



Calanthia Wright

Essential Topics in Statistics

Essential Topics in Statistics

About the Book

Statistics is the science of compilation of numerical data that may represent a social or physical fact. This book on statistics deals with the basic methodology that is followed to derive inferences. Statistics uses the help of aggregation, averages and mean to calculate values of study. Contents included in this book discuss the various techniques through which results on the basis of data, data sets and models can be arrived at. There has been rapid progress in this field and its applications are finding their way across multiple industries. Tools such as sampling, surveys and experiments are a vital part of this field. This book on statistics will be of great help to students and researchers in the fields of economics, number theory and computational statistics.

About the Editor

Calanthia Wright received her MS in pure mathematics and PhD in applied mathematics from the University of Toronto, Canada. Her academic and research interests lie in differential equations, decision sciences and optimization techniques. She is a mathematician and a professor of pure mathematics. Her current work is based on optimization through probability and statistics. Wright's authored and edited writings, which include articles, journal papers, book chapters and conference proceedings, have been published worldwide.

NYRESEARCH
P R E S S
www.nyresearchpress.com



Wright Essential Topics in Statistics

NY RESEARCH
 $\nu_k \nu_s$

Essential Topics in Statistics

Editor: Calanthia Wright

NYRESEARCH
P R E S S

New York

Published by NY Research Press
118-35 Queens Blvd., Suite 400,
Forest Hills, NY 11375, USA
www.nyresearchpress.com

Essential Topics in Statistics
Edited by Calanthia Wright

© 2017 NY Research Press

International Standard Book Number: 978-1-63238-560-4 (Hardback)

This book contains information obtained from authentic and highly regarded sources. Copyright for all individual chapters remain with the respective authors as indicated. All chapters are published with permission under the Creative Commons Attribution License or equivalent. A wide variety of references are listed. Permission and sources are indicated; for detailed attributions, please refer to the permissions page and list of contributors. Reasonable efforts have been made to publish reliable data and information, but the authors, editors and publisher cannot assume any responsibility for the validity of all materials or the consequences of their use.

The publisher's policy is to use permanent paper from mills that operate a sustainable forestry policy. Furthermore, the publisher ensures that the text paper and cover boards used have met acceptable environmental accreditation standards.

Trademark Notice: Registered trademark of products or corporate names are used only for explanation and identification without intent to infringe.

Cataloging-in-Publication Data

Essential topics in statistics / edited by Calanthia Wright.
p. cm.
Includes bibliographical references and index.
ISBN 978-1-63238-560-4
1. Statistics. 2. Mathematical statistics. I. Wright, Calanthia.
QA276 .E87 2017
519.5--dc23

Printed in China.

Essential Topics in Statistics

Preface

I am honored to present to you this unique book which encompasses the most up-to-date data in the field. I was extremely pleased to get this opportunity of editing the work of experts from across the globe. I have also written papers in this field and researched the various aspects revolving around the progress of the discipline. I have tried to unify my knowledge along with that of stalwarts from every corner of the world, to produce a text which not only benefits the readers but also facilitates the growth of the field.

Statistics is the science of compilation of numerical data that may represent a social or physical fact. This book on statistics deals with the basic methodology that is followed to derive inferences. Statistics uses the help of aggregation, averages and mean to calculate values of study. Contents included in this book discuss the various techniques through which results on the basis of data, data sets and models can be arrived at. There has been rapid progress in this field and its applications are finding their way across multiple industries. Tools such as sampling, surveys and experiments are a vital part of this field. This book on statistics will be of great help to students and researchers in the fields of economics, number theory and computational statistics.

Finally, I would like to thank all the contributing authors for their valuable time and contributions. This book would not have been possible without their efforts. I would also like to thank my friends and family for their constant support.

