Statistical Methods for Meta-Analysis

LARRY V. HEDGES
INGRAM OLKIN

Statistical Methods for Meta-Analysis

Larry V. Hedges

Department of Education
University of Chicago
Chicago, Illinois

Ingram Olkin

Department of Statistics
and School of Education
Stanford University
Stanford, California

1985

ACADEMIC PRESS, INC.

(Harcourt Brace Jovanovich, Publishers)

Orlando San Diego New York London
Toronto Montreal Sydney Tokyo

ACADEMIC PRESS, INC.
Orlando, Florida 32887

United Kingdom Edition published by
ACADEMIC PRESS INC. (LONDON) LTD.
24–28 Oval Road, London NW1 7DX

LIBRARY OF CONGRESS CATALOGING IN PUBLICATION DATA

Hedges, Larry V.
Statistical methods for meta-analysis.

Includes index.
1. Social sciences—Statistical methods. I. Olkin, Ingram. II. Title. III. Title: Meta-analysis.
HA29.H425 1985 300'.72 84-12469
ISBN 0-12-336380-2 (alk. paper)

PRINTED IN THE UNITED STATES OF AMERICA

85 86 87 88 9 8 7 6 5 4 3 2 1

Preface

Methodology for combining findings from repeated research studies has a long history. Early examples of combining evidence are found in replicated astronomical and physical measurements. Agricultural experiments particularly lend themselves to replication and led to the development of statistical techniques for merging results.

In recent years a plethora of meta-analyses have emerged in social science research. The need to arrive at policy decisions affecting social institutions fostered the momentum toward summarizing research. But, as with most methodologies, abuse frequently accompanies use. Two central aspects of meta-analysis were quickly recognized. One involved methods for collecting the body of information to be summarized. This pinpointed a variety of problems, pitfalls, and questions. For example, what steps should be taken to guarantee objectivity? Should some studies be omitted because of inadequacies in design or execution?

The second aspect of meta-analysis assumes as a starting point that we have available a set of reasonably well-designed studies that address the same question using similar outcome measures and focuses on the methodology needed for summarizing the data. Because classical statistics primarily addresses the analysis of single experiments, new formulations, models, and methods are required.

The main purpose of this book is to address the statistical issues for integrating independent studies. There exist a number of papers and books that discuss the mechanics of collecting, coding, and preparing data for a meta-analysis, and we do not deal with these.

It is not unusual in the early development of a field for terms to be used that later may be less than adequate or more restrictive than need be. In particular, the term *effect size* has been used to refer to standardized mean differences. In the beginning this usage was very natural in that the particular studies of interest did indeed involve differences between means. However, with more elaborate experimentation and more diverse applications, differences between treatments may depend not only on means but also on variances, medians, correlations, order statistics, distances, etc. Thus it would behoove us to now use the term *effect size* to refer to any such indices. But because this would be contrary to much of the existing literature, we do not, somewhat regrettably, do so. Instead we introduce the term *effect magnitude* to refer to measures in general.

The problem is further compounded. For large samples, quantities such as variances, medians, correlations, etc., will frequently have a normal distribution in which some of the population parameters will indeed be means. Consequently, although we may begin with statistics that are not means, we often end up with statistics that are effect sizes in the original sense.

Because this book concerns methodology, the content necessarily is statistical, and at times mathematical. In order to make the material accessible to a wider audience, we have not provided proofs in the text. Where proofs are given, they are placed as commentary at the end of a chapter. These can be omitted at the discretion of the reader.

We make a number of technical statements such as "The statistic Q has a chi-square distribution with degrees of freedom," or "The statistics Q_1 and Q_2 are independent." Each of these statements warrants a proof or a reference. However, for the sake of simplicity, readability, and accessibility to the materials, we have taken the liberty of omitting many proofs and references. However, as a compromise, we include a few proofs and references for statements that might be considered typical.

At times, mathematical expressions are needed. When possible we provide tables and graphs to make these expressions simple to use.

In our writing we have in mind a prototypical reader who is familiar with basic statistics at an applied level. This normally means the completion of a one-year sequence in statistics at a noncalculus level, and includes the ideas of statistical inference, regression and correlation, and analysis of variance. Concepts such as distribution (cumulative distribution function), expected value, bias, variance, and mean-squared error are generally defined in standard introductory statistics textbooks and are not defined in this book.

