outliers. The rest of the plot indicates that the correlation might not be as strong as the outliers would indicate, which makes sense. Some old houses are very expensive, while some new ones can be inexpensive. The strength of the correlation would also depend on the neighborhood and the recent building that has occurred.

b. The matrix system would be **Db** = **P** or

$$[1\ L\ T\ H\ A] \cdot \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = [P]$$

where **D** is the 12×5 data matrix containing a column of 1s, the lot size, tax amount, house size, and age; **b** is the 5×1 coefficient matrix; **P** is the 12×1 actual price matrix.

4. The regression model would be:

$P = 104524 + 0.198149L + 6.82169T + 13.9178H - 558.264A$.

(If students round the decimals, their results will be different.)

a. Student examples will vary. They might use the given data or create a set of their own. For example, for the first house on the list, $P =$ \$117,739. Students can suggest that the prediction be found using the matrix system **Db**, or they can evaluate the formula for the given set of data.

b. The first house has the largest absolute residual, –4839. The house with the smallest absolute residual is the last one with a residual of –242. For one house where the predicted value is greater than the actual value, the difference between the actual and predicted values using the multiple regression model is \$4838. The smallest difference between actual and predicted values is –\$242.

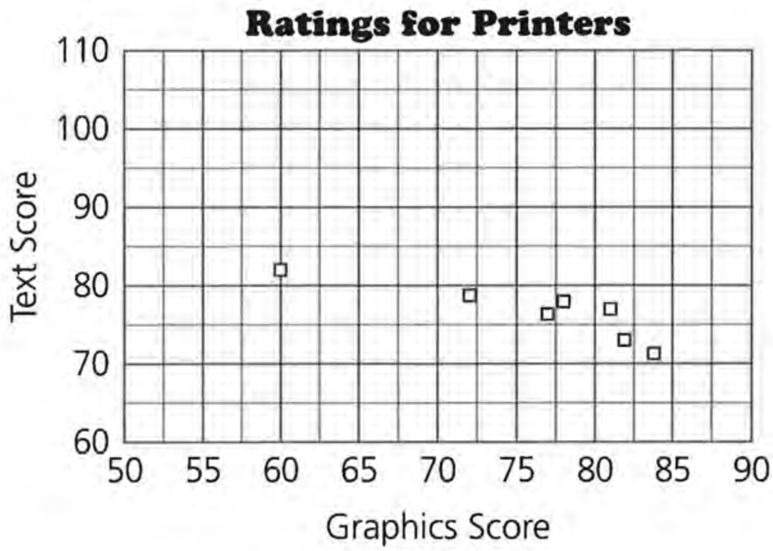
c. The sum of squared residuals is 146,825,374. The root mean squared error is \$3498, so the predicted prices are typically within about \$3498 of the actual price. Students may find the sum of squared residuals different ways, but the most efficient is probably to find the residual matrix R by subtracting the matrix for the predicted prices from the actual price matrix, then calculate R^TR.

5. **a.** The predicted price for the house would be \$154,171.

b. Student responses will vary. The root mean squared error is \$3498, which would indicate a price that typically would vary from \$150,673 to \$157,669. The price of \$165,000 is higher than the predicted price by about 3 times the root mean squared error, so the asking price seems too high.

