Extensions of the Kernel Method of Test Score Equating

Gabriel Wallin

Akademisk avhandling

som med vederbörligt tillstånd av Rektor vid Umeå universitet för avläggande av filosofie doktorsexamen framläggs till offentligt förvar i Hörsal S213H, Samhällsvetarhuset, fredagen den 17 januari, kl. 10:00.
Avhandlingen kommer att förvaras på engelska.

Fakultetsopponent: Dr., Anton A. Béguin,
Cito Institute for Educational Measurement, Arnhem, Netherlands.

Department of Statistics, USBE
Abstract
This thesis makes contributions within the area of test score equating and specifically kernel equating. To equate two test forms using kernel equating, the test score distributions are needed. The first paper of this thesis studies the estimation of those distributions. There are currently two family of models implemented within kernel equating for this purpose, namely log-linear models and item response theory models. The impact of model selection criteria for these models on the equated scores are studied using both empirical and simulated data.

The second paper focuses on the continuization of the estimated score distributions, where different bandwidth selection methods are studied. The study considers multiple data collection designs, sample sizes and score distributions and investigates how large impact the bandwidth selection has on the equated scores.

When the test groups differ in their ability distributions it is necessary to adjust for such differences to make fair comparisons possible. The most common way of adjusting for ability imbalance is by using items that are common for both test forms, known as anchor items. If no such items are available, it has been suggested to use background information about the test-takers instead. However, when the covariate vector of background information increases the combination of covariates with no observations tends to increase as well. The third paper of this thesis therefore suggests to transform the covariate vector into a scalar propensity score. Two equating estimators are suggested under this setting and the standard error of equating (SEE) for both estimators are given.

The SEE for kernel equating has previously been derived using the delta method. The fourth paper revisits the Bahadur representation of sample quantiles to derive the SEE. Both methods of calculating the SEE are compared, and it is shown that they are equivalent for all common data collection designs when the terms of the Bahadur SEE are estimated using Taylor expansions. An implementation of an alternative estimator of the Bahadur SEE for which the equivalence result does not hold is also included to illustrate when the two methods differ.

Keywords
Test equating, Nonequivalent groups, Standard error of equating, bandwidth selection, log-linear models, item response theory