Global Scientific JOURNALS

GSJ: Volume 9, Issue 11, November 2021, Online: ISSN 2320-9186

www.globalscientificjournal.com

TESTS OF MEDIATIONAL EFFECTS ON SAMPLE SIZE USING THE BOOTSTRAPPING APPROACH

[1]Tuggen O. Edward, [2]Agaviezor Kingsley and [3]Ewhrudjakpor Akpojotor

[1,2,3]Foundations/General Studies Unit,

Delta State School of Marine Technology Burutu Nigeria.

Corresponding author

edtuggen@yahoo.com

Abstract

In this paper, we considered the three widely used versions of tests of indirect effect vis-à-vis Sobel, Aroian and Goodman's test. Their test statistics are ratios of indirect effect and standard error of indirect effect. The issue of determining the sample size for these mediational tests is the major work here, baring the associated issues in recent times. This study is aimed at determining the sample size that gives the best mediational effect and the best test of indirect effect based on sample size. The bootstrapping approach was employed for various sample sizes. The comparison of the two methods of effect size calculation shows that effect size is not affected by sample size. The comparison of the three test across the various sample size show that sample size of 100 is the optimum sample size for test of indirect effect. The significance of the tests of indirect effect was observed to increase with sample sizes but reduces at n=200. Statistically, the version of test that is seen on the average to have the highest value of test statistics is the Goodman test.

Keywords: bootstrap, effect size, sample size, mediation.

1.0 Introduction

The term statistical mediation refers to a causal chain in which it is assumed that the effect of one or more independent variables is transmitted to one o more dependent variables through third variables. In the simplest case, the term mediation is used to indicate that the effect of an independent variable (X) is transmitted to a dependent variable (Y) through a third mediator variable (M).

Fairchild & MacKinnon (2009) noted that a mediation model is one that seeks to identify and explain the mechanism or process that underlies an observed relationship between an independent variable and a dependent variable via the inclusion of a third explanatory variable, known as a mediator variable. Rather than hypothesizing a direct causal relationship between the independent variable and the dependent variable, a meditational model hypothesizes that the independent variable influences the mediator variable, which in turn influences the dependent variable. Thus, the mediator variable serves to clarify the nature of the relationship between the independent and dependent variables

Often times, theory suggests that a third variable may improve understanding of the nature of the relationship between the two main variables. This third variable is hypothesized to be linked in a causal chain between the independent and dependent variables. In other words, the independent variable causes the mediator and the mediator causes the dependent variable. The search for these causal variables is called mediation analysis.

The statistical analysis of mediation effects has become an indispensable tool for helping scientists investigate processes thought to be causal. Yet, in spite of many recent advances in the estimation and testing of mediation effects, little attention has been given to methods for communicating effect size and the practical importance of those effect sizes.

According to MacKinnon & Dwyer (1993), the indirect effect in the single - mediator model may be calculated in two ways, as either $\hat{a}\hat{b}$ or $\hat{c} - \hat{c}'$. The value of the mediated or indirect effect estimated by taking the difference in the coefficients, $\hat{c} - \hat{c}'$, corresponds to the reduction in the independent variable effect on the dependent variable when adjusted for the mediator. The difference between the coefficients obtained from the two different regression coefficients $(\hat{c} - \hat{c}')$ is equal to the product of the coefficients ab.

To test for significance of indirect effect, the difference is then divided by the standard error of the difference and the ratio is compared to a standard normal distribution.

There are three widely used versions of tests of indirect effect vis-à-vis Sobel, Aroian and Goodman’s test. Their test statistics are ratios of indirect effect and standard error of indirect effect. Aroian (1947) version adds the third denominator term,$S_a^2 S_b^2$; Goodman (1960) subtracts the third denominator term and one that does not include the third variable at all generally referred to as Sobel test by Sobel (1982). Aroian version was recommended in Baron and Kenny (1986) because it does not make the unnecessary assumption that the product of S_a and S_b is vanishingly small. The Goodman version of the test subtracts the third term for an unbiased estimate of the variance of the mediated effect, but according to Preacher &Leonardelli (2010), this can sometimes have the unfortunate effect of yielding a negative variance estimate.

The issue of determining the sample size for mediational test has been one of the mediational issues in recent days. Even when the sample size is not readily visible in the standard errors of the tests of indirect effect, Preacher and Kelly (2011) stated sample size was already introduced in the denominator of Sobels (1982) test statistic by the inclusion of S^2. This still applies to both Arioan and Goodmans test. The S^2 refers to the variance of paths and variance of coefficient of the mediator in the partial effect model. Hence it is important to determine the best sample size for the mediational effect not from the level of the test, but at the point of the various regression analysis. Furthermore some sample size calculations that already exist are done on predetermined effect sizes. This brings about some restraints on such work as the results are only applicable to the predetermined effect size.

