Correlational analysis of ordinal data: from Pearson's *r* to Bayesian polychoric correlation

Jaehwa Choi · Michelle Peters · Ralph O. Mueller

Abstract Correlational analyses are one of the most popular quantitative methods, yet also one of the mostly frequently misused methods in social and behavioral research, especially when analyzing ordinal data from Likert or other rating scales. Although several correlational analysis options have been developed for ordinal data, there seems to be a lack of didactically written literature illustrating the appropriate use and differences among them. The purpose of this paper is to provide a synthesis of correlational analysis options when analyzing ordinal data. These options span from the traditional methods, such as Pearson's *r*, to more recent developments, such as Bayesian estimation of polychoric correlations. An illustration of these methods utilizing a contemporary dataset is provided.

Keywords Correlation coefficients · Ordinal data analysis · Polychoric correlations · Bayesian analysis

Introduction

Correlational analyses are one of the most popular quantitative methods of analyzing data, yet also one of the mostly frequently misused methods in social and behavioral

J. Choi (✉)
The George Washington University, 2134 G Street NW #106,
Washington, DC 20052, USA
e-mail: jaechoi@gwu.edu

M. Peters
University of Houston, Clear Lake, Houston, TX, USA

R. O. Mueller
The University of Hartford, West Hartford, CT, USA

research (Chen and Popovich 2002). Although various types of correlation indices have been developed over the last century—matching various measurement levels, distributional assumptions, and/or estimation techniques—it still seems common for applied researchers to select inappropriate methods for assessing degrees of association for ordinal data.

There are several correlational analysis options when deciding how to analyze ordinal data. Some argued to treat Likert or similar rating scale data, containing five or more categories, as continuous (Bollen and Barb 1981) and to use Pearson's product-moment correlation coefficient, *r* (Pearson 1957) to analyze such data. Although Pearson's *r* is an appealing choice because it is easy to calculate, interpret, and/or extend to further analyses (e.g., effect size or reliability indices), the validity of Pearson's *r*-based results with ordinal data has been questioned (Chen and Popovich 2002; Poon et al. 2002).

Instead of treating ordinal data as if they were continuous (or metrical), several alternative correlational methods have been developed for ordinal data, such as Spearman's ρ (Kendall and Gibbons 1990), Kendall's τ (Kendall and Gibbons 1990), and polychoric correlations (Olsson 1979). Unlike Pearson's *r*, Spearman's ρ and Kendall's τ are *nonparametric* measures of association and have not been extended to further analyses. Consequently, both are fairly limited as a measure of association. On the other hand, polychoric correlations possess desirable properties similar to Pearson's *r*.

When conducting correlational analyses with ordinal data, researchers must make an appropriate methodological choice in order to yield the most valid (in terms of the extent to which one appropriately accounts for the metric of data) and useful (in terms of the extent to which one can use the measure in further analysis) results. The aim of this

paper is to provide a synthesis of correlational analysis options when analyzing ordinal data, from the traditional (e.g., Pearson's r) to the more recent developments (e.g., polychoric correlation with expected a posteriori estimation). Each reviewed method will be illustrated using a publicly available dataset, *Program for International Student Assessment* (PISA) *2003*.

Measures of association

Pearson's r

The Pearson product–moment correlation coefficient (r) is "probably the most widely used statistic when describing a relationship between two variables" (Chen and Popovich 2002, p. 9). Pearson's r is a measure that assesses the association between two continuous (or metrical) variables. Equation (1) provides a commonly used formula for calculating Pearson's r,

$$r = \frac{\sum \left[\left(\frac{x_i - \bar{x}}{s_X} \right) \left(\frac{y_i - \bar{y}}{s_Y} \right) \right]}{n} \tag{1}$$

where x_i and y_i represent the values of variables X and Y, $\bar{x}$ is the mean of the X values, $\bar{y}$ is the mean of the Y values, s_x is the standard deviation of the X values, s_y is the standard deviation of the Y values, and n represents the number of cases. The values obtained for Pearson's r fall between -1 and 1 (Green and Salkind 2005; Vito and Blankenship 2002). The sign of r represents the direction of the association, while the absolute value of Pearson's r represents the strength of the assumed-to-be *linear* association. The closer r is to -1 or 1, the stronger the linear association or the more interdependent the two variables. Positive values of r represent a positive association between two variables, while negative values of r represent a negative association.

