
EIGHTH EDITION



RESEARCH IN EDUCATION

John W. Best

Butler University, Emeritus

James V. Kahn

University of Illinois at Chicago

Allyn and Bacon

Boston • London • Toronto • Sydney • Tokyo • Singapore

LB
1028
.B4
1998

Vice President, Education: Nancy Forsyth
Editorial Assistant: Cheryl Ouellette
Marketing Manager: Kris Farnsworth
Sr. Editorial Production Administrator: Susan McIntyre
Editorial Production Service: Ruttle, Shaw & Wetherill, Inc.
Composition Buyer: Linda Cox
Manufacturing Buyer: Suzanne Lareau
Cover Administrator: Suzanne Harbison



Copyright ©1998, 1993, 1989, 1986, 1981, 1977, 1970, 1959 by Allyn & Bacon
A Viacom Company
160 Gould Street
Needham Heights, MA 02194

Internet: www.abacon.com
America Online: keyword: College Online

All rights reserved. No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without written permission from the copyright owner.

Library of Congress Cataloging-in-Publication Data

Best, John W.

Research in education / John W. Best, James V. Kahn-8th ed
p. cm.

Includes bibliographical references and index

ISBN 0-205-18697-1

1. Education-Research. I. Kahn, James V., 1948- II. Title.

LB1028.B4 1998

370'.7'2—dc21

96.53399

CIP

Printed in the United States of America

1 0 9 8 7 6 5 4 3 2 1 RRD 04 03 02 01 00 99 98 97



CONTENTS

Preface xiii

PART I *Introduction to Educational Research: Definitions, Research Problems, Proposals, and Report Writing* 1

1	The Meaning of Research	3
	<i>The Search for Knowledge</i>	3
	<i>Science</i>	6
	<i>The Role of Theory</i>	9
	<i>Operational Definitions of Variables</i>	9
	<i>The Hypothesis</i>	10
	<i>The Research Hypothesis</i>	11
	<i>The Null Hypothesis (H_0)</i>	11
	<i>Sampling</i>	72
	<i>Randomness</i>	13
	<i>The Simple Random Sample</i>	13
	<i>Random Numbers</i>	13
	<i>The Systematic Sample</i>	15
	<i>The Stratified Random Sample</i>	15
	<i>The Area or Cluster Sample</i>	16
	<i>Nonprobability Samples</i>	16
	<i>Sample Size</i>	17
	<i>What Is Research?</i>	18
	<i>Purposes of Research</i>	20
	<i>Fundamental or Basic Research</i>	20
	<i>Applied Research</i>	21
	<i>Action Research</i>	21

	<i>Assessment, Evaluation, and Descriptive Research</i>	22
	<i>Types of Educational Research</i>	23
	<i>Summary</i>	24
	<i>Exercises</i>	25
	<i>References</i>	26
2	Selecting a Problem and Preparing a Research Proposal	29
	<i>The Academic Research Problem</i>	30
	Levels of Research Projects	31
	Sources of Problems	31
	Evaluating the Problem	34
	<i>The Research Proposal</i>	36
	<i>Ethics in Human Experimentation</i>	40
	<i>Using the Library</i>	45
	Finding Related Literature	45
	Microfiche	46
	<i>Note Taking</i>	46
	<i>References and Bibliography</i>	48
	Fair Use of Copyrighted Materials	48
	<i>The First Research Project</i>	48
	<i>Submitting a Research Proposal to a Funding Agency</i>	51
	<i>Summary</i>	52
	<i>Exercises</i>	53
	<i>References</i>	54
3	The Research Report	55
	Style Manuals	55
	<i>Format of the Research Report</i>	56
	Main Body of the Report	57
	References and Appendices	60
	<i>The Thesis or Dissertation</i>	60
	<i>Style of Writing</i>	61
	<i>Reference Form</i>	62
	<i>Pagination</i>	64
	<i>Tables</i>	64
	<i>Figures</i>	66
	The Line Graph	67
	The Bar Graph or Chart	67
	The Circle, Pie, or Sector Chart	68
	Maps	70
	Organization Charts	70
	<i>Evaluating a Research Report</i>	70
	<i>Summary</i>	72
	<i>References</i>	72

PART II **Research Methods 73**

4 **Historical Research 77**

<i>The History of American Education</i>	78
<i>History and Science</i>	81
Historical Generalization	82
<i>The Historical Hypothesis</i>	83
Hypotheses in Educational Historical Research	84
Difficulties Encountered in Historical Research	85
<i>Sources of Data</i>	85
Primary Sources of Data	85
Primary Sources of Educational Data	86
Secondary Sources of Data	87
<i>Historical Criticism</i>	87
External Criticism	87
Internal Criticism	88
Examples of Topics for Educational Historical Study	91
<i>Writing the Historical Report</i>	92
<i>Summary</i>	93
<i>Exercises</i>	94
<i>Endnote</i>	94
<i>References</i>	94
<i>Sample Article</i>	96
French Colonial Policy and the Education of Women and Minorities: Louisiana in the Early Eighteenth Century / Clark Robenstine	96

5 **Descriptive Studies: Assessment, Evaluation, and Research 113**

<i>Assessment Studies</i>	115
The Survey	115
Social Surveys	116
Public Opinion Surveys	117
National Center for Education Statistics	119
International Assessment	120
Activity Analysis	121
Trend Studies	121
<i>Evaluation Studies</i>	122
School Surveys	122
Program Evaluation	123
<i>Assessment and Evaluation in Problem Solving</i>	125
The Follow-Up Study	127
Descriptive Research	129
Replication and Secondary Analysis	135
The Post Hoc Fallacy	137
<i>Summary</i>	139
<i>Exercises</i>	139

	<i>References</i>	140
	<i>Sample Article</i>	143
	Perceptions About Special Olympics from Service Delivery Groups in the United States: A Preliminary Investigation / David L. Porretta, Michael Gillespie, Paul Jansma	143
6	Experimental and Quasi-Experimental Research	157
	<i>Early Experimentation</i>	758
	<i>Experimental and Control Groups</i>	159
	<i>Variables</i>	160
	Independent and Dependent Variables	160
	Confounding Variables	161
	<i>Controlling Extraneous Variables</i>	162
	<i>Experimental Validity</i>	764
	Threats to Internal Experimental Validity	164
	Threats to External Experimental Validity	168
	<i>Experimental Design</i>	169
	Pre-Experimental Designs	170
	True Experimental Designs	171
	Quasi-Experimental Designs	175
	Factorial Designs	184
	<i>Summary</i>	188
	<i>Exercises</i>	789
	<i>References</i>	190
	<i>Sample Article</i>	192
	Experiential Versus Experience-Based Learning and Instruction / James D. Laney	192
7	Single-Subject Experimental Research	209
	<i>General Procedures</i>	211
	Repeated Measurement	211
	Baselines	211
	Manipulating Variables	212
	Length of Phases	213
	Transfer of Training and Response Maintenance	214
	<i>Assessment</i>	214
	Target Behavior	215
	Data Collection Strategies	215
	Basic Designs	216
	A-B-A Designs	216
	Multiple Baseline Designs	220
	Other Designs	221
	<i>Evaluating Data</i>	223
	<i>Summary</i>	225
	<i>Exercises</i>	226
	<i>Endnotes</i>	227

References	227
Sample Article	228
Effects of Response Cards on Student Participation and Academic Achievement: A Systematic Replication with Inner-City Students During Whole-Class Science Instruction / Ralph Gardner, William L. Heward, Teresa A. Grossi	228
8 Qualitative Research	239
Themes of Qualitative Research	240
Research Questions	242
Theoretical Traditions	244
Research Strategies	246
Document or Content Analysis	246
The Case Study	248
Ethnographic Studies	250
Data Collection Techniques	253
Observations	253
Interviews	254
Review of Documents	255
Other Qualitative Data Collection Techniques	255
Data Analysis and Interpretation	257
Summary	259
Exercises	259
Endnotes	260
References	260
Sample Article	262
Professionals' Perceptions of HIV in Early Childhood Developmental Center / Norma A. Lopez-Reyna, Rhea F. Boldman, James V. Kahn	262
9 Methods and Tools of Research	275
Reliability and Validity of Research Tools	275
Quantitative Studies	276
Qualitative Studies	278
Psychological and Educational Tests and Inventories	279
Qualities of a Good Test and Inventory	281
Validity	281
Reliability	283
Economy	285
Interest	285
Types of Tests and Inventories	286
Achievement Tests	287
Aptitude Tests	287
Interest Inventories	289
Personality Inventories	289
Projective Devices	290

<i>Observation</i>	291
Validity and Reliability of Observation	294
Recording Observations	295
Systematizing Data Collection	295
Characteristics of Good Observation	298
<i>Inquiry Forms: The Questionnaire</i>	298
The Closed Form	299
The Open Form	300
Improving Questionnaire Items	300
Characteristics of a Good Questionnaire	307
Preparing and Administering the Questionnaire	308
A Sample Questionnaire	310
Validity and Reliability of Questionnaires	310
<i>Inquiry Forms: The Opinionnaire</i>	314
Thurstone Technique	315
Likert Method	315
Semantic Differential	319
<i>The Interview</i>	320
Validity and Reliability of the Interview	321
<i>Q Methodology</i>	322
<i>Social Scaling</i>	324
Sociometry	324
Scoring Sociometric Choices	325
The Sociogram	325
“Guess-Who” Technique	326
Social-Distance Scale	327
<i>Organization of Data Collection</i>	328
Outside Criteria for Comparison	329
<i>Limitations and Sources of Error</i>	330
<i>Summary</i>	331
<i>Exercises</i>	332
<i>References</i>	332

PART III *Data Analysis* 335

10 Descriptive Data Analysis	337
<i>What Is Statistics!</i>	338
<i>Parametric and Nonparametric Data</i>	338
<i>Descriptive and Inferential Analysis</i>	339
<i>The Organization of Data</i>	340
Grouped Data Distributions	341
<i>Statistical Measures</i>	342
Measures of Central Tendency	342
Measures of Spread or Dispersion	347

<i>Normal Distribution</i>	352
Nonnormal Distributions	355
Interpreting the Normal Probability Distribution	355
Practical Applications of the Normal Curve	357
Measures of Relative Position: Standard Scores	357
The T score (T)	359
The College Board Score (Z_{cb})	360
Stanines	360
Percentile Rank	360
<i>Measures of Relationship</i>	362
Pearson's Product-Moment Coefficient of Correlation (ρ)	366
Rank Order Correlation (ρ)	369
Phi Correlation Coefficient (ϕ)	371
<i>Interpretation of a Correlation Coefficient</i>	372
Misinterpretation of the Coefficient of Correlation	373
Prediction	374
<i>Standard Error of Estimate</i>	376
<i>A Note of Caution</i>	378
<i>Summary</i>	379
<i>Exercises (Answers in Appendix I)</i>	380
<i>Endnote</i>	384
<i>References</i>	384
11 Inferential Data Analysis	385
<i>Statistical Inference</i>	385
<i>The Central Limit Theorem</i>	386
<i>Parametric Tests</i>	389
<i>Testing Statistical Significance</i>	389
<i>The Significance of the Difference between the Means of Two Independent Groups</i>	389
<i>The Null Hypothesis (H_0)</i>	390
<i>The Level of Significance</i>	391
<i>Decision Making</i>	392
<i>Two-Tailed and One-Tailed Tests of Significance</i>	393
<i>Degrees of Freedom</i>	395
<i>Student's Distribution (t)</i>	396
<i>Significance of the Difference between Two Small Sample Independent Means</i>	396
<i>Homogeneity of Variances</i>	397
<i>Significance of the Difference between the Means of Two Matched or Correlated Groups (Nonindependent Samples)</i>	400
<i>Statistical Significance of a Coefficient of Correlation</i>	402
<i>Analysis of Variance (ANOVA)</i>	404
<i>Analysis of Covariance (ANCOVA) and Partial Correlation</i>	409
<i>Multiple Regression and Correlation</i>	411

<i>Nonparametric Tests</i>	415
The Chi Square Test (χ^2)	415
The Mann-Whitney Test	420
Summary	422
Exercises (Answers in Appendix I)	423
References	426
12 Computer Data Analysis	427
The Computer	427
Data Organization	429
Computer Analysis of Data	432
Example 1: Descriptive Statistics—SAS:CORR	433
Example 2: Charting-SAS:CHART	433
Example 3: Multiple Regression-SPSS	436
Example 4: Analysis of Variance-SASS-PC+	440
SPSS for Windows Used with Appendix B Data in Chapters 10 and 11 Examples	443
Summary	447
Endnotes	447
Reference	447
Appendix A Statistical Formulas and Symbols	449
Appendix B Sample Data Microsoft Excel Format	455
Appendix C Percentage of Area Lying Between the Mean and Successive Standard Deviation Units under the Normal Curve	459
Appendix D Critical Values for Pearson's Product-Moment Correlation (r)	461
Appendix E Critical Values of Student's Distribution (t)	463
Appendix F Abridged Table of Critical Values for Chi Square	465
Appendix G Critical Values of the F Distribution	467
Appendix H Research Report Evaluation	473
Appendix I Answers to Statistics Exercises	475
Appendix J Selected Indexes, Abstracts, and Reference Materials	479
Author Index	491
Subject Index	494



PREFACE

The eighth edition of *Research in Education* has the same goals as the previous editions. The book is meant to be used as a research reference or as a text in an introductory course in research methods. It is appropriate for graduate students enrolled in a research seminar, for those writing a thesis or dissertation, or for those who carry on research as a professional activity. All professional workers should be familiar with the methods of research and the analysis of data. If only as consumers, professionals should understand some of the techniques used in identifying problems, forming hypotheses, constructing and using data-gathering instruments, designing research studies, and employing statistical procedures to analyze data. They should also be able to use this information to interpret and critically analyze research reports that appear in professional journals and other publications.

