PRIMER *on* EFFECT SIZES, SIMPLE RESEARCH DESIGNS, *and* CONFIDENCE INTERVALS

MARTY SAPP

PRIMER ON EFFECT SIZES, SIMPLE RESEARCH DESIGNS, AND CONFIDENCE INTERVALS

PRIMER ON EFFECT SIZES, SIMPLE RESEARCH DESIGNS, AND CONFIDENCE INTERVALS

By

MARTY SAPP

Professor
Department of Educational Psychology
University of Wisconsin-Milwaukee

CHARLES C THOMAS • PUBLISHER, LTD.
Springfield • Illinois • U.S.A.

Published and Distributed Throughout the World by

CHARLES C THOMAS • PUBLISHER, LTD.
2600 South First Street
Springfield, Illinois 62704

ISBN 978-0-398-09197-2 (paper)
ISBN 978-0-398-09198-9 (ebook)

With THOMAS BOOKS *careful attention is given to all details of manufacturing and design. It is the Publisher's desire to present books that are satisfactory as to their physical qualities and artistic possibilities and appropriate for their particular use.* THOMAS BOOKS *will be true to those laws of quality that assure a good name and good will.*

Printed in the United States of America

TO-C-1

Library of Congress Cataloging-in-Publication Data

Names: Sapp, Marty, 1958- author.
Title: Primer on effect sizes, simple research designs, and confidence intervals / by Marty Sapp, Professor, Department of Educational Psychology, University of Wisconsin-Milwaukee.
Description: Springfield, Illinois : Charles C Thomas, Publisher, [2017] | Includes bibliographical references and index.
Identifiers: LCCN 2017031574 (print) | LCCN 2017046235 (ebook) | ISBN 9780398091989 (ebook) | ISBN 9780398091972 (pbk)
Subjects: LCSH: Psychometrics. | Psychology--Statistical methods. | Psychology--Methodology.
Classification: LCC BF39 (ebook) | LCC BF39 .S2653 2017 (print) | DDC 150.72/1--dc23
LC record available at https://lccn.loc.gov/2017031574

PREFACE

A Primer on Effect Sizes, Simple Research Designs, and Confidence Intervals was designed to help individuals learn to calculate effect sizes for their research designs. Effect sizes allow a clinician or researcher to determine the effect of a treatment. For example, an effect size of zero would indicate that the treatment had no effect, but generally, effect sizes allow researchers to see the degree of effect of some treatment or intervention. Often, researchers and clinicians are not aware that effect sizes are connected to research designs. For years, statisticians have been aware of limits of null hypothesis significance testing (NHST). The Wilkinson Task Force (Wilkinson & Task Force on Statistical Inference, 1999) recommended that researchers report effect sizes and confidence intervals in addition to null hypothesis significance testing (NHST).

A sample effect size allows a researcher or clinician to determine the effect size for his or her sample data, but a confidence interval around an effect size allows one to describe the effect size within a given population. Since 1999, the reporting of effect sizes by researchers has been inconsistent. Also, some researchers and some statisticians have little guidance and understanding of effect sizes. Realistically speaking, there are many effect size measures and researchers need some guidelines on how to calculate these simple statistics and how to interpret them. Recently, in 2015, the journal *Basic and Applied Social Psychology* banned null hypothesis significance testing (NHST). The Publication Manual of the American Psychological Association encourages the reporting of effect sizes with inferential test statistics. Also, this manual encourages the reporting of confidence intervals. The purpose of this book is to provide the connection among effect sizes, confidence intervals, and simple research designs. Also, some commonly used univariate and multivariate statistics are covered. Regression discontinuity designs, simple moderation and mediation designs, power analysis, and fit indices as effect sizes measure are presented. All calculations are demonstrated through a calculator and statistical packages such as Microsoft Excel, SPSS, SAS, and Hayes' Process Analysis. This book covers more than 25 effect sizes that are connected to simple research designs.

This book will be of interest to students taking a statistics class, research methods class, or research design class. Unlike many texts within this area, the current text will give students or researchers the understanding of how to calculate effect sizes with a simple calculator or with a few commands from statistical software programs. Hence, mathematical ability is not a prerequisite for this text. This text provides a nonmathematical treatment of effect sizes within the context of research designs. Finally, to aid understanding, critical material is repeated throughout this book.

