SQMB 2022/23 Formative Assessment F3

EXAM NUMBER

2023-03-13

DEADLINE: THURSDAY 6 APRIL AT NOON

1 Overview

In this assessment, you'll be working with corpus data on the English dative alternation. In English, a number of ditransitive verbs (e.g., *give*, *feed*, *teach*) have two different ways to express the recipient of the action:

- as a **prepositional phrase (PP)**: The child gives a treat *to the dog*.
- as a **noun phrase (NP)**: The child gives *the dog* a treat.

In these examples, *the child* is the giver, *the dog* is the recipient, and *a treat* is the theme.

The data you'll be analysing can be found in `data/dative.csv`. It comes from a 2005 study by Joan Bresnan and colleagues, and it was gathered from the Switchboard corpus and the Treebank Wall Street Journal collection.

The data frame contains the following columns:

- `RealizationRec`: Whether the sentence's recipient is realised as a noun phrase (NP) or a prepositional phrase (PP).
- `Verb`: The verb in the sentence.
- `AnimacyOfRec`: Animacy of the recipient (animate vs. inanimate).
- `AnimacyOfTheme`: Animacy of the theme (animate vs. inanimate).
- `LengthOfTheme`: Number of words that make up the theme.

Your task is to answer the following research question: **What roles do the animacy of the theme and the length of the theme play for the realisation of the recipient?**

Some pointers:

- **Think about the variables that'll go into the analysis:** Which is the outcome variable? What kind of variable is it? And which are the predictors? What kinds of variable are they?
- **Summarise the relevant data:** Pick appropriate summary measures depending on the nature of the variable (mean, median, mode; standard deviation, range).
- **Plot the data:** Use appropriate plots to visualise patterns in the data that are relevant to the research question above. Concisely describe the patterns you see.
- **Prepare the data for analysis:** How will you set up your outcome variable? How will categorical predictors be coded/ordered? What will this mean for their interpretation? Will you transform numeric predictors, and if yes, how?
- **Fit a model to the data:** What distribution family should you use? What model formula will you use? (Optional: What does the mathematical model specification look like?)
- **Report the model and its estimates:** Describe the model that you fit. Summarise its estimates/posteriors in writing and using plots.

- **Address the research question:** Interpret the model's estimates. What do they mean for the research question?

When you are satisfied, you can **submit your project** by:

1. rendering your Rmd file to PDF, and
2. uploading it to Turnitin via Learn.