Eighth Edition

Measurement by the Physical Educator

Why and How

David K. Miller

McGraw Hill Education

Measurement by the Physical Educator

Why and How

Eighth Edition

David K. Miller

University of North Carolina Wilmington (retired)

Mc Graw Hill Education

MEASUREMENT BY THE PHYSICAL EDUCATOR: WHY AND HOW, EIGHTH EDITION

Published by McGraw-Hill Education, 2 Penn Plaza, New York, NY 10121.

This book is printed on acid-free paper.

1 2 3 4 5 6 7 8 9 LCR 21 20 19

ISBN 978-1-259-92242-8 (bound edition)
MHID 1-259-92242-1 (bound edition)

ISBN 978-1-260-39778-9 (loose-leaf edition)
MHID 1-260-39778-5 (loose-leaf edition)

Product Developer: *Francesca King*
Marketing Manager: *Meredith Leo*
Content Project Manager: *Lisa Bruflodt, Emily Windelborn*
Buyer: *Laura Fuller*
Designer: *Beth Blech*
Content Licensing Specialist: *Jacob Sullivan*
Compositor: *Lumina Datamatics, Inc.*

Library of Congress Cataloging-in-Publication Data

Names: Miller, David K. (David Keith) author.
Title: Measurement by the physical educator : why and how / David K. Miller, University of North Carolina Wilmington (retired).
Description: Eighth edition. | New York, NY : McGraw-Hill Education, [2020] | Includes bibliographical references and index.
Identifiers: LCCN 2018051298| ISBN 9781259922428 (hardback) | ISBN 1259922421 (bound edition) | ISBN 9781260397789 (loose-leaf edition) | ISBN 1260397785 (loose-leaf edition)
Subjects: LCSH: Physical fitness–Testing. | Physical fitness–Testing–Statistics. | BISAC: SPORTS & RECREATION / General.
Classification: LCC GV436 .M54 2020 | DDC 613.7–dc23
LC record available at https://lccn.loc.gov/2018051298

mheducation.com/highered

*Dedicated to my wife, Sandra, for your love and support.
You are the joy of my life.*

CONTENTS

Mc Graw Hill Education
5011222

This page intentionally left blank

PREFACE

Purpose and Content

Students in measurement and evaluation classes often are bombarded with an abundance of information. Regrettably, some students complete the class with a little knowledge in many areas but no confidence or skills to perform the procedures and techniques presented in the class. As professionals in school or nonschool settings, these same students often do not measure and assess knowledge, and physical performance, in the proper way.

The purpose of this text is to help the physical education, exercise science, or kinesiology major develop the necessary confidence and skills to conduct measurement techniques properly and effectively. However, more than just measurement techniques are presented. Emphasis is placed upon the reasons for the measurement and the responsibilities after measurement is completed. These inclusions should help the student develop an appreciation of the need for measurement in a variety of settings. In addition, every effort has been made to present all the material in an uncomplicated way, and only practical measurement techniques are included.

Upon successful completion of the chapter objectives, the user of this text should be able to

1. Understand, explain, and use the professional terminology presented in the text.
2. Use and interpret fundamental statistical techniques.
3. Select appropriate knowledge and psychomotor tests.
4. Construct good psychomotor tests.
5. Construct good objective and subjective knowledge tests.
6. Objectively assess and grade students who participate in a physical education class.
7. Administer psychomotor and sports skills tests, interpret the results, and prescribe activities for the development of psychomotor and sports skills.
8. Administer body structure and composition tests, interpret the results, and prescribe scientifically sound methods for attainment of a healthy percentage of body fat.
9. Administer functional fitness tests to older adults, interpret the results, and prescribe activities for the development of functional fitness.
10. Administer psychomotor tests to special-needs populations, interpret the results, and prescribe activities for the development of psychomotor skills.

Audience

At one time, most undergraduate physical education majors planned to teach in grades K through 12. Today many majors in physical education, exercise science, kinesiology, and other similar subject areas anticipate a career in the nonschool environment. This book is designed for use by majors preparing for either environment—school or nonschool. With the exceptions of the construction of knowledge tests and grading skills, all of the competencies presented in this book will be expected in a variety of professional settings.

Additionally, the term *physical educator* is used throughout the text to refer to individuals who perform the professional responsibilities presented in the chapters. When working through the physical attributes to better the lives of others, we all are physical educators.

Organization

The text is organized so that the student will develop fundamental statistics skills early in the course (chapters 1–4). These skills are to be demonstrated throughout the text. Chapter 5, "What Is a Good Test?" describes the criteria of a good test. Since these criteria and related terms are used throughout the text, it is recommended that this chapter be covered before the chapters that follow. Chapter 6, "Construction of Knowledge Tests," and chapter 7, "Assessment and Grading," may be covered in the sequence presented or later in the course. It is recommended that chapter 8, "Construction and Administration of Psychomotor Tests," be presented before any discussion of psychomotor testing. The components of health-related physical fitness and skill-related physical fitness (chapters 9–14) are described before the presentation of health-related and skill-related physical fitness tests (chapter 15, "Physical Fitness") so that the student will better understand these components. Many practical tests are included in chapters 9–15. These chapters may be presented in a different sequence if the instructor wishes to do so. Chapter 16 on functional fitness of "Older Adults"; chapter 17, "Special-Needs Populations"; and chapter 18, "Sports Skills" also may be presented in a different sequence by the instructor.

Approach

The statistics information is presented in a friendly and simplified manner so that it is nonintimidating. In addition, although the text information is sometimes presented in a "nuts-and-bolts" style, it is comprehensive as well as straightforward, accurate, and practical.

This book and related assignments can be completed without the use of a microcomputer, but the microcomputer can be applied in a variety of ways. No instruction for the use of a statistical software program is provided. A variety of such programs are available, however, and the program of choice may be used by the instructor and students. The psychomotor tests presented in chapters 9–15 are practical, inexpensive, and satisfactory for both sexes. They require little, if any, special equipment, and norms are included with many tests. The lack of norms should not limit the use of any test, however. Upon completion of chapter 2, the student should have the ability to construct norms.

Pedagogy

The following features of this book will assist the student in mastering the material:

- The text is readable and understandable.
- Specific objectives are stated at the beginning of each chapter.
- Key words are in bold print.
- Statistical procedures are provided in steps, or cookbook format, and examples related to physical performance are provided.
- Reminders of chapter objectives are placed in the text in the form of "Are you able to do the following?" questions.
- A Point of Emphasis box is used to reinforce particular concepts or practices.
- Review problems to reinforce the chapter objectives are provided at the conclusion of most chapters.
- Review questions are provided at the conclusion of all chapters.

New to This Edition

All chapters in the previous edition have been retained. Included in this edition are the following:

- Description of teacher self-assessment
- Revision and updating of some tables
- Updating of health-related physical fitness tests

Users of edition 7 also will find that minor changes have been made in several chapters.

Instructor Resources

Online ancillaries that accompany this text include PowerPoint slides and test bank questions. These resources are available through Connect.

Acknowledgments

I am grateful to the many students and colleagues who contributed to the development of the eight editions of this text. The production of this text is possible because of your support. To the publishers who permitted use of

their materials and to the McGraw-Hill Higher Education staff members who contributed to eight edition, I say thanks. To Kayci Wyatt, Stacey Lutkoski, Jacob Sullivan, Francesca King, and Nikhil Rajender Kumar Meena I express appreciation for your patience, skill, and guidance.

Eighth Edition Reviewers

Kay Daigle
Southeastern Oklahoma State University

Glenn Steimling
William Penn University

Shelia Lucyle Jackson
Arkansas Tech University

Kristin Heumann
Colorado Mesa University

Brian Brabham
University of Mary Hardin-Baylor

Julie M Knutson
Minnesota State University-Moorhead

Melissa A. Parks
Louisiana State University of Alexandria

Lauri Hauck
California Baptist University

Students—study more efficiently, retain more and achieve better outcomes. Instructors—focus on what you love—teaching.

SUCCESSFUL SEMESTERS INCLUDE CONNECT

FOR INSTRUCTORS

You're in the driver's seat.

Want to build your own course? No problem. Prefer to use our turnkey, prebuilt course? Easy. Want to make changes throughout the semester? Sure. And you'll save time with Connect's auto-grading too.

65%

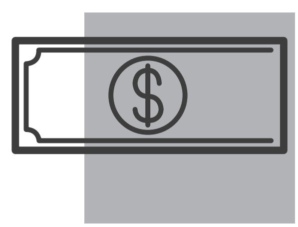
Less Time Grading

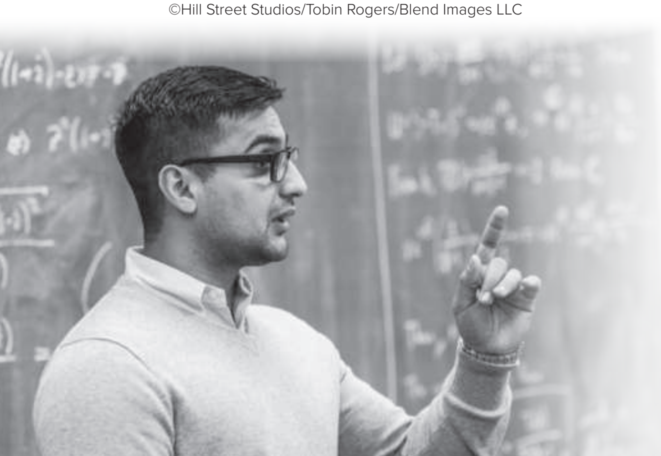
They'll thank you for it.

Adaptive study resources like SmartBook® help your students be better prepared in less time. You can transform your class time from dull definitions to dynamic debates. Hear from your peers about the benefits of Connect at **www.mheducation.com/highered/connect**

Make it simple, make it affordable.

Connect makes it easy with seamless integration using any of the major Learning Management Systems—Blackboard®, Canvas, and D2L, among others—to let you organize your course in one convenient location. Give your students access to digital materials at a discount with our inclusive access program. Ask your McGraw-Hill representative for more information.





Solutions for your challenges.

A product isn't a solution. Real solutions are affordable, reliable, and come with training and ongoing support when you need it and how you want it. Our Customer Experience Group can also help you troubleshoot tech problems—although Connect's 99% uptime means you might not need to call them. See for yourself at **status.mheducation.com**

For Students

©Shutterstock/wavebreakmedia

Effective, efficient studying.

Connect helps you be more productive with your study time and get better grades using tools like SmartBook, which highlights key concepts and creates a personalized study plan. Connect sets you up for success, so you walk into class with confidence and walk out with better grades.

"I really liked this app—it made it easy to study when you don't have your text-book in front of you."

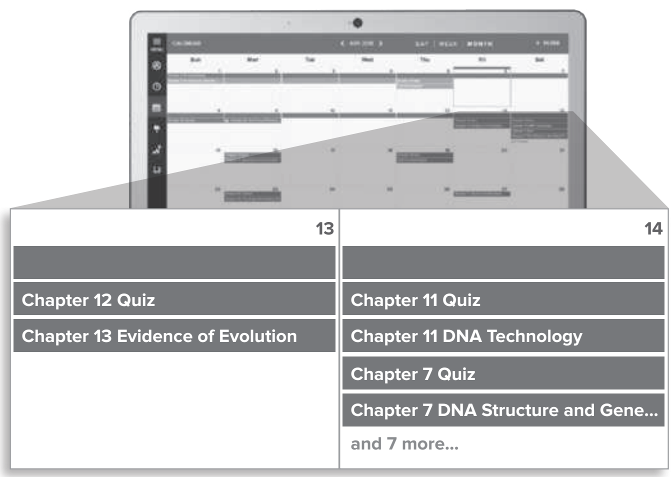
- Jordan Cunningham,
Eastern Washington University

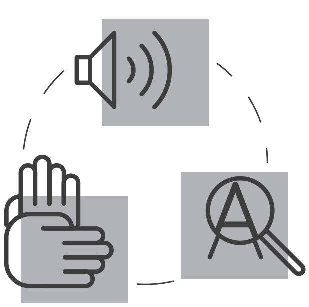
Study anytime, anywhere.

Download the free ReadAnywhere app and access your online eBook when it's convenient, even if you're offline. And since the app automatically syncs with your eBook in Connect, all of your notes are available every time you open it. Find out more at **www.mheducation.com/readanywhere**

No surprises.

The Connect Calendar and Reports tools keep you on track with the work you need to get done and your assignment scores. Life gets busy; Connect tools help you keep learning through it all.





Learning for everyone.

McGraw-Hill works directly with Accessibility Services Departments and faculty to meet the learning needs of all students. Please contact your Accessibility Services office and ask them to email accessibility@mheducation.com, or visit **www.mheducation.com/about/accessibility.html** for more information.

This page intentionally left blank

1 Measurement, Evaluation, Assessment, and Statistics

After completing this chapter, you should be able to

1. Define the term *statistics*.
2. Define the terms *test*, *measurement*, *evaluation*, and *assessment*, and give examples of each.
3. List and describe the reasons for measurement, evaluation, and assessment by the physical educator.
4. State why the ability to use statistics is important for the physical educator.

"Why statistics? I don't need statistics to be a good teacher." "I don't need statistics. I plan to work in a health fitness center."

Perhaps you have made comments similar to these or have heard some of your classmates make them. If you do not plan to perform your responsibilities as they should be performed, and you do not plan to continue your professional growth, you are correct in believing that you do not need statistics. However, if you want to perform your professional responsibilities in exemplary fashion, the study of statistics should be included in your preparation.

Statistics involves the collection, organization, and analysis of numerical data. Statistical methods require the use of symbols, terminology, and techniques that may be new to you, but you should not fear these methods. The idea that statistics is a form of higher mathematics is incorrect. To successfully perform the statistics presented in this book, you need only a basic knowledge of arithmetic and some simple algebra. The most complex formula in statistics can be reduced to a series of logical steps involving adding, subtracting, multiplying, and dividing. If you are willing to study the statistical concepts and perform the provided exercises, you will master the statistics presented to you.

Before finding an answer to "Why statistics?" you should understand the meaning of measurement and evaluation and the reasons for measurement by the physical educator. Measurement is not a new concept to you. You measured your height and weight throughout your growing years. You have read how fast athletes have run, how high some have jumped, and how far a baseball or a golf ball has been hit. All of these are examples of measurement. When you assume a position as a physical educator in the private, public, teaching, or nonteaching sector, you will perform measurement tasks. On many occasions, this measurement will be administered in the form of a test, resulting in a score. For our purposes, a **test** is an instrument or a tool used to make a particular measurement. The tool may be written, oral, mechanical, or another variation. Examples of such tests are cardiorespiratory fitness tests, flexibility tests, and strength tests. On other occasions, measurement may not involve a performance by a person but will consist of the measurement of a particular attribute. Heart rate, blood pressure, and body fat measurements are such examples. You should recognize that in all of the preceding examples, numbers are obtained. So we can say that **measurement** is usually thought of as quantitative; it is the process of assigning a number to a performance or an attribute of a person. Sometimes when you measure, the score is a term or a phrase, but usually measurement will involve the use of numbers. Of course, objects are measured too, but as a physical educator you will be concerned primarily with people.

