Testing Statistical Hypotheses

Second Edition

E. L. LEHMANN

JOHN WILEY & SONS
World Publishing Corporation

Testing Statistical Hypotheses

Second Edition

E. L. LEHMANN
Professor of Statistics
University of California, Berkeley

JOHN WILEY & SONS
World Publishing Corporation

Authorized Reprint of the Original Edition,
Published by John Wiley & Sons, Inc.
This special Reprint Edition is Licensed for sale in China excluding Taiwan Province of China, Hong Kong & Macao
Reprinted by World Publishing Corp. Beijing, 1990
ISBN 7-5062-0506-8

Library of Congress Cataloging-in-Publication Data:

Lehmann, E. L. (Erich Leo), 1917–
Testing statistical hypotheses.

Includes bibliographies and indexes.
1. Statistical hypothesis testing. I. Title.

QA277.L425 1986 519.5'6 85-29469
ISBN 0-471-84083-1

Preface

This new edition reflects the development of the field of hypothesis testing since the original book was published 27 years ago, but the basic structure has been retained. In particular, optimality considerations continue to provide the organizing principle. However, they are now tempered by a much stronger emphasis on the robustness properties of the resulting procedures. Other topics that receive greater attention than in the first edition are confidence intervals (which for technical reasons fit better here than in the companion volume on estimation, *TPE**), simultaneous inference procedures (which have become an important part of statistical methodology), and admissibility. A major criticism that has been leveled against the theory presented here relates to the choice of the reference set with respect to which performance is to be evaluated. A new chapter on conditional inference at the end of the book discusses some of the issues raised by this concern.

In order to accommodate the wealth of new results that have become available concerning the core material, it was necessary to impose some limitations. The most important omission is an adequate treatment of asymptotic optimality paralleling that given for estimation in *TPE*. Since the corresponding theory for testing is less satisfactory and would have required too much space, the earlier rather perfunctory treatment has been retained. Three sections of the first edition were devoted to sequential analysis. They are outdated and have been deleted, since it was not possible to do justice to the extensive and technically demanding expansion of this area. This is consistent with the decision not to include the theory of optimal experimental design. Together with sequential analysis and survey sampling, this topic should be treated in a separate book. Finally, although there is a section on Bayesian confidence intervals, Bayesian approaches to

**Theory of Point Estimation* [Lehmann (1983)].

hypothesis testing are not discussed, since they play a less well-defined role here than do the corresponding techniques in estimation.

In addition to the major changes, many new comments and references have been included, numerous errors corrected, and some gaps filled. I am greatly indebted to Peter Bickel, John Pratt, and Fritz Scholz, who furnished me with lists of errors and improvements, and to Maryse Loranger and Carl Schaper who each read several chapters of the manuscript. For additional comments I should like to thank Jim Berger, Colin Blyth, Herbert Eisenberg, Jaap Fabius, Roger Farrell, Thomas Ferguson, Irving Glick, Jan Hemelrijk, Wassily Hoeffding, Kumar Jogdeo, the late Jack Kiefer, Olaf Krafft, William Kruskal, John Marden, John Rayner, Richard Savage, Robert Wijsman, and the many colleagues and students who made contributions of which I no longer have a record.

Another indebtedness I should like to acknowledge is to a number of books whose publication considerably eased the task of updating. Above all, there is the encyclopedic three-volume treatise by Kendall and Stuart, of which I consulted particularly the second volume, fourth edition (1979) innumerable times. The books by Ferguson (1967), Cox and Hinkley (1974), and Berger (1980) also were a great help. In the first edition, I provided references to tables and charts that were needed for the application of the tests whose theory was developed in the book. This has become less important in view of the four-volume work by Johnson and Kotz: *Distributions in Statistics* (1969–1972). Frequently I now simply refer to the appropriate chapter of this reference work.

There are two more books to which I must refer:

A complete set of solutions to the problems of the first edition was published as *Testing Statistical Hypotheses: Worked Solutions*. [Kallenberg et al. (1984)]. I am grateful to the group of Dutch authors for undertaking this labor and for furnishing me with a list of errors and corrections regarding both the statements of the problems and the hints to their solutions.

The other book is my *Theory of Point Estimation* [Lehmann (1983)], which combines with the present volume to provide a unified treatment of the classical theories of testing and estimation, both by confidence intervals and by point estimates. The two are independent of each other, but cross references indicate additional information on a given topic provided by the other book. Suggestions for ways in which the two books can be used to teach different courses are given in comments for instructors following this preface.

