Henry E. Klugh

STATISTICS
The Essentials for Research

THIRD EDITION

LEA

Statistics:
The Essentials for Research
Third Edition

Statistics:
The Essentials for Research
Third Edition

Henry E. Klugh
Alma College

LEA LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS
1986 Hillsdale, New Jersey London

Lawrence Erlbaum Associates, Inc., Publishers
365 Broadway
Hillsdale, New Jersey 07642

Library of Congress Cataloging in Publication Data

Klugh, Henry E. (Henry Elicker)
Statistics: the essentials for research.

Bibliography: p.
Includes index.
1. Social sciences—Statistical methods. I. Title.
HA29.K62 1986 519.5 86-4593
ISBN 0-89859-627-0

Printed in the United States of America
10 9 8 7 6 5 4 3 2 1

Contents

ACKNOWLEDGMENTS xiii

PREFACE TO THE INSTRUCTOR xv

1 INTRODUCTION 1

1.1 Advice to the Student *1*
1.2 On Calculators *2*
1.3 Why Statistics? *3*

2 GRAPHING DISTRIBUTIONS 5

2.1 Operational Definitions *5*
2.2 Scales of Measurement *7*
2.3 Discrete and Continuous Variables *12*
2.4 Frequency Distributions *12*
2.5 Frequency Polygons *13*
2.6 Histograms *15*
2.7 Bar Charts *15*
2.8 Grouped Frequency Distributions *15*
2.9 Cumulative Distributions *21*
2.10 Review *22*
2.11 Exercises *24*

3 MEASURES OF CENTRAL TENDENCY 26

3.1 Mode *27*
3.2 Rules of Summation *28*
3.3 The Mean *30*
3.4 Mean of Combined Groups *33*
3.5 Transformation Effects on the Mean *34*
3.6 Skewness *35*
3.7 Median *36*
3.8 Centile Points *41*
3.9 Calculating the Centile Equivalent of a Score, Centile Ranks *41*
3.10 On Calculations *42*
3.11 Difference Between Means and Medians *43*
3.12 Review *43*
3.13 Exercises *44*

4 MEASURES OF VARIABILITY 46

4.1 Range *46*
4.2 Variance *46*
4.3 Effect of Linear Transformation on the Variance *51*
4.4 Standard Deviation *53*
4.5 The Normal Distribution *55*
4.6 Standard Scores *57*
4.7 Normalized Standard Scores *63*
4.8 Review *65*
4.9 Exercises *65*

5 MEASURES OF ASSOCIATION 68

5.1 Scattergrams *68*
5.2 The Correlation Coefficient *70*
5.3 The Correlation Ratio *75*
5.4 Rank Order Correlation *77*
5.5 The Point-Biserial Correlation Coefficient *78*
5.6 The $\emptyset$ Coefficient *80*
5.7 Factors Influencing the Interpretation of Correlation *81*
5.8 Assumptions *89*
5.9 Review *89*
5.10 Exercises *90*

6 PREDICTION **92**

6.1 The Regression Line of Y on X *92*
6.2 The Regression Equation with z Scores *95*
6.3 The Regression Line of X on Y *96*
6.4 The Problem of Regression in Experimental Design *100*
6.5 The Regression Paradox *103*
6.6 Errors of Prediction *103*
6.7 Error Variance and Regression *107*
6.8 Using the Standard Error of Estimate *108*
6.9 Interpretations of Correlation *111*
6.10 Restriction of Range *112*
6.11 Multiple Correlation *114*
6.12 Multiple Regression *116*
6.13 Discriminate Function Analysis *119*
6.14 Partial Correlation Coefficients *119*
6.15 Historical Note *121*
6.16 Review *121*
6.17 Exercises *122*

7 THE NATURE OF RESEARCH **124**

7.1 Independent and Dependent Variables *124*
7.2 Populations and Samples *126*
7.3 Sampling Error and Sampling Distributions *130*
7.4 The Null Hypothesis *132*
7.5 Two Kinds of Error *134*
7.6 Levels of Significance *135*
7.7 Glossary of Terms *139*
7.8 Examples *141*
7.9 Exercises *145*

8 PROBABILITY **147**

8.1 A Definition of Probability *147*
8.2 The Product Rule *148*
8.3 The Additive Rule *150*
8.4 Examples Using Both the Additive and Product Rules *152*
8.5 Permutations and Combinations *153*
8.6 The Binomial Expansion and Probability *157*
8.7 Binomial Probabilities when $p \neq q$ *162*
8.8 The Multinomial Distribution *163*
8.9 Exercises *164*

9 THE BINOMIAL DISTRIBUTION AND ITS NORMAL APPROXIMATION 166

9.1 Theoretical Sampling Distributions Based on Binomial Populations *166*
9.2 Approximation for Frequencies *171*
9.3 Correction for Continuity *173*
9.4 Examples *174*
9.5 The Approximation for Proportions *176*
9.6 Hypothesis Testing: One-Tailed Tests *178*
9.7 Hypothesis Testing: Two-Tailed Tests *182*
9.8 Confidence Intervals *185*
9.9 Sampling from Finite Populations *188*
9.10 Power and the Type II Error *189*
9.11 Another Look at the Null Hypothesis *193*
9.12 Assumptions for the Use of the Normal Approximation *193*
9.13 Review *194*
9.14 Exercises *195*

10 CHI SQUARE 197

10.1 The Chi Square Distribution *197*
10.2 One- and Two-Tailed Tests Using Chi Square *201*
10.3 Testing the Significance of a Sample Distribution's Departure from Theoretical Distributions when df > 1 *202*
10.4 Testing the Significance of a Sample Distribution's Departure from a Normal Distribution *205*
10.5 The Significance of Differences Between Proportions *205*
10.6 Partitioning χ^2 *211*
10.7 Restrictions on the Use of χ^2 *217*
10.8 Chi Square for Correlated Observations *218*
10.9 χ^2 with Proportions and Percents *222*
10.10 Power and χ^2 *224*
10.11 Overview *224*
10.12 Exercises *225*

11 THE *t* DISTRIBUTION 230

11.1 Hypotheses About the Population Mean *231*
11.2 Review of the Normal Deviate *z* *236*
11.3 The *t* Distribution and Hypotheses About the Population Mean *237*
11.4 Confidence Interval for the Population Mean *239*

11.5 The Significance of a Difference Between Means of Independent Groups *241*
11.6 The t Test for Paired Observations *246*
11.7 Illustrations of Experimental Designs Using Independent Groups and Matched Pairs t Tests *251*
11.8 Assumptions for t *256*
11.9 Transformations of the Scale *257*
11.10 The Strength of an Effect *258*
11.11 Tests of Significance for Correlation Coefficients *260*
11.12 Summary *265*
11.13 Exercises *266*

12 ANALYSIS OF VARIANCE I 270

12.1 Relationships Between Estimates of the Population Variance *273*
12.2 The F Ratio *274*
12.3 A Demonstration of the Independence of S_b^2 and S_w^2 *275*
12.4 Computing Formulas and Examples *279*
12.5 Planned Comparisons Following a One-Way ANOVA *283*
12.6 Post Hoc Comparisons: Scheffé's Test *287*
12.7 Measuring the Strength of an Effect *288*
12.8 Analysis of Variance and Multiple Regression Analysis *290*
12.9 Assumptions for the Use of Analysis of Variance *292*
12.10 F Test for Sample Variances *295*
12.11 Hartley's $F_{MAX.}$ *296*
12.12 The Relationship Between t and F *297*
12.13 Summary *297*
12.14 Exercises *298*

13 ANALYSIS OF VARIANCE II: FACTORIAL DESIGNS 301

13.1 The Factorial Design *301*
13.2 The Main Effects *302*
13.3 The Interaction *303*
13.4 Independence of Sources of Variation *304*
13.5 Partitioning the Sums of Squares *306*
13.6 Partitioning the Degrees of Freedom *309*
13.7 A Resume of the Analysis *309*
13.8 Computational Formulas and Examples *311*
13.9 Higher Order Factorial Designs *320*
13.10 Equal Numbers of Subjects in the Subgroups *332*
13.11 Independent Observations Revisited *333*

13.12 Models in the Analysis of Variance *333*
13.13 Nested Designs *335*
13.14 Exercises *337*

14 ANALYSIS OF VARIANCE III: REPEATED MEASURES ANALYSIS OF VARIANCE 341

14.1 One-Way Repeated Measures Analysis of Variance *341*
14.2 Assumptions for the Repeated Measures ANOVA *348*
14.3 Computational Formulas and Examples of the Repeated and Randomized Blocks Analysis *348*
14.4 Repeated Measures in a Factorial Design *351*
14.5 Designs with Random Assignment and Repeated Measures; Mixed Designs *357*
14.6 The Nature of Complex Designs in Analysis of Variance *372*
14.7 Exercises *373*

15 NONPARAMETRIC TESTS OF SIGNIFICANCE 376

15.1 The Sign Test *376*
15.2 The Wilcoxon Matched Pairs Signed Ranks Test *378*
15.3 The Median Test *379*
15.4 The Mann-Whitney U Test *380*
15.5 The Kruskall-Wallis One-Way Analysis of Variance by Ranks *382*
15.6 Relative Power of the Nonparametric Tests *382*
15.7 Exercises *383*

ANSWER SECTION 387

APPENDICES 411

Appendix 1
Table A A Table of Random Numbers *412*
Table C Table of χ^2 *414*
Table F Critical Values of *F* Values Marking 5% and 1% of the Right Tail of the *F* Distribution *416*
Table H Critical Values for Hartley's Maximum *F*-Ratio Significance Test for Homogeneity of Variance *420*
Table N Percent of Total Area Under the Normal Curve Between Mean Ordinate and Ordinate at any Given Sigma-Distance from the Mean *422*
Table R Critical Values of the Correlation Coefficient *423*

Table T Critical Values of t for a Two-Tailed Test *424*
Table U Significant Values of U for the Mann-Whitney Test *425*
Table W The Wilcoxon Signed Rank Test. Minimum Significant Values of T for a Two-Tailed Test *426*
Table Z Transformation of r to Z' *427*

Appendix 2 Orthogonal Coefficients *429*

INDEX **431**

Acknowledgments

I thank Mr. Timothy Sipka whose comments and suggestions on portions of the manuscript were particularly helpful. I gratefully acknowledge the assistance of Susan Dinwoody and Betty Shaver whose patiance and typing skills expedited the many drafts leading to this edition of the text and thus contributed substantially to the clarity of the manuscript. I also thank the classes of Alma College students who labored through pre-publication versions of the first and second editions and the more recent classes that provided suggestions incorporated in the present edition.