Once you have completed the measurement of a particular attribute of an individual, you must give meaning to it. For instance, if you administer a cardiorespiratory fitness test to participants of an adult fitness group, they will immediately want to know the status of their cardiorespiratory fitness. Without an interpretation of the quality of the test scores, the test has no meaning to the group. If you perform skinfold measurements in a physical education class, on athletes, or on members of a health club, the individuals will want to know what the sum of the measurements means in relation to body fat; otherwise, the measurements will have no meaning. The same can be said for written tests. There must be an interpretation of the test scores if they are to have meaning. This interpretation of measurement is **evaluation:** that is, a judgment about the measurement. For measurement to be effective, it must be followed by evaluation.

It is at this point that some physical educators stop. They measure an attribute, interpret the results to individuals, and go no further. They fail to use the results of their measurement and evaluation to identify performance and behavior problems and to prescribe how the problems can be corrected. This process—measure, evaluate, identify, and prescribe—is referred to as **assessment.** Let's again use the example of skinfold measurements. Assume that several individuals in the group that you measure are diagnosed as overfat as a result of your measurements. You should attempt to determine the eating and activity habits of the individuals and prescribe the proper diet and exercise program. This process is followed by exercise prescription specialists and athletic trainers. Individuals do not all begin an exercise program at the same level. The prescription level for an individual is determined through measurement, evaluation, and identification of the strengths and weaknesses or needs of the individual. Likewise, athletic trainers follow the same process with injured athletes. The extent of the athlete's injury is determined through measurement and evaluation and then a rehabilitation program is prescribed. Additionally, the athlete usually is not permitted to return to competition until certain tasks or tests can be performed successfully. In the school environment, the concept of authentic assessment is emphasized. This concept will be discussed in a later chapter.

ARE YOU ABLE TO DO THE FOLLOWING?

- Define the term *statistics*.
- Define the terms *test*, *measurement*, *evaluation*, and *assessment*, and give examples of measurement, evaluation, and assessment.

Reasons for Measurement, Evaluation, and Assessment by the Physical Educator

Now that you know what is meant by the terms *measurement, evaluation,* and *assessment,* let's look at ways you will use them in your profession.

Motivation

If used correctly, measurement can highly motivate most individuals. In anticipation of a test, students usually study the material or practice the physical tasks that are to be measured. This study or practice should improve performance. Skinfold measures might encourage overfat individuals in health fitness programs to lose body fat. Older individuals may be motivated to improve their flexibility through the administration of flexibility tests. A sports skills test administered to inform individuals of their ability in the sport might motivate them to improve their skills. This motivation is more likely to occur, however, if you as the instructor provide positive feedback. Always try to keep your evaluation and assessment positive rather than negative.

Finally, most everyone enjoys comparing past performances with current ones. Knowing that a second measurement will take place, students and adults often work to improve on the original score.

Diagnosis

Through measurement you can assess the weaknesses (needs) and strengths of a group or individuals. Measurement before the teaching of a sports skill, physical fitness session, or other events you teach as a physical educator may cause you to alter your initial approach to what you are teaching. For example, you may discover that, before you do anything else in a softball class, you need to teach the students how to throw properly. You also may find that some individuals need more or less attention than others in the group. Identifying those students who can throw with accuracy and good form will enable you to devote more time to the students who cannot perform the skill. If you serve as an adult fitness leader, the identification of individuals with a higher level of fitness than the rest of the group will enable you to begin their program at a different level.

In certain settings, you may be able to prescribe personal exercises or programs to correct the diagnosed weaknesses. *Exercise prescription* is a popular term in fitness programs, but appropriate activities may be prescribed in other programs as well. Diagnostic measurement is valuable also after a group has participated in a class for several weeks. If some individuals are not progressing as you feel they should, testing may help you determine why they are not.

Classification

There may be occasions when you would like to classify individuals into similar groups for ease of instruction. In addition, people usually feel more comfortable when performing with others of similar skill. Sometimes, even in so-called noncontact sports, homogeneous grouping should be done for safety reasons. Also, homogeneous grouping is occasionally necessary in aerobic and fitness classes so that individuals with a low level of fitness will not attempt to perform at the same intensity as individuals with a high level of fitness.

Achievement

The most common reason for measurement and assessment is to determine the degree of achievement of program objectives and personal goals. Most people like to know how far they have progressed in a given period of time. Participants in diet modification and exercise programs like to know their changes in body fat percentage and muscle strength. Students like to know how far they have progressed in sports skills development in a given period of time. You too will need to know the achievement of participants to better evaluate the effectiveness of your instruction. If participants are failing to achieve their stated goals, you may need to revise their program or your method of instruction.

Achievement often is used to determine grades in physical education. If administered properly, performance tests and knowledge tests are appropriate for grading, and they decrease the need for subjective grading of the students. Many physical education teachers, however, mistakenly use tests only for determining grades. The assigning of grades will be discussed at length in chapter 7.

Evaluation of Instruction and Programs

With any responsibility you assume as a physical educator, occasionally you will have to justify the effectiveness of your instruction or program to your employer. For instance, when budget cuts are anticipated in the public schools, physical education and the arts are often the first programs considered for elimination. It is also necessary to justify a program when budget increases are requested. Furthermore, school accreditation studies require assessment of instruction and programs. If measurement and evaluation identify instructional or program problems, correctional procedures are stated. Standardized forms are available for program evaluation, but if program content is professionally sound, the success and effectiveness of instruction and programs are best determined by how well the participants fulfill program objectives. This statement is true for school programs, fitness and wellness programs, and all other professional programs in which you may have responsibilities. You must be able to measure and assess instruction and programs.

Assessment of each student's skill at the beginning of an activity unit helps you determine the effectiveness of previous instruction and programs and at what point you should begin your instruction. If the students do not know basic rules and cannot demonstrate the elementary playing skills of an activity, it will be necessary to begin instruction at that level. In addition, there may be times when you want to compare different methods of teaching sports skills or fitness. If you can be confident that the different groups are of equal initial ability, it is possible to compare the results of test scores at the conclusion of instruction and determine if one method of teaching is better than another. This procedure will be discussed in greater detail in chapter 4.

Prediction

Measurement to predict future performance in sports has increased in popularity, but this type of testing usually requires expertise in exercise physiology and psychology. Maximum oxygen uptake, muscle biopsies, and anxiety level are examples of tests that are used to predict future performance in sports.

Research

Research is used to find meaningful solutions to problems and as a means to expand a body of knowledge. It is of value for program evaluation, instructor evaluation, and improvement in performance, as well as other areas related to physical education. Many opportunities exist for physical educators who wish to perform research.

Now that you are aware of the primary reasons for measurement, evaluation, and assessment in physical education, you are ready to answer the question "Why statistics?"

ARE YOU ABLE TO DO THE FOLLOWING?

- List and describe the reasons for measurement, evaluation, and assessment by the physical educator.

Why Statistics?

Whether you teach, instruct in a fitness center, administrate, or have responsibilities in a corporate setting, the ability to use statistics will be of value to you. Although no attempt will be made in this book to provide an extensive coverage of statistics, after you have completed chapters 2, 3, and 4, you should have the skill to do the following.

Analyze and Interpret Data

The data gathered for any of the measurement reasons described should be statistically analyzed and interpreted. It is a mistake to gather data and make important decisions about individuals without this analysis. Decisions regarding improvement in group performance and differences in teaching methodology should not be made without statistical analysis. Also, if you are willing to statistically analyze and interpret test scores, you can better inform all participants of the test results than you can with a routine analysis of the scores. So, using statistical analysis and interpretation, you can provide a more meaningful evaluation of your measurement.

Interpret Research

As a physical educator you should read research published in professional journals. After completion of this book you will not understand all statistical concepts, but you will understand enough to accurately interpret the results and conclusions of many studies. This ability will enable you to put into practice the conclusions of research. Too many physical educators fail to use research findings because they do not understand them. If you are to continue your professional growth, it is essential that you be able to interpret research related to your professional responsibilities.

Standardized Test Scores

Many measurements performed by the physical educator will be in different units—for example, feet, seconds, and numbers. To compare such measurements, it is best to convert the scores to standardized scores. A popular form of standardized scores is percentile scores (such as reported SAT scores).

Determine the Worth (Validity and Reliability) of a Test

Validity refers to the degree to which a test measures what it claims to measure. *Reliability* refers to the consistency of a test (i.e., the test obtains approximately the same results each time it is administered). These topics may not mean much to you now, but by knowing how to interpret statements about these characteristics, you are more likely to select the appropriate tests to administer to your students, clients, or customers. In addition, you will be able to estimate the validity and reliability of tests that you construct.

ARE YOU ABLE TO DO THE FOLLOWING?

- State why the ability to use statistics is important to the physical educator.

Chapter Review Questions

1. What does the term *statistics* mean?
2. What do the terms *test, measurement, evaluation,* and *assessment* mean? Can you provide an example of a professional role where you might administer a test and perform all assessment responsibilities?
3. Imagine that you are now a professional in your chosen field. Can you provide examples of how you would use measurement, evaluation, and assessment in your field?
4. How might statistics be used in your chosen profession?

2 Describing and Presenting a Distribution of Scores

After completing this chapter, you should be able to

1. Define all statistical terms that are presented.
2. Describe the four scales of measurement, and give examples of each.
3. Describe a normal distribution and four curves for distributions that are not normal.
4. Define the terms measures of *central tendency* and *measures of variability*.
5. Define the three measures of central tendency, identify the symbols used to represent them, describe their characteristics, calculate them with ungrouped data, and state how they can be used to interpret data.
6. Define the three measures of variability, identify the symbols used to represent them, describe their characteristics, calculate them with ungrouped data, and state how they can be used to interpret data.
7. Describe the relationship of the standard deviation and the normal curve.
8. Define *percentile* and *percentile rank,* identify the symbols used to represent them, calculate them with ungrouped data, and state how they can be used to interpret data.
9. Define standard scores, calculate z-scores and T-scores, and interpret their meanings.

Regardless of your employment site, as a physical educator you often will test individuals. You may test for health fitness, sport fitness or skills, subject knowledge, or other areas. After the administration of a test, you will be expected to analyze the test scores and present your analysis to the test takers. The analysis of a set of test scores is referred to as descriptive statistics.

Statistical Terms

Before you begin to develop the skills to use descriptive statistics, become familiar with the following terms. Understanding these terms will be valuable to you in chapters 3 and 4 also.

Data The result of measurement is called data. The term *data* usually means the numerical result of measurement but can also mean verbal information.

Variable A variable is a trait or characteristic of something that can assume more than one value. Examples of variables are cardiovascular endurance, percentage of body fat, flexibility, and muscular strength. Their values will vary from one person to another, and they may not always be the same for one individual.

Population A population includes all subjects (members) within a defined group. All subjects of the group have some measurable or observable characteristic. For example, if you wanted to determine the physical fitness of twelfth-grade students in a particular high school, the population would be all students in the twelfth grade.

Sample A sample is a part or subgroup of the population from which the measurements are actually obtained. Rather than collect physical fitness data on all students in the twelfth grade, you could choose a smaller number to represent the population.

Random sample A random sample is one in which every subject in the population has an equal chance of being included in the sample. A sample could be formed by randomly selecting a group to represent all twelfth graders. The selection could be done by placing the names of all twelfth-grade students in a container and randomly drawing the names out of it or by using a table of random numbers.

Parameter A parameter is a value, a measurable characteristic, that refers to a population. The population mean (the average) is a parameter.

Statistic A statistic is a value, a measurable characteristic, that refers to a sample. The sample mean is a statistic. Statistics are used to estimate the parameters of a defined population. (*Note:* When used in this manner the word *statistics* is plural. If used to denote a subject or a body of knowledge, the word *statistics* is singular.)

Descriptive statistics When every member of a group is measured and no attempt is made to generalize to a larger group, the methods used to describe the group are called descriptive statistics. Conclusions are reached only about the group being studied.

Inferential statistics When a random sample is measured and projections or generalizations are made about a larger group, inferential statistics are used. The correct use of inferential statistics permits you to use the data generated from a sample to make inferences about the entire population. Suppose you were in charge of a physical fitness program at a wellness center and wanted to estimate the physical fitness of all three hundred female adults in the program. You could randomly select thirty of the females, test them, and through the use of inferential statistics, estimate the physical fitness of the three hundred females.

Discrete data Discrete data are measures that can have only separate values. The values are limited to certain numbers, usually whole numbers, and cannot be reported as fractions. Examples of discrete data are the sex of the individual, the number of team members, the number of shots made, and the number of hits in a softball game.

Continuous data Continuous data are measures that can have any value within a certain range. The values can be reported as fractions. Running and swimming events (time) and throwing events (distance) are examples of continuous data.

Ungrouped data Ungrouped data are measures not arranged in any meaningful manner. The raw scores as recorded are used for calculations.

Grouped data Grouped data are measures arranged in some meaningful manner to facilitate calculations.

ARE YOU ABLE TO DO THE FOLLOWING?

- Define statistical terms.
- Provide an example of a study performed with a random sample and generalizations made about a population.
- Provide an example in which you might use descriptive statistics in professional responsibilities.

Scales of Measurement

Variables may be grouped into four categories of scales depending on the amount of information given by the data. Different rules apply at each scale of measurement, and each scale dictates certain types of statistical procedures. As measurement moves from the lowest to the highest scale, the result of measurement is closer to a pure measure of a count of quantity or amount. The lower the level of the data, the less information the data provide.

Nominal Scale

The **nominal scale,** also called categorical, is the lowest and most elementary scale. A naming level only, this scale is used to identify and report the frequency of objects or persons. Names are given to the variables, and categories are exclusive of each other; no comparisons of the categories can be made, and each category is assumed to be as valuable as the others. Some nominal scales have only two categories, but others may have more. Positions on sports teams, level of education, and state of residence are examples of the nominal scale. Numbers may be used to represent the variables, but the numbers do not have numerical value or relationship. For example, 0 may be used to represent the male classification and 1 the female classification.

Ordinal Scale

The **ordinal scale** provides some information about the order or rank of the variables, but it does not indicate how much better one score is than another. No determination can be made of the relative differences from rank to rank. For example, the order of finish in a 10-kilometer race provides information about who is fastest, but it does not indicate how much faster the person who finished in first position is than the person who came in second, or any of the other runners. Examples of the ordinal scale are the ranking of tennis team members from best to worst, class rank in a high school graduating class, team standings in an athletic conference, body frames (large, medium, small), and body fat (lean, normal, overfat).

