e-ISSN: 2176-0756
DOI: 10.5585/ijsm.v14i2.2244
Data de recebimento: 18/01/2015
Data de Aceite: 25/04/2015
Organização: Comitê Científico Interinstitucional
Editor Científico: Fernando Antonio Ribeiro Serra
Avaliação: Double Blind Review pelo SEER/OJS
Revisão: Gramatical, normativa e de formatação

"MORE THAN MEETS THE EYE": A GUIDE TO INTERPRETING THE DESCRIPTIVE STATISTICS AND CORRELATION MATRICES REPORTED IN MANAGEMENT RESEARCH

Arthur G. Bedeian[1]
Louisiana State University

[1] The author of this essay wonders whether in teaching our students the latest analytic techniques we have neglected to emphasize the importance of understanding the most basic aspects of a study's primary data. In response, he provides a 12-part answer to a fundamental question: "What information can be derived from reviewing the descriptive statistics and correlation matrix that appear in virtually every empirically based, nonexperimental paper published in the management discipline?" The seeming ubiquity of strained responses, to what many at first consider to be a vexed question about a mundane topic, leads the author to suggest that students at all levels, seasoned scholars, manuscript referees, and general consumers of management research may be unaware that the standard Table 1 in a traditional Results section reveals "more than meets the eye!"

THIS ARTICLE IS REPRODUCED UNDER PREVIOUS AUTHORIZATION OF THE AUTHOR AND ACADEMY OF MANAGEMENT LEARNING & EDUCATION.
The original article was published by Academy of Management Learning & Education, Vol. 13, n 1, p 121-135.

Modern statisticians are familiar with the notions that any finite body of data contains only a limited amount of information, on any point under examination; that this limit is set by the nature of the data themselves, and cannot be increased by any amount of ingenuity expended in their statistical examination: that the statistician's task, in fact, is limited to the extraction of the whole of the available information on any particular issue (Sir Ronald A. Fisher,1935: 44 – 45)[2].

It has often occurred to me that the purpose of higher education is to make simple things difficult. This thought raced through my mind again when I innocently asked the graduate students in my research-methods course what they could learn from reviewing the descriptive statistics and correlation matrix that appear in virtually every empirically based, nonexperimental paper published in the management discipline. With eyes quickly glazing over, my question was met with blank stares. This struck me as rather curious, as all the students had previously completed a sequence of courses in regression analysis, multivariate statistics, and structural equation modeling. When I had asked questions about any of these techniques, responses came from all around the room. I should add that, in addition to management students of various stripes, there were also marketing, information systems, and statistics majors enrolled in my course.

It thus struck me as rather odd that across students trained in four methods-rich disciplines, not one could provide a comprehensive answer to what I suspect many felt was a vexed question about a mundane topic. What did this say about the quality of the students' graduate education and research preparation? In inquiring further, how- ever, it was evident that, in large part, the students were responding in kind. After all, how many paper presentations had they attended at professional meetings when no more than a few seconds had been spent showing a PowerPoint slide of a study's descriptive statistics and correlation matrix with the only comment being, "All the reliabilities were .70 or greater, and in every case the correlations were in the direction predicted by previous theory and research"? And on to the next slide. I suspect much the same could be said about the vast majority of published papers the students had read in their various disciplines.

[2] The comments of Joshua S. Bendickson, William B. Black, Timothy D. Chandler, Daniel B. Marin, Jean B. McGuire, Hettie A. Richardson, Edward E. Rigdon, Paul E. Spector, David L. Streiner, and, especially, Hubert S. Feild, on earlier drafts are gratefully acknowledged, as is the assistance of Jeremy B. Bernerth, Michael S. Cole, and Thomas H. Greckhamer.

The data reported in this manuscript were extracted from Anita Konieczka Heck (2000), Workplace whining: Antecedents and process of noninstrumental complaining. Unpublished doctoral dissertation, Louisiana State University, Baton Rouge.

BACKSTORY

Following class, I asked a respected colleague the same simple question I had asked my students. After making a few comments related to estimating score reliabilities and range restriction, she acknowledged never having seen a systematic treatment that went much beyond my students' bewildered responses. Come to think of it, neither had I and, as it turned out, neither had any of the other colleagues I was to later canvass. This left me wondering if, as Sir Ronald suggests in the opening epigraph, "any finite body of data contains only a limited amount of information" and a researcher's task is to extract the "whole" of that information, whether in teaching our students the latest analytic techniques we have neglected to emphasize the importance of under-standing the most basic aspects of a study's primary data.

In the ensuing days, I pondered whether the in- ability of my students to respond to what I had thought to be a softball question was a reflection of their preparation or emblematic of graduate education in general. The level of methodological training within the management discipline is hard to estimate. Moreover, the essence of this training varies, as the diverse areas within management differ in their research questions and approaches. The common training offered in core courses (such as I teach) dealing with measurement issues, applied statistics, and data analysis, however, is one aspect of graduate education that unifies our discipline.

In the ensuing days, I pondered whether the inability of my students to respond to what I had thought to be a softball question was a reflection of their preparation or emblematic of graduate education in general.

The last 35 years have been an exciting time for advances in research methods. Starting in the early 1980s, papers applying structural equation modeling, estimating multilevel statistical models, and discussing measurement invariance first began appearing in the Academy of Management Journal and Academy of Management Review. The Academy's Research Methods Division was formed as an interest group in 1985 and received division status in 1988. Signaling a growing appreciation of how enabling methodologies and analytic techniques can shape the questions management re- searchers ask, the Southern Management Association's Journal of Management inaugurated a stand-alone "Research Methods and Analysis (RM&A)" section in 1993. Five years later, RM&A (with the sponsorship of the Research Methods Division and the Academy) evolved into Organizational Research Methods (ORM), our discipline's first journal exclusively devoted to promoting “a more effective understanding of current and new methodologies and their application in organizational settings.” In the ensuing years, the pace of substantive developments in methodologies employed by the various areas within management has quickened, leading to broader and more complex analyses (Lee & Cassell, 2013).

Given the depth of training necessary to master our discipline’s vast methodological armamentarium, time spent understanding data fundamentals may seem a luxury. Such understanding, however, is not only required for assessing the validity of a study’s results, but also provides a foundation for both evaluating and contributing to advances in research methods. At the risk of generalizing from a limited sample, I am concerned that whereas we train our graduate students in the latest analytic techniques, they might not be exposed to the fundamentals necessary to fully understand the nature of the data they zealously collect (and sometimes so mercilessly torture). [3]Consequently, our students may not recognize how their lack of understanding affects the credibility of their conclusions and, in turn, the larger knowledge base of our discipline. Though graduate education intentionally favors sophisticated methodologies, I nevertheless believe that a solid understanding of the most basic aspects of a study’s primary data is required of all students, even if their talents and interests lie elsewhere. In my view, a full appreciation of the information conveyed by the descriptive statistics and relations between a study’s variables is imperative as a precursor to applying techniques as rudimentary as regression analysis or as advanced as multilevel structural equation modeling.

