Welcome!! The syllabus is full of useful information. Please take a careful look before the first class.

Course Description

This course lays the foundations for all of Kellogg’s business analytics courses. Specifically we will develop the statistical techniques that allow for the exploration of relationships between variables primarily through multivariate regression. In addition to teaching basic regression skills, including modeling and estimation, the course will deepen understanding of hypothesis testing and how to make inferences and predictions from data. Students will also learn new principles such as identification and robustness. The course has an intense focus on managerially relevant applications, cases and interpretations.

Motivation and Objectives

In a world that inundates us with data, many managers lack real world experience analyzing 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 subject matter expertise, managerial intuition and understanding of the business issues at stake. Ideally, data-driven evaluation should combine both skills (data analysis and business expertise). Bearing this in mind, this course has several goals:

- To develop an understanding of probability theory for managerial decision making and as the foundational underpinnings for data analytics.
- To introduce you to regression analysis, a powerful tool for exploring relationships among two more or more variables
- To teach you how to build a regression model, interpret regression coefficients, and use the results to test hypotheses and make predictions
- To expose you to advanced techniques in model specification to assure that your analyses are free from unwanted bias and other problems
- To introduce the concept of an “experiment” and demonstrate how to distinguish correlation from causation using a variety of practical approaches to data collection and data analysis
- To increase your comfort level working with different types of data
- To enable you to assess data-driven analyses performed by others, and to perform convincing data-driven analyses of your own
- To increase your ability to work with statisticians as part of an analysis team
- To familiarize you with the software platform used in Kellogg’s elective courses.
**Pedagogy**

Our guiding pedagogical principle is that when it comes to working with data, there is no substitute for hands-on experience. The first part of the course features lectures that present the most relevant concepts from probability theory. The goal is to learn operational skills but also to learn applications of the ideas to practical business decision making. The middle section of the course presents and develops basic elements of regression analysis. During these lectures, we will use real world data to address a variety of management problems. In the final part of the course, we will explore advanced topics in regression analysis. We will work on increasingly more complex data analyses, including cases, providing students with an opportunity to develop modeling skills. This part of the course emphasizes rich business cases that rely on large-scale semi-structured data. These cases will help students refine modeling and presentation skills.

In order to learn these ideas, there is no better way than to practice. There will be a series of drill problems made available on canvas to help you master these ideas. Drill problems are not collected. The problems and solutions are just meant to help you learn how to operationalize the concepts form lectures. The team homeworks will be graded (see below.)

**Course Material and Software**

There is NOT a required textbook to purchase for this course. If you would like a textbook that pushes deeper into the mathematics of probability and statistics, there will be one available at the bookstore:

Vital Statistics, Sandholm et. al.  
**EDITION:** 2017  
**PUBLISHER:** OXF  
**ISBN:** 9780190871048

We will NOT refer to the book in class at all and the treatment is more theoretical than our course. The material in the optional online prep-course is more aligned with our coverage.

The course package contains some of the required reading material for the course, including the textbook manuscript: *Managerial Statistics: A Case-Based Approach: Stata Edition* by Klibanoff, Sandroni, Moselle and Saraniti. This book was developed and written at Kellogg. It was recently redesigned to be used with the software package Stata, the standard statistical software at Kellogg. The course package also includes some HBS cases we will be using. All other course materials may be found on the Canvas course page. You are responsible for regularly checking, downloading, and reading materials posted on the site, as they form an integral part of the class.

---

1 Do not buy the older version of this book, as it does not use the software we will be using – buy the course package!
For those of you seeking more technical coverage of regression, Professor Dranove has written a series of Kellogg Technical Notes (Hereafter KTN) that are posted on canvas. These are more mathematically rigorous than our class but can be helpful IF you have a strong technical background. I have noted the relevant timing of these notes in the course outline below.

**Excel and Stata**

All of you enter the course with at least a passing knowledge of Excel. Excel is of limited value for sophisticated data analyses: Excel has limited capacity for handling large data sets. Perhaps more problematically, Excel permits you to perform only a limited set of analyses, which are often inadequate for the task at hand. For example, Excel does not have a simple way to perform many predictions and hypothesis tests, and does not allow for many tests required to assess for model validity. As a result, analysts who rely on Excel are forced make compromises, presenting results that they are able to generate, rather than the results that they want to generate.