When Pearson's r is used as an inferential statistic, it is generally assumed that a bivariate normal distribution exists for both variables (Green and Salkind 2005). Also, Pearson's r can be used in other statistical applications, such as effect size analysis, reliability analysis, power analysis, utility analysis, or factor analysis (Chen and Popovich 2002). Finally, given that Pearson's r is computed by using the z-scores (e.g., z-score for x_i, $z_i = (x_i - \bar{x})/s_X$; see Eq. 1) for the two variables and z-scores require metrical data, the results obtained from using ordinal-level data would, technically, not be appropriate nor meaningful. There are more appropriate indices of association that applied researchers can chose from to analyze ordinal data, such as Spearman's ρ, Kendall's τ, and polychoric correlations.

In the behavioral and social sciences, it is common to use self-reported questionnaires, often containing Likert or other rating scales to measure participant perceptions or their associated latent (i.e., unobserved, underlying theoretical) constructs. Also, many applied researchers often collect such ordinal data and treat it as though it were metrical. Bollen and Barb (1981) argued that it is acceptable (i.e., negligible amount of attenuation on Pearson's r estimates) to treat ordinal data, with five or more ordinal categories, as if they were metrical. However, Bollen and Barb's study was conducted under a limited simulation design, and the authors could not find evidence of any analytical research conducted on the attenuation of Pearson's r estimates. In our view, the validity of the correlational results using Pearson's r to analyze ordinal data remains questionable. Much research has illustrated the problems resulting from analyzing ordinal data as such (Borgatta and Bohrnstedt 1980; Jöreskog and Moustaki 2001; Kaplan 2000; Mayer 1971; O'Brien 1985; Ware and Benson 1975). On the other hand, there are two commonly used nonparametric measures of association that are recommended for researches who want to analyze ordinal data—Spearman's ρ and Kendall's τ (Chen and Popovich 2002; Kendall and Gibbons 1990).

Spearman's ρ

Spearman's ρ (a special case of Pearson's r) is a *nonparametric* measure that assesses the association using two sets of ranked scores of two variables (Kendall and Gibbons 1990). It is a special case of Pearson's r in that it is Pearson's r on the ranked values of X and Y. Similar to Pearson's r, the range of Spearman's ρ is also from -1 to 1, but Spearman's ρ does not require that the two variables be linearly related and does not assume that the variables are measured on interval or ratio scales. The following equation can be used to calculate Spearman's ρ,

$$\rho = 1 - \frac{6 \sum d_i^2}{n^3 - n} \tag{2}$$

where d_i is the difference between each rank of corresponding values of X and Y, and n is the number of pairs of values, when tied ranks do not exist (Chen and Popovich 2002; Kendall and Gibbons 1990). When the ranks are tied, Pearson's r on the rank values is used instead.

Kendall's τ

Kendall's τ is also a nonparametric measure of association that measures and represents the degree of similarity between two sets of ranked order data by computing the number of concordant pairs and the number of discordant

pairs (Chen and Popovich 2002; Kendall and Gibbons 1990). Kendall's τ can be calculated using equation (3),

$$\tau = \frac{2(P - Q)}{n(n - 1)} \tag{3}$$

where P is the number of concordant pairs (i.e., pairs of bivariate data $\{X_1,Y_1\}$ and $\{X_2,Y_2\}$ such that the sign of $X_2 - X_1$ is equal to the sign of $Y_2 - Y_1$), Q is the number of discordant pairs (i.e., pairs of bivariate data $\{X_1,Y_1\}$ and $\{X_2,Y_2\}$ such that the sign of $X_2 - X_1$ is equal to the opposite sign of $Y_2 - Y_1$), and n is the number of pairs. Note that, if $X_i = X_j$, or $Y_i = Y_j$ or both, the comparison is called a 'tie'. These tied cases are not counted as concordant or discordant. Therefore, when the data contain many tied observations it is recommended to use other types of measures, e.g., the Gamma statistic (Siegel and Castellan 1988). Similar to Pearson's r and Spearman's ρ, the range of values obtained for Kendall's τ is -1 and 1. And the larger the absolute value, the stronger the association or similarity of ranked order between the two variables. Similar to Pearson's r, the sign of Kendall's τ also determines the direction of the association.

