Journal of Modern Applied Statistical Methods

Volume 6 | Issue 1 Article 4

5-1-2007

Ordinal Versions of Coefficients Alpha and Theta for Likert Rating Scales

Bruno D. Zumbo
University of British Columbia, bruno.zumbo@ubc.ca

Anne M. Gadermann
University of British Columbia

Cornelia Zeisser
University of British Columbia

Part of the Applied Statistics Commons, Social and Behavioral Sciences Commons, and the Statistical Theory Commons

Recommended Citation
Zumbo, Bruno D.; Gadermann, Anne M.; and Zeisser, Cornelia (2007) "Ordinal Versions of Coefficients Alpha and Theta for Likert Rating Scales," *Journal of Modern Applied Statistical Methods*: Vol. 6 : Iss. 1 , Article 4.
DOI: 10.22237/jmasm/1177992180

Journal of Modern Applied Statistical Methods
May, 2007, Vol. 6, No. 1, 21-29
Copyright © 2007 JMASM, Inc.
1538 – 9472/07/$95.00

Ordinal Versions of Coefficients Alpha and Theta for Likert Rating Scales



Bruno D. Zumbo   Anne M. Gadermann   Cornelia Zeisser
University of British Columbia

Two new reliability indices, ordinal coefficient alpha and ordinal coefficient theta, are introduced. A simulation study was conducted in order to compare the new ordinal reliability estimates to each other and to coefficient alpha with Likert data. Results indicate that ordinal coefficients alpha and theta are consistently suitable estimates of the theoretical reliability, regardless of the magnitude of the theoretical reliability, the number of scale points, and the skewness of the scale point distributions. In contrast, coefficient alpha is in general a negatively biased estimate of reliability. The use of ordinal coefficients alpha and theta as alternatives to coefficient alpha when estimating the reliability based on Likert response items are recommended. The choice between the two ordinal coefficients depends on whether one is assuming a factor analysis model (ordinal coefficient alpha) or a principal components analysis model (ordinal coefficient theta).

Key words: Internal consistency, reliability, coefficient alpha, coefficient theta.

Introduction

Coefficient alpha is the most widely used index of reliability in the social sciences (Zumbo & Rupp, 2004). There is, however, ongoing debate about the use of alpha for Likert type rating response scales because alpha assumes that the item responses are continuous. Using Likert type response scales, it has been demonstrated that the magnitude of coefficient alpha can be spuriously deflated with less than five scale points. However, reliability was found to level off beyond six points (Gelin, Beasley, & Zumbo, 2003). Likert type data are commonly utilized in psychological and educational settings to measure unobserved continuous variables. Yet, lack of clarity still prevails regarding the statistical impact of various numbers of response scale points on outcomes that are based on a continuous concept. Of course, a special case of coefficient alpha is KR-20, which is computed from binary data.

One can compute estimates of reliability from correlation (or, more generally, covariance) matrices. For example, the Pearson correlation matrix is commonly used to compute coefficient alpha. An important assumption for the use of the Pearson correlation matrix is the assumption of continuity. If this assumption is violated, the

Bruno D. Zumbo is Professor, Measurement, Evaluation, and Research Methodology (MERM) Program, and Department of Statistics and the Institute of Applied Mathematics at the University of British Columbia (UBC), Canada. Email address: bruno.zumbo@ubc.ca. Anne M. Gadermann and Cornelia Zeisser are completing their Ph.D.'s in the MERM Program at UBC. The authors would like to thank Professor Ed Diener for providing the data set that was part of the International College Student Data.

Pearson correlation matrix may be distorted (Rupp, Koh, & Zumbo, 2003). If the data are ordinal, the correlation matrix of choice is the polychoric correlation matrix, which estimates the linear relationship for two unobserved continuous variables given only observed ordinal data (Flora & Curran, 2004). Hence, for Likert type scales it may be useful to investigate reliability estimates based on the polychoric correlation matrix, thereby taking into account the ordinal nature of the data. A special case of the polychoric correlation matrix is the tetrachoric correlation matrix for binary data.