Stata is the standard data analytic software used at Kellogg. Stata is powerful yet user friendly. It has a graphical user (i.e. drop down) interface, and also includes some customizable menus including some made for you at Kellogg. Users who grow comfortable with Stata may prefer to directly type commands instead of relying primarily on menus. Either way, with a bit of practice you will be able to perform advanced data analyses at the drop of a hat. (The biggest challenge is not learning how to get Stata to perform the desired analyses; it is figuring out what analyses are most appropriate to perform and effectively interpreting the results.)

You will need to download and installing the latest version of Stata (currently version 14) and the Kellogg custom menu early in the course. We will rely on it heavily starting in session 5. The software is provided at no cost to you through Kellogg. You get a permanent license so that you may continue to use the installed version even after you have graduated from Kellogg.

It can be quite daunting to learn any new software, even one as user-friendly as Stata. On occasion you will be stymied. A good rule of thumb is that if you cannot figure out how to execute a command within 5 minutes, it is time to seek out help. There are several places you can turn for help:

- There is a guide to frequently used commands posted to Canvas
- There is a Stata mini-manual in the Appendix to the textbook in your course packet
- The Kellogg student Stata support page ([https://www.kellogg.northwestern.edu/rs/software/stata/statamba.aspx](https://www.kellogg.northwestern.edu/rs/software/stata/statamba.aspx)) is a useful resource. There is usually a Kellogg Stata support specialist who is available for all your Stata IT issues (but not for DECS 440 specific problems like “is there a typo on homework problem #2?”)
- The schedule for periodic Stata training sessions can also be found on that page. If you experience software glitches such as Stata crashing or not running at all, then turn to the KIS student support team.
Some of you might prefer to perform your analyses in Stata but create graphics in Excel. Fortunately, it is very easy to transfer data back and forth between Stata and Excel. Please refer to the technical note on Canvas for directions.

**Office Hours, Contact Information**

My office will be in the MEDS department in the Global Hub. On the first day of class we will discuss formal office hours. If you would like to see me and cannot visit during office hours, then write me an e-mail (b-saraniti@kellogg.northwestern.edu) to make an appointment. E-mail is the fastest and most reliable way to reach me outside of class.

For statistical emergencies, my cell: 808-258-7701

**Course Outline**

**Session 1: Friday 6/29 – Probability Basics**

**Readings**
- Kellogg Case: Cost Savings from Pooled Testing
- The Probability of Injustice. (Economist 1/22/04)
- Which Travelers Have Hostile Intent? (Wall Street Journal 8/14/06)
- Very Optional Reference/Background: Sandholm et.al. (Hereafter SS) Chapter 2: Sections 1-5. pp 11-37

**Session 2: 7/2 – Random Variables**

**Readings**
- Kellogg Case: The Value of a Statistical Life
- Kellogg Case: Hole-in-One Insurance
- Market Forecasts (Nature 10/20/2016)
- Geology or Geography? (Economist 10/27/2011)

**Session 3: 7/9 – Correlation, Gauss, Central Limit Theorem**

**Readings**
- The Wisdom of the Crowds – Now with Even More Wisdom (Science 2.0 4/19/2018)
- Very Optional Reference/Background: SS Chapter 4: Sections 1-3. pp 83-99, Chapter 6: Sections 1-4. pp 151-177

**Team Homework 1 Due (on canvas by 1:00 pm)**
Session 4: 7/16 – Sampling Distributions

Readings
- How to Make Jobs Disappear (New York Times 6/19/2012)

Team Homework 2 Due (on canvas by 1:00 pm)

Session 5: 7/23 – Regression: Prediction

Readings
- The Next Job Humans Lose to Robots (Bloomberg 7/19/2017)
- To Predict the Next Box Office Hit... (Wired 6/6/2013)
- Online Textbook: Chapters 3 and 4.

Take Home Midterm Exam I Due (on canvas by 1:00 pm)

Session 6: Friday 7/27 – Regression: Estimating and Benchmarking

Readings
- Online Textbook: Chapter 4.
- KTN: Regression Basics

Session 7: 7/30 – Regression: Slope Dummy Variables

Readings
- Online Textbook: Chapter 5.

