Bootstrapping the Standard Error of the Mediated Effect

Chondra M. Lockwood, Arizona State University
David P. MacKinnon, Arizona State University

ABSTRACT

Mediation analysis seeks to go beyond the question of whether an independent variable causes a change in a dependent variable. Mediation addresses the question of *how* that change occurs. When a third variable is thought to be intermediate in the relationship between two variables, it is called a mediator. In order to test the mediated effect for significance, or to create confidence limits for the effect, the standard error of the effect is needed. Estimates of the standard error can fluctuate widely when sample size is small or when the variables are not normally distributed. Bootstrapping is a method that resamples from an original sample to derive a more accurate estimate than is found through traditional methods. This paper describes a SAS program that estimates the mediated effect of a sample, takes bootstrap samples, and calculates the standard error and confidence limits of the mediated effect. The audience should have an understanding of multiple regression analysis. This program was written with base SAS software and operates in a regular PC environment.

INTRODUCTION

A primary purpose of research is to identify the relationship between two variables, typically referred to as the independent and dependent variables. Often theory suggests that a third variable may improve understanding of the nature of the relationship between the two primary variables. When the third variable is considered a mediator, it is hypothesized to be linked in a causal chain between the independent and dependent variables (James & Brett, 1984). In other words, the independent variable causes the mediator and the mediator causes the dependent variable. The search for intermediate causal variables is called mediation analysis. Mediation analyses are common in social science research, as they elaborate upon other relationships. For example, programs to reduce drug use often target mediators such as resistance skills, hypothesizing that a program that increases resistance skills will decrease drug use (MacKinnon & Dwyer, 1980).

Mediation

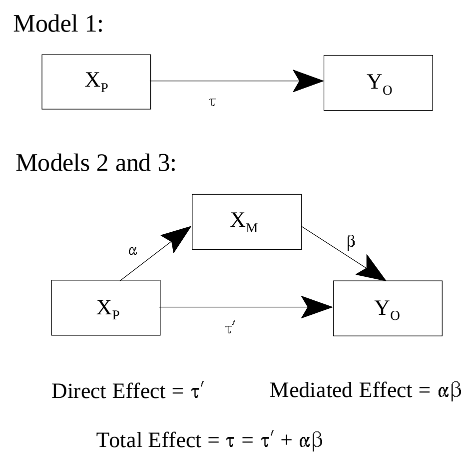
Evidence for mediation occurs when the relationship between two variables can be partially or totally accounted for by an intervening variable, the mediator. The significance of the mediated effect can be tested with a series of regression equations. The mediation model is shown in Figure 1 and is summarized in the following equations, using the notation from MacKinnon, Warsi, and Dwyer (1995).

$$\text{Model 1:} \quad Y_O = \tau X_P + \varepsilon_1$$

$$\text{Model 2:} \quad Y_O = \tau' X_P + \beta X_M + \varepsilon_2$$

Y_O is the outcome or dependent variable, X_P is the program or independent variable, X_M is the mediator, τ codes the relationship between the program to the outcome in the first equation, τ' is the coefficient relating the program to the outcome adjusted for the effects of the mediator, ε_1 and ε_2 code unexplained variability, and the intercept is assumed to be zero. In the first regression, the outcome variable is regressed on the independent variable. In the second regression, the outcome is regressed on the independent variable and the mediator.

Figure 1. The Mediation Model



The mediated effect can be thought of as the difference between the program coefficients ($\tau - \tau'$) in the two regression equations (Judd & Kenny, 1981). Another way to estimate the mediated effect involves Model 2, above, and Model 3, below.

$$\text{Model 3:} \quad X_M = \alpha X_P + \varepsilon_3$$

The α parameter from Model 3 relates the program to the mediator. The β parameter from Model 2 relates the mediator to the outcome. The product of the parameters ($\alpha\beta$) is the mediated or indirect effect. The direct effect is represented by the coefficient that relates the treatment variable to the outcome after adjustment for the mediator (τ'). Computation of the mediated effect in this way reflects the extent to which the program changes the mediator (α) and the extent to which the mediator affects the outcome variable (β). It can be derived mathematically that the $\alpha\beta$ estimate and the τ-τ' estimate are identical (MacKinnon, Warsi, & Dwyer, 1995).