The Sobel test is more accurate than the Baron and Kenny steps explained above, however it does have low statistical power. As such, large sample sizes are required in order to have sufficient power to detect significant effects. This is because the key assumption of Sobel's test is the assumption of normality. Because Sobel's test evaluates a given sample on the normal distribution, small sample sizes and skewness of the sampling distribution can be problematic.

The bootstrap method is becoming the most popular method of testing mediation because it does not require the normality assumption to be met. Bootstrapping involves repeatedly randomly sampling observations with replacement from the data set to compute the desired statistic in each resample. Over hundreds, or thousands, of bootstrap resamples provide an approximation of the sampling distribution of the statistic of interest.

Preacher and Hayes (2004) bootstrapping method provides some advantages to the Sobel's test, mainly because it gives an increased power. The Preacher and Hayes Bootstrapping method is a non-parametric test. As such, the bootstrap method does not violate assumptions of normality and is therefore recommended for small sample sizes. But this non parametric approach may not be applicable when using the regression model in estimating and testing the significance of the effect size.

The distribution of the product term ab is only normal at large sample sizes, since both α and β are assumed to be normally distributed, and the distribution of the product of two normally distributed variables is skewed, unless the means are much larger than the standard deviations (Judd and Kenny, 1981; Preacher and Hayes, 2008) which means that at smaller sample sizes the p-value that is derived from the formula will not be an accurate estimate of the true p-value. If the sample is large enough this will not be a problem, however determining when a sample is sufficiently large is somewhat subjective.

Purpose of the study

The main objective of the study is to determine the sample size that gives the best test of indirect effect. The specific objectives are:

1. To determine the sample size that gives the best 4editational effect

2. To determine the best test of indirect effect based on sample size

1.2 Literature Review

Several studies on mediation and sample size determination are reviewed.

The interest of mediation is in calculation the size of the mediating effect also known as indirect effect. The indirect effect measures the extent to which the dependent variable changes when the independent variable is held fixed; and the mediator variable changes by the amount it would have changed had the independent variable increased by one unit (Judd & Kenny, 1981)

The indirect effect measures the extent to which the dependent variable changes when the independent variable is held fixed; and the mediator variable changes by the amount it would have changed had the independent variable increased by one unit (Judd & Kenny, 1981)

According to MacKinnon & Dwyer (1993), the indirect effect in the single-mediator model may be calculated in two ways, as either $\hat{a}\hat{b}$ or $\hat{c} - \hat{c}'$. The value of the mediated or indirect effect estimated by taking the difference in the coefficients, $\hat{c} - \hat{c}'$, corresponds to the reduction in the independent variable effect on the dependent variable when adjusted for the mediator. The difference between the coefficients obtained from the two different regression coefficients ($\hat{c} - \hat{c}'$) should be equal to the product of the coefficients ab. To test for significance of indirect effect, the difference is then divided by the standard error of the difference and the ratio is compared to a standard normal distribution.

Many studies investigating mediation according to MacKinnon, Fairchild & Fritz (2007) use a randomized experimental design, where participants are randomized to levels of one or more factors in order to demonstrate a pattern of results consistent with one theory and inconsistent with another theory

Differences in means between groups are then attributed to the experimental manipulation of the mediator. The results of the randomized study along with the predictions of different theories are used to provide evidence for a mediation hypothesis and suggest further studies to localize and validate the mediating process.

James et al. (2006) noted that researchers often test whether there is complete or partial mediation by testing whether the c′ coefficient is statistically significant, which is a test of whether the association between the independent and dependent variable is completely accounted for by the mediator. If the c′ coefficient is statistically significant and there is significant mediation, then there is evidence for partial mediation. It is often unrealistic to expect that a single mediator would be explained completely by an independent variable to dependent variable relation.

From the Baron & Kenny (1986) approach, mediation is supported if the partial direct effect for path c is non significantly different from zero and path b is significantly greater than zero. If c is non-significantly different from zero, results are consistent with a *full* mediational model. If path b is significant after controlling for the direct effect of X (path c), but path c is still significant, the model is consistent with *partial* mediation.

Sherman & Gorkin (1980) randomly assigned subjects to solve either (a) a sex-role related brainteaser, or (b) a brainteaser not related to sex roles. The sexist brainteaser condition was designed to evoke cognitive dissonance in the self-identified feminist subjects, while the nonsex-role related condition was not. Participants were then asked to judge the fairness of a legal decision made in an affirmative action trial. The results were consistent with the prediction that participants with strong feminist beliefs were more likely to make extreme feminist judgments in the trial if they failed the sexist brainteaser task, in an attempt to reduce cognitive dissonance. Although results of this experiment were taken as evidence of a cognitive dissonance mediation relation, the mediating variable of cognitive conflict was not measured to obtain more information on the link between the manipulation, cognitive dissonance, and feminist judgments.