Interval Scale

An **interval scale** provides information about the order of the variables, using equal units of measurement. The distance between divisions of the scale is always the same, so it is possible to say how much better one number is than another. However, the interval scale has no true zero point. A good example of an interval scale is temperature. It is possible to say that 90°F is 10° warmer than 80°F and that 55°F is 10° warmer than 45°F, but it cannot be said that 90° is twice as hot as 45°. Since 0°F does not mean the complete absence of heat, there is no true zero point. Many measurements in physical education are in the interval scale. Surveys regarding sportsmanship and attitude toward physical activity are examples of the interval scale. In surveys of these types, a zero score does not mean that a person has absolutely no sportsmanship or no attitude toward activity.

Ratio Scale

A **ratio scale** possesses all the characteristics of the interval scale and has a true zero point. Examples of ratio scales are height, weight, time, and distance. Ten feet is twice as long as 5 feet; 9 minutes is three times as long as 3 minutes; and 20 pounds is four times as heavy as 5 pounds.

In most statistical studies or analyses, interval and ratio measurements are analyzed in the same way. Table 2.1 summarizes the major differences among the four scales of measurement.

ARE YOU ABLE TO DO THE FOLLOWING?

- Define the four scales of measurement, and give examples of each.

TABLE 2.1 Scales of Measurement

Scales	Characteristics	Examples
Nominal	Numbers represent categories. Numbers do not distinguish groups and do not reflect differences in magnitude.	Divisions by gender, race, eye color, or political party
Ordinal	Numbers indicate rank order of measurements, but they do not indicate the magnitude of the interval between the measures.	Order of finish in races, grades for achievement
Interval	Numbers represent equal units between measurements. It is possible to say how much better one measure is than another, but there is no true zero point.	Temperature, year, IQ
Ratio	Numbers represent equal units between measurements, and there is an absolute zero point.	Height, weight, distance, time, heart rate, blood pressure, blood cholesterol

◆ ◆ ◆ POINTS OF EMPHASIS ◆ ◆ ◆

- Measurement is the assignment of numbers to a performance or to an attribute of a person.
- A nominal scale assigns a name or label to different objects or persons; it provides identification.
- Numbers assigned with an ordinal scale identify and rank-order each object or person.
- Interval and ratio scales have all the characteristics of the ordinal scale in addition to providing equal intervals between scores.
- A ratio scale has a true zero point.
- Ordinal, interval, and ratio scales provide measurement data.
- The ratio scale is the only scale in which ratios of measurement can be made. For example, we can correctly state that someone ran twice as far or twice as fast as another person. It also would be correct to say that someone has twice the blood cholesterol level of another person.

Normal Distribution

Most of the statistical methods used in descriptive and inferential statistics are based on the assumptions that a distribution of scores is normal and that the distribution can be graphically represented by the normal (bell-shaped) curve, as shown in figure 2.1. For example, the distribution of the college entrance test scores of all test takers nationally would be normal. The test scores for a particular high school or group of students, however, may not have a normal distribution. Individuals in the school or group may score exceptionally high or low, causing the curve to be skewed. This concept will be discussed later in the chapter. As they are on all graphic representations of frequency distributions, the score values are placed on the horizontal axis, and the frequency of each score is plotted with reference to the vertical axis. The two ends of the curve are symmetrical and represent the scores at the extremes of the distribution.

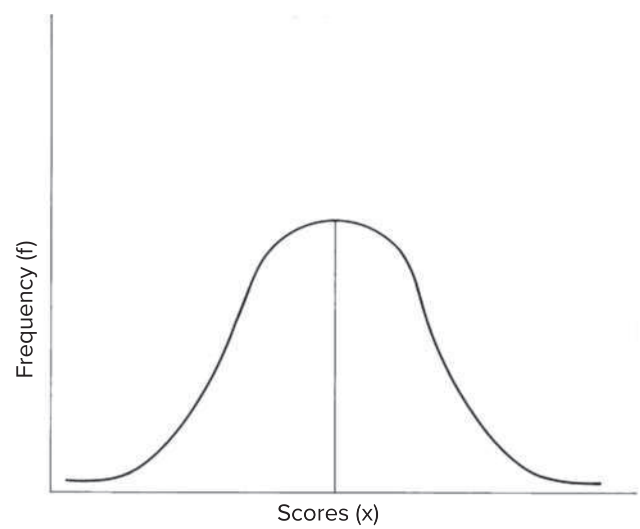


FIGURE 2.1 Normal curve.

The normal distribution is theoretical and is based on the assumption that the distribution contains an infinite number of scores. You will not measure groups of infinite size, so you should not be surprised when distributions do not conform to the normal curve. If the distribution is based on a large number of scores, however, it will be close to normal distribution. You can have a large number of scores if you administer the same test to several groups or if you combine your test scores from several years of

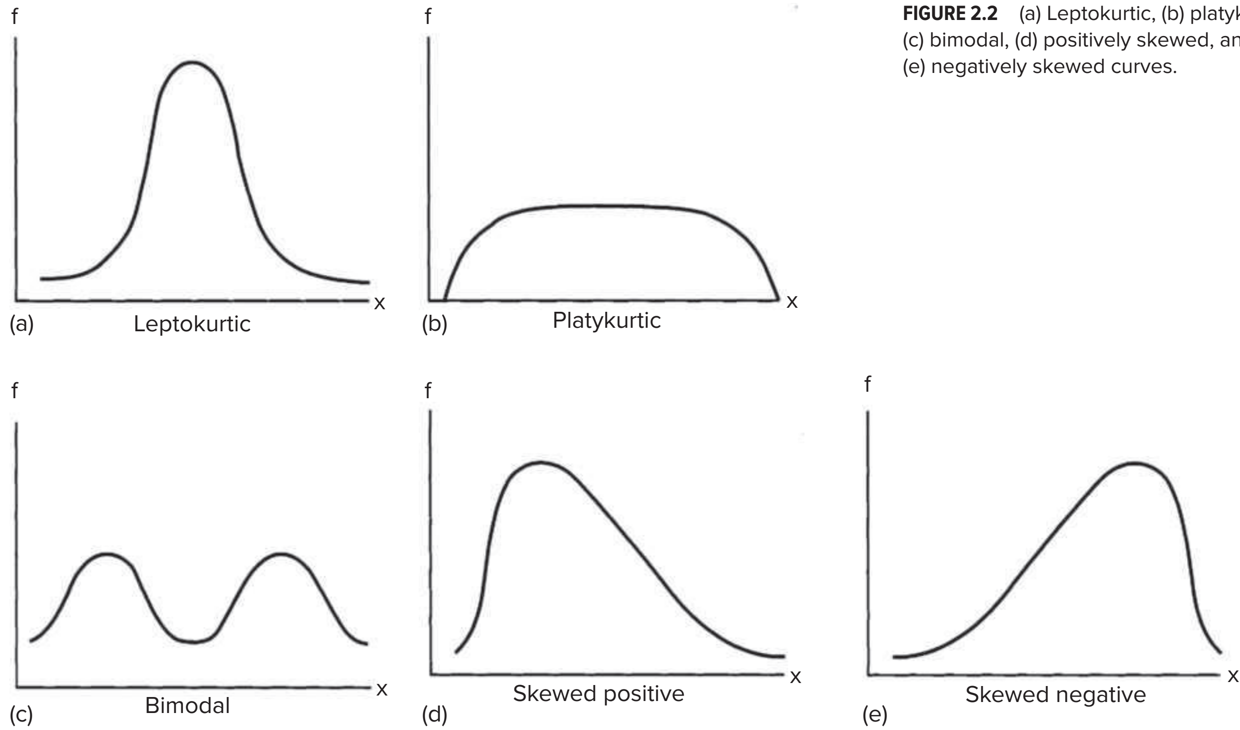


FIGURE 2.2 (a) Leptokurtic, (b) platykurtic, (c) bimodal, (d) positively skewed, and (e) negatively skewed curves.

testing into one distribution. A normal distribution has the following characteristics:

1. A bell-shaped curve.
2. Symmetrical distribution about the vertical axis of the curve; whatever happens on one side of the curve is mirrored on the other.
3. Greatest number of scores found in the middle of the curve, with fewer and fewer found toward the ends of the curve.
4. All measures of central tendency (mean, median, mode) at the vertical axis.

Note: Occasionally, a distribution of scores will produce a curve that is different from the normal curve (see figure 2.2). These curves are very possible in work with small groups. If a group consists of individuals who are very similar in ability (referred to as a homogeneous group), the curve is pointed and is called **leptokurtic.** If the group consists of individuals with a wide range of ability (referred to as a heterogeneous group), the curve is flat and is called **platykurtic.** If a group of scores has two modes, the curve has two high points and is called **bimodal.** Some distributions of scores may have more than two modes, producing a curve with more than two high points. A curve in which the scores are clustered at one end is **skewed.** Skewed curves will be described in greater detail later in this chapter.

ARE YOU ABLE TO DO THE FOLLOWING?

- Describe a normal distribution and four curves for distributions that are not normal.

Analysis of Ungrouped Data

Imagine that you have given a volleyball knowledge test to a group of 30 seventh-grade students, and you want to have a better understanding of the test results as well as interpret the scores to the students. With the aid of an inexpensive calculator, you can fulfill both of those objectives. This portion of the chapter shows you how to use descriptive statistics to analyze and interpret the volleyball knowledge test scores as well as other ungrouped data. A test with high scores has intentionally been selected as an example to demonstrate that the procedures are not difficult. Tables 2.2, 2.3, and 2.4 report the results of the volleyball knowledge test analysis.

TABLE 2.2 Rank of Volleyball Knowledge Test Scores

Rank		Score
1		96
2		95
3		93
4	4.5	92
5		92
6		91
7	6.5	91
8		90
9	9	90
10		90
11		89
12	12	89
13		89
14		88
15		88
16	16	88
17		88
18		88
19		87
20	20	87
21		87
22		86
23	22.5	86
24		85
25	24.5	85
26		84
27	26.5	84
28		83
29		82
30		81

◆ ◆ ◆ POINTS OF EMPHASIS ◆ ◆ ◆

- Descriptive statistics are measures used to determine the current status of scores and/or to describe and summarize a set of data. No treatments are administered, and no generalizations are made about other groups or a larger population.
- A classic example for the use of descriptive statistics is in the school environment. However, descriptive statistics are used in most if not all professions. They can provide valuable information regarding the performance or status of a group.

Score Rank

Although you can perform statistical analysis without putting the scores in rank order, you may first want to carry out this procedure, which provides you with information about each person's rank in the score distribution. Be careful how you use the rank of scores. Sometimes, all the scores will be above what you consider a satisfactory score. Because someone has to be at the bottom when scores are ranked, you may not want to share this information with the group. Do not create alarm or embarrassment when there is no need. Table 2.2 shows the ranking of the thirty volleyball knowledge test scores. The steps for ranking the scores are as follows:

1. List the scores in descending order.
2. Number the scores. The highest score is number 1, and the last score is the number of the total number of scores. (Table 2.2 has thirty scores, so the last score is number 30.)
3. Because identical scores should have the same rank, average the rank, or determine the midpoint, and assign them the same rank.

Measures of Central Tendency

Measures of central tendency are descriptive statistics that describe the middle characteristics of a distribution of scores. The most widely used statistics, they represent

TABLE 2.3 Measures of Central Tendency and Variability and Percentiles (Deciles) Computed from Ungrouped Volleyball Knowledge Test Scores (N = 30)

Score	X^2	cf	Percentile
96	9,216	30	
95	9,025	29	
93	8,649	28	
92	8,464	27	90
92	8,464	26	
91	8,281	25	
91	8,281	24	80
90	8,100	23	
90	8,100	22	
90	8,100	21	70
89	7,921	20	
89	7,921	19	
89	7,921	18	60
88	7,744	17	
88	7,744	16	
88	7,744	15	50
88	7,744	14	
88	7,744	13	
87	7,569	12	40
87	7,569	11	
87	7,569	10	
86	7,396	9	30
86	7,396	8	
85	7,225	7	
85	7,225	6	20
84	7,056	5	
84	7,056	4	
83	6,889	3	10
82	6,724	2	
81	6,561	1	
$\sum X = 2{,}644$	$\sum X^2 = 233{,}398$		

$R = 96 - 81 = 15$

$\bar{X} = \frac{\Sigma X}{N} = \frac{2{,}644}{30} = 88.1$

$Q = \frac{Q_3 - Q_1}{2} = \frac{90 - 85.5}{2} = 2.25$

$P_{50} = 88$

$s = \sqrt{\frac{N\Sigma X^2 - (\Sigma X)^2}{N(N-1)}} = \sqrt{\frac{30(233{,}398) - (2{,}644)^2}{30(30-1)}}$

$Mo = 88$

$s = 3.6$

TABLE 2.4 Standard Deviation Computed from Volleyball Knowledge Test Scores, Deviations, and Squared Deviations (N = 30)

Score	d	d^2	Score	d	d^2
96	7.9	62.41	88	−0.1	0.01
95	6.9	47.61	87	−1.1	1.21
93	4.9	24.01	87	−1.1	1.21
92	3.9	15.21	87	−1.1	1.21
92	3.9	15.21	86	−2.1	4.41
91	2.9	8.41	86	−2.1	4.41
91	2.9	8.41	85	−3.1	9.61
90	1.9	3.61	85	−3.1	9.61
90	1.9	3.61	84	−4.1	16.81
90	1.9	3.61	84	−4.1	16.81
89	0.9	0.81	83	−5.1	26.01
89	0.9	0.81	82	−6.1	37.21
89	0.9	0.81	81	−7.1	50.41
88	−0.1	0.01	$\sum X = 2{,}644$		$\sum d^2 = 373.50$
88	−0.1	0.01			
88	−0.1	0.01			
88	−0.1	0.01			
		$\sum d^2 = 373.5$			

$\bar{X} = 88.1$

$$s = \sqrt{\frac{\Sigma d^2}{N-1}} = \sqrt{\frac{373.5}{29}} = 3.6$$

scores in a distribution around which other scores seem to center. Measures of central tendency are the mean, median, and mode.

The Mean

With any test, the first question usually asked by the students upon knowing their individual score is "What is the class average?" The **mean,** the most generally used measure of central tendency, is the arithmetic average of a distribution of scores. It is calculated by summing all the scores and dividing by the total number of scores. The following are some important characteristics of the mean:

1. It is the most sensitive of all the measures of central tendency. It will always reflect any change within a distribution of scores.
2. It is the most appropriate measure of central tendency to use for ratio data and may be used on interval data.
3. It considers all information about the data and is used to perform other important statistical calculations.

