Homework #3

Bayesian Psychometric Modeling (PSQF 7375, Fall 2024)

Due Friday, November 1st, 2024 at 11:59pm in ICON

## Bayesian Psychometric Models

The objectives of this homework are:

1. To learn how to estimate a confirmatory factor analysis and polytomous item repsonse models with MCMC
2. To see how each type of observed data assumption impacts the latent variable estimates

### General Instructions:

The data for homework 3 were collected at a casino from a sample of persons who were gambling. The content and response options for 24 items measuring pathological gambling are given in the excel file GRI\_Items.xlsx, along with the item responses in the excel file GRI\_Data.xlsx. Use Stan for all models in this assignment.

#### For all questions below, when estimates are requested, report:

* The posterior mean (EAP)
* The posterior standard deviation
* The 95% highest density posterior interval (lower and upper bound)

#### To estimate each model, use Stan with the following options:

* Four chains
* Warmup of 5000 iterations
* Sampling of 5000 iterations
* Starting values drawn from for all factor loading/discrimination parameters

You may work with other students in building the homework, but your answers and syntax must not be copied from anyone else.

### Submission Instructions:

Please submit your homework as an R Markdown file where all R syntax is embedded as chunks and each answer is provided using your words (text answers are required, not just output) following each question in the section marked **My Answer**.

**Name your file with your first and last name and submit your file to ICON.**

### Data File

The data file is available for download at <https://jonathantemplin.com/wp-content/uploads/2022/11/GRI_Data.xlsx>. The list of items are available at <https://jonathantemplin.com/wp-content/uploads/2022/11/GRI_Items.xlsx>

The variables in the data are:

* ID: The ID number of the person
* grii-gri24: The response to a 6-point Likert item for each person

### Section 1

The purpose of section 1 is for you to practice confirmatory factor analysis, including writing the syntax for the model, as well as finding and and interpreting the relevant pieces of output. The CFA model assumes each observed item follows a conditional normal distribution and each factor follows a normal distribution.

Using Stan, estimate a single-factor CFA model for all items. Use the z-score method of identification, in which the factor mean is fixed to 0, the factor variance is fixed to 1, and all item parameters are estimated.

Use Stan’s default prior distributions for each model parameter that is not a factor score.

**1. What is the model for each observed item response (i.e., the linear predictor, the distribution, and any other parameters)**

**My Answer**

**2. Did your Markov chains converge? What evidence do you have to support your answer?**

**My Answer**

**3. What are your estimates for each item intercept?** (see estimate details above for what to report)

**My Answer**

**4. What are your estimates for each factor loading/discrimination parameter** (see estimate details above for what to report)

**My Answer**

**5. What are your estimates for each unique variance (standard deviation) parameter?** (see estimate details above for what to report)

**My Answer**

**6. What are your estimates for the (variance accounted for by the latent factor) for each item?** (see estimate details above for what to report)

**My Answer**

**7. What are your estimates for the factor scores?** (see estimate details above for what to report)

**My Answer**

**8. Create a histogram of the EAP estimates for the factor scores. What range of values of factor scores have the most individuals?**

**My Answer**

**9. Challenge Question: Create a histogram of all sampled values for factor scores across all individuals (i.e., each value from each chain from all people). What does the distribution look like?**

**My Answer**

**10. Challenge Question 2: Create a Q-Q plot for all sampled values for factor scores to evaluate this distribution as following a normal distribution. Hint: use the qqnorm() function in R**

**My Answer**

## Course Project, Part 1: Introduction and Research Questions/Hypotheses (10 Points; Separate Grade from HW)

Create a brief literature review of the area of research that your project addresses and provide a set of research questions and/or hypotheses you seek to answer with your study.

Please keep this section to no more than 10 pages of double-spaced text (not including the title page, figures, tables, or references).

You may wish to preregister your hypotheses: <https://www.cos.io/initiatives/prereg>

If you do not have a real data set you can use, please contact me.