No introductory course can be expected to confer research competence, nor can any book present all relevant information. Research skill and understanding are achieved only through the combination of coursework and experience. Graduate students may find it profitable to carry on a small-scale study as a way of learning about research.

This edition expands and clarifies a number of ideas presented in previous editions. Additional concepts, procedures, and especially examples have been added. Each of the five methodology chapters has the text of an entire published article following it, which illustrates that type of research. Nothing has been deleted from the seventh edition other than a few examples of research that have been replaced with more recent and appropriate examples. An appendix (B) has been added that contains a data set for use by students in Chapters 10, 11, and 12. This edition has been written to conform to the guidelines of the American Psychological Association's (*APA Publications Manual* (4th ed.)). The writing style suggested in Chapter 3 is also in keeping with the *APA manual*.

Many of the topics covered in this book may be peripheral to the course objectives of some instructors. We do not suggest that all of the topics in this book be included in a single course. We recommend that instructors use the topics selectively and in the sequence that they find most appropriate. Students can then use the portions remaining in subsequent courses, to assist in carrying out a thesis, and/or as a reference.

This revision benefited from the comments of Professor Kahn's students, who had used the earlier editions of this text. To them and to reviewers Barbara Boe, Carthage College; John A. Jensen, Boston College; Jerry McGee, Sam Houston State; and Gene Gloekner, Colorado State University, we express our appreciation. We also wish to thank Michelle Chapman and Tara O'Brien who assisted in the preparation of this edition. We wish to acknowledge the cooperation of the University of Illinois at Chicago Library and Computer Center; SPSS, Inc.; and SAS Institute, Inc. Finally, we are grateful to our wives, Solveig Ager Best and Kathleen Cuerdon-Kahn, for their encouragement and support.

J.W.B.

J.V.K.

10

DESCRIPTIVE DATA ANALYSIS

Because this textbook concentrates on educational research methods, the following discussion of statistical analysis is in no sense complete or exhaustive. Only some of the most simple and basic concepts are presented. Students whose mathematical experience includes high school algebra should be able to understand the logic and the computational processes involved and should be able to follow the examples without difficulty.

The purpose of this discussion is threefold:

1. To help the student, as a consumer, develop an understanding of statistical terminology and the concepts necessary to read with understanding some of the professional literature in educational research.
2. To help the student develop enough competence and know-how to carry on research studies using simple types of analysis.
3. To prepare the student for more advanced coursework in statistics.

The emphasis is on intuitive understanding and practical application rather than on the derivation of mathematical formulas. Those who expect and need to develop real competence in educational research will have to take some of the following steps:

1. Take one or more courses in behavioral statistics and experimental design.
2. Study more specialized textbooks in statistics, particularly those dealing with statistical inference (e.g., Glass & Hopkins, 1996; Hays, 1981; Heiman, 1996; Kerlinger, 1986; Kirk, 1995; Siegel, 1956; Shavelson, 1996; Winer, 1971).
3. Read research studies in professional journals extensively and critically.
4. Carry on research studies involving some serious use of statistical procedures.

WHAT IS STATISTICS?

Statistics is a body of mathematical techniques or processes for gathering, organizing, analyzing, and interpreting numerical data. Because most research yields such quantitative data, statistics is a basic tool of measurement, evaluation, and research.

The word statistics is sometimes used to describe the numerical data gathered. Statistical data describe group behavior or group characteristics abstracted from a number of individual observations that are combined to make generalizations possible.

Everyone is familiar with such expressions as “the average family income,” “the typical white-collar worker,” or “the representative city.” These are statistical concepts and, as group characteristics, may be expressed in measurement of age, size, or any other traits that can be described quantitatively. When one says that “the average fifth-grade boy is 10 years old,” one is generalizing about all fifth-grade boys, not any particular boy. Thus, the statistical measurement is an abstraction that may be used in place of a great mass of individual measures.

The research worker who uses statistics is concerned with more than the manipulation of data. The statistical method serves the fundamental purposes of description and analysis, and its proper application involves answering the following questions:

1. What facts need to be gathered to provide the information necessary to answer the question or to test the hypothesis?
2. How are these data to be selected, gathered, organized, and analyzed?
3. What assumptions underlie the statistical methodology to be employed?
4. What conclusions can be validly drawn from the analysis of the data?

Research consists of systematic observation and description of the characteristics or properties of objects or events for the purpose of discovering relationships between variables. The ultimate purpose is to develop generalizations that may be used to explain phenomena and to predict future occurrences. To conduct research, one must establish principles so that the observation and description have a commonly understood meaning. Measurement is the most precise and universally accepted process of description, assigning quantitative values to the properties of objects and events.

PARAMETRIC AND NONPARAMETRIC DATA

In the application of statistical treatments, two types of data are recognized:

1. *Parametric data.* Data of this type are measured data, and parametric statistical tests assume that the data are normally, or nearly normally, distributed. Parametric tests are applied to both interval- and ratio-scaled data.
2. *Nonparametric data.* Data of this type are either counted (nominal) or ranked (ordinal). Nonparametric tests, sometimes known as distribution-free tests, do not rest on the more stringent assumption of normally distributed populations.

TABLE 10.1 Levels of Quantitative Description¹

Level	Scale	Process	Data Treatment	Some Appropriate Tests
4	Ratio	measured equal intervals true zero ratio relationship	parametric	<i>t</i> test analysis of variance analysis of covariance factor analysis Pearson's <i>r</i>
3	Interval	measured equal intervals no true zero		
2	Ordinal	ranked in order	nonparametric	Spearman's ρ (ρ) Mann-Whitney Wilcoxon
1	Nominal	classified and counted		

¹Refer to Chapter 9 for a discussion of the four levels of measurements.

Table 10.1 presents a graphic summary of the levels of quantitative description and the types of statistical analysis appropriate for each level. These concepts will be developed later in the discussion.

However, one should be aware that many of the parametric statistics (*t* test, analysis of variance, and Pearson's *r* in particular) are still appropriate even when the assumption of normality is violated. This robustness has been demonstrated for the *t* test, analysis of variance, and, to a lesser extent, analysis of covariance by a number of researchers including Glass, Peckham, and Sanders (1972), Lunney (1970), and Mandeville (1972). Thus, with ordinal data and even with dichotomous data (two choices such as pass-fail), these statistical procedures, which were designed for use with interval and ratio data, may be appropriate and useful. Pearson's *r*, which can also be used with any type of data, will be discussed later in this chapter.

DESCRIPTIVE AND INFERENTIAL ANALYSIS

Until now we have not discussed the limits to which statistical analysis may be generalized. Two types of statistical application are relevant:

Descriptive Analysis

Descriptive statistical analysis limits generalization to the particular group of individuals observed. No conclusions are extended beyond this group, and any

similarity to those outside the group cannot be assumed. The data describe one group and that group only. Much simple action research involves descriptive analysis and provides valuable information about the nature of a particular group of individuals. Assessment studies (see Chapter 5) also often rely solely or heavily on descriptive statistics.

Inferential Analysis

Inferential statistical analysis always involves the process of sampling and the selection of a small group assumed to be related to the population from which it is drawn. The small group is known as the sample, and the large group is the population. Drawing conclusions about populations based on observations of samples is the purpose of inferential analysis.

A *statistic* is a measure based on observations of the characteristics of a sample. A statistic computed from a sample may be used to estimate a *parameter*, the corresponding value in the population from which the sample is selected. Statistics are usually represented by letters of our Roman alphabet such as X , S , and r . Parameters, on the other hand, are usually represented by letters of the Greek alphabet such as μ , σ , or ρ .

Before any assumptions can be made, it is essential that the individuals selected be chosen in such a way that the small group, or sample, approximates the larger group, or population. Within a margin of error, which is always present, and by the use of appropriate statistical techniques, this approximation can be assumed, making possible the estimation of population characteristics by an analysis of the characteristics of the sample.

It should be emphasized that when data are derived from a group without careful sampling procedures, the researcher should carefully state that findings apply only to the group observed and may not apply to or describe other individuals or groups. The statistical theory of sampling is complex and involves the estimation of error of inferred measurements, error that is inherent in estimating the relationship between a random sample and the population from which it is drawn. Inferential data analysis is presented in Chapter 11.

THE ORGANIZATION OF DATA

The list of test scores in a teacher's grade book provides an example of unorganized data. Because the usual method of listing is alphabetical, the scores are difficult to interpret without some other type of organization.

Alberts, James	60
Brown, John	78
Davis, Mary	90
Smith, Helen	70
Williams. Paul	88

TABLE 10.2 Scores of 37 Students on a Semester Algebra Test

98	85	80	76	67
97	85	80	76	67
95	85	80	75	64
93	84	80	73	60
90	82	78	72	57
88	82	78	70	
87	82	78	70	
87	80	77	70	

$$\text{Range} = 98 - 57 + 1 = 41 + 1 = 42$$

The Ordered Array or Set

Arranging the same scores in descending order of magnitude produces what is known as an *ordered array*.

90
88
78
70
60

The ordered array provides a more convenient arrangement. The highest score (90), the lowest score (60), and the middle score (78) are easily identified. Thus, the range (the difference between the highest and lowest scores, plus one) can easily be determined.

Illustrated in Table 10.2 is a data arrangement of 37 students' scores on an algebra test in ordered array form.

Grouped Data Distributions

Data are often more clearly presented when scores are grouped and a frequency column is included. Data can be presented in frequency tables (see Table 10.3 on page 342) with different class intervals, depending on the number and range of the scores.

A score interval with an odd number of units may be preferable because its midpoint is a whole number rather than a fraction. Because all scores are assumed to fall at the midpoint of the interval (for purposes of computing the mean), the computation is less complicated:

Even interval of four: 8 9 10 11 (midpoint 9.5)

Odd interval of five: 8 9 10 11 12 (midpoint 10)

There is no rule that rigidly determines the proper score interval, and intervals of 10 are frequently used.

TABLE 10.3 Scores on Algebra Test Grouped in Intervals of Five

Score	Interval	Tallies	Frequency (f)	Includes
96-100		11	2	(96 97 98 99 100)
91-95		11	2	(91 92 93 94 95)
86-90		1111	4	etc.
81-85		1111 11	7	
76-80		1111 1111 1	11	
71-75		111	3	
66-70		1111	5	
61-65		1	1	
56-60		11	2	
N = 37				

STATISTICAL MEASURES

Several basic types of statistical measures are appropriate in describing and analyzing data in a meaningful way:

Measures **of** *central tendency or averages*

Mean

Median

Mode

Measures **of** *spread or dispersion*

Range

Variance

Standard deviation

Measures **Of** *relative position*

Standard scores

Percentile rank

Percentile score

Measures **of** *relationship*

Coefficient of correlation

Measures of *Central Tendency*

Nonstatisticians use *averages* to describe the characteristics of groups in a general way. The climate of an area is often noted by average temperature or average amount of rainfall. We may describe students by grade-point averages or by average age. Socioeconomic status of groups is indicated by average income, and the return on

an investment portfolio may be judged in terms of average income *return*. But to the statistician the term average is unsatisfactory, for there are a number of types of averages, only one of which may be appropriate to use in describing given characteristics of a group. Of the many averages that may be used, three have been selected as most useful in educational research: the mean, the median, and the mode.

The Mean (\bar{X})

The mean of a distribution is commonly understood as the arithmetic average. The term *grade-point average*, familiar to students, is a mean value. It is computed by dividing the sum of all the scores by the number of scores. In formula form

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

where \bar{X} = mean

Σ = sum of

X = scores in a distribution

N = number of scores

Example
X
6
5
4
3
2
1
<hr/>
$\Sigma X = 21$
$N = 6$
$\bar{X} = 21/6 = 3.50$

The mean is probably the most useful of all statistical measures, for, in addition to the information that it provides, it is the base from which many other important measures are computed.

