Introduction to Item Response Theory

Course Syllabus

Instructor: Stephen Schilling - (734-355-9801) (schillsg@umich.edu) , Institute for Social Research – Survey Research Center, School of Medicine, School of Education, The University of Michigan

Text: Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage.

Course Description

Over the past half century Item Response Theory (IRT) has revolutionized test analysis and scoring in education, psychology, and medicine. IRT modeling is now the standard for almost all educational assessments, college readiness exams, and patient reported outcomes measures.

IRT involves modeling subjects' responses to individual items, in contrast to Classical Test Theory, which models test scores on complete test forms. IRT offers substantial advantages for many technical problems that arise in creating and using tests, including test design, test equating, assessment of item and test bias, and test scoring. IRT models have the advantage of invariance of person estimates to the collection or sample of items in a particular test, and the invariance of item parameter estimates to the sample of subjects used in test calibration.

This course will begin by comparing Item Response Theory to Classical Test Theory, focusing on the assessment of measurement error. There we will focus on the key components of IRT: the item characteristic curve (ICC), the item information function, and the test information function.

We will then move to a survey of models for unidimensional sets of dichotomously scored items, including the 1-parameter or Rasch model and the 2 and 3-parameter IRT models. We will then look at extensions of IRT to ordinal and nominal data, including the partial credit model, the generalized partial credit model, graded response model, and the nominal response model.

Finally we will examine specific applications of IRT, including test design and equating, assessment of test and item bias (differential item functioning), and test scoring including computerized adaptive testing. Here we will work real world problems and applications of IRT in educational assessment and the assessment of patient reported outcomes. Students will be provided with the knowledge and skills to perform IRT analyses using freely available software available in the R statistical environment and commercially available IRTPRO software. Course work will include three assignments and a final project that requires the students to use IRT analyses on their own data.

Prerequisites: One or more courses in statistics that include basic statistical models, including regression analysis and basic notions of statistical inference and probability. Some familiarity with statistical software such as SPSS and SAS and with the use of EXCEL is required.

Tentative Course Schedule

Introduction to IRT

Lecture 1: Monday, June 11, 9:00 – 10:25. Introduction to Testing, Psychometrics, and Classical Test Theory, IRT Basics – Model Derivation and Assumptions: HSR Chapter 1-2
Lecture 2: Monday, June 11, 10:35 – 12:00. IRT models – The 1, 2, and 3 Parameter Logistic Models, IRT Ability Estimation – MLE, MAP, and EAP: HSR Chapter 2-3
Lab 1: Monday, June 11, 1:00 – 4:00. Introduction to R IRT software, Fitting IRT Models: Class Handout

IRT Parameter Estimation, Model Fit, and Test Construction

Lecture 3: Tuesday, June 12, 9:00 – 10:25. IRT Parameter Estimation – Joint MLE and MMLE Estimation, Assessment of Model/Data Fit: HSR Chapter 3-4
Lecture 4: Tuesday, June 12, 10:25 – 12:00. Item Information and Test Information Functions, IRT Test Construction HSR Chapter 5-7
Lab 2: Tuesday, June 12, 1:00 – 4:00. Model Data Fit, Item and Test Information Functions, and IRT Test Construction: Class Handout.

IRT Applications

Lecture 5: Wednesday, June 13, 9:00 – 10:25. Test Score Equating, Computerized Adaptive Testing: HSR Chapter 9-10
Lecture 6: Wednesday, June 13, 10:35 – 12:00. Item Bias and Differential Item Functioning: HSR Chapter 8
Lab 3: Wednesday, June 13, 1:00 – 4:00. Test Score Equating and Assessing Differential Item Functioning: Class Handout

Polytomous and Multidimensional IRT Models

Lecture 7: Thursday Jun 14, 9:00 – 10:25. Polytomous IRT Models: Class Handout
Lecture 8: Thursday Jun 14, 10:35 – 12:00. Test Score Equating and Differential Item Functioning in Polytomous IRT Models: Class Handout
Lab 4: Thursday Jun 14, 1:00 – 4:00. Fitting Polytomous IRT models in R and Stan: Class Handout

Multidimensional, Multilevel, and Diagnostic Classification Models

Lecture 9: Friday, Jun 15, 9:00 – 10:25. Multidimensional IRT Models and Multilevel IRT Models: Class Handout
Lecture 10: Friday, Jun 15, 10:35 – 12:00. Diagnostic Classification Models: Class Handout
Lab 4: Friday, June 15, 1:00 – 4:00. Fitting Multilevel and Diagnostic Classification Models in Stan: Class Handout

Reference Books

Baker, F. B., & Kim, S. H. (2004). Item Response Theory: Parameter Estimation Techniques. New York, NY: Marcel Dekker.

Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Psychology press.

Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Boston, MA: Kluwer-Nijhoff Publishing.

Holland, P. W., & Wainer, H. (1993). Differential Item Functioning. Hillsdale, NJ: Lawrence Erlbaum.

Kolen, M. J., & Brennan, R. L. (2004). *Test Equating, Saling, and Linking: Methods and Practices, 2nd edition*. New York: Springer-Verlag.

Lord, F.M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Lawrence Erlbaum.

McDonald, R. P. (1999). *Test theory: A Unified Approach*. Mahwah, NJ: Lawrence Erlbaum.

Rupp, A. A., Templin, J., & Henson, R. A. (2010). Diagnostic measurement: Theory, methods, and Applications. New York: Guilford.

Thissen, D., & Wainer, H. (Eds.). (2001). *Test Scoring*. Mahwah, NJ: Lawrence Erlbaum.

van der Linden, W. J., & Hambleton, R. K. (Eds.). (1997). *Handbook of Modern Item Response Theory*. New York, NY: Springer.

Reference Papers and Book Chapters

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46*, 443-459.

Holland, P. W., & Hoskens, M. (2003). Classical test theory as a first-order item response theory: Applications to true-score prediction from a possibly nonparallel test. *Psychometrika*, *68*, 123-149.

Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I Braun (Eds.), *Test validity* (pp. 129-145). Hillsdale, NJ: Erlbaum.

Junker, B. W., & Sijtsma, K. (2001). Nonparametric item response theory in action: An overview of the special issue. *Applied Psychological Measurement*, *25*, 211-220.

Linn, R. L. (1990). Has item response theory increased the validity of achievement test scores? *Applied Measurement in Education, 3*, 115-141.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Meng, X.L. & Schilling, S.G. (1996) Fitting FIIF Models and an Empirical Investigation of Bridge Sampling. *Journal of the American Statistical Association, 91,* 1254-1267

Mislevy, R. J. (1986a). Bayes modal estimation in item response models. *Psychometrika*, *51*, 177-195.

Muraki, E., & Carlson, J. E. (1995). Full-information factor analysis for polytomous item responses. *Applied Psychological Measurement, 19,* 73-90.

Muraki, E. (1992). A generalized partial credit model: application of an EM algorithm. *Applied Psychological Measurement, 16(2),* 159-176.

Reckase, M. D. (1985). The difficulty of test items that measure more than one ability. *Applied Psychological Measurement, 9,* 401-412.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monographs,* No. 17.

Schilling, S. G. & Bock (2005). High-Dimensional Maximum Marginal Likelihood Item Factor Analysis by Adaptive Quadrature. *Psychometrika. 70(3)* 533-555.

Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement, 7,* 201-210.

Thissen, D., & Steinberg, L. (1986) A taxonomy of item response models. *Psychometrika, 31,* 567-577.

Wright, B. D. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement, 14,* 97-116.