As opposed to Pearson's r, further utilities of Spearman's ρ and Kendall's τ are limited in other extended quantitative analyses above and beyond a measure of association. Although Spearman's ρ and Kendall's τ are both measures of association among ordinally scaled variables, they are also different in how they are computed and interpreted. For example, Spearman's ρ can be interpreted as the proportion of variability accounted for with rank data, whereas Kendall's τ can be interpreted as the difference between the probability that concordant pairs versus the probability that discordant pairs. The next section will illustrate a relatively new, versatile, and increasingly popular measure of association with ordinal data known as the polychoric correlation.

Polychoric correlation

Ordinal data do not provide metrical information and, therefore, one needs to analyze frequency information in a contingency table. One analysis option often employed in educational and psychological domains is to hypothesize an *underlying* metrical variable (e.g., academic achievement) that is associated with the *observed* ordinal data (e.g., A–F grading scale). In this strategy of handling ordinal data, each m-category ordinal data, X (e.g., course grade), will be considered to have come from a normally distributed underlying variable, X^* (e.g., academic achievement), with a range from $-\infty$ to $+\infty$. The connection between X and X^* is made through cut points or *thresholds* (e.g., the point distinguishing an A from a B) on the X^* continuum. As can be seen in Fig. 1, the relation between the underlying variable, X^*, and the m-category ordinal data, X, can be established by determining thresholds, v_i.

A correlation between two sets of ordinal data is assessed by extending the univariate case described in the previous paragraph to a bivariate scenario and estimating the *polychoric* correlation between the two respective metrical underlying variables. Two commonly used estimation methods for polychoric correlation are *maximum likelihood* (ML; see, Olsson 1979, for more details on ML estimation of polychoric correlations) and, developed more recently, *expected* a posteriori (EAP; see, Choi et al. in press, for more details of EAP estimation of polychoric correlations). Note that the EAP estimation method uses the same procedure for estimating threshold values, and it also assumes a bivariate underlying normal distribution as ML does. However, EAP estimation follows a Bayesian perspective, treating parameters of interest as random variables that follow a distribution constructed by the observed information from the ordinal data and an a priori knowledge of the correlation (e.g., from research hypotheses or results from previous studies). While the underlying statistical details for EAP and ML estimation are beyond the scope of this paper (consult Bock and Aitkin 1981; Aldrich 1997, for more technical aspects of EAP and ML, respectively), applied researchers should note that, compared to ML, EAP estimation performs better especially for smaller sample sizes (e.g., smaller than 30) and guarantee the parameter estimates that do not depend on a particular choice of starting values (Chen and Choi in press).

Hypothesis testing on measures of association

The hypothesis testing of polychoric correlations with ML estimation is determined by a $(1 - \alpha)$ *confidence* interval: When repeated random samples are drawn from a population infinitely many times, the confidence interval contains the population correlation $(1 - \alpha)*100\%$ of the time. If the confidence interval does not contain zero, a researcher can conclude that the correlation is statistically significant different from zero at an a priori specified α-level. The statistical significance of Pearson's r, Spearman's ρ, and Kendall's τ are interpreted in much the same manner.

In the case of polychoric correlations with EAP estimation, the hypothesis testing result is determined by a *credibility* interval for the correlation. A credibility interval is based on the Bayesian perspective that a population parameter is treated as a random, not a fixed variable. The $(1 - \alpha)$ % *credibility* interval is estimated from the posterior distribution and describes the knowledge (or belief) of the population parameter(s). For example, with $\alpha = .05$, it is 95% credible (or believable) that the interval covers the true underlying population correlation. If the interval

Fig. 1 Underlying Level X^* and Observed Level X

X^* underlying level (e.g., Academic achievement)

ν $\quad$ m-1 thresholds

X observed level (e.g., Course grade)
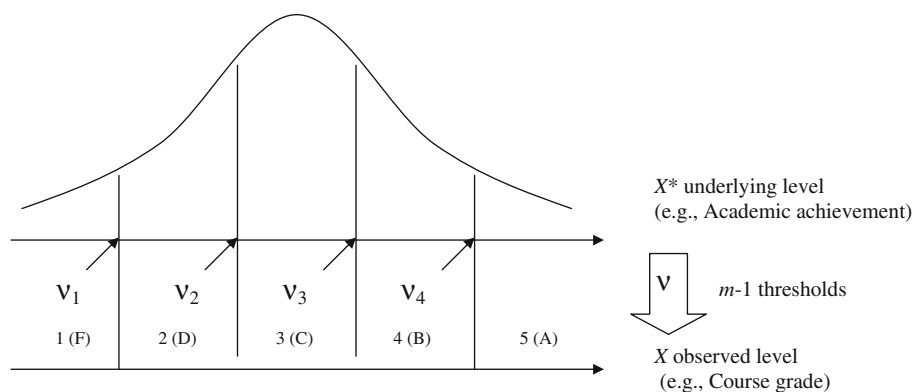
does not contain zero, one would conclude that "I am 95% sure that the underlying population correlation is not zero."