I am particularly indebted to the Literary Executor of the late Sir Ronald A. Fisher, F.R.S. Cambridge, and to Oliver and Boyd Ltd., Edinburgh for permission to reprint Tables III and IV from their book *Statistical Methods for Research Workers*, and for their permission, with that of Dr. Frank Yates, F. R. S. Rothampsted, to reprint Table VI from their book *Statistical Tables for Biological, Agricultural and Medical Research*. Acknowledgments to the authors and original publishers of these and other tables in the Appendix appear following each table.

H. E. K.

Preface to the Instructor

For most undergraduates in the behavioral sciences the required course in statistics is both introductory and terminal. Any text for such a course should accomplish several purposes that are unfortunately, almost mutually exclusive. First, the text should be easy to understand for students with little or no college-level mathematics. It must dispense with mathematical precision when precision unnecessarily increases complexity and reduces "teachability." Second, the text should provide sufficient statistical sophistication to enable the student to read the research literature critically and conduct experiments that are complex enough to be interesting. Finally, the text should provide a foundation for additional work in statistics for students who continue their education in graduate school. Unfortunately, if the first two purposes have not been met, if the student is forced to plod through an unnecessarily rigorous development of concepts, or if the student is presented with procedures rarely encountered in professional journals, enthusiasm for research will not develop and additional courses in statistics will not be needed.

While the text is written for the introductory course, it includes some difficult material. It reflects the changing trends in the kinds of statistical analysis *used* by investigators in the behavioral sciences. The available evidence has suggested for some time that the introductory semester of statistics must go beyond the t test, the correlation coefficient, chi square, and elementary ANOVA if it is to help undergraduates understand contemporary research literature. Edgington[1] writing a decade ago noted substantial increases in the use of ANOVA in the

[1]E. S. Edgington, "A New Tabulation of Statistical Proceduress Used in APA Journals." *American Psychologist*, 1974, *29*, 25–26.

results section of APA journals. He concluded that if undergraduates were to read professional journals with understanding the introductory statistics course would have to cover fairly complex factorial and repeated measures ANOVA. The first two editions of this text have done exactly that.

This, the third edition, will be published almost exactly 20 years after the first edition received its initial classroom tryout, and includes still more ANOVA material. As with the first two editions the student is "talked through" the logic of the process. Mathematical models are not developed rigorously although they are described and the issues involved in their use receive more attention in this edition than in previous ones. I have separated the material on correlation and regression, expanded each topic considerably so that each now constitutes a separate chapter. Multiple regression, an increasingly used procedure, is presented in detail and later its relationship to ANOVA is described by analyzing the same data with both procedures. Much more time has been devoted to the distinction between planned and post hoc tests. Orthogonal contrasts are thoroughly described. I have also included much more material on interpretation and research design and deemphasized routine computations.

Instructors teaching at selective colleges should be able to include most of these topics in their courses. Those sections which go beyond the usual content of a rigorous introductory course have been marked by an asterisk in the Table of Contents. Some, or all, of these sections can be assigned as honors work or used as the statistical portion of a later methods course.

In summary, this edition like its predecessors is designed to teach in a narrative style those statistical techniques most often encountered in the behavioral sciences to *above average* undergraduates.

H. E. Klugh

Statistics:
The Essentials for Research
Third Edition

1 Introduction

1.1 ADVICE TO THE STUDENT

Many students in the behavioral sciences approach their first course in statistics wishing they had majored in some other discipline—*any* other discipline not requiring them to study statistics. They are convinced that their "aptitude for math" is low, and they are prepared to find statistics difficult. Statistics is difficult for many students but not, as a rule, because they have a low aptitude for mathematics. Students who have little or no college work in mathematics suffer from a serious misconception about the rate at which they should be able to read mathematical material. Accustomed to reading a 50-page assignment for other courses in two or three hours, they discover that they have spent that much time before they understand six pages of statistics. When this happens, they become absolutely convinced that they have no aptitude for mathematics and, in despair, they drop statistics and change their majors! If you have had similar thoughts, the following paragraphs are especially for you.

Not even mathematicians read mathematics as rapidly as they read other material. Of course, this depends on the material and on the individual, but even professional mathematicians read unfamiliar mathematics at a considerably slower pace than they read anything else. When you require a great deal of time to understand a page of mathematics, it *does not* reflect unfavorably on your intelligence or mathematical aptitude. You should *expect* to read a statistics textbook much more slowly than you read other textbooks, and you should expect to reread some sections a number of times before the relationships discussed become clear.

One reason for slower reading in a statistics course is that complex ideas are communicated by the use of unfamiliar symbols. In most other courses, you already know the meaning of the words by which any new ideas are communicated. Your problem is to understand and then to remember a novel thought communicated by a new arrangement of largely familiar symbols. In statistics, however, many of the symbols *and* most of the concepts are entirely new! You must begin by learning this new and fairly complex vocabulary of symbols before you can understand the concepts communicated by that vocabulary. For this reason you should make sure you know the meaning of each new symbol or term before you read beyond the section in which it is introduced. And you should expect the study of statistics to take more time, page for page, than you must devote to your other courses. It is very important that you see this process as a challenge which you can meet. The material is much like a crossword puzzle, a chess problem, or a challenging bridge hand. Certainly a portion of your task is to remember, but in this course, that is far less important than to understand, to comprehend. In many college courses understanding and comprehension are automatic. The tough task is remembering specific facts. In statistics the tables are turned; understanding is the primary task. When that is accomplished, retention will be almost automatic!

You can check your comprehension by answering the questions at the end of the chapters, and by reviewing the adequacy of your answers in the answer section before going on to the next question. Answers for most of the problems are supplied, and the procedures by which certain answers are obtained have also been included in the answer section. If you cannot answer one of the questions you should go back and reread the appropriate section of that chapter. Above all you must study the material regularly, but preferably not for more than a few hours at a time. Finding yourself a chapter behind on the day before the test is not a position from which you can recover by an all-night study session. If you are willing to exert consistent effort, you will probably finish the course with much more respect for your "mathematical aptitude" than you had when you began.

1.2 ON CALCULATORS

One piece of equipment that will make your statistics course much easier is the pocket calculator, which sells for as little as $15.00, and is really indispensable for much of the homework you will be assigned. There are, of course, many excellent models and more are appearing all the time so we will not make any recommendations here. Your instructor will have some advice about calculators.

Of course if you have access to a small computer with an appropriate statistical software package, life will be even more pleasant. If this isn't available, don't worry. All calculations can be completed within a reasonable time on a good electronic calculator.

1.3 WHY STATISTICS?

All sciences, including psychology, try to describe and ultimately understand relationships between the empirical events (observations) in their disciplines. In some areas of science (notably physics and chemistry), but also in some sections of psychology, the relationships between these events may be clear cut and easy to demonstrate. For example, the length of time it will take a 1-cubic-inch marble to fall 4 feet can be determined with a fairly high degree of accuracy. If air density is kept constant, and our instruments are in order, we can probably obtain almost exactly the same result with all marbles of similar dimensions. In this example from physics one must consider the density of the medium and the shape of the marble, but for all practical purposes that ends the list of variables that might affect the outcome.

On the other hand, we might wish to know the speed with which a rat will traverse a 4-foot alley for food reward. We can set up instruments for measuring elapsed time which are just as sophisticated as those used in the physics experiment, but it is quite unlikely that the psychologist's rats will produce the consistent speeds produced by the physicist's marbles. The behavioral scientist has a great many more variables to control. Of course, the rats to be compared should all be equally deprived of food, all of the same sex, age, and weight, all receive the same amount and type of food reward on earlier trials, and all be housed under identical conditions. If we carefully observe *all* of these controls, and then compare the running times of two rats chosen by lot, we shall almost certainly find the times to be different; not quite as different as they would have been without the controls, but different nevertheless.

In Table 1.1, in the theoretical column, we have recorded the running times one might expect for a group of "identical" rats, if these were obtainable and, in the observed column, the running time of real rats as they might be recorded in a real experiment.

TABLE 1.1
Time to Traverse an Alley Maze

Rat	*Theoretical Rats*	*Observed Rats*
1	4 sec.	6 sec
2	4 "	8 "
3	4 "	4 "
4	4 "	3 "
5	4 "	2 "

Even if we have exerted every effort to hold constant the unwanted influences on running time, it is still quite safe to assume that we have not controlled them all. Some rats may have been handled a bit more roughly than others; some may have had a fight with their cage mate just before running the alley; one may have noticed an attractive (or repulsive) odor left by the previous occupant of the start box and adjusted his time accordingly. In short, the study of behavior often involves a host of variables, not all of which can be controlled, that act to disguise the relationships between the variables under investigation.

