Three Essential Analytical Techniques for the Behavioral Marketing Researcher: Median Splits, Mean-Centering, and Mediation Analysis

Dawn Iacobucci
Vanderbilt University, Nashville, USA

Deidre L. Popovich
Texas Tech University, Lubbock, USA

Georgios A. Bakamitsos
Stetson University, DeLand, USA

Steven S. Posavac
Vanderbilt University, Nashville, USA

Frank R. Kardes
University of Cincinnati, Cincinnati, USA

the essence of knowledge
Boston — Delft

Foundations and Trends® in Marketing

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

D. Iacobucci, D. L. Popovich, G. A. Bakamitsos, S. S. Posavac, and F. R. Kardes. *Three Essential Analytical Techniques for the Behavioral Marketing Researcher: Median Splits, Mean-Centering, and Mediation Analysis*. Foundations and Trends® in Marketing, vol. 9, no. 2, pp. 83–174, 2014.

This Foundations and Trends® issue was typeset in LaTeX using a class file designed by Neal Parikh. Printed on acid-free paper.

ISBN: 978-1-60198-912-3

Foundations and Trends® in Marketing

Volume 9, Issue 2, 2014

Editorial Scope

Topics

Foundations and Trends® in Marketing publishes survey and tutorial articles in the following topics:

- B2B marketing
- Bayesian models
- Behavioral decision making
- Branding and brand equity
- Channel management
- Choice modeling
- Comparative market structure
- Competitive marketing strategy
- Conjoint analysis
- Customer equity
- Customer relationship management
- Game theoretic models
- Group choice and negotiation
- Discrete choice models
- Individual decision making
- Marketing decisions models
- Market forecasting
- Marketing information systems
- Market response models
- Market segmentation
- Market share analysis
- Multi-channel marketing
- New product diffusion
- Pricing models
- Product development
- Product innovation
- Sales forecasting
- Sales force management
- Sales promotion
- Services marketing
- Stochastic model

Information for Librarians

Foundations and Trends® in Marketing, 2014, Volume 9, 4 issues. ISSN paper version 1555-0753. ISSN online version 1555-0761. Also available as a combined paper and online subscription.

Foundations and Trends® in Marketing
Vol. 9, No. 2 (2014) 83–174

DOI: 10.1561/1700000038

now
the essence of knowledge

Three Essential Analytical Techniques for the Behavioral Marketing Researcher: Median Splits, Mean-Centering, and Mediation Analysis

Dawn Iacobucci
Owen Graduate School of Management
Vanderbilt University
Nashville, TN 37203, USA
dawn.iacobucci@owen.vanderbilt.edu

Deidre L. Popovich
Rawls College of Business
Texas Tech University
Lubbock, TX 79409, USA
Deidre.Popovich@ttu.edu

Georgios A. Bakamitsos
Stetson University, DeLand
FL 32723, USA
bakamitsos@stetson.edu

Steven S. Posavac
Owen Graduate School of Management
Vanderbilt University, Nashville
TN 37203, USA
steve.posavac@owen.vanderbilt.edu

Frank R. Kardes
Lindner College of Business
University of Cincinnati
Cincinnati, OH 45221, USA
kardesfr@ucmail.uc.edu

Contents

Abstract

For the behavioral marketing scholar, experimentation and the analysis of variance are among the most important and frequently relied upon tools of the trade, and many useful texts exist to guide researchers on these topics. This monograph is intended to be a supplemental resource and a helpful guide for conducting three essential analytical techniques that are also frequently useful to the behavioral researcher: (1) we discuss the practice of conducting a median split on a continuous variable to facilitate communication clarity. (2) We demonstrate the practice of centering variables about their means prior to creating product terms to reflect interaction effects in a moderated multiple regression model. (3) We discuss the practice of a mediation analysis to test for the relative impact of direct and indirect effects of predictors on dependent variables.

D. Iacobucci, D. L. Popovich, G. A. Bakamitsos, S. S. Posavac, and F. R. Kardes. *Three Essential Analytical Techniques for the Behavioral Marketing Researcher: Median Splits, Mean-Centering, and Mediation Analysis*. Foundations and Trends® in Marketing, vol. 9, no. 2, pp. 83–174, 2014. Copyright © 2015.
DOI: 10.1561/1700000038.