I am concerned that whereas we train our graduate students in the latest analytic techniques, they might not be exposed to the fundamentals necessary to fully understand the nature of the data they zealously collect (and sometimes so mercilessly torture).

With these thoughts in mind, it is hoped that the following 12-point checklist for reviewing the standard Table 1 (an example of which is reproduced nearby) that is de rigueur for traditional Results sections published in social-science disciplines such as management, industrial/organizational psychology, marketing, information systems, public administration, and vocational behavior, will be of value to students at all levels, as well as seasoned scholars, manuscript referees, and general consumers of management research. Given that the checklist has a didactic flavor, corrections, clarifications, or additions are welcomed. Table 2 summarizes the checklist using a series of questions that may be used as a guide in reviewing descriptive statistics and correlation matrices.

BACKGROUND

The results presented in Table 1 come from a field study of 290 schoolteachers and their principals, representing 22 elementary, middle, and high schools. Study data were collected through traditional paper-and-pencil surveys. The purpose of the study was to explore whether the effects of the independent variables Job Satisfaction (measured with 6 items), Affective Organizational Commitment (6 items), perceived workplace fairness (i.e., Procedural Justice and Distributive Justice; 9 and 6 items, respectively), and Leader–Member Exchange Quality (7 items) on Workplace Complaining (the dependent variable; 5 items) were mediated by self-esteem at work (i.e., Organization-Based Self- Esteem; 10 items). Teachers completed the individual difference and work-related attitude measures. Principals assessed the degree to which teachers complained. To allow for the possibility that teacher self-reports might be confounded by pressure for positive self-presentation, affective feelings, and male–female differences in complaining behavior, Social-Desirability Responding (13 items), Negative Affectivity (11 items), and Gender served as control variables. With the exception of Social Desirability, which was keyed so that true = 1 and false = 0, and Gender, which was recorded using a dummy-coded, dichotomously scored nominal variable, with 0 designating Males and 1 designating Females, participants rated all items with assigned values ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Responses to all multi-item measures were averaged rather than summed (so that they would be on the same metric as their component items) and coded so that higher values signified an increasingly higher level of either agreement or, for Social Desirability, an increased tendency to respond in a self-flattering manner. Averaging (as does summing) presumes that the separate items composing a measure tap the same construct, use the same response format, and have equivalent error score variances.

The variables identified in Table 1 refer to constructs common in OB/HR research. AMLE readers interested in, for instance, strategy or entrepreneurship might be more familiar with business- and industry-level variables such as firm performance, new product quality, and marketplace volatility. A full understanding of the basic aspects of a study’s primary data, however, is no less essential for accurately interpreting results in these areas. As the following checklist is, therefore, equally relevant for reviewing the descriptive statistics and correlation matrices reported through- out our discipline, readers should feel free to substitute variables from their own areas for those listed in Table 1.

[3] An equally nonrandom sample of campus presentations by yet-to-be-fledged PhD job candidates suggests a similar lack of exposure.

Table 1 - Descriptive Statistics and Correlations of Study Variables

Variables	*M*	*SD*	1	2	3	4	5	6	7	8	9	10
Dependent variable												
1. Workplace complaining	1.86	.96	(.96)[a]									
Mediating variable												
2. Organization-based self-esteem	4.01	.49	−.33	(.88)[a]								
Independent variables												
3. Job satisfaction	3.20	.72	−.23	.46	(.80)[a]							
4. Affective organizational commitment	3.63	.73	−.27	.62	.52	(.80)[a]						
5. Procedural justice	3.26	.72	−.26	.63	.64	.56	(.88)[a]					
6. Distributive justice	3.01	.98	−.24	.45	.72	.41	.65	(.94)[a]				
7. Leader–member exchange quality	3.25	.78	−.29	.67	.66	.62	.80	.65	(.90)[a]			
Control variables												
8. Social desirability	.65	.21	−.01	.22	.21	.16	.16	.10	.17	(.70)[b]		
9. Negative affectivity	2.96	.71	.15	−.21	−.13	−.08	−.10	−.07	−.15	−.35	(.86)[a]	
10. Gender (Male = 0; Female = 1)[c]	.82	.38	.06	−.02	.03	.03	−.07	−.02	−.07	−.01	.08	(NA)

Note. n = 290. Abbreviations: Correlations Š |.12| are significant at p < .05 (two-tailed test).
[a] Cronbach’s alpha (a) reliability coefficient.
[b] K-R 20 reliability coefficient.
[c] Point-biserial correlation.

A 12-POINT CHECKLIST

1. Basic Requirements

As a first step in a gaining a full understanding of the basic aspects of a study’s primary data, following Table 1 as an example, it is essential to verify that all relevant variables (including control variables) are listed with (at a minimum) means, standard deviations, number of cases (respondents), and (where appropriate) estimated score reliabilities for each multi-item measure. The total number of unique correlations possible in a study is equal to k * (k *–1)/2, where k is the number of variables. As there are 10 study variables in Table 1, there are 45 correlations to examine. The prespecified significance levels (two-tailed, nondirectional) for all correlations, commonly set at .05 or .01, should be indicated either with a single (*) or double asterisk (**), respectively or, as is done in Table 1, using a general note indicating the absolute magnitude beyond which the correlations are significant. The number of cases (respondents) on which study statistics are based should be considered adequate for interpreting the ensuing analyses with confidence given a study’s goals (for guidance on estimating sample-size requirements relative to desired statistical power, i.e., the probability of finding a relationship when one exists; see Eng, 2003).

A complete correlation matrix (including sample sizes, means, and standard deviations) is necessary as input for others who may wish to reproduce (and confirm) a study’s results, as well as perform secondary analyses (Zientek & Thompson, 2009). Whereas descriptive statistics and correlations should be rounded to two decimal places, recognize that standard zero-order (Pearson product- moment) correlations (r_{xy}) based on fewer than 500 cases lack stability beyond a single digit (Bedeian, Sturman, & Streiner, 2009). Avoid attaching too much importance to any one significant correlation, as it may be the one in 20 that is expected to be significant (at a .05 error rate) by chance alone. Thus, as there are 45 correlations in Table 1, approximately 2–3 would be expected to reach significance due to chance. Which, 2 or 3, however, are flukes and which are attributable to genuine covariations generalizable to a study’s population of interest is impossible to determine. Alternatively, the probability that at least one coefficient in a correlation matrix will be significant by chance alone at the 5% level is 1– 0.95k, where k equals the number of correlations (Streiner & Norman, 2011). Hence, the probability that at a minimum of one out of 20 correlations will be significant at random is > 64%; the probability that at least one out of 45 correlations (as in Table 1) will be significant by chance is > 90%.

