THE OHIO STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

SYLLABUS: GRADTDA 5401
DATA ANALYTICS FOUNDATIONS 1
AUTUMN 2024

Course overview

Instructor

Instructor: Scott Linder
Email address: linder.5@osu.edu
Lectures: Asynchronous video lectures shared via Carmen (1 hour 50 minutes per week) and Synchronously on Zoom **Tuesdays 6:00–6:55 PM EST**
Office hours: Virtual Hours via Zoom (Tuesdays 7pm–8pm, Thursdays 4pm–5pm)

Course Website

Carmen (https://carmen.osu.edu/). All course related materials will be posted here.

Teaching Assistant

Yuxuan Xin (xin.155)
TA office hours: Wednesdays 10am–11am

Course description

This course is the first of a sequence of two courses designed to teach you statistical analysis and data science using R. The first part of the course will focus more on visualizing data, performing exploratory statistical analysis and programming with R, while the second part will introduce some core concepts of statistics. The course will apply a hands-on approach through which you will learn by performing data analysis yourself.

Tentative list of topics we will cover:

1. Introduction to R and data science
    a. Basic R programming, Visualization in base R and ggplot
    b. Data transformations and manipulations
    c. Apply, for loops, iterations
    d. Tibbles, data frames, importing data, merging multiple data tables
    e. Advanced R programming – custom functions
2. Basic statistical inference
    a. Graphical and numerical summaries of data
    b. Introduction to probability, conditional probability and Bayes theorem
    c. Discrete and continuous random variables, probability distributions
    d. Simulation using R
    e. Population, samples, sampling distributions, confidence intervals
3. Regression Analysis
    a. Simple linear regression
    b. Multiple linear regression
    c. Logistic regression

Prerequisite or corequisite:

Students must be graduate students enrolled in the MTDA program.

Course learning outcomes

Upon successful completion of the course, students will be able to

1. Use the statistical programming language R (with Rstudio) to import, clean, and visualize data
2. Generate and interpret graphical and numerical summaries from data
3. Perform exploratory analyses of a data set and begin perform more in-depth statistical analyses (modeling and inference) on data. This includes estimation of population parameters from information in a sample. It also includes the quantification of uncertainty in these estimates through confidence intervals and significance tests.
4. Use regression analysis to understand relationships between two or more variables.
5. Perform programming tasks using R

Course materials

Printed and recorded Materials

We have two required books for the course. Both the books are available to students freely.

- *The R for data science book: https://r4ds.hadley.nz/*
- *Introductory Statistics book: https://openstax.org/details/books/introductory-statistics*

Additional references (not required). These two books are freely available from their websites:

- *Hands on Programming with R* by *Garrett Grolemund* https://rstudio-education.github.io/hopr/index.html
- *R programming for Data Science by Roger D Peng* https://bookdown.org/rdpeng/rprogdatascience/.

Of course, the instructor will post lecture notes and lecture recordings on the Carmen site. Students are expected to watch and understand these materials as a primary responsibility.

Software

- This class requires you to use the statistical software package called R (The R Project for Statistical Computing; http://www.r-project.org/). This software package is available as Free Software.
    - You can download R for Windows, Mac, and Linux, from the CRAN archive at https://cran.r-project.org.
    - An in-depth introduction to R is available at http://cran.r-project.org/doc/manuals/R-intro.pdf
    - You are not expected to know anything about R at the start.
- An easier to use interface to R is available in the software package RStudio. This package is available for Windows, Mac, and Linux and can be downloaded for free from http://rstudio.org. **Note that RStudio operates "over" R, so it requires R to be installed.**

Computer Hardware

You need a reasonably up-to-date computer (OSX or Windows) with a high-speed internet connection, a working web-cam and a working microphone.

Basic Skills

You will be asked to create slides and record a presentation of them using audio narration. You'll also be asked to create pdf files and upload them to Carmen. You'll also need to be able to use Zoom.

Course delivery

The course will be **delivered online in two modes**. (1) Each week approximately two hours of recorded lectures will be posted to the Carmen page. These will typically be broken up into smaller recordings so that they are more manageable. (2) We will also have a 55 minutes synchronous (live) lecture over zoom during the lecture time listed above. The asynchronous content will be posted as videos on the class website in carmen.

You are responsible for viewing and focusing on a week's posted recorded lectures some time before our class meets for the following live (synchronous) meeting. During the live meeting we will have time for questions and will work through examples that apply the material from the recordings.

For the live lectures, attendance is encouraged, but is not required. The live lecture will be recorded and posted in carmen for viewing later.

In addition to the lecture videos, assignments will be posted on the class website. You will be given ample time to complete assignments. The instructor will hold twice weekly office hours via Zoom. The times are given above.

