How Effect Size Impacts Sample Size

Note and Disclaimer:

(1) This PDF is part of YouTube tutorials (https://youtu.be/3yFDBmdAnTs). This PDF is for individual, personal usage only.
(2) The author accepts no responsibility for the topicality, correctness, completeness or quality of the information provided.

The following effect size numbers are based on Jacob Cohen's *Statistical Power Analysis for the Behavioral Sciences* (Second Edition, p. 82)

Effect Size	d	r
Small	0.2	0.1
Medium	0.5	0.3
Large	0.8	0.5

Cohen's *d*

Cohen d is used to quantify the size of the difference between two groups, taking into account the variability within each group.

That is, x is categorical variable, whereas y is numerical variable (i.e., Y is on a continuous scale).

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

Pearson *r*

When both x and y are numerical, we need to use pearson correlation coefficient to measure effect size r.

$$r_{xy} = \frac{cov(x,y)}{s_x s_y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Convert *r* to *d*

$$d = \frac{2r}{\sqrt{1 - r^2}}$$

Convert *d* to *r*

$$r = \frac{d}{\sqrt{4 + d^2}}$$

Effect Size and Sample Size

In general, as the effect size increases, the required sample size decreases, assuming that statistical power (1 - β) and the alpha level (α) are held constant, respectively.

R Code Example 1: t-test

To calculate the needed sample size for a two-sample t-test with a specified power of 0.8, a significance level of 0.05, and an effect size of 0.5, you can use the pwr.t.test function in R.

effect size = 0.50

```
library(pwr)
```

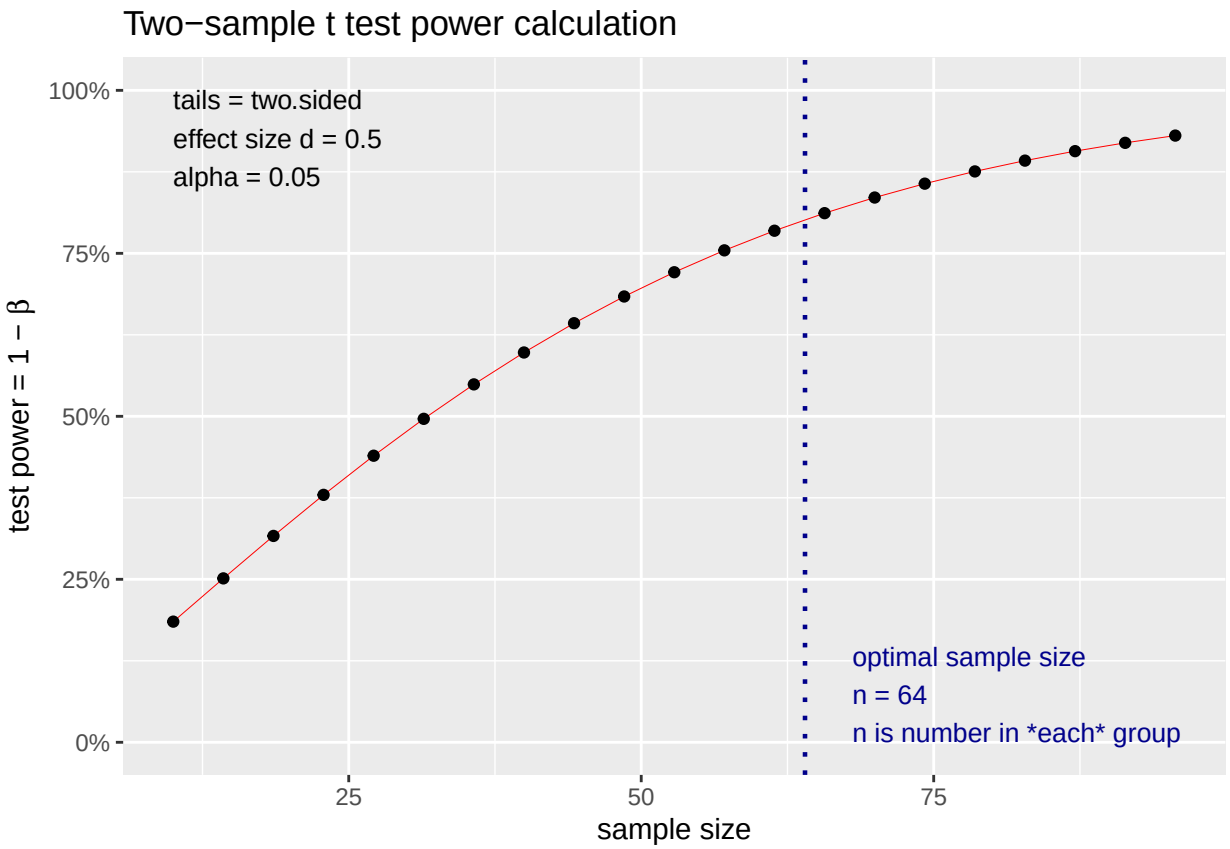
```
## Warning: package 'pwr' was built under R version 4.1.3
```

```
result_1<-pwr.t.test(n=NULL,
           d = 0.5,
           sig.level = 0.05,
           power = 0.80,
           type = "two.sample",
           alternative="two.sided")

print(result_1)
```

```
##
##      Two-sample t test power calculation
##
##               n = 63.76561
##               d = 0.5
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