All scientific observations, even those of physicists, contain a true component and an error component. The true component is equivalent to the theoretical running times of Table 1.1, and the error component is the sum of all the chance, or randomly operating uncontrolled variables that, when added to the true component, give rise to each entry in the observed column. This error component often tends to disguise relationships between events just as static tends to disguise intelligible sound from a radio.

There are two ways to reduce the effect of error: experimentally, by careful laboratory procedures; and statistically, by increasing the number of observations and manipulating the data so that the relationships will be apparent in spite of the random or chance error. Statistics, as a discipline, is concerned with this latter process. In its applied form, as we shall study it in this text, it is concerned with describing and drawing inferences from many observations; observations that are ordinarily translated into measurements or counts.

The study of statistics may be divided into two broad areas. One of these areas is called inferential statistics because it deals with *inferences* about the true nature of the relationships between variables in spite of the ever present chance or error component in their measurement. Most of this book is devoted to inferential statistics. Before we can infer anything from observations, however, they must be described in a systematic fashion. This branch, called descriptive statistics, shows us efficient ways to describe and summarize data, and consequently how to present it in the most usable form. It is this aspect of statistics that we now discuss.

2 Graphing Distributions

In this chapter we discuss the graphical presentation of data, but first we comment briefly on the nature of the data with which the scientist works.

2.1 OPERATIONAL DEFINITIONS

An experiment, in its simplest form, is designed to investigate the effect of one variable upon another. A variable may be defined as any property on which events or objects can take different values. For example, although they represent rather different kinds of variables, IQ, height, sex, and family size are all variables. Scientific convention uses the term "independent variable" to designate any variable presumed to exert the effect, and the term "dependent variable" to designate the variable presumably affected. If we investigate the effect of hunger on activity, hunger is the independent variable and activity is the dependent variable.

If you keep your subjects away from food and observe any systematic changes in their tendency to be active, you have some of the elements of an experiment. You might take notes on the behavior of your subjects, and then summarize your observations in a written description of their behavior. Unfortunately, another investigator conducting the same experiment might write a different report, not necessarily because of differences in the behavior of the animals but, perhaps, because of differences between you and the other investigator regarding the kinds of behavior each considered to be indicative of "activity."

We can increase the objectivity and hence the reliability of such an experiment if we define hunger and activity in a way that permits their measurement. If we define hunger by specifying the operations used to produce or measure it, the definition is called an operational definition. We can define the degree of hunger operationally in terms of the number of hours since food was last available to the animal. Thus, by definition, an animal deprived of food for 24 hours is "hungrier" than one which has been deprived of food for 6 hours. Notice that such a definition does *not* describe the internal stimuli produced by the absence of food, nor does it describe the sensations presumably endured by a hungry animal. In fact, there are a number of ways in which such a definition is deficient, but it *is* an operational definition; it specifies the operations by which "hunger" is produced. If we accept this operational definition of the independent variable we can form different "hunger" subgroups by depriving some animals of food for 6 hours, some for 12 hours, and some for 24 hours.

Similarly, activity can be operationally defined as the number of rotations of an activity wheel made by the animal during a 5-minute test period. Each subject in the different "hunger" subgroups can then be given an activity score, and average activity scores can be compared among the hunger subgroups. If different experimenters repeat this experiment using the same operational definitions of the variables under investigation, and the same procedures and apparatus, they should arrive at essentially the same description of the relationship between these variables. The operational definition is thus an extremely valuable scientific tool; it helps investigators to know if they are really discussing the same phenomena.

Operational definitions increase the objectivity and consequently the reliability or consistency of our experimental results. Unfortunately we achieve this precision at a substantial price—a severe limitation on our ability to generalize our findings. We cannot, without additional research, generalize our findings to the same variables operationally defined in other ways. For example, we might just as well define hunger as the force with which an animal will pull against a tension device in an attempt to reach food placed 10 cm away, and activity as the amount of movement the animal produces in a balance cage or stabilimeter. Also perfectly valid operational definitions, the relationship between this "hunger" and this "activity" may or may not be the same as we might have found with our original operational definitions.

It is probably evident that we can produce many different operational definitions of "hunger" and many different operational definitions of "activity." These groups of definitions determine what we shall call a *construct*. Thus "hunger" and "activity" are constructs, abstractions subject to a variety of definitions. No one definition is "right" but if one investigator uses one definition and another uses a different definition they should not necessarily expect the same results.

If we read that intelligence is related to achievement we must know how each of these constructs was operationally defined or we cannot evaluate the claim. If we read that punishment produces hostility we must know the operations by which punishment was produced and hostility measured or we will not be exactly sure of the statement's meaning. Indeed, if a construct (abstraction) cannot be operationalized, we cannot investigate it scientifically.

2.2 SCALES OF MEASUREMENT

For variables to be operationally defined, the methods by which they are measured must be specified. Consequently, the process of measurement and the various kinds of measurement scales are of considerable importance to the behavioral scientist.

There are four principal types of measurement scales, and we shall discuss each one briefly. The *nominal* scale consists simply of the specification of attributes so that the variable in question can be divided into mutually exclusive categories. For example, political parties constitute a nominal scale. To use such a scale we specify the membership requirements of Republican, Democratic, Socialist, and other parties, so that an individual belonging to one of these groups can be identified and counted. The categories composing the elements of a nominal scale must be exhaustive and mutually exclusive; every individual must fall into one, but only one category. Examples of other nominal scales are martial status, college major, and sex. Nominal scales are sometimes referred to as nonorderable countables. In a nominal scale, such as "marital status," we can only count or enumerate the individuals in each of the following categories.

	Marital Status	Frequency
1.	Single, never married	65
2.	Presently married	330
3.	Divorced—not remarried	29
4.	Widowed—not remarried	10
	Total	434

Although we can assign the numerals 1 to 4 to assist in the designation of the various marital statuses, it is clear that the categories are not intrinsically orderable; the numerals used to *designate* the categories have no quantitative significance. Four single individuals are not equivalent to one widowed person simply because the numeral "1" designates single and "4" designates widowed. Thus, in a nominal scale, where numbers may be used to designate nonorderable categories, the numbers simply take the place of names; they have no other significance. Standard arithmetical operations with these numbers would yield quite meaningless results.

The *ordinal* scale is a somewhat more sophisticated measuring device. The categories in an ordinal scale do imply order. For example, if we wished to measure "friendliness," we could develop a scale consisting of the categories: extremely friendly, very friendly, friendly, slightly friendly, and acquainted. While such a scale has a variety of shortcomings, it does have a definite order such that the categories, from left to right, represent decreasing amounts of friendliness. Numbers can also represent decreasing amounts of a quantity, so we can use numbers to represent this property of the scale. Thus we might have: (5) extremely friendly, (4) very friendly, (3) friendly, (2) slightly friendly, and (1) acquainted. Notice that in the nominal scale, which we discussed previously, numbers were used only as numerals, as *names* for categories. In the ordinal scale we also let the order of the numbers represent the *order* in the categories.

A difficulty now arises because when *numbers* represent the categories of an ordinal scale, the numbers have other properties *not* possessed by the scale itself. For example, the intervals between the numbers 1, 2, 3, *etc.* are all the same. There is exactly one unit between each number. However, we do not know the size of the "friendliness" intervals. We do not now know if it requires as much "friendliness" to move from "acquainted" to "slightly friendly" as it does to move from "friendly" to "very friendly." The intervals between the *numbers* designating these scale categories are the same of course, but that does not mean that there are equal intervals between the *actual scale categories* represented by the chosen numbers. An ordinal scale implies only order. It does not imply that there are equal intervals between the scale categories, or that the scale numbering begins with a true zero point. The scale intervals may not be equal so we cannot meaningfully add or subtract the numbers representing ordinal scale categories unless we arbitrarily *assume* that the intervals between them are equal.

Psychologists make frequent use of ordinal scales. For example, any rank ordering of individuals or objects produces an ordinal scale. A painting ranked first in an art contest is presumably better than a painting ranked second, and it in turn is presumably better than a painting ranked third. But such a scale cannot tell us if the *difference* in artistic value between the first and second ranked paintings is the same as the difference in artistic value between the second and third ranked paintings. The second ranked painting could be very close to the painting ranked first, or very close to the painting ranked third; either way it receives the rank of second. And, of course, the painting in last place cannot be assumed to have zero artistic merit!

Equal intervals between scale values first emerge in the *interval* scale. Temperature measured by degrees Centigrade or Fahrenheit is measured by an interval scale. The difference in temperature between 18° and 19°F is the same as the difference between 180° and 181°F. Similarly, the difference between 1° and 2°C is exactly the same as the difference between any other two consecutive degree marks on a centigrade thermometer.

We have defined an interval scale as one that has equal amounts of the measured variable within consecutive units of the measuring scale. It follows that such scale values can be added and subtracted meaningfully; thus, $6°C - 3°C = 2°C + 1°C$.

Psychologists have not been entirely successful with their efforts to develop interval scales. The earliest attempts were made in the field of psychophysics by using the concept of the "just noticeable difference" (jnd) to provide a constant unit of sensation. More recent attempts have been made to develop interval scales in the measurement of attitudes and intelligence.

While an interval scale permits the operations of addition and subtraction, it does not permit the formation of ratios. This is because the interval scale does not have a true zero point. For example, 0°C does not represent the absence of temperature. It is simply the point at which water freezes. As a result, 4°C, while four times as large a *number* as 1°C, does not represent four times as much temperature. Look at Figure 2.1. Note that the absence of temperature is at −273°C or 0° Kelvin. The units in the Kelvin scale are equivalent to Centigrade units, but the Kelvin zero point is, approximately, the absence of temperature. As we can see from Figure 2.1, 1°C is not 1° of *temperature* but 273° + 1° of temperature; similarly 4°C is not 4° of temperature but 273° + 4° of temperature. Therefore 4°C is, in fact, 277/274 or 1.01 times as much temperature as 1°C. In the Kelvin scale the numerical zero coincides with temperature zero; thus, numerical ratios of degrees Kelvin correspond to ratios of measured temperature.

