Descriptive Statistics

Erin M. Buchanan

Last Update 2022-03-09

Libraries

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(psych)
library(tidyr)
library(rio)
set.seed(58902)
current_year <- 2022

# Suppress summarise info
options(dplyr.summarise.inform = FALSE)
```

Import the Files

```
trials <- import("../05_Data/output_data/trial_data.csv")
participants <- import("../05_Data/output_data/participant_data.csv")
items <- import("../05_Data/output_data/item_data.csv")
priming <- import("../05_Data/output_data/prime_data.csv") #also later for 2.5 and 3.0
```

We will import these files from our data folder for the final analyses. This example analysis examines the Semantic Priming Project data to simulate how to do some of the analyses.

```
SPP_participant <- import("../02_Power/subjectdataLDT.zip")
SPP_participant <- SPP_participant %>%
  filter(target.ACC == 1) %>%  # only correctly answered trials
  filter(rel != "nw") %>% # exclude nonword trials
  filter(!is.na(Ztarget.RT)) %>%  # exclude NAs
  select(rel, Ztarget.RT, target, Subject, Trial)
```

Participant Demographics

- Number of participants excluded
- All this information calculated after participants are excluded for less than 80%
- Separate table provided for the excluded participants (to do later)
- Overall and by language (included later) demographics

```
# gender
table(participants$please_tell_us_your_gender, useNA = "ifany")
```

```
##
## female   male notsay  other
##     76     17      2      3
```

```
# age
describe(current_year - as.numeric(participants$which_year_were_you_born), na.rm = T)
```

```
##    vars  n  mean   sd median trimmed  mad min max range skew kurtosis   se
## X1    1 93 24.89 4.74     23   23.88 2.97  21  41    20 1.77     2.35 0.49
```

```
# education level
table(participants$please_tell_us_your_education_level, useNA = "ifany")
```

```
##
##     college   doctorate high school      master
##          19           7          66           6
```

```
# native language
table(participants$native_language, useNA = "ifany")
```

```
##
##            Bosnian               Dari Dari (afghanistan)             engish
##                  1                  1                  1                  1
##            english    english--- test           Georgisk             german
##                  4                  1                  1                  1
##             German           Japanese              norsk              Norsk
##                  2                  1                  3                  7
##             norway          norwegian          Norwegian             Polish
##                  1                  8                 58                  1
##               sami         Sinhalease               spam            spanish
##                  1                  1                  1                  1
##            Spanish               Urdu
##                  1                  1
```

```
# sample size by lab
table(participants$url_lab, useNA = "ifany")
```

```
##
## 1234 2005 9000 9999 <NA>
##    4    1   90    2    1
```

```
# information about computer
table(participants$meta_platform, useNA = "ifany")
```

```
##
## Linux x86_64     MacIntel        Win32
##            1           51           46
```

```
# information about web browser
participants$browser <- sapply(strsplit(participants$meta_user_agent,
                                        split = " "), tail, 1)
participants <- participants %>%
  separate(col = browser,
           into = c("browser", "browser_version"),
           sep = "/",
           remove = TRUE)

table(participants$browser, useNA = "ifany")
```

```
##
##     Edg Firefox     OPR  Safari
##       6      10       2      80
```

```
# language locale versus language they took it in
table(participants$meta_locale, useNA = "ifany")
```

```
##
##    de de-DE en-GB en-US    ja    nb nb-NO nn-NO    pl
##     1     1     6    11     1    57    19     1     1
```

```
# number of excluded participants
table(participants$keep)
```

```
##
## exclude    keep
##       8      90
```

Trial Level Data

- After excluding participants, calculate these statistics ... you have to leave in the bad trials for accuracy and final time stamp.