```
plot(result_1)
```


Two−sample t test power calculation
tails = two.sided
effect size d = 0.5
alpha = 0.05
optimal sample size
n = 64
n is number in *each* group
100%
75%
50%
25%
0%
test power = 1 − β
25
50
75
sample size

Increasing the effect size from 0.5 to 0.8:

effect size = 0.80

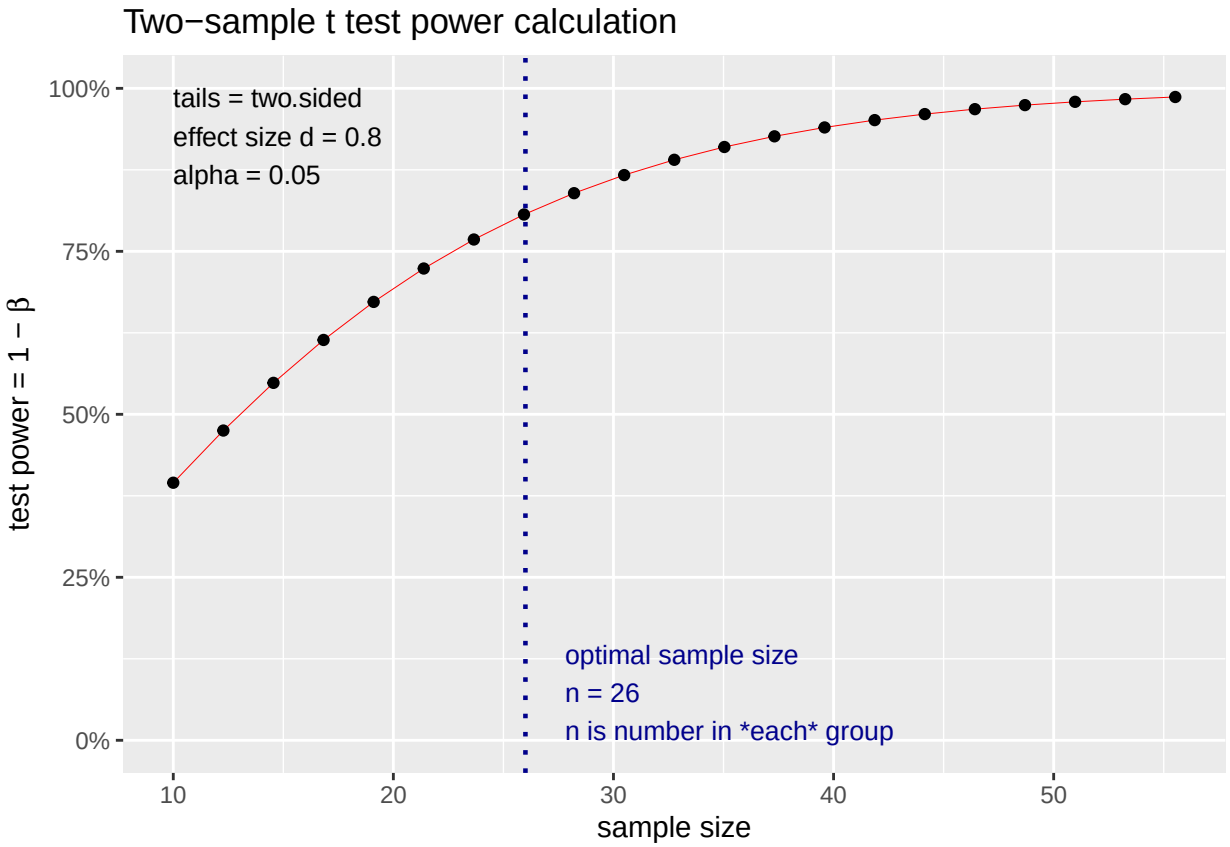
```
library(pwr)

result_2<-pwr.t.test(n=NULL,
           d = 0.8,
           sig.level = 0.05,
           power = 0.80,
           type = "two.sample",
           alternative="two.sided")

print(result_2)
```

```
##
##      Two-sample t test power calculation
##
##               n = 25.52458
##               d = 0.8
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

```
plot(result_2)
```


