



## DEPARTMENT OF STATISTICS AND DATA SCIENCE

STAT 6130

Fall 2022

# Regression Analysis for Business Syllabus

**Professor Abraham (Adi) Wyner** [ajw@wharton.upenn.edu](mailto:ajw@wharton.upenn.edu)

303 Academic Research Building  
215 898-2439 (you can leave a voice message here)  
@adiwyner twitter

### Teaching Assistants:

**Ryan Gross (PhD student)** [rzgross@wharton.upenn.edu](mailto:rzgross@wharton.upenn.edu). Office Hours: Fridays 12-1pm via Zoom  
**Kavya Garikipiti (2Y MBA)** [gkavya@wharton.upenn.edu](mailto:gkavya@wharton.upenn.edu). Office Hours: Tuesdays 6-7pm (zoom)  
**Ivanna Pearlstein (2Y MBA)** [ipearl@wharton.upenn.edu](mailto:ipearl@wharton.upenn.edu). Office Hours: Tuesdays 6-7pm (zoom, same link as Kavya)

---

### Source material

#### *Required*

- JMP 16 (Accessible for free through canvas)
- Peko, Erol *Managers Guide to Statistics* available at the bookstore  
<https://upenn.bncollege.com/my-account/rental-order/orders>
- Class Notes (canvas)

#### Additional Reading Suggestions

- Sall, Creighton, Lehman, *JMP Start Statistics*, 5<sup>th</sup> 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.

## 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.

Although there will be plenty of lecture (some asynchronous), actual class time will also feature in-class activities, student problem-solving, and discussions. **Class will not meet in person on September 26<sup>th</sup> and October 5<sup>th</sup> (Jewish holidays).** I will post videos which can be watched anytime.

For the course to work as well as possible, it is important that you **prepare for class**. Before every class there is a reading and self-study problems (not submitted). It is highly recommended that you prepare for class by do the reading first and then try to solve the problems. Solutions to all the self-study problems are posted to Canvas. If you properly prepare the lecture will be super clear and easy to follow. Towards the end of class there will be breakout problems. This will be submitted but not graded (the TAs will check that they have been submitted on time.) No credit will be given to late submissions.

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

## Background

The statistics background for students in STAT 6130 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 6130 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 “**Manager’s 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 (who were my teachers and advisors 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. An essential component of 6130 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, 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 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/>

### **Class**

Classes will begin with a lecture and often a problem or two sometimes based on the reading. Then we will “flip” the class after about 50-60 mi and break your teams to complete a worksheet of problems. We will then regroup and time permitting a team will present their solution to the class. If time does not permit, then solutions will be posted or solved quickly by the professor. **You have been assigned seats according to your learning teams.**

### **Attendance and Videos**

Use the attendance app to record your participation. If you miss class, you can submit reasons and excuses. If you would like to watch the video of class, you need to ask the TAs for access (which will be granted only for excused absences). There is no check-in when there is no-class and not on the dates of the in-class exam. For information, see [\(For Students\) Wharton Attendance, Absence, and Video Requests Module](#)

Lecture Date	Key Topics
<b>1</b> August 29	<i>How not to be fooled with data!</i> Causality, confounding and Controlled Studies. The method of comparison.
<b>2</b> August 31	<i>The Histogram</i> Portraying variation in the data. Summary measures of the center and spread of the Data (median, mean and IQR). SD and Empirical Rule
<b>3</b> Sep 7	<i>More Data Visualization:</i> Box plots, mosaic plots, <i>More Data Summary Examples</i>
<b>4</b> Sep 12 QUIZ	<i>Normal Distribution:</i> A very useful tool for understanding data and making predictions.
<b>5</b> Sep 14	<i>Telling Statistical Stories with Graphs:</i> The graph-builder in JMP.
<b>6</b> Sep 19	<i>Correlation</i>
<b>7</b> Sept 21 QUIZ	<i>More Correlation</i>
<b>8</b> Sep 26 <u>Rosh Hashanah</u>	<i>Regression Equation, Regression Residuals</i> NO IN PERSON CLASS: WATCH VIDEO
<b>9</b> Sept 28	<i>Multiple Regression Continued</i> Fitting hyperplanes to data. Interpreting Regression Equations. Marginal and Partial Slope.
<b>10</b> Oct 3	<i>Multiple Regression Continued</i> Dummy Variables, Interactions
<b>11</b> Oct 5 Yom Kippur	<i>Multiple Regression Examples. Transformations</i> NO IN PERSON CLASS: WATCH VIDEO
<b>12</b> October 10	Exam I