Editor

Contents

	Preface.....	VII
Chapter 1	A defense against the alleged unreliability of difference scores..... David Trafimow	1
Chapter 2	Bayesian index to compare assumed survival functions of two hazards following exponential distributions..... Yohei Kawasaki, Masaaki Doi, Kazuki Ide and Etsuo Miyaoka	11
Chapter 3	MANOVA, LDA, and FA criteria in clusters parameter estimation..... Stan Lipovetsky	21
Chapter 4	Analysis of a redundant system with priority and Weibull distribution for failure and repair..... Ashish Kumar, Monika Saini and Kuntal Devi	30
Chapter 5	Phase-I design structure of Bayesian variance chart..... Aamir Saghir	40
Chapter 6	Mitigating collinearity in linear regression models using ridge, surrogate and raised estimators..... Diarmuid O'Driscoll and Donald E. Ramirez	53
Chapter 7	Minimax-robust filtering of functionals from periodically correlated random fields..... Iryna Golichenko, Oleksandr Masyutka and Mikhail Moklyachuk	62
Chapter 8	Minimax prediction of random processes with stationary increments from observations with stationary noise..... Maksym Luz and Mikhail Moklyachuk	75
Chapter 9	Minimax-robust filtering problem for stochastic sequences with stationary increments and cointegrated sequences..... Maksym Luz and Mikhail Moklyachuk	92
Chapter 10	SPC methods for time-dependent processes of counts..... Christian H. Weiß	112
Chapter 11	The properties of the geometric-Poisson exponentially weighted moving control chart with estimated parameters..... Aamir Saghir, Zhengyan Lin and Ching-Wen Chen	123
Chapter 12	Response surface designs using the generalized variance inflation factors..... Diarmuid O'Driscoll and Donald E. Ramirez	137

Chapter 13	Sequential Monte Carlo methods for filtering of unobservable components of multidimensional diffusion Markov processes.....	147
	Ellida M. Khazen	
Chapter 14	The iteratile: A new measure of central tendency.....	171
	Blane Hollingsworth	
Chapter 15	On the Bayesianity of minimum risk equivariant estimator for location or scale parameters under a general convex and invariant loss function.....	175
	Amir T. Payandeh Najafabadi	
Chapter 16	A robust unbiased dual to product estimator for population mean through modified maximum likelihood in simple random sampling.....	183
	Sanjay Kumar and Priyanka Chhapparwal	
Chapter 17	Different types of Bernstein operators in inference of Gaussian graphical model.....	195
	Melih Ağraz and Vilda Purutçuoğlu	
Chapter 18	Multilevel multinomial logistic regression model for identifying factors associated with anemia in children 6–59 months in northeastern states of India.....	205
	Sanku Dey and Enayetur Raheem	

Permissions

List of Contributors

Index

A defense against the alleged unreliability of difference scores

David Trafimow^{1*}

*Corresponding author: David Trafimow, Department of Psychology, New Mexico State University, MSC 3452, P.O. Box 30001, Las Cruces, NM 88003-8001, USA

E-mail: dtrafimo@nmsu.edu

Reviewing editor: Hamdi Raïssi, IRMAR-INSA (UEB), France

Abstract: Based on a classical true score theory (classical test theory, CTT) equation, indicating that as the observed correlation between two tests increases, the reliability of the difference scores decreases, researchers have concluded that difference scores are unreliable. But CTT shows that the reliabilities of the two tests and the true correlation between them influence the observed correlation and previous analyses have not taken the true correlation sufficiently into account. In turn, the reliability of difference scores depends on the interaction of the reliabilities of the individual tests and their true correlation when the variances of the tests are equal, and on a more complicated interaction between them and the deviation ratio when the variances of the tests are not equal. The upshot is that difference scores likely are more reliable, on more occasions, than researchers have realized. I show how researchers can predict what the reliability of the difference scores is likely to be, to aid in deciding whether to carry through one's planned use of difference scores.

Subjects: Behavioral Sciences; Communication Studies; Development Studies; Education; Social Sciences; Sports and Leisure

Keywords: reliability of difference scores; classical true score theory; classical test theory; true correlation; deviation ratio

ABOUT THE AUTHOR

David Trafimow is a distinguished achievement professor of psychology at New Mexico State University, a fellow of the Association for Psychological Science, executive editor of the *Journal of General Psychology*, and also for Basic and Applied Social Psychology. He received his PhD in psychology from the University of Illinois at Urbana-Champaign in 1993. His current research interests include attribution, attitudes, cross-cultural research, ethics, morality, methodology and philosophy of science, potential performance theory, and descriptive and inferential statistical analyses.

PUBLIC INTEREST STATEMENT

There are many important issues that depend on difference scores for their elucidation. For example, to find out if an intervention worked, participants might be asked to complete a relevant measure before and after the intervention, with the hope that scores will be more favorable after it than before it. Or participants might complete two or more different measures, such as when high school students take tests in different subject areas to determine relative strengths and weaknesses. Whether the difference of interest pertains to change on a measure, or on the difference between different measures, difference scores are important. However, because of a prevalent belief that difference scores are not reliable, researchers have been discouraged from using them. The present work challenges the blanket assertion that difference scores are inherently unreliable and demonstrates the conditions under which different scores are more reliable or less reliable.