4. It is influenced by extreme scores, especially if the distribution is small. For example, when one or more scores are high or low in relation to the other scores, the mean is pulled in that direction. (See number 1 in characteristics of Median.) This characteristic is the chief disadvantage of the mean.

The symbols used to calculate the mean and other statistics are as follows:

$\bar{X}$ = the mean (called X-bar)

Σ (Greek letter sigma) = "the sum of"

X = individual score

N = the total number of scores in a distribution

The formula for calculating the mean is

$$\bar{X} = \frac{\Sigma X}{N}$$

Simply, to calculate the mean, you add all the scores in a distribution and divide by the number of scores you have. The calculation of $\bar{X}$ from the distribution in table 2.3 is

$$\bar{X} = \frac{2{,}644}{30} = 88.1$$

You probably would report the scores to the students in whole numbers, so you should round the mean to 88.

ARE YOU ABLE TO DO THE FOLLOWING?

- Define the term *measures of central tendency*.
- Identify the symbol for the mean.
- Define the mean.
- Describe the characteristics of the mean.
- Calculate the mean with ungrouped data and use it to interpret the data.

The Median

The **median** is the score that represents the exact middle in the distribution. It is the fiftieth percentile, the score that 50% of the scores are above and 50% of the scores are below. The following are some important characteristics of the median:

1. It is not affected by extreme scores; it is a more representative measure of central tendency than the mean when extreme scores are in the distribution. As an example, consider the height of five individuals. If the heights are 71″, 72″, 73″, 74″, and 75″, the mean (average) height of the five people is 73″, and the median (middle score) height is 73″. Now imagine that we exchange the individual who is 75″ in height for an individual who is 85″ in height. The mean is now 75″, but the median remains at 73″.
2. It is a measure of position; it is determined by the number of scores and their rank order. It is appropriately used on ordinal or interval data.
3. It is not used for additional statistical calculations.

The median may be represented by Mdn or P_{50}. The steps for calculation of P_{50} are as follows:

1. Arrange the scores in ascending or descending order.
2. Multiply N by .50 to find 50% of the distribution.
3. If the number of scores is odd, P_{50} is the middle score of the distribution.
4. If the number of scores is even, P_{50} is the arithmetic average of the two middle scores of the distribution.

The calculation of P_{50} from the distribution in table 2.3 is as follows:

1. .50(30) = 15
2. The fifteenth and sixteenth scores (middle scores of distribution) are 88.
3. $P_{50} = 88$

Because you will be using this same procedure to calculate any percentile, you may want to place a cumulative frequency (cf) column with the ordered listing of scores. The cumulative frequency is an accumulation of frequencies beginning with the bottom score. In the cf column, the highest score will have the same number as N.

ARE YOU ABLE TO DO THE FOLLOWING?

- Identify the symbol for the median.
- Define the median.

- Describe the characteristics of the median.
- Calculate the median with ungrouped data and use it to interpret the data.

The Mode

The **mode** is the score that occurs most frequently. In a normal distribution, the mode is representative of the middle scores. In some distributions, however, the mode may be an extreme score. If a distribution has two modes, it is bimodal, and it is possible for a distribution to be multimodal or to have no mode at all. Because no symbol is used to represent the mode, Mo is sometimes used. Some characteristics of the mode are as follows:

1. It is the least used measure of central tendency. It indicates the score attained by the largest number of subjects. It also is used to indicate such things as the most popular item or product used by a group.
2. It is not used for additional statistical calculations.
3. It is not affected by extreme scores, the total number of scores, or their distance from the center of the distribution.

The mode of the distribution in table 2.3 is 88.

ARE YOU ABLE TO DO THE FOLLOWING?

- Define the mode.
- Describe the characteristics of the mode.
- Calculate the mode with ungrouped data and use it to interpret the data.

Which Measure of Central Tendency Is Best for Interpretation of Test Results?

You have studied the definitions of the three measures of central tendency, calculation procedures, and some characteristics of each. Do you know which of the three is best for interpreting test results to any group that you might be testing? In making your decision, you should consider the following:

1. The mean, median, and mode are the same for a normal distribution (symmetrical curve), but you often will not have a normal curve.
2. The farther away from the mean and median the mode is, the less normal the distribution (i.e., the curve is skewed). Figure 2.3 shows the relationship of these measures to a symmetrical curve, a positively skewed curve, and a negatively skewed curve. In a positively skewed curve the scores are clustered at the lower end of the scale; the longer tail of the curve is to the right, and the mean is higher than the median. In a negatively skewed curve the scores are clustered at the upper end of the scale; the longer tail of the curve is to the left, and the mean is lower than the median. An extremely difficult test, on which most of the scores are low but a few high scores increase the mean, often results in a positively skewed curve. An easy test, on which most of the scores are high but a few low scores decrease the mean, usually results in a negatively skewed curve.
3. The mean and median are both useful measures. If the curve is badly skewed with extreme scores, you may want to use only the median. You must decide how important the extreme scores are. If the curve is approximately normal, use the mean and median.

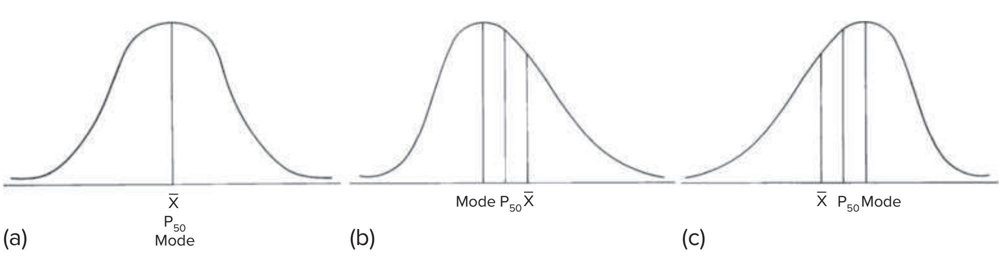


FIGURE 2.3 (a) Normal curve, (b) positively skewed curve, and (c) negatively skewed curve.

◆ ◆ ◆ POINTS OF EMPHASIS ◆ ◆ ◆

- Although the mean is considered the most reliable measure of central tendency and is also used for other statistical calculations, the following is an example of an important use of the median. Suppose that upon your graduation you apply for a job with a small company. The company employs 30 people. You are told that the average salary is $50,000. You ask, "What is the median salary?" You are told that the median salary is $30,000. What does this mean to you?

4. In most testing, the mean is the most reliable and useful measure of central tendency. It also is used in many other statistical procedures.

ARE YOU ABLE TO DO THE FOLLOWING?

- Use the three measures of central tendency to interpret test results to a group.
- Provide an example of a distribution of scores in which the mean is higher than the median.
- Provide an example of a distribution of scores in which the median is higher than the mean.

Measures of Variability

You now are prepared to use the mean, median, and mode to interpret data in relation to a central grouping of scores. However, to provide a more meaningful interpretation you also need to know how the scores spread, or scatter. For example, it is possible for two classes to have the same mean on a skills test, but the spread of the scores can be entirely different. To illustrate this point, consider the following two sets of scores:

Group A: 80, 82, 83, 84, 86 $\bar{X} = 83$

Group B: 65, 75, 90, 90, 95 $\bar{X} = 83$

Both groups have a mean of 83, but the spreads of the scores are very different. The spread, or scatter, of scores is referred to as **variability.** The terms *dispersion* and *deviation* are often used to refer to variability. When groups of scores are compared, measures of variability should be considered as well as measures of central tendency. By knowing the measures of variability, you can determine the amount that the scores spread, or deviate, from the measures of central tendency. The measures of variability are the range, quartile deviation, and standard deviation.

The Range

The **range** is determined by subtracting the lowest score from the highest score. It is the easiest measure of variability to compute, but because it represents only the extreme scores and provides no distribution information, it is also the least useful. Two groups of data may have the same range but have very different distributions. Consider the following scores:

Group A: 97, 95, 89, 87, 86, 85, 83, 80, 75, 72

Group B: 81, 77, 75, 73, 70, 68, 64, 61, 58, 56

Both groups have a range of 25, but the distributions are not similar. It is possible for you to have completely different distributions when you administer the same knowledge or physical performance test to different groups.

Some characteristics of the range are as follows:

1. It is dependent on the two extreme scores.
2. Because it indicates nothing about the variability of the scores between the two extreme scores, it is the least useful measure of variability.

The formula for determining the range (R) is

$$R = \text{High score} - \text{Low score}$$

You may see this formula also:

$$R = H_x - L_x$$

The range for the distribution in table 2.3 is R = 96 − 81 = 15.

POINTS OF EMPHASIS

- Measures of variability are descriptive statistics, and they are reported with the measures of central tendency.
- Measures of variability will increase or decrease in size as the spread of the distribution of scores increases or decreases.
- If a group is similar in ability, the variability of the group usually will be small. If a group includes a wide range of ability, the variability usually will be high.
- In professional and research journals, the standard deviation of a group of scores usually is reported immediately after the mean.

ARE YOU ABLE TO DO THE FOLLOWING?

- Define the term *measures of variability*.
- Identify the letter used to represent the range.
- Define the range.
- Describe the characteristics of the range.
- Calculate the range with ungrouped data and use it to interpret the data.

The Quartile Deviation

Sometimes called the semi-quartile range, the **quartile deviation** is the spread of the middle 50% of the scores around the median. The quartile deviation is not reported often, but it is of value if the distribution is on the ordinal scale. The extreme scores will not affect the quartile deviation; thus, it is more stable than the range. Much like the other measures of variability, the quartile deviation is useful when comparing groups. Additionally, to determine the quartile deviation, the twenty-fifth and seventy-fifth percentiles must be calculated. These values will help you determine if the distribution is symmetrical about the median within the quartile deviation.

The following are some characteristics of the quartile deviation:

1. It uses the seventy-fifth percentile and twenty-fifth percentile to determine the deviation. The difference between these two percentiles is referred to as the interquartile range.
2. It indicates the amount that needs to be added to, and subtracted from, the median to include the middle 50% of the scores.
3. It usually is not used in additional statistical calculations.

The symbols used to calculate the quartile deviation are

Q = quartile deviation

Q_1 = twenty-fifth percentile or first quartile (P_{25} may be used also) = score in which 25% of the scores are below and 75% of the scores are above

Q_3 = seventy-fifth percentile, or third quartile (P_{75} may be used also) = score in which 75% of the scores are below and 25% of the scores are above

The steps for calculation of Q_3 are as follows:

1. Arrange the scores in ascending order.
2. Multiply N by .75 to find 75% of the distribution.
3. Count up from the bottom score to the number determined in step 2. Approximation and interpolation may be required to calculate Q_3 and Q_1 as well as other percentiles. Interpolation is necessary in the calculation of Q_1 from the distribution in table 2.3.

The steps for calculation of Q_1 are as follows:

1. Multiply N by .25 to find 25% of the distribution.

2. Again count up from the bottom score to the number determined in step 1.

To calculate Q, substitute the values in the formula

$$Q = \frac{Q_3 - Q_1}{2}$$

The calculation of Q from the distribution in table 2.3 is as follows:

1. .75(30) = 22.5
   The twenty-second score from the bottom is 90, and the twenty-third score is 90. Midway between these two scores would be the same score, so the score of 90 is at the 75%.
2. .25(30) = 7.5
   The seventh score from the bottom is 85, and the eighth score is 86. Since 7.5 is midway between these two scores, the score of 85.5 is at the 25%. Calculating the average of 85 and 86 would serve the same purpose.
3. $Q = \frac{90 - 85.5}{2} = \frac{4.5}{2} = 2.25$

Now what does 2.25 mean? Remember we said that the quartile deviation is used with the median. So now add 2.25 to 88 (the median in table 2.3) and subtract 2.25 from 88.

$$88 + 2.25 = 90.25$$
$$88 - 2.25 = 85.75$$

Theoretically, the middle 50% of the scores in the distribution should fall between the values 85.75 and 90.25. With thirty scores there should be fifteen scores between 85.75 and 90.25, but you will discover that sixteen scores fall between these two values. The difference is explained by the fact that the distribution in table 2.3 does not fulfill all requirements of normalcy. However, it is similar enough to use Q for interpretation purposes. Most of the test distributions you will use in your professional responsibilities will be similar enough to normalcy.

ARE YOU ABLE TO DO THE FOLLOWING?

- Identify the symbol for the quartile deviation.
- Define the quartile deviation.
- Describe the characteristics of the quartile deviation.
- Calculate the quartile deviation with ungrouped data and use it to interpret the data.

The Standard Deviation

The **standard deviation** is the most useful and sophisticated measure of variability. It describes the scatter of scores around the mean. The standard deviation is a more stable measure of variability than the range or quartile deviation because it depends on the weight of each score in the distribution.

The lowercase Greek letter sigma (σ) is used to indicate the standard deviation of a population, and the letter *s* is used to indicate the standard deviation of a sample. Because you generally will be working with small groups (or samples), the formula for determining the standard deviation will include (N − 1) rather than N. This adjustment produces a standard deviation that is closer to the population standard deviation.

Some characteristics of the standard deviation are as follows:

1. It is the square root of the variance, which is the average of the squared deviations from the mean. The variance is used in other statistical procedures. The population variance is represented as σ^2 and the sample variance is represented as s^2.
2. It is applicable to interval and ratio level data, includes all scores, and is the most reliable measure of variability.
3. It is used with the mean. In a normal distribution, one standard deviation added to the mean and one standard deviation subtracted from the mean include the middle 68.26% of the scores.
4. With most data, a relatively small standard deviation indicates that the group being tested has little variability; it has performed homogeneously. A relatively large standard deviation indicates the group has much variability; it has performed heterogeneously.
5. It is used to perform other statistical calculations. The standard deviation is especially important for comparing differences between means. Techniques for making these comparisons will be presented later in this chapter.

The symbols used to determine the standard deviation are as follows:

s = standard deviation
$\bar{X}$ = mean
Σ = sum of
d = deviation score $(X - \bar{X})$
X = individual score
N = number of scores

Two methods for determining the standard deviation will be presented. The first method requires only the use of the individual scores and a calculator that can compute the square root of a number. The second method requires the use of the squared deviations. The two methods obtain the same results.

