

STATISTICAL MODELING WEEK

bridging the gap between theory and practice

October 17-21, 2011

Boston

Course 1: October 17-18, 2011

Key Driver Regression in Practice:
Challenges and New Solutions
Nicole Krämer, Tony Babinec & Jay Magidson

Course 2: October 19-20, 2011

Latent Class and
Finite Mixture Modeling
Jay Magidson & Tony Babinec



Examples with
CORExpress®,
Latent GOLD®,
LG Choice,
and more!

Course 3: October 21, 2011

Applications of Latent Class Models with
Discrete Choice, MaxDiff, and Rankings Data
Jay Magidson & Gary Bennett

**Three important state-of-the-art workshops that
integrate theory with real world applications!**

Key Driver Regression in Practice: Challenges & New Solutions

Course 1: October 17-18

Instructors: Nicole Krämer, Tony Babinec & Jay Magidson

DESCRIPTION

The identification of key drivers and their contributions to customer satisfaction, loyalty or other dependent variable is an important application of regression modeling. Many challenges exist in model development and interpreting results, especially when dealing with many correlated predictors and multicollinearity. In this course, we review these challenges and show how advances in high dimensional data analysis suggest new measures of variable importance and allow reliable models to be developed even when the number of predictors exceeds the number of cases! Our applications-oriented presentation provides insight into how the new approaches work through examples and by providing an overview of the relevant theory, supplemented by the supporting equations. We use real and simulated data sets to illustrate the different approaches with XLSTAT®, SPSS®, CORExpress® and various R packages.

Day 1

- A. Traditional regression modeling
 - 1. Challenges due to many correlated predictors
 - a. Linear regression
 - b. Logistic regression and ROC curves
 - c. Discriminant analysis
 - d. Failure of stepwise procedures
 - 2. Controls for overfitting
 - 3. Model selection criteria
 - a. p-values
 - b. Information criteria: AIC, BIC
 - c. Cross-validation statistics
 - 4. Correlated predictors and multicollinearity
 - 5. Variable reduction based on variable importance

Day 2

- A. Case studies/applications of key driver regression
 - 1. Orange juice ratings data
 - 2. Job satisfaction data
- B. Importance of including suppressor variables as predictors
- C. Using cross-validation to assess prediction error
 - 1. Penalized regression
 - 2. PLS regression
 - 3. Correlated Component Regression (CCR)
- D. Simultaneous model estimation and variable reduction
 - 1. Enhancing model performance with suppressor variables
 - 2. Scree plots and coefficient path plots
 - 3. Interpreting cross-validation statistics
- E. Segmentation: using hybrid models to capture heterogeneity

WHAT YOU WILL LEARN

- How to develop reliable models, even with extreme multicollinearity and when # predictors > # cases
- Why many popular variable selection techniques are suboptimal
- About the powerful step-down variable reduction technique in CORExpress®
- About free and commercially available software for analyzing and interpreting high dimensional data

Latent Class and Finite Mixture Modeling

Course 2: October 19-20

Instructors: Jay Magidson & Tony Babinec

DESCRIPTION

Interest and usage of Latent Class (LC) models continues to grow, due to the fact that they provide better solutions than traditional approaches to cluster, factor and regression analysis when the population is not homogenous. In this 2-day course we introduce LC as a probability model and describe various applications using the latest version of Latent GOLD®. On day 1, we focus on model fitting strategies and the interpretation of output. On day 2, we consider several advanced topics including identification issues, random effects with continuous factors, repeated observations, hidden (latent) Markov models for latent growth, and the computation of individual-level coefficients.

THE PROGRAM

Day 1

- Basic ideas of latent class analysis
- The concept of local independence
- The general probability model
- Handling nominal, ordinal, continuous and count variables
- LC measurement models
- Discrete vs. continuous factor analysis
- Comparing models and assessing fit
- Inclusion of covariates in LC models
- Identification problems and boundary solutions
- Use of Bayes constants to eliminate boundary solutions
- The problem of local solutions
- Bivariate residuals to diagnose local dependencies
- Case studies and computer demos

Day 2

- LC regression models
- Relationship to random effects regression
- Model-based clustering / latent discriminant analysis
- Repeated measures / conjoint marketing
- LC Growth models
- Model specification using LG-Equations™

