

Binary answers to OR-questions

Hypothesis

There is quite a common pattern when one person ask another a question like "Would you like A or B" and gets an answer like "Yes" or "No". But which argument was meant by the answering person? Can we predict which was the chosen argument? What does this choice depend on? We can suppose that there are some factors influencing the answer, which may be:

- gender of the respondent;
- positiveness / negativeness of the answer;
- the existance of some context before the question.

Assumption :

There are cases when one has to choose from two options and one is much more favorable for the respondent; in this case if that argument is in a comfortable-to-define-by-yes position (let's assume it is the first position), the answer will be "Yes" meaning something like "Yes, the first option"; otherwise it will be "No" meaning "No, not the second option".

Now let's formulate some hypotheses:

- the chosen argument can be predicted by the factors' values;
- context is the most significant factor;
- people of different genders prefer to mean different argument's postion when they use a binary answer.

Research design

Testing of the mentioned hypotheses requires some data which has to be collected manually. The process will be described further. Speaking of how the collected data will look like:

reviewer	gender	question	answer	chosen_position	context_provided
1	m	1	Y	1	N
1	m	2	N	2	Y
2	f	1	Y	2	N

First hypothesis

Three models will be used to check the first hypothesis:

- logistic regression;
- decision tree;
- random forest.

Having a null hypothesis like "We can predict the chosen position using these factors with accuracy more than 80%", the models will be tested to check if the hypothesis has been correct.

Second hypothesis

Most important factor will be found using the same predescribed models.

Third hypothesis

To test the hypothesis formulated as "The distributions of the position choice for different genders is not the same" a two-sample t-test will be used.

Data collection

To collect the data to run the models based on a google form with questions will be made. The gender of the reviewer will be

asked firstly and then there will be questions of this kind:

- Your friend asked you: "What'd you like on your sandwich: cheese or ham?" Assuming you wanted cheese, how would you answer: "Yes" (meaning "Yes, cheese") or "No" (meaning "No, not ham")?

Or this one:

- You've discussed the recent news about your favourite football team. Then a friend yours asked: "Would you like to play football or volleyball?" Assuming you wanted to play football, how would you answer: "Yes" (meaning "Yes, football") or "No" (meaning "No, not volleyball")?

This form of questions has been chosen instead of "... Assuming that you answered 'Yes', which one did you mean?" so that personal preferences won't affect the results.