Available online at www.sciencedirect.com

SciVerse ScienceDirect

JCP JOURNAL OF CONSUMER PSYCHOLOGY

Journal of Consumer Psychology 22 (2012) 582–594

Methods Dialogue

Mediation analysis and categorical variables: The final frontier☆

Dawn Iacobucci

Owen Graduate School of Management, Vanderbilt University, 401 21st Avenue South, Nashville, TN 37203, USA

Received 22 December 2011; accepted 30 March 2012
Available online 12 April 2012

Abstract

Many scholars are interested in understanding the process by which an independent variable affects a dependent variable, perhaps in part directly and perhaps in part indirectly, occurring through the activation of a mediator. Researchers are facile at testing for mediation when all the variables are continuous, but a definitive answer had been lacking heretofore as to how to analyze the data when the mediator or dependent variable is categorical. This paper describes the problems that arise as well as the potential solutions. In the end, a solution is recommended that is both optimal in its statistical qualities as well as practical and easily implemented: compute $z_{Mediation}$.

Keywords: Mediation analysis; Categorical variables; Binary variables

Introduction

Mediation analysis is popular among behavioral researchers as a means of testing hypothetical processes and mechanisms through which an independent variable, X, might elicit a dependent variable, Y, indirectly through the mediating variable, M (Baron & Kenny, 1986; Iacobucci, 2008; MacKinnon, Fritz, Williams, & Lockwood, 2007; MacKinnon, Krull, & Lockwood, 2000; MacKinnon, Lockwood, Brown, Wang, & Hoffman, 2007; MacKinnon, Lockwood, & Williams, 2004; MacKinnon, Warsi, & Dwyer, 1995). For example, attitude toward an advertisement (X) may enhance attitude toward a brand (M), which in turn may positively impact likelihood to purchase the brand (Y). In this paper, we symbolically note the direct effect of X on Y as "$X \rightarrow Y$" and the indirect effect of X on Y through the mediator M as "$X \rightarrow M \rightarrow Y$."

Mediation tests proposed by Baron and Kenny (1986) in their still immensely popular article are conducted by fitting a series of three regular (i.e., "ordinary least squares," OLS) regressions:

$$\hat{Y} = b_{01} + cX \quad (1)$$

$$\hat{M} = b_{02} + aX \quad (2)$$

$$\hat{Y} = b_{03} + c'X + bM. \quad (3)$$

The significance of the parameter estimate c in Eq. (1) indicates whether there exists a direct impact of X on Y (in $X \rightarrow Y$), and Eqs. (2) and (3) are fit to determine whether there exists an indirect effect of X on Y through the mediator M (in $X \rightarrow M \rightarrow Y$). To determine whether there is a significant mediation effect, the researcher extracts estimate a and its standard error s_a from Eq. (2), and the estimate b and s_b from Eq. (3). Mediation is tested via a z-test (Sobel, 1982):

$$z = \frac{a \times b}{\hat{\sigma}_{ab}} = \frac{a \times b}{\sqrt{b^2 s_a^2 + a^2 s_b^2}}, \quad (4)$$

where the numerator $a \times b$ is the estimated size of the indirect effect, $\hat{\sigma}_{ab}$ is its standard error, and $z \sim N(0,1)$; i.e., there is a

☆ The author is grateful to the editor of the *Journal of Consumer Psychology*, Professor C. W. Park, for his consistent support and for the model he sets for all of us in his enthusiasm for life and scholarship. The author is also grateful to these geeky rock stars for feedback on the research and an early draft of the manuscript: Professors Hans Baumgartner (Penn State), Ulf Böckenholt (Kellogg), Sanjay Dhar (Chicago), William Dillon (SMU), Fred Feinberg (Michigan), Zvi Gilula (Chicago), Dominique Hanssens (UCLA), David MacKinnon (ASU), and Robert Peterson (Texas).
E-mail address: dawn.iacobucci@owen.vanderbilt.edu.

1057-7408/$ -see front matter
doi:10.1016/j.jcps.2012.03.006

significant mediated effect if z exceeds |1.96| for 2-tailed tests with $\alpha=0.05$.[1]

Methodological twists

While the basic logic and analyses have stood the test of time, there has been quite a bit of research aimed at improving the methodology. In particular, MacKinnon and his colleagues have made tremendous contributions to improving both the accuracy and precision in mediation procedures (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002; MacKinnon et al., 1995, 2000, 2004).

Mediation modeling has also been extended to incorporate measurement issues. For example, if M is measured via a scale with items, $m_1, m_2, m_3, \ldots, m_m$ (e.g., a series of 7-point or 9-point attitude items), then a structural equation modeling approach allows the researcher to capture the data more optimally than an alternative such as creating a scale for M based on the average of the items (Iacobucci, 2008). Mediation models have also been generalized to allow for nomological networks that are richer than just the three central constructs, X, M, and Y. If there are additional predictors or consequences of any of these, structural equation models are superior (i.e., mathematically statistically optimal given their smaller standard errors), substantively to get a better sense of the bigger theoretical picture, and statistically because the focal associations will be estimated more purely, having other effects partialed out and statistically controlled (Iacobucci, 2008; Sobel, 1990).

Categorical variables

It may seem that every conceivable extension of mediation analyses has been examined in the literature. Yet an elegant, parsimonious treatment of categorical variables in mediation testing has not yet been adequately addressed. A solution would be highly desirable given the frequency with which categorical variables are used by social scientists such as consumer psychologists, as when brand choice is among the battery of responses captured in a study. The quest for sound methods of incorporating categorical variables is perhaps the last dilemma in mediation analysis that lacks a strong solution—it's the 'final frontier' and the topic of this paper.

Before proceeding, let us be clear in terminology. Some statisticians characterize a variable as "continuous" very narrowly, perhaps as a scale that may manifest in all real numbers, or perhaps only integers if they are sufficient in number. The strict definitions result in including as continuous only such measures as, say: "amount spent in dollars and cents," "lengths in millimeters of a mouse click on a sliding scale shown on a computer screen," or "ratings on a 100-point scale." These statisticians would claim that 5-point, 7-point, or even 9-point rating scales are discrete, and that they yield categorical data.

Those definitions are not the ones adopted in this paper. We consider 5-, 7-, and 9-point rating scales as continuous, to be consistent with the assumptions and treatment of such scales as the majority of scholars in the field, e.g., as when we compute means, standard deviations, and correlations on such variables. By categorical in this paper, we mean variables that yield binary responses (e.g., yes, no), nominal categories (e.g., brands A, B, C, or ethnicity groups, etc.), orders such as brand preference rankings, etc.

The question addressed in this paper is how to conduct mediation analyses on data in which X, M, and/or Y are categorical. For simplicity, take the case of a binary variable.[2] If the independent variable, X, is binary, but M and Y are continuous, the standard techniques (in Eqs. (1)–(4)) are perfectly suitable, because X functions only as a predictor variable in the regression equations. The variable X may be treated as a dummy variable and the mechanics of the mediation analyses remain the same.[3]

However, if it is the case that either the mediator M or the dependent variable Y is binary, then traditional approaches are suboptimal because M serves as a dependent variable in Eq. (2) and Y serves as a dependent variable in Eqs. (1) and (3). Statistical theory tells us that OLS regression is better suited for continuous response variables and logit models are better suited for discrete response variables. Thus, in the scenarios in which M and/or Y are categorical, we know it's not ideal to apply OLS regression.

Researchers have proposed various solutions to working with categorical variables or a mix of categorical and continuous variables in mediation. For example, Hayes and Preacher (2011) focus on X and allow it to be multinomial, not just binary. It would be ideal to allow for categorical M and Y as well.

Other researchers make the assumption that while a manifest variable may be discrete, the underlying construct is continuous. For example, Winship and Mare (1983) and Muthén (1984) proposed modifications to structural equation modeling, of which the mediation form is a special case, for categorical variables. Muthén (1984) assumes a continuous latent variable underlying any manifest categorical variable, with observed categorical data arising through a threshold step function. His estimation procedure is based on generalized least squares, normal distributions are assumed, and large sample sizes are recommended. Winship and Mare (1983) also use the threshold approach or categorical variables, and they assume normal or binomial distributions to support the probabilistic approach to

[1] It is true that the direct path need not be fit to test the indirect path (James & Brett, 1984; Zhao, Lynch, & Chen, 2010). Specifically Eq. (1) tests the direct path, and produces neither of the parameters, a or b required to estimate the indirect path. Nevertheless, it is incomplete to fit the indirect path without also fitting the direct path, because they are obviously competing hypotheses. Indeed, most researchers wish to compare the strengths of the two paths, and do so easily via the usual test z-statistic of the form, $z=\left(Parameter_1 - Parameter_2\right)/\sqrt{\left(s.e._1^2+s.e._2^2\right)}$. Eq. (1) is also necessary if one wished to converge on the test of $a\times b$ by comparing c from Eq. (1) and c' from Eq. (3). Doing so assures that Eqs. (1) and (3) could be fit via the same techniques—both OLS regression or both logistics.

