STATISTICS IN LINGUISTIC STUDIES
語言學與統計
Course code number: 1305542

Fall 2024   Thursday 14:10-17:00   文學院 (Humanities) 413

<mark>TENTATIVE (see updated schedule in August)</mark>

Your friendly guide:

James Myers (麥傑)

Office: 文學院 (Humanities) 247

Tel: x31506

Email: Lngmyers *at the university address*

Webpage: https://lngmyers.ccu.edu.tw/

Office hours: Wednesday 10 am - 12, or by appointment (made at least 24 hours ahead)

Goals:

This course will try to teach you the fundamentals of statistical analysis, plus give you a taste of programming and more advanced methods, focusing on linguistic data (phonetics, psycholinguistics, child language, sociolinguistics, corpus analysis, grammar research, language teaching), so that you can apply what you've learned to your own data.

Readings:

Myers, J. (2024). *Yet another statistics-for-linguists book*. National Chung Cheng University ms. [Cf. the similar but not identical Chinese adaptation: 陳宗穎、盧郁安、麥傑 (2024)「再談一回語言學與統計」清華大學、陽明交通大學、中正大學稿。]

Software:

Microsoft Excel

R: <https://www.r-project.org/>

Grading:

25% Graded homework 1 (due 10/17)

25% Graded homework 2 (due 11/14)

25% Graded homework 3 (due 12/19)

25% Statistical report (due 1/10)

Each week before class, you should read (most of) a chapter in my online statistics textbook. (If you like, you can also check the corresponding Chinese version, but the English version is the "official" textbook.) When you're reading, please try out the examples using your own computer. You can skip the "optional" parts on your first reading, though you may need to run the R code in them in order to create/modify R objects used in other sections, and you may also need to look at these sections later to deal with special issues, including in graded homework questions. There will also be ungraded practice exercises each week that we will discuss together using the classroom computer.

There will be three graded homeworks, which are like bigger versions of the weekly practice exercises. You will receive each graded homework two weeks before it is due, and it will cover topics that we have practiced in class up to that point. You can share ideas with your classmates, but you have to write up and hand in your own answers (email me clarification questions, and I'll reply to all, while keeping you anonymous). Note that the exercises, graded homeworks and final report must all use Excel and/or R; for consistency (and to make sure that

you're really learning new things), other statistics programs are not allowed. The graded homeworks (and related files, if any), are due by email by 12 noon.

At the end of the semester you'll submit a brief report (10 pages max for the report itself, in English) analyzing your own data using statistical techniques that you learned in this class, including at least two techniques from after the third graded homework. The report can be based on something that you already wrote (as long as you never analyzed the data statistically before), or you can collect some new data to analyze, or analyze old data in a new way (e.g., corpus data). The grade will be based on your overall logic, reporting style, and use of statistics, not on the linguistic content. The report should be written like a normal linguistics paper (citing statistics in standard format, including graphs or tables), but also include an appendix (after the references) giving explicit information on how you did the statistical analyses (e.g., your R code), plus a text file with the data (anonymized to protect your secrets, if you like), and must be submitted (via email, by 5 pm on 1/10) as a PDF file (with your ID number as part of the filename, and also on the first page). Obviously, do not hand in stuff late and do not plagiarize (including with AI help). Unless you have a really good excuse, you will lose 5 points for each day you are late. Homework or final reports containing plagiarism will receive a score of zero, and you will be reported to the department chair.

Schedule:
I will assume that you've read everything in each week's chapter, except the optional sections.