```
# amount of time the study last line tells you the length
final_trials <-
  trials %>%
  filter(keep_participant == "keep") %>%
  group_by(observation) %>%
  arrange(desc(timestamp)) %>%
  filter(!duplicated(observation))
```

```
describe(final_trials$time_commit / 1000 / 60 )
```

```
##    vars  n  mean   sd median trimmed mad  min   max range skew kurtosis   se
## X1    1 90 18.48 4.96  18.52   18.33 2.9 2.52 37.89 35.38 0.26     3.51 0.52
```

```
# number of trials by word type
# accuracy of trials by word type
trials %>%
  filter(keep_participant == "keep") %>%
  group_by(class) %>%
  summarize(number_trials = n(),
            accuracy_trials = sum(correct, na.rm = T)/n())
```

```
## # A tibble: 2 x 3
##   class   number_trials accuracy_trials
##   <chr>           <int>           <dbl>
## 1 nonword         17142           0.807
## 2 word            35118           0.962
```

- Now we exclude those bad trials and calculate:

```
# response latencies by word type
# SE by word type
describe_trials <-
  trials %>%
  filter(keep_participant == "keep") %>%
  group_by(class) %>%
  filter(keep == "keep")  #also take out bad trials
  #(this is really handled by Z_RT being NA, but doesn't hurt)

describeBy(describe_trials$Z_RT, group = describe_trials$class)
```

```
##
##  Descriptive statistics by group
## group: nonword
##    vars     n mean   sd median trimmed  mad   min  max range skew kurtosis   se
## X1    1 13826 0.36 1.06   0.08    0.18 0.62 -1.69 15.4 17.09  2.9    14.58 0.01
## ------------------------------------------------------------
## group: word
##    vars     n  mean   sd median trimmed  mad   min   max range skew kurtosis
## X1    1 33795 -0.15 0.94  -0.38    -0.3 0.49 -2.03 14.53 16.55 3.63    23.82
##      se
## X1 0.01
```

```
describe_trials <-
  trials %>%
  filter(keep_participant == "keep") %>%
  group_by(class) %>%
  filter(abs(Z_RT) < 2.5) %>%
  filter(keep == "keep")  #also take out bad trials
  #(this is really handled by Z_RT being NA, but doesn't hurt)

describeBy(describe_trials$Z_RT, group = describe_trials$class)
```

```
##
##  Descriptive statistics by group
## group: nonword
##    vars     n mean   sd median trimmed  mad   min max range skew kurtosis   se
## X1    1 13200 0.19 0.67   0.04    0.11 0.57 -1.69 2.5  4.19 1.04     0.89 0.01
## ------------------------------------------------------------
## group: word
##    vars     n  mean   sd median trimmed  mad   min max range skew kurtosis se
## X1    1 33014 -0.24 0.64   -0.4   -0.33 0.47 -2.03 2.5  4.52 1.43     2.49  0
```

```
describe_trials <-
  trials %>%
  filter(keep_participant == "keep") %>%
  group_by(class) %>%
  filter(abs(Z_RT) < 3.0) %>%
  filter(keep == "keep")  #also take out bad trials
  #(this is really handled by Z_RT being NA, but doesn't hurt)

describeBy(describe_trials$Z_RT, group = describe_trials$class)
```

```
##
##  Descriptive statistics by group
## group: nonword
##    vars     n mean   sd median trimmed  mad   min  max range skew kurtosis   se
## X1    1 13397 0.23 0.73   0.05    0.13 0.59 -1.69 2.99  4.68 1.25      1.6 0.01
## ------------------------------------------------------------
## group: word
##    vars     n  mean   sd median trimmed  mad   min  max range skew kurtosis se
```

```
## X1    1 33230 -0.23 0.68  -0.39   -0.32 0.48 -2.03 2.99  5.02 1.63     3.43  0
```

Item Level Data

- All data has been filtered at this point:

```
# average sample size at the item level
describeBy(items$samplesize, group = items$class)
```

```
##
##  Descriptive statistics by group
## group: nonword
##    vars    n  mean    sd median trimmed  mad min max range skew kurtosis  se
## X1    1 1319 10.47 14.39      4    7.29 4.45   0  89    89 1.94     3.52 0.4
## ------------------------------------------------------------
## group: word
##    vars    n  mean    sd median trimmed   mad min max range skew kurtosis   se
## X1    1 1774 19.05 19.88     12   15.95 13.34   0 149   149 1.45      2.2 0.47
```

```
describeBy(items$Z2.5_samplesize, group = items$class)
```

```
##
##  Descriptive statistics by group
## group: nonword
##    vars    n mean    sd median trimmed  mad min max range skew kurtosis   se
## X1    1 1177 11.2 14.26      4    8.13 4.45   1  88    87 1.86     3.22 0.42
## ------------------------------------------------------------
## group: word
##    vars    n mean    sd median trimmed   mad min max range skew kurtosis   se
## X1    1 1765 18.7 19.42     12   15.68 13.34   1 139   138 1.43        2 0.46
```

