Item Response Theory - Final Essay

Marius Keute

September 22, 2022

Contents

1 Introduction **3**

2 Preparing the Data **4**

3 Descriptive Analyses and Dichotomization **5**

4 IRT modeling **7**
- 4.1 Rasch model estimation 7
- 4.2 Model analysis 8
- 4.3 Differential Item Functioning 10

5 Higher-parameterized IRT models **10**

6 Polytomous IRT model **13**

7 Factor models **13**

8 Reliability and Unidimensionality **15**

9 Measurement Invariance **15**

10 Theoretical Part: Key differences between IRT and CTT **15**
- 10.1 Introduction 15
- 10.2 Core Ideas and Terminology 15
- 10.3 Strenghts 16
- 10.4 Limitations 16
- 10.5 Conclusion and Application 16

11 Analysis code **17**

12 References **54**

submitted to:

Dr. Stefano Noventa

University of Tübingen

submitted by:

Marius Keute (QDS, 5991873)

Statutory Declaration: I hereby declare that I composed the present paper independently and that I have used no other resources than those indicated. The text passages which are taken from other works in wording or meaning have been identified as such. I also declare that this work has not been partly or completely used in another examination.

1 Introduction

Understanding sexual habits and behavior can be important for, e.g., improving sex education for adolescents, preventing sexually transmitted diseases (STDs), and identifying high-risk populations for sexual misconduct. The Sexual Compulsivity Scale (SCS) is a 10-item questionnaire constructed to measure hypersexuality and high libido in a given person (Kalichman and Rompa (1995),Kalichman and Rompa (2001)). Each of the 10 items is a statement about sexual habits, feelings, or experiences, and the test-taker can indicate how much they can relate to each statement on a four-level scale ranging from 1 (Not at all like me) to 4 (Very much like me).

The 10 items are (Kalichman and Rompa (2001)):

1. My sexual appetite has gotten in the way of my relationships.
2. My sexual thoughts and behaviors are causing problems in my life.
3. My desires to have sex have disrupted my daily life.
4. I sometimes fail to meet my commitments and responsibilities because of my sexual behaviors.
5. I sometimes get so horny I could lose control.
6. I find myself thinking about sex while at work.
7. I feel that sexual thoughts and feelings are stronger than I am.
8. I have to struggle to control my sexual thoughts and behavior.
9. I think about sex more than I would like to.
10. It has been difficult for me to find sex partners who desire having sex as much as I want to.

In this essay, using data from the original validation cohort (Kalichman and Rompa (2001)), I will provide a thorough analysis of the SCS, using methods derived from Item Response Theory (IRT), and to a lesser extent from Classical Test Theory (CTT). In the final section, I will give an overview over both theories and their key differences.

2 Preparing the Data

The dataset (Kalichman and Rompa (1995)) consists of 3376 observations, the variables being the ten items of the SCS, the sum score, gender and age. From the age variable, three cases where the reported age was 100 years or higher appeared implausible and therefore set to missing values. The remaining cases had a mean age of 30.9 years (median 28 years, range [14, 85]). From the gender variable, 13 values were missing and 15 cases where the reported gender was “3” (other) were set to missing values. Of the remaining cases, 2295 (68.5%) reported male gender (“1”) and 1053 (31.4%) reported female gender (“2”). In the dataset, 133 cases contained at least one missing value.

The pattern of missing SCS items is shown in Figure 1. It can be seen that item Q9 was missing most often, though not by a large margin (Q9: 27 missing values, Q5: 13 missing values). It can be seen that the majority of cases with missing values (118 cases / 88.7%) had only a single missing item, while there were no prominent patterns of items that tended to be jointly missing. Eight cases where more than two SCS items were missing were excluded from all further analyses. For the remaining 3368 cases, the probability of missing values at each SCS variable was modeled as a function of the values in *all other* SCS variables using a logistic regression model:

$$P(M_{i,q} = 1|X_{i,q'}) = \sigma(X_{i,q'}\hat{\beta}),$$

where $M_{i,q}$ is 1 if the i^{th} person has a missing value at item $q \in \{Q1, Q2, ...Q10\}$, $X_{i,q'}$ denotes the item values of all other items except item q, σ is the logistic function $\sigma(x) = \frac{1}{1-e^{-x}}$, and $\hat{\beta}$ are the estimated regression weights (Guan and Yusoff (2011)). Note that each variable’s pattern of missing values could only be predicted based on the observations without missing values in any other variable, since those cases were excluded by the logistic model by default of the implementation. Since the majority of cases had either no or only one variable missing, however, this should not bias the overall picture very much.

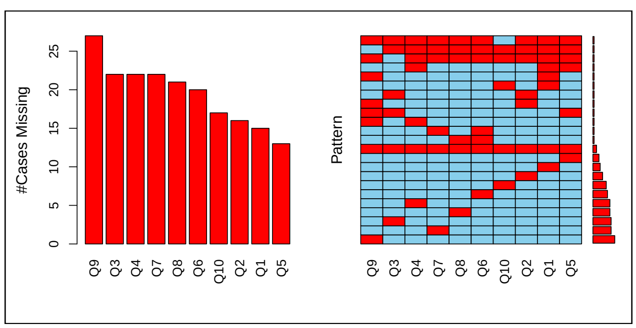


Figure 1: Pattern of missing SCS values.

3 Descriptive Analyses and Dichotomization

The distribution of responses for each item before dichotomization can be seen in Figure 2. All item categories show reasonable coverage of the range of responses (1-4), and there are no obvious flooring or ceiling effects, except for a slight tendency to a flooring effect with item Q6 (few cases with response 1).

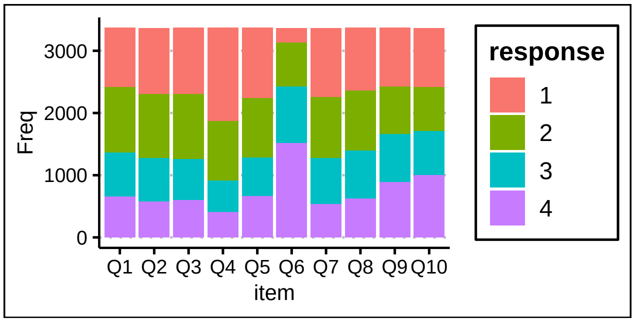


Figure 2: Distribution of non-dichotomized responses per item

For dichotomization of the item data, I considered two options, namely, thresholding each of the 10 items at its own median, to ensure an even distribution of observations into both categories for each item, or finding a common threshold for all items. Since the items have only four levels each, a median split would not necessarily lead to a very balanced dichotomization. Furthermore, the item levels are designed to have the same meaning across all items, therefore I decided to dichotomize at a common threshold of 2, i.e., the dichotomous items $D_q \in \{D_1, D_2, ..., D_{10}\}$ were defined such that

$$D_{i,q} = \begin{cases} 0 \text{ if } Q_{i,q} \in \{1, 2\}, \\ 1 \text{ if } Q_{i,q} \in \{3, 4\}, \end{cases}$$

Of note, simple models in IRT such as the Rasch model (see below) assume that all item responses are either correct or incorrect (or solved / unsolved, respectively). Since a personality test such as the SCS does not have right or wrong responses, it is common to dichotomize the values, as described above, and henceforth treat one of the dichotomous response options as the 'correct' one, in this case, responses greater than 2. This is, however, purely for compliance with IRT terminology and does not imply that the 'correct' dichotomous responses are better than the 'incorrect' ones in any way.

Descriptive characteristics of the 10 SCS items are shown in Table 1, the proportions of correct responses are shown in Figure 2. Since most variables' median was 2, this was not much different from an item-wise median threshold (see Table 1).

Subsequently, I calculated biserial correlations between all pairs of dichotomized items. Moreover, I calculated item discrimination, i.e., each items ability to discriminate between high- and low-scoring individuals, using the adjusted item-total correlation method (Reynolds and Livingston (2021)), i.e., by calculating biserial correlation coefficients between each (dichotomized) item's scores and the sum of all other (dichotomized) items.

Item intercorrelations are shown in Figure 3. It can be seen that all pairs of items show moderate to high positive correlations, indicating that all items measure similar information yet are not redundant. Item easiness (i.e., proportion of correct responses) was between 27% (item Q4) and

Table 1: Descriptive item statistics (mean, median and range *before* dichotomization)

X	stat	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
1	max	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0
2	mean	2.3	2.2	2.2	1.9	2.2	3.1	2.2	2.3	2.5	2.5
3	median	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	3.0
4	min	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

Table 2: Distribution and discrimination of dichotomized items

X	Q1	Q2	Q3	Q4	Q5	Q6	Q7	
item easiness (percent in category 1)	40.50	37.90	37.30	27.00	38.20	71.90	37.80	4
number of cases in category 1	1364.00	1276.00	1255.00	910.00	1287.00	2423.00	1273.00	139
discrimination	0.45	0.45	0.44	0.34	0.29	0.26	0.42	

71.9% (item Q6), item discrimination was between .26 (item Q6) and .45 (items Q1, Q2), i.e., there was no item with a trivial response pattern (e.g., all or no responses correct), and no item was a good representation of the entire scale, since all item discriminations were only moderate in size.

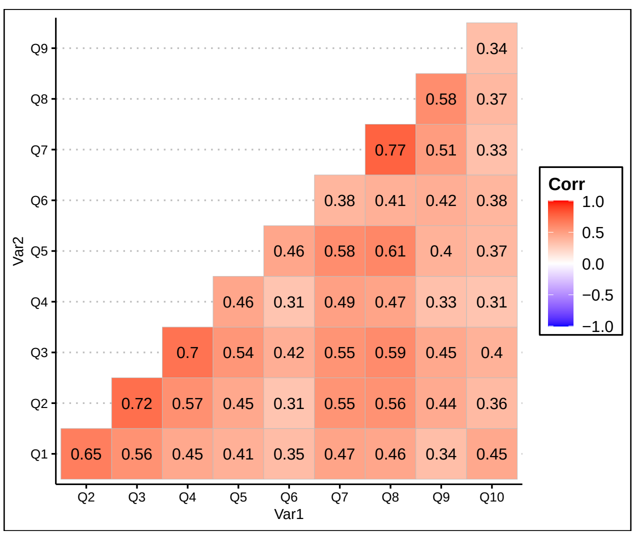


Figure 3: Pattern of missing SCS values.

4 IRT modeling

After analyzing the SCS data using descriptive statistics and concepts derived from CTT, in the following I will fit and discuss different IRT models to the data.

4.1 Rasch model estimation

Next, I estimated a Rasch model for the SCS data, also known as either the one-parameter logistic model or one-parameter normal ogive model, depending on the parameterization.

It models a given person's chances of solving a given item as a logistic function of the difference between the q^{th} item's difficulty β_q and the i^{th} person's ability θ_i, where β_q and θ_i are latent (unobserved) quantities that are estimated from the dichotomous (solved vs. not solved) item data.

The probability for a given person can then be expressed by the logistic function: $P(D_{i,q} = 1|\beta_q, \theta_i) = \sigma(\theta_i - \beta_q)$, where σ is the logistic function as specified above. That is to say, it is purely the difference between item difficulty and person ability that explains the correctness of item responses within the model.

Crucially, the model assumes that this relationship is identical for all items, i.e., the logistic function can only be shifted but not changed in slope across items with different difficulty. Item difficulty is, therefore, the only free parameter of the Rasch model, whereas alternative models (see below) also estimate additional parameters.

To obtain a comprehensive picture, I fitted Rasch models using three different software implementations in `R` 4.1.

The first method was the one implemented in the R package `eRm` (Mair and Hatzinger (2007)). The `eRm::RM` function estimates a Rasch model using conditional maximum likelihood estimation. To make the model identifiable, the user can choose between two model constraints, namely that the model parameters must sum to 0 or that the first item's parameter is fixed to 0. I chose the first (default) option, i.e., forcing item difficulties to sum to 0. Item discriminativity, i.e., the steepest slope of the logistic functions (at $\beta_q = \theta_i$), is fixed to 1 for all items in this implementation.

The second method was the one implemented in the R package `ltm` (Rizopoulos (2006)). The `ltm::rasch` function estimates a Rasch model using approximate marginal maximum likelihood estimation. This package provides the user with more flexibility to impose constraints on the model than `eRm`, I fixed item discriminativity to 1 for all items, to maximize comparability with the `eRm` parameters.

The third method was a structural equation model as implemented in `lavaan` (Rosseel (2012)). Unlike the two previous implementations, `lavaan` requires a more explicitly user-defined model specification, as it does not provide any ready-made function or syntax for Rasch models.

I used a modified copy of the syntax presented in Templin (2022):

```
SCS =~ 1*Q1 + 1*Q2 + 1*Q3 + 1*Q4 + 1*Q5 + 1*Q6 + 1*Q7 + 1*Q8 + 1*Q9 + 1*Q10

Q1 | t1; Q2 | t1; Q3 | t1; Q4 | t1; Q5 | t1; Q6 | t1; Q7 | t1;
Q8 | t1; Q9 | t1;Q10 | t1;

SCS ~ 0;
```

```
SCS ~~ 1*SCS
```

Again, I fixed item discriminativities to 1 for all items. The item parameters `Q1`, ..., `Q10` were subjected to a common threshold `t1`, and the sum of all item parameters (corresponding to the latent variable `SCS`) was fixed to 0, as with `eRm`. Moreover, its variance was fixed to unit. Of note, due to limitations of the implementation, `lavaan` is not able to estimate Rasch models using maximum likelihood estimation, but only using mean- and variance-adjusted weighted least squares (WLSMV) estimation, which limits model fit comparisons. **TODO describe conversion to IRT params**

4.2 Model analysis

The item difficulty parameters of the three models are shown in Figure 4, along with the item difficulty derived from CTT (i.e., the proportion of incorrect responses per item in the data). While the parameters differed between the different models, it is important to note that the parameters from all four models (including CTT) were perfectly correlated for all pairs of models (all r > .999), which indicates that the parameters of one model are simply affine linear transformations of the parameters of any other model, i.e., while numerically different, the models incorporated identical information about the items. The corresponding item-characteristic curves (ICC) are shown in Figure 5. ICCs are generated by calculating the function graph of the item-wise logistic functions parameterized by item difficulty, across a range of possible person ability values on the x-axis.

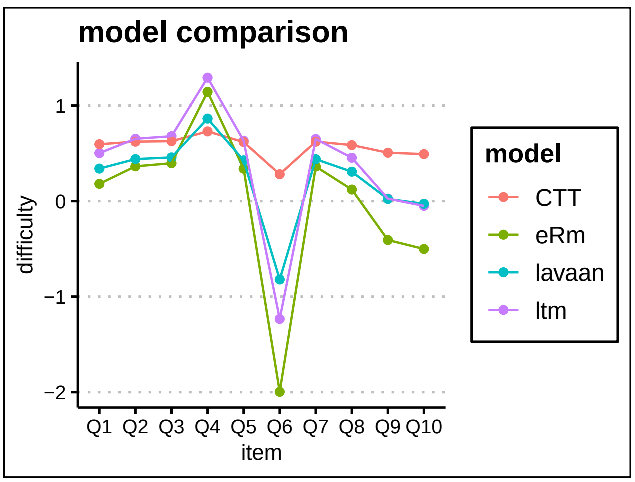


Figure 4: Item difficulties in comparison.

