Conducting and Interpreting t-Tests

Effect sizes for t-tests

Jenine Harris
Brown School

Import and clean the data

```
# import nhanes 2015-2016
nhanes.2016 <- read.csv("/Users/harrisj/Box/teaching/Teaching/Fall2020/d

# add labels to sex and rename variables
nhanes.2016.cleaned <- nhanes.2016 %>%
  mutate(RIAGENDR = recode_factor(.x = RIAGENDR,
                                  `1` = 'Male',
                                  `2` = 'Female')) %>%
  rename(sex = RIAGENDR) %>%
  rename(systolic = BPXSY1) %>%
  rename(systolic2 = BPXSY2) %>%
  mutate(diff.syst = systolic - systolic2)

# check the data
summary(object = nhanes.2016.cleaned)
```

```
##       SEQN          cycle              SDDSRVYR    RIDSTATR      sex
##  Min.   :83732   Length:9544        Min.   :9   Min.   :2   Male  :4676
##  1st Qu.:86222   Class :character   1st Qu.:9   1st Qu.:2   Female:4868
##  Median :88726   Mode  :character   Median :9   Median :2
##  Mean   :88720                      Mean   :9   Mean   :2
##  3rd Qu.:91210                      3rd Qu.:9   3rd Qu.:2
##  Max.   :93702                      Max.   :9   Max.   :2
##
##     RIDAGEYR        RIDAGEMN        RIDRETH1        RIDRETH3        RIDEXMON
##  Min.   : 0.00   Min.   : 0.00   Min.   :1.00   Min.   :1.000   Min.   :1.00
##  1st Qu.: 9.00   1st Qu.: 5.00   1st Qu.:2.00   1st Qu.:2.000   1st Qu.:1.00
```

Computing and interpreting an effect size for significant t-tests

- Small differences between means suggests that reporting statistical significance might be misleading.
- It is useful to have something to report that is about the size of the difference or the strength of the relationship.
- There are several measures of *effect size* for the t-test that are like the Cramer's V, phi, or odds ratio from the chi-squared test.
- Some people have argued that effect sizes are more important than p-values since p-values only report whether a difference or relationship from a sample is likely to be true in the population, while effect sizes provide information about the strength or size of a difference or relationship.
- In addition, in analyses of large samples, p-values usually reach statistical significance, even for very small differences and very weak relationships [@sullivan2012using].

Cohen's d effect size

- For t-tests, the effect size statistic is **Cohen's d**.
- Cohen's d is computed when the test results are **statistically significant** and can be computed for each type of t-test using a slightly different formula.
- The value of d is classified:
- Cohen's d = .2 to d < .5 is a **small** effect size
- Cohen's d = .5 to d < .8 is a **medium** effect size
- Cohen's d ≥ .8 is a **large** effect size

Cohen's d for one-sample t-tests

- The formula for Cohen's d for a one-sample t-test is shown with m_x as the sample mean for x, μ_x as the hypothesized or population mean, and s_x as the sample standard deviation for x.

$$d = \frac{m_x - \mu_x}{s_x}$$

$$d = \frac{120.5 - 120}{18.62} = .027$$

- The effect size was less than .03, which was not even get close to the **small** effect size value of .2.

Cohen's D from R

- Use R to get the value with the `cohensD()` function.
- For the one-sample test, there were two arguments to use: `x =` takes the variable name and `mu =` takes the hypothesized or population mean.

```
# Cohen's d effect size for one sample t
lsr::cohensD(x = nhanes.2016.cleaned$systolic, mu = 120)
```

```
## [1] 0.02897354
```

- There is a small difference between the hand-calculation and the R calculation of d due to rounding error.
- Both the 120.5 and the 18.62 were rounded in the hand-calculation while R uses many digits in memory during all computations.

Revised interpretation of one-sample t-test

- The mean systolic blood pressure in a sample of 7,145 people was 120.54 (sd = 18.62). A one-sample t-test found this mean to be statistically significantly different from the hypothesized mean of 120 [$t(7144) = 2.449$; $p = 0.014$]. The sample likely came from a population with a mean systolic blood pressure not equal to 120. While the sample mean was statistically significantly different from 120, the difference was relatively small with a very small effect size (Cohen's $d = .03$).