Consider another example. Look at Figure 2.2. Suppose we have three sticks of different and unknown lengths, and arrange them all so that they stand on end on some flat surface. We might find that stick S is the shortest stick, stick A is equal to stick S plus 1 inch, and stick B is equal to stick S plus 4 inches. Accordingly, we may let stick $A = S + 1$ in. and $B = S + 4$ in. Stick A is clearly one unit longer than standard, and B is four units longer than standard. However, B is not four times as long as A unless the standard is zero. Note that $4(S + 1 \text{ in.}) = S + 4$ in. *only* when $S = 0$.

When measurements begin at a true zero point and the scale also has equal intervals we have a *ratio* scale. Length, mass, and time are ordinarily measured with ratio scales, but temperature can also be measured with a ratio scale when we record it in degrees Kelvin.

Some of the measurement scales we have just discussed should only be subjected to limited forms of statistical analyses. For example, finding the average of a nominal scale will result in a totally meaningless figure because the individual numbers are only used to stand for names; they have no quantitative referent. In fact, averaging ordinally scaled data is not really appropriate either. It would be much like averaging the lengths of a group of objects measured with a ruler

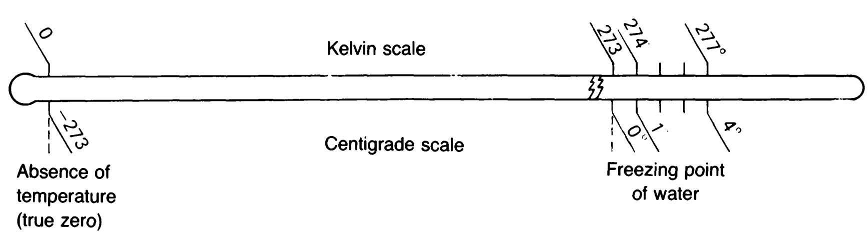

Figure 2.1 A comparison of Centigrade and Kelvin scales.

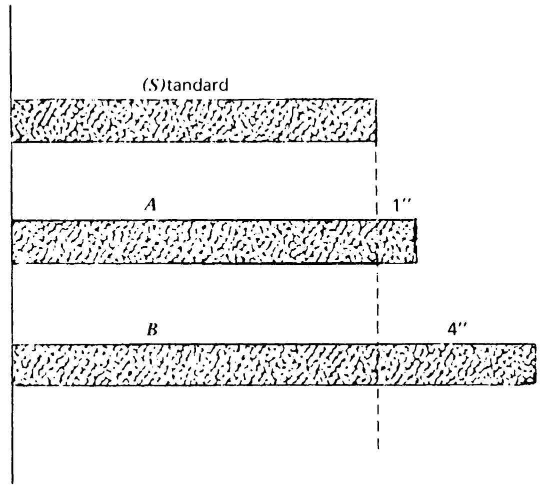

Figure 2.2 Length measured from an arbitrary origin instead of a true origin.

having "inches" of varying and unknown size. We can only be certain that averaging data will result in meaningful numbers when we have interval or ratio scales.

To illustrate this we consider the data in Table 2.1. There we have shown the hypothetical outcome of the judging for a foot race where performances are measured by both ordinal and ratio scales. There are three races—1, 2, and 3—and there are four contestants—*A*, *B*, *C*, and *D*. Assume that each contestant is timed and, in addition, that the contestant's rank order at the finish is recorded. These data appear in the appropriate columns of the table.

First let us determine the overall winner by finding the runner with the lowest overall running time. The summed running times appear in the lower right of the table. Note the order: *B* is fastest with a total time of 30.3 seconds; *B* is followed rather closely by *A*; *D* is considerably slower and is followed closely by *C*. Now suppose the running times are lost, and we must decide the winner on the basis of the rank ordering in each race. If we sum the ranks (lower left corner of the table) instead of running times we find that *A* is now the overall winner followed closely by *B*; *C* is some distance behind followed closely by *D*. Why the discrepancy? When we summed the ranks and ignored the times we lost our measure of the magnitude of the differences between the contestants. Ranks do not take these magnitudes into account. Each difference in rank is treated the same as every other difference in rank in spite of the fact that, as in this example, the equivalent differences in rank may represent very unequal differences in the variable upon which ranking is based. It is not, therefore, logically defensible to add ordinal measurements.

In spite of these logical restrictions, it is quite common to average achievement

TABLE 2.1
A Comparison of Ordinal and Ratio Scales

Race 1		*Race 2*		*Race 3*	
Times	*Rank*	*Times*	*Rank*	*Times*	*Rank*
10.0	B (1st)	10.0	A (1st)	10.0	
10.1		10.1	B (2nd)	10.1	A (1st)
10.2		10.2	D (3rd)	10.2	B (2nd)
10.3		10.3		10.3	
10.4		10.4		10.4	
10.5	A (2nd)	10.5		10.5	
10.6	C (3rd)	10.6		10.6	C (3rd)
10.7	D (4th)	10.7	C (4th)	10.7	D (4th)
Sum of ranks		A = 4		Sum of times	B = 30.3
		B = 5			A = 30.6
		C = 10			D = 31.6
		D = 11			C = 31.9

test scores and even to average the results of rating scales. In such instances we simply *assume* that no serious errors will be incurred and, in most cases, the assumption is probably safe. The ever-present grade point average, or honor point ratio, with which most college students are familiar and which decides probation or graduation with honor, is the result of averaging course grades which were probably based on the use of ordinal scales (course tests). Similar situations may be found when ordinal scaling techniques are used to measure personality characteristics and other psychological dimensions.

One additional point must be made about measurement scales. Do not confuse the scale with the construct. The construct is an abstraction which is measured by the scale. The scale may well be a "sophisticated" ratio scale but that doesn't mean the construct follows the scale. For example, suppose we reconsider the operational definition of hunger which we said earlier was "hours since food was last available." Time is clearly a ratio scale. One hour is half as long as two hours which is half as long as four hours. Unfortunately, time since eating is not a ratio scale of the construct "hunger." It does not seem reasonable to insist that you are six times as hungry one hour after eating as you were ten minutes after eating. We can, of course, form ratios of time, but those ratios do not necessarily correspond linearly to the ratios of other sequential events, e.g. phenomenological increases in hunger.

Similarly, jail terms are a ratio scale, but not necessarily a ratio scale of "punishment." Is a three year sentence three times as much "punishment" as a one year sentence? The matter is not settled simply by reference to time as a measurement scale, but by psychophysical procedures which require other independent definitions of the constructs.

2.3 DISCRETE AND CONTINUOUS VARIABLES

Finally, a distinction of considerable importance to mathematical statisticians is that between continuous and discrete measurement. Continuous measurement occurs when we can infinitely subdivide the units of our measurement "scale." Height, time, and weight are such continuous variables. When we can only count events, such that only whole numbers result, we have a discrete scale. Family size, number of parking tickets, and number of siblings are discrete variables. Unfortunately this distinction can become a bit blurred when we apply it to achievement test scores. If a test has 50 items, it would appear that we have a discrete variable; you can get 39 right or 40 right but not 39.126 right. However, we usually *treat* test scores as if they were continuous variables. We assume that instead of measuring achievement with a 50-item test we could have used a 500-item test or a 5000-item test so that we could, in theory, have an infinitely dividable continuous scale. We will say more about this issue a bit later.

2.4 FREQUENCY DISTRIBUTIONS

Regardless of the scale of measurement used, the data from an experiment must be presented in an orderly fashion. Suppose we wish to compare the effectiveness of two different methods of instruction. We may have test scores from one group of students taught by the lecture method and another group taught by the discussion method, and we may wish to compare the two sets of scores. The data may be compared more easily if we first tabulate the scores into two frequency distributions. A frequency distribution is a listing of all the different score values in order of magnitude with a tally or count of the number of scores at each value. Table 2.2 shows two frequency distributions that might result if our data were presented in this form.

With the scores pictured as they are in Table 2.2 we can see some differences between the distributions. The lecture method seems to produce higher achievement, but the range of scores is about the same. We can also observe that the scores of the lecture students tend to be concentrated toward the top of the distribution, while the scores of the discussion group seem to be more symmetrically distributed about a central value. These differences, rather easily seen in Table 2.2, would be completely obscured if the scores were simply presented in haphazard order.

TABLE 2.2
Frequency Distributions of Examination Scores for Students Taught by Lecture and by Discussion Methods

Lecture Method			*Discussion Method*		
Score	*Tally*	*Frequency*	*Score*	*Tally*	*Frequency*
54	//	2	54		0
53	////	4	53		0
52	𝍸 ///	8	52	/	1
51	𝍸 𝍸 𝍸	15	51	//	2
50	𝍸 𝍸 𝍸 //	17	50	///	3
49	𝍸 𝍸 𝍸 𝍸 //	22	49	𝍸	5
48	𝍸 𝍸 𝍸 ///	18	48	𝍸 //	7
47	𝍸 𝍸	10	47	𝍸 𝍸 𝍸	15
46	𝍸 ////	9	46	𝍸 𝍸 𝍸 ///	18
45	𝍸 //	7	45	𝍸 𝍸 𝍸 𝍸 𝍸	25
44	𝍸	5	44	𝍸 𝍸 𝍸	15
43	///	3	43	𝍸 𝍸 ///	13
42	///	3	42	𝍸 ///	8
41	//	2	41	𝍸	5
40	/	1	40	////	4
39	//	2	39	//	2

2.5 FREQUENCY POLYGONS

The data presented in Table 2.2 as a frequency distribution can also be presented as a graph. Two kinds of graphs are commonly used to illustrate data derived from orderable scales. One of these graphs is called a frequency polygon and the other is called a histogram. When data are based on non-orderable countables (nominal scales), the preferred graph is a bar chart. Thus we might represent intelligence test scores by a frequency polygon or a histogram, but if we have tallied membership in several different political parties the data should be presented in the form of a bar chart.

