DEPARTMENT OF STATISTICS

STAT 613 Fall 2018

Regression Analysis for Business
Syllabus

Instructors:

Abraham (Adi) Wyner	ajw@wharton.upenn.edu	448 JMHH
Emil Pitkin	pitkin@wharton.upenn.edu	444 JMHH

Source material

Required

- Class Notes. These can be downloaded directly from the Stat 613 Canvas e-room.
- JMP 14 (or 13) (software), SAS Institute, downloadable from upenn.onthehub.com
- Stine and Foster, *Statistics for Business*, Addison Wesley. References in this document are to the third edition, though the second edition is very similar.

Optional (on reserve at Lippincott Library)

- Sall, Creighton, Lehman, *JMP Start Statistics*, 5th Edition, SAS Institute.
- Freedman, Pisani and Purves, *Statistics*, 4th edition, Norton.
- Keller, *Statistics for Management and Economics*, 10th edition, 2014, South-Western Cengage Learning.
- Ellenberg, *How Not To Be Wrong: The Power of Mathematical Thinking.* 2014, Penguin Press. (Not on Reserve).

The fundamental material for the class is contained in the Class Notes, which will be discussed and elaborated in the class lectures. The Stine and Foster (SF) textbook elaborates on most (but not all) of the Class Notes. Links to the relevant readings in SF appear throughout the Class Notes.

JMP is the computer package we'll use extensively for statistical calculations and graphics. In particular, an essential component of 613 will be project work requiring substantial use of JMP. Although JMP is merely a tool and not the central point of the course, it is sufficiently useful that you will need it.

For those who would like further background materials, we recommend Sall, Creighton and Lehman (SHL), Freedman, Pisani and Purves (FPP) and Keller (K). SHL is an example-rich guide to statistical analysis with the statistics package JMP. FPP is a highly verbal and conceptual book - an excellent introduction both for "poets" who are unfamiliar with technical readings and for "quants" who would like a better sense of the reasoning behind statistics. K is in the style of a traditional "reference manual" and explains details and provides many formulas for statistical procedures that are not covered in class.

Class Preparation

As soon as possible, you should obtain and install JMP. **Before each class**, you should review the material from the previous class and you should skim the Class Notes that will be covered. This is a course that builds upon itself and it is crucial to not fall behind. The classes focus on critical interpretation of results and analysis of assumptions. We use JMP to carry out the computations, although the software itself is not the main focus of the course.

You should also read the relevant sections of the SF textbook as annotated throughout the Notes and shown in this syllabus. We strongly recommend that you review the exercises that conclude each chapter. The exercises in each chapter of the SF textbook begin with matching, true/false, and conceptual questions. You should routinely skim these exercises in every chapter; they review notation and basic properties of the methods covered in class. In addition, the course outline identifies specific additional "you do it" exercises that require data analysis or computation that is related to examples and topics of lectures. These exercises will not be graded but are useful for review.

Course Overview

This course provides the fundamental methods of statistical analysis, the art and science of extracting information from data. The course will begin with a focus on the basic elements of exploratory data analysis, probability theory and statistic inference. With this as a foundation, it will proceed to explore the use of the key statistical methodology known as regression analysis for solving business problems. These methods and their application will reappear in many other MBA classes and are part of the basic "tool kit" expected of all MBAs in their careers.

Days on which quizzes will be given are marked with an asterisk.

Lecture Date	Key Topics	Reading (SF)	Exercises
1 Aug 29	*Course overview* *Variation* histogram, boxplot, mean, median, interquartile range, standard deviation, skewness, logarithm	Ch 4 SIA p140	4.55-56, 59 1-4, p 153
2 Sep 05	*Probability models* independence, random variable, distribution, expected value, SD and variance	7,9	7.45, 47 9.35, 37, 43, 49 4M, q56 (p223)
3 Sep 10	*Normal models* continuous random variable, quantiles (Empirical Rule), quantile plot	12	12.39,41,43, 49, 4M q51 (p297)
4* Sep 12	*Association* contingency table, mosaic plot, chi-squared, Simpson's paradox, lurking variable	5	5.39, 43, 45, 53
5 Sep 17	*Conditional probability* dependence, Bayes rule	8	8.39, 45, 47, 49
6 Sep 19	*Sampling distributions* simple random sample, iid, Central Limit Theorem	13, 14.1 SIA p298	
7* Sep 24	*Confidence intervals* inference, *t*-distribution, confidence level, margin of error	15	15.39,43,49,51 *Submit Project Installment 1*
8 Sep 26	*Hypothesis testing* Null and alternative hypotheses, Type I and II errors, α-level, *p*-value, break-even analysis	16	16.39,43,45,47
9 Oct 01	*Comparing two samples* Confounding, two-sample *t*-test, confidence interval for difference, experimental design, paired sampling	17	17.29, 31
10 Oct 3	*Dependence and portfolios* Measures of linear association, covariance, correlation, portfolios and the volatility drag	6, 10	
11* Oct 8	*Fitting lines to data* Slope and intercept, fitted values and residuals, r-squared	19	19.39, 41, 43, 47
Oct 10	*Review Lecture* Variation, inference, testing		
Oct 15	***Midterm Exam* 6-8pm**		
Oct 18			*Submit Project Installment 2*