Calculation with ΣX^2

1. Arrange the scores into a series.
2. Find ΣX.
3. Square each of the scores and add to determine the ΣX^2.
4. Insert the values into the formula

$$s = \sqrt{\frac{N\Sigma X^2 - (\Sigma X)^2}{N(N-1)}}$$

The calculation of s from the distribution in table 2.3 is as follows:

1. Scores are in a series.
2. $\Sigma X = 2{,}644$
3. $\Sigma X^2 = 233{,}398$
4.

$$\begin{aligned} s &= \sqrt{\frac{30(233{,}398) - (2{,}644)^2}{30(30-1)}} \\ &= \sqrt{\frac{7{,}001{,}940 - 6{,}990{,}736}{30(29)}} \\ &= \sqrt{\frac{11{,}204}{870}} \\ &= \sqrt{12.8781} \\ &= 3.59 \\ s &= 3.6 \end{aligned}$$

Calculation with Σd^2

1. Arrange the scores into a series.
2. Calculate $\bar{X}$.
3. Determine d and d^2 for each score; then calculate Σd^2.
4. Insert the values into the formula

$$s = \sqrt{\frac{\Sigma d^2}{N-1}}$$

The calculation of s from the distribution in table 2.4 is as follows:

1. Scores are in a series.
2. $\bar{X} = 88.1$
3. $\Sigma d^2 = 373.5$
4.

$$\begin{aligned} s &= \sqrt{\frac{373.5}{30-1}} \\ &= \sqrt{\frac{373.5}{29}} \\ &= \sqrt{12.8793} \\ s &= 3.6 \end{aligned}$$

Note: The s value should be the same value as found in the ΣX^2 method, but owing to the rounding off of $\bar{X}$, there is a slight difference.

To determine the middle 68.26% of the scores, you now add 3.6 to $\bar{X}$ and subtract 3.6 from $\bar{X}$.

$$\bar{X} + 3.6 = 88.1 + 3.6 = 91.7$$
$$\bar{X} - 3.6 = 88.1 - 3.6 = 84.5$$

If the distribution were normal, the middle 68.26% of the scores would be between the values 84.5 and 91.7.

Because each score must be subtracted from the mean, and the mean is often not a whole number, this method can be time-consuming.

Relationship of Standard Deviation and Normal Curve

The use of the standard deviation has more meaning when it is related to the normal curve. On the basis of the probability of a normal distribution, there is an exact relationship between the standard deviation and the proportion of area and scores under the curve. The standard deviation marks off points along the base of the curve. An equal percentage of the curve will be found between the mean plus one standard deviation and the mean minus one standard deviation. The same is true for plus and minus 2.0 or 3.0 standard deviations.

The following observations can be made about the standard deviation and the areas under a normal curve:

1. 68.26% of the scores will fall between +1.0 and −1.0 standard deviations.
2. 95.44% of the scores will fall between +2.0 and −2.0 standard deviations.
3. 99.73% of the scores will fall between +3.0 and −3.0 standard deviations. Generally, scores will not exceed +3.0 and −3.0 standard deviations from the mean. Figure 2.4 shows these observations.

The relationship of the standard deviation and the normal curve provides you a meaningful and consistent way to compare the performance of different groups using the same test and to compare the performance of one individual with the group. In addition, by knowing the value of the mean and of the standard deviation, you can express the percentile rank of the scores. To illustrate how these procedures can be done, consider the following example.

As part of a physical fitness test, a fitness instructor administered a 60-second sit-up test to two fitness classes. She found the mean and the standard deviation for each class to be as follows:

Class 1	*Class 2*
$\bar{X} = 32$	$\bar{X} = 28$
s = 2	s = 4

Figure 2.5 compares the spread of the two distributions. The spread of scores in class 1 (±3 standard deviations from the mean) is much less than the spread of scores for class 2. Class 1 is more homogeneous.

Now consider individual A in class 1, who completed thirty-four sit-ups, and individual B in class 2, who also completed thirty-four sit-ups. Though the two individuals have the same score, figure 2.6 shows that they do not have the same relationship to their respective class means and standard deviations. Individual A is 1 standard deviation above the class 1 mean, and individual B is 1.5 standard deviations above the class 2 mean. Table 2.5 shows that +1 standard deviation above the mean includes approximately 84% of the curve and that +1.5 standard deviations above the mean include approximately 93% of the curve. Though the two scores are the same, they do not have the same percentile score.

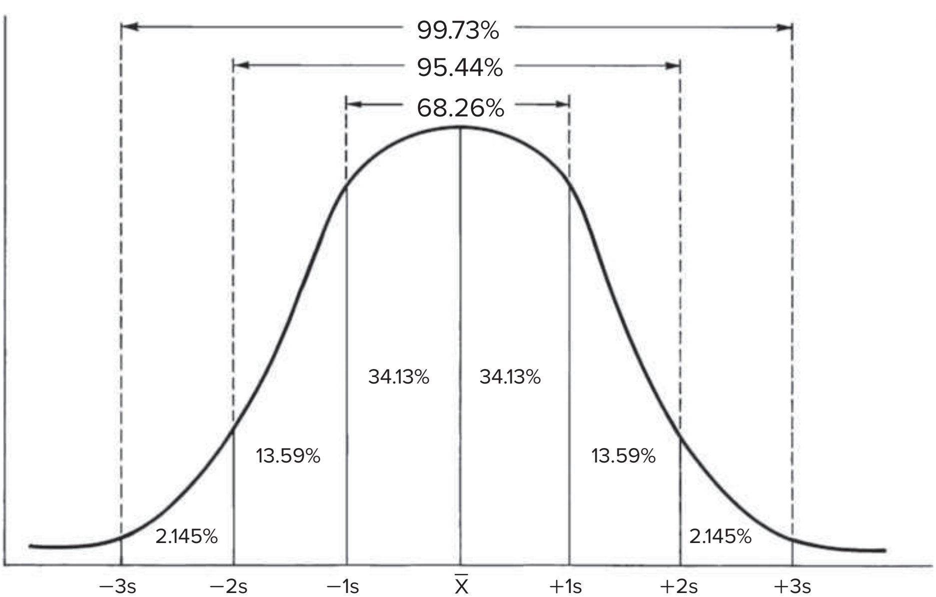


FIGURE 2.4 Characteristics of normal curve.

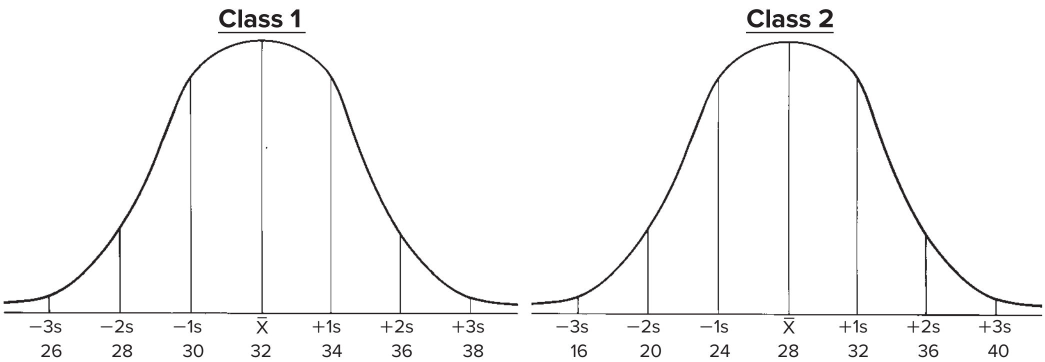


FIGURE 2.5 Comparison of $\bar{X}$ and s for sit-up test.

The steps for calculating the percentile rank through use of the mean and the standard deviation are as follows:

1. Calculate the deviation of the score from the mean:

$$d = (X - \bar{X})$$

2. Calculate the number of standard deviation units the score is from the mean. Some textbooks refer to these units as z-scores. The use and interpretation of z-scores are described later in this chapter.

$$\text{No. of standard deviation units from the mean} = \frac{d}{s}$$

3. Use table 2.5 to determine where the percentile rank of the score is on the curve. On occasions it may be necessary to approximate the percentile rank.

Note: If a negative value is found in step 1, the percentile rank will always be less than 50. If a positive value is found in step 1, the percentile rank will always be more than 50.

ARE YOU ABLE TO DO THE FOLLOWING?

- Identify the symbol for the standard deviation.
- Define the standard deviation.
- Describe the characteristics of the standard deviation.
- Calculate the standard deviation with ungrouped data and use it to interpret the data.
- Describe the normal curve and the relationship of the standard deviation and the normal curve.

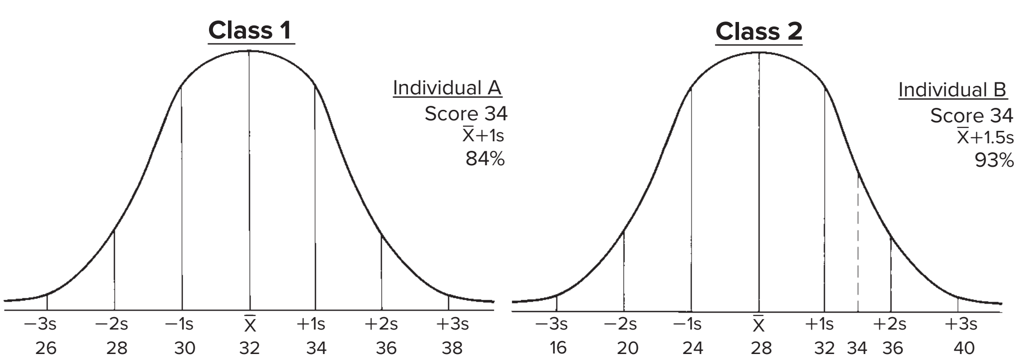


FIGURE 2.6 Comparison of individual performances for sit-up test.

TABLE 2.5 Percentile Scores Based on the Mean and Standard Deviation Units

$\bar{X}$ and s Units	Percentile Rank	T-score	$\bar{X}$ and s Units	Percentile Rank	T-score
$\bar{X}$ + 3.0s	99.87	80	$\bar{X}$ – 0.1s	46.02	49
$\bar{X}$ + 2.9s	99.81	79	$\bar{X}$ – 0.2s	42.07	48
$\bar{X}$ + 2.8s	99.74	78	$\bar{X}$ – 0.3s	38.21	47
$\bar{X}$ + 2.7s	99.65	77	$\bar{X}$ – 0.4s	34.46	46
$\bar{X}$ + 2.6s	99.53	76	$\bar{X}$ – 0.5s	30.85	45
$\bar{X}$ + 2.5s	99.38	75	$\bar{X}$ – 0.6s	27.43	44
$\bar{X}$ + 2.4s	99.18	74	$\bar{X}$ – 0.7s	24.20	43
$\bar{X}$ + 2.3s	98.93	73	$\bar{X}$ – 0.8s	21.19	42
$\bar{X}$ + 2.2s	98.61	72	$\bar{X}$ – 0.9s	18.41	41
$\bar{X}$ + 2.1s	98.21	71	$\bar{X}$ – 1.0s	15.87	40
$\bar{X}$ + 2.0s	97.72	70	$\bar{X}$ – 1.1s	13.57	39
$\bar{X}$ + 1.9s	97.13	69	$\bar{X}$ – 1.2s	11.51	38
$\bar{X}$ + 1.8s	96.41	68	$\bar{X}$ – 1.3s	9.68	37
$\bar{X}$ + 1.7s	95.54	67	$\bar{X}$ – 1.4s	8.08	36
$\bar{X}$ + 1.6s	94.52	66	$\bar{X}$ – 1.5s	6.68	35
$\bar{X}$ + 1.5s	93.32	65	$\bar{X}$ – 1.6s	5.48	34
$\bar{X}$ + 1.4s	91.92	64	$\bar{X}$ – 1.7s	4.46	33
$\bar{X}$ + 1.3s	90.32	63	$\bar{X}$ – 1.8s	3.59	32
$\bar{X}$ + 1.2s	88.49	62	$\bar{X}$ – 1.9s	2.87	31
$\bar{X}$ + 1.1s	86.43	61	$\bar{X}$ – 2.0s	2.28	30
$\bar{X}$ + 1.0s	84.13	60	$\bar{X}$ – 2.1s	1.79	29
$\bar{X}$ + 0.9s	81.59	59	$\bar{X}$ – 2.2s	1.39	28
$\bar{X}$ + 0.8s	78.81	58	$\bar{X}$ – 2.3s	1.07	27
$\bar{X}$ + 0.7s	75.80	57	$\bar{X}$ – 2.4s	0.82	26
$\bar{X}$ + 0.6s	72.57	56	$\bar{X}$ – 2.5s	0.62	25
$\bar{X}$ + 0.5s	69.15	55	$\bar{X}$ – 2.6s	0.47	24
$\bar{X}$ + 0.4s	65.54	54	$\bar{X}$ – 2.7s	0.35	23
$\bar{X}$ + 0.3s	61.79	53	$\bar{X}$ – 2.8s	0.26	22
$\bar{X}$ + 0.2s	57.93	52	$\bar{X}$ – 2.9s	0.19	21
$\bar{X}$ + 0.1s	53.98	51	$\bar{X}$ – 3.0s	0.13	20
$\bar{X}$ + 0.0s	50.00	50			

Which Measure of Variability Is Best for Interpretation of Test Results?

You have studied the definitions of the three measures of variability, how to calculate them, and some characteristics of each. In deciding which of the three is best for interpreting test results of a group you might be testing, you should consider the following:

1. The range is the least reliable of the three, but it is used when a fast method is needed. It does not take into account all the scores and may not represent the true variability of the scores.
2. The quartile deviation is more meaningful than the range, but it considers only the middle 50% of the

scores. Thus, it too is limited in its ability to represent the true variability of the scores.

3. The standard deviation considers every score, is the most reliable, and is the most commonly used measure of variability.

ARE YOU ABLE TO DO THE FOLLOWING?

- Properly use the three measures of variability to interpret test results to a group.
- State why the standard deviation is the most often reported measure of variability.

Percentiles and Percentile Ranks

Though you have learned to calculate percentiles and percentile rank, it will be beneficial to discuss them in greater detail. **Percentile** refers to a point in a distribution of scores below which a given percentage of the scores fall. For example, the sixtieth percentile is the point where 60% of the scores in a distribution are below and 40% of the scores are above. To calculate a percentile, you first multiply N by the desired percentage. You then determine the score that is at that percentile in a distribution. In table 2.3, the sixtieth percentile score is 89.

The **percentile rank** of a given score in a distribution is the percentage of the total scores that fall below the given score. A percentile rank, then, indicates the position of a score in a distribution in percentage terms. Percentile ranks are determined by beginning with the raw scores and calculating the percentile ranks for the scores. In table 2.3, the score of 89 has a percentile rank of 60.

Although percentiles are of value for interpretation of data, they do have weaknesses. The relative distances between percentile scores are the same, but the relative distances between the observed scores are not. Because percentiles are based on the number of scores in a distribution rather than on the size of the raw score obtained, it is sometimes more difficult to increase a percentile score at the ends of the scale than in the middle. The average performers, whose raw scores are found in the middle of the scale, need only a small change in their raw scores to produce a large change in their percentile scores. However, the below-average and above-average performers, whose raw scores are found at the ends of the scale, need a large change in their raw scores to produce even a small change in their percentile scores. This weakness is usually found in all percentile scores.

