```{r setup, include=FALSE}

knitr::opts\_chunk$set(echo = FALSE,

eval = TRUE,

warning = FALSE,

message = FALSE,

fig.width = 5,

fig.height = 4.25)

library(tidyverse)

library(afex)

library(rio)

library(here)

library(knitr)

library(kableExtra)

library(misty)

library(janitor)

library(rstatix)

library(dplyr)

library(psych)

library(MOTE)

theme\_set(theme\_minimal())

```

# Load in the data

```{r}

act\_t1 <- import(here("data", "ACT Suicide Prevention T1.sav"), setclass = "tbl\_df") %>%

characterize() %>%

clean\_names()

act\_t1\_12520 <- import(here("data", "ACT Suicide Prevention T1 1-25-20.sav"), setclass = "tbl\_df") %>%

characterize() %>%

clean\_names()

```

# Joining datasets

```{r}

act\_all <- full\_join(act\_t1, act\_t1\_12520)

```

# Creating df with our variables of interest

```{r}

df <- act\_all %>%

select(id, dem1\_17a, dem1\_02a, bdi1)

df\_1 <- df %>%

drop\_na(dem1\_17a, dem1\_02a, bdi1)

df\_2 <- df\_1[-c(220, 307, 651),]

df\_3 <- df\_2[-c(14, 220),]

```

# Checking variable types

```{r}

str(df)

```

# Frequency table THIS IS NOT MET....NO BALANCED GROUPS

```{r}

df %>%

group\_by(dem1\_02a, dem1\_17a) %>%

summarise(n = n()) %>% #new column "n" = row count of each factor grouping

spread(dem1\_02a, n) # rows "spread" across the top

```

# Checking the assumptions with descriptives

Descriptives:

The average BDI score is 7.68 (7.01), and IS NOT normally distributed (skew=1.77, kurtosis=4.39).

```{r}

describe(df)

```

# Visually inspecting the data

Step 1: Extract means to examine group differences

```{r}

means <- df %>% #creating new object (tibble) with the means

group\_by(dem1\_02a, dem1\_17a) %>% #grouping our results by our two independent variables, 'country' within levels of 'sen2'

summarise(mean\_bdi = mean(bdi1))

means

```

Step 2: Examine the means using a profile plot

```{r}

means %>%

ggplot(aes(dem1\_02a, mean\_bdi, group = dem1\_17a)) +

geom\_point() +

geom\_line(aes(color = dem1\_17a))

```

Visualize using boxplots

```{r}

df %>%

ggplot(aes(x = dem1\_17a, y = bdi1, fill = dem1\_02a)) +

geom\_boxplot() +

theme\_minimal() +

theme(legend.position="bottom")

```

### Check the assumptions

There are three assumptions that are required to be met in order to successfully conduct an ANOVA. These include: Homogeneity of variance, Independence of observations, and normality. All assumptions are met and are further assessed and described below.

### Assessing for homogeneity of variance with a Levene's Test.

```{r}

car::leveneTest(bdi1 ~ dem1\_02a\*dem1\_17a, data = df, center = "mean")

```

### Running the ANOVA

```{r}

anova\_test(df, formula = bdi1 ~ dem1\_02a + dem1\_17a + dem1\_02a:dem1\_17a,

detailed = TRUE, type = 3, effect.size = "pes")

## because we have unbalanced groups anova will not run

### THIS WORKS

anova\_test(df\_3, formula = bdi1 ~ dem1\_02a + dem1\_17a + dem1\_02a:dem1\_17a,

detailed = TRUE, type = 3, effect.size = "pes")

```

2 x 4 factorial ANOVA:

one possible combination:

dating in a dorm

dating in an apartment/house

single in a dorm

single in an apartment/house

other possible combination:

living in dorm

living not in a dorm

living in a dorm with roommates

other possible variables:

(maybe age or year, in state vs out of state)

male

female

BDI

table() function to look at grouping

with(df, table(var1, var2))

table(df$var1, df$var2)

BELOW IS WHERE I STARTED OVER TRYING TO CREATE NEW VARIABLE

### Creating new variables

```{r}

df1 <- act\_all %>%

select(id, dem1\_02a, dem1\_17a, bdi1, dem1\_07a)

df1 <- df1 %>%

rename(livingstat = dem1\_02a,

datingstat = dem1\_17a,

BDI = bdi1,

gender = dem1\_07a)

```

#dropping NA values in rows below

```{r}

df1$living\_dating <- paste(df1$livingstat, df1$datingstat)

df2 <- df1 %>%

drop\_na(BDI, living\_dating, gender)

```

#looking at group blances...this is where there is still an issue with the "I choose not to answer" option. Because of this the ANOVA will not run

```{r}

df3 <- df2[-c(221, 307, 651),]

df3 %>%

drop\_na(df3)

spread <- df3 %>%

group\_by(id, living\_dating, gender, BDI) %>%

summarise(n = n()) %>% #new column "n" = row count of each factor grouping

spread(living\_dating, n)

df4 <- spread %>%

select(1:6, 8)

df5 <- df4 %>%

pivot\_longer(cols = 4:7,

names\_to = "liv\_date",

values\_to = "score") %>%

drop\_na(score)

df5 %>%

drop\_na(df5)

str(df5)

df5 %>%

group\_by(id, liv\_date, gender, BDI) %>%

summarise(n = n()) %>% #new column "n" = row count of each factor grouping

spread(liv\_date, n)

```

```{r}

anova\_test(data = df5, formula = BDI ~ liv\_date + gender + liv\_date:gender,

detailed = TRUE, type = 3, effect.size = "pes")

head(df5)

my\_anova <- aov(BDI ~ liv\_date \* gender, data = df5)

Anova(my\_anova, type = "III")

```