Identifying Inter-subject Difficulties in Norwegian GPA Data Using Item Response Theory

Tony C. A. Tan

Centre for Educational Measurement, University of Oslo

Continuous Draft

Prof Rolf V. Olsen & Dr Astrid M. J. Sandsør

Autumn 2021

Identifying Inter-subject Difficulties in Norwegian GPA Data Using Item Response Theory

Ever since men walked on this Earth, we have always been wondering about one thing: What's for dinner? (Coe, 2008; He et al., 2018; Korobko et al., 2008)

Theoretical Framework

The Norwegian GPA System

Blah blah.

Methods

Sample

Say something about Norwegian registry data.

Missing Value Treatment

Missing patterns are not missing at random. If a candidate chose to do physics, he was also highly likely to have chosen advanced maths. So the presence and absence of data tend to group in clusters.

Marginal Maximum Likelihood

A unidemensional generalised partial credit model (Muraki, 1992) with the probability that Candidate n's score in Subject i (x_{ni}) is Grade j ($j = 0, \ldots, m$) is given by

$$p(x_{ni} = j | d_{ni} = 1; \theta_n) = \frac{\exp\left\{j\alpha_i\theta_n - \sum_{h=1}^{j}\beta_{ih}\right\}}{1 + \sum_{h=1}^{m}\exp\left\{h\alpha_i\theta_h - \sum_{k=1}^{h}\beta_{ik}\right\}}, \tag{1}$$

where θ_n is the unidemensional proficiency parameter that represents the overall proficiency of Candidate n.

In MML, a likelihood function (ℓ) is maximised where the candidates' proficiency parameters (θ) are integrated out of the likelihood. The marginal log-likelihood for a unidemensional GPCM is given by

$$\ell = \sum_{p}\sum_{n|p}\log\int\prod_{i}p(x_{ni} = j | d_{ni}; \theta)g(\theta; \mu_p, \sigma^2)\mathrm{d}\theta, \tag{2}$$

where x_{ni} is the observed grade, $p(x_{ni}|d_{ni};\theta)$ is equal to Equation (1) evaluated at x_{ni} if $d_{ni} = 1$, and $p(x_{ni}|d_{ni};\theta) = 1$ if $d_{ni} = 0$. In addition, $g(\theta; \mu_p, \sigma^2)$ is the normal pdf with mean μ_p and variance σ^2. The model can be identified by choosing standard normal $\mathcal{N}(0, 1)$.

Results

Model 1

Model 2

Model 3

Lots of tables here.

Discussions

What does all this mean? Well, let me make you a cup of tea first.

References

Coe, R. (2008). Comparability of GCSE examinations in different subjects: An application of the Rasch model. *Oxford Review of Education*, *34*(5), 609–636. https://doi.org/10.1080/03054980801970312

He, Q., Stockford, I., & Meadows, M. (2018). Inter-subject comparability of examination standards in GCSE and GCE in England. *Oxford Review of Education*, *44*(4), 494–513. https://doi.org/10.1080/03054985.2018.1430562

Korobko, O. B., Glas, C. A. W., Bosker, R. J., & Luyten, J. W. (2008). Comparing the difficulty of examination subjects with item response theory. *Journal of Educational Measurement*, *45*(2), 139–157. https://doi.org/10.1111/j.1745-3984.2007.00057.x

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *ETS Research Report Series*, *1992*(1), 1–30. https://doi.org/10.1002/j.2333-8504.1992.tb01436.x