Appendix B contains a data set from a population of 100 children (one set in Microsoft Excel and one in SPSS format). The data for each child includes an ID number, the method of teaching reading that was received, the gender, the category of special education in which the child has been classified (LD = learning disabilities; BD = behavior disorder; MR = mild mental retardation), and both pre- and posttest scores. The reader may wish to randomly select a sample of 25 children (or 15 children if recommended by the professor) from the appendix for use in a variety of calculations throughout this chapter. Now calculate the mean for this sample of 25 children's IQ. The mean of the population given in the appendix is 86.12. How does the sample mean compare to the population mean?

The Median (*M_d*)

The median is a point (not necessarily a score) in an array, above and below which one-half of the scores fall. It is a measure of position rather than of magnitude and is frequently found by inspection rather than by calculation. When there are an odd number of untied scores, the median is the middle score, as in the example below:

7	
6	3 scores above
5	
4	— median
3	
2	3 scores below
1	

When there are an even number of untied scores, the median is the midpoint between the two middle scores, as in the example below:

6	
5	3 scores above
4	
-median = 3.50	
3	
2	3 scores below
1	

If the data include tied scores at the median point, interpolation within the tied scores is necessary. Each integer would represent the interval from halfway between it and the next lower score to halfway between it and the next higher score. When ties occur at the midpoint of a set of scores, this interval is portioned out into the number of tied scores and the midpoint or median is found. Consider the set of scores in Figure 10.1.

Because there are four scores tied (75), the interval from 74.5 to 75.5 is divided into four equal parts. Each of the scores is then considered to occupy 0.25 of the interval, and the median is calculated.

One purpose of the mean and the median is to represent the “typical” score; most of the time it is satisfactory to use the mean for this purpose. However, when the distribution of scores is such that most scores are at one end and relatively few are at the other (known as a *skewed distribution*), the median is preferable because it is not influenced by extreme scores at either end of the distribution. In the following examples the medians are identical. However, the mean of Group A is 4,



Group A	Group B
7	50
6	6
5	5
4—Md	4—Md
3	3
2	2
1	0

In a small school with five faculty members, the salaries might be

Total Salaries = \$120,000

The average salary of the group is represented with a different emphasis by the median salary (\$21,400) than by the mean salary (\$24,000), which is substantially

higher than that of four of the five faculty members. Thus, we see again that the median is less sensitive than the mean to extreme values at either end of a distribution.

Using the same 25 children selected from Appendix B to calculate a mean, now calculate the median. How do the two compare? Which is more useful? The median for the population of 100 children is 89.0 (5 scores of 89 fall below the midpoint and 5 above it). How does the sample median compare?

The Mode (Mo)

6	
5	
4	
4	Mode
3	I
2	
1	

The mode is the score that occurs most frequently in a distribution. It is located by inspection rather than by computation. In grouped data distributions the mode is assumed to be the *midscore* of the interval in which the greatest frequency occurs.

For example, if the modal age of fifth-grade children is 10 years, it follows that there are more 10-year-old fifth-graders than any other age. Or a menswear salesman might verify the fact that there are more sales of size 40 suits than of any other size; consequently, a larger number of size 40 suits are ordered and stocked, size 40 being the mode.

In some distributions there may be more than one mode. A two-mode distribution is referred to as bimodal, more than two, multimodal. If the number of auto accidents on the streets of a city were tabulated by hours of occurrence, it is likely that two modal periods would become apparent—between 7 A.M. and 8 A.M. and between 5 P.M. and 6 P.M., the hours when traffic to and from stores and offices is heaviest and when drivers are in the greatest hurry. In a normal distribution of data there is one mode, and it falls at the midpoint, just as the mean and median do. In some unusual distributions, however, the mode may fall at some other point. When the mode or modes reveal such unusual behavior, they do not serve as measures of central tendency, but they do reveal useful information about the nature of the distribution.

Using the data set in Appendix B, the mode of the categories of disability can be determined. Because 50 of the 100 children have learning disabilities (28 have behavior disorders and 22 have mental retardation) as their classification, this is the mode. Now using the data from the 25 children selected for the mean and median calculations above, determine the mode of the sample for disability category. Now determine the mode for IQ of the sample. The mode for the population is 89. How does the sample mode compare?

Measures of Spread or Dispersion

Measures of central tendency describe location along an ordered scale. There are characteristics of data distributions calling for additional types of statistical analysis. The scores in Table 10.4 were made by a group of students on two different tests, one in reading and one in arithmetic.

The mean and the median are identical for both tests. It is apparent that averages do not fully describe the differences in achievement between students' scores on the two tests. To contrast their performance, it is necessary to use a measure of score spread or dispersion. The arithmetic test scores are **homogeneous, with little difference** between adjacent scores. The reading test scores are decidedly heterogeneous, with performances ranging from superior to very poor.

The Range

The range, the simplest measure of dispersion, is the difference between the highest and lowest scores plus one. For reading scores the range is 41 ($95 - 55 + 1$). For arithmetic scores the range is 9 ($79 - 71 + 1$).

The Deviation from the Mean (x)

A score expressed as its distance from the mean is called a deviation *score*. Its formula is

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

TABLE 10.4 Sample Data

Reading			Arithmetic	
Pupil	Score	Academic Grade	Score	Academic Grade
Arthur	95	A	76	C
Betty	90	A	78	C
John	85	B	77	C
Katherine		B	71	C
Charles	80	C	75	C
Larry	70	C	79	C
Donna	65	D	73	C
Edward	60	D	72	C
Mary	55	F	74	C
$\Sigma X = 675$			$\Sigma X = 675$	
$\bar{N} = 9$			$\bar{N} = 9$	
$\bar{X} = \frac{675}{9} = 75$			$\bar{X} = \frac{675}{9} = 75$	
$Md = 75$			$Md = 75$	

If the score falls above the mean, the deviation score is positive (+); if it falls below the mean, the deviation score is negative (-).

Using the same example, compare two sets of scores:

Reading		Arithmetic	
X	$(X - \bar{X})$	X	$(X - \bar{X})$
95	+20	76	+1
90	+15	78	+3
85	+10	77	+2
80	+ 5	71	- 4
75	0	75	0
70	- 5	79	+4
65	-10	73	- 2
60	-15	72	- 3
55	-20	74	-1
$\Sigma X = 675$	$\Sigma x = 0$	$\Sigma X = 675$	$\Sigma x = 0$
$N = 9$		$N = 9$	
$\bar{X} = 75$		$\bar{X} = 75$	

It is interesting to note that the sum of the score deviations from the mean equals zero.

$$\begin{aligned}\Sigma (X - \bar{X}) &= 0 \\ \Sigma x &= 0\end{aligned}$$

In fact, we can give an alternative definition of the mean: The mean is that value in a distribution around which the sum of the deviation score equals zero.

The Variance (σ^2)

The sum of the squared deviations from the mean, divided by N , is known as the variance. We have noted that the sum of the deviations from the mean equals zero ($\Sigma x = 0$). From a mathematical point of view it would be impossible to find a mean value to describe these deviations (unless the signs were ignored). Squaring each deviation score yields a positive score. The scores can then be summed, divided by N , and the mean of the squared deviations computed. The variance formula is

$$\sigma^2 = \frac{\Sigma(X - \bar{X})^2}{N} \quad \text{or} \quad \frac{\Sigma x^2}{N}$$

Thus, the variance is a value that describes how all of the scores in a distribution are dispersed or spread about the mean. This value is very useful in describing the characteristics of a distribution and will be employed in a number of very important statistical tests. However, because all of the deviations from the mean have been squared to find the variance, it is much too large to represent the spread of scores.

The Standard Deviation (σ)

The standard deviation, the square root of the variance, is most frequently used as a measure of spread or dispersion of scores in a distribution. The formula for standard deviation of a population is

$$\sigma = \sqrt{\frac{\Sigma(X - \bar{X})^2}{N}} \quad \text{or} \quad \sqrt{\frac{\Sigma x^2}{N}}$$

In the following example, using the reading scores from Table 10.4, the variance and the standard deviation are computed.

X	x	x^2
95	+20	1400
90	+15	+225
85	+10	+100
80	+ 5	+25
75	0	0
70	- 5	+25
65	-10	+100
60	-15	+225
55	-20	+400
		$\Sigma x^2 = 1500$

$$\text{Variance } \sigma^2 = 1500/9 = 166.67$$

$$\text{Standard deviation } \sigma = \sqrt{1500/9} = \sqrt{166.67} = 12.91$$

As can clearly be seen, a variance of 166.67 cannot represent, for most purposes, a spread of scores with a total range of only 41, but the standard deviation of 12.91 does make sense.

Although the deviation approach (just used in the previous calculation) provides a clear example of the meaning of variance and standard deviation, in actual practice the deviation method can be awkward to use in computing the variances or standard deviations for a large number of scores. A less complicated method, which results in the same answer, uses the raw scores instead of the deviation scores. The number values tend to be large, but the use of a calculator facilitates the computation.

$$\text{Variance } \sigma^2 = \frac{N \Sigma X^2 - (\Sigma X)^2}{N^2}$$

$$\text{Standard deviation } \sigma = \sqrt{\frac{N \Sigma X^2 - (\Sigma X)^2}{N^2}}$$

The following example demonstrates the process of computation, using the raw score method:

X	X^2
95	9025
90	8100
85	7225
80	6400
75	5625
70	4900
65	4225
60	3600
55	3025
$\Sigma X = 675$	$\Sigma X^2 = 52,125$
$N = 9$	
$\sigma^2 = \frac{9(52,125) - (675)^2}{9(9)} = \frac{469,125 - 455,625}{81}$	
$\sigma^2 = \frac{13,500}{81} = 166.67$	
$\sigma = 166.67 = 12.91$	

Standard Deviation for Samples (S)

The variance and standard deviation for a population have just been described. Because most of the time researchers use samples selected from the population, it is necessary to introduce the formulas for the variance S^2 and the standard deviation (S) of a sample. The sample formulas differ only slightly from the population formulas. As will be seen, instead of dividing by N in the deviation formula and by N^2 in the raw score formula, the sample formulas divide by $n - 1$ and $n(n - 1)$, respectively. This is done to correct for the probability that the smaller the sample the less likely it is that extreme scores will be included. Thus the formula for σ , if used with a sample, would underestimate the standard deviation of the population because a randomly selected sample would probably not include the most extreme scores that exist in the population simply because there are so few of them. Dividing by $n - 1$ or $n(n - 1)$ corrects for this bias, more or less depending upon the sample's size. This makes the standard deviation of the sample more representative of the population. In a small sample, say $n = 5$, the correction is rather large, dividing by 4 instead of 5—a reduction of 20% in the denominator. In a large sample, say $n = 100$, the correction is insignificant, dividing by 99 instead of 100—a reduction of 1% in the denominator. Again, this difference in the percent correction is due to the fact that smaller the sample the less likely are extreme scores to be represented.

We should note that these formulas for the standard deviation of the sample are actually inferential statistics and would normally be in the next chapter. However, because these are the formulas used to describe a sample and because sam-

ples are what one normally has to calculate the standard deviation, we believe this is the better place for them.

The two formulas for sample standard deviation with the deviation and the raw score methods of computation, respectively, are

$$S = \sqrt{\frac{(X - \bar{X})^2}{n - 1}} \quad \text{or} \quad \sqrt{\frac{x^2}{n - 1}} \quad \text{and} \quad S = \sqrt{\frac{N \Sigma X^2 - (X)^2}{n(n - 1)}}$$

No doubt the reader can see that the only changes are in the denominator. Thus, if we substitute $n(n - 1)$ for N^2 and calculate S^2 and S using the data from page 350, we would find the following:

$$\begin{aligned} S^2 &= \frac{9(52,125) - (675)^2}{9(8)} = \frac{469,125 - 455,625}{72} \\ &= \frac{13,500}{72} = 187.50 \\ S &= \sqrt{187.50} = 13.69 \end{aligned}$$

These results are quite a change from $\sigma^2 = 166.67$ (change of +20.83) and $\sigma = 12.91$ (change of +.78). These relatively large differences from the population formula to the sample formula are due to the small sample size ($n = 9$), which made a relatively large correction necessary. The correction for calculating the variance and standard deviation is important because, unless the loss of a degree of freedom (discussed in Chapter 11) is considered, the calculated sample variance or standard deviation is likely to underestimate the population variance or standard deviation. This is true because the mean of the squared deviations from the mean of any distribution is the smallest possible value and probably would be smaller than the mean of the squared deviation from any other point in the distribution. Because the mean of the sample is not likely to be identical to the population mean (because of sampling error), the use of $N - 1$ (the number of degrees of freedom) rather than N in the denominator tends to correct for this underestimation of the population variance or standard deviation.

The strength of a prediction or the accuracy of an inferred value increases as the number of independent observations (sample size) is increased. Because large samples may be biased, sample size is not the only important determinant, but if unbiased samples are selected randomly from a population, large samples will provide a more accurate basis than will smaller samples for inferring population values.