Ratings for Printers

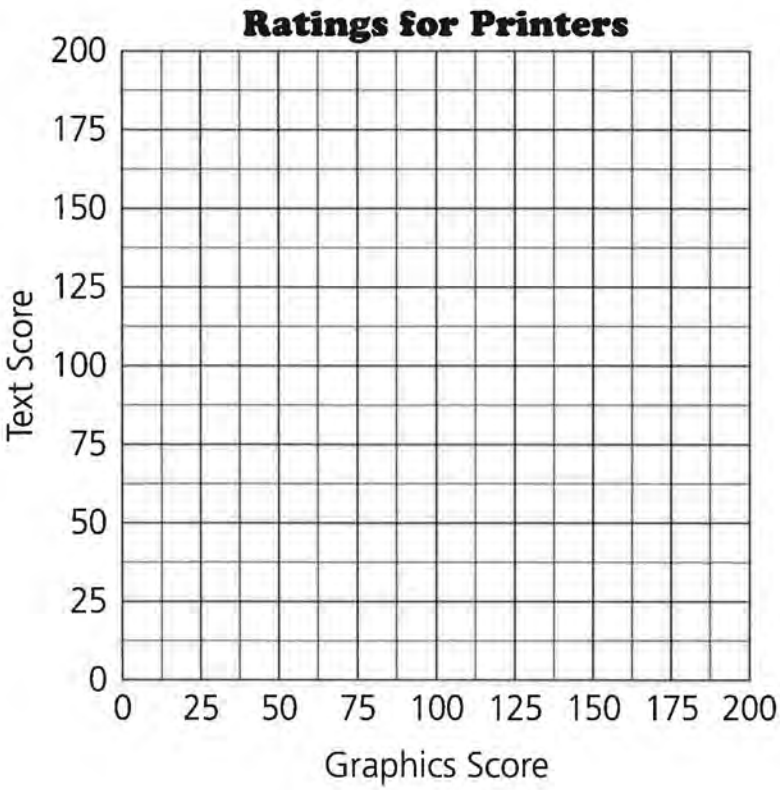
NAME ______________________________



ACTIVITY SHEET 2

Ratings for Printers

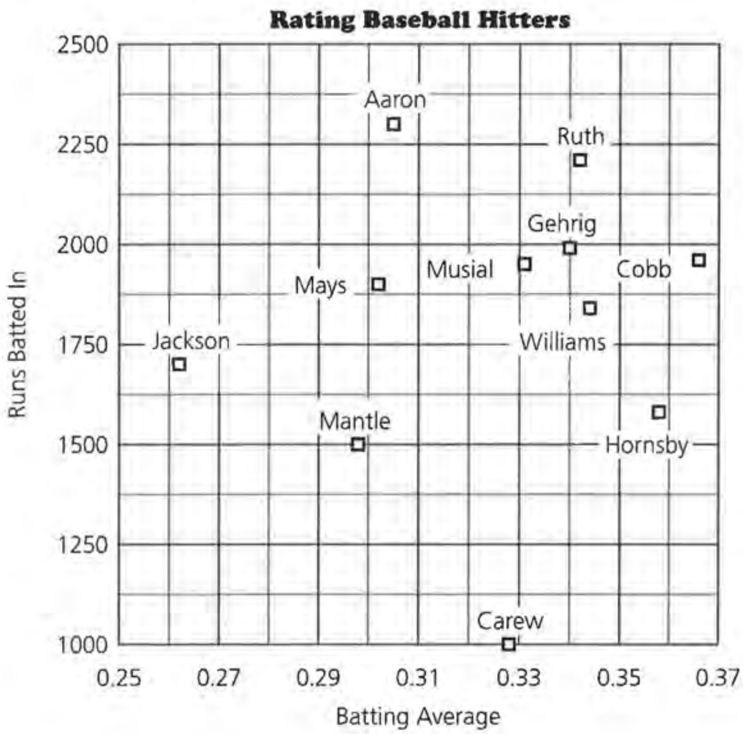
NAME ____

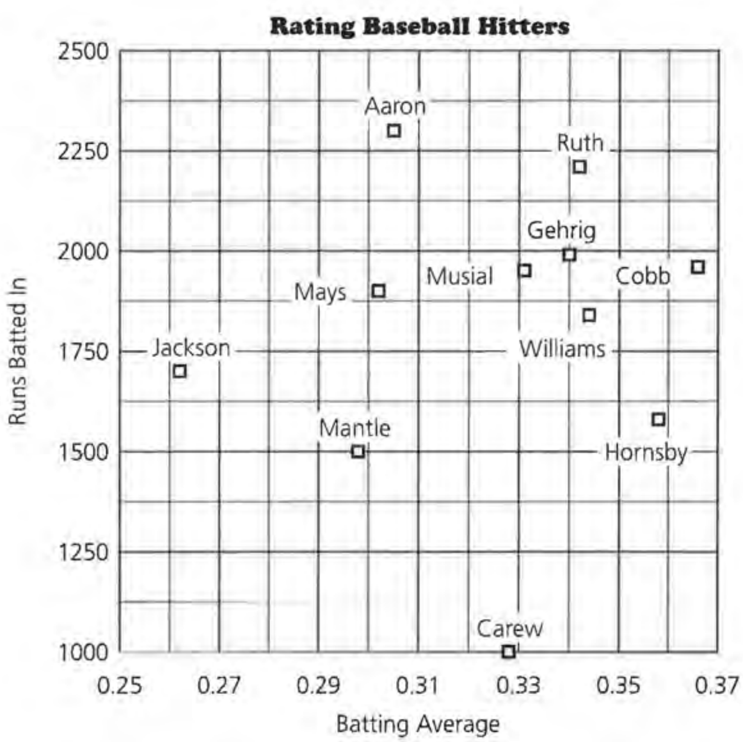


ACTIVITY SHEET 3

Hall of Fame Best Hitters

NAME ______________________________

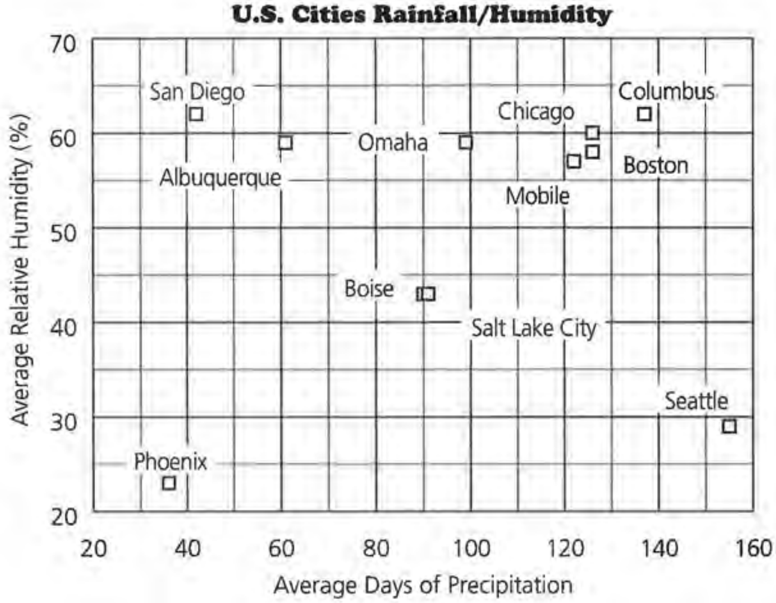




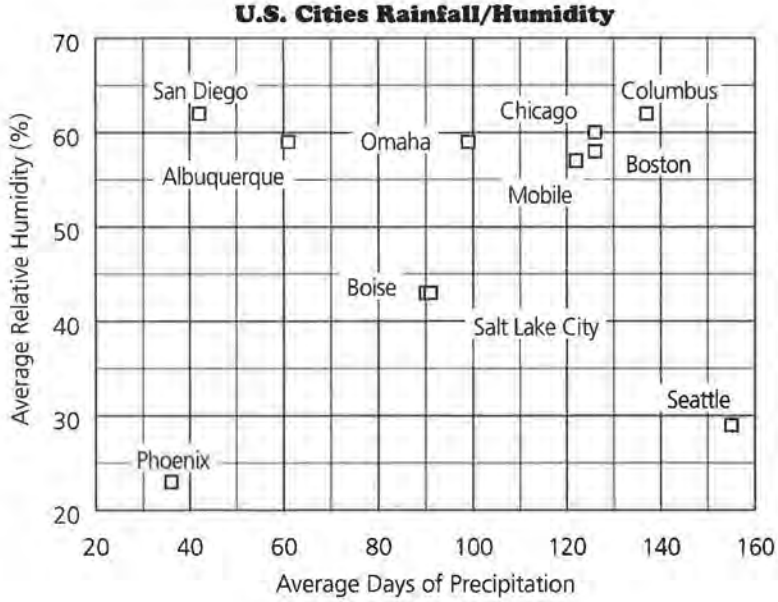
Rainfall and Humidity

NAME ______________________________

U.S. Cities Rainfall/Humidity



U.S. Cities Rainfall/Humidity



Data-Driven Mathematics Procedures for Using the TI-83

I. Clear menus

ENTER will execute any command or selection. Before beginning a new problem, previous instructions or data should be cleared. Press **ENTER** after each step below.

1. To clear the function menu, **Y=**, place the curser any place in each expression, **CLEAR**
2. To clear the list menu, **2nd MEM ClrAllLists**
3. To clear the draw menu, **2nd Draw ClrDraw**
4. To turn off any statistics plots, **2nd STATPLOT PlotsOff**
5. To remove user created lists from the Editor, **STAT SetUpEditor**

II. Basic information

1. A rule is active if there is a dark rectangle over the option. See Figure l.

```
Plot1 Plot2 Plot3
\Y1 = 170 + 7(X–72)
\Y2 = –.35X – 22
\Y3 =
\Y4 =
\Y5 =
\Y6 =
\Y7 =
```

FIGURE 1

On the screen above, Y1 and Plot1 are active; Y2 is not. You may toggle Y1 or Y2 from active to inactive by putting the cursor over the = and pressing **ENTER**. Arrow up to **Plot1** and press **ENTER** to turn it off; arrow right to **Plot2** and press **ENTER** to turn it on, etc.