Rationale and theoretical framework

Coefficient alpha is used as a default for estimating the internal consistency based on the Pearson correlation matrix in widely available software packages such as SPSS and SAS; however, this is done ignoring the Likert response format of the items at hand. The purpose of this article was to introduce two new reliability indices, ordinal coefficient alpha and ordinal coefficient theta, and test their appropriateness as estimates of internal consistency for items with Likert response formats.

Considering only a Pearson correlation matrix and a factor analysis model, McDonald (1985, p. 217) describes how one can compute coefficient alpha from a factor analysis model. For a composite score based on p items coefficient alpha can be computed as

$$\alpha = \frac{p}{p-1}\left[\frac{p(\bar{f})^2 - \bar{f}^2}{p(\bar{f})^2 + \bar{u}^2}\right], \qquad (1)$$

where $\bar{f}$ is the average of the p factor loadings, $\bar{f}^2$ is the average of the squares of the p factor loadings, and $\bar{u}^2$ is the average of the p uniquenesses.

Armor (1974) introduced a reliability estimate, coefficient theta, which was developed to account for multidimensionality in a scale and is based on a principal components model. Coefficient theta for the single factor solution is computed with the following equation (Armor, p. 28):

$$\Theta = [p/(p-1)]*[1-(1/\lambda_1)], \qquad (2)$$

where the only new symbol λ_1 denotes the largest eigenvalue from the principal component analysis of the correlation matrix of the items involved in the composite.

Ordinal coefficient alpha and ordinal coefficient theta are computed by applying equations (1) or (2), respectively, to the polychoric correlation matrix. These reliability estimates are ordinal in the sense that they take into account the ordinal nature of the Likert response data.

In the following, a computer simulation study is reported that investigated the population estimation bias of ordinal coefficients alpha and theta for response scales ranging from two to seven points, with symmetric as well as skewed Likert response distributions, and theoretical reliabilities of .4, .6, .8, and .9. Next, ordinal coefficients alpha and theta were demonstrated with real data. The article closes with discussion of the findings and recommendations.

Methodology

Simulation study

Simulation data were generated to reflect the conditions of theoretical alpha (.4, .6, .8, and .9) as well as skewness conditions of zero and –2 of the item responses. The fundamental equations of factor analysis were used to create a population covariance matrix; this covariance matrix was then used to generate normally distributed item responses. That is, item response data were generated using a factor analysis model. As indicated by Jöreskog (1971) and Henrysson and Wedman (1972), the decomposition of an observed score X into a true score and an error score in classical test theory can be generalized to a factor analytic model with one common factor. The formula X = T + E can be defined as

$$X_i = f_i \xi + u_i \qquad i = 1, 2, \ldots, p, \qquad (3)$$

where X_i denotes the observed scores, f_i denotes the factor loadings, ξ the common factor that can also be regarded as true score, u_i, uniqueness of variables, denotes the error scores, and i indexes the items running from one to p. In a factor model, the reliability of the observed score can be obtained by summing all true score variances and covariances in the matrix and then by dividing this sum by the total variance (Reuterberg & Gustafsson, 1992). Novick and Lewis (1967) showed that coefficient alpha yields an unbiased estimate of reliability when the loadings of each variable on the common factor are equal. The formula for the reliability of a composite score is

$$\rho_{xx} = \frac{\left(\sum_{i=1}^{p} f_i\right)^2}{\left(\sum_{i=1}^{p} f_i\right)^2 + \sum_{i=1}^{p} \mathrm{var}(e)_{ii}}, \qquad (4)$$

where $\mathrm{var}(e)_{ii}$ denotes the error variance in a factor analytical model and all the other symbols are defined above. To obtain the population reliabilities of .4, .6, .8, and .9, factor loadings of .213 .311, .471, and .625, respectively, were computed using the above formula. Therefore, in summary 14 items with continuous (normally distributed) distributions were generated using one common factor model with equal factor loadings across the 14 items. Fourteen items were chosen because it is a typical scale length in health and educational research (Slocum, 2005).

These (underlying) item response distributions were then transformed into Likert responses by applying the thresholds (for the symmetric as well as skewed item responses) as provided in the Appendix. The number of response options was simulated to range from 2 to 7; by including 2 response options, one is also able to investigate how the new reliability estimates perform in the presence of binary data.

