

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

Instructor: Dr. Kubatko

Office: 219 Cockins Hall

Office Hours: WF 9:30-10:30am, other times by appointment

Office Phone: 614-247-8846

E-mail: kubatko.2@osu.edu

Course Meeting Times: Lecture WF 8:00-9:20am, 375 Journalism Building; Recitations M, times vary (4 contact hours per week)

Prerequisite: Stat 3201: Introduction to Probability for Data Analytics

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

Note that this edition of the textbook is **NOT the same as the ISE version!

Supplemental Online Text: *Introduction to Probability and Statistics Using R*, 2nd edition, by G. Jay Kerns, available at <https://github.com/gjkerns/IPSUR>.

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 Canvas periodically for announcements about the class and other class material.

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 Canvas, 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.

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 each exam. No notes are permitted to be used on quizzes.

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

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. Office hours should not be used for instruction on material that has already been covered in class.

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%	

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

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: 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.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: 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; 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 February 3. The last day to drop the course without petitioning is March 24.

LIST OF TOPICS AND READING

Topic	Title	Reading
T0	Introduction to Statistical Inference	WMS Ch. 7, Sec. 8.1 IPSUR Ch. 8
T1	Point Estimation I	WMS Sec. 8.2 - 8.3
T2	Point Estimation II	WMS Sec. 9.3 - 9.5
T3	Point Estimation III	WMS Sec. 9.6 - 9.7 IPSUR Sec. 9.1
T4	Confidence Intervals	WMS Sec. 8.5 - 8.9 IPSUR Sec. 9.2, 9.4, 9.5, 9.7
T5	Bootstrap	IPSUR Sec. 13.1-13.3
T6	Hypothesis Testing	WMS Sec. 10.1 - 10.9 IPSUR Ch. 10
T7	Nonparametric Testing	WMS Sec. 15.1 - 15.3
T8	Linear Statistical Models I	WMS Sec. 11.1 - 11.3 IPSUR Sec. 11.1 - 11.2
T9	Linear Statistical Models II	WMS 11.4 - 11.7
T10	Simple Linear Regression Examples	IPSUR Sec. 11.3 - 11.4
T11	Correlation	IPSUR Sec. 11.3
T12	Bayesian Inference	WMS Ch. 16

LIST OF LAB TOPICS AND REPORT DUE DATES

Lab	Title	Due date
1	Introduction to R and RStudio	1/13*
2	Point Estimation and Sampling Distributions	1/30
3	Confidence Intervals for Proportions (M&M lab)	2/13*
4	Bootstrap Standard Errors and Confidence Intervals	2/27
5	Power for Hypothesis Tests of a Population Mean	3/10
6	Comparison of Methods for Testing Hypotheses	4/3
7	Fitting Linear Models in R	4/3*

* Lab can be completed in class and turned in at the end of the period.

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				
M	1/9	Lab 1: Introduction to R and RStudio Intro to statistical inference, point estimation Point estimation – bias, MSE	Ch. 7, Sec. 8.1 8.2	
W	1/11			T0
F	1/13			T1
Week 2				
M	1/16	MLK Holiday	No class	
W	1/18	Point estimation – consistency	9.3	T2
F	1/20	Point estimation – sufficiency, UMVUE QUIZ 1: T1	9.4, 9.5	T2
Week 3				
M	1/23	Lab 2: Point Estimation and Sampling Distributions Point estimation – MOM, MLE MLE QUIZ 2: T2	9.6, 9.7 9.7	
W	1/25			T3
F	1/27			T3
Week 4				
M	1/30	Work on projects		
W	2/1	Review for Exam		
F	2/3	EXAM 1: T0-T3		
Week 5				
M	2/6	Work on projects		
W	2/8	Interval estimation, CI for means	8.6 - 8.9	T4
F	2/10	CI for means, proportions	8.6 - 8.9	T4
Week 6				
M	2/13	Lab 3: Confidence Intervals for Proportions Bootstrap Hypothesis testing QUIZ 3: T4, T5	handout 10.1,10.2	
W	2/15			T5
F	2/17			T6
Week 7				
M	2/20	Lab 4: Bootstrap SEs and CIs		
W	2/22	Project 1 presentations		
F	2/24	Project 1 presentations		
Week 8				
M	2/27	Lab 5: Power for Hypothesis Tests of a Population Mean Hypothesis testing - type I and type II errors, p-values Hypothesis testing - power, common tests	10.3 - 10.7 10.7 - 10.9	
W	3/1			T6
F	3/3			T6
Week 9				
M	3/6	Lab 5 continued		
W	3/8	Nonparametric tests	16.3 - 16.4	T7
F	3/10	Simulation-based tests QUIZ 4: T6	16.3 - 16.4	T7
SPRING BREAK				

Week 10				
M	3/20	Lab 6: Comparison of Methods for Testing Hypotheses		
W	3/22	Review for Exam 2		
F	3/24	EXAM 2: T4-T7		
Week 11				
M	3/27	Lab 6 continued		
W	3/29	Introduction to statistical modeling	11.1-11.2	T8
F	3/31	Estimation in statistical models	11.3-11.4	T8
Week 12				
M	4/3	Lab 7: Fitting Linear Models in R		
W	4/5	Parameter inference in linear models	11.5 - 11.7	T9
F	4/7	Model-fitting examples, diagnostics		T10
		QUIZ 5: T8, T9		
Week 13				
M	4/10	Work on projects		
W	4/12	Correlation		T11
F	4/14	Bayesian inference	Ch. 16	T12
Week 14				
M	4/17	Work on projects		
W	4/19	Project 2 presentations		
F	4/21	Project 2 presentations		
Week 15				
M	4/24	Review For Final		

FINAL EXAM: Friday, April 28, 8:00-9:45am