Methods for the Behavioral, Educational, and Social Sciences: An R package

KEN KELLEY
Indiana University, Bloomington, Indiana

Methods for the Behavioral, Educational, and Social Sciences (MBESS; Kelley, 2007b) is an open source package for R (R Development Core Team, 2007b), an open source statistical programming language and environment. MBESS implements methods that are not widely available elsewhere, yet are especially helpful for the idiosyncratic techniques used within the behavioral, educational, and social sciences. The major categories of functions are those that relate to confidence interval formation for noncentral t, F, and χ^2 parameters, confidence intervals for standardized effect sizes (which require noncentral distributions), and sample size planning issues from the power analytic and accuracy in parameter estimation perspectives. In addition, MBESS contains collections of other functions that should be helpful to substantive researchers and methodologists. MBESS is a long-term project that will continue to be updated and expanded so that important methods can continue to be made available to researchers in the behavioral, educational, and social sciences.

R is an open source statistical programming language and environment for (essentially) all operating systems that has gained a widespread following in quantitative disciplines (R Development Core Team, 2007b). This following is perhaps most prevalent in the statistical sciences, where many published works now provide R routines so that the proposed methods can be immediately implemented. The speed of implementing many methods in R is astonishing when compared with some of the older "standard" software programs, with which a considerable amount of time can pass between development of a method and its implementation in the program, if indeed it is ever implemented. One of the real benefits of R is that users contribute to the R Project by making packages that contain sets of functions and then contributing the packages to the community of R users. An R package includes a set of functions, often with a particular goal in mind, and a set of help files. Currently, there are approximately 900 packages available that provide specialized routines for a wide variety of topics that are available on the Comprehensive R Archival Network (CRAN; R Development Core Team, 2007a), which is a network of servers around the world that have up-to-date versions of R, R packages, and R documentation. Information on the wide variety of R resources available to the community of R users is available on the R Project Internet home page (R Development Core Team, 2007c).

R has recently gained a following in the applied sciences, where researchers are often interested in implementing new methods and producing high-quality graphics, with the added benefit of working within a relatively complete object-oriented statistical environment. The be-

havioral, educational, and social sciences are examples of applied areas in which R has made tremendous gains in popularity over the last several years. For example, Doran and Lockwood (2006) provide details on using R to fit value-added longitudinal models for behavioral and educational data using R (R Development Core Team, 2007b) with the nlme package (Pinheiro, Bates, DebRoy, & Sarkar, 2006), there is a special issue in *Journal of Statistical Software* on Psychometrics in R (de Leeuw, 2006), and statistical texts used in the applied behavioral, educational, and social sciences are beginning to incorporate R (e.g., Everitt, 2005; Fox, 2002). However, there is not much in the collections of packages specifically designed for the idiosyncratic needs of applied behavioral, educational, and social science researchers. The purpose of the present article is to formally introduce the Methods for the Behavioral, Educational, and Social Sciences (MBESS; Kelley, 2007b) R package, which is designed to implement functions that are especially useful for substantive researchers and methodologists in the behavioral, educational, and social sciences and to highlight some of the user-friendly functions contained within MBESS that are not widely available elsewhere.

Confidence Intervals for Effect Sizes Based on Noncentral Distributions

Distributions such as (central) t, F, and χ^2 are very important for null hypothesis significance testing, because such distributions describe the sampling distribution of t, F, and χ^2 variates when the null hypothesis is true. When the null hypothesis is false, the sampling distribution of t, F, and χ^2 variates is noncentral, where the magnitude of

K. Kelley, kkiii@indiana.edu

the noncentrality parameter indexes the difference between the null and the alternative hypotheses. Central distributions are thus special cases of their corresponding noncentral counterparts when the noncentrality parameter equals zero (i.e., when the null hypothesis is true). When calculating probability values under the null hypothesis, central distributions are necessary because the probability value is conditional on the null hypothesis being true (i.e., the probability of obtaining data as extreme or more extreme than the observed data if the null hypothesis is true).

