Statistics 3301
Statistical Modeling for Discovery I
3-semester-hour course

Autumn 2015 Syllabus

Instructor:	Dr. Christopher Hans	**When:**	MWF 1:50–2:45
Email:	hans@stat.osu.edu	**Where:**	MP 1005
Office:	327 Cockins Hall	**Office Hours:**	Wed. 3:30–4:30
Website:	Carmen		Thur. 2:00–3:00

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

Class format: 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 for 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

Textbook and Other Course Materials

We will use the textbook *Applied Linear Regression, Fourth Edition* (2014) by Sanford Weisberg. The book can be purchased from various online outlets. An electronic version of the book can be accessed through The Ohio State University

Libraries at http://bit.ly/1Q4xbLB. Reading and homework will be assigned from the book throughout the semester. Be sure to use the fourth edition (red/orange cover) and not the third edition (green cover). The material in the textbook will be supplement with additional course notes as necessary throughout the semester. More information about the textbook can be found at http://users.stat.umn.edu/~sandy/alr4ed/.

Students will be required to use the R[1] 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. Once R has been downloaded and installed, students will also need to install the (free) R package **alr4,** which contains the data sets discussed in the textbook.

Assignments

Homework will be assigned regularly throughout the semester, 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 data 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: There will be an end-of-the-semester project where students will work in groups of (roughly) two on an analysis of data. 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 formulating questions that can be answered via data analysis, finding relevant data, and performing an appropriate analysis to answer the questions. Further guidelines will be given mid-way through the semester.

Exams

There will be two in-class midterms that cover material from lecture, the assigned readings and homework. One midterm will be held in early-to-mid October and the other in early-to-mid November. The dates for the midterm exams will be announced two weeks ahead of time.

A cumulative final examination will be given during the university's examination period on December 16, 2015.

---

[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/

Course Grade 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%	20%	25%

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/.

Syllabus Version

8/31/15:   Added information on accessing electronic version of textbook

8/26/15:   Original version