**Introductory Statistics for Linguists**

**Workshop Description and Learning Objectives**: This workshop covers statistical methods for addressing advanced challenges in working with data and communicating statistical ideas. A breadth of topics will be covered: randomized experiments, observational studies, tests of significance, confidence intervals, two-sample tests, simple linear regression, multiple regression, analysis of variance, chi-square tests, nonparametric methods, and logistic regression. Students will use the statistical software R. A major focus for this workshop is the ideas behind, and the methods for, drawing conclusions about a population from a sample in a linguistic context. At the end of this workshop you will be expected to (1) identify the major concepts related to statistical reasoning and to statistical inferences for drawing such conclusions, (2) recognize how these concepts are used in experiments and observational studies related to linguistics, (3) implement the methods yourself in statistical analyses using the methods covered, and (4) explain your methods in sufficient detail so others can reproduce your analysis results. (5) In particular, you are expected to be able to identify the appropriate statistical model or models for a given linguistic analysis, write the model in the correct notation, implement the model in a software package on a given set of data, interpret the output in the context of the study, diagnose model deficiencies, and suggest improvements to the model if necessary. Work will be a balance between understanding the concepts underlying a method, implementation of the method, and interpretation of the results. Students will be encouraged to provide their own data in workshop assignments.

**Workshop Schedule:**

|  |  |
| --- | --- |
| **Week** | **Topic** |
| **1**  Jan 8 – 12 | **Introduction**   * Course overview * Basic stats concepts * *Types of objects in R* |
| **2**  Jan 15 – 19 | **Basic Stats/dplyr**   * Continue review of basic stats * *Exploratory analysis* |
| **3**  Jan 22 – 26 | **Statistics Fundamentals**   * Drawing statistical conclusions * Inference using t-distributions * Assumptions of statistical tests * *Data transformation with dplyr* |
| **4**  Jan 29 – Feb 2 | **Stats Fundamentals cont.**   * Alternatives to t-tools * *Data visualization with ggplot* |
| **5**  Feb 5 – 9 | **Stats Fundamentals cont.**   * Alternatives to t-tools * *Project management in R* * *Group assignment: practice w/t-tools* |
| **6**  Feb 12 – 16 | **Comparisons Among Several Samples**   * ANOVA * *Classes in R (strings, factors, numerics)* |
| **7**  Feb 19 – 23 | **Linear Combinations and Multiple Comparisons**   * *Group assignment: practice w/ANOVA* * *Classes in R (dates and times)* |
| **8**  Feb 26 – March 1 | **Midterm**   * Intro to regression * *How to use pipes to condense code* |
| **9**  March 11 – 15 | **Simple Linear Regression**   * Regression cont. * *Individual project: practice w/regression* * *Writing functions in R* |
| **10**  March 18 – 22 | **Simple Linear Regression cont.**   * Closer look at assumptions of regression * *Iteration in R* |
| **11**  March 25 – 29 | **Multiple Regression**   * Multiple regression * *Group assignment: practice w/multiple regression* * *R markdown lesson* |
| **12**  April 1 – 5 | **Regression cont.**   * Intro to quadratics, log-transformations, and interactions * *Individual project: practice w/ regression and knitting markdowns* * *R markdown continued* |
| **13**  April 8 – 12 | **ANOVA Two-Way Classifications**   * Two-way classifications * *R workshop (make-up day)* * *Group assignment: practice w/ANOVA for two-way classification* |
| **14**  April 15 – 19 | **Intro to Logistic Regression**   * Introduction to logistic regression * *Group assignment: practice w/logistic regression* |
| **15**  April 22 – 26 | **Review**   * Logistic regression cont. * Review of course material * *Open R workshop (Q&A)* |

**Textbooks:**

* *Introduction to Statistics & Data Analysis, 3rd edition* (Peck, Olsen, and Devore)  
  <https://www.spps.org/cms/lib/MN01910242/Centricity/Domain/859/Statistics%20Textbook.pdf>
* *Introductory Statistics* (openstax)  
  https://assets.openstax.org/oscms-prodcms/media/documents/IntroductoryStatistics-OP\_i6tAI7e.pdf
* *R for Data Science* (Grolemund and Wickham)  
  https://r4ds.had.co.nz