<b>13</b> Oct 24	<i>Probability: a “language” to reflect uncertainty</i> <i>Conditional Probability and Bayes Rule</i>
<b>14</b> Oct 26	<i>Random Variables and their Expected Value and SD</i>
<b>15</b> Oct 31	<i>Sampling variation:</i> The law of averages and the standard error of the sample percentage.
<b>16</b> Nov 2 QUIZ	<i>Central Limit Theorem of Statistics</i>
<b>17</b> Nov 7	<i>Confidence Intervals</i>
<b>18</b> Nov 9	<i>Hypothesis Testing</i> One Sample tests, the p-value.
<b>19</b> Nov 14	<i>Hypothesis Testing:</i> Two Sample tests: The “A/B” test
<b>20</b> Nov 16 QUIZ	<i>Confidence Intervals in Single Variable Regression</i> The p-value of a regression.
<b>21</b> Nov 28	<i>Confidence Intervals in Multiple Regression</i>
<b>22</b> Nov 30	<i>Building a regression model</i> Stepwise regression, data mining.
<b>23</b> Dec 05 Optional Quiz 5	<i>How to diagnose problems in regression and what to do about them</i> <i>Wrap-Up !!</i>
<b>24</b> Dec 07	<i>Exam II</i>
<b>Dec 21</b>	<i>Final Project Due</i>

## Homework, Quizzes and Exams

- There will be 6 homework assignments and one final exam-data-analysis project. (5-10 hours of work). Brevity is encouraged! The final data analysis project is individual. Assignments 3 and 5 are individual. Assignments 1,2,4, and 6 are with your teams.
- There will be 4 short quizzes. These will be multiple choice. There will be a fifth quiz that is a make-up for anyone who missed a quiz earlier in the semester.
- There will be two 1.5 hour exams in class.
- One week grade query maximum from the time work has been handed back (go to Ryan Gross, TA).

Many problems will be from the book. All datasets required are posted to canvas. Some of the homework assignments 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 or team** and then uploaded to the “Assignments” section of the course website. Group assignments are submitted once and there is one grade for the entire group. *The content of your homework is to be generated by you (or your team) and you alone. If you copy work from another team or from another source you not only will get severely penalized, you (and your team) will be referred to the Office of Student Conduct for review.*

## Pre-Class Work: Self-Study and Reading

Before most classes, there is a reading assignment from the text and “self-study problems” many based on the readings and lectures. These will not be graded, and they will be simple and they do not need to be submitted.

## Breakout Problems

After every lecture, there will be breakout problems assigned to be completed in groups during class with students sometime presenting their solutions. You are required to submit breakout problem solutions online. You can either type solutions directly into canvas or upload a file or a photo. Completed submissions will get one point (regardless of correctness). Submissions are due on the day they are assigned.

## Teaching Assistants (TAs)

TAs for Stat 6130 will hold office hours throughout the course. Times and locations will be posted on canvas. You can refer questions to any of your TAs. Questions about regrades should be taken to your PhD student TA, Ryan Gross ([rzgross@wharton.upenn.edu](mailto:rzgross@wharton.upenn.edu)).

## **Class Participation Information**

I will construe class participation broadly. There will be plenty of opportunity to participate in class, you don't need to do them all equally well.

1. Answer questions and ask questions during lectures.
2. Submit breakout room solutions
3. Contribute to discussions on Canvas
4. Attend class consistently

## **Grading**

1. The tests and the quizzes are in-class, taken on a laptop using the Lockdown browser. They are closed book, closed notes.
2. Tests last one hour and thirty minutes
3. Quizzes last 10 minutes. There will be a “fifth” quiz that is optional and can be used to replace the lowest quiz or used as a make-up, if an in-class quiz is missed.
4. Late homework is 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 Ryan Gross within one-week of the solutions being posted.
5. There are no extra credit opportunities in the course.

Grades for the course will be based on the following components

- Test 1: 20%
- Test 2: 25%
- Quizzes: 15%
- Homework: 20%
- Take-home final exam data-analysis project: 10%.
- Attendance and participation: 10%.

## **Makeup**

There is no reason to miss a test or a quiz as they are scheduled in class, and students should not schedule anything that conflicts with class time, especially on quiz and exam

dates. Please plan accordingly. Exceptions will only be made for illness, marriages, death etc.