STATISTICS IN LINGUISTIC STUDIES
語言學與統計

Course code number: 1305542

Fall 2026 Wednesday 14:10-17:00 文學院 (Humanities) 413

TENTATIVE SCHEDULE

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: Thursday 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. (2026). *Yet another statistics-for-linguists book*. National Chung Cheng University ms. [Cf. the Chinese adaptation: 陳宗穎、盧郁安、麥傑（in press）「語言算什麼：語言學研究的統計指引」。新竹：國立清華大學出版社。]

Software:

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

Grading:

30% In-class exam (10/14)
35% Take-home exam (11/18)
35% Final report (12/23)

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. There will also be ungraded practice exercises each week that we will discuss together using the classroom computer.

There will be two graded exams, which are like bigger versions of the weekly practice exercises, covering the first and second thirds of the class, respectively. The first exam (10/14) is in class, and believe it or not, has to be done by hand, on paper, without using any computer. You will get all the formulas you need, and the data sets and arithmetic will be kept as simple as possible. If you can't come to class that day, you'll have to schedule another three-hour period to do the exam (e.g., in the back of the classroom the following week). I won't give feedback until everybody is done. The second exam (11/18) will be a take-home exam, but you're not allowed to use AI, just regular computer programs like Excel and R; you'll submit it by email by 12 noon.

On 12/23 (one week after the last class), you'll submit a brief report (10 pages max for the report itself, in English, by 5 pm, by email, as a PDF file, with your ID number as part of the filename, and also on the first page). For this report, you may use AI, but only to ask clarification questions or to check your work, but NOT to write your R scripts or create your Excel files; your report must include a statement on exactly how you used AI (including not using it at all). The report will analyze your own linguistic data using statistical techniques that you learned in this class, including at least two techniques from after the third graded homework. This data can be newly collected data, data from a study that you already conducted (as long as you never analyzed the data statistically before), or public data (e.g., a corpus) that you analyze in a new way. 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 and AI acknowledgement if relevant), plus a text file with the data (anonymized to protect your secrets, if you like).

Obviously, do not hand in stuff late and do not plagiarize (including having AI write your R code, create your Excel files, or write your text). Unless you have a really good excuse, you will lose 5 points for each day you are late. Exams or reports containing plagiarism will receive a score of zero, and you will be reported to the department chair.

Schedule:
*Marks when something graded is due

Week	Topic	Reading
9/9	Why do linguists need statistics?	To be announced (TBA)
9/16	Data analysis software	TBA
9/23	Averages and variation	TBA
9/30	Probability and hypotheses	TBA
10/7	Correlation and modeling	TBA
10/14	***Exam 1** (in class)	TBA
10/21	Comparing category sizes	TBA
10/28	Comparing two continuous variables	TBA
11/4	Comparing multiple continuous variables (1)	TBA
11/11	Comparing multiple continuous variables (2) Get Exam 2 (take-home)	TBA
11/18	***Exam 2 due** (by email by 12 noon) Discuss report ideas or anything else you want	
11/25	Modeling continuous variables	TBA
12/2	Modeling categorical variables	TBA
12/9	Mixing fixed and random variables	TBA
12/16	Flipping statistics on its head **(last class)**	TBA
12/23	***Statistical report due** (by email by 5 pm)	

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.
Desagulier, G. (2017). *Corpus linguistics and statistics with R: Introduction to quantitative methods in linguistics*. 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. (2021). *Statistics for linguistics with R: A practical introduction* (3rd 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.
Rühlemann, C. (2020). *Visual linguistics with R: A practical introduction to quantitative interactional linguistics*. John Benjamins.
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.
吳淑妃（2011）。統計學與 R 軟體的應用。臺中市：滄海。
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。
陳景祥（2010）。R 軟體：應用統計方法。臺北市：臺灣東華。