The MATLAB script leave\_one\_out\_analysis.m performs leave-one-out procedure to model the relationship between multiple predictors and a single outcome. Predictors and outcome can be numerical or categorical. There are 4 types of analysis that are supported:

1. Basic linear regression;
2. Stepwise linear regression;
3. Generalized linear regression (for cases when the outcomes are not distributed normally);
4. Decision tree (a classification method for categorical outcomes).

When you run the script, it asks you for the Excel file which contains the data. The file must contain the sheet called to\_analyze (all other sheets will be ignored). That sheet needs to be organized as follows: the first column is the participant IDs, the last column is the outcomes, and the columns in between are predictors. The top row contains the names of predictors and the outcome. There should be no missing values in the sheet. Please look at the attached examples loo\_regression\_data.xlsx and loo\_classification\_data.xlsx (the first has a numerical outcome, and the second has a categorical outcome).

After selecting the Excel file, the script asks for the analysis type (1 to 4). Then, the script operates in two stages:

1. Leave-one-out analysis

Here, we perform several iterations of the analysis: at each iteration, we remove one observation from the data set, estimate the model using all other observations, and then predict the outcome for the removed observation. Then, we put it back into the data set, remove the next observation, and repeat the procedure until we go through all observations At the end, we have predicted the outcome for all observations; the accuracy of these predictions is a measure of how our analysis generalizes to previously unseen data. This accuracy is reported at the end of the leave-one-out procedure: either as Pearson’s correlation between actual and predicted outcomes (for the three regression methods), or as percentage of correct predictions of outcome category (for the decision tree). In addition, the script also displays the scatterplot of actual versus predicted outcomes (for the regression methods) or the confusion matrix (for the decision tree). Also, the script saves the predicted outcome values in a text file.

For basic and generalized linear regression, we record which predictors were significant (p<0.05) at each iteration. Likewise, for stepwise regression and decision trees, we record which predictors were selected at each iteration. At the end of the leave-one-out procedure, the script reports the predictors which were significant/selected in at least 5% of the iterations. For each predictor, it reports the percentage of iterations when this predictor was significant / selected, and the mean weight of the predictor across these iterations (for the three regression models, this weight is a t statistic). If a variable is very important, it is significant / selected during the majority of iterations, and also has a high absolute weight.

1. Analysis on the full data set

After we are done with leave-one-out procedure, we run the selected method on the full data set, and display the results. For the regression method, we give the full report (F statistic, unadjusted / adjusted R-square, RMS, weights and significance of each predictor, etc.). For the decision tree, we report classification accuracy on the full set (which should be at least as high as the leave-one-out classification accuracy) and the importance of each predictor.

If some predictors are redundant, the script will generate a warning. Redundancy might cause numerical issues and the model estimation might be inaccurate. For example, if the set of predictors contains WAB Spontaneous Speech Score, Auditory Verbal Comprehension Score, Repetition Score, Naming / Word Finding Score, and Aphasia Quotient (AQ), then the last score can be computed from the other four scores and therefore AQ is redundant. In that case, it is recommended to remove AQ from the list of predictors (it does not contain any information that is not already captured by the other predictors).