2. The Home Screen (Figure 2) is available when the blinking cursor is on the left as in the diagram below. There may be other writing on the screen. To get to the Home Screen type **2nd QUIT**. You may also clear the screen completely by typing **CLEAR**.

```
7 → X
                7
Y1
               11
█
```

FIGURE 2

3. The variable x is accessed by the **X, T, Θ, *n*** key.
4. Replay option: **2nd ENTER** allows you to back up to an earlier command. Repeated use of **2nd ENTER** continues to replay earlier commands.
5. Under **MATH**, the **MATH** menu has options for fractions to decimals and decimals to fractions, for taking nth roots and for other mathematical operations. **NUM** contains the absolute value function as well as other numerical operations (Figure 3).

```
MATH NUM CPX PRB
1: abs(
2: round(
3: iPart(
4: fPart(
5: int(
6: min(
7↓max(
```

FIGURE 3

III. The STAT Menus

1. There are three basic menus under the **STAT** key: **EDIT, CALC** and **TESTS**. Data are entered and modified in the **EDIT** mode; all numerical calculations are made in the **CALC** mode; statistical tests are run in the **TEST** mode.
2. **Lists and Data Entry**
   Data is entered and stored in Lists (Figure 4). Data will remain in a list until the list is cleared. Data can be cleared using **CLEAR** L_i or (List name), or by placing the cursor over the List heading and selecting **CLEAR ENTER**. To enter data, select **STAT EDIT** and with the arrow keys move the cursor to the list you want to use.

Type in a numerical value and press **ENTER.** Note that the bottom of the screen indicates the List you are in and the list element you have highlighted. 275 is the first entry in L1. (It is sometimes easier to enter a complete list before beginning another.)

L1	L2	L3
275	67	190
5311	144	120
114	64	238
2838	111	153
15	90	179
332	68	207
3828	94	153
L1 (1) = 275		

FIGURE 4

For data with varying frequencies, one list can be used for the data, and a second for the frequency of the data. In Figure 5 below, the L5(7) can be used to indicate that the seventh element in list 5 is 4, and that 25 is a value that occurs 4 times.

L4	L5	L6
55	1	-------
50	3	
45	6	
40	14	
35	12	
30	9	
25	4	
L5 (7) = 4		

FIGURE 5

3. **Naming Lists**
Six lists are supplied to begin with. L_1, L_2, L_3, L_4, L_5, and L_6 can be accessed also as **2nd L_i**. Other lists can be named using words as follows. Put the cursor at the top on one of the lists. Press **2nd INS** and the screen will look like that in Figure 6.

	L1	L2 1
	-------	-------
Name =		

FIGURE 6

The alpha key is on so type in the name (up to five characters)and press **ENTER** (Figure 7).

PRICE	L1	L2 2
	-------	-------
PRICE(1) =		

FIGURE 7

Then enter the data as before. (If you do not press **ENTER** the cursor will remain at the top and the screen will say error: data type.) The newly named list and the data will remain until you go to Memory and delete the list from the memory. To access the list for later use, press **2nd LIST** and use the arrow key to locate the list you want under the **NAMES** menu. You can accelerate the process by typing **ALPHA P** (for price). (Figure 8) Remember, to delete all but the standard set of lists from the editor, use SetUp Editor from the **STAT** menu.

NAMES OPS MATH
↑PRICE
: RATIO
: RECT
: RED
: RESID
: SATM
↓SATV

FIGURE 8

4. Graphing Statistical Data

General Comments

- Any graphing uses the **GRAPH** key.
- Any function entered in Y1 will be graphed if it is active. The graph will be visible only if a suitable viewing window is selected.
- The appropriate x and y scale can be selected in **WINDOW**. Be sure to select a scale that is suitable to the range of the variables.

Statistical Graphs

To make a statistical plot, select **2nd Y=** for the **STAT PLOT** option. It is possible to make three plots concurrently if the viewing windows are identical. In Figure 9, Plots 2 and 3 are off, Plot 1 is a scatterplot of the data (Costs, Seats), Plot 2 of (L3,L4), and Plot 3 is a boxplot of the data in L3.

