Cue Word Valence Data Example

Erin M. Buchanan

2023-03-28

### Vignette Setup:

knitr::opts\_chunk$set(echo = TRUE)  
  
# Libraries necessary for this vignette  
library(rio)  
library(flextable)  
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(tidyr)  
  
# Function for simulation  
item\_power <- function(data, # name of data frame  
 dv\_col, # name of DV column as a character  
 item\_col, # number of items column as a character  
 sample\_start = 20,   
 sample\_stop = 200,   
 sample\_increase = 5,  
 decile = .5){  
   
 DF <- cbind.data.frame(  
 dv = data[ , dv\_col],  
 items = data[ , item\_col]  
 )  
   
 # just in case  
 colnames(DF) <- c("dv", "items")  
   
 # figure out the "sufficiently narrow" ci value  
 SE <- tapply(DF$dv, DF$items, function (x) { sd(x)/sqrt(length(x)) })  
 cutoff <- quantile(SE, probs = decile)  
   
 # sequence of sample sizes to try  
 samplesize\_values <- seq(sample\_start, sample\_stop, sample\_increase)  
  
 # create a blank table for us to save the values in   
 sim\_table <- matrix(NA,   
 nrow = length(samplesize\_values),   
 ncol = length(unique(DF$items)))  
  
 # make it a data frame  
 sim\_table <- as.data.frame(sim\_table)  
  
 # add a place for sample size values   
 sim\_table$sample\_size <- NA  
  
 # loop over sample sizes  
 for (i in 1:length(samplesize\_values)){  
   
 # temp that samples and summarizes  
 temp <- DF %>%   
 group\_by(items) %>%   
 sample\_n(samplesize\_values[i], replace = T) %>%   
 summarize(se = sd(dv)/sqrt(length(dv)))  
   
 # dv on items  
 colnames(sim\_table)[1:length(unique(DF$items))] <- temp$items  
 sim\_table[i, 1:length(unique(DF$items))] <- temp$se  
 sim\_table[i, "sample\_size"] <- samplesize\_values[i]  
   
 }  
  
 # figure out cut off  
 final\_sample <- sim\_table %>%   
 pivot\_longer(cols = -c(sample\_size)) %>%   
 rename(item = name, se = value) %>%   
 group\_by(sample\_size) %>%   
 summarize(Percent\_Below = sum(se <= cutoff)/length(unique(DF$items)))   
   
 # multiply by correction   
 final\_sample$new\_sample <- round(39.369 + 0.700\*final\_sample$sample\_size + 0.003\*cutoff - 0.694\*length(unique(DF$items)))  
   
 return(list(  
 SE = SE,   
 cutoff = cutoff,   
 DF = DF,   
 sim\_table = sim\_table,   
 final\_sample = final\_sample  
 ))  
  
}

### Project/Data Title:

Cue Word Valence

Data provided by: Krystian Barzykowski

### Project/Data Description:

Participants participated in a voluntary memory task, where they were provided with word-cue in response to which they were about to recall an autobiographical memory. The item set consists of 30 word-cues that were rated/classified by 142 separate participants.

### Methods Description:

They briefly described the content of their thoughts and rated it on a 7-point scale: (a) to what extent the content was accompanied by unexpected physiological sensations (henceforth, called physiological sensation), (b) to what extent they had deliberately tried to bring the thought to mind (henceforth, called effort), (c) clarity (i.e. how clearly and well an individual remembered a given memory/mental content), (d) how detailed the content was, (e) how specific and concrete the content was, (f) intensity of emotions experienced in response to the content, (g) how surprising the content was, (h) how personal it was, and (i) the relevance to current life situation (not included).

### Data Location:

Data included within this vignette.