According to Small (2013), researchers are often interested in mediation analysis to understand how a treatment works, in particular how much of a treatment's effect is mediated by an intermediated variable and how much the treatment directly affects the outcome not through the mediator. The standard regression approach to mediation analysis assumes sequential ignorability of the mediator, which is that the mediator is effectively randomly assigned given baseline covariates and the randomized treatment. The author argued that since the experiment does not randomize the mediator, sequential ignorability is often not plausible.

MacKinnon, Lockwood, and Williams (2004) also compared the bootstrap resampling method with the single sample method and found that the bootstrap method obtained more accurate confidence limits. They further suggested that confidence limits of the mediation effects provided much more information than the estimates themselves.

Preacher and Kelly (2011) outline some general desiderata for effect size measures; described current methods of expressing effect size and practical importance for mediation; used the desiderata to evaluate these methods and develop new method to communicate effect size in the context of mediation analysis. The first new effect size index they described was a residual-based index that quantifies the amount of variance explained in both the mediator and the outcome. The second new effect size index quantifies the indirect effect as the proportion of the maximum possible indirect effect that could have been obtained, given the scales of the variables involved. We supplement our discussion by offering easy-to-use R tools for the numerical and visual communication of effect size for mediation effects.

Imai*et al* (2012) developed a general approach to mediation that offers the definition, identification, estimation, and sensitivity analysis of causal mediation effects without reference to any specific statistical model. Further, their approach explicitly links these 4 elements closely together within a single framework. As a result, the proposed framework can accommodate linear and nonlinear relationships, parametric and nonparametric models, continuous and discrete mediators, and various types of outcome variables. The general definition and identification result also allow for the development of sensitivity analysis in the context of commonly used models, which enables applied researchers to formally assess the robustness of their empirical conclusions to violations of the key assumption.

2.0 Methodology

Given the dependent variable(Y), independent variable (X) and the suspected mediator variable (M), then according to MacKinnon (2000), the test of indirect effect makes use the following three regression equations:

$$Y=c_0+cX+e_1, \qquad (1)$$

$$M_i=\alpha_0+aX+e_2 \qquad (2)$$

$$Y=c'_0+ c'X+ bM+ e_3 \qquad (3)$$

The bootstrap approach is employed in the estimation of these models. This involved resampling and estimating the required parameters over 5000 samples for sample sizes of 5, 10, 20, 30, 50, 100 and 200 so as to determine the sample size that gives the best test of indirect effect. The tests considered are Sobels test, Aroian Test and Goodman test of indirect effect. The data used was the simulated, and the regression equations as well as bootstrapping was done using the SPSS version 20 (see: Igweze & Etaga, 2011 for details)

2.1 Test of Significance of the Indirect Effect

A test of significance of the indirect effect can be constructed using a ratio of the indirect coefficient to its standard error.

The indirect effect is obtained by subtracting coefficient for X in equation (3) (depending on the number of mediator) from the coefficient of X in equation (1): that is

$$b_{indirect} = c - c' \qquad (4a)$$

According to Sobel (1982), an equivalent way to estimate the indirect effect, is to multiply the coefficient of X in equation (2) and the coefficient of Z in equation (3) as follows:

$$b_{indirect} = ab \qquad (4b)$$

The tests of indirect effect tis given as:

$$Z_{indirect} = \frac{b_{indirect}}{S_{(b_{indirect})}} \qquad (5)$$

Where $S_{(b_{indirect})}$ is the standard error for the Sobel test and is given as

$$S_{(b_{indirect})} = \sqrt{b^2S_a^2 + a^2S_b^2} \qquad (6)$$

Aroian test;

$$S_{(b_{indirect}\ )} = \sqrt{b^2 S_a^2 + a^2 S_b^2 + s_a^2 S_b^2} \quad (7)$$

Goodman test:

$$S_{b_{(indirect}\ )} = \sqrt{b^2 S_a^2 + a^2 S_b^2 - s_a^2 S_b^2} \quad (8)$$

Where: b is the unstandardized coefficient for path *b* and *a is* the unstandardized coefficient for path a, S_a and S_b are standard error for a and b.

3.0 Result and Discussion

Table 1: Analysis of path c

n	c	Sc	P(c)
10	0.334	0.034	0.001
20	0.443	0.024	0
30	0.545	0.03	0
50	0.737	0.044	0
100	1.009	0.035	0
200	0.902	0.01	0

The first step in mediation analysis is to test the significance of path c. this is the coefficient of the independent variable when predicting the dependent variable. This is required to ascertain that there is a significant relationship between the dependent and independent variable. The test for path c for each sample size is presented in table 1. The result shows that there is a significant relationship between the dependent variable and independent variable for all sample sizes.