[2] We continue with binary variables without loss of generality in logic; empirically, we could extend to multinomial logits to fit contrasting subsets of categories, e.g., brand A vs. all others, brand A vs. B and C and then B vs. C, etc.

[3] The dummy variable for X may be coded as *0's* and *1's* (or *1's* and *2's*, etc.) or as -1*'s* and *1's* (effects coding). The primary difference is merely being mindful when interpreting the results.

Table 1
All possible combinations of X, M, and Y being continuous or categorical.

Case	X	M	Y	Traditional Mediation Steps:		
				(1) $X \rightarrow Y$ (estimate c)	(2) $X \rightarrow M$ (estimate a)	(3) $X\&M \rightarrow Y$ (estimate b)
				Fit Via:		
i)	Continuous	Continuous	Continuous	SEM */regression	SEM/regression	SEM/regression
ii)	Continuous	Continuous	Categorical	Logistic **	SEM/regression	Logistic
iii)	Continuous	Categorical	Continuous	SEM/regression	Logistic	SEM/regression
iv)	Continuous	Categorical	Categorical	Logistic	Logistic	Logistic
v)	Categorical	Continuous	Continuous	SEM/regression	SEM/regression	SEM/regression
vi)	Categorical	Continuous	Categorical	Logit	SEM/regression	Logistic
vii)	Categorical	Categorical	Continuous	SEM/regression	Logit	SEM/regression
viii)	Categorical	Categorical	Categorical	Logit	Logit	Logit

* SEM stands for "structural equations models".

** Logits and logistic regressions are both applied to categorical dependent variables; a model is called logit if all predictors are also categorical, and logistic if the predictors include any continuous variables.

the manifest categories, and they use nonlinear least squares in estimation.[4] Both approaches are strong theoretically, but require demanding assumptions and difficult implementation. In particular, it would be highly desirable to find a technique that is closer in spirit conceptually and in execution to current mediation testing practice.

MacKinnon and Dwyer (1993) get the closest, when they tackle the difficult challenge of offering a solution to the scenario in which the dependent variable, Y, is categorical. Much like the rescaling of b-weights to betas (e.g., $\beta_{YX} = b_{YX}\frac{s_X}{s_Y}$), they use variance adjusters (e.g., call a' the adjusted a, then $a' = a * (s_X/s_{M'})$, where $s_{M'} = \sqrt{(a^2 \times \sigma_X^2) + \pi^2/3}$) to try to level the scales (cf., McKelvey & Zavoina, 1975, also see MacKinnon, Fritz et al., 2007; MacKinnon, Lockwood et al., 2007). While the resulting standard errors are fairly comparable, the parameter estimates would be best transformed (e.g., from inches to centimeters); that is, one parameter is still linear and the other is still in the form of a natural log of an odds ratio.

In sum, the introduction of categorical variables into mediation analyses is recognized as an important issue. And yet, to date, no single solution has been found. Table 1 shows all possible combinations of X, M, and Y being continuous or categorical. The columns at the right indicate the optimal model for fitting each Eqs. (1), (2), and (3). Note that only in scenarios (i) and (v) may regressions (or ideally, structural equation models) be used to estimate the paths. The other scenarios require a mix of OLS- and logistic-regressions. While OLS regressions and logistic regressions are conceptual analogs, they differ markedly in their particulars, and it would seem a challenge to transfer information easily between the two techniques. It would be ideal to formulate a general analytical solution applicable to any permutation of X, M, and/or Y being continuous or categorical.

Potentially applicable simultaneous models

We begin our search for a single model in which all parameters (indirect and direct) may be fit. Such models offer superior performance because all parameters are estimated simultaneously, so each effect is estimated purely, with all other effects having been partialed out (cf., Sobel, 1990). For example, structural equation models have been shown to provide better parameter estimates than regressions (Iacobucci, Saldanha, & Deng, 2007).[5] Thus, for these statistical reasons—theoretical and practical—it would be ideal to estimate all path coefficients in a single model, regardless of the variables' statuses as continuous or categorical. Yet that goal may be unattainable given the variety of data scenarios in Table 1. It seems daunting to ask that Eqs. (1)–(3) could be reformulated such that all parameter estimates be estimated simultaneously when they comprise a mix of regression and logistic regressions, with their variant requisite assumptions. Nevertheless, there are several methodologies that are potential candidates for simultaneous and parsimonious analysis of such data.

In this section, we see the proposition of three techniques, and their declination, and obtain insight into the logic as to why the ultimate solution is that which is supported later in the paper. Statisticians too often write articles claiming a technique is the best without guiding the non-methodological reader through the logic as to why that claim is made, leaving those

[4] Some researchers use bootstrapping methods which may seem appealing because they make no distributional assumptions (cf., Imai, Keele, & Tingley, 2010). However, like the neural network methods to be discussed shortly, these are approached in a more exploratory manner, and mathematical criteria position analytical solutions as superior and preferred to approximate empirical methods. In addition, a reviewer offers that current empirical Bayes methods dominate bootstrapping methods.

[5] When all three variables, X, M, and Y are continuous, researchers should conduct their mediation tests via structural equations models rather than via a series of regressions. Modeling elegance, parsimony, and efficiency are all important qualities valued by mathematical statisticians, and while it may be perceived as mere opinion that structural equations are more elegant and parsimonious than the regressions, it is inarguable that the structural equations approach is more efficient—the smaller standard errors define the statistically theoretical truism (Iacobucci et al., 2007).

Table 2
Generalized linear model specifications for OLS- and logistic-regression models.

Model: / Specification:	OLS regression	Logistic regression
Dependent variable	Continuous	Dichotomous
Error distribution	Normal	Binomial
Parametric link	Identity, $\theta=\mu$	Logit, $\theta=\ell n(\mu/(1-\mu))$
Variance	σ^2	$\mu(1-\mu)$
Badness-of-fit	$\sum(y_i-\mu_i)^2$	$2\sum\left[y_i\ell n\left(\frac{y_i}{\mu_i}\right)+(m_i-y_i)\ell n\left(\frac{m_i-y_i}{m_i-\mu_i}\right)\right]$

readers to wonder about the comparative alternatives that were rendered less useful in its stead. Thus, the strengths and weaknesses of the techniques are examined en route to the solution: generalized linear models, neural networks, and two stage least squares modeling. Readers less interested in these alternatives, as they will be dismissed, may wish to skip ahead to the section entitled, “The Solution.”

Generalized linear models

Just as “general linear models” subsume the particular model forms of analysis of variance and OLS regression, “generalized linear models” subsume these along with logit, logistic regressions, and several other families of models (Gill, 2001; McCullagh & Nelder, 1983; Nelder & Wedderbun, 1972). The subordinate models are characterized by the type of dependent variable being modeled (e.g., continuous or categorical), the distributions that the errors are assumed to follow (e.g., normal, binomial, poisson, etc.), the canonical (or basic) link between the data and the model parameters (generically called θ), the variance (or “dispersion”), and a badness-of-fit measure (called “deviance”).

Changing the specifications yields quite an array of modeling techniques. For our purposes, we focus on OLS and logistic regressions, and Table 2 lists their specifications. For OLS, the generalized linear model is manifest as modeling a continuous dependent variable and a normal distribution. For logistic regression, generalized linear models posit dichotomous data in a logit form and assume a binomial error sampling distribution. The particulars in Table 2 help to illuminate why OLS and logistic results are not easily traversed.

Direct comparability between these models is also impeded by the fact that OLS regression assumes homoscedasticity (i.e., the variance of Y is relatively constant across varying levels of X). For logits and binary Y, this relationship cannot hold—when a proportion is small (e.g., 0.2) or large (e.g., 0.8), its corresponding variance will also be small (i.e., 0.16), whereas the center of the distribution (e.g., 0.5) poses the maximum variance (i.e., 0.25), and reveals the inherent heteroscedasticity.[6]

Yet the advantage of generalized linear models, for our purposes, is the explicit recognition of the similarities and differences between the two approaches. If X and Y are continuous, but M is binary, it would be ideal to fit Eqs. (1) and (3) via OLS, with its accompanying normal error distribution, identity link, and sums of squared error terms, and Eq. (2) via a logistic regression, with its binomial error distribution, logit parametric formulation, and a product of probabilities as a variance estimator.

