ATLANTIS PRESS

Advances in Social Science, Education and Humanities Research, volume 115

3rd International Conferences on Education in Muslim Society (ICEMS 2017)

The Ability of Estimation Stability and Item Parameter Characteristics Reviewed by Item Response Theory Model

Ilham Falani, Nisraeni, Iyan Irdiyansyah

Post Graduate of Jakarta State University, Jl. Rawamngun Muka, Rawamngun East Jakarta, Indonesia 13220

Corresponding e-mail: ilhamfalani@gmail.com

Abstract

The objective of the study is to determine the ability of estimation stability and item parameter characteristics based on the use of logistic models. The data used are simulated data which generated using WINGEN, as much as 1000 participants sample size with length of test of 40 items. Further estimation is done using 1PL, 2PL, and 3PL model as many as ten replication. Based on result analysis, it can be seen that the parameter capability and parameter characteristics of items produced with the 1PL model, more suitable estimated with the 1PL model, the 2PL model, more suitable estimated with the 2PL model, and the 3PL model, more suitable estimated with the 3PL model. These matches can be seen from median score of the estimated correlation data and the original data generated by WINGEN. Therefore, the ability and the characteristic parameters need to pay attention to the logistic parameter model used, to make the estimation as accurate as possible.

Keywords: ability estimation, item characteristic parameters, logistic model.

1 INTRODUCTION

The objective of Item Response Theory (IRT) is to overcome the weaknesses of classical approaches such as item dependent, sample dependent, oriented test, and the similar measurement error for all test takers (Hambleton, Swaminathan, and Rogers, 1991). The advantages of IRT are known by the item characteristic parameters and the invariant properties, i.e. item characteristics (or problem difficulty levels) that are independent of the group of test participants come from the same population. Similarly, the estimation participants' ability does not depend on the characteristics of the given test. So, it can be done by comparing individual test takers and test items.

The basic idea in IRT is that an observed item response (e.g., choosing the category 'strongly agree' on a 5-point Likert scale) is a function of person properties and item properties. There is tremendous variety contained within the term IRT, but the bulk of these models are non-linear latent variable models which attempt to explain the process by which individuals respond to items. One widely used model is the two-parameter logistic model (2PL), which is appropriate for dichotomous observed responses. The 2PLM is written as :

$$P(x_j = 1|\theta) = \frac{1}{1 + \exp[-a_j(\theta - b_j)]},$$

where x_j is the observed response to item j, a_j is the slope parameter for item j, b_j is the threshold parameter for item j, and θ is the construct being measured (Michael Edwards, 2009):

In the Item response theory, each attribute items can be requested the responses from many subjects or respondents. Based on these responses, the subject parameter ability and the characteristics parameter of the attribute item can be estimated. Durability in the estimation is necessary; it is intended that the ability and characteristics of item parameters can provide a description of the ability level and the characteristics of item. The choice of logistics model in estimation also needs to be considered, to support its accuracy in estimation (Dali, 2017).

2 METHODS

The study begins with generating simulation data using sample size and fixed test lengths, i.e. 1000 and 40. While the model used to generate data is 1PL, 2PL, and 3PL model respectively. Generated data will be a reference on estimation data. Correlation calculation taken from reference data and estimation result on data with 10 replication calculated by WINGEN. Estimation result analysis done by comparing the result of correlation score. The highest score of median correlation represent the most stable estimation.