If you do not remember how to calculate percentiles, you should refer to the sections on the median and quartile deviation. In table 2.3, the percentile scale is divided into deciles (ten equal parts). Deciles are represented as D_1 (tenth percentile), D_2 (twentieth percentile), D_3 (thirtieth percentile), on up to D_9 (ninetieth percentile).

Frequency Distribution

Statistics software programs will group data into frequency distributions. In such a distribution all scores are listed in ascending or descending order, and the number of times each individual score occurs is indicated in a frequency column. The percentage of times that each score occurs and the cumulative percentage (the percentage of scores below a given score) are also presented. Table 2.6 shows the frequency distribution for 100 push-up scores.

◆ ◆ ◆ POINTS OF EMPHASIS ◆ ◆ ◆

- The difficulty in describing and analyzing data or test scores is not calculating the statistics, but knowing why you wish to perform the calculations. You should know the questions for which you are seeking answers. Descriptive statistics will enable you to compare groups or compare the performances of members in the same group after the administration of different tests.

TABLE 2.6 Frequency Distribution of Push-Up Scores

Score	Frequency	Percent	Cumulative Percent
27	1	1.0	1.0
28	1	1.0	2.0
29	1	1.0	3.0
30	2	2.0	5.0
31	3	3.0	8.0
32	3	3.0	11.0
33	4	4.0	15.0
34	3	3.0	18.0
35	5	5.0	23.0
36	6	6.0	29.0
37	6	6.0	35.0
38	5	5.0	40.0
39	6	6.0	46.0
40	8	8.0	54.0
41	6	6.0	60.0
42	6	6.0	66.0
43	7	7.0	73.0
44	6	6.0	79.0
45	5	5.0	84.0
46	4	4.0	88.0
47	4	4.0	92.0
48	3	3.0	95.0
49	2	2.0	97.0
50	2	2.0	99.0
51	1	1.0	100.0
	100		

ARE YOU ABLE TO DO THE FOLLOWING?

- Identify the symbols for percentiles, quartiles, and deciles.
- Define *percentile* and *percentile rank*.
- Calculate percentiles and use them to interpret the data.

Graphs

Data are often presented in graphic form. Well-prepared graphs enable individuals to interpret data without reading the raw data or tables. SPSS and other computer programs can create column, bar, line, pie, area, and other types of graphs. In your future professional responsibilities, you may be asked to provide reports that include much data. If you have the opportunity to develop the knowledge and skill needed to use such computer programs, you should do so.

Figure 2.7 is a histogram (bar graph) of 75 tennis serve scores. For graphing purposes, computer programs will group a large number of scores into groups and plot the midpoint of each group. In figure 2.7, there may be three possible scores in each interval. In the first interval, the score of 49 represents the midpoint of scores 48–50. The frequency column shows that one score occurred in this interval. The next interval is 51–53, and two scores occurred in this interval. The highest number of scores, 10, occurred in the interval 72–74. Figure 2.8 is a frequency polygon (line graph) of the same scores. With these graphs, it is easy to observe the frequency of the serve scores.

Figures 2.9 and 2.10 are a pie chart and a bar chart, respectively, of causes of deaths in the year 2005. These charts show the percentage of each type of death in relation to all deaths. Each chart enables the reader to interpret the data quickly.

Standard Scores

After collecting scores for different performances, you may want to combine or compare scores; but because the scores have no similarities, you cannot perform these functions. For example, suppose you are teaching a physical fitness unit to a high school class, and at the conclusion of the unit you administer a 1-minute sit-up test, the sit and reach test, a 2-mile run, and a physical fitness knowledge test. How do you average the scores of the four tests? You certainly cannot add the scores and divide by four. How do you compare a student's performance on the sit-up test with her performance on the 2-mile run? You cannot tell her that forty sit-ups is a better score than a time of 15:10—unless you have a procedure to convert the raw scores into standard scores. *The conversion of unlike raw scores to standard scores enables you to compare different types of scoring.* Procedures to make such conversions follow.

z-Scores

A **z-score** represents the number of standard deviations by which a raw score deviates from the mean. After calculation of the mean and standard deviation of a distribution, it is possible to determine the z-score for any raw score in the distribution through the use of this formula:

$$z = \frac{X - \overline{X}}{s}$$

Consider the example of the 75 tennis serve scores represented in figures 2.7 and 2.8. The actual scores are as follows:

88 83 75 81 56 82 86 62 87 79 93 58 61 61 75
73 94 48 79 72 81 85 52 73 62 80 73 84 63 61
67 63 75 73 67 72 73 72 77 73 85 82 70 57 58
54 79 68 54 70 77 81 68 83 65 77 90 52 75 62
84 69 56 68 69 63 70 91 70 80 65 70 88 72 63

For these scores, $\overline{X} = 72.05$ and $s = 10.79$. To convert the tennis serve scores into z-scores, you substitute each individual score in the formula. For the scores of 88 and 54, the z-scores would be

$$z = \frac{88 - 72.1}{10.8} \qquad z = \frac{54 - 72.1}{10.8}$$

$$= \frac{15.9}{10.8} \qquad = \frac{-18.1}{10.8} \qquad (\overline{X} \text{ and s are rounded to one decimal place})$$

$$z = 1.47 \qquad z = -1.68$$

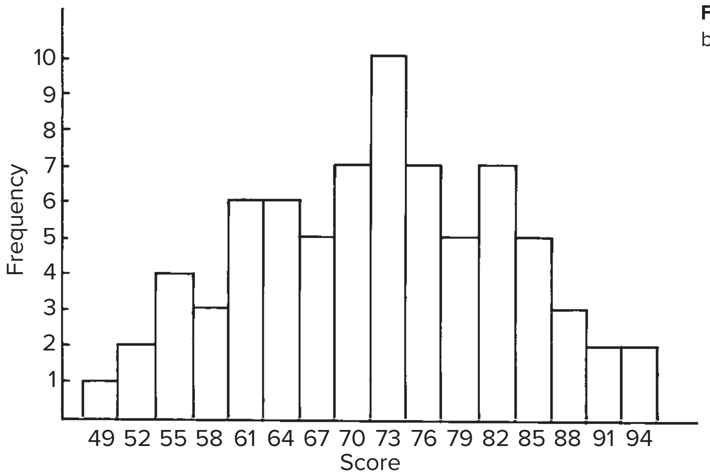


FIGURE 2.7 Histogram of tennis serve scores made by 75 students.

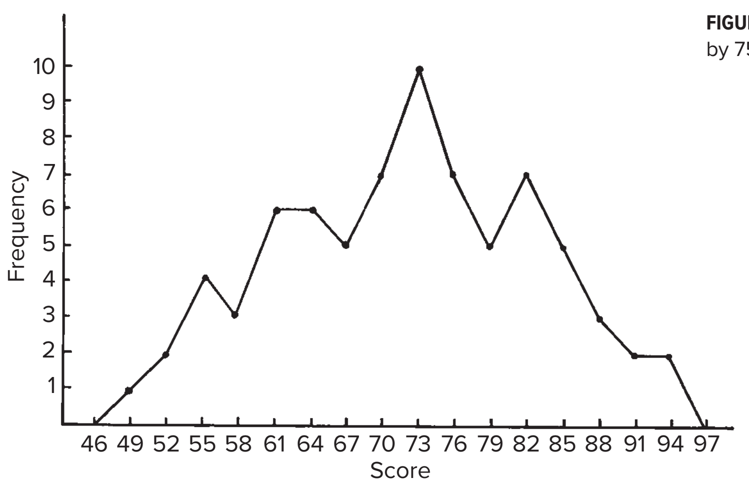


FIGURE 2.8 Frequency polygon of tennis serve scores made by 75 students.

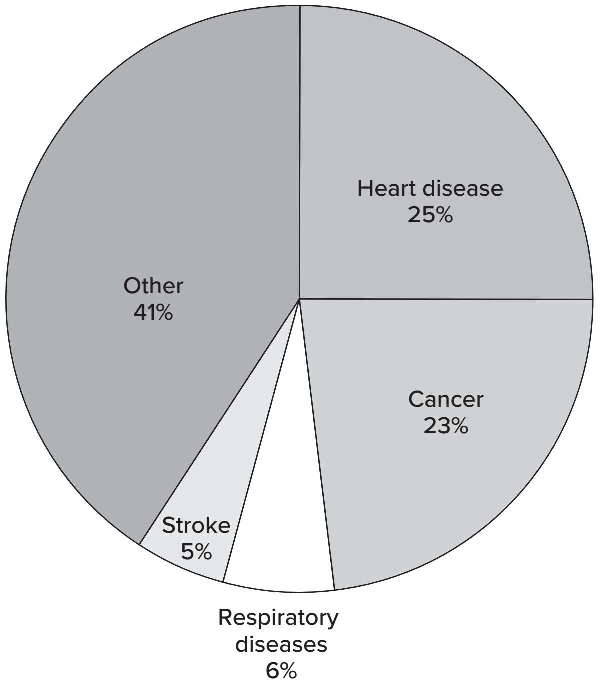


FIGURE 2.9 Pie chart of causes of death in 2009.

How are these z-scores interpreted? The z-scale has a mean of 0 and a standard deviation of 1, and it normally extends from −3 to +3 (plus and minus 3 standard deviations from the mean include 99.73% of the scores). In observing figure 2.11, you see that a z-score of 1.47 is approximately 1.5 standard deviations above the mean, and a z-score of −1.68 is more than 1.5 standard deviations below the mean. Knowing the relationship of the standard deviation and the normal curve, we can state that 1.47 is an excellent z-score and −1.68 is a poor z-score. Also, by referring to table 2.5, we see that 1.5 standard deviations above the mean has a percentile rank of 93, and 1.7 deviations below the mean has a percentile rank of 4.

If other tennis skills tests (e.g., forehand, backhand, and lob tests) had been administered to the same 75 test takers, all scores could be converted to z-scores and averaged for one tennis skill score. Also, each individual's z-score for the four tests could be compared, and the strongest and weakest skills of each individual determined.

All standard scores are based on the z-score. Because z-scores are expressed in small numbers, involve decimals, and may be positive or negative, many testers do not use them.

T-Scores

The **T-scale** has a mean of 50 and a standard deviation of 10. T-scores may extend from 0 to 100, but it is unlikely that any T-score would be below 20 or above 80, since this range includes plus and minus 3 standard deviations. Figure 2.11 shows the relationship of z-scores, T-scores, and the normal curve. The z-score is part of the formula for conversion of raw scores into T-scores. The formula is

$$\text{T-score} = 50 + 10\left(\frac{\left(X - \overline{X}\right)}{s}\right) = 50 + 10\,z$$

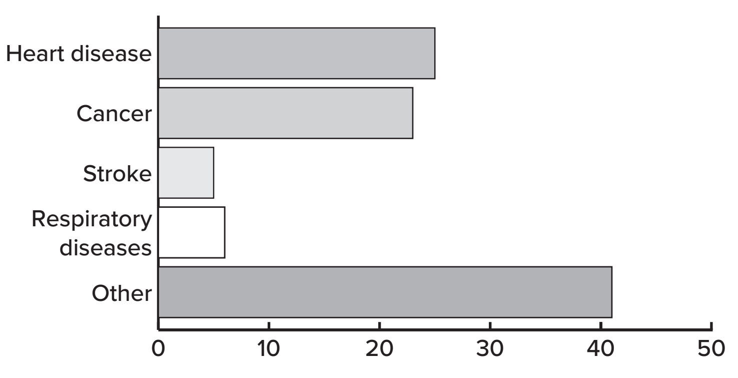


FIGURE 2.10 Bar graph of causes of death in 2009.

◆ ◆ ◆ POINTS OF EMPHASIS ◆ ◆ ◆

- With most performances, z-scores increase as the performance score increases. For some events and measurements, however, a smaller score may be desirable (e.g., timed events and heart rate). When a low score is the better score, the z-score formula is modified as follows:

$$z = \frac{\bar{X} - X}{s}$$

Again, consider the scores of 88 and 54 for the tennis serve test. The T-scores are

$$T_{88} = 50 + 10(1.47) \qquad T_{54} = 50 + 10(-1.68)$$
$$= 50 + 14.7 \qquad = 50 + (-16.8)$$
$$T_{88} = 64.7 = 65 \qquad T_{54} = 33.2 = 33$$

Note: T-scores are reported as whole numbers.

T-scores may be used in the same way as z-scores. However, because only positive whole numbers are reported and the range is 0 to 100, the T-scale is easier to interpret. However, it is sometimes confusing to the individuals being tested when they are told that a T-score of 60 or above is a good score. If you use T-scores, you should be prepared to fully explain their meanings. Table 2.5 shows the relationship of the mean and standard deviation, percentile rank, and T-scores.

You may prefer to convert the raw scores in a distribution to T-scores through the following procedure:

1. Number a column of T-scores from 20 to 80.
2. Place the mean of the distribution of the scores opposite the T-score of 50.
3. Divide the standard deviation of the distribution by 10. The standard deviation for the T-scale is 10, so each T-score from 0 to 100 is one-tenth of the standard deviation.
4. Add the value found in step 3 to the mean and each subsequent number until you reach the T-score of 80.
5. Subtract the value found in step 3 from the mean and each decreasing number until you reach the number 20.
6. Round off the scores to the nearest whole number.

Note: If a low score is better than a high score, subtract the T-values toward 80 and add the values toward 20. Table 2.7 shows the conversion of the 75 tennis serve scores to T-scores. For illustration purposes, only the T-scores 40 to 60 are included in the table. The value found in dividing the standard deviation by 10 (1.079) is rounded to 1.1.

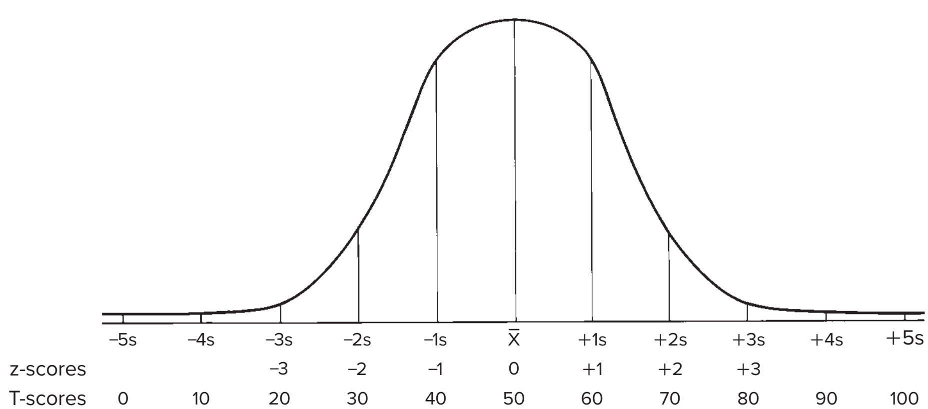


FIGURE 2.11 z-scores and T-scores plotted on a normal curve.

