WCST data analysis - Step 2

# Manipulate data

## Read data frames and manually rated data

We have loaded data frames saved by the Step 1 script, and merged the manually categorized free text answers about the rule with participantdata\_inc.

## Exclude participants

We excluded extra participants if there were more than 78 in a condition.

## Modify data

The subjective difficulty rating supposed to be a number 1-10, but it was also a free text answer, and some participants went outside the recommended range (e.g.: difficulty = 50). We capped these numbers to 10.

# Descriptive statistics

The number of participants was 78 in each condition.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Condition | Number of participants | Number of solvers | Solution rate | Median task time (min) | Aha rate of solvers |
| Letters In | 78 | 78 | 1.00 | 1.16 | 0.82 |
| Letters Out | 78 | 73 | 0.94 | 2.12 | 0.93 |
| No Letters | 78 | 50 | 0.64 | 8.24 | 0.98 |
| Fixed Sequence | 78 | 69 | 0.88 | 4.80 | 0.88 |
| Ambiguous Cards | 78 | 41 | 0.53 | 13.37 | 0.90 |
| Non-Standard | 78 | 70 | 0.90 | 3.20 | 0.76 |
| Letters Only | 78 | 78 | 1.00 | 0.70 | 0.58 |

# Statistical tests

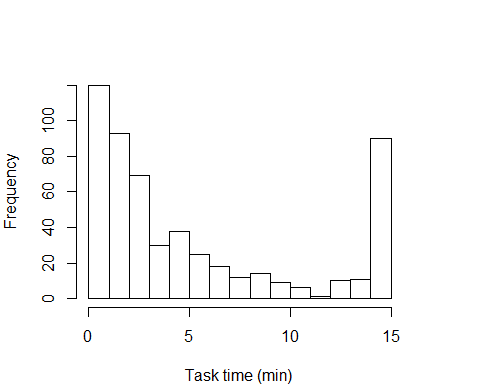
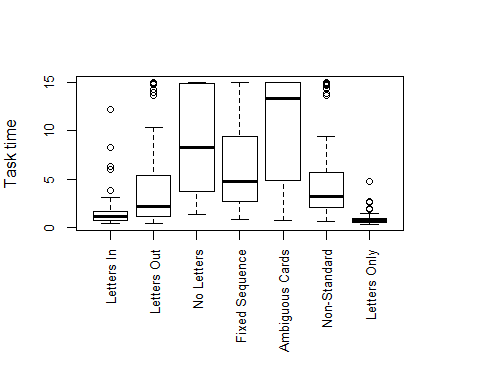
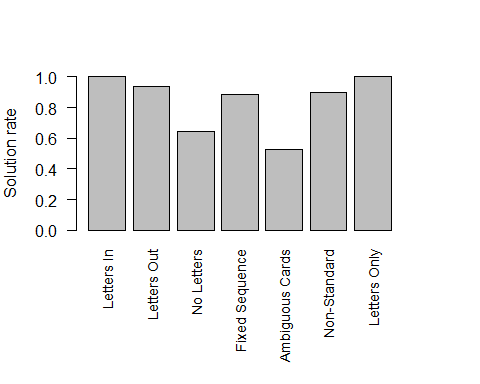
## Difficulty of the task

A task is more difficult in one condition than in another condition, if less participants are able to solve it, or if it takes longer for participants to solve it (even if solution rates are the same). This is why, we compared the difficulty of the task

* first, by Fisher’s exact test on the number of solvers (pairwise comparisons)
* then, if the Fisher’s exact test was not significant, we compared task time by two-sample Wilcoxon test (same as the Mann-Whitney test)
* **ANOVA with Greenhouse-Geisser correction: robust also for non-normally distributed data**
* **We also used binary logistic regression.**

The following plots compare all conditions of Experiment 1 and 2 and the control condition:

* The barplot shows solution rates accross conditions.
* The histogram shows the distribution of task time in all the conditions.
* The boxplot shows the distribution of task time accrosss condition. Circles represent outliers. Whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range from the box. If the notches of two boxes do not overlap this is strong evidence that the two medians differ.



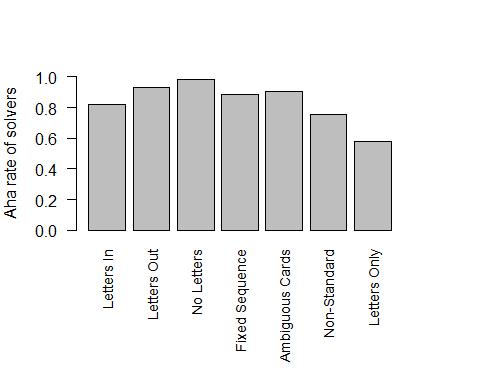
The histogram of task times shows that the data is not normally distributed, so we had to use nonparametric tests. The Kolmogorov-Smirnov normality test on task time also showed that the data was not normally distributed.