1

Introduction: Behavioral Marketing Research Analytics

Our intended audience for this monograph is behaviorally oriented marketers interested in learning more about some core analytical tools of the trade. We believe we can proceed under the assumption that most consumer behavior researchers have a strong, clear foundation in the understanding and execution of experimentation and the analysis of variance (ANOVA). In this monograph, we wish to touch upon several topics that, in practice, are essential complementary analytics that enable behavioral marketing researchers to thoroughly test theories and hypotheses, and thereby advance their respective literatures and contribute to knowledge bases.

Specifically, in Section 2, we discuss the use of median splits: what they are, and given some confusion in the literature, addressing when it is it acceptable and completely appropriate and legitimate to use them. In Section 3, we discuss the use of mean-centering when conducting moderated multiple regressions: again, given a bit of debate in the literature, we will address whether it is worthwhile to adjust one's variables in this manner or not. In Section 4, we discuss mediation analysis: first describing the basic approach, then covering advanced issues around fitting the model, including structural equations models,

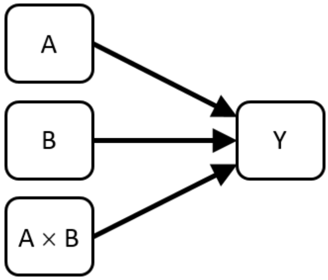

Figure 1.1: The conceptualization of a two-factor experiment and ANOVA.

multi-item scales, and categorical variables. In Section 5, we review and make further recommendations.

To facilitate analytical comparisons, we will provide conceptual diagrams like Figure 1.1 to depict the relationships that a particular analysis is intended to study. Figure 1.1 depicts the familiar ANOVA framework, frequently used for data resulting from an experiment in which two factors were manipulated, factor A and factor B, to see their effects, separately and together, on the dependent variable, Y. The $A \times B$ term represents of course the interaction between factors A and B, which is used to assess whether is there some combination of A and B that provides a particular boost or inhibition to responses on Y. (The $A \times B$ interaction also goes by the name of "moderator," "contingency," or even "boundary condition.") In subsequent sections and alternative models, the categorical factors A and B will be replaced with continuous predictors, X_1 and X_2, for example, and there will be multiple steps when studying a mediated path from X_1 to a mediator M to the dependent variable Y.

To begin simply, consider the following study as an example of the ANOVA model. Imagine the following: a next generation smart phone is about to be launched, and the manufacturer believes its success will depend upon the extent to which the potential adopting consumers understand the nature of its new benefits. Yet the advertising creative is convinced that the phone's design is so sleek as to sell itself, much like the iPhone. So the brand manager wishes to test the comparative effectiveness of two ads. One ad is highly informative, consistent with the manufacturer's expectations. There are several photographs of the phone, with different screen shots illustrating different apps, and a good deal of surrounding text, including the prominent feature of their website for consumers who wish more information. The other ad is consistent with the advertising agency's creative; it is an image-based advertisement, one that features a beauty shot of the phone and is otherwise unencumbered by a great deal of text. This distinction — image ad versus informative ad — will serve as factor A in the experiment.

The marketer wonders if the image ad could be used if complementary information is also made available. Thus factor B will be whether the public relations material that will be issued to preview the product launch would be streamlined, essentially including only a product announcement with few additional details, or whether the expenditure of additional marketing budget will be required to provide more extensive PR information, such as solicited experts' ratings, product testing reports in relevant telecommunications and tech media, etc. Thus, factor B will be whether the PR materials are "minimal" or "more informative."

The advertisements (image or informational) were developed as factor A, as were the PR materials (minimal or more information) as factor B. The two factors comprised a standard orthogonal 2×2 factorial design. (For information on alternative experimental designs, see Cochran and Cox, 1992, Keppel and Wickens, 2004, or Kirk, 2012.)

