Syllabus

Course Information

`Title:` BER 540-321 Statistical Methods in Education

`Course level:` Graduate

`Duration:` @August 18, 2021 → December 3, 2021

`When:` Thursdays, 3:00 pm - 5:50 pm

`Where:` Graves Hall 213

`Instructor:` JoonHo Lee

`Office hours:` Sign up here

For the Live Zoom sessions, please use the following link:

> Join our Cloud HD Video Meeting
>
> Zoom is the leader in modern enterprise video communications, with an easy, reliable cloud platform for video and audio conferencing, chat, and webinars across mobile, desktop, and room systems. Zoom Rooms is the original software-based conference room solution used around the world in board, conference, huddle, and training rooms, as well as executive
>
> https://uasystem.zoom.us/j/9503917422

Course Description

🔗 From the 2020-21 UA Academic Catalog:

This three-hour course covers basic descriptive and inferential statistics, including measures of central tendency and dispersion. Hypothesis testing related to one-sample z-and t-test; independent and dependent samples t-test; correlation; and chi-square and simple regression are included. An introduction SPSS is required. Offered fall, spring, and summer semesters.

👨🏻‍🏫 From the instructor:

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.

Course Objectives

After successfully completing this class, students shall be able to:

- 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.

Outline of Topics

The course is divided into eight learning modules:

Learning Module `#1:` Introduction to data

Learning Module `#2:` Summarizing data

Learning Module `#3:` Probability

Learning Module `#4:` Distributions of random variables

Learning Module `#5:` Foundations for inference

Learning Module `#6:` Inference for categorical data

Learning Module `#7:` Inference for numerical data

Learning Module `#8:` Introduction to linear regression

Course Structure & Materials

For each module, you are expected to go through the following steps:

`Step 1:` **View lecture videos;**

- Watch short videos that help you familiarize yourselves with the content covered in the learning module. Don't worry if you don't understand everything at this step. Just watch them.

`Step 2:` **Go through the lecture slides;**

- After watching the videos, go through the slides. At this step, try to understand WHY to have a clearer understanding of the material.

`Step 3:` **Read the textbook;**

- Read the assigned chapter in the OpenIntro textbook. The textbook will supplement many complex concepts covered in Steps 1 and 2. It won't take long to read the chapter because you've already gone over the same content at Steps 1 and 2.

`Step 4:` **Check out the learning objectives;**

- Make sure you know all of the learning objectives provided in a standalone PDF document.

`Step 5:` **Solve the homework assignments;**

- Attend class sessions. Spend class time using your new content knowledge learned from Steps 1-4 to solve problems while the instructor provides support and enrichment.

`Step 6:` **Lab exercise: go through the interactive R tutorials;**

- Only reading about statistics is about as instructive as reading a lot about hammers or watching someone else wield a hammer. You need to get your hands on a hammer or two.
- The interactive R tutorial will allow you to apply the statistical concepts you learn to data analyses with a real-world dataset.
- Complete the R tutorial and submit it via a survey form at the end of the tutorial

All videos, slides, required readings, and other relevant materials for each learning module will be available at the beginning of the semester. They will be hosted on the UA Box folder:

Box

https://alabama.box.com/s/4ner4y44tirdh86n629abt51tnklj4h8

Textbook

Required textbook

The following two textbook is required for this course:

Diez, D., Cetinkaya-Rundel, M., & Barr, C. (2017). *OpenIntro Statistics (4th ed.)*. https://www.openintro.org/book/os/

The textbook is freely available online. You're welcomed to read on screen or print it out. If you prefer a paperback version you can buy it at the cost of printing ($20) on Amazon.

Optional textbook

The following book is optional, but may be helpful to build your understanding of the material:

Ismay, C., & Kim, A. Y. (2019). *Statistical Inference via Data Science: A ModernDive into R and the Tidyverse*. CRC Press. → freely available online as well.

There will also occasionally be additional articles and videos to read and watch. When this happens, links to these other resources will be included on the reading folder for that week.

Why OpenIntro?

