What influences co-operative decision making?

Dr. Ulrik Beierholm, Associate Professor of Psychology, Durham University, UK

The Dataset

In order to examine the factors influencing co-operative decision making, a researcher collected data from 282 subjects participating in a one-shot Prisoner's Dilemma Game, as well as responses from a set of questionnaires. The Prisoner's Dilemma Game originates from economic game theory where two players are faced with a choice to collaborate or defect, with both players getting an economic payout that depends on the choice of both players.

For some of the games played, the opponent was a human, for some, a computer (see below).

The dataset PDPMetaUni.csv contains responses from the 282 subjects and is similar (but not identical) to the data published in Taheri, Rotshtein & Beierholm et al. 2018 https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0205730. This dataset contains more mixed aged subjects, but only has one response per subject (hence one-shot). The variables included in the dataset are:

Age: Age of the subjects in years.

Anxiety: Measure of anxious attachment style, averaged from the Revised Adult Attachment Scale.

Avoidance: Measure of avoidant attachment style, averaged from the Revised Adult Attachment Scale.

Subject: Unique ID for each subject.

Choice: The choice made by the subject in the Prisoner's Dilemma Game (1 = co-operate, 2 = defect).

TypeHum: Indicates the type of opponent (1 = Human, 0 = Computer).

We are going to look at the data, and compare different generalized linear models, specifically logistic models, to explain the data.

Activity 1

Loading the dataset PDPMetaUni.csv into R is the first essential (and obvious) step in our analysis.

Activity 2

Displaying the first few rows and summarizing the dataset helps us understand its structure, identify any missing values, and get an overview of the variables and their distributions.

Activity 3

There are some issues with the data set. First off, we need to ensure that for our subsequent logistic regression the output variable is binary (zero or one). Secondly, we need to remove the data for which we do not have the age of the participants (encoded as NaN in the data).

Activity 4

Creating various plots, such as histograms and scatter plots, allows us to visualize the distributions and relationships between variables. This is an important part of developing a model, and can often help to avoid later problems. Let’s create a histogram for Age, a plot of Anxiety vs Avoidance, and a barplot for the binary Choice01.

Activity 5

Calculate the Spearman correlation between Anxiety and Avoidance.

Activity 6

Calculate the mean and standard deviation for the three variables.

Activity 7

Our main goal is to construct a logistic regression model with Choice as the outcome variable and other variables as predictors. This will help us understand the relationship between the predictors and the binary outcome. This model can potentially allow us to make predictions and assess the influence of different factors on the subjects’ choices. Reminder: generalized models have almost the same syntax as linear models, but use ‘glm’ instead of ‘lm’! Remember to specify ‘family=binomial’ as our output variable is binary.

Activity 8

It is clear from the model above that not all the variables are actually important for explaining that data. Let’s try a different model that does not include the questionnaire scores on Anxiety and Avoidance, as they seem to have a weak effect. We can then use Akaike Information Criteria (AIC) to compare the models (remember, the smaller the AIC the better).

Activity 9

Plot the logistic regression curve against Age, to see how the model's predictions vary with age and whether there are any trends or patterns. Include the individual data points. To create the curve you may need to create a new data frame with simulated data (use data.frame) and do prediction from the model (use predict).

Activity 10

Plotting the predicted probabilities from the logistic regression model helps visualize the model's output and assess its performance. Normally residual plots can be very useful, but for binary output data they will always show deviations, and they are therefore not recommended. An alternative is to bin the predicted probabilites (e.g. in 10 bins), then for each bin calculate the average residuals. There are many ways to do this (including using some nifty packages) but here we will use 'predict', 'cut' and 'tapply' to first predict values, cut them into bins, and then calculate the mean for each bin.

Concluding remarks

Working with logistic models is a little different, especially in terms of how to evaluate a model. Above, we have given examples of how to examine the output of a model using base R, although binned plots can be more easily created through the ARM package, for example, while evaluating generalized linear models can be done through the DHARMa package.