1. Introduction

Basic and applied researchers in education and psychology often have reason to be interested in the differences between people's scores on two tests. The two tests might have the same items, as in a pretest–posttest design or in other types of within-participants designs. Alternatively, the tests might have different items, such as when a school psychologist identifies that a particular student has a relative weakness in one domain compared to a relative strength in another domain. In either case, difference scores—the difference between scores on two tests—are of interest. But for many decades, psychometric experts have warned that difference scores are not reliable. Bereiter (1963, p. 3) stated the consequence of this warning with admirable clarity:

Although it is commonplace for research to be stymied by some difficulty in experimental methodology, there are really not many instances in the behavioral sciences of promising questions going unresearched because of deficiencies in statistical methodology. Questions dealing with psychological change may well constitute the most important exceptions. It is only in relation to such questions that the writer has ever heard colleagues admit to having abandoned major research objectives solely because the statistical problems seemed to be insurmountable.

Bereiter's observation has retained its force, despite advances in the intervening decades. For example, Chiou and Spreng (1996) noted that difference scores continue to be criticized. Even very recently, Thomas and Zumbo (2012) stated, "There is such doubt in research practice about the reliability of difference scores that granting agencies, journal editors, reviewers, and committees of graduate students' theses have been known to deplore their use" (p. 37). Given the expressed negativity throughout the decades, my goal is to investigate the alleged unreliability of difference scores in a better way than has hitherto been done. The stakes are large. If previous criticisms hold up, this would constitute an important reason to avoid doing research that depends on difference scores. In contrast, if previous criticisms do not hold up, an enormous impediment to research can be removed. I will show that the reliability of difference scores depends on the interaction between the true correlation between the tests, individual test reliabilities, and the deviation ratio, which is a ratio of test variances. A close study of this three-way interaction suggests that the blanket conclusion that difference scores are inherently unreliable is too pessimistic.

1.1. The basis of the criticism

The criticism that difference scores are unreliable comes directly from a theorem based on classical true score theory or classical test theory (CTT) that can be found in many places (e.g. Cronbach, 1990; Lord, 1963; see Gulliksen, 1987 for an excellent review of CTT theorems). CTT features the notion of reliability, which is usually defined as true score variance divided by observed score variance. Equation 1 provides the general formula that expresses the reliability of difference scores ($\rho_{dd'}$) in terms of the reliabilities of the two tests ($\rho_{XX'}$ and $\rho_{YY'}$), the variances of the two tests (σ_X^2 and σ_Y^2), and the correlation between observed scores on the two tests (ρ_{XY}) (Designating a person's score on one test as X_i and on the other as Y_i implies that the difference score is $d_i = Y_i - X_i$).

$$\rho_{dd'} = \frac{\sigma_X^2 \rho_{XX'} + \sigma_Y^2 \rho_{YY'} - 2\rho_{XY} \sigma_X \sigma_Y}{\sigma_X^2 + \sigma_Y^2 - 2\rho_{XY} \sigma_X \sigma_Y} \quad (1)$$

If the variances of the two tests are equal, Equation 1 reduces to Equation 2 below.

$$\rho_{dd'} = \frac{\rho_{XX'} + \rho_{YY'} - 2\rho_{XY}}{2 - 2\rho_{XY}} \quad (2)$$

Under Equation 2, it is easy to see that as the correlation between the two tests increases, keeping the reliabilities of the two tests constant, the reliability of difference scores decreases. In the case of a pretest–posttest design, for example, one would expect the correlation between pretest and posttest

to be substantial, and Equation 2 seems to indicate that this should result in unreliable difference scores. Historically, researchers with psychometric expertise have made similar arguments to militate against the reliability of difference scores.

However, despite widespread criticism, some have defended difference scores. For example, one problem with unreliable difference scores is that such unreliability might render it difficult to obtain statistically significant findings. After reviewing the literature on this criticism, Thomas and Zumbo (2012) suggested that this sort of unreliability is not crucial for inferential statistical procedures, such as t-tests and ANOVAs (also see Gaito & Wiley, 1963). Possible limitations of this defense might be that this alleged lack of importance has not been demonstrated conclusively. Besides that, difference scores are used for many purposes other than significance tests. Suppose a researcher wishes to correlate difference scores with another variable. Or suppose a researcher wishes to base interventions on difference scores. In cases such as these, the unreliability of difference scores, if it holds up, would constitute an important problem, even if it were admitted that inferential statistics are not importantly affected.