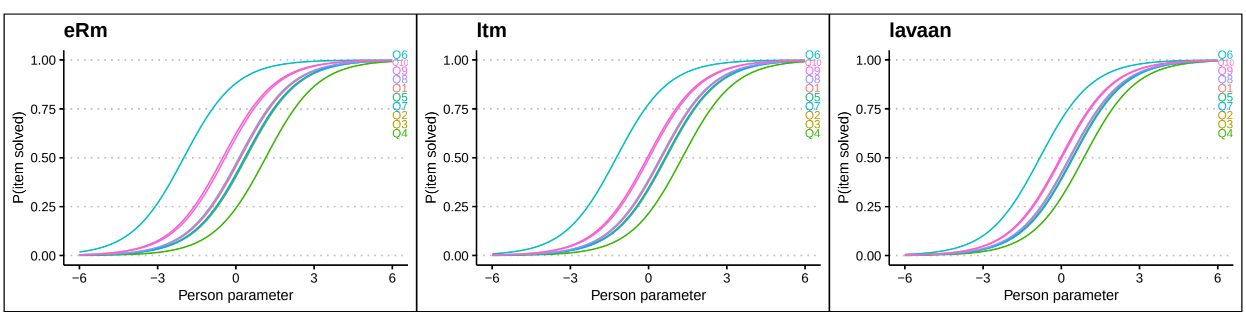


Figure 5: Item-characteristic curves for the three Rasch models.

The overall model fit indices are shown in Table3. Of note, log-likelihood and information criteria

Table 3: Fit indices for Rasch models

X	loglik	npar	AIC	BIC	cAIC
eRm	-17881.68	9	35781.37	35836.47	35845.47
ltm	-18560.47	10	37140.94	37202.16	NA

can only be reported for those models fitted using `ltm` and `eRm`, while the `lavaan` model's fit indices are not comparable, as it was not fitted using maximum likelihood estimation. For brevity, I skip the discussion of the `lavaan` model fit indices. Comparing the fit indices for the `ltm` and `eRm` Rasch models, it can be seen that `eRm` had an overall higher log-likelihood. Since it also had one free parameter less (because of the sum constraint, see above), it was, overall, the preferred model also according to the Akaike and Bayes-Schwarz information criteria.

Finally, to get an impression of how the three models perform for each item, I calculated the mean 0-1-loss per item as:

$\mathcal{L}_q = \frac{1}{n}\sum_{i=1}^{n}|f(\theta_i, \beta_q) - D_{i,q}|$

that is to say, I used the item difficulty parameters β_q and person ability scores θ_i estimated by the models to make predictions for each person and item:

$$f(\theta_i, \beta_q) =: \begin{cases} 1 \text{ if } \sigma(\theta_i - \beta_q) > 0.5, \\ 0 \text{ otherwise} \end{cases}$$

The mean loss is then calculated as the proportion of incorrectly predicted cases for each item. It is shown in Figure 6. It can be seen that the three models, despite differences in parameterization, performed very similarly, with the `eRm` and `lavaan` models performing almost identically, whereas the `ltm` model tended to incur a slightly higher loss, except for items Q6 and Q10. Interestingly, these are the most difficult items, and the fact `ltm` outperformed `eRm` especially for those items might be related to the differences between conditional vs. marginal maximum likelihood estimation, which have the strongest effect in cases where either all or no responses are correct, i.e., for particularly easy or difficult items.

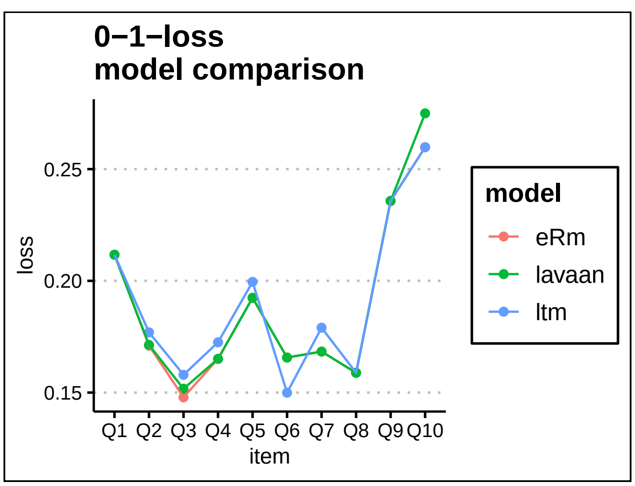


Figure 6: Mean 0-1-loss (proportion of incorrectly predicted responses) per item.

4.3 Differential Item Functioning

I tested for differential item functioning (DIF) using the package `difR` and the procedure outlined in the companion paper (Magis et al. (2010)). DIF is a disadvantageous property of a Rasch model, meaning that item responses differ between subjects from different participant groups, even given the same estimated ability level. The presence of DIF indicates a lack of measurement invariance of the model. The results are displayed in 7. I used the `difLord` method to investigate DIF, but obtained essentially the same results with the `difRaju` method. I discarded `difLRT`, the third recommended IRT-related method, due to its high computational demand. I tested for DIF across genders (male vs. female) and age groups (above median age vs. smaller or equal to median age). Significant DIF (FDR-corrected p-value $< .05$) across the gender groups was detected for items Q5 and Q10, and across the age groups for item Q1, Q5, and Q10. However, the effect sizes were in the negligible range for all but item Q5 in the gender group comparison, where a moderate effect was detected ($\Delta_{Lord} = -1.14$). Since I have only been considering Rasch (one-parameter) models so far, this analysis only tested for uniform DIF.

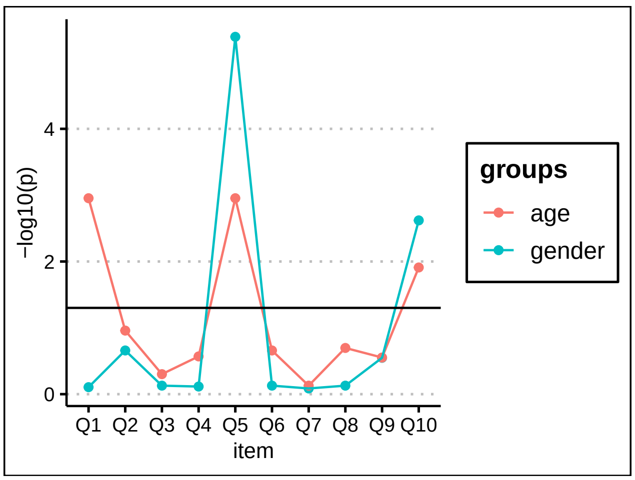


Figure 7: p-values (FDR corrected, negative log-transformed) from DIF analysis. Black line: significance threshold (p $< .05$)

5 Higher-parameterized IRT models

There are several extensions and alternatives to the Rasch model with its restrictive assumptions that differences between items can be described by just one parameter, namely item difficulty, while the Birnbaum model or 2-parameters logistic model also takes into account item discriminativity (corresponding to varying slopes of the item-characteristic curves of different items), and other possible models additionally include a guessing probability term (corresponding to various vertical offsets of the item-characteristic curves of different items) or ceiling probability term (corresponding to clipping the item-characteristic curves from above). For the given dataset, it appears reasonable to estimate a Birnbaum model, whereas 3-PL or 4-PL models seem difficult, since guessing and ceiling probabilities are not easy to operationalize for the dataset, considering that there is no ground truth to the items and we found no prominent ceiling or flooring in any item.

For fitting the 2-PL model, I used the `ltm::ltm` function, and the Rasch model fitted using `ltm::rasch` served as a baseline model for comparison. ICCs and estimated item difficulty parameters are shown in Figure 8 and 9, respectively. It can clearly be seen that item discriminativities, and with them the slopes of the ICC curves, vary considerably between items. A side-effect of two-parameter modeling is that the ICCs now cross each other, i.e., there is no clear order of difficulty between items anymore, but whether one item is more difficult than another can now be dependent on the person ability.

Comparing the two models using a Likelihood Ratio Test, I found the 2-PL model to fit the data significantly better than the Rasch model (log-LR = 1741.55, p < .001).

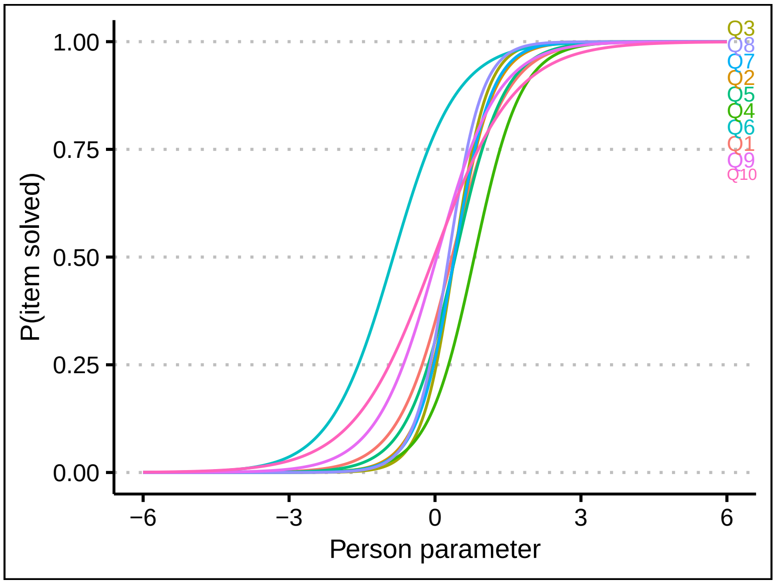


Figure 8: Item-Characteristic Curves from 2-PL model

For further model comparison, I calculated infit and outfit indices for each item and model. Infit and outfit indices are based on model residuals. While outfit (short for outlier-sensitive fit statistic) is particularly sensitive to unexpected responses in cases where item difficulty and person ability are far apart (e.g., a low-ability person unexpectedly solves several very difficult items), infit (short for information-weighted fit statistic) is particularly sensitive to unexpected responses in cases where item difficulty and person ability match (e.g., a person solves far less or far more than half of the items whose difficulty equals their ability). Both are based on the normalized residuals $Z_{iq} = \frac{D_{iq} - \mathbb{E}(D_{iq})}{\sqrt{var(D_{iq})}}$, where D_{iq} is the actual response of person i to item q, $\mathbb{E}(D_{iq}) = P(D_{iq} = 1|\beta_q, \theta_i)$ is the conditional expectation for this person's response to the item, given the model parameters, calculated using the logistic function as shown above. $var(D_{iq})$ can be calculated as $P(D_{iq} = 1|\beta_q, \theta_i)(1 - P(D_{iq} = 1|\beta_q, \theta_i))$. The infit index for item q is then defined as: $Infit_q = \sum_{i=1}^{n}(\frac{var(D_{iq})}{\sum_{i=1}^{n} var(D_{iq})} Z_{iq}^2)$, the outfit index is defined as: $Outfit_q = \sum_{i=1}^{n} \frac{Z_{iq}^2}{n}$. For both indices, values close to 1 indicate good fit, whereas higher values indicate under- and lower values overfitting.

The infit and outfit values for both models are, for the most part, in the acceptable (>0.7 and <1.3) range (see Figure 10). For infit, the 2-PL model consistently outperforms the Rasch model, whereas the outfit values tend to be lower overall, and both models are much closer to each other.

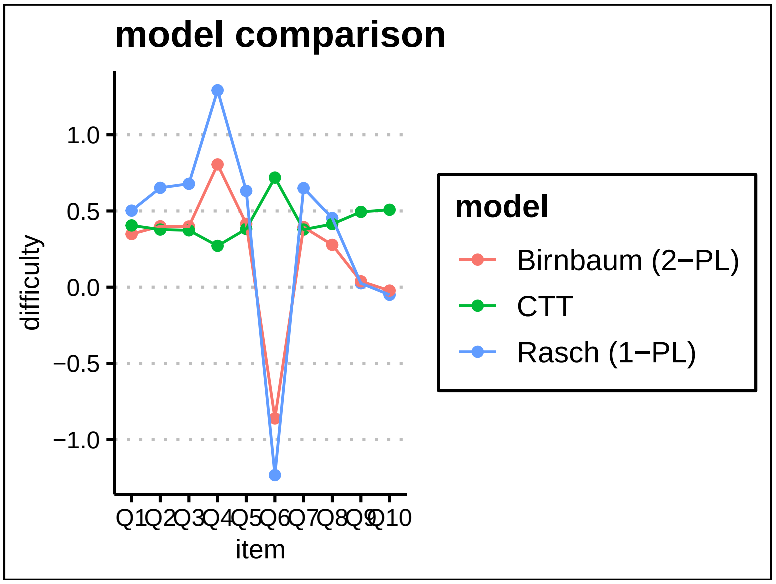


Figure 9: Estimated item difficulties and discriminativities based on CTT, Rasch model and 2-PL model. For both IRT models, error bars indicate standard errors.

Figure 10: Infit and outfit indices for each item in the Rasch and 2-PL models

6 Polytomous IRT model

Procedure outlined in Smyth (2022)

7 Factor models

Following the IRT analysis of the dichotomized data, I went back to the original, non-dichotomized data, to investigate its factorial structure. It has been suggested that the SCS can best be described by two latent factors, one comprising items Q1, Q2, Q3, Q4, and Q10, being related to consequences of sexual behavior and compulsivity to one's lifestyle, and a second one comprising items Q5, Q6, Q7, Q8, and Q9, being related to the compulsivity of one's sexual thoughts without necessarily affecting actual behavior. Using `lavaan::cfa`, I fitted several confirmatory factor analysis (CFA) models to the data to find the latent structure that describes the data best. I specified four candidate latent structures:

The first candidate structure was a unidimensional model, i.e., the data can be explained by a single underlying latent factor.

The second candidate structure was a two-factor correlated-traits model, i.e., two latent factors were specified with item loadings as described above, and correlations between the two latent factors were allowed.

The third candidate structure was a bifactor model, i.e., two latent factors were specified with item loadings as described above, with an additional general factor on which all items load. No correlations between factors were allowed.

The final candidate structure was a hierarchical factor model, which specifies the two item-specific factors, and additionally has them load on a shared second-order factor.