3 RESULT AND DISCUSSION

Based on median correlation calculation done using WINGEN, the data presented as follows:

Table1. Correlation Median for each Model

Data	Med. Estimation Correlation θ with		
	Model 1 PL	Model 2 PL	Model 3PL
Data θ dari Model 1PL	0,935	0,934	0,933
Data θ dari Model 2PL	0,964	0,965	0,964
Data θ dari Model 3PL	0,576	0,571	0,650

The table (first row) shows data θ from the 1PL model is estimated using the three models illustrated in the following figure:

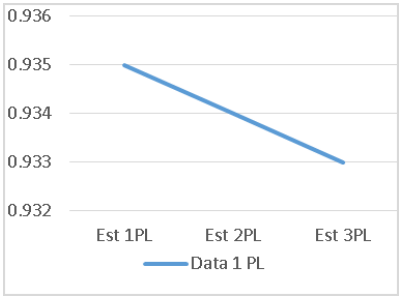

Figure 1. Correlation Median of 1PL data estimation results

The figure above shows the highest correlation median coefficient is generated by estimating the 1 PL model. It indicates that the 1PL generated data would be better estimation using the 1 PL model as well.

The table (first row) shows data θ from Model the 2 PL is estimated using the three models illustrated in the following figure:

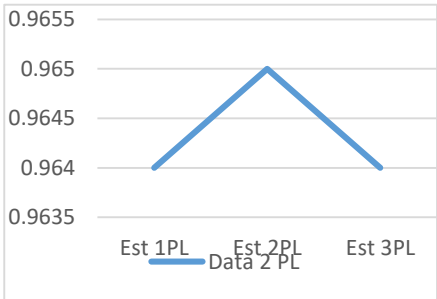

Figure 2. Correlation Median of 2PL data estimation results

The figure shows that the highest correlation median coefficient is generated by estimation using the 2nd PL model. It indicates that the data generated by the 2PL model would be better / estimated using the 2PL model as well.

The table (first row) shows data θ from Model 3PL is estimated using the three models illustrated in the following figure:

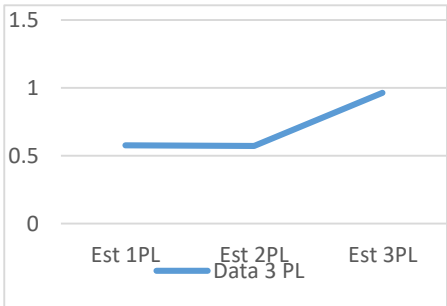

Figure 3. Correlation Median of 3PL data estimation results

The figure shows the highest correlation median coefficients are generated by the 3-PL model estimation. It indicates that the generated data with 3PL model is better/better estimated using the 3PL model as well.

Table 2. Output Resume Parameter Estimation Stability a, b & c

	Med. Corr estimation with								
	1 PL Model			2 PL Model			3 PL Model		
	a	*b*	*c*	*a*	*b*	*c*	*a*	*b*	*c*
1PL Data	0,000	0,998	0,000	0,001	0,998	0,000	0,027	0,089	0,000
2PL Data	0,000	0,996	0,000	0,837	0,999	0,000	0,119	-0,083	0,000
3PL Data	0,000	0,992	0,000	0,764	0,998	0,000	0,127	-0,077	-0,285

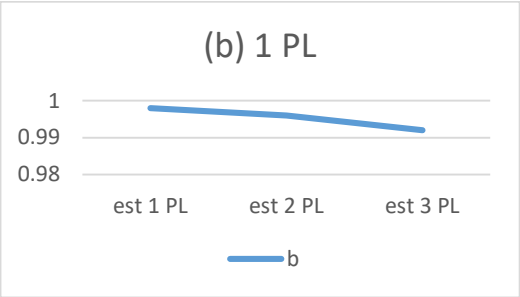

Figure 4. Median correlation of estimation result of 1 PL model

The figure shows b parameter generated by the 1 PL model, it is better/better if it is estimated by the 1 PL model as well. This can be seen from the 1 PL estimation (estimated with the 1PL model) has the highest corrected med correlation coefficient.

