

Syllabus

Data Science for Studying Language & the Mind, Fall 2024

Course description

Data Sci for Lang & Mind is an entry-level course designed to teach basic principles of statistics and data science to students with little or no background in statistics or computer science. Students will learn to identify patterns in data using visualizations and descriptive statistics; make predictions from data using machine learning and optimization; and quantify the certainty of their predictions using statistical models. This course aims to help students build a foundation of critical thinking and computational skills that will allow them to work with data in all fields related to the study of the mind (e.g. linguistics, psychology, philosophy, cognitive science, neuroscience).

There are **no prerequisites beyond high school algebra**. No prior programming or statistics experience is necessary, though you will still enjoy this course if you already have a little. Students who have taken several computer science or statistics classes should look for a more advanced course.

People

- **Instructor:** [Dr. Katie Schuler](#)
- **TAs:** TBD

Lectures

Tuesdays and Thursdays at 10:15am in TBD.

Labs

Hands-on practice, quiz prep, and problem set work guided by TAs.

- 402: Thu at 1:45p in TBD
- 403: Thu at 3:30p in TBD
- 404: Fri at 10:15a in TBD
- 405: Fri at 12:00p in TBD

Office hours

Come see us during office hours! We are here to help.

When	Who	Where
TBD	Katie	Zoom

Problem sets

There are 6 problem sets, due on Mondays to Gradescope by 11:59pm. You may request an extension of up to 3 days for any reason. After solutions are posted, late problem sets can be submitted for half credit (50%). We will drop your lowest problem set.

- Problem set 1 due Sep 9
- Problem set 2 due Sep 23
- Problem set 3 due Oct 14
- Problem set 4 due Oct 28
- Problem set 5 due Nov 11
- Problem set 6 due Dec 9

Exams

There are 2 midterm exams, taken in class on Thursdays. Missed exams cannot be made up except in cases of genuine conflict or emergency (documentation and a [Course Action Notice](#) are required). After solutions are posted, missed exams may be submitted for half credit (50%). You may replace your lowest exam score with the optional final exam.

- Exam 1 in class Oct 1
- Exam 2 in class Nov 21
- Final exam (optional) TBD

Lab exercises

Lab exercises are intended for practice and are not graded.

- Lab 1 on Aug 29-30
- Lab 2 on Sep 5-6
- Lab 3 on Sep 12-13
- Lab 4 on Sep 19-20
- Lab 5 on Oct 10-11
- Lab 6 on Oct 17-18
- Lab 7 on Oct 24-25
- Lab 8 on Nov 1-2
- Lab 9 on Nov 8-9
- Lab 10 on Dec 5-6

Grading

- 40% problem sets (equally weighted, lowest dropped)
- 60% exams (equally weighted, final is optional to replace lowest exam)
- Letter grade minimums: 97% A+, 90% A, 80% B, 70% C, 61% D, else F

Collaborations

Collaboration on problem sets is highly encouraged! If you collaborate, you need to write your own code/solutions, name your collaborators, and cite any outside sources you consulted (you don't need to cite the course material).

Accomodations

We will support any accommodations arranged through Disability Services via the [Weingarten Center](#).

Extra credit

There is no extra credit in the course. However, students can submit any missed problem set or quiz by the end of the semester for half credit (50%). To ensure fair treatment across all students, all students will receive a 1% “bonus” to their final course grade: 92.54% will become 93.54%.

Regrade requests

Regrade requests should be submitted through Gradescope within one week of receiving your graded assignment. Please explain why you believe there was a grading mistake, given the posted solutions and rubric

Resources

In addition to our course website, we will use the following:

- [google colab \(r kernel\)](#) - for computing
- canvas- for posting grades
- gradescope - for submitting problem sets
- ed discussion - for announcements and questions

Please consider using these Penn resources this semester:

- [Weingarten Center](#) for academic support and tutoring.
- [Wellness at Penn](#) for health and wellbeing.

Schedule

Subject to change

	Week of	Topic	Lab	Due this week
1	Aug 26	R Basics	Lab 1	
2	Sep 2	Data import and tidy	Lab 2	
3	Sep 9	Data visualization	Lab 3	Problem set 1
4	Sep 16	Sampling distribution	Lab 4	
5	Sep 23	Hypothesis testing	Exam 1 review	Problem set 2
6	Sep 30	Exam 1 Fall break begins Thursday		
7	Oct 7	Model specification	Lab 5	
8	Oct 14	Model fitting	Lab 6	Problem set 3
9	Oct 21	Model accuracy	Lab 7	
10	Oct 28	Model reliability	Lab 8	Problem set 4
11	Nov 4	Classification	Lab 9	
12	Nov 11	Inference	Exam 2 review	Problem set 5
13	Nov 18	Exam 2		
14	Nov 25	Thanksgiving break (no class this week)		
15	Dec 2	Multilevel models	Lab 10	
16	Dec 9	Last day of classes (no class)		Problem set 6
17	TBD	Final exam (optional)		