Models were fitted with standardized latent variables, i.e., the variance of all latent factors was fixed to unit. The models were specified in `lavaan` syntax as follows:

```
Unidimensional model:
  xi1 =~ Q1+Q2+Q3+Q4+Q5+Q6+Q7+Q8+Q9+Q10

Correlated-traits model:
  xi1 =~ Q1+Q2+Q3+Q4+Q10
  xi2 =~ Q5+Q6+Q7+Q8+Q9
  xi1 ~~ xi2

Bifactor model:
  G =~ Q1+Q2+Q3+Q4+Q5+Q6+Q7+Q8+Q9+Q10
  xi1 =~ Q1+Q2+Q3+Q4+Q10
  xi2 =~ Q5+Q6+Q7+Q8+Q9
  G ~~ 0*xi1
  G ~~ 0*xi2
  xi1 ~~ 0*xi2


Hierarchical model:
  xi1 =~ Q1+Q2+Q3+Q4+Q10
```

Table 4: Model comparison between CFA models

X	model	Df	AIC	BIC
1	hierarchical	33	85624.40	85759.09
2	bifactor	25	85366.00	85549.67
3	correlated traits	34	85622.40	85750.97
4	unidimensional	35	86593.61	86716.05

```
xi2 =~ Q5+Q6+Q7+Q8+Q9
G =~ xi1+xi2
```

The comparison of fits between the four factor models is shown in 4. Models were compared with respect to the Akaike (AIC) and Bayes-Schwarz (BIC) information criteria. Among the four candidate models, the bifactor model was clearly the preferred one among the four candidate models according to AIC as well as BIC.

The ‘winning’ bifactor model is illustrated in Figure 11. Obviously, the fact that the bifactor model is the preferred option among the four candidate models does not mean that it is necessarily a good description of the data in an absolute sense. To understand the absolute goodness-of-fit (not just compared to other models), there is a range of fit indices that we can consider. In particular, the root mean squared error of approximation (RMSEA), standardized root mean squared residual (SRMR), comparative fit index (CFI), and Tucker-Lewis index (TLI) are informative. For the bifactor model, RMSEA was at 0.073 (RMSEA < 0.08 indicating acceptable, RMSEA < 0.05 indicating good fit by convention), SRMR was at 0.028 (SRMR < 0.05 indicating good fit by convention), CFI was at 0.973 (CFI > 0.95 indicating good fit by convention), and TLI was at 0.95 (TLI > 0.95 indicating acceptable, TLI > 0.97 indicating good fit by convention). Overall, the bifactor model was, therefore, an acceptable to good fit for the SCS data.

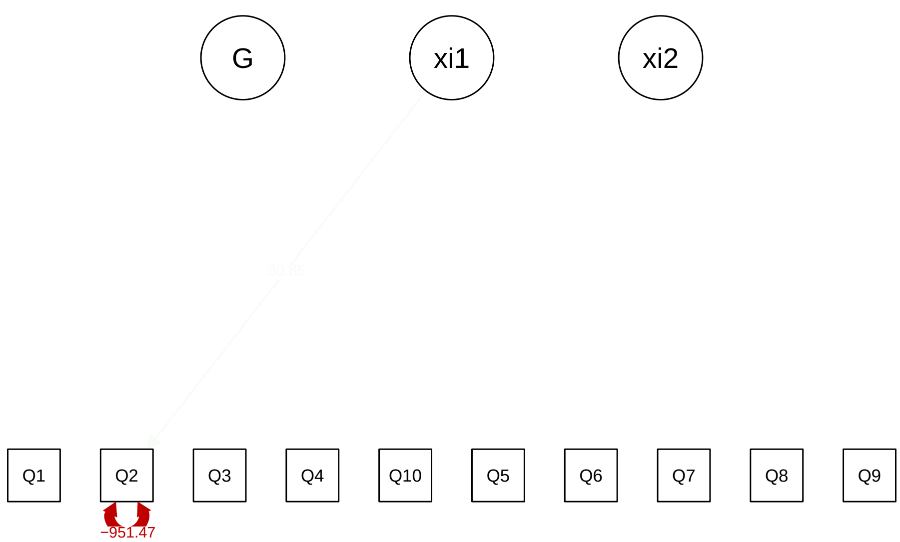


Figure 11: Factor structure and loadings of bifactor model

An open question with respect to the factorial structure is to which subscale item Q10 should belong. While it has been assigned to the first subfactor, its loading on the factor is low (0.18, around half as high as the second-lowest loading item, Q4). To investigate the issue, I fitted two alternative bifactor models, one where item Q10 belonged to the second subfactor, together with items Q5 - Q9, and one where item Q10 constituted its own, third subfactor. However, by all fit indices reported above, the original factor structure was the preferred one, if not by large margins. Looking at the content of the items (see Introduction), we can see that item Q10 is the only item that explicitly involves sex partners and difficulties to find sex partners, while the other items are not specific about sex partners. It could, therefore, be, that responses to item Q10 are not only influenced by sexual compulsivity, but also by a range of social and communicative abilities that might influence whether someone has difficulties finding sex partners or not.

8 Reliability and Unidimensionality

9 Measurement Invariance

10 Theoretical Part: Key differences between IRT and CTT

10.1 Introduction

Unlike some physical quantities, many of the variables of interest in psychology, economics, and other human-centric fields, are latent, i.e., not directly observable. Researchers often try to reconstruct such latent variables by combining several observable variables. In particular, for psychological concepts such as personality traits, a person's score will often be estimated as a combination of item responses in a psychological test. Even though forerunners of psychological tests have been around for centuries, a comprehensive theory of psychological testing only emerged roughly half a century ago. Commonly, Melvin Novick is considered the first author to present a comprehensive account of Classical Test Theory (CTT) (Novick (1965)). Around the same time, a probabilistic view of psychological testing began to emerge, which are now referred to as Item Response Theory (IRT) (Rasch (1960)). It is interesting to note that *Classical* Test Theory does, therefore, not refer to the theory itself being older, but that it rather describes the 'classical' way authors thought about psychological testing from the early 20^{th} century onward, whereas probabilistic approaches became popular only later, when increasing computational capacity made them practical. In the following, I will describe some of the core ideas underlying CTT and ITT, their respective strengths and limitations, and practical applications.

10.2 Core Ideas and Terminology

A test in the sense the CTT as well as IRT use the term is designed to measure a defined *trait* or *state* of a unit (often a person). The trait/state itself is assumed to be unobservable and is captured by combining several *items* that are thought to be reflective of the trait/state. Items in the test-theoretic sense have defined response categories, which can be either dichotomous (e.g., yes/no or solved/unsolved), or ordered (e.g. I… do not agree / agree a little / fully agree), or multinomial (e.g. my favorite color is… red/blue/green). Multinomial items where the responses can not be ordered, nor dichotomized (e.g., into correct and wrong) are more involved from a theoretical point of view and will not be further discussed here. In the end, all items of a test are usually combined by a linear function, i.e., a weighted sum, to form the test *score*.

10.2.1 CTT

The traditional view of psychological tests (Classical Test Theory, CTT) conceived a given person's total score across all items of a test as an additive combination of the person's true score and a testing error: $X_i = \tau_i + \epsilon_i$ (Van der Linden and Hambleton (1997)). (Crocker and Algina (2008)) Reliability, Consistency, Discrimination, Difficulty

10.2.2 IRT

10.3 Strenghts

10.4 Limitations

10.5 Conclusion and Application

11 Analysis code

In the following, the complete analysis code and its output are shown.

```
library(ggplot2)
library(ggthemes)
library(reshape2)
library(readxl)
library(VIM)
```

```
## Loading required package: colorspace

## Loading required package: grid

## VIM is ready to use.

## Suggestions and bug-reports can be submitted at: https://github.com/statistikat/VIM/issues

##
## Attaching package: 'VIM'

## The following object is masked from 'package:datasets':
##
##     sleep
```

```
library(mice)
```

```
##
## Attaching package: 'mice'

## The following object is masked from 'package:stats':
##
##     filter

## The following objects are masked from 'package:base':
##
##     cbind, rbind
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
```

```
##
## Attaching package: 'tidyr'
```

```
## The following object is masked from 'package:reshape2':
##
##     smiths
```

```
library(psych)
```

```
##
## Attaching package: 'psych'

## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

```
library(ggcorrplot)
library(eRm)
```

```
##
## Attaching package: 'eRm'

## The following object is masked from 'package:psych':
##
##     sim.rasch
```

```
library(ltm)
```

```
## Loading required package: MASS

##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##     select

## Loading required package: msm

## Loading required package: polycor

##
## Attaching package: 'polycor'

## The following object is masked from 'package:psych':
##
##     polyserial

##
## Attaching package: 'ltm'

## The following object is masked from 'package:psych':
##
##     factor.scores
```

```
library(patchwork)
```

```
##
## Attaching package: 'patchwork'
```

```
## The following object is masked from 'package:MASS':
##
##     area
```

```
library(difR)
library(semPlot)
library(lavaan)
```

```
## This is lavaan 0.6-11
## lavaan is FREE software! Please report any bugs.
```

```
##
## Attaching package: 'lavaan'
```

```
## The following object is masked from 'package:psych':
##
##     cor2cov
```

```
#####
#part 1: data preparation, descriptive analyses
#####
{
df = read_xlsx("SCS_data.xlsx")
SCS_vars = names(df)[1:10]
#set missing values
print(table(df$gender))
df$gender[df$gender == 3] = NA
df[df==0] = NA

print(unique(df$age))
df$age[df$age >= 100] = NA
mean(df$age,na.rm=T)
median(df$age,na.rm=T)
min(df$age,na.rm=T)
max(df$age,na.rm=T)

sprintf("%i cases are incomplete",sum(!complete.cases(df)))
sprintf("%i cases have incomplete SCS data",sum(!complete.cases(df[,SCS_vars])))


#missing data motifs
# and missing proportion per item
pdf("missingplot.pdf",width = 8, height = 4)
aggr(df[!complete.cases(df[,SCS_vars]),SCS_vars],
     numbers=TRUE, sortVars=TRUE,prop=FALSE,
     labels=SCS_vars,
     ylab=c("#Cases Missing","Pattern"))
box(which = "figure",lwd=2)
dev.off()
```

```
nmissing = rowSums(is.na(df[,SCS_vars]))
table(nmissing[nmissing!=0])
prop.table(table(nmissing[nmissing!=0]))

#missing-at-random analysis
#(check whether missing data points in each variable
#can be jointly predicted by all the other variables)
pvals = data.frame(matrix(ncol = length(SCS_vars), nrow=0))
colnames(pvals) = SCS_vars
for (var in SCS_vars){
  formula = sprintf("I(is.na(%s)) ~ .", var)
  formula0 = sprintf("I(is.na(%s)) ~ 1", var)
  m = summary(glm(formula, data=df[,1:10]))$coefficients
  pvals[var,rownames(m)[2:10]] = m[2:10,"Pr(>|t|)"]
}
min(p.adjust(unlist(pvals), method="fdr"),na.rm=T)


#-> missing at random can be assumed
#remove cases where more than two SCS variables are missing

#15 cases removed
df_clean = df[rowSums(is.na(df[,SCS_vars])) <= 2,]

#use multiple imputation for remaining data
df_clean = complete(mice(df_clean))


#descriptives
df_clean[,1:10] %>% summarise_all(list(mean=mean, median = median,
                                       min = min, max = max)) %>%
  round(1) %>%
  gather(variable, value) %>%
  separate(variable, c("var", "stat"), sep = "\\_") %>%
  spread(var, value) -> descriptives

#fix order of columns in descriptives table
descriptives = descriptives[,c("stat",SCS_vars)]
write.csv(descriptives, "descriptives.csv")
#re-calculate sum score
df_clean$score = rowSums(df_clean[,1:10])

#distribution plot before dichotomization
tmp = melt(
          cbind(data.frame(id=1:nrow(df_clean)),df_clean[,SCS_vars]),
          id.vars="id")
tmp2 = data.frame(table(tmp$variable,tmp$value))
```

```
colnames(tmp2) = c("item","response","Freq")
ggplot(tmp2,aes(x=item, y=Freq, fill=response))+geom_col()+theme_clean()
ggsave("distroplot.pdf",width = 4,height = 2)

#dichotomization
dich = df_clean
dich[,1:10] = data.frame(lapply(df_clean[,1:10],
                                function (x) as.numeric(x > 2)))
dich$score = rowSums(dich[,1:10])

}
```

```
##
##    0    1    2    3
##   13 2295 1053   15
##  [1]  41  50  23  42  36  29  24  35  26  43  21  39  37  64  28  46  34  31  47
## [20]  22  61  16  40  33  30  56  49  51  18  20  45  32  15  27  25  59  58  19
## [39]  14  38  48  44  55 100  65  17  77  57  60  52  53  62  71  78  54  63  67
## [58]  68  72 999  85  69  70  66  84 123  73
```

```
## Warning in plot.aggr(res, ...): not enough vertical space to display frequencies
## (too many combinations)
```

```
##
##  Variables sorted by number of missings:
##  Variable Count
##        Q9    27
##        Q3    22
##        Q4    22
##        Q7    22
##        Q8    21
##        Q6    20
##       Q10    17
##        Q2    16
##        Q1    15
##        Q5    13
##
##  iter imp variable
##   1   1  Q1  Q2  Q3  Q4  Q5  Q6  Q7  Q8  Q9  Q10  gender  age
##   1   2  Q1  Q2  Q3  Q4  Q5  Q6  Q7  Q8  Q9  Q10  gender  age
##   1   3  Q1  Q2  Q3  Q4  Q5  Q6  Q7  Q8  Q9  Q10  gender  age
##   1   4  Q1  Q2  Q3  Q4  Q5  Q6  Q7  Q8  Q9  Q10  gender  age
##   1   5  Q1  Q2  Q3  Q4  Q5  Q6  Q7  Q8  Q9  Q10  gender  age
##   2   1  Q1  Q2  Q3  Q4  Q5  Q6  Q7  Q8  Q9  Q10  gender  age
##   2   2  Q1  Q2  Q3  Q4  Q5  Q6  Q7  Q8  Q9  Q10  gender  age
##   2   3  Q1  Q2  Q3  Q4  Q5  Q6  Q7  Q8  Q9  Q10  gender  age
##   2   4  Q1  Q2  Q3  Q4  Q5  Q6  Q7  Q8  Q9  Q10  gender  age
##   2   5  Q1  Q2  Q3  Q4  Q5  Q6  Q7  Q8  Q9  Q10  gender  age
##   3   1  Q1  Q2  Q3  Q4  Q5  Q6  Q7  Q8  Q9  Q10  gender  age
```

