DECS-431  
Business Analytics II  
Syllabus for Fall 2020  
Prof. Donald Dale  
donald.dale@kellogg.northwestern.edu  

Preliminary Version on August 7, 2020  

**Class Modality:**  
*Hybrid (Zoom + In-Person) Weeks 1-9  
Virtual (Zoom) Week 10*  

*Note that given social distancing and masking requirements, it is likely that due to classroom capacity there will be some form of rotational in-person attendance system, so that any individual student will only be able to attend in-person for at most half of the class sessions. Attendance will be via Zoom for those not in-person at any given session.*

**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 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**

This is the first time Kellogg is using the Hybrid modality, so the course outline will be adapted during the course to ensure the best class experience!

Assignments are due by 10:00am on the day they are listed as due. They will be made available on the course Canvas website and solutions must be submitted as .pdfs through that site (except for Assignment #0).

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
<th>Readings</th>
<th>Assignment Due</th>
</tr>
</thead>
</table>
| 1    | 9/22  | The Linear Regression Model | Prepare for class discussion: Assignment #0 (on Canvas)  
Read before or after class: Text: Chapter 3.1-3.4 | Prepare Assignment #0 for discussion  
(Install Stata with custom menu, print output, and prepare to discuss Autoparts mini-case) |
| 2    | 9/25  | Linear Regression Continued |  
| 3    | 9/29  | Inference | Read before or after class:  
- Text: Review Chapters 1 and 2.1-2.3 and read Chapter 3.5 | Assignment #1 (Individual) |
| 4    | 10/2  | Omitted Variable Bias | Prepare for class discussion:  
- Case: Insert 1 (Energy Costs and Refrigerator Pricing) p. 193-194  
Read before or after class:  
- Text: Chapter 7.5  
- KTN: Omitted Variables | |

1 The text is *Managerial Statistics: A Case-Based Approach (Stata Edition)* by Klibanoff, Sandroni, Moselle and Saraniti, and is included in your course packet 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>Week</th>
<th>Date</th>
<th>Topic</th>
<th>Prepare for class discussion:</th>
<th>Assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>10/6</td>
<td>Prediction Intervals/ Regression Performance</td>
<td>Prepare for class discussion:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Case: Autoparts (B) (on Canvas)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Read before or after class:</td>
<td>Assignment #2 (Individual)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Text: Chapter 4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- KTN: Regression Basics p. 6-10</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>10/9</td>
<td>Dummy Variables</td>
<td>Read before or after class:</td>
<td>Assignment #3 (Individual)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Text: Chapter 5.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- KTN: Regression Basics p.1-5</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>10/13</td>
<td>Slope Dummy Variables</td>
<td>Read before or after class:</td>
<td>Assignment #4 (Individual)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Text: Chapter 5.2-5.3</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>10/16</td>
<td>Slope Dummy Variables Continued/ Choosing Variables for a Regression</td>
<td>Prepare for class discussion:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Case: Nopane (in course packet)</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>10/20</td>
<td>Choosing Variables for a Regression Continued/ Midterm Review</td>
<td>Read before or after class:</td>
<td>Assignment #5 (Group)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Text: Chapter 6.1</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>10/23</td>
<td>Midterm Exam</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>10/27</td>
<td>Building Models</td>
<td>Read before or after class:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Case: ISP (on Canvas)- KTN: Building Models</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>10/30</td>
<td>Building Models Continued</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>11/3</td>
<td>Spurious Correlation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>11/6</td>
<td>Presentations on Assignment #5 Nonlinear Specifications</td>
<td>Read before or after class:</td>
<td>Assignment #5 (Group)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Text: Chapter 8.1-8.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- KTN: Log vs. Linear2</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>11/10</td>
<td>Nonlinear Continued</td>
<td>Read before or after class:</td>
<td>Assignment #6 (Group)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Case: Videogames (on Canvas)</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>11/13</td>
<td>Fixed Effects</td>
<td>Read before or after class:</td>
<td>Assignment #6 (Group)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- KTN: Fixed Effects</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>11/17</td>
<td>Heteroskedasticity Independence</td>
<td>Read before or after class:</td>
<td>Assignment #7 (Individual)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Text Chapter 8.5, 8.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- KTN: Noise, Heteroskedasticity, and Grouped Data</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>11/20</td>
<td>Experiments/ RCTs</td>
<td></td>
<td>Assignment #7 (Individual)</td>
</tr>
<tr>
<td>19</td>
<td>12/1</td>
<td>Experiments Continued Big Data/ML</td>
<td></td>
<td>Assignment #7 (Individual)</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 (Assignment #0 on Canvas) to be prepared for class discussion for the very first class of the course! The first graded assignment (Assignment #1; to be done individually) will be submitted in week 2 (by the deadline listed) in .pdf format through the Assignments section of the course Canvas site.

