

Syllabus – Spring 2024 (B)

Professor: David Juran
<dj114@columbia.edu>

TA: Frederik Fuchs
<FFuchs24@gsb.columbia.edu>

This course is designed for students who wish to increase their capability to build, use, and interpret statistical models for business. It builds on the statistical background gained from the core course in managerial statistics. Students with questions about the course are encouraged to contact the professor.

A primary goal of the course is to enable students to build and evaluate statistical models for managerial use in finance, operations and marketing. The focus is on generating managerially useful information and practical decision-making tools, rather than on statistical theory for its own sake. A number of actual business cases are studied.

Concepts covered include multiple linear regression models and the computer-assisted methods for building them, including stepwise regression and all subsets regression. Emphasis is placed on diagnostic and graphical methods for testing the validity and reliability of regression models.

Course topics include a review of basic statistical ideas, numerical and graphical methods for summarizing data, simple linear and nonlinear regression, multiple regression, qualitative independent and dependent variables, diagnostic methods for assessing the validity of statistical models. The course studies applications of regression to business forecasting and also examines alternative times series forecasting models, including exponential smoothing.

While the primary focus of the course is on regression models, some other statistical models will be studied as well, including cluster analysis, discriminant analysis, analysis of variance, and goodness-of-fit tests.

Term project: A major aspect of course is the opportunity to carry out a practical statistical analysis project of one's own. Students work in teams on a problem of their own choosing. The goal of the project is to develop a useful statistical model for a specific business problem, with the professor providing ongoing guidance and advice during the course of project. The teams will give an oral presentation of their results at the term's end. Examples of previous student projects may be found at <http://www.columbia.edu/~dj114/8114projects.htm>.

Excel is used for basic statistical analysis as well as for developing straightforward regression models. In addition, more advanced commercial statistical software, such as Minitab or SAS, is used to carry out more complex and advanced analyses. In addition to the term project, there will be several computer-based assignments.

Workload and Grading: It is expected that students will attend class regularly and participate fully in class discussions. Since many of these discussions will be based on our analytic assignments (mini-cases), it is important that assigned work be done thoroughly and on time. Assignments can be done individually or in teams of two students. My goal is to maximize the learning that we achieve, and therefore to maximize the grades assigned. The final course grade will be composed of three components:

Attendance and class participation	1/3
Written Assignments	1/3
Term Project	1/3

Homeworks will be graded using the following somewhat subjective criteria:

- Check-Plus: Equivalent to 10 out of 10. Exceeds case expectations by providing unanticipated insight and/or unusually robust analysis compared with the rest of the class.
- Check: 9 out of 10. Fully meets the expectations of a thorough analysis and treatment of the key issues.
- Check-Minus: 8 or less. Below expectation by increasing degrees; significantly inferior to what other teams have done.

Textbooks and Software

Regression analysis is enormous and very much alive research subject. There are many thousands of research papers that have been written on regression and scores of new ones follow each month. In addition there are a hundred or so textbooks and scholarly monographs on the subject. The course, itself, will follow quite closely the text by Chatterjee, Hadi and Price which is listed below. This book strikes a good balance between providing a theoretical understanding and keeping a very concrete focus on applications. In addition to Excel we will use the Minitab statistical package, the software for which comes with a helpful user's manual.

Textbook:

Samprit Chatterjee and Ali Hadi

Regression Analysis by Example, 5th edition (Wiley 2012)

ISBN: 978-0-470-90584-5

Computer Software:

We will use Excel and Minitab. You can download a free 30-day trial version of Minitab from <http://it.minitab.com/en-us/products/minitab/free-trial.aspx> but it would be wise to wait until there are fewer than 30 days until the project presentations.

Tentative Schedule

	Topics	In-Class Examples
21-Mar	Course Objectives & Description Review of Basic Statistical Ideas Intercept, Slope, Correlation, Causality Simple Linear Regression Statistical Model and Concepts Regression in Excel Validity	Computer Repair example
	More Simple Regression Bottom Part of the Output Hypothesis Testing Significance of the slope and intercept Interval Estimation Confidence intervals for the slope and intercept	Gardening example Stock Betas debrief
28-Mar	Multiple Regression Analysis of Variance One-way ANOVA (not regression) Middle Section of the Output Hypothesis Testing Significance of the Whole Model (the F test) Full/Reduced Model Trick	Supervisor Ratings example Neighborhood Interest Rates example
	Summary Measures for the Full Model Top Section of the Output Interval Estimation Nonlinear Regression	Movers debrief Manley's Insurance debrief
04-Apr	Creating the Full Model "From Scratch" Excel Array Functions Qualitative Variables Dummy Variables Interaction Effects	Neighborhood Interest Rates example (B) Cutting Tools example Cars (A) debrief Cigarettes (A) debrief
	Residual Analysis Are they normal? Do they have a common variance? Multicollinearity Serial correlation	Anscombe's Quartet example Money Supply example (A)
11-Apr	Chi-square Goodness-of-Fit Tests Fit to a Normal Simulation Modeling Serial correlation, cont. Runs test Durbin-Watson	Barkevious Mingo (Normal) example Catalog Company (Exponential) example Money Supply example (B) Rick Beck Credit (A)

	Model Building Variable Selection Methods Minitab	
	Forecasting Methods Exponential Smoothing Simple Trend (Holt's Method) Seasonality (Winters' Method) Regression Methods Trend Seasonality Lagged Variables	General Motors example Hotel example Artsy debrief
18-Apr	Two Multivariate Methods Cluster Analysis Discriminant Analysis	B-Schools example Rick Beck Credit (B) example Steam debrief Cars (B) debrief
	Binary Logistic Regression Why? Theoretical and practical difficulties in using regular (continuous) dependent variables How? Minitab procedure Interpreting results Some diagnostics Making predictions Comparison with regular regression model	Rick Beck Credit (C) example Space Shuttle example
25-Apr	Project Presentations	

Schedule of Written Assignment Due Dates

	Deliverables
Thu. 21-Mar	Stock Betas
Thu. 28-Mar	Movers & Insurance
Thu. 04-Apr	Cigarettes (A) and Cars (A)
Thu. 11-Apr	Artsy & Cigarettes (B)
Thu. 18-Apr	Steam and Cars (B)
Thu. 25-Apr	Presentations