BUS 550 - Data and Decision Analytics

Course Instructor: Jeffrey L. Rummel

Associate Professor in the Practice
Information Systems & Operations Management

Email: jrummel@emory.edu (or JEFF.RUMMEL@EMORY.EDU)
Phone: (404) 712-0201
Website: HTTP://JRRUMMEL.COM

ISOM administrative assistant: Jalisa Norton, (404) 727-8698

BUS 550-00A: Tu-Th 8:00am to 9:30am Room 201
BUS 550-00B: Tu-Th 9:45am to 11:15am Room 201
BUS 550-00C: Tu-Th 1:00pm to 2:30pm Room 201
Joint session: August 22, 9:30am to 11:00am Room 130
Joint session: August 29, 12:30pm to 2:00pm Room 130
Joint session: October 3, 8:00am to 9:30am, Room 130

Office: GBS 424
Office Hours: Tu-Th 3:30pm to 6:00pm and by appointment
# Contents

1 Course Overview ........................................ 3

2 Overall Objectives ...................................... 5

3 Teaching Methods ......................................... 5

4 Course Materials ......................................... 6

5 Course Requirements, Grading and the Honor Code ........ 8

6 The Honor Code at Goizueta ........................... 9

7 Schedule Overview ....................................... 9

8 Populations, samples and descriptive statistics ............. 9
   8.1 Class sessions August 12 - August 19 ................ 10
   8.2 Project #1 ........................................... 10

9 Inference and regression modeling ........................ 10
   9.1 Class Session August 21 - September 4 ............... 10
   9.2 Project #2 ........................................... 11
   9.3 Project #3 ........................................... 12
   9.4 Mid-term Exam ....................................... 12

10 Probability and probability distributions .................. 12
   10.1 Class Sessions September 15 - September 30 ....... 12
   10.2 Project #4 ........................................... 13

11 Decision Analysis ....................................... 13
   11.1 Class Sessions October 2 - October 16 ............... 13
   11.2 Project #5 ........................................... 13
   11.3 Final Exam .......................................... 13
1 Course Overview

Seldom affirm, never deny,  
Always distinguish.  
— St. Thomas Aquinas

Business leaders have access to more and more data, and are expected to use it to understand the context for decisions, to understand how systems work, and to assess the risks associated with different courses of action. The advice from St. Thomas Aquinas, a thirteenth century Italian philosopher, about how to answer questions (and make decisions) will serve as one basis for the class. Rarely are business problems so simple that they can be answered unequivocally, with a yes or a no. Rather, we must carefully assess all the complications and see the subtle differences between seemingly similar things.

The overall objective of this course is to improve your ability as a decision maker to use quantitative tools. You will learn to analyze data to support decisions, and how to incorporate uncertainty into decision problems in a more structured way.

Many of you have heard the maxim “there are lies, damned lies and statistics” and as managers, you will be flooded with numbers that claim to support particular positions about decisions to be made in your job. Not all data are well defined or are directly applicable to the decision at hand. Elaborate methods can be employed to analyze data, but the results are dependent on the limitations of the tools. Without a clear understanding of how to evaluate data with appropriate tools, you can make decisions that will cost your firm real dollars and put your company at a disadvantage relative to competitors.

We will discuss how to structure decision problems for analysis. Effective decision making has many aspects, but without clear problem definitions, faulty analytic reasoning is much more likely. Especially when there are risks and tradeoffs, problem structure provide decision makers with a way to test assumptions and to understand the sensitivity of outcomes to various aspects – to know where the real value levers are.

Finally, since the relationships between decisions and outcomes are uncertain, managers must understand the role of model building and inference. With careful use of statistical tools, we can discover the relationships between data and increase our understanding of how systems operate. But there are many pitfalls in building these models, and good decision makers are aware of these limitations in inferential statistics. Sometimes, a model seems to determine causal relationships we are tempted to create a theory that says “if I
change something, I know what will happen." Statistical models are able to see relationships between data, but correlation does not lead directly to causality.