The frequency polygons shown in Figures 2.3 and 2.3a, are based on the data in Table 2.2. In these polygons frequency is graphed as a function[1] of score. The value of each score is recorded on the horizontal axis, or *abscissa,* and the frequency of these values is recorded on the vertical axis, or *ordinate*. The points on the graphs are plotted directly above the midpoints of the intervals on the horizontal axis which represent the scores. You should also remember that the polygon is a closed figure; the ends meet the abscissa one full score unit above

[1]A function in mathematics implies a unique correspondence between x and y such as when $y = x^2$. One and only one value of y occurs for each value of x. For many behavioral scientists this term is used more loosely as when we say intelligence is a function of environment and heredity. This latter usage is most inclusive and is the meaning of the word "function" in this text.

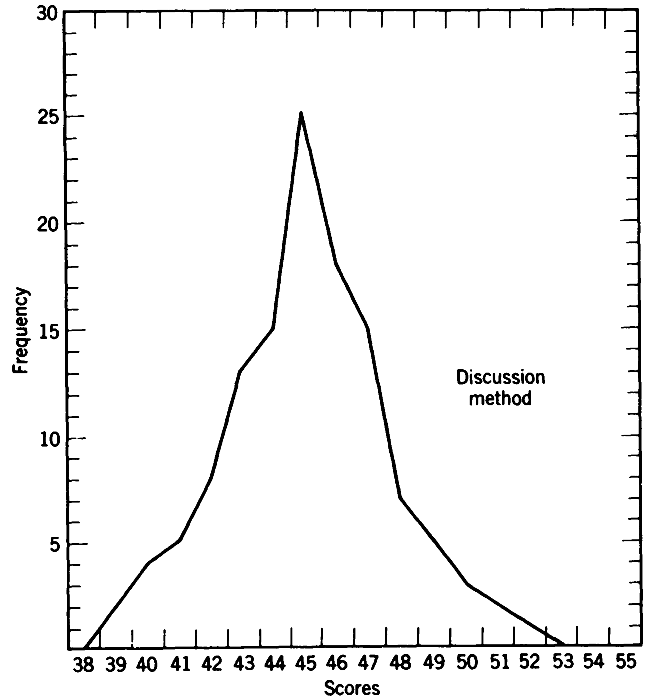

Figure 2.3 Frequency polygon of the examination scores tallied in Table 2.2.

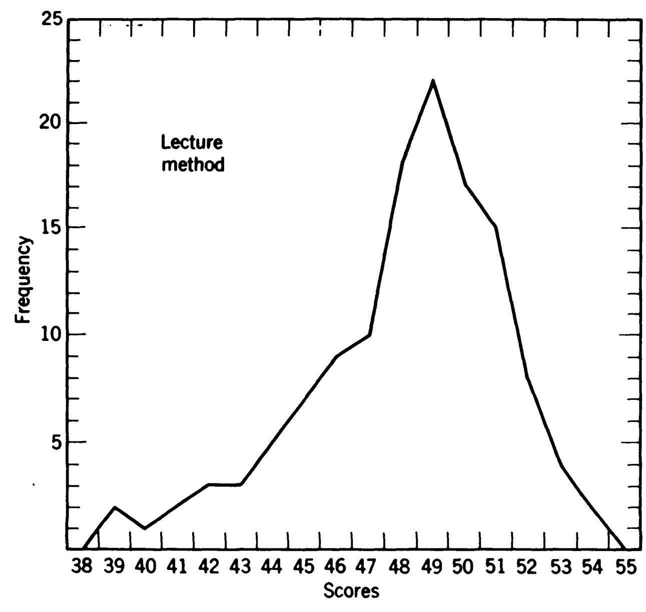

Figure 2.3a Frequency polygon of the examination scores tallied in Table 2.2.

the highest tabulated score and one full score unit below the lowest tabulated score.

2.6 HISTOGRAMS

Figures 2.4 and 2.4a illustrate the histogram. These are also plotted from the data of Table 2.2. The histogram consists of a series of adjacent bars whose heights represent the number of subjects obtaining a score and whose location on the abscissa represents the value of the score. Notice that the vertical lines marking off the bars do not originate from the center of the score interval but from its edges. The edges of the individual bars mark the theoretical limits of the score intervals along the abscissa.

Sometimes frequency polygons, or histograms of two different distributions, will both be plotted on the same set of coordinates. If the differences between the distributions are subtle, this procedure may highlight them. Whether one uses a frequency polygon or a histogram to represent data is largely a matter of personal preference.

2.7 BAR CHARTS

Bar charts are the preferred graphs when data are discrete, that is, when they result from the process of counting. This convention is somewhat fluid in psychology, where ordinal scales are concerned, but it should be followed without exception for nominally scaled data, that is, for nonorderable countables. The bar chart is very much like the histogram except that spaces are left between the bars in the bar chart. Bar charts sometimes use the vertical axis to represent categories and the horizontal axis to represent frequency of occurrence. Study the bar chart in Figure 2.5 where we have graphed the enrollment in introductory courses for science departmennts at a typical college.

2.8 GROUPED FREQUENCY DISTRIBUTIONS

We now consider a more complex kind of frequency distribution called a grouped frequency distribution, but first we call your attention to the approximate nature of all continuous measurements. A length may be measured to the nearest inch, the nearest tenth inch, or the nearest hundredth inch. Each of these measurements is considered to be accurate only within certain limits. The limits are one half of the unit of measure above and below the unit in question.

If we are measuring with a ruler on which an inch is the smallest subdivision, a measurement will be recorded only to the nearest inch. If an object is actually

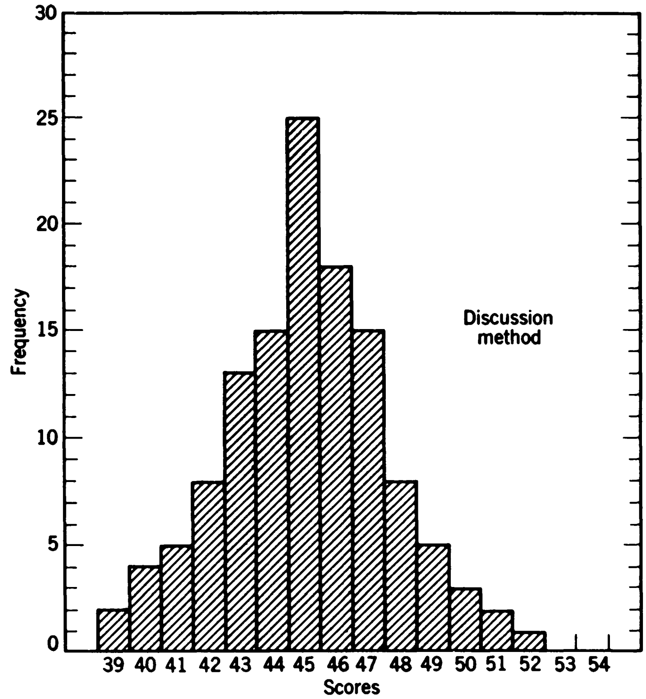

Figure 2.4 Histogram of the examination scores tallied in Table 2.2.

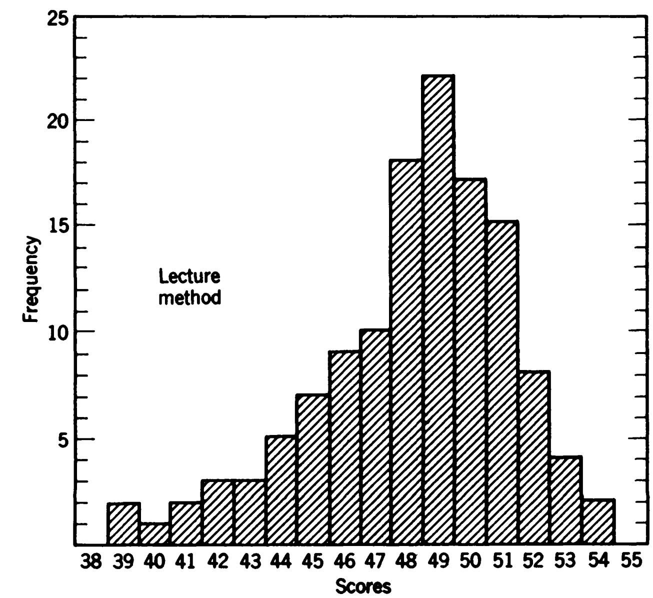

Figure 2.4a A histogram of the examination scores tallied in Table 2.2.

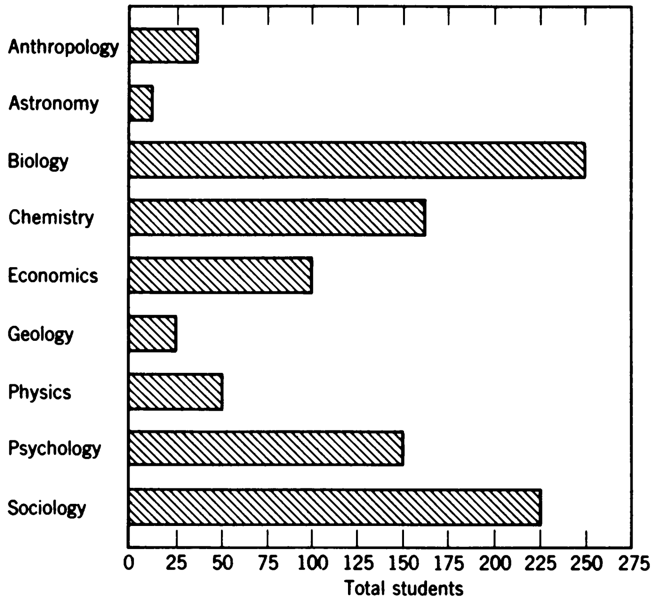

Figure 2.5 A bar chart of enrollment in introductory science courses.