Noncentral distributions are necessary for calculating confidence intervals for many effect sizes that are bounded or that have been standardized (e.g., Cumming & Finch, 2001; Smithson, 2003; Steiger & Fouladi, 1997). A major push in the behavioral, educational, and social sciences is the use of effect sizes and confidence intervals for effect sizes (e.g., Cohen, 1994; Cumming & Finch, 2001; Grissom & Kim, 2005; Meehl, 1997; Rosenthal, Rosnow, & Rubin, 2000; Steiger, 2004; Steiger & Fouladi, 1997; Wilkinson, L., & the American Psychological Association Task Force on Statistical Inference, 1999). It has even been argued that null hypothesis tests should be banned (Schmidt, 1996) and that the future of quantitative research should be based on confidence intervals for effect sizes (Thompson, 2002). Indeed, if significance tests were actually banned and confidence intervals for effect sizes required, at present, many researchers would have a difficult time reporting the results of their research, because such methods are not implemented in standard software for the most commonly used effect sizes.

For effect sizes that follow noncentral distributions, the quantiles of the particular distribution are a function of not only the degrees of freedom, but also a noncentrality parameter. Popular statistical packages that are widely used in the behavioral, educational, and social sciences do not perform such calculations without special programming scripts, making it difficult for many researchers to obtain the critical values necessary for confidence interval formation for effect sizes that follow noncentral distributions. As a result of the added complexity of the probability density function because of the noncentrality parameter, finding confidence limits for noncentral parameters involves complicated iterative routines (Cumming & Finch, 2001; Smithson, 2003; Steiger & Fouladi, 1997).These routines have been implemented in MBESS in the form of easy-to-use R functions for the noncentral t, noncentral F, and noncentral χ^2 distributions.

Due to the importance of working with noncentral distributions for forming confidence intervals for many effect sizes, such as the standardized mean, the standardized mean difference (e.g., Cumming & Finch, 2001; Kelley, 2005; Steiger & Fouladi, 1997), the squared multiple correlation coefficient (e.g., Algina & Olejnik, 2000; Smithson, 2001, 2003), standardized regression coefficients (e.g., Kelley & Maxwell, 2008), the coefficient of variation (e.g., Kelley, 2007c), and the root-mean-square error of approximation (RMSEA; e.g., Browne & Cudeck, 1993), the lack of easy-to-use noncentral distribution functions is clearly a shortcoming of the widely available analytic tools. (See Kelley, 2007a, for a general review of

confidence intervals for standardized effect sizes.) This limitation is overcome with MBESS, which is a package used within the R statistical language and environment. MBESS thus fits within a broader environment for statistical computations that can be used in conjunction with R and all of the other add-on packages.

MBESS implements the three main noncentral distributions in easy-to-use functions. The functions are `conf.limits.nct()`, `conf.limits.ncf()`, and `conf.limits.nc.chisq()` for the noncentral t, F, and χ^2 distributions, respectively.[1] These functions return the noncentral parameters that have at their desired quantiles (e.g., $1-\alpha/2$ and $\alpha/2$) the estimated noncentral parameter, which is how confidence intervals for noncentral parameters are formed (see, e.g., Steiger, 2004, or Steiger & Fouladi, 1997, for a discussion of confidence interval formation procedures for noncentral parameters).

Although there are numerous options that can be used, at minimum, the functions for confidence intervals for noncentral t, F, and χ^2 parameters require users to specify the observed test statistic, the degrees of freedom, and the desired level of confidence. An example call to the `conf.limits.nct()` function would be of the form

```
R> conf.limits.nct(t.value=tObs., df=ν,
conf.level=1-α)
```

for a $(1-\alpha)100\%$ confidence interval for the noncentrality parameter of the t distribution, where $t_{Obs.}$ is the observed t value, ν is degrees of freedom, and $1-\alpha$ is the level of confidence interval coverage. When nonsymmetric, such as one-sided, confidence intervals are of interest, the lower and upper Type I error rate can be specified individually (instead of specifying the $1-\alpha$ confidence level, as above). This is done by using `alpha.lower` and `alpha.upper` in place of `conf.level` in standard applications of the function. An example call to the `conf.limits.ncf()` function for a nonsymmetric confidence interval would be of the form

```
R> conf.limits.ncf(F.value=FObs.,
df.1=ν1, df.2=ν2, alpha.lower=0,
alpha.upper=α)
```

for a $(1-\alpha)100\%$ upper confidence interval, where $F_{Obs.}$ is the observed F value, ν_1 is the numerator degrees of freedom, and ν_2 is the denominator degrees of freedom. An example call to the function `conf.limits.nc.chisq()` for a confidence interval would be of the form

```
R> conf.limits.nc.chisq(Chi.
Square=χ²Obs., conf.level=1-α, df=ν),
```

for a $(1-\alpha)100\%$ confidence interval, where $\chi^2_{Obs.}$ is the observed χ^2 value. Given `conf.limits.nct()`, `conf.limits.ncf()`, and `conf.limits.nc.chisq()` functions, a confidence interval can be formed for noncentrality parameters whose estimates follow noncentral t, F, or χ^2 distributions.