The course will be a survey of various techniques in probability and statistics rather than focusing in depth on one or two topics. We will discuss some details of the techniques, but the emphasis will be on the intuition underlying various statistical tools. In order to be precise, you will see some of the mathematics involved and look at spreadsheet techniques to carry out the mathematics. If you have not used spreadsheets, or feel that your abilities are weak, please contact me and plan to spend some time early in the class familiarizing yourself with spreadsheet modeling.

We will also discuss how various techniques can be misused. Every day in the news there seems to be statistical support for the opposing views (for example, certain economic policies or global warming). Where does this discrepancy come from? Incorrect inferences can arise from correctly using statistical tools on "biased" data, or from the inappropriate application of particular tools. As a manager using data to make decisions, you need to be able to recognize these situations and use the analysis appropriately.

This is not a course in mathematics, although mathematics will be used in the course as the language for formally defining models and techniques. The emphasis in this course is on the structure and logic of the tools and models, not on their mathematical details and proofs. The requirements for particular mathematical operations should be within your capabilities, the most important of which is the ability to deal with abstract symbols and relationships (algebra). If you feel unsure of your ability to deal with the mathematics during a particular part of the course, you should let me know as quickly as possible.
2 Overall Objectives

When the course is finished, you should be able to:

- Explain the role of data and models in the decision making process and how they support (rather than determine) decisions.
- Explain the difficulties of collecting data and why the effectiveness of decision making is dependent on the quality of the data.
- Decide what tools are appropriate in different decision making settings.
- Evaluate the quality of models created with statistical tools.
- Decide how to structure problems so that the essential elements are included and that the model can be analyzed in a timely fashion.
- Execute the decision analysis process for an unstructured problem.
- Translate the results of the model into a statement of the actions to be taken.

3 Teaching Methods

Some topics in decision and information analysis will be new for members of the class, and I will provide lecture material to supplement the course readings and case exercises. My experience is that a course like this suffers if I am the only one taking an active role, and so I will encourage comments and questions in all class meetings.

You need to feel comfortable taking positions about what we are discussing, even if that position turns out to be wrong. As managers, a critical skill is detecting flaws in an analysis given to you, and then working to make appropriate corrections. So please try out ideas even if you are not entirely sure of them. Part of our job as a class is to (gently) help people who are new to this material to become more confident. You should be aware that on occasion I may put forward arguments that are flawed, so you should critically (but kindly) evaluate all comments during the class.
I will provide all of the assignments, material and other details for the class using the learning management system supported by Emory, BLACKBOARD. I may communicate via email from time to time, but that communication will always be available on BLACKBOARD.

I will hold regular office hours on Tuesday and Thursday afternoons. You should feel free to just drop by, or you can contact me via email and arrange for a specific time. I will also have some availability on other days, but that time is somewhat limited. Contact me with a suggested time or two and I will see if those are possible.

The Program Office has also arranged for regular review sessions during the semester. I will not prepare new material for these sessions, but will rather take questions from those attending. There will be some practice problems available on BLACKBOARD, and those are something that can be discussed in the review sessions. Here are the review sessions times (all of them are in Room 130) - the times may change so I will keep the schedule current on BLACKBOARD:

- August 14 at 4:30pm
- August 20 at 4:30pm
- August 27 at 5:00pm
- September 3 at 4:30pm
- September 16 at 4:30pm
- September 24 at 4:30
- October 1 at 4:30pm
- October 8 at 4:30pm
- October 14 at 4:30pm
- October 21 at 1:30pm

There will also be a group of teaching assistants for the class. They will be MBA students who know the material pretty well, but I am not expecting them to be completely expert on all things, and so if they cannot answer all your questions, please come to see me. They will hold regular office hours, and I will announce times and locations on BLACKBOARD after they return from their summer internships.

4 Course Materials

There are many possible textbooks for this class, but no single book covers all the material for the class in a way that supports the work we will be doing. We will use:

**Business Statistics: A First Course** by Levine, Krehbiel and Berenson. This text covers the typical topics in a class about data analysis. We will not cover all of the chapters, but
I think you will find this to be a useful reference book. I will not refer specifically to sections of the book – rather I will consider the book to provide another explanation of the topics in my lecture notes.

Other material for the class will be placed on Blackboard. Cases that have copyrights that you must pay for will be distributed through Study.net.