```
##   3   2  Q1  Q2  Q3  Q4  Q5  Q6  Q7  Q8  Q9  Q10  gender  age
##   3   3  Q1  Q2  Q3  Q4  Q5  Q6  Q7  Q8  Q9  Q10  gender  age
##   3   4  Q1  Q2  Q3  Q4  Q5  Q6  Q7  Q8  Q9  Q10  gender  age
##   3   5  Q1  Q2  Q3  Q4  Q5  Q6  Q7  Q8  Q9  Q10  gender  age
##   4   1  Q1  Q2  Q3  Q4  Q5  Q6  Q7  Q8  Q9  Q10  gender  age
##   4   2  Q1  Q2  Q3  Q4  Q5  Q6  Q7  Q8  Q9  Q10  gender  age
##   4   3  Q1  Q2  Q3  Q4  Q5  Q6  Q7  Q8  Q9  Q10  gender  age
##   4   4  Q1  Q2  Q3  Q4  Q5  Q6  Q7  Q8  Q9  Q10  gender  age
##   4   5  Q1  Q2  Q3  Q4  Q5  Q6  Q7  Q8  Q9  Q10  gender  age
##   5   1  Q1  Q2  Q3  Q4  Q5  Q6  Q7  Q8  Q9  Q10  gender  age
##   5   2  Q1  Q2  Q3  Q4  Q5  Q6  Q7  Q8  Q9  Q10  gender  age
##   5   3  Q1  Q2  Q3  Q4  Q5  Q6  Q7  Q8  Q9  Q10  gender  age
##   5   4  Q1  Q2  Q3  Q4  Q5  Q6  Q7  Q8  Q9  Q10  gender  age
##   5   5  Q1  Q2  Q3  Q4  Q5  Q6  Q7  Q8  Q9  Q10  gender  age
```

```
#####
#part 2: CTT-style item analysis
#####
{

  #biserial correlations
  biserial_cor = biserial(dich[,SCS_vars],dich[,SCS_vars])
  ggcorrplot(biserial_cor, type = "lower", lab = TRUE)+theme_clean()
  ggsave("biserial_cor_mat.pdf",width = 6, height = 6)
  #dichotomous item statistics (percent and N correct, discriminativity)
  dich.distro = rbind(as.character(round(100*unlist(lapply(dich[,SCS_vars],
                                                           mean)),1)),
                      as.character(as.integer(unlist(lapply(dich[,SCS_vars],
                                                            sum)))))
  rownames(dich.distro) = c("item easiness\n(percent in category 1)",
                            "number of cases in category 1")

  discrimination = c()
  for (item in 1:10){
    itemname = SCS_vars[item]
    discrimination[itemname] = as.character(round(biserial(
      rowSums(dich[,-item]),dich[,item]),2))
  }

  dich.stats = rbind(dich.distro, discrimination)
  write.csv(dich.stats,"dich_stats.csv")
}
```

```
## Warning in biserialc(x[, j], y[, i], j, i): For x = 1 y = 1 x seems to be
## dichotomous, not continuous

## Warning in biserialc(x[, j], y[, i], j, i): For x = 2 y = 2 x seems to be
## dichotomous, not continuous

## Warning in biserialc(x[, j], y[, i], j, i): For x = 3 y = 3 x seems to be
```

```
## dichotomous, not continuous

## Warning in biserialc(x[, j], y[, i], j, i): For x = 4 y = 4 x seems to be
## dichotomous, not continuous

## Warning in biserialc(x[, j], y[, i], j, i): For x = 5 y = 5 x seems to be
## dichotomous, not continuous

## Warning in biserialc(x[, j], y[, i], j, i): For x = 6 y = 6 x seems to be
## dichotomous, not continuous

## Warning in biserialc(x[, j], y[, i], j, i): For x = 7 y = 7 x seems to be
## dichotomous, not continuous

## Warning in biserialc(x[, j], y[, i], j, i): For x = 8 y = 8 x seems to be
## dichotomous, not continuous

## Warning in biserialc(x[, j], y[, i], j, i): For x = 9 y = 9 x seems to be
## dichotomous, not continuous

## Warning in biserialc(x[, j], y[, i], j, i): For x = 10 y = 10 x seems to be
## dichotomous, not continuous
```

```
#####
#part 2: estimate and analyze Rasch model
#####
{
  #approach 1: eRm
  #prepare data for eRm estimation
  #(just item data in wide format)
  rasch_model_eRm = RM(dich[,SCS_vars])

  #approach 2: ltm
  #constraint fixes item discriminativity to 1
  rasch_model_ltm = rasch(dich[,SCS_vars],
                          constraint = cbind(length(SCS_vars) + 1, 1))
  smr_ltm = summary(rasch_model_ltm)



  #TODO check syntax
  #aproach 3: lavaan
  #modified copy from https://jonathantemplin.com/wp-content/uploads/2022/02/
                     #EPSY906_Example05_Binary_IFA-IRT_Models.nb.html
  lavaansyntax = "

    # loadings/discrimination parameters:
    SCS =~ 1*Q1 + 1*Q2 + 1*Q3 + 1*Q4 + 1*Q5 + 1*Q6 + 1*Q7 + 1*Q8 + 1*Q9 + 1*Q10

    # threshholds use the | operator and start at value 1 after t:
    Q1 | t1; Q2 | t1; Q3 | t1; Q4 | t1; Q5 | t1; Q6 | t1; Q7 | t1;
    Q8 | t1; Q9 | t1;Q10 | t1;
```

```
  # factor mean:
  SCS ~ 0;

    # factor variance:
  SCS ~~ 1*SCS

  "

rasch_model_lavaan = sem(model = lavaansyntax, data =  dich[,SCS_vars],
                         ordered = SCS_vars, mimic = "Mplus",
                         estimator = "WLSMV", std.lv = TRUE,
                         parameterization = "theta")
smr_lavaan = summary(rasch_model_lavaan, fit.measures = TRUE)

convertTheta2IRT = function(lavObject){
  #modified copy from
  #https://jonathantemplin.com/wp-content/uploads/2022/02/
      #EPSY906_Example05_Binary_IFA-IRT_Models.nb.html

  if (!lavObject@Options$parameterization == "theta") {
    stop("your model is not estimated with parameterization='theta'")
    }

  output = inspect(object = lavObject, what = "est")
  if (ncol(output$lambda)>1) { stop("IRT conversion is only valid
           for one dimensional factor models.
           Your model has more than one dimension.")
    }
  a = output$lambda
  b = output$tau/output$lambda
  return(list(a = a, b=b))
}

#make ICC plot function
ICC_plot = function(difficulty, discriminativity = 1){
  if (length(discriminativity)==1){
      discriminativity = rep(discriminativity, length(difficulty))
    }
  df = data.frame(x=seq(-6,6,.01))
  for (i in 1:length(difficulty)){
    df[[SCS_vars[i]]] = logistic(x=df$x, d=difficulty[i],
                                 a=discriminativity[i])
  }

  df = melt(df, id.vars = "x")
  colnames(df)[2] = "item"
  plt=ggplot(df, aes(x = x, y = value, color = item, label = item)) +
```

```
    geom_line() + theme_clean() + xlab("Person parameter") +
    ylab("P(item solved)")
  return(directlabels::direct.label(plt, "last.qp"))
}



#make ICC plots
difficulties_eRm = -rasch_model_eRm$betapar
iccplot_eRm=ICC_plot(difficulties_eRm)+ggtitle("eRm")



#lme4 difficulties are shifted by .42 from eRm difficulties, why?

difficulties_ltm = smr_ltm$coefficients[1:10,"value"]
iccplot_ltm = ICC_plot(difficulties_ltm)+ggtitle("ltm")


difficulties_lavaan =   convertTheta2IRT(lavObject = rasch_model_lavaan)$b

iccplot_lavaan=ICC_plot(difficulties_lavaan)+ggtitle("lavaan")



difficulties = rbind( data.frame(model="eRm",
                          item=factor(SCS_vars),
                          difficulty=as.numeric(difficulties_eRm)),
              data.frame(model="ltm",
                         item=factor(SCS_vars),
                         difficulty=as.numeric(difficulties_ltm)),
              data.frame(model="lavaan",
                         item=factor(SCS_vars),
                         difficulty=as.numeric(difficulties_lavaan)),
              data.frame(model="CTT",
                         item=factor(SCS_vars),
                         difficulty=1-as.numeric(dich.distro[1,])/100))
difficulties_plot = ggplot(difficulties,aes(x=item,y=difficulty,
                                         color=model,group=model)) +
  geom_point() + geom_line() + theme_clean() + ggtitle("model comparison")+
  scale_x_discrete(breaks=SCS_vars,limits=SCS_vars)


difficulties_plot
ggsave("diffcfig.pdf",width = 4,height = 3)

#arrange plots vertically and save
```

```r
iccplot_eRm|iccplot_ltm|iccplot_lavaan

ggsave("iccfig.pdf",width = 12,height = 3)


#compare fits
  #select second line of output (corresponding to marginal MLE)
eRm_fit = IC(person.parameter(rasch_model_eRm))[[1]][2,]

ltm_fit = c()
ltm_fit['value'] = smr_ltm$logLik
ltm_fit['npar'] = 10
ltm_fit['AIC'] = smr_ltm$AIC
ltm_fit['BIC'] = smr_ltm$BIC
ltm_fit['cAIC'] = NA

rasch_model_fits = rbind(eRm_fit,ltm_fit)
rownames(rasch_model_fits) = c("eRm","ltm")
colnames(rasch_model_fits)[1] = "loglik"
write.csv(rasch_model_fits,"rasch_model_fits.csv")



#calculate loss per item

predict_responses = function(item_dffc,person_params,item_discr=1){
  item_dffc = as.numeric(item_dffc)
  if (length(item_discr)==1) item_discr = rep(item_discr,length(item_dffc))
  person_params = as.numeric(person_params)
  preds = matrix(nrow=length(person_params),ncol=length(item_dffc))
  for (p in 1:length(person_params)){
    for(i in 1:length(item_dffc)){
      preds[p,i] = logistic(x=person_params[p],d = item_dffc[i], a = item_discr[i])
  }}
  return(preds)
}


#extract latent person abilities
person_params_eRm = person.parameter(rasch_model_eRm)$theta.table[,"Person Parameter"]
person_params_ltm=factor.scores(rasch_model_ltm,dich[,SCS_vars])[[1]][,"z1"]
person_params_lavaan = as.numeric(predict(rasch_model_lavaan))

#make predictions for individual persons and items
preds_ltm = predict_responses(difficulties_ltm,person_params_ltm)>.5
preds_eRm = predict_responses(difficulties_eRm,person_params_eRm)>.5
preds_lavaan = predict_responses(difficulties_lavaan,person_params_lavaan)>.5
```

```r
#calculate and plot mean 0-1-loss per item
itemloss_eRm = colMeans(dich[,SCS_vars]!=preds_eRm)
itemloss_ltm = colMeans(dich[,SCS_vars]!=preds_ltm)
itemloss_lavaan = colMeans(dich[,SCS_vars]!=preds_lavaan)
itemloss = rbind(data.frame(model="eRm",item=SCS_vars,loss=itemloss_eRm),
                 data.frame(model="ltm",item=SCS_vars,loss=itemloss_ltm),
                 data.frame(model="lavaan",item=SCS_vars,loss=itemloss_lavaan))

ggplot(itemloss,aes(x=item,y=loss,
                        color=model,group=model)) +
  geom_point() + geom_line() + theme_clean() + ggtitle("0-1-loss\nmodel comparison")+
  scale_x_discrete(breaks=SCS_vars,limits=SCS_vars)

ggsave("itemlossfig.pdf",width = 4,height = 3)

}
```

```
## lavaan 0.6-11 ended normally after 7 iterations
##
##   Estimator                                         DWLS
##   Optimization method                             NLMINB
##   Number of model parameters                          10
##
##   Number of observations                            3368
##
## Model Test User Model:
##                                               Standard      Robust
##   Test Statistic                              2415.307    1517.187
##   Degrees of freedom                                45          45
##   P-value (Chi-square)                           0.000       0.000
##   Scaling correction factor                                  1.610
##   Shift parameter                                           17.059
##        simple second-order correction (WLSMV)
##
## Model Test Baseline Model:
##
##   Test statistic                             39116.190   24311.807
##   Degrees of freedom                                45          45
##   P-value                                        0.000       0.000
##   Scaling correction factor                                  1.610
##
## User Model versus Baseline Model:
##
##   Comparative Fit Index (CFI)                    0.939       0.939
##   Tucker-Lewis Index (TLI)                       0.939       0.939
##
```

```
##   Robust Comparative Fit Index (CFI)                             NA
##   Robust Tucker-Lewis Index (TLI)                                NA
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                         0.125       0.099
##   90 Percent confidence interval - lower        0.121       0.094
##   90 Percent confidence interval - upper        0.129       0.103
##   P-value RMSEA <= 0.05                         0.000       0.000
##
##   Robust RMSEA                                                  NA
##   90 Percent confidence interval - lower                        NA
##   90 Percent confidence interval - upper                        NA
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                          0.109       0.109
##
## Weighted Root Mean Square Residual:
##
##   WRMR                                          6.627       6.627
##
## Parameter Estimates:
##
##   Standard errors                          Robust.sem
##   Information                                Expected
##   Information saturated (h1) model       Unstructured
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   SCS =~
##     Q1                1.000
##     Q2                1.000
##     Q3                1.000
##     Q4                1.000
##     Q5                1.000
##     Q6                1.000
##     Q7                1.000
##     Q8                1.000
##     Q9                1.000
##     Q10               1.000
##
## Intercepts:
##                    Estimate  Std.Err  z-value  P(>|z|)
##     SCS               0.000
##    .Q1                0.000
##    .Q2                0.000
##    .Q3                0.000
```

```
##    .Q4                0.000
##    .Q5                0.000
##    .Q6                0.000
##    .Q7                0.000
##    .Q8                0.000
##    .Q9                0.000
##    .Q10               0.000
##
## Thresholds:
##                    Estimate  Std.Err  z-value  P(>|z|)
##     Q1|t1             0.340    0.031   11.015    0.000
##     Q2|t1             0.440    0.031   14.137    0.000
##     Q3|t1             0.457    0.031   14.685    0.000
##     Q4|t1             0.863    0.033   26.391    0.000
##     Q5|t1             0.426    0.031   13.726    0.000
##     Q6|t1            -0.821    0.032  -25.252    0.000
##     Q7|t1             0.438    0.031   14.103    0.000
##     Q8|t1             0.308    0.031    9.984    0.000
##     Q9|t1             0.022    0.031    0.724    0.469
##     Q10|t1           -0.028    0.031   -0.930    0.352
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)
##     SCS               1.000
##    .Q1                1.000
##    .Q2                1.000
##    .Q3                1.000
##    .Q4                1.000
##    .Q5                1.000
##    .Q6                1.000
##    .Q7                1.000
##    .Q8                1.000
##    .Q9                1.000
##    .Q10               1.000
##
## Scales y*:
##                    Estimate  Std.Err  z-value  P(>|z|)
##     Q1                0.707
##     Q2                0.707
##     Q3                0.707
##     Q4                0.707
##     Q5                0.707
##     Q6                0.707
##     Q7                0.707
##     Q8                0.707
##     Q9                0.707
##     Q10               0.707
```

