SEM with Categorical Variables

Definitions and Distinctions

Categorical variables are those with two values (i.e., binary, dichotomous) or those with a few ordered categories (typically less than five) require special estimation considerations in structural equation modeling. Examples might include sex, dead vs. alive, hospitalized vs. not, or variables with few response options like “never,” “sometimes,” or “always.” Social scientists generally treat variables measured on a ratio or interval scale, such as temperature, height, or income in dollars, or ordinal variables with 5 or more categories as continuous (see the handout “SEM with Nonnormal Continuous Variables” and the handout from my Univariate Statistics course called “Levels of Measurement and Choosing the Correct Statistical Test” for references and details). Similar to regression analysis, measured variables which are not predicted by other variables in the model (including not predicted by latent variables) do not require special estimation for categorical variables (e.g., experimental vs. control group).

When variables predicted by other variables (an endogenous variable in a model or an indicator of a latent variable) are measured on an ordinal scale and there are relatively few categories, 2-4 categories, estimation methods specifically designed for categorical variables are recommended (Finney & DiStefano, 2013). This includes nominal binary variables (e.g., pass/fail, divorced, yes/no, heart attack vs. no heart attack). For ordinal variables with several values, a categorical analysis approach will have the greatest advantage (less bias) compared with standard ML when the following conditions hold (Rhemtulla, Brosseau-Liard, & Savalei, 2012): (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—not entirely unrelated to (1); and (3) when the ordinal variable is skewed or kurtotic.

Estimators for Binary and Ordinal Variables

WLSMV. The most common and potentially the best approach to analysis of binary and ordinal variables (with few categories) is what is referred most to most commonly as diagonally weighted least squares (DWLS) approach (Muthén, du Toit, & Spisic, 1997). This estimation method is called weighted least squares mean and variance adjusted (WLSMV) in Mplus and the R package `lavaan` (it is invoked by `estimator = WLSMV`). The WLSMV approach seems to work well if sample size is 200 or better (Bandalos, 2014; Flora & Curran, 2004; Muthén et al., 1997; Rhemtulla, Brosseau-Liard, & Savalei, 2012).[1] In LISREL and EQS, a similar approach that uses WLS together with polychoric correlations and asymptotic covariance matrices is used.

WLSMV works by first creating a matrix of polychoric correlations. Polychoric correlations estimate the what the association between two continuous, normally distributed variables would be if they been converted to dichotomous or ordinal observed variables.[2] Polychoric correlations, which Finney and DiStefano (2013) refer to as “latent correlations” represent the correlation between the unobserved underlying continuous variables frequently called y^* (“y star”; some authors use eta, η, but that potentially confuses it with what we use for multiple indicator latent variables). In the estimation process, the polychoric correlations are used to create an asymptotic covariance matrix that serves as the weight matrix for the WLS estimation. The resulting path estimates represent the change in y^* on a standardized z scale for each unit change in the predictor. These are the same as what is obtained with probit regression. The distribution of y* requires an arbitrary scaling constraint, much like a latent variable. In Mplus, there are two versions of WLSMV estimates that have different approaches to setting the scaling of the y^* distribution, *delta parameterization* and *theta parameterization* (Muthén & Asparouhov, 2002). Delta parameterization is the default and provides a scaling constraint by setting the measurement residual to 1.0, with no measurement residual variance estimated, and theta parameterization sets the scaling of the y^* variance to 1.0, estimating the

[1] There is another option for weighted least squares, called WLSM in Mplus and lavaan. Both WLSM and WLSMV adjust the standard errors in the same way, but the WLSM method only does a mean adjustment when correcting the chi-square. Because the WLSMV method seems to work better (Asparouhov & Muthen, 2010a; Muthen et al., 1997), I do not recommend using the WLSM option.
[2] There are three related definitions for this type of corrected correlation: tetrachoric, for binary variables, polyserial for binary and continuous, and polychoric for ordinal variables. The term polychoric correlation is often used as a general way of referring to any of these three related correlations.

measurement residual variance. Both can be called variants on the probit model, but theta parameterization corresponds more exactly to the probit regression estimates in which the y^* distribution is assumed to be standardized. These scaling choices are arbitrary in the sense that the chi-square for the model and the significance tests of the parameter estimates will be equal. WLSMV works well in many situations, although because WLSMV is not a full information method it has stricter assumptions for missing data (Asparouhov & Muthén, 2010).

