

Data Analysis and Statistical Inference

Introduction

Sta 104 - Summer 2015

Duke University, Department of Statistical Science

May 13, 2015

1. Introductions
2. General info
3. Goals
4. Course structure and components
5. Support
6. Policies
7. Tips for success

- ▶ Teaching team:
 - Professor: Dr. Mine Çetinkaya-Rundel
 - TA: Andrew Wong
- ▶ Student teams:
 - Team 1: Alex, Logan, Kaia
 - Team 2: Matthew, Jaidev, Melissa
 - Team 3: Kathleen, Frannie, Albert
 - Team 4: Abhi, Max, Samhita, Caitlin

1. Introductions
2. General info
3. Goals
4. Course structure and components
5. Support
6. Policies
7. Tips for success

- ▶ OpenIntro Statistics, 2nd Edition:
[*https://www.openintro.org/*](https://www.openintro.org/)
- ▶ (optional) Calculator

<http://bit.ly/sta104su15>

Component	Weight
Attendance & participation + peer evaluation	10%
Problem sets	10%
Labs	10%
Readiness assessments	10%
Performance assessments	5%
Project 1	10%
Midterm 1	20%
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.

1. Introductions
2. General info
- 3. Goals**
4. Course structure and components
5. Support
6. Policies
7. Tips for success

- ▶ 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.

- ▶ 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.

- ▶ 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.

- ▶ 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.

- ▶ 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.

- ▶ 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.

- ▶ 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.

- ▶ 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.

1. Introductions
2. General info
3. Goals
4. Course structure and components
5. Support
6. Policies
7. Tips for success

- ▶ 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

- ▶ 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

- ▶ 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

- ▶ 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

- ▶ 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

- ▶ 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
- ▶ Seek help from your team members, but first attempt individual assignments before asking for help
- ▶ 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: 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
- ▶ Up to two excused absences (regardless of reason)

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

- ▶ Due Monday nights
- ▶ Questions from the textbook
- ▶ Show all your work to receive credit
- ▶ Submit on Sakai
 - PDF submission ensures we receive and see what you intended to submit
 - Word submission is less reliable
 - Writing your answers in the text box is acceptable but make sure to save often (Sakai doesn't automatically save for you)
- ▶ Welcomed and encouraged to work with others, but turn in your own work
- ▶ No make-ups, excused absences (e.g. STINF) do not excuse problem sets
- ▶ 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

- ▶ Software: R/RStudio accessed on the web via NetID
- ▶ Due dates on course website
- ▶ No make-ups, excused absences (e.g. STINF) do not excuse labs
- ▶ Lowest lab score will be dropped

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

- ▶ Dates on course website
- ▶ 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 on Sakai, 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.

- ▶ Due dates on course website
- ▶ 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

- ▶ 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
- ▶ Must complete both project and score at least 30% of the points on each project in order to pass this class

Exam	Date
Midterm	May 29, 11am - 12:30pm (in class)
Final	June 24, 11am - 2pm

- ▶ 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
- ▶ If you have a medical excuse on the day of an exam, work with your Dean
- ▶ Calculator + cheat sheet allowed

1. Introductions
2. General info
3. Goals
4. Course structure and components
- 5. Support**
6. Policies
7. Tips for success

- ▶ I will regularly send announcements by email, so 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, 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: Generally TR 12:30 - 1:30pm
- ▶ Andrew: Generally MW 12:30 - 1:30pm
- ▶ Changes will be announced at the beginning of the week
- ▶ This week: No OH today, but I'll stick around to answer questions after class, Andrew will hold office hours tomorrow (Thursday) 12:30-1:30pm, and we will both answer questions on Piazza throughout the week

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>

1. Introductions
2. General info
3. Goals
4. Course structure and components
5. Support
- 6. Policies**
7. Tips for success

- ▶ 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 project: 20% off for each day late

Regrade requests must be made *within 2 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 turn in the projects in order to pass this course

- ▶ 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

1. Introductions
2. General info
3. Goals
4. Course structure and components
5. Support
6. Policies
7. Tips for success

- ▶ Complete the reading before a new unit begins, and then review again after the unit is over.
- ▶ Be an active participant during class.
- ▶ Ask questions - during class or office hours, or by email. Ask me, your TA, and your classmates.
- ▶ Do the problem sets - start early and make sure you attempt and understand all questions.
- ▶ Start your projects early and allow adequate time to complete them.
- ▶ Give yourself plenty of 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.

- ▶ Complete the reading before a new unit begins, and then review again after the unit is over.
- ▶ Be an active participant during class.
- ▶ Ask questions - during class or office hours, or by email. Ask me, your TA, and your classmates.
- ▶ Do the problem sets - start early and make sure you attempt and understand all questions.
- ▶ Start your projects early and allow adequate time to complete them.
- ▶ Give yourself plenty of 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.

- ▶ Complete the reading before a new unit begins, and then review again after the unit is over.
- ▶ Be an active participant during class.
- ▶ Ask questions - during class or office hours, or by email. Ask me, your TA, and your classmates.
- ▶ Do the problem sets - start early and make sure you attempt and understand all questions.
- ▶ Start your projects early and allow adequate time to complete them.
- ▶ Give yourself plenty of 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.

- ▶ Complete the reading before a new unit begins, and then review again after the unit is over.
- ▶ Be an active participant during class.
- ▶ Ask questions - during class or office hours, or by email. Ask me, your TA, and your classmates.
- ▶ Do the problem sets - start early and make sure you attempt and understand all questions.
- ▶ Start your projects early and allow adequate time to complete them.
- ▶ Give yourself plenty of 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.

- ▶ Complete the reading before a new unit begins, and then review again after the unit is over.
- ▶ Be an active participant during class.
- ▶ Ask questions - during class or office hours, or by email. Ask me, your TA, and your classmates.
- ▶ Do the problem sets - start early and make sure you attempt and understand all questions.
- ▶ Start your projects early and allow adequate time to complete them.
- ▶ Give yourself plenty of 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.

- ▶ Complete the reading before a new unit begins, and then review again after the unit is over.
- ▶ Be an active participant during class.
- ▶ Ask questions - during class or office hours, or by email. Ask me, your TA, and your classmates.
- ▶ Do the problem sets - start early and make sure you attempt and understand all questions.
- ▶ Start your projects early and allow adequate time to complete them.
- ▶ Give yourself plenty of 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.

- ▶ Complete the reading before a new unit begins, and then review again after the unit is over.
- ▶ Be an active participant during class.
- ▶ Ask questions - during class or office hours, or by email. Ask me, your TA, and your classmates.
- ▶ Do the problem sets - start early and make sure you attempt and understand all questions.
- ▶ Start your projects early and allow adequate time to complete them.
- ▶ Give yourself plenty of 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.