The standard deviation for IQ of the population in Appendix B is 11.55, using the formula for the population (it would be 11.61 if the sample formula were used). The reader should calculate the standard deviation (using the formula for a sample) for the sample. How does it compare with the standard deviation of this population?

The standard deviation is a very useful device for comparing characteristics that may be quite different or may be expressed in different units of measurement.

The following discussion shows that when the normality of distributions can be assumed it is possible to compare the proverbial apples and oranges. The standard deviation is independent of the magnitude of the mean and provides a common unit of measurement. To use a rather farfetched example, imagine a man whose height is one standard deviation below the mean and whose weight is one standard deviation above the mean. Because we assume that there is a normal relationship between height and weight (or that both characteristics are normally distributed), a picture emerges of a short, overweight individual. His height, expressed in inches, is in the lowest 16% of the population, and his weight, expressed in pounds, is in the highest 16%. In this chapter only the standard deviation of a population is discussed.

But before using the standard deviation to describe status or position in a group is discussed, the normal distribution needs to be examined.

NORMAL DISTRIBUTION

The earliest mathematical analysis of the theory of probability dates to the 18th century. Abraham DeMoivre, a French mathematician, discovered that a mathematical relationship explained the probabilities associated with various games of chance. He developed the equation and the graphic pattern that describes it. During the 19th century a French astronomer, Laplace, and a German mathematician, Gauss, independently arrived at the same principle and applied it more broadly to areas of measurement in the physical sciences. From the limited applications made by these early mathematicians and astronomers, the theory of probability, or the curve of distribution of error, has been applied to data gathered in the areas of biology, psychology, sociology, and other sciences. The theory describes the fluctuations of chance errors of observation and measurement. It is necessary to understand the theory of probability and the nature of the curve of normal distribution to comprehend many important statistical concepts, particularly in the area of standard scores, the theory of sampling, and inferential statistics.

We should keep in mind that “the normal distribution does not actually exist. It is not a fact of nature. Rather, it is a mathematical model—an idealization—that can be used to represent data collected in behavioral research” (Shavelson, 1996, p. 120). The law of probability and the normal curve that illustrates it are based on the law of chance or the probable occurrence of certain events. When any body of observations conforms to this mathematical form, it can be represented by a bell-shaped curve with definite characteristics (see Figure 10.2).

1. The curve is symmetrical around its vertical axis—50% of the scores are above the mean and 50% below the mean.
2. The mean, median, and the mode of the distribution have the same value.
3. The terms cluster around the center—most scores are near the mean, median, and mode with fewer scores as the score is further from the center.
4. The curve has no boundaries in either direction, for the curve never touches the base line, no matter how far it is extended. The curve is a curve of probability, not of certainty.

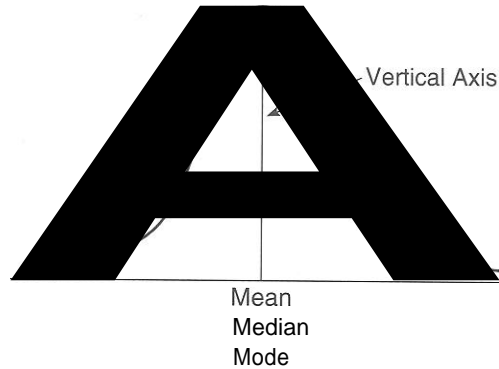


FIGURE 10.2 The Normal Curve

5. One way to think of the normal curve (or the nonnormal curves described shortly) is to view it “as a solid geometric figure made up of all the subjects and their different scores” (Heiman, 1996, p. 53). That is, the curve is a smoothed, curved version of a bar graph that represents each possible score and the number of persons who got that score.

Researchers often consider one standard deviation from the mean to be a particularly important point on the normal curve. This is for both a practical and a mathematical reason. The practical reason is that this results in approximately 68% (slightly over two-thirds) of the population falling between one standard deviation above and one standard deviation below the mean. Perhaps more important, this is the point at which the curve changes from a downward convex shape to an upward convex shape. Thus, mathematically, this is the point at which the direction of the curve changes. As will be discussed later, ± 1.96 standard deviations from the mean will result in 95% of the population. This is another critical point in the curve, which is often rounded to 2 standard deviations from the mean.

The operation of chance prevails in the tossing of coins or dice. It is believed that many human characteristics respond to the influence of chance. For example, if certain limits of age, race, and gender were kept constant, such measures as height, weight, intelligence, and longevity would approximate the normal distribution pattern. But the normal distribution does not appear in data based on observations of samples. There just are not enough observations. The normal distribution is based on an infinite number of observations beyond the capability of any observer; thus, there is usually some observed deviation from the symmetrical pattern. But for purposes of statistical analysis, it is assumed that many characteristics do conform to this mathematical form within certain limits, providing a convenient reference.

The concept of measured intelligence is based on the assumption that intelligence is normally distributed throughout limited segments of the population. Tests are so constructed (standardized) that scores are normally distributed in the large group that is used for the determination of norms or standards. Insurance companies determine their premium rates by the application of the curve of probability

Basing their expectation on observations of past experience, they can estimate the probabilities of survival of a man from age 45 to 46. They do not purport to predict the survival of a particular individual, but from a large group they can predict the mortality rate of all insured risks.

The total area under the normal curve may be considered to approach 100% probability. Interpreted in terms of standard deviations, areas between the mean and the various standard deviations from the mean under the curve show these percentage relationships (see Figure 10.3).

Note the graphic conformation of the characteristics of the normal curve:

1. It is symmetrical-the percentage of frequencies is the same for equal intervals below or above the mean.
2. The terms or scores “cluster” or “crowd around the mean”—note how the percentages in a given standard deviation are greatest around the mean and decrease as one moves away from the mean.

\bar{X} to $\pm 1.00z$	34.13%
± 1.00 to $\pm 2.00z$	13.59%
± 2.00 to $\pm 3.00z$	2.15%

3. The curve is highest at the mean-the mean, median, and mode have the same value.
4. The curve has no boundaries-a small fraction of 1% of the space falls outside of ± 3.00 standard deviations from the mean.

The normal curve is a curve that also describes probabilities. For example, if height is normally distributed for a given segment of the population, the chances are

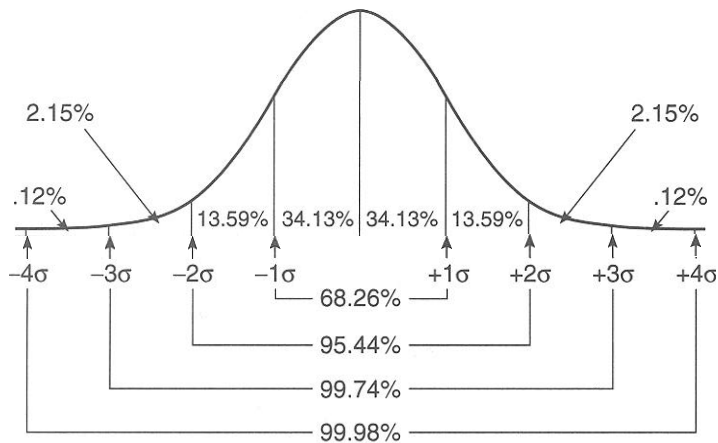


FIGURE 10.3 Percentage of Frequencies in a Normal Distribution Falling within a Range of a Given Number of Standard Deviations from the Mean

$\frac{34.13}{100}$ that a person selected at random will be between the mean and one standard deviation above the mean in height, and $\frac{34.13}{100}$ that the person selected will be between the mean and one standard deviation below the mean in height—or $\frac{68.26}{100}$ that the selected person will be within one standard deviation (above or below) the mean in height. Another interpretation is that 68.26% of this population segment will be between the mean and one standard deviation above or below the mean in height.

An example may help the reader understand this concept. IQ (intelligence quotient) is assumed to be normally distributed. The Wechsler Intelligence Scale for Children-Revised (WISC-R) has a mean of 100 and a standard deviation of 15. Thus, a WISC-R IQ score that is one standard deviation above the mean is 115, and a score of 85 is one standard deviation below the mean. From this information it is known that approximately 68% of the population should have WISC-R scores between 85 and 115.

For practical purposes the curve is usually extended to ± 3 standard deviations from the mean ($\pm 3z$). Most events or occurrences (or probabilities) will fall between these limits. The probability is $\frac{99.74}{100}$ that these limits account for observed or predicted occurrences. This statement does not suggest that events or measures could not fall more than three standard deviations from the mean but that the likelihood would be too small to consider when making predictions or estimates based on probability. Statisticians deal with probabilities, not certainty, and there is always a degree of reservation in making any prediction. Statisticians deal with the probabilities that cover the normal course of events, not the events that are outside the normal range of experience.

Nonnormal Distributions

As mentioned earlier in the discussions of parametric and nonparametric data and the relative usefulness of the mean and median, not all distributions, particularly of sample data, are identical to or even close to a normal curve. There are two other types of distributions that can occur: *skewed* and *bimodal*. In skewed distributions the majority of scores are near the high or low end of the range with relatively few scores at the other end. The distribution is considered skewed in the direction of the tail (fewest scores). In Figure 10.4 on page 356 distribution A is skewed positively, and distribution B is skewed negatively. Skewed distributions can be caused by a number of factors, including a test that is too easy or hard or an atypical sample (very bright or very low intelligence).

Bimodal distributions have two modes (see distribution C in Figure 10.4) rather than the single mode of normal or skewed distributions. This often results from a sample that consists of persons from two populations. For instance, the height of American adults would be bimodally distributed, females clustering around a mode of about 5 feet 4 inches and males around a mode of about 5 feet 10 inches.

Interpreting the Normal Probability Distribution

When scores are normally or near normally distributed, a normal probability table is useful. The values presented in the normal probability table in Appendix B are

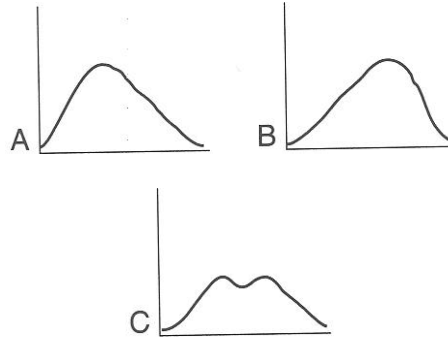


FIGURE 10.4 Nonnormal Distributions

critical because they provide data for normal distributions that may be interpreted in the following ways:

1. The percentage of total space included between the mean and a given standard deviation (z) distance from the mean
2. The percentage of cases, or the number when N is known, that fall between the mean and a given standard deviation (z) distance from the mean
3. The probability that an event will occur between the mean and a given standard deviation (z) distance from the mean

z = number of standard deviations from the mean

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

Figure 10.5 demonstrates how the area under the normal curve can be divided. In a normal distribution the following characteristics hold true:

1. The space included between the mean and $+1.00z$ is .3413 of the total area under the curve.
2. The percentage of cases that fall between the mean and $+1.00z$ is .3413.
3. The probability of an event's occurring (observation) between the mean and $+1.00z$ is .3413.
4. The distribution is divided into two equal parts, one half above the mean and the other half below the mean.
5. Because one half of the curve is above the mean and .3413 of the total area is between the mean and $+1.00z$, the area of the curve that is above $+1.00z$ is .1587.

Because the normal probability curve is symmetrical, the shape of the right side (above the mean) is identical to the shape of the left side (below the mean). Because the values for each side of the curve are identical, only one set of values is presented in the probability table, expressed to one-hundredth of a sigma (standard deviation) unit.

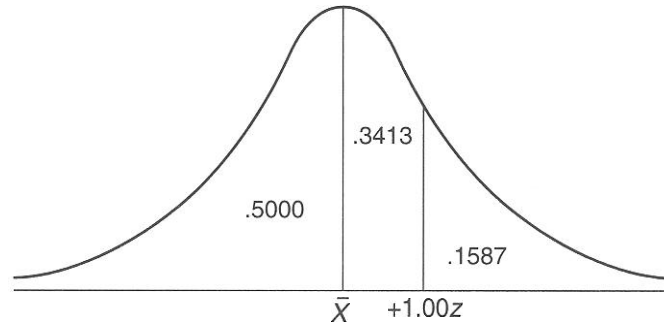


FIGURE 10.5 The Space Included Under the Normal Curve Between the Mean and $\pm 1.00z$

The normal probability table in Appendix C provides the proportion of the curve that is between the mean and a given sigma (z) value. The remainder of that half of the curve is beyond the sigma value.

		Probability
Above the mean	.5000	50/100
Below the mean	.5000	50/100
Above $+1.96z$	$.5000 - .4750 = .0250$	2.5/100
Below $+1.32z$	$.5000 + .1255 = .6255$	62.5/100
Below $-1.32z$	$.5000 - .1255 = .3745$	37.5/100

Practical Applications of the Normal Curve

In the field of educational research the normal curve has a number of practical applications:

1. To calculate the percentile rank of scores in a normal distribution.
2. To normalize a frequency distribution, an important process in standardizing a psychological test or inventory.
3. To test the significance of observed measures in experiments, relating them to the chance fluctuations or errors inherent in the process of sampling and generalizing about populations from which the samples are drawn.

