

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



Academic Press
San Diego New York Boston
London Sydney Tokyo Toronto

Contents

Preface

xv

1. Introduction

A. The Use of Statistical Procedures for Combining the Results of Research Studies in the Social Sciences	2
A.1 The Misuse of Statistical Significance in Reviewing Research	3
A.2 Statistical Procedures for Combining Estimates of Effect Magnitude	6
B. Failings in Conventional Statistical Methodology in Research Synthesis	7
B.1 Goals of Statistical Procedures in Research Synthesis	7
B.2 Conventional Analyses for Effect Size Data	9
B.3 Conceptual Problems with Conventional Analyses	10
B.4 Statistical Problems with Conventional Analyses	11
C. Statistics for Research Synthesis	12

2. Data Sets

A. Cognitive Gender Differences	16
B. Sex Differences in Conformity	21
C. The Effects of Open Education	23

D. The Relationship between Teacher Indirectness and Student Achievement	26
3. Tests of Statistical Significance of Combined Results	
A. Preliminaries and Notation	28
B. General Results on Tests of Significance of Combined Results	31
C. Combined Test Procedures	33
C.1 Methods Based on the Uniform Distribution	34
C.2 The Inverse Chi-Square Method	37
C.3 The Inverse Normal Method	39
C.4 The Logit Method	40
C.5 Other Procedures for Combining Tests	42
C.6 Summary of Combined Test Procedures	43
D. The Uses and Interpretation of Combined Test Procedures in Research Synthesis	45
E. Technical Commentary	46
4. Vote-Counting Methods	
A. The Inadequacy of Conventional Vote-Counting Methodology	48
B. Counting Estimators of Continuous Parameters	52
C. Confidence Intervals for Parameters Based on Vote Counts	53
C.1 Use of Normal Theory Approach	54
C.2 Use of Chi-Square Theory	54
D. Choosing a Critical Value	56
E. Estimating an Effect Size	56
F. Estimating a Correlation	63
G. Limitations of Vote-Counting Estimators	67
H. Vote-Counting Methods for Unequal Sample Sizes	69
H.1 The Large Sample Variance of the Maximum Likelihood Estimator	70
H.2 Estimating Effect Size	71
5. Estimation of a Single Effect Size: Parametric and Nonparametric Methods	
A. Estimation of Effect Size from a Single Experiment	76
A.1 Interpreting Effect Sizes	76
A.2 An Estimator of Effect Size Based on the Standard Mean Difference	78

A.3	An Unbiased Estimator of Effect Size	81
A.4	The Maximum Likelihood Estimator of Effect Size	81
A.5	Shrunken Estimators of Effect Size	82
A.6	Comparing Parametric Estimators of Effect Size	82
B.	Distribution Theory and Confidence Intervals for Effect Sizes	85
B.1	The Asymptotic Distribution of Estimators of Effect Size	85
B.2	Confidence Intervals for Effect Sizes Based on Transformations	88
B.3	Exact Confidence Intervals for Effect Sizes	91
C.	Robust and Nonparametric Estimation of Effect Size	92
C.1	Estimates of Effect Size that are Robust Against Outliers	93
C.2	Nonparametric Estimators of Effect Size	93
C.3	Estimators Based on Differences of Control versus Treatment Proportions	95
C.4	Estimators Based on Gain Scores in the Experimental Group Relative to the Control Group	97
C.5	Nonparametric Estimators Involving Only Posttest Scores	98
C.6	Relationships between Estimators	99
D.	Other Measures of Effect Magnitude	100
D.1	The Correlation Coefficient and Correlation Ratio	101
D.2	The Intraclass Correlation Coefficient	102
D.3	The Omega-Squared Index	103
D.4	Problems with Variance-Accounted-For Measures	103
E.	Technical Commentary	104
6.	Parametric Estimation of Effect Size From a Series of Experiments	
A.	Model and Notation	108
B.	Weighted Linear Combinations of Estimates	109
B.1	Estimating Weights	110
B.2	Efficiency of Weighted Estimators	113
B.3	The Accuracy of the Large Sample Approximation to the Distribution of Weighted Estimators of Effect Size	114
C.	Other Methods of Estimation of Effect Size from a Series of Experiments	117
C.1	The Maximum Likelihood Estimator of Effect Size from a Series of Experiments	118