7.25 inches long and another object 6.57 inches long, they will both be classified as 7 inches long because they are seen as closer to 7 than to either 6 or 8. That is the most accurate measurement we can make if we can read only to the nearest inch. Therefore, the measurements labeled "7 inches" really include all those from 6.5 to 7.5 inches. We may say that a measurement of 7 inches could really *extend* from 6.5 to (but not including) 7.5 inches, and that 6.5 and 7.5 are the *theoretical* limits of that unit, or measurement interval. This is the situation when the unit of measurement is an inch; the measuring device is presumed accurate only to within a half unit above and below the unit of measurement.

Similarly if the ruler is marked off in tenths of inches, a measurement of 6.8 inches will include measurements from 6.75 to 6.85 inches. If we can read hundredths of inches on our ruler, a measurement of 6.88 will include all lengths from 6.875 to 6.885. Even when we can measure to the hundredth of an inch, our accuracy is within plus or minus half the unit of measurement. In each case we think of the measurement as extending one-half unit above and below the recorded value, from one theoretical limit to the next.

This situation is assumed to exist whether we are measuring inches, seconds, or achievement. A score of 176 on an achievement test is assumed to have an accuracy extending one half unit to either side of the obtained score, from 175.5 to 176.5. A score of 176 is thought of occupying an interval from 175.5 to 176.5 just as a measure of 8 inches occupies the interval from 7.5 to 8.5 inches. Scores, then, are thought of as occupying intervals, and these intervals have theoretical limits. A score of 77 occupies the interval 76.5 to 77.5, and 76.5 is the lower theoretical limit of the interval while 77.5 is the upper theoretical limit of the interval. A score of 78 has a lower theoretical limit of 77.5 (the same as the upper limit of the next lower score interval) and an upper theoretical limit of 78.5. The midpoint of the score interval is the recorded value of the score itself. When we constructed a frequency polygon we made use of these ideas. Notice that the scale on the abscissa in Figure 2.3 is continuous, the upper theoretical limit of one interval coinciding with the lower theoretical limit of the next. The points forming the outline of the polygon were plotted above the midpoints of the intervals.

Let's return for a moment to Table 2.2 and apply these ideas to the data in just one interval of the frequency distribution for the lecture class. We shall assume that the 22 students in the lecture class who had scores of 49, might have had, with a more accurate measuring device, scores spread over the interval from 48.5 to 49.5. We will, therefore, assume that a score of 49 encompasses that distance, that it includes those hypothetically finer measurements from 48.5 to 49.5. We have assumed that the achievement measured is continuous even though our measuring device (the number of right answers) is not.

One should notice that the size of the unit of measurement is quite arbitrary and depends on the measuring device at hand. For example, we could transform the data of Table 2.2 by using a coarser grouping; this is analogous to the use of a less accurate scale of measurement. We could group together scores of 39, 40, and 41; scores of 42, 43, and 44; scores of 45, 46, and 47, *etc.*, throughout the distribution. This is a *grouped* frequency distribution. If we were to group together scores of 39, 40, and 41 into a new expanded interval, the new interval would extend from 38.5 to 41.5, and the midpoint of the interval would be 40.

We shall now distinguish between the score limits of a grouped interval and the theoretical limits of a grouped interval. The score limits of an interval are the highest and lowest scores within it which can actually be obtained. In the interval consisting of scores of 39, 40, and 41, the score limits are 39 and 41. The theoretical limits of the interval extend a half unit of measurement above the upper score limit and below the lower score limit. Thus, in the interval consisting of scores of 39, 40, and 41, the theoretical limits of the interval are 38.5 and 41.5. In the interval formed by grouping together scores of 42, 43, and 44, the score limits are 42 and 44; the theoretical limits are 41.5 and 44.5. If the lower *theoretical* limit is subtracted from the upper *theoretical* limit of any interval the grouping interval size (i) is obtained, which in this example is three. Three original intervals of one unit each are used to form each grouped interval. In any distribution, i is usually a constant; the grouping interval is the same size throughout the distribution.

The utility of a grouped frequency distribution is not apparent from the data in Table 2.2, but look at Table 2.3, which includes all of the scores from a real examination in introductory psychology. Suppose we had to make up a frequency

TABLE 2.3
Test Scores from a Final Examination in Introductory Psychology

86	74	66	63	58	54	51	45
85	74	66	62	58	54	51	45
84	73	66	62	57	53	50	45
84	73	66	62	57	53	50	45
84	73	66	62	57	53	50	45
84	72	66	61	57	53	49	43
83	72	66	61	56	53	49	43
82	72	65	61	56	53	49	43
82	71	65	61	56	53	49	43
80	71	65	61	56	53	49	42
79	71	64	61	56	53	48	41
79	70	64	61	55	53	48	41
78	70	64	61	55	53	48	41
78	70	64	60	55	53	47	41
78	67	64	60	55	52	47	39
77	67	64	60	55	52	47	38
76	67	64	58	55	52	47	38
76	67	64	58	54	52	46	37
76	67	63	58	54	52	46	
75	67	63	58	54	52	46	
75	67	63	58	54	51	46	
75	67	63	58	54	51	46	

distribution for these scores. Notice that the range of scores covers 50 *different* score magnitudes. An ungrouped frequency distribution would require tallies for 50 different intervals and, obviously, would be rather clumsy. We simplify our task considerably if we construct a grouped frequency distribution. The first decision involves a choice of size for the grouping interval. How many of the single unit intervals should be grouped together to form each new grouped interval?

You must decide how much grouped intervals will reduce work but at the same time preserve the essential configurations of the distribution. Statisticians normally use about 10 intervals when constructing frequency distributions, and i, the size of the group interval, should be chosen to obtain about that number. If we have 50 original intervals, grouping by fours or by fives ($i = 4$ or $i = 5$) will result in either 10 or 13 grouped intervals. It seems that either of these values for i will do, but a second consideration is to try to use an interval size that is either five or a multiple of ten. This is common practice because of our use of a decimal notation, but is by no means a requirement. Therefore we will let $i = 5$ and group together scores of 35, 36, 37, 38 and 39, and so on through 85, 86, 87, 88, 89. Note that we began grouping with a score which is a multiple of i. This is also simply a convenience for the reader.

One more important note about i, the size of the grouping interval: if we group together scores of 40–43, 44–47 *etc.* you may mistakenly calculate i = 3 by subtracting 40 from 43 or 44 from 47. To obtain i you must subtract the lower *theoretical* limit from the upper *theoretical* limit. Thus we have $i = 43.5 - 39.4 = 4$! Be careful. When you are calculating i from a grouped frequency distribution, use theoretical score limits.

Table 2.4 illustrates a grouped frequency distribution (and some other things to be discussed in the next section) based on the data in Table 2.3. This is certainly a clearer picture of the class performance than one can get from a listing of scores. Note that the length of the string of tallies in each interval is proportional to the frequency in that interval and an indication of what a histogram based on these data would look like.

Grouped frequency distributions are a convenient way of summarizing data. There are few firm rules for constructing them but one of these is that the adjacent grouping intervals cannot overlap. Intervals of 35–40, 40–45, 45–50 will not do because the interval into which scores of 40, 45 and so on are to be tallied is ambiguous. Grouping intervals must be mutually exclusive. They must be so

TABLE 2.4
Grouped Frequency, Cumulative Frequency, and Cumulative Proportion Distributions Based on Data of Table 2.3

Score Limits (i = 5)	*Tally*	*Theoretical Limits*	*Frequency*	*Cumulative Frequency*	*Cumulative Proportion*
		89.5		172	1.000
85–89	||		2		
		84.5		170	.988
80–84	𝍸 |||		8		
		79.5		152	.942
75–79	𝍸 𝍸 ||		12		
		74.5		150	.872
70–74	𝍸 𝍸 𝍸		14		
		69.5		136	.791
65–69	𝍸 𝍸 𝍸 |||		18		
		64.5		118	.686
60–64	𝍸 𝍸 𝍸 𝍸 𝍸 |||		28		
		59.5		90	.523
55–59	𝍸 𝍸 𝍸 𝍸 |||		23		
		54.5		67	.390
50–54	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 ||		32		
		49.5		35	.203
45–49	𝍸 𝍸 𝍸 𝍸 ||		22		
		44.5		13	.076
40–44	𝍸 |||		9		
		39.5		4	.023
35–39	𝍸		4		

constructed that a score can be tallied in only one interval. This rule is rigid. Beyond that you have considerable leeway to group the data so that the shape of the distribution will be clear to your reader.

Once data have been cast into a grouped frequency distribution one can then construct frequency polygons or histograms using the procedures described earlier. The abscissa is formed from the grouping intervals instead of from individual scores; otherwise the procedures are exactly the same.