DF <- import("barzykowski\_data.xlsx") %>%   
 bind\_rows(import("barzykowski\_data.xlsx", sheet = 2)) %>%   
 bind\_rows(import("barzykowski\_data.xlsx", sheet = 3)) %>%   
 bind\_rows(import("barzykowski\_data.xlsx", sheet = 4)) %>%   
 bind\_rows(import("barzykowski\_data.xlsx", sheet = 5)) %>%   
 bind\_rows(import("barzykowski\_data.xlsx", sheet = 6)) %>%   
 bind\_rows(import("barzykowski\_data.xlsx", sheet = 7)) %>%   
 bind\_rows(import("barzykowski\_data.xlsx", sheet = 8)) %>%   
 bind\_rows(import("barzykowski\_data.xlsx", sheet = 9)) %>%   
 bind\_rows(import("barzykowski\_data.xlsx", sheet = 10)) %>%   
 bind\_rows(import("barzykowski\_data.xlsx", sheet = 11)) %>%   
 bind\_rows(import("barzykowski\_data.xlsx", sheet = 12)) %>%   
 bind\_rows(import("barzykowski\_data.xlsx", sheet = 13)) %>%   
 bind\_rows(import("barzykowski\_data.xlsx", sheet = 14)) %>%   
 bind\_rows(import("barzykowski\_data.xlsx", sheet = 15)) %>%   
 bind\_rows(import("barzykowski\_data.xlsx", sheet = 16)) %>%   
 bind\_rows(import("barzykowski\_data.xlsx", sheet = 17)) %>%   
 bind\_rows(import("barzykowski\_data.xlsx", sheet = 18)) %>%   
 bind\_rows(import("barzykowski\_data.xlsx", sheet = 19)) %>%   
 bind\_rows(import("barzykowski\_data.xlsx", sheet = 20)) %>%   
 bind\_rows(import("barzykowski\_data.xlsx", sheet = 21)) %>%   
 bind\_rows(import("barzykowski\_data.xlsx", sheet = 22)) %>%   
 bind\_rows(import("barzykowski\_data.xlsx", sheet = 23)) %>%   
 bind\_rows(import("barzykowski\_data.xlsx", sheet = 24)) %>%   
 bind\_rows(import("barzykowski\_data.xlsx", sheet = 25)) %>%   
 bind\_rows(import("barzykowski\_data.xlsx", sheet = 26)) %>%   
 bind\_rows(import("barzykowski\_data.xlsx", sheet = 27)) %>%   
 bind\_rows(import("barzykowski\_data.xlsx", sheet = 28)) %>%   
 bind\_rows(import("barzykowski\_data.xlsx", sheet = 29)) %>%   
 bind\_rows(import("barzykowski\_data.xlsx", sheet = 30))   
  
str(DF)

## 'data.frame': 918 obs. of 11 variables:  
## $ Participant's ID : num 12 13 14 15 16 17 18 19 20 21 ...  
## $ Cue no : num 1 1 1 1 1 1 1 1 1 1 ...  
## $ Physiological reaction: num 1 1 1 1 1 1 2 2 4 2 ...  
## $ Effort : num 4 4 7 4 1 2 4 4 2 3 ...  
## $ Vividness : num 5 6 1 4 2 7 3 6 3 3 ...  
## $ Clarity : num 2 4 1 4 6 7 4 5 4 4 ...  
## $ Detailidness : num 5 5 1 4 4 7 2 3 3 2 ...  
## $ Concretness : num 2 4 3 5 4 7 5 5 3 2 ...  
## $ Emotional Intensity : num 3 1 2 2 1 6 2 1 2 1 ...  
## $ How surprising : num 5 2 1 1 4 4 3 2 5 1 ...  
## $ Personal nature : num 2 1 4 3 1 5 1 1 2 1 ...

### Date Published:

No official publication, see citation below.