Two−sample t test power calculation
tails = two.sided
effect size d = 0.8
alpha = 0.05
optimal sample size
n = 26
n is number in *each* group
test power = 1 − β
100%
75%
50%
25%
0%
10
20
30
40
50
sample size

R Code Example 2: Correlation

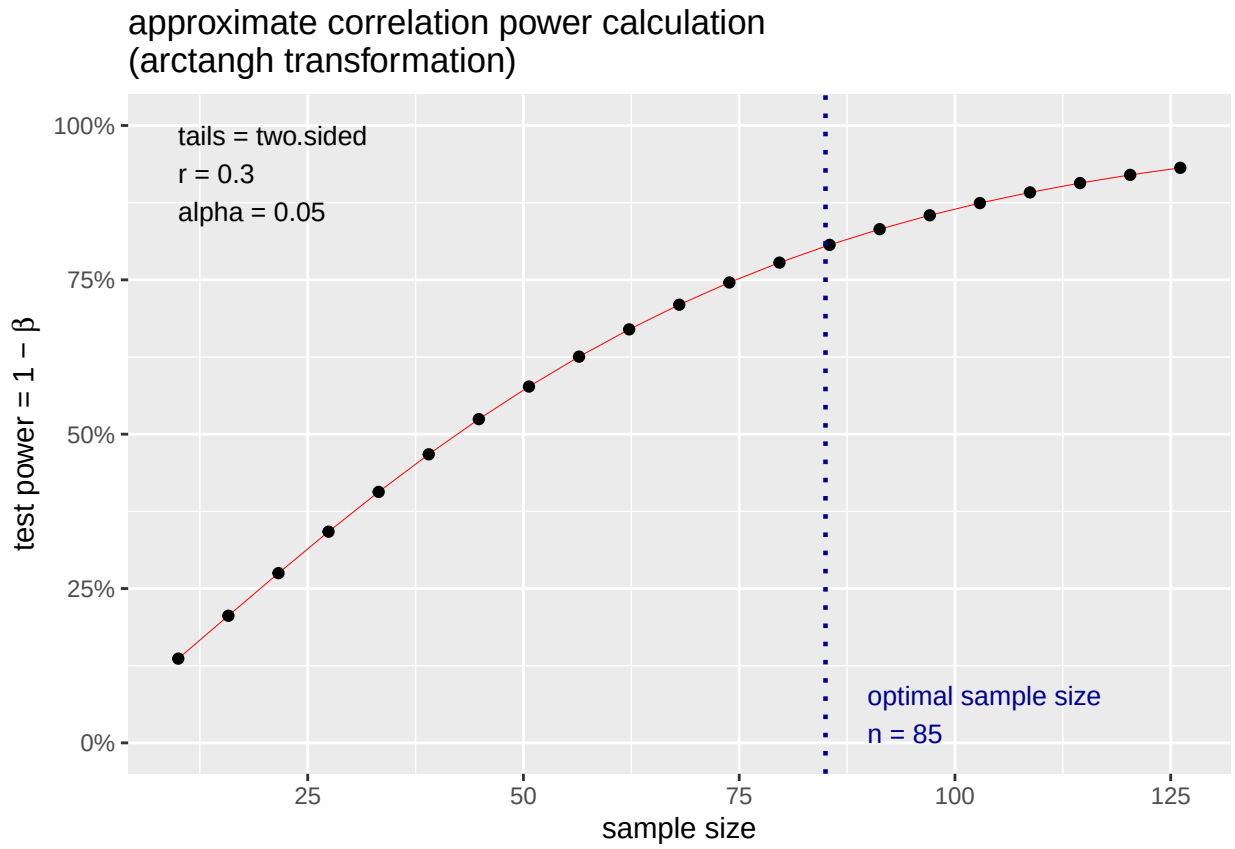
effect size = 0.30

```
result_3<-pwr.r.test(n = NULL,
           r = 0.3,
           sig.level = 0.05,
           power = 0.80,
           alternative = "two.sided")

print(result_3)
```

```
##
##      approximate correlation power calculation (arctangh transformation)
##
##               n = 84.07364
##               r = 0.3
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
```

```
plot(result_3)
```



effect size = 0.50

```
result_4<-pwr.r.test(n = NULL,
           r = 0.5,
           sig.level = 0.05,
           power = 0.80,
           alternative = "two.sided")

print(result_4)
```

```
##
##      approximate correlation power calculation (arctangh transformation)
##
##                n = 28.24841
##                r = 0.5
##        sig.level = 0.05
##            power = 0.8
##      alternative = two.sided
```

```
plot(result_4)
```