2.9 CUMULATIVE DISTRIBUTIONS

Psychological data are sometimes usefully presented in the form of cumulative proportion graphs. Table 2.4 contains columns headed "Cumulative Frequency" and "Cumulative Proportion." The cumulative frequency column lists, opposite the upper theoretical limit of each interval, the cumulative frequency, or total, of all scores below that point. Consider the interval whose score limits are 45–49. Opposite the upper theoretical limit of that interval, 49.5, we have recorded a cumulative frequency of 35. The 35 was obtained by adding (cumulating) the frequency of all measurements below this upper theoretical limit; that is 22, 9, and 4. Opposite the upper theoretical limit 64.5 we have a cumulative total of 118. This total results from adding the frequencies in all of the intervals up to and including 60–64. At the top of the cumulative frequency column, opposite the upper theoretical limit of the highest interval, we will always find the total number of cases in the distribution (N).

The entries in the column headed *cumulative proportion* are obtained by dividing each cumulative frequency entry by N (the total number of measures) and listing this figure as a decimal. Thus the highest entry is 172/172 or 1.000. The entry opposite the upper theoretical limit of the interval whose score limits are 65–69 is .791. This means that the proportion of scores below 69.5 is .791 or about 79 percent. This proportion as obtained by adding the frequencies in all the intervals below this *theoretical* limit and then dividing the sum by N.

A cumulative frequency and cumulative proportion polygon is shown in Figure 2.6. This is constructed by plotting the cumulative frequency and cumulative proportion opposite the *upper theoretical limit* of the intervals listed along the abscissa. Note this subtle but critical difference between noncumulative and cumulative polygons; the plot of points for noncumulative polygons are always above the midpoints of intervals, for cumulative polygons they are always above the upper theoretical limits of the intervals.

Once the cumulative proportion graph has been drawn we can determine the score values which separate different proportions of the distribution. For example, if we want to obtain the score point which just separates the distribution into two equal halves, we locate the .50 point on the ordinate, construct a horizontal line until it intersects the graph, then drop a vertical line and read the

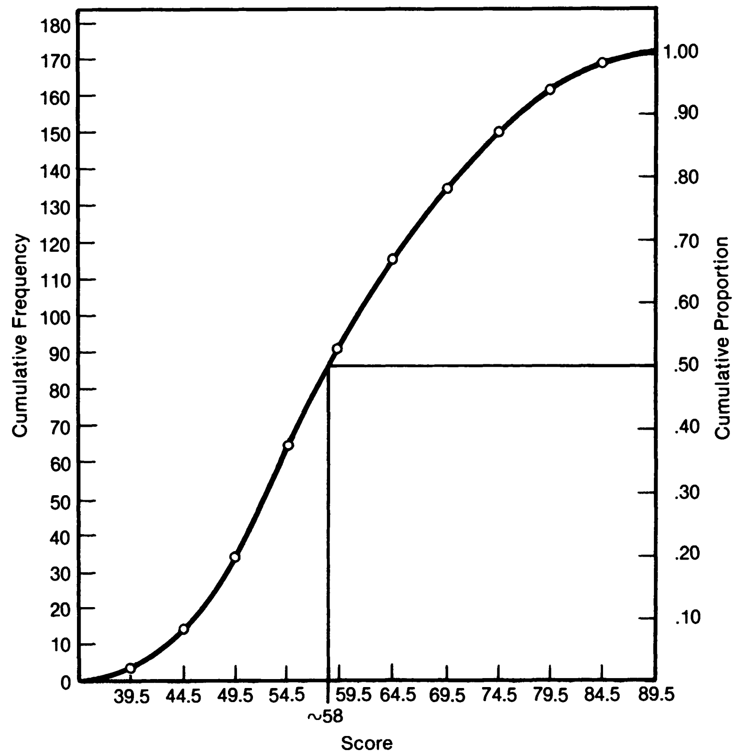

Figure 2.6 Cumulative frequency and cumulative proportion graph of data in Table 2.4.

score point from the abscissa. These lines have been drawn in Figure 2.6 and the score point is approximately 58. To check this answer, count up 172/2 (or 86 scores) from the lowest score recorded in Table 2.3; the value should be 58. An analagous procedure allows us to estimate the percentage of cases falling below any score value.

This discussion has not included enough material to make you an expert on tabulating and graphing data, but it should enable you to present a group of measurements in the form of a grouped frequency distribution, frequency polygon, histogram, cumulative frequency, or cumulative proportion graph.

2.10 REVIEW

Constructs are most usefully defined scientifically by specifying the procedures for their measurement. These operational definitions may result in nominal scales in which numerals are substituted for names and have no other quantitative significance; ordinal scales in which numerical increases signify increasing amounts of the construct but difference between consecutive numbers do not necessarily represent equal amounts of the variable; interval scales in which differences between consecutive numbers do represent equal amounts of the variable; ratio scales which, in addition to equal intervals between consecutive scale numbers, have a true zero point so that meaningful ratios can be formed.

When a fairly large number of observations have been recorded, the characteristics of the data are often made clearer if they are cast into a frequency distribution (Tables 2.2 and 2.4). In this form the different values are arranged from highest to lowest, and a tally (or count) of each value is recorded. Once this has been accomplished, graphs can be constructed. Graphs are typically constructed with the abscissa or horizontal axis, representing the values of the scores or measurements, and the ordinate representing their frequency. The histogram (Figure 2.4) is a type of graph constructed of adjacent vertical bars; the height of the bar represents the frequency with which the score occurred. The frequency polygon is a closed figure drawn by connecting points plotted above the midpoints of the score intervals which are located along the abscissa. The distance of the points above the abscissa represents the frequency of the score's occurrence.

A measurement is assumed to occupy an interval and this interval extends one-half unit of measurement above and below the score. These are called the theoretical limits of the score interval. When many different measurements are obtained, it is often convenient to group several adjacent score intervals together. Such grouping intervals have score limits determined by the highest and lowest scores in them, and theoretical limits determined by the theoretical limits of these same scores. The grouped intervals must be mutually exclusive. The preference is to use 5 or 10 of the original score intervals to form the new grouped intervals, and to choose a value for this interval size (i) which will result in enough new grouped intervals to accurately depict the distribution. Grouping should proceed so that the lowest score in an interval is an even multiple of i. Such a grouping of score units into intervals, and the tally of measurements within these separate intervals, yields a grouped frequency distribution.

Data may also be graphed in the form of a cumulative frequency curve or a cumulative proportion curve (Figure 2.6). These curves are obtained by representing the cumulative frequency of measurements below the upper theoretical limit of each interval on the ordinate, and representing the intervals on the abscissa. The maximum value on the ordinate of a cumulative frequency distribution will always equal N, the total number of observations. A cumulative proportion graph is the same as a cumulative frequency graph, except that each cumulative frequency is divided by N to obtain a proportion. These cumulative proportions are then plotted opposite the upper theoretical limit of the appropriate intervals.

2.11 EXERCISES

*1. Draw cumulative proportion graphs of the data in Table 2.2, and by means of a graphical solution find the score value at the midpoint of each distribution.[2]

2. (a) Using the conventions for grouped frequency distributions, construct a grouped frequency distribution for the data below.

11	17	22	26	30	33	36	40	48	57
12	18	23	27	30	33	37	41	49	58
13	18	23	27	30	34	37	43	50	58
14	19	24	28	31	34	38	44	51	59
15	20	25	29	31	35	38	46	53	60
16	21	25	29	31	35	39	46	55	61
16	21	25	30	32	36	39	47	56	61

(b) Construct a cumulative frequency polygon and by a graphical solution estimate the score below which 25%, 50%, and 75% of the data fall.

3. Give two operational definitions of each of the following.
(a) Sociability
(b) Creativity
(c) Leadership

4. Using $i = .005$, construct a grouped frequency distribution for the data below.

1.699	1.695	1.693	1.686	1.678	1.671
1.699	1.694	1.692	1.685	1.678	1.671
1.697	1.694	1.692	1.683	1.677	1.670
1.697	1.694	1.689	1.681	1.675	1.669
1.696	1.694	1.687	1.679	1.674	1.668
1.695	1.693	1.687	1.679	1.671	1.667

*5. What are the theoretical limits of each of the two lowest intervals in Exercise 3?

6. What scale of measurement is represented by the following:
(a) Numbers on football players' jerseys.
(b) Rank of naval officers.
*(c) Jail sentences given by a criminal court.

---

[2]Questions preceded by an asterisk are answered in the answer section at the end of the text.

(d) Scores on a statistics test.
*(e) Typing errors.

3 Measures of Central Tendency

Up to this point our concern has been with the use of graphs that picture the characteristics of a distribution of measurements. We can also use more exact methods and actually measure the characteristics of a distribution just as a score within a distribution measures some characteristic of an individual.

One characteristic of a distribution is its size. This is symbolized by N, the number of measurements that make up the distribution. If we have recorded the IQ scores of 100 students, $N = 100$. Figure 3.1 shows two distributions based on the intelligence test scores of students majoring in fields A and B. The two distributions pictured are based one equal numbers of subjects; therefore, $N_A = N_B$. This equality is reflected in the equal areas under the two curves. The area enclosed by a frequency polygon or histogram is directly proportional to the number of cases on which it is based. When frequency polygons and histograms are constructed, each case contributes an increase in the height of the figure opposite the appropriate score interval on the abscissa. Consequently, each measurement that goes into the distribution adds an increment to the area encompassed by the figure.

While the N's are equal and the areas are therefore the same, the distributions in Figure 3.1 differ in several ways. Although the scores overlap, there seems to be a tendency for students majoring in A to have lower IQ scores than students majoring in B. This difference between the distributions is reflected in their relative location on the abscissa or score continuum. Notice that the majority of scores in each distribution cluster, or "pile up," in a particular region. The location of this clustering or central tendency provides a useful method for comparing distributions. Central tendency can be measured in several ways.