Even though there are similarities between the confidence and the credibility intervals, they are based on quite different perspectives. The interpretation of the credible interval is more straightforward than that of the confidence interval because the confidence interval requires, theoretically, infinitely many samplings for the interpretation. Furthermore, credibility intervals can be applicable for smaller samples than the ML confidence intervals (which are only asymptotically true). In the next section, illustrations of the aforementioned methods are presented.

Illustrations

Using a publicly available dataset, *Program for International Student Assessment* (PISA) *2003*, we examined the differences in analyzing ordinal data, using five correlational analysis options (Pearson's r, Spearman's ρ, Kendall's τ, and polychoric correlations using ML and EAP estimation). PISA is organized by the Organization for Economic Cooperation and Development (OECD), an intergovernmental organization of industrialized countries and is a system of international assessments, which focus on the capabilities of 15-year-olds within three domains of literacy (mathematical, scientific, and reading) (NCES 2008). Starting in 2000, PISA has been administered every 3 years. In 2003, the United States had 274 schools and 5,456 students participate.

Within the PISA framework, there is the Pupil Assessment, a Student Questionnaire, and a School Questionnaire for principals. For purposes of this study, data were analyzed from the Student Questionnaire. Specifically, we analyzed Question 24, "Thinking about what you have learned in school," consisting of four items examining students' attitudes toward school (see Table 1). Participants were asked to respond to each item by endorsing on a 4-point Likert scale ($0 = $ *Strongly Agree*, $1 = $ *Agree*,

Table 1 Survey items for PISA's 2003 index of student's attitudes

Q24: Thinking about what you have learned in school

1. School has done little to prepare me for adult life when I leave school.
2. School has been a waste of time
3. School helped give me confidence to make decisions
4. School has taught me things, which could be useful in a job

$2 = $ *Disagree*, $3 = $ *Strongly Disagree*) the extent to which they agreed with the following four statements.

For purposes of this illustration, we compared a large and small sample scenario to investigate the differences in correlational analysis options over two different sample sizes. All correlational analyses were conducted by a stand-alone Windows software program—BayesPCC (v 3.0.0; Choi et al. 2009) which can be obtained upon request to the first author of this paper. Given that the current version of BayesPCC is incapable of handling missing data, it was necessary to use a complete dataset. Therefore, the large sample scenario ($n = 5,244$) consisted of all students who provided answers on all four items, and the small sample scenario ($n = 75$) included only students, who provided answers on all four items, within the first three schools ($SCHOOLID = 00001$, 00002, and 00003).

Large sample scenario

First, we compared five correlational analysis options for ordinal data in terms of the signs and magnitudes of the coefficient estimates and the pattern of hypothesis test results with a large sample ($n = 5,244$). Table 2 displays the correlation analysis results for the large sample scenario.

As shown in this table, for all of the analyses, the signs of the coefficient estimates were consistent, and the hypothesis test results of all estimates were statistically significant as expected because of the large sample size. In

Table 2 Correlational analysis results across methods using a large sample size example ($n = 5,244$)

PISA Items	Pearson's r			Spearman's ρ			Kendall's τ			PCML			PCEAP		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
2	.407*			.417*			.386*			.502** (.478, .526)			.502** (.465, .525)		
3	−.235*	−.362*		−.252*	−.369*		−.229*	−.343*		−.289** (−.318, −.259)	−.450** (−.477, −.423)		−.288** (−.323, −242)	−.449** (−.485, −.424)	
4	−.260*	−.392*	.467*	−.280*	−.405*	.451*	−.257*	−.382*	.421*	−.335** (−.365, −.306)	−.498** (−.525, −.472)	.570** (.547, .594)	−.335** (−.364, −.303)	−.498** (−.525, −.465)	.570** (.546, .606)

* $p < .05$

** When a 95% interval estimate does not contain the null value, 0

terms of magnitude of estimates, the strongest relationship existed between items 3 and 4 for all analysis options. The ML polychoric correlation estimates displayed the greatest magnitudes, while Kendall's τ produced the smallest estimates among the five options. Furthermore, the magnitude differences between the ML and the EAP polychoric correlation estimates were negligibly small. In summary, the signs and hypothesis test results of the correlational estimates were all the same over the five correlational analysis options for the large sample size scenario. In terms of magnitude, the ML and EAP polychoric correlation estimates were the largest over all of the options.