### Dataset Citation:

The cues were used in study published here: Barzykowski, K., Niedźwieńska, A., & Mazzoni, G. (2019). How intention to retrieve a memory and expectation that it will happen influence retrieval of autobiographical memories. Consciousness and Cognition, 72, 31-48. DOI: <https://doi.org/10.1016/j.concog.2019.03.011>

### Keywords:

cue-word, valence, memory retrieval

### Use License:

Open access with reference to original paper (Attribution-NonCommercial-ShareAlike CC BY-NC-SA)

### Geographic Description - City/State/Country of Participants:

Poland, Kraków

### Column Metadata:

metadata <- import("barzykowski\_metadata.xlsx")  
  
flextable(metadata) %>% autofit()

| Variable Name | Variable Description | Type (numeric, character, logical, etc.) |
| --- | --- | --- |
| Participant's ID | Participants’ identification number | Numeric |
| Cue no | Number of the specific cue they saw | Numeric |
| Physiological reaction | To what extent the content was accompanied by unexpected physiological sensations (henceforth, called physiological sensation) | Numeric (1 to 7 scale) |
| Effort | To what extent they had deliberately tried to bring the thought to mind (henceforth, called effort) | Numeric (1 to 7 scale) |
| Vividness | How vivid the thought was | Numeric (1 to 7 scale) |
| Clarity | Clarity (i.e. how clearly and well an individual remembered a given memory/mental content) | Numeric (1 to 7 scale) |
| Detailidness | How detailed the content was | Numeric (1 to 7 scale) |
| Concretness | How specific and concrete the content was | Numeric (1 to 7 scale) |
| Emotional Intensity | Intensity of emotions experienced in response to the content | Numeric (1 to 7 scale) |
| How surprising | How surprising the content was | Numeric (1 to 7 scale) |
| Personal nature | How personal it was | Numeric (1 to 7 scale) |

### AIPE Analysis:

Note that the data is already in long format (each item has one row), and therefore, we do not need to restructure the data.

#### Stopping Rule

In this example, we have multiple variables to choose from for our analysis. We could include several to find the sample size rules for further study. In this example, I’ll use the variables with the least and most variability and take the average of the 50% decile as suggested in our manuscript. This choice is somewhat arbitrary - in a real study, you could chose to use only the variables you were interested in and pick the most conservative values or simply average together estimates from all variables.

apply(DF[ , -c(1,2)], 2, sd)

## Physiological reaction Effort Vividness   
## 1.691892 1.577815 1.651065   
## Clarity Detailidness Concretness   
## 1.651066 1.696818 1.646454   
## Emotional Intensity How surprising Personal nature   
## 1.753200 1.500973 1.889276

These are Likert type items. The variance within them appears roughly equal. The lowest variance appears to be How surprising, and the max appears to be Personal nature.

Run the function proposed by the manuscript:

# set seed  
set.seed(8548)  
# Function for simulation  
var1 <- item\_power(data = DF, # name of data frame  
 dv\_col = "How surprising", # name of DV column as a character  
 item\_col = "Cue no", # number of items column as a character  
 sample\_start = 20,   
 sample\_stop = 100,   
 sample\_increase = 5,  
 decile = .5)  
  
var2 <- item\_power(DF, # name of data frame  
 "Personal nature", # name of DV column as a character  
 item\_col = "Cue no", # number of items column as a character  
 sample\_start = 20,   
 sample\_stop = 100,   
 sample\_increase = 5,  
 decile = .5)

What the usual standard error for the data that could be considered for our stopping rule?

# individual SEs for how surprising   
var1$SE

## 1 2 3 4 5 6 7 8   
## 0.2654025 0.1843656 0.2395449 0.2463527 0.2478174 0.2702634 0.2385878 0.1781742   
## 9 10 11 12 13 14 15 16   
## 0.2643095 0.2784668 0.2803546 0.2595053 0.2568655 0.2387265 0.3268158 0.2692327   
## 17 18 19 20 21 22 23 24   
## 0.3114397 0.3007016 0.2886410 0.2815891 0.2677375 0.1979609 0.2478808 0.2986490   
## 25 26 27 28 29 30   
## 0.3589899 0.2982066 0.2921602 0.3035231 0.2643255 0.3249033

