PSYR6003 Assignment 1 Write Up

Nicole Basso

## Load Libraries Needed For Assignment

library(tidyverse)

── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
✔ dplyr 1.1.4 ✔ readr 2.1.5  
✔ forcats 1.0.0 ✔ stringr 1.5.1  
✔ ggplot2 3.4.4 ✔ tibble 3.2.1  
✔ lubridate 1.9.3 ✔ tidyr 1.3.0  
✔ purrr 1.0.2   
── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
✖ dplyr::filter() masks stats::filter()  
✖ dplyr::lag() masks stats::lag()  
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(ggplot2)  
library(haven)  
library(pwr)  
library(TOSTER)  
library(effectsize)  
library(skimr)

## Question 1

## Load “Avengers” CSV file dataset, assigning it to object “avengers”

## then previewing the dataset using “glimpse” for inspection of variable types and names

avengers <- read.csv("avengers.csv") %>%   
 glimpse()

Rows: 814  
Columns: 15  
$ iq <int> 118, 109, 90, 114, 116, 137, 114, 102, 98, 99, 112, …  
$ agility <int> 27, 54, 56, 63, 30, 62, 57, 63, 70, 52, 47, -11, 33,…  
$ speed <dbl> 5.05, 4.98, 4.84, 5.24, 5.01, 4.79, 4.96, 5.39, 5.14…  
$ strength <int> 400, 782, 569, 535, 663, 435, 418, 630, 480, 663, 49…  
$ damage.resistance <dbl> 2.10, 2.09, 1.99, 2.19, 2.33, 2.05, 2.00, 2.05, 1.71…  
$ superpower <chr> "no", "no", "no", "no", "no", "no", "no", "no", "no"…  
$ flexibility <dbl> -1.9, -2.2, -1.0, -0.3, -2.6, -0.9, -1.5, -0.6, -0.2…  
$ willpower <int> 50, 38, 67, 58, 77, 97, 57, 75, 92, 74, 102, 52, 52,…  
$ ptsd <dbl> 4.9, 3.8, 4.2, 3.5, 4.2, 2.6, 4.2, 3.1, 3.9, 4.3, 3.…  
$ north\_south <chr> "south", "north", "south", "north", "south", "north"…  
$ died <chr> "no", "no", "no", "no", "no", "no", "yes", "no", "no…  
$ kills <int> 0, 22, 1, 10, 1, 1, 1, 2, 5, 1, 2, 0, 0, 0, 2, 9, 1,…  
$ injuries <int> 4, 4, 2, 3, 3, 5, 5, 2, 5, 2, 3, 5, 3, 2, 3, 4, 2, 3…  
$ minutes.fighting <dbl> 7.7, 14.3, 13.9, 10.4, 11.9, 9.5, 9.8, 15.4, 11.9, 1…  
$ shots.taken <int> 134, 125, 147, 104, 94, 111, 148, 119, 139, 130, 121…

## Glimpse showed there was no Participant ID column included in the original dataset; adding an ID column “PID” at the beginning of the dataset to aid in Data Inspection & Cleaning

avengersID <- avengers %>%   
 mutate(PID = row\_number(), .before = 1)

##subsetting cases with missing values and opening the dataset to inspect cases & PIDs

## Looking at summary statistics to inspect for unusual values or cases in the data

missing <- avengersID %>%   
 filter(!complete.cases(avengersID))  
missing

PID iq agility speed strength damage.resistance superpower flexibility  
1 784 132 77 NA NA NA "yes" NA  
2 814 109 100000 1e+05 100000 NA NA  
 willpower ptsd north\_south died kills injuries minutes.fighting shots.taken  
1 NA NA NA NA NA NA  
2 NA NA NA NA NA NA

skim\_tee(avengersID)

── Data Summary ────────────────────────  
 Values  
Name data   
Number of rows 814   
Number of columns 16   
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
Column type frequency:   
 character 3   
 numeric 13   
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
Group variables None   
  
── Variable type: character ────────────────────────────────────────────────────  
 skim\_variable n\_missing complete\_rate min max empty n\_unique whitespace  
1 superpower 0 1 0 5 1 4 0  
2 north\_south 0 1 0 5 2 3 0  
3 died 0 1 0 3 2 3 0  
  