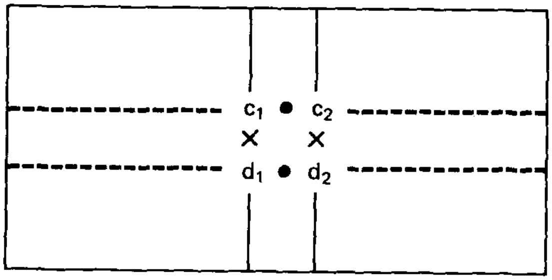
Occasionally, we use more advanced statistical concepts, such as consistency, efficiency, invariance, or asymptotic distributions. Because these concepts are used infrequently, we do not give formal definitions and on occasion give only a brief explanation or no explanation at all. Elementary expositions of these concepts can be found in more advanced statistics books. Our main reason for not discussing these concepts is that they are not essential for an understanding of the principles. A lengthy explanation would destroy readability. On the other hand, these comments can be useful for those readers familiar with the concepts.

Throughout the book we describe computational procedures whenever required. Many computations can be completed on a hand calculator, whereas some require the use of a standard statistical package such as SAS, SPSS, or BMD. Readers with experience using a statistical package or who conduct analyses such as multiple regression or analysis of variance should be able to carry out the analyses described with the aid of a statistical package.

Because of the inclusion of so many tables, a commentary on interpolation may be in order. For any two-way table (see figure), the simplest method of interpolation is linear in each direction; and in general, linear interpolation will suffice for most practical purposes. For more accurate interpolation, the values can be plotted horizontally or vertically. Thus, for example, vertical plots give interpolated values between c_1 and d_1 and between c_2 and d_2, denoted by crosses. Similarly, horizontal plots given interpolated values between c_1 and c_2 and between d_1 and d_2, denoted by dots. Subsequent linear interpolation in the other direction will provide a more accurate result than two-way linear interpolation. Of course, plotting a complete row or column of interpolated values gives still more accurate results, but this may be more effort than is warranted in practice.

A comment is in order concerning the calculations presented in the examples. Individual terms are presented after rounding, whereas totals are computed without roundoff. Consequently, discrepancies in the last decimal may exist in some of the computations.

There is an inherent difficulty in trying to use a single letter to denote a particular characteristic. For example, if we denote a population mean by μ, then how shall we denote the mean of the means of an experiment and control group? If we label the two means as μ^E and μ^C and $\mu = ½ (\mu^E + \mu^C)$, then we have denoted by μ two different types of means. Such inconsistencies are inherent in the subject and occur throughout the book, so it is important to make quite explicit the underlying context and thereby remove potential confusion. Because



We write $a \approx b$ to mean that a and b are "approximately" equal. We use the symbol $a \equiv b$ to signify a definition.

Independent multiple opinions and replications of experiments are but two examples of corroborative evidence which are currently in vogue, and which we believe will increase in frequency during the next decade. We hope that the present development of methodology will provide some guidelines for the rigorous interpretation of data from independent sources.

Acknowledgments

It is a pleasure to express our appreciation to Dr. Betsy Becker (Michigan State University), Mr. Brad Hanson (Stanford University), Dr. Nan Laird (Harvard University), Ms. Therese Pigott (University of Chicago), Dr. Marie Tisak (University of California, Berkeley), and Dr. Margaret Uguroglu (DeVry, Inc.) for reading the manuscript and providing us with numerous comments and suggestions. The authors are particularly grateful to Dr. Betsy Becker for carrying out the simulation studies and for the computations in many of the tables. Therese Pigott (University of Chicago) checked the computations in all of our examples. Although it would be comforting to be able to blame these readers for errors, honesty compels us to say that any shortcomings are solely ours.

Unfortunately, the first draft of this book was not typed onto a word processor, thereby necessitating many retypings. For this task we thank Julie Less, Irene Miura, Jerri Rudnick and Tonda West who managed to maintain a calm exterior at all times.

We are grateful to the Spencer Foundation and to the National Science Foundation for their continued support of our research. Their generous support has been essential in making this work possible.

The authors wish to acknowledge, with thanks, the following associations and societies for kindly granting us permission to use published materials:

The American Psychological Association for Tables 1, 2, 3, and Fig. 1 of Chapter 4; Tables 4, 5, 6 and Figure 3 of Chapter 5; Tables 3, 4, and 7 of Chapter

6; Tables 1, 2, 3, 4, 5, and 6 of Chapter 13; and Table 4 of Chapter 14.
The American Educational Research Association for Table 2 of Chapter 5, Tables 1, 2 and 3 of Chapter 8, and Table 1, 2 and 3 of Chapter 14.
The Institute of Mathematical Statistics for Table 1 of Chapter 11.
The Biometrika Trust for Appendix E and Appendix G.

Contents