Introdução à TRI

Ricardo Primi USF 2019



Capítulo 6



Quais propriedades um teste ideal deveria ter ?

- **Confiável** (precisão)
- Útil para algum propósito (validade)
- "Stable frame of reference", "invariance"
 - Medidas pelo menos no nível intervalar (qualitativo nominal, qualitativo ordinal, quantitativo intervalar e quantitativo de razão
- Com métrica com sentido substantivo (não arbitrária)
- TCT -> medidas ordinais







Ability Estimates Based on Raw Scores





93





Item Characteristic Curve









Linking Students and Items



Estimando o escore de um aluno com a TRI



Fig. 6.8 Items at a student's ability level—about 50% correct



Fig. 6.9 Items located below a student's ability level-more than 50% correct



Fig. 6.10 Items located above a student's ability level—less than 50% correct



Fig. 6.11 Given item response pattern, find student ability

Pattern of Student Responses

Invariância e medida intervalar



-

test and a hard test

Fig. 6.12 Easy items are not administered



Rasch Score on Easy Test

Additional Notes

IRT Viewed as a Transformation of Raw Scores

The Rasch model is a particular IRT model. The Rasch model can be viewed as applying a transformation to the raw scores so that distances between the locations of two students can be preserved independent of the particular items administered. The curved line in Fig. 6.2 will be "straight-ened" through this transformation. Figure 6.13 shows an example of this transformation. Note that the distance between A and C on the easy test (horizontal axis) is the same as the distance between A and C on the hard test (vertical axis).

A crude transformation from raw test score to an IRT ability score is

$$\theta = \log\left(\frac{p}{1-p}\right)$$

where θ is IRT ability and p is the raw score in percentage (e.g., p = 0.8, if the raw score is 80% correct on the test).

A number of points can be made about IRT (Rasch) transformation of raw scores:

- The transformation preserves the order of raw scores. That is, Rasch scores do not alter the ranking of students according to their raw scores. Technically, the transformation is said to be monotonic. If one is only interested in ordering students in ability, or items in difficulty, then raw scores will serve just as well. No IRT is needed.
- There is a one-to-one correspondence between raw scores and Rasch scores if every student is administered the *same* test. That is the pattern of correct/incorrect responses does not play a role in determining the Rasch score (see Chap. 7 for more details). However, if students take *different* tests, as illustrated above with easy and hard tests, and within a computer adaptive testing environment, then the raw scores and Rasch scores will not have a one-to-one correspondence. The Rasch scores will take the item difficulties of the overall test into account.
- When students take the same test, the correlation between raw score and Rasch score will be close to 1, as a result of the property of the Rasch model. Occasionally, one sees researchers plotting Rasch scores against raw scores. The high correlation between these two scores has sometimes been taken as indications of good fit of the data to the model. This is a misconception. Actually, even if data mis-fit the model, the correlation between Rasch scores and raw scores will still be close to one.

How About Other Transformations of Raw Scores, for Example, Standardised Score (Z-Score) and Percentile Ranks? Do They Preserve "Distances" Between People?

Using classical test theory approach, raw scores are sometimes transformed to z-scores or percentile ranks. For z-scores, a transformation is applied to make the mean of the raw scores equal to zero, and the standard deviation equal to 1. This transformation is linear, so the relative distance between two points will be the same whether raw scores or z-scores are used. For example, if A and C are further apart than C and B in raw scores, then the z-scores will also reflect the same relative difference. Consequently, z-scores suffer from the same problem as raw scores. That is, z-scores on an easy test and a hard test will not necessarily preserve the same relative distances between students.

Transforming raw scores to percentile ranks will solve the problem of producing differing distances between two people on two different tests. This is because percentile ranks have relinquished the actual distances between students, and turned the scores to ranks (ordering) only. So, on the one hand, the percentile ranks of people on two different tests may indeed be the same, on the other hand, we have lost the actual distances between students. Raw scores, while not quite providing an interval scale, offer more information than just ordinal scales.

Exercício 2

Hands-on Practices

Task 1

Use simulation to generate raw scores for students on an easy test and a hard test.

Q1. Plot the two test scores on a graph

- Q2. Apply a logistic transformation to the raw scores as follows:
- Step 1: Compute percentage correct from the raw scores (raw score divided by possible maximum score). Let p denote percentage correct.
- Step 2: Compute transformed score by applying transformation, $\log(p/(1 p))$, where log is the natural logarithm. The ratio, p/(1 p), is referred to as an "odds". The results from the transformation of $\log(p/(1 p))$ are said to be in the "log of odds unit" (abbreviated as "logit")
- Step 3: Plot the two transformed scores on a graph

Discuss the shapes of the two graphs in terms of measurement invariance. Which graph is closer to a straight line?

Note: This hands-on practice is to demonstrate IRT as viewed as a transformation of the raw scores. However, the actual mathematical modelling of IRT is at the individual item and individual person level, not at the test score level. In IRT software programs, often logistic transformations applied to the test scores or to item scores (percentage of students getting an item right), as shown in this hands-on practice, are used to provide initial values of person and item parameters.

Task 2

Investigate the relationship between raw scores and transformed logit scores. For example, if a test has a maximum score of 30, plot raw scores (between 0 and 30) against transformed scores. What are your observations in terms of the distances between raw scores and between logit scores? Is the relationship between raw scores a linear one? If not, is there a range between which the relationship is approximately linear?