The mediated effect can be tested for significance with an estimate of its standard error. The most commonly used estimate of the standard error ($\hat{\sigma}_{\alpha\beta}$) is the approximate formula derived by Sobel (1982) using the multivariate delta method, $\sqrt{\hat{\alpha}^2\hat{\sigma}^2_\beta + \hat{\beta}^2\hat{\sigma}^2_\alpha}$. The mediated effect divided by its standard error yields a z-score of the mediated effect. Typically, if the z-score is greater than 1.96 we conclude the effect is larger than would be expected by chance and call the effect significant. However, simulation studies of the testing of statistical significance using $\hat{\alpha}\hat{\beta}/\hat{\sigma}_{\alpha\beta}$ suggests that this formula may not be accurate due to the incorrect assumption that this function is normally distributed (MacKinnon & Lockwood, 1998).

The standard error can also be used to construct confidence intervals around the mediated effect. Ninety-five percent confidence intervals are calculated by adding and subtracting the product of 1.96 and the standard error from the mediated effect.

Bootstrapping

Traditional statistical inference is based on assumptions about the distribution of the population from which the sample is drawn. In standard multiple regression analysis, the estimates of the standard errors of parameter estimates are derived based on the assumption of normality. Bootstrapping uses computer intensive resampling to make inferences rather than making assumptions about the population. In other words, bootstrapping treats a given sample as the population.

Bootstrapping can be used for any statistic, but this paper focuses on the mediated effect, its standard error, and its confidence limits (also see Bollen & Stine, 1990, for a treatment of bootstrapping a mediated effect). The steps of a bootstrap procedure outlined by Mooney and Duval (1993, p. 10) are as follows:

1. Construct an empirical probability distribution from the sample in which the probability of each observation is 1/*n*, where *n* is sample size.

2. Draw a random sample with replacement from the empirical probability distribution, a "resample" of size *n*. In other words, an observation is drawn at random into the resample, but the observation also remains in the pool with the possibility of being drawn again. This procedure is repeated until you have *n* observations.

3. Calculate the statistic, $\hat{\theta}$, from the resample. In this case, the statistic of interest is the mediated effect ($\alpha\beta$).

4. Steps 2 and 3 are repeated some large number of times (*B*). Efron and Tibshirani (1985) recommend that *B* should be at least 50-200 when estimating the standard error of $\hat{\theta}$, and at least 1000 when constructing confidence intervals around $\hat{\theta}$.

5. Finally, construct a probability distribution from the all of the resampled estimates, with a probability of 1/*B* for each $\hat{\theta}^*_b$. This distribution is the bootstrapped estimate of the sampling distribution of $\hat{\theta}$.

The probability distribution constructed in step 5 becomes the basis for making inferences. Thus, the standard deviation of this distribution is the bootstrapped standard error of the mediated effect. This sampling distribution of $\hat{\theta}$ can be used to calculate confidence intervals. For a 95% confidence interval, the lower limit will be at the 2.5th percentile and the upper limit will be at the 97.5th percentile.

SAMPLE DATA

Samples representing four effect sizes (null, small, medium, and large) and three sample sizes (N=50, 100, and 1000) were generated. The α and β paths were simulated to be zero for the null effect, .14 for the small effect, .39 for the medium effect, and .59 for the large effect. In each case the direct effect (τ') was set to zero. With four effect sizes and three sample sizes, twelve samples were simulated. For each of these samples the mediated effect and its standard error were estimated as described in the mediation section. One thousand bootstrap samples were then taken from each sample, and each of these was estimated as described in the bootstrapping section.

As shown in Table 1, the bootstrap estimate of the standard error is sometimes quite different than the multivariate delta estimate. For example, the sample of 50 with a small mediated effect has a delta estimate of the standard error of .0250. The bootstrap standard error is .0330, over 30% larger. The general tendency was for the bootstrap standard error to be larger than the delta estimate of the standard error.