Course grades

Your course average will be computed using the following category weights:

Assignment or category	Percentage
Homework Assignment Average	**40**
Project 1	**20**
Project 2	**20**
Take home Exam	**20**
Total	**100**

Your course average will be rounded up to the nearest tenth of a percentage point, and then the following "minimum grade" scale will be applied to determine your course grade

93–100	A	77 – 79.9	C+
90–92.9	A-	73 – 76.9	C
87–89.9	B+	70 – 72.9	C–
83–86.9	B	67 – 69.9	D+
80–82.9	B-	60 – 66.9	D
		< 60	E

Assignment information

Homework: Homework assignments will be assigned regularly (about 5-6 assignments over the semester). These will consist of data analysis and computer programming/simulation problems. You may discuss the problems with your fellow classmates, tutors/TA, and with the instructor, but each student must hand in their own work, written in their own words. Do not copy any part of another student's homework including computer output. Use of homework solutions distributed in previous offerings of the course or available on the web constitutes academic misconduct and will be handled according to university rules. **All homework must be typed and submitted online as a PDF file through the class website (Carmen).** Please be sure that the questions are clearly labeled, all supporting work (including computer code) can be easily identified, and that all figures/tables are referenced and interpreted in the text.

Please turn assignments in on time. Prior instructor approval is required for late submissions. No assignment will be accepted more than 72 hours after it is due because assignment solutions need to be posted for the class. Your lowest assignment score will be dropped.

Projects: In addition to homework assignments, there will be two group projects assigned. Each project will ask you to apply analysis methods studied in the course to a data set of your selection using R. For each project your group will write a report and record an oral presentation of results obtained. Students will work in groups of 1–2 as assigned by the instructor. There will be time to request work with a student you prefer to work with. Information about due dates, project guidelines and group selection is provided on the Carmen page.

Exams: There will be one take home exam. Coverage includes lecture material, assigned reading, and homework. Further details will be given in advance of the take home exam. The format of the exam is similar to other homework assignments and consists of data analysis problems. The rules regarding academic integrity are different from those of other homework assignments: You are not allowed to discuss the problems with your fellow classmates. You may discuss the exam with the instructor, of course.

Take home exam period: 11.25 – 12.8.

Faculty feedback and response time

The following is intended to give you an idea of instructor availability throughout the course. (Remember that you can call **614-688-HELP** at any time if you have a technical problem.)

Grading and feedback

For weekly assignments, you can generally expect graded homework back within **14 days**.

E-mail

The most important rule is: Always send email to the instructor **through the Carmen page**. If you do that, you'll hear back from the instructor quickly – often the same day, usually within 24 hours, and always as soon as possible. If you don't send email through the Carmen page, you risk not having your message seen for several days.

Assignment help by email is not prohibited, but it is not favored. Instead, if you have questions on understanding the course materials or regarding the homework assignments please visit Zoom office hours for the instructor or the TA.

Student participation requirements

Because this is a distance-education course, your attendance is based on your online activity and participation. The following is a summary of everyone's expected participation:

- **Logging in**: **AT LEAST ONCE PER WEEK**
  Be sure you are logging in to the course in Carmen each week, including weeks with holidays or weeks with minimal online course activity. (During most weeks you will probably log in many times.) If you have a situation that might cause you to miss an entire week of class, discuss it with the instructor *as soon as possible.*
- **In live lectures: REGULAR ATTENDANCE ENCOURAGED BUT NOT REQUIRED.** Students are encouraged to participate, discuss, and answer questions in online live lectures.
- **Recorded video lectures: REQUIRED.** Students are expected to watch the video lectures I will upload regularly at their own time.
- **Office hours**: **OPTIONAL OR FLEXIBLE**
  All office hours, are optional. If you are required to discuss an assignment with me, please contact me at the beginning of the week if you need a time outside my scheduled office hours.

Other information

Course technology

For help with your password, university e-mail, Carmen, or any other technology issues, questions, or requests, contact the OSU IT Service Desk. Standard support hours are available at https://ocio.osu.edu/help/hours, and support for urgent issues is available 24x7.

- **Self-Service and Chat support:** http://ocio.osu.edu/selfservice
- **Phone:** 614-688-HELP (4357)
- **Email:** 8help@osu.edu
- **TDD:** 614-688-8743

Potential disruptions to instruction

- As much as is possible, students will have access to material online if they are unable to attend class because of positive diagnosis, symptoms, or quarantine required following contact tracing.
- If the instructor is unable to be present in person because of positive diagnosis, symptoms, or quarantine following contact tracing a new instructor will be assigned to the course. Details will be given on the course website

Student academic services

Student academic services offered on the OSU main campus
http://advising.osu.edu/welcome.shtml.

Student support services

Student support services offered on the OSU main campus http://ssc.osu.edu.

Academic integrity policy

Policies for this online course

- **Exam**: You must complete the exam yourself, without any external help or communication with anybody apart from the instructor.
- **Written assignments**: Your written assignments should be your own original work. You are permitted to work together in assignments with fellow students as long as you write your solutions yourself and as long as your work is essentially your own.
- **Reusing past work**: In general, you are prohibited in university courses from turning in work from a past class to your current class, even if you modify it. If you want to build on past research or revisit a topic you've explored in previous courses, please discuss the situation with me. This rule is relevant to our projects: You cannot use project work from another course for this course.
- **Falsifying research or results**: All research you will conduct in this course is intended to be a learning experience; you should never feel tempted to make your results or your research look more successful than it was.
- **Collaboration and informal peer-review**: The course includes many opportunities for formal collaboration with your classmates. The group projects give you opportunity to work together in a group on a project. You are also encouraged to form study groups and peer-review for the projects. While group study for assignments is permitted, remember that your answers on homework assignments should not be an identical copy of another students' answers. If you're unsure about a particular situation, please feel free just to ask ahead of time.