12 Oct 22	*Fitting curves to data* Transformations (logarithm, reciprocal), elasticity	20	20.33, 35, 37
13 Oct 24	*Simple regression model* Parameters, assumptions, basic diagnostics	21.1-2	
14 Oct 29	*Remedies for common problems* Nonlinearity, dependence, heteroscedasticity, outliers	22	22.37,39,45 4M (q49, p628)
15* Oct 31	*Inference for the Simple Regression Model* Tests, confidence intervals, prediction intervals	21.3-4	21.39,41,43,47
16 Nov 05	*Multiple regression* Scatterplot matrix, marginal and partial slope, path diagram	23.1-2	
17 Nov 07	*Multiple regression model* R^2, *F*-statistic, model profile, diagnostic plots	23.3-5	23.39, 41, 43, 47
18* Nov 12	*Collinearity in multiple regression*	24	24.33, 35, 37, 41 *Submit Project Installment 3*
19 Nov 14	*Using categorical variables in regression* Dummy variable, partial *F*-test, model profile	25.1-4	25.39, 41, 43, 47
	Week of Thanksgiving: No STAT 613 Lectures		
20 Nov 26	*More categorical predictors*	25.5	
21* Nov 28	Review: building a regression model Stepwise regression, data mining	SIA p815	
22 Dec 03	*Forecasting with regression models* Lagged variable, auto-regression, Durbin-Watson, seasonality	27.2-3	27.37,39,43
23 Dec 05	*Installment 4 project review*		
Dec 16	*Installment 4 project due*		*Submit Final Project Installment 11:59 PM*
Dec 20	***Final Exam* 9-11am**		

Attendance

Attendance is an important aspect of the Wharton commitment. Wharton students are admitted in part because of the experiences they bring to the community that they can add to class discussions. Without attending, learning as a collaborative process cannot exist. Accordingly, absences are only appropriate in cases of personal emergency. In addition,

late arrival is disruptive to the learning environment and promptness is expected. Please make note of the start of the term and the time of deliverables and exams as you make travel plans. In case of illness, we require a letter of confirmation from Student Health Services. If you find yourself in a conflict due to your career search or recruiting activity, you should work with the MBA Career Management Office to find a resolution. Absences due to recruiting are not excused. ***Employers are prohibited from requiring recruiting-related activities (e.g., interviews, events or travel) that conflict with a student's academic commitments.*** An employer's inflexibility on this issue is a violation of Wharton's recruiting policies.

Why use JMP?

There are many statistical packages, including SAS, Minitab and Excel. If you are generous, you can even include Excel. JMP does not have a large user base and thus it is not likely to be the tool of choice when you return to the work force. But JMP has many advantages. It is **extremely** powerful. Indeed, its most recent versions are equipped with an unsurpassed suite of artificial intelligence tools including natural language processing, machine learning and internet data acquisition. It can be used entirely through a "point and click" interface which is super easy and highly conducive to exploration. Thus you can concentrate on understanding. You will also be doing very powerful analyses very quickly.

This year, for the first time, we are going to give the learning teams options. If you choose, you many do some or all of the statistical analyses in Excel. We will not provide any technical support (that's ok, you can use Google) and frankly Excel is not equipped to do the final project easily or even at all. So proceed with caution. Alternatively, for the truly skilled, you can use the free programming language "R". This is the statistical programming language for all upper level classes. We lack the resources to support teams that work with R.

Exercises, Quizzes and Exam

There will be weekly exercises as indicated in the course syllabus. These exercises will not be collected, but they are essential for the learning process and you should treat them as a requirement. The textbook supplies brief answers to these questions and office hours are available for further questions.

There will be six in-class quizzes throughout the course. See the Canvas calendar for dates.

There will be a two-hour midterm and a two-hour final exam.

One week grade query maximum from the time work has been handed back.

Learning Team Project

A project will be assigned to each learning team during the course. It will entail the statistical analysis of a data for a business application that your team will describe in four installments. It will be possible to complete these installments before the listed due dates, and you are encouraged to submit them early. This project must reflect the work of only your learning team. You are strictly forbidden from discussing this project with anyone outside your learning team.

Please note that your team for Stat 613 may differ somewhat from your ordinary learning team. We will assign teams after the second week of class.

Teaching Assistants (TAs)

TAs for Stat 613 will hold office hours throughout the course. Times and locations will be posted in the 613 Canvas e-room.

The TAs for Stat 613 are:

- Raiden Hasegawa (raiden@wharton.upenn.edu)
- Gemma Moran (full time after midterm) (gmoran@wharton.upenn.edu)

Classroom Expectations - Concert Rules

- Class starts and ends on time.
- Sit according to the seating chart (posted on line in Canvas).
- Late entry or reentry only under exceptional circumstances.
- Name tents displayed.
- Phones, laptops and other electronic devices turned off. Tablets (e.g., an iPad, Surface etc.) can be used to take notes in class.

Grading

Grades for the course will be based on the following components

Final Examination	35%
Midterm Examination	25%
In-class Quizzes (6, lowest score dropped)	15%
Project (4 installments)	20% (3, 4, 5 and 8% respectively)
Concert rules, including attendance	5%

Instructor Office Hours

- Professor Wyner: Mondays 3-4 PM and Thursday 4-5 PM in 448 JMHH.
- Dr. Emil Pitkin: Mondays, 3-5 PM in 444 JMHH.
- Raiden Hasegawa: Wednesday 3-4 pm in JMHH F96.
- Gemma Moran: TBA. JMHH F96.