```r
#DIF
{
  data_dif_age = dich[,SCS_vars]
  data_dif_age$age = dich$age > median(dich$age)
  dif_ageL = difLord(data_dif_age,"age",FALSE,"1PL")
  dif_ageR = difRaju(data_dif_age,"age",FALSE, "1PL")


  data_dif_gender= dich[,c(SCS_vars,"gender")]
  dif_genderL = difLord(data_dif_gender,"gender",1, "1PL")
  dif_genderR = difRaju(data_dif_gender,"gender",1, "1PL")

  difstats=data.frame(
    p=-log10(p.adjust(
      c(dif_genderL$p.value,dif_ageL$p.value),method="fdr")),
    item = c(dif_genderL$names,dif_ageL$name),
    groups = c(rep("gender",10),rep("age",10))
    )


  ggplot(difstats, aes(x=item, y=p, group=groups,col=groups)) +
    geom_point() + geom_line() + theme_clean() + ylab("-log10(p)")+
    geom_hline(yintercept=-log10(.05))+scale_x_discrete(breaks=SCS_vars,
                                                        limits=SCS_vars)

  ggsave("DIF_pvals.pdf",width = 4,height = 3)
  }

#alternative model: 2PL
{

  #fit 1PL and 2PL, compare fit
  twoPL_model = ltm(dich[,SCS_vars] ~ z1, IRT.param = TRUE)
  difficulties_2PL = coef(twoPL_model)[,"Dffclt"]
  discriminativities_2PL = coef(twoPL_model)[,"Dscrmn"]
  ICC_2PL = ICC_plot(difficulty = difficulties_2PL,
                     discriminativity = discriminativities_2PL)


  Rasch_vs_twoPL_comparison = anova(rasch_model_ltm, twoPL_model)

  difficulties_1vs2PL = rbind( data.frame(model="Rasch (1-PL)",
                                          item=factor(SCS_vars),
                                          difficulty=as.numeric(difficulties_ltm)),
                               data.frame(model="Birnbaum (2-PL)",
                                          item=factor(SCS_vars),
                                          difficulty=as.numeric(difficulties_2PL)),
```

```
                        data.frame(model="CTT",
                                   item=factor(SCS_vars),
                                   difficulty=as.numeric(dich.distro[1,])/100))
ggplot(difficulties_1vs2PL,aes(x=item,y=difficulty,
                                                     color=model,group=model)) +
  geom_point() + geom_line() + theme_clean() + ggtitle("model comparison")+
  scale_x_discrete(breaks=SCS_vars,limits=SCS_vars)
ggsave("difficulties_plot_2PL.pdf",width = 4,height = 3)


#calculate item-wise infit and outfit

get_outfit = function(ltm_model){
  X=ltm_model$X
  personscores = factor.scores(ltm_model,X)[[1]][,"z1"]
  dffc = coef(ltm_model)[,"Dffclt"]
  discr = coef(ltm_model)[,"Dscrmn"]
  expected = predict_responses(dffc, personscores, discr)
  var_X = expected * (1-expected)
  Z_ij = (X-expected)/sqrt(var_X)
  chisq = colSums(Z_ij**2)
  #divide chisq by n
  return(chisq/nrow(ltm_model$X))
    }

get_infit = function(ltm_model){
  X=ltm_model$X
  personscores = factor.scores(ltm_model,X)[[1]][,"z1"]
  dffc = coef(ltm_model)[,"Dffclt"]
  discr = coef(ltm_model)[,"Dscrmn"]
  expected = predict_responses(dffc, personscores, discr)
  var_X = expected * (1-expected)
  Z_ij = (X-expected)/sqrt(var_X)
  infit = c()
  for (i in 1:length(dffc)){
    infit[i] = sum((var_X[,i] * (Z_ij[,i]**2))/sum(var_X[,i]))}
  return(infit)
}

outfit_rasch = get_outfit(rasch_model_ltm)
outfit_2PL = get_outfit(twoPL_model)
infit_rasch = get_infit(rasch_model_ltm)
infit_2PL = get_infit(twoPL_model)

inoutfit = rbind(data.frame(model="Rasch",fit="outfit",
                            item=names(outfit_rasch), value=outfit_rasch),
                 data.frame(model="2-PL",fit="outfit",
```

```
                              item=names(outfit_2PL), value=outfit_2PL),
                   data.frame(model="Rasch",fit="infit",
                              item=names(outfit_rasch),value=infit_rasch),
                   data.frame(model="2-PL",fit="infit",
                              item=names(outfit_2PL),value=infit_2PL))

  ggplot(inoutfit,aes(x=item,y=value,
                                 color=model,group=model)) + facet_wrap(~fit)+
    geom_point() + geom_line() + theme_clean() + ggtitle("model comparison")+
    scale_x_discrete(breaks=SCS_vars,limits=SCS_vars)
  ggsave("inoutfit_plot_2PL.pdf",width = 8,height = 3)

}

#factorial models:
{

  covdat = cov(df_clean[,SCS_vars])
  N=nrow(df_clean)


  #unidimensional model
  unidimensional_model <- '
  xi1 =~ Q1+Q2+Q3+Q4+Q5+Q6+Q7+Q8+Q9+Q10

'

  unidimensional_cfa <- cfa(unidimensional_model,
                    sample.cov=covdat,
                    sample.nobs=N,
                    std.lv=T)


  #correlated traits
  correlated_traits_model <- '
  xi1 =~ Q1+Q2+Q3+Q4+Q10
  xi2 =~ Q5+Q6+Q7+Q8+Q9
  xi1 ~~ xi2
'

  correlated_traits_cfa <- cfa(correlated_traits_model,
                               sample.cov=covdat,
                               sample.nobs=N,
                               std.lv=T)


  #bifactor model (general factor and two item-specific factors)
```

```
  bifactor_model <- '
  G =~ Q1+Q2+Q3+Q4+Q5+Q6+Q7+Q8+Q9+Q10
  xi1 =~ Q1+Q2+Q3+Q4+Q10
  xi2 =~ Q5+Q6+Q7+Q8+Q9
  G ~~ 0*xi1
  G ~~ 0*xi2
  xi1 ~~ 0*xi2
'

  bifactor_cfa <- cfa(bifactor_model,
            sample.cov=covdat,
            sample.nobs=N,
            std.lv=T)





  #hierarchical model

  hierarchical_model <- '
  xi1 =~ Q1+Q2+Q3+Q4+Q10
  xi2 =~ Q5+Q6+Q7+Q8+Q9
  G =~ xi1+xi2
'

  hierarchical_cfa <- cfa(hierarchical_model,
                    sample.cov=covdat,
                    sample.nobs=N,
                    std.lv=T)

  #
  smr_hierarchical = summary(hierarchical_cfa, fit=T)$FIT
  smr_bifactor = summary(bifactor_cfa, fit=T)$FIT
  smr_correlated_traits = summary(correlated_traits_cfa, fit=T)$FIT
  smr_unidimensional = summary(unidimensional_cfa, fit=T)$FIT
#
#   cfaaov_df = data.frame(model=c("hierarchical","bifactor","correlated traits",
#                                  "unidimensional"),
#                          Df = c(smr_hierarchical["df"],
#                                 smr_bifactor["df"],
#                                 smr_correlated_traits["df"],
#                                 smr_unidimensional["df"]),
#                          AIC = c(smr_hierarchical["aic"],
#                                  smr_bifactor["aic"],
#                                  smr_correlated_traits["aic"],
#                                  smr_unidimensional["aic"]),
```

```
#                          BIC = c(smr_hierarchical["bic"],
#                                  smr_bifactor["bic"],
#                                  smr_correlated_traits["bic"],
#                                  smr_unidimensional["bic"]))
#
#   write.csv(cfaaov_df,"cfaaov_df.csv")



  pdf("semplot_bifactor.pdf", width = 8,height = 4)
  semPaths(bifactor_cfa, "std")
  dev.off()




  #fit alternative bifactor model (item Q10 belongs to subscale 2)
  bifactor_model_2 <- '
  G =~ Q1+Q2+Q3+Q4+Q5+Q6+Q7+Q8+Q9+Q10
  xi1 =~ Q1+Q2+Q3+Q4
  xi2 =~ Q5+Q6+Q7+Q8+Q9+Q10
  xi3 =~ Q10
  G ~~ 0*xi1
  G ~~ 0*xi2
  xi1 ~~ 0*xi2
  xi1 ~~ 0*xi3
  xi2 ~~ 0*xi3
  G ~~ 0*xi3

'

  bifactor_cfa_2 <- cfa(bifactor_model_2,
                        sample.cov=covdat,
                        sample.nobs=N,
                        std.lv=T)

  #fit alternative bifactor model (item Q10 is its own subscale)
  #-> does not converge
  bifactor_model_3 <- '
  G =~ Q1+Q2+Q3+Q4+Q5+Q6+Q7+Q8+Q9+Q10
  xi1 =~ Q1+Q2+Q3+Q4
  xi2 =~ Q5+Q6+Q7+Q8+Q9
  G ~~ 0*xi1
  G ~~ 0*xi2
  xi1 ~~ 0*xi2
```

```
'

bifactor_cfa_3 <- cfa(bifactor_model_3,
                      sample.cov=covdat,
                      sample.nobs=N,
                      std.lv=T)


smr_bif1=summary(bifactor_cfa,fit=T)$FIT
smr_bif2=summary(bifactor_cfa_2,fit=T)$FIT
smr_bif3=summary(bifactor_cfa_3,fit=T)$FIT

print(rbind(smr_bif1,smr_bif2,smr_bif3))

}
```

```
## Warning in lavaan::lavaan(model = bifactor_model, sample.cov = covdat, sample.nobs = N, : la
##     the optimizer warns that a solution has NOT been found!

## Warning in lav_model_vcov(lavmodel = lavmodel, lavsamplestats = lavsamplestats, : lavaan WAR
##     Could not compute standard errors! The information matrix could
##     not be inverted. This may be a symptom that the model is not
##     identified.

## lavaan 0.6-11 ended normally after 30 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of model parameters                        22
##
##   Number of observations                          3368
##
## Model Test User Model:
##
##   Test statistic                               872.955
##   Degrees of freedom                                33
##   P-value (Chi-square)                           0.000
##
## Model Test Baseline Model:
##
##   Test statistic                             16319.834
##   Degrees of freedom                                45
##   P-value                                        0.000
##
## User Model versus Baseline Model:
##
##   Comparative Fit Index (CFI)                    0.948
##   Tucker-Lewis Index (TLI)                       0.930
##
```

```
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)            -42760.187
##   Loglikelihood unrestricted model (H1)    -42323.709
##
##   Akaike (AIC)                              85564.373
##   Bayesian (BIC)                            85699.059
##   Sample-size adjusted Bayesian (BIC)       85629.155
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                         0.087
##   90 Percent confidence interval - lower        0.082
##   90 Percent confidence interval - upper        0.092
##   P-value RMSEA <= 0.05                         0.000
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                          0.041
##
## Parameter Estimates:
##
##   Standard errors                            Standard
##   Information                                Expected
##   Information saturated (h1) model         Structured
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   xi1 =~
##     Q1                0.316       NA
##     Q2                0.364       NA
##     Q3                0.380       NA
##     Q4                0.307       NA
##     Q10               0.249       NA
##   xi2 =~
##     Q5                0.335       NA
##     Q6                0.224       NA
##     Q7                0.367       NA
##     Q8                0.395       NA
##     Q9                0.314       NA
##   G =~
##     xi1               2.156       NA
##     xi2               2.157       NA
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)
##    .Q1                0.612       NA
##    .Q2                0.400       NA
```

```
##    .Q3                0.347       NA
##    .Q4                0.541       NA
##    .Q10               1.057       NA
##    .Q5                0.609       NA
##    .Q6                0.650       NA
##    .Q7                0.380       NA
##    .Q8                0.299       NA
##    .Q9                0.783       NA
##    .xi1               1.000
##    .xi2               1.000
##     G                 1.000
##
## lavaan 0.6-11 did NOT end normally after 10000 iterations
## ** WARNING ** Estimates below are most likely unreliable
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of model parameters                        30
##
##   Number of observations                          3368
##
## Model Test User Model:
##
##   Test statistic                                    NA
##   Degrees of freedom                                NA

## Warning in .local(object, ...): lavaan WARNING: fit measures not available if model did not

##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Latent Variables:
##                    Estimate   Std.Err  z-value  P(>|z|)
##   G =~
##     Q1                0.714       NA
##     Q2                0.830       NA
##     Q3                0.901       NA
##     Q4                0.743       NA
##     Q5                0.705       NA
##     Q6                0.488       NA
##     Q7                0.714       NA
##     Q8                0.764       NA
##     Q9                0.628       NA
##     Q10               0.625       NA
##   xi1 =~
```

```
##     Q1                 0.004       NA
##     Q2                33.130       NA
##     Q3                 0.001       NA
##     Q4                -0.001       NA
##     Q10               -0.003       NA
##   xi2 =~
##     Q5                 0.353       NA
##     Q6                 0.197       NA
##     Q7                 0.501       NA
##     Q8                 0.578       NA
##     Q9                 0.397       NA
##
## Covariances:
##                    Estimate   Std.Err  z-value  P(>|z|)
##   G ~~
##     xi1                0.000
##     xi2                0.000
##   xi1 ~~
##     xi2                0.000
##
## Variances:
##                    Estimate   Std.Err  z-value  P(>|z|)
##    .Q1                 0.666       NA
##    .Q2             -1097.160       NA
##    .Q3                 0.351       NA
##    .Q4                 0.521       NA
##    .Q5                 0.624       NA
##    .Q6                 0.656       NA
##    .Q7                 0.379       NA
##    .Q8                 0.262       NA
##    .Q9                 0.787       NA
##    .Q10                1.016       NA
##     G                  1.000
##     xi1                1.000
##     xi2                1.000
##
## lavaan 0.6-11 ended normally after 20 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of model parameters                        21
##
##   Number of observations                          3368
##
## Model Test User Model:
##
##   Test statistic                               872.955
##   Degrees of freedom                                34
```

```
##   P-value (Chi-square)                           0.000
##
## Model Test Baseline Model:
##
##   Test statistic                             16319.834
##   Degrees of freedom                                45
##   P-value                                        0.000
##
## User Model versus Baseline Model:
##
##   Comparative Fit Index (CFI)                    0.948
##   Tucker-Lewis Index (TLI)                       0.932
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)             -42760.187
##   Loglikelihood unrestricted model (H1)     -42323.709
##
##   Akaike (AIC)                               85562.373
##   Bayesian (BIC)                             85690.937
##   Sample-size adjusted Bayesian (BIC)        85624.210
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.086
##   90 Percent confidence interval - lower         0.081
##   90 Percent confidence interval - upper         0.091
##   P-value RMSEA <= 0.05                          0.000
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                           0.041
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   xi1 =~
##     Q1                0.751    0.017   43.821    0.000
##     Q2                0.865    0.016   54.304    0.000
##     Q3                0.904    0.016   57.393    0.000
##     Q4                0.729    0.016   44.793    0.000
##     Q10               0.592    0.020   29.291    0.000
##   xi2 =~
```