One-sample Kolmogorov-Smirnov test: participantdata\_inc %>% select(task\_time)

|  |  |  |
| --- | --- | --- |
| Test statistic | P value | Alternative hypothesis |
| 0.6558 | 0 \* \* \* | two-sided |

## Aha feelings

We used Fisher’s exact tests to make pairwise comparison between conditions regarding the Aha-feelings of solvers (whether they reported having aha-feelings or not).



# Results - Experiment 1

## Difficulty of the task - Experiment 1

We analyzed the contingency table containing the number of solvers and non-solvers in pairs of conditions with Fisher’s exact test. A p<0.05 means that the row/column association is statistically significant:

WLIN-WLOUT: **not significant**

Fisher’s Exact Test for Count Data: partic\_groups2[c("wlin", "wlout"), c("Nbof\_solvers", "Nbof\_nonsolvers")]

|  |  |  |
| --- | --- | --- |
| P value | Alternative hypothesis | odds ratio |
| 0.05851 | two.sided | Inf |

WLOUT-WNOL: **significant**

Fisher’s Exact Test for Count Data: partic\_groups2[c("wlout", "wnol"), c("Nbof\_solvers", "Nbof\_nonsolvers")]

|  |  |  |
| --- | --- | --- |
| P value | Alternative hypothesis | odds ratio |
| 8.206e-06 \* \* \* | two.sided | 8.07 |

WLIN-WLOUT task times: **significant**

Wilcoxon rank sum test with continuity correction: wlin\_time and wlout\_time

|  |  |  |
| --- | --- | --- |
| Test statistic | P value | Alternative hypothesis |
| 1624 | 5.055e-07 \* \* \* | two.sided |

We conclude, that the difficulty of conditions in Experiment 1 was:

* WLIN < WLOUT
* WLOUT < WNOL

## Aha feelings - Experiment 1

We analyzed the contingency table containing the number of solvers who reported Aha-feelings and number of solvers who did not report Aha-feelings in pairs of conditions with Fisher’s exact tests. We compared all conditions to the control condition (Letters Only). A p<0.05 means that the row/column association is statistically significant:

WLIN > WONLY: **significant**

Fisher’s Exact Test for Count Data: partic\_groups2[c("wlin", "wonly"), c("Nbof\_ahas\_solvers", "Nbof\_noahas\_solvers")]

|  |  |  |
| --- | --- | --- |
| P value | Alternative hypothesis | odds ratio |
| 0.001533 \* \* | two.sided | 3.326 |

WLOUT > WONLY: **significant**

Fisher’s Exact Test for Count Data: partic\_groups2[c("wlout", "wonly"), c("Nbof\_ahas\_solvers", "Nbof\_noahas\_solvers")]

|  |  |  |
| --- | --- | --- |
| P value | Alternative hypothesis | odds ratio |
| 4.403e-07 \* \* \* | two.sided | 9.826 |

WNOL > WONLY: **significant**

Fisher’s Exact Test for Count Data: partic\_groups2[c("wnol", "wonly"), c("Nbof\_ahas\_solvers", "Nbof\_noahas\_solvers")]

|  |  |  |
| --- | --- | --- |
| P value | Alternative hypothesis | odds ratio |
| 7.237e-08 \* \* \* | two.sided | 35.22 |

Levels:

WLOUT > WLIN: **not significant**

Fisher’s Exact Test for Count Data: partic\_groups2[c("wlout", "wlin"), c("Nbof\_ahas\_solvers", "Nbof\_noahas\_solvers")]

|  |  |  |
| --- | --- | --- |
| P value | Alternative hypothesis | odds ratio |
| 0.05001 | two.sided | 2.955 |

WNOL > WLOUT: **not significant**

Fisher’s Exact Test for Count Data: partic\_groups2[c("wnol", "wlout"), c("Nbof\_ahas\_solvers", "Nbof\_noahas\_solvers")]

|  |  |  |
| --- | --- | --- |
| P value | Alternative hypothesis | odds ratio |
| 0.3994 | two.sided | 3.572 |

## Described rules - Experiment 1

## data frame with 0 columns and 1 row

# Results - Experiment 2

## Difficulty of the task - Experiment 2

Our predictions about the difficulty of the task were:

