

## Practical Approaches to Dealing with Nonnormal and Categorical Variables

### Definitions and Distinctions

First, it is important to distinguish between categorical variables and continuous variables. Categorical variables are those with two values (i.e., binary, dichotomous) or those with a few ordered categories. Examples might include gender, dead vs. alive, audited vs. not audited, or variables with few response options like “never,” “sometimes,” or “always.” Continuous variables are variables measured on a ratio or interval scale, such as temperature, height, or income in dollars.

Ordinal variables with many categories, such as 7-point Likert-type scales of agreement, are usually safely treated as “continuous.” In practice, most researchers treat ordinal variables with 5 or more categories as continuous, and there is some evidence to suggest this is not likely to result in much practical impact on results (e.g., Johnson & Creech, 1983). If ordinal variables with many categories are nonnormal, then data analytic techniques for nonnormal continuous variables should be used (see below).

When variables are measured on an ordinal scale and there are relatively few categories, 2-4 categories, estimation methods specifically designed for categorical variables are recommended. This includes nominal binary variables, because binary variables can be considered ordinal for the purpose of meaningful comparisons between the two groups (e.g., gender). A categorical analysis approach will have the greatest advantage (less bias) compared with standard ML when the following conditions hold: (1) when the values between categories are not equidistant; (2) when the relationship between the categorical measured variable and the theoretical variable it is supposed to measure is not a linear relationship—another way of stating (1); (3) when the ordinal variable is skewed or kurtotic.

### Detection of Multivariate Nonnormality

So, how do you know your data are multivariate normal? The first step is to carefully examine univariate distributions and skew and kurtosis. West, Finch, & Curran (1995) recommend concern if skewness  $> 2$  and kurtosis  $> 7$ . Kurtosis is usually a greater concern than skewness. If the univariate distributions are nonnormal, then the multivariate distribution will be nonnormal. One can have multivariate nonnormality (i.e., the joint distributions of all the variables is a nonnormal joint distribution) even when all the individual variables are normally distributed (although this is probably relatively infrequent in practice). Therefore, one should also examine multivariate kurtosis and skewness. However, tests of multivariate normality are only available in EQS and Lisrel. Mardia's multivariate skewness and kurtosis tests are distributed normally (z-test) in very large samples, so can be evaluated against a t or z-distribution. EQS also provides a “normalized estimate” of Mardia's kappa. Bentler and Wu (2002) suggest that a normalized estimate greater than 3 will lead to chi-square and standard error biases. Lawrence DeCarlo (1997) has developed macros for SPSS and SAS to calculate a variety of multivariate nonnormality indices (available at <http://www.columbia.edu/~ld208/>).

### Recommendations for Continuous Nonnormal Variables

In practice, many structural equation models with continuous variables (and generally including ordinal variables of five categories or more) will not have severe problems with nonnormality. The effect of violating the assumption of nonnormality is that chi-square is too large (so too many models are rejected) and standard errors are too small (so significance tests of path coefficients will result in Type I error).

The scaled chi-square and “robust” standard errors using the method developed by Satorra and Bentler (1994) appears to be a good general approach to dealing with nonnormality (Hu, Bentler, & Kano, 1992; Curran, West, & Finch, 1996). Adjustments are made to the chi-square (and to relative fit indices in some packages, such as Mplus and EQS) and standard errors based a weight matrix derived from an estimate of multivariate kurtosis. Mplus prints this kurtosis adjustment, referred to as the “scaling correction factor” (SCF). The scaling correction factor is the standard chi-square divided by the scaled chi-square. The ratio is derived from a multivariate kurtosis estimate used to adjust the chi-square and standard errors. When data are multivariate normal, this scaling correction factor is 1.0, and there is no adjustment to the standard ML chi-square. The more multivariate kurtosis, the larger this scaling correction factor will be (e.g., 1.6

suggests the ML chi-square is approximately 60% higher than the scaled chi-square). At this point, no one has suggested a conventional value for the scaling correction factor that would indicate problematic levels of nonnormality, but I become more concerned when the chi-square inflation is greater than 5 or 10% (SCF of 1.05 or 1.10).

Depending on the complexity of the model and the severity of the problem, sample sizes of 200-500 may be sufficient for good estimates with these “robust” statistics, but, to be safe, sample sizes of over 500 may be best. This approach is now available in Lisrel (ML Robust), EQS (ML Robust), and Mplus (MLM for maximum likelihood mean adjusted). The most recent version of Mplus has made the Satorra-Bentler estimates the default, and this could be a concern with the smallest sample sizes.