Before proceeding, let us be clear and state that the data we are about to analyze are not from an experiment on real people (although such a data set would have worked as well). The data were simulated, drawn merely as a statistically random sample from the computer, and here is why that is a good thing. If we had worked with a "real" data set and showed certain properties in the analytic results, readers would not know if there was something peculiar about the data set or sample, whereas now they know this demonstration was made on just some random sample. In addition, we did not want to construct a data set by hand in such a way as to show any property of data analysis to any advantage, because we wanted data that could look as much like any data set that readers might be working with in analyses of their own.

Indeed, we encourage readers to emulate our steps (in every section) on their own data sets to convince themselves in a roughly inductive way, that the principles we will demonstrate hold more broadly beyond the particular, albeit random, data set with which we are working (after all, any given random sample may be peculiar as well.) Accordingly, we simply drew a random sample (using SAS's vnormal subroutine in proc IML) from univariate normal distributions (with means varying with condition) to serve as the dependent variable Y, distributing the data across the four experimental conditions per random assignment. Using this procedure, readers can therefore be confident that the observations we make will hold for most data sets.

To continue, imagine a random sample of 80 consumers was drawn from an online web survey service. Each consumer would be randomly assigned to one of the four conditions; each shown one combination of ad (factor A level 1 for image, or level 2 for informational), and PR materials (factor B level 1 for minimal, or level 2 for more information). After reviewing the ad and PR materials, each consumer would be asked to make a simple judgment that will serve as the dependent variable, Y: on a 9-point scale, how likely is it that they would purchase the new phone (1 = unlikely to 9 = very likely). The data appear in Appendix A and are available from the authors for readers who wish to verify their facility with the models we use and the results we present.

Table 1.1 contains the ANOVA table, and it indicates that both main effects contribute significantly to the consumers' perceptions,

Table 1.1: The ANOVA table for the smart-phone data in Appendix A.

Source of variation	DF	Sums of squares	Mean squares	F value	p-value
A	1	186.05	186.05	104.51	< 0.0001
B	1	180.00	180.00	101.11	< 0.0001
A× B	1	92.45	92.45	51.93	< 0.0001
Error	76	135.30	1.79		
Total	79	593.80			

and there is an interaction — a joint effect of factors A and B together on the dependent variable as well. The means for factor A are: $\bar{X}_{\text{image}-\text{ad}} = 2.42$ versus $\bar{X}_{\text{informational}-\text{ad}} = 5.48$. For factor B, $\bar{X}_{\text{PR}-\text{minimal}} = 2.45$ versus $\bar{X}_{\text{PR}-\text{information}} = 5.45$. These main effects tell us that the informational ad was more persuasive, and the PR packet with more information was more effective than the PR packet with less information.

The interaction means ($\bar{X}_{\text{image}-\text{ad, PRmin}} = 2.00$, $\bar{X}_{\text{image}-\text{ad,PRinfo}} = 2.85$, $\bar{X}_{\text{info}-\text{ad,PRmin}} = 2.90$: $\bar{X}_{\text{info}-\text{ad,PRinfo}} = 8.05$) may be understood quickly and efficiently in the form of the plot in Figure 1.2. The manufacturer's hunch seems to be borne out: consumers were least interested in the phone when they were presented with an image ad and minimal PR materials. Consumers who saw a little information — either the image ad with the more informative PR materials, or the informational ad with minimal PR information — were persuaded only a little bit more. The winning combination appears to be the informational ad with the more informative PR materials.

That description would suffice if these data were analyzed for a marketing manager. However, the marketing journals would require that the implied comparisons be tested. Those test of contrasts within an interaction are called simple effects. Table 1.2 shows the results testing the simple effects. Specifically, when the effect of the ad (image

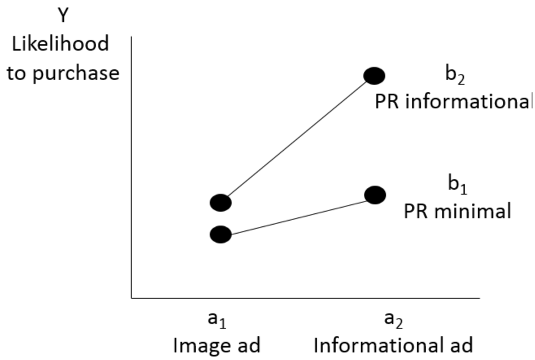

Figure 1.2: The interaction plot for the ANOVA results in Table 1.1.