Another defense involves a switch of focus from Equations 2 to 1. The idea here is that although difference scores are unreliable under the assumption of equal variances, if the variances are unequal, difference scores do not have to be unreliable. Chiou and Spreng (1996) provided some examples of how the unreliability of difference scores is substantially mitigated by having sufficiently unequal variances of the two tests. Thus, their analysis suggested that a potential answer to the problem is to use tests with unequal variances. A limitation, however, is that it might be difficult to arrange matters in this way, particularly if the same test is given twice.

Another defense traces back to Bereiter (1963), who pointed out that as the reliabilities of the individual tests increase, so does the reliability of the difference scores. This can be seen easily in Equation 2.

In his insightful analysis, Bereiter (1963) pointed out that increasing the reliabilities of the individual tests has contradictory effects on the reliability of the difference scores. To understand these contradictory effects, it is necessary to recall that as the reliability of the tests increase, so does the observed correlation. Therefore, on the one hand, increasing the reliabilities of the individual tests decreases the reliability of the difference scores by increasing the correlation between observed scores over what it otherwise would be. But, on the other hand, as Equation 2 shows, increasing the reliabilities of the individual tests increases the reliability of the difference scores.

Unfortunately, Bereiter did not proceed further with his analysis, and researchers have not attended to these contradictory effects. Given that Equation 1, Equation 2, and the contradictory effects of varying the reliabilities of the individual tests, all depend on CTT, this lack of attention is surprising. My goal is to take the contradictory effects seriously to investigate their consequences for difference score reliability. In the analyses that follow, I intend to return to the roots of CTT to discern exactly the extent of the deleterious vs. beneficial effects of increasing the reliabilities of the individual tests on the reliability of the difference scores. In so doing, I am not committing to an actual belief or disbelief in CTT. Rather, because the criticisms stem from CTT, an investigation from the point of view of CTT directly addresses them. It also seems worthwhile to note that although more advanced theories have been proposed and favored by many researchers, such as generalized ability theory and item response theory, these reduce down to CTT in their simplest incarnations (see Cronbach, Gleser, Nanda, & Rajaratnam, 1972; Hulin, Drasgow, & Parsons, 1983, respectively).

1.2. The importance of CTT's true correlations

Consider the attenuation formula from CTT, written below as Equation 3, where $\rho_{T_x T_y}$ represents the true correlation between two tests (i.e. the correlation between true scores).

$$\rho_{XY} = \rho_{T_x T_y} \sqrt{\rho_{XX'} \rho_{YY'}} \quad (3)$$

Equation 3 makes clear the importance of the reliabilities of the two tests. As the reliabilities of the tests increase, the observed correlation becomes increasingly similar to the true correlation. In the limiting case of perfectly reliable tests, the observed correlation equals the true correlation. When the tests are not perfectly reliable, the observed correlation is less than the true correlation. Thus, Equation 3 clarifies Bereiter's (1963) statement that increasing the reliabilities of the tests increases the correlation between observed scores, and hence decreases the reliability of the difference scores. To reiterate, however, Equation 2 shows a contradictory effect whereby increasing the reliability of the individual tests increases the reliability of the difference scores.

Let us again consider Equation 3, and the mathematical fact that the observed correlation depends on two things: the reliabilities of the individual tests and the true correlation coefficient. I have discussed the reliabilities of the individual tests above, but what about the true correlation coefficient? Equation 3 shows that, in general, as the true correlation increases, so does the observed correlation. An exception would be if the reliability of one or both of the individual tests equals zero, in which case, the observed correlation also would equal zero, regardless of the true correlation. Well, then, going back to Equation 2, we see that because increasing the true correlation tends to increase the observed correlation, and because increasing the observed correlation decreases the reliability of the difference scores, it follows that the reliability of the difference scores depends partly on the size of the true correlation. Specifically, the larger the true correlation, keeping the reliabilities of the individual tests constant, the less reliable the difference scores.

