Data Analysis and Statistical Inference
Introduction

Sta 101 - Fall 2015

Duke University, Department of Statistical Science

Dr. Çetinkaya-Rundel   Slides posted at *http://bit.ly/sta101_f15*

- ► Professor: Dr. Mine Çetinkaya-Rundel - *mine@stat.duke.edu*
- ► TAs:
  - – Erika Ball
  - – David Clancy
  - – Reuben McCreanor
  - – Anne Driscoll
  - – Megan Robertson

- ► OpenIntro Statistics, 3rd Edition
- ► i>clicker2 - See Google Doc for a list of students selling used clickers (link emailed)
- ► (optional) Calculator (just something that can do square roots)

http://bit.ly/sta101_f15

- ▶ Recognize the importance of data collection, identify limitations in data collection methods, and determine how they affect the scope of inference.
- ▶ Use statistical software to summarize data numerically and visually, and to perform data analysis.
- ▶ Have a conceptual understanding of the unified nature of statistical inference.
- ▶ 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.
- ▶ Model numerical response variables using a single or multiple explanatory variables.
- ▶ Interpret results correctly, effectively, and in context without relying on statistical jargon.
- ▶ Critique data-based claims and evaluate data-based decisions.
- ▶ Complete two research projects: one that focuses on statistical inference and one that focuses on modeling.

- ▶ *Unit 1 - Intro to data:* Observational studies & non-causal inference, principles of experimental design & causal inference, exploratory data analysis, introduction to simulation-based statistical inference.
- ▶ *Unit 2 - Probability & distributions:* Basics of probability and chance processes, Bayesian perspective in statistical inference, the normal and binomial distributions.
- ▶ *Unit 3 - Framework for inference:* CLT, sampling distributions, and introduction to theoretical inference.
    - – Midterm 1
- ▶ *Unit 4 - Statistical inference for numerical variables*
- ▶ *Unit 5 - Statistical inference for categorical variables*
    - – Project 1 & Midterm 2
- ▶ *Unit 6 - Simple linear regression:* Bivariate correlation and causality, introduction to modeling.
- ▶ *Unit 7 - Multiple linear regression:* More advanced modeling with multiple predictors.
    - – Project 2 & Final

- ▶ Set of learning objectives and required and suggested readings, videos, etc. for each unit.
- ▶ Prior to beginning the unit, watch the videos and/or complete the readings and familiarize yourselves with the learning objectives.
- ▶ Begin a new unit with a readiness assessment: individual, then team.
- ▶ Class time: split between lecture, discussion/application, and lab.
- ▶ Complement your learning with problem sets.
- ▶ Wrap up a unit with a performance assessment.

- ▶ Highly functional teams of learners based on survey and pre-test.
- ▶ Team members first point of contact.
- ▶ Application exercises, labs, team readiness assessments, projects.
- ▶ Study together, but anything that is not explicitly a team assignment must be your own work.
- ▶ Peer evaluations 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 ASAP (don't wait till things get worse).

Objective: Two-way communication and instant feedback.

► Readiness assessments (graded for accuracy)
► Questions throughout lecture (graded for participation)
  – Get credit for the day you by responding to at least 75% of the questions.
  – Up to three unexcused late arrivals or absences.
► Register your clicker at
  https://www1.iclicker.com/register-clicker (Student ID = Net ID)
  – If you bought a used clicker, the registration process might ask for a payment. You don't need to pay, you'll be able to register in class.

Objective: Make you an active participant and help me pace the class.

► Attendance and participation during class, as well as your activity on Piazza make up a non-insignificant portion of your grade in this class.
► Might sometimes call on you during the class discussion, however it is your responsibility to be an active participant without being called on.

Objective: Help you develop a more in-depth understanding of the material and help you prepare for exams and projects.

► Questions from the textbook.
► Show <u>all</u> your work to receive credit.
► *Submission instructions:*
  – Prepare in a word processor and submit as PDF.
  – Can submit Word documents, but note that if we can't open your file you'll get a 0.
  – You can also type your answers in the text box on Sakai and attach any plots/images as separate files, but if you choose this option make sure to save often (there's no auto-save on Sakai).
► Welcomed and encouraged to work with others, but turn in your own work
► No make-ups, excused absences (e.g. STINF) do not excuse homework
► Lowest PS score will be dropped

Objective: Give you hands on experience with data analysis using statistical software and provide you with tools for the projects.