Usually, however, what is of interest is the confidence interval for the effect size parameter, not the confidence interval for the noncentrality parameter. By transforming the limits of the confidence intervals for the noncentrality

parameter to the scale of the effect size, using the confidence interval transformation principle that allows confidence limits from one metric to be transformed into confidence limits of another metric under certain conditions (e.g., Steiger, 2004; Steiger & Fouladi, 1997), confidence intervals for the population effect size of interest can be obtained. A set of functions exist that determine the limits of the confidence intervals directly (these functions call upon the noncentral functions discussed above) for several important and widely used effect size measures.

For example, a confidence interval for the standardized mean difference can be obtained with the `ci.smd()` function. An example call to the `ci.smd()` function could be of the form

```
R> ci.smd(smd=d, n.1=n_1, n.2=n_2,
conf.level=1-α),
```

where `smd` is the standardized mean difference argument and d the observed standardized mean difference, `n.1` and `n.2` are the sample size arguments for Groups 1 and 2, with n_1 and n_2 being the sample sizes of Group 1 and Group 2, respectively, and `conf.level` is the confidence interval argument, with $1-\alpha$ being the desired level of confidence. For example, suppose a researcher performs an experiment in which the observed standardized mean difference between the treatment group and the control group is .525, where there were 64 participants in each of the two groups. The `ci.smd()` function could be used as

```
R> ci.smd(smd=.525, n.1=64, n.2=64,
conf.level=.95),
```

which yields a 95% confidence interval,

$$\text{CI}_{.95} = [.17152 \leq \delta \leq .87647],$$

where $\text{CI}_{.95}$ represents a 95% confidence interval for δ, the population standardized mean difference.[2]

The structure of the confidence interval procedure for the standardized mean, a standardized contrast in ANOVA or ANCOVA context, the squared multiple correlation coefficient (based on random or fixed regressors), the signal-to-noise ratio in an ANOVA context, the coefficient of variation, standardized regression coefficients, and the root mean square error of approximation are of the same form as for the standardized mean difference. Confidence intervals for these effect sizes can be obtained with the `ci.sm()`, `ci.sc()`, `ci.sc.ancova()`, `ci.R2()`, `ci.snr()`, `ci.cv()`, `ci.src()` [or `ci.rc()` for unstandardized regression coefficients] and `ci.rmsea()` functions, respectively.[3] The ease with which these functions can be used is greater than in many programs, some of which require complicated code and others of which are stand-alone programs, with the added bonus that all of the functions from MBESS are part of the larger R statistical computing environment. These confidence intervals are all based on exact confidence intervals for noncentral distributions, which is of course the ideal approach. For effect sizes that do not have stand-alone functions, the use of the confidence interval functions for noncentral parameters can many times be used, where the limits of the confidence interval for the noncentrality parameter are transformed into the limits of the effect size of interest (e.g., see the discussion in Steiger, 2004; Steiger & Fouladi, 1997). Steiger (2004) provides an important and timely review of many important standardized effect sizes, all of which can be implemented with MBESS either directly or with a small number of additional (hand or R) calculations after obtaining the confidence limits for the noncentrality parameter with the noncentral confidence interval functions.

Sample Size Planning

Although there are some functions within R and within certain R packages for planning sample size from a power analytic perspective, where the goal is to obtain results that reach statistical significance, the MBESS package contains functions for planning sample size from the power analytic perspective (e.g., Cohen, 1988; Kraemer & Thiemann, 1987; Lipsey, 1990; Murphy & Myors, 2004) as well as from the accuracy in parameter estimation (AIPE) perspective (e.g., Hahn & Meeker, 1991; Kelley & Maxwell, 2003, 2008; Kelley, Maxwell, & Rausch, 2003; Kelley & Rausch, 2006; Kupper & Hafner, 1989). Whereas the goal of the power analytic approach is to plan sample size so that there is some desired probability of rejecting a false null hypothesis, the goal of the AIPE approach is to obtain a sufficiently narrow confidence interval with some desired probability.

