Available online at www.sciencedirect.com

ScienceDirect

Journal of Consumer Psychology 25, 4 (2015) 652 – 665

JCP JOURNAL OF CONSUMER PSYCHOLOGY

Research Article

Toward a more nuanced understanding of the statistical properties of a median split☆

Dawn Iacobucci [a,*], Steven S. Posavac [a], Frank R. Kardes [b], Matthew J. Schneider [c], Deidre L. Popovich [d]

[a] Owen Graduate School of Management, Vanderbilt University, Nashville, TN 37203, USA
[b] Lindner College of Business, University of Cincinnati, Cincinnati, OH 45221, USA
[c] Samuel C. Johnson Graduate School of Management, Cornell University, Ithaca, NY 14853, USA
[d] Department of Marketing, Goizeuta School of Business, Emory University, Atlanta, GA 30322, USA

Received 21 June 2014; received in revised form 26 November 2014; accepted 3 December 2014
Available online 10 December 2014

Abstract

Some behavioral researchers occasionally wish to conduct a median split on a continuous variable and use the result in subsequent modeling to facilitate analytic ease and communication clarity. Traditionally, this practice of dichotomization has been criticized for the resulting loss of information and reduction in power. More recently, this practice has been criticized for sometimes producing Type I errors for effects regarding other terms in a model, resulting in a recommendation of the unconditional avoidance of median splits. In this paper, we use simulation studies to demonstrate more thoroughly than has been shown in the literature to date when median splits should not be used, and conversely, to provide nuance and balance to the extant literature regarding when median splits may be used with complete analytical integrity. For the scenario we explicate, the use of a median split is as good as a continuous variable. Accordingly, there is no reason to outright reject median splits, and oftentimes the median split may be preferred as more parsimonious.

Keywords: Median split; Median-split; Dichotomization; Categorization

Introduction

Consumer behavior researchers often consider the relation between continuous independent variables and some criterion. For example, Haugtvedt, Petty, Cacioppo, and Steidley (1988) studied the role in persuasion of the "need for cognition," an individual difference scale reflecting the tendency to engage in and enjoy thinking. These researchers split their sample based on whether participants were above or below the median in need for cognition, and showed that compared to consumers low in need for cognition, those high in need for cognition were more influenced by the quality of an argument in an advertisement and less by the peripheral cue of endorser attractiveness.

Dividing a sample into two groups based on whether each score on a continuous predictor variable is above or below the median prior to conducting analyses is referred to as a median split. For a number of reasons we elaborate shortly, the use of median splits is quite popular. MacCallum, Zhang, Preacher, and Rucker (2002) found that 15.8% of the articles published in a contemporary three-year span of the premier journal, *Journal of Personality and Social Psychology* used median splits. A Google Scholar search of "median split" yields "about 518,000 results," with "median-split" adding another 48,600 and "dichotomization" another 25,300. Even granting that those hits include multiple median splits within a paper, and papers

☆ The authors are grateful to the Editor, Associate Editor, and anonymous reviewers for providing such helpful feedback.
* Corresponding author.
E-mail addresses: Dawn.Iacobucci@owen.vanderbilt.edu (D. Iacobucci), Steve.Posavac@owen.vanderbilt.edu (S.S. Posavac), Frank.Kardes@uc.edu (F.R. Kardes), mjs533@cornell.edu (M.J. Schneider), Deidre.Popovich@Emory.edu (D.L. Popovich).

http://dx.doi.org/10.1016/j.jcps.2014.12.002
1057-7408

critical of median splits (though these have been few in number), the technique is obviously popular. A search within the *Journal of Consumer Psychology* since its inception yielded 178 hits for "median split" or "median-split," and there were 213 hits in the *Journal of Consumer Research*, 1049 for the *Journal of Marketing Research*, and 438 for *Marketing Science*.

Median splits are frequently encountered in many fields. In recent articles published in *The New England Journal of Medicine*, median splits were used to examine the effects of patient conditions and treatments in myocardial infarctions (Kastrati et al., 2011), and the effects of corticosteroid treatment of asthma in children (Lemanske et al., 2010). The usefulness of a median split is easy to understand—a physician might acknowledge that a patient has a degree of symptomatology, yet the question is whether the symptoms surpass a threshold to induce action: prescribe treatment or not. In recent articles published in the *Journal of the American Statistical Association*, median splits were used to examine the driving behavior of teenagers whose friends engaged in risky behaviors (Kim, Chen, Zhang, Simons-Morton, & Albert, 2013), the effects of age and comparative treatments for patients with prostate cancer (Farewell, Tom, & Royston, 2004), and markers to determine the symptoms necessary for establishing a possible new classification of "complicated grief" for the clinical psychologist's *Diagnostic Statistical Manual of Mental Disorders* (Wang et al., 2013).

In still other arenas, median splits were used to study the effectiveness of inpatient treatments for pathological gamblers in Germany (Buchner et al., 2014), condom use among men being treated for sexually transmitted diseases (Crosby & Charnigo, 2013), the effect of friendships vs. isolation on patients' perceptions of lower back pain and quality of life (Hawthorne, de Morton, & Kent, 2013), gesture sequences in chimpanzees (McCarthy, Jensvold, & Fouts, 2013), effective measurement of ability and achievement in football, golf, hockey, and other sports (Robertson, Burnett, & Cochrane, 2014), the effect of time and previous injuries to concussion recoveries (Silverberg et al., 2013), cocaine-induced motor responses in rats (Yamamoto et al., 2013), and the effects of visual stimuli on reaction times and electrophysiological brain activity (Wiebel, Valsecchi, & Gegenfurtner, 2014). In top finance journals, researchers have used median splits to study countries' debt levels and asset values (Kalteier & Posch, 2013), and the effect of venture capital on subsequent corporate financial performance (Hsu, 2013).

As popular as median splits are, researchers have noted potential problems with their use (Cohen, 1983; Fitzsimons, 2008; Humphreys, 1978; Irwin & McClelland, 2003; MacCallum et al., 2002; Maxwell & Delaney, 1993). Although we agree that there is the potential for misleading results when using median splits, we suggest that the cautions against using them need to be updated. In his book, *Statistics as Principled Argument*, Abelson (1995) repeatedly made the point that there are many misconceptions about statistics, and we might argue that misconceptions about median splits should be added to Abelson's list. The core issue is that while current thinking suggests that median splits will always produce inferior analytic conclusions, the reality is much more nuanced. In this article we demonstrate when median splits will, and importantly, will not produce erroneous conclusions.

Specifically, we shall show that median splits, when accompanied by multicollinearity, can cause problems in the analysis of variance (ANOVA) or in multiple regression. While the modeling choice of ANOVA versus regression in and of itself is not the culprit, it is nevertheless typically the case that regression is usually conducted on several continuous predictors, most of which are likely to be at least somewhat correlated, whereas ANOVA is the analytical tool of choice for experiments and orthogonal factors. Experiments are an extremely popular and useful tool for examining the effects of one or more manipulated factors on some focal dependent variable, and are still considered by philosophers of science to be the gold standard for testing causal relationships. In particular for our purposes, a very frequently employed methodology in behavioral research is for a researcher to manipulate one (or more) factors and measure another variable (one but not more) on which is conducted a median split, then an ANOVA is run on this combination of manipulated and measured factors, and the contrasts tested and means plotted. We shall show that this popular scenario, including the use of the median split, is a completely legitimate means of analyzing data, with none of the problems implied by the literature.

This paper is structured as follows. After presenting a summary of the benefits and criticisms of median splits, we report the results of two simulation studies that replicate concerns expressed in prior research, but also demonstrate when median splits are appropriate. Our conclusion is that conducting median splits is perfectly fine when using some research designs.

Why median splits are attractive

As implied by the opening scenario of Haugtvedt et al. (1988) and myriad other published articles, researchers often find using median splits to be attractive (MacCallum et al., 2002). Indeed, Maxwell and Delaney (1993, p. 181) say, "the ubiquitous median split has retained its popularity in many areas of psychology." To try to understand the popularity of median splits, DeCoster, Iselin, and Gallucci (2009) queried samples of theoretical and methodological researchers and identified several reasons that researchers may wish to conduct median splits. Some researchers justified the dichotomization of a variable by saying that doing so "makes analyses easier to conduct and interpret" (DeCoster et al., 2009, p. 350), in particular, allowing the use of the familiar ANOVA model. MacCallum et al. (2002, p. 19) had also stated that researchers believe that "analyses or presentation of results will be simplified." (We reiterate: while the choice to dichotomize or not is distinct from the choice to use a regression or ANOVA, as DeCoster et al. (2009, p. 360) state, the "analysis of continuous variables typically uses regression, and the analysis of dichotomized variables typically uses ANOVA.")