Table 2 - A 12-Point Guide for Reviewing Descriptive Statistics and Correlation Matrices

Point 1: Basic requirements	• Are all relevant study variables (including control variables) listed with (at a minimum) means, standard deviations, number of cases (respondents), and (where appropriate) estimated reliability scores for each multi-item measure? Discussion and conclusions: Complete correlation matrix (including sample sizes, means, and standard deviations) is necessary as input for others to reproduce (and confirm) study's results, as well as perform secondary analyses. • Are significance levels (two-tailed, nondirectional) for all correlations indicated? • Is number of cases (respondents) on which study statistics are based adequate for interpreting ensuing analyses with confidence, given study's goals? • Are descriptive statistics and correlations rounded to two decimal places? Discussion and conclusions: Correlations based on fewer than 500 cases lack stability beyond a single digit.
Point 2: Frequency distributions	• Are means for all variables measured on a unidirectional scale less than twice their standard deviations, indicating underlying frequency distributions are likely asymmetrical, suggesting mean is not a typical or representative score? Discussion and conclusions: The mean is not a typical or representative score of an asymmetrical distribution. Moreover, if mean value is reported for dummy-coded dichotomously scored nominal variable (Male = 0 and Female = 1), value should not be interpreted as a measure of central tendency, but (assuming complete data) as proportion of females in a study sample, with a value > .5, indicating more women than men.
Point 3: Standard deviations	• Do standard deviations reported for study variables exceed their maximum possible values? Discussion and conclusions: Maximum standard deviation of variable created by averaging responses across items using 1–5 scoring continuum is half the range or (5 — 1)/2 = 2. If item responses are summed (rather than averaged), maximum possible reported standard deviation for a 6-item measure with a 5-point response format, would be half the range or (30 — 6)/2 = 12. • Are there any small standard deviations that may limit correlations between study variables?
Point 4: Reliabilities	• Are appropriate reliability estimates (e.g., Kuder-Richardson's K-R 20 coefficient of reliability for dichotomous scored items, Cronbach's coefficient alpha reliability for polytomous scored items) reported for each multiple-item measure composed of theoretically correlated items? Discussion and conclusions: Kuder-Richardson's K-R 20 coefficient and Cronbach's coefficient alpha are affected by measure's length and will either under- or overestimate score reliabilities depending on extent to which error components of all pairs of items composing a measure are correlated or uncorrelated. • Are estimated score reliabilities of acceptable magnitude considering a study's purpose, sample composition (e.g., gender, race, age, ethnicity, and education level), number of cases (respondents), and the specific conditions under which results were obtained? Discussion and conclusions: Reliability is a property of scores in hand, not a given measure per se; can seldom be compared across samples, settings, and time. To the extent these considerations promote greater score variance, they will yield a higher score reliability. Moreover, unless measure's item content is interpreted similarly by respondents who differ, in gender, race, age, ethnicity, education level, measure is unlikely to tap same common factor, thereby is meaningless to compare estimated score reliabilities across samples.
Point 5: Correlations	• Do any reported correlations exceed their maximum possible value given estimated score reliabilities of their individual correlates? Discussion and conclusions: Per the classical true-score theory of measurement, the maximum *possible* observed correlation between two variables (X and Y) cannot exceed product of square roots of their estimated reliabilities. Whatever their magnitude, reported correlations are not assumed to be representative of either all or most of study's respondents and, by extension, all or most individuals within a defined population.
Point 6: Correlate pairs	• Are there reasons for differences in the magnitude of correlate pairs? • Are variables in one or both correlate pairs nonlinearly related? Discussion and conclusions: Two variables may be "zero correlated" and, unless their joint (bivariate) distribution is normal, have a perfect (curvilinear) relationship. • Are relationships between variables, in either or both of the X-Y pairs, similar across their full range of scores?

Table 2
(Continued)

	Discussion and conclusions: Zero-order correlations assume the relationship between variable X and variable Y is of similar magnitude for all values of both. Violations in this assumption typically result in confidence intervals that are too wide or too narrow, thus, misrepresenting the set of values that likely includes an unknown population correlation. • Are there differences in the range of one or more of a correlate pair's constituent variables? Discussion and conclusions: Correlations may either be weakened by "range restriction" (wherein the scores on one or both of the variables being correlated cover only a portion of the variables' possible scores) or enhanced (wherein potential scores on a variable or variables in a correlate pair are restricted to extremes). Any form of range restriction (i.e., shrinkage or expansion) will also bias estimates of score reliabilities (as assessed by Kuder-Richardson's K-R 20 coefficient or Cronbach's coefficient alpha) by misrepresenting the true homogeneity/heterogeneity of underlying variable scores, with subsequent effects on Type I and Type II errors. • Do the estimated score reliabilities of the individual correlates comprising variables in one or both of correlate pairs reflect greater measurement error? Discussion and conclusions: The observed correlation between two variables cannot exceed product of square roots of their estimated score reliabilities. Imprecise measurement generally attenuates relationship estimates between variables, increasing the probability of Type II errors.
Point 7: Common-method variance/data dependency	• Are there indications of potential common-method variance given the source(s) from which study data were collected? Discussion and conclusions: Some differential covariance between variables may result from sharing same measurement approach. • Is it possible that some differential covariance between variables may be due to interdependence among ratings or raters? Discussion and conclusions: To the extent that ratings or raters are interdependent, there will be an underestimation of standard error estimators and an increase in risk of Type I errors.
Point 8: Sign reversals	• Are there unexpected sign reversals such as a negative correlation in a matrix of otherwise positive correlations? Discussion and conclusions: Unexpected sign reversals may indicate errors in data editing or coding that could easily produce spurious results. A mixture of signs may also hint at possible suppression effects.
Point 9: Collinearity	• Are there any correlations between predictor variables > |.70|, thus, suggesting collinearity may be a problem? Discussion and conclusions: When predictors are highly correlated (i.e., collinear), coefficient estimates (and their variances) in regression-type analyses will be inflated, elevating the risk of Type I errors.
Point 10: Point-biserial correlations	• Are point-biserial correlations properly identified? Discussion and conclusions: Whereas both Pearson product-moment and point-biserial correlations are a function of underlying relationship being estimated, point-biserial correlations are also a function of proportion of observations in each category of dichotomized variable, reaching their maxima when proportions in categories are equal. • Are category coding values for point-biserial correlations indicated? Discussion and conclusions: A point-biserial correlation cannot be interpreted without knowing how its dichotomized categories were coded. Thus, relative proportions in defining the dichotomous variable should be considered because as the difference between number of observations in each category increases, r_{pb} decreases, increasing likelihood of Type II errors.
Point 11: Missing data	• Are descriptive statistics and correlations between study variables based on complete (or incomplete) data for all cases (respondents) or computed using missing data imputation? Discussion and conclusions: If number of cases used to estimate correlations between study variables is not the same for each pair of correlates, the power of reported statistical tests may vary, resulting in correlations of identical magnitude being significant in one instance and not in another; correlations based on different subsets of cases are rarely comparable.
Point 12: Sampling	• Were targeted participants randomly chosen from a defined population? • If a representative (probability) sample has been drawn from a clearly defined population, is number of cases (respondents) sufficient to make statistical inferences about sampling frame from which they were drawn and adequate for eliciting an effect size of importance? Discussion and conclusions: To obtain true estimates of population parameters (including estimated score reliabilities) and to apply standard likelihood methods for generalizing a study's results, obtain a representative (probability) sample from a clearly defined population. Simulations aside, some error is always present in sampling, as even random samples are rarely perfectly representative. Random samples are nonetheless virtually always more representative than nonprobability samples. Whereas nonresponse may not necessarily bias a study's data, a single nonresponse renders a probability sample nonrandom and, thus, introduces ambiguity into the inferences that can be made.

