Anne L. Rothstein

RESEARCH DESIGN & STATISTICS FOR PHYSICAL EDUCATION

RESEARCH DESIGN AND STATISTICS FOR PHYSICAL EDUCATION

Anne L. Rothstein

Herbert H. Lehman College
City University of New York

PRENTICE-HALL, INC., ENGLEWOOD CLIFFS, NEW JERSEY 07632

Library of Congress Cataloging in Publication Data

ROTHSTEIN, ANNE L.
Research design and statistics for physical education.

Bibliography: p.
Includes index.
1. Physical education and training—Statistical methods. 2. Statistics. I. Title.
GV362.R67 1985 613.7'028 84-15014
ISBN 0-13-774142-1

Editorial/production supervision and
interior design: Dee Josephson
Cover design: Ben Santora
Manufacturing buyer: Harry Baisley

Printed in the United Staes of America

10 9 8 7 6 5 4 3

ISBN 0-13-774142-1 01

Prentice-Hall International, Inc., *London*
Prentice-Hall of Australia Pty. Limited, *Sydney*
Editora Prentice-Hall do Brasil, Ltda., *Rio de Janeiro*
Prentice-Hall Canada Inc., *Toronto*
Prentice-Hall of India Private Limited, *New Delhi*
Prentice-Hall of Japan, Inc., *Tokyo*
Prentice-Hall of Southeast Asia Pte. Ltd., *Singapore*
Whitehall Books Limited, *Wellington, New Zealand*

This book is dedicated to my father, William Rothstein, and to my brother, Harold Rothstein, with love and appreciation. Most especially it is dedicated to the memory of my mother, Rose Rothstein.

PREFACE

This book is intended as a generic text on research design and statistics which incorporates examples from physical education and sport. It emphasizes application, learning by doing, and learning through example. It is intended for use in connection with formal study and by individuals with minimal knowledge of statistics who require a resource book in this area. It focuses on the thinking, reasoning, and problem-solving nature of research design and statistics, and it views research as a way of looking at the world, and as a mindset that can be learned.

The text focuses primarily on how rather than why. It includes a wide variety of statistical tests, indicates how basic factorial designs can be extended for additional factors and for unequal groups, provides annotated examples of microcomputer and statistical package programs, facilitates the statistical decision-making process by presenting and incorporating a logical, step-by-step approach, and includes an introduction to meta-analysis, that is, statistical synthesis of research findings for the purpose of application.

Chapter 1 provides an overview and preliminary example and an introduction to research design and statistics. Chapters 2 and 3 introduce concepts about measurement and descriptive statistics. Chapters 4 through 8 cover topics in research design: Chapter 4 clarifies the research process and Chapter 5 provides an overview of the experimental study. The other chapters in this section, 6, 7, and 8, present concepts of experimental design as they relate to control, validity, and sampling. Chapter 9 explains statistical inference and uses computer-generated tables to illustrate the concepts of statistical inference. Chapter 10 focuses on probability and the normal curve and uses computer-generated examples and extensive diagrams to facilitate student understanding of the concepts presented. All of these are particularly useful to the beginning researcher.

Chapters 11 through 19 present the statistical tests. Each chapter provides a conceptual framework for the test, includes a detailed, step-by-step example, uses the decision-making process outlined in Chapter 9, and contains additional exercises. Answers to all exercises are provided in Appendix D. An important and unique feature of the book is the inclusion, in Appendix B, of annotated microcomputer programs for most of the tests presented in the book and the presentation of annotated input and output for three commonly used statistical packages, BMDP, SAS, and SPSS. These programs can be used to check answers to the exercises in the book or to analyze data. In Chapter 16, Analysis of Factorial Designs, there is an easily understood introduction to the concept of analysis of variance; Chapter 17 contains sections on interaction and extension of the methods presented to other, more complicated designs. Chapter 19 focuses on follow-up tests, and covers both individual comparisons and simple effects.

It is important to note that in selecting or designing the studies to be used as examples or exercises in this textbook the goal was simplicity rather than perfection. Thus there was no attempt to consider all the complex problems involved in research design for each example. Rather, the goal was to use examples and exercises that beginning researchers might relate to most readily.

Chapters 20 and 21, respectively, cover meta-analysis and use of computers. The chapter on meta-analysis takes the approach that research can influence practice if appropriate synthesis methods are used. A variety of techniques, which come under the general heading of meta-analysis, are described and examples from physical education and related areas are presented. An introduction to principles and procedures is included and the interested reader could perform a meta-analysis based on the information provided. A chapter on computers is viewed as essential in a text on research design and statistics because microcomputers have facilitated applications that were not feasible several years ago and promise to continue to affect the conduct of research and statistical analysis in important ways. Chapter 21 provides an easily understood introduction to computers and computer applications. A general section on computers includes information on the use of computers in several areas within physical education and sport. Annotated research and statistical analysis programs are provided in Appendix B and are available on disk for several microcomputers. Annotated input and output for several commonly used statistical packages is also presented.

