Computer assignment 2 in RStudio

Sigrún Ósk Jakobsdóttir

(Worked with Guðrún Carstensdóttir, Helga Margrét Ólafsdóttir, Katrín Mjöll Halldórsdóttir, and Laufey Ásta Guðmundsdóttir)

Reykjavik University

Question 1 - Item Response Functions and Person Estimates

(a) Which item was the easiest item and which item was the hardest?

Item 5 was the easiest item, with an item difficulty of -1.63 and item 10 was the hardest item with an item difficulty of 0.53.

(b) Provide a 95% confidence interval for the easiest item and interpret it.

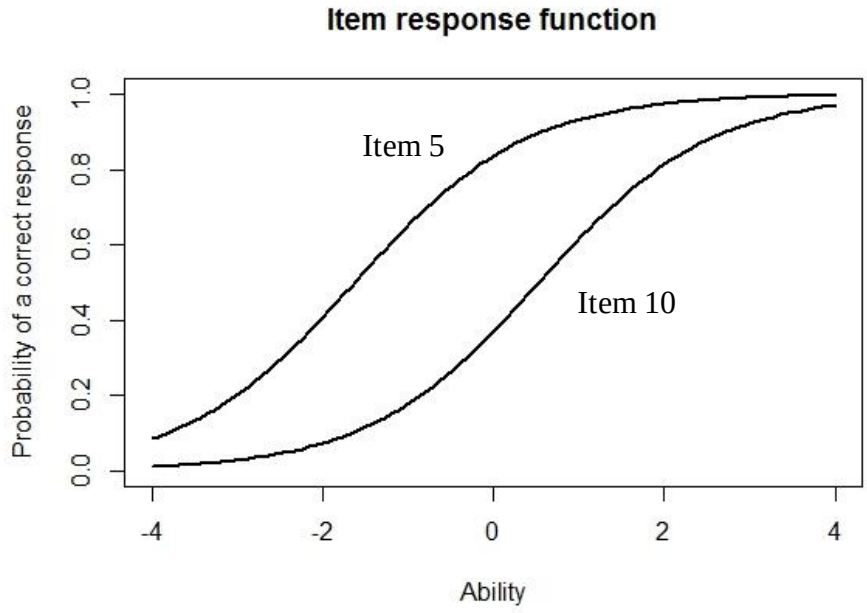
The formula for a 95% confidence interval is: $b \pm 1.96 \cdot SE_b$, where b is the item difficulty and SE_b is the standard error for the item difficulty. The standard error for item 5 was 0.1320 and the item difficulty was -1.6267.

Lower bound: -1,63 – 1.96 · 0.13 = -1,89

Upper bound: -1.63 + 1.96 · 0.13 = -1,37

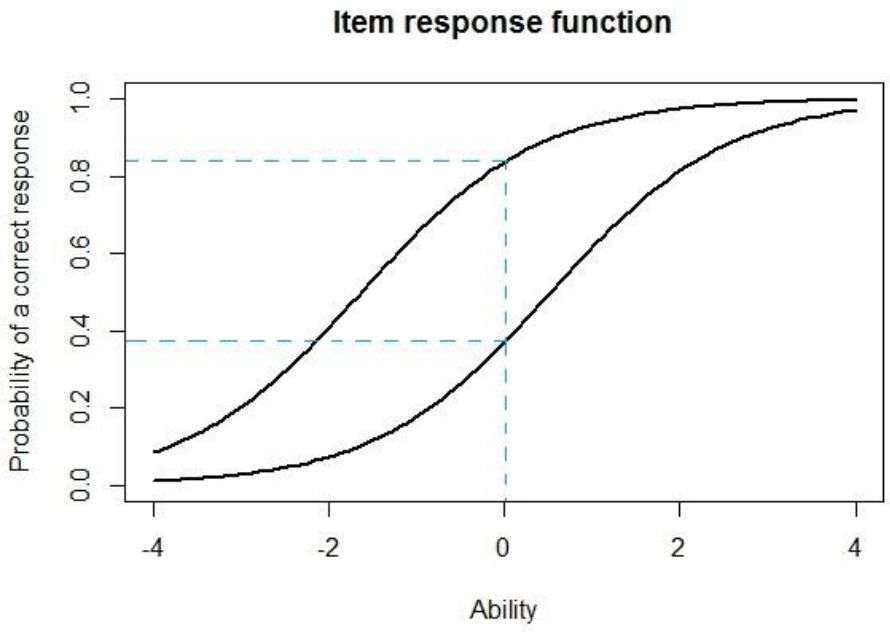
This means that there is a 95% probability that the item difficulty is between -1,37 and -1,87.

(c) Provide a plot that contains both the easiest and the hardest item.



