

## 11.4: Two-sample effect size

### Introduction

An effect size is a measure of the strength of the difference between two samples. The effect size statistic is calculated by subtracting one sample mean from the other and dividing by the pooled standard deviation.

Measures of effect size, **Cohen's  $d$**

$$d = \frac{\bar{X}_1 - \bar{X}_2}{s_{pooled}}$$

where  $s_{pooled}$  is the **pooled standard deviation** for the two sample means. An equation for pooled standard deviation was provided in [Chapter 3.3](#), but we'll give it again here.

$$s_{pooled} = \sqrt{\frac{s_1^2 + s_2^2}{2}}$$

An alternative version of Cohen's  $d$  is available for the t-test test statistic value:

$$d = \frac{2t}{\sqrt{df}}$$

A  $d$  of one (1) indicates the effect size is equal to one standard deviation; a  $d$  of two (2) indicates the effect size between two sample means is equal to two standard deviations, and so on. Note that effect sizes complement inferential statistics such as p-values.

### What makes a large effect size?

Cohen cautiously suggested that values of  $d$

0.2 – small effect size

0.5 – medium effect size

0.8 – large effect size

That is, if the two group means don't differ by much more than 0.2 standard deviations, then the magnitude of the treatment effect is small and unlikely to be biologically important, whereas a  $d = 0.8$  or more would indicate a difference of 0.8 standard deviations between the sample means and, thus, likely to be an important treatment effect. Cohen (1992) provided these guidelines based on the following argument. The small effect 0.2 comes from the idea that it is much worse to conclude there is an effect when in fact there is no effect of the treatment rather than the converse (conclude no effect when there is an effect). The ratio of the Type II error (0.2) divided by the Type I error (0.05) gives us the penalty of 4. Similarly, for a moderate effect, 0.5/0.05 equals 10. Clearly, these are only guidelines (see Lakens 2013).

### Examples

The difference in average body size between 6 week old females of two strains of lab mice is 0.4 g (Table 11.4.1), and increases to 1.38 g by 16 weeks (Table 11.4.2).

Table 11.4.1. Average body weights of 6 week old female mice of two different inbred strains.†

Strain	$\bar{X}$	$s$
C57BL/6J	18.5	0.9
CBA/J	18.1	1.27

†Source: Jackson Laboratories: [C57BL/6J](#); [CBA/J](#)

Table 11.4.2. Average body weights of 16 week old female mice of two different inbred strains.†

Strain	$\bar{X}$	$s$



Strain	$\bar{X}$	$s$
C57BL/6J	23.9	2.3
CBA/J	25.38	3.76

†Source: Jackson Laboratories: [C57BL/6J](#); [CBA/J](#)

The descriptive statistics are based on weights of 360 individuals in each strain ([Jackson Labs](#)).

The differences are both statistically significant from a independent t-test, i.e., p-value less than 0.05. I'll show you how to calculate the independent t-test given summary statistics (means, standard deviations), for Table 11.4.1 data, then I will ask you to do this on your own in Questions.

Write an R script, using example data from Table 11.4.1:

```
sdd1 = 0.9
var1 = sdd1^2
sdd2 = 1.27
var2 = sdd2^2
mean1 = 18.5
mean2 = 18.1
n1 = 360
n2 = 360
dff = n1+n2-2
pooledSD <-sqrt((var1+var2)/2)
pooledSEM <-sqrt(var1/n1 + var2/n2); pooledSEM
tdiff<-(mean1-mean2)/pooledSEM; tdiff
pt(tdiff, df=dff, lower.tail=FALSE)
#get two-tailed p-value
2*0.0000006675956
#get cohen's d
2*tdiff/sqrt(dff)
```

Results from the calculations we report (value of the test statistic, degrees of freedom, p-value), and the effect size, then are

```
t = 4.875773, df = 718, p-value = 0.0000006675956
cohen's d = 0.364
```

Now, I'm from the school of “don't reinvent the wheel” or “someone has already solved your problems” (Freeman et al 2008), when it comes to coding problems. And, as you would expect, of course someone has written a function to calculate the t-test given summary statistics. In addition to base R and the `pwr` package (see [Chapter 11.5](#)), the package `BSDA` contains several nice functions for power calculations.

To follow this example, install `BSDA` , then run the following code:

```
require(BSDA)
tsum.test(mean1, sdd1, n1, mean2, sdd2, n2, alternative = "two.sided", mu = 0, var.eq)
```

R output:

```
Standard Two-Sample t-Test
```



```
data: Summarized x and y
t = 4.8758, df = 718, p-value = 0.000001335
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
0.2389364 0.5610636
sample estimates:
mean of x mean of y
18.5 18.1
```

Similarly, Cohen's  $d$  is available from a package called `effsize`.

#### Note:

One reason to “re-invent the wheel”: I only needed the one function; the BSDA package contains more 330 different objects/functions. A simple way to check how many objects in a package, e.g., BSDA, run

```
ls("package:BSDA")
```

BSDA stands for “Basic Statistics and Data Analysis,” and was intended to accompany the 2002 book of the same title by Larry Kitchens.

And of course, if using someone else's code, give proper citation!

## Questions

1. We needed an equation to calculate pooled standard error of the mean ( `pooledSEM` in the R code). Read the code and write the equation used to calculate the pooled SEM.
2. Calculate the t-test and the effect size for the Table 11.4.1 data, but at three smaller sample sizes. Change 360 to  $n_1 = n_2 = 20$ , repeat for  $n_1 = n_2 = 50$ , and finally, repeat for  $n_1 = n_2 = 100$ . Use your own code, or use the `tsum.test` function from the `BSDA` package.
3. Calculate Cohen's effect size  $d$  for each new calculation based on a different sample size.
4. Create a table to report the p-values from the t-tests, the effect size, for each of the four  $n_1 = n_2 = (20, 50, 100, 360)$ .
5. True or false. The mean difference between sample means remains unaffected by sample size.
6. True or false. The effect size between sample means remains unaffected by sample size.
7. Based on comparisons in your table, what can you conclude about p-value and “statistical significance?” About effect size?
8. Repeat questions 2 – 7 for Table 11.4.2

This page titled [11.4: Two-sample effect size](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Michael R Dohm](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.