Measures of Relative Position: Standard Scores

Standard scores provide a method of expressing any score in a distribution in terms of its distance from the mean in standard deviation units. The utility of this conversion of a raw score to a standard score will become clear as each type is introduced and illustrated. Three types of standard scores are considered.

1. Z score (Sigma)
2. T score (T)
3. College board score (Z_{cb})

Remember that the distribution is assumed to be normal when using any type of standard score.

The Z Score (Sigma)

In describing a score in a distribution, its deviation from the mean-expressed in standard deviation units-is often more meaningful than the score itself. The unit of measurement is the standard deviation.

$$z = \frac{X - \bar{X}}{\sigma} \quad \text{OR} \quad \frac{x}{\sigma}$$

where X = raw score

\bar{X} = mean

σ = standard deviation

$x = (X - \bar{X})$ score deviation from the mean

Example A	Example B
$X = 76$	$X = 67$
$\bar{X} = 82$	$\bar{X} = 62$
$\sigma = 4$	$\sigma = 5$
$z = \frac{76 - 82}{4} = \frac{-6}{4} = -1.50$	$z = \frac{67 - 62}{5} = \frac{5}{5} = +1.00$

The raw score of 76 in Example A may be expressed as a Z score of -1.50 , indicating that 76 is 1.5 standard deviations below the mean. The score of 67 in Example B may be expressed as a sigma score of $+1.00$, indicating that 67 is one standard deviation above the mean.

In comparing or averaging scores on distributions where total points may differ, the researcher using raw scores may create a false impression of a basis for comparison. A Z score makes possible a realistic comparison of scores and may provide a basis for equal weighting of the scores. On the sigma scale the mean of any distribution is converted to zero, and the standard deviation is equal to 1.

For example, a teacher wishes to determine a student's equally weighted average (mean) achievement on an algebra test and on an English test.

Subject	Test Score	Mean	Highest Possible Score	Standard Deviation
Algebra	40	47	60	5
English	84	110	180	20

It is apparent that the mean of the two raw test scores would not provide a valid summary of the student's performance, for the mean would be weighted overwhelmingly in favor of the English test score. The conversion of each test score to a sigma score makes them equally weighted and comparable, for both test scores have been expressed on a scale with a mean of zero and a standard deviation of one.

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

$$\text{Algebra } z \text{ score} = \frac{40.5 - 47}{5} = \frac{-7}{5} = -1.40$$

$$\text{English } z \text{ score} = \frac{84 - 110}{20} = \frac{-26}{20} = -1.30$$

On an equally weighted basis, the performance of the student was fairly consistent: 1.40 standard deviations below the mean in algebra and 1.30 standard deviations below the mean in English.

Because the normal probability table describes the percentage of area lying between the mean and successive deviation units under the normal curve (see Appendix C), the use of sigma scores has many other useful applications to hypothesis testing, determination of percentile ranks, and probability judgments.

The reader may wish to select one score from the sample of 25 children selected earlier and calculate the z score for that person in relation to the sample. The population mean (86.12) and standard deviation (11.55) in the formula could then be used to calculate the z for the same child. How do these two z scores compare?

The T Score (T)

$$T = 50 + 10 \frac{(X - \bar{X})}{\sigma} \text{ or } 50 + 10z$$

Although the z score is most frequently used, it is sometimes awkward to have negatives or scores with decimals. Therefore, another version of a standard score, the T score, has been devised to avoid some confusion resulting from negative z scores (below the mean) and also to eliminate decimal values.

Multiplying the z score by 10 and adding 50 results in a scale of positive whole number values. Using the scores in the previous example, $T = 50 + 10z$:

$$\text{Algebra } T = 50 + 10(-1.40) = 50 + (-14) = 36$$

$$\text{English } T = 50 + 10(-1.30) = 50 + (-13) = 37$$

T scores are always rounded to the nearest whole number. A z score of +1.27 would be converted to a T score of 63.

$$T = 50 + 10(+1.27) = 50 + (+12.70) = 62.70 = 63$$

Convert the z scores just calculated for the person selected from the sample into T scores.

The College Board Score (Z_{cb})

The College Entrance Examination Board and several other testing agencies use another conversion that provides a more precise measure by spreading out the scale (see Figure 10.6).

$$Z_{cb} = 500 + 100 \frac{(X - \bar{X})}{\sigma} = 500 + 100z$$

The mean of this scale is 500.

The standard deviation is 100.

The range is 200–800.

Stanines

A stanine is a standard score that divides the normal curve into nine parts, thus the term *stanine* from *sta* of *standard* and nine. The 2nd to 8th stanines are each equal to one-half standard deviation unit. Thus, stanine 5 includes the center of the curve and goes one-quarter (.25) standard deviations above and below the mean. Stanine 6 goes from the top of stanine 5 to .75 standard deviations above the mean, whereas stanine 4 goes from the bottom of stanine 5 to .75 standard deviations below the mean and so on. Stanine 1 encompasses all scores below stanine 2, and stanine 9 encompasses all scores above stanine 8. Figure 10.6 demonstrates the stanine distribution and compares it to the other standard scores.

Percentile Rank

Although the percentile rank is not usually considered a standard score, it is pertinent to this discussion. It is often useful to describe a score in relation to other scores; the percentile rank is the point in the distribution below which a given percentage of scores fall. If the 80th percentile rank is a score of 65, 80% of the scores fall below 65. The median is the 50th percentile rank, for 50% of the scores fall below it.

When N is small, the definition needs an added refinement. To be completely accurate, the percentile rank is the score in the distribution below which a given percentage of the scores falls, plus one half the percentage of space occupied by the given score.

SCOWS

50

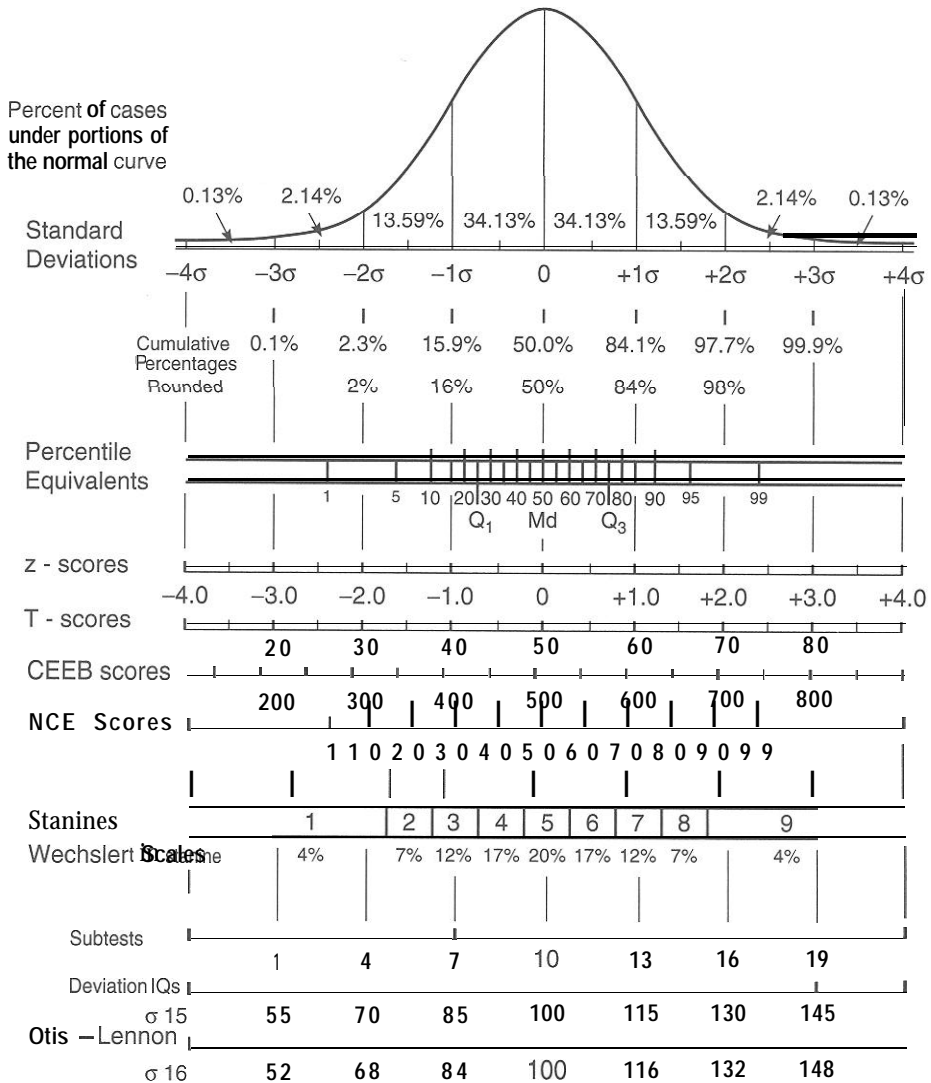
47

43

39

30

On inspection it is apparent that 43 is the median, or occupies the 50th percentile rank. Fifty % of the scores should fall below it, but in fact only two out of



The Normal Curve, Percentiles, and Selected Standard Scores

FIGURE 10.6 Illustration of Various Standard Score Scales

(Test Service Notebook 148, The Psychological Corporation, NY.)

five scores fall below 43. That would indicate 43 has a percentile rank of 40. But by adding the phrase “plus one half the percentage of space occupied by the score,” the calculation is reconciled:

40% of scores fall below 43; each score occupies 20% of the total space
 40% + 10% = 50 (true percentile rank)

When N is large, this qualification is unimportant because percentile ranks are rounded to the nearest whole number, ranging from the highest percentile rank of 99 to the lowest of zero.

High schools frequently rate their graduating seniors in terms of rank in class. Because schools vary so much in size, colleges find these rankings of limited value unless they are converted to some common basis for comparison. The percentile rank provides this basis by converting class rank into a percentile rank.

$$\text{Percentile rank} = 100 - \frac{(100RK - 50)}{N}$$

where RK = rank from the top

Jones ranks 27th in his senior class of 139 students. Twenty-six students rank above him, 112 below him. His percentile rank is

$$100 - \frac{(2700 - 50)}{139} = 100 - 19 = 81$$

In this formula 50 is subtracted from $100RK$ to account for half the space occupied by the individual's score. What is the percentile rank of the person you selected in order to calculate z and T scores?

MEASURES OF RELATIONSHIP ---

Correlation

Correlation is the relationship between two or more paired variables or two or more sets of data. The degree of relationship is measured and represented by the coefficient of correlation. This coefficient may be identified by either the letter r , the Greek letter rho (ρ), or other symbols, depending on the data distributions and the way the coefficient has been calculated.

Students who have high intelligence quotients tend to receive high scores in mathematics tests, whereas those with low IQs tend to score low. When this type of relationship is obtained, the factors of measured intelligence and scores on mathematics tests are said to be positively correlated.

Sometimes variables are negatively correlated when a large amount of one variable is associated with a small amount of the other. As one increases, the other tends to decrease.

When the relationship between two sets of variables is a pure chance relationship, we say that there is no correlation.

These pairs of variables are usually positively correlated: As one increases, the other tends to increase.

- | | |
|--------------------------|----------------------|
| 1. Intelligence | Academic achievement |
| 2. Productivity per acre | Value of farm land |
| 3. Height | Shoe size |
| 4. Family income | Value of family home |

These variables are usually negatively correlated: As one increases, the other tends to decrease.

- | | |
|---------------------------|-------------------------------|
| 1. Academic achievement | Hours per week of TV watching |
| 2. Total corn production | Price per bushel |
| 3. Time spent in practice | Number of typing errors |
| 4. Age of an automobile | Trade-in value |

There are other traits that probably have no correlation

- | | |
|----------------|----------------|
| 1. Body weight | Intelligence |
| 2. Shoe size | Monthly salary |

The degree of linear correlation can be represented quantitatively by the coefficient of correlation. A perfect positive correlation is $+1.00$. A perfect negative correlation is -1.00 . A complete lack of relationship is zero (0). Rarely, if ever, are perfect coefficients of correlations of $+1.00$ or -1.00 encountered, particularly in relating human traits. Although some relationships tend to appear fairly consistently, there are variations or exceptions that reduce the measured coefficient from either a -1.00 or a $+1.00$ toward zero.

A definition of perfect positive correlation specifies that for every unit increase in one variable there is a proportional unit increase in the other. The perfect negative correlation specifies that for every unit increase in one variable there is a proportional unit decrease in the other. That there can be no exceptions explains why coefficients of correlation of $+1.00$ or -1.00 are not encountered in relating human traits. The sign of the coefficient indicates the direction of the relationship, and the numerical value its strength.

The Scattergram and Linear Regression Line

When the relationship between two variables is plotted graphically, paired variable values are plotted against each other on the X and Y axis.

