Sta 101: Data Analysis and Statistical Inference   Spring 2015

Professor: Dr. Mine Çetinkaya-Rundel - mine@stat.duke.edu
                        Old Chemistry 213

Teaching   Radhika Anand
Assistants: Haosheng Luo
                        Jialiang Mao
                        Anthony Weishampel

Lecture: Mon and Wed, 1:25 pm - 2:40 pm, Bio Sci 111
Lab: Tues at Old Chem 101
                        10:05-11:20am, 11:45am-1pm, 1:25-2:40pm, 3:05-4:20pm, 4:40-5:55pm

Required materials:

Textbook   OpenIntro Statistics, Diez, Barr, Çetinkaya-Rundel
                CreateSpace, 2nd Edition, 2011, ISBN: 978-1478217206
                PDF available for free at `http://www.openintro.org` or paperback copy on
                Amazon.

Clicker   i>clicker2. ISBN: 1429280476, available at the Duke textbook store or for
                slightly cheaper on the i>clicker website or on Amazon. For a list of former
                students selling their used clickers, click here.

Calculator   (Optional) You might need a four function calculator that can do square roots
                for this class. There is no limitation on the type of calculator you can use.

Course website: `http://bit.ly/sta101sp15`

Support:

Office hours   Mon, Tue, Thur 3-4pm
                I'm also available to talk after class and by appointment. You're highly encouraged
                to stop by with any questions or comments about the class, or just to say hi and
                introduce yourself. If you're coming by OH for reviewing that week's problem set, I
                recommend that you attempt all problems prior so that you can come to office hours
                with questions.

SECC   Sunday - Thursday 4pm - 9pm (Old Chemistry 211A)
                The statistics education center has upper level statistics students available to help
                you. For more information and a schedule see `http://stat.duke.edu/courses/`
                `resources-students`.

Exams: Midterm 1: Wed, Feb 18, in class
                        Midterm 2: Wed, Mar 25, in class
                        Final: Sat, May 2 (2-5pm)

Course goals & objectives:

This course introduces students to the discipline of statistics as a science of understanding and analyzing data. Throughout the semester, students will learn how to effectively make use of data in the face of uncertainty: how to collect data, how to analyze data, and how to use data to make inferences and conclusions about real world phenomena.

The course goals are as follows:

1. Recognize the importance of data collection, identify limitations in data collection methods, and determine how they affect the scope of inference.

2. Use statistical software to summarize data numerically and visually, and to perform data analysis.

3. Have a conceptual understanding of the unified nature of statistical inference.

4. Apply estimation and testing methods to analyze single variables or the relationship between two variables in order to understand natural phenomena and make data-based decisions.

5. Model numerical response variables using a single or multiple explanatory variables.

6. Interpret results correctly, effectively, and in context without relying on statistical jargon.

7. Critique data-based claims and evaluate data-based decisions.

8. Complete two research projects: one that focuses on statistical inference and one that focuses on modeling.

Tips for success:

1. Complete the reading before a new unit begins, and then review again after the unit is over.

2. Be an active participant during lectures and labs.

3. Ask questions - during class or office hours, or by email. Ask me, your TAs, and your classmates.

4. Do the problem sets - start early and make sure you attempt and understand all questions.

5. Start your projects early and and allow adequate time to complete them.

6. Give yourself plenty of time time to prepare a good cheat sheet for exams. This requires going through the material and taking the time to review the concepts that you're not comfortable with.

7. Do not procrastinate - don't let a unit go by with unanswered questions as it will just make the following unit's material even more difficult to follow.

Course structure:

The course is divided into seven learning units. For each unit a set of learning objectives and required and suggested readings, videos, etc. will be posted on the course website. You are expected to watch the videos and/or complete the readings and familiarize yourselves with the learning objectives. We will begin the unit with a readiness assessment: 10 multiple choice questions that you answer using your clickers in class. You will then re-take this assessment in teams. The rest of the class time will be split between discussion of the material and application exercises that you'll complete in teams. Slides and other relevant materials for each class and lab will be available on the schedule page before each class. Videos and other relevant prep materials for each unit will be available on the resources page. Within each unit you will complement your learning with problem sets and labs, and wrap up each unit with a performance assessment.

Work load:

You are expected to put in about 4-6 hours of work / week outside of class. Some of you will do well with less time than this, and some of you will need more.

Grading:

Your final grade will be comprised of the following.

Attendance&participation+peer eval - 7.5%	Project 1 - 5%
Problem sets - 10%	Project 2 - 10%
Labs - 10%	Midterm 1 - 10%
Readiness assessments - 10%	Midterm 2 - 10%
Performance assessments - 2.5%	Final - 25%

Grades may be curved at the end of the semester. Cumulative numerical averages of 90 - 100 are guaranteed at least an A-, 80 - 89 at least a B-, and 70 - 79 at least a C-, however the exact ranges for letter grades will be determined after the final exam. The more evidence there is that the class has mastered the material, the more generous the curve will be.