2. Frequency Distributions

Compare the mean and standard deviation for each study variable. If a variable is measured on a unidirectional scale using low-to-high positive integers, such as 1 to 5 (as opposed to a bidirectional scale using minus-to-plus integers such as —3 to +3 with zero in the middle), and its mean is less than twice its standard deviation, the variable’s underlying frequency distribution is likely asymmetric, suggesting that the mean is neither a typical nor representative score (Altman & Bland, 1996). If a mean value is reported for a dummy- coded dichotomously scored nominal variable such as male = 0 and female = 1, this value should not be interpreted as a measure of central tendency, but (assuming complete data) as the proportion of females in a study sample, with a value > .5 indicating more women than men. In Table 1, the mean value of .82 signifies that 82% of the study sample is female. The accompanying standard deviation is equal to the square root of the proportion of males times the proportion of females or $\sqrt{.18\times.82} = .38$. As there are, however, only two possible values for a dichotomously scored variable, the standard deviation of the observed scores as a measure of variability is not very meaningful.

3. Standard Deviations

Confirm that the standard deviations reported for study variables do not exceed their maxima. Alternatively, be alert to any small standard deviations, as they may limit correlations between study variables. As noted, in the study on which Table 1 is based, responses to all multi-item measures were averaged and, with the exception of Social Desir- ability and Gender, coded such that higher values signify an increasingly higher level of agreement. Thus, as Job Satisfaction was assessed using a 6-item measure, with assigned values ranging from 1 (*strongly disagree*) to 5 (*strongly agree*), the maximum possible standard deviation is 2, half the range or (5–1)/2 = 2. Similarly, the maximum possible standard deviation of a variable created by averaging responses across items using a 1–7 scoring continuum is 3. In instances were item responses are summed (rather than averaged), the maximum possible reported standard deviation for a 6-item measure with a 5-point response format, is 12, half the range or (30 – 6)/2 = 12.

4. Reliabilities

Inspect the estimated score reliabilities for each multiple-item measure composed of theoretically correlated items. In the present context, *reliability* is defined on a conceptual level as the degree that respondents’ scores on a given measure are free from random error. Be sure that the appropriate estimators (e.g., Kuder-Richardson’s K-R 20 coefficient of reliability for dichotomous scored items, Cronbach’s coefficient alpha reliability for polytomous scored items) are reported and, as reliability is a property of the scores in hand rather than a given measure per se, are of acceptable magnitude considering a study’s goals, sample composition (e.g., gender, race, age, ethnicity, and education level), number of cases (respondents), and the specific conditions under which results were obtained. [4]To the extent sample composition, number of cases, and the specific conditions under which results were obtained promote greater variability in a measure’s scores, they will yield a higher estimated reliability (Rodriguez & Maeda. 2006). Because reliability is a property of scores derived from a measure and not of the measure itself, estimated reliabilities can seldom be compared across samples, settings, and time (for further de- tails, see Helms, Henze, Sass, & Mifsud, 2006). As a further complication, unless a measure’s item con- tent is interpreted similarly by respondents who differ, for example, in gender, race, age, ethnicity, and education level, it is unlikely that the mea- sure will tap the same common factor, in which case it is meaningless to compare estimated score reliabilities across samples (Raykov & Marcoulides, 2013).

Be aware that Kuder-Richardson’s (K-R 20) coefficient and Cronbach’s coefficient alpha are affected by a measure’s length. [5]If a measure contains 15 or more items, even if it is not composed of theoretically correlated items, both of these estimators may nevertheless be substantial (Cortina, 1993). Further, to the extent Kuder-Richardson’s K-R20

[4] Although Cronbach’s coefficient alpha remains the most established approach to estimating score reliability, several alternatives are available for other types of data and analyses. For instance, in contrast to coefficient alpha, which is based on item correlations, estimates of “composite reliability” have become increasingly popular. Composite-reliability estimates are computed using factor loadings, which are typically parameter estimates from a structural equation model or, alternatively, derived in studies conducted to estimate a measure’s factorial validity. As such, they are not customarily included in a standard Table 1 correlation matrix of study variables. For further details, see Peterson and Kim (2013). Note, too, Kuder- Richardson’s K-R 20 coefficient of reliability and Cronbach’s coefficient alpha reliability should only be used to provide reliability estimates for raw (summed) scores or scores that have been linearly transformed (e.g., averaged scores or linearly standardized scores). For specifics on estimating the reliability of nonlinearly transformed and norm-based scores, see Almehrizi (2013) and the references therein.

[5] In general, as the number of items in a measure increases, coefficient alpha increases. The exception occurs when items added to a measure are so weakly correlated with prior items that their negative effect on the average correlations among the items exceeds their positive influence on the total number of items, thereby, decreasing the estimated reliability of a measure’s scores (cf. DeVillis, 2006: S53).

coefficient and Cronbach's coefficient alpha register only specific-factor (i.e., content-specific) error associated with the items that compose a measure, they are lower bound reliability estimates. On the other hand, Kuder-Richardson's K-R20 coefficient and Cronbach's coefficient alpha overestimate score reliabilities when they incorporate item-error components that positively correlate due to, for instance, extraneous conditions (such as variations in feelings and mood) that carry over across items or item covariances that overlap because they measure a common factor (Gu, Little, & Kingston, 2013). Whether Kuder- Richardson's K-R 20 coefficient and Cronbach's co- efficient alpha under- or overestimate reliability depends on which set of contingencies is more pronounced (Huysamen, 2006). As this is impossible to know, and population parameters can only be estimated when using sample data, both Kuder-Richardson's K-R 20 coefficient and Cronbach's co-efficient alpha are at best *approximations* of true score reliability (Miller, 1995).

As noted in Table 1, with one exception, Cronbach's coefficient alpha reliability is provided for each multi-item study variable. Given that Social Desirability was measured using a true or false format, the Kuder-Richardson formula 20 measure of reliability is reported. Reliability coefficients theoretically range from 0 — 1.00. Returning to Table 1, the .86 reliability coefficient for Negative Affectivity indicates that, for the sample in question, 86% of the estimated observed score variance is attributable to "true" variance as opposed to random error.

