

Mgmt 469

Empirical Methods in Strategy

Lies, Damned Lies, and Statistics

Despite the obvious importance of data-driven analyses for business decision making, most managers lack experience working with data, above and beyond the basics of spreadsheet analysis. As a result, managers either perform inadequate evaluations of the data themselves, or turn the data over to statisticians who may lack the relevant institutional knowledge. Ideally, data-driven evaluation should combine both skills (data analysis and business expertise). This course is intended to help you obtain skills with data analysis. Bearing this in mind, this course has several goals:

- To familiarize you with the process of assembling and analyzing a data set
- To expose you to good research methods
- To reinforce statistical techniques learned in your DS class and also introduce new statistical techniques
- To increase your comfort level working with data
- To introduce Stata, a powerful yet user-friendly empirical research software package
- To enable you to assess the research of others and to perform convincing research of your own.
- To help you distinguish convincing research from damnable lies.

Pedagogy

This is a *clinical* course. You may be familiar with the educational process in other clinical fields, such as medicine. A famous dictum in medical education is "see one, do one, teach one." We will adhere to the same philosophy. I will give several lectures during the quarter, but you will spend most of your classroom time (and virtually all of your time outside the classroom) performing and discussing empirical research. You will design and implement three research projects that I have chosen to highlight a range of key research issues. You will also be asked to read and discuss 3 empirical research papers. Such "learning by doing" is surely the best way to become an accomplished empirical analyst.

Roughly speaking, the course is divided into four modules. The first module consists of lectures that build rapidly from basic to advanced research methods. The centerpiece of each remaining module is the analysis of a real world data set. Each data set allows you to answer an important managerial question, such as "Identify entry opportunities for internet service providers (ISPs)."

Working in small groups, you will answer the managerial question by developing and implementing an empirical model. During each module we will conduct “lab sessions” in the classroom to discuss proposed research designs. Each module will end with several group presentations of their research findings.

Statistical software

Your statistical analyses must be performed in **Stata**. Stata Version 9 has been installed on computers in the Kellogg computer lab. I will have much more to say about Stata (including providing a Stata primer) in a later class. While Stata is very easy to use when compared with comparably powerful software, you will find yourself spending a disproportionate amount of time during the first two or three weeks mastering the basics of Stata. Some students find this to be the most challenging part of the course. I hope that the upfront fixed time investments pay off by dramatically reducing the marginal cost of doing good empirical work down the road.

Why learn Stata when you know Excel? Excel is not designed for sophisticated analyses you will want to perform in practice.¹ Moreover, you cannot readily program in Excel; this makes it difficult for you (or others) to reproduce your work.² Reproducibility is a hallmark of good research! There is no excuse for you to perform inadequate work merely because your software does not permit you to do something better. In fact, most firms that do a lot of empirical research, including all the brand-name consulting firms, use powerful statistical packages like Stata. Due to its relative ease of use and the network effect of a growing community of users, Stata has become the package of choice among empirical researchers. (Stata is by far the package of choice among Kellogg faculty.)

Stata offers a graphical user interface (GUI). Most students prefer to use old-fashioned typed commands. This is due to Stata’s programming power and the many options associated with each command (making GUI rather complex to use.) I will rely exclusively on typed commands in class.

Stata has fine graphics but it is not in the same league as Excel. Thus, you may want to use Excel for graphics. I will show you how to quickly move data back and forth between Stata and Excel. Some of you may be interested to know that a software called Stat/Transfer enables you to convert data from one statistical format to another (e.g. from SAS to Stata)

If you have never used a statistical software package before, you will find it a bit difficult to adjust to Stata at first. You will need to learn a new “language” and “rules of grammar”. Rest assured that by

¹ At least one study claims that the Excel algorithm for computing even the simplest regression coefficients and standard errors is incorrect.

