Available online at www.sciencedirect.com

ScienceDirect

Journal of Consumer Psychology 25, 4 (2015) 666 – 678

JCP JOURNAL OF CONSUMER PSYCHOLOGY

Research Dialogue

A researcher's guide to regression, discretization, and median splits of continuous variables☆

Derek D. Rucker [a,*], Blakeley B. McShane [a], Kristopher J. Preacher [b]

[a] *Marketing Department, Kellogg School of Management, Northwestern University, USA*
[b] *Department of Psychology & Human Development, Vanderbilt University, USA*

Received 24 April 2015; accepted 27 April 2015
Available online 6 May 2015

Abstract

We comment on Iacobucci, Posavac, Kardes, Schneider, and Popovich (2015) by evaluating the practice of discretizing continuous variables. We show that dichotomizing a continuous variable via the median split procedure or otherwise and analyzing the resulting data via ANOVA involves a large number of costs that can be avoided by preserving the continuous nature of the variable and analyzing the data via linear regression. As a consequence, we recommend that regression remain the normative procedure both when the statistical assumptions explored by Iacobucci et al. hold and more generally in research involving continuous variables. We also discuss the advantages of preserving the continuous nature of the variable for graphical presentation and provide practical suggestions for such presentations.

Keywords: Continuous; Discretization; Categorization; Dichotomization; Median split; Regression

Introduction

Iacobucci, Posavac, Kardes, Schneider, and Popovich (IPKSP) have brought the general issue of the discretization (or categorization) of continuous variables—and the specific issue of discretization via the median split procedure—to the forefront of consumer psychology. We appreciate their call for "a more nuanced understanding of the statistical properties of a median split." It is useful and constructive to discuss best practices in research, and we thank the editors of the *Journal of Consumer Psychology* for allowing us to offer our own perspective and contribution.

In this commentary, we provide a "researcher's guide" to regression, discretization, and median splits of continuous variables. Our commentary is targeted at both those who are unfamiliar with the core issues involved in discretization as well as those who wish to better understand them. To preview the perspective proffered in this commentary, it is helpful to recall the following statement from IPKSP:

> 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.

In contrast to IPKSP's near exclusive focus on Type I error, we propose a more holistic and integrative view of the costs and perceived benefits associated with discretization. In particular, we believe that Type II error is of considerable importance and suggest that the relative cost of Type I versus Type II error (which varies by research setting) should be given due consideration. Consequently, we emphasize the use of efficient statistical

☆ The authors thank Ulf Böckenholt for helpful comments on this manuscript.
* Corresponding author at: Marketing Department, Kellogg School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208, USA.
E-mail address: d-rucker@kellogg.northwestern.edu (D.D. Rucker).

http://dx.doi.org/10.1016/j.jcps.2015.04.004
1057-7408

procedures in order to increase statistical power (i.e., decrease Type II error) while keeping Type I error fixed at α, the size of the test (i.e., the maximum probability of a Type I error or the minimum significance level; typically $\alpha = 0.05$). We also discuss additional costs of discretization such as the loss of individual-level variation, reduced predictive performance, and inefficient effect size estimates.

In the remainder of this commentary, we provide an in depth treatment of issues pertaining to the discretization of a continuous variable under the statistical assumptions explored by IPKSP. We explore the issues associated with various statistical approaches to continuous variables and discretization and show that preserving the continuous nature of the variable and analyzing the data via linear regression both is more informative and has greater power. As noted, the rationale for this perspective is based on more general considerations beyond the Type I error issues explored by IPKSP. We aim to make it transparent that discretization is associated with a large number of costs. Thus, we recommend that regression remain the normative procedure both when the statistical assumptions explored by IPKSP hold and more generally in research involving continuous variables. We also discuss the advantages of preserving the continuous nature of the variable for graphical presentation. Finally, we comment briefly on several additional considerations and provide a brief summation.

Analysis strategies in the default case

Statistical assumptions

IPKSP rightly note that it is quite common for a "researcher to manipulate one (or more) factors and measure another [continuous] variable (one but not more)" and thus we make this case the focus of our commentary. We further assume that the sample size per (manipulated) condition is not unreasonably small because (i) one generally does not attempt to account for a measured variable (whether by regression or other means) when one has very few subjects per condition and (ii) reasonable sample sizes ensure the measured variable is independent of the (generally randomized) treatment manipulation provided either the measured variable is assessed prior to the manipulation or it is assessed after the manipulation but is unaffected by it; this assumption does not seem unreasonable as the field is moving in the direction of larger sample sizes (Asendorpf et al., 2013; Brandt et al., 2014; Cumming, 2014; McShane & Böckenholt, 2014, in press; Pashler & Wagenmakers, 2012).

We further assume that the linear model holds. That is, for each manipulated condition, the continuous dependent variable Y is a linear function of the measured variable X.

The assumptions stated thus far are entirely consistent with those of IPKSP. We make the additional assumption that the number of manipulated conditions is two and they are labeled "Treatment" and "Control." As such, the treatment status can be represented by a binary treatment variable T that is zero for subjects in the control condition and one for those in the treated condition. We make this last assumption for illustration purposes only and our comments hold for any natural number of conditions (e.g., four in a 2 × 2 study design).

These assumptions define the "default case" for this commentary and we explicitly note whenever we make comments that depart from this case.

Linear regression

We illustrate linear regression in the default case in Fig. 1. The data underlying the figure comes from a simulated two-condition study with two hundred subjects per condition. The points indicate the raw data, the *x*-axis indicates the measured variable, the *y*-axis indicates the dependent variable, and the color indicates the treatment variable. Finally, the lines indicate a linear regression fit to the data; this can be fit by regressing the dependent variable Y on the treatment variable T, the measured variable X, and their product $T \cdot X$.

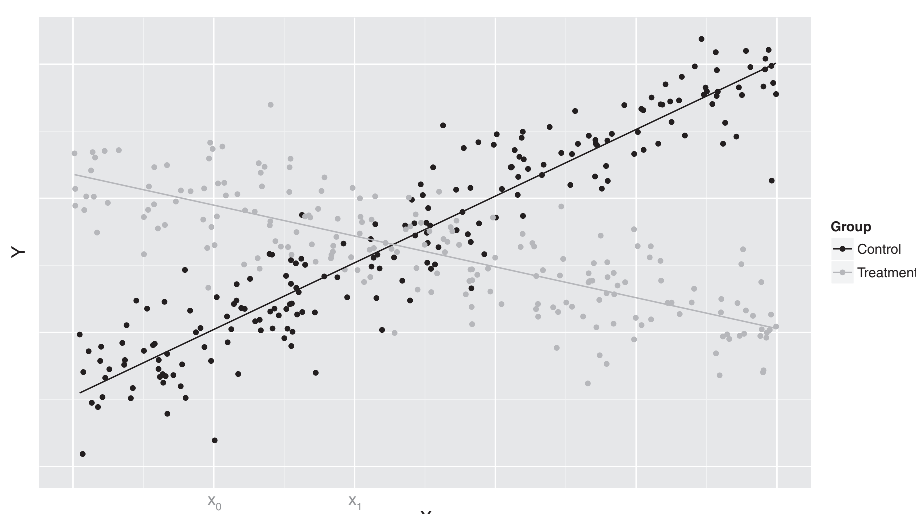

Fig. 1. Linear regression in the default case. The points indicate the raw data, the *x*-axis indicates the measured variable, the *y*-axis indicates the dependent variable, and the color indicates the treatment variable. The lines indicate a linear regression fit to the data.