Some of the weekly assignments later in the term will be group assignments. The class will be divided into groups (assigned by me), 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 a virtual Midterm Exam in the second class of Week 5. The Final Exam will be virtual 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:

- Individual Homework 13% (Assignments #1-4, 2.5% each; #7, 3%)
- Group Homework/Cases 20% (Assignments #5-6, 6% each; #8, 8%)
- Midterm Exam 22%
- Final Exam 35%
- Class Participation and Attendance 5%
- Peer Evaluation 5%

Participation and homework is vital to the learning experience, even on Zoom! 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.
If you expect any personal circumstance to limit your ability to participate during class when on Zoom (e.g., due to weak internet connection, loud environment), please let me know by sending me an email.

**Course Material and Software**

The course packet 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 packet 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.

**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 unlike other analytical programs every student can start using it from the very first class with minimal supervision (it is truly “Day 1 Ready”) and it can be used to run the most advanced analyses we conduct (we will even use it to run machine learning algorithms).

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. The biggest challenge 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.

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!
Installing the latest version of Stata and the Kellogg custom menu is part of Assignment #0. The software is provided at no cost to you through 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 many 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 of the textbook in the course packet
- The Kellogg 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.

**Office Hours, Contact Information and Weekly Review Sessions**

**My office hours will be on Zoom on TBD from TBD.** If you would like to see me and cannot visit during office hours, then send me an email to make an appointment. E-mail is the fastest and most reliable way to reach me outside of class. Questions about the material or about general course administrative issues should be posted in the course’s discussion board (see below).

The class TA is Roman Acosta: roman.acosta@kellogg.northwestern.edu. Feel free to reach out to him with any question too.

There are **review sessions** on Zoom each Wednesday, starting from the first week of class at TBD and again at TBD run by the TA (links available on Canvas). These sessions are an excellent resource and are often highly recommended by past students.

There are two other times that you can meet with a TA. First, a TA will be holding weekly course-wide office hours intended for those students who may be having an especially difficult time with the material. These are TBD, starting the first week of classes, from TBD on Zoom (links available on Canvas).

Second, for those students with a solid grasp of the material desiring further/deeper exploration than the constraints of the class allow and/or exposure to more advanced topics and techniques, a TA will be holding weekly course-wide office hours on TBD, starting the first week of classes, from TBD on Zoom (links available on Canvas).

**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

ELECTRONICS: You will need to use your laptop in class to participate on Zoom and complete exercises in Stata. However, no other use of your laptop (e.g., tracking your portfolio, web surfing, instant messaging, emailing, tweeting, etc.) during class is allowed. No other electronic devices should be used during class (with the exception of tablets for notes or any medical devices). All ringers should be switched off to minimize distractions to you and others. If you are attending in person, you will need to bring headphones with a microphone so that you can converse with students who are connecting remotely through Zoom.

ATTENDANCE AND PUNCTUALITY: You are required to attend class and be on time. To help encourage this, any student who misses or is late to more than two classes will lose some participation credit. There are exceptions in accordance with Kellogg policy for religious holidays, funeral attendance and student/dependent hospitalization/severe illness.

Kellogg Honor Code


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. **In other words, you should not discuss the solutions to the problems with other students before the assignment is due.**

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.

**Accommodations and Resources**

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 on the NU help website and app.

**Class Recordings**

The class and TA review sessions will be recorded by the instructor each week. These recordings will be shared only with students enrolled in the course and will be deleted at the end of the course. Your instructor will communicate how you can access the recordings. Unauthorized student recording of classroom or other academic activities (including advising sessions or office hours) is prohibited.