WHAT YOU WILL LEARN

- How to specify LC cluster, factor and regression / segmentation models
- Interpreting graphical displays of results
- What to look for when examining output
- Basic and advanced uses of the Latent GOLD program
- Strategies for assessing model fit with sparse and non-sparse data
- How to isolate the scale effects in ratings data
- How to develop models with mixed scale types
- Why LC models improve over K-means clustering

Applications of Latent Class Models with Discrete Choice, MaxDiff, and Rankings Data

Course 3: October 21

Instructors: Jay Magidson & Gary Bennett

DESCRIPTION

Latent Class (LC) models are natural tools to analyze discrete choice, MaxDiff and rankings data to identify segments with differing preferences. These models are widely used to forecast market share, design optimal products and services, and more. This course begins by introducing the theory and practical applications of these models in traditional choice, rating, ranking, MaxDiff and constant sum experiments in conjunction with the latest version of the Latent GOLD Choice program. We then show how to extend these models to improve interpretation by separating out potentially confounding scale factors and incorporating additional data from the survey into the model to determine absolute as well as relative preferences. Examples include two case studies, one covering a trade-off among potential customers with a selection of 'best' and 'worst' from alternative configurations for a new product, and the other from a more traditional choice design.

THE PROGRAM

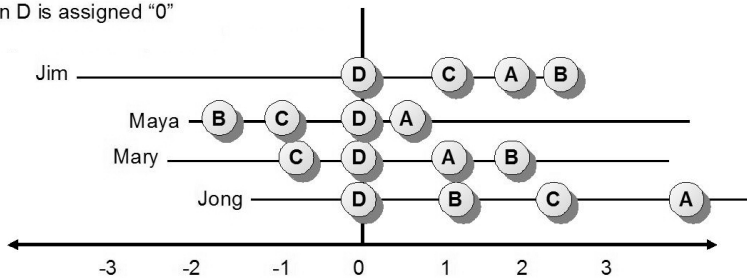
- Development of Excel-based simulators
- Stated preference vs. revealed preference data
- Experimental designs for stated preference
- Independence of Irrelevant Alternatives (IIA)
- Accounting for segment differences - Latent Class vs. HB

WHAT YOU WILL LEARN

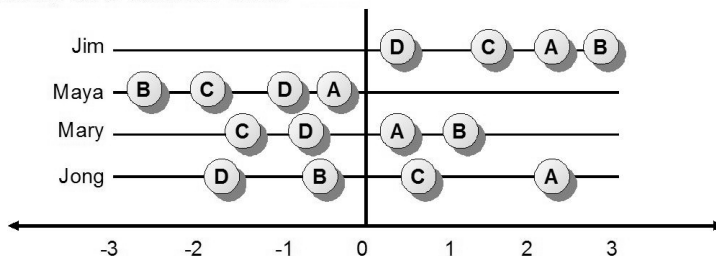
- Overview of the theory of choice modelling
- The LC Choice approach – comparison with HB
- The family of choice model specifications/types
- Alternative approaches to experimental designs
- How to specify, estimate, and interpret the results from choice models
- Advances in Choice Modelling

Importance on Relative Scales

Option D is assigned "0"



Importance on a Common Scale



LG-Equations™ Syntax

```
//LC MAXDIFF WITH 2 SCALE FACTORS
variables
  caseid ID;
  repscale sweight;
  choicesetid setid2 ;
  dependent Q5 ranking;
  independent set nominal inactive;
  attribute attr nominal;
  latent
    Class nominal 5, sClass nominal 2
    coing=first, scale continuous;
equations
  Class <- 1;
  sClass <- 1;
  Scale <- (1)1 + (+) sClass;
  (0) Scale;
  Class <-> sClass;
  Q5 <- attr scale | Class;
```

LG Choice Syntax Model

Enrollment & Information

Attendance is limited to allow for your active involvement and productive interaction with other participants and seminar instructors

Tuition

PREPAID TUITION, which includes lunch and all course materials is \$895 for the 1-day seminar (Course 3) and \$1,490 for the 2-day seminars (Course 1 and 2).

Discounts are available for multiple course registrations. After enrolling in the first course for an undiscounted tuition, you deduct \$50 for each additional course you sign up for (the maximum discount would be \$100 if all 3 courses are taken).