As noted above, the unidimensionality and equal factor loadings provide a strict condition where empirical alpha should equal theoretical alpha. It was confirmed that the simulation methodology worked correctly because the theoretical alpha was obtained when analyzing the continuous data. It should be noted that, given the simulation design, there was no interest in the sample-to-sample variability in the estimates but rather the focus was on accuracy (bias) of the estimates. Therefore, population analogues of the empirical reliability estimates were computed with a sample size of 10,000 simulees in each cell of our simulation design.

The following steps were followed for the analysis. The data were simulated and coefficient alpha was obtained using SPSS. The simulated data were then read into PRELIS. In order to compute ordinal coefficient alpha the polychoric correlation matrix was factor analysed using the MINRES procedure. The resulting factor loadings and uniquenesses were then used to compute ordinal coefficient alpha. In addition, the eigenvalues of the polychoric correlation matrix among the items were computed from the principal components analysis and used to compute ordinal theta.

Results

The reliability estimates for the simulated data are displayed in Tables 1 to 4, for theoretical reliability of 0.4 to 0.9, respectively. As can be seen from these tables, coefficient alpha is consistently a negatively biased estimate of the theoretical reliability. Note that in the case of equal factor loadings and unidimensionality coefficient alpha should equal the reliability; that is, it is not a lower bound. The negative bias of alpha was even more evident under the condition of negative skewness; for example, in the case of theoretical reliability of .6 and 3 response options alpha underestimates the theoretical reliability by .175. These results highlight that coefficient alpha, likewise KR-20 for binary data, gives one a downwardly biased estimate of the theoretical reliability with Likert data. With regard to the number of scale points our finding is a replication of the finding of Gelin, Beasley, and Zumbo (2003) that showed that alpha computed from Likert item response data approaches its theoretical value as the number of scale points increases, and levels off at about 6 scale points.

Table 1. Reliability Estimates for Theoretical Alpha of .4 (all factor loadings are .213)

	Skewness					
# of response options	0			-2		
	Alpha	Ordinal Alpha	Ordinal Theta	Alpha	Ordinal Alpha	Ordinal Theta
2	.288	.393	.395	.211	.389	.391
3	.328	.401	.400	.233	.383	.387
4	.356	.399	.400	.258	.379	.382
5	.377	.406	.408	.255	.384	.387
6	.378	.398	.400	.291	.382	.387
7	.386	.401	.404	.303	.391	.391

Table 2. Reliability Estimates for Theoretical Alpha of .6 (all factor loadings are .311)

	Skewness					
# of response options	0			-2		
	Alpha	Ordinal Alpha	Ordinal Theta	Alpha	Ordinal Alpha	Ordinal Theta
2	.488	.608	.609	.379	.596	.596
3	.527	.609	.609	.425	.603	.603
4	.561	.608	.609	.421	.598	.600
5	.576	.607	.609	.452	.597	.598
6	.587	.609	.609	.459	.599	.600
7	.589	.606	.607	.477	.598	.598

Table 3. Reliability Estimates for Theoretical Alpha of .8 (all factor loadings are .471)

	Skewness					
# of response options	0			-2		
	Alpha	Ordinal Alpha	Ordinal Theta	Alpha	Ordinal Alpha	Ordinal Theta
2	.702	.802	.802	.629	.806	.806
3	.732	.799	.799	.655	.798	.798
4	.762	.800	.800	.668	.803	.804
5	.773	.798	.798	.689	.800	.800
6	.783	.801	.801	.709	.803	.804
7	.785	.798	.798	.725	.804	.804

Table 4. Reliability Estimates for Theoretical Alpha of .9 (all factor loadings are .625)

	Skewness					
# of response options	0			-2		
	Alpha	Ordinal Alpha	Ordinal Theta	Alpha	Ordinal Alpha	Ordinal Theta
2	.826	.897	.897	.778	.899	.899
3	.849	.899	.899	.806	.899	.899
4	.872	.897	.897	.810	.898	.898
5	.882	.897	.897	.830	.899	.899
6	.886	.898	.898	.840	.900	.900
7	.891	.898	.898	.852	.900	.900