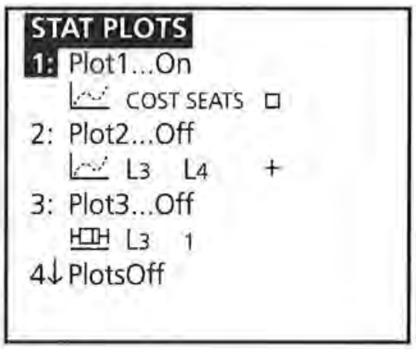


FIGURE 9

Activate one of the plots by selecting that plot and pressing **ENTER.**

Choose **ON**, then use the arrow keys to select the type of plot (scatter, line, histogram, box plot with outliers, box plot, or Normal probability plot). (In a line plot, the points are connected by segments in the order in which they are entered. It is best used with data over time.) Choose the lists you wish to use for the plot. In the window below, a scatter plot has been selected with the x-coordinate data from COSTS, and the y-coordinate data from SEATS. (Figure 10) (When pasting in list names, **2nd LIST,** press **ENTER** to activate the name and **ENTER** again to locate the name in that position.)

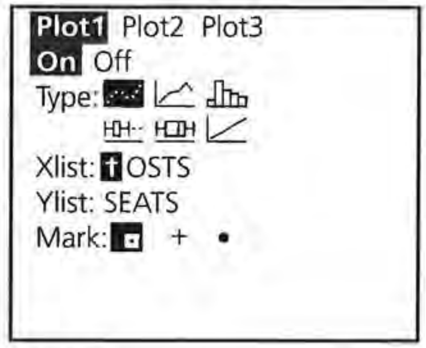


FIGURE 10

For a histogram or box plot you will need to select the list containing the data and indicate whether you used another list for the frequency or are using 1 for the frequency of each value. The x scale selected under **WINDOW** determines the width of the bars in the histogram. It is important to specify a scale that makes sense with the data being plotted.

5. Statistical Calculations

One variable calculations such as mean, median, maximum value of the List, standard deviation, and quartiles can be found by selecting **STAT CALC 1-Var Stats** followed by the list in which you are interested. Use the arrow to continue reading the statistics. (Figures 11, 12, 13)

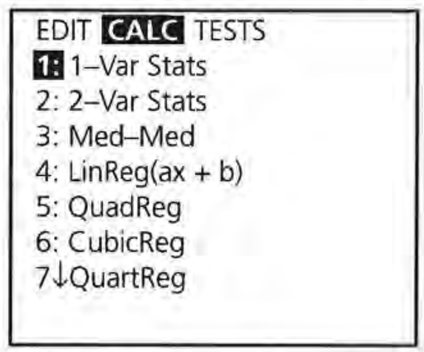


FIGURE 11

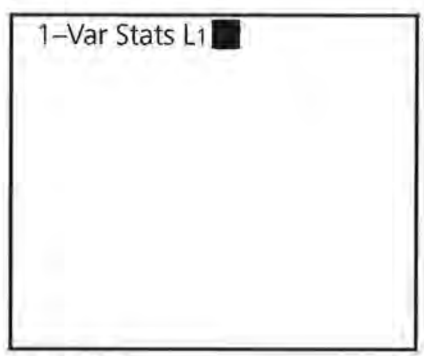


FIGURE 12

```
1–Var Stats
x̄ = 1556.20833
Σx = 37349
Σx² = 135261515
Sx = 1831.353621
σx = 1792.79449
↓n = 24
■
```

FIGURE 13

Calculations of numerical statistics for bivariate data can be made by selecting two variable statistics. Specific lists must be selected after choosing the **2-Var Stats** option. (Figure 14)

```
2–Var Stats L1, L
2■
```

FIGURE 14

Individual statistics for one or two data sets can be obtained by selecting **VARS Statistics,** but you must first have calculated either 1-Var or 2-Var Statistics. (Figure 15)

```
X/Y Σ EQ TEST PTS
1: n
2: x̄
3: Sx
4: σx
5: ȳ
6: Sy
7↓σy
```

FIGURE 15

6. **Fitting Lines and Drawing their graphs**
Calculations for fitting lines can be made by selecting the appropriate model under **STAT CALC: Med-Med** gives the median fit regression; **LinReg** the least squares linear regression and so on. (Note the only difference between **LinReg (ax+b)** and **LinReg (a+bx)** is the assignment of the letters a and b.) Be sure to specify the appropriate lists for x and y. (Figure 16)

```
Med–Med LCal, LFA
CAL■
```