Team Homework 3 Due (on canvas by 1:00 pm)

Session 8: 8/6 – Multiple Regression: Multicollinearity & Omitted Variable Bias

Readings
- Online Textbook: Chapter 7.4, 8.1-8.4.
- KTN: Building Models
• **KTN: Omitted Variables**

**Take Home Midterm Exam II Due (on canvas by 1:00 pm)**

**Session 9: 8/13 – Log vs. Linear Models, Heteroskedasticity, Independence**

**Readings**
- **KTN: Log vs. Linear**
- **KTN: Noise, Heteroskedasticity, and Grouped Data**

**Team Homework 4 Due (on canvas by 1:00 pm)**

**Session 10: 8/13 – Fixed Effects Models & Exam Review**

**Readings**
- **KTN: Fixed Effects**

**Team Homework 5 Due (on canvas by 1:00 pm)**
Grading and Assignments

As can be seen from the course outline above, there are homework assignments most weeks which will be posted on the course webpage, accessible through the Canvas site at Northwestern (https://courses.northwestern.edu/webapps/login). Assignments are due by the start of class on the day they are listed as due. They will be made available on the course Canvas website and solutions must be submitted through that site.

All of the collected assignments will be group assignments. The class will be divided into groups and you should do the group homework assignments with your team (i.e., all group members work together on the assignment, one submission per group, submitted one time by one member of the group with all the group members’ names on it).

The final exam will take place according to Kellogg’s exam schedule and will be three hours long. Computer use will be required for the exam.

Your course grade will be based on the following:

<table>
<thead>
<tr>
<th>Assignment Type</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Team Homeworks</td>
<td>20%</td>
</tr>
<tr>
<td>Take Home Midterm I</td>
<td>25%</td>
</tr>
<tr>
<td>Take Home Midterm II</td>
<td>25%</td>
</tr>
<tr>
<td>Final Exam</td>
<td>30% OR 55%</td>
</tr>
</tbody>
</table>

[If your Final Exam Score is higher than your lower Midterm Score, I will throw out your low midterm grade and “overweight” the final. So instead of [20, 25, 25, 30] your scores will be weighted [20, 0, 25, 55] or [20, 25, 0, 55] whichever is highest.

Lastly, I reserve the right to adjust your raw score by up to 10% based on class participation, attendance, and peer reviews. This is rare and will only be done in exceptional cases.

Participation is vital to the learning experience. Class participation and attendance includes being present and on time for class, the quality of your in-class contributions to discussion and also responses to in-class questions. I do “cold call” students in class. You should keep up with what has been going on in class and in the homework and be prepared for questions. Be sure to listen to your classmates. Helping others become stronger through your questions and comments is a key purpose of participation.

Classroom behavior and Norms

General classroom behavior should follow the guidelines in the Kellogg Code of Classroom Etiquette that can be found at the link below:
http://www.kellogg.northwestern.edu/stu_aff/policies/etiquette.htm

Some specific policies that apply to this class are:
ELECTRONICS: You are welcome (and will sometimes need) to bring your laptop to class in order to use Stata to perform class-related calculations and data analysis. However, no other use of your laptop (e.g., note-taking, tracking your portfolio, web surfing, instant messaging, e-mailing, tweeting, or blogging etc.) during class is allowed. Please see: [https://www.scientificamerican.com/article/students-are-better-off-without-a-laptop-in-the-classroom/](https://www.scientificamerican.com/article/students-are-better-off-without-a-laptop-in-the-classroom/)

No other electronic devices may be used in class with the exception of tablets used exclusively to take notes. All ringers, speakers, etc. should be switched off. This will minimize distractions to you and to others.

ATTENDANCE AND PUNCTUALITY: You are expected to attend class and be on time.

**KELLOGG HONOR CODE**

The Kellogg Honor Code applies. The complete text of the Honor Code is available on the Honor Code website: [http://www.kellogg.northwestern.edu/stu_aff/policies/honorcode.htm](http://www.kellogg.northwestern.edu/stu_aff/policies/honorcode.htm)

The Honor Code is enforced at Kellogg and violations are subject to disciplinary sanctions. The following discussion in this syllabus of the Honor Code does not cover all applications of the Honor Code but only highlights some very important aspects of it. If you believe something is unclear or has been omitted, please do not hesitate to speak to me.

Some specific policies applying to the work for this class are:

ASSIGMENTS: Write-ups must be your original work. You may not use materials containing solutions or partial solutions to the assignments (including solutions prepared by current or former Kellogg students). If your analysis contains information from outside sources, then you must properly cite the sources.

Group assignments should be the work of only the individuals in your group, and you may not consult with other students outside your group.

EXAMS: No assistance may be given or received during an exam. Regardless of when you take the exam, you may not discuss the exam with any other person before the graded exams are returned (in case there are students who have not yet taken the exam). Even casual statements, such as “it was easy” or “it was hard” are not permitted. Exams must be completed within the allotted time and only approved materials may be used.