

PORTLAND STATE UNIVERSITY
Systems Science Ph.D. Program
Professor Martin Zwick
(503) 725-4987

Winter 2008
Tues-Thurs 4:40-6:30
104 Harder House
zwick@pdx.edu

MODEL almost ANY DATA!

SySc 551/651: **DISCRETE MULTIVARIATE MODELING**

In this course, information theory is used as a framework for modeling and data mining: for analyzing static or dynamic relations among discrete* variables, for detecting complex interaction effects, and for discovering nonlinearities in continuous variables made discrete by binning.

In the systems literature, these information-theoretic and related set-theoretic methods, used together with graph theory techniques, are called “Reconstructability Analysis” (RA). RA overlaps with and extends log-linear modeling in the social sciences, Bayesian networks and graphical models in machine learning, decomposition techniques in multi-valued logic design, Fourier methods for compression, and other modeling approaches. It can be used for confirmatory and exploratory statistical modeling as well as for non-statistical applications.

Because of their applicability to both qualitative and quantitative variables, RA methods are very general. They are usable in the natural sciences, social sciences, engineering, business, and other professional fields. The ideas of RA define “structure,” “complexity,” “holism,” and other basic notions, and are foundational for systems science. For course-related research and publications, see http://www.pdx.edu/sysc/research_dmm.html.

*Discrete variables are typically nominal (categorical, symbolic), but may be ordinal or integer.

Prerequisites: Background in probability/statistics. SySc511 is helpful but not essential.

Texts: 1-2 at bookstore; 3 (packet) at Smart Copy, 1915 SW 6th Ave, 227-6137

1. Krippendorff, Klaus (K). *Information Theory: Structural Models for Qualitative Data*. Series: Quantitative Applications in the Social Sciences, Paper # 62, Sage Publications, Beverly Hills, California, 1986. (ISBN 0-8039-2132-2, paperback)
2. Knoke, David and Burke, Peter J. (K & B). *Log-Linear Models*. Series: Quantitative Applications in the Social Sciences, paper # 20. Sage Publications, Beverly Hills, California, 1980. (ISBN 0-8039-1492-X, paperback)
3. Xeroxed articles and selections from books. (Articles available on the web are collected in a second – optional – packet, for possible convenience.)

Grades will be based on midterm and final exams and either a computational project (e.g., data analysis using DMM software or software development) or a term paper.

Outline (K, K & B = texts; author names = xeroxed articles; [] = optional or skim)

- Jan 8 Course overview
Zwick (Overview) [Wholes & Parts]
- Jan 10,15 **I. Basic concepts** (information-theoretic): static/dynamic multivariate relations in categorical data; uncertainty, Transmission
Miller; K: 1-32; McGill & Quastler; Ashby; Cover & Thomas Shannon 1-52, 62-70, 87-91; [Garner & McGill; Hosseini et al]]
- Jan 17 Basic concepts: **exercises**
- Jan 22,24 **II. Structures**; degrees of freedom
K: 32-43, 47-53 [70-81]; Klir: 233-244; K&B: 36-37
- Jan 29 Structures: **exercises**
- Jan 31, Feb 5 **III. Information-theoretic reconstructability analysis**
generalized transmission, information distance; uncertainty maximization; iterative proportional fitting
K: 43-62, 66-67, 70-88; K&B: 30-34, 38-42 [Zwick (CU, except 157-160), Klir]
- Feb 7 Information-theoretic reconstructability analysis: **exercises**
- Feb 12 MIDTERM EXAM
- Feb 14,19 Information-theoretic RA, continued: statistical issues; over exam
- Feb 21,26 Information-theoretic RA: **exercises**
IV. OCCAM software; binning & mask analysis
Willett & Zwick (paper, manual); Zwick, Shu, & Koch
- Feb 28, Mar 4,6 **V. Supplementary topics**: reconstruction vs. identification; state-based RA; set-theoretic RA; log-linear modeling
Johnson & Zwick, Zwick & Johnson; Conant; Zwick (CU 157-160); Zwick & Shu; Jones; K&B: 5-29, 42-63
- Mar 11 Supplementary topics: **exercises**
- Mar 13 **Review; PROJECTS/PAPERS DUE**; short project presentations
- Mar 18 FINAL EXAM [will cover material from entire course].

Xeroxed articles & sections of books

The required SmartCopy reader contains the readings below EXCEPT for articles available on the web. However, if for convenience you want a printed copy of these web articles, you can obtain a second (optional) reader at SmartCopy.

