Syllabus for LIN 538

Statistics for Linguists – Fall 2020

MW 14:40-16:00 via ZOOM

Draft as of August 24, 2020

CLASS ZOOM: Meeting ID: 913 1978 1082 Password: 3f707i
(click the box above to connect by zoom)

Course Website: `http://jeffreyheinz.net/classes/20F/`

What is this course? Like many areas of mathematics, statistics and probability have found widespread use in many scientific disciplines today, including linguistics. This course covers four units on statistics and probability as they can be applied to the study of language. The first unit is a short review of basic concepts in probability and statistics. The second unit covers regression analyses for use in experimental and corpus linguistics. The third unit covers language modelling, which is the development and evaluation of models that predict the next items in sequences. The last unit covers probabilistic graphical models (PGMs) for developing statistical arguments for identifying causal factors in models and making inferences.

Course outcomes. Students will learn how

- understand strengths and weaknesses of of analyzing linguistic phenomena with probabilistic and statistical methods;
- understand and perform regression analyses for analyzing experimental data, corpus data, and other types of data;
- understand language modelling and to create, use, and evaluate different kinds of generative grammars;
- understand probabilistic graphical models and to create, use, and evaluate different kinds of PGMs for causal and inferential statistical analysis.

Required readings. There is no required textbook. The following sources are recommended.

- Bodo Winter. *Statistics for Linguists: An Introduction Using R.* Routledge, 2019
- Daniel Jurafsky and James Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics.* Prentice-Hall, Upper Saddle River, NJ, 2nd edition, 2008
- Daphne Koller and Nir Friedman. *Probabilistic Graphical Models: Principles and Techniques.* MIT Press, 2009

Grading policy. The final course grade depends on the following work.

40%	Assignments	There will be four assignments, one for each unit.
40%	Research Paper	Conduct a research project and write a research report
10%	Presentation	Present your research to the class in weeks 13 or 14
10%	Participation	Regular participation in classes and lead one lecture
100%		

Assignments. There will be four assignments, one for each unit. These units are Basics, Regression, Language Modelling, and Probabilistic Graphical Models. These assignments will involve some combination of thinking, writing, problem-solving, R scripting, Python programming, command line usage, and so on.

Research Paper. A research paper is due by noon on Monday December 14. Ideally, the research problem will utilize methods from one of the units in the class.

Presentation. In week 13 or 14, students will present on the current state of their research. What is the question/problem? What is the analysis/solution? How does this analysis/solution compare with others?

Participation. Attendance and participation in class is required. There are 28 class sessions. If you cannot attend class or recitation, please notify the instructors in advance (if that is not possible please notify them as soon as possible). Students are encouraged to participate in class by asking and answering questions, making informed comments and observations, using the chat box, and indicating reactions to ongoing discussion using the participant functions in ZOOM.

Virtual Office Hours. Office Hours will be on Tuesdays from 1-4pm and by appointment. Office Hours are conducted by ZOOM and connection details are below. If you plan to attend office hours, it helps me if you sign up for a slot in advance.

Sign up for Office Hour slot(s)

(click the box above to go to the calendar with sign-up slots)

However, you are also always welcome to just stop by. If you want to meet another day/time, please email me and we can set up an appointment.

ZOOM OFFICE HOURS: Meeting ID: 475 814 4071 Password: 788198

(click the box above to connect by zoom)

Tentative Schedule.

Week	Dates	Topics
01	08/24 26 28	Basics
02	08/31 09/02 04	Basics
03	09/07 09 11	Linear Regression
04	09/14 16 18	Multiple Linear Regression
05	09/21 23 25	Language Modelling
06	09/28 30 10/02	Hidden Markov Models
07	10/05 07 09	Non-deterministic FSMs
08	10/12 14 16	Deterministic FSMs
09	10/19 21 23	Probabilistic CFGs
10	10/26 28 30	Bayesian Networks
11	11/02 04 06	Causal Networks
12	11/09 11 13	Probabilistic Graphical Models
13	11/16 18 20	Presentations
Thanksgiving Break		
14	11/30 12/02 04	Presentations

University Policies and Services

Student Accessibility Support Center Statement If you have a physical, psychological, medical, or learning disability that may impact your course work, please contact the Student Accessibility Support Center, 128 ECC Building, (631) 632-6748, or at `sasc@stonybrook.edu`. They will determine with you what accommodations are necessary and appropriate. All information and documentation is confidential.

Academic Integrity Statement Each student must pursue his or her academic goals honestly and be personally accountable for all submitted work. Representing another person's work as your own is always wrong. Faculty are required to report any suspected instances of academic dishonesty to the Academic Judiciary. Faculty in the Health Sciences Center (School of Health Technology & Management, Nursing, Social Welfare, Dental Medicine) and School of Medicine are required to follow their school-specific procedures. For more comprehensive information on academic integrity, including categories of academic dishonesty please refer to the academic judiciary website at `http://www.stonybrook.edu/commcms/academic_integrity/index.html`

Critical Incident Management Stony Brook University expects students to respect the rights, privileges, and property of other people. Faculty are required to report to the Office of University Community Standards any disruptive behavior that interrupts their ability to teach, compromises the safety of the learning environment, or inhibits students' ability to learn. Faculty in the HSC Schools and the School of Medicine are required to follow their school-specific procedures. Further information about most academic matters can be found in the Undergraduate Bulletin, the Undergraduate Class Schedule, and the Faculty-Employee Handbook.

Additional resources To access mental health services, call Counseling and Psychological Services at 631-632-6720; Counselors are available to speak with 24/7. For updated information on the Academic Success and Tutoring Center please check `www.stonybrook.edu/tutoring` for the most up-to-date information. For IT Support: Students can visit the Keep Learning website at `https://sites.google.com/stonybrook.edu/keeplearning` for information on the tools you need for alternative and online learning. Need help? Report technical issues at `https://it.stonybrook.edu/services/itsm`or call 631-632-2358.

For information on Library services and resources please visit the Continuity of Library Operations guide.