FIGURE 16

To graph a regression line on a scatter plot, follow the steps below:

- Enter your data into the Lists.
- Select an appropriate viewing window and set up the plot of the data as above.
- Select a regression line followed by the lists for x and y, **VARS Y-VARS Function** (Figure 17, 18) and the Yi you want to use for the equation, followed by **ENTER.**

```
VARS Y-VARS
1: Function...
2: Parametric...
3: Polar...
4: On/Off...
```

FIGURE 17

```
Med–Med _CAL, LFA
CAL, Y1
```

FIGURE 18

The result will be the regression equation pasted into the function Y1. Press **GRAPH** and both the scatter plot and the regression line will appear in the viewing window. (Figure 19, 20).

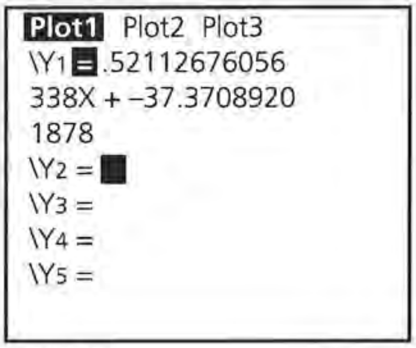


FIGURE 19

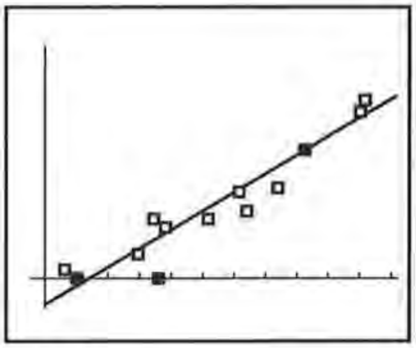


FIGURE 20

- There are two cursors that can be used in the graphing screen.

TRACE activates a cursor that moves along either the data (Figure 21) or the function entered in the Y variable menu (Figure 22). The coordinates of the point located by the cursor are given at the bottom of the screen.

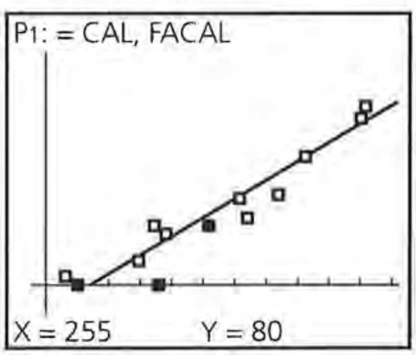


FIGURE 21

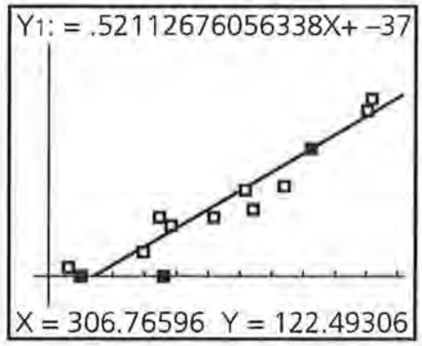


FIGURE 22

Pressing **GRAPH** returns the screen to the original plot. The up arrow key activates a cross cursor that can be moved freely about the screen using the arrow keys. See Figure 23.

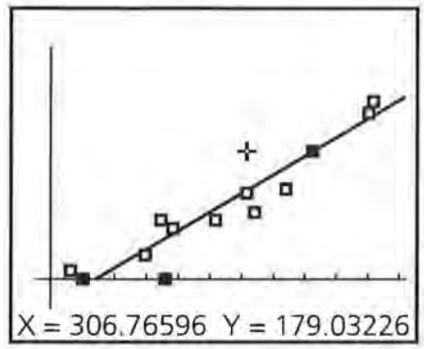


FIGURE 23

Exact values can be obtained from the plot by selecting **2nd CALC** Value. Select **2nd CALC Value ENTER.** Type in the value of x you would like to use, and the exact ordered pair will appear on the screen with the cursor located at that point on the line. (Figure 24)

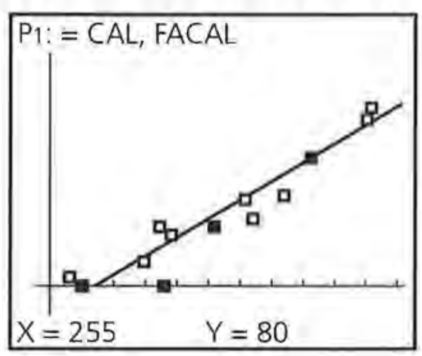


FIGURE 24

IV. Evaluating an expression:

To evaluate $y = .225x\text{-}15.6$ for $x = 17$, 20 and 24 you can:

1. Type the expression in Y1, return to the home screen, 17 **STO X,T,Θ,n ENTER, VARS Y1=1VARS Function Y1 ENTER ENTER**. (Figure 25)

```
17 → X
                 17
Y1
        −11.75833333
█
```

FIGURE 25

Repeat the process for $x = 20$ and 24.