5. Correlations

Ensure that the reported correlations do not exceed their maximum possible value. Following the classical true-score theory of measurement, the observed correlation between two variables (*X and Y*) cannot exceed the product of the square roots of their estimated score reliabilities (Bedeian, Day, & Kelloway, 1997). [6]Thus, if scores on the measures used to operationalize the variables each have an estimated reliability of .80, their maximum possible observed correlation (r_{xy}) will equal .80 or $\sqrt{.80} \times \sqrt{.80} = .80$. If the scores for one have a reliability of .60 and the other .80, their maximum possible observed correlation equals .69 or $\sqrt{.60} \times \sqrt{.80} = .69$. Referring to Table 1, given their estimated score reliabilities, the maximum possible correlation between Organization-Based Self-Esteem and Distributive Justice is $\sqrt{.88} \times \sqrt{.94} = .91$. Do recognize, however, whatever their magnitude, it should not be assumed that the reported correlations are representative of either all or even most of a study's respondents and, by extension, all or most of the individuals within a defined population. Simply put, associations that hold in the aggregate may not hold for either individual respondents within a sample or specific individuals within a sample's referent population and vice versa (Hutchinson, Kamakura, & Lynch, 2000).

6. Correlate Pairs

When comparing zero-order correlations between study variables recognize that one possible explanation for differences in magnitude may be the variables in one or both of the correlate pairs are not linearly related. Because zero-order correlations only measure the degree of linear (straight- line) association between two variables (X and Y), they underestimate the relationship between variables that nonlinearly covary. Indeed, it is possible for two variables to be "zero correlated" and, unless their joint (bivariate) distribution is normal, have a perfect (curvilinear) relationship (Good, 1962).

Differences in the magnitude of correlate pairs may also result if the strength of the relationship between the *X-Y* variables, in one or both of the pairs, varies across their corresponding scores. Zero-order correlations assume that the relation- ship between *X* and *Y* is of similar magnitude for all values of both variables. Referred to as *homoscedasticity*, when this assumption holds, the strength of the relationship between any given value of *X* will be the same for each of the possible values of *Y*, and the strength of the relationship between any given value of *Y* will be the same for each of the possible values of *X*. Thus, if there is a strong (weak) correlation between *X* and *Y*, the strong (weak) relationship will exist across all values of both variables (cf. Sheskin, 2011: 1285). If, however, there is more variation in Y for high values of *X* than for low values of *X*, a zero-order correlation will underestimate the relationship between *X* and *Y* for low values of *X* and overestimate the relationship for high values of *X* and vice versa (cf. Evans & Rooney, 2011: 312). By extension, the magnitudes of different correlate pairs are only comparable to the extent that the strength of the relationship between variables, in either or both of the *X-Y* pairs, is similar across their full range of scores. Violations in homoscedasticity may be caused by non-normality in the underlying distribution of either *X* or *Y* scores or by the indirect

[6] The classical true-score theory of measurement assumes complete independence among true- and error-score components. When this assumption does not hold, the observed correlation between two variables may exceed the product of the square roots of their estimated reliabilities and, in fact, be greater than 1.00. This is a common pitfall when correcting observed correlations for attenuation due to measurement error. For further details, see Nimon, Zientak, and Henson, 2012. Whereas the Pearson r also assumes that the joint distribution of two variables (X and Y) is bivariate normal, it has been shown to be insensitive to even extreme violations of this assumption (Havlicek & Peterson, 1977).

effect of a third variable, and typically result in confidence intervals that are either too wide or too narrow, thereby, misrepresenting the set of values that likely includes an unknown population correlation.

Correlate pairs may further vary in magnitude due to differences in the range of one or more of their constituent variables. Correlations are usually weakened by “range restriction,” wherein the scores on one or both of the variables being correlated cover only a portion of the variables’ possible scores (e.g., scores are either all generally high or all generally low or mostly in the middle with a few extremes). Consequently, the variance of the scores is reduced, which may decrease their correlation. Conversely, the opposite may occur if the range of scores on one or both of the variables being correlated is artificially expanded, thereby increasing the variance in scores and enhancing their correlation. Known as “reverse range restriction” or “range enhancement” this would *typically* happen when scores on a variable or variables in a correlate pair are restricted to extremes; for ex- ample, when only the highest and lowest third of scores are entered into an analysis and, as a result, deletion of the middle third increases the variance in scores (as scores around the mean are excluded). The qualifiers “usually,” “may,” and “typically” in the preceding sentences reflect the fact that in those rare instances where the association between two variables is perfectly linear, range restriction will not affect their correlation, as the relationship between the variables is constant across all values. As an aside, as estimated score reliabilities are partially a function of the variance for the summed scores all items composing a mea- sure, any form of range restriction (i.e., shrinkage or expansion) will also bias estimates of score reliabilities (as assessed by Kuder-Richardson’s K-R 20 coefficient or Cronbach’s coefficient alpha) by misrepresenting the true homogeneity/heterogeneity of underlying variable scores, with subsequent effects on Type I (falsely identifying an effect in a sample that does not exist in a defined population) and Type II (failing to identify an existing population effect within a study sample) errors (Weber, 2001). For a complete discussion of range- restriction issues, see Bobko (2001) and Wiberg and Sundström (2009).

Finally, as mentioned, following the classical true-score theory of measurement, the observed correlation between two variables cannot exceed the product of the square roots of their estimated score reliabilities. Thus, an additional explanation for differences in magnitude when comparing correlations between study variables may be that the estimated score reliabilities of the individual correlates comprising the variables in one or both of the correlate pairs reflect greater measurement error. Imprecise measurement generally attenuates relationship estimates between variables, increasing the probability of Type II errors.

7. Common-Method Variance/Data Dependency

Check for potential common-method variance, wherein some of the differential covariance between variables results from sharing the same measurement approach. Taking self-report measures as an example, evidence of common-method variance is present if the magnitudes of a disproportionate share of the observed correlations be- tween self-reported variables are higher than be- tween those collected using other methods. In the opposite way, there is support for the correlations between self-reported variables not being biased due to common-method variance if the magnitudes of a similar proportion of observed correlations between self-reported variables are no greater than those collected using nonself-reports. That said, other-report data (including interviews with workplace collaterals, behavioral observations by supervisors and peers, professional assessment reports, and archival records) should not automatically be presumed to be more valid than self- reports. Indeed, if the estimated correlation be- tween two variables differs depending on whether the variables have been measured using self- report or other-source ratings, which estimate is more valid is inconclusive, as both self-report and other-source ratings are susceptible to many of the same attributional and cognitive biases. In turn, if the correlations are similar, the likelihood of a constant inflation effect due to common-method variance is reduced. In Table 1, principals’ ratings of teachers’ Workplace Complaining is the only nonself-report measure. Consequently, though common-method variance is likely reduced given the different rating sources from which the study data were collected, the extent to which common- method variance may still be present is unknown. For a further discussion of method variance as an artifact in data reporting, see Chan (2009).

It should also be noted that some of the differential covariance between variables may likewise be due to interdependence among either ratings or raters (Kenny & Judd, 1996). Such interdependencies might occur for many reasons. In considering the variables presented in Table 1, each of the 22 participating principals assessed the degree to which teachers at their schools complained. Consequently, each principal’s ratings are nested in a priori groupings (viz., teachers within schools). To the extent that the principals’ ratings of the teachers’ complaining behaviors are clustered by school (and therefore dependent by virtue of coming from a common source), there will be an underestimation of the true standard errors and an increase in the risk of Type I bias. Ratings may also be dependent when raters interact with one another. For example, given that the teachers at the schools from which the data in Table 1 were collected shared their work-related experiences with each other, their perceptions of Leader–Member

Ex- change Quality and Distributive Justice may like-wise be clustered by school.