Bootstrapping is an increasingly popular and promising approach to correcting standard errors, but it seems that more work is needed to understand how well it performs under various conditions (e.g., specific bootstrap approach, sample sizes needed). The simulation work that has been done (Fouladi, 1998; Hancock & Nevitt, 1999; Nevitt & Hancock, 2001) suggests that, in terms of bias, a standard “naïve” bootstrap seems to work at least as well as robust adjustments to standard errors. However, the Nevitt and Hancock (2001) results suggest that standard errors may be erratic for sample size of 200 or less and samples of 500 to 1,000 may be necessary to overcome this problem. The complexity of the model should be taken into account as their simulations were based on a moderately complex factor model (i.e., smaller sample sizes may be acceptable for simpler models). An alternative bootstrapping approach, the Bollen-Stine bootstrap approach, is usually recommended for estimation of chi-square. The Bollen-Stine chi-square approach seems to adequately control Type I error but there is some cost to power (Nevitt & Hancock, 2001). Bootstrapping approaches have now been incorporated in most major SEM packages.

### **Recommendations for Categorical Variables**

There seems to be growing consensus that the best approach to analysis of categorical variables (with few categories) is the CVM approach implemented in Mplus. This approach, usually referred to as a robust weighted least squares (WLS) approach in the literature (estimator = WLSMV or WLSM in Mplus). The WLSMV approach seems to work well if sample size is 200 or better (Muthen, du Toit, & Spisic, 1997; Flora & Curran, 2004). In Lisrel and EQS, a similar approach that uses WLS together with polychoric correlations and asymptotic covariance matrices is used. In the most recent edition of Amos, an alternative approach to categorical variables has been added. The Bayesian approach requires an iterative process known as the Markov Chain Monte Carlo (MCMC). At this point in time, there is little information on the performance of this approach with SEM with respect to fit estimation, the optimal algorithms to use, and standard errors under various conditions (cf. Lee & Yang, 2006), so I cannot recommend this approach to categorical variables yet.

### **Fit Indices**

Relatively little simulation work on alternative fit indices (e.g., RMSEA, IFI, CFI) derived from robust approaches to nonnormal continuous variables (Satorra-Bentler robust approach or bootstrapping) is currently available. The user should use some caution, because programs do not always recalculate incremental fit indices such as the CFI, TLI, or the IFI using the scaled chi-square for the tested model or the null model (I know that Mplus and EQS do use the scaled chi-squares in their calculation). Relative fit indices will likely be problematic when scaling corrections to the null model are not used (Hu & Bentler, 1999).

To date, we also know relatively little about how fit indices perform with CVM under various circumstances—certainly not with the same level of precision on which Hu and Bentler based their recommendations about fit with continuous variables. The robust WLSMV chi-square used by Mplus seems to perform pretty well (Flora & Curran, 2004), although there is still likely to be a practical problem with using chi-square as a sole measure of fit because of its sensitivity to sample size. There is some evidence that RMSEA performs reasonably well with categorical model estimation (CVM; Hutchinson & Olmos, 1998). Relative fit indices may give valid information about fit if the robust WLSMV chi-square is used to calculate them, but I have not seen any data on this point. CFI as computed in Mplus, for example, appears to give fit

information that is consistent with the RMSEA and WRMR in my experience testing models, but I have not run any simulations to examine it systematically.

The Weighted Root Mean Square Residual (WRMR) is a measure that Muthen has recommended for fit of models with categorical observed variables. Yu and Muthen (2002) recommend good fitting model have a WRMR less than 1.0 indicates good fit (I have also seen the value of .9 recommended). The Bayesian Information Criterion (BIC) is sometimes suggested as a measure of fit for categorical models, but there is no consistently used cutoff for good fit and the BIC may be most practical for comparing fit of different models. In sum, my best recommendation at this point in time for evaluating model fit with the CVM approach is to use the WRMR (approximately less than 1) and RMSEA (approximately less than .06) indices.

## References

- Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1*, 16-29.
- DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods, 2*, 292-307.
- Fouladi, R.T. (1998, April). *Covariance structure analysis techniques under conditions of multivariate normality and non-normality—modified and bootstrap based test statistics*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
- Hancock, G.R., & Nevitt, J. (1999). Bootstrapping and the identification of exogenous latent variables within structural equation models. *Structural Equation Modeling, 6*, 394-399.
- Nevitt, J., & Hancock, G.R. (2001). Performance of bootstrapping approaches to model test statistics and parameter standard error estimation in structural equation modeling. *Structural Equation Modeling, 8*, 353-377.
- Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.
- Hu, L., Bentler, P.M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin, 112*, 351-362.
- Lee, S-Y., & Tang, N.-S. (2006). Bayesian analysis of structural equation models with mixed exponential family and ordered categorical data. *British Journal of Mathematical and Statistical Psychology, 59*, 151-172
- Muthen, B.O, du Toit, S., & Spisic, D. (1997). *Robust infoerince using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes*. Unpublished manuscript.
- Bentler, P.M., & Wu, E.J.C. (2002). *EQS for Windows user's guide*. Encino, CA. Multivariate Software, Inc.
- Johnson, D.R., & Creech, J.C. (1983) Ordinal measures in multiple indicator models: A simulation study of categorization error. *American Sociological Review, 48*, 398-407.
- Recommended reading:** Finney, S.J., & DiStefano, C. (2006). Non-normal and categorical data in structural equation modeling. In G.R. Hancock & R.O. Mueller (Eds.), *Structural equation modeling: A second course*. Greenwich, CT: Information Age Publishing.