The Wharton School
University of Pennsylvania

MKTG 940 Measurement and Data Analysis - Part A
MKTG 941 Measurement and Data Analysis - Part B
Spring 2017

Monday, 9:00-12:00 PM

Christophe Van den Bulte
Office: 759 JMHH
Office hours: By appointment (send an e-mail)
Tel: 898-6532 Fax: 898-2534
E-mail: vdbulte@wharton.upenn.edu

Course objectives

The purpose of these two sequenced courses is to provide students an understanding and working knowledge of statistical data analysis techniques commonly used in marketing. The focus is on techniques that provide insights in how one variable is predicted (and possibly caused) by other variables.

The courses are designed to complement MKTG 942/943, Research Methods in Marketing (A/B), which focuses on (i) the general linear model (linear regression and ANOVA) and on (ii) experimental data. The present courses extend the students' tool kit in two directions:

1. Analyzing binary data, choice data, count data, and duration data using Generalized Linear Models.
2. Identifying and tackling challenges when analyzing non-experimental data.

Prerequisites

For MKTG 940: MKTG 942/943, or a graduate-level course on linear regression or ANOVA.
For MKTG 941: MKTG 940.

Course format

The class will meet once a week, on Mondays 9:00-12:00 PM.

I think of and designed the 940/941 sequence as a single one-semester course. But, since the sequence is administratively split up into two units, it is possible to take MKTG 940 without also taking 941.

There will be several problem sets / homework assignments. Students will also analyze a data set from a project they are working on or re-analyze the data from a published paper, present the process they went through on the final day of class, and write up a report.

A typical class session will consist of (i) a debrief on the recently submitted problem set or homework assignment (if applicable), (ii) a lecture on a given topic, (iii) examples and discussion of practical estimation issues. Occasionally, there will also be a discussion of assigned application readings.

List of topics

Regression-type models for analyzing …

1. Binary data
2. Multinomial / choice data
3. Count data
4. Duration data
5. Other censored/truncated data; Data with self-selection

Identifying and tackling challenges in non-experimental data

1. Influence points, collinearity / ill-conditioning, and missing data
2. Selectivity and other sources of endogeneity
3. Designs and statistical methods to strengthen causal identification
4. Application: Peer influence in social networks

Statistical software

I will be using SAS in class. Students are welcome to use other statistical packages they are familiar with (e.g., Stata, R, SPSS, JMP) for assignments. All analyses we'll cover can be performed using SAS, Stata and R. I am not quite sure about SPSS or JMP, but most analyses definitely can. I promise complete support only for SAS.

Course materials

There is no assigned textbook. Readings, data sets, etc. will be made available on Canvas.

As preparation for the session on analyzing censored, truncated, and self-selected data, I suggest you read the following little book retailing for about $14:

> Breen, Richard. 1996. *Regression Models: Censored, Sample Selected, or Truncated Data.* Sage Publications, Thousand Oaks, CA.

Reference materials: General

While there is no assigned (must buy) textbook, the following books provide additional details and background:

Long, J. Scott, 1997. *Regression Models for Categorical and Limited Dependent Variables*. Sage Publications, Thousand Oaks, CA.

Cameron, A. Colin and Pravin Trivedi. 2005. *Microeconometrics: Methods and Applications*. Cambridge University Press, Cambridge, UK.

Dobson, Annette J. and Adrian G. Barnett. 2008. *An Introduction to Generalized Linear Models, 3rd ed.* Chapman & Hall / CRC Press, Boca Raton, FL.

Train, Kenneth. 2009. *Discrete Choice Methods with Simulation, 2nd ed*. Cambridge University Press, Cambridge, UK. It is available for free here - http://elsa.berkeley.edu/books/choice2.html

Maddala, G.S. 1983. *Limited-dependent and Qualitative Variables in Econometrics*. Cambridge University Press, Cambridge, UK.

Morgan, Sephen L. and Christopher Winship. 2007. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge University Press, Cambridge, UK.

When looking for a very easy primer on a specific topic or technique, consider the "little green books" in the Quantitative Applications in the Social Sciences published by Sage. I found several of them quite useful when I was a student. They are listed here: http://srmo.sagepub.com/browse?doctype=qass

Reference materials: Software guides

There are several books providing hands-on guidance that you may find useful, also after the course.

For SAS, I have found the following two books by Penn Sociology professor Paul Allison especially useful:

Allison, Paul D. 2012. *Logistic Regression Using SAS: Theory and Application, 2nd ed.* SAS Institute, Cary, NC.

Allison, Paul D. 2010. *Survival Analysis Using SAS: A Practical Guide, 2nd ed.* SAS Institute, Cary, NC.

Similar books exist for Stata and R. Examples for Stata are:

Cameron, A. Colin and Pravin Trivedi. 2010. *Microeconometrics Using Stata, Revised Edition*. Stata Press, College Station, TX.

Long, J. Scott and Jeremy Freese. 2014. *Regression Models for Categorical Dependent Variables Using Stata, 3rd ed.* Stata Press, College Station, TX..

Cleves, Mario, William Gould, Roberto G. Gutierrez, and Yulia V. Marchenko. 2010. *An Introduction to Survival Analysis Using Stata, 3rd ed.* Stata Press, College Station, TX.

Classroom

Our meetings will be held in 741 JMHH.

Grading

Problem sets / Assignments	35%
Project Write-up & Presentation*	15%
In-class contributions	10%
Take-Home Final Examination	40%

* The presentation is an opportunity for you to get feedback before finalizing your project. I expect your presentation to be coherent and clear, but do not grade it for rigor / correctness. Unless your presentation is exceptionally poor or exceptionally good, that 15% of the grade is driven by the write-up only.

Plan of Sessions

1.	Jan. 11*	Introduction – Beyond Classical Linear Regression
	Jan. 16	*No class – MLK Day*

A. Generalized Linear Models

2.	Jan. 23	Beyond Classical Linear Regression & Binary data I
3.	Jan 30	Binary data II
4.	Feb. 6	Binary data III
5.	Feb. 13	Multinomial data I
6.	Feb. 20	Multinomial data II & Ordered response data
7.	Feb. 27	Count data
	Mar. 6	*No class – Spring Break*
8.	Mar. 13	Duration data

B. Challenges with Non-Experimental Data

9.	Mar. 20	Censored, truncated, and self-selected data
10.	Mar. 27	Endogenous regressors
11.	Apr. 3	Repeated/clustered observations and unobserved heterogeneity
12.	Apr. 10	Ill-conditioning, outliers, and influence points

C. Topical Applications

13.	Apr. 17	Peer influence in networks
14.	Apr. 24	Presentation of student analyses

* This is a Wednesday.