The linear regression allows researchers to compare the treatment and control groups at various levels of the measured variable X. For instance, at $X = x_0$ the treatment group scores substantially higher than the control group on the dependent variable Y while at $X = x_1$ they score similarly. Regression also allows researchers to obtain standard errors for the differences between the two groups at a given level of X and conduct null hypothesis significance tests of whether the differences are, for example, statistically significantly different from zero.

Linear regression also allows researchers to compute differences in differences, for example the difference between (i) the difference between the treatment and control groups at $X = x_1$ and (ii) the difference between the treatment and control groups at $X = x_0$. This is of course the interaction and it allows researchers to assess the magnitude of the difference in the slopes of the two groups.

Additionally, linear regression allows researchers to compare how the dependent variable Y changes as the measured variable X changes for a given group (i.e., treatment or control). We emphasize caution here because, in contrast to the comparisons illustrated in the prior two paragraphs, this type of comparison is merely associational and not causal (i.e., experimental). This is so because the measured variable was measured rather than manipulated. As a consequence, although it is correct to state that an increase in the measured variable is *associated* with a decrease (increase) in the dependent variable for the treatment (control) group, it is not necessarily correct to state that this change in the dependent variable was *caused* by the increase in the measured variable. This limitation aside, this analysis is informative as it allows researchers to make predictions about the change in Y that is expected for a given change in X for each group. For example, if X is an attitude measure assessed on an 11-point continuous scale and Y is a behavioral measure such as purchase frequency, researchers can use the regression to make a prediction about the purchase frequency of a given subject with an attitude score of 6.5 and compare it to the prediction for another subject with an attitude score of 9.5.

Finally, researchers can engage in comparisons of simple effects via spotlight analysis (Aiken & West, 1991; Irwin & McClelland, 2001) and floodlight analysis (Johnson & Neyman, 1936; Spiller, Fitzsimons, Lynch, & McClelland, 2013) and comparisons of slopes via simple slopes analysis (Aiken & West, 1991). These techniques can be used to, among other things, obtain predicted means for individuals at critical points (e.g., one standard deviation below and above the mean of the measured variable); the predicted means come directly from the regression line and can be used for direct comparison.

Dichotomization and the median split procedure

Having discussed linear regression in the default case, we now turn to dichotomization. Dichotomization involves splitting the measured variable X at some fixed value to form two categories that can be described as "Low" and "High." One popular choice for the split point is the sample median, and dichotomization at the sample median is called a median split. Another choice for the split point is the midpoint of the measurement scale of the measured variable; in this case, we suggest calling the procedure a midpoint split to distinguish it from a median split. Other split points are also possible.

We refer to discrete groupings based on manipulated variables as conditions or groups and those based on the discretization of measured variables as categories. We introduce this terminology in order to keep conceptually distinct (i) treatment conditions based on manipulated variables and (ii) categories created from measured variables. This terminology has the benefit of emphasizing the distinction between causal comparisons (between different levels of manipulated variables) and associational comparisons (between different levels of measured variables).

Given a dichotomization via the median split procedure or otherwise, we have four groupings: (Low, High) × (Treatment, Control). The typical practice is to analyze these groupings via ANOVA; indeed, following IPKSP we will assume that the data will be analyzed via linear regression when and only when the continuous nature of the variable is preserved and that it will be analyzed via ANOVA when and only when the variable is dichotomized.

Using the data plotted in Fig. 1, we illustrate the typical presentation of a median split in Fig. 2. The median split allows researchers to compare how Y varies across the treatment and control groups given a fixed value of the new binary X variable. For example, the treatment group scores higher on the dependent variable than the control group in the low category while the treatment group scores lower in the high category. The median split also allows researchers to consider the interaction via a difference in differences. Finally, the median split allows researchers to compare how Y varies across the two categories implied by the binary X given a fixed value of the treatment variable; as above, we caution that this comparison is merely associational and not causal. In sum, the median split procedure appears to allow researchers to make comparisons that are similar to those they would make by preserving the continuous nature of the variable (e.g., simple comparisons, interactions). As such, the primary difference between the two approaches appears to be in the modeling choice (i.e., linear regression versus ANOVA).

A much more complete and informative illustration of the median split is provided in Fig. 3. The points indicate the raw data, the x-axis indicates the measured variable, the y-axis indicates the dependent variable, the color indicates the treatment variable, and the lines indicate the fit implied by the median split procedure. Although Fig. 3 provides the same fit as Fig. 2, it makes clear that the median split procedure assumes that, for each manipulated condition, the dependent variable Y is modeled as a step function with a single step at the sample median of the measured variable X and that the predicted value below (above) the sample median is given by the sample mean of the dependent variable Y among the points with measured variable X below (above) the sample median. It also makes clear that this yields a poor fit to the data in the default case of linearity. For example, consider the control group data; for low and moderately high (moderately low and high) values of the measured variable, the fit implied by the median split procedure is too high (low).

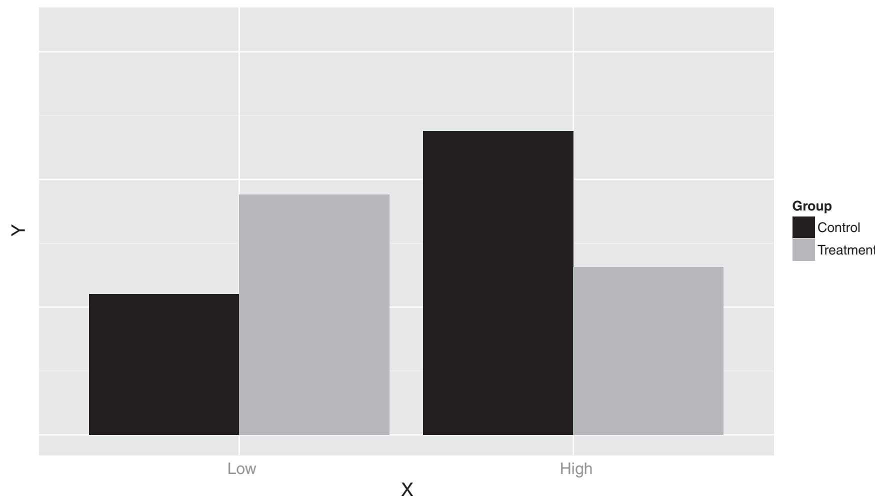

Fig. 2. Median split in the default case. The bars indicate the mean of the dependent variable for each of the four groupings: (Low, High) × (Treatment, Control).

Costs and perceived benefits of dichotomization

Costs

Loss of individual-level variation

Dichotomization into low and high categories via the median split procedure or otherwise discards the potentially rich variation in individual scores. For example, subjects with measured *X* just above the median, moderately above the median, and substantially above the median are all treated as identical by the median split procedure as illustrated in Figs. 2 and 3 (i.e., they are all classified in the high category). However, when the linear model holds, treating these subjects identically is simply incorrect as the raw data plotted in Figs. 1 and 3 show: subjects just above the median are clearly not identical to subjects moderately or substantially above the median in terms of the dependent variable when the slope is nonzero. Consequently, the variation in the treatment effect from substantially positive to roughly zero to substantially negative observed across the range of the measured variable *X* in Fig. 1 is collapsed into a simple positive (negative) effect in the low (high) category in Figs. 2 and 3. This is a cost inherent to dichotomization whether via the median split procedure or otherwise.