Analyses of contrasts and interactions in particular seem to be better understood via the ANOVA model than regression. For example, DeCoster et al. (2009, p. 361) state that in ANOVA, the researcher can simply provide "the means of each of the groups along with post hoc analyses indicating which groups are significantly different from each other. …[whereas in contrast, the analysis] of a relation between a continuous independent variables and a continuous dependent variable is slightly more complicated. There are no group means to present; instead, the best that a researcher can do is present a regression line illustrating the relation. …[These can be] more difficult to interpret because there are no post hoc tests specifying which values of the independent variable are significantly different from each other."

Thus, many researchers perceive that a median split will make their subsequent analyses more straightforward and readily understood. DeCoster et al. (2009, p. 352) cite that several researchers also mentioned that "it is easier to present the results from analyses with a dichotomized independent variable." That is, because median split analyses are comparatively easy to comprehend, many researchers feel that they are especially useful in communicating complex theoretical ideas to audiences of varying sophistication, including researchers, practitioners, and members of the media. Accordingly, when considering analytical strategies, in some instances researchers may feel that performing a median split may increase the impact of their work.

Another class of explanations that researchers offered for the popularity of median splits was that "analyses conducted with dichotomized indicators may better match the theoretical purpose of the research" (DeCoster et al., 2009, p. 351; cf., MacCallum et al., 2002, pp. 19, 22). For example, some researchers are more interested in group differences rather than individual differences. For these researchers, the contrast between two types of people matches more closely on how they think about their theories and the phenomena they study than any concerns about possible heterogeneity within groups. We might also add that some variables of interest, such as self-esteem or depression, are not easily or ethically manipulated, so their measurement is essential.

In fact, there are numerous constructs that, while being measured on continuous rating scales, are conceptually more discrete, viz dichotomous (MacCallum et al., 2002). For example, locus of control (Srinivasan & Tikoo, 1992) is usually discussed with an emphasis on internal versus external, people are said to be low or high on "self-monitoring" (Becherer & Richard, 1978), and people are said to be low or high in their "need for closure" (Silvera, Kardes, Harvey, Cronley, & Houghton, 2005). Such personality typologies abound: introversion and extraversion, gender identity, type A and B personalities, liberal and conservative, and so forth. When researchers think in terms of groups, or study participants having relatively more or less of a characteristic, it is natural that they would seek an analytical method that is isomorphic, so the data treatment may optimally match the construct conceptualization.

Thus, there are numerous reasons why researchers may wish to use median splits. The popularity of this data treatment is not irrational, nor does it reflect an unwillingness or inability of researchers to learn a different modeling and data presentation technique. We have presented a number of benefits that median splits afford. For a balanced perspective, we next present the criticisms of using median splits, in fact presenting a more thorough investigation of the conditions under which problems arise than has been offered in the literature to date. In doing so, we also identify the conditions under which conducting a median split is perfectly acceptable. In those scenarios, the choices of working with a median split variable (in all likelihood, via ANOVA) or a continuous variable (probably via regression) are statistically equivalent; one is not superior to the other. If two methods are stochastically equivalent (within circumscribed parameters to be explicated), then the choice between them rests on other, non-statistical criteria, such as parsimony and other qualities that many researchers believe favor median splits.

Concerns about median splits

In this section, we consider whether the use of median splits might undermine the veracity of researchers' analytical conclusions. The traditional concern with median splits involves the loss of information about individual variability (Farewell et al., 2004; Humphreys, 1978; MacCallum et al., 2002; Neelamegham, 2001). Indeed, individuals who are above the median are classified in the "high" group regardless if they are only slightly above the median or extremely high. Similarly, individuals who are only slightly below the median, as well as those with extremely low scores, are aggregated into the "low" group. A potentially problematic implication of losing individual-level information is that subsequent analyses may be less likely to find support for hypotheses because median splits may increase the likelihood of Type II errors by reducing effect sizes and experimental power (Humphreys, 1978; Lagakos, 1988). Cohen (1983) demonstrated the effect of median splits on correlation coefficients. For two bivariate normal variables, X and Y, if one variable is dichotomized, the resulting correlation will be only 0.798 (or roughly 80%) of the size of the original. For example, observed correlations between continuous X and continuous Y of values 0.8, 0.6, and 0.4 would be reduced to values of 0.638, 0.479, and 0.319, respectively if one of the variables was dichotomized.

Given that median splits may reduce power and increase the potential for Type II errors, the question arises as to the severity of this problem. Although researchers may perceive that median splits facilitate their aims of analytic and communication simplicity, these benefits come at a cost. However, the solution is relatively easy—researchers can simply draw large enough sample sizes to offset any reduction in power. Of course, given that null hypotheses are never "accepted," and experiments featuring null results do not form the basis of peer-reviewed publications, there is no material risk to science posed by median splits pertaining to Type II errors.

If the only untoward consequences of median splits were to make analyses more conservative, there would be little concern. Indeed, although median splits may be perceived as suboptimal from the perspective of power, if there was no possibility that they could produce misleading support for apparent relations between variables that, in truth, are spurious, their use would not be a problem. However, if median splits can produce Type I errors (the false conclusion of an effect), their use would be inappropriate.

Let us examine this concern, that median splits may increase the likelihood of Type I errors. Irwin and McClelland (2003, p. 370) say, "…in multiple regression, dichotomization of several continuous predictor variables can exacerbate or disguise model misspecification and can induce models that do not accurately reflect the underlying data." Specifically, "…one predictor variable may appear significant in the multiple regression when it is dichotomized but not when it is not dichotomized (i.e., a spurious effect)," and "…predictor variables can appear significant when they are not significant in the original [continuous] data." To provide support for these points, they direct readers to Maxwell and Delaney (1993), who reported analyses of a hypothetical data set that produces results consistent with the idea that conducting median splits increases the risk of Type I error (which we examine shortly).

MacCallum et al. (2002, p. 28) take a similar position based on results reported by Maxwell and Delaney (1993), and say that, "…although under many conditions dichotomization of two independent variables will result in loss of effect size for main effects and interaction [i.e., the traditional conservative results], it was shown that under some conditions dichotomization can yield a spurious main effect." Specifically, a spurious main effect, "…is likely to occur after dichotomization of both predictors" when "…the independent variables are correlated with each other."

Fitzsimons' (2008, p. 5) editorial seems to express the judgment wide-spread throughout the field. He references Irwin and McClelland (2003) and Maxwell and Delaney (1993), and argues that median splits should never be used because doing so, "…can at times create spurious significant results if the independent variables are correlated."

If Maxwell and Delaney's (1993) core claim that a median split facilitates spurious findings in subsequent analyses is general and robust, then the avoidance of median splits would be a foregone conclusion. Our question is whether this finding is indeed an apt basis for the summary dismissal of median splits. Our examination into the question is important because conventional wisdom in the field is based largely on their results.

To obtain this central finding, Maxwell and Delaney (1993) constructed a hypothetical data set to show that median splits can produce misleading results that erroneously suggest the presence of an effect. Their data set had a sample size of 16, and three variables, X_1, X_2, and Y. A regression found only one significant main effect. When Maxwell and Delaney (1993) performed median splits on both X_1 and X_2, a subsequent ANOVA produced significant results for both main effects, which they attributed to median splits creating spurious results.

Median split detractors may certainly argue that one instance of spurious findings using median splits proves the existence of increased Type I error risk, and that accordingly using them should be avoided. However, it is important to examine whether a data set like the one Maxwell and Delaney (1993) created would be encountered in actual behavioral research. In describing their data, we may note that: i) theirs was an extremely small sample size, ii) they conducted median splits on both predictor variables (rather than the more common practice of conducting a median split on just one variable), and iii) their data were created with unrealistically high correlations. In particular, their first explanatory variable is essentially equivalent to their dependent variable: $r_{X_1Y} = 0.996$. Their second explanatory variable is also highly correlated with the dependent variable: $r_{X_2Y} = 0.742$, and there is considerable multicollinearity: $r_{X_1X2} = 0.745$. Thus, although Maxwell and Delaney's (1993) hypothetical data do demonstrate a case in which median splits produce spurious results, the data set is so contrived that whether their results have any relevance to actual behavioral research is an open question. It is a little surprising that their finding biased estimates under such limited conditions has had such prescriptive impact.

To investigate whether these issues arise more generally, we conducted two Monte Carlo studies. The purpose of these studies is to provide a more thorough and comprehensive understanding of where median splits cause problems and where they do not.

Study 1

To understand the precise nature of the impact of a median split, we ran a Monte Carlo simulation on four parameters. We varied sample size and the correlations among X_1, X_2, and Y. Sample size took on values N = 50, 100, and 200. The correlations for $\rho_{X_1,Y}$ and $\rho_{X_2,Y}$ took on values 0.0, 0.1, 0.3, 0.5, and 0.7. The correlation between the predictors X_1 and X_2 varied as well, from 0.0, 0.1, 0.3, 0.5, and 0.7 to allow for varying levels of multicollinearity. Together, these factors resulted in a 3 × 5 × 5 × 5 factorial design. Given that the full design yields results requiring a large table with 375 cells, we plot a subset of results. We will examine the results for the full design shortly as well, presenting the subset of results for ease of illustration.