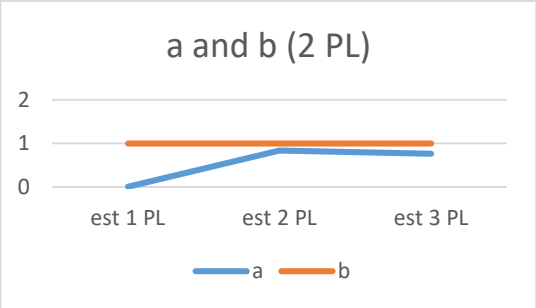

Figure 5. Median correlation of estimation results of a and b 2 PL model

The figure shows the parameters a and b generated by the 2-PL model, it would be preferable if it is estimated with 2-PL models as well. It can be seen from the 2 PL estimation (estimated with 2PL model) has the highest correlation coefficient.

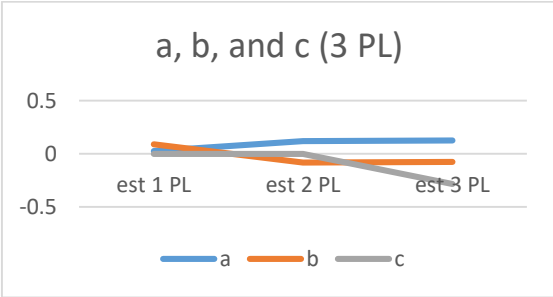

Figure 6. Median correlation of estimation results of a, b, and c 3 PL model

The figure shows the parameters a, b, and c generated by the 3-PL model, are better/ better if it is estimated with the 3-PL models as well. It shows from the 3 PL estimation (estimate with 3PL model) has the highest med correlation coefficient.

3 CONCLUSION

The ability estimation and item characteristic parameters in the response theory items need to pay attention to the sample size information, the length of the test, and the model used in the calculation of reference data to be estimated. The data generated using the 1PL model, more suitable estimation is the 1PL model. The data generated using the 2PL model, more suitable estimation is the 2PL model and the data generated using the 3PL, more suitable estimation is the 3PL model.

4 REFERENCES

Cohen, A.S. & Kane, M.T. (2001). The Precision of simulation study results**.** *Applied Psychological Measurement Journal*, 25 (2), pp. 136-145.

Dali, S.Naga (2017). *Psychometrika pada Pengukuran Model Rasch*. Jakarta: Nagarani Citrayasa.

Embretson S.E & Reise, S.P. (2000). *Item Response Theory for Psychologists*. New Jersey: Lawrence Erlbaum Associates Publishers

Eric Loken & Kelly L.Rulison.(2010). Estimation of a four parameter item response theory model. *British Journal of Mathematical and Statistical Psychology*, 63, pp. 509-525.

Hambelton, R. K. (1987). *The Three-Parameter Logistic Model*. In D.L Mc Arthur(Ed.), *Alternative approces to the assesment of Achievement.* Boston, MA: Kluwe Academic Publisher

Hambleton, R.K, Swaminathan H, & Roger, H.J. (1991). *MMSS Fundamental Of Item Response Theory (Volume 2).* California: Sage Publications.

Hambleton, R. K., & Jones, R. W. (1993). Comparison Of Classical Test Theory And Item Response Theory And Their Applications To Test Development. *Educational Measurement: Issues and Practice*, 12(3), pp. 253-262.

Nathan A.Thompson.(2009). *Ability Estimation with Item Response Theory.* Minessota: Assesment System Corporation. USA: Scientific Software International, Inc.

Mathilda du Toit.(2013). IRT from SSI: Bilog-MG Multilog, Parcale, TestFact.

Mislevy, R.J. & Bock,R.D. (1990). *BILOG 3 : Item analysis & test scoring with binary logistic models*. Moorseville : Scientific Sofware Inc.

Mutasem Akour & Hassan Al Omari.(2013). Empirical Investigation of the Stability of IRT Item-Parameters Estimation. *International online Journal of Educational Sciences,* 5(2), pp. 291-301.

Stabilitas Estimasi Parameter pada Regresi Logistik (*Prosiding).* Seminar Matematika Dan Pendidikan Matematika 2006, pp. 24-48. Yogyakarta: Universitas Negeri Yogyakarta.