Reduced predictive performance

A cost of dichotomization related to the loss of individual-level variation is diminished precision in the predictions researchers can make as the measured variable *X* varies and the concomitant reduction in the variance of the dependent variable explained. A researcher using linear regression with a continuous variable can make predictions about the dependent variable *Y* along the continuum of the measured variable *X*. In contrast, a researcher using dichotomization can make predictions only about how the low and high categories differ ignoring distinctions

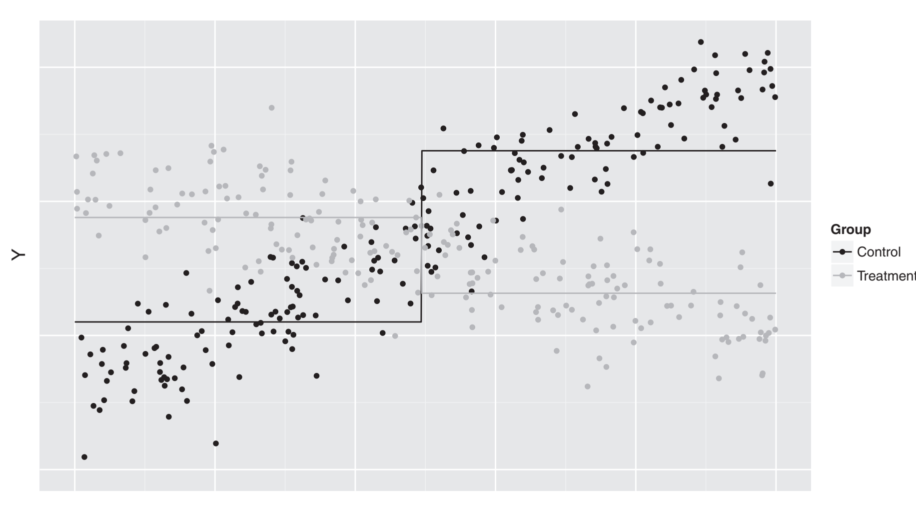

Fig. 3. Median split in the default case, scatterplot presentation. The points indicate the raw data, the *x*-axis indicates the measured variable, the *y*-axis indicates the dependent variable, and the color indicates the treatment variable. The lines indicate the fit implied by the median split procedure.

based on the relative score within each category. Although these less precise associational predictions may not be entirely useless, they are clearly dominated by those provided by the regression.

To illustrate, assume that X is an attitude measure assessed on an 11-point continuous scale, that the median in the sample is 5.5, that Y is a behavioral measure such as purchase frequency, and that a one unit increase in the attitude score X is associated with a one-half unit increase in purchase frequency Y for, say, the control group. Imagine comparing the predicted purchase frequency for subjects in the control group given an observed attitude score of 6.5 versus an observed attitude score of 9.5. Regression makes it clear that we should expect the purchase frequency to increase by one-and-a-half. In other words, regression predicts differently for $X = 6.5$ than for $X = 9.5$. In contrast, the median split considers these two subjects identical: both are lumped into the high category. A prediction can still be made but it will be the same for both values of X.

Increased Type I error

IPKSP note that the median split procedure can, in some cases, increase Type I error focusing on two cases in particular:

1. Replicating prior research (MacCallum, Zhang, Preacher, & Rucker, 2002; Maxwell & Delaney, 1993), they note that when two continuous variables are both dichotomized, this can create spurious effects particularly when the two variables are collinear.
2. In their first simulation, they note that even when only a single continuous variable is dichotomized, it can create spurious effects in a second continuous variable (that is left continuous) provided the two variables are collinear.

These facts are noteworthy and we concur with IPKSP about the danger of the median split procedure in these settings. We note that both of these cases fall outside the default case as they require two continuous variables and collinearity.

IPKSP also provide a more nuanced understanding of the potential risk of increased Type I error resulting from dichotomization in the default case of (i) one or more manipulated variables, (ii) a single measured variable, and (iii) a linear relationship between the dependent variable and the measured variable for each manipulated condition. In particular, they demonstrate via simulation that, at least in the cases simulated and reported, the median split procedure applied to the single continuous predictor did not result in increased Type I error. Based on this, IPKSP conclude “a median split is absolutely as good as, and not one iota less appropriate than, a continuous variable” in the default case. This conclusion fails to consider (i) that costs other than increased Type I error are important and (ii) that increased Type I error is a genuine risk in many settings because, for example, researchers often possess multiple measured variables that are collinear at least to some degree.

Increased Type II error

Many researchers do—and we argue all should—seek not just to avoid claiming that effects exist when they do not (i.e., Type I error) but also to avoid claiming that effects do not exist when they do (i.e., Type II error). To these researchers, we emphasize that dichotomization via the median split procedure or otherwise increases Type II error (i.e., reduces power; Cohen, 1983; Irwin & McClelland, 2003; MacCallum et al., 2002; Maxwell & Delaney, 1993) and this holds both in the default case and otherwise. For example, consider a researcher who conducts a study that is adequately powered (i.e., at 80%; Cohen, 1992) to detect a correlation between two continuous variables. If the researcher chooses to employ the median split procedure on one of the variables and analyze the data via ANOVA rather than to preserve the continuous nature of both variables and analyze the data via linear regression, power drops from 80% to about 60% (Cohen, 1983): the median split procedure turns an adequately powered study into one that is little better than a coin toss. Put differently, a researcher who runs studies like this and analyzes the data using linear regression will detect 33% (i.e., (80%–60%)/60%) more true effects than one who employs the median split procedure and analyzes the data via ANOVA. A similar loss of power from 80% to just under 60% occurs when a researcher conducts a study that consists of a manipulated variable with two conditions and a continuous variable and that is adequately powered to detect the difference between the slopes in the two conditions (i.e., the interaction). Thus, researchers interested in interactions will detect many more true effects when they preserve the continuous nature of their variables and analyze their data via linear regression.

As a remedy for the decreased power associated with dichotomization, IPKSP glibly suggest that “researchers can simply draw large enough sample sizes to offset any reduction in power.” We find this conclusion unsatisfying on at least two dimensions. First, it fails to consider the real costs (monetary and otherwise) associated with larger sample sizes. Second, the better and more proper comparison is the one discussed in the prior paragraph, namely that between linear regression and dichotomization holding the sample size fixed; it is clear that regression offers greater power for a fixed sample size.

That dichotomization can reduce power (i.e., increase Type II error) is a clear problem that undermines IPKSP’s conclusion that “a median split is absolutely as good as, and not one iota less appropriate than, a continuous variable.” Indeed, the consequences of this loss of power are manifold. For example, a researcher who dichotomizes may abandon the pursuit of a research paradigm because evidence for an effect is lacking in a given dataset. Similarly, when testing competing hypotheses, authors and reviewers may ask whether sufficient power exists to test whether an effect is moderated by an alternative variable. To the extent that a researcher dichotomizes this variable and the presence of an effect would have provided an alternative explanation for the data, the loss of power undermines theory testing; put differently, when researchers are hoping for the absence of an effect, it is self-serving to use statistical procedures with lower power.

In short, Type I and Type II error are both important and it is their relative cost in a particular research setting that requires consideration. In some settings, researchers may be relatively more tolerant of Type I error while in others they may be relatively more tolerant of Type II error. For example, consider various legal standards for the burden of proof under the presumption of innocence (i.e., the null hypothesis is that one is "innocent until proven guilty"); the standard of "proof beyond a reasonable doubt" that applies in criminal proceedings is relatively less tolerant of Type I error (i.e., convicting the innocent) and thus more tolerant of Type II error (i.e., letting the guilty go free) as compared to the standard of the "preponderance of the evidence" that applies in most civil proceedings. Similarly, a terminal patient may be willing to take a potentially unsafe experimental drug while a pharmaceutical company fearing legal risk may be unwilling to offer it. Regardless, preserving the continuous nature of the variable and analyzing the data via linear regression results in Type I error that is no greater and Type II error that is less compared to dichotomizing the variable and analyzing the data via ANOVA.