2. Type $17^2 - 4$ for $x = 17$, **ENTER** (Figure 26) Then use **2nd ENTRY** to return to the arithmetic line. Use the arrows to return to the value 17 and type over to enter 20.

```
.225*17 − 15.6
                −11.775
.225*2█ − 15.6
```

FIGURE 26

You can also find the value of x by using the table command. Select **2nd TblSet** (Figure 27). (Y1 must be turned on.) Let **TBlMin** = 17, and the increment **ΔTbl** = 1.

```
TABLE SETUP
  TblMin = 17
  ΔTbl = █
Indpnt:  Auto  Ask
Depend:  Auto  Ask
```

FIGURE 27

Select **2nd TABLE** and the x and y-values generated by the equation in Y1 will be displayed. (Figure 28)

X	Y1	
17	-11.76	
18	-11.53	
19	-11.31	
20	-11.08	
21	-10.86	
22	-10.63	
23	-10.41	
X = 17		

FIGURE 28

V. Operating with Lists

1. A list can be treated as a function and defined by placing the cursor on the label above the list entries. List 2 can be defined as L1 + 5. (Figure 29)

L1	L2	L3
275	-------	190
5311		120
114		238
2838		153
15		179
332		207
3828		153
L2 = L1 + 5		

FIGURE 29

Pressing **ENTER** will fill List 2 with the values defined by L1+5. (Figure 30)

L1	L2	L3
275	280	190
5311	5316	120
114	119	238
2838	2843	153
15	20	179
332	337	207
3828	3833	153
L2(1) = 280		

FIGURE 30

2. List entries can be cleared by putting the cursor on the heading above the list, and selecting **CLEAR** and **ENTER.**

3. A list can be generated by an equation from Y= over a domain specified by the values in Li by putting the cursor on the heading above the list entries. Select **VARS Y-VARS Function Y1 ENTER (L1) ENTER.** (Figure 31)

L1	L2	L3
120	12	--------
110	14	
110	12	
110	11	
100	7	
100	6	
120	9	
L3 = Y1(L1)		

FIGURE 31

4. The rule for generating a list can be attached to the list and retrieved by using quotation marks (**ALPHA** +) around the rule (Figure 32). Any change in the rule (Y1 in the illustration) will result in a change in the values for L1. To delete the rule, put the cursor on the heading at the top of the list, press **ENTER,** and then use the delete key. (Because L1 is defined in terms of **CAL,** if you delete **CAL** without deleting the rule for L1 you will cause an error.)

CAL	FACAL	L1 5
255	80	-------
305	120	
410	180	
510	250	
320	90	
370	125	
500	235	
L1 = "Y1(LCAL)"		

FIGURE 32

VI. Using the DRAW Command

To draw line segments, start from the graph of a plot, press **2ND DRAW** and select **Line(.** (Figure 33)

DRAW POINTS STO
1: ClrDraw
2: Line(
3: Horizontal
4: Vertical
5: Tangent(
6: DrawF
7↓Shade(

FIGURE 33

This will activate a cursor that can be used to mark the beginning and ending of a line segment. Move the cursor to the beginning point and press **ENTER;** use the cursor to mark the end of the segment, and press **ENTER** again. To draw a second segment, repeat the process. (Figure 34)

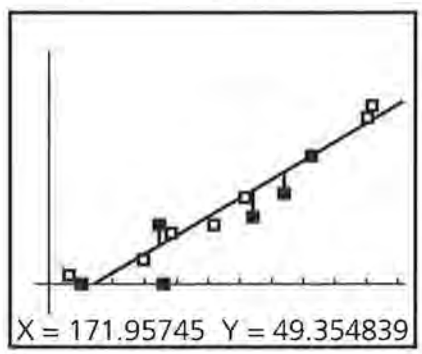


FIGURE 34

VII. Random Numbers

To generate random numbers, press **MATH** and **PRB**. This will give you access to a random number function, **rand**, that will generate random numbers between 0 and 1 or **randInt(** that will generate random numbers from a beginning integer to an ending integer for a specified set of numbers. (Figure 35) In Figure 36, five random numbers from 1 to 6 were generated.

```
MATH NUM CPX PRB
1: rand
2: nPr
3: nCr
4: !
5: randInt(
6: randNorm(
7: randBin(
```

