

## Syllabus for Stat 3202: Introduction to Statistical Inference for Data Analytics

**Instructor:** Yoonkyung Lee

**Office:** 440B Cockins Hall

**Office Hours:** W 4:20-5:15pm; R 2:00-2:55pm; and by appointment

**Office Phone:** 614-292-9495

**E-mail:** yklee@stat.osu.edu

---

**Lectures:** WF 3:00-4:20pm in Cockins Hall (CH) 218

**Recitations:** M, either 3:00-3:55pm in East Annex (EA) 285 or 4:10-5:05pm in EA 295

**Credit Hours:** 4

**Course Teaching Assistant:** Your TA is Mr. Min Ho Cho. During recitation sessions, Mr. Cho will assist you in working through suggested problems, and he will work with you to perform some activities using the R software. You are also encouraged to visit office hours if you have further questions.

**Office Hours:** TW 1:50-2:50pm; W 4:10-6:20pm (CH 122)

**E-mail:** cho.829@osu.edu

---

**Prerequisite:** Stat 3201: Introduction to Probability for Data Analytics

**Required Text:** *Mathematical Statistics with Applications*, 7<sup>th</sup> edition, by Wackerly, Mendenhall, and Scheaffer, Brooks/Cole, Cengage Learning, 2008.

---

**Course Description and Learning Outcomes:** The course covers foundational inferential methods for learning about populations from samples, including point and interval estimation, and the formulation and testing of hypotheses. Statistical theory is introduced to justify the approaches. The course emphasizes challenges that arise when applying classical ideas to big-data, partially through the use of computational and simulation techniques.

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

1. Describe the role of a parameter in a statistical model and its relationship to observed data
2. Use data to estimate and describe uncertainty about the parameters of a statistical model
3. Translate scientific hypotheses about a population into mathematical statements about parameters in a statistical model
4. Formulate statistical procedures to test a hypothesis about parameters in a statistical model, and interpret the results in both statistical and application-specific terms
5. Explain the difference between statistical and practical significance in massive data settings
6. Appreciate the effect of missing data on statistical inference
7. Evaluate and compare different statistical procedures for answering the same question

**Course attendance policy:** You are expected to attend all lectures and recitations. Formal attendance records will not be kept; however, students are responsible for all material covered in class.

**Homework:** Homework problems will be assigned for each topic covered in the course, and solutions to all assigned problems will be posted. Homework will not be collected or graded.

**Lab:** Lab exercises using the R software will be carried out in approximately half of the scheduled recitation sessions. These lab exercises will be turned in via Carmen, and will together account for 10% of the overall grade.

**Quizzes:** Short quizzes (5 in total) will be given approximately every other week in class. Each quiz will consist of one problem taken directly from the assigned homework, and will account for 5% of your overall grade. The lowest quiz score will be dropped from the recorded grades. No notes may be used on quizzes.

**Projects:** There will be two data analysis projects, which will be completed in groups of 3-4 students. Each group will be provided with a data set and several questions related to it. The first project will focus on modeling and parameter estimation; the second project will focus on applying and extending the statistical methods learned in class. Students will present their work in class (about 10 minute presentations) and also submit a write-up of their results (approximately 3 pages of text, plus any additional figures and tables).

**Exams:** There will be two in-class exams and a final exam. Statistical tables will be provided as needed. Calculators may be used on the exams, but the calculators on cell phones, PDAs, or any other communication device are NOT allowed. You may use one 8.5 x 11 inch sheet of paper (both sides), with whatever facts, formulas, or explanations you find helpful, for each exam.

**Makeup exams:** If you absolutely need a makeup exam and have a valid excuse, please see me (not your recitation instructor) for the necessary arrangements. However, you must notify me in advance in such a situation. A make-up exam may be a bit harder than the regularly scheduled exam and must be taken within a week of the missed exam. Exceptions to this policy will be permitted only in extreme situations such as serious injury immediately prior to an exam or severe illness requiring hospitalization.

**Final Grade:** Your final course grade will be based on the following weighting of assessment components:

Quizzes (4) – 5% each	Projects (2) – 10% each
Midterm exams (2) – 15% each	Lab grade - 10%
Final exam – 20%	

Effort, class attendance, and class participation will be taken into account in borderline cases.

**Grading Scale:** The following grading scale will be used:

>90%	A/A-
80-90%	B-/B/B+
70-80%	C-/C/C+
60-70%	D/D+
<60%	E

**Website:** The course website will be available at <http://www.carmen.osu.edu>. Check the website periodically for announcements about the class and other class material.

**R Software:** The R (<https://www.r-project.org/>) and RStudio (<https://www.rstudio.com/>) software will be used for the lab portion of the course. This software is installed in the recitation classroom, as well as in most computer labs on campus. It is free software that you can download and install on your personal machines as well. Your TA will help you learn to use R for statistical analysis during recitation, but you should also expect to put in time outside of recitation doing data analysis with R for homework.

---

**Academic Misconduct:** Please help us to maintain an academic environment of mutual respect, fair treatment, and personal growth. You are expected to produce original and independent work for exams. Although students are encouraged to work together on homework and lab assignments, all students must submit their own written work IN THEIR OWN WORDS. 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/>.

**E-mail Correspondence:** In order to protect your privacy, all course e-mail correspondence must be done through a valid OSU name.nm account.

**Special Accommodations:** Students with disabilities (including mental health, chronic or temporary medical conditions) that have been certified by the Office of Student Life Disability Services will be appropriately accommodated and should inform the instructor as soon as possible of their needs. The Office of Student Life Disability Services is located in 098 Baker Hall, 113 W. 12th Avenue; telephone 614-292-3307, [slds@osu.edu](mailto:slds@osu.edu); [slds.osu.edu](http://slds.osu.edu).