It is interesting that although the argument that difference scores have low reliability has its origin in CTT, and that CTT depends on the notion of a true correlation (which, in turn, depends on the notion of a true score), nobody to my knowledge has actually included true correlations in their investigation of the reliability of difference scores. In addition, my foregoing comments implicate the importance of the reliabilities of the individual tests as having complicated effects on the reliability of difference scores. Therefore, one of my main goals is to investigate how the true correlation interacts with the reliabilities of the two tests to determine the reliability of difference scores. It is not necessary to derive equations that are radically different from existing ones; rather, it is sufficient to express existing CTT equations in a form that includes true correlations and reliabilities of individual tests.

In addition, there is no law stating that the variances of the two tests have to be equal. As Chiou and Spreng (1996) suggested, the sizes of variances between the two tests also matter. Possibly, the reliability of difference scores might be influenced by the interaction between the variances of the two tests, their reliabilities, and the true correlation. Therefore, my second goal is to investigate this potential triple interaction.

2. Analyses

I will present two types of analyses. First, I will assume equal variances and explore the reliability of difference scores as a function of the interaction between the reliabilities of the individual tests and the true correlation. Second, I will take into account the possibility of unequal variances and explore the reliability of difference scores as a function of three variables. These are the reliabilities of the individual tests, the true correlation, and the ratio of variances or standard deviations between the two tests.

2.1. Equal variances

To perform the analyses of interest, it is necessary to render the reliability of the difference scores in Equation 2 as a function of the reliabilities of the individual tests and the true correlation. At present, however, Equation 2 includes the observed correlation rather than the true correlation. Because the observed correlation is influenced, in part, by the reliabilities of the two tests, there is no way to vary these reliabilities and the observed correlation independently of each other. This inconvenience can be remedied simply by substituting Equation 3 into Equation 2. The result is Equation 4 below that features the true correlation, as opposed to the observed correlation.

$$\rho_{dd'} = \frac{\rho_{XX'} + \rho_{YY'} - 2\rho_{T_x T_y} \sqrt{\rho_{XX'} \rho_{YY'}}}{2 - 2\rho_{T_x T_y} \sqrt{\rho_{XX'} \rho_{YY'}}} \quad (4)$$

Based on Equation 4, Figure 1 provides an illustration of how the reliabilities of the two tests interact with the true correlation to determine the reliability of the difference scores. In Figure 1, the reliabilities of the two tests were set equal to each other, and allowed to range from 0 to 1 along the horizontal axis. In addition, the curves represent the cases where the true correlation equals .1 (top curve), .3 (second curve), .5 (third curve), .7 (fourth curve), and .9 (bottom curve). Figure 1 illustrates three effects. Most obviously, as the true correlation increases, the reliability of the difference scores (along the vertical axis) decreases. There also is an effect of the reliabilities of the tests on the reliability of the difference scores. Figure 1 illustrates the totality of the contradictory effects I discussed earlier whereby increasing the reliabilities of the tests has both a deleterious effect on the reliability of difference scores (by increasing the true correlation) and a beneficial effect on the reliability of difference scores (by direct entry into Equation 2). The upward trend of all five curves in Figure 1, as the reliabilities of the individual tests increase, shows that the net effect of the contradictory forces is positive for the reliability of the difference scores. That is, as the reliabilities of the individual tests increase, so does the reliability of the difference scores.

Finally, the interaction between the true correlation and the reliabilities of the individual tests deserves observation. As the true correlation increases, the effect of increasing the reliabilities of the individual tests on the reliability of the difference scores becomes increasingly nonlinear. For example, when the true correlation is .9, the reliabilities of the individual tests have to be quite impressive to result in a respectable reliability of the difference scores.

Well, then, are difference scores unreliable, as we so often have been told? Figure 1 shows that the answer is, "It depends." If the true correlation is large (e.g. .9), then it takes extremely impressive reliabilities of the individual tests to result in a reasonably sized reliability of difference scores; the shape of the bottom curve is a real problem. As an example, when the true correlation is .9, even if the reliabilities of both individual tests also equal .9, the reliability of the difference scores is an unimpressive .47. The reliabilities of both individual tests must be ratcheted up to .96 for the reliability of the difference scores to exceed an arbitrary cut-off of .7.

On the other hand, the problem is substantially mitigated if the true correlation is at lesser levels. For example, for the reliability of the difference scores to pass the cut-off of .7 when the true correlation is .7, it is only necessary for the reliabilities of the two tests to exceed .89—a difficult but far from impossible requirement. And as the true correlation decreases further, it is tolerable to have lesser reliabilities of the individual tests.