Input from and assistance of others has been essential in the initiation and completion of this book, although the final responsibility for the material contained here is my own. I must thank A. M. Gentile, whose intuitive grasp of statistics and method of sharing her knowledge and insights with her students provided the foundation upon which this book is structured. Ree Arnold, a good friend and colleague, has contributed to this book through our planning for workshops and our discussions about research, statistics, statistical analysis, and computers. Preliminary versions of this text have been field-tested through the efforts of a former student who is now a good friend and colleague, Emily Wughalter, who provided valuable assistance with the SAS statistical examples in Appendix B and also some interesting research design and statistics questions and answers. My many senior honors students and graduate students at Lehman College and my friends and colleagues completing theses and dissertations presented provocative research problems that contributed to my thinking and also provided some of the examples used in this text. The graduate and undergraduate students in my research classes, who challenged me to make difficult concepts clear, and who used and commented on early drafts of this book, deserve thanks. Tina Caraszi, who assisted with the typing of manuscript and tables, played an important role in the completion of this project.

I particularly appreciated the time and efforts of Jo Safrit, whose thorough review of the manuscript led to many worthwhile changes and additions. Dee Josephson, production editor, has played an important role in the completion of this book and I appreciate her assistance, concern, and support. To my close friends, who understood why I often had to work instead of play, and to Jan Anderheggen, whose support helped me to finish this project, go my appreciation and thanks. To Carol Wolfgram, whose encouragement, confidence, and support have been unflagging, go may deepest appreciation and gratitude.

A.L.R.

CONTENTS

CHAPTER ONE RESEARCH DESIGN AND STATISTICS: AN OVERVIEW

This book will approach research design and statistics with an emphasis on doing and finding out and with the view that statistics can be enjoyable. Statistics and research design can help us to explain observations about teaching, learning, and performance in physical education and sport. They can help us to answer questions about physical education. At the very least, familiarity with these topics can help us to understand the ways in which statistics are used by the media to influence us.

The notion that statistics is a form of higher mathematics is a myth. In fact, success at statistics takes little more than the ability to add, subtract, divide, and multiply in logical fashion. The most complex formula in statistics can be reduced to a series of logical operations involving these four procedures.

Research design and statistics involves the study of certain tools that enable us to systematically investigate, describe, and interpret phenomena in physical education. Statistics is one of the tools. Research design is another. Measurement techniques is the third. Each of these tools thus represents a way of finding out about sport and physical education. You have had or will have courses in kinesiology, exercise physiology, motor learning, sport psychology, and/or sport sociology. Associated with each of these subjects are facts and theories that elucidate behavior in sport. The facts and theories were derived through the use of measurement, research, and statistics tools.

Further, you use these tools every day. Much of the information you receive is based upon measurement, research, and statistics. Among the common uses of these tools are baseball statistics, grade-point average, gross national product, Gallup polls, fitness testing, the Census, and the cost-of-living index. In addition, you are bombarded with claims based upon "research and statistics" every day in advertising. "Nine out of ten dentists who recommend gum, recommend Trident." How many dentists did they question to find ten who recommended chewing gum to their patients? "The group that used *X* toothpaste had 47% fewer cavities." Fewer than what? What percent fewer cavities did the other group have? When you get into your car in the morning and it won't start, you consider alternative reasons—maybe the battery is dead, no gas, bad starter—and systematically eliminate each possibility until you discover the cause. You have done research. You have "guessed" at the cause and have tested your "guess." When you discover that you are correct, you act.

Before continuing, let us briefly define each of the terms we have used.

> **Measurement** A numerical score derived from an observation which generally reflects the extent to which an individual or object possesses a particular characteristic.
>
> **Research** Systematic procedures for observing and recording information about phenomena. These procedures are detailed in such a way that anyone could carry out the same observations and obtain the same information.
>
> **Statistics** A single numerical score which characterizes an entire set of scores. Thus, if you measure the height of a group of individuals and take the average height, the average height is a statistic which characterizes the entire group of people you measured.

So, research is a plan for systematically observing some phenomenon which you will measure. Research dictates that you measure one or more individuals and that, in some cases, you compare the scores of one person or group to those of another person or group. (Research involving one person is called a *case study*.) In these instances you must calculate statistics—representative scores—for each group. A simple example which demonstrates the use of measurement, research, and statistics should be helpful.

