The Wharton School
University of Pennsylvania

MKTG 940 Measurement and Data Analysis - Part A
MKTG 941 Measurement and Data Analysis - Part B
Spring 2022

Monday, 8:30-11:45 PM

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

Course objectives

This sequence of two half-semester courses provides students an understanding and working knowledge of statistical data analysis techniques commonly used in marketing. The focus is on techniques that provide insights into 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). The latter focuses on (i) linear modeling (linear regression and ANOVA) and on (ii) experimental data. The present courses extend the students' tool kit in two directions:

1. Analyzing binary data, ordered response data, choice data, count data, truncated or censored data, and duration data using Generalized Linear Models.
2. Identifying and tackling challenges with making causal claims in non-experimental designs.

In short, MKTG940/941 is about "funny Y's and messy X's."

Prerequisites

For MKTG 940: MKTG 942/943, or a graduate / advanced undergraduate level course on regression.
For MKTG 941: MKTG 940.

Course format

The class meets once a week, on Mondays 8:30-11:45 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—typically a weekly assignment for the first 10 weeks. 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 consists of (i) a debrief on the recently submitted homework assignment (if applicable), (ii) a lecture on a given topic, and (iii) examples and discussion of practical estimation and interpretation issues.

List of topics

Regression-type models for analyzing …
1. Binary data
2. Ordered response data
3. Multinomial / choice data
4. Count data
5. Duration data
6. Other censored/truncated data; Data with selectivity

Identifying and tackling challenges in non-experimental data
1. Potential Outcomes / Rubin Causal Model framework to causal inference
2. Designs and statistical methods to strengthen causal identification
    - Diff-in-Diff; Matching; Regression Discontinuity; Instrumental variables
3. Endogeneity in mediation analysis

Megastudies & A/B/n tests

Statistical software

I will be using SAS in class. Students are welcome to use any statistical packages they are familiar with. All analyses we cover can be performed using SAS, Stata, and R. I am not quite sure about Python, SPSS or JMP, but most analyses definitely can.

Course materials

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

As preparation for the session on analyzing censored, truncated, and self-selected data, I ask that 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.

It is also available online through the Penn Library: http://hdl.library.upenn.edu/1017.12/2246772

Suggestions in case you really want some textbooks

For the section on Generalized Linear Models, the following provides additional details and background, roughly at the level of the course:

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

For the section on causal inference from non-experimental data, the following provide additional details and background, be it at a somewhat higher level than the course:

Lee, Myoung-Jae. 1997. *Matching, Regression Discontinuity, Difference in Differences, and Beyond.* Oxford University Press, Oxford, UK.
Available online through the Penn Library:
https://franklin.library.upenn.edu/catalog/FRANKLIN_9977626093203681

Angrist, Joshua D. and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press, Princeton, NJ.
Available online through the Penn Library:
http://hdl.library.upenn.edu/1017.12/2550435

Classroom

Our first day of class will be on a Wednesday, and will meet via Zoom. Subsequent classes are currently planned to be held in person, in JMHH 741 (seminar room in the Marketing Suite).

Grading

Problem sets / Assignments	45%
In-class contributions	10%
Project Write-up & Presentation*	15%
Take-Home Final Examination	30%

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

Plan of Sessions

1.   Jan. 12*   Introduction – Beyond Classical Linear Regression

A. Regression-type models for funny Y's
(Generalized Linear Models)

2.   Jan 24   Beyond Classical Linear Regression & Binary data I
3.   Jan 31   Binary data II
4.   Feb. 7   Binary data III
5.   Feb. 14   Multinomial data I
6.   Feb. 21   Multinomial data II & Ordered response data
7.   Feb. 28   Count data
8.   Mar. 14   Duration data
9.   Mar. 21   Censored, truncated, and self-selected data

B. Causal inference with messy X's
(Non-experimental Data)

10.   Mar. 28   Potential outcomes framework of causal inference ('Rubin causal model')
                        Matching & Regression discontinuity]
11.   Apr. 4   Differences-in-Differences & Repeated/clustered observations
12.   Apr. 11   Instrumental variables & Mediation analysis

C. Special Topics / Topical Applications

13.   Apr. 18   Megastudies & A/B/n tests
14.   Apr. 25   Presentation of student analyses

*This is a Wednesday.

Appendix: Additional reference materials

General

When looking for an accessible 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

Here are some other books you might find useful.

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

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

Morgan, Stephen L. and Christopher Winship. 2014. *Counterfactuals and Causal Inference: Methods and Principles for Social Research, 2nd ed.* 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.

Hands-on software guides

For SAS, the following two books by Penn Sociology professor Paul Allison are quite 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:

Moffatt, Peter G. 2016. *Experimetrics: Econometrics for Experimental Economics*. Palgrave Macmillan, New York.

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.