```
##     Q5                0.797    0.017   45.885    0.000
##     Q6                0.532    0.016   32.984    0.000
##     Q7                0.871    0.016   55.446    0.000
##     Q8                0.938    0.016   60.409    0.000
##     Q9                0.745    0.019   40.038    0.000
##
## Covariances:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   xi1 ~~
##     xi2               0.823    0.008   96.991    0.000
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)
##    .Q1                0.612    0.017   36.174    0.000
##    .Q2                0.400    0.013   30.990    0.000
##    .Q3                0.347    0.012   28.406    0.000
##    .Q4                0.541    0.015   35.846    0.000
##    .Q10               1.057    0.027   39.309    0.000
##    .Q5                0.609    0.017   35.903    0.000
##    .Q6                0.650    0.017   38.928    0.000
##    .Q7                0.380    0.012   30.985    0.000
##    .Q8                0.299    0.011   26.365    0.000
##    .Q9                0.783    0.021   37.572    0.000
##     xi1               1.000
##     xi2               1.000
##
## lavaan 0.6-11 ended normally after 16 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of model parameters                        20
##
##   Number of observations                          3368
##
## Model Test User Model:
##
##   Test statistic                              1857.314
##   Degrees of freedom                                35
##   P-value (Chi-square)                           0.000
##
## Model Test Baseline Model:
##
##   Test statistic                             16319.834
##   Degrees of freedom                                45
##   P-value                                        0.000
##
## User Model versus Baseline Model:
##
```

```
##   Comparative Fit Index (CFI)                    0.888
##   Tucker-Lewis Index (TLI)                       0.856
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)             -43252.366
##   Loglikelihood unrestricted model (H1)     -42323.709
##
##   Akaike (AIC)                               86544.732
##   Bayesian (BIC)                             86667.174
##   Sample-size adjusted Bayesian (BIC)        86603.625
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.124
##   90 Percent confidence interval - lower         0.120
##   90 Percent confidence interval - upper         0.129
##   P-value RMSEA <= 0.05                          0.000
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                           0.053
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   xi1 =~
##     Q1                0.706    0.017   40.901    0.000
##     Q2                0.808    0.016   49.890    0.000
##     Q3                0.848    0.016   53.005    0.000
##     Q4                0.694    0.016   42.484    0.000
##     Q5                0.779    0.017   44.869    0.000
##     Q6                0.530    0.016   33.122    0.000
##     Q7                0.821    0.016   51.371    0.000
##     Q8                0.876    0.016   55.116    0.000
##     Q9                0.717    0.019   38.440    0.000
##     Q10               0.601    0.020   30.192    0.000
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)
##    .Q1                0.678    0.018   38.052    0.000
##    .Q2                0.495    0.014   35.662    0.000
##    .Q3                0.445    0.013   34.423    0.000
```

```
##    .Q4                 0.591    0.016   37.723    0.000
##    .Q5                 0.638    0.017   37.165    0.000
##    .Q6                 0.652    0.017   39.292    0.000
##    .Q7                 0.465    0.013   35.109    0.000
##    .Q8                 0.411    0.012   33.396    0.000
##    .Q9                 0.825    0.021   38.505    0.000
##    .Q10                1.046    0.026   39.636    0.000
##     xi1                1.000
```

```
## Warning in lav_model_vcov(lavmodel = lavmodel, lavsamplestats = lavsamplestats, : lavaan WAR
##     Could not compute standard errors! The information matrix could
##     not be inverted. This may be a symptom that the model is not
##     identified.
```

```
## Warning in lav_model_vcov(lavmodel = lavmodel, lavsamplestats = lavsamplestats, : lavaan WAR
##     Could not compute standard errors! The information matrix could
##     not be inverted. This may be a symptom that the model is not
##     identified.
```

```
## Warning in lav_object_post_check(object): lavaan WARNING: some estimated ov
## variances are negative
```

```
## lavaan 0.6-11 did NOT end normally after 10000 iterations
## ** WARNING ** Estimates below are most likely unreliable
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of model parameters                        30
##
##   Number of observations                          3368
##
## Model Test User Model:
##
##   Test statistic                                    NA
##   Degrees of freedom                                NA
```

```
## Warning in .local(object, ...): lavaan WARNING: fit measures not available if model did not
```

```
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Latent Variables:
##                    Estimate   Std.Err  z-value  P(>|z|)
##   G =~
##     Q1                 0.714       NA
##     Q2                 0.830       NA
```

```
##     Q3                 0.901       NA
##     Q4                 0.743       NA
##     Q5                 0.705       NA
##     Q6                 0.488       NA
##     Q7                 0.714       NA
##     Q8                 0.764       NA
##     Q9                 0.628       NA
##     Q10                0.625       NA
##   xi1 =~
##     Q1                 0.004       NA
##     Q2                33.130       NA
##     Q3                 0.001       NA
##     Q4                -0.001       NA
##     Q10               -0.003       NA
##   xi2 =~
##     Q5                 0.353       NA
##     Q6                 0.197       NA
##     Q7                 0.501       NA
##     Q8                 0.578       NA
##     Q9                 0.397       NA
##
## Covariances:
##                    Estimate   Std.Err  z-value  P(>|z|)
##   G ~~
##     xi1                0.000
##     xi2                0.000
##   xi1 ~~
##     xi2                0.000
##
## Variances:
##                    Estimate   Std.Err  z-value  P(>|z|)
##    .Q1                 0.666       NA
##    .Q2             -1097.160       NA
##    .Q3                 0.351       NA
##    .Q4                 0.521       NA
##    .Q5                 0.624       NA
##    .Q6                 0.656       NA
##    .Q7                 0.379       NA
##    .Q8                 0.262       NA
##    .Q9                 0.787       NA
##    .Q10                1.016       NA
##     G                  1.000
##     xi1                1.000
##     xi2                1.000
##
## lavaan 0.6-11 ended normally after 59 iterations
##
##   Estimator                                         ML
```

```
##   Optimization method                           NLMINB
##   Number of model parameters                        31
##
##   Number of observations                          3368
##
## Model Test User Model:
##
##   Test statistic                               597.783
##   Degrees of freedom                                24
##   P-value (Chi-square)                           0.000
##
## Model Test Baseline Model:
##
##   Test statistic                             16319.834
##   Degrees of freedom                                45
##   P-value                                        0.000
##
## User Model versus Baseline Model:
##
##   Comparative Fit Index (CFI)                    0.965
##   Tucker-Lewis Index (TLI)                       0.934
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)             -42622.601
##   Loglikelihood unrestricted model (H1)     -42323.709
##
##   Akaike (AIC)                               85307.201
##   Bayesian (BIC)                             85496.985
##   Sample-size adjusted Bayesian (BIC)        85398.484
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.084
##   90 Percent confidence interval - lower         0.078
##   90 Percent confidence interval - upper         0.090
##   P-value RMSEA <= 0.05                          0.000
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                           0.033
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
```

```
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   G =~
##     Q1                0.791       NA
##     Q2                0.850       NA
##     Q3                0.932       NA
##     Q4                0.735       NA
##     Q5                0.669       NA
##     Q6                0.463       NA
##     Q7                0.685       NA
##     Q8                0.733       NA
##     Q9                0.594       NA
##     Q10               0.585       NA
##   xi1 =~
##     Q1                0.568       NA
##     Q2                0.087       NA
##     Q3               -0.186       NA
##     Q4               -0.154       NA
##   xi2 =~
##     Q5                0.418       NA
##     Q6                0.253       NA
##     Q7                0.539       NA
##     Q8                0.606       NA
##     Q9                0.449       NA
##     Q10               0.128       NA
##   xi3 =~
##     Q10               0.635       NA
##
## Covariances:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   G ~~
##     xi1               0.000
##     xi2               0.000
##   xi1 ~~
##     xi2               0.000
##     xi3               0.000
##   xi2 ~~
##     xi3               0.000
##   G ~~
##     xi3               0.000
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)
##    .Q1                0.228       NA
##    .Q2                0.417       NA
##    .Q3                0.261       NA
##    .Q4                0.510       NA
##    .Q5                0.622       NA
```

```
##    .Q6                 0.654       NA
##    .Q7                 0.379       NA
##    .Q8                 0.276       NA
##    .Q9                 0.784       NA
##    .Q10                0.647       NA
##     G                  1.000
##     xi1                1.000
##     xi2                1.000
##     xi3                1.000
##
## lavaan 0.6-11 ended normally after 8799 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of model parameters                        29
##
##   Number of observations                          3368
##
## Model Test User Model:
##
##   Test statistic                               658.531
##   Degrees of freedom                                26
##   P-value (Chi-square)                           0.000
##
## Model Test Baseline Model:
##
##   Test statistic                             16319.834
##   Degrees of freedom                                45
##   P-value                                        0.000
##
## User Model versus Baseline Model:
##
##   Comparative Fit Index (CFI)                    0.961
##   Tucker-Lewis Index (TLI)                       0.933
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)             -42652.974
##   Loglikelihood unrestricted model (H1)     -42323.709
##
##   Akaike (AIC)                               85363.949
##   Bayesian (BIC)                             85541.489
##   Sample-size adjusted Bayesian (BIC)        85449.343
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.085
##   90 Percent confidence interval - lower         0.079
```

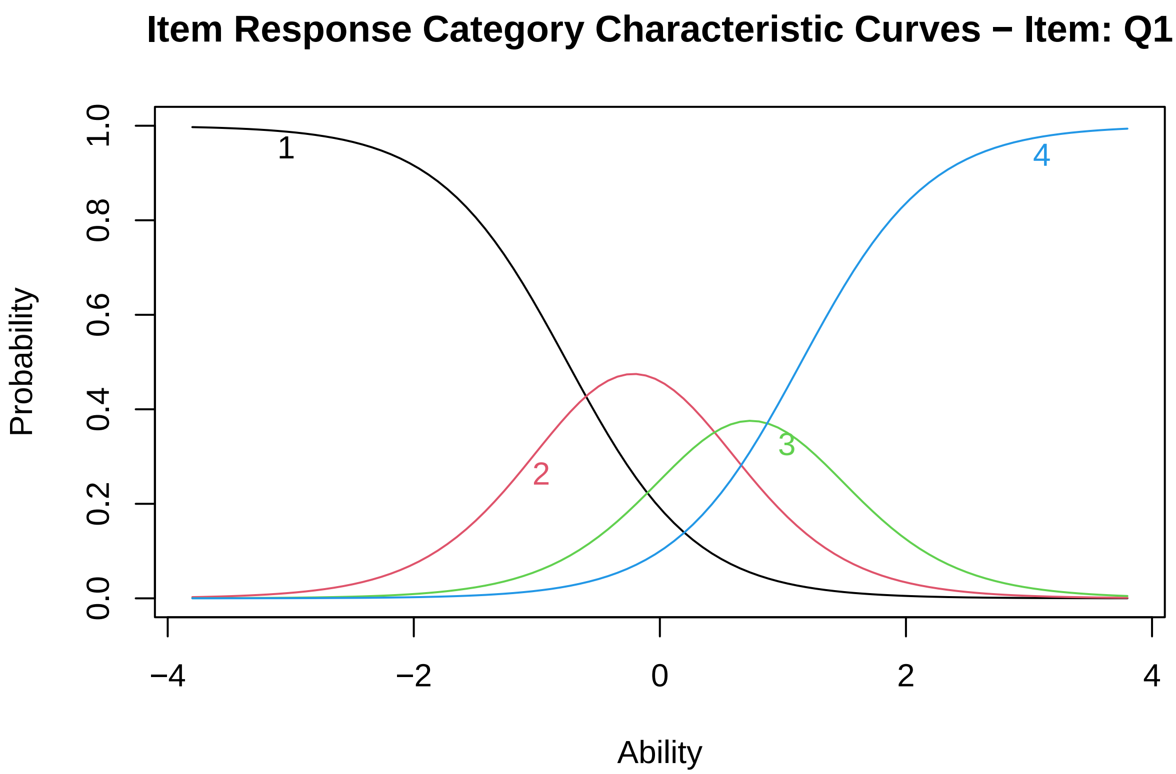
```
##   90 Percent confidence interval - upper         0.091
##   P-value RMSEA <= 0.05                          0.000
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                           0.035
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   G =~
##     Q1                0.713       NA
##     Q2                0.793       NA
##     Q3                0.900       NA
##     Q4                0.742       NA
##     Q5                0.717       NA
##     Q6                0.500       NA
##     Q7                0.718       NA
##     Q8                0.766       NA
##     Q9                0.626       NA
##     Q10               0.615       NA
##   xi1 =~
##     Q1                0.007       NA
##     Q2               23.363       NA
##     Q3                0.003       NA
##     Q4                0.000       NA
##   xi2 =~
##     Q5                0.332       NA
##     Q6                0.177       NA
##     Q7                0.493       NA
##     Q8                0.580       NA
##     Q9                0.398       NA
##
## Covariances:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   G ~~
##     xi1               0.000
##     xi2               0.000
##   xi1 ~~
##     xi2               0.000
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)
```

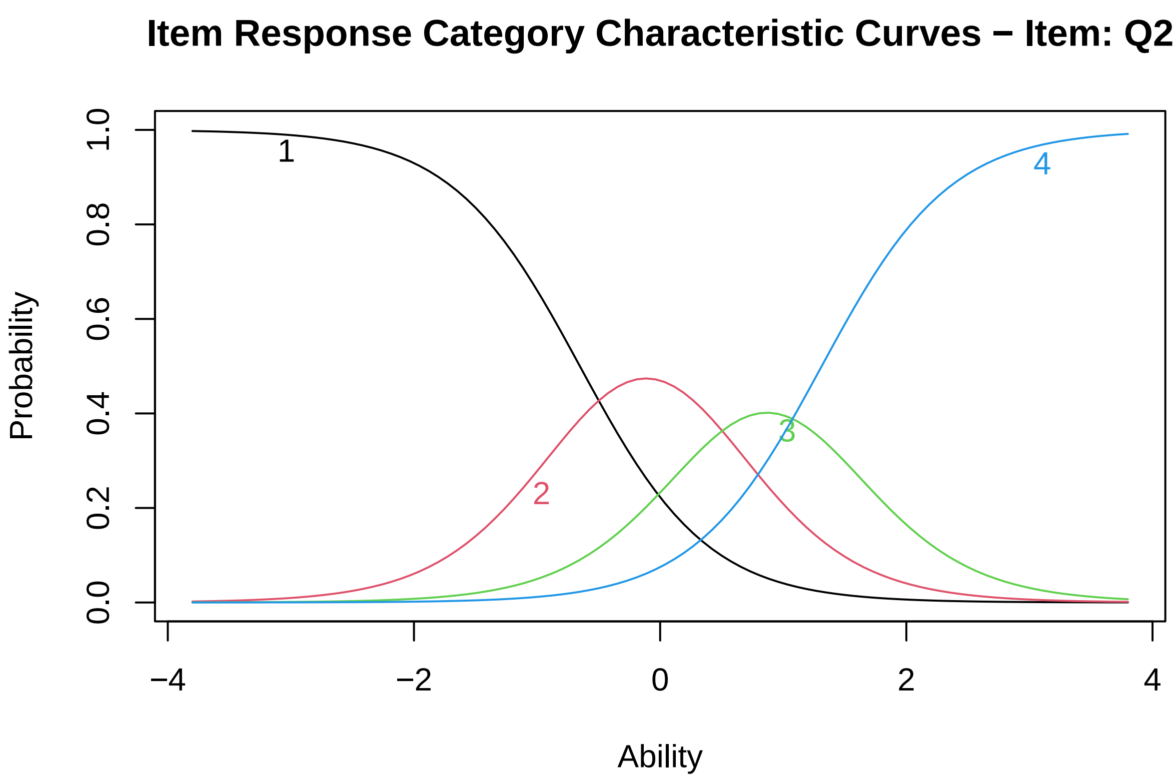
```
##    .Q1                0.667       NA
##    .Q2             -545.320       NA
##    .Q3                0.354       NA
##    .Q4                0.522       NA
##    .Q5                0.620       NA
##    .Q6                0.652       NA
##    .Q7                0.381       NA
##    .Q8                0.257       NA
##    .Q9                0.788       NA
##    .Q10               1.029       NA
##     G                 1.000
##     xi1               1.000
##     xi2               1.000
##
##           npar       fmin    chisq df pvalue baseline.chisq baseline.df
## smr_bif2   31 0.08874453 597.7831 24      0       16319.83          45
## smr_bif3   29 0.09776288 658.5308 26      0       16319.83          45
##           baseline.pvalue       cfi       tli      logl unrestricted.logl
## smr_bif2                0 0.9647441 0.9338953 -42622.60         -42323.71
## smr_bif3                0 0.9611344 0.9327327 -42652.97         -42323.71
##                aic      bic ntotal     bic2      rmsea rmsea.ci.lower
## smr_bif2 85307.20 85496.99   3368 85398.48 0.08425239     0.07847940
## smr_bif3 85363.95 85541.49   3368 85449.34 0.08499006     0.07944116
##           rmsea.ci.upper rmsea.pvalue       srmr
## smr_bif2     0.09016582 5.329071e-15 0.03256561
## smr_bif3     0.09066755 5.329071e-15 0.03470250
```