CONTENTS

PRIMER ON EFFECT SIZES, SIMPLE RESEARCH DESIGNS, AND CONFIDENCE INTERVALS

Chapter 1

INTRODUCTION: R AND D EFFECT SIZES

HISTORY OF EFFECT SIZES

Huberty (2002) found that the history of effect size started around 1940. The correlation ratio or eta coefficient was proposed during the 1940s. The correlation ratio is used to measure curvilinear relationships. In addition, eta measures the relationship between a grouping variable and a dependent or outcome variable. During this period, eta squared was connected to ANOVA to show the variance accounted for on a dependent variable. Suppose 20 participants were randomly assigned to four groups, and let us assume the groups and data are the following:

Group	*Dependent Variable*
1	53
1	54
1	52
1	55
1	54
2	53
2	56
2	57
2	55
3	57
3	56
3	54
3	58
3	59
3	58
4	62

4	62
4	61
4	60
4	56

The groups 1 through 4 is the grouping variable. In other words, a grouping variable is the levels of an independent variable. Eta squared equals between sum of squares divided by the total sum of squares. Eta Squared=118.050/172.800=.683.

Eta=.826. Cohen characterized eta squared of .01 as a small effect size, an eta squared of .06 as a medium effect size, and an eta squared of .14 as a large effect size.

The .683 is the variance accounted for on the dependent variable, and .826 is the correlation of the group identifications with the dependent variable. Ronald A. Fisher (1890–1961), in 1924, derived the probability of eta in the context of ANOVA. Truman (1935) Kelley (1884–1961) proposed an adjustment to the eta squared within the context of ANOVA. Some statisticians refer to this as the partial eta squared. The psychologist William L. Hays (1926–1995) in his popular textbook, proposed omega squared as an alternative to eta squared (Hays, 1981). Omega squared is said to be derived through unbiased estimates. Omega squared =SSB-(K-1)MSW/(SST+MSW). Where SSB equals the sum of squares between and K equals the number of groups. MSW is the mean squares within, and SST is the total sum of squares. Generally, omega squared and eta squared will not differ much. If the levels of the grouping variable (independent variable) are random, in contrast to being fixed, the intraclass correlation coefficient can be used as an effect size. The formula for the intraclass correlation R is the following:

$$R=(MSB-MSW)/[MSB+(n-1)MSW]$$

MSB and MSW are the numerator and denominator from an F statistic or test and n equals the number of participants per group.

In summary, at least three strengths of relationship effect sizes were proposed between 1935 to 1963: eta squared, omega squared, and the intraclass correlation coefficient. Karl Pearson (1857–1936), in 1910, proposed the biserial correlation coefficient. It is used when a continuous variable is forced into a discrete variable and is correlated with a continuous variable. For example, suppose we were interested in the correlation between hypnotizablity and creative imagination. Both of these variables are continuous, but we forced the hypnotizability scores into high and low hypnotizability. The correlation between these two variables would be the biserial correlation coefficient. The biserial correlation coefficient cannot be used in regression in order to predict y values or dependent variables. Also, confidence intervals cannot be placed around the biserial correlation coefficient. Finally, the biserial correlation coefficient is less reliable than the Pearson correlation coefficient, and it is not recommended as an effect size (Sapp, 2015).

Jacob Cohen, in 1969, proposed an effect size for a two group mean comparison, and Huberty (2002) referred to these as group differences indices. Cohen defined his effect size as the differences between means divided by the pool standard deviation across the two groups. Like Cohen, the statistician Gene V. Glass also proposed a d effect size as the differences between means divided by the control group standard deviation. In addition, the statistician Larry V. Hedges (1982) took exception with Cohen and Glass, and he proposed an adjusted d that he called g (Huberty, 2002). Cohen also proposed a standard difference type of effect size for multiple groups or multiple means context (ANOVA), and he used the letter f as this effect size, and it is the following formula:

$$f=[(K-1)F/N]^{1/2}$$

K is the number of groups, and F is the F statistics from ANOVA. N is the total group size. When using Cohen's power tables, the average group size is used or the harmonic mean if the group sizes are unequal. The f effect size can be seen as the standard deviation of the standardized means, or the variability of the group means relative to the standard deviation (Huberty, 2002). Cohen (1977) characterized f equals .10 as a small effect size, $f=.25$ as a medium effect size, and $f>.40$ as a large effect size.

Huberty (2002) discussed another effect size based on overlap indices. Within a two-group situation, if two have a large amount of overlap the effect size will be small. Cohen (1977) also defined d as the percent of non-overlap of the treatment group scores with those of the untreated group. For example, the following shows various effect sizes and percent of non-overlap.