How Sample Size Impacts Statistical Power

Note and Disclaimer

(1) This PDF is part of YouTube tutorials (https://youtu.be/3uSe5N7gYN0). 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.

Statistical Power

Correct Action	H0 is True Should Not Reject H0	H0 is False Should Reject H0
A Test Rejects H0 (Positive)	α	$1-\beta$
A Test Doesn't Reject H0 (Negative)	$1-\alpha$	β

α = probability of Type I error, known as a "false positive"

β = probability of Type II error, known as a "false negative"

$1-\beta$ is also called power, or statistical power. It is the probability that, null hypothesis is false and we correctly reject the null hypothesis.

Sample Size

Sample size refers to the number of observations or individuals measured or included in a study.

Effect Size

The following effect size numbers are from Jacob Cohen's *Statistical Power Analysis for the Behavioral Sciences*.

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

Remarks

Remark 1: For a given effect size, as sample size increases, the power also increases.

Remark 2: For a given effect size, as sample size increases, the observed p-value will decrease, meaning that we are more likely to reject null hypothesis.

Examples: Impact of Sample Size on Statistical Power

We are going to use R code to demonstrate *Remark 1*, namely as sample size increases, the power also increases.

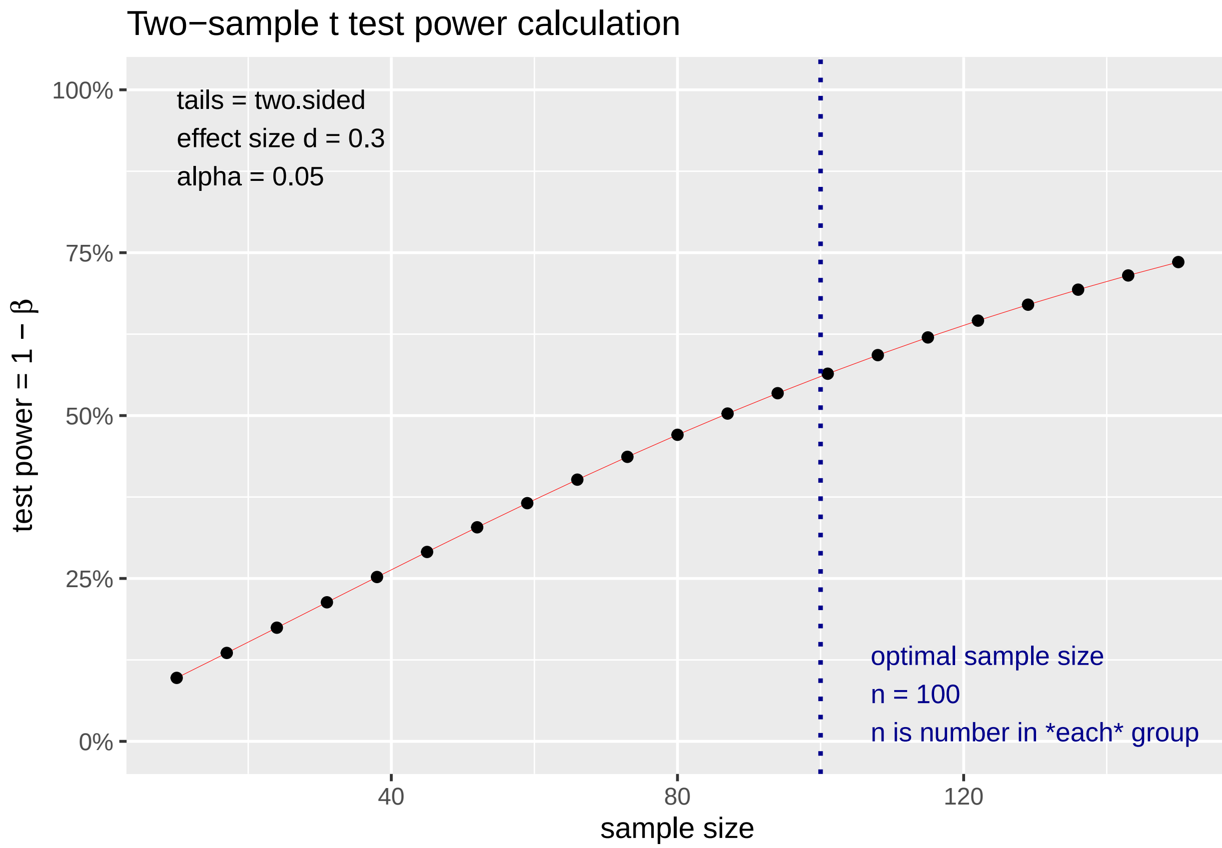
```
library(pwr)

p_output_1<-pwr.t.test(n=100,
           d = 0.3,
           sig.level = 0.05,
           power = NULL,
           type = "two.sample",
           alternative="two.sided")

print(p_output_1)
```

```
##
##      Two-sample t test power calculation
##
##               n = 100
##               d = 0.3
##       sig.level = 0.05
##           power = 0.5600593
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