── Variable type: numeric ──────────────────────────────────────────────────────  
 skim\_variable n\_missing complete\_rate mean sd p0 p25  
 1 PID 0 1 408. 235. 1 204.   
 2 iq 0 1 110. 8.04 85 105   
 3 agility 0 1 173. 3503. -11 41   
 4 speed 1 0.999 128. 3507. 4.25 4.89  
 5 strength 1 0.999 622. 3493. 22 402   
 6 damage.resistance 2 0.998 2.00 0.210 1.35 1.85  
 7 flexibility 2 0.998 -1.00 1.00 -3.9 -1.7   
 8 willpower 2 0.998 60.0 18.0 -7 48   
 9 ptsd 2 0.998 4.00 0.600 1.8 3.6   
10 kills 2 0.998 3.83 10.8 0 0   
11 injuries 2 0.998 3.49 1.26 0 2   
12 minutes.fighting 2 0.998 12.4 8.45 1.1 8.7   
13 shots.taken 2 0.998 130. 30.0 58 111   
 p50 p75 p100 hist   
 1 408. 611. 814 ▇▇▇▇▇  
 2 110 115 137 ▁▃▇▃▁  
 3 51 61 100000 ▇▁▁▁▁  
 4 5.01 5.12 100000 ▇▁▁▁▁  
 5 505 600 100000 ▇▁▁▁▁  
 6 1.99 2.14 2.59 ▁▅▇▅▁  
 7 -1 -0.3 2.1 ▁▆▇▅▁  
 8 59.5 72 115 ▁▂▇▅▁  
 9 4 4.4 7.4 ▁▇▇▁▁  
10 1 3 176 ▇▁▁▁▁  
11 4 5 5 ▁▆▆▆▇  
12 10.4 13.1 93.7 ▇▁▁▁▁  
13 126 146 267 ▂▇▃▁▁

##Subsetting data to only include cleaned cases ##Participant 784 and 814 excluded for incomplete data

clean.avengers <- filter(avengersID, PID != 784 & PID != 814)

##Question 2 ##Creating a custom variable called “CombatEffectiveness”, representing the sum of agility ##using head to check that the variable computed correctly

appended.avengers <- clean.avengers %>%   
 mutate(CombatEffectiveness = rowSums(select(clean.avengers, c(speed, agility, strength, willpower))))  
  
head(appended.avengers)

PID iq agility speed strength damage.resistance superpower flexibility  
1 1 118 27 5.05 400 2.10 no -1.9  
2 2 109 54 4.98 782 2.09 no -2.2  
3 3 90 56 4.84 569 1.99 no -1.0  
4 4 114 63 5.24 535 2.19 no -0.3  
5 5 116 30 5.01 663 2.33 no -2.6  
6 6 137 62 4.79 435 2.05 no -0.9  
 willpower ptsd north\_south died kills injuries minutes.fighting shots.taken  
1 50 4.9 south no 0 4 7.7 134  
2 38 3.8 north no 22 4 14.3 125  
3 67 4.2 south no 1 2 13.9 147  
4 58 3.5 north no 10 3 10.4 104  
5 77 4.2 south no 1 3 11.9 94  
6 97 2.6 north no 1 5 9.5 111  
 CombatEffectiveness  
1 482.05  
2 878.98  
3 696.84  
4 661.24  
5 775.01  
6 598.79

##Question 3 ## Creating a subsetted copy of the dataset in both .sav and .csv formats that only includes avengers that did not have a superpower and have died ## Then summarizing the newavengers dataset (M, SD, range) for the values combat effectivenesss, kills, and injuries for the overall sample and by battle location

NewAvengers <- appended.avengers %>%   
 filter(superpower == "no" & died == "yes")  
 head(NewAvengers)

PID iq agility speed strength damage.resistance superpower flexibility  
1 7 114 57 4.96 418 2.00 no -1.5  
2 12 109 -11 4.25 22 1.62 no -1.3  
3 19 116 36 4.75 282 1.81 no -1.3  
4 29 104 47 4.92 352 1.67 no -1.9  
5 38 109 48 5.19 554 1.94 no -3.1  
6 46 113 25 4.65 236 1.85 no -1.4  
 willpower ptsd north\_south died kills injuries minutes.fighting shots.taken  
1 57 4.2 north yes 1 5 9.8 148  
2 52 4.0 south yes 0 5 4.5 263  
3 33 4.4 north yes 0 5 5.0 164  
4 53 3.6 north yes 0 5 7.6 151  
5 82 3.2 south yes 1 5 10.4 113  
6 93 3.3 north yes 0 5 7.0 173  
 CombatEffectiveness  
1 536.96  
2 67.25  
3 355.75  
4 456.92  
5 689.19  
6 358.65

write.csv(NewAvengers, "newavengers.csv", row.names = F,na="")   
write\_sav(NewAvengers, "newavengers.sav")   
  
