

Statistics 3301

Statistical Modeling for Discovery I

3-semester-hour course

Autumn 2017 Syllabus

Instructor:	Dr. Christopher Hans	When:	MWF 1:50-2:45
Email:	hans@stat.osu.edu	Where:	CH 218
Office:	CH 327 (Cockins Hall)	Office Hours:	Wed. 5:00-6:00
Website:	Carmen		Thur. 2:00-3:00

Grader:	Yiyang Lin	Tutor room hours and	Mon. 3:00-5:00
		location:	Wed. 3:00-4:00
Email:	lin.1459@osu.edu		CH 122

Prerequisite: C- or better in Stat 3202. Prereq or concur: Math 2568.

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. An electronic version of the book can be accessed for free through The Ohio State University Libraries at <http://bit.ly/1Q4xbLB> (although there may only be limited copies available at any given time). 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 supplemented 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 are required to use the R 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.

We will occasionally use a graphics package called **ggplot2**. Examples of how to use this package to create statistical graphics will be given in class. The book “ggplot2: Elegant Graphics for Data Analysis” (Second Edition) by Hadley Wickham provides a detailed description of the software. If you are on campus, you should be able to download the book for free at the following link: <http://www.springer.com/us/book/9783319242750>. I recommend reading the book if you are interested in developing your ggplot2 skills, but you are not required to read the book for Stat 3301.

Lecture notes will be posted on Carmen along with the R Markdown source used to generate the notes. You will likely find it helpful to print out these notes ahead of time and bring them with you to class so you can annotate them as we discuss the material. The R Markdown source will contain all of the R code used to generate plots and analyses shown in the notes. You can refer to this R code when doing the homework assignments and studying the course material. The source code will be helpful if you are interested in learning how to use R Markdown (not required for this class).

Tutor Room

The Department of Statistics will operate a tutor room for students in the courses Stat 3201, 3202 and 3301 in **Cockins Hall (CH) 122**. The room will be staffed by graders for the three courses. You can visit the tutor room for help with homework assignments and to ask questions about R. The hours during which the grader for Stat 3301 will be available are provided at the top of this syllabus.

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.

Labs: Two take-home "lab" exercises will be assigned during the semester. These will be guided exercises that are intended to introduce you to the graphics packages ggplot2 and ggvis. The lab assignments will be graded complete/incomplete.

Data Analysis: There will be an end-of-semester data analysis assignment that will tie together the concepts learned throughout the semester. A written report describing the analysis will be required. Details will be provided midway through the semester.

Exams

There will be two in-class midterms that cover material from lecture, the assigned readings and homework. The first midterm will be held on Monday, October 2 and the second will be held on Wednesday, November 8.

A cumulative final examination will be given during the university's examination period on Wednesday, December 13, 2017 **from 2:00-3:45** in our usual classroom. No early or-makeup exams will be given except in cases of documented emergencies (serious illness, etc.).

Course Grade Information

The final course grade will be based on homework assignments, the labs, the end-of-term data analysis, the two midterms and a comprehensive final examination.

The weights for each component of the grade are:

Homework	Labs	Midterm 1	Midterm 2	Data Analysis	Final Exam
15%	5%	20%	20%	15%	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

8/27/17: Added info on tutor room location/hours

8/23/17: Original version