C.2 Estimators of Effect Size Based on Transformed Estimates	119
D. Testing for Homogeneity of Effect Sizes	122
D.1 Small Sample Significance Levels for the Homogeneity Test Statistic	124
D.2 Other Procedures for Testing Homogeneity of Effect Sizes	124
E. Computation of Homogeneity Test Statistics	127
F. Estimation of Effect Size for Small Sample Sizes	128
F.1 Estimation of Effect Size from a Linear Combination of Estimates	129
F.2 Modified Maximum Likelihood Estimation of Effect Size	131
G. The Effects of Measurement Error and Invalidity	131
G.1 The Effects of Measurement Error	132
G.2 The Effect of Validity of Response Measures	138
7. Fitting Parametric Fixed Effect Models to Effect Sizes: Categorical Models	
A. An Analogue to the Analysis of Variance for Effect Sizes	149
B. Model and Notation	149
C. Some Tests of Homogeneity	153
C.1 Testing Whether All Studies Share a Common Effect Size	153
C.2 Testing Homogeneity of Effect Sizes Across Classes	154
C.3 Testing Homogeneity within Classes	155
C.4 An Analogy to the Analysis of Variance	156
C.5 Small Sample Accuracy of the Asymptotic Distributions of the Test Statistics	156
D. Fitting Effect Size Models to a Series of Studies	157
E. Comparisons among Classes	159
E.1 Simultaneous Tests for Many Comparisons	160
F. Computational Formulas for Weighted Means and Homogeneity Statistics	163
8. Fitting Parametric Fixed Effect Models to Effect Sizes: General Linear Models	
A. Model and Notation	168
B. A Weighted Least Squares Estimator of Regression Coefficients	169

C. Testing Model Specification	172
D. Computation of Estimates and Test Statistics	173
E. The Accuracy of Large Sample Approximations	174
F. Other Methods of Estimating Regression Coefficients	183
F.1 Maximum Likelihood Estimators of Regression Coefficients	183
F.2 Estimators Based on Transformations of Sample Effect Sizes	184
G. Technical Commentary	187
9. Random Effects Models for Effect Sizes	
A. Model and Notation	191
B. The Variance of Estimates of Effect Size	193
C. Estimating the Effect Size Variance Component	193
D. The Variance of the Effect Size Variance Component	195
E. Testing That the Effect Size Variance Component Is Zero	197
F. Estimating the Mean Effect Size	198
G. Empirical Bayes Estimation for Random Effects Models	200
10. Multivariate Models for Effect Sizes	
A. Model and Notation	206
B. The Multivariate Distribution of Effect Sizes	209
C. Estimating a Common Effect Size from a Vector of Correlated Estimates	210
C.1 Testing Homogeneity of Correlated Effect Sizes	210
C.2 Estimation of Effect Size From Correlated Estimates	212
D. Estimating a Common Effect Size and Testing for Homogeneity of Effect Sizes	213
D.1 An Estimator of Effect Size	213
D.2 Testing Homogeneity of Effect Sizes	215
E. Estimating a Vector of Effect Sizes	216
F. Testing Homogeneity of Vectors of Effect Sizes	219
G. Is Pooling of Correlated Estimators Necessary?	221
11. Combining Estimates of Correlation Coefficients	
A. Estimating a Correlation from a Single Study	224
A.1 Point Estimation of a Correlation from a Single Study	224
A.2 Approximations to the Distribution of the Sample Correlation Coefficient	226

A.3	Exact Confidence Intervals for Correlations	228
B.	Effects of Measurement Error	228
C.	Estimating a Common Correlation from Several Studies	229
C.1	Weighted Estimators of a Common Correlation	230
C.2	The Maximum Likelihood Estimator of a Common Correlation	232
D.	Testing Homogeneity of Correlations across Studies	234
D.1	A Test of Homogeneity Based on Fisher's z-Transform	235
D.2	The Likelihood Ratio Test of Homogeneity of Correlations	236
E.	Fitting General Linear Models to Correlations	237
E.1	Model and Notation	237
E.2	Estimating Regression Coefficients	238
E.3	Testing Model Specification	240
E.4	Computation of Estimates and Test Statistics	241
F.	Random Effects Models for Correlations	242
F.1	Model and Notation	242
F.2	Testing That the Variance of Population Correlations is Zero	243
F.3	An Unbiased Estimate of the Correlation Variance Component	244

