Syllabus for STAT 3201: Introduction to Probability for Data Analytics

Instructor: Dr. Sebastian Kurtek
Office: 440H Cockins Hall
My Office Hours: M 10:00-11:00AM, WF 3:00-4PM, other times by appointment
TA Office Hours: T 2:00-4:00PM, Cockins Hall 217
E-mail: kurtek.1@stat.osu.edu
Lecture Location: Cockins Hall 240
Lecture Time: MWF 1:50PM-2:45PM

Required text and other course materials: The required textbook for the course is *Mathematical Statistics with Applications (7th edition)* by Wackerly, Mendenhall and Scheaffer. The book is available for purchase at the official University bookstore (ohiostate.bncollege.com) and elsewhere online. The book is available on reserve in the 18th Ave. Library. Students will be required to use R software for statistical computing and graphics. R can be downloaded for free at http://www.r-project.org. Instructions for using this software will be given in class. Also, we will supplement course materials with the freely available book titled *Introduction to Probability and Statistics using R* by Kerns (available at https://cran.r-project.org/web/packages/IPSUR/vignettes/IPSUR.pdf). This book provides many excellent R exercises directly related to course material.

Course description: This course provides an introduction to probability and its role in statistical methods for data analytics. Equal emphasis will be placed on analytical and simulation-based methods for quantifying uncertainty. Approaches to assessing the accuracy of simulation methods will be discussed. Applications of probability and sampling to big data settings will also be given.

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

1. Quantify uncertainty about events using mathematical descriptions of probability.
2. Quantify uncertainty about events using simulation methods.
3. Assess the quality and accuracy of simulation based descriptions of uncertainty.
4. Update a description of uncertainty based on new information.
5. Identify appropriate probability models for experiments/data and summarize expected outcomes from such models.
6. Use correlation and conditional expectation to describe the relationship between two random variables.
7. Quantify uncertainty about summary statistics for large data sets.

Course website: Please visit http://www.carmen.osu.edu/. Check Carmen periodically for announcements about the class and other class material.

Assignments:

1. *Homework*: Homework will be assigned (approximately) weekly, will be due on dates announced in class and will be graded. Assignments will consist of a mix of several

problems selected from the textbook, problems motivated by data analytics applications, and small computer simulation problems.

2. *Project*: Suggested project topics will be provided; group-initiated topics will be allowed with approval of the instructor. Group presentations of the results will be given in class and a written report of the results must be submitted.
3. *Exams*: There will be three in-class exams that cover material from lecture, the assigned readings and homework. The tentative exam dates are provided on the schedule attached to this syllabus. 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 devices are NOT allowed. You may use one 8.5 x 11 inch handwritten sheet of paper (both sides) with formulas.

Note on makeup exams: If you absolutely need a makeup exam and have a valid excuse, please see me 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.

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

Grading Policy: Your final course grade will be based on the following weighting of assessment components: homework = 20%, exams = 20% each, project = 20%. The following rubric will be used for determining final grades: A = 93-100, A- = 90-92.9, B+ = 87-89.9, B = 83-86.9, B- = 80-82.9, C+ = 77-79.9, C = 73-76.9, C- = 70-72.9, D+ = 67-69.9, D = 60-66.9, E = below 60.

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 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/coam.html)

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

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 098 Baker Hall (phone: 292-3307) to coordinate reasonable accommodations for the course. ODS forms must be given to your instructor as early in the semester as possible to be filled out and returned to you.

Note: Except for changes that substantially affect implementation of the evaluation (grading) statement, this syllabus is a guide for the course and is subject to change with advanced notice.

Tentative Course Schedule

Date	Lecture Topic	Textbook Reading
8/24	Orientation/Course Introduction	1.1
8/26	Introduction to R	-
8/29	Characterizing Data Using Numerical and Graphical Summaries	1.2-1.3
8/31	Introduction to Probability	2.1-2.5
9/2	Introduction to Probability and Counting Methods	2.1-2.5, 2.6
9/5	Labor Day	
9/7	Conditional Probability and Independence, Probability Laws	2.7-2.9
9/9	Conditional Probability and Independence, Probability Laws	2.7-2.9
9/12	Bayes Theorem	2.10
9/14	Introduction to Simulation and Monte Carlo (MC) Estimation	-
9/16	Discrete Random Variables and Probability Distributions	3.1-3.2
9/19	Expected Value and Variance	3.3
9/21	Bernoulli and Binomial Distributions	3.4
9/23	Negative Binomial, Geometric and Hypergeometric Distributions	3.5-3.7
9/26	Negative Binomial, Geometric and Hypergeometric Distributions Continued	3.5-3.7
9/28	Poisson Distribution	3.8
9/30	Review for Exam 1	-
10/3	Exam 1	-
10/5	Continuous Random Variables, Density and Distribution Functions	4.1-4.2
10/7	Percentiles and Expected Values	4.3
10/10	Uniform Distribution, Normal Distribution	4.4-4.5
10/12	Uniform Distribution, Normal Distribution	4.4-4.5
10/14	Autumn Break	-
10/17	Gamma, Exponential and Beta Distributions	4.6-4.7
10/19	Simulation via Probability Integral Transform, Simulation of Transformed Random Variables	6.1-6.3

10/21	Simulation via Probability Integral Transform, Simulation of Transformed Random Variables	6.1-6.3
10/24	Joint Distribution of Two Random Variables (Discrete Only)	5.1-5.2
10/26	Conditional Distributions, Expected Value and Variance	5.3, 5.5
10/28	Review for Exam 2	-
10/31	Exam 2	-
11/2	Covariance and Correlation	5.7
11/4	Independent and Identically Distributed Random Variables, Joint Probability Density Function	5.4
11/7	Random Samples and Statistics	2.12, 7.1
11/9	Simulating From Sampling Distribution of a Statistic	7.2
11/11	Veterans Day	
11/14	Law of Large Numbers	-
11/16	Distribution of the Sample Mean	7.2
11/18	Samples from Normal Populations; χ^2 and t Distributions	7.2
11/21	Distribution of the Sample Variance	7.2
11/23	Thanksgiving Break	
11/25	Thanksgiving Break	
11/28	Central Limit Theorem	7.3
11/30	Central Limit Theorem	7.3
12/2	Review for Exam 3	-
12/5	Project Group Presentations	
12/7	Project Group Presentations	
12/14 2:00PM-3:45PM	Exam 3	