This course is about the family of statistical data analysis tools called regression analysis. The course can be viewed as a logical successor to the core Managerial Statistics course, and is most frequently taken by 2nd-year MBA students wishing to solidify and extend their quantitative skills.

The purpose of a regression analysis is to build a mathematical/statistical representation of the relationships between variables that can be used for prediction of outcomes and enhanced understanding of causes. When regression is used in a business context the ultimate managerial goal is, of course, better business decisions.

Linear regression models are widely used in the business world, as well as in many other fields such as economics, engineering, social research and in the health and biological sciences. In the business world regressions have been used in marketing analyses of customer behavior, in financial analyses of investment opportunities, in human resources to test the fairness of employment policies, in operations to identify the determinants of output quality, and in strategic planning to create sales forecasts.

Contemporary computing hardware and modern statistical software has made it extraordinarily easy to produce regression models. For example, Microsoft Excel has a quite powerful regression tool that is very easy to use with absolutely no knowledge of the underlying concepts. Though it has become child’s play to “run a regression,” it is quite a challenge to create a regression model that is really useful and reliable. One might argue that regression is almost too easy to do – it being very easy to mechanically produce “bad” regressions. The goal of this course is to learn how to create good regressions, and how to judge if a good regression is even possible. The course is based on the premise that successful applications of regression require sound understanding of both the underlying theory and the practical problems that are encountered when building and using models of real-life situations with serious consequences. Therefore this course seeks to blend theory and applications effectively, thereby avoiding the extremes of presenting theory in isolation and of giving elements of applications without the foundation concepts that are needed for practical understanding.

The course will deal with three topics in an interwoven way. First and most basic, is an approach to data and data analysis that is based on statistical theory, the scientific method and on some very pragmatic epistemology. Second, is regression analysis itself, including extensions to the basic linear model such as logistic regression and some basic multivariate models. Third, is forecasting of time series from historical data. The title of our textbook is descriptive of our approach: regression by example. Each concept and procedure shall be explained, indeed, whenever possible introduced by an example. Moreover, we will emphasize examples in which the business context matters and may influence the analytic approach taken.

Computing

The course will be very computationally hands-on from the very first lecture. Your laptop computer will be used extensively for data analysis. Most of this work can be done in Excel and we assume a basic familiarity with the data analysis tools of Excel – including the regression tool. However, even though most of the work can be done in Excel, there can be some advantages to using a statistical software package. Moreover, several of the special and important analytic tools — stepwise regression and logistic regression — cannot be done in Excel. Therefore we will supplement Excel with the Minitab statistical analysis system. Minitab gives us professional statistical analysis capabilities while being easy
to learn and use. Any version of Minitab that can do regression, stepwise regression and logistic regression will be adequate. Students who already are familiar with another software package that has the aforementioned capabilities are welcome to use it. (SPSS, SAS, R, Stata)

**Conduct of the Course**

**Course Project:** A major part of the course work will be a data analysis project. The project will be a significant data analysis in a real business context. I will provide a standard project. However, I strongly suggest that students who have particular interests propose their own project. This can be a way of increasing the value you get out of the course. Projects will be due and presented in class in the last class session.

**Workload and Grading:** It is expected that students will attend class regularly and participate fully in class discussions. Since many of these discussions will be based on our analytic assignments (mini-cases), it is important that assigned work be done thoroughly and on time. Assignments can be done individually or in teams of two students. The final course grade will be composed of three components:

<table>
<thead>
<tr>
<th>Component</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attendance and class participation</td>
<td>1/3</td>
</tr>
<tr>
<td>Written Assignments</td>
<td>1/3</td>
</tr>
<tr>
<td>Term Project</td>
<td>1/3</td>
</tr>
</tbody>
</table>

**Textbooks and Software**

The course will follow the same general outline as the text by Chatterjee, Hadi and Price which is listed below. This book strikes a good balance between providing a theoretical understanding and keeping a very concrete focus on applications. I will, in general use different examples than those in the text. It will thereby offer a second view on most issues and procedures. I recommend purchasing it. In addition to Excel we will use the Minitab statistical package, the software for which comes with a helpful user’s manual.

**Textbook:**

Samprit Chatterjee and Ali Hadi  
*Regression Analysis by Example, 5th edition* (Wiley 2012)  

Note: The textbook is recommended, not required. It is entirely possible to do very well in this course without buying the textbook. However, it is a very good resource on this subject, and goes into greater depth than we will have time for in the class meetings.

**Computer Software:**

Minitab  