

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 3302
Course Title Statistical Modeling for Discovery II
Transcript Abbreviation Stat Model Disc 2
Course Description Statistical models for data analysis and discovery in big-data settings. The regression methods developed in Stat 3301 are extended to data settings with binary and multi-category outcomes. An introduction to some of the most commonly used statistical methods for exploring and analyzing multivariate data is provided. 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 Stat 3301
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

- Build, fit and interpret statistical models for binary outcomes
- Understand the difference between nominal and ordinal outcomes and build regression models that are appropriate for each
- Recognize the types of questions that can be answered by regression models for multi-category data and structure models to answer those questions
- Comprehend the statistical principles that underlie basic methods of multivariate data analysis

Content Topic List

- Logistic regression
- regression for data with multi-category outcomes
- introduction to multivariate data
- multivariate normal distribution
- principal components analysis
- linear and quadratic discriminant analysis

Attachments

- 3302_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 02:51 PM	Submitted for Approval
Approved	Craigmile,Peter F	10/13/2013 06:08 PM	Unit Approval
Approved	Hadad,Christopher Martin	10/14/2013 06:50 AM	College Approval
Pending Approval	Vankeerbergen,Bernadette Chantal Nolen,Dawn Jenkins,Mary Ellen Bigler Hogle,Danielle Nicole Hanlin,Deborah Kay	10/14/2013 06:50 AM	ASCCAO Approval

Statistics 3302

Statistical Modeling for Discovery II

3-semester-hour course

Prerequisite: Stat 3301 (Statistical Modeling for Discovery I)

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. The regression methods developed in Stat 3301 are extended to data settings with binary and multi-category outcomes. An introduction to some of the most commonly used statistical methods for exploring and analyzing multivariate data is provided. Interpretation and communication of the results of analyses is emphasized. Upon successful completion of the course, students will be able to

1. Build, fit and interpret statistical models for binary outcomes
2. Understand the difference between nominal and ordinal outcomes and build regression models that are appropriate for each
3. Recognize the types of questions that can be answered by regression models for multi-category data and structure models to answer those questions
4. Comprehend the statistical principles that underlie basic methods of multivariate data analysis

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>. 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>.

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/>

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. Regression models for binary outcomes
2. Logistic regression models
3. Challenges in large data settings
4. Regression models for polytomous (multi-category) outcomes
5. Models for nominal responses
6. Models for ordinal responses
7. Introduction to multivariate data and the multivariate normal distribution
8. Principal Components Analysis and Regression
9. Linear and Quadratic Discriminant 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/>.