² You can write macros of any length to create “programs”, but must do so in Visual Basic. Stata is much more user friendly for programming.

mid-quarter, you will be performing sophisticated analyses at lightning speed. In fact, most people who use Stata swear by it. It is not cheap (\$150-\$200), but it is a worthwhile, long-term investment. I will offer instructions on how to order Stata (and which version is best for you) in the introductory class.

Grading

Your course grade will be based on the projects, class participation, and a final take home quiz, broken down as follows:

Yogurt	20%
Sports Teams	20%
Benchmarking	20%
Quiz	25%
Class Participation	15%

Participation

Class participation is very important and is intended to promote the collaborative nature of empirical research. In fact, participation is vital to the learning experience in this course. Be sure to listen to your classmates as they discuss their work. If you can help others become stronger researchers through your questions and comments, then you will receive a strong participation grade. If you say nothing, then you are not helping others and you will receive a low participation score. Also be sure to come to each “Research Paper” class session having more than a passing knowledge of the paper that has been assigned.

Projects

You will work in groups of 4-5 members (depending on total class enrolment). You are free to choose the members of your group – bear in mind that the projects require extensive group work so select your group accordingly. Each group will have to submit a project report (~15 pages) along with a brief presentation (~6 slides). I will provide further details in class. Late submissions will incur a 2 point penalty.

Quiz

The final exam for the class will be in the form of a closed notes take home quiz, which will be made due on or around the 26th of May. The quiz will be handed out a week before the due date and is to be completed in 120 minutes (at a time of your choosing). You may refer to two 2-sided cheat sheets which will have to be attached to your answer sheets. The cheat sheets are to be prepared individually. As with projects, late submissions will incur a 2 point penalty.

Laptop Use Policy

There is no restriction on Laptop use in class – in fact, I would encourage you to use your laptop computers especially during the “lab” sessions. If you would like to reserve a temporary laptop (for the duration of a class) through KIS, then please let me know in advance so that I can do the needful.

Office Hours

I will hold office hours every Tuesday from 4:30 pm onwards.

Course Outline

Topics in *italics* are lectures. "Project start" indicates the start date for each of the research projects. I will discuss research papers in class as examples of good (and sometimes bad) research methods.

<i>Session</i>	<i>Date</i>	<i>Topic</i>	<i>Project Start</i>	<i>Research Papers</i>
1	3/28	Introduction to Course <i>Ten Steps towards Convincing Research</i>		
2	3/28	<i>Using Stata</i> <i>Regression Basics</i>	CAPM (Stata tutorial)	
3	4/4	<i>Regression Basics</i> <i>Getting Your Data Ready</i>	Yogurt 1	
4	4/4	<i>Choosing the Right Variables</i>		
5	4/11	<i>Fixed Effect Models</i> <i>Nonlinear Relationships</i>		
6	4/11	Yogurt "Lab" <i>Programming in Stata</i>	Yogurt 2	
7	4/18	<i>Nonlinear Relationships</i>		
8	4/18	Yogurt "Lab"		
9	4/25	Present Yogurt Results <i>Maximum Likelihood</i>	Sports Teams	
10	4/25	<i>Noisy Variables</i> <i>Discrete Variables</i>		
11	5/2	<i>Discrete Variables</i>		
12	5/2	Sports Team "Lab"		
13	5/9	Accessing Census Data A Research Paper		Measuring Competence? (Henderson/Cockburn, <i>SMJ</i> 1994)
14	5/9	Present Sports Team results	Hospitals	
15	5/16	<i>Causality</i> <i>Benchmarking Methods</i>		
16	5/16	Review Session		
17	5/23	<i>Some useful Stata tricks</i> Hospital "lab"		
18	5/23	A Research Paper		Optimal Scale for R&D (Kuemmerle, <i>Research Policy</i> 1998)
---	5/26	Take Home Quiz due*		
19	5/30	Present Hospital Results		

* The Take Home Quiz will be handed out in class (or emailed) roughly a week in advance of the due date