Sta 101.002: Data Analysis and Statistical Inference

Classroom: Social Sciences 136
Time: Tuesdays and Thursdays 2:50pm - 4:05pm

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

Teaching Assistants:
Monika Jingchen Hu - jh309@stat.duke.edu
Brittany Cohen - bjc0405@gmail.com
Fu Tan - fu.tan@duke.edu

Required materials:
- Textbook: OpenIntro Statistics, Diez, Barr, Çetinkaya-Rundel
  PDF available for free at http://www.openintro.org or paperback copy on Amazon.
- Calculator: You will 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://stat.duke.edu/courses/Fall11/sta101.02

Office hours:
- Professor: Mondays 2:30 pm - 4:30 pm and Thursdays 12:30pm - 2:30pm
  I am also available to talk after class, by appointment, or anytime I am in my office with the door open. You are highly encouraged to stop by with any questions or comments about the class, or just to say hi and introduce yourself. Note that most homework assignments will be due on Tuesday. I recommend that you attempt all homework problems over the weekend so that you can come to my Monday office hours with questions.

  TAs: Sunday - Thursday 4pm - 9pm at the SECC (Old Chemistry 211A)
  For more information and a schedule see http://www.stat.duke.edu/secc

Exams: Midterm 1: Thursday, October 6
Midterm 2: Tuesday, November 8
Final: Tuesday, December 13, 2011, 2:00pm - 5:00pm

Holidays: Monday, October 10 and Tuesday, October 11 - Fall Break
Thursday, November 24 - Thanksgiving
Course goals & objectives:

The overall goal of this class is to introduce you to the discipline of statistics as a science of understanding and analyzing data and not as a branch of mathematics. This class is designed to provide you with the tools you need for solving real world problems using statistics and a better understanding of the process of scientific research and statistical inference.

By the end of this class you should be able to interpret statistical results in context and critique news stories and journal articles that include statistical information. We expect you to be comfortable with concepts such as association and causation, random sampling and random assignment, statistical bias and statistical significance, and to understand and appreciate why real data beats anecdotes.

We plan to achieve these goals by introducing you to the relevant statistical knowledge, teaching you how to use an open source (i.e. free!) statistical software called RStudio to perform data analysis, and having you engage in problem solving, application, analysis, and synthesis of statistical information through homework, labs, quizzes and exams.

Tips for success:

1. Read the assigned sections before a new week begins. And then read again after the lectures.
2. Be an active participant during lectures and labs.
3. Ask questions - during class or office hours, or by email. Ask me, the TAs, and/or your classmates.
4. Do the homework - start early and make sure you attempt and understand all questions.
5. Start your projects early and and allow adequate time for working on 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 week go by with unanswered questions as it will just make the following week’s material even more difficult to follow.

Grading:

Your final grade will be comprised of the following.

<table>
<thead>
<tr>
<th>Component</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clicker questions</td>
<td>10%</td>
</tr>
<tr>
<td>Homework</td>
<td>10%</td>
</tr>
<tr>
<td>Online quizzes</td>
<td>5%</td>
</tr>
<tr>
<td>Labs</td>
<td>5%</td>
</tr>
<tr>
<td>Project 1</td>
<td>10%</td>
</tr>
<tr>
<td>Project 2</td>
<td>10%</td>
</tr>
<tr>
<td>Midterms (2 x 15%)</td>
<td>30%</td>
</tr>
<tr>
<td>Final Exam</td>
<td>20%</td>
</tr>
</tbody>
</table>

Cumulative numerical averages of 90 - 100 are guaranteed at least an A-, 80 - 89 at least a B-, and 70 - 79 are at least a C-. These ranges may be lowered, but they will not be raised, e.g., if everyone has averages in the 90s, everyone gets at least an A-. 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.

Work load:

You are expected to put in 6-8 hours of work outside of class. A few of you will do well with less time than this, and a few of you will need more.
Lectures & clicker questions:

Lecture slides will be posted by 8am the day of the lecture. You should print these out and bring them with you to class.

Throughout the lectures you will use clickers to answer mostly conceptual multiple choice questions. In order to receive credit for the clicker questions you must register your clicker at http://iclicker.com/support/registryourclicker. In the Student ID field enter your Net ID, and in the Remote ID field enter the 8-character alphanumeric code printed below the barcode on the back of your remote. Clicker grading will start during the second week of classes and will be equally divided between two types of questions:

- **Review questions**: On material covered in previous classes, and you get credit for answering correctly. The objective of these questions is to motivate you to keep up with the material. Skimming the previous days’ class notes before each lecture is highly recommended.

- **New questions**: On new material introduced in class that day, and you get credit for clicking in, regardless of whether you have the correct answer. The objective of these questions is to help make you an active participant and gauge your understanding of the material.

You are required to bring your clicker and your calculator to every lecture and it is your responsibility to show up to class on time as there will usually be a review question within the first 5 minutes of class. 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 two 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.

Homework:

These will be assigned weekly on the course webpage and will be comprised of problems from the book and the “on your own” part at the end of your weekly lab. The objective of the homework assignments is to help you develop a more in-depth understanding of the material covered in the lectures and help you prepare for exams and projects.

