04 CSI online aphasia: Spoken - Descriptives

Kirsten Stark

16 März, 2022

Load packages

```
library(dplyr)
```

```
##
## Attache Paket: 'dplyr'

## Die folgenden Objekte sind maskiert von 'package:stats':
##
##     filter, lag

## Die folgenden Objekte sind maskiert von 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)

rm(list = ls())
```

Load and preprocess data

```
options( "encoding" = "UTF-8" )

# input
input <- "aphasia_patient_TeaP.csv"

# load data
df <- read.csv(here::here("data", "transient_data_files", input), sep = ",",  na = "")
```

Duration of the experiment

```
print("Total duration, not outlier corrected")
```

```
## [1] "Total duration, not outlier corrected"
```

```r
mean(df$timetotal, na.rm=TRUE) # 36,27 min
```

```
## [1] 36.27857
```

```r
sd(df$timetotal)
```

```
## [1] 26.48572
```

```r
range(df$timetotal) # 19 to 126 min
```

```
## [1]  19.23333 125.98333
```

```r
print("Total duration, split by session")
```

```
## [1] "Total duration, split by session"
```

```r
df %>% group_by(session) %>% summarise(mean = mean(timetotal),
                                       sd = sd(timetotal),
                                       min = min(timetotal),
                                       max = max(timetotal))
```

```
## # A tibble: 3 x 5
##   session  mean    sd   min   max
##     <int> <dbl> <dbl> <dbl> <dbl>
## 1       1  28.6  5.05  22.4  39.5
## 2       2  39.0 35.6   19.4 126.
## 3       3  41.3 26.8   19.2  90.6
```

Description of participants

Gender:

```r
df <- df %>% mutate(gender_char = case_when(gender == 1 ~ "female",
                                            gender == 2 ~ "male"))
table(df$gender_char)/160 # 1 = female, 2 = male, 3 = diverse
```

```
##
## male
##   21
```

```r
print("percentage female:")
```

```
## [1] "percentage female:"
```

```r
sum(df$gender == 1)/nrow(df)
```

```
## [1] 0
```

Age:

```r
print("mean:"); mean(df$age)
```

```
## [1] "mean:"
```

```
## [1] 53.57143
```

```r
print("sd:"); sd(df$age)
```

```
## [1] "sd:"
```

```
## [1] 5.925579
```

```r
print("range:"); range(df$age)
```

```
## [1] "range:"
```

```
## [1] 45 64
```

Handedness:

```r
# 1 = left handed, 2 = right handed, 3 = ambidexter/both
df <- df %>% mutate(handedness_char = case_when(handedness == 1 ~ "left-handed",
                                                handedness == 2 ~ "right-handed"))
table(df$handedness_char)/160
```

```
##
##  left-handed right-handed
##            3           18
```

```r
print("percentage right-handed:")
```

```
## [1] "percentage right-handed:"
```

```r
sum(df$handedness == 2)/nrow(df)
```

```
## [1] 0.8571429
```

Mother tongue (experiment was restricted to native German speakers): This seems to have worked

```r
table(df$language) # 1 = yes (mother tongue is German), 2 = no
```

```
##
##    1
## 3360
```

```
table(df$language.test) # 1 = der, 2 = die, 3 = das (das is correct)
```

```
## < table of extent 0 >
```

Attention checks

1) Item vs. non-item

```
## Item vs. non-item
# CH01_01 (Taube), CH01_02 (Apfel), CH02_01 (Luftballon) and CH02_02 (Biene) are items and 2 should be
# CH02_03 (Radio), CH02_04 (Sparschwein), CH02_03 (Laptop) and CH02_04 (Wattestäbchen) are non-items an
## Did participants cheat
# CH03 = 1 - yes, I worked through it till the end,
# CH03 = 2 - no, I stopped or cheated midway
# CH03 = -9 - no answer

attcheck <- data.frame(subject = unique(df$subject))

df <- df %>% mutate(itemvsnonitem1 =
                      case_when(CH01_01==2 & CH01_02==2 & CH01_03==1 & CH01_04==1 ~2,
                                CH01_01==2 || CH01_02==2 ~1,
                                CH01_01!=2 & CH01_02!=2 ~0)) %>%
  mutate(itemvsnonitem2 =
                      case_when(CH02_01==2 & CH02_02==2 & CH02_03==1 & CH02_04==1 ~2,
                                CH02_01==2 || CH02_02==2 ~1,
                                CH02_01!=2 & CH02_02!=2 ~0))
table(df$itemvsnonitem1)/160
```

```
##
##  1  2
##  4 17
```

```
table(df$itemvsnonitem2)/160
```

```
##
##  1  2
##  3 18
```

```
# attcheck <- data.frame(subject = unique(df$subject))
#
#
# data <- data %>% mutate(attcheck =
#                   ifelse(CH01_01 == 2 & CH01_02 == 2 & CH02_01 == 2 & CH02_02 == 2 &
#                   CH01_03 == 1 & CH01_04 == 1 & CH02_03 == 1 & CH02_04 == 1, 1, 0)) %>%
#                    mutate(cheat = ifelse(CH03 == 1,1,ifelse(CH03 == 2,2,0)))
# data.frame(data$subject, )
# table(data$attcheck)
# table(data$cheat)
#
# # get prolific IDs of participants who failed the attention check
# #pretest %>% subset(attcheck == 0 & cheat == 2 ) %>%
```

```
# #  pull(SD24_01) # SD24_01 is prolific ID
#
# # subset to participants who passed only
# valid <- pretest  %>% filter(attcheck == 1 & cheat != 2)
#
# inspect <- data.frame(df$subject, df$word, df$fam_typed)
```

2) Cheating

```
df <- df %>% mutate(CH03 = case_when(CH03 == 1 ~
                                       " Ja, ich habe alles bis zum Ende bearbeitet.",
                                       CH03 == 2 ~
                                       "Nein, ich habe zwischendurch aufgehoert oder geschummelt."))
table(df$CH03)/160
```

```
##
##  Ja, ich habe alles bis zum Ende bearbeitet.
##                                         21
```

Comments

Comments don't indicate any problems that should lead to participant exclusion:

```
table(df$comments)/160
```

```
##
##  2 NA
##  1  2
```

Fully anonymize data and reduce data frame

```
df_a <- df %>% dplyr::select(!"gender" & !starts_with("language") &
                               !starts_with("TIME") &
                               !starts_with("handedness") &
                               !starts_with("fingers") &
                               !starts_with("browser") &
                               !starts_with("operator_system") &
                               !starts_with("system") &
                                !starts_with("KB") &
                               !starts_with("CH0") & !starts_with("X.") &
                               !starts_with("screen") & !starts_with("os") &
                               !starts_with("questionnaire") & !"OR01_01" &
                               !"comments" &
                               !"FINISHED" & !"Q_VIEWER" &
                               !"LASTPAGE" & !"MAXPAGE" &
                               !"DEG_TIME"& !"type" &
                               !"gender_char" & !"itemvsnonitem1" &
                               !"itemvsnonitem2" &
                               !"QUESTNNR" & !"STARTED" & !"OR02_01" &
                                !"SD22_BID" & !starts_with("MC") &
                                !"audio"& !"name" & !"audio_vot")
```

```
write.csv(df_a, here::here("data","data_long_anonymous.csv"))
```