Unfortunately, when one adopts the generalized linear modeling framework, and begins to fit data specifying a particular error distribution, parametric link, etc., those specifications are applied to all relationships estimated in the model. This issue is not simply a problem with extant software packages—in fact, the comprehensiveness of the assumptions is inherent to the modeling framework. For our purposes, this issue means that we cannot fit the equivalent of a structural equation model wherein one path is a logit link and another is OLS regression.

Neural networks

Neural networks are structural models with input nodes, i.e., explanatory variables (such as X), output nodes (response variables such as Y), and in between, one or more layers of so-called hidden nodes which specify the functional form that relates the independent and dependent variables (Briesch & Rajagopal, 2010; Fausett, 1993). If one hidden node was linear in form and another nonlinear, our simultaneous fitting challenge may be solved.

One of the reasons that neural networks have become popular among marketers (albeit more so in industry than academia) is because they are massively powerful data-mining tools. Prior to model fitting, very little is specified of the model's structure, including the functional forms of the hidden layers. They are generally used as nonparametric methods, with no restrictions regarding distributional assumptions. It is part of the data-mining exploration itself to identify such qualities as the functions that characterize the hidden nodes.

For our analytical needs, we do not seek a technique to discover whether a relationship is linear or not; rather, we know the nature of ideal relationships and we need a technique that allows us to control over that specification. Neural nets are fundamentally exploratory whereas in mediation analyses, the researcher is quite purposeful, testing linkages that are theoretically posited between the constructs or variables. In addition, given the exploratory nature of neural networks, that technique requires very large data sets, first to train the model (that is, to identify its relational structures), and then subsequently to assess the fit on the remaining data.

Thus, neural networks may not be optimally suitable.[7] Yet the notion of the contingent nature of the relationships among variables and the functions that tie them together is appealing, so we will carry forward this logic.

[6] A reviewers recommended clarification regarding heteroscedasticity in this context: namely a note that the errors and issues relevant to the models in this paper are not the same as those in the “utility space” or those derived from utility-based theories which require and assume homoscedastic and fixed, untestable errors, per Guadagni and Little (1983).

[7] A reviewer pointed me to an intriguing development; early research combining neural networks with a structural equations model logic. As yet however, while the software neusrel is available, the underpinnings of the model are not, nor, of course, have they been vetted by the academic blind review process.

Two-stage least squares

Two-stage least squares (2SLS) is a technique that econometricians use more frequently than their behavioral counterparts (Kiviet & Phillips, 2000). The setting is this: a relationship is being estimated, $M \rightarrow Y$, yet there is some uncertainty as to whether the relationship has been isolated (with other effects having been partialed out). There is a concern that instead, perhaps it could be the case that some other common factor or error contributes to both M and Y, which thereby creates a spurious relationship, or magnifies a weak one.

To try to tease out the common factor or error, an "instrumental variable," X, is inserted into the model $X \rightarrow M$. The instrument X is deemed useful in 2SLS if X and M are correlated, but X is not supposed to be correlated with Y, otherwise X would be essentially yet another common factor between M and Y, which simply exacerbates the problem rather than solving it. If such an X is obtained, then the overall model is fit via two-stages, first, $X \rightarrow M$ and the resulting $\hat{M}$s are used to predict Y; i.e., $\hat{M} \rightarrow Y$.

There are at least two limitations of 2SLS which renders the technique inapplicable for our purposes. First, note that both stages of estimation are regressions; hence, the flexibility to model logits or logistic regressions is unavailable. Second, most mediation analyses cannot guarantee that there is no relationship between X and Y; indeed there usually is some direct association and the research questions include a comparison of the effect size of the direct and indirect paths.

As with the other techniques, however, we can bring forward the strengths of this method. In particular, it will be very useful to estimate the models recursively; i.e., in two stages.

In sum, none of the three classes of analytics just described is quite adequate for the goal of being able to fit all the component paths simultaneously. This conclusion is hardly surprising, or researchers would already be using the alternative methodology. Nevertheless, each of the modeling techniques had desirable qualities, e.g., we should aim to fit parts of mediation analyses via regressions and other parts via logit models, as exactly as possible, with minimal compromise on optimal assumptions regarding error distributions and function forms. Further, perhaps modeling in two stages of recursive iterations can create a happy medium between the old-fashioned piece-meal series of three regressions and the simultaneous fitting of the structural equation modeling approach.

Potential solution in a macro view using R^2

If a technique like structural equation models cannot be found to fit all paths simultaneously when a mix of OLS and logistic regressions are required, then perhaps the next best thing would be a solution configured at a relatively macro view of the models, e.g., in comparing overall fit statistics, than in the particulars of a micro view, e.g., parameter estimates that we know will vary more dramatically.

Indeed, some modelers recommend comparing R^2s rather than β weights, given the likely multicollinearity in the data (Lehmann, 2001). Mediation research seems to concur. Several studies comparing indices of effect sizes, show, for example, that proportions and ratios of indirect and direct effects relative to totals behave statistically problematically, and these researchers recommend the use of measures of explained variance, namely R^2s (Fairchild, MacKinnon, Taborga, & Taylor, 2009; MacKinnon et al., 2002; Preacher & Kelley, 2011). R^2s have a long, strong history being used as a legitimate measure of effect size, and it may be useful here—we know it could obviate the multicollinearity problem and we need to examine its utility across classes of variables (continuous or categorical) and requisite models (OLS and logistic regression).

Thus, in this section, we examine the potential of R^2, and we demonstrate the relationships in the form of equations as well as in an example. The illustration is intended to make the results more palatable for readers not interested in proof-like analytics. Thus, consider the small data set in Table 3, which comprises three continuous variables and three binary variables. We will proceed by fitting each of Eqs. (1)–(3), to note the equalities that arise in simple regressions (in Eqs. (1) and (2)) and in multiple regressions (in model (3)).

Table 3
Example data set.

X	M	Y	Categorical X	Categorical M	Categorical Y
1	2	3	0	0	0
5	6	7	0	0	0
10	11	12	0	1	1
13	16	19	1	1	1
5	7	10	0	0	1
11	6	6	1	0	0
4	6	8	0	0	0
16	13	16	1	1	1
12	8	10	1	0	1
13	6	2	1	0	0
19	14	13	1	1	1
10	12	5	0	1	0
5	2	6	0	0	0
1	7	3	0	0	0
8	9	11	0	0	1
18	12	16	1	1	1
14	8	11	1	0	1
8	8	1	0	0	0
5	1	9	0	0	0
5	7	1	0	0	0
19	11	3	1	1	0
18	18	15	1	1	1
9	21	23	0	1	1
12	20	6	1	1	0

Equalities in a simple regression

Let us begin with the simplest case of X, M, and Y all being continuous variables (i.e., the first three columns in Table 3). Eq. (1) tests the direct relationship between X and Y. In preparation, note that the simple correlation between X and Y, r_{xy}=0.41449, thus we know that the variance in Y accounted for by its linear relationship with X is: r_{xy}^2=0.1718. Fitting Eq. (1) yields the regression weights:

$\hat{Y} = 4.55906 + 0.44225X$ in terms of raw regression weights, b's, or

$z_{\hat{Y}} = 0.41449 z_X$ scaled as standardized regression weights, β's, with the accompanying $R^2_{xy}=0.1718$, which equals r^2_{xy} because there is only a single predictor variable. For the same reason (one predictor),[8] the simple correlation noted previously, $r_{xy}=0.41449$, is also equal to the standardized regression coefficient, $\beta_{xy}=0.41449$.

In addition, recall the following relationships between the data, X and Y, and the predicted values and their error terms:

$Y=b_0+cX+\varepsilon$ is the equation for the data written in terms of the model and error,
$\hat{Y}=b_0+cX$ is the notation for the model, and
$Y=\hat{Y}+\varepsilon$ is the data re-written using the notation: model plus error.

The correlation between Y and $\hat{Y}$, $r_{y\hat{y}}$, is equal to the correlation between X and Y, r_{xy}. That is:

$$r_{y\hat{y}}=r_{xy}=\beta_{xy}=0.41449.$$

Naturally, the obverse relationship also follows—the residuals $\left(Y_i-\hat{Y}_i\right)$ are by definition completely uncorrelated with X, $r_{x\varepsilon}\equiv 0$. By construction of a regression, every ounce of shared variability between X and Y has been incorporated into $\hat{Y}$. Thus,

$$\begin{aligned} Y-\hat{Y} &= (b_0+cX+\varepsilon)-(b_0+cX) \\ &= b_0+cX+\varepsilon-b_0-cX \\ &= b_0-b_0+cX-cX+\varepsilon=\varepsilon. \end{aligned}$$

These equalities and relationships hold because the $X\rightarrow Y$ model is a simple regression; i.e., it has a single predictor. Therefore, the analogous equalities and relationships also hold when fitting Eq. (2), which tests the relationship from the independent variable X to the mediator M (one predictor also). The results for fitting $X\rightarrow M$ follow: $\beta=0.59493$, $b=0.57974$, $R^2=0.3539$, with $R^2=r^2_{xm}$ and $r_{xm}=\beta_{xm}=r_{m\hat{m}}$.