In contrast to coefficient alpha, ordinal coefficients alpha and theta were consistently found to be suitable estimates of reliability regardless of the magnitude of the theoretical reliability and number of scale points. In addition, it should be noted that the skewness of the item response distribution affects coefficient alpha, whereas ordinal coefficients alpha and theta remain unaffected by skewness. Specifically, ordinal coefficients alpha and theta are still suitable in the presence of skewed data; however, coefficient alpha becomes more biased with skewness. A comparison between the two ordinal estimates shows that they are almost exactly identical. In the following, ordinal coefficients alpha and theta are compared to coefficient alpha in the context of real data.

Real data examples

The real data examples are based on two samples. The data of the first sample was collected between 1995-1996 by Professor Ed Diener and his collaborators worldwide with College students from 42 nations. The following scales were used. The Satisfaction with Life Scale (SWLS) (Diener, Emmons, Larsen, & Griffin, 1985) is a 5-item instrument designed to measure global cognitive judgments of one's life using a 7-point Likert-type scale ranging from 1 (strongly disagree) to 7 (strongly agree). Diener’s Affect Balance Scale (Veenhoven, 2004) is an 8-item instrument designed to measure positive and negative affect (each being one dimension with four items; this was supported in the present study by a principal component analysis of the polychoric correlation matrix) using a 7-point Likert-type response scale ranging from 1 (never) to 7 (always). The Positive and Negative Affect Schedule (PANAS) (Watson, Clark, & Tellegen, 1988) consists of two 10-item scales with a 5-point Likert type response scale, ranging from 1 (very slightly) to 5 (extremely). In the present study only the Positive Affect Schedule (PAS) was used. Sample sizes for these questionnaires ranged between 6958 and 7014.

The data of the second sample was collected in 1993 by the first author at a Canadian university. The Eysenck Personality Questionnaire (EPQ) (Eysenck & Eysenck, 1975) was administered to 922 students. This questionnaire consists of four subscales with a binary response scale with 0 (no) and 1 (yes). For the present study only the neuroticism (23 items) and extraversion (21 items) subscales were used.

Coefficient alpha was computed for the (sub)scales using SPSS. The data were entered into PRELIS to obtain the polychoric correlation matrix to compute ordinal coefficients alpha and theta as described above. The items of the SWLS exhibited a skewness ranging from -.56 to .18, with an average skewness of -.27. The positive items of the affect scale exhibited a skewness ranging from -.06 to .53, with an average skewness of .17. The negative items of the affect scale exhibited a skewness ranging from .90 to 1.27, with an average skewness of 1.04. The items of the PAS exhibited a skewness ranging from -.39 to .05, with an average skewness of –.21. The items of the extraversion scale exhibited a skewness ranging from –3.27 to .56, with an average skewness of –1.02. The items of the neuroticism scale exhibited a skewness of –1.88 to .89, with an average skewness of -.32.

The reliability estimates, coefficient alpha and ordinal coefficients alpha and theta, for the scales are provided in Table 5. Table 5 shows that ordinal coefficients alpha and theta display a larger reliability estimate than coefficient alpha for all scales. However, for the four scales with the 5- and 7-point Likert type response scales, the difference between coefficient alpha and ordinal coefficients alpha and theta is small. In contrast, for the scales with the binary response format the difference between coefficient alpha and ordinal coefficients alpha and theta is more prominent. This is in accordance with the findings of the simulation study, which showed that with increased number of response options, coefficient alpha and the ordinal estimates become closer. Based on the findings from the simulation study, where ordinal coefficients alpha and theta were consistently demonstrated to be more precise estimates, this finding can be interpreted as showing that ordinal coefficients alpha and theta are closer to the theoretical alpha of the scales.