For every combination of the design factors, N, ρ_{X_1Y}, ρ_{X_2Y}, and ρ_{X_1X2}, in each of the 375 cells in the 3 × 5 × 5 × 5 factorial, a sample was drawn from a multivariate normal distribution to obtain N observations, with population parameters $\rho_{X_1,Y}$, $\rho_{X_2,Y}$, and $\rho_{X_1,X2}$. A multiple regression model was used to analyze the three variables, and the estimates for β_1, β_2, and β_3 (for the interaction term) were obtained. Next, a median split was performed on X_1, and another regression was run and new β estimates were obtained. The βs from the fully continuous model and those from the model with the median split variable were obtained for comparison. This procedure was repeated 10,000 times in each cell of the design. (See the appendix.)

Fig. 1 presents the results for β_1. The plotted means have been averaged over 10,000 replications in each cell as well as levels of N (which impacts *p*-values but not estimates of β_1) and $\rho_{X_1,Y}$ (β_1 is larger for larger $\rho_{X_1,Y}$, as expected). Fig. 1 shows the classic finding, that the median split weakens the apparent effect of X_1 on Y; the entire line for the median split results falls beneath that for the continuous variable. Indeed the ratio of the median split effects to the continuous effect sizes are close to what Cohen (1983) predicted (i.e., 0.240 / 0.300 = 0.80, 0.178 / 0.230 = 0.77, and 0.142 / 0.199 = 0.71). Multicollinearity has only a minor impact on the estimate of β_1, and in all cases, the median split yields the traditional conservative result.

Fig. 2 presents the results for β_2. In the presence of any multicollinearity, the results indicate a lift for β_2 when X_1 is a median split compared to when X_1 is continuous. Thus the results that Maxwell and Delaney (1993) illustrated exist even for our data in which multicollinearity is far less extreme than in their dataset. What seems to be occurring is that the relationship between X_1 and Y is diminished given the median split on X_1, so the burden of explaining variance of Y falls more on variable X_2. That is, X_2 compensates for the relatively weaker X_1.

Note that there are no spurious results when X_1 and X_2 are uncorrelated. That is, a sharing of variance—a change in the effects of X_2 due to a modification in the treatment of X_1—is only possible when the correlation between X_1 and X_2 is nonzero. Thus, potentially spurious results are not a function of median splits alone, but rather they are due to a combination of median splits and multicollinearity. We acknowledge that the absence of multicollinearity is an unlikely scenario in regression studies—researchers are likely to model the effects of, say, attitudes toward the ad and brand on purchase likelihood, and the attitudes will almost certainly be at least modestly correlated. However, just as we granted that Maxwell and Delaney (1993) found a case in which median splits can lead to erroneous conclusions, in the interest of science and truth, it is equally important to understand when median splits may perform equally well.

Note also that while median splits increased the size of some effects, those increases were very small. Fig. 2 shows that the enhancement to β_2 is quite modest. Specifically, when

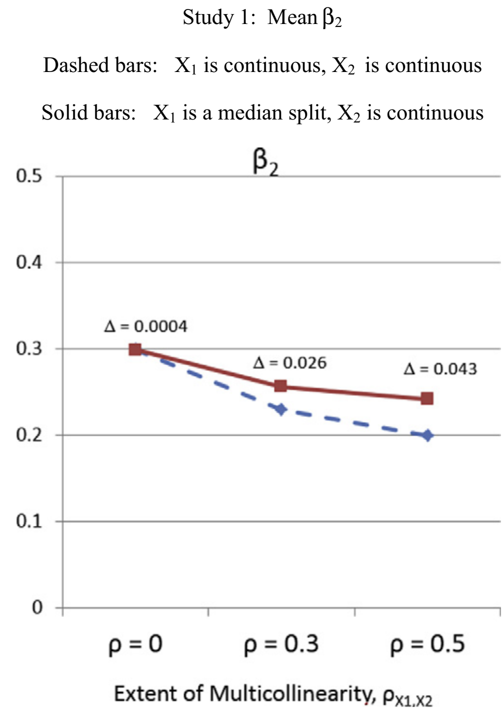

Fig. 2. Study 1: mean β_2. Dashed bars: X_1 is continuous, X_2 is continuous. Solid bars: X_1 is a median split, X_2 is continuous.

$\rho_{X_1,X2} = 0.0$, the mean for β_2 when X_1 was treated as a median split was 0.299 and when X_1 was a continuous variable, the mean β_2 was 0.2994, for a difference of −0.0004. When $\rho_{X_1,X2} = 0.3$ the mean β_2s for the median split and continuous X_1 were 0.256 and 0.230, respectively, for a difference of 0.026. Finally, for still greater multicollinearity, with $\rho_{X_1,X2} = 0.5$, the mean β_2s for the median split and continuous X_1 were 0.242 and 0.199, respectively, for a difference of 0.043.

Fig. 3 presents the results for β_3, the interaction term. The findings are similar to those on β_1 in Fig. 1, in that the continuous model is more powerful, detecting a stronger effect than in the model in which X_1 is treated as a median split. This finding, and that for β_1, are in the classic direction of being more conservative, and hence are not overly problematic to the scientific pursuit. This finding, that the interaction terms are not disturbed, is perhaps even more important than the finding that the main effects are not spurious, given the central importance of interaction predictions and results in many research papers.

When the data from the full 3 × 5 × 5 × 5 design are analyzed, every effect is significant (due to there being 10,000 observations in each cell). Therefore a more telling index is η^2, the proportion of variance attributed to each predictor; for example, the strength of the relationship between X_1 and Y alone explains 67.1% of the variance in β_1, and the relationship between X_2 and Y explains 67.0% of the variance in β_2.

In sum, Study 1 first confirms the effect that a median split will make more conservative effects involving the median split variable itself. Study 1 secondly verifies, and shows more comprehensively than in the literature to date, the possibility of a median split on one variable producing spurious findings on

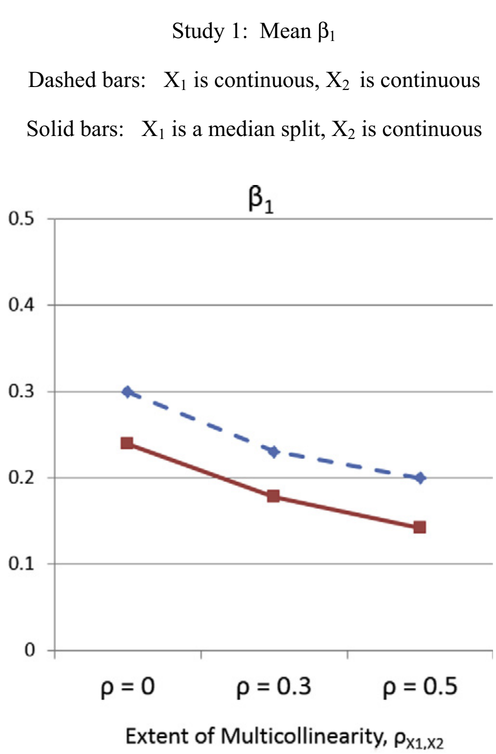

Fig. 1. Study 1: mean β_1. Dashed bars: X_1 is continuous, X_2 is continuous. Solid bars: X_1 is a median split, X_2 is continuous.

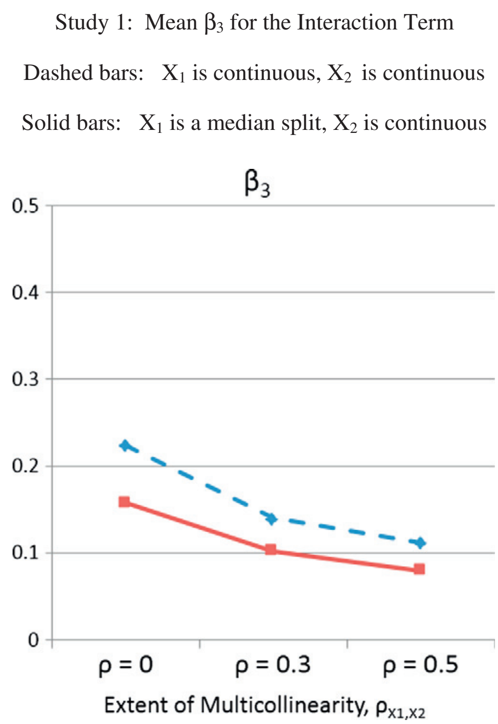

Fig. 3. Study 1: mean β_3 for the interaction term. Dashed bars: X_1 is continuous, X_2 is continuous. Solid bars: X_1 is a median split, X_2 is continuous.

another. This reversal occurred only when the predictor variables showed multicollinearity, and even then, the effects were very small. Median splits had no deleterious effects when the predictor variables were uncorrelated with each other.

