Final project

Mandi Ward & Marielena McWhirter

5/5/2022

2 x 3 factorial ANOVA: IV: Dating, married, or single IV: living in a house or living in a dorm

DV: BDI scores

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

THIS CODE WORKS # Load in the data

```
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

```
act_all <- full_join(act_t1, act_t1_12520)
```

Creating df with our variables of interest

```
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(220),]

df_4 <- df_3 %>%
  rename(livingstat = dem1_02a,
         datingstat = dem1_17a,
         BDI = bdi1)
```

Describing Data

```
describe(df)
```

```
##           vars   n     mean       sd  median  trimmed      mad   min   max
## id           1 972 28980.03 14627.19 24326.5 28492.83 14941.64 10018 51798
## dem1_17a*    2 972     2.78     1.47     4.0     2.86     0.00     1     4
## dem1_02a*    3 972     2.42     0.50     2.0     2.40     0.00     1     3
## bdi1         4 967     7.68     7.01     6.0     6.58     5.93     0    53
##           range  skew kurtosis     se
## id        41780  0.36    -1.18 469.17
## dem1_17a*     3 -0.39    -1.85   0.05
## dem1_02a*     2  0.29    -1.78   0.02
## bdi1         53  1.77     4.39   0.23
```

Frequency table

```
df_4 %>%
  group_by(livingstat, datingstat) %>%
  summarise(n = n()) %>% #new column "n" = row count of each factor grouping
  spread(livingstat, n) # rows "spread" across the top
```

```
## # A tibble: 3 × 3
##   datingstat `In a dorm` `In an apartment or house`
##   <chr>            <int>                      <int>
## 1 Dating             208                        183
## 2 Married             NA                          1
## 3 Single             352                        219
```

Checking the assumptions with descriptives

Descriptives: The average BDI score is 7.68 (6.99), and IS NOT normally distributed (skew=1.77, kurtosis=4.44).

```
describe(df_4)
```

```
##             vars   n     mean       sd median  trimmed      mad   min   max
## id             1 963 29036.15 14667.85  24371 28563.60 15076.56 10018 51798
## datingstat*    2 963     2.19     0.98      3     2.23     0.00     1     3
## livingstat*    3 963     1.42     0.49      1     1.40     0.00     1     2
## BDI            4 963     7.68     6.99      6     6.58     5.93     0    53
##             range  skew kurtosis     se
## id          41780  0.36    -1.19 472.66
## datingstat*     2 -0.38    -1.86   0.03
## livingstat*     1  0.33    -1.89   0.02
## BDI            53  1.77     4.44   0.23
```

```
describeBy(df_4, group = "livingstat")
```

```
## 
##  Descriptive statistics by group 
## livingstat: In a dorm
##             vars   n     mean       sd  median  trimmed      mad   min   max
## id             1 560 29018.76 14551.55 30199.5 28542.46 14951.28 10018 51762
## datingstat*    2 560     1.63     0.48     2.0     1.66     0.00     1     2
## livingstat*    3 560     1.00     0.00     1.0     1.00     0.00     1     1
## BDI            4 560     7.13     6.59     6.0     6.10     4.45     0    53
##             range  skew kurtosis     se
## id          41744  0.32    -1.17 614.92
## datingstat*     1 -0.53    -1.72   0.02
## livingstat*     0   NaN      NaN   0.00
## BDI            53  2.10     7.06   0.28
## ------------------------------------------------------------ 
## livingstat: In an apartment or house
##             vars   n     mean       sd median  trimmed      mad   min   max
## id             1 403 29060.31 14846.06  23164 28591.09 16361.97 10030 51798
## datingstat*    2 403     2.09     1.00      3     2.11     0.00     1     3
## livingstat*    3 403     1.00     0.00      1     1.00     0.00     1     1
## BDI            4 403     8.44     7.46      6     7.37     5.93     0    43
##             range  skew kurtosis     se
## id          41768  0.39    -1.24 739.53
## datingstat*     2 -0.18    -1.97   0.05
## livingstat*     0   NaN      NaN   0.00
## BDI            43  1.41     2.13   0.37
```

Visually inspecting the data

Step 1: Extract means to examine group differences

