

Item Response Theory and Computerized Adaptive Testing in a collective environment

*Dionisio Alves da Silva Neto*¹, *Dr. Héilton Ribeiro Tavares*²

¹ Federal University of Pará

² Federal University of Pará

Abbreviated abstract: The main objective of this research is to compare the traditional application of IRT models and the agility of a CAT. We imported the data from the Human Sciences fragment of ENEM 2019. After estimating all abilities throughout the traditional IRT model, we developed a CAT for an individual considering two Standard Errors (SE) as the stop rule, $SE = 0.30$ and $SE = 0.35$. The results showed that to predict the same ability with a $SE = 0.30$, we needed 12 items, whereas $SE = 0.35$, settled 15 items. In conclusion, the CAT offers a way to assess, in an equivalent way, the real latent trace in a process less onerous and exhausting.

Related publications: (up to 2 references)

– SILVA NETO, D. A. Teoria de Resposta ao Item e Testes Adaptativos Computadorizados em Ambientes Coletivos. In: XXIV Simpósio Nacional de Probabilidade e Estatística, 2022, Gramado.



dionisio.neto@icen.ufpa.com - 1



4th Conference on
**Statistics and
Data Science**
Salvador, Brazil (online)
December 1-3, 2022

Problem, Data, Previous Works

1. The Challenge of many items

- In Brazil, the National Exam of High School (ENEM, an acronym in Portuguese) is the principal way to get enrolled into qualified public institutions (universities).
- However, preparing an assessment containing 180 items divided into 4 areas of knowledge is exhausting for students and expensive to the Ministry of Education of Brazil (MEC, acronym in Portuguese).
- To solve this, some recent research in Item Response Theory (IRT) proposed the implementation of a more individualized type of evaluation, which is completely different from the paper-based exams.

2. The innovative Computerized Adaptive Testing

- The objective of this work is to compare the traditional estimation of Three-parameter IRT model, adopted in ENEM, with the agility of Computerized Adaptive Testing (CAT), while we estimate the distribution of the number of items administered.
- The data came from the Human Sciences fragment (pink notebook) of the 21st edition of ENEM, which occurred in 2019. The information contained the response pattern of 233.151 candidates for 45 items.



Methods

- The first step after analyzing the data was to estimate the parameters a (discrimination), b (difficulty) and c (guessing) for the 45 items. After that, we predicted all the abilities under the *Expected a Posteriori* method.

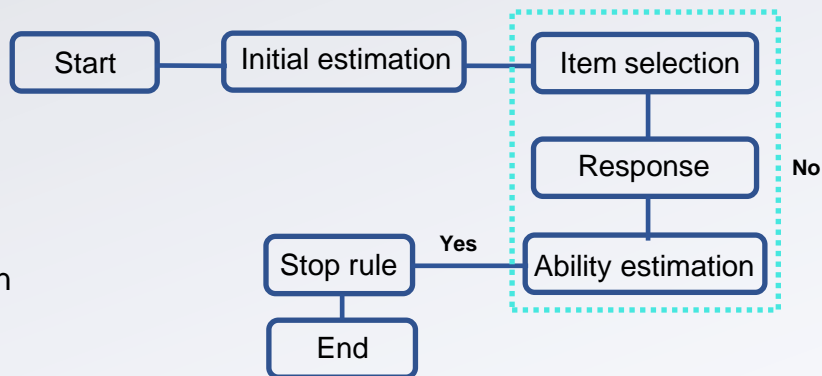
$$P(U_{ij} = 1 | \theta_j) = c_j + (1 - c_j) \frac{1}{1 + e^{-Da_{ij}(\theta_j - b_i)}}$$

- The rules of CAT's scheme included:

- 5 median difficult items to obtain the initial estimation
- Expected a Posteriori* method along the process
- Fisher Information to select the next item
- Two standard errors to end the exams ($SE = 0.30$ and $SE = 0.35$).

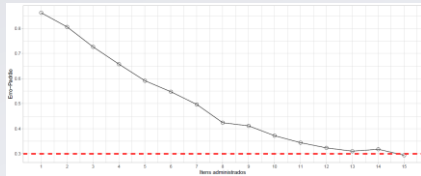
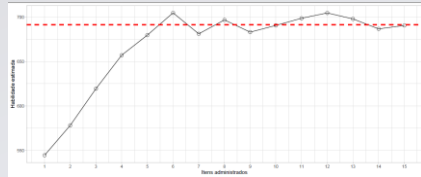
- In a collective environment, multiple adaptive tests were generated by the same design in order to study the number of items regulated to end 1,000 abilities estimations for $SE = 0.30$ and $SE = 0.35$.

- To construct the Computerized Adaptive Test for a random Student, we based the exam on the following algorithm:

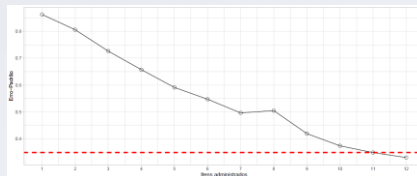
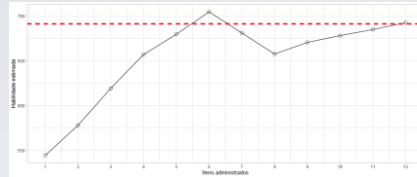


Results and Conclusions

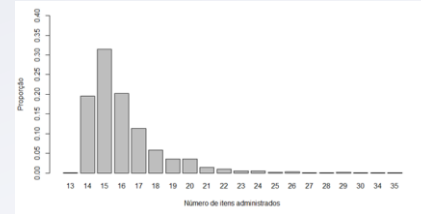
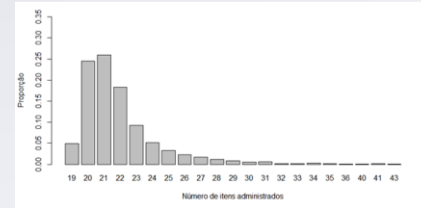
- The results for the CAT with $SE = 0.30$ showed that we needed only 15 items.



- Whereas, the results for the CAT with $SE = 0.35$ showed that we needed only 12 items.



- In an environment where we generated 1,000 CATs for both standard errors, the majority of CATs ended up from 14 to 24 items.



The conclusions of this research point out that the CAT has the potential to change the way that we evaluate exams, such as ENEM, in a process that is less onerous and exhausting for students and test centers.

