Tentative Syllabus for
LING 411: Linguistic Methods

Pavel Logačev

Fall 2017

E-mail: pavel.logacev@boun.edu.tr

Office Hours: by appointment

Office: Room 308, John Freely Hall

Web: Moodle

Class Hours:	Mon, 15:00-17:00	JF 333
	Tue, 13:00-14:00	JF 331
	Wed, 11:00-12:00	EF 203

Course Description

The aim of this course is to help you become a capable used and critical consumer of current methods of statistical inference. In this course, you will learn

- how to answer a research question empirically
- how to summarize data (*descriptive statistics*)
- how to find out to what degree a dataset warrants certain conclusions (*inferential statistics*)
- the use of the statistics software packages R (r-project.org) and Stan (mc-stan.org)
- other computational methods for dealing with language data

At the end of the course, you should be able to

- translate research questions into testable hypotheses or estimation problems
- conduct (some) statistical analyses in R/Stan/brms/lme4, and interpret their results
- independently interpret typical statistical analyses presented in research papers

Course Structure and Grading

- **Attendance and participation.** Attendance and participation will account for 10% of your course grade. Please actively participate in class.

- **Readings.** Please read the assigned chapters for each week. I will not check whether you have read the material, but keeping up with the reading will be essential to follow the course.
- **Programming assignments.** There will be several (very) short programming assignments every week. The quality of the submitted solutions will determine 20% of your grade.
- **Statistics problems.** Further 20% of the grade will be based on two short problems to be submitted every week.
- **Final project & term paper.** The final 50% of the grade will be based on a final project.

Core Readings

- McEreath, Richard (2016). *Statistical Rethinking: A Bayesian course with examples in R and Stan.* CRC Press.

Additional Readings

- Kruschke, John K. (2015). *Doing Bayesian Data Analysis.* Academic Press.
- Baayen, Harald (2008). *Analyzing Linguistic Data: A practical introduction to statistics using R.* Cambridge University Press.

Case Studies

1. **Alice: Language and gender**
   - Does gender influence how much you speak?

   → Switchboard corpus
2. **Bob: Phonotactics of consonant clusters in Turkish**
   - Are there any consonant clusters in Turkish?
   - Are there any at the beginning of a word? (as in *'fren'*)
   - In which environments do they occur?

   → probably TS corpus
3. **Clara 1: Phoneme inventories**
   - Are the sizes of consonant and vowel inventories related?

   → World Atlas of Language Structures
4. **Clara 2: Implicational universals**
   - Does verb-finality imply the presence of postpositions?

   → World Atlas of Language Structures

5. **Donald: Lexical decision**
   - Are frequent words shorter than infrequent words?
   - Do frequency and length influence the speed with which we recognize words?
   - If so, Do both contribute independently to reaction times?

   → English Lexicon Project
6. **Emily: Questionnaire Study**
   - Does usage of low-frequent words or loan-words affect the acceptability of a sentence?
   - If so, does knowledge of a foreign language play a role in this effect?

Course Policy

- Please bring your laptop with a copy of RStudio (https://www.rstudio.com/products/rstudio/download/).
- Please switch off your phones, or set them silent mode.
- Don't cheat and try to learn stuff.

Tentative Schedule

Week 1, 17.09 - 21.09: Introduction to R

optional reading: Kruschke (2015), Chapters 3

Week 2, 24.09 - 28.09: Introduction to Statistical Modeling

required reading: McElreath (2016), Chapter 1
optional reading: Kruschke (2015), Chapters 2

Week 3, 01.10 - 05.10: Probability

required reading: McElreath (2016), Chapter 2
optional reading: Kruschke (2015), Chapter 4

Week 4, 08.10 - 12.10: Fundamentals of Bayesian Inference

required reading: McElreath (2016), Chapter 3
very optional reading: Kruschke (2015), Chapters 5 & 6

Week 5, 15.10 - 19.10: Linear Regression

required reading: McElreath (2016), Chapter 4
very optional reading: Kruschke (2015), Chapter

Week 6, 22.10 - 26.10: Multivariate Linear Regression

required reading: McElreath (2016), Chapter 5
very optional reading: Kruschke (2015), Chapter

Week 7, 29.10 - 02.11: Parsimony in Estimation & MCMC

required reading: McElreath (2016), Chapters 6 & 8
optional reading: Lynch (2007), Chapters 4 & 5

Week 8, 05.11 - 09.11: Interactions between Variables

required reading: McElreath (2016), Chapter 7

Week 9, 12.11 - 16.11: Generalized Linear Model I: Count and Binomial Data

required reading: McElreath (2016), Chapters 9, 10
very optional reading: Kruschke (2015), Chapter 15

Week 10, 19.11 - 23.11: Generalized Linear Model II: Ordered Categorical Data

required reading: McElreath (2016), Chapter 11

Week 11, 26.11 - 30.11: Multilevel Models I

required reading: McElreath (2016), Chapter 12
very optional reading: Kruschke (2015), Chapter 9

Week 12, 03.12 - 07.12: Multilevel Models II

Week 13, 10.12 - 14.12:

required reading: Kruschke (2015), Chapters 11 & parts of 12