Practical exercise (recuperación): Pronoun interpretation

Instructions

This is a *recuperación* for the practical exercise. You can do this exercise even if you passed the previous one. The highest grade will count. If you are already satisfied with the outcome of the first practical exercise, then don't hand in anything. Use the remainder of your time to work on your final report. Make sure you have all its details worked out by the end of the course.

As before, you will find a dataset prepared for you if you modify the **bolfaced** portion of the following link:

https://raw.githubusercontent.com/MetodesEmpirics/MetodesEmpirics.github.io/main/material/2022q3/session08/data/csvs/**insert_your_student_ID**.csv

Each student has their own dataset, identified by their student ID. For instance, if your student ID is **u103672**, then your link is:

https://raw.githubusercontent.com/MetodesEmpirics/MetodesEmpirics.github.io/main/material/2022q3/session08/data/csvs/**u103672**.csv

You can then load your data as

```
df <- read.csv('replace_this_with_URL_to_your_data.csv')
```

The data is described below.

Submit your answers to the questions as a PDF file, through Aula Global (upload task in **grup gran**). Round all answers to the second decimal after the comma.

Data description

Variation in referential form, such the use of either a proper name (Peter) or a more reduced pronominal expression (he), is an integral part of language use. The question what drives speakers' choices and listeners' interpretations to decide for one expression or interpretation over another is a long-standing and –to this date– open debate. On the hearers' side, the literature suggests that the boldfaced referent is more likely to be interpreted as the intended referent of the pronoun in the sentence in (1) than in (2).

1. **Sally** amazed Sara because she won the race.
2. Sally amazed **Sara** because she won the race.

In other words, people generally take *she* to refer to the subject in this construction: Sally is more often interpreted as the one who won the race. However, matters are reversed for (3) and (4):

3. **Sally** scolded Sara because she was late.
4. Sally scolded **Sara** because she was late.

That is, in this construction, people tend to take *she* to refer to the object: Sara is more often interpreted as the one who was late. In other words, verb type seems to matter when interpreting a pronoun. In the literature, verbs like *scold* are called ICV2 verbs. These are implicit causality verbs that bias the interpretation toward the object. Analogously, verbs like *amaze* are called ICV1 verbs. They are also implicit causality verbs, but they bias the interpretation toward the subject. Beyond verb type, most research (but not all!)

suggests that there should be stronger bias to interpret a pronoun as refering to the object if the sentence is passive instead of active.

Your colleague has collected data from subjects that were asked to read sentences like (1) and (3), and asked whether they think the pronoun refers to the subject or the object. The experiment had 5 different pairs of ICV1 and ICV2 sentences. Each sentence pair consisted of an active and a passive version of the sentence. For instance, the sentence in (1) and its passive alternative "Sara was amazed by Sally because she won the race" were one of five pairs of ICV1 sentences. The sentence in (3) and its passive alternative was one of the five ICV2 sentences. 12 different subjects read each pair. Your colleague recorded whether they interpreted the ambiguous pronoun as refering to the subject (1) or the object (0); and saved the results in a CSV file.

Here's how the data may look like:

```
##   sentence.ID  status subject type
## 1           1  active       1 icv1
## 2           1 passive       1 icv1
## 3           2  active       1 icv1
## 4           2 passive       1 icv1
## 5           3  active       1 icv1
## 6           3 passive       1 icv1
```

The column `sentence.ID` identifies which of the 5 sentence pairs the interpretation is for. The column `status` shows whether the interpretation is for the active (`active`) or passive (`passive`) version of a pair. The column `subject` shows the interpertation of the pronoun: 1 if it was interpreted as refering to the subject; 0 if it was interpreted as the object. The column `type` specifies whether this is a ICV1 (`icv1`) or an ICV2 construction (`icv2`).

Questions

Help your colleague analyze the data from her experiments! Remember to round all answers to the second decimal after the comma.

Descriptive statistics (0.5 points each)

1. What is the median interpretation of pronouns as refering to the subject in active ICV1 sentences?
2. What is the variance of the interpretation rate of pronouns as refering to the subject in active ICV1 sentences?
3. What is the median interpretation of pronouns as refering to the subject in passive ICV1 sentences?
4. What is the variance of the interpretation rate of pronouns as refering to the subject in passive ICV1 sentences?
5. What is the median interpretation of pronouns as refering to the subject in active ICV2 sentences?
6. What is the variance of the interpretation rate of pronouns as refering to the subject in active ICV2 sentences?
7. What is the median interpretation of pronouns as refering to the subject in passive ICV2 sentences?
8. What is the variance of the interpretation of pronouns as refering to the subject in passive ICV2 sentences?

Inferential statistics (1 point each)

Using two regression models, with `status` as predictor and `subject` as outcome, estimate

1. The difference between passive and active ICV1 sentences; the standard error of this difference; and the probability of a pronoun referring the the object in a passive sentence.
2. The difference between passive and active ICV2 sentences; the standard error of this difference; and the probability of a pronoun referring to the object in a passive sentence.
3. Based on your answer to (1): Do you think there is evidence to suggest a difference between active and passive ICV1 sentences? Why (not)?
4. Based on your answer to (2): Do you think there is evidence to suggest a difference between active and passive ICV2 sentences? Why (not)?
5. Your colleague suggests to compare the models used for (1) and (2) using the Akaike Information Criterion. Is this a good idea? If so, which model performs better? If not, why and how could you compare the two models performance instead?
6. Do your findings agree with the expectations from the literature? Explain why (not).