Week	Topic	Optional	Deadlines
9/12	Why do linguists need statistics? [ch. 1]	2, 3.3	
9/19	Data analysis software [ch. 2]	4.4.3, 4.4.4, 5.3	
9/26	Quantifying some familiar ideas: Averages and variation [ch. 3]	2.4.1, 4.2, 4.3	
10/3	Probability and hypotheses [ch. 4] **[distribute homework 1]**	2.1, 2.3, 3.2, 4.3, 5.3	
10/10	雙十節 **(no class)**		
10/17	**Discuss homework 1**		**Homework 1**
10/24	Correlation and modeling [ch. 5]	2.4, 3.3.2-3, 4.2-4	
10/31	Comparing two continuous variables [ch. 6] **[distribute homework 2]**	3.2-3, 4	
11/7	Comparing more than two continuous variables: Introduction to ANOVA [ch. 8]	5	
11/14	**Discuss homework 2**		**Homework 2**
11/21	More ANOVA: Repeated measures [ch. 9]	2.3.2, 2.4, 3.1, 3.3.2, 3.4	
11/28	Modeling continuous variables: Multiple regression [ch. 10] **[distribute homework 3]**	2.4, 3.2.1, 3.3, 4.2.2	
12/5	Modeling categorical variables: Logistic regression [ch. 11]	2.6, 3.1-3, 4	
12/19	**Discuss homework 3 and report plans**		**Homework 3**
12/26	Beyond ANOVA, regression, and chi-squared: Mixed-effects modeling [ch. 12]	2.4, 3.2, 4.1-3	
1/2	The future of statistics: Bayesian modeling [ch. 13] **[last class]**	2.2.3-4, 3.2-3	
1/10 (Fri)	**Statistical report due (email by 5 pm)**		**Report**

Some other statistics books:

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.

Brown, J. D. (1988). *Understanding research in second language acquisition: A teacher's guide to statistics and research design*. Cambridge: Cambridge University Press.

Crawley, M. J. (2005). *Statistics: An introduction using R*. Wiley.

Dalgaard, P. (2002). *Introductory statistics with R*. Springer.

Eddington, D. (2015). *Statistics for linguists: A step-by-step guide for novices*. Cambridge Scholars Publishing.

Gonick, L., & Smith, W. (1993). *The cartoon guide to statistics*. Harper Perennial. [鄭惟厚譯（2003）。看漫畫，學統計。天下遠見。]

Gries, S. T. (2013). *Statistics for linguistics with R: A practical introduction* (2nd edition). Berlin: De Gruyter. [1st edition is in our library]

Hatch, E. and Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. Newbury House Publishers.

Jaisingh, L. (2000). *Statistics for the utterly confused*. McGraw-Hill.

Johnson, K. (2008). *Quantitative methods in linguistics*. Wiley.

Kruschke, J. K. (2011). *Doing Bayesian data analysis*. Academic Press.

Larson-Hall, J. (2015). *A guide to doing statistics in second language research using SPSS and R* (second edition). Routledge.

Levshina, N. (2015). *How to do linguistics with R: Data exploration and statistical analysis*. John Benjamins.

McGrayne, S. B. (2011). *The theory that would not die: How Bayes' rule cracked the Enigma code, hunted down Russian submarines, & emerged triumphant from two centuries of controversy*. Yale University Press.

Navarro, D. (2014). *Learning statistics with R: A tutorial for psychology students and other beginners*. University of Adelaide ms.

Salsburg, D. (2001). *The lady tasting tea: How statistics revolutionized science in the twentieth century*. Henry Holt and Company. [薩爾斯伯格（2001）。統計，改變了世界。 天下文化。]

Spiegelhalter, D. (2019). *The art of statistics: Learning from data*. Pelican.

Vernoy, M., & Kyle, D. J. (2002). *Behavioral statistics in action*. McGraw-Hill.

Winter, B. (2019). *Statistics for linguists: An introduction using R*. Routledge.

Woods, A., Fletcher, P., & Hughes, A. (1986) *Statistics in language studies.* Cambridge University Press.

王文中（2004）。統計學與 Excel 資料分析之實習應用（第五版）。台北：博碩。

吳淑妃（2011）。統計學與 R 軟體的應用。臺中市：滄海。

陳景祥（2010）。R 軟體：應用統計方法。臺北市：臺灣東華。