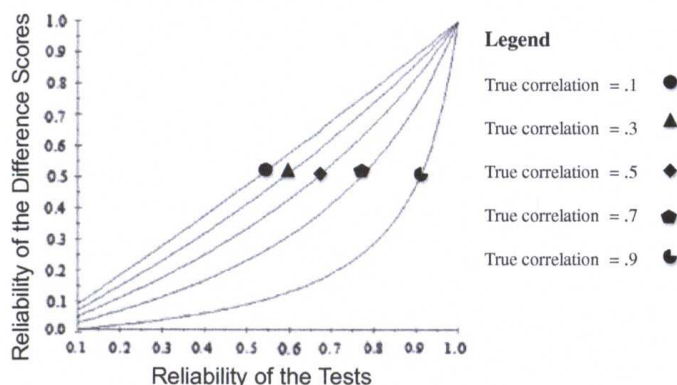


Figure 1. The reliability of difference scores is presented as a function of the reliability of the individual tests and the true correlation.

2.2. Unequal variances

Before conducting the analyses, it is again necessary to obtain an equation based on the true correlation as opposed to the observed correlation. In addition, it is necessary to include a ratio of standard deviations or variances (I will use standard deviations). The net effect of these considerations is to render a slightly more complicated derivation than in the previous section.

To commence, I define the deviation ratio (a) as the standard deviation of test X scores divided by the standard deviation of test Y scores, as is indicated in Equation 5 below. And Equation 6 is simply a rearrangement of Equation 5.

$$a = \frac{\sigma_X}{\sigma_Y} \quad (5)$$

$$\sigma_X = a\sigma_Y \quad (6)$$

Substituting Equation 6 into Equation 1 renders Equation 7.

$$\rho_{dd'} = \frac{a^2 \sigma_Y^2 \rho_{XX'} + \sigma_Y^2 \rho_{YY'} - 2\rho_{XY} a \sigma_Y^2}{a^2 \sigma_Y^2 + \sigma_Y^2 - 2\rho_{XY} a \sigma_Y^2} \quad (7)$$

Cancelling σ_Y^2 renders Equation 8.

$$\rho_{dd'} = \frac{a^2 \rho_{XX'} + \rho_{YY'} - 2\rho_{XY} a}{a^2 + 1 - 2\rho_{XY} a} \quad (8)$$

Although Equation 8 is nice to have, it is not sufficient in one respect. Specifically, it still retains the observed correlation, whereas we would like it to include the true correlation. The solution is to substitute Equation 3 into Equation 8 to render Equation 9.

$$\rho_{dd'} = \frac{a^2 \rho_{XX'} + \rho_{YY'} - 2\rho_{T_x T_y} \sqrt{\rho_{XX'} \rho_{YY'}} a}{a^2 + 1 - 2\rho_{T_x T_y} \sqrt{\rho_{XX'} \rho_{YY'}} a} \quad (9)$$

Figure 2 illustrates how Equation 9 manifests where Panels A, B, and C represent when the true correlation is .1, .5, and .9, respectively; the reliabilities of the two tests are set equal to each other and vary from 0 to 1 (as in Figure 1); and where the deviation ratio is 4 (top curve), 3, (second curve), 2 (third curve), or 1 (bottom curve). As in Figure 1, the reliability of the difference scores is indicated on the vertical axis.

Figure 2 shows that there are three trends that replicate what we saw in Figure 1. Comparing across panels in Figure 2 shows that as the true correlation increases, the reliability of the difference scores decreases. Also, the upward curves in all three panels replicate that as the reliabilities of the individual tests increase, so does the reliability of the difference scores. Finally, these two effects qualify each other, as we saw earlier.

But some new observations also emerge. First, there is a general trend, which can be seen most easily in Panel C of Figure 2, that as the deviation ratio increases, so does the reliability of the difference scores. This effect supports the argument by Chiou and Spreng (1996) in favor of high deviation ratios if one wishes to increase the reliability of difference scores. Second, there is an interesting interaction between the deviation ratio and the reliabilities of the individual tests that is particularly easy to see in Panel C of Figure 2. Harkening back to Figure 1, recall the unfavorable shape of the curve when the true correlation was .9, which is illustrated again by the bottom curve in Panel C of Figure 2. But as the deviation ratio increases, note that the shape (and altitude) of the curve becomes increasingly less unfavorable. Thus, even in the worst-case scenario of a large true correlation, it is possible to obtain a respectable number for the reliability of the difference scores so long as the deviation ratio is high. Put