**CONFIRMATORY STRATEGIES – NOTES 2020\_10\_21**

**SIMULATION DATA**

* The problem in the code is in the coding of the true guilty suspect – fix that (make it as in the human data)
* The problem seems to be that there is a column in the human data that is not there in the simulated data
* Also columns in Step4 that do not match the data:
  + Suspect (is the number coding reversed?)
  + Action\_type: there is one row always missing for every file
  + Count\_red\_samples and Count\_blue\_samples do not match the data
* Suggestion: make a separate version of step 2 called Step2\_Simulation that takes care of all the differences with human data, after that step 3 and step 4 should be exactly the same as the human files
* Add comments for steps 2-3-4 to add more details about what happens

**HUMAN DATA**

* Run the BATCH file to include the new datasets
* Fix the problem with the maximum number of investigation rounds – that corresponds to more than 60 columns (60 search + 1 accuse + 1 move to next trial + X evidence found, sometimes more than one). Keep it to 70 to be on the safe side for steps 1-3

**GRAPHS**

* Add the script that automatically saves the images as PDF
* Find a way save the individual results (e.g. table of values to display) in an output file (all of them in a single summary file, stored under different variable names, not one file for every analysis)

**2-D line [to show the relation between two continuous variables]**

EXAMPLE: Pr(accuse red) as a function of Pr(red)

DESCRIPTION: show the probability of accusing the red suspect as a function of the posterior (in the accusation action). Show the overall function (black) and also show separately the function for the trials in which the prior is <0.4 (blue) and >0.6 (red)

Input:

* GENERAL:
  + Title = Action probability
  + range x-axis = [0, 1]
  + range y-axis = [0, 1]
  + tick x-axis = [0:0.25:1]
  + tick y-axis = [0:0.25:1]
  + label x-axis = Updated Pr(Red)
  + label y-axis = Pr(Accuse Red)
* LINES:
  + x-values coordinates: 0:0.05:1 (vector of 21 numbers equally spaced)
    - NOTE: we will use this x-values coordinates for ALL the lines. We use these 21 bins to group continuous variables into easier-to-show values
  + values-matrix (each row in the matrix is a line to show)
    - From the SummaryStatistics script, create the matrix with 3 rows (lines to draw) and 21 columns (21 points for every line). Select only the accusation actions, and start by saving the action (accuse red or blue) and the posterior (that is a number between 0 and 1)
    - The posterior is from a continuous space, but we want to use only 21 bins (described above as x-values coordinates). In order to do so, group the continuous values based on the closest bin-center (so assign 0 to all the values below 0.025, assign 0.05 to all the values between 0.025 and 0.075, etc.)
    - In this way you generate the row for the line to be displayed in black
    - Repeat the procedure for the two subsets with low (blue) and high (red) initial probability of red being guilty
* line-colors (one color for every line): [black, red, blue]
* line- styles (one style for every line): [solid, solid, solid]
* line- width (one width for every line): [2, 1, 1]

**Barplot [to show the comparison between different groups, e.g. human vs model]**

EXAMPLE: Human vs Model prediction for the probability of finding the final evidence

DESCRIPTION: compare the probability of finding the evidence for humans and SimulatedDatasetA

Input:

* GENERAL:
  + Title = Probability of finding Evidence
  + range x-axis = [0.3, 1.7]
  + range y-axis = [0, 1]
  + tick x-axis (position of the bars) = [1, 2]
  + names tick x-axis (labels) = [Human, Bayes]
  + tick y-axis = [0:0.25:1]
  + label y-axis = Pr(Evidence)
* BARPLOT:
  + values-vector (vector, one value per bar)
    - From the SummaryStatistics script, values for the human data and the Simulation A (neutral decision maker)
* barplot-colors (one color for every bar): [red, blue]
  + width = 0.4

**Histogram [to show distributions – vertical lines represent comparison between different groups, similar to the values comparisons shown in a barplot]**

EXAMPLE: Distribution of the average number of samples

DESCRIPTION: for every participant, calculate the average number of samples collected

Input:

* GENERAL:
  + Title = Average number of samples
  + range x-axis = [0, 10]
  + range y-axis = [0, 50]
  + tick x-axis = [0:1:10]
  + tick y-axis = [0:10:50]
  + label x-axis = Number of samples
  + label y-axis = Frequency
* HISTOGRAM:
  + values-vector (vector, one value per participant)
    - Modify the SummaryStatistics script in order to get these values at the subject level, not only at the aggregate level
  + number of bins (integer number): 21
  + histogram-color: grey
* VERTICAL LINES (one line only)
  + x-values (vector): average of all the values displayed in the histogram
  + colors (one color for every line): red
  + line width (one for every line): 2
  + line style (one for every line): solid line

**Heatmap [to display a 3-dimensional distribution]**

EXAMPLE: Probability of voting RED as a function of prior (x-axis) and cost (y-axis).

DESCRIPTION: Calculate for every COST/PRIOR combination the probability of voting red

Input:

* GENERAL:
  + Title = Pr(Vote Red)
  + tick x-axis = [0.1:0.2:0.9]
  + tick y-axis = [5, 10, 20, 40, 80]
  + label x-axis = Initial Pr(Red guilty)
  + label y-axis = Search Cost
* HEATMAP:
  + values (matrix, each cell is a value to display)
    - Add this analysis to the SummaryStatistics, calculate the average action within each condition
  + color palette (default, red=1, blue=0)

**Scatterplot [to display the relation between two variables]**

EXAMPLE: Pr(confirmatory) vs Avg number of samples  
DESCRIPTION: for every participant, calculate the probability of using a confirmatory evidence (across all the rounds) and the average number of samples collected

Input:

* GENERAL:
  + Title = Subjects Heterogeneity Analysis
  + range x-axis = [0, 1]
  + range y-axis = [0, 10]
  + tick x-axis = [0.1:0.2:0.9]
  + tick y-axis = [0:1:10]
  + label x-axis = Pr(confirmatory)
  + label y-axis = Avg number of samples
* SCATTERPLOT:
  + values (matrix with two columns=variables, each row is a participant)
    - Modify the SummaryStatistics script in order to get these values at the subject level, not only at the aggregate level
  + dot size = 10
  + dot color = grey