

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

**Instructor:** Dr. Kubatko

**Office:** 219 Cockins Hall

**Office Hours:** W 2:00PM-3:00PM, R 11:00AM-12:00PM, other times by appointment

**Office Phone:** 247-8846

**E-mail:** kubatko.2@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

**Website:** Please visit <http://www.carmen.osu.edu/>. Check Carmen periodically for announcements about the class and other class material.

**Homework:** Homework problems will be assigned and graded for each topic covered in the course. Homework must be turned in during lecture on the date it is due. If you are unable to attend lecture when the homework is due, you must bring it to me in my office **earlier** in the day. Please write your name on the top of each page of your assignment, and staple the pages together.

**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.

**Notes for use on the exams:** You may use one 8.5 x 11 inch sheet of paper (both sides), with whatever facts, formulas, or explanations you find helpful, for the first exam. Two sheets of paper (as described

above) may be brought to the second exam and to the final 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.

**Full credit on homework and exam problems:** You need to show your justification for or work on each homework or exam problem. Answers without work will not receive full credit.

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

Homework – 20%	Projects (2) – 10% each
Midterm exams (2) – 20% each	Final exam – 20%

**R Software:** The R and RStudio 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.

**Course Teaching Assistant:** Your TA is Mr. Andrew Bean. During recitation sessions, Mr. Bean 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.

**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 quizzes and exams. Although students are often encouraged to work together on homework assignments, all students must submit their own written work **IN THEIR OWN WORDS**. Academic misconduct will not be tolerated and will be dealt with procedurally in accordance with University Rule 3335-31-02. (This policy can be found at <http://oaa.osu.edu/procedures/1.0.html>.)

**E-mail Correspondence:** In order to protect your privacy, all course e-mail correspondence must be done through a valid OSU name.nn account. If you have not activated your OSU email account, you can activate your account at <https://acctmgt.service.ohio-state.edu/cgi-bin/KRB1EntryAdd>.

**Special Accommodations:** All students who feel they may need accommodations based on the impact of a disability should contact the instructor privately to discuss their specific needs. Students with documented disabilities must also contact the Office of Disability Services (ODS) in 150 Pomerene Hall (phone: 292-3307) to coordinate reasonable accommodations for the course. ODS forms must be given to your instructor as early in the quarter as possible to be filled out and returned to you.

**Drop dates:** The last day to drop the course without a “W” appearing on your record is Friday, February 6. The last day to drop the course without petitioning is Friday, March 27.

## TENTATIVE LECTURE AND RECITATION SCHEDULE

Please note that topics covered, particularly for the recitation, may change depending on the pace of the course.

Week 1			
T	1/13	Introduction to Statistical Inference Review of sampling distributions Introduction to R and RStudio	Sec. 8.1 Ch. 7
R	1/15		
F	1/16		
Week 2			
T	1/20	Sampling distributions in complex settings Point estimation - bias, MSE, efficiency Simulating sampling distributions in complex settings	8.2-8.3, 9.2
R	1/22		
F	1/23		
Week 3			
T	1/27	Point estimation - consistency, sufficiency, UMVUE Point estimation - MOM, MLE Computing point and interval estimates in R	9.3-9.5 9.6-9.7
R	1/29		
F	1/30		
Week 4			
T	2/3	Problem solving / Review for Exam Exam 1 Go over exam 1	
R	2/5		
F	2/6		
Week 5			
T	2/10	Interval estimation Bootstrap Bootstrap activity	8.6 - 8.9
R	2/12		
F	2/13		
Week 6			
T	2/17	Missing data Project 1 presentations Missing data activity	
R	2/19		
F	2/20		
Week 7			
T	2/24	Project 1 presentations Introduction to hypothesis testing Problem solving	10.1-10.2
R	2/26		
F	2/27		
Week 8			
T	3/3	Hypothesis testing - type I and type II errors, p-values Hypothesis testing - power, common tests Hypothesis testing in R	10.3 - 10.7 10.7 - 10.9
R	3/5		
F	3/6		
Week 9			
T	3/10	Review Exam 2 Go over exam 2	
R	3/12		
F	3/13		
SPRING BREAK			

Week 10			
T	3/24	Simulation-based hypothesis testing	
R	3/26	Simulation-based hypothesis testing	
F	3/27	Simulation-based testing activity	
Week 11			
T	3/31	Introduction to statistical modeling	11.1-11.2
R	4/2	Estimation in statistical models	11.3-11.4
F	4/3	Fitting linear models in R	
Week 12			
T	4/7	Parameter inference in linear models	11.5 - 11.7
R	4/9	Model comparison	
F	4/10	Correlation activity	
Week 13			
T	4/14	Project 2 presentations	
R	4/16	Project 2 presentations	
F	4/17	Problem solving session	
Week 14			
T	4/21	Case studies: data analytics	
R	4/22	Case studies: data analytics	
F	4/23	Review for final exam	

**FINAL EXAM:** Monday, May 4 8:00 AM - 9:45 AM