

Wisconsin School of Business
Fall session - 2018
GB704, Data to Decisions

Instructor

Erwan Quintin, US Bank Professor of Finance

Office: 5257 Grainger Hall

Office hours: Tuesdays, Thursdays and Fridays, 1-4pm, and by appointment

E-mail : equintin@bus.wisc.edu

URL: erwan.marginalq.com

I do not use Canvas. What I need you to download for the course will be posted on my public access website.

Teaching assistants: Christian Gero and Felix Nguyen

Course objective

This course introduces and applies the statistical inference and data analytics tools commonly used in business applications. We will review the relevant foundations of statistics and probability theory but the emphasis throughout the class will be on applying the resulting concepts to canonical business examples, using both Excel and R. The course also provides an introduction to the language and methods of machine learning.

Resources

The main source of material for this course is my power point slides. No textbook is required since this is an area where free resources are good, plentiful and easily available. The list of topics below details the readings and video-watching I suggest you do for each chapter of the class. Videos, for the most part, are from the Khan academy (KA hereafter.) If I were you I would go through the entire inferential statistics section as you complete this class but, at the same time, I have tried to select the videos that most closely track what we will do in class. I also point you to various notes posted on stat-help.com (SH hereafter.) You may want to bookmark that site for future use. It provides a lot of free tools and references.

I will also expect you to become proficient with some advanced features of Excel, by

far the most common tool for number crunching in Business. On-line resources in this respect are ubiquitous but feel free to ask me if you want specific recommendations. In addition, we will replicate all our estimations using R. All students should download R and R studio on their laptop. Instructions on how to download those free resources are in the following **web book** put together by Professor Margie Rosenberg. The book also provides a full R tutorial on a level that is adequate for our class.

Grading

Your numerical course grade is a weighted average of homework score (15%), midterm score (35%), final score (50%). As per Core MBA rules, the average GPA for the class will be no higher than 3.5. I will implement this objective as follows:

- A Top 25% of course grades
- AB 35-50%
- B 10-25%
- BC 0-5%
- C 0-5%
- D/F 0-5%

The midterm will take place on 9/27, in class, while the final will take place on 10/18, also in class. Exams are open books and open internet. You are allowed to use any resource you want and you will perform all calculations on your own laptop. **Under no circumstances will I take computer malfunction into account in my grading. There is no exception.** Backup laptops will be present in the room for each class and each exam as a favor but, again, **under no circumstances will I take computer malfunction into account in my grading.**

Group problem sets will be assigned more or less on a weekly basis. Your team is your core group. The problems are meant to make you practice/replicate what I will show you in class. While I only need to receive one set of answers per group, each group member must make sure that they fully understand the tasks it contains. Being ready for my exams means first and foremost understanding the homework fully. Several of my exam questions will be variations on homework assignments and will require the use of Excel. Assignments are due at the start of lecture on the due date. No late assignment will be accepted, barring a documented emergency or an exception within the University's guidelines.

For issues pertaining to academic conduct and the honor code, please refer to the **Wisconsin MBA Honor Code**.

Teaching assistants

Each section will have its own teaching assistant. TAs are responsible for grading homework assignments. But their number-one responsibility is to assist you as you work on these assignments and, when the time comes, help you be ready for my exams – which I, alone, will grade.

TAs are available to take your questions by email and in individual and group meetings. I cannot encourage you enough to take advantage of this option.

Course learning outcomes

1. Read, clean, create new datasets
2. Describe the resulting data by computing and interpreting summary statistics, and via visualization tools
3. Describe uncertainty and risk using the language of probability theory
4. Perform hypothesis tests common in business applications
5. Estimate regression models for forecasting purposes and select among competing models via cross-validation
6. Apply machine learning methods in the context of simple examples

Credit hours

This is a 2-credit class. These are met according to the traditional Carnegie Definition: 3 times 75 minutes of classroom time and a minimum of two hours of out of class student work per hour of class over approximately 8 weeks.

List of topics and associated resources

1. Preliminaries
 - (a) SH Intro chapters 2, 5, 7, 9 (both), 19

- (b) KA inference
 - i. Introduction to the normal distribution
 - ii. Standard error of the mean
 - iii. Confidence interval 1
 - iv. Hypothesis testing and p-values
 - (c) KA regression
 - i. Fitting a line to data
 - ii. Squared error of regression line
2. From probability to statistics, and back
- (a) My RE notes (section 3.1 only)
 - (b) KA probability
 - i. Random variables
 - ii. Law of large numbers
 - (c) KA inference
 - i. Central limit theorem
 - ii. Sampling distribution of the sample mean
 - iii. Chi-square probability distribution
 - (d) Critical values for key distributions
3. Hypothesis testing
- (a) SH Intro chapters 17, 18, 19, 20, 22
 - (b) SH Linear regression chapter 1 (review chapter, nice check-point)
 - (c) KA inference
 - i. Confidence interval 1
 - ii. Hypothesis testing and p-values
 - iii. T-statistic confidence interval
 - iv. Hypothesis test for difference for means
 - v. Contingency table chi-square test
 - (d) The bootstrap (We'll implement the bootstrap approach with Excel in class)
4. Regression analysis: a primer

- (a) SH Linear regression chapters 2, 3, 4, 5 (*, don't worry about the SPSS talk in there)
 - (b) KA regression
 - i. Fitting a line to data
 - ii. Correlation and causality
 - iii. R-squared or coefficient of determination
5. Regression analysis for categorical variables
- (a) Stata probit example
6. Model selection
- (a) Kocherlakota's brilliant piece (*, a bit off topic and heavy on the economics)
7. Fundamentals of machine learning