

# Statistical Methods in Linguistics

LING 214 / LING 414 ■ Fall 2026

University of Rochester  
Department of Linguistics

## Instructor

 **Aaron Steven White**  
 511A Lattimore Hall  
 [aaronstevenwhite.io](https://aaronstevenwhite.io)  
 [Office Hours Scheduler](#)

## Course Meetings

 Monday & Wednesday  
12:30 PM to 1:45 PM  
 Lattimore 513  
 [Course Zulip](#)

## 1 Course Overview

### 1.1 Course Communication

 **Important:** All course communication will be conducted through Zulip

 **Course Zulip:** [stats-in-ling-fall-2026.zulipchat.com](https://stats-in-ling-fall-2026.zulipchat.com)

- Post general questions to the appropriate channel
- Use direct messages (DMs) for private matters
- Do not use email for course-related communication

If you are registered for the course but cannot access the workspace, please contact me.

### 1.2 Course Description

This course provides an introduction to probability and statistics for linguistics, serving as an essential foundation for linguistics students who aim to analyze experimental and corpus linguistic data. The course covers:

- Elementary probability theory
- Descriptive and inferential statistics
- Fixed and mixed effects models
- Model evaluation and comparison
- Analysis of corpus data
- Analysis of judgment data
- Analysis of reading time data
- Analysis of acoustic measures

### 1.3 Prerequisites

This course is restricted to students who have completed **LING 110** with a grade of C- or better.

### 1.4 Credit Hours

This 4-credit course includes 150 minutes per week of direct instruction (two 75-minute sessions) and a minimum of 480 minutes per week of out-of-class student work, including reading assignments, problem sets, data analysis, and project work. This aligns with the University of Rochester's [credit hour policy](#), where each credit hour requires 50 minutes of instruction and 120 minutes of supplementary work.

As a cross-listed course, LING414 students are expected to demonstrate advanced content mastery, rigor, and requirements beyond the LING214 level. Graduate students will complete a comprehensive final project that demonstrates independent research capabilities consistent with graduate-level scholarship.

## 2 Learning Objectives and Outcomes

---

### 2.1 Course Objectives

By the end of this course, students will:

1. **Master** fundamental concepts in probability and statistics
2. **Apply** statistical methods to linguistic data analysis
3. **Design** and analyze linguistic experiments
4. **Interpret** statistical results in linguistics research
5. **Implement** analyses using R programming
6. **Construct** mixed effects models for complex datasets

## 3 Course Materials

---

### 3.1 Textbooks

#### **An Introduction to Bayesian Data Analysis for Cognitive Science**

Bruno Nicenboim, Daniel Schad, and Shravan Vasishth

[bruno.nicenboim.me/bayescogsci](https://bruno.nicenboim.me/bayescogsci)

#### **Statistics for Linguists: An Introduction Using R**

Bodo Winter

[appliedstatisticsforlinguists.org/bwinter\\_stats\\_proofs.pdf](https://appliedstatisticsforlinguists.org/bwinter_stats_proofs.pdf)

#### **Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan**

John K. Kruschke, second edition (2015)

[Elsevier book page](#)

## 3.2 Software

-  **R** | Statistical computing language  
[www.r-project.org](http://www.r-project.org)
-  **RStudio** | Integrated development environment  
[posit.co/download/rstudio-desktop](https://posit.co/download/rstudio-desktop)
-  **Visual Studio Code** | Integrated development environment  
[code.visualstudio.com/](https://code.visualstudio.com/)

## 3.3 Readings

 **Zipf's word frequency law in natural language: A critical review and future directions**  
Steven T. Piantadosi  
[doi.org/10.3758/s13423-014-0585-6](https://doi.org/10.3758/s13423-014-0585-6)

 **Population samples**  
Isabelle Buchstaller and Ghada Khattab  
[doi.org/10.1017/CBO9781139013734.006](https://doi.org/10.1017/CBO9781139013734.006)

 **Acoustic characteristics of American English vowels**  
James Hillenbrand, Laura A. Getty, Michael J. Clark, and Kimberlee Wheeler  
[doi.org/10.1121/1.411872](https://doi.org/10.1121/1.411872)

 **Cross-validation**  
Bruno Nicenboim, Daniel Schad, and Shravan Vasishth  
[bruno.nicenboim.me/bayescogsci/ch-cv.html](https://bruno.nicenboim.me/bayescogsci/ch-cv.html)

 **The Natural Stories Corpus**  
Richard Futrell, Edward Gibson, Harry Tily, Idan Blank, Anastasia Vishnevetskaya, Steven T. Piantadosi, and Evelina Fedorenko  
[aclanthology.org/L18-1012](https://aclanthology.org/L18-1012)