(d) What would we expect the probability of a correct response would be for someone who had an ability score of 0 for these two items? (2 points)

For a person with an ability score of 0, the probability that he/she would get item 5 correct is approximately .85 or 85% (see image below). The probability that he/she would get item 10 correct is approximately .35 or 35% (see image below).



(e) What was the score of the person who did the best on the test? What was the score of the person who did the worst on the test?

Student number 49 did the best on the test and had an estimated ability score of 3.999921. Student number 393 did the worst on the test and had an estimated ability score of -3.999947.

(f) Provide a 95% confidence interval for the estimated ability for the student who did the best on the test and interpret it. (2 points)

The formula for a 95% confidence interval is: $\theta \pm 1.96 \cdot SE_m$, where θ is the ability score and SE_m is the standard error of measurement. The estimated ability score for student number 49 was 4.00 and the standard error of measurement was 2.20.

Lower bound: $4.00 - 1.96 \cdot 2.20 = -0.31$

Upper bound: $4.00 + 1.96 \cdot 2.20 = 8.31$

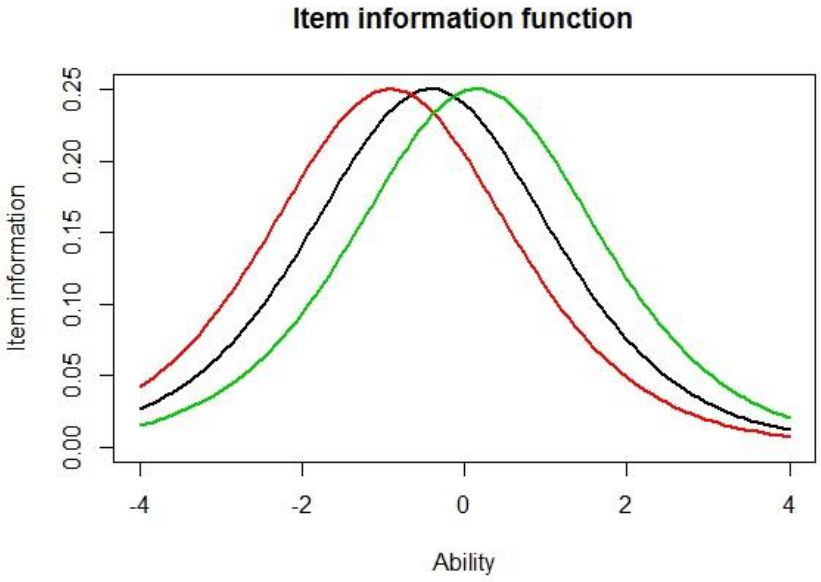
This means that there is a 95% probability that the ability score for student number 49 is between -0.31 and 8.31.

Question 2 - Information

(a) Please state the three items you selected.

I selected items number 7, 9 and 13.

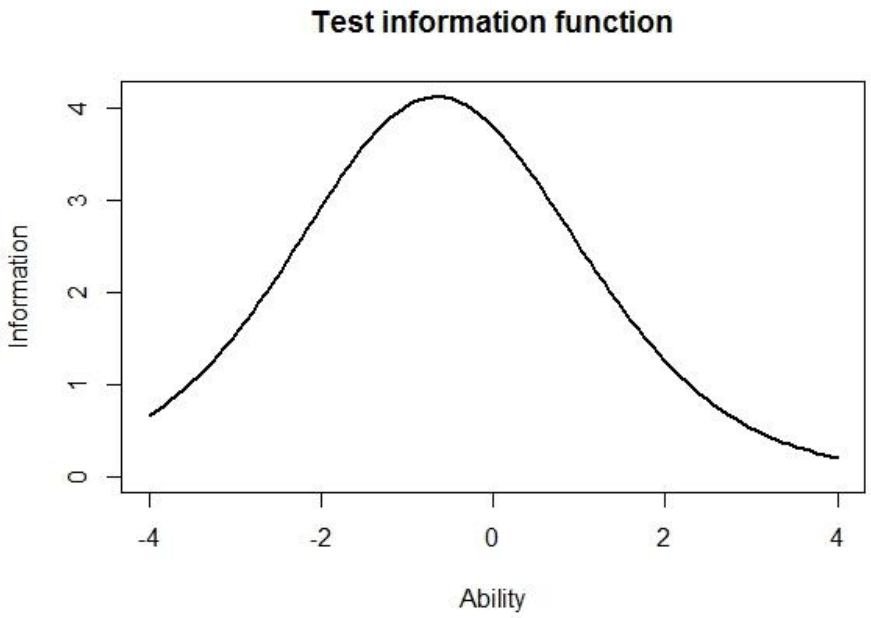
(b) Provide a plot that contains these three items' information functions.



