Introduction

Parametric and nonparametric item response theory in survey research

Introduction

Measurement is one of the most important step in research that allows to describe a wide variety of psychological attributes, such as preferences, attitudes, personality and behaviour. Measurement instruments used by researchers are called tests when the assessment of an achievement is of interest, and inventories or questionnaires when the assessment of typical behaviour is of interest. Achievement refers to a maximum performance, for example when respondents are tested for their arithmetic ability or applicants for their intelligence, and one has to solve as many problems correctly as possible, relying on the person's abilities. Measurement instruments traditionally consist of a set of problems or a set of questions or statements.

Since the beginning of time, skill in the practice has been passed on from generation to generation. Nowadays, we experience assessment in a wide range of our activities and areas. Tests and surveys are used almost in every branch of our life starting from school admission, to business competences and qualification evaluations in our professional career. The types of assessment can take many forms, such as, interviews, examinations, multiple-choice, or one-choice tests. A test consists of collection of test items. Test items are one of the most important features in the measurement theory. Osterlind [1990] defines a test item in an examination of mental attributes as a unit of measurement with a stimulus and a perspective form for answering; and it is intended to yield a response from an examinee from which performance in some psychological construct (such as ability, predisposition, or trait) may be inferred. Within many scientific areas survey instruments play a prominent role in the operationalisation of construct central to the advancement of theory, research and application. Of central importance is test score validity and test itself.

Each of the test items is complex in its own way. Early procedures for test analysis were mostly based on very simple methods such as counting the number of correct responses in the scored set of items. However, in tests there are always some items that are more difficult than other items and such simple counting does not allow to cover those differences. To capture those observed differences in test items more advanced methods of item analysis were developed to overcome these difficulties. Modern statistical methods include discriminating power, the proportion of persons choosing each alternative, and selection of other indicators

to describe the function of test item. One of methods going much deeper in the analysis that allows to overcome these difficulties and limitations is item response theory (IRT).

Item analysis refers to the process of statistically analysing assessment data to evaluate the quality and performance of the test items. This is an important step in the test development cycle, not only because it helps improve the quality of the test, but because it provides documentation for validity. When we deal with tests, the most interesting from the researcher's point of view is item analysis. It is usually focused on finding the items that are not performing well (we can study difficulty and discrimination), and also to figure out why those items are not performing well. On the other hand, economic research require surveys to measure consumers' satisfaction, respondents' attitude toward a product or a service, or preferences measurement.

Using item analysis we can distinguish two approaches. One is classical test theory (CTT) and the other is item response theory (IRT). Classical test theory is based on the idea that a person's observed or obtained score on a test is the sum of a true score (error-free score) and an error score, does not allow for latent variable analysis, and is simply the computation of a proportion of correct answers to the total number of questions. On the other hands there is item response theory that provides a wider and more detailed information including ability test characteristic and model based on latent variables. In this monograph we focus on the second approach and we will try to prove the hypothesis that item response models have more advantages over classical test theory and allow for more detailed analysis. We use model-based method to present the interaction between respondent's ability and test's characteristics as a probability of giving a correct answer. This may be done using the item response theory methods. Item response theory (IRT) is a measurement framework for the development of tests and the scoring of item responses on tests. Key aspects of the IRT framework include the focus on items as the units of observed measurement, the fitting of parametric statistical models to categorical item response data, the estimation of a latent trait variable, and the conditional nature of reliability and the standard error of measurement.

The hypothesis to be verified in this monograph is that item response theory present the relationship between a latent trait and test items and relate observed and latent variables in any space of latent variable of interest. There are many methods in statistics designed for latent variables (e.g. factor analysis). As confirmatory factor analysis and item response theory are related, the model structure is different. CFA characterises the relationship based on a linear model, whereas IRT use non-linear model presenting the probability of particular item response. An IRT model estimates the likelihood of different responses to items by respondents with different levels of the latent trait being measured. Item response models provide a flexible model-based approach to examine the factor structure of instruments used in survey research. Such models embody a broad class of statistical models that seek to express the probability that an examinee will select a particular item

response. Specifically item response theory posit that a respondent's answer is based on specific item and individual characteristics. IRT postulates that examinee test performance can be predicted or explained by a set of factors called traits, latent traits, or abilities, and that relationship between examinee item performance and these traits can be described by a monotonically increasing function. This function is called an item characteristic function and it specifies that examinees with higher scores on the traits have higher expected probabilities for answering an item correctly in comparison to the examinees with lower scores on the traits. In applying item response theory to measurement problems, a common assumption is made that there is one dominant factor or ability which can account for item performance. This so-called ability which the test measures can be also described as aptitude, achievement, or personality variable. Item response theory (IRT) is concerned with accurate test scoring and development of test items. We often want to design test items to measure various kinds of abilities (such as math ability, physical abilities, proficiency), traits, behavioral characteristics or attitudes and preferences.