 **Limits on lexical prediction during reading**  
Steven G. Luke and Kiel Christianson  
[doi.org/10.1016/j.cogpsych.2016.06.002](https://doi.org/10.1016/j.cogpsych.2016.06.002)

 **The Provo Corpus**  
Steven G. Luke and Kiel Christianson  
[doi.org/10.3758/s13428-017-0908-4](https://doi.org/10.3758/s13428-017-0908-4)

 **Analyzing linguistic diversity using Poisson regression; adding exposure variables**

Bodo Winter

[appliedstatisticsforlinguists.org/bwinter\\_stats\\_proofs.pdf](https://appliedstatisticsforlinguists.org/bwinter_stats_proofs.pdf)

 **French liaison is allomorphy, not allophony: Evidence from lexical statistics**

Enrico Storme

[doi.org/10.1007/s11525-024-09429-8](https://doi.org/10.1007/s11525-024-09429-8)

 **Factivity, presupposition projection, and the role of discrete knowledge in gradient inference judgments**

Julian Grove and Aaron Steven White

[doi.org/10.1007/s11050-025-09244-9](https://doi.org/10.1007/s11050-025-09244-9)

 **The independence assumption; dealing with non-independence via experimental design and averaging**

Bodo Winter

[appliedstatisticsforlinguists.org/bwinter\\_stats\\_proofs.pdf](https://appliedstatisticsforlinguists.org/bwinter_stats_proofs.pdf)

 **Predicting the Dative Alternation**

Joan Bresnan, Anna Cueni, Tatiana Nikitina, and R. Harald Baayen

[web.stanford.edu/~bresnan/qs-submit.pdf](https://web.stanford.edu/~bresnan/qs-submit.pdf)

 **The CommitmentBank: Investigating projection in naturally occurring discourse**

Marie-Catherine de Marneffe, Mandy Simons, and Judith Tonhauser

[ojs.ub.uni-konstanz.de/sub/article/view/601](https://ojs.ub.uni-konstanz.de/sub/article/view/601)

 **Final devoicing before it happens: A large-scale study of word-final obstruents in French**

Adèle Jatteau et al.

[doi.org/10.16995/labphon.10855](https://doi.org/10.16995/labphon.10855)

 **Aspectual similarity predicts sense similarity**

Aaron Steven White, Scott Grimm, and Lelia Glass

[Course PDF](#)

 **Control methods used in a study of the vowels**

Gordon E. Peterson and Harold L. Barney

[doi.org/10.1121/1.1906875](https://doi.org/10.1121/1.1906875)

 **Frequency, acceptability, and selection: A case study of clause-embedding**

Aaron Steven White and Kyle Rawlins

[doi.org/10.5334/gjgl.1001](https://doi.org/10.5334/gjgl.1001)

## 📖 Decomposing and recomposing event structure

William Gantt, Lelia Glass, and Aaron Steven White

[aclanthology.org/2022.tacl-1.2](https://aclanthology.org/2022.tacl-1.2)

## 4 Course Structure and Requirements

### 4.1 Assessment Overview

| Component       | LING 214 | LING 414 | Details                                           |
|-----------------|----------|----------|---------------------------------------------------|
| 📝 Problem Sets  | 65%      | 40%      | Five coding assignments with in-class assessments |
| 🎓 Midterm Exam  | 25%      | 20%      | Comprehensive exam on Oct. 28                     |
| 👥 Participation | 10%      | N/A      | Attend office hours twice                         |
| 📋 Final Project | N/A      | 40%      | <a href="#">See rubric</a>                        |

### 4.2 Problem Set Grading

#### 📝 Problem Set Assessment Structure

Each problem set will be graded on two components:

- **50%: Functionality and Completeness:** Code works correctly and addresses all requirements
- **50%: In-Class Presentation:** Randomly selected students present their homework solution and explain how it works to the class. All students must be prepared to present on any assessment day following the problem set due date.

#### ⚠️ Important Notes:

- Students will not know in advance if they will present on a given day
- All students must attend class and be ready to present their solutions
- **LLM Usage:** While LLMs may be used for problem sets, you must fully understand your code to present it effectively

### 4.3 Final Project Requirements (LING 414 Only)

#### 🎓 Final Project Components (40% of total grade)

- **Pre-proposal meeting (5%):** Meeting with me to discuss proposal (by Oct. 16)
- **Proposal (5%):** Proposal outlining research question and methods (due Oct. 28)
- **Presentation (20%):** In-class presentation with professional slides (Dec. 7, 9, or 14)
- **Code Submission (20%):** All code structured following the rubric (due Dec. 14); evaluated via a 10 to 20 minute 🗣️ oral assessment scheduled during finals week
- **Paper Submission (50%):** Research paper (due Dec. 14)

**Project Options:**

1. **Corpus Analysis:** using open source or LDC corpora with external NLP tools
2. **Experiment Design:** judgment/reading experiment, including data collection instrument, with simulated data analysis
3. **Mechanistic Interpretation:** deep learning model analysis

 See the [Final Project Rubric](#) for detailed requirements and grading criteria

#### 4.4 Grading Scale

| A   | A- | B+ | B  | B- | C+ | C  | C- | D+ | D  | E   |
|-----|----|----|----|----|----|----|----|----|----|-----|
| 93+ | 90 | 87 | 83 | 80 | 77 | 73 | 70 | 67 | 60 | <60 |

## 5 Course Policies

### 5.1 Office Hours

 **By appointment** Please schedule using this form: [app.squarespacescheduling.com](https://app.squarespacescheduling.com)

I am available to meet if and only if a time is listed as available on the scheduling form.

