9.59 Lab in Psycholinguistics: Problem Set #4

In this assignment, you will analyze data from an experiment that was run on Amazon Mechanical Turk. In this experiment we tested how and whether the placement of a verb particle (i.e., the position of a particle preposition relative to a verb, for verbs like "put up", "throw out", ...) depends on the length (in number of words) of the direct object of the verb. We show participants four kinds of sentences, in which the particle either comes early or late, and the direct object is either long or short. This is called a 2x2 design. The independent variables are: Particle position (early or late), and object length (long or short).

Example Sentences:

- Joe threw the documents out. (late-short)
- Joe threw the very important documents that he brought home out. (late-long)
- Joe threw out the documents. (early-short)
- Joe threw out the very important documents that he brought home. (early-long)

Your task is to produce a full analysis of this experiment and write a paragraph fit for publication in the results section of a cognitive science journal paper, with plots. You should use simple linear regression to analyze the data.

1. Load the pset4_particle_shift_data.csv file into R. Exclude all participants from the analysis i) whose home country is not USA (Answer.country), whose native language is not English (Answer.English), iii) and who did not answer at least 90% of the item comprehension questions correctly (Correct). Also, throw out any individual data points (not the whole participants!) that have NA for Answer.Rating. Report the number of remaining rows in the data frame after excluding these participants and data points.

2. The column Condition simultaneously encodes the two independent variables. It would be better to have one independent variable per column. Use `separate()` to split the column Condition into two columns based on the position of the character.

3. Transform the grammaticality ratings (Answer.Rating) into z-scores with means and standard deviations estimated within subjects.

4. Make a bar plot with the means for each condition and their 95% confidence intervals (plot raw means, not z-scores). Map one independent variable to the x axis and a fill color, and split the data by the other independent variable using `facet_wrap` or `facet_grid`.

5. Define two dummy-coded predictors based on the independent variables (early vs late, long vs not long). What will their coefficients fit? What will the coefficient of the intercept fit?

6. Fit a least squares regression (using `lm()` or `glm()`) to the data predicting z-scored judgments from the dummy coded predictors based on the independent variables and their interaction. Use the summary function applied to the output of the `lm` function to get the model output. Briefly describe the result.

7. Use the coefficients the model outputs to calculate the predicted group means for the four cells of the design. How far off are they from the actual group means?

9.59J/24.905J Lab in Psycholinguistics
Spring 2017

For information about citing these materials or our Terms of Use, visit .