Table 1. Classical and Bootstrap Estimates of the Mediated Effect and Its Standard Error

		Delta		Bootstrap	
Model	Sample Size	Estimate	SE	Average	SE
Null	50	0.0152	0.0498	-0.0085	0.0516
	100	0.0014	0.0229	-0.0044	0.0300
	1000	-0.0001	0.0011	0.0005	0.0015
Small	50	0.0168	0.0250	0.0188	0.0330
	100	0.0294	0.0233	0.0327	0.0275
	1000	0.0201	0.0063	0.0212	0.0066
Medium	50	0.0098	0.0432	0.0155	0.0433
	100	0.2388	0.0652	0.2203	0.0629
	1000	0.1459	0.0174	0.1506	0.0175
Large	50	0.1439	0.0718	0.1435	0.0718
	100	0.3898	0.0906	0.4205	0.0853
	1000	0.3637	0.0274	0.3749	0.0289

Note: Bootstrap estimates are based on 1000 bootstrap samples.

Table 2. Classical and Bootstrap Estimates of Confidence Limits of the Mediated Effect

		Delta		Bootstrap	
Model	Sample Size	LCL	UCL	LCL	UCL
Null	50	-0.0825	0.1128	-0.1171	0.1003
	100	-0.0435	0.0462	-0.0654	0.0602
	1000	-0.0024	0.0021	-0.0023	0.0042
Small	50	-0.0322	0.0658	-0.0395	0.0942
	100	-0.0162	0.0751	-0.0076	0.0957
	1000	0.0078	0.0324	0.0100	0.0351
Medium	50	-0.0749	0.0945	-0.0650	0.1204
	100	0.1109	0.3666	0.1052	0.3547
	1000	0.1119	0.1800	0.1171	0.1876
Large	50	-0.0125	0.3003	0.0198	0.3000
	100	0.2121	0.5675	0.2570	0.5950
	1000	0.3099	0.4175	0.3197	0.4312

Note: LCL is the lower confidence limit and UCL is the upper confidence limit. Bootstrap estimates come from 1000 bootstrap samples.

Table 2 contains the confidence limits from delta estimation and from the bootstrapping procedure. The bootstrap confidence intervals tend to be larger than the delta confidence intervals. This difference reflects the difference in the standard error estimates. What is more interesting is that the bootstrap confidence intervals are not symmetric. This illustrates what is perhaps the greatest advantage of bootstrap estimation. With large sample sizes, the delta estimates of the standard error are virtually identical to the bootstrap estimates. The bootstrap confidence intervals, however, remain asymmetrical even with large samples. The classical technique of constructing confidence intervals assumes a normal distribution of the mediated effect, forcing symmetric confidence intervals.

BOOTME PROGRAM

The BOOTstrap of MEdiation program for SAS® is found in Appendix A. The BOOTME program will:

1. call in the original sample to serve as the empirical probability distribution,

2. generate a bootstrap resample,

3. estimate the resample and save the results,

4. repeat the first three steps for the desired number of bootstrap resamples, and

5. use the probability distribution of the resample estimates to calculate the standard error and confidence interval of the mediated effect.

The first part of the program reads in an existing sample and calculates the mediated effect and an estimate of the standard error, using output from the REG procedure. The bootstrap loop is accomplished by using the %DO, %END statements in the MACRO procedure. PROC MEANS calculates the average estimate of the mediated effect and the bootstrap standard error. Confidence intervals are obtained by constructing a bootstrap distribution using PROC IML. The output consists of the estimates of the mediated effect, standard error, and confidence limits from the original sample, as well as each corresponding bootstrap estimate.

This program can be easily adapted to bootstrap any other statistic by simply replacing the estimation portions with the desired computations.

CONCLUSION

The sample data indicate a tendency for bootstrap estimates of the standard error of the mediated effect to be somewhat larger than the corresponding delta estimate, and for the bootstrap confidence intervals to be asymmetric. This is consistent with Bollen and Stine's (1990) examples of bootstrapping mediated effects, in which they used three previously published data sets. Taken together, these results show that bootstrapping the standard error is indicated when sample size is low and that bootstrapping is indicated whenever confidence intervals will be constructed around the estimate of the mediated effect. The bootstrapping procedure may be especially useful in situations where the observed data are not normally distributed.