Cohen's d for independent samples t-tests

The formula for Cohen's d for an independent samples t-test is computed using:

$$d = \frac{m_1 - m_2}{s_{pooled}} = \frac{m_1 - m_2}{\sqrt{\frac{s_1^2 + s_2^2}{2}}}$$

- Where m_1 and m_2 are the sample means and s_1^2 and s_2^2 are the sample variances.

Computing Cohen's d for independent samples t-test

```
# compare means of systolic by sex
nhanes.2016.cleaned %>%
  drop_na(systolic) %>%
  group_by(sex) %>%
  summarize(m.sbp = mean(systolic),
            var.sbp = var(systolic))
```

```
## # A tibble: 2 x 3
##   sex    m.sbp var.sbp
##   <fct>  <dbl>   <dbl>
## 1 Male    122.    329.
## 2 Female  119.    358.
```

$$d = \frac{122.1767 - 118.9690}{\sqrt{\frac{329.2968+358.2324}{2}}} = .173$$

- The effect size from the calculations was .173, which is close to the **small** effect size value of .2 but the effect of sex on systolic blood pressure is not quite even a small effect.

Using R for Cohen's d in independent samples t-tests

- Enter a formula for `x =` and the data frame name for the `data =` argument.

```
# cohen's d effect size for indep sample t
lsr::cohensD(x = systolic ~ sex,
             data = nhanes.2016.cleaned)
```

```
## [1] 0.1729286
```

- There was a slight difference in the values between the hand-calculation and the R calculation due to rounding.

Adding Cohen's d to t-test interpretation

- There was a statistically significant difference [$t(7142.9989031) = 7.31$; $p < .05$] between the mean systolic blood pressure for males ($m = 122.18$) and females ($m = 118.97$) in the sample. The sample was taken from the US population indicating that males in the US likely have a different mean systolic blood pressure than females in the US. The difference between male and female mean systolic blood pressure was 3.21 in the sample; in the population this sample came from, the difference between male and female mean blood pressure was likely to be between 2.35 and 4.07 ($d = 3.21$; 95% CI: 2.35-4.07). The effect size for the relationship between sex and systolic blood pressure was small (Cohen's $d = .17$).

Cohen's d for paired samples t-tests

- The formula for Cohen's d for a paired samples t-test is:

$$d = \frac{m_d - 0}{s_d}$$

- Where m_d is mean difference between the two measures, in this case `systolic` and `systolic2`, and s_{diff} is the standard deviation of the differences between the two measures.

Computing Cohen's d for paired samples

```
# var and sample size of the difference variable
nhanes.2016.cleaned %>%
  drop_na(diff.syst) %>%
  summarize(m.sbp = mean(diff.syst),
            sd.sbp = sd(diff.syst))
```

```
##        m.sbp   sd.sbp
## 1 0.5449937 4.898043
```

$$d = \frac{.5449937 - 0}{4.898043} = .111$$

- The effect size after the calculations was .111, which was not quite up to the **small** effect size value of .2.

Computing Cohen's d for paired samples in R

- Enter each blood pressure measure as a separate vector and adding a `method = "paired"` argument:

```
# cohen's d effect size for indep sample t
lsr::cohensD(x = nhanes.2016.cleaned$systolic,
             y = nhanes.2016.cleaned$systolic2,
             method = "paired")
```

```
## [1] 0.1112676
```

Interpret the paired t-test with Cohen's d

- The mean difference between two measures of systolic blood pressure was statistically significantly different from zero [$t(7100) = 9.38$; $p < .05$]. The positive difference of 0.54 indicated that systolic blood pressure was significantly higher for the first measure compared to the second measure. While the mean difference in the sample was .54, the mean difference between the first and second measures in the population was likely between 0.43 and 0.66 (m_d = 0.54; 95% CI: 0.43-0.66). The effect size for the comparison of the two systolic blood pressure measures was very small (Cohen's d = .11).