Table 1.2: Simple effects contrasts added to the ANOVA Table 1.1.

Source of variation	DF	Sums of squares	Mean squares	F value	p-value
Simple effects					
A at b1	1	8.10	8.10	4.55	0.0361
A at b2	1	270.40	270.40	151.89	< 0.0001
Or					
B at a1	1	7.22	7.22	4.06	0.0475
B at a2	1	265.22	265.22	148.98	< 0.0001

or informational) is tested for those consumers exposed to the minimal PR, the means are significantly different ($F_{1,76} = 4.55$, $p = 0.0361$) and when the effect of the ad is tested conditional upon exposure to more informative PR, the means are even more different ($F_{1,76} = 151.89$, $p < 0.0001$). The interaction may be deconstructed for the flip-side simple effects (though note that only one pair would be reported). Here we see that testing the effect of PR (minimal or informative) is significant for the image ad ($F_{1,76} = 4.06$, $p = 0.0475$) and the effect of PR is even stronger for the informational ad ($F_{1,76} = 148.98$, $p < 0.0001$). (For details on how to coax tests of simple effects out of statistical computing packages, see Iacobucci, 1994.)

As this example has illustrated, the ANOVA model is easy to implement, understand, and communicate. We use it as a benchmark basis of comparison for the models that follow in the rest of this monograph.

Before concluding this section, we might mention one other issue because it arises frequently, and that is the tactic to be used when the data are "unbalanced." In the ANOVA context, data are balanced if all the cell sizes are exactly equal. In the 2×2 data in Appendix A, there are $n = 20$ study participants in each cell. When this balance is not true, for example, even if the values of n were: 20, 20, 20, 19, then the design is (slightly) unbalanced. Unbalanced data can introduce bias; depending on its extent, it essentially creates a form of multicollinearity. The good news is that the well agreed upon solution is easy: be sure to report the F-statistics and p-values (and sums of squares and mean squares, if these are reported) from the analysis of "Type III sums of squares," from the computer printout. Basically, there are several ways to estimate the elements of the ANOVA table, and the Type III sums of squares behave statistically the best. They are produced by default in SAS's proc glm, and SPSS's glm. (To learn more about the effects of unbalanced designs, see Iacobucci, 1995, Little and Rubin, 1987, Perreault and Darden, 1975, Searle, 1987.) In the context of types of sums of squares, there is often also discussion of the computation of different types of means, but here it is best to keep it simple. That is, report the regular default means (that you could compute by hand or in Excel), not the "lsmeans" (the least squares means), which estimate what the means would be if the data were balanced.

In the following sections, we will see how to incorporate into the ANOVA model (depicted in Figure 1.1) a variable that measures some individual difference construct, such as consumers' attitudes toward an advertisement or their familiarity with a brand, or even their age or household income. We will also see the slight shift in the form of the model when the primary predictors are both continuous scale measures rather than the discrete factors implemented in an experiment. Lastly, we will reformulate the model further to test whether an effect of some variable on the dependent variable is direct, or indirect through some mediating mechanism.

References

P. D. Allison. Testing for interaction in multiple regression. *American Journal of Sociology*, 83(1):144–153, 1977.

H. J. Arnold and M. G. Evans. Testing multiplicative models does not require ratio scales. *Organizational Behavior and Human Performance*, 24:41–59, 1979.

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

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

D. Baumrind. Specious causal attributions in the social sciences. *Journal of Personality and Social Psychology*, 45(6):1289–1298, 1983.

P. Bentler. Mediation. *Journal of Consumer Psychology*, 10(1, 2):84, 2001.

G. W. Bohrnstedt and T. Michael Carter. Robustness in regression analysis. *Sociological Methodology*, 3:118–146, 1971.

G. W. Bohrnstedt and A. S. Goldberger. On the exact covariance of products of random variables. *Journal of the American Statistical Association*, 64 (328):1439–1442, 1969.

