

Term Information

Effective Term Autumn 2014

General Information

Course Bulletin Listing/Subject Area Statistics
Fiscal Unit/Academic Org Statistics - D0694
College/Academic Group Arts and Sciences
Level/Career Undergraduate
Course Number/Catalog 3301
Course Title Statistical Modeling for Discovery I
Transcript Abbreviation Stat Model Disc 1
Course Description Statistical models for data analysis and discovery in big-data settings, with primary focus on linear regression models. The challenges of building meaningful models from vast data are explored, and emphasis is placed on model building and the use of numerical and graphical diagnostics for assessing model fit. Interpretation and communication of the results of analyses is emphasized.
Semester Credit Hours/Units Fixed: 3

Offering Information

Length Of Course 14 Week
Flexibly Scheduled Course Never
Does any section of this course have a distance education component? No
Grading Basis Letter Grade
Repeatable No
Course Components Lecture
Grade Roster Component Lecture
Credit Available by Exam No
Admission Condition Course No
Off Campus Never
Campus of Offering Columbus

Prerequisites and Exclusions

Prerequisites/Corequisites Math 2568 and Stat 3202
Exclusions

Cross-Listings

Cross-Listings

Subject/CIP Code

Subject/CIP Code 27.0501
Subsidy Level Baccalaureate Course
Intended Rank Junior, Senior

Requirement/Elective Designation

Required for this unit's degrees, majors, and/or minors

Course Details

Course goals or learning objectives/outcomes

- Formulate regression models that describe relationships between variables and understand the models' statistical foundations
- Perform a complete regression analysis and communicate the results in both statistical and problem-specific terms
- Use linear regression methods to build models for large data sets and use the results of the analysis to recommend actions
- Evaluate and compare different regression models using formal statistical methods and graphical techniques
- Formulate a regression model for data collected over time and identify whether more advanced time series modeling techniques are required

Content Topic List

- Simple and multiple linear regression
- Model building
- Diagnostic checks
- Constructing and choosing useful predictors
- Computation for big data sets
- Introduction to data collected over time

Attachments

- 3301_Syllabus.pdf
(Syllabus. Owner: Hans, Christopher M)

Comments

- This is a required course for the proposed major in Data Analytics. *(by Craigmile, Peter F on 10/11/2013 03:18 PM)*

Workflow Information

Status	User(s)	Date/Time	Step
Submitted	Hans, Christopher M	10/09/2013 02:50 PM	Submitted for Approval
Approved	Craigmile, Peter F	10/13/2013 06:08 PM	Unit Approval
Approved	Hadad, Christopher Martin	10/14/2013 06:52 AM	College Approval
Pending Approval	Vankeerbergen, Bernadette Chantal Nolen, Dawn Jenkins, Mary Ellen Bigler Hogle, Danielle Nicole Hanlin, Deborah Kay	10/14/2013 06:52 AM	ASCCAO Approval

Statistics 3301

Statistical Modeling for Discovery I

3-semester-hour course

Prerequisite: Math 2568 (Linear Algebra) and Stat 3202 (Introduction to Statistical Inference for Data Analytics)

Exclusions:

Class distribution: Three 55-minute lectures per week

Course Description and Learning Outcomes

Statistical models for data analysis and discovery in big-data settings, with primary focus on linear regression models. The challenges of building meaningful models from vast data are explored, and emphasis is placed on model building and the use of numerical and graphical diagnostics for assessing model fit. Interpretation and communication of the results of analyses is emphasized.

Upon successful completion of the course, students will be able to

1. Formulate regression models that describe relationships between variables and understand the models' statistical foundations
2. Perform a complete regression analysis and communicate the results in both statistical and problem-specific terms
3. Use linear regression methods to build models for large data sets and use the results of the analysis to recommend actions
4. Evaluate and compare different regression models using formal statistical methods and graphical techniques
5. Understand the challenges of regression modeling for data collected over time

Required Text and Other Course Materials

The required textbook for the course is (books currently under review). The book is available for purchase at the official University bookstore (ohiostate.bkstore.com) and elsewhere online. The book is available on reserve in the 18th Avenue Library.

Students will be required to use the R¹ software environment for statistical computing and graphics. R can be downloaded for free at <http://www.r-project.org>.

¹ For information on the use of R in data analytics, see:

- <http://www.revolutionanalytics.com/why-revolution-r/whitepapers/r-is-hot.php>
- <http://techcrunch.com/2012/10/27/big-data-right-now-five-trendy-open-source-technologies/>
- <http://www.nytimes.com/2009/01/07/technology/business-computing/07program.html>
- <http://bits.blogs.nytimes.com/2009/01/08/r-you-ready-for-r/>

Instructions for using the software will be given in class. Many students prefer to use RStudio, an IDE designed for use with R. RStudio is available for free at <http://www.rstudio.com>.

Assignments

Homework will be assigned (approximately) bi-weekly, will be due on the dates announced in class and will be graded. Assignments will consist of a mix of technical questions to assess students' understanding of the statistical models, and questions asking students to perform analyses of data sets. The grade for the analysis portion of each assignment will be based on both the accurateness and appropriateness of the analysis, as well as the clarity of the description of the analysis and results.

Project: Each student will be responsible for completing an individual project. Proposals for project ideas will be due mid-way through the semester, and the project will be due near the end of the semester. The project will consist of finding a data set, formulating questions that can be answered with the data, and performing an appropriate analysis to answer the questions.

Exams

There will be two in-class midterms that cover material from lecture, the assigned readings and homework.

A cumulative final examination will be given during the university's examination period.

Grading Information

The final course grade will be based on homework assignments, two projects, two midterms and a comprehensive final examination. The weights for each component of the grade are:

Homework	Midterm 1	Midterm 2	Project	Final Exam
15%	20%	20%	15%	30%

Outline of topics

1. Introduction to statistical models
 - a. Additive models
 - b. Multiplicative models
 - c. Sources of variation
 - d. Hierarchical models and structured dependence
2. Linear regression models
 - a. Simple linear regression
 - b. Review of matrix algebra
 - c. Multiple linear regression
 - d. Building models with categorical predictor variables
 - e. Model building and analysis: constructing useful predictors from unfiltered information, diagnostics, case analysis, outliers, prediction
3. Big data challenges in regression modeling
 - a. Building models with meaningful causal relationships in large-data settings
 - b. Issues of scale and aggregation, weights
 - c. Robustness considerations
 - d. Computation
4. Translating models to actions
 - a. Case study
5. Regression models for data collected over time
 - a. Lagged regression
 - b. Regression models with seasonal effects
 - c. Identifying correlated errors
6. Overview of change-point analysis

Statement on Academic Misconduct

It is the responsibility of the Committee on Academic Misconduct to investigate or establish procedures for the investigation of all reported cases of student academic misconduct. The term “academic misconduct” includes all forms of student academic misconduct wherever committed; illustrated by, but not limited to, cases of plagiarism and dishonest practices in connection with examinations. Instructors shall report all instances of alleged academic misconduct to the committee (Faculty Rule 3335-5-487). For additional information, see the Code of Student Conduct <http://studentlife.osu.edu/csc/>.

Special Accommodations

Students with disabilities that have been certified by the Office for Disability Services will be appropriately accommodated and should inform the instructor as soon as possible of their needs. The Office for Disability Services is located in 150 Pomerene Hall, 1760 Neil Avenue; telephone 292-3307, TDD 292-0901; <http://www.ods.ohio-state.edu/>.