Business Analytics II (DECS 431)
Alireza Tahbaz-Salehi

E-mail: alirezat@kellogg.northwestern.edu
Office: Kellogg Global Hub 4133
Quarter: Fall 2019
Lectures: Saturdays, 1:30pm–6:15pm
Office hours: Saturdays 12:00pm–1:00pm

Course Description
This sequel to DECS 430-5 (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 or more variables
- teach you how to build a regression model, interpret regression coefficients, and use the results to test hypotheses and make predictions
- expose you to advanced techniques in model specification to assure that your analyses are free from unwanted bias and other problems
- introduce the concept of an “experiment” and demonstrate how to distinguish correlation from causation using a variety of practical approaches
- enable you to assess data-driven analyses performed by others, and to perform convincing data-driven analyses of your own
- 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 more 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:

<table>
<thead>
<tr>
<th>Session</th>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10/19 (@3:30pm)</td>
<td>The linear regression model</td>
</tr>
</tbody>
</table>
| 2       | 10/26      | Inference  
Prediction intervals/regression performance  
Dummy variables (1/2) |
| 3       | 11/2       | Dummy Variables (2/2)  
Omitted variable bias |
| 4       | 11/9       | ISP case  
Building models  
Spurious correlation  
Multicollinearity  
Midterm review |
| 5       | 11/16      | Midterm exam (in-class exam)  
Fixed effects  
Nonlinear specifications (1/2) |
| 6       | 11/23      | Nonlinear specifications (2/2)  
Heteroskedasticity  
Independence  
Experiments/RCTs |
| 7       | 12/07      | Big Data/ML  
Final Review |
Assignments
Homework assignments are posted on the course webpage, accessible through Canvas. Assignments are due by 1:30pm on the day they are listed as due. Solutions must be submitted as .pdfs through that site.

The weekly assignments are group assignments. The class will be divided into groups of 4–5, and you should do the 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).

Exams and Grading
There will be an in-class midterm exam on November 16th. The final exam will take place on December 14th during regular class time 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:

- Homeworks: 35%
- Midterm Exam: 25%
- Final Exam: 35%
- Class Participation/Attendance: 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.

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. The course package also includes some HBS cases we will be using. All other course materials may be found on Canvas. You are responsible for regularly checking, downloading, and reading materials posted on the course site, as they form an integral part of the class.

1Do not buy the older version of this book, as it does not use the software we will be using — buy the course package!
**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-added for your career.

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:

- There are a 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 is a useful resource. ([http://www.kellogg.northwestern.edu/rs/software/stata/statamba.aspx](http://www.kellogg.northwestern.edu/rs/software/stata/statamba.aspx))
- 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.

**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.

**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).

**Support for Accommodations**

Any student requesting accommodations related to a disability or other condition is required to register with AccessibleNU (accessiblenu@northwestern.edu; 847-467-5530) and provide professors with an accommodation notification from AccessibleNU, preferably within the first two weeks of class. All information will remain confidential.

Students can find useful resources for safety and security, academic support, and mental and physical health and well-being at the NU help website and app.