These equations make it obvious that it is very helpful in OLS regression to be able to work interchangeably between the macro level (R^2 model fit) and the micro level estimates (b or β). If the same kinds of relationships held in logistic regressions, then it wouldn't matter that the parameter estimates in logits (i.e., the equivalents to b or β), were in the natural log scale; instead we could simply work with the logit equivalent of R^2. We will consider logit equivalents to R^2 shortly.

Equalities in a multiple regression

While the equalities that hold for simple regressions offer flexibility with Eqs. (1) and (2), we are nevertheless left with the conundrum that Eq. (3) is a multiple regression. Eq. (3) has two predictors—the independent variable (X) and the mediator (M), $X\&M\rightarrow Y$. In fitting this multiple regression, we obtain:

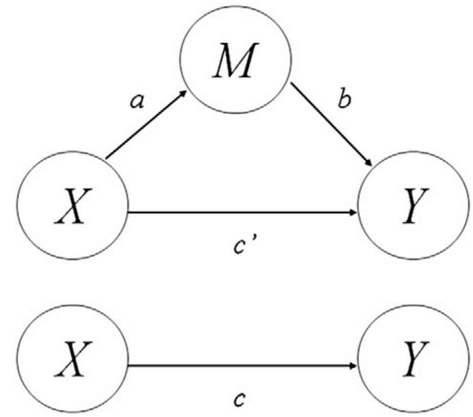

Fig. 1. Standard Trivariate Mediation: X = Independent Variable, M = Mediator, Y = Dependent Variable.

$\hat{Y}=2.23047+0.08731\ X+0.61223\ M$ in terms of raw regression weights, b's, or
$z_{\hat{Y}}=0.08183z_X+0.55915z_M$ scaled as standardized weights, β's.

While there are two b's or two β's, there is of course only one $R^2=0.3738$. Hence, the equality that held in simple regression between r^2 and R^2, or r and β cannot hold true for a single measure R^2 with two input β's. Yet 2SLS has already given us a hint as to how this 2-predictor model may be mapped onto a 1-predictor variant.

Consider that in fitting Eq. (3), we seek primarily the estimate of the path coefficient "b" in Fig. 1. That is, we wish to assess the effect of M on Y, having statistically controlled for the effect of X. The shared variance between M and Y is depicted in Fig. 2 as the union of areas A3 and A4, or A3∪A4. We adjust for the relationship between X and Y, or A2∪A3, as well as that between X and M, or A3∪A6. Logically: A3∩A4 – [(A2∪A3)∩(A3∪A6)]=A3∪A4–A3=A4. We can get a statistical estimate of that area.[9]

If we proceed as in 2SLS, we first fit $X\rightarrow M$. In the second stage, 2SLS takes the predicted values forward, using $\hat{M}$ (represented by the variance common to X and M in areas A3 and A6) to predict Y. Of A3 and A6, the variance shared with Y is A3 and that is not the area we seek. Instead, in our second

[8] Similarly, the t-statistic testing the significance of the slope parameter associated with X (that is, b_{xy} or β_{xy}), with "error" degrees of freedom (N-2, for the estimation of the slope and intercept) can be squared to obtain the F-statistic for the overall model (testing R) on "1" and "error" degrees of freedom.

[9] Recall from regression: a partial correlation coefficient is: $r_{YM\cdot X}=(r_{YM}-r_{YX}r_{XM})\Big/\left(\sqrt{1-r^2_{YX}}\sqrt{1-r^2_{XM}}\right)$. Its square is $r^2_{YM\cdot X}=(R^2_{Y\cdot XM}-r^2_{YX})/(1-r^2_{YX})$, which expresses the logical equivalent of areas $A4/(A4+A1)$. The area we seek to isolate is A4. The lesser known, or less frequently used, semi-partial correlation coefficient is: $sp_M=(r_{YM}-r_{YX}r_{XM})\Big/\left(\sqrt{1-r^2_{XM}}\right)$ and its square, $sp^2_M=R^2_{Y\cdot XM}-r^2_{YX}$, represents area A4. (For comparison, the more familiar βs from multiple regression are partial regression coefficients, cf., $\beta_{YM\cdot X}=(r_{YM}-r_{YX}r_{XM})/(1-r^2_{XM})$.) (Fairchild et al. (2009) conducted an analogous decomposition, focusing on area A3, the heart of the overlap among all three variables, the area they called R^2_{Med}. As they show, an estimate of area A3 could be obtained in several ways, e.g., begin with r^2_{MY}, which captures area (A3+A4) and subtract off area A4, as per the above semi-partial, altogether yielding their $R^2_{Med}=r^2_{MY}-(R^2_{Y\cdot XM}-r^2_{YX})$. We had been focusing on area A4, seeking a pure estimate of the path coefficient from M to Y (labeled A2, in Fig. 1).) What our three coefficients have in common is their numerator—we begin with the relationship between M and Y, and we adjust for the relationships that X has with both M and Y.

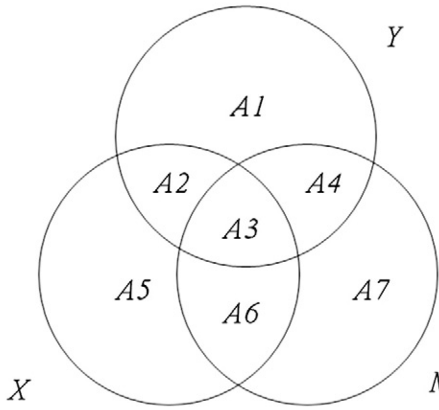

Fig. 2. Venn Diagram for Multiple Regression Depicting Variances and Covariances Among X, M, and Y (Notation is such that A = Area; e.g., A1 = Area 1).

stage, if we take the residuals forward, using $\left(M-\hat{M}\right)$ (represented by areas A4 and A7) to predict Y, then the overlap of areas A4 and A7 with Y is A4, the area we sought to estimate. Thus, we proceed as follows: in stage 1 we fit $X \rightarrow M$, we extract the predicted values $\hat{M}$, compute the residuals $\left(M-\hat{M}\right)$, and use them to predict Y, per $\left(M-\hat{M}\right) \rightarrow Y$, then the single resulting regression coefficient b reflects the relationship between M and Y with the effects of X partialed out.[10] For the data in Table 3, $\hat{Y}=9.0+0.61223 \left(M-\hat{M}\right)$, thus capturing precisely in a single coefficient the relationship we sought to isolate. The standard error is incorrect (as is known in the 2SLS literature, cf., Kiviet & Phillips, 2000) and as a result, the standardized parameter estimate β is also incorrect, but once the standard error is adjusted, β is also easily corrected: $\beta = b(s_{mresids}/s_y)$. (It is also recognized that 2SLS increase Type I error rates by bringing forward and exacerbating error, so we should be mindful and conservative in choices for α.) These relationships could hold the promise of making OLS and logistic regression results comparable if there is an equivalent to a macro measure of model fit, R^2, in logits and logistic regressions.

Is there a logit equivalent of R^2?

The R^2 in OLS regression is exceedingly useful. It is a single summary fit index, and its interpretation as the proportion of variance in the response variable by the predictors is highly intuitive. A corresponding statistic would be desirable in logistic regressions. Unfortunately, the situation is more complicated, because there are multiple criteria for "fit" in categorical data analysis.

For simplicity, we stay with Eq. (1), but change the scenario to one that works with the binary variables in Table 3. If we begin simply with Y (coded 1 and 0), and X (also coded 1 and 0, though the logic also works for continuous X) and if we were to fit the OLS regression model to predict Y, the fact that there are no restrictions on the b's implies that the $\hat{y}$'s could range from $-\infty$ to $+\infty$, when in fact the actual Y's are 0's and 1's. The b's will be unbiased, but they will not necessarily have the smallest standard errors (i.e., they are statistically inefficient; note, however, that this quality implies that OLS is at least a conservative approach). Thus instead of modeling Y, we model the odds of Y. Take p to be short for $p(Y=1)$, the probability that Y takes on the value 1 rather than 0. The odds are then stated: $p/(1-p)$, which represents the probability that $Y=1$ for a given X value, versus (divided by) the probability that $Y=0$ for that X value. This quantity is now free to range at the high end to $+\infty$, but it is still bounded at the lower end at 0. If we tweak once more and model the natural log of the odds, $\ell n[p/(1-p)]$, or the "logit," then the predicted logits have no restrictions and may range from $-\infty$ to $+\infty$. This logic is the standard introduction to logit models, to enhance the comparability between OLS and logistic regression models.

