Analysis of Variance

Effect sizes

Jenine Harris
Brown School

Importing and cleaning the data

```
# load GSS rda file
load(file = "/Users/harrisj/Box/teaching/Teaching/Fall2020/data/gss2018.

# assign GSS to gss.2018
gss.2018 <- GSS
# remove GSS
rm(GSS)

# recode variables of interest to valid ranges
gss.2018.cleaned <- gss.2018 %>%
  select(HAPPY, SEX, DEGREE, USETECH, AGE) %>%
  mutate(USETECH = na_if(x = USETECH, y = -1)) %>%
  mutate(USETECH = na_if(x = USETECH, y = 999)) %>%
  mutate(USETECH = na_if(x = USETECH, y = 998)) %>%
  mutate(AGE = na_if(x = AGE, y = 98)) %>%
  mutate(AGE = na_if(x = AGE, y = 99)) %>%
  mutate(DEGREE = na_if(x = DEGREE, y = 8)) %>%
  mutate(DEGREE = na_if(x = DEGREE, y = 9)) %>%
  mutate(HAPPY = na_if(x = HAPPY, y = 8)) %>%
  mutate(HAPPY = na_if(x = HAPPY, y = 9)) %>%
  mutate(HAPPY = na_if(x = HAPPY, y = 0)) %>%
  mutate(SEX = factor(x = SEX, labels = c("male","female"))) %>%
  mutate(DEGREE = factor(x = DEGREE, labels = c("< high school",
                                                "high school", "junior c
                                                "college", "grad school"

  mutate(HAPPY = factor(x = HAPPY, labels = c("very happy",
                                              "pretty happy",
```

Visualizing the groups

```
# graph usetech
gss.2018.cleaned %>%
  drop_na(USETECH) %>%
  ggplot(aes(y = USETECH, x = DEGREE)) +
  geom_jitter(aes(color = DEGREE), alpha = .6) +
  geom_boxplot(aes(fill = DEGREE), alpha = .4) +
  scale_fill_brewer(palette = "Spectral", guide = FALSE) +
  scale_color_brewer(palette = "Spectral", guide = FALSE) +
  theme_minimal() +
  labs(x = "Highest educational attainment",
       y = "Percent of time spent using technology",
       title = "Distribution of time spent using technology\nuse by educ
```

Group means

```
# mean and sd of age by group
use.stats <- gss.2018.cleaned %>%
  drop_na(USETECH) %>%
  group_by(DEGREE) %>%
  summarize(m.techuse = mean(USETECH),
            sd.techuse = sd(USETECH))
use.stats
```

```
## # A tibble: 5 x 3
##   DEGREE         m.techuse sd.techuse
##   <fct>              <dbl>      <dbl>
## 1 < high school       24.8       36.2
## 2 high school         49.6       38.6
## 3 junior college      62.4       35.2
## 4 college             67.9       32.1
## 5 grad school         68.7       30.2
```

ANOVA results

```
# conduct ANOVA for technology use by degree category with oneway.test
techuse.by.deg <- oneway.test(formula = USETECH ~ DEGREE,
                              data = gss.2018.cleaned,
                              var.equal = TRUE)
techuse.by.deg
```

```
##
##      One-way analysis of means
##
## data:  USETECH and DEGREE
## F = 43.304, num df = 4, denom df = 1404, p-value < 2.2e-16
```

```
# conduct ANOVA for technology use by degree category with aov
techuse.by.deg.aov <- aov(formula = USETECH ~ DEGREE,
            data = gss.2018.cleaned)
summary(object = techuse.by.deg.aov)
```

```
##               Df  Sum Sq Mean Sq F value Pr(>F)
## DEGREE         4  221301   55325    43.3 <2e-16 ***
## Residuals   1404 1793757    1278
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 936 observations deleted due to missingness
```

Computing and interpreting effect sizes for ANOVA

- Chi-squared has Cramer's V and t-tests have Cohen's d effect sizes.
- For ANOVA, eta-squared, η^2 is the proportion of variability in the continuous outcome variable that is explained by the groups and is the commonly used effect size for ANOVA.
    - However, eta-squared has a known positive bias but was still used widely because it was the effect size that was easily available in some statistical software programs.
- Another statistic, omega-squared (ω^2) has the same general meaning, is adjusted to account for the positive bias, and is more stable when assumptions were not completely met.
- In the omega-squared equation, n is the number of observations and k is the number of groups; the F is from the ANOVA results.

$$\omega^2 = \frac{F - 1}{F + \frac{n-k+1}{k-1}}$$

Computing omega-squared

- The functions used so far, `oneway.test()` and `aov()`, do not compute omega-squared as part of the output.
- However, there are R packages that do compute it.
- Even so, using the output from `aov()` to compute omega-squared is recommended.

```r
# ANOVA model from earlier
summary(object = techuse.by.deg.aov)
```

```
##               Df  Sum Sq Mean Sq F value Pr(>F)    
## DEGREE         4  221301   55325    43.3 <2e-16 ***
## Residuals   1404 1793757    1278                   
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 936 observations deleted due to missingness
```

- Everything needed to compute omega-squared is in the output:

$$\omega^2 = \frac{F-1}{F+\frac{n-k+1}{k-1}} = \frac{43.3-1}{43.3+\frac{1409-5+1}{5-1}} = .107$$

Cutoffs for small, medium, large effects

- Cutoffs for omega-squared effect size:
    - $\omega^2 = .01$ to $\omega^2 < .06$ is a small effect
    - $\omega^2 = .06$ to $\omega^2 < .14$ is a medium effect
    - $\omega^2 \geq .14$ is a large effect

Interpreting effect size

- The mean time spent on technology use was significantly different across educational attainment groups [$F(4,1404) = 43.3$; $p < .05$] indicating these groups likely came from populations with different mean time spent on technology use. The highest mean was 68.7% of time used for technology for those with graduate degrees. The lowest mean was 24.8% of the time for those with less than a high school diploma. A set of planned comparisons found that the mean time spent using technology was statistically significantly ($p < .05$) lower for (1) those with < *high school* education (m = 24.8) compared to those with *high school or junior college* (m = 51.7), (2) those with a high school education (m = 49.61) compared to those with any college (m = 67.0), (3) those with a junior college degree (m = 62.4) compared to those with more college than that (m = 68.2), and (4) those with a bachelor's (m = 67.9) compared to those with a master's degree (m = 68.7). Overall the patterns show statistically significant increases in time spent using technology for those with more education. The strength of the relationship between degree and time using technology was medium ($\omega^2 = .11$).