POINTS OF EMPHASIS

- As with z-scores, the T-score formula should be modified when a low score is desirable. For such events, the formula is as follows:

$$\text{T-score} = 50 + 10\left(\frac{\bar{X} - X}{s}\right)$$

TABLE 2.7 Conversion of Tennis Serve Scores to T-Scores

T-Score	Observed Score
60	83.1 = 83
59	82.0 = 82
58	80.9 = 81
57	79.8 = 80
56	78.7 = 79
55	77.6 = 78
54	76.5 = 77
53	75.4 = 75
52	74.3 = 74
51	73.2 = 73
50	72.1 = 72
49	71.0 = 71
48	69.9 = 70
47	68.8 = 69
46	67.7 = 68
45	66.6 = 67
44	65.5 = 66
43	64.4 = 64
42	63.3 = 63
41	62.2 = 62
40	61.1 = 61

$\bar{X} = 72.1$

$s = 10.79$

$$\frac{s}{10}\left(\begin{matrix}\text{value added to and}\\ \text{subtracted from } \bar{X}\end{matrix}\right) = \frac{10.79}{10} = 1.079 = 1.1$$

Percentiles

Percentile scores are also standard scores and may be used to compare scores of different measurements. Because they change at different rates (remember the comparison of low and high percentile scores with middle percentiles), they should not be averaged to determine one score for several different tests. For this reason, you may prefer to use the T-scale when converting raw scores to standard scores. The calculation of percentiles was previously described, so it will not be described here.

Statistics Software

Numerous statistics software programs—varying in cost, statistical procedures, computer requirements, and ease of use—may be purchased. In addition, student versions of many of these programs are available. Statistical Package for Social Science (SPSS) is one such program. It can be used to perform the statistical procedures presented in this chapter and chapters 3 and 4. Microsoft Excel also may be used to perform the descriptive statistics presented in this chapter. Additionally, Microsoft Excel provides supplemental statistical software programs that are used with the Excel spreadsheet to perform more complex statistical procedures.

You should become familiar with a statistical software program during your college experience. Use of the microcomputer and statistics software will significantly reduce the time needed to perform statistical procedures, and you will feel more confident in the results of your work than if you perform the calculations yourself. The use of statistics software will be especially helpful in regard to the statistics presented in chapters 3 and 4. Another reason for developing the ability to use statistics software is that potential employers may expect you to have this skill.

Information about statistics programs is available in campus computing centers or bookstores and from computer software retailers. In addition, most statistics software producers have a Web home page.

ARE YOU ABLE TO DO THE FOLLOWING?

- Describe the purposes of standard scores.
- Convert raw scores into z-scores, T-scores, and percentiles and interpret them.

Review Problems

1. Mrs. Block completed the volleyball unit with her 2 ninth-grade classes by administering the same volleyball skills test to both classes. One of the test items was the volleyball serve. Calculate the range, $\bar{X}$, P_{50}, mode, Q, s, and deciles for the two groups of scores. Were the performances of the classes different in any way? Was one class more homogeneous?

   Class A

   42, 50, 57, 45, 56, 69, 45, 43, 46, 51, 61, 55, 40, 47, 59, 47, 46, 30, 48, 53, 40, 40, 64, 48, 41

   Class B

   51, 43, 43, 37, 33, 53, 44, 44, 38, 34, 37, 51, 20, 24, 39, 34, 38, 50, 10, 37, 39, 27, 25, 29, 23

2. Prior to working with a group of twenty adults, ages forty to forty-nine, Mr. Fitness administered a sit and reach test (flexibility of lower back and posterior thighs). After a six-week flexibility program, Mr. Fitness tested the group again. Compare the medians, means, quartile deviations, and standard deviations of the two test administrations. Did the flexibility of the group change? If so, describe the changes.

   Test Administration 1

   10, 9, 11, 9, 10, 11, 12, 8, 9, 10, 8, 7, 9, 8, 10, 10, 7, 9, 11, 9

   Test Administration 2

   12, 11, 13, 11, 11, 14, 12, 12, 14, 12, 11, 11, 13, 12, 13, 14, 10, 11, 14, 12

3. Calculate the $\bar{X}$ and s for these 2-minute sit-up test scores:

   42, 41, 53, 60, 84, 49, 57, 65, 61, 48, 33, 57, 55, 50, 65, 54, 55, 57, 65, 45, 58, 54, 52, 40, 55

4. Given the following information, use the normal curve and determine the percentile rank for the three observed scores:

$$\bar{X} = 33$$
$$s = 2$$
$$X_1 = 36$$
$$X_2 = 29$$
$$X_3 = 32$$

5. Mr. Bird administered a badminton clear test to sixty students. Determine the mean and standard deviation, and convert the scores into T-scores. The scores are as follows:

74	75	64	86	73	74	75	70	69	67
80	78	61	81	77	78	65	65	70	69
85	84	83	62	84	74	75	81	66	74
63	73	77	64	66	72	80	77	80	70
66	69	72	83	69	67	73	72	75	75
70	67	65	73	73	72	72	73	74	73

Chapter Review Questions

1. Which scales of measurement include how much the variables differ from each other?
2. Which scale of measurement has a true zero point?
3. If you ask other students to name their major, which scale of measure are you using?
4. What are the four characteristics of a normal distribution?

5. If you administered a performance test to a homogeneous group, what type of distribution curve probably would result? What curve probably would result if you administered the same test to a heterogeneous group?

6. What information is provided when you rank the scores in a distribution?

7. What information do measures of central tendency provide about a distribution of scores?

8. What is the average score in a distribution?

9. What score represents the exact middle in a distribution?

10. With the scores 3, 6, 9, 12, and 15, what are the mean and median values? If the score of 15 is changed to a score of 18, what are the mean and median values? Why will the median value be the same for both distributions?

11. What do the symbols $\bar{X}$, X, Σ, N, and P_{50} represent?

12. What is the most frequently occurring score in a distribution?

13. Which measure of central tendency is not influenced by extreme scores?

14. What can be said about the values of the mean, median, and mode in a normal distribution?

15. In what type of distribution is the mean lower than the median? When would the mean be higher than the median?

16. Which way is the tail of the curve pointed if a distribution of scores produces a negatively skewed curve? In a negatively skewed distribution, is the mean or median a higher value? If this type of curve is produced after the administration of a test, what can you say about the test?

17. What information do the measures of variability provide about a distribution of scores?

18. What do the symbols Q, Q_1, Q_2, Q_3, d, and s represent?

19. What is the weakness of the range measurement?

20. With which measure of central tendency is the quartile deviation used? Which percentage of scores is included in the quartile deviation?

21. With which measure of central tendency is the standard deviation used?

22. In a normal distribution, what percentage of the curve is found between the mean plus and minus one standard deviation, between the mean plus and minus two standard deviations, and the mean plus and minus three standard deviations?

23. In a normal distribution, what percentage of the curve is found between the mean plus one standard deviation?

24. In a normal distribution of scores, a score is one standard deviation above the mean. What percentage of scores is below the given score?

25. In a distribution of scores, the $\bar{X} = 30$ and the $s = 2$. What percentage of the scores is found between the scores 28 and 32? What percentage is found between the scores 26 and 34?

26. If two different groups are administered the same physical performance test, what can you say about the group that has the largest standard deviation?

27. Which measure of variability is the most reliable and most commonly used?

28. What does it mean if someone's test score is at the seventy-fifth percentile?

29. If your score for a test is one standard deviation above the mean, what is your percentile rank?

30. What is the purpose of standard scores?

31. With a z-scale, what is the value of the mean score? How many standard deviations above the mean is a z-score of 1? What is the percentile rank of a z-score of 1?

32. With a T-scale, what is the value of the mean score? Suppose you score one and one-half standard deviations above the mean on a test. What would be your T-score?

33. Normally, what are the highest and lowest z-scores and T-scores that will occur? Why?

This page intentionally left blank

3 Investigating the Relationship between Scores

After completing this chapter, you should be able to

1. Define *correlation* and *linear correlation*, interpret the correlation coefficient (statistical significance and coefficient of determination), and use the Spearman rank-difference and Pearson product-moment methods to determine the relationship between two variables.
2. Construct a scattergram and interpret it.

In chapter 2, we discussed descriptive statistics. Although it is important that you have the skills to use these statistics, your professional responsibilities will require that you be knowledgeable about other statistical procedures. You also should be able to demonstrate the relationship of scores. Skills in this procedure will enable you to better interpret professional literature and to conduct beneficial research.

Linear Correlation

Correlation is a statistical technique used to express the relationship between two sets of scores (two variables). In this chapter, we will discuss **linear correlation,** or the degree to which a straight line best describes the relationship between two variables. A relationship between two variables may be curvilinear, meaning that the relationship is best described with a curved line. Linear relationship is the simplest and most common correlation, however. In our profession, many opportunities exist for determining if there is a relationship between two variables. For example, is there a relationship between athletic participation and academic achievement? Do individuals with a high level of physical fitness earn higher academic grades? Is there a relationship between arm strength and golf driving distance? Is there a relationship between percentage of body fat and the ability to run 2 miles? Also, correlation techniques are used to determine the validity, reliability, and objectivity of tests. These techniques will be described in chapter 5.

The number that represents the correlation is called the **correlation coefficient.** Two techniques for determining the correlation coefficient will be presented here. Regardless of the technique used, correlation coefficients have several common characteristics. (Statements 2 and 3 are general statements; the size and significance of the coefficient must be considered.)

1. The values of the coefficient will always range from +1.00 to −1.00. It is rare that the coefficients of +1.00, −1.00, and .00 are found, however.
2. A positive coefficient indicates direct relationship; for example, an individual who scores high on one variable is likely to score high on the second variable, and an individual who scores low on one variable is likely to score low on the second variable.
3. A negative coefficient indicates inverse relationship. The individual who scores low on one variable is likely to score high on the second variable, and the individual who scores high on the first variable is likely to score low on the second.

4. A correlation coefficient of, or very close to, .00 indicates no relationship. An individual who scores high or low on one variable may have any score on the second variable.

5. The number indicates the degree of *relationship,* and the sign indicates the type of relationship. The number +.88 indicates the same degree of relationship as the number −.88. The signs indicate that the directions of the relationship are different.

6. A correlation coefficient indicates relationship. After determining a correlation coefficient, you cannot infer that one variable causes something to happen to the other variable. If a high, positive correlation coefficient is found between participation in school sports and high academic grades, it cannot be said that participation in school sports causes a person to earn good grades. It can only be said that there is a high, positive relationship.

Scattergram

A **scattergram** is a graph used to illustrate the relationship between two variables. To prepare a scattergram, perform the following steps:

1. Determine the range for each variable.

2. Designate one variable as the X score and the other variable as the Y score.

3. Draw and label the axes. Represent the X scores on the horizontal axis and the Y scores on the vertical axis. Begin with the lower X scores at the left portion of the X axis and the lower Y scores at the lower portion of the Y axis.

4. Plot each pair of scores on the graph by placing a point at the intersection of the two scores.

Figure 3.1 is a scattergram of the relationship between isometric and isotonic strength scores.

The scattergram can indicate a positive relationship, a negative relationship, or a zero relationship. If a positive relationship exists, the points will tend to cluster along a diagonal line that runs from the lower left-hand corner of the scattergram to the upper right-hand corner. In a negative relationship, the points tend to do the opposite; they move from the upper left-hand corner to the lower right-hand corner. With a positive or negative line, the closer the points cluster along the diagonal line, the higher the correlation. In a zero relationship, the points are scattered throughout the scattergram. Figure 3.2 illustrates the three types of relationships.

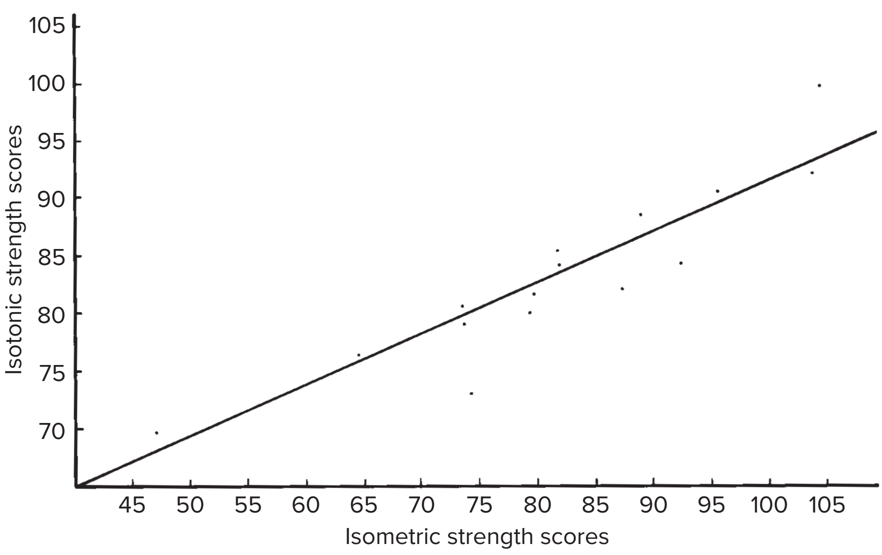


FIGURE 3.1 Scattergram of relationship between isometric and isotonic strength scores.

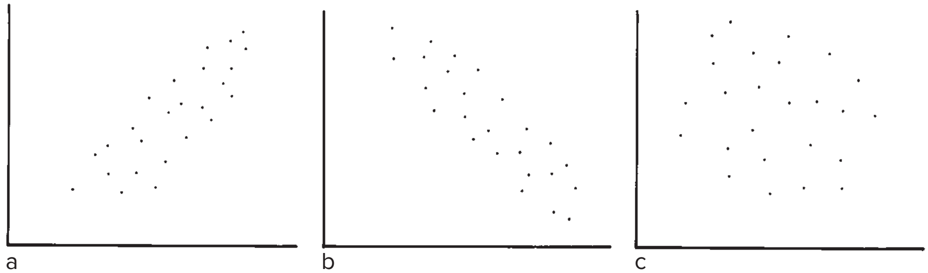


FIGURE 3.2 Scattergrams showing (a) positive, (b) negative, and (c) zero correlation between two variables.

◆ ◆ ◆ POINTS OF EMPHASIS ◆ ◆ ◆

The Spearman rank-difference correlation and the Pearson product-moment correlation are presented in this chapter. To use these procedures:

- The data used must be paired; this means you must have two scores for each participant in your group.
- The data must be bivariate, which means you must have measures for two different variables; two measures cannot be made on the same variable.