(c) What is the same about these items' information functions? What is different? Hint: This can be a very short answer.

The shape of the functions is the same but the item location is different. Item 9 is located at -.91, item 7 is located at -.41 and item 13 is located at .15.

(d) Provide a plot of the test information function.



(e) Where is the majority of the information for this test located?

The majority of information on this test is around -0.6. This means that it is possible to discriminate with more precision between students with ability scores around -0.6.

Question 3 - Comparing the 2-PL

(a) Which item had the highest discrimination? Which one had the lowest discrimination?

Item 8 had the highest discrimination, 2.29, and item 12 had the lowest discrimination, 0.33.

(b) Are the items that were the easiest and hardest in the Rasch model, also the easiest and hardest in the 2-PL? (1 point)

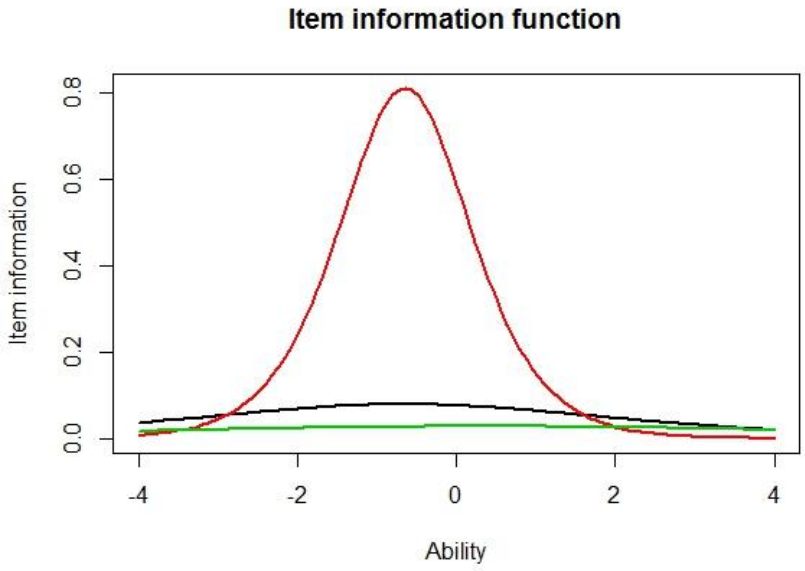
In the Rasch model the easiest item was item 5, with an item difficulty of -1.63. However, in the 2-PL model the easiest was item 1 with an item difficulty of -2.00. The hardest item in the Rasch model, item 10, was also the hardest item in the 2-PL model. According to the Rasch model item 10 had an item difficulty of .53 but according to the 2-PL model the difficulty of item 10 was .49.

(c) What is the correlation between the ability estimates on the Rasch model and the 2-PL? If your interest was solely on estimating person abilities, do you think you would draw the same conclusions from both models? Why? (2 points)

The correlation between ability estimates on the Rasch model and the 2-PL was 0.97. Given that the variance in ability scores is the same using the Rasch model and the 2-PL model, we would expect to draw the same conclusions from both models. But if the variance in ability scores is different, we could not be sure to draw the same conclusions.

A perfect correlation does not necessarily mean that both models return the same ability scores. One model could estimate three individuals' ability scores to be 1, 2, and 3, while another model would estimate the scores to be 1, 3, 5 and the correlation between the ability estimates from these two models would have a perfect correlation. So even though the correlation is close to perfect, we cannot be sure to draw the same conclusions from both models.

(d) Provide a plot of the item information function for the three items you selected in Question 2 but this time for the 2-PL model.



(e) For the 2-PL model, how do the item information functions for these items differ? How do the 2-PL item information functions from these items differ from their Rasch item information functions? (2 point)

The main difference between the item information functions for these items is the item discrimination parameter, which can be seen in the height of the curves. Item number 9 (red line) has the highest discrimination, 1.80, and therefore has the most item information. Item number 7 (black line) has a discrimination of .56 and item 13 (green line) has the lowest discrimination, .34. Items 7 and 9 have approximately the same location (-.65) but item 13 is located at .36. These different locations can be seen in the different locations of the peaks of the curves relative to the x-axis (ability).

When comparing the 2-PL item information functions to the Rasch item information functions the biggest difference is the shape of the functions. The Rasch functions all had the similar bell-shaped curve indicating that they all had a similar discrimination parameter. In the 2-PL functions the different discrimination parameters become obvious. Item locations

also differ, although it is not as apparent as the difference in item discrimination. According to the Rasch model, items 9, 7, and 13 were located at -.91, -.41, and .15 but according to the "-PL model they were located at -.65, -.65, and .36.