# var 1 cut off  
var1$cutoff

## 50%   
## 0.2684851

# individual SEs for personal nature  
var2$SE

## 1 2 3 4 5 6 7 8   
## 0.2870153 0.2839809 0.2432371 0.3319512 0.3600411 0.4064413 0.3389511 0.3191424   
## 9 10 11 12 13 14 15 16   
## 0.2502489 0.3696715 0.3708976 0.2800560 0.2674643 0.2832563 0.4580862 0.2836613   
## 17 18 19 20 21 22 23 24   
## 0.3588157 0.3122039 0.3498098 0.3617125 0.3435939 0.3702658 0.3156137 0.2744988   
## 25 26 27 28 29 30   
## 0.4125648 0.3298670 0.3760496 0.3258969 0.2797701 0.3615385

# var 2 cut off  
var2$cutoff

## 50%   
## 0.3309091

# overall cutoff  
cutoff <- mean(var1$cutoff, var2$cutoff)  
cutoff

## [1] 0.2684851

The average SE cutoff across both variables is 0.268.

#### Minimum Sample Size

How large does the sample have to be for 80% to 95% of the items to be below our stopping SE rule?

final\_sample <- bind\_rows(  
 var1$final\_sample %>%   
 mutate(variable = "How surprising"),   
 var2$final\_sample %>%   
 mutate(variable = "Personal nature")  
)  
  
flextable(final\_sample %>%   
 filter(Percent\_Below >= .80) %>%   
 arrange(Percent\_Below, new\_sample)) %>%   
 autofit()

| sample\_size | Percent\_Below | new\_sample | variable |
| --- | --- | --- | --- |
| 40 | 0.8000000 | 47 | How surprising |
| 40 | 0.8333333 | 47 | Personal nature |
| 45 | 0.9333333 | 50 | How surprising |
| 45 | 0.9333333 | 50 | Personal nature |
| 50 | 0.9666667 | 54 | How surprising |
| 55 | 0.9666667 | 57 | How surprising |
| 50 | 1.0000000 | 54 | Personal nature |
| 55 | 1.0000000 | 57 | Personal nature |
| 60 | 1.0000000 | 61 | How surprising |
| 60 | 1.0000000 | 61 | Personal nature |
| 65 | 1.0000000 | 64 | How surprising |
| 65 | 1.0000000 | 64 | Personal nature |
| 70 | 1.0000000 | 68 | How surprising |
| 70 | 1.0000000 | 68 | Personal nature |
| 75 | 1.0000000 | 71 | How surprising |
| 75 | 1.0000000 | 71 | Personal nature |
| 80 | 1.0000000 | 75 | How surprising |
| 80 | 1.0000000 | 75 | Personal nature |
| 85 | 1.0000000 | 78 | How surprising |
| 85 | 1.0000000 | 78 | Personal nature |
| 90 | 1.0000000 | 82 | How surprising |
| 90 | 1.0000000 | 82 | Personal nature |
| 95 | 1.0000000 | 85 | How surprising |
| 95 | 1.0000000 | 85 | Personal nature |
| 100 | 1.0000000 | 89 | How surprising |
| 100 | 1.0000000 | 89 | Personal nature |

In this scenario, we could go with the point in which they both meet the 80% criterion, which is *n* = 47 for the personal variable and *n* = 47 for the surprising variable. In these scenarios, it is likely better to estimate a larger sample.

#### Maximum Sample Size

If you decide to use the 95% power as your criterion, you would see that items need somewhere between 54 and 54 participants for both variables. In this case, you could choose to make the larger value for participants your maximum sample size to ensure both variables reach criterion.

#### Final Sample Size

You should also consider any potential for missing data and/or unusable data given the requirements for your study. Given that participants are likely to see all items in this study, we could use the minimum, stopping rule, and maximum defined above.