Ohio State's academic integrity policy

It is the responsibility of the Committee on Academic Misconduct to investigate or establish procedures for the investigation of all reported cases of student academic misconduct. The term "academic misconduct" includes all forms of student academic misconduct wherever committed; illustrated by, but not limited to, cases of plagiarism and dishonest practices in connection with examinations. Instructors shall report all instances of alleged academic misconduct to the committee (Faculty Rule 3335-5-487). For additional information, see the Code of Student Conduct http://studentlife.osu.edu/csc/.

Copyright disclaimer

The materials used in connection with this course may be subject to copyright protection and are only for the use of students officially enrolled in the course for the educational purposes associated with the course. Copyright law must be considered before copying, retaining, or disseminating materials outside of the course.

Statement on title IX

Title IX makes it clear that violence and harassment based on sex and gender are Civil Rights offenses subject to the same kinds of accountability and the same kinds of support applied to offenses against other protected categories (e.g., race). If you or someone you know has been sexually harassed or assaulted, you may find the appropriate resources at http://titleix.osu.edu or by contacting the Ohio State Title IX Coordinator at titleix@osu.edu

Accessibility accommodations for students with disabilities

The university strives to make all learning experiences as accessible as possible. In light of the current pandemic, students seeking to request COVID-related accommodations may do so through the university's request process, managed by Student Life Disability Services. If you anticipate or experience academic barriers based on your disability (including mental health, chronic, or temporary medical conditions), please let me know immediately so that we can privately discuss options. To establish reasonable accommodations, I may request that you register with Student Life Disability Services. After registration, make arrangements with me as soon as possible to discuss your accommodations so that they may be implemented in a timely fashion. SLDS contact information: slds@osu.edu; 614-292-3307; http://slds.osu.edu; 098 Baker Hall, 113 W. 12th Avenue.

Accessibility of course technology

This online course requires use of Carmen (Ohio State's learning management system) and other online communication and multimedia tools. If you need additional services to use these technologies, please request accommodations with your instructor.

- Carmen (Canvas) accessibility
- Streaming audio and video
- Synchronous course tools

Your mental health

As a student you may experience a range of issues that can cause barriers to learning, such as strained relationships, increased anxiety, alcohol/drug problems, feeling down, difficulty concentrating and/or lack of motivation. These mental health concerns or stressful events may lead to diminished academic performance or reduce a student's ability to participate in daily activities. The Ohio State University offers services to assist you with addressing these and other concerns you may be experiencing. If you or someone you know are suffering from any of the aforementioned conditions, you can learn more about the broad range of confidential mental health services available on campus via the Office of Student Life's Counseling and Consultation Service (CCS) by visiting ccs.osu.edu or calling 614- 292-5766. CCS is located on the 4th Floor of the Younkin Success Center and 10th Floor of Lincoln Tower. You can reach an on call counselor when CCS is closed at 614-292-5766 and 24 hour emergency help is also available through the 24/7 National Suicide Prevention Hotline at 1-800-273- TALK or at suicidepreventionlifeline.org

Disclaimer

This syllabus should be taken as a fairly reliable guide for the course content. However, you cannot claim any rights from it and in particular we reserve the right to change due dates or the methods of grading and/or assessment if necessary. Any changes will be communicated to you through official course announcements.

Course schedule (tentative)

The dates indicate the week over which the material will be delivered while the last date is the live lecture date.

Week	Dates	Topics, Readings, Assignments, Deadlines
1	Aug 20 - Aug 27	Introduction to R and base R programming
2	Aug 28 - Sep 3	Types of data, Graphical and numerical summaries of data
3	Sep 4 - Sep 10	Data transformation using R tidyverse/dplyr
4	Sep 11 - Sep17	Visualization using R ggplot
5	Sep 18 - Sep 24	Exploratory data analysis in R and data import
6	Sep 25 - Oct 1	R topics: import and tidy messy data, strings
7	Oct 2 - Oct 8	R topics: merging multiple data tables
8	Oct 9 - Oct 15	Sample correlation, Autumn break
9	Oct 16 - Oct 22	Introduction to probability, methods for counting
10	Oct 23 - Oct 29	Conditional probability, Bayes theorem
11	Oct 30- Nov 5	Discrete random variables, expectations, simulation
12	Nov 6 - Nov 12	Advanced programming in R – loops, apply, custom functions
13	Nov 13 - Nov 19	Simple Linear regression – model, estimation, inference on parameters
14	Nov 20 – Nov 26	Simple Linear regression – diagnostics and remedies, inference on model
15	Nov 27- Dec 3	Multiple linear regression