Week 10 - Item Response Theory

# Introduction

Welcome to Week 10. This worksheet will introduce you to Item Response Theory (AKA Latent Variable theory) and how to conduct this in Jamovi. This is *not essential* for the assignment, but it is useful to know alternatives to Classical Test Theory, their strengths and weaknesses.

Item Response Theory was initially developed for educational testing, the initial focus was in assessing ability rather than other latent traits. Since ability is commonly assessed with 'right' or 'wrong' answers, IRT started as a method for questions with 'dichotomous' answers. Using a method called *Logistic Functions*, IRT provides an estimate of the probability that a person will questions of increasing difficulty correct. This probability is used to estimate a persons 'position' on a latent variable distribution.

# Accessing IRT in Jamovi

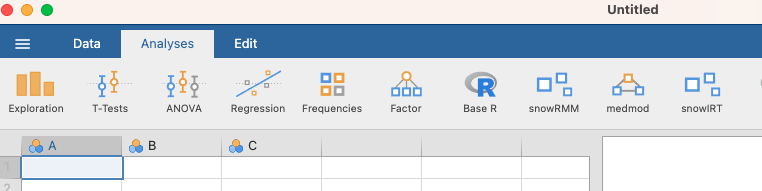
Graphical user interface, application

Description automatically generatedIRT analyses is not available by default in Jamovi, but there is an add-on that you can access. Select *Modules -> Jamovi Library*

Graphical user interface, text, application, email

Description automatically generatedScroll down to *SnowIRT - Item response Theory for Jamovi* and click *INSTALL*

A new option will now appear in the Analyses menu bar:



# Importing Data

IRT is quite computationally heavy. I have tried running analyses on the full VIQT dataset (n=12000+) and it was still chugging along after 30 minutes on a Macbook Pro - so there is not much hope for the university computers. Instead I took a random sample of 400 people from the dataset - this data is available in the Week 10 materials (VIQT\_Short.csv).

Open up this dataset as you usually do in Jamovi.

Since this is based on the 'basic' VIQT dataset where it is only indicated whether the participant got the question correct or incorrect there should just be 400 rows of the following variables:

* Q1 to Q45: 0 = incorrect, 1 = correct
* S1 to S30: Survey items (not relevant for this task)
* Education: Not relevant (See Codebook if you are interested)
* Urban: Not relevant (See codebook if you are interested)
* Gender: 0 = not provided, 1= male, 2 = female, 3 = 'other'
* engnat: Is English your native language? 1= Yes, 2 = No
* age: age in years

# Preparing the Data

For dichotomous responses, the IRT module needs all questions to be labelled as 'continuous'. By default, the questions are imported as nominal, so we need to change the datatype. Select all of the columns Q1 to Q45 (click on Q1, hold CTRL, scroll to and click on Q45 - the columns should be shaded blue)

Table

Description automatically generated with medium confidence

Graphical user interface, text, application

Description automatically generatedSelect *Data -> Setup* from the menu bar.

On the *Measure type* drop down box select *Continuous*. This will change all of the columns to Continuous data types.

# Running an IRT Analysis

## Item Statistics

We are now ready to take a look at the questions. Only a small subset will be covered in this document because otherwise there will be dozens of pages of graphs to include.

From the *Analysis* menu select *SnowIRT -> Dichotomous Rasch Model*.

Graphical user interface, application

Description automatically generated

This will bring up the familiar Jamovi analysis interface. First expand the IRT Analysis menu and make sure *Proportion* and *Measure* from *Item Statistics*, and *Reliability* from *Model Fit* are selected.

Graphical user interface, application, Word

Description automatically generated

In the left box we have our columns, and in the right the variables that are to be analysed. Select the question items (Q1-45) and click the Right Arrow to move the items over to the *Variables* panel.

Graphical user interface, application

Description automatically generated

Graphical user interface

Description automatically generated In the right hand panel we are provided with two tables by default: *Model Fit* and *Item Statistics*

We will start with the *Model Fit* table. At the moment there is only one value, *Person Reliability*. This is analogous to the reliability coefficients we looked at in Classical Test Theory and essentially estimates the accuracy with which estimates of ability are made using these items (Adams, 2005).

In the item statistics table we have two estimates - the first, *Proportion,* is the proportion of the norm group who selected the correct answer. We can see that the vast majority of people (99.8%) got the first question correct, around 91% of people got the second question correct, and so on.