Ensuring every student has access

> "65% of students have opted out of buying a college textbook due to its high price", and 94% of those students know they suffer academically

> from that decision.

📌 Ref #1: NY Times - How professors help rip off students: textbooks are too expensive

📌 Ref #2: Why college textbooks are so expensive

📌 Ref #3: Survey show students opting out of buying high-cost textbooks

When using an OpenIntro textbook, every student has immediate, perpetual, and free access to the PDF. The textbook is also available in print for $20 or less.

Grading Policy

You (the student) and I (the instructor) should care the most about what you *learn*, not what numerical/letter summary of that learning you get at the end of the semester. So I would love to not have grades at all, but unfortunately we humans are very good at procrastinating on our good intentions when there is no incentive not to. Thus, we have grades to help solve this commitment problem and to encourage you to put effort into learning the course material.

Students will have the opportunity to earn up to **140 total points** in this course:

Grading Table

Aa Category	# Points can be earned	# Points per each assignment
8 Homework Assignments	48	6
6 Interactive R Tutorials	48	8
Final Exam / Data Analysis Project	32	32
Extra credits (based on class participation, etc.)	12	12

Bump-up policy: Please note there will also be an opportunity to earn up to `12 points` of extra credits at the instructor's discretion. I reserve the right to “bump up” the grades of students who

have made valuable contributions to the course in the session or discussion. This also applies to students who show tremendous progress and growth over the semester.

Grading will be assigned using the following grading scale:

90+ (`A`) ; 80+ (`B`) ; 70+ (`C`) ; 60+(`D`); below 60 (`F`)

Grades may be curved at the end of the semester. Cumulative numerical averages of 90 - 140 are guaranteed 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 or data analysis project. The more evidence there is that the class has mastered the material, the more generous the curve will be.

Expectation & Workload

In this course, you will be expected to

- complete EIGHT weekly or biweekly **homework assignments**,
- complete SIX weekly or biweekly **interactive R tutorials**,
- complete ONE **final take-home exam** or **final data analysis project** (choose either one)

You are expected to put in about 4-8 hours of work per week outside of class. Some of you will do well with less time than this, and some of you will need more.

`Tips for success:` Don't let a learning module go by with unanswered questions as it will just make the following module's material even more difficult to follow.

Homework Assignments

Homework assignments will be assigned weekly and comprised of similar problems from the textbook. Each assignment will list roughly ten to twelve problem sets. The objective of the homework assignment is to help you develop a more in-depth understanding of the material and help you prepare for exams and the project.

While you are expected to complete them on time, **they will be graded based on completion not on how correct the answers are**. That said, going slowly and making sure that you understand each question and its answer is a good practice to ensure you are mastering the course material. You are welcomed, and encouraged, to work with each other on the problems, but you must turn in your own work. The solutions will be uploaded after their due dates.

Submission instructions:

You will turn in your homework assignments in your individual Box folder, which will be assigned to you at the beginning of the semester. Only you, the instructor, and the teaching assistant will have access to the folder. I recommend saving your work as PDF and submitting the PDF. This will ensure that what we read is exactly what you intended to submit. All assignments will be time-stamped and late work will be penalized based on this time stamp.

Lab Exercise: Interactive R Tutorials

In each learning module, I have included the interactive R tutorials — a collection of short video lectures, R coding exercises, and quizzes. The objective of the lab exercises delivered through the tutorials is to give you hands-on experience with data analysis using modern statistical software R. At the beginning of the term, these tutorials will focus on getting up to speed in R, and over the course of the term, they will focus more on the theoretical aspects of data analysis. The R tutorials will be graded based on completion as well.