Inefficient effect size estimates

Though the null hypothesis significance testing paradigm is the dominant statistical paradigm in academic training and reporting in the biomedical and social sciences (see, for example, Gigerenzer (1987), Gill (1999), Morrison and Henkel (1970), Sawyer and Peter (1983)), this paradigm has received no small degree of criticism over the decades (see, for example, Cohen (1994), Gigerenzer (2004), McShane and Gal (in press), Meehl (1978), Rosnow and Rosenthal (1989), Rozenboom (1960)) and many have argued for a greater focus on effect sizes, their variability, and the uncertainty in estimates of them (see, for example, Cohen (1990), Fidler, Thomason, Cumming, Finch, and Leeman (2004), Gelman (2015), Iacobucci (2005), Kelley and Preacher (2012)). Without entering into this debate here, we note that preserving the continuous nature of the variable and analyzing the data via linear regression is superior to dichotomizing the variable and analyzing the data via ANOVA on both dimensions: the former performs as well or better on Type I and Type II error as discussed above and yields more efficient estimates of effect sizes (Cox, 1957; Gelman & Park, 2009; Lagakos, 1988; Morgan & Elashoff, 1986). Researchers interested in effect sizes will naturally favor linear regression as its primary purpose is to facilitate the estimation and interpretation of effect sizes.

Perceived benefits

Perceived simplicity

IPKSP echo the argument of DeCoster, Iselin, and Gallucci (2009, p. 350) that dichotomization via the median split procedure or otherwise "makes analyses easier to conduct and interpret" particularly via ANOVA. However, we question whether this really is the case, and, if so, easier compared to what? *Prima facie*, linear regression seems easier to conduct and interpret relative to dichotomization given that it both respects the continuous nature of the measured variable X and does not require the specification of a split point. At the very least, with contemporary statistical software linear regression seems no more difficult than dichotomization; indeed, both seem rather easy.

Related to the argument for simplicity is IPKSP's claim that a "median split may be preferred as more parsimonious" than alternative techniques such as linear regression. We fail to see how this could be the case. A linear regression requires the estimation of two parameters for each experimental group (i.e., an intercept and a slope). Although dichotomization also requires the estimation of two parameters for each experimental group (i.e., the mean of the dependent variable in the low category and in the high category), it also requires the estimation or specification of the split point (e.g., the median). Thus, dichotomization requires at least one additional parameter. As linear regression offers superior predictive performance and requires fewer parameters, Occam's razor (also known as the law of parsimony) favors it over dichotomization. Further, although the median split procedure is obviously not statistically efficient relative to linear regression, perhaps surprisingly it is also not even statistically efficient relative to alternative dichotomization procedures (see the section that follows entitled Dichotomization for the general audience).

Preference

Some researchers may opt for dichotomization because of a preference for ANOVA over linear regression. We speculate that, given the myriad costs of dichotomization, this preference results primarily from a lack of familiarity with plotting continuous variables and analyzing them via linear regression. We believe that lack of familiarity with a technique is generally not a reason to abandon it—particularly when it is more informative, is statistically superior, and is easy both conceptually and in terms of implementation via software. Thus, we strongly urge researchers who find themselves opting for dichotomization solely because of a preference for ANOVA to learn more about and become comfortable with regression instead of shoehorning their data into an ANOVA. Greater familiarity with regression will allow such researchers to provide a more informative analysis with greater statistical power.

Focus on group (category) differences

IPKSP suggest that when researchers are studying group differences (or category differences in the terminology of this manuscript) and collinearity is not present, median splits are appropriate (IPKSP Table 1). However, MacCallum et al. (2002) note that researchers who wish to treat a variable as if it truly consists of taxa (i.e., distinct groups or categories) should provide evidence that the variable does indeed consist of taxa rather than simply assuming it does so, for example, because they deem it convenient. Indeed, desiring that a variable consists of taxa does not make it so and treating it as such can be misleading to readers, discard meaningful data, and reduce power. Whether a measured variable consists of taxa is also independent of whether the measured variable is easy or ethical to manipulate.

In addition, even when researchers are truly working with taxa, the median is not necessarily the ideal split point and indeed could be far from it. For example, if the true distribution of the taxa is 70/30, splitting at the median mischaracterizes the categories and will likely impair efforts to demonstrate distinct consequences. Alternatively, if the variable consists of more than two distinct taxa, dichotomization clearly mischaracterizes the categories. Further, settings exist where regression can be a superior analytic approach compared to discretization even when a variable consists of taxa (e.g., when the measured variable is measured with error). We will not entertain these issues in greater detail here, but the core points are that (i) neither the desire to study distinct categories nor the potential convenience of them implies that a construct truly consists of taxa, (ii) even when the variable consists of taxa the median may not be the appropriate split point and two may not be the appropriate number of categories, and (iii) even when the variable consists of taxa linear regression may still be a superior approach. Consequently, researchers desiring to work with taxa should perform due diligence to show the variable truly consists of taxa and to identify them.

In addition to performing taxometric analyses to support the argument that one is working with a variable that consists of taxa, a simple "gut check" may be useful for researchers wrestling with whether their construct is best represented as an individual difference or a category difference. For example, consider an attitude measure assessed on an 11-point continuous scale where the median in the sample is 5.5. Further, suppose the goal is to predict a behavioral measure such as purchase frequency. If one were considering performing a median split to identify the taxa, does one believe that a subject with an attitude score of 6.5 will purchase as frequently as a subject with an attitude score of 9.5? Does one believe that the difference in purchase frequency for a subject with an attitude score of 4.5 and a subject with an attitude score of 6.5 will be identical to the difference in purchase frequency for a subject with an attitude score of 1 and a subject with an attitude score of 11? Does one believe the purchase frequency of a subject with an attitude score of 4.5 is more like one with an attitude score of 1 than one with an attitude score of 6.5? If a researcher is uncomfortable answering in the affirmative to all of these questions, this suggests that category difference is not central to their research.

Finally, for researchers who opt to dichotomize a variable because they have a preference for discussing low and high categories, we note that such discussion is possible when the continuous nature of the variable is preserved and the data is analyzed via linear regression. This is possible as discussed previously via, for example, spotlight analysis, floodlight analysis, and simple slopes analysis. Thus, a preference for discussing low and high categories does not require dichotomization.

Recommendation

Given the various costs of dichotomization as well as the dubiousness of the various perceived benefits discussed in the prior two subsections, we recommend that preserving the continuous nature of the variable and analyzing the data via linear regression remain the normative procedure both when the statistical assumptions explored by IPKSP hold and, as discussed below, more generally in research involving continuous variables. We elaborate on this point in our concluding section.

Graphical presentation of continuous data

To this point, we have compared the costs associated with dichotomizing a measured variable and analyzing the data via ANOVA as opposed to preserving its continuous nature and analyzing the data via linear regression. Here, we consider the graphical presentation of data. Graphical communication is paramount in facilitating an understanding and appreciation of data (Cleveland, 1993, 1994; Robbins, 2013; Tufte, 2001; Tukey, 1977). Researchers have argued for increased transparency in reporting of measures, conditions, and exclusion criteria (Simmons, Nelson, & Simonsohn, 2011), and we believe informative graphs help with data transparency.

