Statistics for Linguistics

Session 01

Welcome & Organisational Stuff

Who are we?

Dominic Schmitz

- Master of Arts in Linguistics @ University of Cologne, Germany
- PhD Student & Research Assistant @ HHU Düsseldorf, Germany
- PhD Project: *Final /s/ in English – Acoustic Detail in Morphology*
- Areas of Expertise:
  - phonetics, phonology, morpho-phonetics, morpho-phonology, morphology, linear discriminative learning (LDL)
  - statistical analysis of empirical data in linguistics
- Part of the research group ‘Gauging the 'generic' in masculines with discriminative learning’

Who are we?

Janina Esser

- Mathematician
- Studied Business Mathematics @ University of Cologne, Germany, and @ TU Dortmund, Germany
- Areas of Expertise:
  - analysis, linear algebra, mathematical programming, numerical analysis, stochastics
- Author of mathematical non-fiction & Assistant for Virtual Reality Software Development
- Part of the research group ‘Gauging the 'generic' in masculines with discriminative learning’

Who are you?

Tell us about yourself:

- Name
- Affiliation
- What do you study? / What's your area of expertise?
- What's your personal experience with statistics?
- **What do you wish to learn during our workshop?**

Organisational Stuff

- Please use a headset if available
- Zoom sessions
  - open at 9.30am (BST)
  - start at 10.00am (BST)
- Mute your microphone whenever you are not speaking
- We will have bio breaks regularly

Organisational Stuff

- Internet issues / connection problems
  - on your side: please try to re-connect; message us via e-mail in case of long-term problems
  - on our side: please try to re-connect within 10 minutes; e-mail in case of long-term problems
- We will use breakout sessions for group work

Learning Objectives

- Introduction to R and RStudio
  - Scripts, variables, vectors, functions, data frames, packages, etc.
- Data Visualisation
  - 'The Grammar of Graphics' in R – working with 'ggplot2'
- Simple and Multiple Linear Regression
  - Mathematical background, modelling procedure, diagnostics
- Linear Mixed-Effects Regression
  - Mathematical background, modelling procedure, issues, diagnostics

Learning Objectives

- Introduction to R and RStudio – **Session 02**
  - Scripts, variables, vectors, functions, data frames, packages, etc.
- Data Visualisation – **Session 03**
  - 'The Grammar of Graphics' in R – working with 'ggplot2'
- Simple and Multiple Linear Regression – **Session 04 & 05**
  - Mathematical background, modelling procedure, diagnostics
- Linear Mixed-Effects Regression – **Sessions 06 & 07**
  - Mathematical background, modelling procedure, issues, diagnostics

Learning Objectives

- Introduction to R and RStudio – **Session 02**
  - Scripts, variables, vectors, functions, data frames, packages, etc.
- Data Visualisation – **Session 03**
  - 'The Grammar of Graphics' in R – working with 'ggplot2'
- Simple and Multiple Linear Regression – **Session 04 & 05**
  - Mathematical background, modelling procedure, diagnostics

Day 1

- Linear Mixed-Effects Regression – **Sessions 06 & 07**
  - Mathematical background, modelling procedure, issues, diagnostics

Day 2

Materials

- All materials (slides, markdown files with explanations and exercises, data sets, references, etc.) are available as part of our R package:

Schmitz, Dominic & Esser, Janina. (2021). SfL: Statistics for Linguistics. R package version 0.1. URL: https://github.com/dosc91/SfL

- We will introduce the package and its usage in our “Introduction to R and RStudio” session

GatherTown

- We offer a GatherTown space for
  - meet-ups of participants during breaks, before and after sessions
  - for meetings in-between days 1 and 2

- To access our GatherTown space, use the following link:

  https://gather.town/app/SxqgXn09XGiXCBU3/FLsc2021

Q&A

- We offer a Q&A platform to ask questions (anonymously) in-between the first and the second day of the workshop:

  https://app.sli.do/event/nhwqxw5j/live/questions

- Others can upvote questions already asked, i.e. this shows us (the lecturers) what questions are potentially interesting to most of you
- The most upvoted / important questions are considered in the recap part of day 2
- Questions not answered during the recap will be answered via email and/or in individual meetings

Questions?