Next, just as Maxwell and Delaney (1993) examined both ANOVA and regression, we will do the same, having seen the effects of median splits on regression in Study 1, and next, turning to examine the effects on ANOVA in Study 2. Most researchers who conduct median splits are more interested in working with ANOVA than regression. Given the relationship between ANOVA and regression as both being variants of the general linear model, one might expect the results to be similar. Yet similar findings are not a foregone conclusion given that one way in which ANOVA and regression differ in practice is that ANOVA is often applied in experimental settings, in which manipulated factors are orthogonal. That is to say, ANOVA is usually used when the predictors are uncorrelated with each other, and we saw in Fig. 2 that in this circumstance, even regression produced acceptable, unbiased results. It is not the case that ANOVA is superior to regression, only that it is more frequently used in situations wherein independent variables are uncorrelated and therefore would produce correct results. Study 2 addresses the question of what effect median splits have in the experimental and ANOVA modeling context.

Study 2

Study 2 follows Study 1's Monte Carlo procedure. The experimental design varied sample size (N = 50, 100, 200), and the correlation between X_1 and the dependent variable Y, with $\rho_{X_1,Y}$ translating to mean differences for factor A (the median split of X_1) of μ_{A1} vs. μ_{A2}: 0.00 vs. 0.00, −0.2 vs. 0.2, −0.4 vs. 0.4, −0.6 vs. 0.6, and −0.8 vs. 0.8. That is, factor A had effect sizes of 0.0, 0.4, 0.8, 1.2, or 1.6 standard deviation units. Next, a manipulated variable, X_2 was created by simulating a "heads or tails" toss of the coin, to mimic the random assignment of study participants into one of the two experimental conditions. We did impose the restriction, toward the end of the random assignment draws, that the cell sample sizes be equal, an action akin to conducting a real experiment (this restriction is not critical to our results).

For each sample, ANOVA was used to model the median split on X_1 as factor A, X_2 (the heads or tails factor) as factor B, and an interaction term A × B. Population mean differences were built into the Monte Carlo study for factor A, but population effects for B and the interaction were null—the point of the study was to determine whether the implementation of a median split on A would introduce spurious results on B or the interaction, hence, they had to begin equated, i.e., $\mu_{B1} = \mu_{B2}$, and $\mu_{A1B1} = \mu_{A1B2} = \mu_{A2B1} = \mu_{A2B2}$. The *p*-values indicating whether each of the main effects and the interaction was significant or not (and which serve as rough estimates of effect sizes) were noted. This analysis was repeated 10,000 times in each cell of the design.

Fig. 4 presents the average *p*-values for the main effect of factor A (the median split X_1), the main effect for factor B (the heads–tails assigned condition), and the interaction term, A × B. The *p*-values for A show that when there are no mean differences in the population between μ_{A1} and μ_{A2} (for the effect size of 0.0 standard deviations), then as expected, for approximately 50% of the samples, M_1 is greater than M_2, and for the other 50% the reverse is true, such that the average *p*-value is 0.5. The data are null, as they should be. As the effect size grows, from μ_{A1} and μ_{A2} differing by 0.4 standard deviations to 1.6 standard deviations, the results show the predictable effect of approaching significance, particularly for a larger N.

What is even more reassuring is that creating a median split for factor A does not markedly affect the results for the other main effect, B, or the A × B interaction. The *p*-values for factor B are essentially 0.500; while they veer downward slightly for increasing effects on A, they do not approach significance even for the largest effect size (the largest mean differences). The *p*-values for the interaction term A × B are similarly 0.500, as is appropriate given that there are no mean differences in the population. Together, these results indicate that a median split on factor A does not increase the likelihood of finding a significant effect for the other main effect or the interaction.

These results are excellent news for the researcher who wishes to conduct a median split on a continuous variable in order to use it as a factor in ANOVA. Doing so will not yield misleading results on the factor itself. When a result should be null, it will be null, and in contrast, larger effects will be detected. It is also extremely important to note that conducting a median split to create one factor will not result in spurious findings for another factor, nor for the interaction between the two. Researchers can be confident in the integrity of their results should they choose to use a median split on one

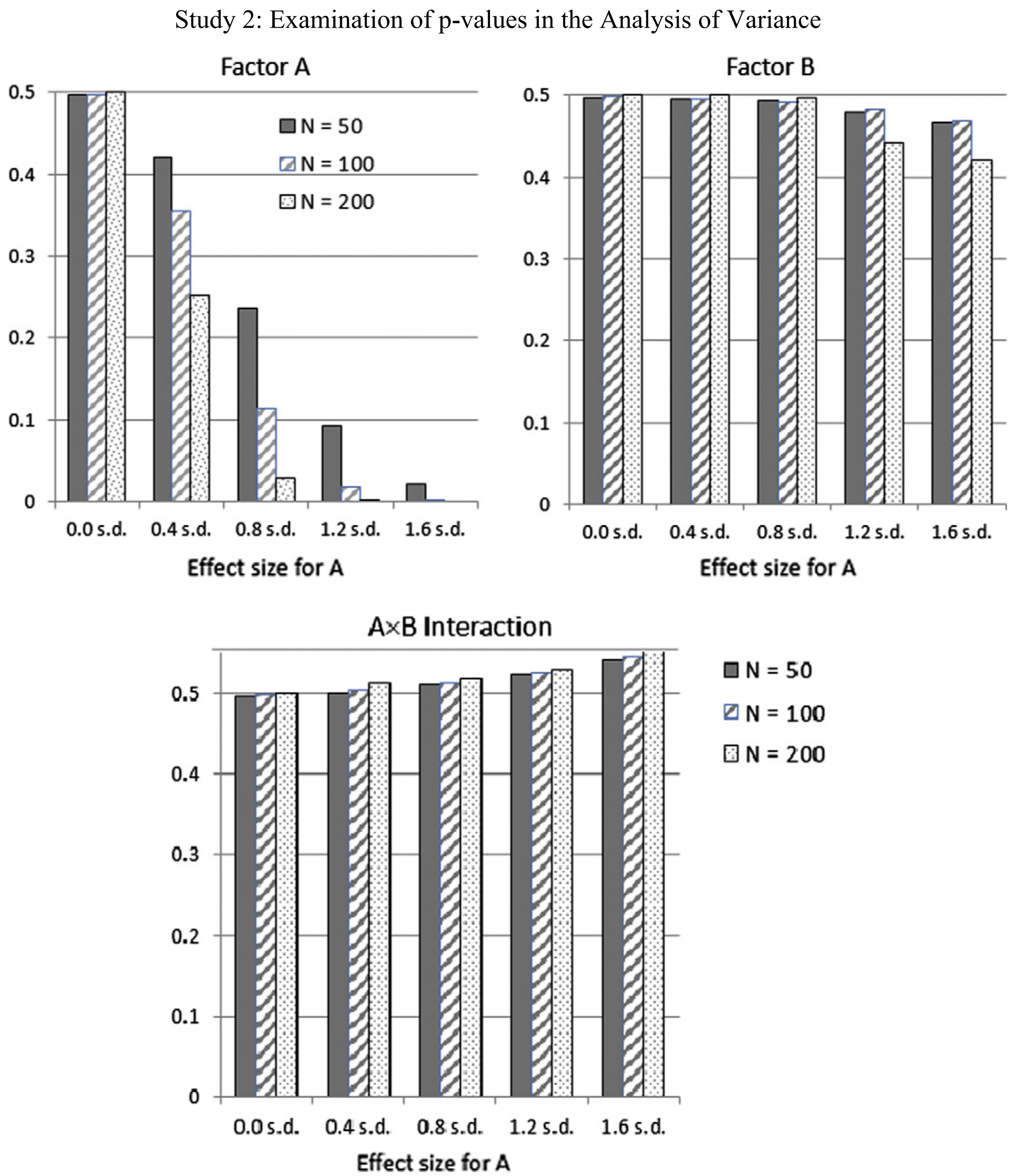

Fig. 4. Study 2: examination of *p*-values in the analysis of variance.

explanatory factor when analyzing data in conjunction with an orthogonal, experimentally manipulated factor via ANOVA.

To be clear, it is not that ANOVA is superior to regression. It is simply the case that when using regression models, the predictors are usually at least somewhat correlated, so altering one variable can affect the other relationships. ANOVA is simply more frequently used on data with orthogonal predictors, so altering one variable will not impact the other relationships. Even for regression, when there was no multicollinearity, Study 1 had shown that the regression results were accurate and not inflated.

Discussion

In recent years the use of median splits has been largely discouraged. Indeed, given current conventional wisdom, at present a researcher who submits a paper that includes a median split is almost certain to provoke the ire of the review team. Our results demonstrate that such criticism would be unwarranted—the outright rejection of the use of median splits is not warranted. Specifically, although we replicated (and extended) the core findings of research that advises against the use of median splits, our results also demonstrate the circumstances in which statistical conclusions based on median splits are perfectly legitimate.

In our paper, Study 1 replicated the most easily recognized risk of conducting median splits; that of diminishing the apparent relationship between the dichotomized X_1 and Y. Study 1 also demonstrated that when multicollinearity was present, a median split on X_1 can impact the relationship between X_2 and Y, although Type I errors were minor.