I owe very special thanks to two people. My wife, Juliet Shaffer, critically read the new sections and gave advice on many other points. Wei Yin Loh read an early version of the whole manuscript and checked many of the new problems. In addition, he joined me in the arduous task of reading the complete galley proofs. As a result, many errors and oversights were corrected.

The research required for this second edition was supported in part by the National Science Foundation, and I am grateful for the Foundation's continued support of my work. Finally, I should like to thank Linda Tiffany, who converted many illegible pages into beautifully typed ones.

REFERENCES

Berger, J. O.
(1980). *Statistical Decision Theory*, Springer, New York.

Cox, D. R. and Hinkley, D. V.
(1974). *Theoretical Statistics*, Chapman & Hall.

Ferguson, T. S.
(1967). *Mathematical Statistics*, Academic, New York.

Johnson, N. L. and Kotz, S.
(1969–1972). *Distributions in Statistics*, 4 vols., Wiley, New York.

Kallenberg, W. C. M., et al.
(1984). *Testing Statistical Hypotheses: Worked Solutions*, Centrum voor Wiskunde en Informatica, Amsterdam.

Kendall, M. G. and Stuart, A.
(1977, 1979). *The Advanced Theory of Statistics*, 4th ed., vols. 1, 2, Charles Griffin, London.

Kendall, M. G., Stuart, A., and Ord, J. K.
(1983). *The Advanced Theory of Statistics*, 4th ed., vol. 3, Charles Griffin, London.

Lehmann, E. L.
(1983). *Theory of Point Estimation*, Wiley, New York.

E. L. LEHMANN

Berkeley, California
February 1986

Preface to the First Edition

A mathematical theory of hypothesis testing in which tests are derived as solutions of clearly stated optimum problems was developed by Neyman and Pearson in the 1930s and since then has been considerably extended. The purpose of the present book is to give a systematic account of this theory and of the closely related theory of confidence sets, together with their principal applications. These include the standard one- and two-sample problems concerning normal, binomial, and Poisson distributions; some aspects of the analysis of variance and of regression analysis (linear hypothesis); certain multivariate and sequential problems. There is also an introduction to nonparametric tests, although here the theoretical approach has not yet been fully developed. One large area of methodology, the class of methods based on large-sample considerations, in particular χ^2 and likelihood-ratio tests, essentially has been omitted because the approach and the mathematical tools used are so different that an adequate treatment would require a separate volume. The theory of these tests is only briefly indicated at the end of Chapter 7.

At present the theory of hypothesis testing is undergoing important changes in at least two directions. One of these stems from the realization that the standard formulation constitutes a serious oversimplification of the problem. The theory is therefore being reexamined from the point of view of Wald's statistical decision functions. Although these investigations throw new light on the classical theory, they essentially confirm its findings. I have retained the Neyman–Pearson formulation in the main part of this book, but have included a discussion of the concepts of general decision theory in Chapter 1 to provide a basis for giving a broader justification of some of the results. It also serves as a background for the development of the theories of hypothesis testing and confidence sets.

Of much greater importance is the fact that many of the problems, which traditionally have been formulated in terms of hypothesis testing, are in reality multiple decision problems involving a choice between several decisions when the hypothesis is rejected. The development of suitable procedures for such problems is at present one of the most important tasks of statistics and is finding much attention in the current literature. However, since most of the work so far has been tentative, I have preferred to present the traditional tests even in cases in which the majority of the applications appear to call for a more elaborate procedure, adding only a warning regarding the limitations of this approach. Actually, it seems likely that the tests will remain useful because of their simplicity even when a more complete theory of multiple decision methods is available.

The natural mathematical framework for a systematic treatment of hypothesis testing is the theory of measure in abstract spaces. Since introductory courses in real variables or measure theory frequently present only Lebesgue measure, a brief orientation with regard to the abstract theory is given in Sections 1 and 2 of Chapter 2. Actually, much of the book can be read without knowledge of measure theory if the symbol $\int p(x)\,d\mu(x)$ is interpreted as meaning either $\int p(x)\,dx$ or $\Sigma p(x)$, and if the measure-theoretic aspects of certain proofs together with all occurrences of the letters a.e. (almost everywhere) are ignored. With respect to statistics, no specific requirements are made, all statistical concepts being developed from the beginning. On the other hand, since readers will usually have had previous experience with statistical methods, applications of each method are indicated in general terms, but concrete examples with data are not included. These are available in many of the standard textbooks.