At this point, the right-hand side of the logit model resembles the OLS regression model: $\ell n[p/(1-p)] = b_0 + b_1 x + \varepsilon$. The b's are interpretable analogously to (but not identically to) regression coefficients in OLS; in logit models, a one unit change in X implies a change of b_1 in the dependent variable which is a logit, or the natural log of the odds that Y will be $=1$ (vs. 0) for a given X value.[11] To illustrate, for the categorical X and Y in Table 3, the logit estimates were: $\ell n[p/(1-p)] = -0.8109 + 1.3705\, x$. For the $X=1$ cases in the data, the log odds $= -0.8109 + 1.3705(1) = 0.5596$. For the $X=0$ cases, the logit $= -0.8109 + 1.3705\,(0) = -0.8109$. In English, for the $X=1$ cases, the log of the odds that $Y=1$ is greater than the log of the odds that $Y=1$ for the $X=0$ cases.

If the log scale seems uninterpretable, the pieces may be transformed. If $\ell n[p/(1-p)] = b_0 + b_1 x$, then taking e to the power of both sides tells us $[p/(1-p)] = e^{b_0+b_1 x}$. For the example of $\ell n[p/(1-p)] = -0.8109 + 1.3705\, x$, the transformation takes us to $[p/(1-p)] = e^{-0.8109+1.3705\, x}$. For $X=1$, we get $e^{-0.8109+1.3705} = e^{0.5596} = 1.750$ and for $X=0$, $e^{-0.8109} = 0.4445$. Translated, the odds that $Y=1$ are higher for the $X=1$ cases than for the $X=0$ cases.

If odds are still a problematic scale, we can solve for $p = [(e^{b_0+b_1 x})/(1+e^{b_0+b_1 x})]$. To continue the example, for $X=1$, we obtain $p = [e^{(-0.8109+1.3705(1))}/(1+e^{(-0.8109+1.3705(1))})] = [e^{0.5596}/(1+e^{0.5596})] = (1.750/2.750) = 0.6364$. When $X=1$, the probability that $Y=1$ is 63.64%. For $X=0$, the probability that $Y=1$ is $p = [e^{(-0.8109)}/(1+e^{(-0.8109)})] = (0.4445/1.4445) = 0.3077$, or 30.77% likely. In sum, being in the $X=1$ set enhances the probability of Y occurring.

Regardless of whichever transformation is preferred, the coefficients of the model yield predicted values, $\hat{Y}$'s. Those

[10] Alternatively stated: fitting $X \rightarrow M$ yields $\hat{m} = \beta_{mx} x$, using standardized coefficients without loss of generality. Let the residuals $\lambda = (m - \hat{m})$, and use $\lambda \rightarrow Y$. That model is: $y = \beta(\lambda) = \beta(m-\hat{m}) = \beta_1 m - \beta_2 \hat{m}$, per difference scores. Call $\beta_1 = \beta_{ym}$ and $\beta_2 = \beta_{ym} = \beta_{yx}$. Then $y = \beta_{ym} m - \beta_{yx} \hat{m} = \beta_{ym} m - \beta_{yx} \beta_{mx} x$, the essence of the numerator of $\beta_{ym \cdot x}$.

[11] Alternatively, one standard deviation change in X implies a change of standardized β_1 in the logit. Conversely the model predicting the $Y=0$ case would be written: $\ell n[(1-p)/p] = b_0 + \varepsilon$.

predicted values may be compared to the actual values of Y, per $\sum (y-\hat{y})^2$, which in OLS contributes to the overall model fit, $R^2=1-\left[\sum\left(Y_i-\hat{Y}_i\right)^2/\sum(Y_i-\bar{Y})^2\right]$. In logistic regressions, there are several candidates for an analog to R^2.

First, we could compute the standard R^2 as from OLS. There are (at least) two reasons why this might be acceptable, and (at least) one reason why it might not be. On the plus side, the index is remarkably robust, consistently demonstrating good empirical performance, even in application to binary data (DeMaris, 2002; Efron, 1978; McKelvey & Zavoina, 1975; Mittlböck & Schemper, 1996). In addition, as we have alluded previously, statisticians do not dispute the usefulness of R^2 as an indicator of effect size. Unfortunately (for our needs), since Efron (1978), researchers have also made clear that while a value comparing Y and $\hat{Y}$ would be mathematically equivalent to the OLS R^2, it is conceptually different. The objective function in logistic regression does not minimize the sum of squared errors, and accordingly, R^2_{OLS} cannot be interpreted as variance explained, or error variance reduced, in describing a logistic model.

As an alternative for R^2s in logistic regressions, Cox and Snell (1989) proposed an R^2-like index: $R^2_{CoxSnell}=1-(L_0/L_M)^{(2/n)}$, where L_0 is the likelihood statistic for the null model (the model with only an intercept) and L_M is the likelihood statistic for the focal model (with one or more predictors). This index cannot achieve a maximum value of 1.0, so the adjustment offered by Nagelkerke (1991) such that the maximum R^2 of 1.0 may be obtained is usually preferred,[12] in a third version of an R^2: $R^2_{Nagel}=[R^2_{CS}/(1-L_0^{(2/n)})]$.

A fourth alternative was recommended by Menard (2000), based on its conceptual similarity to R^2 in OLS as well as its empirical performance in a simulation comparing it to R^2_{OLS}, R^2_{CS}, and R^2_N. This index is based on the likelihood ratio test statistic comparing the performance of the null and focal models, hence is eminently embraceable from a theoretical statistical vantage (and also supported by McFadden, 1974; Mittlböck & Schemper, 1999).[13] It is defined: $R^2_{LRT}=1-[\ell n(L_M)/\ell n(L_0)]$. Menard explains that R^2_{LRT} is a good analog to R^2_{OLS} because conceptually, the R^2_{OLS} is interpretable as the proportional reduction in the error function, or as the statistic that maximize the likelihood function. In logistic regression models, R^2_{LRT} similarly reflects the proportional reduction in the error function. Granted, the error function is different from the OLS error function, but it is the statistic that maximizes the likelihood function. If a superior solution cannot be found, we might wish to return to these omnibus measures, but for now, we simply note the complications that the various R^2s are comparable conceptually or mathematically, but not both for the same statistic.

[12] These indices, R^2_{CS} and R^2_N are produced in SAS's Proc Logistic as "R-Square" and "Max-rescaled R-Square," respectively.

[13] While SAS does not compute R^2_{LRT} currently, it does print L_M and L_O in its default output. Both appear under the header, "Model Fit Statistics." L_O is the value labeled "-2 Log L" for the "Intercept Only" model, and L_M is labeled "-2 Log L" for the "Intercept and Covariates" model.

Potential solution in a micro view using βs

Eqs. (1)–(3) allow us to answer the mediation questions: "Is there a significant direct path in the data, $X\rightarrow Y$," and "Is there a significant mediated (indirect) path in the data, $X\rightarrow M\rightarrow Y$?" To estimate and test a mediated (indirect) path, we must combine estimates of a in Eq. (2) and b in Eq. (3), as per Eq. (4) or in some comparable manner. If X, M, and Y are all continuous, we know how to proceed, e.g., via a structural equation model. However, consider the challenge of getting the OLS and logistic regression models to talk to each other. If estimate a is obtained from a logit model and estimate b is derived from a regression, the question is how can they be combined in a valid and sensible manner?

To be more specific, consider one of the mixed cases, say scenario (iii) from Table 1, in which M is a categorical variable, and X and Y are continuous. Proceeding with data of this nature, we could answer the research question about the presence of a direct path via Eq. (1), and do so in a simple regression setting, $\hat{Y}=b_{01}+cX$. To address the research question about a mediated path, we must fit Eqs. (2) and (3), or something like them, and combine the estimates in some manner. Eq. (3) can be fit via multiple regression, given Y's status as a continuous variable, and we will extract b and s_b from its results, $\hat{Y}=b_{03}+c'X+bM$. Given M's status as categorical, we know it to be optimal to fit Eq. (2) via logistic regression, $\hat{M}=b_{02}+aX$. That model will produce an estimate of a and its standard error s_a. At this point, we would have all the components, but it should be obvious that we cannot simply mash them into the z-test in Eq. (4) given that the estimates arise from different statistical machines. So let's see what can be done, and we can begin by understanding what we've got.