We consider three distinct graphical approaches to displaying data in the default case: a scatterplot with regression lines superimposed as depicted in Fig. 1, a median split plot as depicted in Fig. 2, and a simple slopes plot (i.e., plotting the regression line for each condition as the measured variable X varies from, for example, one standard deviation below its mean to one standard deviation above its mean) as depicted in Fig. 4. Comparison of these plots yields several insights.

The median split plot and the simple slopes plot do not seem particularly different. As alluded to by IPKSP, both involve critical points that are selected somewhat arbitrarily (i.e., above and below the median for the median split plot; one standard deviation above and below the mean for the simple slopes plot). Indeed, we view these plots as more similar than different, and both seem relatively easy to discuss with a peer audience. That said, the simple slopes plot both preserves and presents the continuous nature of the measured variable, which is to some degree superior.

A much more complete presentation is provided by the scatterplot. As in the simple slopes plot, the regression line for each group is presented. However, whereas the simple slopes plot involves presenting the line over a somewhat arbitrary range, the scatterplot necessarily presents it over the range of the measured variable X thereby allowing the easy observation of the difference between the treatment and control group at each level of the measured variable X as well as the slope for each group. Further, the scatterplot presents the individual datapoints thereby depicting the variability of the data. This allows one to observe minima and maxima, the degree to which points deviate from the regression line, potential outliers, and other characteristics.

Perhaps most important, the scatterplot allows one to easily assess whether the assumption of linearity (or any other functional form) is reasonable. For instance, it is clear from Fig. 1 that linearity is a reasonable assumption for the data. On the other hand, Fig. 3 demonstrates that the functional form assumed by the median split procedure yields a poor fit.

In sum, the scatterplot is more informative than either the median split barplot or the simple slopes lineplot. Further, should one desire, it is possible to overlay on it markers of deviation (e.g., the points representing one standard deviation below and above the mean present in the simple slopes plot can be easily added to the scatterplot). Importantly, this graph is possible only by preserving the continuous nature of the measured variable. We suggest that researchers make greater use of scatterplots in reporting their data.

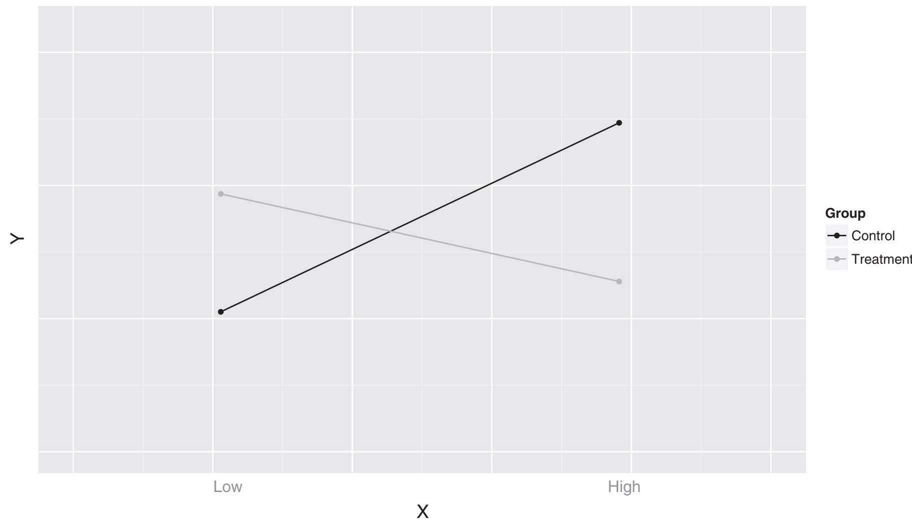

Fig. 4. Simple slopes plot in the default case. The *x*-axis indicates the measured variable, the *y*-axis indicates the dependent variable, and the color indicates the treatment variable. The lines indicate a linear regression fit to the data and the points indicate one standard deviation below and above the mean of the measured variable respectively. The points are selected somewhat arbitrarily and other values (e.g., two standard deviations below and above the mean of the measured variable, the minimum and maximum of the measured variable) are possible.

Dichotomization for the general audience

Statistical analysis via linear regression and graphical presentation via scatterplots allow researchers to quickly communicate a rich set of information to their peers who are or should be well-versed in regression. Of course, researchers do not speak only to fellow experts. For example, both readers of and writers for the popular press may find regressions or scatterplots difficult to understand. Thus, while we challenge the perceived simplicity or convenience of dichotomization when communicating to a peer audience, we believe that supplementing regression analyses with analyses or plots that make use of dichotomization may sometimes be valuable when speaking to a general audience. For example, a general audience may have an easier time understanding the notion of low and high categories as opposed to the intercept and slope of a regression line or predictions made on the basis of them.

That said, we emphasize that, even for researchers interested in communicating to general audiences, the median split is a statistically rather poor way of choosing low and high categories. Gelman and Park (2009) suggest that, when speaking to a general audience, it would be better to split the data into three categories such that the low and high categories consist of 25% or 33% of the data each and to discard the middle category (this recalls the "27 Percent Rule" in psychology (Cureton, 1957; D'Agostino & Cureton, 1975; Jensen, 1928; Kelley, 1939; Ross & Weitzman, 1964)). Gelman and Park (2009) summarize the statistical rationale behind this approach as follows:

> We make the general recommendation that the high and low categories each be set to contain 1/4 to 1/3 of the data, which results in comparisons with approximately 80%–90% of the efficiency of linear regression if the predictor X follows a uniform or normal distribution. A loss of 10%–20% of efficiency is not minor, and so we do not recommend that the comparisons replace regressions but rather that they be considered as useful supplementary summaries, especially for the goal of communicating to a general audience… discretizing X into three categories claws back about half the efficiency lost by dichotomizing the predictor [at the median], while retaining the simple interpretation as a high versus low comparison.

In sum, their three-category procedure is just as interpretable as the two-category median split procedure (because the middle category is discarded) and is substantially more efficient from a statistical perspective. In addition, by setting the low and high categories to each contain 25% or 33% of the data as opposed to 50% of the data as in the median split procedure, one gains additional confidence that those in the low and high categories are in fact relatively distinct from one another. This addresses, in part, the concern about loss of individual-level variation and reduced predictive performance discussed earlier.

Despite the benefit of their three-category procedure relative to the median split procedure, we emphasize Gelman and Park (2009)'s comment that, from a statistical perspective, the three-category approach is not as efficient as linear regression: "a loss of 10%–20% of efficiency is not minor" (see also Preacher, Rucker, MacCallum, and Nicewander (2005)). Consequently, we reiterate that statistical analysis via linear regression and graphical presentation via scatterplots allows researchers to quickly communicate a rich set of information to their peers in, for example, peer-reviewed journals and that the occasional use of dichotomization for a general audience need not and ought not preclude that. Thus, we suggest that researchers use regression and scatterplots in peer-reviewed journals and that they consider the procedure of Gelman and Park (2009) instead of the median split procedure when communicating to less sophisticated audiences such as the popular press.

Additional considerations

Nonlinear effects

Sometimes the relationship between the measured continuous variable *X* and the dependent variable *Y* is not linear. Scatterplots like those of Fig. 1 are extremely helpful here because they can alert researchers to this issue—yet another important rationale for presenting them. When faced with a nonlinear relationship, there are several analytic strategies available to researchers. As in the linear case, these can be divided into two classes: strategies that preserve the continuous nature of the measured variable and strategies that involve discretization.