  
allsummary <- NewAvengers %>%   
 summarize(north\_south = "Total Sample",  
 Mean.CombatEffectiveness = mean(CombatEffectiveness),  
 SD.CombatEffectiveness = sd(CombatEffectiveness),  
 Min.CombatEffectiveness = (min(CombatEffectiveness)),   
 Max.CombatEffectiveness = (max(CombatEffectiveness)),  
 Mean.kills = mean(kills),  
 SD.kills = sd(kills),  
 Min.kills = (min(kills)),   
 Max.kills = (max(kills)),  
 Mean.injuries = mean(injuries),  
 SD.injuries = sd(injuries),  
 Min.injuries = (min(injuries)),   
 Max.injuries = (max(injuries))   
 )  
groupedsummary <- NewAvengers %>%   
 group\_by(north\_south) %>%   
 summarize(Mean.CombatEffectiveness = mean(CombatEffectiveness),  
 SD.CombatEffectiveness = sd(CombatEffectiveness),  
 Min.CombatEffectiveness = (min(CombatEffectiveness)),   
 Max.CombatEffectiveness = (max(CombatEffectiveness)),  
 Mean.kills = mean(kills),  
 SD.kills = sd(kills),  
 Min.kills = (min(kills)),   
 Max.kills = (max(kills)),  
 Mean.injuries = mean(injuries),  
 SD.injuries = sd(injuries),  
 Min.injuries = (min(injuries)),   
 Max.injuries = (max(injuries)))  
SummaryStats <- bind\_rows(allsummary, groupedsummary)  
SummaryStats <- mutate(SummaryStats, Group = north\_south, .keep = "unused", .before = 1)  
  
SummaryStats

Group Mean.CombatEffectiveness SD.CombatEffectiveness  
1 Total Sample 497.5348 177.5593  
2 north 499.7815 174.0656  
3 south 491.6771 189.5299  
 Min.CombatEffectiveness Max.CombatEffectiveness Mean.kills SD.kills  
1 67.25 946.89 2.554455 8.807355  
2 130.68 897.06 1.712329 4.565813  
3 67.25 946.89 4.750000 14.992900  
 Min.kills Max.kills Mean.injuries SD.injuries Min.injuries Max.injuries  
1 0 79 4.554455 0.7412860 2 5  
2 0 34 4.602740 0.6819035 2 5  
3 0 79 4.428571 0.8789123 2 5

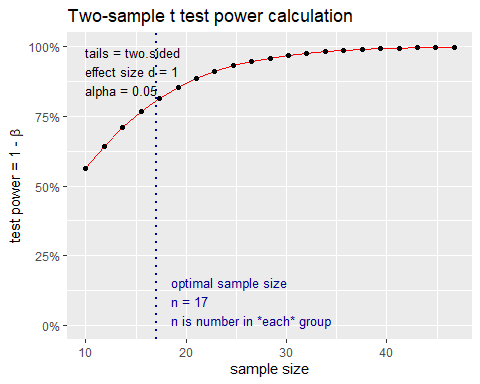
##Question 8 ## Power analysis for secondary analysis regarding superpowers and IQ ## Powering for a cohen’s d of 1, 80% power, and an alpha of 0.05 ##Two sided, two sample test

##Plotting a power curve to see if 80% power is optimal

Power8.ttest <- pwr.t.test(d = 1, power = .8, sig.level = 0.05, type = "two.sample", alternative = "two.sided")  
Power8.ttest

Two-sample t test power calculation   
  
 n = 16.71472  
 d = 1  
 sig.level = 0.05  
 power = 0.8  
 alternative = two.sided  
  
NOTE: n is number in \*each\* group

plot(Power8.ttest)



##Question 9 #Estimating power required for an equivalence test in our secondary analyses following the same parameters as the t test

Power9.EQ <- powerTOSTtwo(alpha=0.05, statistical\_power=.8, low\_eqbound\_d=- -1, high\_eqbound\_d= 1)

Warning: `powerTOSTtwo()` was deprecated in TOSTER 0.4.0.  
ℹ Please use `power\_t\_TOST()` instead.

The required sample size to achieve 80 % power with equivalence bounds of 1 and 1 is 18 per group, or 36 in total.

Power9.EQ

[1] 17.12769

##Question 10 #Calculating the Cohen’s d effect size using t-test results (test statistic 4.25; df = 810)

EffectSize.TTest <- t\_to\_d(4.25, 810, paired = F, ci = 0.95, alternative ="two.sided")  
EffectSize.TTest

d | 95% CI  
-------------------  
0.30 | [0.16, 0.44]