```
means <- df_4 %>% #creating new object (tibble) with the means
  group_by(livingstat, datingstat) %>% #grouping our results by our two independent variables, 'country' within l
evels of 'sen2'
  summarise(mean_bdi = mean(BDI))
means
```

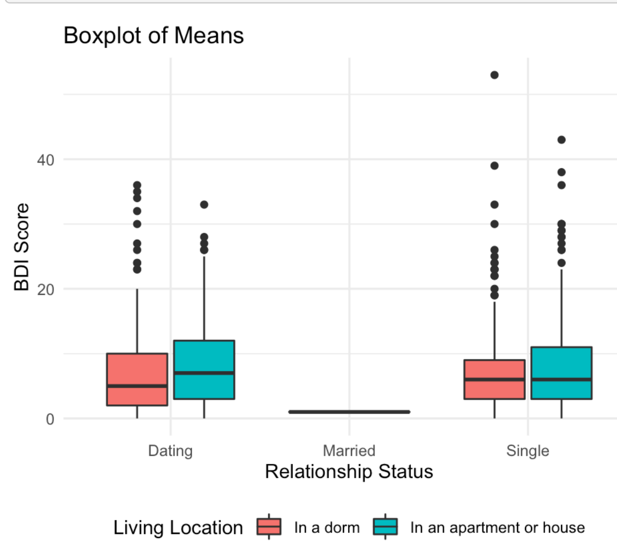
```
## # A tibble: 5 × 3
## # Groups:   livingstat [2]
##   livingstat                datingstat mean_bdi
##   <chr>                     <chr>         <dbl>
## 1 In a dorm                 Dating         7.25
## 2 In a dorm                 Single         7.05
## 3 In an apartment or house Dating         8.70
## 4 In an apartment or house Married        1   
## 5 In an apartment or house Single         8.25
```

Step 2: Examine the means using a profile plot

```
means %>%
  ggplot(aes(livingstat, mean_bdi, group = datingstat)) +
  geom_point() +
  geom_line(aes(color = datingstat)) +
  labs(title = "Profile Plot of Means", x = "Living Location", y = "BDI Score", fill = "Relationship Status")
```

Visualize using boxplots

```
df_4 %>%
  ggplot(aes(x = datingstat, y = BDI, fill = livingstat)) +
  geom_boxplot() +
  theme_minimal() +
  theme(legend.position="bottom") +
  labs(title = "Boxplot of Means", x = "Relationship Status", y = "BDI Score", fill = "Living Location")
```



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

Levene's Test on the mean of the variables is significant. This indicates that the variables do not meet the assumption of homogeneity in variance. However, experiemental study designs have more flexability with Levene's Test, and homogeneity in variance is not required to be considered for ANOVA analyses.

```
car::leveneTest(BDI ~ livingstat*datingstat, data = df_4, center = "mean")
```

```
## Levene's Test for Homogeneity of Variance (center = "mean")
##        Df F value   Pr(>F)   
## group   4  4.0102 0.003122 **
##       958                    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ANOVA

```
### THIS DOES NOT WORK
#anova_test(df_3, formula = bdi1 ~ dem1_02a + dem1_17a + dem1_02a:dem1_17a,
                #detailed = TRUE, type = 3, effect.size = "pes")


## THIS ONE WORKS
res.aov2 <- aov(BDI ~ livingstat + datingstat + livingstat:datingstat, data = df_4)
summary(res.aov2)
```

```
##                        Df Sum Sq Mean Sq F value  Pr(>F)   
## livingstat              1    404   403.7   8.305 0.00404 **
## datingstat              2     78    38.9   0.801 0.44904   
## livingstat:datingstat   1      4     3.5   0.073 0.78736   
## Residuals             958  46562    48.6                   
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Effect size for interaction