Teams:

To construct highly functional teams of learners, you are asked to complete a short survey as well as a pre-test to gauge your previous exposure to statistics and statistical literacy. If you haven□t yet done so please complete these items as soon as possible.

You will be assigned to teams of 3-5 students based on the results of the survey and the pre-test. Once team assignments have been made there is no option for changing teams, other than under extraordinary circumstances. You will work in these teams during application exercises and team portions of the readiness assessments. In addition, your team members will be your first point of contact in this class.

You are encouraged to study with your team members and other classmates. But remember that anything that is not explicitly a team assignment must be your own work.

Clickers:

Throughout the lectures you will use clickers to both answer conceptual questions and for data collection/class surveys. In order to receive credit for the clicker questions you must register your clicker at http://iclicker.com/support/registeryourclicker. In the Student ID field enter your Net ID, and in the Remote ID field enter the alphanumeric code printed below the barcode on the back of your clicker.

These questions are on material introduced in class that day, and you get credit for clicking in, regardless of whether you have the correct answer. To get credit for the day you must respond to at least 75% of the questions. The objective of these questions is to help make you an active participant and gauge the class□ mastery of the material. Participation data from clickers will become a part of your attendance/participation grade.

You are required to bring your clicker to every lecture and it is your responsibility to show up to class on time. Most importantly, it is your responsibility to come to class. I realize that occasionally you may be

late, forget your clicker, or need to miss class. Up to three unexcused late arrivals or absences will not affect your clicker grade. If one person is simultaneously using two or more clickers, the all owners of the clickers will receive a 0 for an overall clicker grade, and will be reported to the Office of Student Conduct.

Attendance, participation, and peer evaluation:

You are expected to be present at class meeting and actively participate in the discussion. Your attendance and participation during class, as well as your activity on the discussion forum on Sakai will make up a non-insignificant portion of your grade in this class. While I might sometimes call on you during the class discussion, it is your responsibility to be an active participant without being called on.

Throughout the semester you will also be asked to complete a few peer evaluations. These will be used to ensure that all team members contribute to the success of the group and to address any potential issues early on. If you feel that there are issues within your team, you are encouraged to discuss it with your team members and to bring it to my or your TA☐s attention.

Problem sets:

These will be assigned (approximately) weekly on the course webpage and will be comprised of problems from the textbook. Each assignment will list roughly five to seven problems from the book to be turned in for grading, and roughly 10 practice problems. You do not need to turn in the practice problems, and the solutions can be found in the back of the book.

The objective of the problem sets is to help you develop a more in-depth understanding of the material and help you prepare for exams and projects. Grading will be based on completeness as well as accuracy. In order to receive credit you must show all your work. Lowest score will be dropped.

You are welcomed, and encouraged, to work with each other on the problems, but you must turn in your own work. If you copy someone else's work, both parties will receive a 0 for the problem set grade as well as being reported to the Office of Student Conduct. Work submitted on Sakai will be checked for instances of plagiarism prior to being graded.

Submission instructions: You will turn in your problem sets on Sakai using one of the following methods:

– Type your answers in the text box on Sakai and attach any plots/images as separate files, properly named, or

– Attach a PDF (**not** Word, Google Doc, etc.) of your answers If you choose the former, you might want to also maintain a copy of your work in a word processor. No other formats will be accepted. It is your responsibility to follow these instructions. If the TAs cannot view your work, or read your handwriting, you will lose points accordingly.

All assignments will be time stamped and late work will be penalized based on this time stamp (see late work policy below).

Labs:

The objective of the labs is to give you hands on experience with data analysis using modern statistical software. The labs will also provide you with tools that you will need to complete the projects successfully. We will use a statistical analysis package called RStudio, which is a front end for the R statistical language.

In class your TAs will give a brief overview of the lab and learning goals, and guide you through some of the exercises. You will start working on the lab during the class session, but note that the labs are designed to take more than just the class time, so you will meet up with your team at a later time to finish the lab before the due date (which will be the following lab session). In order to get credit for a lab report, you must be in the lab session on the day that lab is introduced. Lowest score will be dropped.

Readiness assessments:

Readiness assessments will be given at the beginning of a unit. These are 10 question multiple choice assessments comprised of conceptual questions addressing the learning objectives of the new unit. You are not expected to master all topics in the unit ahead of time, but you are responsible for completing the reading assignment, understanding how the material fits in the greater framework of the course, and acquire a conceptual understanding of the learning objectives. As described above, you will first take the individual readiness assessment using your clickers, and then re-take the same assessment in teams using scratch off sheets. Your performance on both assessments will factor into your final grade (3/4 individual score, 1/4 team score). In addition, your input during the team portion will factor into your participation grade. Lowest score will be dropped.