Methods to plan sample size for standardized effect sizes from the AIPE perspective—so that the widths of the confidence intervals are sufficiently narrow—are beginning to receive attention in the literature as a way to overcome limitations of null hypothesis significance tests and as a general way to learn about the phenomenon of interest from data. Although such methods have not historically received much attention for standardized effect sizes (cf. Algina & Olejnik, 2000; Bonett & Wright, 2000), this is in part due to the difficulties of forming confidence intervals for standardized effect sizes, as discussed in the previous section. Going beyond forming the intervals and planning sample size in order to achieve sufficiently narrow confidence intervals can be quite difficult due to the necessary iterative nature of the procedures. The planning of sample size from the AIPE perspective takes two forms: one where the expected width of the confidence interval is sufficiently narrow, and one where there will be some desired degree of assurance (i.e., a probability) that the confidence interval observed in a particular study will be sufficiently narrow. The former is generally simpler than the latter, because incorporating a desired degree of assurance into the procedure requires finding a value to base sample size that will not lead to a wider than desired confidence interval width with more than the desired probability.

Example functions for sample size planning from the AIPE perspective that are available in MBESS are the standardized mean difference (Kelley & Rausch, 2006) using the `ss.aipe.smd()` function, the squared multiple correlation coefficient (fixed or random regressors) using the `ss.aipe.R2()` function (Kelley, 2007d), the coefficient of variation using the `ss.aipe.cv()`

function (Kelley, 2007c), for unstandardized regression coefficients using the `ss.aipe.rc()` and for standardized regression coefficients using the `ss.aipe.src()` function (Kelley & Maxwell, 2008), and the standardized ANOVA and ANCOVA contrasts using the `ss.aipe.sc()` and `ss.aipe.sc.ancova()` functions (Lai & Kelley, 2007). The available functions will be augmented as the AIPE approach is extended to other effect sizes.

As an example of the AIPE approach to sample size planning, suppose a researcher would like to have a 99% degree of assurance that the width of the 95% confidence interval for the standardized mean difference will be no wider than .30 units. For the hypothesized population value (`delta`) of .50, the necessary sample size can be obtained with the `ss.aipe.smd()` function as follows:

```
R> ss.aipe.smd (delta=.50, conf.
level=.95, width=.30, degree.of.
assurance=.99),
```

which yields a necessary sample size of 362 (Kelley & Rausch, 2006). Thus, with 362 participants per group, a researcher can have 99% assurance that the width of the confidence interval for δ will be no larger than .30 units. The other functions that implement the AIPE approach to sample size planning are used in an analogous fashion.

Planning sample size usually requires one to specify one or more population parameters as if they were known exactly. In almost all circumstances, the estimated population value will differ from the actual population value. This affects the appropriateness of the sample size because of the misspecified population value. A series of functions for many of the sample size functions discussed have corresponding sensitivity analysis functions, where the effects of misspecifying the population value(s) and the effect of a particular sample size can be determined. The effect of misspecifying the population value(s) is determined by specifying a particular population effect size and an estimated effect size(s). Another option is to specify a particular sample size to assess the effect of sample size on properties of the confidence intervals. A Monte Carlo simulation study is conducted within the functions for the cases discussed, on the basis of the specifications given. When the specified effect size and the population effect size differ, the sample size is based on the specified effect size given, but the data are generated from a distribution where the effect size is the population value specified. This allows one to determine the effect of misspecification of the population value on the characteristics of the results (e.g., proportion of confidence intervals less than some desired width, mean or median confidence interval width, proportion of confidence intervals that do not contain zero, etc.). In some power analytic contexts, this information is available analytically. However, it is generally not available for applications of AIPE. The same is true when specifying a particular value of sample size to discern the effect on the characteristics of the results. The functions that implement a sensitivity analysis in the AIPE context within MBESS are `ss.aipe.smd.sensitivity()` for the standardized mean difference, `ss.aipe.R2.sensitivity()` for the squared multiple correlation coefficient, `ss.aipe.rc.sensitivity()` for an unstandardized regression coefficient, `ss.aipe.src.sensitivity()` for a standardized regression coefficient, `ss.aipe.cv.sensitivity()` for the coefficient of variation, and `ss.aipe.sc.sensitivity()` and `ss.aipe.sc.ancova.sensitivity()` for ANOVA and ANCOVA standardized contrast, respectively.