S. J. Breckler. Applications of covariance structure modeling in psychology: Cause for concern? *Psychological Bulletin*, 107(2):260–273, 1990.

W. G. Cochran and G. M. Cox. *Experimental Designs*. Wiley, New York, 2nd edition, 1992.

J. Cohen. The cost of dichotomization. *Applied Psychological Measurement*, 7(3):249–253, 1983.

J. Cote. Mediation. *Journal of Consumer Psychology*, 10(1, 2):93–94, 2001.

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

L. J. Cronbach. Statistical tests for moderator variables: Flaws in analyses recently proposed. *Psychological Bulletin*, 102(3):414–417, 1987.

J. DeCoster, A.-M. Iselin, and M. Gallucci. A conceptual and empirical examination of justifications for dichotomization. *Psychological Methods*, 14(4): 349–366, 2009.

W. P. Dunlap and E. R. Kemery. Failure to detect moderating effects: Is multicollinearity the problem? *Psychological Bulletin*, 102(3):418–420, 1987.

R. Echambadi and J. D. Hess. Mean-centering does not alleviate collinearity problems in moderated multiple regression models. *Marketing Science*, 26 (3):438–445, 2007.

A. G. Glen, L. M. Leemis, and J. H. Drew. Computing the distribution of the product of two continuous random variables. *Computational Statistics and Data Analysis*, 44:451–464, 2004.

R. V. Hogg and A. T. Craig. *Mathematical Statistics*. Prentice-Hall, Englewood Cliffs, NJ, 5th edition, 1995.

L. G. Humphreys. Research on individual differences requires correlational analysis, not ANOVA. *Intelligence*, 2:1–5, 1978a.

L. G. Humphreys. Doing research the hard way: Substituting analysis of variance for a problem in correlational analysis. *Journal of Educational Psychology*, 70(6):873–876, 1978b.

D. Iacobucci. Analysis of experimental data. In Richard Bagozzi, editor, *Principles of Marketing Research*, pages 224–278, Cambridge, MA, 1994. Blackwell.

D. Iacobucci. Analysis of variance for unbalanced data. In David W. Stewart and Naufel J. Vilcassim, editors, *AMA Winter Educators' Conference: Marketing Theory and Practice*, volume 6, pages 337–343, Chicago, 1995. AMA.

D. Iacobucci. *Mediation Analysis*. Sage, Thousand Oaks, CA, 2008.

D. Iacobucci. Meditation analysis and categorical variables: The final frontier. *Journal of Consumer Psychology*, 22:582–594, 2012.

D. Iacobucci and G. A. Churchill, Jr. *Marketing Research: Methodological Foundations.* Earlie Lite Books, Inc, Nashville, TN, 11th edition, 2015.

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

D. Iacobucci, S. S. Posavac, F. R. Kardes, M. J. Schneider, and D. L. Popovich. Toward a more nuanced understanding of the statistical properties of a median split. *Journal of Consumer Psychology*, 25(4):652–665, 2015a.

D. Iacobucci, S. S. Posavac, F. R. Kardes, M. J. Schneider, and D. L. Popovich. The median split: Robust, refined, and revived. *Journal of Consumer Psychology*, 25(4):690–704, 2015b.

D. Iacobucci, M. J. Schneider, D. L. Popovich, and G. A. Bakamitsos. Mean centering helps alleviate 'Micro' but not 'Macro' multicollinearity. *Behavior Research Methods,* forthcoming, 2015c.

J. R. Irwin and G. H. McClelland. Misleading heuristics and moderated multiple regression models. *Journal of Marketing Research*, 38:100–109, February 2001.

J. R. Irwin and G. H. McClelland. Negative consequences of dichotomizing continuous predictor variables. *Journal of Marketing Research*, 40:366–371, August 2003.

J. Jaccard, C. K. Wan, and R. Turrisi. The detection and interpretation of interaction effects between continuous variables in multiple regression. *Multivariate Behavioral Research*, 25(4):467–478, 1990.

