

Data Science for studying language and the mind

Fall 2024

Announcements

- No Wesley lab on THurs Nov 14! Brittany's still on!
- Exam 2 next Thursday! Nov 21!

Welcome to Data Science for Studying Language & the Mind! The Fall 2024 course information and materials are below. Course materials from previous semesters are archived [here](#).

1 Syllabus

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).

Prerequisites: 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.

Instructor: [Dr. Katie Schuler \(she/her\)](#)

TAs: [Brittany Zykoski](#) and [Wesley Lincoln](#)

Lectures: Tuesdays and Thursdays from 12 - 1:29pm in COHN 402.

Labs: Hands-on practice and exam prep guided by TAs.

- 402: Fri at 3:30p in WILL 201 with Brittany
- 403: Thu at 1:45p in WILL 321 with Wesley
- 404: ~~Thu at 3:30p~~ **CANCELLED**
- 405: Fri at 12:00p in ~~WILL 316~~ **WILL 306** with Wesley
- 406: Thu at 5:15p in TBD with Brittany

Office Hours: You are welcome to attend any office hours that fit your schedule. The linguistics department is located on the 3rd floor of 3401-C Walnut street, between Franklin's Table and Modern Eye.

- Katie Schuler: 11:30-12:30 on Fridays in 314C
- Brittany Zykoski: 10:30-11:30 on Wednesdays in 325C
- Wesley Lincoln: 2-3 on Mondays in 325C

Grading:

- 40% Homework (equally weighted, lowest dropped)
- 60% exams (equally weighted, final is optional to replace lowest exam)

Collaboration: 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).

Accommodations: We will support any accommodations arranged through Disability Services via the [Weingarten Center](#). Please make arrangements as soon as possible (1-2 weeks in advance).

Extra credit: There is no extra credit in the course. However, students can submit any missed problem set or exam by the end of the semester for half credit (50%). To ensure fair treatment, 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

2 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

Other helpful materials and resources:

- [markdown tutorial](#)
- [tidyverse cheatsheets](#)

Please consider using these Penn resources this semester:

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

3 Materials

3.1 Study guides

Study guides include weekly study guides and additional resources from each week, including slides, demos, and further reading.

- [Week 1: R Basics](#)
- [Week 2: Data visualization](#)
- [Week 3: Data import, tidy, wrangle](#)
- [Week 4: Sampling distribution](#)
- [Week 5: Hypothesis testing](#)
- [Week 6: Exam 1 review](#) ([practice exam](#), [solutions](#))
- [Week 7: Model specification](#) ([slides](#), [demo](#))
- [Week 8: Applied model specification](#) ([slides](#), [demo](#))
- [Week 9: Model fitting](#) ([slides](#), [demo](#))
- [Week 10: Model accuracy](#) ([slides](#), [demo](#))
- [Week 11: Model reliability](#) ([slides](#), [demo](#))
- [Week 12: Classification](#) ([slides](#), [demo](#))
- [Week 13: Exam 2 review](#) ([slido](#), [practice exam](#), [solutions](#))
- [Week 15: Multilevel Models](#) ([slides](#), [demo](#))

3.2 Problem sets

There are 6 problem sets, due to Gradescope by noon on the following Mondays. You may request an extension of up to 3 days for any reason. After solutions are posted, late problem sets can still be submitted for half credit (50%). If you submit all 6 problem sets, we will drop your lowest.

- [Problem set 1 due Sep 9, solutions](#)
- [Problem set 2 due Sep 23, solutions](#)
- [Problem set 3 due ~~Oct 14~~ Oct 21, solutions](#)

- [Problem set 4 due ~~Oct 28~~ Nov 4](#)
- [Problem set 5 due Nov 18](#) (cancelled due to missed week)
- Problem set 6 due Dec 9

3.3 Exams

There are 2 midterm exams, taken in class on the following dates. Exams cannot be rescheduled, except in cases of genuine conflict or emergency (documentation and a [Course Action Notice](#) are required). However, you can submit any missed exam by the end of the semester for half credit (50%). You may also replace your lowest midterm exam score with the optional final exam.

- [Exam 1 in class Tuesday Oct 1](#), [solutions](#)
- [Exam 2 in class Thursday Nov 21](#), [solutions](#)
- Final exam (optional) Dec 19

3.4 Lab exercises

Lab exercises are intended for practice and are not graded.

- [Lab 1 on Aug 29 or 30](#)
- [Lab 2 on Sep 5 or 6](#)
- [Lab 3 on Sep 12 or 13](#)
- [Lab 4 on Sep 19 or 20](#)
- [Lab 5 on Sep 26 or 27](#)
- [Lab 6 on Oct 10 or 11](#)
- [Lab 7 on Oct 17 or 18](#)
- [Lab 8 on Oct 24 or 25](#)
- [Lab 9 on Nov 1 or 2](#)
- [Lab 10 on Nov 14 or 15](#)

4 Schedule

Week	Begins	Topic	Practice	Due on Monday
1	Aug 26	R Basics	Lab 1	
2	Sep 2	Data visualization	Lab 2	
3	Sep 9	Data import, tidy, wrangle	Lab 3	Problem set 1
4	Sep 16	Sampling distribution	Lab 4	
5	Sep 23	Hypothesis testing (Thu) Exam 1 review (Fri)	Lab 5	Problem set 2

Week	Begins	Topic	Practice	Due on Monday
6	Sep 30	Exam 1 (Tue in class) Fall break (Thu)		
7	Oct 7	Model specification	Lab 6	
8	Oct 14	Applied model specification	Lab 7	
9	Oct 21	Model fitting	Lab 8	Problem set 3
10	Oct 28	Model accuracy	Lab 9	
11	Nov 4	Election day & conference (no class)	No lab	Problem set 4
12	Nov 11	Model reliability (Tue) Classification (Thu)	Lab 10	
13	Nov 18	Exam 2 review (Tue) Exam 2 (Thu)		Problem set 5
14	Nov 25	Thanksgiving break (no class)		
15	Dec 2	Model extrapolation (Tue) Multilevel models	Lab is Pset 6 support	
16	Dec 9	Last day of classes (no class)		Problem set 6
17	Dec 19	Final exam at noon (optional)		