The course will make heavy use of Microsoft Excel. There will be Excel “toolkit workshops” during the fall and they are excellent ways to improve your skills. In addition, there are a number of tutorials available on the Microsoft web site and a number of third-party training resources. I have some helpful material on my website, http://jrummel.com.

Unfortunately, using Excel has become a complicated issue, with different organizations using Office 2013, Office 2010, Office 2007, or even Excel 2003. Mac users may be running one of a number of versions of Excel, all of which have some compatibility problems with the Windows versions. You are free to use any of these, but I will focus on the version presently supported in the classrooms at GBS: the Windows version that is part of Office 2010. If you wish to use another version, you will need to be responsible for finding how to accomplish tasks that I demonstrate using Excel. Please note that for the program-supplied tablet computers, only Office 2013 was available. I will discuss the use of Excel on the tablets during the first residency.

There are many techniques that we will explore in the class, and so I have arranged for some additional software that will be distributed. The first is a package called Decision Tools. This has been the standard Excel add-in used at Goizueta, but some of you may be familiar with a competitive product called Crystal Ball. The two tools are quite similar, with certain advantages and disadvantages for each.

Finally, there are a number of statistical software packages that implement the techniques we will use. The list is long and includes SAS, SPSS, Minitab, and R. For this class, we will use the SAS Institute product called JMP. It integrates very nicely with Excel, but provides many more tools for data analysis and model building. Excel alone can be used for all of the cases and discussions in class, but I will also demonstrate some of the power JMP.
5 Course Requirements, Grading and the Honor Code

Grades for the course will be based on the following:

**Class participation (30%).** I will assess your participation during most class sessions (the joint sessions are difficult to assess). Higher evaluations will go to those who contribute *constructively* to our discussions, not just for how much you contribute. This can be in the form of providing clear insights for the rest of the class (you got the answer), but also in posing questions that indicate an attempt to more clearly understand the decision at hand. These “I don’t understand” contributions should always include a description of why you have the question and what the answer to the question will provide. I will provide regular feedback about this part of the class, but feel free to contact me if you have concerns or questions. To help me with the assessment, bring your name plate to every class and use the seating arrangement that I will publish.

**Team case assignments (30%).** For a number of the case studies, I will have teams prepare and submit one document that addresses the decision problem of the case. These will not need to be very long (5-7 pages is typically enough, plus attachments). For the early cases, I will provide questions to guide your work, but in later cases you will need to answer not only the questions that I provide, but the others that arise during your analysis.

**Midterm exam (20%).** The midterm will be given on

```
Monday September 8, 8:00am to 11:00am
```

in rooms arranged by the program office. The exam will cover all the material from the first module.

**Final examination (20%).** The final exam will be given on

```
Wednesday October 22, 1:00pm to 4:00pm
```

in rooms arranged by the program office. The exam will emphasize the material from the second module, but it will be comprehensive.
Please note that a passing level of performance on each of these components is required in order to receive a passing grade for the class. Also recognize that the MBA program at Goizueta has an official grading policy that limits the number of the highest grades. There will be sufficient work to allow me to implement this policy and I will clearly communicate during the class where you are located in the grade distribution. If you have any questions or concerns about this policy, please feel free to contact me at any time during the class.

6 The Honor Code at Goizueta

Academic misconduct of any sort will not be tolerated and everyone should understand that I take this issue quite seriously. I will pursue violations of the honor code according to University policies, and all students in this class must comply with the Goizueta Honor Code. Please make time to talk with me or someone in the MBA office if you are unsure about these policies. Also understand that a student who knowingly assists another student in committing an act of academic misconduct will be equally accountable for the violation and will be subject to sanctions.

7 Schedule Overview

The course will break into four natural sections. The first will introduce (or reintroduce) the basic language of descriptive statistics. The second extends these tools to inference and model building. Probability theory and distributions will be covered in the third section. The last section will be an introduction to decision analysis and decision trees.

I will work with the class during the term and adjust dates if it becomes necessary. That will always mean pushing dates further into the future, never making something due sooner than shown below.

8 Populations, samples and descriptive statistics
8.1 Class sessions August 12 - August 19

**Basic definition of statistics.** We will discuss the terms used in statistical analysis, and why statistics is an important part of the BBA curriculum.

