**Power**

From earlier:

In hypothesis testing, we will make the assumption that Ho is true and then try to prove it to be incorrect using evidence gathered in the sample. Thus, it is important to define

P(Reject Ho | Ha is TRUE).

This is called the power of the test. Notice where this result falls in the above table and it has a probability of 1 – β.

A power analysis is often performed BEFORE taking a sample to determine how likely it is to reject Ho under a number of conditions. If your research involves some type of hypothesis test, you will often be asked to provide a power analysis to justify your research design. In fact, many (most?) grant funding agencies will not fund a research proposal without some type of power analysis.

Consider the case of a left-tail test:

Ho: μ≥μ0  
Ha: μ<μ0

We reject Ho if t < -tα, n-1. A power analysis examines

P(T < -tα,n-1)

assuming that Ha is true.

Because Ho is no longer true, μ0 is NOT the population mean and



no longer has a t distribution with n – 1 degrees of freedom. Let μa be the actual population mean. This leads to T having a non-central t distribution with n – 1 degrees of freedom and a “non-centrality” parameter of .

The probability distribution can be written as f(t) = \_\_ just like the regular t distribution, but the formula is much more complicated, so this is why I am not showing it here.

Below are example non-central t distributions for n = 30 (see non-central\_t.R for code):



Notes:

* The standard deviation of Y1, …, Yn, σ, is needed for this calculation. Why is this a problem?
*  is often referred to as the effect size. Notice the absolute value in the formula.
* The t distribution itself has a non-centrality parameter of 0. Notice what happens to the non-centrality parameter when . What would be the “power” in this situation?

For the case of a right-tail test,

Ho: μ≤μ0  
Ha: μ>μ0

we reject Ho if t > tα, n-1. The power is then P(T > tα,n-1), where Ha is true and T has a non-central t distribution. For the case of a two-tail test,

Ho: μ=μ0  
Ha: μ≠μ0

we reject Ho if |t| > tα/2,n-1. The power is then   
P(|T| > tα/2,n-1), where Ha is true and T has a non-central t distribution.

The power.t.test() function calculates these probabilities in R.

Example: Tire life (tire\_life.R)

The hypotheses were

Ho: μ≥22,000  
Ha: μ<22,000

with α = 0.01. What is the probability of rejecting the null hypothesis if the true mean is actually 21,500?

We need to calculate P(T < -t0.01,99) where T has a non-central t distribution with 99 degrees of freedom and a “non-centrality” parameter of . What should we use for σ?

Suppose σ = 1,000. Below are the necessary calculations from R:

> diff.in.means <- 20500 - 21000

> effect.size <- abs(diff.in.means)

> sigma <- 1000

> alpha <- 0.01

> n <- 100

> power.t.test(n = n, d = effect.size, sd = sigma,

sig.level = alpha, type = "one.sample", alternative =

"one.sided")

One-sample t test power calculation

n = 100

delta = 500

sd = 1000

sig.level = 0.01

power = 0.9954011

alternative = one.sided

> crit.val <- qt(p = 0.99, df = n - 1)

> pt(q = -crit.val, df = n - 1, ncp = sqrt(n) \*

diff.in.means / sigma)

[1] 0.9954011

> #Equivalent using effect.size and a positive critical

Value

> #This is what power.t.test() does

> 1-pt(q = crit.val, df = n - 1, ncp =

sqrt(n)\*effect.size/sigma)

[1] 0.9954011

The probability of reject Ho in this setting is quite high. However, note that we had to make a number of assumptions to calculate this probability.

What if there was more than one effect size of interest? Alternatively, what if the effect size was set at one value, but the sample size was not known? Whenever an effect size or some other “assumed” value is not set, a power curve can be graphed which plots the power versus a varying value like effect or sample size.

Example: Tire life (tire\_life.R)

Continuing the last example, below is a power curve plotting power vs. μa – μ0:

> curve(expr = pt(q = -crit.val, df = n - 1, ncp =

sqrt(n)\*x/sigma), xlim = c(-500, 0), col = "red", xlab

= expression(mu[a] - mu[0]), ylab = "Power", main =

expression(paste("Power vs. ", mu[a] - mu[0])),

panel.first = grid(lty = "dotted", col = "lightgray"))



Below is a power curve plotting power vs. effect size using the power.t.test() function in the expr argument instead.