One approach to nonlinear relationships that preserves the continuous nature of the measured variable is to directly model a particular functional form (e.g., exponential, logarithmic, trigonometric) that is rooted in theory. For example, Fechner's law in psychophysics states that the subjective perception of a stimulus is proportional to the logarithm of the intensity of the stimulus, thereby suggesting the logarithmic model $Y = \beta_0 \log(\beta_1 X)$ rather than the linear model $Y = \beta_0 + \beta_1 X$. When theory supports a particular functional form, this is the best approach to modeling nonlinear relationships.

Unfortunately, such theory is often lacking, thereby necessitating more empirical or statistical approaches. One such approach that also preserves the continuous nature of the measured variable is to employ transformations (e.g., logarithm, square root) of the measured variable, the dependent variable, or both that result in a linear relationship between the transformed variables. This is similar to the functional form approach discussed in the prior paragraph except that the transformation is selected based on the data rather than based on theory. Two popular methods for choosing transformations are the Box–Cox transformation (Box & Cox, 1964) and the Bulging Rule (Mosteller & Tukey, 1977). Transformations are particularly effective when the nonlinear relationship is monotone.

Yet another approach that preserves the continuous nature of the measured variable is to fit a nonlinear function directly. One popular version of this approach is to fit a polynomial (i.e., add terms $X^2, X^3, \ldots, X^d$ to the regression). We caution against this as several superior alternatives such as splines, wavelets, kernel smoothers, and local regression exist. Though these techniques vary in their implementation, they all attempt to fit a function that best fits the data. For more details, we refer the interested reader to texts such as Bates and Watts (1988), Hastie, Tibshirani, and Friedman (2009), James, Witten, Hastie, and Tibshirani (2013), Lindsey (2001), McCullagh and Nelder (1989), and Seber and Wild (2003).

Discretization represents another approach to nonlinear relationships. When the relationship is nonlinear, we note that dichotomization via the median split procedure or otherwise is simply not appropriate as dichotomization necessarily conceals nonlinear relationships: more than two categories are necessary to reflect a nonlinear relationship while two suffice for a linear one. Instead, discretization in this setting requires splitting the continuous variable into a relatively large number of categories (e.g., four, five, ten, or more); split points are typically chosen using either (i) the quantiles of the observed data as split points so that an equal number of subjects falls into each category or (ii) equally-spaced values as split points so that each category is of equal width (see, for example, Gal & McShane, 2012). The mean of the observations in each category is then used for prediction. An alternative approach is to choose the splits points in a more strategic sense that reflects both the measured variable as well as the dependent variable so as to provide a better fit to the data; popular versions of this include tree regression (Breiman, Friedman, Olshen, & Stone, 1984; Hastie et al., 2009; James et al., 2013) and the method of O'Brien (2004).

Ordinal data

When both the measured variable and the dependent variable are truly or approximately continuous (e.g., assessed on a 100-point slider scale or a composite of many items assessed on an 11-point scale), we recommend analysis via regression and presentation via scatterplots. However, when one or more is ordinal (e.g., assessed on a single 11-point scale), some additional considerations are in order. In terms of graphical presentation, we suggest a jittered scatterplot (Gelman & Hill, 2006) in place of a standard one; jittering prevents datapoints with identical values of the discrete variables from being plotted on top of one another (a phenomenon known as overplotting). In terms of analysis, if only the measured variable is ordinal, researchers can generally use regression as if it were continuous; scatterplots will alert them to departures from linearity that require more sophisticated modeling of the ordinal measured variable (e.g., nonlinear modeling approaches, discretizing in a manner that respects the ordinal nature of the data, treating the data as nominal). When the dependent variable is ordinal, however, researchers may want to consider ordinal regression techniques (Gelman & Hill, 2006; Lindsey, 2001; McCullagh & Nelder, 1989) that account for the ordinal nature of the data; a similar consideration applies for binary dependent variables.

Measurement error

When the measured variable is measured with error, standard regression models are biased. For example, when there is a single measured variable, as in the default case, the bias is towards zero and is called attenuation bias. Thus, we suggest that researchers consider errors-in-variables regression models (Buonaccorsi, 2010) that account for this.

We also note that some of the costs associated with discretization discussed above can be exacerbated in this setting. For example, when researchers discretize their data, subjects just below and just above a split point are placed into separate categories. However, when there is measurement error, subjects assessed near the split point seem about as likely to truly belong to the lower category as to the higher category; this applies even if a categorical representation of the data is correct and the split point accurately distinguishes between the categories. In this setting, the approach of Gelman and Park (2009) of splitting the data into three groups and discarding the middle group seems even more advantageous for researchers considering dichotomization.

Assessing measured variables

IPKSP note that the dichotomization of a single continuous variable can create spurious effects when collinearity exists. Thus, we appreciate their recommendation that researchers interested in employing the median split procedure on a single continuous measured variable first verify that it is uncorrelated with the manipulated treatment variable(s).

While we reiterate the advantages of regression regardless of the correlation between the measured variable and manipulated variable(s), we note that this correlation is often likely to be small in practice. In fact, it should only be large either when the sample size is quite small (i.e., because with small samples the measured variable might be unbalanced across the conditions) or when the measured variable is assessed after the manipulation but is affected by it. The remedy for the former situation is rather straightforward: use reasonable sample sizes. In the latter situation, researchers interested in understanding how the measured variable moderates the treatment effect face a dilemma and generally have to turn to alternative experimental designs; for example, they might consider manipulating the construct associated with the measured variable or measuring it either long before or after the manipulation.

A related concern is that when the measured variable is assessed before the dependent variable is assessed, the measurement of the measured variable could affect the measurement of the dependent variable. Similarly, when the measured variable is assessed after the dependent variable is assessed, the measurement of the measured variable could be affected by the measurement of the dependent variable. As above, alternative experimental designs are generally required to address these concerns.

One solution to these issues involving measurement is to consider only measured variables that (i) are assessed both after the manipulation and after the dependent variable is assessed and (ii) are relatively “objective” and therefore are unlikely to be measured with error or to be affected by the manipulation or the measurement of the dependent variable. Examples of such measured variables might include age, sex, and years of education. Unfortunately, theoretical considerations constrain the choice of measured variables and it may not be possible to meet these two criteria in some research paradigms.

Multiple measured variables

IPKSP replicate prior research and show that the dichotomization of multiple continuous variables can create spurious effects. We thus appreciate their recommendation against the dichotomization of multiple continuous variables.

We add to their discussion of multiple continuous variables by noting that regression offers many strategies for quantifying and dealing with collinearity (e.g., variance inflation factors, penalized regression, principal components regression) that can be used if necessary. Simply ignoring collinearity is one such strategy that is often quite reasonable in practice.

We also return to a point made earlier, namely that comparisons between different levels of a manipulated variable are causal while comparisons between different levels of a measured variable are merely associational. Such comparisons can be more difficult to properly assess when there are multiple measured variables—particularly when they are collinear. Consequently, we reiterate that comparisons between different levels of a measured variable should be interpreted cautiously, and that greater caution is necessary when there are multiple measured variables particularly when they are collinear.

Finally, we note that multiple measured variables are common in research: quite seldom do researchers possess only a single measured variable in addition to the manipulated variable(s) even if they report or focus only on a single measure. The costs of dichotomization discussed above thus apply to each of these variables. These facts should be borne in mind when evaluating the default case of IPKSP.