**Summary measures of a data set.** There are choices about how to represent data sets and reasons for using different summary statistics.

**Using graphs to summarize a data set.** Graphs are often useful to paint a “picture” of the data, and there are ways to construct and interpret graphs in different ways.

**Pitfalls and paradoxes in descriptive statistics.** Unfortunately, graphs and statistics can be used to mislead as well as to enlighten, and we will discuss some of these issues.

**Collecting samples from populations.** Sometimes we have access to all of the data, but in business it is more common that we only have a sample of the data. We will discuss how to be able to use samples effectively.

**Collecting survey data.** Asking for opinions is a common way businesses learn from their customers, and there are aspects of surveys that require some care in order for the data to be useful.

**Readings** The most important chapters are 1, 2 and 3. We will also talk about chapter 7.

8.2 Project #1

The first case will be due on August 19.

9 Inference and regression modeling

9.1 Class Session August 21 - September 4

**Assumptions and derivation of regression modeling.** Model building relates one set of variables to another variable that is related. This allows us to make judgments and forecasts.
Summary statistics for a regression model. The quality of a regression model depends on certain key statistics that are calculated by regression software.

Introduction to multiple regression. When there are many variables that are used to predict another variability, some complication arise that need to be addressed.

Building models from data collections. We will look at the process of starting with a data set and refining it to produce a useful model. Not all data will be helpful so we will look at methods to create models that use the “right” amounts of data. Regression tools also can give indications of missing data.

Confidence interval construction and interpretation. When we make statements from collected data, there is the need to characterize the accuracy (or uncertainty) of those statements.

Simple hypothesis testing. We also need to make statements about whether there are differences or preferences, and with statistical data, there is always the possibility of making wrong judgments. Hypothesis testing allows us to characterize our confidence in those statements.

Analysis of variance (ANOVA), and other types of statistical testing. More complex statements require more complex statistical tool, and we will look at a few more examples of using data to make decisions (as, for example, the FDA does for new drug applications).

Additional regression topics. Regression makes some assumptions about the variables and we can relax them to allow us to incorporate non-linear relationships or categorical data.

Readings Chapters 12 and 13 are the most important. We will cover some material from chapters 8, 9 and 10.

9.2 Project #2

The second case will be due on August 26.
9.3 Project #3

The third case will be due on September 2.

9.4 Mid-term Exam

The midterm exam will take place on Monday September 8, 8:00am to 11:00am. There will be multiple rooms for the exam and I will assign you to a particular room. If you have any special requests for the exam, please contact me via email prior to September 2.

10 Probability and probability distributions

10.1 Class Sessions September 15 - September 30

Basic concepts of probability. There are different ways to conceptualize probability and calculate values for events of interest. We will explore mostly simple cases that involve enumeration.

Discrete probability distribution functions. For many business applications, direct enumeration is difficult, and so we will use probability distributions that convert a situation into the corresponding probability. Some distributions deal with discrete quantities, like counting processes.

Continuous probability distribution functions. Other problems require measurements, and these are often continuous values, and so the nature of probability distribution functions becomes slightly different.

Sampling distributions. Special kinds of distributions relate samples and populations and we will look at those distributions.

Modeling probabilities with spreadsheets. All of the probability calculations will be done using spreadsheet formulas so that you will be able to utilize probabilities in spreadsheet models that we will develop later.

Readings We will cover chapters 4, 5, 6 and 7.
10.2 Project #4

The fourth case will be due on September 30.

11 Decision Analysis

11.1 Class Sessions October 2 - October 16

Conditional probability. Often, we can collect information that helps us revise our assessment of probabilities, and there are useful techniques to systematize this probability adjustment.

Decision trees and decision analysis. When making decisions were we don't know what exactly will happen, there are methods for structuring the problem to help reveal the best decisions. These decision “trees” also allow us to evaluate additional information about the decision.

Readings We will cover chapters 4. I will provide additional background reading material for this topic.

11.2 Project #5

The fifth case will be due on October 14.

11.3 Final Exam

The final exam will take place on Wednesday October 22, 1:00pm to 4:00pm. There will be multiple rooms for the exam and I will assign you to a particular room. If you have any special requests for the exam, please contact me via email prior to October 14.