Sample size planning for the power analytic approach is available for certain tests in R by default and with other specialized packages (e.g., the pwr package for some general linear model tests [Champely, 2006], and asypow for the asymptotic power of likelihood tests [Brown, Lovato, Russel, & Halvorsen, 2006], and MBESS). R, MBESS, pwr, and asypow together provide a large set of functions that can be used to plan an appropriate sample size for a wide variety of effects under the power analytic perspective (e.g., the effects discussed in Cohen, 1988).

Installation and Help

As mentioned, R is available for essentially all modern operating systems (e.g., Microsoft Windows, Macintosh, Linux, Unix). At the R Project Web page (www.r-project.org/), there is a link to CRAN (cran.r-project.org/), where R and the available add-on packages can be downloaded. Actually, after downloading and installing R, either in source form or in executable format, MBESS and other packages can easily be installed with the automatic package installation feature from within a Microsoft Windows environment. From R's toolbar, the Package option allows packages to be installed with the Install Package(s) option. From the Install Package(s) option, the desired package(s) can be selected and installed (after a download site has been selected).

Although this article introduces many of the important and useful functions contained within MBESS, there are optional specifications not outlined in the article. A set of help files also accompanies MBESS. For any function in MBESS (or R more generally), the help file can be displayed with the help function, `help()`. For example, the associated help files for the `ci.smd()` function can be called on using

```
R> help(ci.smd).
```

Alternatively, one can search for functions and help files by using the `help.search()` function. For example, suppose one is interested in functions that pertain to the standardized mean difference, but is not aware of what those functions are. The `help.search()` function could be called on as

```
R> help.search("standardized mean
difference"),
```

where functions that mention the string in quotes are returned (seven functions from MBESS, in this case). The R Project Web site (R Development Core Team, 2007c) also has a search feature that can be useful when learning what functions relate to different methods or techniques. On the R Project Web site (R Development Core Team, 2007c), there are many freely available manuals, books, and technical reports (under the Manuals and Other links,

respectively, in the Documentation section). Of particular importance is "An Introduction to R" (Venables, Smith, & The R Development Core Team, 2007), which is a very informative manual for users of R. There are also several active R mailing lists, which provide a forum for discussion related to R. Furthermore, Kelley, Lai, and Wu (2008) provide an R tutorial that also uses MBESS, with examples of statistical methods useful for researchers working in the behavioral, educational, and social sciences.

As with all other packages, there is an MBESS page on CRAN that has the most up-to-date released version of the package and its corresponding manual (via the MBESS Web page: cran.at.r-project.org/src/contrib/Descriptions/MBESS.html). Although Microsoft Windows and Macintosh executable files are provided on (most) R package pages, executable files are not provided for Unix or Linux. The source files (i.e., set of files and code that defines MBESS), however, are operating system independent and are available on the MBESS page. Detailed directions on how to install packages from source files for Macintosh, Unix, and Linux are given on the R Web page.

Looking to the Future

I have not outlined all of the features available with MBESS. In addition to the confidence interval formation and sample size planning issues, there are functions to perform other techniques, such as visualization techniques for longitudinal data [see the visualizing individual trajectories functions, vit() and vit. fitted()], visualizing techniques for interactions in a multiple regression context [see the interaction plotting functions intr.plot() for 3-D representations of interactions and intr.plot.2d() for 2-D representations], and the estimation of various effect sizes (as well as their variances and unbiased analogs, where appropriate).

MBESS is an ongoing project that will continue to be developed and expanded so that important quantitative methods can be made available to researchers in the behavioral, educational, and social sciences. The goal is that MBESS will be a useful tool for researchers—both substantive ones and methodologists—in the behavioral, educational, and social sciences. Actually, the methods implemented within MBESS are more general than serving only the needs of researchers working within the behavioral, educational, and social sciences. For example, many of the methods implemented within MBESS are directly applicable to various aspects of business and managerial, biological and medical, and wildlife and fisheries research. MBESS is an open source project, and thus, anyone can obtain and modify the source code in any way that he or she sees fit. Because MBESS is part of the R environment, the methods within MBESS can be used in conjunction with the basic R installation and the many other R packages that are freely available.

AUTHOR NOTE

REFERENCES

ALGINA, J., & OLEJNIK, S. (2000). Determining sample size for accurate estimation of the squared multiple correlation coefficient. *Multivariate Behavioral Research*, **35**, 119-136.

BONETT, D. G., & WRIGHT, T. A. (2000). Sample size requirements for estimating Pearson, Kendall and Spearman correlations. *Psychometrika*, **65**, 23-28.