The line drawn through, or near, the coordinate points is known as the “line of best fit,” or the regression line. On this line the sum of the deviations of all the coordinate points has the smallest possible value. As the coefficient approaches zero (0), the coordinate points fall further from the regression line (see Figure 10.7 on page 364 for examples of different correlations’ scattergrams).

When the coefficient of correlation is either $+1.00$ or -1.00 , all of the coordinate points fall on the regression line, indicating that, when $r = +1.00$, for every increase in X there is a proportional increase in Y; and when $r = -1.00$, for every increase in X there is a proportional decrease in Y. There are no individual exceptions. If we know a person’s score on one measure, we can determine his or her exact score on the other measure.

The slope of the regression line, or line of best fit, is not determined by guess or estimation but by a geometric process that will be described later.

There are actually two regression lines. When $r = +1.00$ or -1.00 , the lines are superimposed and appear as one line. As r approaches zero, the lines separate further.

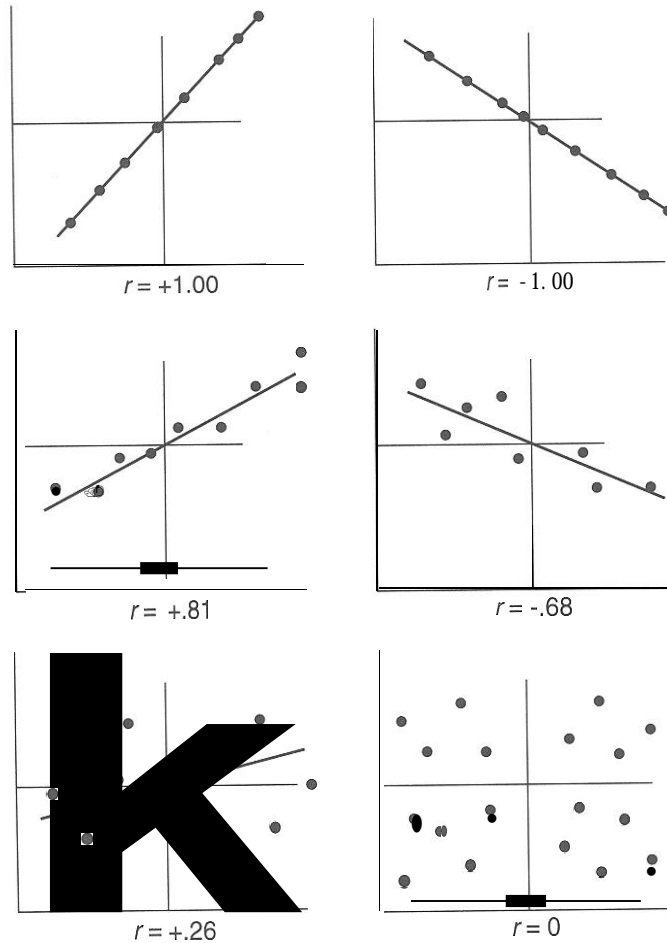


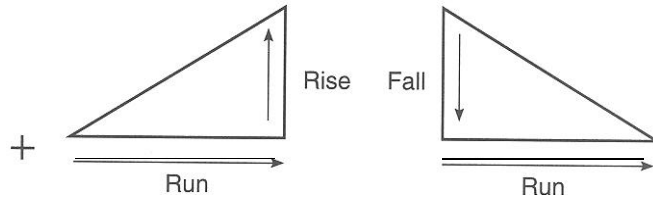
FIGURE 10.7 Scatter Diagrams Illustrating Different Coefficients of Correlation

Only one of the regression lines is described in this discussion, the Y on X (or Y from X) line. It is used to predict unknown Y values from known X values. The X values are known as the predictor variable, and the Y values, the predicted variable. The other regression line (not described here) would be used to predict X from Y.

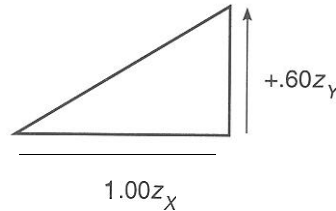
Plotting the Slope of the Regression Line

The slope of the regression (Y from X) line is a geometric representation of the coefficient of correlation and is expressed as a ratio of the magnitude of the *rise* (if r is +) to the run, or as a ratio of the *fall* (if r is -) to the run, expressed in standard

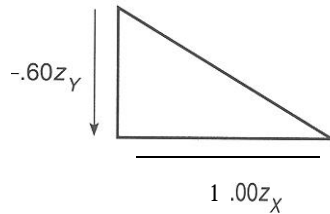
deviation units. The geometric relationship between the two legs of the right triangle determines the slope of the hypotenuse, or the regression line.



For example, if $r = +.60$, for every sigma unit increase (run) in X, there is a .60 sigma unit increase (rise) in Y.



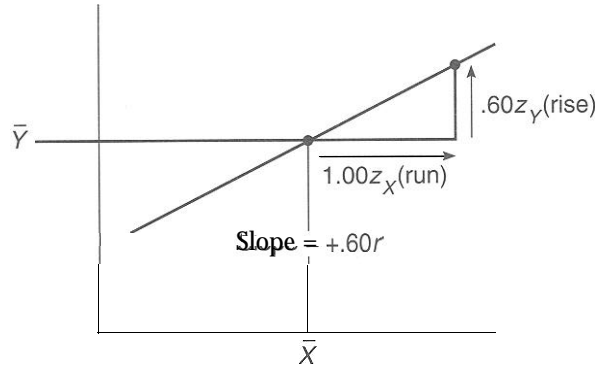
If $r = -.60$, for every sigma unit increase (run) in X, there is a .60 sigma unit decrease (fall) in Y.



Because all regression lines pass through the intersection of the mean of X and the mean of Y lines, only one other point is necessary to determine the slope. By measuring one standard deviation of the X distribution on the X axis and a .60 standard deviation of the Y distribution on the Y axis, the second point is established (see Figures 10.8 and 10.9 on page 366).

The regression line (r) involves one awkward feature: all values must be expressed in sigma scores (z) or standard deviation units. It would be more practical to use actual scores to determine the slope of the regression line. This can be done by converting to a slope known as b . The slope of the b regression line Y on X is determined by the formula

$$b = r \frac{\sigma_Y}{\sigma_X}$$

FIGURE 10.8 A Positive Regression Line, $r = +.60$

For example, if $r = +.60$

and

$$\sigma_y = 6$$

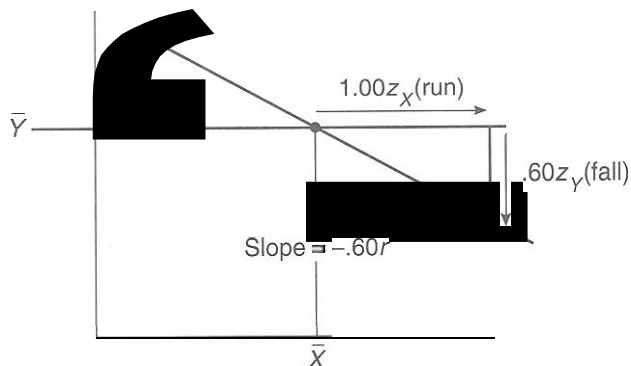
$$\sigma_x = 5$$

$$b = +.60 \frac{6}{5} = \frac{3.60}{5} = +.72$$

Thus an r of $+.60$ becomes $b = +.72$. Now the ratio run has another value and indicates a different slope line (Figure 10.10).

Pearson's Product-Moment Coefficient of Correlation (r)

The most often used and most precise coefficient of correlation is known as the Pearson's Product-Moment coefficient (r). This coefficient may be calculated by

FIGURE 10.9 A Negative Regression Line, $r = -.60$

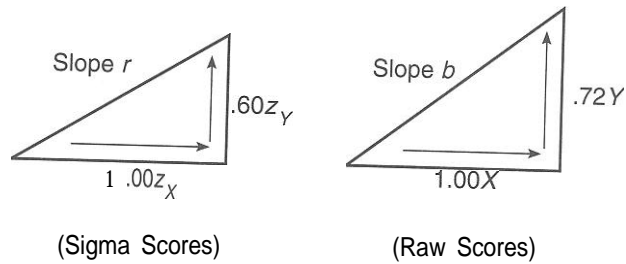


FIGURE 10.10 Two Regression Lines, r and b
 An r of $+.60$ is converted to a b of $+.72$ by the formula

$$b = r \frac{\sigma_Y}{\sigma_X}$$

converting the raw scores to sigma scores and finding the mean value of their cross-products.

$$r = \frac{\Sigma(z_x)(z_y)}{N}$$

z_x	z_y	$(z_x)(z_y)$
+1.50	+1.20	+1.80
+2.00	+1.04	+2.08
-.75	-.90	+.68
+.20	+.70	+.14
-1.00	+.20	-.20
-.40	+.30	-.12
+1.40	+.70	+.98
+.55	+.64	+.35
-.04	+.10	-.00
-.10	+.30	-.03
		$\Sigma(z_x)(z_y) = 5.68$
		$r = \frac{+5.68}{10} = +.568$

If most of the negative values of X are associated with negative z values of Y , and positive values of X with positive values of Y , the correlation coefficient will be positive. If most of the paired values are of opposite signs, the coefficient will be negative.

positive correlation $(+)(+) = +$ high on **X**, high on **Y**
 $(-)(-) = +$ low on **x**, low on **Y**

negative correlation $(+)(-) = -$ high on **X**, low on **Y**
 $(-)(+) = -$ low on **X**, high on **Y**

The *z* score method is not often used in actual computation because it involves the conversion of each score into a sigma score. Two other methods, a deviation method and a raw score method, are more convenient, more often used, and yield the same result.

The deviation method uses the following formula and requires the setting up of a table with seven columns.

$$r = \frac{\Sigma xy}{\sqrt{(\Sigma x^2)(\Sigma y^2)}}$$

where Σx^2 = the sum of the \bar{X} subtracted from each X score squared
 $(X - \bar{X})^2$

Σy^2 = the sum of the \bar{Y} subtracted from each Y score squared
 $(Y - \bar{Y})^2$

Σxy = the cross product of the mean subtracted from that score
 $(X - \bar{X})(Y - \bar{Y})$

Using the data from Table 10.4, with reading scores being the X variable and arithmetic scores being the Y variable, the researcher calculates *r* like this:

Variables						
X	Y	<i>x</i>	<i>x</i> ²	<i>y</i>	<i>y</i> ²	<i>xy</i>
95	76	20	400	1	1	+20
90	78	15	225	3	9	+45
85	77	10	100	2	4	+20
80	71	5	25	-4	16	-20
75	75	0	0	0	0	0
70	79	-5	25	4	16	-20
65	73	-10	100	-2	4	+20
60	72	-15	225	-3	9	+45
55	74	-20	400	-1	1	+20
$\Sigma X = 675$	$\Sigma Y = 675$		$\Sigma x^2 = 1500$		$\Sigma y^2 = 60$	$\Sigma xy = 130$
$\bar{X} = 75$	$\bar{Y} = 75$					

$$r = \frac{130}{\sqrt{(1500)(60)}} = \frac{130}{\sqrt{90,000}} = \frac{130}{300} = .433$$

The raw score method requires the use of five columns, as illustrated below using the same data.

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

where ΣX = sum of the X scores
 ΣY = sum of the Y scores
 ΣX^2 = sum of the squared X scores
 ΣY^2 = sum of the squared Y scores
 ΣXY = sum of the products of paired X and Y scores
 N = number of paired scores

Variables				
X	Y	X^2	Y^2	XY
95	76	9025	5776	7220
90	78	8100	6084	7020
85	77	7225	5929	6545
80	71	6400	5041	5680
75	75	5625	5625	5625
70	79	4900	6241	5530
65	73	4225	5329	4745
60	72	3600	5184	4320
55	74	3025	5476	4070
$2X=675$	$\Sigma Y = 675$	$\Sigma X^2 = 52,125$	$\Sigma Y^2 = 50,685$	$\Sigma XY = 50,755$

$$r = \frac{9(50,755) - (675)(675)}{\sqrt{9(52,125) - (675)^2} \sqrt{9(50,685) - (675)^2}}$$

$$= \frac{456,795 - 455,625}{\sqrt{469,125 - 455,625} \sqrt{456,165 - 455,625}}$$

$$= \frac{1170}{\sqrt{13,500} \sqrt{540}}$$

$$= \frac{1170}{(116.19)(23.24)}$$

$$= \frac{1170}{2700.26} = .433$$

Now take the 25 children selected earlier and calculate the correlation of IQ with pretest scores. The correlation for IQ with pretest scores for the entire population of 100 children is +.552. How does the sample's correlation relate to the correlation for the population? Now calculate the correlation of the pretest and posttest scores. The correlation for the population of 100 children between their pretest scores and their posttest scores is +.834. How does the sample's correlation relate to the correlation for the population?

