1 Analysis plan

1.1 Data preparation

Goal

Prepare data for regression analysis.

Steps

- Code all categorical variables as factors, taking into account all possible categories (including those that do not appear in any of the answers)
- Consistent coding of `NA`s
- Transform all categorical variables into sets of binary (dummy) variables
  - Keep a list to define which binary variables are part of which categorical variable
- Define groups of variables:
  - Response variables
  - Explanatory variables:
    - Binary variables
    - Continuous variables
  - Demographic and auxiliary covariates (gender, age, occupation, etc.)
  - Other variables not taken into account for statistical analysis (comments, etc.)

Result

Analysis-ready data containing ID, response variables, explanatory variables, and covariates.

1.2 Data exploration

1.2.1 Correlation

Goal

Arrive at a first understanding of relationships between variables, and potential clusters of interrelated variables.

Steps

- Pairwise correlation coefficients (Spearman's) between:
- response variables and other variables
- Among all explanatory variables and covariates

Results

- Heat map for each response variable
- Heat map for explanatory variables and covariates
- List of most influential variables (when considered in isolation)

1.2.2 Variable selection

Goal

Define a subset of explanatory variables to be considered for further statistical analysis, based on the strength of their relationship with the response variables.

Steps

- For each response variable, set up a Bayesian multinomial model with all explanatory variables (covariates are excluded).
- Define a horseshoe prior over the explanatory variables (using a proportion of ??)
- For all binary variables with non-zero effect: Identify corresponding categorical variable

Result

List of explanatory variables to be taken into account for further statistical analysis.