BROWN, B. W., LOVATO, J., RUSSEL, K., & HALVORSEN, K. (2006). Asymptotic power package (Version 1.2) [Computer software and manual]. Retrievable from www.cran.r-project.org/.

BROWNE, M. W., & CUDECK, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

CHAMPELY, S. (2006). Power package (Version 1.0) [Computer software and manual]. Retrievable from www.cran.r-project.org/.

COHEN, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

COHEN, J. (1994). The earth is round ($p < .05$). *American Psychologist*, **49**, 997-1003.

CUMMING, G., & FINCH, S. (2001). A primer on the understanding, use, and calculation of confidence intervals that are based on central and noncentral distributions. *Educational & Psychological Measurement*, **61**, 532-574.

DE LEEUW, J. (2006). *R in psychometrics and psychometrics in R*. Paper presented at the UseR! conference, Vienna.

DING, C. G. (1996). On the computation of the distribution of the square of the sample multiple correlation coefficient. *Computational Statistics & Data Analysis*, **22**, 345-350.

DORAN, H. C., & LOCKWOOD, J. R. (2006). Fitting value-added models in R. *Journal of Educational & Behavioral Statistics*, **31**, 205-230.

EVERITT, B. S. (2005). *An R and S-Plus companion to multivariate analysis*. New York: Springer.

FOX, J. (2002). *An R and S-Plus companion to applied regression*. Thousand Oaks, CA: Sage.

GATSONIS, C., & SAMPSON, A. R. (1989). Multiple correlation: Exact power and sample size calculations. *Psychological Bulletin*, **106**, 516-524.

GRISSOM, R. J., & KIM, J. J. (2005). *Effect sizes for research: A broad practical approach*. Mahwah, NJ: Erlbaum.

HAHN, G., & MEEKER, W. (1991). *Statistical intervals: A guide for practitioners*. New York: Wiley.

KELLEY, K. (2005). The effects of nonnormal distributions on confidence intervals around the standardized mean difference: Bootstrapping as an alternative to parametric confidence intervals. *Educational & Psychological Measurement*, **65**, 51-69.

KELLEY, K. (2007a). Confidence intervals for standardized effect sizes: Theory, application, and implementation. *Journal of Statistical Software*, **20**, 1-24.

KELLEY, K. (2007b). Methods for the Behavioral, Educational, and Social Sciences (MBESS) [Computer software and manual]. Retrievable from www.cran.r-project.org/.

KELLEY, K. (2007c). Sample size planning for the coefficient of variation from the accuracy in parameter estimation approach. *Behavior Research Methods*, **39**, 755-766.

KELLEY, K. (2007d). *Sample size planning for the squared multiple correlation coefficient: Accuracy in parameter estimation via narrow confidence intervals*. Manuscript submitted for publication.

KELLEY, K., LAI, K., & WU, P.-J. (2008). Using R for data analysis: A best practice for research. In J. Osbourne (Ed.), *Best practices in quantitative methods* (pp. 535-572). Newbury Park, CA: Sage.

KELLEY, K., & MAXWELL, S. E. (2003). Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. *Psychological Methods*, **8**, 305-321.

KELLEY, K., & MAXWELL, S. E. (2008). Sample size planning for multiple regression: Power and accuracy for omnibus and targeted effects. In J. Brannon, P. Alasuutari, & L. Bickman (Eds.), *Sage handbook of social research methods*. Thousand Oaks, CA: Sage.

KELLEY, K., MAXWELL, S. E., & RAUSCH, J. R. (2003). Obtaining power

or obtaining precision: Delineating methods of sample size planning. *Evaluation & the Health Professions*, **26**, 258-287.

KELLEY, K., & RAUSCH, J. R. (2006). Sample size planning for the standardized mean difference: Accuracy in parameter estimation via narrow confidence intervals. *Psychological Methods*, **11**, 363-385.

KRAEMER, H. C., & THIEMANN, S. (1987). *How many subjects?* Beverly Hills, CA: Sage.

KUPPER, L. L., & HAFNER, K. B. (1989). How appropriate are popular sample size formulas. *American Statistician*, **43**, 101-105.

LAI, K., & KELLEY, K. (2007). *Sample size planning for standardized ANCOVA and ANOVA contrasts: Obtaining narrow confidence intervals*. Manuscript submitted for publication.

