Fundamentals of Item Response Theory

Psychology 821

Spring 2007

Lecture:	Tuesdays, 10am-12:30pm
	Psychology Building (PS) Room 217
Instructor:	Michael C. Edwards
	Lazenby 226
	phone: 614-688-8030
	email: edwards.134@osu.edu
Readings:	All readings will be made available on the course website in pdf format.
Website:	This course will use Carmen.

Course Overview

Item response theory (IRT) has become increasingly popular in the past few decades in a wide variety of fields. Whether it is being used to create computerized adaptive tests for health related quality of life, allow for seamless equating in K-12 educational testing, or to create more precise measurement instruments in psychological testing, IRT has become an essential feature of the modern measurement landscape. This graduate level seminar will be broken into three sections. The first section will provide a brief overview of classical test theory (CTT), focusing on the assumptions typically made in CTT and how IRT allows us to relax those assumptions. The second portion of the class will cover the basic IRT models (2- & 3-parameter logistic, graded response model, etc.), with the emphasis on understanding the parameters and the basic concepts involved. This second part will also explores different software packages available for estimating the parameters of these models. Finally, we will briefly discuss several advanced topics to provide an overview of the wide variety of potential uses for IRT. Topics in this section will include: measurement invariance, computerized adaptive testing, linking, equating, and multidimensional IRT models.

Software

We will primarily use the IRT software package `MULTILOG` (Thissen, 1991). Free demonstration versions will be distributed for the purposes of the course. In addition to `MULTILOG`, we will briefly discuss factor analysis software capable of performing exploratory and confirmatory analyses with categorical measured variables. These software packages will include `CEFA` (Browne, Cudeck, Tateneni, & Mels, 2004) and `LISREL` (Jöreskog & Sörbom, 2003).

Grading Policies

There are no exams. There will be two homework assignments involving data analysis and a (brief) report. These reports should contain tables with relevant output, graphics where helpful, and a verbal description of the results. Each assignment accounts for 20% of your final grade. There will be a final project, which will account for the remaining 60% of your final grade. The final project can be one of two flavors. Your first option is an IRT analysis of data you have which you would like to analyze. From time to time I will also be able to make data sets available to students who wish to pursue this option but do not have their own data. The second option is a review paper on some topic in IRT - either a topic we have covered in greater depth or a topic we do not cover. In either case, I expect that a satisfactory project will run somewhere between 10 and 20 pages. We will discuss the final project in greater detail as we progress through the quarter.

Class Participation

There will be weekly readings on each topic. Please complete the readings during the week before class and submit two questions for clarification or discussion by 5pm the Sunday before class. Please email your questions to me (edwards.134@osu.edu) using the subject line "QUESTIONS FOR IRT CLASS" (in all caps).

Students with Disabilities

Any student who feels s/he may need an accommodation based on the impact of a disability should contact me privately to discuss your specific needs. Please contact the Office for Disability Services at 614-292-3307 in room 150 Pomerene Hall to coordinate reasonable accommodations for students with documented disabilities.

Academic Misconduct

All students at the Ohio State University are bound by the Code of Student Conduct (see http://studentaffairs.osu.edu/resource_csc.asp). Suspected violations of the code in this class will be dealt with according to the procedures detailed in the code.

Tentative Schedule

Date		Topic	Readings
Sept	26	Classical Test Theory	Wainer & Thissen, 2001
			Crocker & Algina, 1986, Ch.6
Oct	3	IRT Overview & Background	Hambleton, Swaminathan, & Rogers, 1991, Ch.1
			Hambleton & Swaminathan, 1985, Ch.1
			Hambleton & Swaminathan, 1985, Ch.2
			Bock, 1997a
	10	IRT for Dichotomous Responses	Thissen & Orlando, 2001, pp. 73-98
			Steinberg & Thissen, n.d.-a, Ch.1
	17	IRT for Polytomous Responses	Thissen, Nelson, Rosa, & McLeod, 2001, pp. 141-150
			Steinberg & Thissen, n.d.-b, Ch.3
			Bock, 1997b
			Samejima, 1997
	24	Estimation & Scoring	Thissen & Orlando, 2001, pp. 98-140
			Thissen et al., 2001, pp. 150-186
			Wainer & Mislevy, 2000
			Bock & Aitkin, 1981
	31	`MULTILOG`	
Nov	7	Differential Item Functioning	Thissen, Steinberg, & Gerrard, 1986
			Thissen, Steinberg, & Wainer, 1993
			Steinberg, 2001
	14	Linking & Equating	Kolen & Brennan, 2004, Ch. 6
	21	Computerized Adaptive Testing	Wainer, 2000
			Mills & Steffen, 2000
			Thissen & Mislevy, 2000
			Edwards & Thissen, 2003
			Wainer & Eignor, 2000
	28	Multidimensional IRT	Reckase, 1997
			Ackerman, 2005
			Edwards, 2005

References

Ackerman, T. A. (2005). Multidimensional item response theory modeling. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics* (p. 3-26). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Bock, R. D. (1997a). A brief history of item response theory. *Educational Measurement: Issues and Practice*, *16*, 21-33.