Small sample scenario

Given the common practice in the behavioral and social sciences to work with small sample sizes (e.g., less than 100), it would be most interesting to compare all five options in a small sample size ($n = 75$). Table 3 displays the correlation coefficient estimates for the small sample scenario.

Although the signs of the coefficient estimates were the same for all options, the patterns of hypothesis test results were not. For Pearson's r, the four coefficient estimates (r_{12}, r_{23}, r_{24}, and r_{34}) were statistically significant, and the magnitude of Pearson's r between items 2 and 3 was the largest among the estimates.

When interpreting the Pearson's r correlation coefficients, it is imperative to note that the validity of this analysis greatly depends on whether this Likert scale data can be treated as though it were metrical (interval) data. Although rating and metrical scale data possess the same property of order as the numbers in the real number system, ordinal data, unlike metrical data, lack the property of equal distances between units or categories (Crocker and Algina 1986). Given that Pearson's r requires metrical data, one must assume that equal distance exists among consecutive response categories of in order to warrant any meaningfulness of the correlation coefficient.

In this case, student participants were asked to endorse the extent to which they agreed (0 = *Strongly Agree*, 1 = *Agree*, 2 = *Disagree*, 3 = *Strongly Disagree*) with

Table 3 Correlational analysis results across methods using a small sample size example ($n = 75$)

PISA Items	Pearson's r			Spearman's ρ			Kendall's τ			PCCML			PCCEAP		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
2	.389*			.425*			.394*			.492** (.273, .710)			.465** (.222, .667)		
3	−.216	−.404*		−.245*	−.378*		−.220*	−.357*		−.270** (−.531, −.01)	−.497** (−.723, −.271)		−.251 (−.485, .020)	−.468** (−.667, −.202)	
4	−.189	−.270*	.288*	−.185	−.217	.213	−.169	−.206	.198	−.243 (−.515, .030)	−.324** (−.597, −.050)	.329** (.063, .595)	−.226 (−.485, .061)	−.301** (−.546, −.020)	.307** (.020, .546)

* $p < .05$

** When a 95% interval estimate does not contain the null value, 0

four items examining their attitudes toward school. In order for participant responses to be considered metrical data, one must assume that (in a psychological continuum) equal distance exists between all participants' responses over the four consecutive categories. In this example, it should be obvious that this is not a trivial assumption. For example, on the third item, if Student A endorsed 3 (*Strongly Disagree*) and Student B endorsed 1 (*Agree*), Pearson's r requires us to assume the Student A disagreed exactly three times more than Student B. Given that participants are not viewed as identically calibrated instruments, it should not be assumed that the distance represented between or the meaning attached to the different categories is the same across the participants (Kampen and Swyngedouw 2000). Given the difficulties of assuming these measurements are interval, readers are strongly encouraged to exercise caution when interpreting Pearson's r coefficients and should consider other measures of association more appropriate for analyzing ordinal data, such as: Spearman's ρ, Kendall's τ, and Polychoric correlations.

Results of the Spearman's ρ found three of the correlation coefficients to be statistically significant (ρ_{12}, ρ_{13}, and ρ_{23}), with the greatest magnitude existing between items 1 and 2. Although Spearman's ρ coefficients were all larger in magnitude than Kendall's τ, they both had the same (a) signs for all estimates, (b) set of statistically significant estimates (τ_{12}, τ_{13}, and τ_{23}), and (c) greatest magnitude (or strongest) estimates existing between items 1 and 2. Importantly, while Spearman's ρ and Kendall's τ estimates displayed the same signs as Pearson's r, they did not share the same hypothesis test results. For example, the Spearman's ρ and Kendall's τ estimates were significant between items 1 and 3, while the Pearson's r estimates were not. Furthermore, Spearman's ρ and Kendall's τ estimates did not agree with Pearson's r estimates on which items displayed the greatest magnitude (or strongest relationship) as shown in Tables 2 and 3. Given that Pearson's r assumes a bivariate normal distribution and captures the linear dependency between two variables, there might be some possibility that there is a nonlinear relationship between items 1 and 2. As opposed to Pearson's r, further utilities of Spearman's ρ and Kendall's τ are limited in other extended quantitative analyses above and beyond a measure of association. For example, we can further analyze this attitude toward school measure by computing the popular internal consistency measure (Crocker and Algina 1986), which is an application of the Pearson's r estimates. Also, Spearman's ρ and Kendall's τ coefficient estimates are not versatile beyond their own interpretation. As a result, we move to the interpretation of the polychoric correlations.