K. Jöreskog and D. Sörbom. *Lisrel 8: User's Reference Guide.* SSI Scientific Software International, Chicago, IL, 1997.

G. Keppel and T. D. Wickens. *Design and Analysis: A Researcher's Handbook.* Pearson, Upper Saddle River, NJ, 4th edition, 2004.

R. E. Kirk. *Experimental Design: Procedures for the Behavioral Sciences.* Sage, Los Angeles, 4th edition, 2012.

S. W. Lagakos. Effects of mismodelling and mismeasuring explanatory variables on tests of their association with a response variable. *Statistics in Medicine*, 7:257–274, 1988.

R. J. A. Little and D. B. Rubin. *Statistical Analysis with Missing Data.* Wiley, New York, 1987.

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

R. C. MacCallum, S. Zhang, K. J. Preacher, and D. D. Rucker. On the practice of dichotomization of quantitative variables. *Psychological Methods*, 7(1): 19–40, 2002.

D. P. MacKinnon and J. H. Dwyer. Estimating mediated effects in prevention studies. *Evaluation Review*, 17(2):144–158, 1993.

D. P. MacKinnon, G. Warsi, and J. H. Dwyer. A simulation study of mediated effect measures. *Multivariate Behavioral Research*, 30(1):41–62, 1995.

D. P. MacKinnon, J. L. Krull, and C. M. Lockwood. Equivalence of the mediation, confounding and suppression effect. *Prevention Science*, 1(4): 173–181, 2000.

D. P. MacKinnon, C. M. Lockwood, J. M. Hoffman, S. G. West, and V. Sheets. A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, 7(1):83–104, 2002.

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

D. P. MacKinnon, M. S. Fritz, J. Williams, and C. M. Lockwood. Distribution of the product confidence limits for the indirect effect: Program PRODCLIN. *Behavioral Research Methods*, 39(3):384–389, 2007.

D. W. Marquardt. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics*, 12(3):591–612, 1970.

S. E. Maxwell and H. D. Delaney. Bivariate median splits and spurious statistical significance. *Psychological Bulletin*, 113(1):181–190, 1993.

G. H. McClelland, J. G. Lynch Jr., J. R. Irwin, S. A. Spiller, and G. J. Fitzsimons. Median splits, type II errors, and false positive consumer psychology: Don't fight the power. *Journal of Consumer Psychology*, 25(4): 679–689, 2015.

D. Muller, C. M. Judd, and V. Y. Yzerbyt. When moderation is mediated and mediation is moderated. *Journal of Personality and Social Psychology*, 89(6):852–863, 2005.

R. Neelamegham. Treating an individual difference predictor as continuous or categorical. *Journal of Consumer Psychology*, 10(1, 2):49–51, 2001.

R. Netemeyer. Mediation. *Journal of Consumer Psychology*, 10(1, 2):83–84, 2001.

W. D. Perreault and W. R. Darden. Unequal cell sizes in marketing experiments. *Journal of Marketing Research*, 12:333–342, 1975.

K. J. Preacher and A. F. Hayes. SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, 36(4):717–731, 2004.

P. Royston, D. G. Altman, and W. Sauerbrei. Dichotomizing continuous predictors in multiple regression: A bad idea. *Statistics in Medicine*, 25: 127–141, 2006.

D. D. Rucker, B. B. McShane, and K. J. Preacher. A researcher's guide to regression, discretization, and median splits of continuous variables. *Journal of Consumer Psychology*, 25(4):666–678, 2015.

S. Searle. *Linear Models for Unbalanced Data*. Wiley, New York, 1987.

M. E. Sobel. Asymptotic confidence intervals for indirect effects in structural equation models. In Samuel Leinhardt, editor, *Sociological Methodology*, pages 290–312, San Francisco, 1982. Jossey-Bass.

S. J. Spencer, M. P. Zanna, and G. T. Fong. Establishing a causal chain: Why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology*, 89(6):845–851, 2005.

A. Vargha, T. Rudas, H. D. Delaney, and S. E. Maxwell. Dichotomization, partial correlation, and conditional independence. *Journal of Educational and Behavioral Statistics*, 21(3):264–282, 1996.