Table 5. Reliability Estimates for Real Data with the SWLS, Positive and Negative Affect, PAS, Extraversion and Neuroticism Scales

Scale	Alpha	Ordinal Alpha	Ordinal Theta
SWLS	.814	.835	.836
Positive Affect	.709	.735	.738
Negative Affect	.667	.684	.686
PAS	.824	.845	.846
Extraversion	.819	.908	.916
Neuroticism	.830	.905	.910

Conclusion

In summary, it was found that coefficient alpha computed from Likert response data results in a negatively biased estimate of the theoretical reliability. Because it is a special case of coefficient alpha, KR-20 also shows this bias when used with binary response data. It should be noted that coefficient alpha (and KR-20) are correlation-based statistics and hence assume continuous data. What is noteworthy about the coefficient alpha findings is that the measurement model used in the simulation involves all of the assumptions of coefficient alpha, so that alpha would equal the conceptual/theoretical reliability. However, it was found that coefficient alpha is rather drastically affected by Likert data – e.g., imagine a 14 item scale comprised of a 3-point Likert response format with a skewness of –2; the resulting coefficient is .66 when the theoretical reliability is .80.

Ordinal coefficients alpha and theta, on the other hand, were found to be suitable alternatives to coefficient alpha when a researcher is confronted with having to compute a reliability estimate with Likert response data. It should be noted that with advances in statistical software, these ordinal coefficients are easy to calculate using the newly developed and freely available software FACTOR (Lorenzo-Seva & Ferrando, in press) or with widely available software such as PRELIS that provide polychoric correlation matrices. Depending on how they are computed, polychoric correlation matrices can be non positive-definite – i.e., pairwise estimation of the elements of a polychoric correlation matrix is problematic because it can lead to non positive-definite correlation matrices; as opposed to estimating all the correlations in the matrix simultaneously.

The matter of how to estimate polychoric correlation matrices to avoid non positive-definiteness is an open area of research that needs further study but in the meantime a solution to this potential problem, when a non positive-definite matrix is found, is to use software, e.g., EQS, that estimates the polychoric correlations in a manner that reduces the concern for non positive-definite matrices.

In the present study, ordinal coefficients alpha and theta performed equally well. A direction for future research would be to compare ordinal coefficients alpha and theta in the presence of multidimensional items because theta was originally developed to account for multidimensionality in an item set.

Based on the present study, the following recommendations are presented:

1. Use either ordinal coefficient alpha or ordinal coefficient theta to correct for the negative bias in coefficient alpha, and of course KR-20, due to Likert or binary response data.

2. In terms of which of these two ordinal reliability coefficients to use, the decision should be based on whether one is assuming a factor analysis model (ordinal coefficient alpha) or a principal components model (ordinal coefficient theta). For a distinction between principal components analysis and factor analysis the reader is referred to Fabrigar, Wegener, MacCallum, & Strahan (1999) or Zumbo (2007).

It should be noted that the strategy of using the polychoric correlation could be applied to any reliability estimate that can be computed from a correlation matrix. For example, although it is not described herein, one would have an ordinal version of the McDonald's coefficient omega, yet another reliability estimate, by applying the equation described by McDonald (1985, p. 217), or of Revelle's reliability coefficient beta (Zinbarg, Revelle, Yovel, & Li, 2005). Future research should explore these other coefficients as well.

References

Armor, D. J. (1974). Theta reliability and factor scaling. In H. Costner (Ed.), *Sociological methodology* (pp. 17-50). San Francisco: Jossey-Bass.

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, *49*, 71-75.

Eysenck, H. J., & Eysenck, S. B. G. (1975). *Manual of the Eysenck personality questionnaire*. London: Hodder & Stoughton.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*, 272-299.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9*, 466-491.

Gelin, M. N., Beasley, T. M., & Zumbo, B. D. (2003, April). *What is the impact on scale reliability and exploratory factor analysis of a Pearson correlation matrix when some respondents are not able to follow the rating scale?* Paper presented at the annual meeting of the American Educational Research Association (AERA) in Chicago, Illinois.

Henrysson, S., & Wedman, I. (1972). Analysis of the inter-item covariance matrix. *Scandinavian Journal of Educational Research, 16*, 25-35.

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika, 36*, 109-133.

Lorenzo-Seva, U., & Ferrando, P. J. (in press). FACTOR: A computer program to fit the exploratory factor analysis model. *Behavior Research Methods.*

McDonald, R. P. (1985). *Factor analysis and related methods*. Hillsdale NJ: Erlbaum.