Bock, R. D. (1997b). The nominal categories model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (p. 33-50). New York, NY: Springer.

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

Browne, M. W., Cudeck, R., Tateneni, K., & Mels, G. (2004). CEFA: Comprehensive Exploratory Factor Analysis, Version 2.00 [Computer software]. Retrieved from http://quantrm2.psy.ohio-state.edu/browne/.

Crocker, L., & Algina, J. (1986). *Introduction to classical & modern test theory.* Belmont, CA: Wadsworth Group.

Edwards, M. C. (2005). *A Markov chain Monte Carlo approach to confirmatory item factor analysis.* Unpublished doctoral dissertation, University of North Carolina at Chapel Hill.

Edwards, M. C., & Thissen, D. (2003). *Finding mult-stage CAT designs associated with uniform item exposure and precise measurement.* (Unpublished manuscript)

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications.* Norwell, MA: Kluwer Academic Publishers.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park, CA: Sage Publications.

Jöreskog, K. G., & Sörbom, D. (2003). LISREL 8.54 [Computer software]. Chicago, IL: Scientific Software International, Inc.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking.* New York, NY: Springer.

Mills, C. N., & Steffen, M. (2000). The GRE computer adaptive test: Operational issues. In W. J. van der Linden & C. A. W. Glas (Eds.), *Computerized adaptive testing: Theory and practice* (p. 75-99). Boston, MA: Kluwer Academic Publishers.

Reckase, M. D. (1997). The past and future of multidimensional item response theory. *Applied*

Psychological Measurement, *21*, 25-36.

Samejima, F. (1997). Graded respose model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (p. 85-100). New York, NY: Springer.

Steinberg, L. (2001). The consequences of pairing questions: Context effects in personality measurement. *Journal of Personality and Social Psychology*, *81*, 332-342.

Steinberg, L., & Thissen, D. (n.d.-a). *Chapter 1: An intellectual history of item response theory: Models for binary items.* From a draft of *Item response theory for psychological research.*

Steinberg, L., & Thissen, D. (n.d.-b). *Chapter 3: A graded response model.* From a draft of *Item response theory for psychological research.*

Thissen, D. (1991). MULTILOG: Multiple cateogry item analysis and test scoring using item reponse theory [Computer software]. Chicago, IL: Scientific Software International, Inc.

Thissen, D., & Mislevy, R. J. (2000). Testing algorithms. In H. Wainer et al. (Eds.), *Computerized adaptive testing: A primer* (2nd ed., p. 101-132). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Thissen, D., Nelson, L., Rosa, K., & McLeod, L. (2001). Item response theory for items scored in more than two categories. In D. Thissen & H. Wainer (Eds.), *Test scoring* (p. 141-186). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Thissen, D., & Orlando, M. (2001). Item response theory for items scored in two categories. In D. Thissen & H. Wainer (Eds.), *Test scoring* (p. 73-140). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Thissen, D., Steinberg, L., & Gerrard, M. (1986). Beyond group mean differences: The concept of item bias. *Psychological Bulletin*, *99*, 118-128.

Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In P. W. Holland & H. Wainer (Eds.), *Differential item funtioning* (p. 67-113). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Wainer, H. (2000). Introduction and history. In H. Wainer et al. (Eds.), *Computerized adaptive testing: A primer* (2nd ed., p. 1-20). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Wainer, H., & Eignor, D. (2000). Caveats, pitfalls, and unexpected consequences of implementing large-scale computerized testing. In H. Wainer et al. (Eds.), *Computerized adaptive testing: A primer* (2nd ed., p. 271-298). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Wainer, H., & Mislevy, R. J. (2000). Item response theory, item calibration, and proficiency estimation. In H. Wainer et al. (Eds.), *Computerized adaptive testing: A primer* (2nd ed., p.

61-100). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Wainer, H., & Thissen, D. (2001). True score theory: The traditional method. In D. Thissen & H. Wainer (Eds.), *Test scoring* (p. 73-140). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.