

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

Winter 2010
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/syssc/research-discrete-multivariate-modeling>

*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.