Friedrich Schiller University Jena   Dr. Adam J.R. Tallman
Department of English and American Studies   Office Hours: By appointment only
Winter term 2023-3024   email: adam.james.ross.tallman@uni-jena.de

Statistics for Linguists

Lecturer:	Adam J.R. Tallman
Course type:	WS 2023
Time:	Wednesday 12:00-14:00
Venue:	Ernst-Abbe-Platz 8

Course description

This course introduces statistical analysis and basic programming with R. It is graduate level. No background in statistics or programming is required.

Requirements and marking system

i. The final exam
ii. Talking to me about your final paper and reading relevant statistical literature
iii. Writing a brief paper/project description with an annotated bibliography
iii. Writing a draft of your paper (allows you to qualify for part iv)
iv. Writing a revised version of your paper (this is what you are graded on)

The final mark for this class is based on the final paper. You will write a draft of your final paper which I will ask you to revise and submit again. I will give homeworks and I recommend that you complete them. I will give you feedback, but they won't be part of your final grade.

Try to do the reading before class. I am going to attempt to teach you the following: (i) descriptive statistics & exploratory data analysis; (ii) frequentist versus bayesian statistics; (iii) the meaning of p-values and frequentist statistical methods (anova, chi-square, etc.); (iv) linear models; (v) logistic models; (vi) multilevel model; (vii) some machine learning or exploratory data analysis depending on interest.

I expect you to read beyond the class material on your own. I will provide some recommended readings, but you are expected to self-study. In addition to the course textbooks, it is recommended that you find a general introduction to statistics to reinforce the concepts (introductory biostatistics or statistics for psychologists are often good).

The course management system of our department, called MOODLE (https://moodle.uni-jena.de/course/view.php?id=32946), will be the online platform of this class. The purpose of MOODLE is to allow you to download the weekly texts, follow-ups handouts and materials etc. Important information of the organisation of the course and its lectures (e.g., announcements, etc.) will also be distributed view MOODLE (to your email address).

Course textbooks

"Gentle" conceptual background
Stigler, Stephen M. 2016. *The Pillars of Statistical Wisdom*. Harvard: Harvard University Press.
**gentle conceptual introduction to statistics

Frequentist statistics and R for Linguistics

Levshina, Natalia. 2015. *How to do Linguistics with R: Data exploration and statistical analysis.* Amsterdam: John Benjamins Publishing Company. **might be the easiest textbook with a focus on linguistics

Baayen, R. H. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics using R.* Cambridge: Cambridge University Press. **somewhat more difficult

Rough schedule (but you should design your own study schedule with my consultation)

	Date	Topic & Activity	Reading
	18.10.2022	Introduction to course, installing R and basics to R language	
	25.10.2022	R language	Levshina Ch.1
	01.11.2022	Descriptive statistics	Stigler Ch. 1-2, Levshina Ch.2
	08.11.2022	Descriptive statistics	Levshina Ch. 3-4
	15.11.2022	p-values, chi-square tests, t-tests, anova (and Bayes' rule)	Stigler Ch. 3-4
	22.11.2022	p-values, chi-square tests, t-tests, anova (and Bayes' rule)	Levshina Ch. 5
	29.11.2022	Linear model ("regression")	Stigler Ch. 5-7
	06.12.2022	(Multivariate) Linear model ("regression")	Levshina Ch. 6-9
	13.12.2022	Project descriptions due with annotated bibliography No class (there will be a make up class where more programming and data organization skills will be taught)	
	03.01.2023	Logistic regression	Levshina Ch. 11-12
	10.01.2023	Interactions & Causal inference	TBA
	17.01.2023	Multilevel models	TBA
	24.01.2023	Multilevel models	TBA
	31.01.2023	Exploratory data analysis / machine learning	TBA