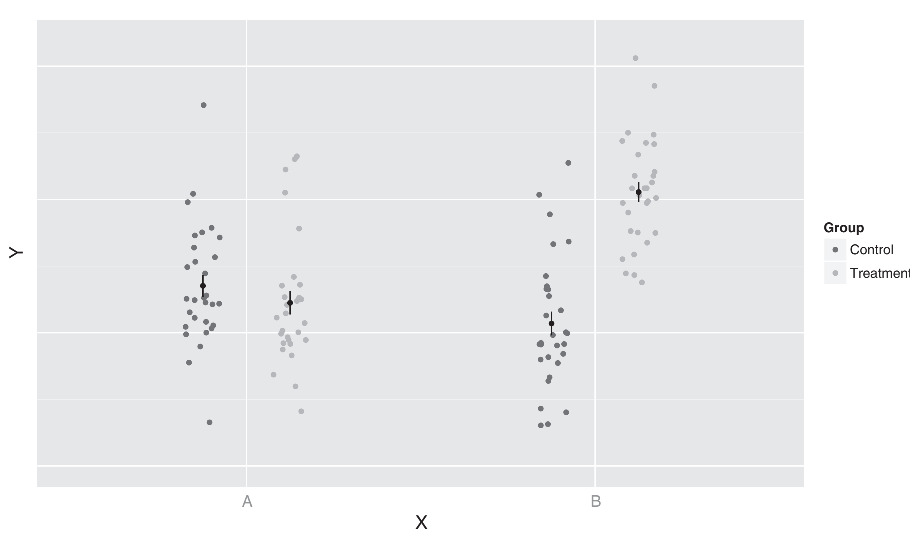

Fig. 5. Dotplot for data with discrete groupings. The points indicate the raw data, the *x*-axis indicates the discrete measured variable, the *y*-axis indicates the dependent variable, and the color indicates the treatment variable. The points are jittered horizontally to prevent overplotting. The black points indicate the mean in each grouping and the black lines indicate plus and minus one standard error.

Graphical presentation of discrete groupings

We note for researchers working with truly discrete groupings—whether these groupings have arisen because the researchers are working with manipulated experimental conditions or because the researchers are working with a combination of manipulated experimental conditions and variables that consist of taxa—far more informative methods of graphical presentation than the standard barplot of Fig. 2 are available. For example, consider the dotplot (of different data) presented in Fig. 5. The points indicate the raw data, the *x*-axis indicates the discrete measured variable (e.g., here consisting of two taxa), the *y*-axis indicates the dependent variable, and the color indicates the treatment variable. Finally, the black points indicate the mean in each grouping and the black lines indicate plus and minus one standard error. This dotplot provides a much fuller presentation of the data and allows for the easy observation of the pairwise differences between each of the four groupings. The dotplot also depicts important features of the data such as minima, maxima, the degree to which points deviate from the mean, potential outliers, and other characteristics. We recommend that researchers use such plots in place of the much less informative barplot.

Summation and best practices

IPKSP have brought the important topic of the discretization of continuous variables to the forefront of consumer psychology. IPKSP focus primarily on Type I error when there is one or more manipulated treatment (or experimental) variables, a single measured continuous variable, and a linear relationship between the dependent variable and the measured variable for each manipulated condition. In this setting, they suggest that concerns about increased Type I error are misguided when considering dichotomization in the absence of collinearity, and thus they "giv[e] the green light to researchers who wish to conduct a median split."

Were researchers solely concerned with avoiding Type I error, IPKSP would provide some degree of comfort. However, even in the default case, dichotomization is riddled with numerous costs including loss of individual-level variation, reduced predictive performance, increased Type II error, and inefficient effect size estimates. Outside the default case, not only are these costs often present but also increased Type I error can be a concern (e.g., when there are multiple dichotomized variables, when there are multiple collinear variables, or when the relationship between the dependent variable and the continuous variable is nonlinear). Indeed, there is a large literature summarized in Table 1 that notes these costs. Limiting ourselves to quoting from only the *titles* of these works, we note they caution dichotomization is "a practice to avoid" (Dawson & Weiss, 2012), "a bad idea" (Royston, Altman, & Sauerbrei, 2006), and a "heartbreak" (Streiner, 2002); call for its "death" (Fitzsimons, 2008); point out its "cost[s]" (Cohen, 1983), "perils" (Lemon, 2009), and "negative consequences" (Irwin & McClelland, 2003); and simply ask "why" (Owen & Froman, 2005).

One might argue that it is the prerogative of researchers to choose whether to bear these costs. For instance, if researchers wish to choose statistical procedures with lower power, they will bear the costs in terms of finding fewer true effects and presumably publishing fewer papers. We disagree. Researchers impose these costs not just on themselves but on the field as a whole: the research community has a vested interested in lower Type I and Type II error, more accurately reported and more efficient effect size estimates, and richer and more informative analyses and thus has the right—and perhaps the obligation—to demand as much from its members. In addition, when the absence of an effect is in a researcher's interest, it is clearly self-serving to use statistical procedures with lower power.

As such, we believe regression ought to be the normative tool of analysis when working with continuous variables both when the statistical assumptions explored by IPKSP hold and more generally. Simply put, regression more accurately represents the data and it lacks the numerous costs associated with dichotomization. Further, scatterplots with regression lines superimposed like Fig. 1 provide a powerful format for graphical presentation of data and results, and we encourage researchers to make ample use of such plots.

In light of this, we echo and extend IPKSP's call for greater justification of the use of dichotomization: merely claiming that one is interested in category differences and demonstrating that

Table 1
Costs of dichotomization. Selected costs of dichotomization with references. Iacobucci et al. (2015) note that the increase in Type I error is eliminated when (i) there are one or more manipulated variables, (ii) there is a single measured variable, and (iii) the relationship between the dependent variable and the measured variable is a linear function for each manipulated condition.

Cost	Selected references
Loss of individual-level variation	Altman and Royston (2006), Butts and Ng (2009), MacCallum et al. (2002)
Reduced predictive performance	Cohen (1983), Dawson and Weiss (2012), MacCallum et al. (2002)
Increased Type I error	Altman and Royston (2006), Butts and Ng (2009), Dawson and Weiss (2012), Fitzsimons (2008), Iacobucci et al. (2015), Lemon (2009), MacCallum et al. (2002), Owen and Froman (2005), Vargha, Rudas, Delaney, and Maxwell (1996)
Increased Type II error	Altman and Royston (2006), Cohen (1983), Dawson and Weiss (2012), Farewell, Tom, and Royston (2004), Humphreys (1978), Iacobucci et al. (2015), Irwin and McClelland (2003), Lemon (2009), Maxwell and Delaney (1993), Owen and Froman (2005), Royston et al. (2006), Streiner (2002), Vargha et al. (1996)
Inefficient or distorted effect size estimates	Cox (1957), Gelman and Park (2009), Hunter and Schmidt (1990), Lagakos (1988), Morgan and Elashoff (1986)
Misrepresentation of continuous constructs as fictional groups	MacCallum et al. (2002)
Concealment of nonlinearity	Altman and Royston (2006), MacCallum et al. (2002)

collinearity is absent should not be accepted as sufficient justification for dichotomization given the costs associated with it. We also reiterate that employing dichotomization because one is more comfortable with ANOVA than with regression is a poor justification. The choice of a statistical procedure should largely be based on its statistical properties. Thus, we strongly urge researchers who perceive regression as more difficult than ANOVA to learn more about regression, to embrace it, and to benefit from its superior performance in terms of statistical inference and presentation of results.

In sum, we are far less sanguine than IPKSP with regard to dichotomization. Rather than giving researchers a "green light" to dichotomize, we hope we have put them in the mindset of carefully thinking about what they are doing and whether the costs are worth bearing. We hope that for many researchers, the benefits of regression will be recognized and realized. For this reason, we encourage researchers, as both authors and reviewers, to remain skeptical of the use of dichotomization and more generally of discretization without proper justification.