```
 omega.F(dfm = 1, dfe = 4,
      Fvalue = 0.073, n = 963, a = .05)
```

```
## $omega
## [1] -0.0009635443
## 
## $omegalow
## [1] 0
## 
## $omegahigh
## [1] 0.9999952
## 
## $dfm
## [1] 1
## 
## $dfe
## [1] 4
## 
## $F
## [1] 0.073
## 
## $p
## [1] 0.800385
## 
## $estimate
## [1] "$\\omega^2$ = 0.00, 95\\% CI [0.00, 1.00]"
## 
## $statistic
## [1] "$F$(1, 4) = 0.07, $p$ = .800"
```

Estimate marginal means

```
model <- lm(BDI ~ livingstat + datingstat, df_4)

means_1 <- emmeans(model, specs = "livingstat", by = "datingstat")

pairs(means_1) # display the means produced by emmeans() in pairs
```

```
## datingstat = Dating:
##  contrast                              estimate    SE  df t.ratio p.value
##  In a dorm - In an apartment or house      -1.3 0.457 959  -2.854  0.0044
## 
## datingstat = Married:
##  contrast                              estimate    SE  df t.ratio p.value
##  In a dorm - In an apartment or house      -1.3 0.457 959  -2.854  0.0044
## 
## datingstat = Single:
##  contrast                              estimate    SE  df t.ratio p.value
##  In a dorm - In an apartment or house      -1.3 0.457 959  -2.854  0.0044
```

```
means_2 <- emmeans(model, ~livingstat*datingstat) #use same model but use different specifications

pairs(means_2)
```

```
##  contrast                                                         estimate
##  In a dorm Dating - In an apartment or house Dating                 -1.305
##  In a dorm Dating - In a dorm Married                                7.628
##  In a dorm Dating - In an apartment or house Married                 6.323
##  In a dorm Dating - In a dorm Single                                 0.312
##  In a dorm Dating - In an apartment or house Single                 -0.993
##  In an apartment or house Dating - In a dorm Married                 8.932
##  In an apartment or house Dating - In an apartment or house Married  7.628
##  In an apartment or house Dating - In a dorm Single                  1.617
##  In an apartment or house Dating - In an apartment or house Single   0.312
##  In a dorm Married - In an apartment or house Married               -1.305
##  In a dorm Married - In a dorm Single                               -7.316
##  In a dorm Married - In an apartment or house Single                -8.621
##  In an apartment or house Married - In a dorm Single                -6.011
##  In an apartment or house Married - In an apartment or house Single -7.316
##  In a dorm Single - In an apartment or house Single                 -1.305
##     SE  df t.ratio p.value
##  0.457 959  -2.854  0.0502
##  6.981 959   1.093  0.8844
##  6.980 959   0.906  0.9451
##  0.459 959   0.679  0.9842
##  0.675 959  -1.472  0.6823
##  7.012 959   1.274  0.7993
##  6.981 959   1.093  0.8844
##  0.620 959   2.608  0.0963
##  0.459 959   0.679  0.9842
##  0.457 959  -2.854  0.0502
##  6.980 959  -1.048  0.9014
##  7.013 959  -1.229  0.8226
##  6.977 959  -0.862  0.9554
##  6.980 959  -1.048  0.9014
##  0.457 959  -2.854  0.0502
## 
## P value adjustment: tukey method for comparing a family of 6 estimates
```

Effect size for main effect

```
omega.F(dfm = 1, dfe = 404,
      Fvalue = 8.305, n = 963, a = .05)
```

```
## $omega
## [1] 0.007528561
## 
## $omegalow
## [1] 0
## 
## $omegahigh
## [1] 0.03324983
## 
## $dfm
## [1] 1
## 
## $dfe
## [1] 404
## 
## $F
## [1] 8.305
## 
## $p
## [1] 0.004164549
## 
## $estimate
## [1] "$\\omega^2$ = 0.01, 95\\% CI [0.00, 0.03]"
## 
## $statistic
## [1] "$F$(1, 404) = 8.30, $p$ = .004"
```

ADD IN QQ PLOT VISUAL IN ADDITION TO SKEW AND KURTOSIS in discussion discuss implications and not meeting criteria and just that it may impact the validity of the results.

```
df_4 %>%
  ggplot(aes(sample = BDI)) +
    stat_qq()+
    stat_qq_line(color = "violet") +
    theme_classic() +
    labs(title = "QQ Plot of Mean Depression Scores")
```