```
describeBy(items$Z3.0_samplesize, group = items$class)
```

```
##
##  Descriptive statistics by group
## group: nonword
##    vars    n  mean    sd median trimmed  mad min max range skew kurtosis   se
## X1    1 1183 11.31 14.41      4    8.21 4.45   1  88    87 1.85     3.13 0.42
## ------------------------------------------------------------
## group: word
##    vars    n  mean    sd median trimmed   mad min max range skew kurtosis   se
## X1    1 1767 18.81 19.55     12   15.77 13.34   1 144   143 1.44     2.11 0.47
```

```
# accuracy of trials by word type
describeBy(items$accuracy, group = items$class)
```

```
##
##  Descriptive statistics by group
## group: nonword
##    vars    n mean   sd median trimmed mad min max range  skew kurtosis   se
## X1    1 1319  0.8 0.32      1    0.88   0   0   1     1 -1.69     1.45 0.01
## ------------------------------------------------------------
## group: word
##    vars    n mean  sd median trimmed mad min max range  skew kurtosis se
## X1    1 1774 0.96 0.1      1    0.99   0   0   1     1 -5.08    34.78  0
```

```
# response latencies by word type
describeBy(items$avgZ_RT, group = items$class)
```

```
##
##  Descriptive statistics by group
## group: nonword
##    vars    n mean   sd median trimmed  mad   min  max range skew kurtosis   se
## X1    1 1198 0.46 0.82   0.28    0.36 0.62 -1.11 5.06  6.18 1.66     4.35 0.02
## ------------------------------------------------------------
## group: word
##    vars    n  mean   sd median trimmed  mad   min  max range skew kurtosis   se
## X1    1 1768 -0.17 0.51  -0.27   -0.23 0.39 -1.34 3.18  4.52 1.71     5.33 0.01
```

```
describeBy(items$Z2.5_avgZ_RT, group = items$class)
```

```
##
##  Descriptive statistics by group
## group: nonword
##    vars    n mean   sd median trimmed  mad   min  max range skew kurtosis   se
## X1    1 1177 0.28 0.55   0.21    0.24 0.51 -1.11 2.49   3.6 0.82     1.16 0.02
## ------------------------------------------------------------
## group: word
##    vars    n  mean  sd median trimmed  mad   min  max range skew kurtosis   se
## X1    1 1765 -0.25 0.4  -0.31   -0.28 0.34 -1.34 2.29  3.62 1.23     3.78 0.01
```

```
describeBy(items$Z3.0_avgZ_RT, group = items$class)
```

```
##
##  Descriptive statistics by group
## group: nonword
##    vars    n mean  sd median trimmed  mad   min  max range skew kurtosis   se
## X1    1 1183 0.31 0.6   0.23    0.26 0.54 -1.11 2.73  3.84 0.98     1.53 0.02
## ------------------------------------------------------------
## group: word
##    vars    n  mean   sd median trimmed  mad   min  max range skew kurtosis   se
## X1    1 1767 -0.23 0.43   -0.3   -0.27 0.35 -1.34 2.76  4.09 1.48      5.3 0.01
```

```
# SE by word type
describeBy(items$seZ_RT, group = items$class)
```

```
##
##  Descriptive statistics by group
## group: nonword
##    vars   n mean   sd median trimmed  mad  min  max range skew kurtosis   se
## X1    1 890 0.28 0.33   0.19    0.22 0.14 0.01 4.89  4.87 5.19    48.94 0.01
## ------------------------------------------------------------
## group: word
##    vars    n mean   sd median trimmed  mad min  max range skew kurtosis se
## X1    1 1566  0.2 0.18   0.14    0.17 0.09   0 1.57  1.57    3    12.62  0
```

```
describeBy(items$Z2.5_seZ_RT, group = items$class)
```

```
##
##  Descriptive statistics by group
## group: nonword
##    vars   n mean   sd median trimmed mad  min  max range skew kurtosis   se
## X1    1 878  0.2 0.16   0.15    0.17 0.1 0.01 1.23  1.23 2.24     7.01 0.01
## ------------------------------------------------------------
## group: word
##    vars    n mean   sd median trimmed  mad min  max range skew kurtosis se
## X1    1 1566 0.16 0.13   0.12    0.14 0.07   0 1.26  1.26 3.15    15.75  0
```