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park: Sage Publications, Inc.

Altman, D. G., & Royston, P. (2006). The cost of dichotomising continuous variables. *British Medical Journal*, *332*(7549), 1080.

Asendorpf, J. B., Conner, M., Fruyt, F. D., Houwer, J. D., Denissen, J. J. A., Fiedler, K., et al. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality*, *27*(2), 108–119.

Bates, D. M., & Watts, D. G. (1988). *Nonlinear regression analysis and its applications.* Wiley.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, *26*(2), 211–252.

Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., et al. (2014). The replication recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology*, *50*, 217–224.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees.* New York: Wadsworth.

Buonaccorsi, J. P. (2010). *Measurement error: Models, methods, and applications.* Boca Raton, FL: Chapman and Hall/CRC.

Butts, M. M., & Ng, T. W. (2009). *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences.* Chopped liver? OK. Chopped data? Not OK. New York: Routledge.

Cleveland, W. S. (1993). *Visualizing data.* Summit, NJ: Hobart Press.

Cleveland, W. S. (1994). *The elements of graphing data* (2nd ed.). Summit, NJ: Hobart Press.

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

Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, *45*(12), 1304–1312.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *155–159*.

Cohen, J. (1994). The earth is round (p < .05). *American Psychologist*, *49*, 997–1003.

Cox, D. R. (1957). Note on grouping. *Journal of the American Statistical Association*, *52*, 543–547.

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*(1), 7–29.

Cureton, E. E. (1957). The upper and lower twenty-seven percent rule. *Psychometrika*, *22*, 293–296.

D'Agostino, R. B., & Cureton, E. E. (1975). The 27 percent rule revisited. *Educational and Psychological Measurement*, *35*(1), 47–50.

Dawson, N. V., & Weiss, R. (2012). Dichotomizing continuous variables in statistical analysis: A practice to avoid. *Medical Decision Making*, *32*, 225.

DeCoster, J., Iselin, A. -M. 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.

Fidler, F., Thomason, N., Cumming, G., Finch, S., & Leeman, J. (2004). Editors can lead researchers to confidence intervals, but can't make them think: Statistical reform lessons from medicine. *Psychological Science*, *15*(2), 119–126.

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

Gal, D., & McShane, B. B. (2012). Can small victories help win the war? Evidence from consumer debt management. *Journal of Marketing Research*, *XLIX*, 487–501.

Gelman, A. (2015). The connection between varying treatment effects and the crisis of unreplicable research: A Bayesian perspective. *Journal of Management*, *41*(2), 632–643.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models.* New York, NY: Cambridge University Press.

Gelman, A., & Park, D. K. (2009). Splitting a predictor at the upper quarter or third and the lower quarter or third. *The American Statistician*, *63*(1), 1–8.

Gigerenzer, G. (1987). *The probabilistic revolution. Vol. II: Ideas in the Sciences*, *Vol. II*, Cambridge, MA: MIT Press.

Gigerenzer, G. (2004). Mindless statistics. *Journal of Socio-Economics*, *33*, 587–606.

Gill, J. (1999). The insignificance of null hypothesis significance testing. *Political Research Quarterly*, *52*(3), 647–674.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). New York: Springer.

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.

Hunter, J. E., & Schmidt, F. L. (1990). Dichotomization of continuous variables: The implications for meta-analysis. *Journal of Applied Psychology*, *75*(3), 334–349.

Iacobucci, D. (2005). From the editor. *Journal of Consumer Research*, *32*(1), 1–6.

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

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

Irwin, J. R., & McClelland, G. H. (2003). Negative consequences of dichotomizing continuous predictor variables. *Journal of Marketing Research*, *40*(3), 366–371.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning.* New York: Springer.

Jensen, M. B. (1928). *Genetic psychology monographs: Child behavior, differential and genetic psychology.* Objective differentiation between three groups in education (teachers, research workers, and administrators)*Vol. 3*. (pp. 333–454), 333–454.

Johnson, P. O., & Neyman, J. (1936). Tests of certain linear hypotheses and their application to some educational problems. *Statistical Research Memoirs*, *1*, 57–93.

Kelley, T. L. (1939). The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology*, *30*(1), 17–24.

Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological Methods*, *17*(2), 137–152.

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.

Lemon, J. (2009). On the perils of categorizing responses. *Tutorials in Quantitative Methods for Psychology*, *5*(1), 35–39.

Lindsey, J. K. (2001). *Nonlinear models for medical statistics.* Oxford: Oxford University Press.

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.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). Boca Raton, FL: Chapman and Hall/CRC.

McShane, B. B., & Böckenholt, U. (2014). You cannot step into the same river twice: When power analyses are optimistic. *Perspectives on Psychological Science*, *9*(6), 612–625.

McShane, B. B., & Böckenholt, U. (2015). Planning sample sizes when effect sizes are uncertain: The power-calibrated effect size approach. *Psychological Methods* (in press).

McShane, B. B., & Gal, D. (2015). Blinding us to the obvious? the effect of statistical training on the evaluation of evidence. *Management Science* (in press).

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, sir Ronald, and the slow progress of soft psychology. *Journal of Counseling and Clinical Psychology*, *46*, 806–834.

Morgan, T. M., & Elashoff, R. M. (1986). Effect of categorizing a continuous covariate on the comparison of survival time. *Journal of the American Statistical Association*, *81*(396), 917–921.

Morrison, D. E., & Henkel, R. E. (1970). *The significance test controversy.* Chicago: Aldine.

Mosteller, F., & Tukey, J. W. (1977). *Data analysis and regression.* Pearson.

O'Brien, S. M. (2004). Cutpoint selection for categorizing a continuous predictor. *Biometrics*, *60*, 504–509.

Owen, S. V., & Froman, R. D. (2005). Why carve up your continuous data? *Research in Nursing and Health*, *28*, 496–503.

Pashler, H., & Wagenmakers, E. -J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, *7*, 528–530.

Preacher, K. J., Rucker, D. D., MacCallum, R. C., & Nicewander, W. A. (2005). Use of the extreme groups approach: A critical reexamination and new recommendations. *Psychological Methods*, *10*(2), 178–192.

Robbins, N. B. (2013). *Creating more effective graphs.* Hoboken, NJ: Wiley-Interscience.

Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist*, *44*(10), 1276–1284.

Ross, J., & Weitzman, R. A. (1964). The twenty-seven percent rule. *The Annals of Mathematical Statistics*, *35*, 214–221.

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

Rozenboom, W. W. (1960). The fallacy of the null hypothesis significance test. *Psychological Bulletin*, *57*, 416–428.

Sawyer, A. G., & Peter, J. P. (1983). The significance of statistical significance tests in marketing research. *Journal of Marketing Research*, *20*(2), 122–133.

Seber, G. A. F., & Wild, C. J. (2003). *Nonlinear regression.* Hoboken, NJ: Wiley-Interscience.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366.

Spiller, S. A., Fitzsimons, G. J., Lynch, J. G., Jr., & McClelland, G. H. (2013). Spotlights, floodlights, and the magic number zero: Simple effects tests in moderated regression. *Journal of Marketing Research*, *50*(2), 277–288.

Streiner, D. L. (2002). Breaking up is hard to do: The heartbreak of dichotomizing continuous data. *Canadian Journal of Psychiatry*, *47*(3), 262–266.

Tufte, E. R. (2001). *The visual display of quantitative information* (2nd ed.). Cheshire, CT: Graphics Press.

Tukey, J. W. (1977). *Exploratory data analysis.* Pearson.

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