* WNOLFS < WNOLA: The sequence rule can be used in both conditions but the exclusion rule can only be used in the WNOLFS condition, which might help.
* MOONSQ < WNOLA: Only the sequence rule can be used in both conditions. Removing distracting visual cues makes finding the sequence rule easier

We analyzed the contingency table containing the number of solvers and non-solvers in pairs of conditions with Fisher’s exact test. A p<0.05 means that the row/column association is statistically significant:

WNOLFS-WNOLA: **significant**

Fisher’s Exact Test for Count Data: partic\_groups2[c("wnolfs", "wnola"), c("Nbof\_solvers", "Nbof\_nonsolvers")]

|  |  |  |
| --- | --- | --- |
| P value | Alternative hypothesis | odds ratio |
| 1.213e-06 \* \* \* | two.sided | 6.829 |

MOONSQ-WNOLA: **significant**

Fisher’s Exact Test for Count Data: partic\_groups2[c("moonsq", "wnola"), c("Nbof\_solvers", "Nbof\_nonsolvers")]

|  |  |  |
| --- | --- | --- |
| P value | Alternative hypothesis | odds ratio |
| 3.694e-07 \* \* \* | two.sided | 7.786 |

## Aha feelings - Experiment 2

We compared all conditions of Experiment 2 with the control condition (Letters Only).

WNOLFS > WONLY: **significant**

Fisher’s Exact Test for Count Data: partic\_groups2[c("wnolfs", "wonly"), c("Nbof\_ahas\_solvers", "Nbof\_noahas\_solvers")]

|  |  |  |
| --- | --- | --- |
| P value | Alternative hypothesis | odds ratio |
| 3.758e-05 \* \* \* | two.sided | 5.526 |

WNOLA > WONLY: **significant**

Fisher’s Exact Test for Count Data: partic\_groups2[c("wnola", "wonly"), c("Nbof\_ahas\_solvers", "Nbof\_noahas\_solvers")]

|  |  |  |
| --- | --- | --- |
| P value | Alternative hypothesis | odds ratio |
| 0.0003037 \* \* \* | two.sided | 6.687 |

MOONSQ > WONLY: **significant**

Fisher’s Exact Test for Count Data: partic\_groups2[c("moonsq", "wonly"), c("Nbof\_ahas\_solvers", "Nbof\_noahas\_solvers")]

|  |  |  |
| --- | --- | --- |
| P value | Alternative hypothesis | odds ratio |
| 0.02419 \* | two.sided | 2.273 |

# Other ideas

## Language

## Restructuring

Sliding average for move time

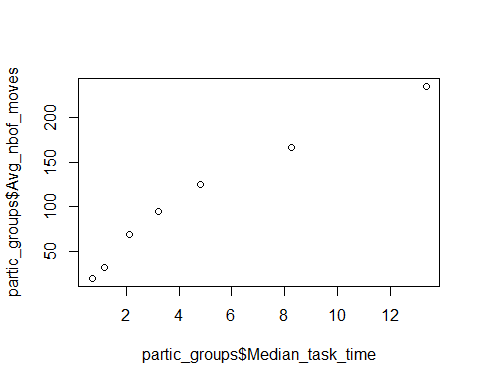
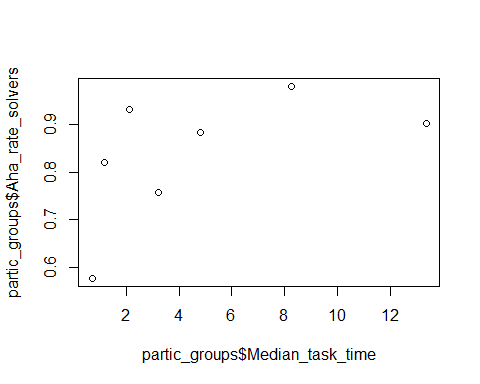
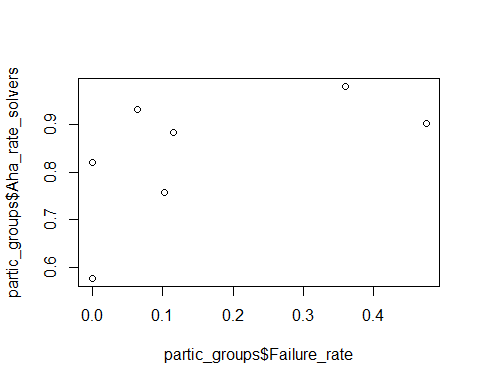
## Rules used (matching rule for moves)

## Correct rule

Does it correlate with problem difficulty or aha

## Aha-rate of non solvers

Correlates with difficulty? Number of false ahas vs subjective difficulty or solution time



## Goal of the experiment