Power analysis via for t-test based upon precision of effect size estimate

Crispin Jordan

07/06/2021

Revisiting our previous power analysis for a t-test

In a previous file, we learned to conduct a power analysis for data analysed with a t-test, where we defined a 'successful' experiment as one that could detect (based on `p < 0.05`) an effect size that we designated. In that previous example, the mean values of the groups equaled 5 and 6, giving an effect size of 1 unit (i.e., 6-5 = 1). We used this code:

```
#It can be a good idea to run this code before the simulations; it deletes all objects
#from R, leaving you with a 'clean' working environment
rm(list=ls())

mean.1 <- 5
mean.2 <- 6
sd.both <- 0.5
sample.size <- 5
counter <- 0
nsims <- 10000

for(i in 1:nsims){
  group.1 <- rnorm(sample.size, mean.1,sd.both)
  group.2 <- rnorm(sample.size, mean.2,sd.both)
  t.out <- t.test(group.1,group.2,var.equal = TRUE)
  if(t.out$p.value < 0.05){counter <- counter + 1}
}

counter/nsims
```

```
## [1] 0.7863
```

Conducting a power analysis for a different definition of 'success'

Finding `p < 0.05` can be one definition of success in the context of a power analysis. But we can use alternative criteria for define 'success', as well.

Here, we'll use the ability to estimate an effect size with a desired level of precision as our criterion of 'success'. As we have discussed in lectures (videos), `p < 0.05` provides limited biological interpretation of data, whereas measurement of an effect's size can provide highly relevant insight into one's data. It is knowing how big or small an effect might be that can reveal whether an effect is likely biologically important. Therefore, we very often wish to estimate an effect size and to do so as precisely as possible.

This file demonstrates how to conduct a power analysis when we use the ability to estimate an effect size to a given level of precision as our criterion for 'success'. To do so, we will use the estimate of the **standard error (SE) for the effect size** to determine an experiment's 'success'.

Consider the example, above, where we set the effect size in our experiment to 1. Imagine that we wished to estimate the effect size for our experiment so that the 95% confidence intervals were +/- 0.7. i.e., imagine that obtaining 95% CI's for our effect size that were equal to or less than 0.07 above and below the mean constituted a 'successful' experiment. How can we obtain the 95% CI for the effect size in order to judge whether our experiment was a success? There are several ways, and we'll explore one.

Recall that, roughly speaking, 95% CI's for a measurment equal approximately 2*SE for that measurement (i.e., 2 times the standard error). When R implements the function, `t.test()`, it calculates the standard error (SE) for the effect size and stores it in the output. We obtain this SE by appending `$stderr` to an object that contains output from `t.test()`. For example, drawing from the code above:

```
t.out <- t.test(group.1,group.2,var.equal = TRUE)
t.out$stderr
```

```
## [1] 0.4425656
```

Knowing this, it is a simple matter to alter our code, above, to determine the power with which (i.e., probability that) we estimate an effect size with SE below (i.e., as precise or more precise than) some specified value. Below, we alter the code to:

- set the desired `stderr` to 0.35; recall that this equals half of our desired 95% CI's (0.7). We set this value at the top of our code by establishing the parameter, `max.SE`
- alter the `if()` statement that determines whether the experiment was a 'success': we no longer compare a p-value vs. 0.05. Instead, we compare the SE vs. the desired maximum SE for the effect size (`max.SE`).

```
#It can be a good idea to run this code before the simulations; it deletes all objects
#from R, leaving you with a 'clean' working environment
rm(list=ls())

max.SE <- 0.35
mean.1 <- 5
mean.2 <- 6
sd.both <- 0.5
sample.size <- 5
counter <- 0
nsims <- 10000

for(i in 1:nsims){
  group.1 <- rnorm(sample.size, mean.1,sd.both)
  group.2 <- rnorm(sample.size, mean.2,sd.both)
  t.out <- t.test(group.1,group.2,var.equal = TRUE)
  if(t.out$stderr < max.SE){counter <- counter + 1}
}

counter/nsims
```

```
## [1] 0.7261
```

The probability, given as output, above, represents the experiment's **power** to estimate an effect size with an SE equal to `max.SE`, or smaller. If this probability equals, say, 0.71, then we say the experiment has 71% power to estimate an effect size with an SE of `max.SE`, or smaller (remember: smaller SE's mean that we estimate something with greater precision).

Something cool.

I want you to see something cool.

Let's try changing our effect size to something really dramatic. Let's change the effect size from a unit of 1 (i.e., 6-5) to a unit of a million (i.e., 1000005 - 5). See what happens:

```
#It can be a good idea to run this code before the simulations; it deletes all objects
#from R, leaving you with a 'clean' working environment
rm(list=ls())

max.SE <- 0.35
mean.1 <- 5
mean.2 <- 1000005
sd.both <- 0.5
sample.size <- 5
counter <- 0
nsims <- 10000

for(i in 1:nsims){
  group.1 <- rnorm(sample.size, mean.1,sd.both)
  group.2 <- rnorm(sample.size, mean.2,sd.both)
  t.out <- t.test(group.1,group.2,var.equal = TRUE)
  if(t.out$stderr < max.SE){counter <- counter + 1}
}

counter/nsims
```

```
## [1] 0.7182
```

Notice that the power has hardly changed at all. The power from our simulations with effect size of 1 and of 1000000 is actually the same, but we obtained slightly different numbers for these two sets of simulation due to stochasticity inherent in simulations. (Try running this code, yourself, several times - you'll see that power fluctuates slightly among sets of simulations). What's going on here? Why has power changed so little?

For some types of data and for some types of analyses, the SE of an effect size is independent from the magnitude of the effect size, itself. This is true for t-tests, and this quality is very useful: it means that, **if we design an experiment to have a desired power to detect an effect size to a specified precision, we may not need to guess a meaningful effect size ahead of time**. Whether we do need to guess a meaningful effect size ahead of time will depend on whether the SE of the effect size is independent of the magnitude of the effect size in your given the analysis. If you're unsure (for your own data/analyses) whether the SE of an effect size is independent of the size of the effect, you can perform the simple experiment we used here: change the effect size to test whether power changes, too.

Cool, eh?