PORTLAND STATE UNIVERSITY
Systems Science Ph.D. Program
Professor Martin Zwick

Spring 2021
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 items listed in the Discrete Multivariate Modeling category in my Selected Works website, https://works.bepress.com/martin_zwick/ .

This is the theory course that goes with the project course, SySc 431/531 **Data Mining with Information Theory**, next offered in Winter 2022. It is also the theory course for the **Occam software**, recently been made open source; see https://www.occam-ra.io/

*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 (at the PSU bookstore):
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) *Main* text for course.

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) *Secondary* text for course.

Grades will be based on midterm and final exams and a micro-project.*

*If any aspect of this course presents a barrier to your learning, please notify me.

Outline ([ ] = optional or skim)

March 29 Course overview
Zwick (Overview) *[Wholes & Parts] [Introduction to RA]*
https://media.pdx.edu/media/t/1_16prfr5f

March 31 **I Basic concepts** (information-theoretic): static/dynamic multivariate
https://media.pdx.edu/media/t/1_3jn2qog2

April 5 relations in categorical data; uncertainty, transmission; *K: 1-32*
https://media.pdx.edu/media/t/1_3ml3ej8n

April 7 **I Basic concepts** (continued)
https://media.pdx.edu/media/t/1_gbxf0uxn

April 12 **I Basic concepts**: **exercises**
https://media.pdx.edu/media/t/1_j0rbgvds

April 14 **II Structures** (graph theoretic); degrees of freedom
K: 32-43, 47-53 [70-81]; K&B: 36-37
https://media.pdx.edu/media/t/1_2hkkr56k

April 19 **II Structures** (graph theoretic); degrees of freedom
https://media.pdx.edu/media/t/1_fbqlxt75

April 21 **III Information-theoretic reconstruction** (Putting I & II together)
Generalized transmission, information distance; uncertainty maximization; iterative proportional fitting
K: 43-62, 66-67, 70-88; K&B: 30-34, 38-42 [Zwick (CU, RAE)]
https://media.pdx.edu/media/t/1_is3vhd03

April 26 **II Structures: exercises**
https://media.pdx.edu/media/t/1_lptqpoy5

April 28 **MIDTERM EXAM** on TOPICS I & II only
https://media.pdx.edu/media/t/1_q9kz1msd

May 3 Over midterm exam
https://media.pdx.edu/media/t/1_27k5jdew

May 5 **III Information-theoretic reconstruction**, statistical issues
https://media.pdx.edu/media/t/1_lisgpdxz

May 10 **III Information-theoretic reconstruction**, statistical issues
https://media.pdx.edu/media/t/1_q019nh5u

May 12 **IV OCCAM software**
Willett & Zwick; Fusion, Willet & Zwick; Kramer et al
https://media.pdx.edu/media/t/1_by6uones

May 17 **III Information-theoretic reconstruction: exercises**
https://media.pdx.edu/media/t/1_zdx8l1q8

May 19 **V Supplementary topics:**
Binning *Zwick, Shu, & Koch,*
Set-theoretic RA *Zwick (CU_4), Zwick & Shu; Zwick (O_II.3)*
State-based RA *Johnson & Zwick, Zwick & Johnson, Zwick (O_II.2)*
https://media.pdx.edu/media/t/1_82syd2rv

May 24 **IV OCCAM software: exercises**
https://media.pdx.edu/media/t/1_unhmbgr4

May 26 **V Supplementary topics (continued) & exercises (on May 19 topics)**
Log-linear modeling *K&B: 5-29, 42-63*
Reconstruction vs. identification, masks, k-systems
https://media.pdx.edu/media/t/1_yhw2f5sm

May 31 UNIVERSITY HOLIDAY

June 2 **General review; MICRO-PROJECTS DUE**
https://media.pdx.edu/media/t/1_ps9nlk11

June 7 **FINAL EXAM**: will cover material from entire course,
except additional V topics discussed on May 26.
https://media.pdx.edu/media/t/1_ajn2wpon

To be distributed in class or available on web/D2L:

Chi-square table, Midterm and final exams: 2016, 2018, 2019

For March 29: Zwick, M. "Reconstructability Analysis Overview for DMM Class" In D2L.

Associated Papers

For March 29

Zwick, M. "Overview of Reconstructability Analysis (O)." Kybernetes, 33 (5-6), 2004, pp. 877-905. https://works.bepress.com/martin_zwick/57/

Zwick, M. "Wholes and Parts in General Systems Methodology." The Character Concept in Evolutionary Biology, Gunter Wagner, ed., Academic Press. pp. 237-256, 2001. https://works.bepress.com/martin_zwick/52/

For April 21

Zwick, M. "Control Uniqueness in Reconstructability Analysis (CU: not 4)." Int. J. General Systems, vol. 24 (1-2), 1996, pp.151-162. https://works.bepress.com/martin_zwick/61/

Zwick, M. "Reconstructability Analysis of Epistasis (RAE)." Annals Human Genetics 75(1): 157-171. https://works.bepress.com/martin_zwick/3/

For May 17

Willett, K. & Zwick, M. "A Software Architecture for Reconstructability Analysis." Kybernetes, 33 (5-6), 2004, pp. 997-1008. https://works.bepress.com/martin_zwick/55/

Zwick, M (2016). "OCCAM: A Reconstructability Analysis Program." OCCAM Manual at https://works.bepress.com/martin_zwick/164/

Kramer, P., Westaway, S., Zwick, M., and Shervais, S. (2012) "Reconstructability Analysis of Genetic Loci Associated with Alzheimer Disease." IEEE 6th International Conference on Soft Computing & Intelligent Systems &13th International Symposium on Advanced Intelligent Systems, SCIS-ISIS2012, Kobe, Japan, Nov. 20-24, 2012. https://works.bepress.com/martin_zwick/92/

For May 19

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. https://works.bepress.com/martin_zwick/148/

Zwick, M. & Shu, H. "Set-Theoretic Reconstructability of Elementary Cellular Automata." Advances in Systems Science and Applications (1995), Special issue 1, pp. 31-36. https://works.bepress.com/martin_zwick/79/

Johnson, M.S. and Zwick, M. "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. https://works.bepress.com/martin_zwick/74/

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. https://works.bepress.com/martin_zwick/47/