While both ML and EAP polychoric correlation estimates displayed the same signs, the ML estimates were larger in magnitude. Most interestingly, the five ML estimates ($PCCML_{12}$, $PCCML_{13}$, $PCCML_{23}$, $PCCML_{24}$, and $PCCML_{34}$) were statistically significant compared to only four ($PCCEAP_{12}$, $PCCEAP_{23}$, $PCCEAP_{24}$, and $PCCEAP_{34}$) for the EAP method. Furthermore, the statistical significance of the PCC using ML estimation is determined using the 95% confidence interval, while PCC using EAP estimation is determined using the 95% credibility interval. Although there are similarities between the confidence and the credibility intervals, the interpretation for the confidence interval requires infinitely many samples, while credibility intervals can be applicable for smaller sample sizes. In this illustration, the relationship between items 1 and 3 was statistically significant for PCC using ML estimation, but not for PCC using EAP estimation. For this particular PCC estimate, for the ML case, we can say that "If we randomly sample and estimate confidence intervals infinitely many times, we believe interval estimates capture the population value 95% of the time. Because the confidence interval estimates ($-.531$, $-.01$) did not capture the null value, zero, we conclude that the population value is not zero with 95% confidence." For the EAP case, "It is 95% credible that the population value is within the credible interval estimate ($-.485$, $.020$). Because this credible interval estimate does capture zero, with 95% credibility, we retain the null hypothesis such that the population value is zero."

For the purpose of endorsing one method over the other, applied researchers should be aware that the sample size requirements for polychoric correlations, utilizing the ML estimation, have yet to be fully studied. Given that it cannot be concluded whether the sample size of this dataset ($n = 75$) is sufficient enough to be used with ML estimation, the validity of the hypothesis testing results might be questionable. As a result, caution should be used when interpreting the ML results with such a small sample size and count more on the EAP method, which is known to have the capability of analyzing small sample sizes. In other words, the EAP polychoric correlation method has the advantage of yielding much more statistically valid results with smaller sample sizes in lieu of the ML method (Chen and Choi in press).

Summary and conclusion

From the traditional and most popular method, Pearson's r, to the more recent development of polychoric correlation using EAP estimation, the current paper offered a didactic comparison of correlational analysis methods appropriate for ordinal data. A synthesis of the methodological options for estimating measures of association when data are ordinal is provided in Table 4. The differences between the

Table 4 Summary table of correlational analysis methods for ordinal data

	Pearson's r	Kendall's τ	Spearman's ρ	PCCML	PCCEAP
Inferential statistics	O	O	O	O	O
Interpretation in terms of linear dependency	O	X	X	O	O
Extent in further statistical analyses	Versatile	Limited	Limited	Versatile	Versatile
Bivariate Normal Assumption required	O	X	X	O	O
Minimum sample size as inferential statistics	30*	19*	10*	Large sample size only	Possibility with small sample size
Level of measurement	Metrical	Ordinal	Ordinal	Ordinal	Ordinal

PCCML polychoric correlation using the maximum likelihood (ML) estimation method, *PCCEAP* polychoric correlation using the expected a posteriori (EAP) estimation method

* Values from Chen and Popovich (2002, p. 20, 32, 84)

various statistics are outlined by the (a) interpretations in terms of linear dependency, (b) extensibilities in further statistical analyses, (c) assumptions needed to be carried out for estimation, (d) minimum sample size requirement recommendations as inferential statistics, and (e) required level of measurement of data.

Although it is common practice to use Pearson's r to estimate the association between two sets of ordinal data, the validity of the results from such analysis is quite questionable. To ensure the validity of the correlational results, Pearson's r should only be considered when both variables are measured on an interval or ratio scale and with a minimum sample size of 30 (Chen and Popovich 2002). On the other hand, Kendall's τ, Spearman's ρ, and polychoric correlation are more appropriate correlational analysis options for ordinal data. Spearman's ρ can be used with a sample size as small as 10, while Kendall's τ requires a minimum sample size of 19 (Chen and Popovich 2002).