LEE, Y. S. (1971). Tables of the upper percentage points of the multiple correlation. *Biometrika*, **59**, 175-189.

LIPSEY, M. W. (1990). *Design sensitivity: Statistical power for experimental research*. Newbury Park, CA: Sage.

MEEHL, P. E. (1997). The problem is epistemology, not statistics: Replace significance tests by confidence intervals and quantify accuracy of risky numerical predictions. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 393-426). Mahwah, NJ: Erlbaum.

MURPHY, K. R., & MYORS, B. (2004). *Statistical power analysis: A simple and general model for traditional and modern hypothesis tests* (2nd ed.). Mahwah, NJ: Erlbaum.

PINHEIRO, J., BATES, D., DEBROY, S., & SARKAR, D. (2006). NLME: Linear and nonlinear mixed effects models [Computer software and manual]. Retrieved from www.cran.r-project.org/.

R DEVELOPMENT CORE TEAM (2007a). The comprehensive R archive network. Vienna: R Foundation for Statistical Computing. Available at www.cran.r-project.org.

R DEVELOPMENT CORE TEAM (2007b). R: A language and environment for statistical computing [Computer software and manual]. R Foundation for Statistical Computing. Retrieved from www.r-project.org.

R DEVELOPMENT CORE TEAM (2007c). R Project Internet home page: www.r-project.org/. R Foundation for Statistical Computing.

RENCHER, A. C. (2000). *Linear models in statistics*. New York: Wiley.

ROSENTHAL, R., ROSNOW, R. L., & RUBIN, D. B. (2000). *Contrasts and effect sizes in behavioral research: A correlational approach*. New York: Cambridge University Press.

SAMPSON, A. R. (1974). A tale of two regressions. *Journal of the American Statistical Association*, **69**, 682-689.

SCHMIDT, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods*, **1**, 115-129.

SMITHSON, M. (2001). Correct confidence intervals for various regression effect sizes and parameters: The importance of noncentral distributions in computing intervals. *Educational & Psychological Measurement*, **61**, 605-632.

SMITHSON, M. (2003). *Confidence intervals*. Thousand Oaks, CA: Sage.

STEIGER, J. H. (2004). Beyond the *F* test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods*, **9**, 164-182.

STEIGER, J. H., & FOULADI, R. T. (1997). Noncentrality interval estimation and the evaluation of statistical methods. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 221-257). Mahwah, NJ: Erlbaum.

THOMPSON, B. (2002). What future quantitative social science research could look like: Confidence intervals for effect sizes. *Educational Researcher*, **31**, 25-32.

VENABLES, W. N., SMITH, D. M., & THE R DEVELOPMENT CORE TEAM. (2007). *An introduction to R*. Vienna: R Development Core Team.

WILKINSON, L., & THE AMERICAN PSYCHOLOGICAL ASSOCIATION TASK FORCE ON STATISTICAL INFERENCE (1999). Statistical methods in psychology: Guidelines and explanations. *American Psychologist*, **54**, 594-604.

NOTES

1. All R code is typeset using a Courier font similar to that of a typewriter. R code that is directly executed is preceded with the R> prompt.

2. Notice that even though the observed effect size was essentially "medium" ($\delta = .50$) in magnitude according to Cohen's (1988) recommendations, the lower limit was smaller than "small" ($\delta = .20$) and the upper limit was larger than "large" ($\delta = .8$)! This particular example illustrates the wide range of plausible values (i.e., those contained within the confidence interval) for some point estimates. The sample size used yields a statistical power of .80 under the assumption that the effect is truly medium in the population. As can be seen, when there is sufficient statistical power in the study design and even when statistical significance is reached, the confidence interval may be very wide. This is an issue dealt with in the next section, where sample size planning is discussed with the goal of obtaining narrow confidence intervals.

3. The distribution theory of the observed squared multiple correlation coefficient is different when the null hypothesis is false for regressors that are random versus regressors that are fixed. See, for example, Sampson (1974), Gatsonis and Sampson (1989), or Rencher (2000) for information regarding the distributional differences, Algina and Olejnik (2000), Ding (1996), or Lee (1971) for information on confidence interval formation for the random predictor case, and Smithson (2001) or Steiger (2004) for information on confidence interval formation in the case of fixed regressors. The `ci.R2()` function implements the confidence interval methods under both cases by specifying `Random.Regressors=TRUE` (the default) or `FALSE`, as appropriate, given the design of the study.