



## DEPARTMENT OF STATISTICS AND DATA SCIENCE

STAT 6130

Fall 2022

### Regression Analysis for Business Syllabus: Sections 001, 002 and 003

**Richard Waterman**

waterman@wharton.upenn.edu

315 WARB

Office hours: M/W 12p – 1pm. 315 WARB and  
Zoom <https://upenn.zoom.us/j/5136434021>

PhD TA: Zirui Fan

[ziruifan@wharton.upenn.edu](mailto:ziruifan@wharton.upenn.edu)

---

### Source material

#### *Required*

- Class Notes. These can be downloaded directly from the Stat 6130 Canvas site.
- JMP 16 (software), SAS Institute, downloadable from the Canvas site.
- Pekoz, Erol, *Managers Guide to Statistics* (2020 edition) [University of Pennsylvania Official Bookstore \(bncollege.com\)](https://www.bncollege.com)  
Ebook (2018 edition) :  
[https://play.google.com/store/books/details/The\\_Managers\\_Guide\\_to\\_Statistics\\_2018\\_Edition?id=f5j6DwAAQBAJ&hl=en\\_US&gl=US](https://play.google.com/store/books/details/The_Managers_Guide_to_Statistics_2018_Edition?id=f5j6DwAAQBAJ&hl=en_US&gl=US)

#### *Optional*

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

The fundamental material for the class are contained in the Class Notes, which will be discussed and elaborated on in the lectures. JMP is the computer package we'll use extensively for statistical calculations and graphics. In particular, an essential component of 6130 will be project work requiring the use of the JMP software. The Pekoz textbook, which is conceptual and discursive in nature elaborates on most (but not all) of the Class Notes. For a more traditional statistics textbook, the optional Stine and Foster serves well.

### 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 important not to fall behind.

Ideally, you should read the relevant sections of the Pekoz textbook before class as annotated in the Course Schedule. We strongly recommend that you review the exercises that conclude each chapter. Relevant exercises from the Pekoz book are identified. These exercises will not be graded but are useful for review and will reinforce the class learnings.

## Course overview

This course provides the fundamental methods of statistical analysis, the art and science of extracting information from data. The emphasis is on a conceptual understanding of the material together with the practical skill of implementing the methods with software. The key technique to be taught is the use of multiple regression modeling. This will be introduced initially as a descriptive and interpretative tool, and then later in the course as an inferential one. These core methods and their application will reappear in many other MBA classes, especially finance, marketing, and operations, and are a part of the basic “tool kit” expected of all MBAs in their careers.

## Course schedule

Class #	Date	Topics
1	8/29	Introduction to the course. Avoiding big mistakes: counts and rates, causality, confounding and controlled studies.
2	8/31	Data visualization I. Displaying variation in data. Numerical measures of center and spread.
3	9/7	The normal distribution and the Empirical Rule. Heavy tails. VaR.Normal quantile plots.
4	9/12 Quiz 1.	Data visualization II. * Mosaic plots. * Scatterplot. * Scatterplot matrix. * Time-series. * Smoothing. * Color coding.
5	9/14 HW 1 due.	Correlation I
6	9/19	Correlation II

7	9/21. Quiz 2.	Simple regression equations and residuals. Graphing residuals.
8	9/26*	The market model and dummy variables.
9	9/28	Non-linearities: dealing with curvature. Logs, powers and elasticity.
10	10/3 HW 2 due.	Multiple regression: Interpretation, marginal and partial slopes.
11	10/5*	Categorical variables and interactions in the MRM.
12	10/10	TEST I in class
13	10/24	Categorical variables and interactions in the MRM. Collinearity,
14	10/26	Probability distributions.
15	10/31	Conditioning and Bayes rule.
16	11/2	Random variables, expected value, standard deviation, and covariance.
17	11/7	Sampling variation: the Law of Averages and standard error.
18	11/9	The Central Limit Theorem. Confidence intervals.
19	11/14	Hypothesis testing. One-sample test. P-values.
20	11/16	Comparative analytics: Two-sample tests for comparisons and A/B testing. Non-parametric tests.
21	11/28	Simple Regression Model (SRM) assumptions. Confidence intervals, inference, and p-values in the Simple Regression Model (SRM).
22	11/30	The Multiple Regression Model (MRM), confidence intervals, inference, and p-values. Diagnostics – residuals and leverage plots.
23	12/5	Multiplicity concerns. Building a multiple regression model with stepwise and the train and test paradigm.
24	12/7	TEST II in-class.

## **Quizzes and Tests**

There will be five 10 minute in-class quizzes during the course. See the Canvas calendar for dates. There are no quiz makeups, and you can drop the lowest of the 5. The in-class tests last one hour and twenty minutes. They are closed book/closed notes and use the LockDown browser. You can bring a one-page cheat sheet to the test and use both sides. You will need a basic scientific calculator.

## **Assignments**

There will be four graded assignments in the semester. Assignments 1 and 3 will be individual and assignments 2 and 4 will be done with your learning team. The homeworks must reflect the work of only you and your learning team. You are strictly forbidden from discussing the homeworks with anyone outside your learning team. Individualized data sets will be provided for the assignments.

Late assignments are penalized by 25% for up to 24 hours late and 50% for up to 48 hours late. At that point solutions will be posted and no further submissions will be accepted. Any homework grade queries must be made via email, to the PhD TA within one-week of the solutions being posted.

## **Final take-home project**

This final deliverable will involve students making their own multiple-regression model and using it to predict outcomes on a hold-out dataset.

## **Attendance and participation**

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. We will be using the attendance app to track attendance. Some classes will involve a learning team based hands-on exercise, often using JMP. A learning team may be called on to present their solution to the class. The teams work should also be uploaded to Canvas where its presence/absence will contribute to the class participation grade.

## **Exercises (not graded)**

There will be weekly exercises available as indicated in the Readings and exercises document. These exercises will not be collected, but they are an important part of the learning process, and you should treat them as a requirement.

## **Teaching Assistants (TAs)**

TAs for Stat 6130 will hold office hours throughout the course. Times and locations will be posted in the 6130 Canvas site.

## **Classroom Expectations - Concert Rules**

- Class starts and ends on time.
- Sit with your learning team.
- 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:

Test 1	20%
Test 2	25%
In-class Quizzes (5, lowest score dropped)	15%
Homeworks (4 installments)	20% (3, 4, 5 and 8% respectively)
Take-home final project	10%
Attendance and participation	10%

There are no extra credit opportunities in the course. Grades are not rounded.