

Problem Set 2:

LING280 Experimental methods Winter 2013

Instructions

Submit your response to mwagers@ucsc.edu by January 22, 2pm. Please format it legibly and intelligently.

Background

In their paper on the dative alternation, Bresnan & Nikitina* consider what factors condition the production of dative versus double object constructions, as in:

1. John shot the goalie the puck.
2. John shot the puck to the goalie.

Some relevant factors include argument pronominality, discourse accessibility, definiteness, person and animacy. Bresnan and Nikitina first annotated a corpus for any factors which might matter (based on the proposals in the literature). They then created a statistical model which predicted the likelihood of a given form occurring, given the values of the various factors. For example, if the recipient argument is a pronoun, then it is less likely to appear as the object of PP, compared to if the recipient were a noun. By the end of the course, you will be able to build a model like Bresnan and Nikitina's. But first we must understand what is meant by a statistical model and what its building blocks are.

In this problem set, you will sharpen your understanding of the binomial distribution. We use the binomial distribution when the event we measure in our experiment has two states (or can be conceived of as having two states). For example, the VP headed by *give* is either V DP DP or V DP PP, in the same way a coin is either heads or tails. The binomial distribution forms the link between observable outcomes in an experiment to the underlying binary event. It thus allows us to quantify the likelihood of their occurrence. For example: if we toss a coin 100 times, and 60 heads come up, would we want to hold on to the hypothesis that the coin is fair? Suppose you tossed a coin 10 times, and got 6 heads?

Circling back to B&N, one way of thinking about their achievement is to say: suppose I think of producing a GOAL/THEME VP like flipping a fair coin, where heads corresponds to one alternant (say, V DP PP) and tails corresponds to the other. Now, as I start learning details about the arguments involved – that the theme is given, that the recipient DP is a much longer constituent than the theme DP – I will begin weighting the coin. Finally, once I've added in all the factors, I flip the coin.

In this problem set, you will pick one of the factors that B&N mention in their article or a factor that interests you otherwise. You will then present 10 - 20 individuals with six dative-alternating sentence pairs and ask them to select which they prefer. Finally you will estimate the effect your factor has on producing the dative.

Step 1: Run an experiment

First, pick a potential factor. In this example, I will vary whether the goal/recipient is a pronoun or not. Then, create six versions of the alternation that vary along that factor.

1. Version 1

- | | |
|--|---------------|
| (a) John gave Bill a book, John gave a book to Bill. | GOAL: NOUN |
| (b) John gave him a book, John gave a book to him. | GOAL: PRONOUN |

2. Version 2

- | | |
|---|---------------|
| (a) The secretary sent Ellen a letter, The secretary sent a letter to Ellen | GOAL: NOUN |
| (b) The secretary sent me a letter, The secretary sent a letter to me | GOAL: PRONOUN |

*Joan Bresnan and Tatiana Nikitina. 2009. "The Gradience of the Dative Alternation." In *Reality Exploration and Discovery: Pattern Interaction in Language and Life*, edited by Linda Uyechi and Lian Hee Wee, 161–184. Stanford: CSLI Publications.

3. Version 3

- (a) The president handed the general the award, The president handed the award to the general GOAL: NOUN
- (b) The president handed her the award, The president handed the award to her GOAL: PRONOUN

Think carefully about how you select the words that are not directly affected by the manipulation. With your 12 sentence pairs in hand, create two lists. One list should consist of alternating a/b pairs, e.g., Version 1, Pair A; Version 2, Pair B; Version 3, Pair A, etc Another list should consist of alternating b/a pairs, e.g., Version 1, Pair B; Version 2, Pair A, etc. Thus neither list should contain both pairs from a given sentence version.

Find 10 to 20 individuals and ask them to choose which form they prefer for each of the pairs on a list. Present half of your respondents with the pairs from one list; half with pairs from the other list. I leave you some latitude in how you select the individuals and how you present the question to them. But you should record and justify the details in your write-up. Importantly you should obtain a set of binary preference responses from each individual.

Step 2: Estimate

Combine responses across individuals. Compute the proportion of V-DP-PP outcomes for each value of the factor you manipulated.

We will set aside differences among individuals and sentence versions in this assignment. However if you observe any interesting variation among participants or items, please comment on it.

Step 3: Simulation Exercise #1

Now consider one level of your factor as a 'baseline.' Simulate many draws from a very large population for which the true likelihood of a V-DP-PP outcome is identical to the 'baseline' proportion you computed in Step 2. The size of each draw should correspond to the number of responses you collected (throughout the entire experiment).

What is the mean simulated outcome? How does this relate to the underlying population likelihood?

Based on the simulation outcome, compute the cumulative density function for your baseline value. Where does the observed outcome for the other value of your factor fall along this function?[†] What is the interval that includes 95% of the simulated outcomes around the mean outcome?

What can you conclude about the effect of your factor on realization of the dative alternation?

Step 4: Simulation Exercise #2

Take for granted that the differences you observed in your experiment are real. In other words, assume there are two underlying "populations": realizations of the dative alternation when, e.g., the recipient is a pronoun, versus realizations of the dative alternation when the recipient is a noun.

What is the fewest number of observations you would need to collect so that you could assert, with a conventional level of confidence, that the populations are non-identical? Argue for your answer via simulation.

General Guidelines

Write up this assignment using clear, connected and coherent prose which incorporates effective graphical supports. Throughout reflect on how you would design an experiment similar to this one differently in the future. Append an R transcript which has been commented for clarity. Append your experimental materials.

[†]If you find identical response patterns for the two values of your factor (i.e., the same proportion of V DP PP responses), set a hypothetical reference value, e.g., 0.5. In other words, shift your argument from whether your factor had an effect, to whether realizing the dative alternation is like flipping a fair coin.