List of the interactive R tutorials (links in parentheses):

Tutorial `#1:` Introduction to R (https://joonho.shinyapps.io/01-intro-to-R/)

Tutorial `#2:` Introduction to Tidyverse (https://joonho.shinyapps.io/intro-to-tidyverse/)

Tutorial `#3:` Foundations of probability (https://joonho.shinyapps.io/foundations-of-probability/)

Tutorial `#4:` Sampling (https://joonho.shinyapps.io/sampling/)

Tutorial `#5:` Hypothesis testing (https://joonho.shinyapps.io/hypothesis-testing/)

Tutorial `#6:` Correlation and regression (https://joonho.shinyapps.io/correlation-and-regression/)

Submission instructions:

Once you have completed the tutorial and are happy with your solutions, go to the `Submit` section at the end of the tutorial. Then, clicking the `Generate` button should give you the "hash" data. The hash is a compressed text that contains all your progress while solving exercises in the tutorial. You need to copy your hash data to the clipboard and paste it to the answer box of the survey form. The detailed instruction is available at the Box folder (How to submit the completed interactive R tutorials).

Final Take-Home Exam / Final Data Analysis Project

For the final comprehensive assessment, you can choose either (1) the final take-home exam or (2) the final data analysis project.

The final take-home exam will be similar to a homework in format and in the sense that it will be open book and open internet, but you will not be allowed to collaborate with other students or be able to communicate with any humans about the exam. You will be given several days to complete the exam. I will provide more information about the exam as it approaches.

The final project for the course will be a data analysis project where students will find a dataset of interest, state an interesting research question about that data, and answering this question using that data. The objective of the project is to give you independent applied research experience using real data and statistical methods. Project details will be posted later.

Class Sessions

I will hold weekly live class meetings to perform an in-class activity related to the homework assignments. In these sessions, I will usually walk through homework assignments, and you will be expected to follow along. I will also answer any questions to clarify any misconceptions. These meetings will take place on Thursday at 3:00 pm for three hours.

You should generally:

- attend the class sessions on `Thursday` at 3:00 pm; and

- complete and submit the homework assignments and the interactive R tutorials by `Sunday` at 11:59 pm.

Policy on Missed Exams and Coursework

Late submissions will not be accepted. In any emergency circumstances, please get in touch with the instructor via email if you need flexibility on deadlines.

Attendance Policy

The individual Box folder and the interactive R tutorials will automatically monitor your attendance at the weekly learning modules. For the scheduled live class sessions, attendance is encouraged but not required. But you are responsible for completing your own homework assignments and the R tutorials. In the live class sessions, the instructor will help students with the assigned work, such as when they get stuck on a problem or want clarification to some questions.

Statistical computing: Why R?

This course will expose you to R—one of the most popular, sought-after, and in-demand statistical programming languages. Armed with the foundation of R skills you'll learn in this class, you'll know enough to be able to find how to analyze any data-based problem in the future.

Many people seem to believe that R is particularly technically challenging and difficult to master. This probably stems from its extreme flexibility; it is a fully functional programming language as well as a statistical analysis package. R can do things that many other software packages (I'm looking at you, SPSS) essentially can't. These more involved things are frequently hard to do because they require you to think like a programmer rather than a data analyst. As a result, R is perceived as a "hard" language to use.

However, for straightforward and off-the-shelf analyses, R is arguably as easy to learn and use in many ways as programs such as SPSS or Stata. This is especially true when R is combined with RStudio for a better programming experience. An incomplete list of reasons to use R instead of SPSS or Stata are as follows:

- R is free. This might not matter much now, but it could matter depending on what future career you pick.
- R is a very marketable skill. Stata is used in certain organizations, and SPSS is primarily used for instructional purposes. But R is quickly becoming a language of choice for research

organizations around the world.

- You can analyze multiple data sets of different formats and types at once. This allows for rapid exploration. Using a flexible statistical environment is also mind-expanding.
- R has many more packages than SPSS or Stata, and the packages it has usually work better/do more. This means that you will have access to the cutting-edge methods being developed, which gives you greater control over your data.
- Working in R develops better programming habits and relies less on the hacky solutions other statistical software encourages.
- R works a lot better with automatic report generation. Report generation is an excellent way for coupling a data analysis with the write-up, which makes reproducibility a snap. This is an increasingly important issue in the sciences.

NOTE: If you want to learn SPSS rather than R, this course is not the right one for you.