Wharton
University of Pennsylvania

DEPARTMENT OF STATISTICS

STAT 613 Fall 2020

Regression Analysis for Business
Syllabus

Professor Abraham (Adi) Wyner ajw@upenn.edu

Teaching Assistant :
Ryan Gross (PhD student in Statistics) rzgross@upenn.edu
Nikki Bowser (2nd Year MBA, Stat Major) anbowser@wharton.upenn.edu

Source material

Required

- JMP 15 (Accessible for free through canvas)
- Pekoz, Erol *Statistics for Managers*. ProbabilityBookstore.com, Boston, MA ISBN 978-0-9795704-2-1
- Class Notes and Videos

Additional Reading Suggestions

- 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.
- Stine and Foster, *Statistics for Business*, Addison Wesley.

Where to get the book:

- https://www.amazon.com/gp/offer-listing/0979570425/ (scroll down until you find Amazon as the seller)
- https://www.barnesandnoble.com/w/the-managers-guide-to-statistics-2018-edition-erol-pekoz/1133600196

- https://upenn.bncollege.com/shop/BNCB_TextbookDetailView?displayStoreId=10056&urlRequestType=Base&catalogId=10001&productId=600008835548&langId=-1&partNumber=MBS_2210659&storeId=10056&sectionId=96837696&item=N **(Penn bookstore)**
- Or search any other book retailer for ISBN 978-097957042

Course Pedagogy

The traditional way of teaching a statistics class is lecture. The professor presents formulas, justifies the formulas and then perhaps illustrates their use by solving problems in class. Students ask questions about confusing points.

I will adopt a much more active learning approach. Although there will be plenty of lecture (some asynchronous), actual class time will also feature in-class activities, student problem-solving, and discussions. Synchronous class time (zoom sessions) will be shorter if a lecture has been assigned to be watched before class. There will always be a synchronous class at your regular class time except on September 28th where there will be no synchronous class (because of the Jewish holiday of Yom Kippur).

For the course to work as well as possible, it is important that you **prepare for class.** Each week you will do some combination of watching videos, reading sections from the textbook, doing self-study problems (submitted), and participating in asynchronous group and individual work.

The class is organized into modules on canvas where all the assignments posted.

Background

The statistics background for students in STAT613 is quite variable. Some of you have never had any courses and others have had many. *I will assume this is your first course in statistics*. For those of you who are more advanced, I will sprinkle in optional material in "math supplement" designed to stretch you a bit.

The Wharton MBA program has a strong analytical focus and the Statistics 613 class has always reflected that orientation. Sometimes the urgency to teach skills obscures the need for conceptual understanding. To that end, I am assigning the book "**Managerial Guide to Statistics"** by Boston University Statistics Professor Erol Pekoz. The book is highly conceptual; it is in many ways a managerial version of class text "Statistics" by Freedman, Pisani and Purves. Erol and I were disciples of Professor Freedman at Berkeley. I think you will enjoy the book. If you are new to statistics (or it has been years and years) the reading will be an essential part of the learning process.

JMP (pronounced "jump") is the computer package we'll use extensively for statistical calculations and graphics. In particular, an essential component of 613 will be homework that use JMP.

Why use JMP?

There are many statistical packages, including SAS, Minitab and even Excel (if you are generous). JMP has a growing user base and but 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. It is also used in many upper level stat classes at Wharton and UPenn.

Can I learn R?

The book "Statistics for Managers" has optional material that teaches you to use the free statistical programming language R. If majoring in Statistics is something you are seriously thinking about, then learning to use R will be important. Use of R is optional in this course but it will be supported by the TAs. You may install and experiment with the free R software package and the corresponding R studio editor using the two links here:

- R: https://cran.rstudio.com/
- R studio: https://www.rstudio.com/products/rstudio/download/

Synchronous Learning

There will always be a class at the regularly scheduled class time. A few classes will be regular lectures. Other classes (perhaps most) will begin with a short review of the material **prepared before class** and a practice problem or two will be solved. Then we will "flip" the class and break you into random teams to complete a worksheet of problems. We will then regroup and select teams will present their solutions to the class.