Table 2: Analysis of path a

n	a	S_a	P(a)
10	1.11	0.171	0.284
20	1.272	0.074	0
30	1.204	0.042	0
50	1.197	0.027	0
100	1.095	0.18	0
200	0.984	0.012	0

Next is the analysis of path a, this is required in order to determine the value of *a*, and the standard error for path a, S_a, which will be used in the calculation of the indirect effect using the product of coefficients. Again it is required to test the ability of the mediator variable to absorb the effect of the independent variable. Here only sample size of 10 had a p-value greater than 0.05.

Table 3: Partial Effect Model

n	c'	b	$S_{c'}$	S_b
10	0.25	0.076	0.497	0.541
20	0.246	0.155	0.131	0.104
30	0.552	-0.006	0.199	0.163
50	0.86	-0.103	0.411	0.338
100	1.69	-0.622	0.229	0.219
200	0.877	0.026	0.127	0.128

The partial effect is a model of both independent variable and the mediator variable predicting the dependent variable. The partial effect model is required to obtain the coefficient of X in the partial effect model known as the direct effect (C′) and the coefficient of *b* to be used in computation of the effect size using a product of coefficient approach.$S_{c'}$ and S_b are the standard error for the direct effect (C′) and b respectively.

Table 4: Comparison of Indirect Effect

N	c	c'	c-c'	a	b	ab
10	0.334	0.25	**0.084**	1.11	0.076	**0.08436**
20	0.443	0.246	**0.197**	1.272	0.155	**0.19716**
30	0.545	0.552	**-0.007**	1.204	-0.006	**-0.00722**
50	0.737	0.86	**-0.123**	1.197	-0.103	**-0.12329**
100	1.009	1.69	**-0.681**	1.095	-0.622	**-0.68109**
200	0.902	0.877	**0.025**	0.984	0.026	0.025584

The comparison of the two methods of computation of indirect effect: product of coefficient and difference of coefficient gave approximately the same result. This shows that the effect size is not affected by sample size.

Table 5: Tests of Indirect Effect

	SOBEL	AROIAN	GOODMAN

10	0.140	0.1388	0.142
20	1.485	1.482	1.487
30	-0.0379	-0.0379	-0.0379
50	-0.2655	0.2654	-0.2655
100	**-2.573**	**-02.5454**	**-2.6025**
200	0.2031	0.2031	0.2031

The comparison of the three test of indirect effect across the various sample sizes shows that significance increases as sample size increases. However the tests were not significant at sample size of 200. On the average the Goodman version of the test gives the highest test statistics value.

4.0 Conclusion

The study on sample size and tests of indirect effect was conducted to determine the effect of sample size on effect size and tests of indirect effect. The comparison of the two methods of effect size calculation shows that effect size is not affected by sample size, as the two methods gave approximately the same result across all sample sizes.

The comparison of the three tests across the various sample size show that sample size of 100 is the optimum sample size for test of indirect effect, as it has the highest level of significance. Significance of the test of indirect effect was observed to increase with sample sizes but reduces at n=200. The Goodman version of the test is seen on the average to have the highest value of test statistics.

References

[1] Fairchild, A. J and MacKinnon, D. P. (2009). A general model for testing mediation and moderation effects. Prevention Science, 10, 87-99.

[2] MacKinnon, D. P., & Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. Evaluation Review, **17**, 144-158.

[3] Goodman, L. A. (1960). On the exact variance of products. Journal of the American Statistical Association, 55, 708-713.

[4] Sobel, M. E. (1982). Asymptotic Confidence Intervals For Indirect Effects In Structural equations models. In S. Leinhart (Ed.), Sociological methodology 1982 (pp. 290-312). San Francisco: Jossey-Bass.

[5] Sobel, M. E. (1986). Some new results on indirect effects and their standard errors in covariance structure models. In N. Tuma (Ed.), Sociological Methodology 1986 (pp. 159-186). Washington, DC: American Sociological Association.

[6] Preacher, C. J &Leonardelli, G. J (2010). Calculation for the Sobel test: An interactive calculation tool for Mediation tests, http://quantpsy.org/sobel/sobel.htm

[7] Preacher, K. J and Kelley, K. (2011). Effect Size Measures for Mediation Models:Quantitative Strategies for Communicating Indirect Effects. Psychological Methods, Vol16(2)pp:93–115

[8] Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. Behavior Research Methods, Instruments, & Computers, **36**, 717-731.

[9] Judd CM, Kenny DA. (1981). Process analysis: Estimating mediation in treatment evaluations. Evaluation. Review. (5) pp:602–619.

[10] Mackinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence Limits For the indirect effect: Distribution of the product and resampling methods. Multivariate Behavioral Research, **39**, 99-128.