REFERENCES

Bollen, K. A., & Stine, R. (1990). Direct and indirect effects: Classical and bootstrap estimates of variability. *Sociological Methodology*, *20*, 115-140.

Efron, B., & Tibshirani, R. (1985). The bootstrap method for assessing statistical accuracy. *Behaviormetrika*, *17*, 1-35.

James, L. R., & Brett, J. M. (1984). Mediators, moderators and tests for mediation. *Journal of Applied Psychology*, *69*, 307-321.

Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review, 5*, 602-619.

MacKinnon, D. P., & Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review*, *17*, 144-148.

MacKinnon, D. P., & Lockwood, C. M. (1998). *The power to detect mediated effects in experimental and correlational studies*. Manuscript submitted for publication.

MacKinnon, D. P., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research*, *30*, 41-62.

Mooney, C. Z., & Duval, R. D. (1993). *Bootstrapping: A nonparametric approach to statistical inference*. Newbury Park, CA: Sage Publications, Inc.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology*, *12*, 290-312.

ACKNOWLEDGEMENT

This research was supported by the National Institute on Drug Abuse grant number 1 R01 DA09757.

FOR FURTHER INFORMATION CONTACT:

Chondra M. Lockwood
Department of Psychology, Box 871104
Arizona State University
Tempe, AZ 85287
(602) 727-6127
chondra@asu.edu

David P. MacKinnon
Department of Psychology, Box 871104
Arizona State University
Tempe, AZ 85287
(602) 727-6126
davidpm@asu.edu