**Diversity:** The Ohio State University affirms the importance and value of diversity in the student body. Our programs and curricula reflect our multicultural society and global economy and seek to provide opportunities for students to learn more about persons who are different from them. We are committed to maintaining a community that recognizes and values the inherent worth and dignity of every person; fosters sensitivity, understanding, and mutual respect among each member of our community; and encourages each individual to strive to reach his or her own potential. Discrimination against any individual based upon protected status, which is defined as age, color, disability, gender identity or expression, national origin, race, religion, sex, sexual orientation, or veteran status, is prohibited.

**Drop dates:** The last day to drop the course without a “W” appearing on your record is Friday, February 2. The last day to drop the course without petitioning is Friday, March 23.

## LIST OF TOPICS AND READING

Topic	Title	Reading
T0	Introduction to Statistical Inference	Ch. 7, Sec. 8.1
T1	Point Estimation I	Sec. 8.2 - 8.3
T2	Point Estimation II	Sec. 9.3 - 9.5
T3	Point Estimation III	Sec. 9.6 - 9.7
T4	Confidence Intervals	Sec. 8.5 - 8.9
T5	Bootstrap	Handout
T6	Hypothesis Testing	Sec. 10.1 - 10.9
T7	Nonparametric Testing	Sec. 15.1 - 15.4
T8	Linear Statistical Models I	Sec. 11.1 - 11.4
T9	Linear Statistical Models II	Sec. 11.5 - 11.7
T10	Bayesian Inference	Sec. 16.1 - 16.4

## LIST OF LAB TOPICS AND REPORT DUE DATES

Lab exercises will be turned in via Carmen and due by midnight on the date listed.

Lab	Title	Due date
1	Introduction to R and RStudio	1/12
2	Point Estimation and Sampling Distributions	1/29
3	Confidence Intervals for Proportions (M&M lab)	2/14
4	Bootstrap Standard Errors and Confidence Intervals	2/26
5	Power for Hypothesis Tests of a Population Mean	3/7
6	Comparison of Methods for Testing Hypotheses	4/2
7	Fitting Linear Models in R	4/4

## TENTATIVE LECTURE AND RECITATION SCHEDULE

Please note that topics covered may change depending on the pace of the course. Any change in due dates or exam dates will be officially announced in class and also posted on the course website.

Week 1				
M	1/8	Lab 1: Introduction to R and RStudio Intro to statistical inference, point estimation Point estimation – bias, MSE; <b>Lab 1 due</b>	Ch. 7, Sec. 8.1 8.2	T0
W	1/10			
F	1/12			T1
Week 2				
M	1/15	MLK Holiday Point estimation – consistency Point estimation – sufficiency <b>QUIZ 1: T1</b>	No class 9.3 9.4, 9.5	T2
W	1/17			
F	1/19			T2
Week 3				
M	1/22	Lab 2: Point Estimation and Sampling Distributions Point estimation – MOM, MLE MLE <b>QUIZ 2: T2</b>	9.6, 9.7 9.7	T3
W	1/24			
F	1/26			T3
Week 4				
M	1/29	Work on projects; <b>Lab 2 due</b> Interval estimation, CI for means Review for Exam 1	8.6 - 8.9	T4
W	1/31			
F	2/2			
Week 5				
M	2/5	Work on projects <b>EXAM 1: T0-T3</b> CI for means, proportions	8.6 - 8.9	T4
W	2/7			
F	2/9			
Week 6				
M	2/12	Lab 3: Confidence Intervals for Proportions Bootstrap; <b>Lab 3 due</b> Hypothesis testing <b>QUIZ 3: T4, T5</b>	handout 10.1, 10.2	T5
W	2/14			
F	2/16			T6
Week 7				
M	2/19	Lab 4: Bootstrap SEs and CIs <b>Project 1 presentations</b> <b>Project 1 presentations</b>		
W	2/21			
F	2/23			
Week 8				
M	2/26	Lab 5: Power for Hypothesis Tests of a Population Mean; <b>Lab 4 due</b> Hypothesis testing - type I and type II errors, p-values Hypothesis testing - power, common tests	10.3 - 10.7 10.7 - 10.9	T6
W	2/28			
F	3/2			T6
Week 9				
M	3/5	Lab 5 continued Nonparametric tests; <b>Lab 5 due</b> Nonparametric tests <b>QUIZ 4: T6</b>	15.1 - 15.4 15.1 - 15.4	T7
W	3/7			
F	3/9			T7
SPRING BREAK				

Week 10				
M	3/19	Lab 6: Comparison of Methods for Testing Hypotheses		
W	3/21	Review for Exam 2		
F	3/23	<b>EXAM 2: T1-T7</b>		
Week 11				
M	3/26	Lab 6 continued		
W	3/28	Introduction to statistical modeling	11.1 - 11.2	T8
F	3/30	Estimation in statistical models	11.3 - 11.4	T8
Week 12				
M	4/2	Lab 7: Fitting Linear Models in R; <b>Lab 6 due</b>		
W	4/4	Parameter inference in linear models; <b>Lab 7 due</b>	11.5 - 11.7	T9
F	4/6	Model-fitting examples, diagnostics		T9
		<b>QUIZ 5: T8, T9</b>		
Week 13				
M	4/9	Work on projects		
W	4/11	Bayesian inference	16.1 - 16.4	T10
F	4/13	Bayesian inference	16.1 - 16.4	T10
Week 14				
M	4/16	Work on projects		
W	4/18	<b>Project 2 presentations</b>		
F	4/20	<b>Project 2 presentations</b>		
Week 15				
M	4/23	Review for Final		
R	4/26	4:00-5:45pm		
		<b>FINAL EXAM: T0-T10</b>		