MLR (robust marginal maximum likelihood). Marginal maximum likelihood estimation (or sometimes “maximum marginal likelihood” or just “full maximum likelihood”) is a special maximum likelihood approach for binary and ordinal variables (Bock & Atkin, 1981; Christoffersson, 1975; Muthén & Christoffersson, 1981). This method, which is less commonly employed with SEM models than the WLSMV method, uses the frequency tables in the analysis so can be distinguished from the ML estimation process used for continuous variables. Commonly used in Item Response Theory (IRT) measurement analysis, the default categorical ML estimation yields logistic parameter estimates that can be converted to odds ratios.[3] This approach is not available in many SEM software programs, but Mplus uses the marginal maximum likelihood estimation approach this when ESTIMATOR=ML is used in conjunction with dependent variables identified as categorical on the CATEGORICAL statement. A robust version of ML for categorical variables, which uses robust standard error estimates, is called MLR in Mplus (Note that this MLR is different from the estimation approach obtained when variables are not declared as categorical). The R package `lavaan` currently has limited categorical ML estimation capabilities. An important drawback to the MLR estimation is that the traditional chi-square nor most of the usual alternative fit indices are available to judge fit. Instead Mplus prints a Pearson chi-square and likelihood ratio chi-square based on the contingency tables for the model implied and obtained frequencies. WLSMV probit and ML logistic estimates will often be quite similar in terms of their statistical conclusions. Work by Bandalos (2014) indicates that robust MLR performs better than the unadjusted ML and that MLR performed similarly to the WLSMV method. Compared with WLSMV, MLR has somewhat less power but better control of Type I error in smaller samples. Bandalos's work also suggests that sample sizes of 150 may be too small with either method, especially where distributions of the categorical variables are asymmetric. Newsom and Smith (2020) also found that the two estimators performed similarly but that with MLR had important convergence advantages and that WLSMV had some advantages for the small sample sizes in the context of binary growth curve models. See the handout from this class “Alternative Estimation Methods” for more details on categorical estimation.

Bayesian. Amos does not use either of the above estimation approaches for categorical variables, but the most recent editions allow a Bayesian approach (Lee, 2007). The Bayesian approach, also available in Mplus and the R package `blavaan`, requires an iterative process known as the Markov Chain Monte Carlo (MCMC). To date, there is less information on the performance of this approach with SEM with respect to fit estimation, the best priors settings, the optimal algorithms to use, and standard errors under various conditions (cf. Lee & Yang, 2006). The Bayesian estimation process depends upon the distributional priors used and some artful judgment in the testing process. The Bayesian structural modeling approach has not become a widely used alternative thus far for general conditions but may increase in popularity at least for certain circumstances in the future (see Kaplan & Depaoli, 2012 for an introduction; see also Depaoli, 2021; and the handout from this class “Alternative Estimators”).

Fit Indices

Less is known about how fit indices perform with WLSMV and categorical MLR under various circumstances—certainly not with the same level of precision on which Hu and Bentler and others have based their recommendations about fit with continuous variables under a variety of conditions. The WLSMV chi-square used by Mplus (see Asparouhov & Muthén, 2010a) seems to perform pretty well, even for sample sizes as small as 100 (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. Although some work has supported use of RMSEA, TLI, and CFI with categorical model estimation (WLSMV; Beauducel &

[3] Probit parameter estimates in Mplus can also be requested with `link = probit`.

Herzberg, 2006; Hutchinson & Olmos, 1998; Yu & Muthén, 2002), Savalei (2021) provides evidence that fit is overestimated with these indices and suggests some computational adjustments (using continuous ML with polychoric correlations) that appear to work well. McNeish and Wolf (2024) suggest a dynamic fit approach with relative fit index cutoffs tailored to specific models.

Until the most recent version (Version 8), Mplus reported the Weighted Root Mean Square Residual (WRMR) for fit of models with categorical observed variables (Yu & Muthén, 2002 recommended WRMR of less than 1.0 as indicative of good fit). The most recent versions of the program no longer use WRMR because of data suggesting poor performance with larger samples and larger models (DeStefano, Liu, Jiang, & Shi, 2018) and use a modified computation of the SRMR instead (Asparouhov & Muthén, 2018), where the usual cutoff of ≤ .08 is suggested to indicate good approximate fit. Recent evidence (Savalei, 2021; Shi, Maydeu-Olivares, & Rosseel, 2020) suggests that SRMR is not subject to the overfit bias of the CFI and RMSEA.

Nested Tests

Nested tests (likelihood ratio test) require special attention for robust estimation, including WLSMV (i.e., estimator = WLSMV in Mplus and `lavaan`). The scaling correction factor (scf) must be used to weight the difference (Satorra, 2000; Satorra & Bentler, 2001). Asparouhov and Muthén (2006; 2018b) have adapted the tests developed by Satorra (2000) and Satorra and Bentler (2001) that computes the estimated ratio of the weighted likelihoods of two models using WLSMV estimation for ordinal variables. Mplus provides automated nested tests with the DIFFTEST command that can be used for several estimation or robust methods (Asparouhov & Muthén, 2013, 2018b; Bryant & Satorra, 2012; Satorra & Bentler, 2010). See the handout "Examples of Chi-square Difference Tests with Nonnormal and Categorical Variables" for an illustration.

Missing Data Estimation with Non-normal or Categorical Data

WLSMV (`method = WLSMV`) estimation uses a partially pairwise deletion estimation and is not a full information estimation approach, and therefore it may not work as well with missing data as categorical MLR unless there is only one dependent variable and values are at least MAR (Asparouhov & Muthén, 2006; Asparouhov & Muthén, 2010b).

