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

Course Description

This sequel to DECS-430 (Business Analytics I) extends the statistical techniques learned in that course to 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. The course has an intense focus on managerially relevant applications, cases, and interpretations.

Motivation and Objectives

In a world that inundates us with data, most managers lack experience in analyzing data 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, 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:

- 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 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 data scientists as part of a team

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 basic elements of regression analysis. Even during these lectures, we will use real-world data to address a variety of management problems. In the second part of the course, we will explore more advanced topics in regression analysis. We will work on increasingly complex data analyses, including cases, providing students with an opportunity to develop data skills. The cases in the second part of the course will help students develop and refine both modeling and presentation skills.

Course Outline

Assignments are due by 1:00pm on the day they are listed as due. They will be made available on the course Canvas website and (except for Assignment #0) solutions must be submitted as .pdfs through that site.

<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Topic</th>
<th>Readings</th>
<th>Assignment Due</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4/7</td>
<td>The Linear Regression Model</td>
<td>Assignment #0 (on Canvas) Text: Chapter 3.1-3.4</td>
<td>Prepare Assignment #0 for discussion (Stata installed with custom menu; printed output; prepared to discuss Autoparts mini-case)</td>
</tr>
<tr>
<td></td>
<td>116</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>4/14</td>
<td>Inference</td>
<td>Text: Review Chapters 1 and 2.1-2.3 and read Chapter 3.5 Read after class: Handout -- Making Inferences Using Someone Else’s Data</td>
<td>Assignment #1 (Individual)</td>
</tr>
<tr>
<td></td>
<td>116</td>
<td>Prediction Intervals/ Regression Performance</td>
<td>Text: Chapter 4 Autoparts (B) (on Canvas) KTN: Regression Basics pp 6-10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>116</td>
<td>Dummy Variables (1/2)</td>
<td>Text: Chapter 5.1 KTN: Regression Basics pp 1-5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>116</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4/21</td>
<td>Dummy Variables (2/2)</td>
<td>Text: Chapter 5.2-5.3</td>
<td>Assignment #2 (Individual)</td>
</tr>
</tbody>
</table>

1 The text is Managerial Statistics: A Case-Based Approach (Stata Edition) by Klibanoff, Sandroni, Moselle and Saraniti, and is included in your course pack along with some HBS cases. KTN refers to Kellogg Technical Notes, and these will be available through the Canvas site. Additional readings may also be assigned and posted to Canvas as warranted.
<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>5/5</td>
<td>ISP case Building models</td>
<td>ISP case background (on Canvas) KTN: Building Models Text: Chapter 6.1</td>
</tr>
<tr>
<td>5/12</td>
<td>Midterm Exam (1:30h in-class midterm) Fixed Effects Nonlinear specifications (1/2) Refer to #1</td>
<td>KTN: Fixed Effects Text: Chapter 8.1-8.4 KTN: Log vs. Linear</td>
</tr>
<tr>
<td>5/19</td>
<td>Nonlinear specifications (continued) Heteroskedasticity Independence Types of Data/RCT</td>
<td>Videogames Report (on Canvas) Text Chapter 8.5, 8.7 KTN: Noise, Heteroskedasticity, and Grouped Data</td>
</tr>
<tr>
<td>5/19</td>
<td>Presentations on Assignment #5 Presentation tips Benchmarking Final Review</td>
<td>Assignment #5 (Group)</td>
</tr>
</tbody>
</table>

2 The material on the “Box-Cox” test can be safely ignored for this course.
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 (http://mycourses.northwestern.edu).

Please note that there is an initial assignment to be prepared for class discussion for the very first class of the course (Assignment #0 on Canvas!)

The weekly assignments in the second part of the term will be group assignments. The class will be divided into groups of 3-5, and you should do the group homework assignments with your group (i.e., all group members work together on the assignment, one submission per group, submitted by one member of the group with all the group members’ names on it).