Researchers working on substantive theoretical questions are no strangers to theory, but many seem surprised to find that methodological research is guided by mathematical statistical theoretical axioms as well. For example, it is known, or defined to be the case, that under the null hypothesis, the square of a standard normal, i.e., $z\sim N(0,1)$, is a central chi-square on one degree of freedom, $X^2\sim\chi^2_1$. When we examine statistical tests in OLS regression to determine the significance of the predictors, each test statistic is a $t=\hat{b}/\hat{s_b}$, and to simplify our world, let us assume sufficient sample size that degrees of freedom equal or exceed 30, such that the t-statistic may be referred to more simply as a z-test, with no subsequent concern for degrees of freedom (even small scale consumer psychology studies tend to have sample sizes at least this large).

In logistic regression models, the significance of each predictor variable is evaluated via Wald's $X^2=\left(\hat{b}/\hat{s_b}\right)^2$. The square root of the X^2 is a t on the same degrees of freedom as the X^2 had been, which, if sufficient, exceeding 30, we can refer to the t-test as a z-test.

Together, then, in the OLS regression corner, we would have from Eq. (3) $z_b=\hat{b}/\hat{s_b}$, and in the logistic regression corner, we would have from Eq. (2) $z_a=\hat{a}/\hat{s_a}$. The coefficients, $\hat{b}$ from the OLS regression and $\hat{a}$ from the logistic regression are still on different scales, no more comparable than apples and oranges. However, their z's are of the same class and may be compared head to head. In addition, while we cannot insert the a, b, s_a, and s_b estimates directly into a z-test like Eq. (4), we can use their standardized forms, as we see next.

The numerator of Eq. (4), that which tests the significance of the mediated (indirect) effect of X on Y is simply the parameter a (from Eq. (2)) times b (from Eq. (3)). We replace $a \times b$ with the product of their standardized form, $z_a \times z_b$. Two questions now arise: first, what is the standard error of this product (that we can use in the denominator), and second, what is the theoretical statistical distribution of this product, so we can have a comparison basis against which to judge probability and significance. These answers are well-known for the sum of two z's, but are not as familiar for the product of two normally distributed random variables, so they are worth elucidating here.

For A defined as the product of two independent normally distributed random variables, $A = \frac{B}{s_B} \times \frac{C}{s_C} = \frac{BC}{s_B s_C}$, its expected value (population mean) is $E(A) = \mu_B \mu_C$ with variance $V(A) = \mu_B^2 \sigma_C^2 + \mu_C^2 \sigma_B^2 + \sigma_B^2 \sigma_C^2$ (Craig, 1936; Glen, Leemis, & Drew, 2004; Hogg & Craig, 1995; Lomnicki, 1967). For standardized variables, the variance term simplifies to $V(A) = \mu_B^2 + \mu_C^2 + 1$. For mediation, we'll call the product estimate $z_{a \times b} = z_a z_b = \frac{a}{s_a} \times \frac{b}{s_b}$ with variance $(z_{a \times b}) = z_a^2 + z_b^2 + 1$. The z-test in Eq. (4) is adjusted, becoming:

$$z_{Mediation} = \frac{z_a z_b}{\hat{\sigma}_{z_{ab}}} = \frac{\frac{a}{s_a} \times \frac{b}{s_b}}{\sqrt{z_a^2 + z_b^2 + 1}}. \tag{5}$$

Note that with the exception of the 1.0 adjuster in the denominator, this z-test and the z-test in Eq. (4) (the so-called Sobel) are identical test statistics under the condition of standardized coefficients.

With a statistical test in hand, we now simply require a distribution to determine the significance of $z_{Mediation}$. Aroian (1947) showed that the product of two normals is asymptotically normal if either component (here, z_a or z_b or both) are large.[14] Presumably researchers testing for mediation expect it, thus at least one of these z's, indeed probably both, should be large. From another vantage, we can invoke a multivariate null hypothesis, $z_a = z_b$ such that the numerator of Eq. (5) is a z^2 random variable, and the denominator is the square root of the sum of two z^2's (plus one, which fades in contribution relative to large z_a or z_b) statistically equated with a chi-square with two degrees of freedom. Together, the distributional forms of the ratio would be approximately $z^2 \Big/ \sqrt{X^2} = z^2/z = z$, again converging on the likely normal shape of this distribution. Thus, once $z_{Mediation}$ is calculated per Eq. (5), it may be evaluated as a z-statistic, standard normal, i.e., the mediation effect is significant at $\alpha = 0.05$ if $z_{Mediation}$ exceeds |1.96| (for 2-tailed tests and $\alpha = 0.05$).

[14] The distribution is a modified Bessel, and is ill-defined at zero (which should not be a frequent problem in mediation studies because presumably most researchers estimating mediation effects are anticipating that they are non-zero). Otherwise, the product normal resembles a slightly leptokurtic t- or z-distribution.

What is important about $z_{Mediation}$ is that in using this statistic, OLS and logistic regression results may be combined to test mediation, no matter whether X, M, and Y are continuous or categorical. This estimator carries the added bonus that when tested in the context of OLS on continuous variables, it fared among the best with regard to performance on criteria like unbiasedness, sufficient power, etc., in part because it provides a better approximation to the normal distribution, particularly for large samples (MacKinnon, Fritz et al., 2007; MacKinnon, Lockwood et al., 2007).

While the analytical results reviewed in this paper must hold theoretically, it is frequently of interest to view an empirical demonstration of a test statistic's performance. Accordingly, Table 4 presents the results of a simulation study. This study comprised a 4×5 factorial, with 1000 replications per cell. Sample size varied, from N = 30 (minimally acceptable) to 50, 100, and 300. The extent to which the population covariance matrix contains direct and indirect effects all varied, from full to no mediation.

Specifically, the extent of mediation ranged from 100% indirect effect to 0% mediation, with intermediate levels of 25%, 50%, and 75% mediation (or 75%, 50%, 25% direct effect). The strengths of mediation can manifest from a variety of population parameters, and those used in this simulation follow. For 100% $\rho_{XM} = 0.707$, $\rho_{MY} = 0.707$, and $\rho_{XY} = 0.000$, which yield estimates $a = 0.707$, $b = 1.414$, and therefore $ab = 1.000$. Population rhos were used to generate pseudo-random multivariate normal deviates for the 1000 samples. All μ's were 0.0 and categorical variables were created at that cut-point. For 75% mediation, population correlations of $\rho_{XM} = 0.612$, $\rho_{MY} = 0.612$, and $\rho_{XY} = 0.125$ in essence provide $0.612 \times 0.612 = 0.3745$, a 3:1 ratio to 0.125, hence akin to 75%. These correlations yield path coefficients of $a = 0.612$, $b = 0.856$, for a product of $ab = 0.524$. For 50% mediation, the population rhos were $\rho_{XM} = 0.500$, $\rho_{MY} = 0.500$, and $\rho_{XY} = 0.250$, with $0.5 \times 0.5 = 0.25$ for 50:50 shared variance, yielding coefficients: $a = 0.500$, $b = 0.500$, $ab = 0.250$. For 25% mediation, population parameters were $\rho_{XM} = 0.354$, $\rho_{MY} = 0.354$, and $\rho_{XY} = 0.375$, which gives $0.354 \times 0.354 = 0.1253$, 1:3 compared to 0.375, and these yield $a = 0.354$, $b = 0.253$, and $ab = 0.089$. Finally, for 0% mediation, population parameters were, $\rho_{XM} = 0.000$, $\rho_{MY} = 0.000$, and $\rho_{XY} = 0.500$, which yield $a = 0.000$, $b = 0.000$, and $ab = 0.000$ compared to $c = 0.500$.

The large blocks of rows in the table depict the analytical treatments of the variables. For example, in case (i), all 3 variables were continuous, and regressions were used to fit Eqs. (1)–(3). For case (ii), Y was categorical, logistic regressions were used for Eqs. (1) and (3) and a regression used for Eq. (2).

With cell sizes of 1000, all effects were significant, thus cell means are presented for the interaction between sample size and extent of mediation. The main effects are also easily discerned and (almost all results are) as expected; namely, the average $z_{Mediation}$ rises with greater presence of mediation (e.g., 100% vs. 0%), and with the power accrued to larger samples. The test statistic, $z_{Mediation}$ is supposed to capture mediation effects, hence it should exceed |1.96| (for 2-tailed tests with $\alpha = 0.05$) for the columns labeled 100% and 75% mediation. It should not

Table 4
$z_{Mediation}$: Cell entries are means over N = 1000 samples*.