Other Types of Categorical Variables

In addition to binary and ordinal variables, Mplus also has estimation approaches for count variables, including Poisson, negative binomial, zero-inflated Poisson and negative binomial, nominal (multnomial logistic regression), and continuous survival analysis. Few other SEM packages have features for these types of variables and, although the extension from regression is relatively straightforward, there has been little simulation work thus far to examine how these algorithms perform with model fit and standard error estimation in the SEM context.

References

Asparouhov, T., & Muthén, B. (2010a). *Simple second order chi-square correction.* Unpublished manuscript. https://www.statmodel.com/download/WLSMV_new_chi21.pdf
Asparouhov, T., & Muthén, B. (2010b). *Weighted least squares estimation with missing data.* Unplublished technical report, retrieved from https://www.statmodel.com/download/GstrucMissingRevision.pdf
Asparouhov, T. and Muthen, B. (2013). Computing the strictly positive Satorra-Bentler chi-square test in Mplus. Unpublished manuscript. https://www.statmodel.com/examples/webnotes/SB5.pdf
Asparouhov, T., & Muthén, B. (2018a). SRMR in Mplus. Unpublished manuscript. https://www.statmodel.com/download/SRMR2.pdf
Asparouhov, T., & Muthén, B. (2018a) Nesting and equivalence testing in Mplus. https://www.statmodel.com/chidiff.shtml
Babakus, E., Fergnson, C. E. & Joreskog, K G, (1987). The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions. *Journal of Marketing Research, 24*, 2228.
Bentler, P.M., & Wu, E.J.C. (2002). *EQS for Windows user's guide.* Encino, CA. Multivariate Software, Inc.
Bock, R. D., & Atkin, M. (1981). Marginal maximum likelihood estimation of item parameters. An application of an EM algorithm. *Psychometrika, 46*, 443–459.
Bryant, F. B., & Satorra, A. (2012). Principles and practice of scaled difference chi-square testing. *Psychometrika, 66*, 507–514.
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.
Christoffersson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika, 40*, 5–32.
DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods, 2*, 292-307.
Depaoli, S. (2021). *Bayesian structural equation modeling.* Guilford.
DiStefano C., Liu J., Jiang N., & Shi D. (2018) Examination of the Weighted Root Mean Square Residual: Evidence for Trustworthiness? *Structural Equation Modeling, 25*, 453-466.
Dolan, C. V. (1994), Factor analysis of variables with 2, 3, 5, and 7 response categories: A comparison of categorical variable estimators using simulated data. *British journal of Mathematical and Statistical Psychology, 47*, 309-326.
Finney, S.J., & DiStefano, C. (2013). Non-normal and categorical data in structural equation modeling. In G.R. Hancock & R.O. Mueller (Eds.), *Structural equation modeling: A second course, 2nd Edition* (pp. 439–492). Charlotte, NC: Information Age Publishing.
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.
Hutchinson, S, R., & Olmos, A (1998). Behavior of descriptive fit indexes in confirmatory factor analysis using ordered categorical data. *Structural Equation Modeling: A Multidisciplinary Journal*, *5*, 344-364.

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.
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.
Kaplan, D., & Depaoli, S. (2012). In R.H. Hoyle (Ed.). *Handbook of structural equation modeling* (pp., 650-673). New York: Guilford Press.
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
Lee, S-Y. (2007). *Structural Equation Modeling: A Bayesian Approach.* New York: Wiley.
McNeish, D., & Wolf, M. G. (2024). Direct discrepancy dynamic fit index cutoffs for arbitrary covariance structure models. *Structural Equation Modeling: A Multidisciplinary Journal, 31*(5), 835-862.
Muthén, B., & Christoffersson, A. (1981). Simultaneous factor analysis of dichotomous variables in several groups. *Psychometrika, 46*, 407–419.
Muthén , B.O, du Toit, S., & Spisic,, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes.* Unpublished manuscript.
Muthén, B., & Asparouhov, T. (2002). Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus. http://www.statmodel.comexamples/webnote.shtml#web4.
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.
Newsom, J.T., & Smith, N.A (2020): Performance of Latent Growth Curve Models with Binary Variables. *Structural Equation Modeling: A Multidisciplinary Journal,* 1-20. DOI: 10.1080/10705511.2019.1705825
Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological methods*, *17*, 354-373.
Satorra, A., & Bentler, P.M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. *1988 Proceedings of the Business and Economic Statistics Section of the American Statistical Association*, 308-313.
Satorra, A., & Bentler, P.M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye and C.C. Clogg (eds.), *Latent Variable Analysis: Applications to Developmental Research* (pp. 399-419). Newbury Park: Sage.
Savalei, V. (2021). Improving fit indices in structural equation modeling with categorical data. *Multivariate Behavioral Research, 56*(3), 390-407.
Shi, D., Maydeu-Olivares, A., & Rosseel, Y. (2020). Assessing fit in ordinal factor analysis models: SRMR vs. RMSEA. *Structural Equation Modeling: A Multidisciplinary Journal, 27*(1), 1-15.
Yu, C.-Y., & Muthén, B. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes.* Doctoral dissertation. Retrieved from http://www.statmodel.com/download/Yudissertation.pdf