There will be an in-class midterm exam in Week 5. The final exam will be take-home and will be three hours long. Exams are individual, open-book, open-notes, and timed; computer use will be required.

Your course grade will be based on the following components:

- Homework 20%
- Midterm Exam 20%
- Final Exam 50%
- Class Participation and Attendance 5%
- Peer Evaluation 5%

Participation and homework 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. 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.

Want to boost your participation grade? Send me newspaper or magazine articles on business topics that closely relate to a topic we have studied in the course. Bonus points for international settings or a relationship to “Big Data”.

Course Material and Software

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.\(^3\) The course package also includes some HBS cases we will be using. All other course materials may be found on the Canvas course page which you can access through the Northwestern Canvas homepage (http://mycourses.northwestern.edu). You are responsible for regularly checking, downloading, and reading materials posted on the course site, as they form an integral part of the class.

**Part of preparing for our first class is completing Assignment 0 as posted in the assignments section of the course Canvas site and as listed in the syllabus and calendar on Canvas.**

**Excel and Stata**

In this class, we will use Excel for creating graphs and cleaning small data sets. However, we will learn that it has limitations for data cleaning. It has trouble handling larger data sets and does not record any of the changes that we make to the data.

More importantly, it permits you to perform only a limited set of analyses, which will make it inadequate for most of the cases we will encounter. For example, Excel does not have a simple way to perform many predictions and hypothesis tests and does not allow for many of the tests required to assess model validity.

Instead, we will use Stata, which is a commercial statistical analysis program. 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 by 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.

We use Stata in this class because it is both easy to use and powerful. If you choose to specialize in data analysis, you may encounter other analytical software programs along the way. Do not fear! Once you become comfortable with Stata, it will be easy for you to transition to a new program. The biggest challenge in data analysis is not learning how to use a specific program. Instead, it is figuring out what analyses are most appropriate to perform and effectively interpreting the results. This is also the highest value-add for your career.

Installing the latest version of Stata (currently version 14.2) and the Kellogg custom menu is part of Assignment #0. 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 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:

---

\(^3\) Do not buy the older version of this book, as it does not use the software we will be using – buy the course package!
There are a very large number of Q&A forums for Stata online, so try Googling your question first!

- There is a guide to frequently used commands posted to the course Canvas site
- There is a Stata mini-manual in the Appendix to the textbook in your course packet
- The Kellogg student Stata support page (http://www.kellogg.northwestern.edu/rs/software/stata/statamba.aspx) is a useful resource.
- If you experience software glitches such as Stata crashing or not running at all or the custom menu not installing, then turn to the KIS student support team

Excel does one thing at least as well as Stata: graphics. 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.

**Office Hours, and Contact Information**

My office hours are 12:30pm-1:15pm (Room TBA). If you would like to see me and cannot visit during office hours, then write me an e-mail (e-shmaya@kellogg.northwestern.edu) to make an appointment.

**Discussion Board**

If you have any questions about the course material or the assignments that are not personal in nature, please post them using the class Discussion Board on Canvas. I will be checking the board regularly to answer the questions. Writing an email will not speed up the process, as I will ask you to post any non-personal question on the Discussion Board. This is to assure that all students receive the same help with the class. Also, this will save you time: any time you have a question, by checking the recent posts on the board, you may find the answers quickly! You are also encouraged to answer questions posted by other students; this is another form of class participation. All students should both subscribe to the “General DECS-431 (Business Analytics II) questions” discussion topic and adjust their Canvas notification settings to ensure that they get ASAP email notification with the content of new discussion posts and replies.

**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. 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.

SEATING: Please also bring your nameplate to class and display it throughout the term.

**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:

ASSIGNMENTS: 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.

Regarding individual assignments, while discussing the general ideas behind the assignment’s questions with your colleagues is permitted and encouraged, formulating and writing solutions should be completely individual. This leaves some room for ambiguity – you should practice common sense.

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.