Spearman Rank-Difference Correlation Coefficient

The **Spearman rank-difference correlation coefficient,** also called rank-order, is used when one or both variables are ranks or ordinal scales. The difference (D) between the ranks of the two sets of scores is used to determine the correlation coefficient. The following examples are relationships that could be determined through utilization of the rank-difference correlation coefficient:

- The ranking of participants in a badminton class and their order of finish in tournament play.
- The ability to serve well and the order of finish in a tennis tournament.
- Vertical jump scores and speed in the 100-meter run.

Of course, many other relationships can be determined with the rank-difference correlation coefficient.

The symbol for the rank-difference coefficient is the Greek rho (ρ) or $r_{\rho\eta o}$. To determine ρ, perform the following steps:

1. List each set of scores in a column.
2. Rank the two sets of scores.
   Note: This procedure was described in chapter 2 (see table 2.2).
3. Place the appropriate rank beside each score.
4. Head a column D and determine the difference in rank for each pair of scores.
   Note: The sum of the D column should always be 0. If it is not, check your work.
5. Square each number in the D column and sum the values (ΣD^2).
6. Calculate the correlation coefficient by substituting the values in the formula

$$\rho = 1.00 - \frac{6(\Sigma D^2)}{N(N^2 - 1)}$$

Table 3.1 illustrates the calculation of the rank-difference correlation coefficient for sit-up and push-up scores.

TABLE 3.1 Rank-Difference Correlation Coefficient for Sit-Up and Push-Up Scores

Student	Sit-Up Score	Rank of S-U Score	Push-Up Score	Rank of P-U Score	D	D^2
A	28	19.5	19	18.0	1.5	2.25
B	31	16.5	22	15.5	1.0	1.00
C	32	13.5	25	11.0	2.5	6.25
D	35	7.0	23	13.5	–6.5	42.25
E	32	13.5	27	8.5	5.0	25.00
F	40	1.0	35	1.0	0.0	0.00
G	33	10.0	29	4.5	5.5	30.25
H	35	7.0	28	6.5	0.5	0.25
I	32	13.5	26	10.0	3.5	12.25
J	36	4.5	29	4.5	0.0	0.00
K	35	7.0	31	3.0	4.0	16.00
L	31	16.5	33	2.0	14.5	210.25
M	36	4.5	24	12.0	–7.5	56.25
N	30	18.0	21	17.0	1.00	1.00
O	37	2.5	18	19.0	–16.5	272.25
P	37	2.5	27	8.5	–6.0	36.00
Q	33	10.0	22	15.5	–5.5	30.25
R	28	19.5	17	20.0	–0.5	0.25
S	32	13.5	28	6.5	7.0	49.00
T	33	10.0	23	13.5	–3.5	12.25
						$\sum D^2 = 803.00$

$N = 20$

$$\rho = 1.00 - \frac{6(\Sigma D^2)}{N(N^2 - 1)}$$

$$= 1.00 - \frac{6(803)}{20(400 - 1)}$$

$$= 1.00 - \frac{4{,}818}{7{,}980}$$

$$= 1.00 - .60$$

$$\rho = .40$$

Pearson Product-Moment Correlation Coefficient

The **Pearson product-moment correlation coefficient,** also called Pearson r, is used when measurement results are reported in interval or ratio scale scores. The Spearman correlation coefficient actually is a Pearson r computed on the ranks. The Pearson r gives a more precise estimate of relationship because the actual scores, rather than the ranks of the scores, are taken into account. (This popular correlation coefficient has many variations, but only one method will be described.) The symbol for the product-moment correlation coefficient is r.

Observed scores are used with this calculation method of the product-moment correlation coefficient. The steps for calculation are as follows:

1. Label columns for name, X, X^2, Y, Y^2, and XY.
2. Designate one set of scores as X, designate the other set as Y, and place the appropriate paired scores by the individual's name.
3. Find the sums of the X and Y columns (ΣX and ΣY).
4. Square each X score, place squared scores in the X^2 column, and find the sum of the column (ΣX^2).
5. Square each Y score, place squared scores in the Y^2 column, and find the sum of the column (ΣY^2).
6. Multiply each X score by the Y score, place the product in the XY column, and find the sum of the column (ΣXY).
7. Substitute the values in the formula

$$r = \frac{N(\Sigma XY) - (\Sigma X)(\Sigma Y)}{\sqrt{N(\Sigma X^2) - (\Sigma X)^2}\sqrt{N(\Sigma Y^2) - (\Sigma Y)^2}}$$

Table 3.2 illustrates the calculation of the Pearson product-moment correlation coefficient for isometric and isotonic strength scores.

Interpretation of the Correlation Coefficient

After calculating the correlation coefficient, you must interpret it. This interpretation should be done with caution. For example, a correlation of .70 may be considered quite high in one analysis but low in another. Also, in your interpretation of the coefficient, you must remember that a high correlation does not indicate that one variable causes something to happen to another variable. Correlation is about the *relationship* of variables. The purpose for which the correlation coefficient is computed must be considered when making a decision about how high or low a coefficient is. Keeping the purpose of the correlation study in mind, use the following ranges as general guidelines for interpretation of the correlation coefficient. Note that negative values are considered in the same manner.

r = below .20 (extremely low relationship)

r = .20 to .39 (low relationship)

r = .40 to .59 (moderate relationship)

r = .60 to .79 (high relationship)

r = .80 to 1.00 (very high relationship)

Significance of the Correlation Coefficient

Your interpretation of a correlation coefficient should not be limited to the general interpretation described earlier. The **statistical significance,** or reliability, of the correlation coefficient should be considered also. In determining the coefficient significance, you are answering the following question: If the study were repeated, what is the probability of obtaining a similar relationship? You want to be sure that the relationship is real and did not result from a chance occurrence.

When r is calculated, the number of pairs of scores is important. With a small number of paired scores, it is possible that a high r value can occur by chance. For this reason, when a small number of paired scores is recorded, the r value must be large to be significant. On the other hand, if the number of scores is large, it is less likely that a high r value will occur by chance. Thus, a small r can be significant with a large number of paired scores.

A table of values is used to determine the statistical significance of a correlation coefficient. (In this discussion, .05 and .01 levels of significance and degrees of freedom are used. These terms will be explained in greater detail in the discussion of t-tests in chapter 4.) Before using the values table, you must first calculate the degrees of freedom (df) for the paired scores. The degrees of freedom equal $N - 2$ (N is the number of paired scores). In the calculation of r for isometric and isotonic strength scores in table 3.2, the degrees of freedom equal $N - 2 = 15 - 2 = 13$. To determine the statistical significance of $r = .90$, we will use the values in appendix A. Find the number 13 in the degrees of freedom column in appendix A. You see that a correlation of .514 is required for significance at the .05 level and .641 at the .01 level. These numbers indicate that, for 13 degrees of freedom, a correlation as high as .514 occurs only 5 in 100 times by chance, and a correlation as high as .641 occurs only 1 in 100 times by chance. Because the r value of .90 in table 3.2 is greater than both

TABLE 3.2 Product-Moment Correlation Coefficient for Isometric and Isotonic Strength Scores

Name	Isometric Score X	X^2	Isotonic Score Y	Y^2	XY
A	63	3,969	76	5,776	4,788
B	78	6,084	80	6,400	6,240
C	46	2,116	70	4,900	3,220
D	103	10,609	102	10,404	10,506
E	74	5,476	73	5,329	5,402
F	82	6,724	87	7,569	7,134
G	95	9,025	92	8,464	8,740
H	103	10,609	93	8,649	9,579
I	87	7,569	83	6,889	7,221
J	73	5,329	79	6,241	5,767
K	78	6,084	82	6,724	6,396
L	89	7,921	90	8,100	8,010
M	82	6,724	85	7,225	6,970
N	73	5,329	81	6,561	5,913
O	92	8,464	85	7,225	7,820
	1,218	102,032	1,258	106,456	103,706

N = 15

$$r = \frac{N(\Sigma XY) - (\Sigma X)(\Sigma Y)}{\sqrt{N(\Sigma X^2) - (\Sigma X)^2}\sqrt{N(\Sigma Y^2) - (\Sigma Y)^2}}$$

$$= \frac{15(103{,}706) - (1{,}218)(1{,}258)}{\sqrt{15(102{,}032) - (1{,}218)^2}\sqrt{15(106{,}456) - (1{,}258)^2}}$$

$$= \frac{1{,}555{,}590 - 1{,}532{,}244}{\sqrt{1{,}530{,}480 - 1{,}483{,}524}\sqrt{1{,}596{,}840 - 1{,}582{,}564}}$$

$$= \frac{23{,}346}{\sqrt{46{,}956}\sqrt{14{,}276}}$$

$$= \frac{23{,}346}{(216.69)(119.48)}$$

$$= \frac{23{,}346}{25{,}890.12}$$

$$r = .90$$

of these values, we conclude that the correlation is significant at the .01 level.

In summary, the obtained r is compared with the appropriate table values. If the obtained r is larger than the values found at the .05 and .01 level, r is significant at the .01 level. If the obtained r falls between these two table values, r is significant at the .05 level. If the obtained r is smaller than both table values, it is not significant. Significant correlation coefficients lower than .50 can be useful for indicating nonchance relationships among variables, but they probably are not large enough to be useful in predicting individual scores.

You should note that the correlation needed for significance decreases with the increased number of paired scores. Remember the earlier statement that a large number of scores decreases the likelihood that a high r will occur by chance. The table values reflect the influence of chance. You also should note that a higher correlation is needed for significance at the .01 level than at the .05 level. The reason is that we are reducing the odds that the correlation is due to chance (1 in 100 vs. 5 in 100).

Coefficient of Determination

The statistical significance of correlation is important, but to better determine the relationship of two variables, the **coefficient of determination** should be utilized. The coefficient of determination is the square of the correlation coefficient (r^2). It represents the common variance between two variables, or the proportion of variance in one variable that can be accounted for by the other variable. For example, imagine we administered a standing long jump test and a leg strength test to a group of male high school students, and we calculated a correlation coefficient of .85 between the two tests. We calculate that r^2 equals .72. We interpret this value to mean that 72% of the variability in the standing long jump scores is associated with leg strength. Or, we might say that 72% of both standing long jump ability and leg strength comes from common factors. We also can say that the two tests have common factors that influence the individuals' scores.

Using the coefficient of determination shows that a high correlation coefficient is needed to indicate a substantial to high relationship between two variables. In addition, coefficients of determination can be compared as ratios, whereas correlation coefficients cannot. For example, an r of .80 is not twice as large as an r of .40. By using the coefficient of determination, we see that an r of .80 is four times stronger than an r of .40 ($r = .80$, $r^2 = .64$; $r = .40$, $r^2 = .16$).

Negative Correlation Coefficients

There are occasions when a negative correlation coefficient is to be expected. When a smaller score that is considered to be a better score is correlated with a larger score that also is considered to be a better score, the correlation coefficient usually will be negative. The relationship between maximum oxygen consumption and the time required to run 2 miles is an example. In general, individuals with high values for maximum oxygen consumption will have lower times for the 2-mile run than individuals with low values for maximum oxygen consumption. In addition, a negative correlation will probably occur with performance that requires support of the body (weight correlated with pull-ups or dips).

Correlation, Regression, and Prediction

Recall that linear correlation refers to the relationship of two variables. It tells us how close the relationship between two variables is to a straight line. Once a relationship between two variables is found, if we have the score for one variable we can predict the score for the other variable. Consider the example provided in table 3.2. A correlation of .90 was found between isometric strength and isotonic strength. Through linear regression analysis, it is possible to predict an individual's isotonic score once the isometric score is known. Also consider the question asked at the beginning of the chapter about a possible relationship between physical fitness and academic grades. If a high relationship between these two variables were found, it would be possible to predict grades based on physical fitness scores. Rarely, however, will a predicted score be the actual score for an individual. Only if the correlation is -1 or $+1$ can we be confident that the predicted score will be the actual score. In correlation studies, we can determine the standard error of estimate (a numerical value that indicates the amount of error to expect in a predicted score) and how much confidence (referred to as *confidence limits*) we have in the prediction. To determine our confidence limits, we use the standard error of estimate in the same way as the standard deviation is used with a group of scores. Again, consider the isometric and isotonic scores in table 3.2. With linear regression analysis, an isotonic score may be predicted based upon an individual's isometric score. The standard error of estimate is then calculated to determine how much confidence we have in our predicted score. The standard error of estimate is added to and subtracted from the predicted score to determine the confidence limits. The confidence limits are:

68.26% of the time the predicted score is located ± 1 standard error of estimate from the predicted score

95.44% of the time the predicted score is located ± 2 standard error of estimates from the predicted score

99.73% of the time the predicted score is located ±3 standard error of estimates from the predicted score

Correction studies report these values.

Often correlation studies involve more than two variables. These studies, referred to as *multiple correlation-regression studies,* enable investigators to predict a score using several other scores. For example, many universities admit students based on their predicted freshman year grade point average. Values or scores for such variables as class rank, high school grade point average, and SAT or ACT score are used in a multiple regression equation to predict the freshman year grade point average. Also, many health-related or lifestyle studies involve multiple correlation-regression analysis. Such studies predict the health problems or diseases that may occur as a result of factors involved in a person's lifestyle.

Although linear and multiple correlation-regression techniques will not be presented in this text, you should be aware of their use and purpose. Your professional literature will include these studies.

ARE YOU ABLE TO DO THE FOLLOWING?

- Define *correlation, linear correlation, correlation coefficient*, and *coefficient of determination*.
- Interpret the correlation coefficient.
- Construct a scattergram and interpret it.
- Determine the Spearman rank-difference correlation coefficient and interpret it (statistical significance and coefficient of determination).
- Determine the Pearson product-moment correlation coefficient and interpret it (statistical significance and coefficient of determination).

Review Problems

1. Using the rank-difference correlation coefficient method, determine the relationship between the height (inches) and weight (pounds) measurements of twenty individuals. After you have determined r, calculate the coefficient of determination, interpret the obtained value, and use appendix A to interpret the significance of r.

Student	Height	Weight
A	68	170
B	68	160
C	66	140
D	67	150
E	69	155
F	65	145
G	70	168
H	71	182
I	74	208
J	65	144
K	68	165
L	67	160
M	71	195
N	71	190
O	70	195
P	66	150
Q	69	170
R	66	148
S	65	140
T	76	210

2. Using the product-moment correlation coefficient method, determine the relationship between a 2-minute sit-up test and physical fitness test scores. After you have determined r, calculate the coefficient of determination, interpret the obtained value, and use appendix A to interpret the significance of the correlation coefficient.

Student	Sit-Up Score	PF Score
A	45	81
B	51	94
C	30	91
D	55	75
E	32	65
F	42	82
G	41	85
H	52	65
I	61	73
J	39	85
K	38	93