Rank Order Correlation (ρ)

A particular form of the Pearson product-moment correlation that can be used with ordinal data is known as the *Spearman rank order coefficient* of correlation. The

symbol ρ (rho) is used to represent this correlation coefficient. The paired variables are expressed as ordinal values (ranked) rather than as interval or ratio values. The correction lends itself to an interesting graphic demonstration.

In the following example, the students ranking highest in IQ rank highest in mathematics, and those lowest in IQ, lowest in mathematics.

Pupil	IQ Rank	Achievement in Mathematics Rank
A	1	1
B	2	2
C	3	3
D	4	4
E	5	5

Perfect positive coefficient of correlation

$$\rho = +1.00$$

In the following example the students ranking highest in time spent in practice rank lowest in number of errors.

Pupil	Time Spent in Practice Rank	Number of Typing Errors Rank
A	1	5
B	2	4
C	3	3
D	4	2
E	5	1

Perfect negative coefficient of correlation

$$\rho = -1.00$$

In the following example, there is probably little more than a pure chance relationship (due to sampling error) between height and intelligence.

Pupil	Height Rank	IQ Rank
A	1	3
B	2	4
C	3	2
D	4	1
E	5	5

Very low coefficient of correlation

$$\rho = +.10$$

To compute the Spearman rank order coefficient of correlation, this rather simple formula is used:

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

where D = the difference between paired ranks

$\sum D^2$ = the sum of the squared differences between ranks

N = number of paired ranks

If the previously used data were converted to ranks and calculated Spearman's ρ , it would look like this:

Pupil	Rank in Reading	Rank in Arithmetic	D	D^2
	1	4	-3	9
Betty	2	2	0	0
John	3	3	0	0
Katherine	4	9	-5	25
Charles	5	5	0	0
Larry	6	1	5	25
Donna	7	7	0	0
Edward	8	8	0	0
Mary	9	6	3	9
				$\sum D^2 = 68$

$$\begin{aligned}\rho &= 1 - \frac{6(68)}{9(81 - 1)} = 1 - \frac{408}{9(80)} \\ &= 1 - \frac{408}{720} = 1 - .567 \\ &= +.433\end{aligned}$$

As has been just demonstrated, Spearman's ρ and Pearson's r yielded the same result. This occurs when there are no ties. When there are ties, the results will not be identical, but the difference will be insignificant.

The Spearman rank order coefficient of correlation computation is quick and easy. It is an acceptable method if data are available only in ordinal form. Teachers may find this computation method useful when conducting studies using a single class of students as subjects.

Phi Correlation Coefficient (ϕ)

The data are considered dichotomous when there are only two choices for scoring a variable (e.g., pass-fail or female-male). In these cases each person's score usually would be represented by a 0 or 1, although sometimes 1 and 2 are used instead.

The Pearson product-moment correlation, when both variables are dichotomous, is known as the phi (ϕ) coefficient. The formula for ϕ is simpler than for Pearson's r but algebraically identical. Because there are rarely two dichotomous variables of interest of which the researcher wants to know the relationship, the formula will not be presented here. This brief mention of ϕ is to make the reader aware of it. Those wishing more detail should refer to one of the many statistics texts available (e.g., Heiman, 1996; Glass & Hopkins, 1996).

INTERPRETATION OF A CORRELATION COEFFICIENT

Two circumstances can cause a higher or lower correlation than usual. First, when one person or relatively few people have a pair of scores differing markedly from the rest of the sample's scores, the resulting r may be spuriously high. When this occurs, the researcher needs to decide whether to remove this individual's pair of scores (known as an *outlier*) from the data analyzed. Second, when all other things are equal, the more homogeneous a group of scores, the lower their correlation will be. That is, the smaller the range of scores, the smaller r will be. Researchers need to consider this potential problem when selecting samples that may be highly homogeneous. However, if the researcher knows the standard deviation of the heterogeneous group from which the homogeneous group was selected, Glass and Hopkins (1984) and others describe a formula that corrects for the restricted range and provides the correlation for the heterogeneous group.

There are a number of ways to interpret a correlation coefficient or adjusted correlation coefficient, depending on the researcher's purpose and the circumstances that may influence the correlation's magnitude. One method that is frequently presented is to use a crude criterion for evaluating the magnitude of a correlation:

Coefficient (r)	Relationship
.00 to .20	Negligible
.20 to .40	LOW
.40 to .60	Moderate
.60 to .80	Substantial
.80 to 1.00	High to very high

Another interpretative approach is a test of statistical significance of the correlation, based on the concepts of sampling error and tests of significance described in Chapter 11.

Still another way of interpreting a correlation coefficient is in terms of variance. The variance of the measure that the researcher wants to predict can be divided into the part that is explained by, or due to, the predictor variable and the part that is explained by other factors (generally unknown) including sampling error. The researcher finds this percentage of explained variance by calculating r^2 ,

known as the *coefficient of determination*. The percentage of variance not explained by the predictor variable is then $1 - r^2$.

An example may help the reader understand this important concept. In combining studies using IQ to predict general academic achievement, Walberg (1984) found the overall correlation between these variables to be .71. We can use this correlation to find $r^2 = .50$. This means that 50% of the variance in academic achievement (how well or poorly different students do) is predictable from the variance of IQ. This also obviously means that 50% of the variance of academic achievement is due to factors other than IQ, such as motivation, home environment, school attended, and test error. Walberg also found that the correlation of IQ with science achievement was .48. This means that only 23% (r^2) of variance in science achievement is predictable by IQ and that 77% is due to other factors, some known and some unknown. Finally, the correlation of IQ and posttest scores reported earlier for the 100 children in our data set in Appendix B is +.638 and between the pre- and posttests +.894. Thus, 41% ($.638^2$) of the variance in posttest scores is predicted by IQ while 80% ($.894^2$) is predicted by pretest scores.

There are additional techniques, some too advanced for this introductory text, that allow researchers to use more than one variable. Thus, it is possible, for example, to use a combination of IQ, pretest scores, and other measures such as motivation and a socioeconomic scale to predict academic achievement (posttest scores). This multiple correlation would increase the correlation, which would, in turn, increase the percent of variance of academic achievement that is explained by known factors. The next chapter (11) discusses how multiple regression results in multiple correlations.

Misinterpretation of the Coefficient of Correlation

Several fallacies and limitations should be considered in interpreting the meaning of a coefficient of correlation. The coefficient does not imply a cause-and-effect relationship between variables. High positive correlations have been observed between the number of storks' nests and the number of human births in northwestern Europe and between the number of ordinations of ministers in the New England colonies and the consumption of gallons of rum. These high correlations obviously do not imply causality. As population increases, both good and bad things are likely to increase in frequency.

Similarly, a zero (or even negative) correlation does not necessarily mean that no causation is possible. Glass and Hopkins (1996) point out, "Some studies with college students have found no correlation between hours of study for an examination and test performance. [This is likely due to the fact that] some bright students study little and still achieve average scores, whereas some of their less gifted classmates study diligently but still achieve an average performance. A controlled experimental study would almost certainly show some causal relationship" (p. 139).

Prediction

An important use of the coefficient of correlation and the Y on X regression line is for prediction of unknown Y values from known X values. Because it is a method for estimating future performance of individuals on the basis of past performance of a sample, prediction is an inferential application of correlational analysis. It has been included in this chapter to illustrate one of the most useful applications of correlation.

Let us assume that a college's admissions officers wish to predict the likely academic performance of students considered for admission or for scholarship grants. They have built up a body of data based on the past records of a substantial number of admitted college students over a period of several years. They have calculated the coefficient of correlation between their high school grade-point averages and their college freshman grade-point averages. They can now construct a regression line and predict the future college freshman GPA for any prospective student, based on his or her high school GPA.

Let us assume that the admissions officers found the coefficient of correlation to be $+ .52$. The slope of the line could be used to determine any Y values for any X value. This process would be quite inconvenient, however, for all grade-point averages would have to be entered as sigma (z) values.

A more practicable procedure would be to construct a regression line with a slope of b so that any college grade-point average (Y) could be predicted directly from any high school grade-point average. The b regression line and a carefully drawn graph would provide a quick method for prediction. For example

$$\begin{array}{ll} \text{If } r = +.52, & \text{then } b = r \frac{(S_Y)}{(S_X)} \\ S_Y = .50 & b = +.52 \frac{(.50)}{(.60)} \\ S_X = .60 & b = +.43 \end{array}$$

X_A is student A's high school GPA, Y_A his predicted college GPA.

X_B is student B's high school GPA, Y_B her predicted college GPA.

Figure 10.11 uses these data to predict college GPA from high school GPA.

Another, and perhaps more accurate, alternative for predicting unknown Vs from known Ks is to use the regression equation rather than the graph. The formula for predicting Y from X is

$$\hat{Y} = a + bX$$

where \hat{Y} = the predicted score (e.g., college freshman GPA)

X = the predictor score (e.g., high school GPA)

b = slope

a = constant, or Y intercept

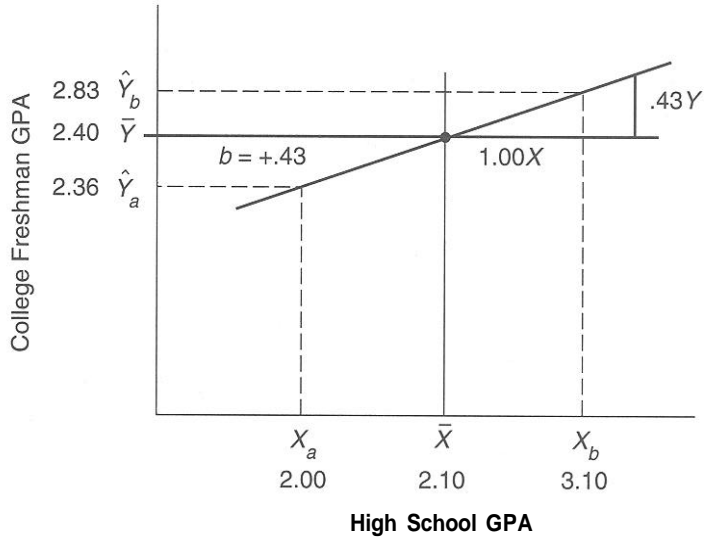


FIGURE 10.11 A Regression Line Used to Predict College Freshman GPA from High School GPA

We have already seen that $b = S_Y/S_X$. We can find a by $a = \bar{Y} - b\bar{X}$. Given the following data, we can then find the most likely freshman GPA for two students.

$b = .43$ (found earlier)

$$\bar{X} = 2.10$$

$$\bar{Y} = 2.40$$

$$a = 2.40 - 2.10(.43) = 2.40 - .90 = 1.50$$

$$X_a \text{ (student A's high school GPA)} = 2.00$$

$$X_b \text{ (student B's high school GPA)} = 3.10$$

$$\begin{aligned}\hat{Y}_a &= 1.50 + .43(X_a) \\ &= 1.50 + .43(2.00) \\ &= 1.50 + .86 \\ &= 2.36\end{aligned}$$

$$\begin{aligned}\hat{Y}_b &= 1.50 + .43(X_b) \\ &= 1.50 + .43(3.10) \\ &= 1.50 + 1.33 \\ &= 2.83\end{aligned}$$

For student A, whose high school GPA was below the mean, the predicted college GPA was also below the mean. For student B, whose high school GPA was well above the mean, the predicted GPA was substantially above the mean. These results are consistent with a positive coefficient of correlation in general: high in X, high in Y; low in X, low in Y.

STANDARD ERROR OF ESTIMATE

When the coefficient of correlation based on a sufficient body of data has been determined as ~ 1.00 , there will be no error of prediction. Perfect correlation indicates that for every increase in X, there is a proportional increase (when +) or proportional decrease (when -) in Y. There are no exceptions. But when the magnitude of r is less than $+1.00$ or -1.00 , error of prediction is inherent because there have been exceptions to a consistent, orderly relationship. The regression line does not coincide or pass through all of the coordinate values used in determining the slope.