Novick, M. R., & Lewis, M. R. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika, 32*, 1-13.

Reuterberg, S. E., & Gustafsson, J. E. (1992). Confirmatory factor analysis and reliability: Testing measurement model assumptions. *Educational and Psychological Measurement, 52*, 795-811.

Rupp, A., Koh, K. & Zumbo, B.D. (2003, April). *What is the impact on exploratory factor analysis results of a polychoric correlation matrix from LISREL/PRELIS and EQS when some respondents are not able to follow the rating scale.* Paper presented at the annual meeting of the American Educational Research Association (AERA) in Chicago, Illinois.

Slocum, S. L. (2005). *Assessing unidimensionality of psychological scales: Using individual and integrative criteria from factor analysis*. Unpublished Doctoral Dissertation, University of British Columbia, Canada.

Veenhoven (2004). World database of happiness. Retrieved September 14 from http://www2.eur.nl/fsw/research/happiness/hap_quer/hqs_fp.htm

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS Scales. *Journal of Personality and Social Psychology, 54*, 1063-1070.

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and McDonald's ω_H: There relations with each other and two alternative conceptualizations of reliability. *Psychometrika, 70,* 123-133.

Zumbo, B. D., & Rupp, A. A. (2004). Responsible modelling of measurement data for appropriate inferences: Important advances in reliability and validity theory. In D. Kaplan (Ed.), *The SAGE Handbook of Quantitative Methodology for the Social Sciences* (pp. 73-92). Thousand Oaks, CA: Sage Press.

Zumbo, B.D. (2007). Validity: Foundational Issues and Statistical Methodology. In C.R. Rao and S. Sinharay (Eds.) *Handbook of Statistics, Vol. 26: Psychometrics* (pp. 45-79). Elsevier Science B.V.: The Netherlands.

Appendix: Thresholds for Symmetric and Skewed Likert Responses

Thresholds for symmetric scale point distribution:

1. Two-point scale: (Lowest thru 0=1) (ELSE=2)

2. Three-point scale: (Lowest thru -1=1) (-.9999 thru 1=2) (ELSE=3)

3. Four-point scale: (Lowest thru -1.5=1) (-1.4999 thru 0=2) (0.0001 thru 1.5=3) (ELSE=4)

4. Five-point scale: (Lowest thru -1.8=1) (-1.7999 thru -0.6=2) (-0.5999 thru 0.6000=3) (0.6001 thru 1.8=4) (ELSE=5)

5. Six-point scale: (Lowest thru -2=1) (-1.9999 thru -1.0=2) (-0.9999 thru 0 =3) (0.0001 thru 1=4) (1.0001 thru 2=5) (ELSE=6)

6. Seven-point scale: (Lowest thru -2.14286=1) (-2.14285 thru -1.28571=2) (-1.28570 thru -0.42857 =3) (-0.42857 thru 0.428571=4) (0.428572 thru 1.28571=5) (1.28571 thru 2.14286=6) (else =7)

Thresholds for scale point distribution with skewness of –2:

1. Two-point scale: (Lowest thru -1.06251930227=1) (ELSE=2)

2. Three-point scale: (Lowest thru 0.9002=3) (0.9003 thru 1.29883663264=2) (ELSE=1)

3. Four-point scale: (Lowest thru 0.8508=4) (0.8509 thru 1.086=3) (1.087 thru 1.2816 =2) (ELSE=1)

4. Five-point scale: (Lowest thru 0.6808=5) (0.6809 thru 1.036=4) (1.037 thru 1.2816 =3) (1.2817 thru 1.6546=2) (ELSE=1)

5. Six-point scale (Lowest thru 0.5008=6) (0.5009 thru 1.036=5) (1.037 thru 1.0816 =4) (1.0817 thru 1.4546=3) (1.4547 thru 1.8002=2) (ELSE=1)

6. Seven-point scale: (Lowest thru 0.4008=7) (0.4009 thru 0.8360=6) (0.8361 thru 1.1816 =5) (1.1817 thru 1.4546=4) (1.4547 thru 1.8002=3) (1.8003 thru 2.1002=2) (ELSE=1)