Performance assessments:

Performance assessments will be given at the end of a unit. These are very similar to the readiness assessments in format, however you will be taking them outside of class on Sakai. Outstanding performance will require mastery of all topics in the unit. Lowest score will be dropped.

Projects:

The objective of the projects is to give you independent applied research experience using real data and statistical methods. Both projects will be completed in teams.

- Project 1: For a parameter of interest to you, you will describe the relevant data, compute a confidence interval and conduct a hypothesis test, and summarize your findings in a written, fully reproducible, data analysis report.

- Project 2: You will use all (relevant) techniques learned in this class to analyze a dataset provided by me, and share your results in a poster session.

Further details on the projects will be provided as due dates approach. Note that you must complete both projects and score at least 30% of the points on each project in order to pass this class.

Exams:

There will be two midterms and one final in this class. See course info for dates and times of the exams. Exam dates cannot be changed and no make-up exams will be given. If you can□t take the exams on these dates you should drop this class. You can□t pass the class if you do not take the final exam. You are allowed to use one sheet of notes ("cheat sheet□) to the midterm and the final. This sheet must be no larger than 8 1/2 x 11, and must be prepared by you. You may use both sides of the sheet.

Email & Forum (Piazza):

I will regularly send announcements by email, please make sure to check your email daily.

Any non-personal questions related to the material covered in class, problem sets, labs, projects, etc.

should be posted on Piazza forum. Before posting a new question please make sure to check if your question has already been answered. The TAs and myself will be answering questions on the forum daily and all students are expected to answer questions as well. Please use informative titles for your posts.

Note that it is more efficient to answer most statistical questions "in person☐ so make use of OH.

Other learning resources:

Aside from the SEC, your TAs and the professor's office hours, you can also make use of the Academic Resource Center.

Students with disabilities:

Students with disabilities who believe they may need accommodations in this class are encouraged to contact the Student Disability Access Office at (919) 668-1267 as soon as possible to better ensure that such accommodations can be made.

Academic integrity:

Duke University is a community dedicated to scholarship, leadership, and service and to the principles of honesty, fairness, respect, and accountability. Citizens of this community commit to reflect upon and uphold these principles in all academic and non-academic endeavors, and to protect and promote a culture of integrity. Cheating on exams and quizzes, plagiarism on homework assignments and projects, lying about an illness or absence and other forms of academic dishonesty are a breach of trust with classmates and faculty, violate the Duke Community Standard, and will not be tolerated. Such incidences will result in a 0 grade for all parties involved as well as being reported to the Office of Student Conduct. Additionally, there may be penalties to your final class grade. Please review the Duke☐s Academic Dishonesty policies.

Excused Absences:

Students who miss graded work due to a scheduled varsity trip, religious holiday or short-term illness should fill out an online NOVAP, RHoliday or short-term illness form respectively. If you cannot complete an assignment on the due date due to a short-term illness, you have until noon the following day to complete it at no penalty. Then the regular late work policy will kick in. Those with a personal emergency or bereavement should seek a Dean☐s Excuse; check with your academic dean for details.

Policies:

- Late work policy for problem sets and labs reports:

    - next day: lose 30% of points

    - later than next day: lose all points

- Late work policy for projects: 10% off for each day late.

- There will be no make-up for readiness or performance assessments, labs, problem sets, project, or exams. If the midterm exam must be missed, absence must be officially excused in advance, in which case the missing exam score will be imputed using the final exam score. This policy only applies to the midterm. All other missed assessments will receive a grade of 0. The final exam must

be taken at the stated time. You must take the final exam and turn in the project in order to pass this course.

– Regrade requests must be made within 3 days of when the assignment is returned, and must be submitted in writing. These will be honored if points were tallied incorrectly, or if you feel your answer is correct but it was marked wrong. No regrade will be made to alter the number of points deducted for a mistake. There will be no grade changes after the final exam.

– Clickers may not be shared, and the clicker registered to a person may only be used by that person. Failure to abide by this will result in a 0 clicker grade for everyone involved.

– Use of disallowed materials (textbook, class notes, web references, any form of communication with classmates or other persons, etc.) during exams will not be tolerated. This will result in a 0 on the exam for all students involved, possible failure of the course, and will be reported to the Office of Student Conduct. If you have any questions about whether something is or is not allowed, ask me beforehand.

Green Classroom:

This course has achieved Duke's Green Classroom Certification. The certification indicates that the faculty member teaching this course has taken significant steps to green the delivery of this course. Your faculty member has completed a checklist indicating their common practices in areas of this course that have an environmental impact, such as paper and energy consumption. Some common practices implemented by faculty to reduce the environmental impact of their course include allowing electronic submission of assignments, providing online readings and turning off lights and electronics in the classroom when they are not in use. The eco-friendly aspects of course delivery may vary by faculty, by course and throughout the semester. Learn more at http://sustainability.duke.edu/action/certifications/classroom/index.php.

It is easy to lie with statistics. It is hard to tell the truth without it.
Andrejs Dunkels