The problems at the end of each chapter, many of them with outlines of solutions, provide exercises, further examples, and introductions to some additional topics. There is also given at the end of each chapter an annotated list of references regarding sources, both of ideas and of specific results. The notes are not intended to summarize the principal results of each paper cited but merely to indicate its significance for the chapter in question. In presenting these references I have not aimed for completeness but rather have tried to give a usable guide to the literature.

An outline of this book appeared in 1949 in the form of lecture notes taken by Colin Blyth during a summer course at the University of California. Since then, I have presented parts of the material in courses at Columbia, Princeton, and Stanford Universities and several times at the University of California. During these years I greatly benefited from comments of students, and I regret that I cannot here thank them individually. At different stages of the writing I received many helpful suggestions from W. Gautschi, A. Høyland, and L. J. Savage, and particularly from Mrs. C. Striebel, whose critical reading of the next to final version of the manuscript resulted in many improvements. Also, I should like to mention gratefully the benefit I derived from many long discussions with Charles Stein.

It is a pleasure to acknowledge the generous support of this work by the Office of Naval Research; without it the book would probably not have been written. Finally, I should like to thank Mrs. J. Rubalcava, who typed and retyped the various drafts of the manuscript with unfailing patience, accuracy, and speed.

E. L. LEHMANN

Berkeley, California
June 1959

Comments for Instructors

The two companion volumes, *Testing Statistical Hypotheses* (*TSH*) and *Theory of Point Estimation* (*TPE*), between them provide an introduction to classical statistics from a unified point of view. Different optimality criteria are considered, and methods for determining optimum procedures according to these criteria are developed. The application of the resulting theory to a variety of specific problems as an introduction to statistical methodology constitutes a second major theme.

On the other hand, the two books are essentially independent of each other. (As a result, there is some overlap in the preparatory chapters; also, each volume contains cross-references to related topics in the other.) They can therefore be taught in either order. However, *TPE* is somewhat more discursive and written at a slightly lower mathematical level, and for this reason may offer the better starting point.

The material of the two volumes combined somewhat exceeds what can be comfortably covered in a year's course meeting 3 hours a week, thus providing the instructor with some choice of topics to be emphasized. A one-semester course covering both estimation and testing can be obtained, for example, by deleting all large-sample considerations, all nonparametric material, the sections concerned with simultaneous estimation and testing, the minimax chapter of *TSH*, and some of the applications. Such a course might consist of the following sections: *TPE*: Chapter 2, Section 1 and a few examples from Sections 2, 3; Chapter 3, Sections 1–3; Chapter 4, Sections 1–4. *TSH*: Chapter 3, Sections 1–3, 5, 7 (without proof of Theorem 6); Chapter 4, Sections 1–7; Chapter 5, Sections 1–4, 6–8; Chapter 6, Sections 1–6, 11; Chapter 7, Sections 1–3, 5–8, 11, 12; together with material from the preparatory chapters (*TSH* Chapter 1, 2; *TPE* Chapter 1) as it is needed.

Contents

Theory of Point Estimation

Contents

CONTENTS

CHAPTER 1

The General Decision Problem

1. STATISTICAL INFERENCE AND STATISTICAL DECISIONS

The raw material of a statistical investigation is a set of observations; these are the values taken on by random variables X whose distribution P_θ is at least partly unknown. Of the parameter θ, which labels the distribution, it is assumed known only that it lies in a certain set Ω, the *parameter space*. *Statistical inference* is concerned with methods of using this observational material to obtain information concerning the distribution of X or the parameter θ with which it is labeled. To arrive at a more precise formulation of the problem we shall consider the purpose of the inference.

The need for statistical analysis stems from the fact that the distribution of X, and hence some aspect of the situation underlying the mathematical model, is not known. The consequence of such a lack of knowledge is uncertainty as to the best mode of behavior. To formalize this, suppose that a choice has to be made between a number of alternative actions. The observations, by providing information about the distribution from which they came, also provide guidance as to the best decision. The problem is to determine a rule which, for each set of values of the observations, specifies what decision should be taken. Mathematically such a rule is a function δ, which to each possible value x of the random variables assigns a decision $d = \delta(x)$, that is, a function whose domain is the set of values of X and whose range is the set of possible decisions.

In order to see how δ should be chosen, one must compare the consequences of using different rules. To this end suppose that the consequence of taking decision d when the distribution of X is P_θ is a *loss*, which can be expressed as a nonnegative real number $L(\theta, d)$. Then the long-term average loss that would result from the use of δ in a number of repetitions