Zwick, M. "Overview of Reconstructability Analysis." *Kybernetes*, 33 (5-6), 2004, pp. 877-905; *Proceedings of 12th International World Organization of Systems and Cybernetics and 4th International Institute for General Systems Studies Workshop*, Pittsburgh, March 24-26, 2002. <http://www.syssc.pdx.edu/download/papers/wholesg.pdf>

Zwick, Martin. "Wholes and Parts in General Systems Methodology." *The Character Concept in Evolutionary Biology*, Gunter Wagner, ed., Academic Press. pp. 237-256, 2001. <http://www.syssc.pdx.edu/download/papers/wholesg.pdf>

Miller, George A. "What is Information Measurement." *American Psychologist* 8 (1963) 3-11.

Klaus Krippendorff, *Information Theory: Structural Models for Qualitative Data*. Sage Quantitative Applications in the Social Sciences, Paper #62, London, 1986, pp. 1-32.

McGill, William and Quastler, H. "Standardized Nomenclature: An Attempt." From *Information Theory in Psychology*. The Free Press, 1955, pp. 83-92.

Ashby, W. Ross. "Incessant Transmission." Ch. 9 (pp. 161-191) of *An Introduction to Cybernetics*. Methuen & Co., London, 1964.

Cover, Thomas and Thomas, Joy A., Ch. 1 & 2 of *Elements of Information Theory*. John Wiley & Sons, Inc., New York, 1991.

Shannon, Claude E. and Weaver, Warren. *The Mathematical Theory of Communication*. University of Illinois Press, 1975 (first published in 1949), pp. 1-52, 62-70, 87-91

Garner, W.R. and McGill, William J. "The Relation Between Information and Variance Analyses." *Psychometrika*, Vol. 21, No. 3, 1956, pp. 219-228.

Hosseini, Jamshid C., Harmon, R. Robert, and Zwick, Martin. "An Information Theoretic Framework for Exploratory Multivariate Market Segmentation Research." *Decision Sciences*, Vol. 22, No. 3, 1991, pp. 663-677
<http://www.syssc.pdx.edu/download/papers/inftheoretic.pdf>

Klir, George J. *Architecture of Systems Problem Solving*. Plenum Press, New York, 1985, pp. 212-266, 281-284.

Zwick, Martin. "Control Uniqueness in Reconstructability Analysis (CU)." *Int. J. General Systems*, vol. 24 (1-2), 1996, pp.151-162.

<http://www.syssc.pdx.edu/download/papers/control.pdf>

Willett, K. & Zwick, M. "A Software Architecture for Reconstructability Analysis." *Kybernetes*, 33 (5-6), 2004, pp. 997-1008.

<http://www.syssc.pdx.edu/download/papers/kenpitf.pdf>

Willett, K. & Zwick, M. "OCCAM: A Reconstructability Analysis Program." OCCAM Manual at <http://www.syssc.pdx.edu/download/papers/woccaman.pdf>

Zwick, M., Shu, H., & Koch, R. "Information-Theoretic Mask Analysis of Rainfall Time Series Data." *Advances in Systems Science and Applications* (1995), Special issue 1, pp. 154-159. http://www.syssc.pdx.edu/download/papers/mz_assa_it.pdf

Conant, Roger C. "Set-Theoretic Structure Modeling." *International Journal of General Systems*, 1981, vol.7, pp.93-107.

Zwick, M. & Shu, H. "Set-Theoretic Reconstructability of Elementary Cellular Automata." *Advances in Systems Science and Applications* (1995), Special issue 1, pp. 31-36. http://www.syssc.pdx.edu/download/papers/mz_assa_ca.pdf

Johnson, Michael S. and Zwick, Martin. "State-Based Reconstructability Modeling for Decision Analysis." *Proceedings of World Congress of the Systems Sciences & 44th Annual Meeting of the International Society for the Systems Sciences*, Toronto, July 16-22, 2000. http://www.syssc.pdx.edu/download/papers/iss_jo_zw.pdf

Zwick, M. & Johnson, M. "State-Based Reconstructability Analysis." *Kybernetes*, 33 (5-6), 2004, pp. 1041-1052; *Proceedings of 12th International World Organization of Systems and Cybernetics* and *4th International Institute for General Systems Studies Workshop*, Pittsburgh, March 24-26.

<http://www.syssc.pdx.edu/download/papers/mjpitf.pdf>

Jones, Bush. "Reconstructability Analysis for General Functions." *Int. J. General Systems*, 1985, Vol. 11, pp. 133-142.

At the end of the binder are:

Klir, George J. "Bibliography of Reconstructability Analysis." *Int. J. of General Systems*, vol. 24, pp. 225-229, 1996. <http://www.syssc.pdx.edu/download/papers/bibliography.pdf>

Chi-square table

Midterm and final exams: Winter 2005, 2006, 2007