Ling 104 - Syllabus

Instructor:	Sandra Auderset
E-Mail:	auderset@umail.ucsb.edu
Office:	SH 5431H
Office Hours:	Wednesday 3-5pm (or by appointment)
Class Time:	Tuesday 10am-12:20pm
Class Room:	PSY East 1805

Syllabus (subject to change!)

Please bring your laptop to every class. There are computers in the class room, but it is easier if you have everything on your own device.

Week 1 – 10/02/2018: Intro to statistics & basics of R and RStudio
What is statistics? Why should we use it? What are R and RStudio? How do we use them?
Types of variables, formulating hypotheses, samples, types of R objects, basic programming operations

　　Please install this on your computer before class:
- R　　　　　https://www.r-project.org/
- RStudio　　https://www.rstudio.com/
- to test if it's installed properly, open a new file in RStudio (File>new R Script), type `1+1` on the first line and hit `command+enter` (Mac) or `control+enter` (Windows). This should give the following result on the second line: `[1] 2`

　　Homework:
- complete chapter 1 (or more!) of this free R tutorial:
  https://www.datacamp.com/courses/free-introduction-to-r
- try any other R tutorial(s) (see online materials below)

　　Readings:
- Levshina 2015: 1.1-1.3 (10p) and 2.1-2.5 (12p)
- Gries 2013: 1.1-1.3.1 (optionally 1.3.2) (10/20p) and 2.1 (5p)
- (Baayen 2008: chapter 1 - optional)

Week 2 – 10/09/2018: Data organization & basics of statistical tests
How to find/collect data, how to store it, how to read it into R; What are statistical tests? What kinds are there? What are p-values?

　　Preparation for class:
- Please make sure you have a program on your laptop that can open spreadsheets.
　　I use LibreOffice, which is open source and can be downloaded here:
　　https://www.libreoffice.org/

Of course, if you have OpenOffice or Excel from Microsoft Office, that works as well.
- You should also make sure you have a text editor. I recommend TextMate for Mac and SublimeText for Windows, but feel free to use something else.

Homework: posted on GauchoSpace (fyi: we need it for a class acitivity)

Readings:
- Gries 2013: 2.2-2.5 (33p)
- Levshina 2015: 1.4-1.5 (4p) and 2.6-2.7 (7p)

Week 3 – 10/16/2018: Descriptive statistics for numeric variables
Measures of central tendency, measures of dispersion, testing for normality, basics of visualization

Homework: posted on GauchoSpace

Readings:
- Gries 2013: 3.1 (32p)
- Levshina 2015: chapter 3 (27p)

Week 4 – 10/23/2018: Descriptive statistics for categorical variables
frequency tables, proportions, more visualization; preparation and discussion of assingment 1; practicing code

Homework: posted on GauchoSpace

Readings:
- Gries 2013: 3.2 (10p)
- Levshina 2015: chapter 4 (16p)

Week 5 – 10/30/2018: Intro to regression analysis & linear regression with one predictor
basic concepts of regression analysis; doing linear regression with R; linear regression with one predictor

! Draft of Assignment 1 due on Monday, 10/29/2018 at 11:55pm (optional) !
! Assignment 1 due on Friday, 11/02/2018 at 11:55pm !

NO Homework

Readings:
- Levshina 2015: 7.1-7.2.5 (23p)
- (Gries 2013: 5.2 (23p) - optional)

Week 6 – 11/06/2018: Linear regression with multiple predictors & model selection
linear models with more than one predictor, model selection based on maximum likelihood

   ! Mini-Midterm in class (first 30 minutes) – please bring a green scantron !

   Homework: finishing/working on the exercise code from lecture

   Readings:
   ○ Levshina 2015: 7.2.6-end of chapter 7 (8p)

Week 7 – 11/13/2018: Generalized linear models
generalized regression models for categorical, binary, ordinal, and frequency dependent variables, implementation in R; assignment 2 will be posted after class

   Homework: posted on Gaucho Space

   Readings:
   ○ Gries 2013: 5.3 and 5.4 (33p)
   ○ (Levshina 2015: 12.2.2 and 13 – optional)

Week 8 – 11/20/2018: Classification and regression trees
classification and regression trees and what they are used for; implementation in R; practicing code in general

   ! Draft of Assignment 2 due on Wednesday, 11/21/2018 at 11:55pm (optional) !

   NO Homework

   Readings:
   ○ Levshina 2015: chapter 14 (10p)

Week 9 – 11/27/2018: Cluster analysis
What is cluster analysis and what can it tell you?; Implementation and visualization in R

   ! Assignment 2 due on Monday, 11/26/2018 at 11:55pm !

   NO Homework

   Readings:
   ○ Levshina 2015: chapter 15 (21p)

Week 10 – 12/04/2018: Recap and outlook on other statistical methods

! Final Project draft due on Thursday, 12/06/2018 at 11:55pm !

Homework: posted on Gaucho Space

NO Readings

Finals Week

! Final Project due on Thursday, 12/13/2018 at 11:55pm !

References

Baayen, R. Harald. 2008. *Analyzing Linguistic Data: A practical introduction to statistics using R.* Cambridge: Cambridge University Press

Desagulier, Guillaume. 2017. *Corpus Linguistics and Statistics with R: Introduction to Quantitative Methods in Linguistics.* Cham, Switzerland: Springer

Gries, Stefan Th. 2013. *Statistics for Linguistics with R: a practical introduction.* Berlin: De Gruyter Mouton. 2nd edition

Johnson, Keith. 2008. *Quantitative Methods in Linguistics.* Malden, MA: Blackwell

Levshina, Natalia. 2015. *How to do Linguistics with R: Data exploration and statistical analysis.* Amsterdam: John Benjamins

Selected online materials

- basic info:
  - https://www.statmethods.net/
  - https://cran.r-project.org/manuals.html

- cheat sheets:
  - https://www.rstudio.com/resources/cheatsheets/
  - https://www.r-graph-gallery.com/cheatsheet/
  - https://www.datacamp.com/community/data-science-cheatsheets

- tutorials:
  - https://www.datacamp.com/courses/free-introduction-to-r
  - https://www.pluralsight.com/courses/r-programming-fundamentals
  - https://www.statmethods.net/r-tutorial/index.html
  - http://www.r-tutor.com/

- forums:
  - https://stackoverflow.com/questions
  - https://stats.stackexchange.com/questions

- companion website to Levshina 2015 (super useful!): https://benjamins.com/sites/z.195/