The *Measure* value provides an estimate for the difficulty of that item. This is defined as the point at which a person has 50% chance of getting the question correct (i.e. by chance). It is inversely related to the *Proportion* column, so they higher the proportion correct the 'easier' the item is.

|  |  |
| --- | --- |
| Question 1 |  |
| What is the most difficult item on the VIQT? |  |

You might be wondering why there are no scale descriptive statistics outside of the reliability. This is because the outputs are scaled to a standard normal distribution - i.e. they always have a mean of 0 and a standard deviation of 1, like z scores.

|  |  |
| --- | --- |
| Question 2 (recap) |  |
| What percentage of scores call between ±1 in a standard normal distribution? |  |

### Graphing Items

In addition to numerical information we can also plot items using *Item Characteristic Curves*. These curves tell us what range of abilities this question is accessible to, the probability that people at different abilities will get the question correct, and can also tell us how distinct different items are - i.e. how informative they are about peoples ability.

To plot an *ICC* click *Plots* and *Expected scores curve*. This may take some time...

We'll look at Question 1 first:

Chart, box and whisker chart

Description automatically generated

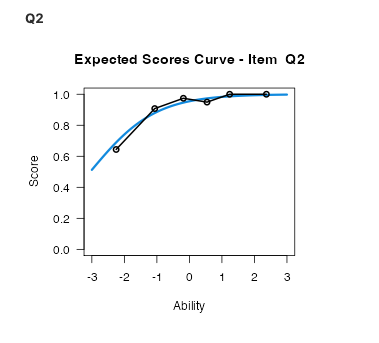
The X axis refers to a persons position on the latent variable 'Ability', and the Y axis gives us the probability that a person with each level of ability will get the question correct. Because so many people got this question correct the ICC is essentially straight. There is a small dip at an ability of around -1 which tells us that the person who got this item correct probably had an overall ability of about -1.

Question 2 gives us a slightly more usual curve

Chart, line chart

Description automatically generated

We see here that people at lower abilities are less likely to get the question correct, and at an ability of -3 are actually only just above chance level. The probability increases with ability with people of higher ability being almost certain to get the question correct. Another thing to note is that the relationship between Ability and probability of a correct answer is not linear. A linear curve would look like this:



Here a small change in ability leads to an equally small change in probability of correct answer and the rate stays the same at all levels. In the ICC for Question 2 the relationship is *Non-Linear* meaning that ability has a large influence of probability at lower levels, and less of an influence at higher levels.

|  |  |
| --- | --- |
| Question 3 |  |
| Find the ICC for the most difficult question. Paste it in the next box and describe the shape. |  |

## Person Statistics

In addition to the item statistics, we can also get scores for individual people (it is here that you'll probably get why you've only got 400 people and not 12,000 in the example dataset). Expand the *Person Statistics* option and select *Total Score* and *Measure*.

Table

Description automatically generated

The table will have one row for each participant in the dataset and one column for each estimate. The *Total score* is the total number of items they got correct, *Measure* is their position on the latent 'Ability' distribution, and *S.E. Measure* is the amount of variation in the estimate of their position on the latent scale. Remember that a *Measure* value of 0 does not mean that the person got no questions correct. It means their Ability is average compared with the norm group.

The *Measure* parameter is calculated not just from the number of items answered correctly (like in Classical Test Theory), but also from the difficulty of the items. So if somebody only got really difficult items correct then their ability would be estimated higher than someone who only got easy questions correct.

|  |  |
| --- | --- |
| Question 4 |  |
| Which participant has the highest 'Ability'? |  |

Unfortunately there is no way in Jamovi to pull descriptive statistics such as percentiles from these tables. But you can copy and paste the table into Excel, save as a CSV, and then re-open in Jamovi.

### Visualising Item and Person Statistics

We have seen how to plot ICCs, but we can also plot participant ability and item difficulty using a *Wright Map.*

Chart, histogram

Description automatically generated

On the left we have a histogram of the distribution of people at different ability levels (this terminology comes from the educational setting that IRT is usually used in - our norm group is likely not made up of just students). On the right we have each item mapped to its corresponding level of difficulty.

References

Adams, R. J. (2005). Reliability as a measurement design effect. Studies in Educational Evaluation, 31(2-3), 162-172.