A PRELIMINARY EXAMPLE

Teacher A at B high school designs a new way to teach the tennis serve. The teacher believes this new method is much better than the old one but wishes to subject it to a test. The teacher decides that the Hewitt Tennis Achievement Test (Hewitt, 1966) will be a good way to measure serve performance. So the measure of tennis performance is the score the student obtains on the test. Having taken care of the problem of measurement of tennis serve performance, the teacher turns attention to the research design. In order to have a fair test, students who are equal in serve performance (possibly all beginners) should be randomly assigned to either the *traditional* serve instruction group or the *new* serve instruction group. Having made this random assignment, the teacher instructs both groups. After the instruction period and the practice period are over (these periods can be as long as the teacher wishes but must be equal for both groups), the teacher administers the Hewitt Tennis Achievement Test and records the scores of each student. Finally, using the scores of the students in each group, the teacher calculates statistics. The one you are most familiar with is the *average* (statisticians like to call this the *mean*). If the mean score for the students in the traditional group is lower than that for the students in the new group, the teacher may conclude that the new method is better and may use it in the future instead of the traditional method.

In testing the effectiveness of the traditional versus the new method of teaching the tennis serve, the teacher did the following:

Identified the problem. (How can I teach the tennis serve more effectively?)

Reviewed available information and past experience with teaching the tennis serve.

Decided that the *new* method was better than the *traditional* method.

Designed a research study to test the *new* method against the *traditional* method. (This included selecting a method of measurement.)

Conducted the study according to previously chosen systematic procedures to assure a fair test.

Gathered scores on the test of serve performance.

Analyzed the scores. (By calculating the average for each group.)

Reached a conclusion about which was the better method.

In fact, the analysis of the scores or measures obtained would normally have a more stringent basis than just the average scores, and in the third section of the book we will consider other ways of evaluating research studies. For now the average is the easiest statistic to calculate.

So, we see in the example that research, statistics, and measurement are intertwined and that together they enable us to answer questions about sport and physical education. These questions may center on the acquisition of skill (motor learning), the behavior of individuals in group settings (sport sociology), the characteristics of individuals (sport psychology), attaining and maintaining physical fitness (exercise physiology), efficiency of movement (kinesiology), prevention and treatment of injury (sports medicine), design and implementation of learning experiences (curriculum), or the clarification and interpretation of past events (sports history). The design and conduct of experimental research will be considered in the second section of this book.

TYPES OF STATISTICS

Traditionally the study of statistics has been divided into two parts, descriptive statistics and inferential statistics. *Descriptive statistics* is concerned with characterizing a particular set of observations which have been made on things or people. We may describe the average height of college seniors, the frequency of incomplete grades in various academic departments, the frequency of males and females in different majors, the median income of college professors, or the variability of ages of college students. In describing the items indicated we might use numerical, tabular, or graphic devices. However, the purpose of descriptive statistics is to summarize information and observations in a simple and meaningful way. Chapter 3 discusses the mean (average score), median (middle score), and mode (most frequent score) as well as the standard deviation, the range, and the interquartile range. These are all examples of descriptive statistics. Each statistic tells us something important about the group of scores it describes; in combination these statistics are very powerful, but we must know how to interpret their meaning.

Descriptive statistics provide information to characterize groups of measures, but in research we need more. The researcher must be able to generalize the scores of a small group of individuals (a sample) to the larger group from which they come (the population). *Inferential statistics* provides the tools to make these generalizations. The researcher, typically, wants to make predictions about a large group of individuals by studying only a small sample of the total group. The purpose of inferential statistics is to predict or estimate characteristics of a population from a knowledge of only a sample of the population. This can be accomplished with a known margin of error; that is, the researcher knows what the probability is that the estimate or prediction is wrong.

Inferential statistics always involves:

1. Statement of a guess about what will happen if you do something to a group of individuals.
2. Collection of data from a sample of the population to be studied.
3. Analysis of the results using some inferential statistical test.

The second section of this book (Chapters 4 to 8), which is concerned with research techniques and methods, details what happens in steps 1 and 2. The third section (Chapters 9 to 19) focuses on the various statistical tests you will use in step 3 to draw inferences from a sample about a population. A final section, Chapters 20 and 21, considers two special topics in research, meta-analysis and computers.

CHAPTER TWO MEASUREMENT

A measurement is a numerical score derived from observation of a behavior or characteristic. A behavior or characteristic that can take different values is called a *variable*. If, for example, in your exercise physiology class each student takes the Harvard Step Test to determine cardiovascular endurance, then we say we have data on the variable cardiovascular endurance. If we determine the eye color of each person in this class, then we have data on the variable eye color. ("Data" is just a fancy name for a set of measurements.) In similar manner we can obtain data on the variables of motivation, social status, class in school, college major, height, weight, tennis ability, strength, endurance.

LEVELS OF MEASUREMENT

In measurement, we derive numerical scores by assigning numbers to observations according to preset rules. When your performance on the Harvard Step Test is measured, you are assigned a number which reflects your cardiovascular endurance. Your scoring in a tennis game is guided by rules. All measurement is (or should be) guided by explicit or implicit rules.

Different rules apply at different levels of measurement. There are four such levels, each useful in its own way and each dictating certain types of statistical procedures. These levels are nominal, ordinal, interval, and ratio.

Nominal Measurement

Nominal measurement may be thought of as measurement "in name only." At this, the lowest level of measurement, people or objects are categorized by