> here::i\_am("BDD\_vs\_practice.qmd") #setting working directory

here() starts at /Users/elizabethhogg/Downloads

> data\_pathBDD <- here("BDD\_vs\_v2/BDD")

> data\_pathOCD <- here("BDD\_vs\_v2/OCD")

>

> # get a list of all BDD csv files

> csv\_filesBDD <- list.files(path = data\_pathBDD, pattern = "\*.csv")

>

> # get a list of all OCD csv files

> csv\_filesOCD <- list.files(path = data\_pathOCD, pattern = "\*.csv")

>

> # read in all BDD files together into one big dataframe

> all\_BDD <- csv\_filesBDD %>%

+ map\_df(~read\_csv(here(data\_pathBDD, .)))

>

> # remove end of file and NA responses

> all\_BDD <- all\_BDD %>% filter(`Event Index` != "END OF FILE", !is.na(`Event Index`))

> # remove end of file and NA responses

> all\_OCD <- all\_OCD %>% filter(`Event Index` != "END OF FILE", !is.na(`Event Index`))

>

>

> # create new variable based on BDD ranking all correct

> all\_BDD <- all\_BDD %>%

+ mutate(BDD\_correct = case\_when(

+ `BDD v Dep object-30 Quantised` == 1 & `BDD vs ED object-75 Quantised` == 1 & `GAD v BDD object-46 Quantised` == 2 & `OCD vs BDD object-67 Quantised` == 2 & `SAD vs BDD object-58 Quantised` == 2 ~ 1,

+ TRUE ~ 0))

>

> # merge two cols in OCD data set where you need to collapse them into eachother because NAs due to diff wording

> all\_OCD <- all\_OCD %>% mutate(ED\_v\_OCD\_71\_Quantised\_Merged = coalesce(`ED v OCD object-71 Quantised`, `ED v OCD object-71 Quantised`)) %>%

+ select(-c(`ED v OCD object-71 Quantised`, `ED v OCD object-71 Quantised`))

>

> # create new variable based on OCD ranking all correct

> all\_OCD <- all\_OCD %>%

+ mutate(OCD\_correct = case\_when(

+ `OCD v dep object-26 Quantised` == 1 & `OCD v SAD object-54 Quantised` == 1 & `OCD vs BDD object-67 Quantised` == 1 & ED\_v\_OCD\_71\_Quantised\_Merged == 2 & `GAD vs OCD object-42 Quantised` == 2 ~ 1,

+ TRUE ~ 0))

>

>

> # test <- read\_csv(here(data\_pathBDD, "BDD\_vs\_v2\_BDDmale\_2nd.csv"))

> # test2 <- read\_csv(here(data\_pathBDD, "BDD\_vs\_v2\_BDDmale\_1st.csv"))

> # test3 <- read\_csv(here(data\_pathBDD, "BDD\_vs\_v2\_BDDfemale\_2nd.csv"))

> # test4 <- read\_csv(here(data\_pathBDD, "BDD\_vs\_v2\_BDDfemale\_1st.csv"))

> #

> # # check which column names are different

> # all\_BDD %>%

> # select(which(!(colnames(all\_BDD) %in% colnames(test2))))

>

>

> # Now loading in v1 OCD and BDD data

>

> here::i\_am("BDD\_vs\_practice.qmd") #setting working directory

here() starts at /Users/elizabethhogg/Downloads

> data\_pathBDD <- here("BDD\_vs\_v1/BDD\_v1")

> data\_pathOCD <- here("BDD\_vs\_v1/OCD\_v1")

>

> # get a list of all BDD csv files

> csv\_filesBDD <- list.files(path = data\_pathBDD, pattern = "\*.csv")

>

> # get a list of all OCD csv files

> csv\_filesOCD <- list.files(path = data\_pathOCD, pattern = "\*.csv")

>

> # read in all v1 BDD files together into one big dataframe

> all\_BDD\_v1 <- csv\_filesBDD %>%

+ map\_df(~read\_csv(here(data\_pathBDD, .)))

>

> # remove end of file and NA responses

> all\_BDD\_v1 <- all\_BDD\_v1 %>% filter(`Event Index` != "END OF FILE", !is.na(`Event Index`))

> # remove end of file and NA responses

> all\_OCD\_v1 <- all\_OCD\_v1 %>% filter(`Event Index` != "END OF FILE", !is.na(`Event Index`))

>

> # create new variable based on BDD ranking all correct in BDD v1

> all\_BDD\_v1 <- all\_BDD\_v1 %>%

+ mutate(BDD\_correct = case\_when(

+ `BDD v Dep object-30 Quantised` == 1 & `BDD vs ED object-75 Quantised` == 1 & `GAD v BDD object-46 Quantised` == 2 & `OCD vs BDD object-67 Quantised` == 2 & `SAD vs BDD object-58 Quantised` == 2 ~ 1,

+ TRUE ~ 0))

>

> # merge two cols in OCD v1 data set where you need to collapse them into eachother because NAs due to diff wording

> all\_OCD\_v1 <- all\_OCD\_v1 %>% mutate(ED\_v\_OCD\_71\_Quantised\_Merged = coalesce(`ED v OCD object-71 Quantised`, `ED v OCD object-71 Quantised`)) %>%