In addition to discussing other forms of data dependency, Bliese and Hanges (2004; Bliese, 2000) review various procedures for estimating interdependence among observations (e.g., ratings and raters) and advise that even if only individual- level relationships are of interest, such procedures should be applied whenever observations may be dependent. A traditional Table 1 reports raw correlations without corrections for data dependency. Whenever the observed correlations and associated significance tests in a Table 1 are suspected of being biased due to non-independence, they should be interpreted with caution until properly modeled. When non-independence is present, appropriate statistical analyses (e.g., hierarchical linear models, heteroscedasticity-consistent standard- error estimators for ordinary least squares regression) should be used to control for a lack of independence in any subsequent analyses.

8. Sign Reversals

Look for unexpected sign reversals, such as a negative correlation in a matrix of otherwise positive correlations. This may indicate an error in data editing or coding that could easily produce spurious results. A mixture of signs may also hint at possible suppression effects, in which a third variable (e.g., verbal ability) unrelated to a designated outcome variable (e.g., job performance) removes (suppresses) outcome-irrelevant variance in one or more predictors (e.g., a paper-and-pencil test of job performance), thereby enhancing the overall explanatory or predictive power of a hypothesized model (cf. Cohen, Cohen, West, & Aiken, 2003: 78). For a detailed treatment of suppression in its classic form, as well as other types of suppression, see Pandy and Elliott (2010).

9. Collinearity

Check for potential collinearity between predictor (explanatory) variables. When predictors are highly correlated (i.e., collinear), coefficient estimates (and their variances) in regression-type analyses will be inflated, elevating the risk of Type I errors. Collinearity is typically associated with a redundancy (overlap) in the information contained in predictor variables (e.g., age and years of work). Its general effect is to obscure the role of individual predictors and, hence, may lead to the potential misidentification of relevant effects in a hypothesized model (Tu, Kellett, Clerehugh, & Gilthorpe, 2005). Though there is no specific cut-off, if the correlation between two predictor variables is between —0.70 and +0.70 (suggesting 50% shared variance), collinearity is unlikely to be a problem. As indicated in Table 1, collinearity could be a threat to conclusions drawn from, for instance, a multiple regression in which either both Job Satisfaction and Distributive Justice or Leader–Member Exchange Quality and Procedural Justice were used to predict Workplace Complaining.

10. Point-Biserial Correlations

Note that if a reported correlation is between a continuous variable *X* and a truly dichotomous variable *Y* (e.g., Male/Female, stayers/leavers, present/absent, employed/unemployed), it is not a standard zero-order (Pearson product-moment) correlation (r_{xy}), but a point-biserial correlation (r_{pb}) and should be identified as such. Whereas both Pearson product-moment and point-biserial correlations are a function of the underlying (linear) relationship being estimated, point-biserial correlations are also a function of the proportion of observations in each category of the dichotomous variable, reaching their maxima when the proportions in the categories are equal. As the difference between the proportions in each category of the dichotomous variable increases, r_{pb} decreases, in-creasing the likelihood of Type II errors. Thus, in interpreting a point-biserial correlation, the rela- tive proportions in the two categories defining the dichotomous variable should be considered. In- deed, given the limits imposed by differing propor- tions in the categories composing the dichotomous variable, researchers must also consider the goal of an analysis and the context in which results are to be understood when assessing the practical value of estimating a point-biserial correlation (McGrath & Meyer, 2006).

Finally, a point-biserial correlation cannot be interpreted without knowing how its dichotomized categories were coded. If the categories were coded 0 for Male and 1 for Female, as in Table 1, r_{pb} would fall in the range —1 to +1 and be construed in the same manner as r_{xy}. [7]Although the assignment of category values is arbitrary (as in the pre- ceding example; it would have been equally acceptable to code 1 for Male and 0 for Female), which category is coded 1 and which is coded 0 does affect the sign of the observed correlations.

Thus, with reference to Table 1 and the association between Gender and other study variables, a correlation with a positive sign indicates a stronger relationship for the category coded 1 (Female), and a negative sign signifies a weaker relationship for the category coded 0 (Male). The across-the-board low correlations observed for Gender (range —.07 to .08),

[7] A perfect correlation can only occur between two variables with the same shaped (both in skewness and kurtosis) distribution of scores. Because continuous and dichotomous variables inherently have different distributions, the true range of the point-biserial correlation only approaches ± 1 (cf. Karabinus, 1975: 279).

however, suggest that the associations in question do not substantially vary for males and females.

11. Missing Data

Determine whether the descriptive statistics and correlations between study variables were based on complete (or incomplete) data for all cases (respondents) or computed using missing data imputation. In the absence of complete data, if the number of cases is the same for all variables (as in Table 1), it is possible that either listwise deletion (in which study respondents missing even a single observation are eliminated from all statistical analyses) or a more advanced procedure was employed to replace missing observations by imputing plausible values predicted from avail- able data.

If the number of cases, however, is different across variables, pairwise deletion was used to deal with missing data. In contrast to listwise deletion, pairwise deletion only drops from analysis pairs of variables (not respondents) for which an observation is missing. Thus, in computing correlations and other statistics, all cases in which X and Y are observed are used regardless of whether observations on other variables are missing. If missing data were handled using pairwise deletion and, thus, a different number of cases was used to estimate the correlations between different study variables, the range that includes the lowest and highest number of cases should be reported (e.g., n = 297–312). As the number of cases used to estimate the correlations between study variables may not be the same for each pair of correlates, the power of the reported statistical tests may vary, resulting in correlations of identical magnitude being significant in one instance and not in another. Moreover, because such correlations are based on different subsets of cases, they will rarely be com- parable. Note, although the number of cases on which a correlation is computed will partially determine its statistical significance, by itself, sample size, as contrasted with, say, the amount of variability in a data set, does not directly affect the magnitude of a correlation (Goodwin & Goodwin, 1999). At the same time, other things being equal, the likelihood of finding a spurious correlation is greater for small than for large sample sizes, as the latter will be more representative of a defined population (Kozak, 2009). See Point 12, “sampling,” for the appropriate caveats in this regard.

Pairwise deletion is generally only considered appropriate when the number of cases is large and there are relatively few missing data randomly distributed across cases and variables. Both pair- wise and listwise deletion assume that data are missing completely at random, meaning that missing values for a particular variable are unrelated to other study variables or the underlying values of the variable itself. If this assumption is violated, the sample-derived standard error estimates of the true standard errors will be biased, calling into question the validity of statistical tests and confidence intervals (Venter & Maxwell, 2000). See Baraldi and Enders (2010) and Johnson and Young (2011) for further specifics on handling missing data.