Discounts are also available for additional people who sign up from your organization. Their tuition is \$795 for the 1-day seminar and \$1,390 for each 2-day seminar. **Make checks payable to Statistical Innovations, Registrar.**

Program Hours

9:00-5:00, first day of each course
(registration from 8:30-9:00)

8:30-4:30, second day of Course 1 & 2

Register Online:

www.statisticalinnovations.com

or

Call: (617) 489-4490

Payment & Cancellation Policy

In order to confirm your place in a course, we require payment in full. Tuition is fully refundable for cancellations made by September 17. Cancellations made after that date until October 1 are subject to a \$150 service charge. For cancellations made after October 1, tuition is not refundable. Registration for any course may be transferred to another member of the same organization without penalty.

Note: All workshop staff have agreed to participate in these workshops. We reserve the right to substitute for staff who are unable to attend for health or other emergencies. We will attempt to notify all registrants if any changes in staffing occur.

Workshop Site and Hotel

Accommodations

Workshop Site

Holiday Inn Boston at Beacon Hill
5 Blossom St.
Boston, MA 02114
(617) 742-7630

Holiday Inn Boston at Beacon Hill Special Room rate: \$229

(to guarantee this rate, rooms must be booked by September 16, 2011).

For reservations contact the hotel at (617) 742-7630

Boston hotel space goes quickly, so please reserve your hotel room as soon as possible. Remember to mention **Statistical Innovations** to take advantage of the special rate.

Contact Us:

Statistical Innovations Inc.
(617) 489-4490
will@statisticalinnovations.com

Enrollment Fees

(1) Key Driver Regression in Practice: Challenges and New Solutions	(10/17-18)	\$ 1,490
(2) Latent Class and Finite Mixture Modeling	(10/19-20)	\$ 1,490
(3) Applications of Latent Class Models with Discrete Choice, MaxDiff, and Rankings Data	(10/21)	\$ 895

Seminar Leaders



Jay Magidson is founder and president of Statistical Innovations Inc., a Boston based consulting, training and software development firm specializing in innovative applications of statistical modeling. His clients have included A.C. Nielsen Co., Chemical Bank, National Geographic Society, and Pfizer. He taught statistics at Tufts and Boston University and is widely published on the theory and applications of multivariate statistical methods. Dr. Magidson designed SPSS CHAID, GOLDMineR® and CORExpress®, and is the co-developer (with Jeroen Vermunt) of the Latent GOLD® and Latent GOLD® Choice programs.



Gary Bennett is founder of Logit Research, a statistical consultancy formed in April 2004. Gary is an applied Market Researcher with 20 years experience in the UK Marketing Research Industry, including five years in director roles at leading agencies. He has managed hundreds of discrete choice projects across a diverse array of sectors from Pharmaceutical and Medical to Media and Political applications. He is an advanced user of Latent GOLD and has incorporated many of its flexible features into his modelling work. In 2002 he studied at the EPFL in Switzerland, under the guidance of among others, Prof. Daniel McFadden, the Nobel Laureate who pioneered the use of Choice Modelling Techniques.



Tony Babinec, President of AB Analytics, specializes in the application of statistical and data mining methods to the solution of business problems. Tony has multiple degrees from the University of Chicago, where he specialized in Advanced Statistics and Survey Research. Prior to AB Analytics, Tony worked for over two decades at SPSS Inc. in a variety of technical and managerial roles. He has presented at numerous professional meetings. He is on the Board of Directors of the Chicago Chapter of the American Statistical Association, where he has held various offices including President. He is on the Editorial Board of the Journal of Targeting, Measurement and Analysis for Marketing.



Nicole Krämer is a senior researcher at the Department of Mathematical Statistics at Technische Universität München, Germany. Her research concentrates on the statistical analysis of high-dimensional data. She is a leading expert in Partial Least Squares modeling, and is the recipient of the 2011 Addinsoft award for significant contributions to this area. Recent research projects also include variable selection methods and copula models. She teaches courses on data mining, multivariate statistics, biostatistics, linear models, machine learning, and statistical computing.

Transforming *Data Overload* into Meaningful Information and Improved Predictions

Dear Researcher/Research Manager:

A major problem facing researchers today is that commonly accessible statistical methods have not kept pace with the vast amounts of available data and related challenges. The challenges of extracting meaningful information and obtaining reliable predictions can no longer be met by traditional techniques. Our courses are designed to expedite the process of learning the important new innovative breakthrough methods by showing how these methods can be used effectively in important real-world applications.