Traditionally, IRT models have been used to analyse data in psychological assessments and educational testing, however nowadays they are used in various fields in many areas, such as economic research. With the use of IRT models, one may not only improve scoring accuracy, but also economise test administration by adaptively using only the discriminative items. These features might explain why in recent years IRT models have become increasingly popular in many other fields, such as medical research, health sciences, quality-of-life research, and even marketing research. Using computer software that allows to conduct item response analysis makes computational part easier and more accessible for researchers from many areas.

In this monograph we present statistical analysis of item. This book consists of 4 chapters. The first chapter covers historical background of the measurement theory and its role in science starting from ancient China, where tests and examinations were used to select candidates working for the Chinese Emperor. We also present the development of the measurement methods introducing the developments in the 18^{th} century with its roots in mathematics, and most crucial achievements in the 19^{th} and the 20^{th} century that has its origins in psychology (especially experimental psychology). In this chapter we briefly describe classical test theory and we present advantages and disadvantages of this approach in relations to an alternative approach of item analysis – item response theory.

The second chapter covers nonparametric IRT models. Nonparametric item response theory is a family of item response models for ordinal person and item measurement. The main difference between nonparametric and parametric IRT models is that the nonparametric IRT models rest on assumptions about people responding to items in a test or a questionnaire that are more liberal than the assumptions parametric models make. Nonparametric IRT models assume that the relation between the probability of a patient giving a positive response to an

item indicating ease of climbing the stairs and the underlying attribute of physical functioning is monotonne – the better physical functioning, the more ease climbing the stairs – and parametric IRT models assume the relation not only is monotone but also logistic. This extra condition renders the relation more restrictive and the fit of the IRT model to the data more problematic. The distinctive feature that makes an item response model nonparametric is that in a test either each item response function or the set of all item response functions is restricted by some monotonicity condition, without specifying a parametric family of monotone functions such as the logistic. Instead, item response functions are estimated from the test data and the hypothesised monotonicity condition is evaluated. Several nonparametric IRT models have been defined to evaluate the fit of a model to the data. In this monograph we present the theory, assumptions and descriptions of basic nonparametric models such as Mokken scale and unifold process. We present NIRT analysis on the example of the secondary data using **R** software. We also provide graphical presentation of the analysis using scale for simulated data and the distribution of θ's (person parameter) on the latent scale. The book finishes with the summary and reference list.

In the third chapter we present description of basic types of IRT models for binary and ordinal data. The basic term introduced first is latent variable referred to an examinee's or respondent's ability. Such variable is related to an unobserved and hypothetical construct playing crucial role in tests and surveys in categorical data analysis and many other methods related to item analysis. In this chapter we present history, early and classical IRT models, assumptions, theory, mathematical formulas, estimation method and graphical presentation of item response theory. We describe unidimensional IRT models including one parmeter Rasch model, two- and three parameter Birnbaum model and four parameter IRT model for nominal data. This chapter also covers the topic of graphical methods designed for IRT models such item information curve, test information curve, item characteristic curve, and test characteristic curve. For nominal model we also introduce functions such as category boundary response function (CBRF) and item category response function (CRF). Due to the fact there are adjacent response categories, there are two types of polytomous IRT model: Rasch and Thurstone/Samejima. This chapter includes also description of IRT models for polytomous data such as: polytomous Rasch model, partial credit model (PCM), nominal response model (NRM), graded response model (GRM), rating scale model (RSM), and generalised partial credit model (GPCM). We also present the description of estimation of parameters, ability and jointly estimation of parameters and ability. In practical part of this chapter we covered detail, **R** procedures used for economic knowledge testing and satisfaction from on-line learning, as well as a variety of secondary data available in **R** software.

Fourth chapter of this monograph covers multidimensional IRT models which are a special case of unidimensional IRT model. Such models are based on assumption that the latest variable describing an examinee ability or proficiency is multidimensional. This chapter gives a brief review of multidimensional IRT

history, assumption, parameter estimation and visualisation. In such models we include multidimensional latent trait, we describe their connection to one another (compensatory and non compensatory MIRT models), and their relationship with the item responses. This chapter also covers descriptive MIRT statistics including the latent space description, the item response surface, and information function. In this chapter we describe multidimensional parameters, as well as graphical presentation of relations between abilities levels and the probability of giving a correct answer using three dimensional surface. We also demonstrate how to apply multivariate IRT models for testing economic knowledge and satisfaction based on own research and studies using **R**.

I would like to express my appreciation, respect and deep affection for prof. Adam Sagan – the reviewer of this book, for inspiring me with his passion of statistics and science. I am deeply indebted to prof. Sagan for all comments and suggestions on draft book.

My sincerest gratitude goes to my mentor – prof. Eugeniusz Gatnar – for his never-ending support, belief, and inspiration during my professional life at the University of Economics in Katowice.

I also appreciate the financial effort and support I received from the University's Authorities, thanks to whom I was able to achieve my scientific goals and publish this book.

I want to thank to my family, for love, support and strength.