A measure for estimating this prediction error is known as the standard *error of estimate* (S_{est}).

$$S_{\text{est } Y} = S_Y \sqrt{1 - r^2}$$

As the coefficient of correlation increases, the prediction error decreases. When $r = \pm 1.00$

$$s_{\text{est } Y} = S_Y \sqrt{1 - r^2} = S_Y \sqrt{1 - (1)^2} = S_Y(0) = 0$$

When $r = 0$

$$S_{\text{est } Y} = S_Y \sqrt{1 - (0)^2} = S_Y(1) = S_Y$$

When $r = 0$ (or when the coefficient of correlation is unknown), the best blind prediction of any Y from any X is the mean of Y. This is true because we know that most of the scores in a normal distribution cluster around the mean and that about 68% of them would probably fall within one standard deviation from the mean. In this situation the standard deviation of Y may be thought of as the standard error of estimate. When $r = 0$, $S_{\text{est } Y} = S_Y$

If the coefficient of correlation is more than zero, this blind prediction can be improved on in these ways:

1. By plotting Y from a particular X from the regression line (see Figure 10.12)
2. By reducing the error of prediction of Y by calculating how much S_Y is reduced by the coefficient of correlation

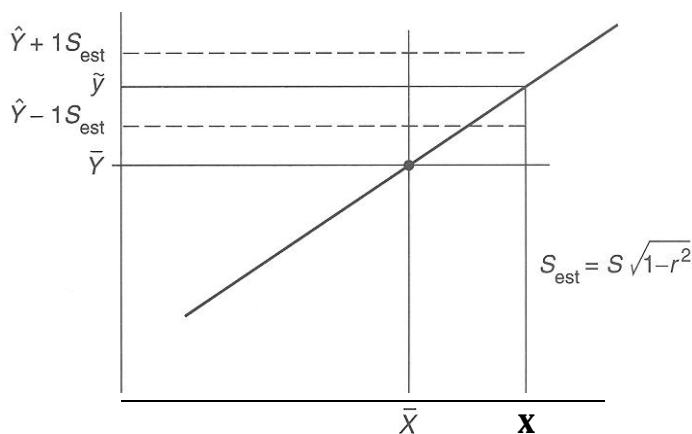


FIGURE 10.12 A Predicted Y Score from a Given X Score, Showing the Standard Error of Estimate

For example, when $r = \pm .60$

$$\begin{aligned} S_{\text{est } Y} &= S_Y \sqrt{1 - (r)^2} = S_Y \sqrt{1 - (.60)^2} = S_Y \sqrt{1 - .36} \\ &= S_Y \sqrt{.64} = .80 S_Y \end{aligned}$$

Thus the estimate error of Y has been reduced from S_Y to $.80 S_Y$. Interpretation of the standard error of estimate is similar to the interpretation of the standard deviation. If $r = +.60 S_Y$, the standard error of estimate of Y will be $.80 S_Y$. An actual performance score of Y would probably fall within a band of $+.80 S_Y$ from the predicted Y in about 68 of 100 predictions. In other words, the probability is that the predicted score would not be more than one standard error of estimate from the actual score in about 68% of the predictions.

In addition to the applications described, the coefficient of correlation is indispensable to psychologists who construct and standardize psychological tests and inventories. A few of the basic procedures are briefly described.

Computing the coefficient of correlation is the usual procedure used to evaluate the degree of validity and reliability of psychological tests and inventories (see Chapter 9 for a more detailed description of these concepts).

The Coefficient of Validity

A test is said to be valid to the degree that it measures what it claims to measure, or, in the case of predictive validity, to the extent that it predicts accurately such types of behavior as academic success or failure, job success or failure, or stability

or instability under stress. Tests are often validated by correlating test scores against some outside criteria, which may be scores on tests of accepted validity, successful performance or behavior, or the expert judgment of recognized authorities.

The Coefficient of Reliability

A test is said to be reliable to the degree that it measures accurately and consistently, yielding comparable results when administered a number of times. There are a number of ways of using the process of correlation to evaluate reliability:

1. Test-retest--correlating the scores on two or more successive administrations of the test (administration number 1 versus administration number 2)
2. Equivalent forms--correlating the scores when groups of individuals take equivalent forms of the test (form L versus form N)
3. Split halves--correlating the scores on the odd items of the test (numbers 1, 3, 5, 7, and so forth) against the even items (numbers 2, 4, 6, 8, and so forth). This method yields lower correlations because of the reduction in size to two tests of half the number of items. This may be corrected by the application of the *Spearman-Brown prophecy formula*.

$$r = \frac{2r}{1 + r}$$

If $r = \pm .60$,

$$r = \frac{1.20}{1 + .60} = +.75$$

A NOTE OF CAUTION

Statistics is an important tool of the research worker, and an understanding of statistical terminology, methodology, and logic is important for the consumer of research. A number of limitations, however, should be recognized in using statistical processes and in drawing conclusions from statistical evidence:

1. Statistical process, a servant of logic, has value only if it verifies, clarifies, and measures relationships that have been established by clear, logical analysis. Statistics is a means, never an end, of research.
2. A statistical process should not be employed in the analysis of data unless it adds clarity or meaning to the analysis of data. It should not be used as window dressing to impress the reader.
3. The conclusions derived from statistical analysis will be no more accurate or valid than the original data. To use an analogy, no matter how elaborate the mixer, a cake made of poor ingredients will be a poor cake. All the refinement of elaborate statistical manipulation will not yield significant truths if the data

result from crude or inexact measurement. In computer terminology this is known as GI-GO, “garbage in-garbage out.”

4. AU treatment of data must be checked and double-checked frequently to minimize the likelihood of errors in measurement, recording, tabulation, and analysis.
5. There is a constant margin of error wherever measurement of human beings is involved. The error is increased when qualities or characteristics of human personality are subjected to measurement or when inferences about the population are made from measurements derived from statistical samples.

When comparisons or contrasts are made, a mere number difference is not in itself a valid basis for any conclusion. A test of statistical significance should be employed to weigh the possibility that chance in sample selection could have yielded the apparent difference. To apply these measures of statistical significance is to remove some of the doubt from the conclusions.

6. Statisticians and liars are often equated in humorous quips. There is little doubt that statistical processes can be used to prove nearly anything that one sets out to prove if the procedures used are inappropriate. Starting with false assumptions, using inappropriate procedures, or omitting relevant data, the biased investigator can arrive at false conclusions. These conclusions are often particularly dangerous because of the authenticity that the statistical treatment seems to confer. Of course, intentionally using inappropriate procedures or omitting relevant data constitutes unethical behavior and is quite rare.

Distortion may be deliberate or unintentional. In research, omitting certain facts or choosing only those facts favorable to one's position is as culpable as actual distortion, which has no place in research. The reader must always try to evaluate the manipulation of data, particularly when the report seems to be persuasive.

SUMMARY

This chapter deals with only the most elementary descriptive statistical concepts. For a more complete treatment the reader is urged to consult one or more of the references listed.

Statistical analysis is the mathematical process of gathering, organizing, analyzing, and interpreting numerical data and is one of the basic phases of the research process. Descriptive statistical analysis involves the description of a particular group. Inferential statistical analysis leads to judgments about the whole population, to which the sample at hand is presumed to be related.

Data are often organized in arrays in ascending or descending numerical order. Data are often grouped into class intervals so that analysis is simplified and characteristics more readily noted.

Measures of central tendency (mean, median, and mode) describe data in terms of some sort of average. Measures of position, spread, or dispersion describe data in terms of relationship to a point of central tendency. The range, deviation,

variances, standard deviation, percentile, and Z (sigma) score are useful measures of position, spread, or dispersion.

Measures of relationship describe the relationship of paired variables, quantified by a coefficient of correlation. The coefficient is useful in educational research in standardizing tests and in making predictions when only some of the data are available. Note that a high coefficient does not imply a cause-and-effect relationship but merely quantifies a relationship that has been logically established prior to its measurement.

Statistics is the servant, not the master, of logic; it is a means rather than an end of research. Unless basic assumptions are valid; unless the right data are carefully gathered, recorded, and tabulated; and unless the analysis and interpretations are logical, statistics can make no contribution to the search for truth.

EXERCISES (ANSWERS IN APPENDIX I)

1. More than half the families in a community can have an annual income that is lower than the mean income for that community. Do you agree or disagree? why?
2. The median is the midpoint between the highest and the lowest scores in a distribution. Do you agree or disagree? Why?

3. Compute the mean and the median of this distribution:

74
72
70
65
63
61
56
51
42
40
37
33

4. Determine the mean, the median, and the range of this distribution:

88
86
85
80
80
77
75
71
65
60
58

5. Compute the variance (σ^2) and the standard deviation (σ) using the formula for the population (as indicated by the Greek letters) and then for a sample (S and S^2 , respectively) for this set of scores:

27
27
25
24
20
18
16
16
14
12
10
7

6. The distribution with the larger range is the distribution with the larger standard deviation. Do you agree or disagree? Why?
7. If five points were added to each score in a distribution, how would this change each of the following:
- the range
 - the mean
 - the median
 - the mode
 - the variance
 - the standard deviation
8. Joan Brown ranked 27th in a graduating class of 367. What was her percentile rank?
9. In a coin-tossing experiment where $N = 144$ and P (probability) = .50, draw the curve depicting the distribution of probable outcomes of heads appearing for an infinite number of repetitions of this experiment. Indicate the number of heads for the mean, and at 1, 2, and 3 standard deviations from the mean, both positive and negative.
10. Assuming the distribution to be normal with a mean of 61 and a standard deviation of 5, calculate the following standard score equivalents:

x	x	z	T
66			
58			
70			
61			
52			

11. Using the normal probability table in Appendix C, calculate the following values:
- below $-1.25z$ _____%
 - above $-1.25z$ _____%
 - between $-1.40z$ and $+1.67z$ _____%

- d. between $+1.50z$ and $+2.50z$ _____ %
- e. 65th percentile rank _____ z
- f. 43rd percentile rank _____ z
- g. top 1% of scores _____ z
- h. middle 50% of scores _____ z to _____ z
- i. not included between $-1.00z$ and $+1.00z$ _____ %
- j. 50th percentile rank _____ z
12. Assuming a normal distribution of scores, a test has a mean score of 100 and a standard deviation of 15. Compute the following scores:
- a. score that cuts off the top 10% _____
- b. score that cuts off the lower 40% _____
- c. percentage of scores above 90 _____ %
- d. score that occupies the 68th percentile rank _____
- e. score limits of the middle 68% _____ to _____
13. Consider the following table showing the performance of three students in algebra and history:

	Mean	σ	Tom	Donna	Harry
Algebra	90	30	60	100	85
History	20	4	25	22	19

Who had:

- a. the poorest score on either test? _____
- b. the best score on either test? _____
- c. the most consistent scores on both tests? _____
- d. the least consistent scores on both tests? _____
- e. the best mean score on both tests? _____
- f. the poorest mean score on both tests? _____
14. The coefficient of correlation measures the magnitude of the cause-and-effect relationship between paired variables. Do you agree or disagree? why?
15. Using the Spearman rank order coefficient of correlation method, compute ρ .

	X Variable	Y Variable
Mary	1	3
Peter	2	4
Paul	3	1
Helen	4	2
Ruth	5	7
Edward	6	5
John	7	6

16. Two sets of paired variables are expressed in z (sigma) scores. Compute the coefficient of correlation between them.

Z_X	Z_Y
+.70	+.55
-.20	-.32
+1.50	+2.00
+1.33	+1.20
-.88	-1.06
+.32	-.40
-1.00	+.50
+.67	+.80
-.30	-.10
+1.25	+1.10
+.50	-.20

17. Using the Pearson product-moment raw score method, compute the coefficient of correlation between these paired variables:

X	Y	X^2	Y^2	XY
66	42			
50	55			
43	60			
8	24			
12	30			
35	18			
24	48			
20	35			
16	22			
54	38			

18. A class took a statistics test. The students completed all of the questions. The coefficient of correlation between the number of correct and the number of incorrect responses for the class was _____
19. There is a significant difference between the slope of the regression line r and that of the regression line b . Do you agree? Why?
20. Compute the standard error of estimate of Y from X when:
- $S_Y = 6.20$
 $r = +.60$
21. Given the following information, predict the Y score from the given X , when $X = 90$, and:
- a. $r = +.60$
 $X = 80$ $S_X = 12$
 $Y = 40$ $S_Y = 8$
- b. $r = -.60$

ENDNOTE

1. N represents the number of subjects in the population; n represents the number of subjects in a sample.

REFERENCES

- Glass, G. V., & Hopkins, K. D. (1996). *Statistical methods in education and psychology* (3rd ed.). Boston: Allyn and Bacon.
- Glass, C. V., Peckham, P. D., & Sanders, I. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance. *Review of Educational Research*, 42, 237-288.
- Hays, W. L. (1981). *Statistics* (3rd ed.). New York: Holt, Rinehart & Winston.
- Heiman, G. W. (1996). *Basic statistics for the behavioral sciences*. (2nd ed.). Boston: Houghton Mifflin.
- Kerlinger, F. N. (1986). *Foundations of behavioral research*. (3rd ed.). New York: Holt, Rinehart, and Winston.
- Kirk, R. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole.
- Lunney, G. H. (1970). Using analysis of variance with a dichotomous dependent variable: An empirical study. *Journal of Educational Measurement*, 7, 263-269.
- Mandeville, G. K. (1972). A new look at treatment differences. *American Educational Research Journal*, 9, 311-321.
- Shavelson, R. J. (1996). *Statistical reasoning for the behavioral sciences* (3rd ed.). Boston: Allyn and Bacon.
- Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill.
- Walberg, H. J. (1984). Improving the productivity of America's schools. *Educational Leadership*, 41, 19-30.
- Winer, B. J. (1971). *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill.