Quantitative Methods in Linguistics (Ling 147)

Course Description & Syllabus

Spring 2014

Contents

1	Org	anizational matters	1
	1.1	Office hours	2
	1.2	Web access	2
	1.3	Weekly schedule	2
	1.4	Important note	2
2	Ger	eral Description	3
3	Cou	urse Requirements	5
	3.1	Grade calculation	5
			-
4	Imp	oortant general advice and the main computational tools	7
4 5	Imp Bas	ic course structure (subject to change!)	7 10
4 5	Imp Bas 5.1	ic course structure (subject to change!) Introduction: What this course is about – Week 1	7 10 10
4 5	Imp Bas 5.1 5.2	ic course structure (subject to change!) Introduction: What this course is about – Week 1	10 10 10
4 5	Imp Bas 5.1 5.2 5.3	ic course structure (subject to change!) Introduction: What this course is about – Week 1	10 10 10 11
4 5	Imp Bas 5.1 5.2 5.3 5.4	ic course structure (subject to change!) Introduction: What this course is about – Week 1	10 10 10 11
4 5	Imp Bas 5.1 5.2 5.3 5.4	ic course structure (subject to change!) Introduction: What this course is about – Week 1	10 10 10 11 11
45	Imp Bas 5.1 5.2 5.3 5.4 5.5	ic course structure (subject to change!) Introduction: What this course is about – Week 1	10 10 10 11 11 11
4 5	Imp Bas 5.1 5.2 5.3 5.4 5.5 5.6	ic course structure (subject to change!) Introduction: What this course is about – Week 1	10 10 10 11 11 12 12
45	Imp Bas 5.1 5.2 5.3 5.4 5.5 5.6 5.7	ic course structure (subject to change!) Introduction: What this course is about – Week 1	7 10 10 10 11 11 12 12
45	Imp Bas 5.1 5.2 5.3 5.4 5.5 5.6 5.7	ic course structure (subject to change!) Introduction: What this course is about – Week 1	 7 10 10 10 11 11 12 12 12

1 Organizational matters

- Class: TuTh 4:00AM–5:45PM, Soc Sci 1 Rm 135 (the computer lab)
- Instructor: Adrian Brasoveanu
- Teaching Assistant: Clara Sherley-Appel

- Section: Wed: 9:30AM–10:40PM, Soc Sci 1 Rm 135 (the computer lab) the same room as the lectures.
- Sections and lectures for this class are mandatory. Attendance will be taken in both lectures and sections. On occasion, we might decide to make the section on a particular week optional, depending on how the course progresses. If so, you will be informed about this via eCommons or in-class.

1.1 Office hours

- Adrian Brasoveanu: Tue 1:45PM-3:15PM, Stevenson 259, abrsvn@ucsc.edu
- Clara Sherley-Appel: Mon 11:00AM-12:00PM, Stevenson 217 (The Meeting & Analysis Room for the Ling. Labs) csherley@ucsc.edu

1.2 Web access

- There is an eCommons site for this course. To access it, go to: https://ecommons.ucsc.edu/xsl-portal.
- The syllabus, hw assignments, lecture notes and any other materials will be posted under Resources. CHECK REGULARLY, e.g., before/after each class, every day an assignment will/might be posted etc.

1.3 Weekly schedule

- Assignments will generally be posted on Thur every other week, starting the first week of classes (see eCommons>Resources). Assignments are due 2 weeks later (unless otherwise specified), at the beginning of class, i.e., print them, solve them and bring them to class on their due date.
- The first hw assignment will be posted on the first day of classes and should be due the Thur of the 2nd week of classes, but we will extend this deadline by one week (so it will be due the Thur of the 3rd week of classes). The second assignment will be posted the Thur or the 2nd week of classes and will be due 2 weeks later at the beginning of the class. We will generally stick to this biweekly schedule after that.
- The assignments will have to be typed, formatted as closely as possible to the way the lecture notes are formatted, and stapled together properly. If the assignment does not follow this style very closely, we will not grade your assignment and you will get no points for it.

1.4 Important note

If you qualify for classroom accommodations because of a disability, please get an Accommodation Authorization from the Disability Resource Center (DRC) and submit it to me in person outside of class (e.g., office hours) within the first two weeks of the quarter. Contact

DRC at 459-2089 (voice), 459-4806 (TTY), or http://drc.ucsc.edu for more information on the requirements and/or process.

2 General Description

This course provides an introduction to data analysis for linguistics focusing on categorical data and continuous data, and using R. No previous knowledge of probability, statistical methods or statistical computing is presupposed. We will, however, presuppose basic theoretical knowledge of phonetics, phonology, morphology, syntax and semantics as taught in the following level 1 courses offered by the UCSC Department of Linguistics: Phonology 1, Semantics 1 and Syntax 1.

This course is useful for students if they want to pursue research projects in any linguistic subfield and will also be useful in their future careers – data analysis is a highly translatable skill.

There are many textbooks out there (and other resources) aimed at teaching statistics to linguists or teaching statistics *simpliciter*. There is no required textbook for this class, all the required content will be provided in the lecture notes, slides and related materials, and *viva voce* in class or section. The lecture notes are based on various (text)books, including but not limited to:

- the following three, substantial, 'statistics for linguists' textbooks: Baayen (2008), Johnson (2008), and Gries (2013)
- parts of these textbooks will be recommended as additional readings (but not required) and the data sets associated with them, which are freely available, will be on occasion used in class, hw assignments, and could be used for your final research projects (upon consultation with the instructor and/or TA)
- other truly excellent resources for (computational) statistics in general and statistics in the cognitive sciences in particular are: Abelson (1995), Miles and Shevlin (2001), Faraway (2004), De Veaux et al. (2005), Braun and Murdoch (2007), Gelman and Hill (2007), Wright and London (2009), Gries (2009), Kruschke (2011), Diez et al. (2013) (note that Diez et al. 2013 is freely available online; see the references at the end of the syllabus for a link to that textbook and related materials)
- there is no required textbook for this class, but if you feel you need more exposure more incremental and/or from a different perspective to the content of the class, you should probably take a look at the corresponding chapters in one of the following three textbooks:
 - De Veaux et al. (2005): note that this is an old edition, which you might be able to buy used for a fraction of the cost of the new edition; it doesn't matter that it's an older edition, the material usually covered in intro-to-stats courses has not changed for many years now and the presentation is already very thoughtful and polished in this edition; it emphasizes the 'probabilistic modeling' view of data analysis that we will try to pursue in this course too

- Gonick and Smith (1993): it might sound a bit frivolous, but it's as 'for real' as many other textbooks out there or even more so (the guide really doesn't shy away from the math), and shorter/more to the point than many comparable texts; plus it's inexpensive (check out its Amazon page), it has a lot of very good visual aids, and will provide some entertainment if you're into that sort of thing (just go for one of the many other options if you aren't); there are additional resources for Gonick and Smith (1993) on eCommons: an errata and a chapter-by-chapter study guide that is really good definitely read the study guide for chapter 1 to see what it's all about
- Johnson (2008): a balanced, very thoughtful introduction to quantitative methods for linguistics that is thorough (again, doesn't completely shy away from the math), and makes a careful and compelling argument for quantitative methods in linguistics; it's a bit light on the practical R aspects and it doesn't emphasize the 'dynamic / interactive / incremental model development' view of statistical analysis in the focused way in which we will try to, so it's not the required textbook for the class; but it is a really great and fairly inexpensive choice (particularly when compared to regular intro-to-stats textbooks)

The course will revolve around a suitable mix-and-match selection from many of the above (text)books and various other sources. But the content of the course won't simply be a 'collage' of disparate bits and pieces: the overarching goal of the course is to give you the tools and insight to perform data analysis on a data set in the same way you would build an explanatory linguistic theory to account for a set of generalizations.

- The first step is always to take a close look at the data and understand its main qualitative features and its overall structure.
- The second step is to come up with a theory / model / hypothesis / conjecture that you think could account for the data, or at least for an important part of it.
- The third step is to check and evaluate how well the theory / model accounts for / fits the data and identify weak points of the theory / model that are in need of revision or further specification.
- The fourth and final step is to iterate over the second and third steps until we think we have a sensible, and hopefully somewhat insightful theory that accounts for the data to a sufficient extent.

Data analysis and statistical modeling done right can be just as deep, interesting and open-ended as 'higher-level' theory construction. We should not think of them as recipes that we mindlessly apply, or spell-checking procedures that we add at the end of our investigation to make it look right, or put-in-the-data-and-turn-the-crank procedures that yield some estimates and *p*-values that show we're doing real science. The same kind of creativity, attention to detail, rigorous thinking and insight involved in any other scientific-theory construction activity is needed to do data analysis and statistical modeling properly. The reason for the mix-and-match selection from the above mentioned (and other) sources is to communicate this kind of message in the most explicit and clearest way possible, for each of the specific topics we will end up focusing on.

3 Course Requirements

The learning (and grading) tools are: lectures, section meetings, more or less biweekly problem sets, a final research project.

- Class attendance is a necessary part of this course (attendance will be taken). Reading the posted notes and doing the assigned readings is crucial but does not substitute for class attendance. Speaking up in class is strongly encouraged. If you cannot make a class, it is your responsibility to find out from a classmate what happened in the class you missed. Handouts / lecture notes etc. will be posted on the eCommons site the night before class or the day of class. Reading them is important but does not substitute for class attendance.
- The TA conducts mandatory section meetings (attendance will be taken). Sections for this course will be particularly useful to help you understand the material and use the computational tools. The purpose of the sections is to go over homework problems in detail and to discuss whatever issues you feel need further attention. If you cannot make the section, that will prove to be a significant problem. Please talk to your instructor as soon as possible about that.
- Written work for the course consists of more or less biweekly problem sets, posted and due according to the weekly schedule above, and the final research project. Homeworks will NOT be handed out in class. Homeworks should NOT be submitted by email—they must be submitted in person at the beginning of the class (on the day they are due). Again: please type your hw using a format that follows the format of the posted lecture notes as closely as possible, print it, and bring it to class on its due date.
- It is an excellent idea to form study groups and discuss the problem sets in your meetings. However, **YOU SHOULD WRITE UP THE ASSIGNMENTS ON YOUR OWN**. If you do form study groups, please list the people you discussed the problem set with at the beginning of the assignment. Turning in identical homeworks counts as plagiarism for all students involved.
- Homework policy: no late homework is accepted unless by prior arrangement (or because of a health problem properly documented to the satisfaction of the instructor). A student who misses more than 2 assignments automatically fails the course.
- There will be no midterm for this class, only a final research project.

3.1 Grade calculation

• Lecture & section attendance and participation: around 15%. Please remember that the lectures are mandatory – attendance will be taken. Sections are also mandatory and attendance will be taken in sections too, unless the instructor and the TA decide

to explicitly designate a particular section as optional (depending on how the course progresses).

- Completing 1 (online or in the lab) experiment in the first 5 weeks of classes is a requirement for this course. The deadline is the Thur of the 5th week of classes (May 1st, 2014), 4 pm. We will focus mostly on data analysis in this course and we will usually use data already collected by other people or data easily gathered from corpora. But data generation (experimental design etc.) is an important aspect of the scientific investigation process, and you need to have some first-hand experience with this part of the scientific investigation 'pipeline'.
 - go to https://ucsc-ling.sona-systems.com/
 - if you already have an account on SONA, just log in normally, select any experiment and select Ling147 (Quantitative Methods) as the course associated with that experiment
 - if you do not have an account on SONA, please first request a new account here https://ucsc-ling.sona-systems.com/student_new_user.aspx and associate it with Ling147
 - please complete 1 experiment of your choice in the first 5 weeks of classes
 - this will give you hands-on experience with experimental data collection and will provide a good basis for the data analysis techniques we will dedicate most of the quarter to
- If you are not able or willing to complete an experiment, please let me know and we will arrange for you to complete an alternative task by the same deadline as the experiment; the alternative task might consist of reading and summarizing a short paper (each summary should be approx. 2 pages), doing some linguistically interesting corpus searches (e.g., on COCA, www.americancorpus.org) and summarizing the results, or some other suitable alternative.
- Completing 5 problem sets, assigned and due according to the biweekly schedule above; please note that schedule changes might occur, and some assignments might be due at different times; this will be discussed and announced in class and/or in section and/or on eCommons: around 60% (usually 12% per problem set, but changes might occur and some assignments might deviate from this).
- Completing a final research project: around 25% (i.e., basically worth around 2 hw assignments or a bit more in terms of size, and work).
 - You should have a clear idea about your research project by the 8th week of classes at the latest.
 - You should discuss it with the instructor and/or the TA before embarking on it.
 - The final research project has to include the presentation (including goal of the experiment, experimental design and method, participants, descriptive summaries, data visualization etc.) and data analysis of a linguistically relevant data set.

- You can choose one of the data sets associated the three 'stats for ling' textbooks mentioned above, or a data set that you have generated yourself from corpus or experimental investigations.
- You can also simulate several data sets that vary in linguistically interesting ways and run data analyses on them (including descriptive summaries, data visualization etc.)
- You should discuss both the choice of data set and the planned data analysis with your instructor and/or TA before starting the project.
- The final project should strictly follow the same style, formatting conventions, level of detail and explicitness etc. as the lecture notes and the hw assignments.
- The final project will be due on the day finals are officially scheduled for our class time; there will be no final for this class, but the final research project has to be turned in at that time.

4 Important general advice and the main computational tools

This is a hard class that will require a lot of independent work on your part because you will have to learn both (i) new, difficult material (probability and statistics) and (ii) the computational tools needed to thoroughly understand and apply that knowledge to specific problems of (practical) interest.

You should think of probability and statistics on one hand, and R on the other, as two new (related) languages that you need to learn so well as to be able to use smoothly in scientific 'conversations' that are mathematically and computationally explicit.

You need to **PRACTICE** 'speaking' / using both these languages **EXTENSIVELY ON YOUR OWN** to be able to participate at an adequate level in the 'conversations' that the hw assignments and the final research project will require you to have – in addition to the actual conversations in lectures and sections.

Why do we try to do two hard things at once rather than just one? For two reasons, both pedagogical and both having to do with the expected background and goals of the participants in this class:

i. statistics properly studied requires fairly advanced mathematical knowledge (basically, calculus and linear algebra); but we can gain a fairly deep and correct (and also 'gut-level', which is really important) understanding of statistics without those skills if we instead make regular use of programming and computer-based probabilistic simulations;

we don't need to understand and know how to implement all the relevant mathematical functions, although we will do a little bit of that; we can instead develop good intuitions about the nature and behavior of probabilistic functions – and models that assemble multiple such functions together – by just looking at a wide range of inputs for those functions and ask the computer to compute and show us the corresponding outputs;

it is therefore crucial that we use a statistical *programming* environment like R that allows us to do create new functions out of simpler ones in unrestricted ways (compositionality at work!) rather than a graphical-menu-driven software package that has a limited, pre-canned set of routines / procedures (and for which you would probably need to pay)

ii. statistics is important to all of us because it can be applied to real data and solve real problems; for any realistic set of problems and realistic data set, we need computers because only fairly simple (often unrealistically simple) problems have a 'closed-form' solution (http://en.wikipedia.org/wiki/Closed-form_expression);

so sooner or later – in fact, sooner rather than later – you will need to use a computer to analyze your data anyway; and if this is the only course about quantitative methods you will ever take, you should probably have some practical skills at the end of it, backed by a good intuitive understanding of the statistical models you use (even if your mathematical understanding lags a bit behind)

Mathematics and computation in general, and probability / statistics and related computation (data summarization, visualization, data analysis, modeling etc.) in particular, are not spectator sports. You will not be able to provide satisfactory solutions to the biweekly hw assignments, to have a final research project of satisfactory quality at the end of the class, and to participate regularly and meaningfully in lectures and sections **if you do not put in the required (and significant) amount of time and energy needed to learn these two new 'languages'**.

This means, among other things, that you need to invest a lot of time:

- (re)reading the lecture notes and related materials
- typing all the R code discussed in lectures and sections on your own at home (i.e., in your own R session) type it, no copy-paste! examining the output, invoking the extensive R help / documentation that is available within R and on the web, trying variations of the provided R code etc.
- if you do not do this, you probably won't be able to solve the hw assignments, or complete the final project

In general, you should get in the habit of firing up R for any little computation task you need to do, no matter how small (the tax deadline is coming up :-)). R is a complex tool and you need to practice a lot – until it becomes second nature. Extensive practice with R will give you a kind of hands-on experience with random variation and gut-level understanding of probability that would be very hard to get otherwise. If you like playing video games, seriously think about not playing them anymore and just do batches of random draws from one of R's many built-in probability distributions and think of interesting ways to plot those draws to enjoy and appreciate the random variation. After all, the interesting and fun aspects of video games are really just random draws from some probability distribution dressed up to look nice \ldots

I will try really hard to provide detailed and practical lecture notes, discuss them in class extensively, improve them as needed, help with hw assignments, the final project etc. In addition, the TA will also try really hard to go over difficult issues again in section, help with the hw assignments and the final project etc. But you should also try really hard to put in the necessary amount of work. The more work you put in, the better you will do in this course, but most importantly: the better you will learn these notions and computational tools that have wide application both within linguistics and in other cognitive sciences, or other sciences in general!

We will use R and RStudio in this class. They are both free (as in 'free speech'—the source code is publicly available, and also as in 'free of charge'), and are installed on all the computers in the lab. You should feel free to bring your laptops to class and use them instead of or in addition to the lab computers. This will be useful to you if you want to practice at home, and also to have easy access to these tools after the class is over. If you decide to use your own laptops, please download and install R from here (or some other mirror): http://cran.cnr.berkeley.edu/. Also, please download and install RStudio from here (after you install R): http://rstudio.org/download/desktop.

RStudio has a lot of the R-specific integrated development environment (IDE) functionality out of the box. There are some other choices out there that you might want to pursue later on, but RStudio is a great place to start: it's available for all platforms (Win, Mac, Linux), it's stable and requires very little customization to set up a pleasant and functional working environment. We'll go over that customization at the beginning of the course.

Alternatives to RStudio. If you decide RStudio doesn't agree with you, your can also simply launch the R console on its own (not via RStudio) and use a separate text editor to write / edit R scripts and execute the scripts in batch mode, or one line / chunk at a time by copy-pasting the line / chunk from the editor window into the R console.

If you decide to go this way, what text editors should you use? R comes with its own basic text editor that you can use. But you can also use other, more 'professional'-level editors that provide extensive R support, i.e., syntax highlighting and possibly a convenient way to run pieces of code in R. Here are some resources – Geany and Gedit are probably the easiest editors to get started with:

- Geany (http://www.geany.org/) is a lightweight editor; go here http://www.geany. org/Download/Releases to install it on Windows (consider choosing the full installer with GTK included) and here http://wiki.geany.org/howtos/osx/running to install it on Mac OSX
- Gedit is another good editor that is straightforward to use and has support for all major operating systems; go here https://projects.gnome.org/gedit/ and follow the relevant link in the 'Downloads' sidebar on the right-hand side of the page
- an excellent (I hear) Emacs package for R is available here http://ess.r-project. org/
- an excellent (I can vouch for that) Vim plugin for R is available here http://www.vim.org/scripts/script.php?script_id=2628
- more info about other editors and IDEs is available here http://www.r-bloggers.com/r-guis-ides-and-text-editors/ and here http://www.sciviews.org/_rgui/projects/Editors.html among other places on the web

Speaking of editors. Emacs (http://www.gnu.org/software/emacs/) and Vim (http: //www.vim.org/) require some time to learn. Coming from a 'Word'-like world, that might seem like a lot of work for a task that should require very little work if any: just start the editor and start typing. Not dealing with this extra complication is a perfectly fine choice for this course. But if you expect to spend a lot of time in your future career generating and editing technical text, be it research papers / reports in LaTeX, scripting web pages, online surveys or apps and / or writing any other kind of software, learning to use one of these editors will definitely be worth your while.

Both Emacs and Vim have built-in tutorials that can get you started, and there are plenty of resources on the web. Read some reviews online to make a choice or choose whatever your friends use. The choice is less important, what's more important is to stick with it for 2 weeks or so. Learn these editors like you learn something new in a medium-difficulty linguistics course, don't expect to just 'get it' and start typing. Alternatively, think of these editors as a new kind of fairly complicated computer game and play with them for a while (say, about 1 hour a day for 2 weeks). Don't try to do any serious work for the first few days (like writing a paper). Things will become much easier after several days and you will be productive after 1 week or so of learning / playing around with the editor. This might seem like a long time to spend learning a text editor, but chances are that later on you'll be very glad you did it.

5 Basic course structure (subject to change!)

The following is a general outline of the progress of the course. The selection of topics and material covered each week and within each topic are subject to change. We might decide to make changes to this, for example: cover these topics in a slightly different order, emphasize frequentist and/or Maximum Likelihood and/or Bayesian approaches to any of these sub(topics), remove some of the (sub)topics to be able to add other topics of interest – e.g., an introduction to multinomial and/or ordinal regression models, a more in-depth discussion of mixed-effects / hierarchical / multilevel regression models etc. – or make various other changes that will improve the learning experience and will provide a better fit to the background knowledge and goals of the participants in the class.

5.1 Introduction: What this course is about – Week 1

- the scientific method
- statistics as modeling, i.e., as theory construction and testing

5.2 Types of data (numeric vs. categorical) and fundamentals of R – Week 1, 2, 3

- introduction to R and installation
- functions and arguments
- vectors (usually continuous data); generating them, loading & saving them, editing them

- factors, i.e., categorical data; generating them, loading & saving them, editing them
- data frames (multiple variables, as obtained in an experiment or observational study); generating them, loading & saving them, editing them

5.3 Displaying and summarizing data – Week 2, 3, 4

- one variable:
 - Scatterplots and line plots, bar plots, histograms, boxplots
 - measures of central tendency: mode, median, mean (arithmentic, geometric)
 - measures of dispersion: range, quantiles, quartiles, standard deviation
 - summary functions in R
 - centering and standardization (z-scores)
- two variables:
 - cross-tabulation, bar plots, mosaic plots, line plots, scatterplots
 - measures of association: correlation

5.4 Experimental vs. observational studies, probability and uncertainty, random sampling – Week 4, 5

- controlled randomized studies vs. observational studies
- taking samples
- types of experimental design
- simulating data for different types of experimental design
- properties of probability
- probability as long-run frequency
- betting odds and probability as degree of belief
- joint probabilities (two-way probability tables)
- conditional probability
- law of total probability
- Bayes' rule

5.5 The family of (general) linear models – Week 5, 6, 7

- the nature of statistical modeling and its place in linguistic theory development
- statistical modeling for continuous data
- we will focus on the family of (general) linear models for continuous data (which subsumes t-tests, ANOVAs, simple linear regression, and multiple linear regression), the estimation of the parameters of linear models, hypothesis testing, the proper interpretation of the estimated coefficients, model assumptions, model criticism, model comparison
- we will also learn how to simulate data for various experimental designs and associated types of linear models

5.6 Generalized linear models (GLMs) – Week 7, 8

- statistical modeling for categorical data, with a primary focus on binary categorical data
- we will discuss generalized linear models, and how they generalize even further the linear models we discussed before; we will primarily but not exclusively focus on generalized linear models for binomial (binary categorical) data (which subsumes t-tests, ANOVAs, and regression for proportions), and in particular on logistic regression models; we will briefly discuss the estimation of the parameters, but quickly move on to focus on the proper interpretation of the estimated coefficients, model assumptions, model criticism and model comparison
- we will also learn how to simulate binary categorical data for various experimental designs and associated types of generalized linear models

5.7 Introduction to linear mixed-effects models (LMMs) and generalized linear mixed-effects models (GLMMs) – (hopefully) Week 9, 10)

5.8 Back to the beginning – Week 10

- the scientific method and the place of statistical modeling in the development of scientific theories
- review

References

Abelson, R.P. (1995). Statistics as Principled Argument. L. Erlbaum Associates.

- Baayen, R. Harald (2008). Analyzing Linguistic Data: A Practical Introduction to Statistics Using R. Cambridge University Press.
- Braun, J. and D.J. Murdoch (2007). A First Course in Statistical Programming with R. Cambridge University Press.
- De Veaux, R.D. et al. (2005). Stats: Data and Models. Pearson Education, Limited.

- Diez, D. et al. (2013). *OpenIntro Statistics: Second Edition*. CreateSpace Independent Publishing Platform. URL: http://www.openintro.org/stat/textbook.php.
- Faraway, J.J. (2004). *Linear Models With R.* Chapman & Hall Texts in Statistical Science Series. Chapman & Hall/CRC.
- Gelman, A. and J. Hill (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Analytical Methods for Social Research. Cambridge University Press.
- Gonick, L. and W. Smith (1993). Cartoon Guide to Statistics. HarperCollins.
- Gries, S.T. (2009). Quantitative Corpus Linguistics with R: A Practical Introduction. Taylor & Francis.
- (2013). Statistics for Linguistics with R: A Practical Introduction, 2nd Edition. Mouton De Gruyter.
- Johnson, K. (2008). Quantitative methods in linguistics. Blackwell Pub.
- Kruschke, John K. (2011). Doing Bayesian Data Analysis: A Tutorial with R and BUGS. Academic Press/Elsevier.
- Miles, J. and M. Shevlin (2001). Applying Regression and Correlation: A Guide for Students and Researchers. SAGE Publications.
- Wright, D.B. and K. London (2009). Modern regression techniques using R: A practical guide for students and researchers. SAGE.