In our paper, Study 2 offers a compelling defense of median splits because performing a median split on a single variable used in conjunction with an orthogonal experimental design factor and modeled via ANOVA introduced no problematic results on any effect in the model; the median split used to create factor A did not create spurious results for B, or the A × B interaction. We are not claiming that median splits in ANOVA are a superior means of modeling data compared to continuous scales in regression. Rather, the predictors in ANOVA are more likely to be uncorrelated than those in regression models.

We believe that our research is important and it implies guidelines for researchers going forward, which we elaborate shortly. Our findings are also important because they indicate that researchers need not summarily question the validity of published articles that had implemented median splits. Let us consider for a moment the implications of our research for those published articles.

Past research

Our research provides insights regarding articles already published that have used median splits. If a journal article reported a study in which a median split was used in conjunction with one or more experimental factors, the results are completely valid and should not be questioned. If a median split was used as one variable among correlated predictors, any potential estimation bias would have been minimal. Let us explain.

We begin by saying that one obvious conclusion that may be drawn from our findings is that while it is true that in past research median splits may have indeed spuriously enhanced some findings when predictor variables were correlated, when they occurred, their impact would have been negligible. We acknowledge that even with the tiny sizes of the effects we documented, it is conceivable that a median split treatment of one variable might have enhanced the size of the effect of another, and with sufficient sample size, an otherwise borderline effect may have been bumped by 0.043 (per the largest delta in Fig. 2). Yet it is unlikely that the bump would have helped the effect surpass the dichotomous threshold of significance given that the requisite sample size for an effect of size 0.043 would be vastly larger than typical behavioral studies. Even so, imagine the increase happened, and the article had been published with that particular result being spurious. Taking a broader perspective, we remind the reader that the answer to the question as to whether some published findings might actually be Type I errors is: yes, presumably 5% or so are overstated, whether due to median splits, unreliable measures, improper experimental execution, human error—both subject and experimenter, quantitative model misspecification, etc. Presumably those spurious effects will not be replicated in future studies that seek to build on and extend the previous findings. The scientific process is imperfect; nevertheless, scientific progress is made.

A second conclusion that may be drawn from our research is that the occurrence of possible spurious findings would only have occurred in the presence of multicollinearity. Not all articles publish the correlation matrix of variables included in the analyses, so it would be difficult to assess the data from published articles. Nevertheless, given our findings, a not improbable conjecture would be that if a median split variable had been used as a factor in an experiment with one or more experimental factors in an orthogonal design, the findings that were reported are probably valid. Correlated variables can arise through the choice of some experimental designs. For example, incomplete experimental designs (entirely missing cells) can introduce non-orthogonal effects, though the practice of creating incomplete designs is rare, used mostly for very large-scale engineering and conjoint problems that require the manipulation of many variables, and even for those researchers, balanced versions of incomplete designs are sought (Cochran & Cox, 1992; Kirk, 2012). Other forms of experimental designs also necessitate care; e.g., nested factors can be orthogonal if median splits were conducted within nested groups, otherwise the nesting factor may be correlated with the median split factor. Of course this concern is true for any experimental manipulation as well, that it should be orthogonal if implemented within groups. Thus, many types of experimental designs bring their own challenges; luckily many of them are also rarely used. Presumably there is consensus that the most frequently implemented experimental design is the 2 × 2 factorial, which is an extraordinarily robust machine. Finally, aside from the experimental design issues, multicollinearity, in general, in all statistical models, regression included, has the overall effect of dampening effect sizes, and making it less likely that effects are significant. Thus, we suspect that if a median split had been used in the presence of other correlated variables, it would be less likely that the study would have been successful in yielding interesting significant findings. Instead, the study is probably sitting unpublished in the researcher's "file drawer."

We turn next to the consideration of some related issues that have arisen, from reviewers and colleagues as natural extensions or applications of our research. We address the concerns of: 1) naturally occurring vs. experimenter created groups, 2) experimental design issues that can help, 3) measurement issues that can help, 4) the effect of skewness or non-normal distributions, 5) observations regarding two median splits, and 6) the recommendation of empirical verification of lack of multicollinearity.

Related concerns

First, there is a debate among personality theorists referred to in MacCallum et al. (2002) regarding the distinct phenomenon of naturally occurring groups (e.g., men vs. women) compared to groups that are created by experimenters when they seek to study people who are "high" vs. "low" on some measure (e.g., need for cognition). MacCallum et al. (2002) suggested that dichotomization may be defensible if there exists a true binary typology. They mention the literature that distinguishes a taxon, defined as a genetically determined categorization, from groups that are defined as superimposed by a researcher. While we find this debate interesting, it is independent of and theoretically separate from the research we have presented. We have tried to focus as cleanly as possible on the empirical behavior of some statistics—continuous and median split variables in ANOVA and regression models. We have shown circumstances for which median splits are and are not acceptable on statistically theoretical grounds with empirical testing and demonstrations. Under those circumstances for which median splits are acceptable, on statistical grounds, they may be used to compare two groups whether those groups are defined by a genetic distinction or by a researcher interested in the differences between those who score "high" or "low" on some marker.

Second, certain experimental procedures may enhance the likelihood that the experimental factors are uncorrelated with the median split on an individual difference measure. Typically, an experiment unfolds with the measurement of the central dependent variable(s), then perhaps manipulation checks, then likely the individual attitudes, personality, and demographic covariates. While certain attitudes and personality traits are presumably stable and likely to remain unaffected by an experimental manipulation or the measurement of its effectiveness (as might be a concern with some sort of halo ratings), it might be worthwhile to put somewhat later in the survey the measure of the variable to be used as a median split (i.e., not immediately following the dependent variable or manipulation checks).

In addition, one can assume in the design stage (and verify once data have been collected) that certain constructs and measures are more and less likely to be correlated. For example, in a recent *Journal of Consumer Psychology* article, Packard and Wooten (2013) found that greater consumers' knowledge discrepancy (defined as the difference between the consumers' ideal and actual product category knowledge) resulted in consumers expecting to share product reviews with more close acquaintances (friends, family, coworkers), compared to strangers when the researchers tested for this interaction in a regression involving a continuous measure of knowledge discrepancy and plotted the data at plus and minus one standard deviation, a perfect analysis. Imagine another set of researchers seeking to replicate Packard and Wooten's (2013) basic findings, but replacing the closeness of the intended recommendation targets with a measure of another construct. If the new construct were something like sensitivity to product country of origin, we might expect that it would not be highly correlated with knowledge discrepancy, compared to the scenario if the new construct were brand expertise.

Third, certain measurement practices may be helpful when using median splits. If a researcher were to use a rating scale with few integers, such as a 3- or 4-point scale, for example, the empirical median would likely be, say 2 or 3, and the resulting distribution of coding a "2 (or 3) or higher" as "high" and lower scores as "low" might yield sample sizes that might not be even roughly equal empirically, in a way that a median split is intended to do theoretically. This issue is not as problematic for continuous scales (e.g., number of past purchases of a focal product, or hours spent online in a typical week reading product-related materials), or medians to be constructed on consumers' means over multi-item scales (e.g., need for cognition and self-monitoring); scales with more choice points are likely to contribute to a more precise balance. If a median split is to be conducted over a single integer scale, it would probably behoove the researcher to use a 7-point, 9-point, or higher number of points along a rating scale to enhance the likelihood of equal cell sizes after the median split is conducted. In this same realm, a researcher is free to choose between two coding schemas when creating a median split variable: either "all values on the continuous variable at or below the median are low and all values above are high," or "all values on the continuous variable below the median are low and all values at or above the median are high." The researcher would be best served to choose between the coding rules that which creates a split as close to 50:50 as empirically possible.

Another issue related to both measurement and statistical theory involves interaction plots. Consider this example of a recent *Journal of Consumer Psychology* article: Goodman, Broniarczyk, Griffin, and McAlister (2013) found that consumers with more developed preferences experienced greater decision difficulty when confronted by conflicting information as measured by longer decision times and larger consideration sets when the researchers tested for this interaction in a regression involving a continuous measure of preference development and plotted the data at plus and minus one standard deviation, a perfect analysis. Imagine another set of researchers seeking to replicate these basic findings, but they wish to replace the concept of preference development with an established scale of expertise. They think of the construct of expertise as yielding two groups—novices and experts, so they wish to present their findings by creating a median split on expertise, testing the significance of effects via ANOVA, and plotting the means in each group. With knowledge of the properties of a standard normal curve, we can see that these sets of researchers are engaging in highly similar behavior. The quartiles of a normal curve are -0.67, 0.0, and 0.67, and the means of the values below and above 0.0 are -0.8 and $+0.8$. The first set of researchers plotted regression predictions at $\pm$ 1.0 standard deviation units; perfectly fine. The second set of researchers is considering plotting means at ± 0.8 standard deviation units; not substantially different.

