

Prospectus for LING 5620: Quantitative analysis of linguistic variation
Meredith Tamminga

This course on regression methods and statistical programming is designed to meet the applied quantitative research needs of graduate students in linguistics or closely related fields. The central goal of the course is to equip participants with a set of up-to-date tools to implement and interpret common quantitative analyses of linguistic data. The course is taught exclusively in an R environment with an emphasis on the Tidyverse. Key topics include data visualization, exploratory data analysis, contrast coding systems, interpreting interaction terms, and hierarchical (i.e. mixed-effects) regression modeling.

The course is aimed at PhD students in Linguistics or a closely related field (psycholinguistics, computational linguistics) who have taken STAT 5000 or equivalent and now wish to apply that background to methods that are currently in use for linguistic research. You should already know something about the basics of probability theory, the premises of frequentist statistics, and the mechanics of classical (non-hierarchical) linear or logistic regression models. Informal experience with applied statistical modeling (for example, from working in a lab) might also provide enough background to succeed. Importantly, however, this course should not be your first exposure to statistics.

If you are not a PhD student in the Linguistics department, **please email me** (tamminga@ling.upenn.edu) before enrolling in this course. **This course is very unlikely to be appropriate for undergraduates.**