12. Sampling

For studies in which targeted participants were randomly chosen from a defined population, con- firm that the number of cases (respondents) is sufficient to make statistical inferences about the sampling frame from which they were drawn and adequate for eliciting an effect size of importance (i.e., whether the variance explained by a hypothesized model is “big enough” relative to unexplained variability to be judged practically significant). For guidance on determining an effective number of cases for achieving an effect size of interest, see Lenth (2001). Furthermore, to obtain true estimates of population parameters (including estimated score reliabilities) and to apply standard likelihood methods for the purpose of generalizing a study’s results, it is necessary to obtain a representative (probability) sample from a clearly defined population. Note, though, outside of simulations, some error is virtually always present in sampling, as even random samples are rarely perfectly representative. Random samples are nonetheless almost always more representative than nonprobability samples, which tend to systematically differ from a referent population on certain characteristics (cf. Johnson & Christensen, 2012: 217). Moreover, whereas nonresponse may not necessarily bias a study’s data, a single nonresponserenders a probability sample nonrandom and, thus, introduces ambiguity into the inferences that can be made (Wainer, 1999).

AFTERTHOUGHTS

In reflecting further on the bewildered responses of both my students and the colleagues I consulted in seeking an answer to what was meant as an innocent question, several additional thoughts beyond the content of our students’ graduate education and research preparation came to mind. An initial thought was sparked by Sherman’s (1990) observation that graduate programs in psychology have come to place an increasing emphasis on publications as a means of enhancing the future placement of their PhD recipients. In doing so, many have begun to immerse their students in research projects beginning in their first semester of course work. Sherman notes, however, that this “immersion in research” approach all too often comes without considering whether the students have taken the courses necessary to possess a full understanding of the fundamentals of sound re- search. I suspect much the same is true in our own discipline, where the pressure to establish one’s research spurs prior to entering the job market is no less extreme (Miller, Taylor, & Bedeian, 2011).

This initial thought led to the realization that whereas the pace of substantive developments in methodologies employed by the various areas within management has quickened, leading to broader and more complex analyses, as noted supra, there is a notable absence of information regarding the actual level of methodological training in our discipline. A survey of management doctoral programs (perhaps under the sponsorship of the Academy’s Research Methods Division) to discern the depth of students’ research preparation would be a welcome first step in estimating the content and level of contemporary methodological training. In particular, information regarding which analytic techniques the diverse areas within management require their students to master would provide insights into what different pro- grams consider necessary for embarking upon a successful career. Further, I would be curious to know the extent to which our doctoral programs depend on courses offered “across campus” to train graduate students in newer analytic techniques. I suspect that programs offering the “best” methodological training access resources across a variety of curricula, including psychology, sociology, and economics. In addition, an increasing percentage of new PhDs are awarded outside North America. If there are differences in methodological training between North American and other graduate programs, it would be informative to know the bases on which these differences rest.

Course work, however, is not the only way for graduate students to learn the rudiments of good research. Proseminars and brown-bag sessions in which accepted research practices are discussed are also helpful. Moreover, workshops and tutorials offering instruction in new methodological developments are regularly held at professional meetings on both the regional and national levels. Such supplements are valuable for at least two reasons. First, with the rapid advancement in sophisticated methodologies, we can no longer pro- vide our students with classroom instruction that offers more than an overview of the vast range of data collection and analytic techniques now avail- able. Second, for faculty members who have fallen behind, such informal means represent a way for updating their methodological training. In this connection, it has been estimated that most faculty members acquire 80% of the knowledge necessary to sustain their careers after they have completed their formal education. For this reason, it has been advised, “When one submits to the temptation to jump from a research report’s abstract to its conclusion, bypassing the methods section, it is time to go back to school” (Bedeian, 1996: 8).

A final thought concerns the continuing advancement of management as a discipline. For the purpose of methodological training, Muthén (1989:186) has identified three types of students: “those who emphasize substantive interest, those who emphasize methodological interest but do not aspire to contribute to methodology, and those who place a strong emphasis on methodology and have aspirations to in some way enhance... methodology.” The first type constitutes the majority of “the users” (students and faculty) in any discipline and only requires a conceptual understanding of advanced techniques. These users are best served by working closely with colleagues who have inti- mate knowledge of emerging methodological developments. The second type is composed of users who combine a strong grasp of methods with a good understanding of their substantive interest. These users will be capable of making meaningful contributions to their discipline’s understanding with minor assistance from more quantitatively adept colleagues. The third type is made up of a relatively small number of users interested in becoming specialized methodologists. These users aspire to master not only the latest methodological developments, but to someday be at the forefront of advancing their discipline’s research methods. As with other disciplines, our continued success in furthering management learning and education will require the combined efforts of all three types of users. Regardless of inclination, however, to be successful in their chosen career paths, all users require a full appreciation of the information conveyed by the descriptive statistics and relations between a study’s variables.

CODA

In contemplating the natural growth and development of a garden as it moves through the sea- sons, poet Rudyard Kipling (1911: 249) observed, “The Glory of the Garden lies in more than meets the eye.” As the preceding checklist illustrates, the glory of a standard correlation matrix with its accompanying descriptive statistics also “lies in more than meets the eye,” being more revealing than it may first appear. Thinking back on the blank stares I encountered with the simple question—*What do you see when you look at a standard correlation matrix with its accompanying descriptive statistics?*—I continue to wonder if, in a similar fashion, as we educate our students in the glory of the latest analytic techniques, we have overlooked Sir Ronald’s admonition that ingenuity in methods is no substitute for a complete understanding of the most basic aspects of one’s data.

REFERENCES

Almehrizi, R. S. 2013. Coefficient alpha and reliability of scale scores. Applied Psychological Measurement, 37: 438 – 459.

Altman, D. G., & Bland, J. M. 1996. Detecting skewness from summary information. BMJ, 313: 1200.

Baraldi, A. N., & Enders, C. K. 2010. An introduction to modern missing data analyses. Journal of School Psychology, 48: 5–37.

Bedeian, A. G. 1996. Lessons learned along the way: Twelve suggestions for optimizing career success. In P. J. Frost & M. S. Taylor (Eds.), Rhythms of academic life: Personal ac- counts of careers in academia: 3–9. Thousand Oaks, CA: Sage.

Bedeian, A. G., Day, D. V., & Kelloway, E. K. 1997. Correcting for measurement error attenuation in structural equation models: Some important reminders. Educational Psychological Measurement, 57: 785–799.

Bedeian, A. G., Sturman, M. C., & Streiner, D. L. 2009. Decimal dust, significant digits, and the search for stars. Organizational Research Methods, 12: 687– 694.

Bliese, P. D. 2000. Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. J. Klein and S. W. Kozlowski (Eds.), Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions: 349 –381. San Fran- cisco: Jossey-Bass.

Bliese, P. D., & Hanges, P. J. 2004. Being both too liberal and too conservative: The perils of treating grouped data as though they were independent. Organizational Research Methods, 7: 400 – 417.