Fourth, we have focused on data that may have arisen from normal distributions, easily defensible for individual difference variables and certainly for averages computed over multi-item scales due to the normalizing effect of the central limit theorem. A natural question seeks to understand the extent to which our findings generalize to other distributions. For example, in a recent *Journal of Consumer Psychology* article, Mok and Morris (2013) found that consumers high in bicultural identity integration were more likely to examine individualistic (vs. collectivistic) information after an American (vs. Asian) prime when they tested for this interaction in a regression involving a continuous measure of bicultural identity integration and plotted the data at plus and minus one standard deviation, a perfect analysis. Imagine another group of researchers desiring to replicate and extend Mok and Morris's (2013) findings, but perhaps they are scholars at universities where a minority population is rare so their sample is skewed—are they at risk if they created a median split? To investigate the effect of different distributions, we reran the simulation that gave rise to Figs. 1 and 2, varying the shape to one of the four alternative population distributions: we created skew in X_1 by squaring it, a different skew by taking the natural logarithm, a bimodal distribution, and a uniform distribution. The differences between the β_1 for a continuous treatment of X_1 and its median split version were: −0.001 for the squared skew, 0.056 for the natural log skew, 0.014 for the bimodal distribution, and −0.003 for the uniform distribution. The differences on β_2 were all 0.000, and the differences on β_3 (for an interaction term) were also negligible, 0.002. It would appear that our findings are more robust than we had even anticipated, holding true across multiple forms of distributional shapes.

We ran a final form of skewness—we created a population with naturally occurring groups of proportions 75% and 25%. If the natural groupings are distributed 75:25, it would be artificial to superimpose a 50:50 split. Such unbalanced designs (unequal cell sizes) can, but do not necessarily, contribute to multicollinearity. One way to maintain orthogonality among the factors is to randomly assign to levels of the experimental factor within the 75% group and separately, randomly assign members in the 25% group. When we created even this extreme distribution, our results continued to show modest and negligible differences between the continuous treatment of X_1 and its median split version on β_1 of 0.080, and 0.003 on β_2.

Fifth, we have tried to be clear in our support of conducting a median split on one variable to be used in conjunction with other, orthogonal experimental factors. We have said little about the sensibility of using two median split variables in the same model. Theoretically, we can state that more than one median split might be used if it can be shown that the median splits are not significantly correlated with each other and the experimental factors, but we suspect such scenarios to be unlikely empirically. The problem is that by definition, a median split on X_1 assigns 50% of the observations to the "low" group and 50% to the "high" group. However, conducting a median split on X_1 and another on X_2 does not guarantee that the cell sample sizes in the resulting 2 × 2 table would be distributed evenly 25% across the cells. Indeed, given that median splits are usually conducted on continuous variables representing individual differences (e.g., personality traits and brand knowledge), those individual differences scales are likely to be correlated at least $r = 0.30$, a pattern which would certainly end up reflected in the 2 × 2 table. Hence the relationship between two median split factors is not likely to be orthogonal, and therefore, ANOVA (or regression) could yield somewhat spurious results.

Yet to put that concern in perspective, MacCallum et al. (2002) had documented that while median splits were frequently used, e.g., 15.8% prevalence in the *Journal of Personality and Social Psychology*, the simultaneous use of two median splits was rare—found in only 1.4% of the articles. Furthermore, a truism regarding two median splits is somewhat ironic—the reduction in power that would result from two median splits, 64% (or 80% for each split, Cohen, 1983) would reduce the correlation between two continuous variables in their median split form, thereby reducing the multicollinearity between them. For example, if constructs X and Y measured in their continuous form were correlated 0.5, say, then their median split forms would be correlated a mere 0.32. A smaller continuous correlation of 0.3, representing the beginnings of multicollinearity, would be reduced to 0.19. Lessening the multicollinearity enhances the likelihood that two median splits would not boost spurious effects and yield proper results. Researchers could report the phi coefficient—the Pearson product–moment correlation coefficient on the 2 × 2 table formed by the two median splits—to defend their use if the correlation was insignificant.

Sixth, more generally, given that orthogonality is important to the proper functioning of median splits (compared to the problematic results that arise from median splits coupled with multicollinearity), it may be worthwhile to demonstrate in studies that use median splits that the median split factor is uncorrelated with the experimental factors with which it is crossed. That is, on occasion, a median split on X_1 (to obtain a factor "A") may yield a slight imbalance in the representation of the "low" and "high" respondents across conditions of factor B. To assure the results we have explicated hold, researchers could verify that no unintended multicollinearity arose in the study's design. To do so and lend additional confidence to their results, after creating a median split, researchers might run a correlation between the median split and the experimental factor (each may be coded 0/1, or 1/2, or −1/1, etc.) to verify that the design is statistically approximately orthogonal. Specifically, for a total sample size of say N = 40 (10 observations per cell in a 2 × 2 design), the resulting correlation will not be significantly different from zero if it less than r = 0.31 in magnitude. For a total sample size of N = 100, the correlation will be statistically zero if it is less than 0.19 in size. (By default, statistical packages usually produce *p*-values corresponding to a test of the null hypothesis of zero correlation; the *r* and *p*-value could be reported.) Equivalent statistical tests would include the *z*-test for proportions, i.e., testing whether the proportion of "high" X_1 (versus "low") in factor B level 1 is statistically approximately the same as the proportion of "high" X_1 in factor B level 2, or finally the X^2 test of independence on the 2 × 2 table resulting from cross-classifying the median split X_1 with the experimental factor B (although the X^2 is notoriously hyper-sensitive to power and sample size, the relationship between the correlation and X^2 being: $X^2 = N * r$).

In this paper, we have used the terms "dichotomize" and "median split" essentially interchangeably. Of course there are some times when researchers wish to dichotomize a measure or variable at a meaningful cut-point, such as an established blood sugar level at which patients are more prone to cardiovascular disease (Leite, Huang, & Marcoulides, 2008), or an age at which cognitive functioning begins to decline (Conijn, Emons, van Assen, Pedersen, & Sijtsma, 2013). In these circumstances, while the frequency distributions in the two groups would not likely be 50:50, the researcher can still employ research practices to ensure or heighten the likelihood of the dichotomous variable being uncorrelated with the other predictors, such as by using the dichotomous variable as a blocking factor, and randomly assigning study participants to experimental conditions within each block, that is, within each portion of the dichotomous variable. Of course, it is frequently the case that the variable of interest is a single or multi-item scale of attitudes or propensities, with no inherent cut-point, thus the median split is frequently employed.

Recommendations

We might summarize our findings in a set of recommendations: We are not interested in persuading researchers to use median splits who do not wish to do so. Researchers interested in modeling individual differences would be best served by retaining the continuity of measures of attitudes and preferences (see Table 1). In the unlikely event that the predictors are uncorrelated, a median split might be used, if there is a concern that the regression is behaving overly sensitively to outliers or to nonlinearities per assumptions required by the model. In the more likely event that the predictors are at least somewhat correlated, then median splits should not be used, and the usual concerns regarding multicollinearity apply, e.g., likely suppressed effect sizes and diminished power. Furthermore, such researchers will have to continue to be careful in creating plots to examine interaction terms, and in selecting data points to highlight and report.

Our message is intended for a different audience. Specifically, researchers more interested in examining group differences, such as individuals who score “high” or “low” on attitudes and preference measures. Such researchers interested in group differences may use a median split variable in conjunction with one or more orthogonal experimental factors. Given the concerns regarding median splits, it would probably benefit researchers to establish their effects and reduce counter-arguments to report the correlation coefficient and *p*-value testing the relationship between the median split variable and the experimental factor(s). That is, we suggest the proof be in the pudding and encourage researchers to offer empirical demonstration that the correlation between their experimental factor and their median split variable is zero, by including the supporting statistical test, e.g., H_0: $\rho = 0$. If the correlation is not significant, the researcher may proceed to analyze the median split variable with the experimental factors and their interactions, e.g., via ANOVA, reporting means and contrasts, per usual (see Table 1).

The bottom line is that there are some situations in which median splits are completely acceptable. What many perceive as a blanket prohibition against using median splits is not warranted. Under the conditions we have explicated, a median split is absolutely as good as, and not one iota less appropriate than, a continuous variable. Furthermore, many researchers may wish to use the median split due to its greater parsimony.

Conclusion

Our article’s main contribution is giving the green light to researchers who wish to conduct a median split on one of their continuous measures to be used as one of the factors in an orthogonal experimental design, such as a factorial, and then use ANOVA to model and communicate results. If a study focuses on group differences versus individual heterogeneity, group differences as represented by median split results are often closer to many researchers’ mental models. In addition, an article investigating group differences may be more impactful if median splits are used because communication clarity of the core concepts is facilitated. We suggest proceeding as follows: 1) construct the median split, 2) report the correlation coefficient and *p*-value demonstrating the orthogonality between the median split variable and the experimental factor(s), and 3) proceed to analyze the median split variable and the experimental factors and their interactions, e.g., via ANOVA, reporting the means and contrasts. Researchers who perceive advantages of using median splits may use their preferred analytic strategy knowing that doing so is perfectly legitimate and doing so yields results with statistical integrity.