12. Diagnostic Procedures for Research Synthesis Models

A.	How Many Observations Should Be Set Aside	249
B.	Diagnostic Procedures for Homogeneous Effect Size Models	251
B.1	Graphic Method for Identifying Outliers	251
B.2	The Use of Residuals to Locate Outliers	253
B.3	The Use of Homogeneity Statistics to Locate Outliers	256
C.	Diagnostic Procedures for Categorical Models	257
D.	Diagnostic Procedures for General Linear Models	257
D.1	The Use of Residuals to Locate Outliers in General Linear Models	258
D.2	Changes in Regression Coefficients	260
D.3	The Use of Model Specification Statistics to Locate Outliers	261
E.	Diagnostic Procedures for Combining Estimates of Correlation Coefficients	262
F.	Technical Commentary	262

13. Clustering Estimates of Effect Magnitude

- A. Theory for Clustering Unit Normal Random Variables 266
 - A.1 Disjoint Clustering 267
 - A.2 Overlapping Clustering 269
- B. Clustering Correlation Coefficients 271
- C. Clustering Effect Sizes 273
- D. The Effect of Unequal Sample Sizes 276
 - D.1 The Effect on the Disjoint Clustering Procedure 277
 - D.2 The Effect on the Overlapping Clustering Procedure 278
 - D.3 An Alternative Method for Handling Unequal Sample Sizes 279
- E. Relative Merits of the Clustering Procedures 281
- F. Computation of Tables 282
 - F.1 The Significance of Gaps 282
 - F.2 Significance Values for the Overlapping Clustering Procedure 283

14. Estimation of Effect Size When Not All Study Outcomes Are Observed

- A. The Existence of Sampling Bias in Observed Effect Size Estimates 286
- B. Consequences of Observing Only Significant Effect Sizes 287
 - B.1 Model and Notation 288
 - B.2 The Distribution of the Observed Effect Size 288
- C. Estimation of Effect Size from a Single Study When Only Significant Results Are Observed 290
 - C.1 Estimation of Effect Size 291
 - C.2 The Distribution of the Maximum Likelihood Estimator 292
- D. Estimation of Effect Size from a Series of Independent Experiments When Only Significant Results are Observed 297
 - D.1 Estimation of Effect Size Using Counting Procedures 298
 - D.2 The Maximum Likelihood Estimator of Effect Size 301
 - D.3 Weighted Estimators of Effect Size 302
 - D.4 Applications of the Methods That Assume Censoring 303
- E. Other Methods for Estimating Effect Sizes When Not All Study Outcomes Are Observed 304
 - E.1 Assessing the Number of Studies (with Null Results) Needed to Overturn a Conclusion 305
 - E.2 Random Effects Models When Not All Study Outcomes Are Observed 306

F. Technical Commentary	307
15. Meta-Analysis in the Physical and Biological Sciences	
A. A Multiplicative Model	314
B. Estimating Displacement and Potency Factor	318
C. A Multiplicative Model with Scaling	321
D. An Additive Model for Interlaboratory Differences	322
E. An Additive Model With a Control	323
Appendix	
A. Table of the Standard Normal Cumulative Distribution Function	328
B. Percentiles of Chi-Square Distributions	330
C. Values of $z = \frac{1}{2} \log[(1+r)/(1-r)]$ and $r = (e^{2z} - 1)/(e^{2z} + 1)$ and Graph of Fisher's z -Transformation	333
D. Values of $\sqrt{2} \sinh^{-1} x$ and $\sinh(x/\sqrt{2})$	335
E. Confidence Intervals of the Parameter p of the Binomial Distribution	337
F. Nomographs for Exact Confidence Intervals for Effect Size δ When $2 \leq n \leq 10$	341
G. Confidence Intervals for the Correlation Coefficient for Different Sample Sizes	342
References	347
Index	361