Whether you work in marketing, health care, biomedical research or some other area, you will learn more than the techniques - you will also gain the theoretical insights needed to apply the techniques properly.

Take an important step towards mastering your data overload and related challenges by signing up today for one or more of these important sessions.

Sincerely,

Jay Magidson, Ph.D., Program Chair

Prerequisites for Attendees

Two courses in statistics and some experience performing regression analysis in practice.

Special Features of These Seminars

Limited Attendance

Attendance is limited at each session to allow for active involvement and encourage productive interaction with other participants and the seminar leaders.

Seminar Leaders

Seminars are conducted by internationally recognized authorities who are experienced in seminar leadership.

Comprehensive Seminar Workbooks

What you see at the seminar is in the workbook! These comprehensive workbooks will help you learn more while at the program, and will become an invaluable reference source when you return to your responsibilities on the job.

Optional Software Demonstrations

To bring the applications to life, selected software programs will be used to demonstrate applications with real data.

Lunch

Lunch is included and presents a valuable opportunity to interact with both course leaders and other attendees.

A sampling of institutions that have participated in previous workshops:

A.C. Nielsen	Mercedes Benz
Aetna	Merck & Co.
Allstate	MOBIL Oil Corp.
American Express	National Geographic Society
AT&T	National Institutes of Health
BancOne	Naval Surface Warfare Center
Blue Cross Blue Shield	Novartis
Boston Edison	Northwestern University
Bureau of Labor Statistics	Pharmacia
Capital One	Pfizer Pharmaceuticals
CBS	Price Waterhouse
Chase Manhattan Bank	Pinceton Survey Research
Ciba-Geigy Pharmaceuticals	Prodigy Services Company
Citibank	Quaker Oats
Columbia University	Raytheon Company
Dean Witter	RJR Nabisco
Dow Jones & Co.	Scotia Bank
Dun & Bradstreet	Siemens Medical Research
Eastman Kodak	Simon Schuster
Equifax	Smithsonian Institute
Franklin Mint	Stanford University
Gartner Custom Research	Swissair
General Electric	The Polk Co.
General Motors Corp.	Transportation Systems Center
GlaxoSmithKline	University of Michigan
Hallmark Cards	University of Wisconsin
Harvard University	U.S. Automobile Association
HBO	U.S. Dept. of Agriculture
IBM	Veterans Administration
Kellogg Co.	Wang Laboratories
Liz Claibourne	Westinghouse Health Systems
Marriott Corp.	World Bank
MCI	Yale University

What people are saying about Statistical Modeling Week:

"Eye-opening introduction to a new class of models"
Richard Popper, Vice President - R&D, Peryam & Kroll

"I have discovered a brand new way of extracting meaning from my data"
Richard Shewchuk, Professor, University of Alabama

"Perfect level of in-depth statistical thinking and practical applications"
Matthew Gray, Manager, Quantitative Commercial Insight, AstraZeneca

Online Courses

Can't attend Statistical Modeling Week?

We also offer Online Courses.
Learn about modeling techniques
from the comfort of your own home.

For more information, go to
www.statisticalinnovations.com



Statistical Modeling Week 2011
Oct. 17 - Oct. 21, 2011
Space is limited! Register today!

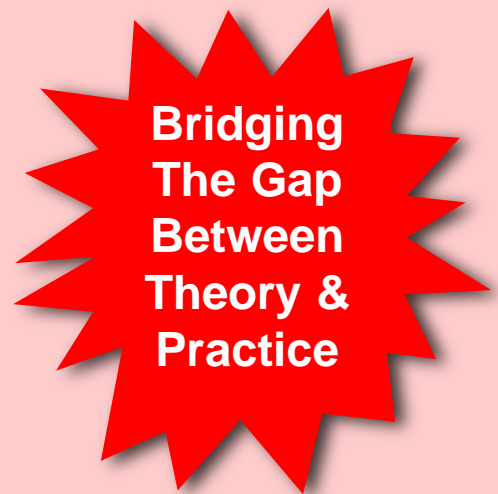
P.O. Box #1
Belmont, MA 02478



STATISTICAL MODELING WEEK 2011

Oct. 17 - Oct. 21, 2011
Boston, MA

Call us at **(617) 489-4490** or
visit our website at
www.statisticalinnovations.com
for more information or to register.



**Three important state-of-the-art workshops that
integrate theory with real world applications!**