



DEPARTMENT OF STATISTICS

STAT 621

Fall 2021

Accelerated Regression Analysis for Business Syllabus

Instructor:

Emil Pitkin

pitkin@wharton.upenn.edu 329 Academic Research Building

Source material

Required

- Class Notes. These can be downloaded directly from the Stat 621 Canvas page.
- JMP 14 (Download for free through Canvas).

Recommended

- Stine and Foster, *Statistics for Business*, Addison Wesley. References in this syllabus are to the third edition, though the second edition is very similar.
- Huff, *How to Lie with Statistics*, Norton

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.

Materials

The fundamental material for the class, expounded on in lecture, is contained in the class notes. For the interested reader, the Stine and Foster (SF) textbook delves into further detail on the topics we cover; references to SF are in the class notes. *How to Lie with Statistics* belongs on everybody's bookshelf.

JMP is the computer package we'll use in class for statistical calculations, analysis, and graphics. It's the easiest package to learn quickly and use well. Most students prefer

to use it for their regression assignments, but you're welcome to use any program you like, especially if you have prior and extensive familiarity with it. ¹

How to prepare for class

Before each class, you should (1) review your notes from the previous class, and (2) skim the class notes for the class ahead. The best way to learn is to learn twice, and since the course builds upon itself and refers back to itself, you'll succeed best if you reinforce what you've learned, ask about what you don't understand, and treat lecture as an explanation rather than an exposition.

The relevant sections of the SF textbook, as annotated throughout the notes and shown in this syllabus, will help you consolidate what you've learned in class and give you more experience applying what you've freshly learned. The exercises in each chapter begin with matching, true/false, and conceptual questions – you should skim these exercises in every chapter, since they review notation and basic properties of the methods covered in class. The “you do it” exercises that require data analysis or computation are not graded but are useful for review.

Course Overview

This course equips you with the theoretical foundations and practical applications of regression analysis, always with an eye toward solving business problems. These methods and their application will reappear in many other MBA classes, and form part of the essential armory of every graduating MBA.

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

Lecture Date	Key Topics	Reading (SF)	Exercises
1 Aug 30	<i>Fitting lines to data</i> Slope and intercept, fitted values and residuals, R-squared	19	19.39, 41, 43, 47
2* Sep 01	<i>Fitting curves to data</i> Transformations (logarithm, reciprocal), elasticity	20	20.33, 35, 37
3 Sep 08	<i>Simple regression model</i> Parameters, assumptions, basic diagnostics	21.1-2	

¹ Among the books on reserve: Hall, Creighton and Lehman is an example-rich guide to statistical analysis with JMP. Freedman, Pisani and Purves is a loquacious but well grounded exposition of statistics. Keller is a traditional “reference manual,” explaining details giving formulas for statistical procedures not covered in class.

4* Sep 13	<i>Remedies for common problems</i> Nonlinearity, dependence, heteroscedasticity, outliers	22	22.37,39,45 4M (q49, p628)
5* Sep 15	<i>Inference for the Simple Regression Model</i> Tests, confidence intervals, prediction intervals	21.3-4	21.39,41,43,47
6* Sep 20	<i>Multiple regression</i> Scatterplot matrix, marginal and partial slope,	23.1-2	
7* Sep 22	<i>Multiple regression model</i> R^2 , F -statistic, diagnostic plots	23.3-5	23.39, 41, 43, 47 <i>Submit Project Installment 1</i>
8* Sep 27	<i>Collinearity in multiple regression</i>	24	24.33, 35, 37, 41
9* Sep 29	<i>Using categorical variables in regression</i> Dummy variable, partial F -test	25.1-4	25.39, 41, 43, 47
10* Oct 04	<i>Multiple categorical predictors</i>	25.5	
11* Oct 06	<i>Forecasting with regression models</i> Lagged variable, auto-regression, Durbin-Watson, seasonality		
12 Oct 11	Final Exam review		
Oct 12	Final Exam (7pm-9pm)		
Oct 15	<i>Installment 2 due</i>		

Attendance

Prompt attendance is required except, of course, in case of personal emergency and COVID-19 restrictions.

Why JMP?

We chose JMP among R, SAS, Excel, and other standard software because it's the easiest to use, which means that you can more quickly attempt more types of – and more powerful -- analyses and spend more time focusing on interpretation. Because it's “point and click” software, you can explore any dataset at your fingertips. When I do consulting work, I always start my data explorations in JMP.

Quizzes and Exam

There will be daily quizzes, beginning with the second class. For some of the quizzes, the questions will be sent out ahead of time. For the others, you'll see the questions for the first time in class.

A two-hour final exam will cover any material covered in class, but only material covered in class.

Data analysis regression project

The capstone of the course is a full-fledged multiple regression analysis of a dataset you haven't seen before. It will entail the statistical analysis of a genuine business application that you'll work on over two installments. It will be possible to complete these installments before the listed due dates, and you are encouraged to submit them early. The work must be done solo.

Teaching Assistants (TAs)

TAs for Stat 621 will hold office hours in person and over Zoom throughout the course. Times and locations will be posted on Canvas.

Classroom Expectations

Simple: come on time, volunteer answers, and keep all distractors turned off. Taking out a phone without getting permission over email or in person (it'll be granted for reasonable needs like if you're running a company, have a friend or relative that's not well, expect an important career-related email coming in, etc.) caps the final grade at a C. Tablets (iPad, Surface, etc.) can be used to take notes in class.

Grading

Grades will be computed from:

Installment 1	10%
Installment 2	25%
Quizzes	25%
Final Exam	25%
Participation	10%

Instructor Office Hours

- Dr. Emil Pitkin: times will vary weekly to accommodate different schedules, and will be posted on Canvas.