```
describeBy(items$Z3.0_seZ_RT, group = items$class)
```

```
##
##  Descriptive statistics by group
## group: nonword
##    vars   n mean   sd median trimmed mad  min  max range skew kurtosis   se
## X1    1 883 0.21 0.16   0.16    0.18 0.1 0.01 1.13  1.13 2.05     5.51 0.01
## ------------------------------------------------------------
## group: word
##    vars    n mean   sd median trimmed  mad min  max range skew kurtosis se
## X1    1 1566 0.17 0.14   0.12    0.14 0.07   0 1.26  1.26 3.18    15.27  0
```

Priming Level Data

```
# average priming
describe(priming$avgZ_prime)
```

```
##    vars   n mean   sd median trimmed  mad  min max range  skew kurtosis   se
## X1    1 321 0.11 0.63   0.11     0.1 0.48 -3.9 3.6   7.5 -0.11     7.73 0.03
```

```
describe(priming$avgZ_RT_unrelated)
```

```
##    vars   n mean   sd median trimmed mad   min  max range skew kurtosis   se
## X1    1 570 -0.2 0.58  -0.34   -0.28 0.4 -1.13 3.26  4.39 2.05     6.74 0.02
```

```
describe(priming$avgZ_RT_related)
```

```
##    vars   n  mean   sd median trimmed  mad   min  max range skew kurtosis   se
## X1    1 571 -0.32 0.56  -0.43    -0.4 0.39 -1.52 3.38   4.9    2     7.06 0.02
```

```
describe(priming$samplesize_unrelated)
```

```
##    vars   n  mean   sd median trimmed  mad min max range skew kurtosis   se
## X1    1 570 11.06 14.1      4    8.04 4.45   1  61    60 1.67     1.68 0.59
```

```
describe(priming$samplesize_related)
```

```
##    vars   n  mean    sd median trimmed  mad min max range skew kurtosis   se
## X1    1 571 11.01 14.16      4    7.96 4.45   1  60    59 1.67     1.66 0.59
```

```
# total number of words with priming > 0
sum(priming$avgZ_prime > 0, na.rm = T)
```

```
## [1] 192
```

```
# also do this for the 2.5 and 3.0 trials
```

Split Half Reliability (Item)

We will split the data in half and correlate the items together. This procedure will be repeated 100 times to avoid random poor splits. We will perform this on the trial level data - here is an example calculated on the SPP data.

```
save_correlation <- rep(NA, 100)

for (i in 1:100){

  # split data in half
  SPP_participant$split <- sample(1:2, nrow(SPP_participant),
                                  replace = T)

  # summarize the data
  SPP_summary <- SPP_participant %>%
    group_by(target, rel, split) %>% #calculate by item, relation, split
    summarize(meanZ_RT = mean(Ztarget.RT))

  # pivot
  SPP_wide <- SPP_summary %>%
    pivot_wider(id_cols = target,
                names_from = c(rel, split),
                values_from = meanZ_RT) %>%
    mutate(prime_1 = un_1 - rel_1,
           prime_2 = un_2 - rel_2)

  save_correlation[i] <- cor(SPP_wide$prime_1, SPP_wide$prime_2)

}

describe(save_correlation)
```

```
##    vars   n mean   sd median trimmed  mad  min  max range skew kurtosis se
## X1    1 100 0.24 0.02   0.25    0.25 0.02 0.19 0.28  0.09 -0.7     0.08  0
```

Split Half Reliability (Person)

```
SPP_participant$Trial_even <- (SPP_participant$Trial) %% 2 == 0

SPP_summary <- SPP_participant %>%
  group_by(Subject, Trial_even, rel) %>%
  summarize(meanZ_RT = mean(Ztarget.RT)) %>%
  pivot_wider(id_cols = Subject,
              names_from = c(Trial_even, rel),
              values_from = meanZ_RT) %>%
  mutate(prime_1 = FALSE_un - FALSE_rel,
         prime_2 = TRUE_un - TRUE_rel)

cor(SPP_summary$prime_1, SPP_summary$prime_2)
```

```
## [1] 0.3600168
```