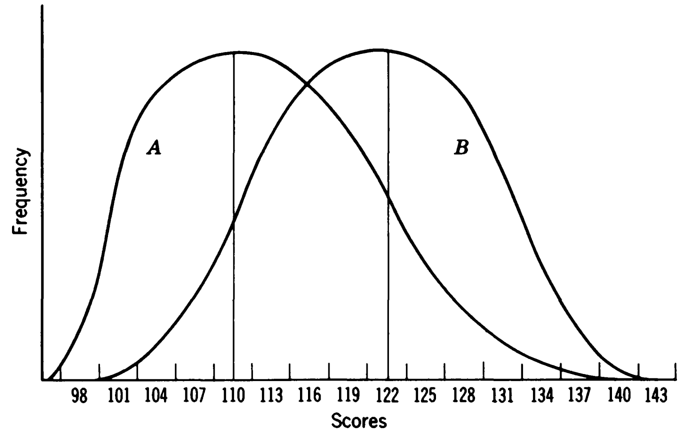

Figure 3.1 Distribution of intelligence test scores of students majoring in departments A and B.

3.1 MODE

One measure of central tendency is called the mode. The mode is the most frequent score in a distribution. In Table 2.2 the mode for the "Lecture Method" is 49, because that score occurs more often than any other. Similarly, the mode for the "Discussion Method" is 45. Notice that the mode is the *value* of the score with the highest frequency; it is not the frequency of that score.

When measures have been grouped into intervals, we call the interval with the highest frequency the modal interval. In Figure 3.1 the abscissa is labeled with the midpoints of the grouping intervals. The modal interval for each distribution can be determined from inspection. Frequency is represented on the ordinate, so the modal interval will fall directly below the highest point of the curve. The modal interval of distribution A is 109–111, with a midpoint of 110; the modal interval of distribution B is 121–123, with a midpoint of 122.

Some distributions have two modes, which are not necessarily intervals with equal frequencies (f), but separated intervals, each with higher frequencies than the intervals adjacent to them. These are called bimodal distributions. When a distribution is bimodal the measurement (or measurement interval) with the greater frequency is called the major mode or major modal interval, and the measurement (or measurement interval) with the second greatest frequency is called the minor mode or minor modal interval.

Bimodal distributions usually result from including two different kinds of subjects in the same distribution. If we make up a single frequency distribution from the intelligence test scores of freshmen and first year medical students at a university, we will probably obtain a bimodal distribution such as that in

Figure 3.2. The major mode, contributed by the large number of freshmen, is lower on the score continuum than the minor mode contributed by the smaller number of generally brighter medical students.

3.2 RULES OF SUMMATION

This section, rules of summation, reviews some basic algebra and introduces some statistical notation, both necessary for your understanding of the concepts to follow.

Four different rules of summation are applied when we sum, or add, a series of terms. These rules all follow directly from elementary algebra, but they involve a new symbol for addition: the summation sign (Σ), which is a capital Greek sigma. This sign directs us to sum all the measures $X_1, X_2, X_3, X_4 \ldots X_N$. Each of these measures is called a variate, and the dimension measured is called a variable. If X stands for the variable IQ, and if we have five variates, the IQ scores of five individuals, so that $X_1 = 100, X_2 = 100, X_3 = 105, X_4 = 110$, and $X_5 = 120$, then $\Sigma X = X_1 + X_2 + X_3 + X_4 + X_5 = 535$. If Y represents running time in a maze for N rats, then $\Sigma Y = Y_1 + Y_2 + Y_3 + \ldots + Y_N$.

There are occasions when we may wish to sum only a part of a series, for example, the first 10 members. If this is the case, we can place limits on the summation sign thus $\sum_{i=1}^{i=10} X_i$. This means that we are to sum the first ten values of X, from $X_{i=1}$ to $X_{i=10}$. If we wished to be very formal about it we should show the limits of summation even when all N members of a series are to be summed. Thus, $\sum_{i=1}^{N} X_i$ instructs us to sum all N variates of the series from $X_{i=1}$ to X_N. In this text we will ordinarily require the summation of all members of

Figure 3.2 A bimodal distribution.

any series, so it will be understood that when ΣX or ΣY appear they require the summation of all variates, unless otherwise noted.

The first rule of summation we shall discuss concerns the summation of a constant. A constant, by definition, can have only one value. In the formula for the area of a circle, $A = \pi r^2$, A and r symbolize variables while π symbolizes a constant. If we have a constant, C, repeated N times, and we wish to sum these N instances of the constant, we can find the sum two ways; by adding the N instances of C, or by multiplying N times C. If these equivalent methods are stated in statistical notation we have $C + C + C + \ldots + C = \Sigma C = NC$. *The sum of a constant equals N times the constant.*

The second rule of summation is used when we sum a series in which each term consists of a constant multiplied by a variate. The series might consist of the terms $CX_1 + CX_2 + CX_3 + \ldots + CX_N$. Of course we can replace the plus signs by rewriting the series as ΣCX, but a further simplification is possible. By factoring the constant C from the series we have $C(X_1 + X_2 + X_3 + \ldots X_N)$. The terms within the parenthesis can now be replaced by ΣX and the entire expression becomes $C \Sigma X$. Thus, $CX_1 + CX_2 + CX_3 + \ldots + CX_N = \Sigma CX = C\Sigma X$. *The sum of a constant times a variable equals the constant times the sum of the variable.*

The third rule applies to the situation in which the series is composed of terms which are themselves the sum of scores on two or more variables. The series might consist of the terms $(X_1 + Y_1) + (X_2 + Y_2) + (X_3 + Y_3) + \ldots + (X_N + Y_N)$. This series of terms could be regrouped so that it would appear as $(X_1 + X_2 + X_3 + \ldots + X_N) + (Y_1 + Y_2 + Y_3 + \ldots + Y_N)$. The summed series within the first parenthesis may be symbolized by ΣX, and the summed series within the second parenthesis by ΣY, so the sum of these two series may be given as $\Sigma X + \Sigma Y$. Therefore, $\Sigma(X + Y) = \Sigma X + \Sigma Y$. *All this means is that we can sum or add numbers in any order we please and the result will be the same.*

The fourth rule concerns the summation of a series in which each term is composed of a variate plus a constant. Such a series might consist of $(X_1 + C) + (X_2 + C) + (X_3 + C) + \ldots + (X_N + C)$. We can regroup these additions in the form $(X_1 + X_2 + X_3 + \ldots + X_N) + (C + C + C + \ldots + C)$. The terms in the first parenthesis are given by ΣX and those in the second as ΣC, or NC. Consequently, the sum of the entire series may be written as $\Sigma X + NC$. Thus, $(X_1 + C) + (X_2 + C) + (X_3 + C) + \ldots + (X_N + C) = \Sigma X + \Sigma C = \Sigma X + NC$. *The sum of a variable plus a constant is equal to the sum of the variable plus N times the constant.*

In each of the six examples below be sure you understand why the terms to the left of the equality are equivalent to the terms on the right. Each example illustrates one or more of the rules we have just discussed and in each example N and C are constants.

1. $\Sigma(X + 6C) = \Sigma X + 6NC$
2. $\Sigma(X + Y + C) = \Sigma X + \Sigma Y + NC$
3. $\Sigma(X/C + Y/C) = 1/C(\Sigma X + \Sigma Y)$
4. $\Sigma(X + C)(X - C) = \Sigma(X^2 - C^2)$
   $= \Sigma X^2 - NC^2$
5. $\Sigma(X + Y)^2 = \Sigma(X^2 + 2XY + Y^2)$
   $= \Sigma X^2 + 2\Sigma XY + \Sigma Y^2$
6. $\Sigma(X + C)^2 = \Sigma(X^2 + 2CX + C^2)$
   $= \Sigma X^2 + 2C\Sigma X + NC^2$
7. $\Sigma(X + CY)^2 = \Sigma X^2 + 2C\Sigma XY + C^2\Sigma Y^2$

3.3 THE MEAN

Now that you are familiar with these rules of summation we can discuss the mean, another very widely used measure of central tendency. The mean is determined by summing all of the measures in a distribution and then dividing this sum by the number of measures.

If we let $X_1, X_2, \ldots X_N$ symbolize the various IQ scores in a distribution, then the formula for the mean of this distribution of IQs becomes:

$$\mu = \frac{\Sigma X}{N}$$ Formula 3.1 The mean[1]

This is a general formula; it will give us the mean of a distribution regardless of the kinds of measures represented by the Xs. If we calculate the mean of a distribution consisting of the scores 8, 7, 6, and 5, the mean will be 6.5, that is: $\Sigma X = 26$, $N = 4$, and $\mu = 26/4 = 6.5$.

If data are in the form of a frequency distribution, some short cuts are possible. Remember that a frequency distribution consists of a list of scores with their frequencies. We can find the sum of the scores by finding $\Sigma(f_1X_1 + f_2X_2 + f_3X_3 + \ldots + f_NX_N)$ where f is the frequency of each score and X is the magnitude of the score. We can rewrite this expression ΣfX. Note that it is *not* equivalent to $f\Sigma X$ because f is not a constant; f will probably have a different value for each different score magnitude. When ΣfX is divided by N we have a formula for the mean of the data which have been cast in a frequency distribution. This formula allows us to make fewer calculator entries by doing a little multiplying. The formulas $\mu = \Sigma X/N$ and $\mu = \Sigma fX/N$ are exactly equivalent.

---

[1]When statisticians refer to the mean of a population they ordinarily designate it by the Greek letter μ (pronounced mū), or perhaps M. Sample means, however, are usually symbolized by $\bar{X}$. The methods we describe in this chapter and the next for determining central tendency and variability assume that we are dealing with *populations*. In Chapter 7 we distinguish between samples and populations, and discuss the estimation of population parameters.