APPENDIX A

```
LIBNAME SAMPLE 'C:\BTSTRAP\';
%MACRO BTSTRAP(NBOOT,FILE,NOBS);

*========= call in original sample ============;

DATA ORIGSAMP; SET SAMPLE.&FILE;

*========= estimate original sample ==========;

PROC REG DATA=ORIGSAMP COVOUT
OUTEST=OUT1 MSE NOPRINT;
MODEL M = X;

DATA B1; SET OUT1;
IF _TYPE_='PARMS';
RSQ1=_RSQ_ ; A=X; KEEP RSQ1 A;
DATA COV1; SET OUT1;
IF _NAME_='X';
VARA=X; KEEP VARA;

PROC REG DATA=ORIGSAMP COVOUT
OUTEST=OUT2 MSE NOPRINT;
MODEL Y = X M;

DATA B2; SET OUT2;
IF _TYPE_='PARMS';
RSQ2=_RSQ_; B=M; G=X; MSE=_MSE_;
KEEP RSQ2 B G MSE;
DATA COV2; SET OUT2;
IF _NAME_='M';
VARB=M; COVBG=X; KEEP VARB COVBG;
DATA COV3; SET OUT2;
IF _NAME_='X';
VARG=X; KEEP VARG;

PROC REG DATA=ORIGSAMP COVOUT
OUTEST=OUT3 MSE NOPRINT;
MODEL Y = X;

DATA B3; SET OUT3;
IF _TYPE_='PARMS';
RSQ3=_RSQ_; XONLY=X; KEEP RSQ3 XONLY;
DATA COV4; SET OUT3;
IF _NAME_='X';
VARXONLY=X; KEEP VARXONLY;

DATA ALL; MERGE B1 COV1 B2 COV2 COV3 B3
COV4;
MED=A*B;
SEMED=SQRT(VARB*A*A+VARA*B*B);
LCL=MED-(1.96*SEMED);
UCL=MED+(1.96*SEMED);
DATA ORIG; SET ALL;
KEEP MED SEMED LCL UCL;

PROC DATASETS LIB=WORK MT=DATA NOLIST;
DELETE B1 B2 B3  COV1 COV2 COV3 COV4 OUT1
OUT2 OUT3 ALL;

*========== take bootstrap samples ===========;

DATA BTSTRAP; SET _NULL_;
%DO OBS=1 %TO &NBOOT;
DATA BOOT;
DO OBS=1 TO &NOBS;
NUM=MOD(RANNOR(0),1);
IF NUM<0 THEN NUM=-NUM;
SELECT=CEIL((NUM)*(&NOBS));
SET ORIGSAMP POINT=SELECT;
IF _ERROR_=1 THEN STOP;
OUTPUT;
END;
STOP;
RUN;

*======== estimate each bootstrap sample =======;

PROC REG DATA=BOOT COVOUT OUTEST=OUT1
MSE NOPRINT;
MODEL M = X;

DATA B1; SET OUT1;
IF _TYPE_='PARMS';
RSQ1=_RSQ_ ; A=X; KEEP RSQ1 A;
DATA COV1; SET OUT1;
IF _NAME_='X';
VARA=X; KEEP VARA;

PROC REG DATA=BOOT COVOUT OUTEST=OUT2
MSE NOPRINT;
MODEL Y = X M;

DATA B2; SET OUT2;
```

```
IF _TYPE_='PARMS';
RSQ2=_RSQ_; B=M; G=X; MSE=_MSE_;
KEEP RSQ2 B G MSE;
DATA COV2; SET OUT2;
IF _NAME_='M';
VARB=M; COVBG=X; KEEP VARB COVBG;
DATA COV3; SET OUT2;
IF _NAME_='X';
VARG=X; KEEP VARG;

PROC REG DATA=BOOT COVOUT OUTEST=OUT3
MSE NOPRINT;
MODEL Y = X;

DATA B3; SET OUT3;
IF _TYPE_='PARMS';
RSQ3=_RSQ_; XONLY=X; KEEP RSQ3 XONLY;
DATA COV4; SET OUT3;
IF _NAME_='X';
VARXONLY=X; KEEP VARXONLY;

DATA ALL; MERGE B1 COV1 B2 COV2 COV3 B3
COV4;
BMED=A*B;

DATA NEW; SET BTSTRAP;
DATA BTSTRAP; SET NEW ALL;

%END;

PROC MEANS DATA=BTSTRAP MEAN STD
NOPRINT;
VAR BMED;
OUTPUT OUT=OUT;

DATA STRAP1A; SET OUT;
IF _STAT_='MEAN';
BMEANMED=BMED;
KEEP BMEANMED;
DATA STRAP1B; SET OUT;
IF _STAT_='STD';
BSEMED=BMED;
KEEP BSEMED;
DATA STRAP1; MERGE STRAP1A STRAP1B;

DATA STRAP2; SET BTSTRAP;
PROC FREQ;
TABLE BMED/ NOPRINT  OUT=OUT;
DATA B; SET OUT;
PROC IML;
USE B;
READ ALL VAR {BMED COUNT} INTO BMED;
TABLE1=J(&NBOOT,2,0);
PERC=1;
A=0;
COUNT=0;
DO PERC=1 TO &NBOOT;
  A=A+1;
  DIV=&NBOOT/100;
  COUNT=COUNT+BMED[A,2];
  TABLE1[PERC,1]=BMED[A,1];
  TABLE1[PERC,2]=COUNT/DIV;
END;
LABELS1={BMED CUMPCT};
CREATE BOOTME FROM
TABLE1[COLNAME=LABELS1];
APPEND FROM TABLE1;
QUIT;
DATA C; SET BOOTME;
IF CUMPCT=2.5 THEN BLCI=BMED;
IF CUMPCT=97.5 THEN BUCI=BMED;
DATA D; SET C;
IF BLCI NE '.';
KEEP BLCI;
DATA E; SET C;
IF BUCI NE '.';
KEEP BUCI;
DATA STRAP2; MERGE D E;
DATA STRAP; MERGE STRAP1 STRAP2;

*===== output classical and bootstrap estimates ====;

DATA FINAL; MERGE ORIG STRAP;
PROC PRINT;

%MEND;

*===== provide # of bootstrap samples, ==========;
*===== file that contains original data,  ==========;
*===== and sample size                     ==========;

%BTSTRAP(NBOOT=1000,FILE=TEST,NOBS=50);

RUN;
QUIT;
```