Bobko, P. 2001. Correlation and regression: Applications for industrial organizational psychology and management, (2nd ed). Thousand Oaks, CA: Sage.

Chan, D. 2009. So why ask me? Are self-report data really bad? In C. E. Lance & R. J. Vandenberg (Eds.), Statistical and methodological myths and urban legends: 309 –336. New York: Routledge.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. 2003. Applied multiple regression/correlation analysis for the behavioral sciences, (3rd ed). Mahwah, NJ: Erlbaum.

Cortina, J. M. 1993. What is coefficient alpha? An examination of theory and applications. Journal of Applied Psychology, 78: 98 –104.

DeVellis, R. F. 2006. Classical test theory. Medical Care, 44: S50 –S59.

Eng, J. 2003. Sample size estimation: How many individuals should be studied? Radiology, 227: 309 –313.

Evans, A. N., & Rooney, B. J. 2011. Methods in psychological research, (2nd ed). Thousand Oaks, CA: Sage.

Fisher, R. A. 1935. The design of experiments. Edinburgh, Scot- land: Oliver and Boyd.

Good, I. J. 1962. A classification of fallacious arguments and interpretations. Technometrics, 4: 125–132.

Goodwin, L. D., & Goodwin, W. L. 1999. Measurement myths and misconceptions. School Psychology Quarterly, 14: 408 – 427.

Gu, F., Little, T. D., & Kingston, N. M. 2013. Misestimation of reliability using coefficient alpha and structural equation modeling when assumptions of tau-equivalence and uncorrelated errors are violated. Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 9: 30 – 40.

Havlicek, L. L., & Peterson, N. L. 1977. Effect of the violation of assumptions upon significance levels of the Pearson r. Psychological Bulletin, 84: 373–377.

Heck, A. K. 2000. Workplace whining: Antecedents and process of noninstrumental complaining. Unpublished doctoral dissertation, Louisiana State University, Baton Rouge.

Helms, J. E., Henze, K. T., Sass, T. L., & Mifsud, V. A. 2006. Treating Cronbach’s alpha reliability coefficients as data in counseling research. Counseling Psychologist, 34: 630 – 660.

Hutchinson, J. W., Kamakura, W. A., & Lynch, J. G., Jr. 2000. Unobserved heterogeneity as an alternative explanation for “reversal” effects in behavioral research. Journal of Consumer Research, 27: 324 –344.

Huysamen, G. K. 2006. Coefficient alpha: Unnecessarily ambiguous, unduly ubiquitous. South Africa Journal of Industrial Psychology, 32: 34 – 40.

Johnson, D. R., & Young, R. 2011. Toward best practices in analyzing datasets with missing data: Comparisons and recommendations. Journal of Marriage and the Family, 73: 920 –945.

Johnson, R. B., & Christensen, L. B. 2012. Educational research: Quantitative, qualitative, and mixed approaches, (4th ed). Thousand Oaks, CA: Sage.

Karabinus, R. A. 1975. The r-point biserial limitation. Educational Psychological Measurement, 34: 277–282.

Kenny, D. A., & Judd, C. M. 1996. A general procedure for the estimation of interdependence. Psychological Bulletin, 119: 138 –148.

Kipling, R. 1911. The glory of the garden. In C. L. R. Fletcher and R. Kipling (Eds.) A school history of England: 249 –250. Ox- ford, England: Clarendon Press.

Kozak, M. 2009. How to show that sample size matters. Teaching Statistics, 31: 52–54.

Lee, B., & Cassell, C. M. 2013. Research methods and research practice: History, themes and topics. International Journal of Management Reviews, 15: 123–131.

Lenth, R. V. 2001. Some practical guidelines for effective sample size determination. American Statistician, 55: 187–193.

McGrath, R. E., & Meyer, G. J. 2006. When effect sizes disagree: The case of r and d. Psychological Methods, 11: 386 – 401.

Miller, A. N., Taylor, S. G., & Bedeian, A. G. 2011. Publish or perish: Academic life as management faculty live it. Career Development International, 16: 422– 445.

Miller, M. B. 1995. Coefficient alpha: A basic introduction from the perspectives of classical test theory and structural equation modeling. Structural Equation Modeling, 2: 255– 273.

Muthén, B. O. 1989. Teaching students of educational psychology new sophisticated statistical techniques. In M. C. Wittrock & F. Farley (Eds.), The future of educational psychology: 181–189. Hillsdale, NJ: Erlbaum.

Nimon, K., Zientak, L. R., & Henson, R. K. 2012. The assumption of a reliable instrument and other pitfalls to avoid when considering the reliability of data. Frontiers in Quantitative Psychology and Measurement, 3: 1–13. Available from http://www.frontiersin.org/Quantitative_Psychology_and_Measurement/10.3389/fpsyg.2012.00102/abstract. Accessed on January 6, 2013.

Pandy, S., & Elliott, W. 2010. Suppressor variables in social work research: Ways to identify in multiple regression models. Journal of the Society for Social Work Research, 1: 28 – 40.

Peterson, R. A., & Kim, Y. 2013. On the relationship between coefficient alpha and composite reliability. Journal of Applied Psychology, 98: 194 –198.

Raykov, T., & Marcoulides, G. A. 2013. Meta-analysis of scale reliability using latent variable modeling. Structural Equation Modeling, 20: 338 –353.

Rodriguez, M. C., & Maeda, Y. 2006. Meta-analysis of coefficient alpha. Psychological Methods, 11: 306 – 322.

Sherman, S. J. 1990. Commentary on “graduate training in statistics, methodology, and measurement in psychology: A survey of PhD programs in North America,” American Psychologist, 45: 729.

Sheskin, D. J. 2011. Handbook of parametric and nonparametric statistical procedures, (5th ed). Boca Raton, FL: Chapman & Hall/CRC.

Streiner, D. L., & Norman, G. R. 2011. Correction for multiple testing: Is there a resolution? Chest, 140: 16 –18.

Tu, Y. K., Kellett, M., Clerehugh, V., & Gilthorpe, M. S. 2005. Problems of correlations between explanatory variables in multiple regression analyses in the dental literature. British Dental Journal, 199: 457– 461.

Venter, A., & Maxwell, S. E. 2000. Issues in the use and application of multiple regression analysis. In H. E. A. Tinsley & S. D. Brown (Eds.), Handbook of applied multivariate statistics and mathematical modeling: 152–182. San Diego, CA: Academic Press.

Wainer, H. 1999. The most dangerous profession: A note on nonsampling error. Psychological Methods, 4: 250 –256.

Weber, D. A. 2001. Restriction of range: The truth about consequences and corrections. (ERIC Document Reproduction No. Service ED449213).

Wiberg, M., & Sundström, A. 2009. A comparison of two approaches to correction of restriction of range in correlation analysis. Practical Assessment Research and Evaluation, 14: 1–9.

Zientek, L. R., & Thompson, B. 2009. Matrix summaries improve research reports: Secondary analyses using published literature. Educational Researcher, 38: 343–352.