+ select(-c(`ED v OCD object-71 Quantised`, `ED v OCD object-71 Quantised`))

>

> # create new variable based on OCD ranking all correct in OCD v1

> all\_OCD\_v1 <- all\_OCD\_v1 %>%

+ mutate(OCD\_correct = case\_when(

+ `OCD v dep object-26 Quantised` == 1 & `OCD v SAD object-54 Quantised` == 1 & `OCD vs BDD object-67 Quantised` == 1 & ED\_v\_OCD\_71\_Quantised\_Merged == 2 & `GAD vs OCD object-42 Quantised` == 2 ~ 1,

+ TRUE ~ 0))

>

> # merging BDD v1 and v2 data frames

> BDD\_v1\_and\_v2 <- rbind(all\_BDD, all\_BDD\_v1)

>

> # merging OCD v1 and v2 data frames

> OCD\_v1\_and\_v2 <- rbind(all\_OCD, all\_OCD\_v1)

>

> # merging BDD and OCD v1 and v2 data frames using participant public ID

> merged\_BDD\_OCD\_both\_versions <- merge(BDD\_v1\_and\_v2, OCD\_v1\_and\_v2, by = "Participant Public ID")

>

> # selecting random sample of 20% of participants to run chi-square for power analysis.

>

> set.seed(123) # For reproducibility so selects same 20% if rerun

>

> # Calculate 20% of 161 participants

> sample\_size <- floor(0.2 \* 161) # = 32

>

> # Randomly select 32 participants (rows)

> BDD\_OCD\_random\_20\_percent <- merged\_BDD\_OCD\_both\_versions[sample(nrow(merged\_BDD\_OCD\_both\_versions), sample\_size), ]

>

> # Create 2x2 contingency table in the 32 ps

> contingency\_table <- table(BDD\_OCD\_random\_20\_percent$BDD\_correct,

+ BDD\_OCD\_random\_20\_percent$OCD\_correct)

> rownames(contingency\_table) <- c("BDD\_Incorrect", "BDD\_Correct")

> colnames(contingency\_table) <- c("OCD\_Incorrect", "OCD\_Correct")

>

> # View the 2x2 table

> print(contingency\_table)

OCD\_Incorrect OCD\_Correct

BDD\_Incorrect 7 6

BDD\_Correct 7 12

>

> # Chi-square test using the 2x2 table

> chisq\_result <- chisq.test(contingency\_table)

> print(chisq\_result)

Pearson's Chi-squared test with Yates' continuity correction

data: contingency\_table

X-squared = 0.34754, df = 1, p-value = 0.5555

>

> # intall power package

>

> install.packages("pwr")

trying URL 'https://cran.rstudio.com/bin/macosx/big-sur-x86\_64/contrib/4.4/pwr\_1.3-0.tgz'

Content type 'application/x-gzip' length 160704 bytes (156 KB)

==================================================

downloaded 156 KB

The downloaded binary packages are in

/var/folders/sk/smlfwt9n0nqbl7mzmmky5s700000gn/T//RtmpLJyV5G/downloaded\_packages

> library(pwr)

>

> # calculated power with actual 32 sample size and medium effect size

> pwr.chisq.test(w = 0.3, df = 1, N = 32, sig.level = 0.05)

Chi squared power calculation

w = 0.3

N = 32

df = 1

sig.level = 0.05

power = 0.3964384

NOTE: N is the number of observations

>

> # our power was 0.3964384

>

> # calculated effect size with actual 32 sample size and medium effect size

> chisq\_val <- 0.3401

> N <- 32

>

> w <- sqrt(chisq\_val / N)

> w

[1] 0.1030928

>

> # our effect size was 0.1030928

>

> # based on this effect size, calculated sample size would need (using below it is 739)

> result <- pwr.chisq.test(w = w, df = 1, sig.level = 0.05, power = 0.8)

> total\_N\_needed <- ceiling(result$N)

> total\_N\_needed

[1] 739

>

> # given very high number of participants, I calculated sample size would need with 0.3 (medium effect size), 0.05 sig level and 0.8 power

> pwr.chisq.test(w = 0.3, df = 1, sig.level = 0.05, power = 0.8)

Chi squared power calculation

w = 0.3

N = 87.20954

df = 1

sig.level = 0.05

power = 0.8

NOTE: N is the number of observations

>

> # would need 87 participants based on above

>

> # Calculated sample size would need with 0.2 (small-medium effect size), 0.05 sig level and 0.8 power

> pwr.chisq.test(w = 0.2, df = 1, sig.level = 0.05, power = 0.8)

Chi squared power calculation

w = 0.2

N = 196.2215

df = 1

sig.level = 0.05

power = 0.8

NOTE: N is the number of observations

>

> # would need 196 participants based on above

>

> # Calculated sample size would need with 0.15 (small effect size), 0.05 sig level and 0.8 power

> pwr.chisq.test(w = 0.15, df = 1, sig.level = 0.05, power = 0.8)

Chi squared power calculation

w = 0.15

N = 348.8382

df = 1

sig.level = 0.05

power = 0.8

NOTE: N is the number of observations

>

> # would need 349 participants based on above