Case	*X*	*M*	*Y*	% Mediation: sample size	100%	75%	50%	25%	0%
i)	Continuous	Continuous	Continuous	30	5.495	2.973	1.740	0.817	0.018
				50	7.080	4.003	2.450	1.295	0.022
				100	9.966	5.789	3.598	1.952	0.000
				300	17.275	10.236	6.447	3.598	0.000
ii)	Continuous	Continuous	Categorical	30	3.852	1.947	1.292	0.613	0.022
				50	5.017	2.642	1.838	1.022	0.016
				100	7.023	3.859	2.738	1.518	0.000
				300	12.247	6.833	4.894	2.882	0.000
iii)	Continuous	Categorical	Continuous	30	2.117	1.645	1.072	0.505	0.021
				50	2.848	2.240	1.568	0.853	0.031
				100	4.135	3.307	2.381	1.377	0.000
				300	7.273	5.896	4.339	2.665	0.000
iv)	Continuous	Categorical	Categorical	30	1.612	1.327	0.865	0.379	0.024
				50	2.285	1.881	1.303	0.712	0.025
				100	3.362	2.798	2.011	1.115	0.002
				300	5.952	5.018	3.698	2.245	0.000
v)	Categorical	Continuous	Continuous	30	3.324	2.316	1.540	0.784	0.028
				50	4.374	3.134	2.157	1.237	0.036
				100	6.223	4.524	3.188	1.887	0.002
				300	10.855	8.057	5.737	3.525	0.001
vi)	Categorical	Continuous	Categorical	30	1.917	1.673	1.177	0.609	0.023
				50	2.657	2.297	1.685	0.995	0.032
				100	3.893	3.368	2.551	1.529	0.000
				300	6.907	6.008	4.601	2.942	0.003
vii)	Categorical	Categorical	Continuous	30	1.928	1.490	1.010	0.507	0.017
				50	2.744	2.118	1.490	0.827	0.035
				100	4.025	3.131	2.294	1.360	0.000
				300	7.096	5.652	4.194	2.676	0.000
viii)	Categorical	Categorical	Categorical	30	0.069 → 1.565	0.973	0.784	0.390	0.019
				50	0.010 → 2.262	1.664	1.245	0.694	0.019
				100	0.085 → 3.274	2.639	1.972	1.136	0.012
				300	0.153 → 5.771	4.822	3.646	2.319	0.000

**Note*: The $z_{Mediation}$ means that exceed 1.96 are underlined.

indicate the presence of mediation in the right column for 0% mediation. These results are properly reflected in the table, with variations attributable to powerful large samples.

The interaction is seen in the cascading of significance, with stronger results to the left (for larger proportions of mediation) and falling off in the table toward the right, in steps associated with diminishing sample size. For example, depending on one's research context, 25% mediation may or may not seem overly impressive, but whether it is important or not is distinct from whether it is significant or not, and the latter depends in part on sample size. Hence, with the 25% column, the *z*-test exceeds 1.96 when N = 300. Conversely, within the 100% column, the *z*-test almost always exceeds 1.96, failing when samples are smaller, particularly for the cases when *Y* is categorical. Such patterns are also not unexpected, consistent with the traditional warnings of lost information and reduced measures of association for categorical rather than continuous variables. Of note, the strongest $z_{Mediation}$ appears for cases (i) and (v), when both *M* and *Y* are continuous, which makes sense, given that these are the two variables that serve as dependent variables throughout the system of Eqs. (1)–(3). The worst performing z's, meaning they are small even in the presence of full mediation in the population (the 100% column) are those for cases (ii) and (viii).

The results in Table 4 indicate (almost perfectly) that $z_{Mediation}$ will properly identify a mediation path as significant, the stronger the mediation effect (i.e., 100%, 75%, and perhaps 50%) and the larger the sample (N = 300, 100, and perhaps 50). These patterns also largely hold no matter whether the analytics were based on continuous variables or categorical, or OLS or logistic regressions. The flexibility of this powerful and largely well-behaved statistic is an important advance for mediation researchers.

One exception merits further comment. The performance of $z_{Mediation}$ is peculiar for the case of full mediation on wholly categorical data (viii). In scenario "viii", in the left column labeled "100%", the four means on the left (0.069 to 0.153) are woeful underestimates. In terms of statistical estimation and computation, the result is not difficult to explain; when full mediation is expected, the logit models are estimating relationships with dampened strengths due to the categorical nature of the variables, solving for estimates with probabilities near 0 and 1 boundaries. (Indeed, the computer will generate warnings about separation and quasi-separation of the data.) The problem with estimates at extreme bounds is further exacerbated by the estimates being squared, thus the near zero estimates become even smaller, near statistical computing mimima (which usually, and in this case, has the effect of blowing up standard errors as well). In this scenario, the condition of 100% mediation guarantees data sets in which the $X \rightarrow M$ and $M \rightarrow Y$ paths will be strong. The $X \rightarrow M$ relationship further necessarily results in extreme multicollinearity when fitting the $X \& M \rightarrow Y$ model (Eq. (3)). For a different perspective or insight, imagine that these categorical data are in the form of a 2×2 matrix or table in preparation for the computation of a phi coefficient (a Pearson product–moment correlation on two binary variables). Further, imagine all of the data observations are in one cell—this scenario is that which is occurring in the 100% mediation coupled with fully categorical variables conditions; e.g., there are $X \rightarrow M$ and $M \rightarrow Y$ linkages or correlations but no data linking $X \rightarrow Y$. Under such circumstances, while the parameter estimate b can appear reasonable, the standard errors for b are very high, quite out of range of most of the other cells in the table (e.g., on the order of 100 or 200 compared with other s_b's around 0.50 to 3.50), which is a typical, early diagnostic of a statistical computing problem. Then in turn, the large standard error necessarily reduces z_b to a very small number, such that the product in the numerator of $z_{Mediation}$ is reduced to those around the means observed in Table 4. When standard errors are adjusted, as per the four means to the right in that cell (1.565, 2.262, 3.274, and 5.771), the $z_{Mediation}$ index behaves more properly, consistent with the table at large. (For comparison, the denominator of $z_{Mediation}$ is very well-behaved, constantly ranging from 1.0 to 5.0 across the entire Table 4.)

Nevertheless, understanding results and accepting them are different matters. The results in the lower left block of Table 4 render $z_{Mediation}$ lacking in that particular circumstance—when a researcher is working with X, M, and Y all being categorical and expecting full mediation. Granted, most journal articles reporting mediation effects conclude partial, not full, mediation, so this outlying result should be rare. Nevertheless, if a researcher is working with all categorical data and expecting very, very strong mediation effects,[15] $z_{Mediation}$ should still be computed (given that the data may yield effects like 75% or 50% mediation, smaller than but more likely than the exuberant expectations), but admittedly, computing an alternative in addition would be prudent. That alternative would be to fit the equivalent of a structural equation model in the log linear modeling framework (see Iacobucci, 2008). All estimates and statistical tests would be conducted within the categorical and logit frameworks.

It would have been nice if $z_{Mediation}$ performed perfectly under all conditions. Indeed, it mostly did—behaving as it should across diverse conditions of 5 levels of mediation, 4 levels of sample size, and 8 permutations of variable status and analytical applications. For nearly all of these 160 combinations derived from radically varying descriptors, the statistic performed well. Thus, we can acknowledge the power and flexibility this little statistic affords us. We can assess mediation effects of any strength, with any sample size (though 50 or larger is ideal), with X, M, and Y being continuous, categorical, or any mix, and doing so using OLS and logistic regressions. The $z_{Mediation}$ test is an important advance—it is powerful, generally well-behaved, and it extends greatly the flexibility of application for mediation study.

Finally, as a reviewer commented, it is always good practice to create a hold-out sample. The mediation model may be tested on the larger of the two portions of data, and then confirmed (or at least re-tested) in the (usually smaller) hold-out sample. Many mediation tests in the literature are indeed performed on large data sets. For those conducted on smaller samples, this robustness test may simply be an ideal, a direction for subsequent research.

Remarks will follow this article from commentators who are interested in still further extensions. Several are offered, including a consideration of empirical Bayesian estimation techniques.

The solution

In sum, if all variables, X, M, and Y are continuous, fit a structural equation model. If not, here are the new steps to test for mediation. For an example of a continuous variable, consider "dollar amount spent," and for an example of a categorical variable, consider "brand choice" or "ethnicity." Feel free to use SAS, SPSS, or one's favorite statistical computing package.

(1) If Y is continuous, fit the following equation via regression. If Y is categorical, fit it via a logistic regression.[16] The value "b_{01}" is merely an intercept (as always), and the slope c will be produced along with its standard error. This model will estimate the strength of the direct path.

$$\hat{Y} = b_{01} + cX.$$

(2) If M is continuous, fit the following equation via regression. If M is categorical, fit it via a logistic regression:

$$\hat{M} = b_{02} + aX.$$

Collect the parameter estimate a, and its standard error, s_a.

