

Analysis

Michal Pospěch

07 December, 2021

Load libraries

```
if(!require(tidyverse)){install.packages("tidyverse"); library(tidyverse)} # bundle of packages commonl
if(!require(here)){install.packages("here"); library(here)} # package for handling paths in the project
if(!require(ez)){install.packages("ez"); library(ez)} # package for running ANOVAs
if(!require(saccades)){install.packages("saccades"); library(saccades)} # package for detecting fixation

if(!"devtools" %in% rownames(installed.packages())) install.packages("devtools")
# Install the stable development versions from GitHub
devtools::install_github("crsh/papaja")
library(papaja)

source("utils.R") # functions are in separate file for easier debugging
```

Download data

Download zip file from the web page. Description of the dataset is on the page ETRA dataset description

```
data_pth <- "data"
if(!dir.exists(data_pth)) {
  dir.create(data_pth)
}
# uncomment following file for downloading data
if(!file.exists(here(data_pth, "ETRA2019Challenge.zip"))){
  options(timeout=300)
  download.file("http://smc.neuralcorrelate.com/ETRA2019/ETRA2019Challenge.zip", here(data_pth, "ETRA2019
  unzip(here(data_pth, "ETRA2019Challenge.zip"), exdir = data_pth)
}
```

Than unzip data and move them to the folder. The directory *data* should therefore contain following directories/files

- data
- images
- DataSummary.csv

Load data

You can use *readr* package, which is supplied within *tidyverse* package bundle. Multiple files can be read using these lines of code. For example, we want only load Free viewing condition in Puzzle subtask for subject 9

```
local_data_pth <- here(data_pth, "data", "009")
files <- dir(local_data_pth, pattern = ".*Fixation_Puzzle_*.csv") # get file names
```

```
df_fix_puzzle <- files %>%
  map(read_data) %>%
  reduce(rbind) %>% as_tibble()

local_data_pth <- here(data_pth, "data", "009")
files <- dir(local_data_pth, pattern = ".*Fixation_Waldo_*.csv") # get file names

df_fix_waldo <- files %>%
  map(read_data) %>%
  reduce(rbind) %>% as_tibble()
```

You can easily adjust this code for your hypotheses.

Hypotheses

In this section, state 2-3 hypotheses. For example, we might want to test, whether fixation duration would differ between Freeviewing and Where is Waldo. The hypothesis would be therefore stated as

1. There will be differences in fixation duration for participant 9 between Puzzle subtask and Where is Waldo subtask

Of course, the hypotheses should be slightly more complex (we want to do similar test for all participants, not just 1).

Data manipulation

This is an optional section, in which you could describe, what did you do with the data to obtain given format. In our example, we just need to merge the data. Additionally, we want to detect fixations

```
df_hyp1_samples <- rbind(df_fix_puzzle,df_fix_waldo) %>%
  rename(time = Time, trial = trial_id, x = LXpix, y = LYpix) %>%
  group_by(participant_id, trial) %>%
  mutate(time = time - (min(time))) %>%
  ungroup()

df_hyp1_fix <- df_hyp1_samples %>%
  group_by(participant_id, trial) %>%
  do(detect.fixations(.))

df_hyp1_fix <- df_hyp1_samples %>% select(participant_id, trial, fv_fixation, task_type, stimulus_id) %>%
```

Some subtask might have different columns, so you need to use dplyr *select* or *mutate*, otherwise rbind won't work. But hey, data manipulation takes significant portion of time in each analysis.

Results

In this section, describe statistical test that you used for testing your hypotheses. In general, the selection of statistical test depends on the type of variable.

There are following types of variables:

- Continuous - when the variable behaves as number. All fixation durations, pupil size, time are continuous variables
- Ordinal - variables do not behave as numbers, but you can order them. Grades in school are typical example. You can't say how many times is 1 better than 2, but but you can say that 1 is better grade than 2. There are no nominal variables in this dataset, so I added this description only for completeness.

- Nominal - when variables are qualitative different. Type of task is an example of nominal variable

In the case of two variables, there are following options * both variables continuous - regression or correlation (in R function `lm()` or `cor/cor.test`) * both variables nominal - contingency tables a chi square test * one variable nominal, other continuous - this is very common, this type is used, when we compare differences two condition - in this case, we use t.tests

There are three main types of t.tests

- Independent t-test - both groups contain independent data points (each data point is a different entity)
- Paired t-test - data points are linked to each other. This is typical example, when we measure same subjects multiple times
- One sample t-test - we are testing the sample against some theoretically interesting number

In case of more than two variables, we need to use ANOVAs. Usually, one variable will be dependent (outcome, the one which levels interest us) and others will be independent (predictors, the one, that we manipulate). * For more than two groups, we use between-subject ANOVA * For more than two measurement of same subject, we use within-subject ANOVA * We can combine multiple between- and within-subject factors into mixed ANOVA

In our case, simplest way how to test that is to aggregate data per each trial and use t-tests

using t-tests

Because we have multiple data points from each participant, we can first aggregate data for each trial

```
t_test_result <- df_hyp1_fix %>%
  group_by(task_type, trial) %>%
  summarize(dur = mean(dur)) %>%
  t.test(dur~task_type,.)
```

`summarise()` has grouped output by 'task_type'. You can override using the `.groups` argument.

Results show that there is no difference between Fixation duration in Where is Waldo and Puzzle ($t(27.92) = -0.03$, $p = .977$)