

The Wharton School
University of Pennsylvania

MKTG 9400 Measurement and Data Analysis - Part A
MKTG 9410 Measurement and Data Analysis - Part B
Spring 2024

Monday, 8:30-11:30 AM
741 JMHH

Christophe Van den Bulte
Office: 759 JMHH
Office hours: By appointment (send an e-mail)

Tel: (215) 898-6532
E-mail: vdbulte@wharton.upenn.edu

Course objectives

This sequence of two half-semester courses provides students an understanding and working knowledge of statistical data analysis techniques commonly used in academic marketing research. 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 9420/9430, 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, MKTG9400/9410 is about “funny Y’s and messy X’s.” We will also dedicate one session to challenges with analyzing data from A/B/n tests, intervention tournaments and megastudies.

Prerequisites

For MKTG 9400: MKTG 9420/9430, or a course on regression.
For MKTG 9410: MKTG 9400.

Working knowledge of a general-purpose statistics platform / package like R, Stata or SAS.

Course format

The class meets once a week, on Mondays 8:30-11:30 AM.

I designed the 9400/9410 sequence as a single one-semester course. But since the sequence is administratively split up into two units, it is possible to take MKTG 9400 without also taking 9410.

There will be a weekly homework assignment for the first 10-11 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 (1) a debrief on the recently submitted homework assignment, (2) a lecture on a given topic, and (3) 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 sample selection

Megastudies, intervention tournaments, and A/B/n tests

Making causal claims from non-experimental data

1. Potential outcomes framework to causal inference ('Rubin Causal Model')
2. Methods to strengthen causal identification
[Matching; Diff-in-Diff; Regression Discontinuity; Instrumental variables]
3. Endogeneity in mediation analysis

Statistical software

I will be using SAS in class, but students are welcome to use any statistical packages they are familiar with. All analyses we cover can be performed using R, Stata, and SAS. 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:

Breen, Richard. 1996. *Regression Models: Censored, Sample Selected, or Truncated Data*. Sage Publications, Thousand Oaks, CA. <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 provide additional details and background, roughly at the level of the course. The second includes sample R and Stata code.

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

Smithson, Michael & Merkle, Edgar C. 2014. *Generalized Linear Models for Categorical and Continuous Limited Dependent Variables*. CRC Press, Boca Raton, FL.

Classroom

We will meet in JMHH 741, the large 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

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

1. Jan. 22 Introduction & Beyond Classical Linear Regression I
2. Jan. 29 Beyond Classical Linear Regression II & Binary data I
3. Feb. 5 Binary data II
4. Feb. 12 Binary data III
5. Feb. 19 Multinomial data I
6. Feb. 26 Multinomial data II
7. Mar. 11 Ordered response data & Count data
8. Mar. 18 Duration data
9. Mar. 25 Censoring, truncation, and sample selection

B. Special topic

10. Apr. 1 Megastudies, Intervention tournaments, and A/B/n tests

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

11. Apr. 8 Mediation analysis & Instrumental variables
12. Apr. 15 Potential outcomes framework of causal inference ('Rubin causal model')
Matching & Regression discontinuity
13. Apr. 22 Differences-in-Differences & Repeated/clustered observations

D. Topical Applications

14. Apr. 29 Presentation of student analyses

Appendix: Additional reference materials

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.

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

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.

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

For the section on causal inference from non-experimental data, the following provide additional details and background, be it at a 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>