(3) If Y is continuous, fit the following equation via regression. If Y is categorical, fit it via a logistic regression:

$$\hat{Y} = b_{03} + c'X + bM.$$

Collect the parameter estimate b, and its standard error, s_b.

(4) Using the parameter estimates a and b and their standard errors, s_a and s_b, compute the standardized elements:

$$z_a = \hat{a}/\hat{s_a},$$

$$z_b = \hat{b}/\hat{s_b},$$

[15] One warning sign in practice would be if the variables are all categorical and Eq. (1), involving the estimation of the direct effect, is not significant, indeed negligible, on a respectable sample size, and mediation is expected. Then the researcher can examine whether all the explained variance in Y is being carried by the indirect, mediated path.

[16] Technically, again, the model would be a "logit" if X is categorical, and a "logistic regression" if X is continuous, but they would be fit via the same procedure in the statistical computing packages (e.g., "proc logistic" in sas, "logistic regression" in spss or xlstat, etc.).

their product: $z_{a\times b}=z_a z_b$,
and their collected standard error: $\sqrt{z_a^2+z_b^2+1}$.

(5) Finally, compute the pièce de résistance, the z-test that combines results from OLS and OLS, or logistic and logistic, or OLS and logistic, to indicate whether there is a significant mediation effect:

$$z_{Mediation}=\frac{z_a z_b}{\hat{\sigma}_{z_{ab}}}=\frac{\frac{a}{s_a}\times\frac{b}{s_b}}{\sqrt{z_a^2+z_b^2+1}}.$$

Test $z_{Mediation}$ against a standard normal, i.e., it is significant at the $\alpha=0.05$ level if it exceeds |1.96| (for a 2-tailed test with $\alpha=0.05$). Report $z_{Mediation}$ and whether it is significant.[17] Notice that the mechanics of testing for the significance of mediation is very familiar to current practice.

Even more important is that the mechanics of testing for mediation do not change whether the variables are continuous or categorical or some mix, or whether the Eqs. (1)–(3) are OLS regressions or logistic regressions or some combination. In every case, we create the standardized coefficient z_a from Eq. (2), z_b from Eq. (3) and plug them into Eq. (5).

In close, it is worth reiterating the importance of this test: Eq. (5) allows for the translation and combination of OLS and logistic regression results. Mediation analyses can now be conducted no matter whether X, M, and/or Y are continuous or categorical variables.

References

Aroian, Leo A. (1947). The probability function of a product of two normally distributed variables. *Annals of Mathematical Statistics*, *18*, 265–271.

Baron, Reuben M., & Kenny, David A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182.

Briesch, Richard, & Rajagopal, Priyali (2010). Neural network applications in consumer behavior. *Journal of Consumer Psychology*, *20*, 381–389.

Cox, David R., & Snell, E. J. (1989). *The analysis of binary data* (2nd ed.). London: Chapman and Hall.

Craig, Cecil C. (1936). On the frequency function of *xy*. *Annals of Mathematical Statistics*, *7*, 1–15.

DeMaris, Alfred (2002). Explained variance in logistic regression: A Monte Carlo study of proposed measures. *Sociological Methods & Research*, *31*(1), 27–74.

Efron, Bradley (1978). Regression and ANOVA with zero-one data: Measures of residual variation. *Journal of the American Statistical Association*, *73*(361), 113–121.

Fairchild, Amanda J., MacKinnon, David P., Taborga, Marcia P., & Taylor, Aaron B. (2009). R^2 effect-size measures for mediation analysis. *Behavior Research Methods*, *41*(2), 486–498.

Fausett, Laurene V. (1993). *Fundamentals of neural networks: Architectures, algorithms, and applications.* Upper Saddle River, NJ: Prentice-Hall.

[17] To extreme diligence, we might add that if all variables are categorical, in addition to steps (1) through (5), it is worth fitting the direct and indirect associations simultaneously in a single log linear model (cf. Iacobucci, 2008). In this circumstance, if the log linear model indicates significance and the $z_{Mediation}$ does not, the log linear model results are probably those closer to truth.

Gill, Jeff (2001). *Generalized linear models: A modified approach.* Thousand Oaks, CA: Sage.

Glen, Andrew G., Leemis, Lawrence M., & Drew, John H. (2004). Computing the distribution of the product of two continuous random variables. *Computational Statistics & Data Analysis*, *44*, 451–464.

Guadagni, Peter M., & Little, John D. C. (1983). A logit model of brand choice calibrated on scanner data. *Marketing Science*, *2*(3), 203–238.

Hayes, Andrew F. and Kristopher J. Preacher (2011), "*Indirect and Direct Effects of a Multicategorical Causal Agent in Statistical Mediation Analysis*," Unpublished manuscript.

Hogg, Robert V., & Craig, Allen T. (1995). *Mathematical statistics* (5th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Iacobucci, Dawn (2008). *Mediation analysis.* Thousand Oaks, CA: Sage.

Iacobucci, Dawn, Saldanha, Neela, & Deng, Jane Xiaoyan (2007). A meditation on mediation: Evidence that structural equations models perform better than regressions. *Journal of Consumer Psychology*, *17*(2), 140–154.

Imai, Kosuke, Keele, Luke, & Tingley, Dustin (2010). A general approach to causal mediation analysis. *Psychological Methods*, *15*(4), 309–334.

James, Lawrence R., & Brett, Jeanne M. (1984). Mediators, moderators, and tests for mediation. *Journal of Applied Psychology*, *69*(2), 307–321.

Kiviet, Jan F., & Phillips, Garry D. A. (2000). The bias of the 2SLS variance estimator. *Economic Letters*, *66*, 7–15.

Lehmann, Donald (2001). Mediation. *Journal of Consumer Psychology*, *10*(1&2), 90–92.

Lomnicki, Z. A. (1967). On the distribution of products of random variables. *Journal of the Royal Statistical Society: Series B: Methodological*, *29*(3), 513–524.

MacKinnon, David P., & Dwyer, James H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review*, *17*, 144–158.

MacKinnon, David P., Fritz, Matthew S., Williams, Jason, & Lockwood, Chondra M. (2007). Distribution of the product confidence limits for the indirect effect: Program prodclin. *Behavior Research Methods*, *39*(3), 384–389.

MacKinnon, David P., Krull, Jennifer L., & Lockwood, Chondra M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science*, *1*(4), 173–181.

MacKinnon, David P., Lockwood, Chondra M., Brown, C. Hendricks, Wang, Wei, & Hoffman, Jeanne M. (2007). The intermediate endpoint effect in logistic and probit regression. *Clinical Trials*, *4*, 499–513.

MacKinnon, David P., Lockwood, Chondra M., Hoffman, Jeanne M., West, Stephen G., & Sheets, Virgil (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, *7*(1), 83–104.

MacKinnon, David P., Lockwood, Chondra M., & Williams, Jason (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, *39*(1), 99–128.

MacKinnon, David P., Warsi, Ghulam, & Dwyer, James H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research*, *30*(1), 41–62.

McCullagh, Peter, & Nelder, John A. (1983). *Generalized linear models.* New York: Chapman and Hall.

McFadden, Daniel (1974). The measurement of urban travel demand. *Journal of Public Economics*, *3*, 303–328.

McKelvey, Richard, & Zavoina, William (1975). A statistical model for the analysis of ordinal dependent variables. *Journal of Mathematical Sociology*, *4*, 103–120.

Menard, Scott (2000). Coefficients of determination for multiple logistic regression analysis. *The American Statistician*, *54*(1), 17–24.

Mittlböck, Martina, & Schemper, Michael (1996). Explained variation for logistic regression. *Statistics in Medicine*, *15*, 1987–1997.

Mittlböck, Martina, & Schemper, Michael (1999). Computing measures of explained variation for logistic regression models. *Computer Methods and Programs in Biomedicine*, *58*, 17–24.

Muthén, Bengt (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, *49*(1), 115–132.

Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, *78*(3), 691–692.

Nelder, John A., & Wedderbun, Robert W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A*, *135*(3), 370–384.

Preacher, Kristopher J., & Kelley, Ken (2011). Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological Methods*, *16*, 93–115.

Sobel, Michael E. (1982). asymptotic confidence intervals for indirect effects in structural equation models. In Samuel Leinhardt (Ed.), *Sociological methodology* (pp. 290–312). San Francisco: Jossey-Bass.

Sobel, Michael E. (1990). Effect analysis and causation in linear structural equation models. *Psychometrika*, *55*(3), 495–515. Winship, Christopher, & Mare, Robert D. (1983). Structural equations and path analysis for discrete data. *The American Journal of Sociology*, *89*(1), 54–110.

Zhao, Xinshu, Lynch, John G., Jr., & Chen, Qimei (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research*, *37*, 197–206 (August).