Piazza

This term we will be using Piazza for class discussion. The system is highly catered to getting you help fast and efficiently from classmates, the TA, and myself. Rather than emailing questions to the teaching staff, I encourage you to post your questions on Piazza. If you have any problems or feedback for the developers, email team@piazza.com.

Find our class signup link at: https://piazza.com/upenn/fall2020/stat613

Lecture Date	**Key Topics**
1 Sep 2	*How not to be fooled with data!* Causality, confounding and Controlled Studies. The method of comparison.
2 Sep 09	*The Histogram* Portraying variation in the data. Summary measures of the center and spread of the Data (median, mean and IQR). SD and Empirical Rule
3 Sep 14	*More Data Visualization:* Box plots, mosaic plots, *More Data Summary Examples*
4 Sep 16	*Normal Distribution:* A very useful tool for understanding data and making predictions.
5 Sep 21	*Telling Statistical Stories with Graphs:* The graph-builder in JMP.
6 Sep 23	*Correlation*
7 Sep 28 **Yom Kippur**	***No Live Class. Class will be recorded.*** *More Correlation*
8 Sep 30	*Regression Equation, Regression Residuals*
9 Oct 5	*Multiple Regression Continued* Fitting hyperplanes to data. Interpreting Regression Equations. Marginal and Partial Slope.
10 Oct 7	*Multiple Regression Continued* Dummy Variables, Interactions
11 Oct 12	*Multiple Regression Examples*
12 October 14	Midterm Exam Review

13 Oct 26	*Probability: a "language" to reflect uncertainty* *Conditional Probability and Bayes Rule*
14 Oct 28	*Random Variables and their Expected Value and SD*
15 Nov 2	*Sampling variation:* The law of averages and the standard error of the sample percentage.
16 Nov 4	*Central Limit Theorem of Statistics*
17 Nov 9	*Confidence Intervals*
18 Nov 11	*Hypothesis Testing* One Sample tests, the p-value.
19 Nov16	*Hypothesis Testing:* Two Sample tests: The "A/B" test
20 Nov 18	*Confidence Intervals in Single Variable Regression* The p-value of a regression.
21 Nov 30	*Confidence Intervals in Multiple Regression*
22 Dec 2	B*uilding a regression model* Stepwise regression, data mining.
23 Dec 07	*How to diagnose problems in regression and what to do about them!*
24 Dec 09	*Finish Up and Review*

Homework, Quizzes and Exam

- There will be seven homework assignments and one final data analysis project. **The homework is not intended to be difficult and you should be able to get full marks or close to if you apply yourselves at least little.** Brevity is encouraged!
- The final project will be more substantial.
- There will be 2 short quizzes.
- 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 (go to Ryan Gross, TA).

Many problems will be from book. All datasets required are posted to canvas. Homework problems assigned must be done individually and submitted online by the due date and time. You may discuss the problems with other students in the class, but **your final write-up should be done individually** and then uploaded to the "Assignments" section of the course website. The only exception is twice during the semester your assigned group will work together to solve a homework problem in detail to be shared with the class as a model.

I will also be assigning "self-study problems" usually from the textbook based on the readings and asynchronous lectures. These will not be graded and they will be simple but they will be submitted online and may be used towards class participation.

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. Your PhD student TA is Ryan Gross (rzgross@upenn.edu). Your MBA TA is Nikki Bowser (anbowser@wharton.upenn.edu).

If you have questions about the grading of homework or exams please go to Ryan.

Classroom Expectations

Please arrive to the zoom room a few minutes early. Turn your video on and your sound to mute. If you want to ask a questions raise your virtual hand. I will then call on you and you can unmute yourself.

Grading

Grades for the course will be based on the following components

Final Examination	30%
Midterm Examination	25%
Graded Quizzes	5%
Final Project	15%
All other Homework	15%
Class Participation	10%