```
plot(p_output_1)
```



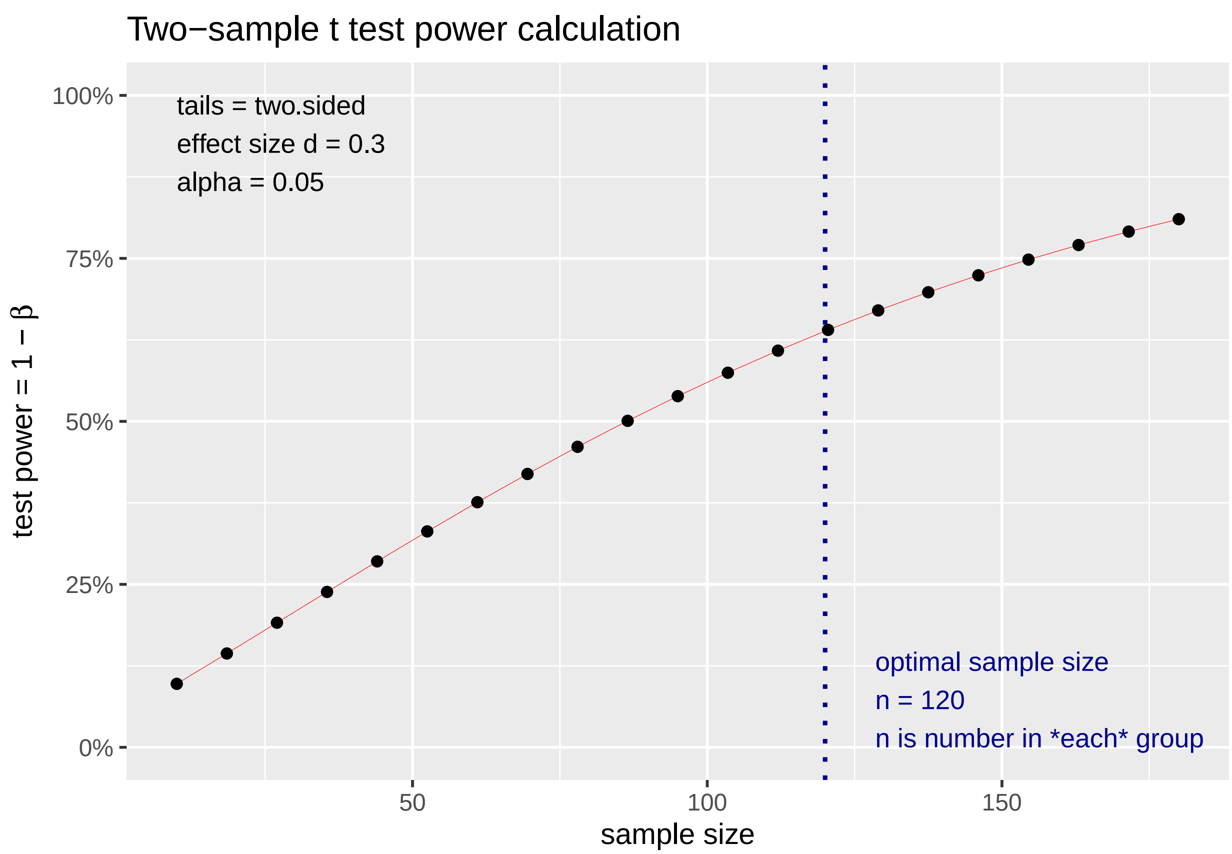
Two−sample t test power calculation
tails = two.sided
effect size d = 0.3
alpha = 0.05
optimal sample size
n = 100
n is number in *each* group
100%
75%
50%
25%
0%
test power = 1 − β
40
80
120
sample size

```r
p_output_2<-pwr.t.test(n=120,
            d = 0.3,
            sig.level = 0.05,
            power = NULL,
            type = "two.sample",
            alternative="two.sided")

print(p_output_2)
```

```
##
##      Two-sample t test power calculation
##
##               n = 120
##               d = 0.3
##       sig.level = 0.05
##           power = 0.6385054
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

```r
plot(p_output_2)
```



Example: Impact of Sample Size on p-value

Illustrate *Remark 2.*That is, for a given effect size, as sample size increases, the observed p-value will decrease, meaning that we are more likely to reject null hypothesis.

Typically, we reject null hypothesis when p-value is smaller than 0.05, which is the alpha level (i.e., false positive).

Predetermined Effect Size:

$$d = \frac{4}{10} = 0.4$$

Sample size = 200

```r
# Set the seed for reproducibility
set.seed(123)

# Specify parameters
n1 <- 100  # Sample size for group 1
n2 <- 100  # Sample size for group 2
mean_diff <- 4  # Desired mean difference between groups
sd <- 10  # Common standard deviation for both groups

# Generate random samples for two groups
group1 <- rnorm(n1, mean = 0, sd = sd)
group2 <- rnorm(n2, mean = mean_diff, sd = sd)

# Perform a two-sample t-test
t.test(group1, group2)
```

```
##
##  Welch Two Sample t-test
##
## data:  group1 and group2
## t = -1.5194, df = 197.35, p-value = 0.1303
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4.6428618  0.6019159
## sample estimates:
## mean of x mean of y
## 0.9040591 2.9245320
```

Sample size = 240

```r
# Set the seed for reproducibility
set.seed(123)

# Specify parameters
n1 <- 120  # Sample size for group 1
n2 <- 120  # Sample size for group 2
mean_diff <- 4  # Desired mean difference between groups
sd <- 10  # Common standard deviation for both groups

# Generate random samples for two groups
group1 <- rnorm(n1, mean = 0, sd = sd)
group2 <- rnorm(n2, mean = mean_diff, sd = sd)

# Perform a two-sample t-test
t.test(group1, group2)
```

```
##
##  Welch Two Sample t-test
##
## data:  group1 and group2
## t = -2.9529, df = 235.89, p-value = 0.003466
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -5.974416 -1.192717
## sample estimates:
## mean of x mean of y
## 0.1544151 3.7379818
```