How much is too much? Item Response Theory procedures to shorten tests

Ottavia M. Epifania[1] and Livio Finos[2]

[1] Department of Psychology and Cognitive Science, University of Trento, IT
ottavia.epifania@unitn.it
[2] Université de Paris-Sud, Laboratoire d'Analyse Numérique, Bâtiment 425, F-91405 Orsay Cedex, France

Abstract. The abstract should summarize the contents of the paper using at least 70 and at most 150 words. It will be set in 9-point font size and be inset 1.0 cm from the right and left margins. There will be two blank lines before and after the Abstract. . . .

Keywords: Item response theory, careless error, information functions, short test forms

1 Introduction

As a general rule of thumb, the higher the number of items in a test, the better the measurement in terms of validity and reliability. However, there is a trade-off between the number of administered items and the response quality (giuro che ho della letteratura in merito). Quindi non torturiamo le persone che dopo di un po' non ne possono più. Item Response Theory (IRT) provides an ideal framework for shortening existing tests (or for developing tests from item banks) given the detailed information that they provide with respect to the measurement precision of each item considering different levels of the latent trait. In this contribution, we present a new algorithm for shortening tests that take into accounts the number of administered items by considering the "tiredness" of the respondents for each of the items included in the short test form (STF). The performance of the algorithm in retrieving an ideal target information function (i.e., the measurement precision that would be obtained by administering all the items in a test if the respondents would never get tired) is investigated in a simulation study. il concetto c'è mancano le parole

2 Item response theory e mortacci

In Item Response Theory (IRT) models for dichotomous responses (e.g., correct vs. incorrect), the probability of observing a correct response on item i by person p depends on both the charcteristics of the respondent (as described by their latent trait level, θ_p) and on the characteristics of the item, which can be described by different parameters. IRT models differentiate according to the number of parameters used for describing the characteristics of the items. According to the 4-parameter logistic model (4-PL), the probability of a correct can be formalized as:

$$P(x_{pi} = 1|\theta_p, b_i, a_i, c_i, d_i) = c_i + (d_i - c_i) + \frac{\exp[a_i(\theta_p - b_i)]}{1 + \exp[a_i(\theta_p - b_i)]}, \tag{1}$$

where θ_p is the latent trait level of person p, b_i is the location of the item on the latent trait (i.e., difficulty parameter, the higher the value, the higher the difficulty of the item), a_i describes the ability of i to discriminate between respondents with different latent trait levels (i.e.,discrimination parameter, the higher the value, the higher the discrimination ability of the item), and c_i and d_i describe the probability of observing a correct response when $\theta \to -\infty$ and $\theta \to +\infty$, respectively. When $\theta \to -\infty$, the probability of observing a correct response should tend to 0. Likewise, when $\theta \to +\infty$, the probability of observing a correct response should tend to 1. However, there might be instances where respondents with θ levels below the difficulty of the item, whom are hence expected not to respond correctly, might provide the correct response out of luck. The lucky guess parameter c_i describes the probability of endorsing the item even if the θ level is below the difficulty of the item, such that the probability of observing a correct response for $\theta \to -\infty$ tends to c_i instead of 0. The same but inverse consideration applies for the upper asymptote that describes the probability of giving the correct response for $\theta \to +\infty$. The d_i parameter describes the probability of not endorsing the item given that the latent trait is above the location of the item, such that the probability of observing a correct response for $\theta \to +\infty$ tends to d_i instead of 1.

By constraining $\forall i \in B$, $d_i = 1$ (where B is the set of items in a set), the 3-Parameter logistic (3-PL) model is obtained. From 3-PL, the 2-parameter logistic model (2-PL) is obtained by constraining $\forall i \in B$, $c_i = 0$, and the 1-parameter logistic model (1-PL, equivalent to the Rasch model) is obtained by constraining $\forall i \in B$, $a_i = 1$.

2.1 Information Functions

3 Simulation study

Va detto da qualche parte che:

- La stanchezza dei soggetti viene operazionalizzata come probabilità di non careless error (va beh l'asintoto)

- la tif target è definita come forma ideale dell'informatività che si avrebbe se le persone non si stancassero mai

- la funzione di stanchezza è quella che mi ha dato chatgpt

- va descritto l'algoritmo e va descritto anche Frank (Frank ha solo un passaggio in meno rispetto a Leon perché non considera la stanchezza, basta descrivere leon e poi dire che Frank fa la stessa roba ma non mette la penalizzazione)

References

1. Smith, T.F., Waterman, M.S.: Identification of common molecular subsequences. J. Mol. Biol. 147, 195?197 (1981). doi:10.1016/0022-2836(81)90087-5
2. May, P., Ehrlich, H.-C., Steinke, T.: ZIB structure prediction pipeline: composing a complex biological workflow through web services. In: Nagel, W.E., Walter, W.V., Lehner, W. (eds.) Euro-Par 2006. LNCS, vol. 4128, pp. 1148?1158. Springer, Heidelberg (2006). doi:10.1007/11823285_121
3. Foster, I., Kesselman, C.: The Grid: Blueprint for a New Computing Infrastructure. Morgan Kaufmann, San Francisco (1999)
4. Czajkowski, K., Fitzgerald, S., Foster, I., Kesselman, C.: Grid information services for distributed resource sharing. In: 10th IEEE International Symposium on High Performance Distributed Computing, pp. 181?184. IEEE Press, New York (2001). doi:10.1109/HPDC.2001.945188
5. Foster, I., Kesselman, C., Nick, J., Tuecke, S.: The physiology of the grid: an open grid services architecture for distributed systems integration. Technical report, Global Grid Forum (2002)
6. National Center for Biotechnology Information. http://www.ncbi.nlm.nih.gov