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

Course Description

This sequel to DECS-430 extends the statistical techniques learned in that course to allow for the exploration of relationships between variables, primarily through multivariate regression. In addition to learning basic regression skills, including modeling and estimation, students will deepen their 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 real world 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, managerial intuition and understanding of the business issues at stake. 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:

- 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
**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 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. The final weeks of the course emphasize two rich business cases that rely on large-scale semi-structured data. These cases will help students refine modeling and presentation skills.

**Course Material and Software**

The course package contains some required reading material for the course. The textbook manuscript: “Managerial Statistics: A Case-Based Approach: Stata Edition” by Klibanoff, Sandroni, Moselle and Saraniti is included in the course package. 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. In exchange for getting the textbook in manuscript form, you are saving a substantial amount of money compared to having to purchase a bound book. **Do not buy the older version of this book, as it does not use the software we will be using – buy the course package!** The course package also includes some HBS cases we will be using. The case packet can be obtained at Study.Net. All other course materials may be found at the canvas course page (http://www.it.northwestern.edu/education/learning-management/login.html). There you will find Handouts, Assignments, Data Sets, The HBS Cases, answers to selected questions from the manuscript, and a Stata Command Guide (with the most useful Stata Commands). You are responsible for regularly checking, downloading, and reading materials posted on the site, as they form an integral part of the class. In addition, **you should install a copy of the latest version of Stata** provided at no cost to you through Kellogg. Go to the KIS Stata page (https://kis.kellogg.northwestern.edu/Pages/StataStatisticalPackage.aspx) for instructions on installing the software. The license you get for Stata is permanent, so that you may continue to use the installed version even after you have graduated from Kellogg. Once you have installed Stata, you should go to the Kellogg student Stata support page (http://kellogg.northwestern.edu/researchcomputing/sec/stata-support.htm) and install the Kellogg custom core statistics menu (doing this is also part of preparing for our first class and is described in the Autoparts (A) reading for that class on Canvas as well). There are no other required materials for this course.

**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, however, for several reasons. 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.

Fortunately, there are many sophisticated data analytic software programs. In Kellogg classrooms we rely on Stata. 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.)

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 (http://kellogg.northwestern.edu/researchcomputing/sec/stata-support.htm) is a useful resource.

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 one hour before class starts. If you cannot visit during office hours, then write me an e-mail (sandroni@kellogg.northwestern.edu) to make an appointment. E-mail is the fastest and most reliable way to reach me outside of class. My office phone is 1-3729.

Questions about the material or administrative issues that are relevant to others should be posted in the course’s discussion board.

The class TAs are Sanket Ramakant Patil <sanket.patil@kellogg.northwestern.edu> and Krzysztof Kalisiak <k.kalisiak@gmail.com>.
Discussion Board

If you have any questions about the course material or the assignments that are not personal in nature, post them using the class Discussion Board on Blackboard. Writing an email will not speed up the process, as I will ask you to post the 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 are likely to find the answers right away! You are also encouraged to answer questions posted by other students; this is another form of class participation. **All students should subscribe to the discussion board by going to the course Canvas site.** Just click through to the discussion board and then click the forum titled “General Questions.” Near the top of the page a button labeled “subscribe” appears, click it to subscribe. When a message is posted to the discussion forum you should automatically receive an email alert with the content of the post. So, this will ensure that you are notified by email of discussion board posts.

Course Outline

Assignments are due **by 6:00pm 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.

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
<th>Readings¹</th>
<th>Assignment Due</th>
</tr>
</thead>
</table>
| 1/07  | Review of BA I and Introduction to Regression | Assignment #0 (on Canvas)  
Text: Chapter 1 and 2 (for review)  
Chapter 3 (for regression) | Prepare Assignment #0.  
Due Day 1/07 |
| 1/14  | Inference and Prediction Intervals         | Text: Chapter 4  
Autoparts (B) | Assignment #1 (Individual)  
Due day 1/14 |
| 1/21  | Regression Performance Dummy Variables     | Text: Chapter 5 | Assignment #2 (Individual)  
Due day 1/21 |
| 1/28  | Dummy Variables (Continued) and Omitted Variable Bias | Text: Case Insert 1 (Energy Costs and Refrigerator Pricing):  
Chapter 7.5 | Assignment #3 (Individual)  
Due day 1/28 |
| 2/04  | Spurious Correlation; Midterm Review       |                                              | Assignment #4 (individual)  
Due day 2/04 |
| 2/11  | Midterm Exam                               |                                              | Assignment #5 (Group) |
| 2/18  |                                             | Text: Chapter 6.1                           |                                |

¹ The text is *Managerial Statistics: A Case-Based Approach (Stata Edition)* by Klibanoff, Sandroni, Moselle and Saraniti, and is included in your coursepack along with some HBS cases.
Grading and Assignments

As can be seen from the course outline above, there are homework assignments most classes which will be posted on the course webpage, accessible through the Canvas site at Northwestern (https://courses.northwestern.edu/webapps/login).

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)! Additionally, the first assignment to be handed in (Assignment #1; to be done individually) is due (submitted in .pdf format only through the Assignments section of the course Canvas site) by 6pm, Monday January 14.

Some of the assignments later in the term will be group assignments. The class will be divided into study 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). In several occasions, some groups will be selected randomly to present their work in front of the class. The groups not presenting will be asked to challenge the group who is presenting. Hence, we will have a class debate over these assignments. Feedback will be given to the written solutions, but a significant part of the grade will be based on the quality of the class presentations and on the quality of the challenge to the presenters.

There will be an in-class midterm exam in class 6. The final exam will take place according to Kellogg’s exam schedule 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 15%
- Group Homework/Cases 20%

<table>
<thead>
<tr>
<th>Date</th>
<th>Assignment</th>
<th>Text</th>
<th>Due date</th>
</tr>
</thead>
<tbody>
<tr>
<td>2/25</td>
<td>Logs in regression</td>
<td>Text: Chapter 8.1-8.4</td>
<td>Assignment #6 (Group) Due date 2/25</td>
</tr>
<tr>
<td>3/4</td>
<td>Fixed Effects Heteroskedasticity</td>
<td>Text Chapter 8.5,8.7</td>
<td>Assignment #7 (Group) Due date 3/4</td>
</tr>
<tr>
<td>3/11</td>
<td>Case Studies</td>
<td></td>
<td>Assignment #8 (Group) Due date 3/11</td>
</tr>
<tr>
<td></td>
<td>Final Exam (as per Kellogg Exam Schedule)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Midterm Exam
Midterm Exam 20%

### Final Exam
Final Exam 35%

### Class Participation and Attendance
Class Participation and Attendance 10%

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 or other software to perform class-related calculations and data analysis. However, no other use of your laptop (e.g., track your portfolio, web surf, instant messaging, e-mailing, tweeting, or blogging etc.) during class is allowed. No other electronic devices (iPads, tablets, mobile phones, Blackberries, etc.) may be used in class unless the instructor directs students to do so and 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.

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 after seats are assigned will lose some participation credit. There are exceptions in accordance with Kellogg policy for religious holidays, funeral attendance and student/dependent hospitalization/severe illness. Note, however, that weddings, interviews, delayed flights etc. are what the two "free passes" are meant to cover and will be counted as absences.

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
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.