```
#reliability, unidimensionality
{



}
```

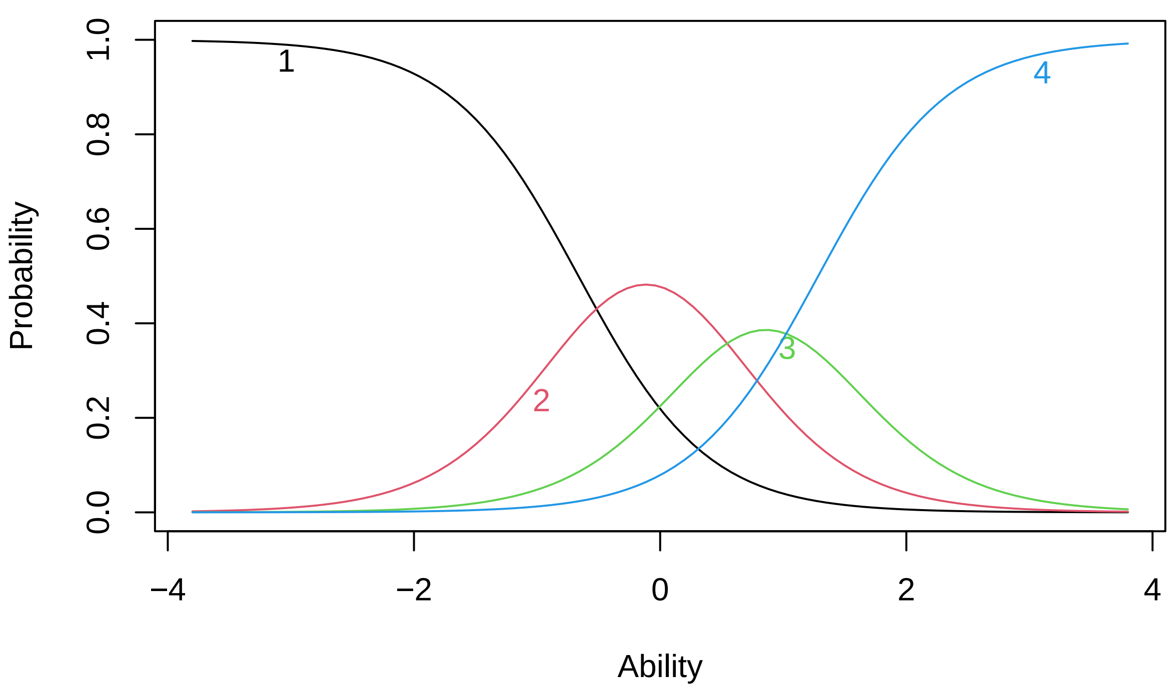
```
## NULL
```

```
#polytomous IRT model
{
  grm_model = grm(df_clean[,SCS_vars],constrained=T)
  plot(grm_model)
}
```

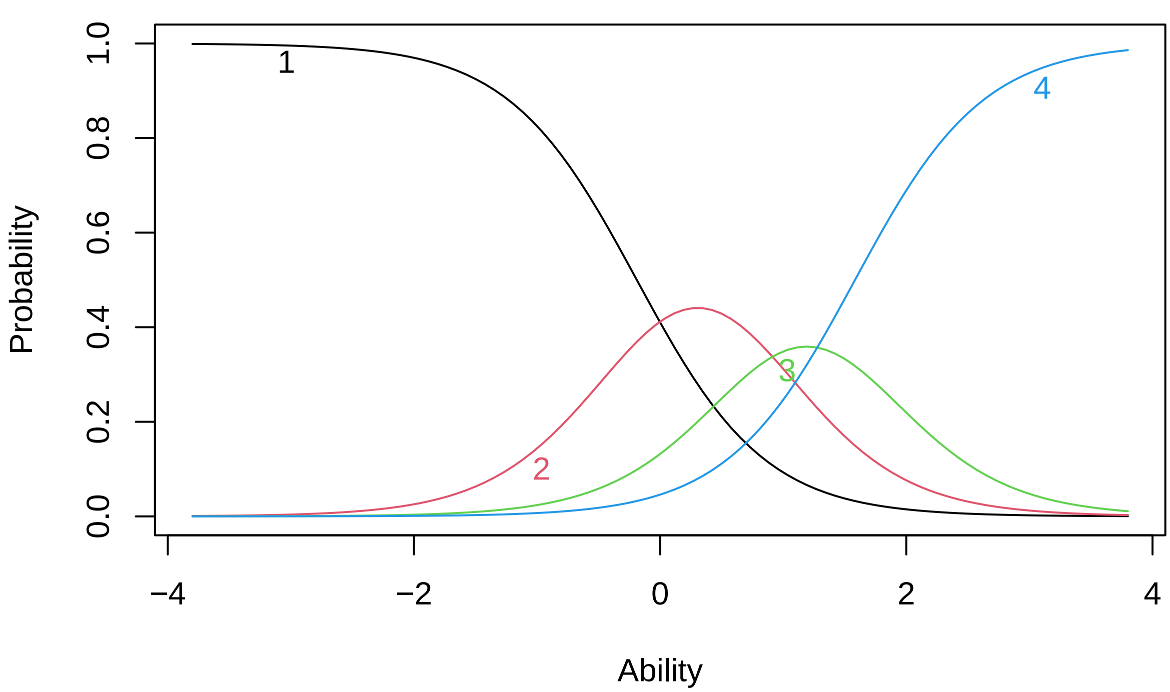

Item Response Category Characteristic Curves − Item: Q1
Probability
0.0
0.2
0.4
0.6
0.8
1.0
1
2
3
4
−4
−2
0
2
4
Ability

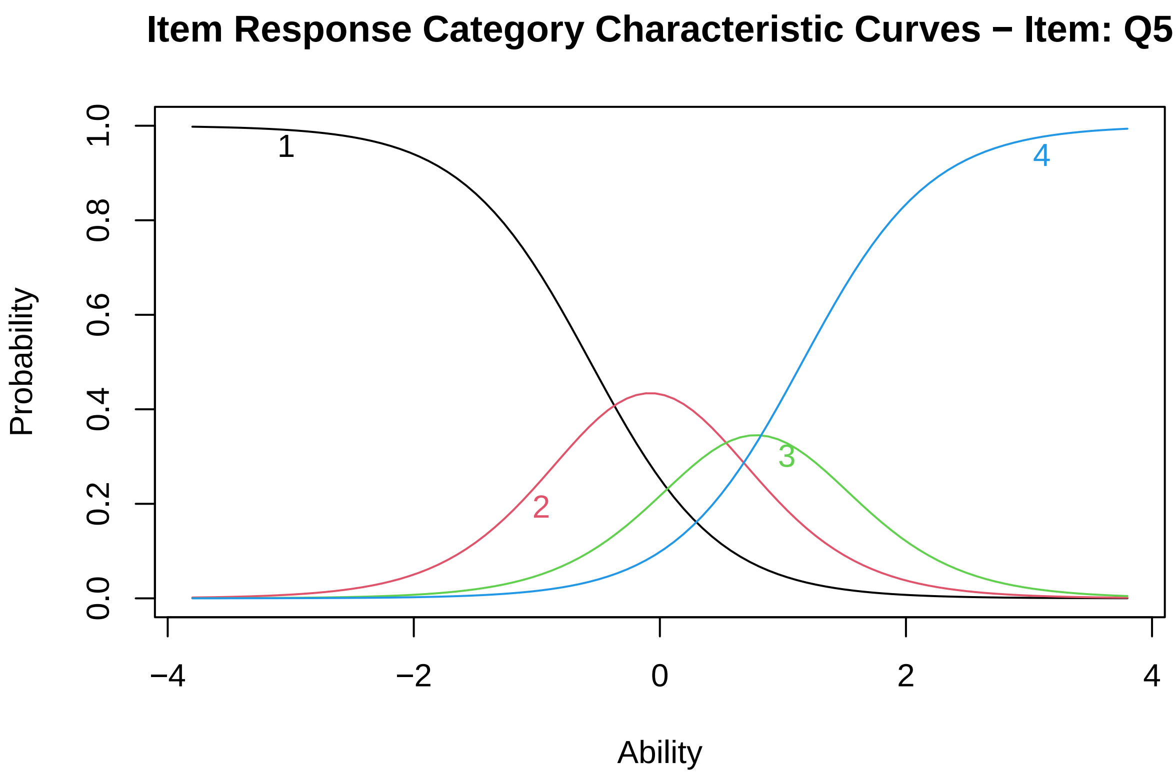

Item Response Category Characteristic Curves − Item: Q2
Probability
0.0
0.2
0.4
0.6
0.8
1.0
1
2
3
4
−4
−2
0
2
4
Ability

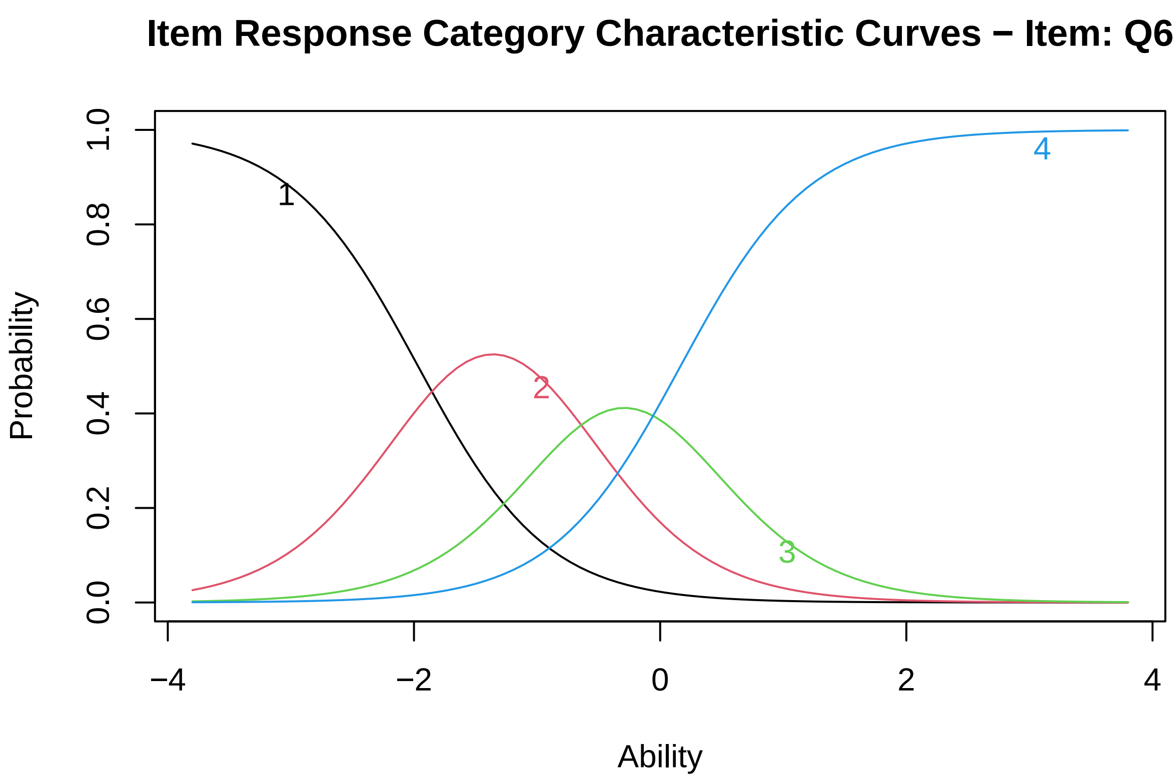
Item Response Category Characteristic Curves − Item: Q3

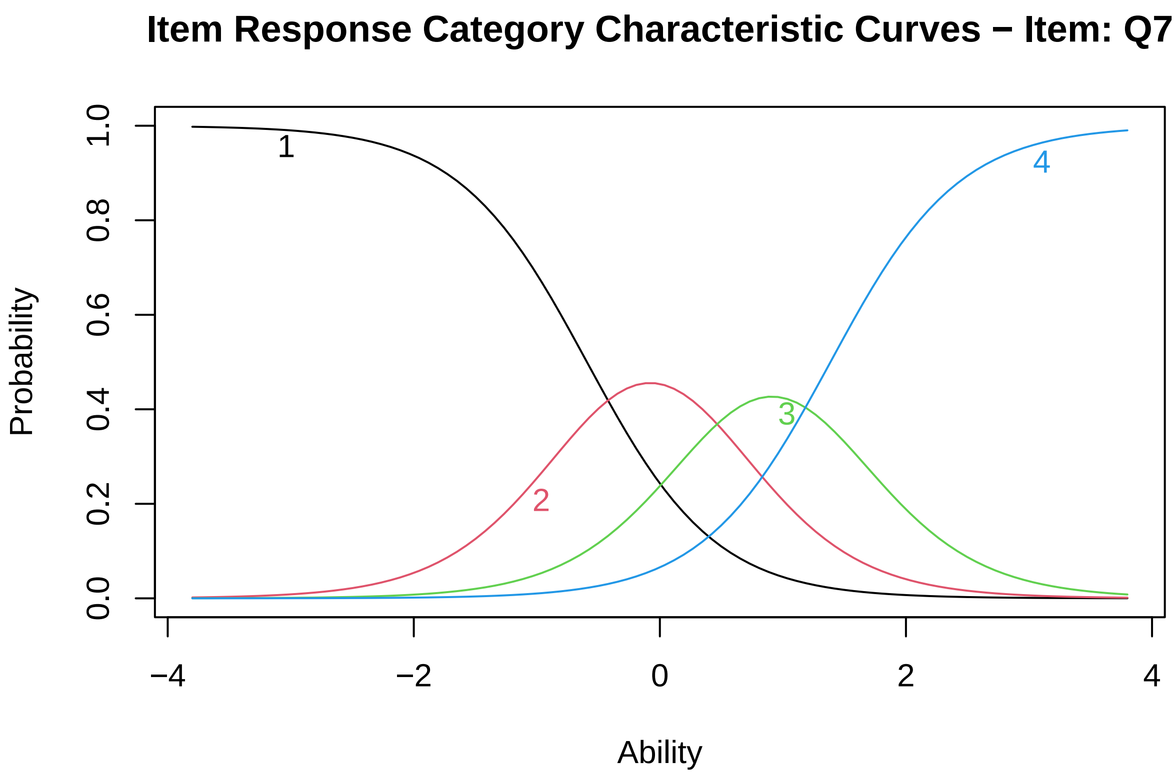


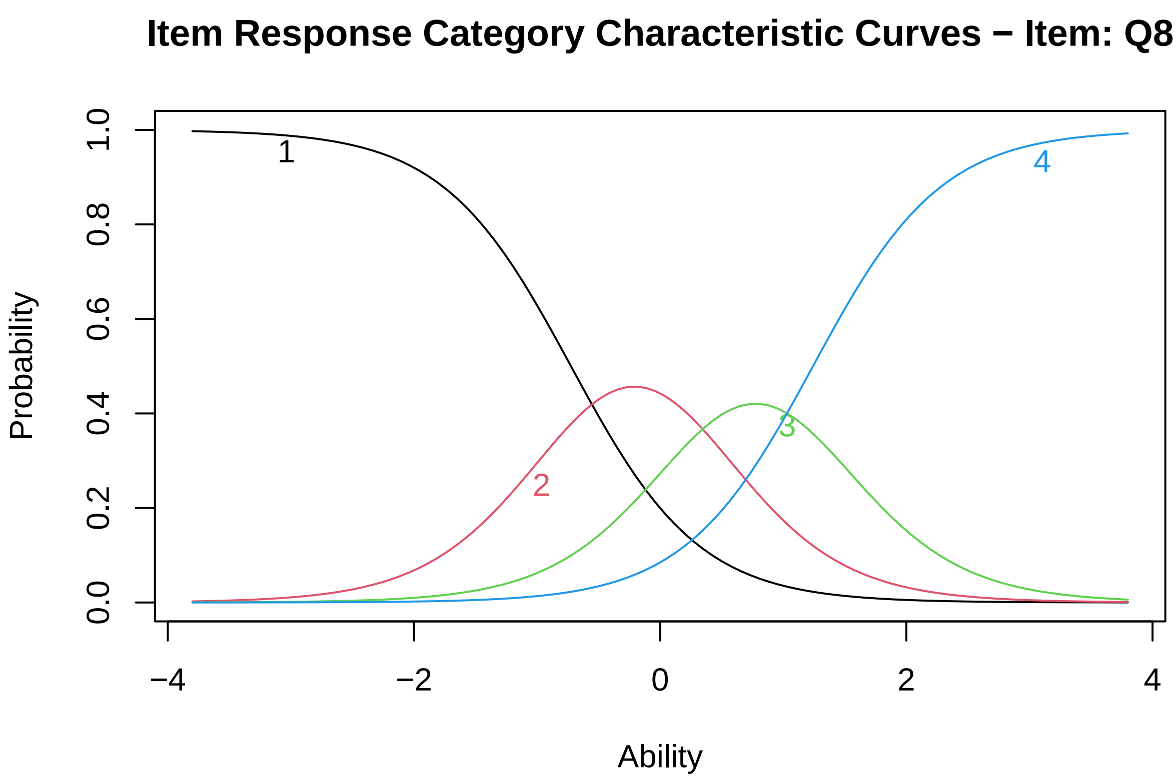
Item Response Category Characteristic Curves − Item: Q4

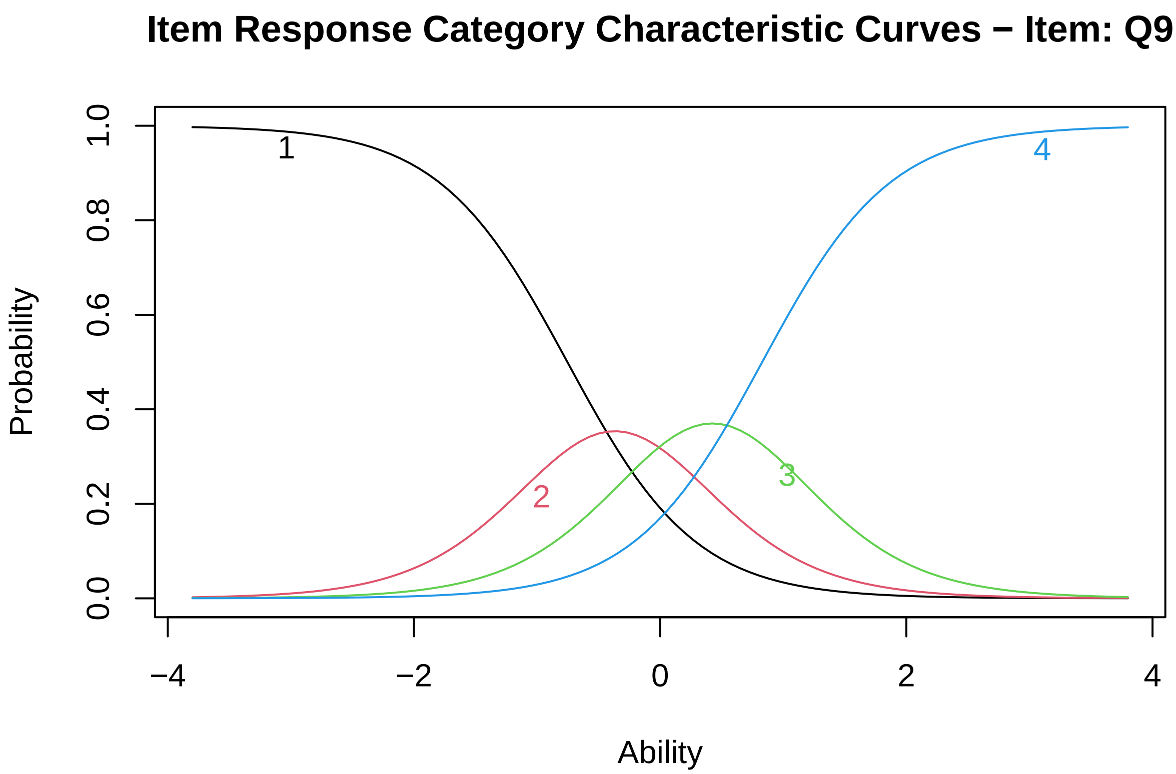


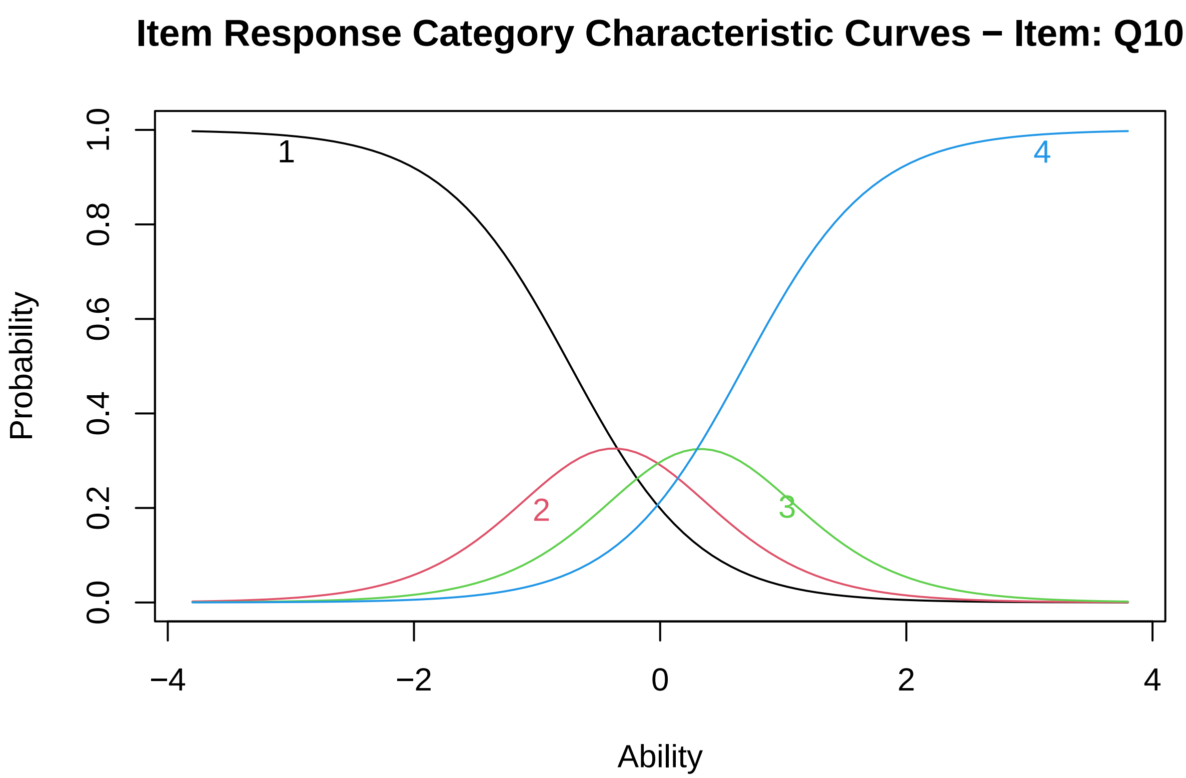

Item Response Category Characteristic Curves − Item: Q5
Probability
1.0
0.8
0.6
0.4
0.2
0.0
−4
−2
0
2
4
Ability
1
2
3
4


Item Response Category Characteristic Curves − Item: Q6
Probability
1.0
0.8
0.6
0.4
0.2
0.0
−4
−2
0
2
4
Ability
1
2
3
4


Item Response Category Characteristic Curves − Item: Q7
Probability
0.0
0.2
0.4
0.6
0.8
1.0
1
2
3
4
−4
−2
0
2
4
Ability


Item Response Category Characteristic Curves − Item: Q8
Probability
0.0
0.2
0.4
0.6
0.8
1.0
1
2
3
4
−4
−2
0
2
4
Ability


Item Response Category Characteristic Curves − Item: Q9
Probability
Ability
0.0
0.2
0.4
0.6
0.8
1.0
−4
−2
0
2
4
1
2
3
4


Item Response Category Characteristic Curves − Item: Q10
Probability
Ability
0.0
0.2
0.4
0.6
0.8
1.0
−4
−2
0
2
4
1
2
3
4

12 References

Crocker, Linda, and James Algina. 2008. *Introduction to Classical and Modern Test Theory.* Cengage Learning.

Guan, Ng Chong, and Muhamad Saiful Bahri Yusoff. 2011. "Missing Values in Data Analysis: Ignore or Impute?" *Education in Medicine Journal* 3 (1).