> save.power <- power.t.test(n = 100, d = 200, sd = 1000,

sig.level = 0.01, type = "one.sample", alternative =

"one.sided")

> names(save.power)

[1] "n" "delta" "sd" "sig.level"

[5] "power" "alternative" "note" "method"

> save.power$power

[1] 0.3618013

> curve(expr = power.t.test(n = 100, d = x, sd = 1000,

sig.level = 0.01, type="one.sample", alternative =

"one.sided")$power, xlim = c(0, 500), col = "red",

xlab = expression(mu[a] - mu[0]), ylab = "Power",

main = expression(paste("Power vs. ", mu[a] - mu[0])),

panel.first = grid(lty = "dotted", col = "lightgray"))



Question: What is the power at effect size = 0?

In many situations, you will want to determine a particular sample size that gives you a certain power level to detect a specified effect size. Below is the code that can be used to find the sample size to achieve a power level of 0.80 when the effect size is 500 and σ = 1000.

> save.power <- power.t.test(n = seq(from = 10, to = 100,

by = 10), d = effect.size, sd = sigma, sig.level =

alpha, type = "one.sample", alternative = "one.sided")

> save.power

One-sample t test power calculation

n = 10, 20, 30, 40, 50, 60, 70, 80, 90, 100

delta = 500

sd = 1000

sig.level = 0.01

power = 0.1654013, 0.4018035, 0.6117925, 0.7659006,

0.8666483, 0.9274354, 0.9619760, 0.9806991,

0.9904663, 0.9954011

alternative = one.sided

> names(save.power)

[1] "n" "delta" "sd" "sig.level"

"power" "alternative" "note"

[8] "method"

> data.frame(n = save.power$n, power =

round(save.power$power, 2))

n power

1 10 0.17

2 20 0.40

3 30 0.61

4 40 0.77

5 50 0.87

6 60 0.93

7 70 0.96

8 80 0.98

9 90 0.99

10 100 1.00

> power.t.test(n = 43, d = effect.size, sd = sigma,

sig.level = alpha, type = "one.sample", alternative =

"one.sided")

One-sample t test power calculation

n = 43

delta = 500

sd = 1000

sig.level = 0.01

power = 0.8011974

alternative = one.sided

> curve(expr = power.t.test(n = x, d = 500, sd = 1000,

sig.level = 0.01, type = "one.sample", alternative =

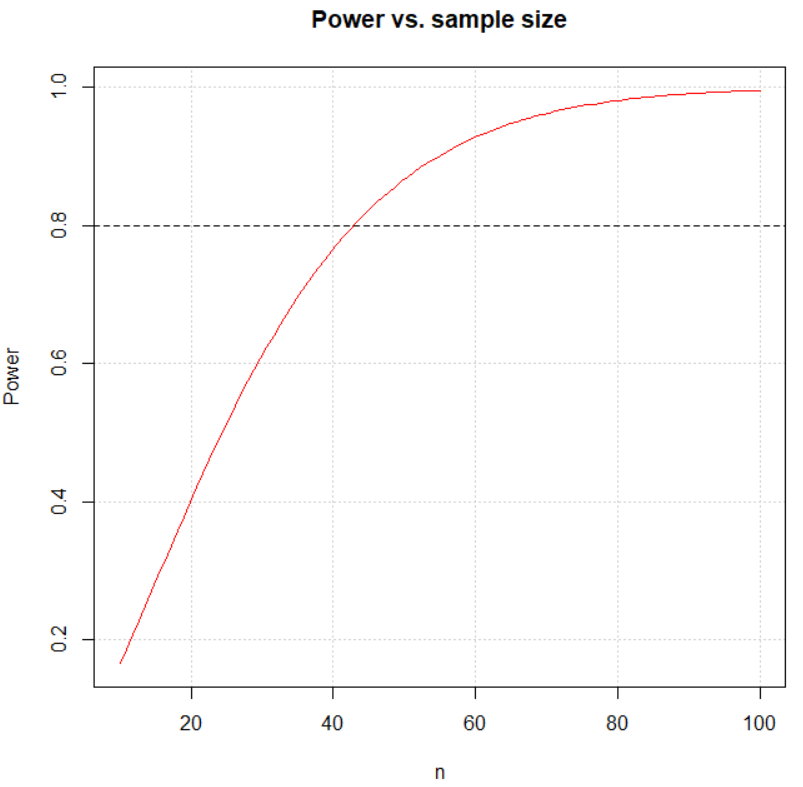
"one.sided")$power, xlim = c(10, 100), col = "red",

xlab = "n", ylab = "Power", main = "Power vs. sample

size", panel.first = grid(lty = "dotted", col =

"lightgray"))

> abline(h = 0.8, lty = "dashed")



The sample size needs to be at least 43 to obtain a power of 0.80 or above.