Table 1
Recommendation for using a median split.
Based on the presence of multicollinearity and research interests.

Research interest	Multicollinearity present?	
	Yes	No
Individual differences	Continuous variable	Continuous variable
Group differences	Continuous variable	Median split

Appendix A

Use this program to derive Figs. 1 and 2:

```
proc iml;
*design parameters;
n={ 50 100 200} ;
rho={ .0 .1 .3 .5 .7 } ;
nreps=10000;
sizedesign=3*5*5*5;                summaries=
j(sizedesign,10,0); q=0;
*generating the X data;
do h=1 to 5; r12=rho[ h] ;
do i=1 to 3; nn=n[ i] ;
do j=1 to 5; ry1=rho[ j] ;
do k=1 to 5; ry2=rho[ k] ;
sigma=i(3);   sigma[ 1,2] =ry1;   sigma[ 2,1] =
sigma[ 1,2] ; *cols are y, x1, x2;
sigma[ 1,3] =ry2;     sigma[ 3,1] =sigma[ 1,3] ;
sigma[ 2,3] =r12; sigma[ 3,2] =sigma[ 2,3] ;
   outreg=j(nreps,6,0);
do nr=1 to nreps;
call vnormal(x,,sigma,nn,);
*———— make median; mdn=j(nn,1,0);
do mm=1 to nn; if x[ mm,2] <= 0 then mdn[ mm,1] =
0; else mdn[ mm,1] =1; end;
x=x||mdn; *cols are y, x1, x2, x1mdn;
*———— regression part of simulation ————;
nroww=nrow(x);    jj=j(nroww);    xbar=jj*x/
nroww; x=x-xbar; covar=x`*x/(nroww-1);
gdiag=diag(covar);    gg=sqrt(inv(gdiag));
corrmatrix=gg*covar*gg;
*standardized betas;
beta1=(corrmatrix[ 1,2]   -   (corrmatrix[ 1,3]
*corrmatrix[ 2,3] )) /
(1-(corrmatrix[ 2,3] **2));
```

```
beta2=(corrmatrix[1,3] - (corrmatrix[1,2]
*corrmatrix[2,3]))/
(1-(corrmatrix[2,3]**2));
beta1mdn=(corrmatrix[1,4] - (corrmatrix
[1,3]*corrmatrix[3,4]))/
(1-(corrmatrix[3,4]**2));
beta2mdn=(corrmatrix[1,3] - (corrmatrix
[1,4]*corrmatrix[3,4]))/
(1-(corrmatrix[3,4]**2));
interc=j(nn,1,1);    regx=interc||x[,2]||x
[,3]; regy=x[,1];
b=inv((regx`)*regx)*((regx`)*regy);
resid=regy-(regx*b);    dff=nn-3;    ssq=
((resid`)*resid)/dff;
betavar=inv((regx`)*regx)*ssq;     betasd=
sqrt(vecdiag(betavar)); t=b/betasd; pval=
1-probf(t#t,1,dff); p1=pval[2]; p2=pval
[3];
outreg[nr,1]=beta1;    outreg[nr,2]=pval
[2]; outreg[nr,3]=beta2; outreg[nr,4]=
pval[3];
regx=interc||x[,4];     b=inv(regx`*regx)
*(regx`)*regy; resid=regy-(regx*b); dff=
nn-3; ssq=((resid`)*resid)/dff;
betavar=inv(regx`*regx)*ssq;        betasd=
sqrt(vecdiag(betavar)); t=b/betasd;
pval=1-probf(t#t,1,dff);    outreg[nr,5]=
beta1mdn; outreg[nr,6]=beta2mdn;
*———;
end;
mbeta1=outreg[+,1]/nreps;  mbeta1p=outreg
[+,2]/nreps;    mbeta2=outreg[+,3]/nreps;
mbeta2p=outreg[+,4]/nreps;
mbeta1mdn=outreg[+,5]/nreps;    mbeta2mdn=
outreg[+,6]/nreps;
q=q+1;
summaries[q,1]=r12;    summaries[q,2]=nn;
summaries[q,3]=ry1;   summaries[q,4]=ry2;
summaries[q,5]=mbeta1;     summaries[q,6]=
mbeta1p; summaries[q,7]=mbeta2; summaries
[q,8]=mbeta2p;
summaries[q,9]=mbeta1mdn; summaries[q,10]
=mbeta2mdn;
end; end; end; end;
create xdat from summaries[colname={"r12"
"nn" "ry1" "ry2" "mbeta1" "mbeta1p"
"mbeta2" "mbeta2p" "mbeta1mdn" "mbeta2mdn"
}];
append from summaries;
quit;
run;
```

Next, we were asked to add the effect on interaction terms (Fig. 3), and that's done by modifying the above program as follows.

Replace previous rho vector with `rho={.0 .1 .3 .5 };`

Change "sizedesign" to (3*4*4*4*4), and replace the 5's on the do loops with 4's.

After the "do k=..." line, add:

```
do int1=1 to 4; ry3=rho[int1];
```

Next, change "sigma=i(3);" to "sigma=i(4);"

After the "sigma[1,3]..." line, add:

```
sigma[1,4]=ry3;      sigma[4,1]=sigma[1,4];
sigma[2,4]=r12; sigma[4,2]=sigma[2,4];
sigma[3,4]=r12;      sigma[4,3]=sigma[3,4];
*could vary but keeping multicoll among x1,
x2,int about same;
```

Replace the regression part of the simulation above with this:

```
interc=j(nn,1,1);  regy=x[,1];  interact=x
[,4];
regx=interc||x[,2]||x[,3]||interact;
b=inv((regx`)*regx)*((regx`)*regy);   b1=b
[2]; b2=b[3]; bint=b[4];
mx1=((interc`)*x[,2])/nn; mx2=((interc`)*x
[,3])/nn;   mint=((interc`)*interact)/nn;
my=((interc`)*regy)/nn;
mx1=interc*mx1;    mx2=interc*mx2;    mint=
interc*mint; my=interc*my;
diff1=(x[,2]-mx1)##2;   diff2=(x[,3]-mx2)
##2;   diffi=(interact-mint)##2;   diffy=
(regy-my)##2;      sdy=((interc`)*diffy)/
(nn-1);  sdy=sqrt(sdy);  sdx1=((interc`)
*diff1)/(nn-1); sdx1=sqrt(sdx1);
sdx2=((interc`)*diff2)/(nn-1);        sdx2=
sqrt(sdx2);      sdint=((interc`)*diffi)/
(nn-1);     sdint=sqrt(sdint);      beta1=
b1*(sdx1/sdy);       beta2=b2*(sdx2/sdy);
betaint=bint*(sdint/sdy);   outreg[nr,1]=
beta1;  outreg[nr,2]=beta2;  outreg[nr,3]=
betaint;    interact=x[,4]#x[,5];    regx=
interc||x[,5]||x[,3]||interact;
b=(inv(regx`*regx))*((regx`)*regy);   b1=b
[2]; b2=b[3]; bint=b[4];
mx1=((interc`)*x[,5])/nn; mx2=((interc`)*x
[,3])/nn;   mint=((interc`)*interact)/nn;
my=((interc`)*regy)/nn;
mx1=interc*mx1;    mx2=interc*mx2;    mint=
interc*mint; my=interc*my;
diff1=(x[,5]-mx1)##2;   diff2=(x[,3]-mx2)
##2;   diffi=(interact-mint)##2;   diffy=
(regy-my)##2;      sdy=((interc`)*diffy)/
(nn-1);  sdy=sqrt(sdy);  sdx1=((interc`)
*diff1)/(nn-1);  sdx1=sqrt(sdx1);  sdx2=
((interc`)*diff2)/(nn-1);             sdx2=
sqrt(sdx2);      sdint=((interc`)*diffi)/
(nn-1); sdint=sqrt(sdint);
```

```
beta1mdn=b1*(sdx1/sdy); beta2mdn=b2*(sdx2/sdy); betaintmdn=bint*(sdint/sdy);
outreg[nr,4]=beta1mdn; outreg[nr,5]=beta2mdn; outreg[nr,6]=betaintmdn;
*———;
```

After the "end;" and before the "q=q+1;" replace the "mbeta1" and "mbeta1mdn" lines with these lines:

```
mbeta1=outreg[+,1]/nreps; mbeta2=outreg[+,2]/nreps; mbetaint=outreg[+,3]/nreps;
mbeta1mdn=outreg[+,4]/nreps; mbeta2mdn=outreg[+,5]/nreps; mbetaintmdn=outreg[+,6]/nreps;
```

Replace the "summaries[q,5]" line and that which follows with these:

```
summaries[q,5]=mbeta1; summaries[q,6]=mbeta2; summaries[q,7]=mbetaint;
summaries[q,8]=mbeta1mdn; summaries[q,9]=mbeta2mdn; summaries[q,10]=mbetaintmdn;
```

Where you see 4 "end;" add one more.