► Work in teams: author / discussants
► Must be present in lab session to get credit
► Lowest lab score will be dropped

Activity: Get started with R/RStudio

► Go to the course website, *http://bit.ly/sta101_f15*, click on the RStudio link (top right)
  – Make sure you're on the Duke network, not visitor
► Log in using your Net ID and password
► In the Console, generate a random number between 1 and 5, and introduce yourself to that many people sitting around you:
  `sample(1:5, size = 1)`

Objective: Encourage you to watch the videos and/or complete the reading assignment and review the learning objectives prior to coming to class as well as evaluate your conceptual understanding of the unit's material.

► 10 multiple choice questions, at the beginning of a unit
► Conceptual questions addressing the learning objectives of the new unit, assessing familiarity and reasoning, not mastery
► Take the individual RA using clickers, then re-take in teams
► Individual RA score 3/4 of grade, team RA score 1/4 & your input during the team portion will factor into your participation grade
► Lowest RA score will be dropped

Objective: Evaluate your mastery of the material by the end of a unit and give you instant feedback on your performance.

► 10 multiple choice questions, at the end of a unit
► Taken individually on Sakai
► Lowest PA score will be dropped

Objective: Give you independent applied research experience using real data and statistical methods.

► 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: Use all (relevant) techniques learned in this class to analyze a dataset provided by me, and share your results in a poster session
► Must complete both projects and score at least 30% of the points on each project in order to pass this class

Midterm 1	Mon, Oct 5
Midterm 2	Mon, Nov 9
Final	Thur, Dec 10 (2-5pm)

► 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
► Calculator + cheat sheet allowed

- I will regularly send announcements by email, so make sure to check your email daily.
- All content related (non-personal) questions should be posted on Piazza.
- Before posting a new question please make sure to check if your question has already been answered, and answer others' questions.
- Use informative titles for your posts.
- It is more efficient to answer most statistical questions "in person" so make use of OH.

- Prof. Çetinkaya-Rundel: Tue + Thur 4:30-6pm
- TAs: TBA

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.

http://www.access.duke.edu/students/requesting/index.php

Component	Weight
Attendance & participation + peer evaluation	7.5%
Problem sets	10%
Labs	10%
Readiness assessments	10%
Performance assessments	2.5%
Project 1	5%
Project 2	10%
Midterm 1	10%
Midterm 2	10%
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.

- Late work policy for problem sets and labs reports:
  - Next day: lose 30% of points (within 24 hours of due date)
  - Later than next day: lose all points

- Late work policy for projects: 20% off for each day late

Regrade requests must be made *within 3 days* of when the assignment is returned, and must be submitted to me 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.

- No make-up for attendance, individual and team readiness assessments, labs, problem sets, projects, or exams.
- If the midterm exam must be missed due to a documented medical excuse, absence must be officially excused *in advance*, in which case the missing exam score will be imputed using the final exam score.
- The final exam must be taken at the stated time.
- You must take the final exam and earn at least 30% in the projects in order to pass this course.

- 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.

Any form of academic dishonesty will result in an immediate 0 on the given assignment and will be reported to the Office of Student Conduct. Additional penalties may also be assessed if deemed appropriate. If you have any questions about whether something is or is not allowed, ask me beforehand.
Some examples:

► Use of disallowed materials (including any form of communication with classmates or accessing the web) during exams and readiness assessments

► Plagiarism of any kind

► Use of outside answer keys or solution manuals for the homework

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

► Be an active participant during lectures and labs.

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

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

► Take each PA and complete practice quizzes (on Coursera) for each unit, and review the feedback for questions you miss.

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

► 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.

► 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.

► Download or purchase the textbook

► Obtain and register your clicker
   – *https://www1.iclicker.com/register-clicker* (Student ID = Net ID) or in class

► Complete the pretest and the getting to know you survey

► Read the syllabus and let me know if you have any questions

► Watch/Read/Review the resources for Unit 1