Unlike Kendall's τ and Spearman's ρ, polychoric correlations, similar to Pearson's r, assume an underlying bivariate normal distribution, capture the linear dependency between two underlying variables and can be used in many other statistical applications (e.g., Cronbach alpha application with polychoric correlation by Zumbo et al. 2007; factor analysis with polychoric correlation, Jöreskog and Moustaki 2001). The estimation technique that one chooses to calculate polychoric correlations will impact the sample size requirements. The traditional maximum likelihood estimation method requires large sample sizes, while the EAP estimation for polychoric correlations can have stable estimates of this correlation even when sample sizes are small (e.g., smaller than 30). In light of the current synthesis, it is our hope that when applied researchers consider analyzing ordinal data, they choose a correlational analysis method which yields not only methodologically valid but also practically useful results. Further extensions of polychoric correlations to more advanced quantitative analyses

using other Bayesian estimation method (e.g., maximum a posteriori; Bock and Aitkin 1981) with simulation study with various factors (e.g., nonnormal underlying distributions or options for interval estimates) would be most desirable and remains to be addressed in future work.

References

Aldrich, J. (1997). R.A. Fisher and the making of maximum likelihood 1912–1922. *Statistical Science, 12*, 162–176.

Bock, R., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of the EM algorithm. *Psychometrika, 46*, 443–459.

Bollen, K., & Barb, K. (1981). Pearson's r and coarsely categorized measures. *American Sociological Review, 46*, 232–239.

Borgatta, E., & Bohrnstedt, G. (1980). Level of measurement once over again. *Sociological Methods & Research, 9*, 147–160.

Chen, J., & Choi, J. (in press). A comparison of maximum likelihood and expected a posteriori estimation for polychoric correlation using Monte Carlo simulation. *Journal of Modern Applied Statistical Methods.*

Chen, P., & Popovich, P. (2002). *Correlation: Parametric and nonparametric measures*. Thousand Oaks, CA: Sage Publications.

Choi, J., Chen, J., & Kim, S. (2009). BayesPCC (v 3.0.0): A stand-alone Microsoft Windows software program for estimating polychoric correlation matrices using Bayesian estimation methods [Computer software].

Choi, J., Chen, J., & Kim, S. (in press). EAPPCC: A Matlab subroutine for estimating polychoric correlation matrices using an expected a posteriori estimation method. *Applied Psychological Measurement.*

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Belmont, CA: Wadsworth Group.

Green, S., & Salkind, N. (2005). *Using SPSS for windows and Macintosh: Analyzing and understanding data*. Upper Saddle River, NJ: Prentice Hall.

Jöreskog, K., & Moustaki, I. (2001). Factor analysis of ordinal variables: A comparison of three approaches. *Multivariate Behavioral Research, 36*, 347–387.

Kampen, J., & Swyngedouw, M. (2000). The ordinal controversy revisited. *Quality & Quantity, 34*, 87–102.

Kaplan, D. (2000). *Structural equation modeling: Foundations and extensions*. Thousand Oaks, CA: Sage.

Kendall, M., & Gibbons, J. (1990). *Rank correlation methods*. New York: Oxford University Press.

Mayer, L. (1971). A note on treating ordinal data as interval data. *American Sociological Review, 36*, 519–520.

NCES (2008). *Program for international student assessment (PISA)*. Retrieved November 7, 2008, from http://nces.ed.gov/surveys/pisa/index.asp.

O'Brien, R. (1985). The relationship between ordinal measures and their underlying values: Why all the disagreement? *Quality & Quantity, 19*, 265–277.

Olsson, U. (1979). Maximum likelihood of the polychoric correlation coefficient. *Psychometrika, 44*, 443–460.

Pearson, K. (1957). *The grammar of science*. New York: The Meridian Library.

Poon, W., Leung, K., & Lee, S. (2002). The comparison of single item. Constructs by relative mean and relative variance. *Organizational Research Methods, 5*, 275–298.

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.

Vito, G., & Blankenship, M. (2002). *Statistical analysis in criminal justice and criminology: A user's guide*. Upper Saddle River, NJ: Prentice-Hall, Inc.

Ware, W., & Benson, J. (1975). Appropriate statistics and measurement scales. *Science Education, 59*, 575–582.

Zumbo, B., Gadermann, A., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. *Journal of Modern Applied Statistical Methods, 6*, 21–29.