### 5.2 Communication

**All communication must be done through the [Fall 2026 course Zulip](#).**

-  Post general questions to appropriate Zulip channels
-  Use Zulip DMs for grade questions, course absence notifications, etc.
-  Response time: 1-2 business days during work hours
-  No monitoring after 5 PM or weekends

### 5.3 Late Work Policy

- Problem sets: 10% deduction per day (max 3 days)
- Project components (414 only): No late submissions without prior approval
- Extensions require 48-hour advance notice (except emergencies)

### 5.4 Academic Integrity

Collaboration on problem sets with classmates is encouraged, but each student must write their own solutions. See the generative AI policy below and see: [rochester.edu/college/honesty](https://rochester.edu/college/honesty)

### 5.5 Generative AI Policy

Students may use generative AI tools (e.g., ChatGPT, Claude, GitHub Copilot) for coding assistance on problem sets and projects. However, students must:

- Fully understand and be able to explain all generated code. This will be assessed during in-class presentations and oral assessments.
- Cite in code comments any AI tools used, including the specific prompts
- Not use AI tools for written assignments

## 5.6 Accessibility

The University of Rochester welcomes students with disabilities. Students seeking accommodations must request them through the Office of Disability Resources: [rochester.edu/disability](https://rochester.edu/disability)

## 5.7 Title IX

All faculty are mandatory reporters for Title IX issues. Students can contact the Title IX Office directly at [titleix@rochester.edu](mailto:titleix@rochester.edu) or 585-275-1654. More information: [rochester.edu/sexualmisconduct](https://rochester.edu/sexualmisconduct)

# 6 Course Schedule

| Week | Dates           | Topics                                                                                                     | Due                           |
|------|-----------------|------------------------------------------------------------------------------------------------------------|-------------------------------|
| 1    | Aug. 31, Sep. 2 | Linguistic data and representation; probability spaces                                                     |                               |
| 2    | Sep. 7, 9       | <i>Labor Day</i> ; events and probability measures                                                         |                               |
| 3    | Sep. 14, 16     | Random variables; discrete distributions                                                                   |                               |
| 4    | Sep. 21, 23     | Continuous and joint distributions; dependence                                                             |                               |
| 5    | Sep. 28, 30     | Populations, samples, estimands, and uncertainty                                                           |                               |
| 6    | Oct. 5, 7       | Paired inference; conditional means and regression                                                         | <b>PS1 (Oct. 7)</b>           |
| 7    | Oct. 12, 14     | <i>Fall Break</i> ; multiple regression and model criticism                                                |                               |
| 8    | Oct. 19, 21     | Out-of-sample prediction, loss, and grouped validation                                                     | <b>PS2 (Oct. 21)</b>          |
| 9    | Oct. 26, 28     | Review;  <b>Midterm</b> | <b>Proposal (414)</b>         |
| 10   | Nov. 2, 4       | GLMs: binary, count, and bounded responses                                                                 |                               |
| 11   | Nov. 9, 11      | Experimental design, dependence, and generalization                                                        | <b>PS3 (Nov. 11)</b>          |
| 12   | Nov. 16, 18     | Mixed models: ordinal and bounded responses                                                                |                               |
| 13   | Nov. 23, 25     | Clustering and mixture models; <i>Thanksgiving recess</i>                                                  | <b>PS4 (Nov. 23)</b>          |
| 14   | Nov. 30, Dec. 2 | Factorization and missing-data validation; custom model design                                             |                               |
| 15   | Dec. 7, 9       |  Presentations (414)    | <b>PS5 (Dec. 7)</b>           |
| 16   | Dec. 14         |  Presentations (414)    | <b>Paper &amp; code (414)</b> |
|      | Dec. 18 to 23   |  Oral assessments (414) |                               |

# 7 Student Resources

## 7.1 Academic Support

- Center for Excellence in Teaching
- Writing, Speaking & Argument Program
- Q&i Statistics Help
- Library Research Support

## 7.2 Health & Wellness

- University Counseling: 585-275-3113
- University Health: 585-275-2662
- CARE Network
- IT Help Desk: 585-275-2000