You should generally have sufficient time to complete the lab portion during your lab session on Mondays. Homework assignments will be graded out of 10 points and grading will be based on completeness as well as accuracy. In order to receive credit you must show all your work. You are welcomed, and encouraged, to work with each other on the homework problems, but you must turn in your own work. If you copy someone else’s work, both parties will receive a 0 for the homework grade.

Your homework must be stapled, legible, and contain your name and lab section number and is due at the beginning of class on the due date (see late work policy below). If you cannot make it to class the day homework is due, please email me to make arrangements to drop off your homework earlier. There are a total of 10 homework assignments, and you will be graded out of 9, i.e. lowest homework score will be dropped.

Online quizzes:

These weekly quizzes are designed to help you find any problem areas, and to help me judge how to pace the course. You will have 1 hour to complete each quiz and you must take the quizzes by yourself. Quizzes will be available on Fridays from 8 am to Mondays at 8 am and will cover the previous and the coming week’s material.

These quizzes can be found on the OpenIntro website (http://www.openintro.org). You should register for this course on the website immediately using the course number and the access code sent out in an
email. Note that the quiz interface does not allow you to save your quiz and come back to finish it later, so make sure you submit your quiz once you start taking it.

The quizzes will be graded on whether or not you took them, not your actual score on the quiz. However I will be reported some individual statistics, such as how long you took to complete the quiz. If I suspect that you’re simply clicking through the questions and not actually attempting to solve the problems, I will discuss the issue with you and you will lose credit for those quizzes. There are a total of 12 online quizzes (CAOS Pretest + 11 quizzes on the OpenIntro website) and you will be graded out of 10, i.e. you can miss up to two quizzes without any penalty. The CAOS Posttest will be given for extra credit.

As an extra incentive, one question from an online quiz will appear on the final exam.

Labs:

The objectives of the labs are to teach you how to do data analysis using a statistical software and to give you hands on experience with the topics we cover in lecture. We will use a statistical analysis package called RStudio, which is really a front end for a statistical language called R. You can use RStudio on any computer with a browser, Mac or PC.

To get an RStudio Account, please email Josh Paulson at josh@rstudio.org. Include your full gmail address and reference Professor Mine Çetinkaya-Rundel. You will receive a welcome email once your account has been set up. If you don’t have a gmail account, you can create one at https://www.google.com/accounts/NewAccount.

At the end of each lab session you will be asked to save your workspace to a file and submit it on Sakai (http://sakai.duke.edu) within 15 minutes of the end time of your lab session to get credit for that lab. This means that you must go to the lab section you are enrolled in to receive credit for labs. While the grading is on 0/1 scale (you receive credit if you submit your workspace), we will be doing some random spot checking to make sure you have been following along with the lab.

The “on your own” part of the labs will be turned in with the week’s homework. These will be graded for accuracy.

Projects:

- Project 1 will be done individually. Proposals are due Thursday, October 13, and the project is due on Thursday, November 3. For a parameter of interest to you, you will describe the relevant data, compute a confidence interval (using both bootstrapping and distributional methods), conduct a hypothesis test (using both randomization methods and distributional methods), and summarize your findings in a report up to 5 pages in length.

- Project 2 will be done in groups. Groups should be formed by Monday, November 7 and can only consist of students from your lab section. Proposals are due Thursday, November 17. You will present your project in lab on Monday, December 5 and get some feedback on your work. All group members must be present for the presentation. The write up of the project is due on Thursday, December 8. This will be a regression project on a dataset of interest to you, involving both simple and multiple linear regression. You are also welcomed, but not required, to do make use of ANOVA.

Your projects must be typed, stapled, legible, and contain your name and lab section number and is due at the beginning of class on the due date (see late work policy below).

Exams:

First midterm is on Thursday, October 6 and second midterm is on Tuesday, November 8. Final Exam is a comprehensive 3 hour exam that will be administered on Tuesday, December 13 from 2pm - 5pm.
Exam dates cannot be changed. No make-up exams will be given. If you cannot take the exams on these dates you should drop this class.

You are allowed to bring one sheet of notes (“cheat sheet”) to the midterms and the final. This sheet must be no larger than $8\frac{1}{2} \times 11''$, and must be prepared by you. You may use both sides of the sheet.

Email:

I will regularly send announcements by email, so make sure to check your email daily. While email is the quickest way to reach me outside of class, note that it is much more efficient to answer most statistical questions in person.

Other learning resources:

Aside from the TAs and the professor’s office hours, you can also make use of the Academic Skill Center. For more information, see http://web.duke.edu/arc.

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. Additionally, there may be penalties to your final class grade along with being reported to the Undergraduate Conduct Board.

Please review the Academic Dishonesty policies at http://www.studentaffairs.duke.edu/conduct/resources/academicdishonesty.

Excused Absences:

Students who miss tests 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 and will be given the grade of their Final Exam for those tests.

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 homework assignments and projects:
  - late but during class: lose 20% of points
  - after class on due date: lose 30% of points
  - next day: lose 50% of points
  - later than next day: lose all points

- There will not be make-ups for any of the online quizzes, homework, labs, or exams.

- All regrade requests on homework assignments, labs, and exams must be discussed with the professor within one week of receiving your grade. There will be no grade changes after the final exam.