



Introduction to Business Statistics

Instructor and TA

Robert Stine	444 Huntsman Hall	stine@wharton
Linjun Zhang		linjunz@wharton

Professor Stine's office hours follow classes on Tuesday and Thursday, from 4:30-6 pm.
Mr. Zhang will have office hours in F96 JMHH:

Monday	4:00-5:30pm
Wednesday	9:00-10:30am
Friday	10:00-11:30am

Overview

This course explores the use of the statistical method known as regression analysis for the study of problems in business, economics, society, and the sciences. Regression analysis is perhaps the most important and heavily used collection of statistical methods. Its popularity arises from its ability to “explain” variation, find patterns, and make predictions. The models that underlie regression analysis form the basis for even the most recent techniques used in predictive modeling, such as neural networks used in deep learning.

The essential prerequisite for this course is familiarity with the contents of Statistics 101. Topics from Stat 101 include data displays (boxplots, histograms, quantile plots, and scatterplots), summary statistics (such as the mean, standard deviation, and correlation), and properties of statistical estimates and tests (including standard error, confidence intervals, p-values). The first few weeks of Stat 102 review these concepts, focusing on applications and assumptions. If you need to refresh your knowledge, consult materials from Stat 101 and your textbook.

At each class, I expect that you will have read the assigned portions of the textbook. Classes focus on a simple four-step paradigm for answering questions from data:

Frame the problem (motivation)

What's the question I'm trying to answer? What data can help me?

Pick the appropriate statistics (method)

Which statistics should I compute? Do I have a model for the problem?

Are the assumptions of the model/method well-matched to the problem?

Do the data analysis (mechanics)

Software does the heavy lifting here, but you need to know how to use it and make sense of the resulting output.

Interpret the results (message)

If you cannot explain the results to someone who understands the question, but doesn't know statistics, then those stats aren't going to be useful.

Software does the computations, but the key role of the data scientist is to identify an appropriate approach and to interpret its results within the context of the problem. The software itself is not a main focus of the course. I will use JMP for classroom examples. You're welcome to use something else if you prefer, but neither I nor the course TA can be expected to offer help with that choice.

Grading

Your grade for this course is composed of graded homework assignments (MyStatLab), attendance in class, two in-class midterm exams, and a final exam.

Assigned textbook exercises	15%
Attendance	10%
Midterm exam 1	20%
Midterm exam 2	20%
Final Exam	35%

Assignments are due roughly weekly and are graded on-line. Attendance will be sampled through occasional interactive class activities. All questions about grades for any exercises or exams must be resolved within 1 week of posting grades.

Both midterm exams will be held during the scheduled classroom times during the semester. The following calendar has the specific dates. If your current plans conflict with those dates, consider changing your plans. You will not be excused from an exam due to a scheduling conflict with a job interview or other elective events. Plan accordingly.

Rules governing the timing of the final exams at Penn are overseen by the Provost's Office. You can find the rules here:

http://www.upenn.edu/registrar/pdf_main/provost-rules.pdf.

The schedule for homework assignments will be available through Canvas and MyStatLab. You can get access to MyStatLab via Canvas.

Texts and Software

Stine and Foster, *Statistics for Business: Decision Making and Analysis, Second Edition*. Pearson. You will need MyStatLab access as used in Stat 101.

JMP, Version 12 or newer

Planned schedule of lectures, quizzes and exams

Follows on the next page.

Date	Topic	Reading
Aug 30	Overview, sampling variation	Ch 3-6, 11-12
Sep 1	Standard error, central limit theorem	Ch 13-14
Sep 6	Confidence intervals and tests	Ch 15-16
Sep 8	Two-sample comparison	Ch 17
Sep 13	Chi-squared, dependence, and Simpson's paradox	Ch 5, 18
Sep 15	Linear patterns	Ch 19
Sep 20	Residual analysis	Ch 19
Sep 22	Data transformations	Ch 20
Sep 27	Review	
Sep 29	Mid-term Exam 1	Midterm 1
Oct 4	Finding patterns in data	Ch 20, 27
	... Fall Break ...	
Oct 11	Simple Regression Model	Ch 21
Oct 13	Inference and prediction in the SRM	Ch 21
Oct 18	Regression diagnostics	Ch 22
Oct 20	Dependence and time series	Ch 22, 27
Oct 25	Multiple Regression Model	Ch 23
Oct 27	Inference in multiple regression	Ch 23
Nov 1	Collinearity in multiple regression	Ch 24
Nov 3	Review	
Nov 8	Mid-term Exam 2	Midterm 2
Nov 10	Building regression models	Ch 24
Nov 15	Categorical variables in regression	Ch 25
Nov 17	Interactions in regression	Ch 25
Nov 22	More complex categorical features	Ch 25
Nov 29	Trends in time series	Ch 27
Dec 1	Models for time series	Ch 27
Dec 6	Categorical responses in regression	Handout
Dec 8	Review	
Dec 21	Final Exam, 6-8 pm	