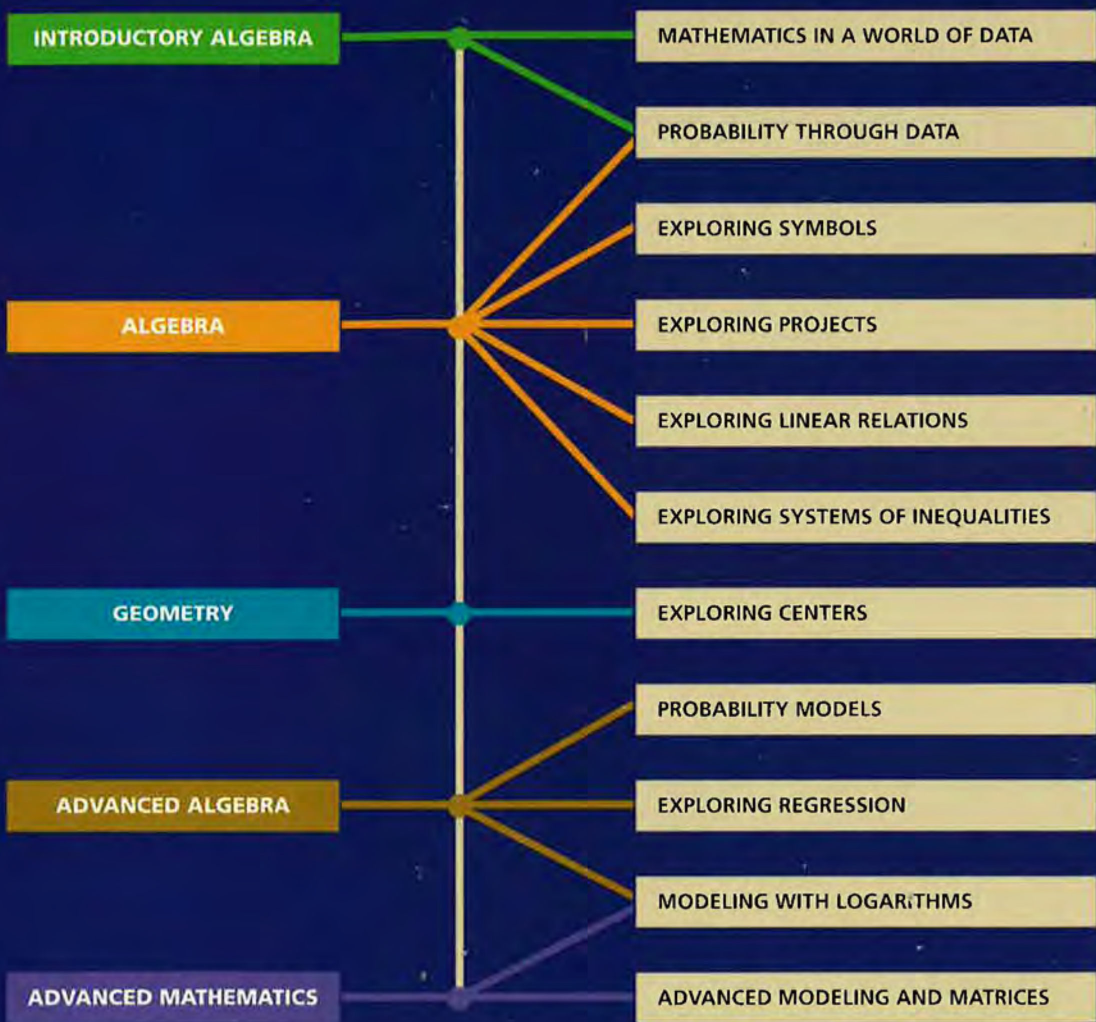
FIGURE 35

```
randInt(1, 6, 5,)
      (2 4 2 5 3)
```

FIGURE 36

Pressing **ENTER** will generate a second set of random numbers.

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 322592
ISBN 1-57232-259-4
21190