Kalichman, Seth C, and David Rompa. 1995. "Sexual Sensation Seeking and Sexual Compulsivity Scales: Validity, and Predicting HIV Risk Behavior." *Journal of Personality Assessment* 65 (3): 586–601.

———. 2001. "The Sexual Compulsivity Scale: Further Development and Use with HIV-Positive Persons." *Journal of Personality Assessment* 76 (3): 379–95.

Magis, David, Sébastien Béland, Francis Tuerlinckx, and Paul De Boeck. 2010. "A General Framework and an r Package for the Detection of Dichotomous Differential Item Functioning." *Behavior Research Methods* 42 (3): 847–62.

Mair, Patrick, and Reinhold Hatzinger. 2007. "Extended Rasch Modeling: The eRm Package for the Application of IRT Models in r."

Novick, Melvin R. 1965. "The Axioms and Principal Results of Classical Test Theory." *ETS Research Bulletin Series* 1965 (1): i–31.

Rasch, Georg. 1960. "Studies in Mathematical Psychology: I. Probabilistic Models for Some Intelligence and Attainment Tests."

Reynolds, Cecil R, and RA Livingston. 2021. *Mastering Modern Psychological Testing.* Springer.

Rizopoulos, Dimitris. 2006. "Ltm: An r Package for Latent Variable Modelling and Item Response Theory Analyses." *Journal of Statistical Software* 17 (5): 1–25. https://doi.org/10.18637/jss.v017.i05.

Rosseel, Yves. 2012. "Lavaan: An r Package for Structural Equation Modeling." *Journal of Statistical Software* 48: 1–36.

Smyth, Rachael. 2022. "Item Response Theory for Polytomous Items." https://www.uwo.ca/fhs/tc/labs/12.Polytom

Templin, Jonathan. 2022. "IRT Estimation with r Packages Mirt and Lavaan." https://jonathantemplin.com/irt-estimation-packages-mirt-lavaan/.

Van der Linden, Wim J, and RK Hambleton. 1997. "Handbook of Item Response Theory." *Taylor & Francis Group.* 1 (7): 8.