Institutionen för Lärande, Informatik, Management och Etik (LIME)

A '3 step' IRT strategy for evaluation of the use of sum scores in small studies with questionnaires using items with ordered response levels

AKADEMISK AVHANDLING
som för avläggande av medicine doktorsexamen vid Karolinska Institutet offentligen försvaras i sal John, Widerströmska huset, Tomtebodavägen 18A, KI, Stockholm

Fredagen de 28 november, 2014, kl. 9.00 av **Ulf B. Brodin**

Huvudhandledare:
Professor Uno Fors
Institutionen för data och systemvetenskap
Stockholms universitet

Fakultetsopponent:
Professor Curt Hagquist
Centrum för forskning om barns och ungdomars psykiska hälsa
Karlstads universitet

Betygsnämnd:
Professor Petter Gustavsson
Institutionen för klinisk neurovetenskap (CNS)
Karolinska Institutet

Professor Ove Almkvist
Institutionen för psykologi
Stockholms universitet

Professor Max Scheja
Institutionen för pedagogik och didaktik
Stockholms universitet

Abstract

This study is focused on a strategy for a basic evaluation of a questionnaire at an early state (i.e. only a limited sample from the target population is available).
Several questionnaires are constructed within the medical research to investigate phenomenon which cannot be explicitly measured (latent variables). In many cases, these questionnaires are applied on a limited sample (less than 100 subjects), without any proper evaluation of its basic characteristic.
This thesis presents a '3- step' strategy for evaluation of questionnaires, where only a limited sample is available for the investigation. Only questionnaires, containing items with a common set of answer alternatives, are considered. The answer alternatives are in terms of an ordered scale to measure the underlying, latent, variable. In all cases, the intention is that a sum score will be a relevant measure of the status of a subject. The higher the score, the more of the latent characteristic is attached to the subject. This implies that all items should represent one common dimension.
The investigation is carried out in three steps and is focused on whether the following criteria, at least approximately, correspond to the intentions which the researcher was supposed to have in mind, at the construction of the questionnaire:

1. All items cooperate.
2. All items work together towards a common dimension.
3. Non-coherent/irrelevant items, as well as non-coherent answer profiles can be identified.
4. The subjects can be reasonably ranked, based on the sum score, on the latent scale.
5. The sum score can be transformed, via a statistical model, to a relevant interval scaled measure.
6. The set of items covers reasonably the intended population.
7. The item set is sufficient for an estimate of a person's position on the latent scale.
8. Defined subgroups perceive the questionnaire in the same way.
9. A straight forward sum score is a sufficient measure (sufficient statistic). Otherwise, an elaborated model, with item specific discrimination, is considered.

Step 1: Non parametric statistical analysis according to 'The Mokken scale analysis' (1 - 4).
Step 2: A parametric model according to the Rasch approach (3, 5 - 8).
Step 3: Can further information be gained from an extended model? (9).

This strategy was applied on 5 studies, all carried out with the intention to use the sum score as a relevant measure of persons' status on the underlying latent scale. Each study involves more than one questionnaire. The '3 – step' strategy was applied on 13 questionnaires within these 5 studies, where the intention was to use a sum score.

Results: Already Step 1 reveals most of the basic weaknesses of the questionnaire, such as weak or contradictory items, deficient correlations between items and a violation of an invariant ordering of the item across low to high scoring persons. These findings are also confirmed in later steps, where further characteristics can be revealed. It also turned out that a well behaved questionnaire according to the 'Mokken requirements' is a prerequisite for a reasonable parametric approach. Furthermore, in most cases the number of items appears to be too small and that the item set has an insufficient coverage to reasonably estimate a person measure for all subjects belonging to the intended population. But, the '3 steps' can constitute a comprehensive message for a basic improvement of the questionnaire.