Person-Fit Statistics with Model Misspecification: A Monte Carlo Study

Ruben Castaneda, Nicole Zelinsky, and Michelle Turitz.

Background

- Person fit:
  - Evaluation of model-data fit at the individual respondent level across the set of test items
- Person misfit
  - Occurs when a person's response pattern is inconsistent with what the model predicts
  - For example, low ability person scoring high on a set of difficult items
- Person-fit Monte Carlo studies
  - Focused on the statistics' abilities to capture aberrant response patterns
  - Context of unidimensional IRT but not multidimensional IRT to the authors' knowledge
- Many psychological constructs are inherently multidimensional, but collapsed to unidimensional for application
  - For example, scaling a personality measurement using unidimensional 2PL
- Because the *Zh* statistic is the most common, pre-packaged person-fit statistic, we used it for our simulation

Objective

- To assess the performance of the *Zh* person-fit statistic under null conditions when the model is misspecified with 1, 2 and 3 factors.

Drasgow & Levines L_z statistic

Person fit is evaluated using Drasgow & Levines L_z statistic. Here X_g is the binary item response and P_g is the item response function:

- l_0
  - $\sum_{g=1}^{k}\{X_g \ln P_g(\theta) + (1 - Xg)\ln[1 - Pg(\theta)]\}$
- $E(l_0)$
  - $\sum_{g=1}^{k} \{Pg(\theta)\ln[P_g(\theta)] + [1 - Pg(\theta)]\ln[1 - Pg(\theta)]\}$
- $Var(l_0)$
  - $\sum_{g=1}^{k} P_g(\theta)[1 - Pg(\theta)]\left[ln\frac{P_g(\theta)}{1-Pg(\theta)}\right]^2$
- L_z
  - $\frac{l_0 - E(l_0)}{[Var(l_0)]}$

Simulation Conditions

To examine the performance of the Zh person-fit statistic we used the following conditions:

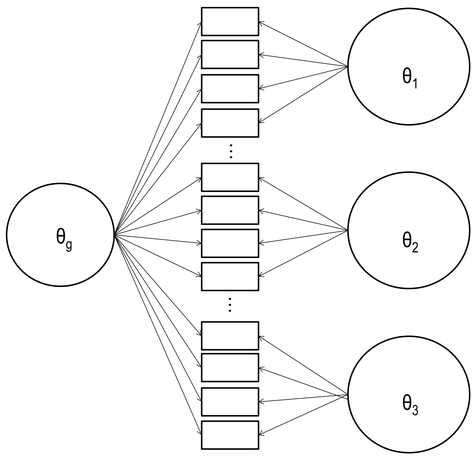
- Model Specification
  - Correctly specified (Bifactor Model)
  - Incorrectly specified (1, 2, or 3 Factors)
- Sample Size (200, 500, or 1000)
- Number of Items (5, 10, 15, or 20)

Proportion of Converged Cells

		Number of Items per Factor			
		5	10	15	20
Sample Size	200	0.139	0.539	0.853	0.607
	500	0.472	0.452	0.914	0.674
	1000	0.315	0.361	0.340	0.229

Data Generating Model

Figure 1. Bifactor Model



Contact Information

- **Ruben Castaneda**: rcastaneda2@ucmerced.edu
- **Nicole Zelinsky**: nzelinsky@ucmerced.edu
- **Michelle Turitz**: mturitz@ucmerced.edu

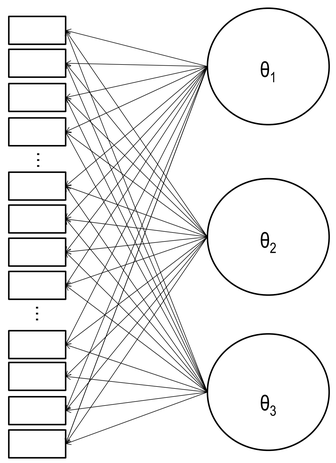
University of California, Merced
5200 Lake Rd, Merced, CA 95343

Methods

- Data for this study were simulated using the R software (R Core Team, 2014) and IRT estimation used the package `mirt`.
  - Simulated item responses were generated by a 2PL, bifactor model
  - Slopes were generated randomly from a log normal distribution with a mean = 0 and standard deviation = 0.5
  - Intercepts were randomly generated from a standard normal distribution
- After simulating the data, we applied the correctly specified model to the data as well as the incorrectly specified 1, 2, and 3 factor models.
  - Only used results when models converged
  - Kept running cell until at least 1,000 converged

Misspecified Model

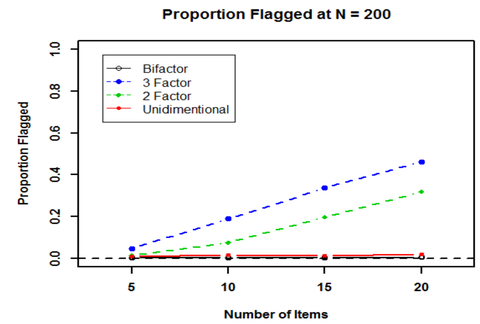
Figure 2. Multidimensional Model

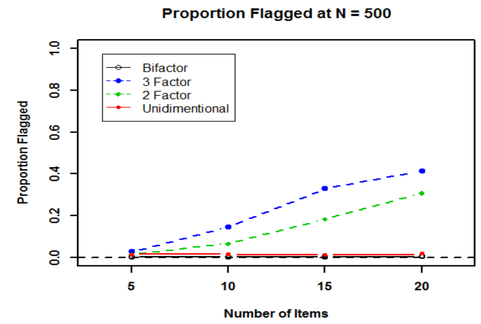


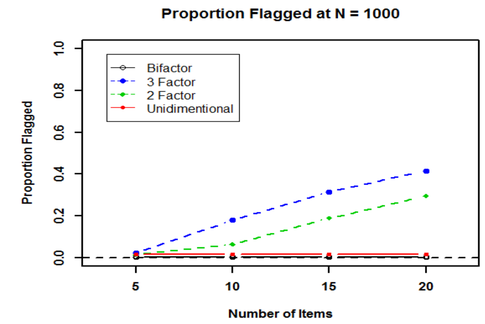
References

- Drasgow, F., Levine, M. V., Williams, E. A. (1985). Appropriateness measurement with polychotomous item response models and standardized indices.
- Meijer, R.R., Sijtsma, K. (2001). Methodology Review: Evaluating Person Fit.
- R Core Team, R. (2014). R: A language environment for statistical computing. Vienna, Austria.: R Foundation for Statistical Computing.0

Results







Discussion and Future Directions

- Bifactor and Unidimensional models flagged for person misfit at a rate below alpha
- Models with 2 factors and 3 factors models erroneously flagged person misfit as the number of items increased
- Stayed constant as the sample size changed
- Further simulation work with person fit statistics with misspecified, multidimensional models

UNIVERSITY OF CALIFORNIA
UCMERCED