# Statistics for Linguists
# 08 July 2022

| | |
|---|---|
| 10:00 | Workshop introduction |
| 10:15 | Loading and exploring datasets |
| 10:45 | Data transformation and coding |
| 11:15 | Practical exercise |
| 12:15 | Review of practical |
| 12:30 - 13:30 | LUNCH BREAK |
| 13:30 | lmer and glmer |
| 14:30 | Post-hoc analysis and model visualization |
| 15:00 | Practical exercise |
| 16:00 | Review of practical |
| 16:15 | Model building |
| 17:00 | End of workshop |

# Statistics for Linguists

# Practical exercise 1

Margreet Vogelzang – mv498@cam.ac.uk

# Practical exercise 1

1. Load the dataset *psycholinguistics_data.csv*
2. Remove the first two columns (names with ’X’)
3. Which conditions are there? What do you think they may mean?
4. How many trials did the experiment have?
5. Calculate the mean reading time per condition, per participant, and per item
6. ’ trialID’ should be a categorical variable, not an integer. Recode this

# Practical exercise 1

7. Plot a boxplot of the reading times. How many outliers does this function identify?
8. How many outliers are there if you take 2.5 * SD (standard deviation) as the cutoff point?
9. Code the levels of the capitalization and determiner variables so that they are sum coded and their baselines (reference levels) are *cap* and *det*
10. Examine the distribution of the reading times measure through density plots. How are the data skewed?
11. Which transformation might be good to apply? Try this out.