Relabel the "create" line to:

```
create xdat from summaries[colname={"r12" "nn" "ry1" "ry2" "mbeta1" "mbeta2" "mbetaint" "mbeta1mdn" "mbeta2mdn" "mbetaintmdn"}];
```

Lastly, for the anovas, to create Fig. 4:

Begin with the first program (not the second). Right after "proc iml;" statement, define the rank macro:

```
start matrank(xx); matrank=round(trace(xx*ginv(xx)));
return(matrank); finish;
```

Replace "outreg=j(…" with "outanov=j(nreps,3,0);"

Replace the regression part of the simulation with this anova part:

```
y=x[,1]; mdn=mdn[,1]; nn2=nn/2; b1=j(nn2,1,1); b2=j(nn2,1,2); bb=b1//b2;
mdnintb=mdn#bb; xa=design(mdn); xb=design(bb); xab=design(mdnintb); xx=xa||xb||xab;
nroww=nrow(xx); rankx=matrank(xx); ranka=matrank(mdn); rankb=matrank(xb);
ssa=y`*(xa*ginv(xa)-(1/nroww)*J(nroww))*y; ssb=y`*(xb*ginv(xb)-(1/nroww)*J(nroww))*y;
sse=y`*(I(nroww)-xx*ginv(xx))*y; sst=y`*(I(nroww)-(1/nroww)*J(nroww))*y;
ssab=sst-ssa-ssb-sse; ssm=sst-sse;
dfa=1; dfb=1; dfab=1; dfe=nroww-3; dft=nroww-1; msa=ssa/dfa; msb=ssb/dfb; msab=ssab/dfab; mse=sse/dfe; fa=msa/mse; fb=msb/mse; fab=msab/mse; if fa<.0 then fa=.0; if fb<.0 then fb=.0; if fab<.0 then fab=.0;
pvaluea=1-probf(fa,dfa,dfe); pvalueb=1-probf(fb,dfb,dfe); pvalueab=1-probf(fab,dfab,dfe); source={"A", "B", "AB"}; ss=ssa||ssb||ssab; ss=ss`; df=dfa||dfb||dfab; df=df`; ms=msa||msb||msab; ms=ms`; f=fa||fb||fab; f=f`; pvalue=pvaluea||pvalueb||pvalueab; pvalue=pvalue`;
outanov[nr,1]=pvaluea; outanov[nr,2]=pvalueb; outanov[nr,3]=pvalueab;
```

Prior to "q=q+1;" place:

```
pvala=outanov[+,1]/nreps; pvalb=outanov[+,2]/nreps; pvalab=outanov[+,3]/nreps;
```

Then don't forget to put pvala, pvalb, pvalab into summaries (take out the betas), and change the labels in the "create" line.

References

Abelson, R. P. (1995). *Statistics as principled argument.* New York: Psychology Press.

Becherer, R. C., & Richard, L. M. (1978). Self-monitoring as a moderating variable in consumer behavior. *Journal of Consumer Research*, *5*(3), 159–162.

Buchner, U. G., Erbas, B., Stürmer, M., Arnold, M., Wodarz, & Wolstein, J. (2014). Inpatient treatment for pathological gamblers in Germany: Setting, utilization, and structure. *Journal of Gambling Studies*, *2013*, 1–23, http://dx.doi.org/10.1007/s10899-013-9430-5.

Cochran, W. G., & Cox, G. M. (1992). *Experimental designs* (2nd ed.). New York: Wiley.

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

Conijn, J. M., Emons, W. H. M., van Assen, M. A. L. M., Pedersen, S. S., & Sijtsma, K. (2013). Explanatory, multilevel person-fit analysis of response consistency on the Spielberger state-trait anxiety inventory. *Multivariate Behavioral Research*, *48*, 692–718.

Crosby, R., & Charnigo, R. J. (2013). A comparison of condom use perceptions and behaviours between circumcised and intact men attending sexually transmitted disease clinics in the United States. *International Journal of STD & AIDS*, *24*, 175–178.

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

Farewell, V. T., Tom, B. D. M., & Royston, P. (2004). The impact of dichotomization on the efficiency of testing for an interaction effect in exponential family models. *Journal of the American Statistical Association*, *99*(467), 822–831.

Fitzsimons, G. J. (2008). Editorial: Death to dichotomizing. *Journal of Consumer Research*, *35*(1), 5–8.

Goodman, J. K., Broniarczyk, S. M., Griffin, J. G., & McAlister, L. (2013). Help or hinder? When recommendation signage expands consideration sets and heightens decision difficult. *Journal of Consumer Psychology*, *23*(2), 165–174.

Haugtvedt, C., Petty, R. E., Cacioppo, J. T., & Steidley, T. (1988). Personality and ad effectiveness: Exploring the utility of need for cognition. In M. J.

Houston (Ed.), *Advances in Consumer Research*, *Vol. 15*. (pp. 209–212). Provo, UT: Association for Consumer Research.

Hawthorne, G., de Morton, N., & Kent, P. (2013). Back pain and social isolation: Cross-sectional validation of the friendship scale for use in studies on low back pain. *Clinical Journal of Pain*, *29*(3), 245–252.

Hsu, H. S. (2013). Technology timing of IPOs and venture capital incubation. *Journal of Corporate Finance*, *19*, 36–55.

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

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

Kalteier, E., & Posch, P. N. (2013). Sovereign asset values and implications for the credit market. *Review of Financial Economics*, *22*, 53–60.

Kastrati, A., Neumann, F. -J., Schulz, S., Massberg, S., Byrne, R. A., Ferenc, M., et al. (2011). Abciximab and heparin versus bivalirudin for non-ST-elevation myocardial infarction. *The New England Journal of Medicine*, *365*(21), 1980–1989.

Kim, S., Chen, Z., Zhang, Z., Simons-Morton, B. G., & Albert, P. S. (2013). Bayesian hierarchical Poisson regression models: An application to a driving study with kinematic events. *Journal of the American Statistical Association*, *108*(502), 494–503.

Kirk, R. (2012). *Experimental design: Procedures for the behavioral sciences*. Los Angeles: Sage.

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

Leite, W. L., Huang, I., & Marcoulides, G. A. (2008). Item selection for the development of short forms of scales using an ant colony optimization algorithm. *Multivariate Behavioral Research*, *43*, 411–431.

Lemanske, R. F., Mauger, D. T., Sorkness, C. A., Jackson, D. J., Boehmer, S. J., Martinez, F. D., et al. (2010). Step-up therapy for children with uncontrolled asthma receiving inhaled corticosteroids. *The New England Journal of Medicine*, *362*(11), 975–985.

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

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

McCarthy, M. S., Jensvold, M. L. A., & Fouts, D. H. (2013). Use of gesture sequences in captive chimpanzee (*Pan troglodytes*) play. *Animal Cognition*, *16*, 471–481.

Mok, A., & Morris, M. W. (2013). Bicultural self-defense in consumer contexts: Self-protection motives are the basis for contrast versus assimilation to cultural cues. *Journal of Consumer Psychology*, *23*(2), 175–188.

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

Packard, G., & Wooten, D. B. (2013). Compensatory knowledge signaling in consumer word-of-mouth. *Journal of Consumer Psychology*, *23*(4), 434–450.

Robertson, S. J., Burnett, A. F., & Cochrane, J. (2014). Tests examining skill outcomes in sport: A systematic review of measurement properties and feasibility. *Sports Medicine*, *44*, 501–518.

Silvera, D. H., Kardes, F. R., Harvey, N., Cronley, M. L., & Houghton, D. C. (2005). Contextual influences on omission neglect in the fault tree paradigm. *Journal of Consumer Psychology*, *15*, 117–126.

Silverberg, N. D., Lange, R. T., Millis, S. R., Rose, A., Hopp, G., Leach, S., et al. (2013). Post-concussion symptom reporting after multiple mild traumatic brain injuries. *Journal of Neurotrauma*, *30*, 1398–1404.

Srinivasan, N., & Tikoo, S. (1992). Effect of locus of control on information search behavior. In J. F. SherryJr., & B. Sternthal (Eds.), *Advances in Consumer Research*, *Vol. 19*. (pp. 498–504). Provo, UT: Association for Consumer Research.

Wang, Y., Chen, H., Zeng, D., Mauro, C., Duan, N., & Shear, K. (2013). Auxiliary marker-assisted classification in the absence of class identifiers. *Journal of the American Statistical Association*, *108*(502), 553–565.

Wiebel, C. B., Valsecchi, M., & Gegenfurtner (2014). Early differential processing of material images: Evidence from ERP classification. *Journal of Vision*, *14*(7), 1–13.

Yamamoto, D. J., Nelson, A. M., Mandt, B. H., Larson, G. A., Rorabaugh, J. M., Ng, C. M. C., et al. (2013). Rats classified as low or high cocaine locomotor responders: A unique model involving striatal dopamine transporters that predicts cocaine addiction-like behaviors. *Neuroscience and Biobehavioral Reviews*, *36*, 1738–1753.