

# Sample Size and Power

*(your name here)*

*27 September 2019*

## Sample Size and Power Tools

For this exercise, we'll be giving you some very general practice in using a few key R packages for power and sample size calculations. Though many other tools exist online and in the additional resources we've noted in the class powerpoint, it's always best to try a few different resources before you settle on the one you like best. Note that you won't need any underlying data for this exercise, though if you want to conduct more complicated power calculation procedures in the future, you'll eventually need to start simulating datasets!

**(dont forget to insert your name at the top)**

## Part 0. Install packages and run libraries

First, start by installing the needed packages (you may already have 'ggplot2' installed):

```
## Install Power Packages

#install.packages('pwr', dependencies = TRUE)
#install.packages("clusterPower")
#install.packages("ggplot2")

library(pwr)
library(clusterPower)
library(ggplot2)

## Be sure to take a few moments to explore options within the various packages:
? power.t.test           ## The built-in power command(s) for individual RCTs
? `pwr-package`         ## The pwr package for individual RCTs
? `clusterPower-package` ## The clusterPower package for CRCTs, click "Index"
? plot.power.htest       ## Visualizations of sample size vs. test power
```

## Part I. Power and sample size calcs for individual-level RCTs

### *example one*

First, lets explore the most basic of the commands, `power.t.test()`. If you want to find one of the parameters (sample size; MDE; power), just leave that one out. In this example, we're looking for the sample size:

```
# ex i.
power.t.test(n = ,                ## Required sample size (leave one out)
             delta = .8,          ## Minimum detectable effect size (delta)
             sd = 1,              ## Standard deviation (default = 1)
             sig.level = 0.05,    ## Significance level (alpha, default = 0.05)
             power = 0.8,         ## Power (1-beta)
             type = "two.sample", ## type of t test: one-, two- or paired-sample
             alternative = "two.sided") ## one.sided vs. two.sided
```

```
##
##      Two-sample t test power calculation
##
##              n = 25.52463
##              delta = 0.8
##              sd = 1
##              sig.level = 0.05
##              power = 0.8
##      alternative = two.sided
##
## NOTE: n is number in *each* group
```

Looks like the sample size required (n) is 26 in each group (always round up with decimals).

Question 1. Now, use both `'power.t.test()'` and `'pwr.t.test()'` to calculate the minimum number of samples required for an effect size (d) of 0.3, standard deviation of 1, a significance level of 0.01, and a power of 0.8 for a two-sample test (note that there is no parameter for sd in the `'pwr'` command):

```
# 1.
```

a. What is the difference in MDE between the two tests? What does that tell you?

Question 2. Next, using `power.t.test`, explore the impact of estimates of the standard deviation. For a two-armed trial with  $\alpha = 0.05$ , power = 0.80, to detect a MDE of 0.5, what sample size is necessary if the standard deviations are 1?

What if standard deviation is .75?

And if standard deviation is .5?

# 2.

a. What do your findings tell you about the importance of the standard deviation in estimating the required sample size?

*example ii.*

Now lets look at another command, `pwr.t2n.test()`. This is the power calculation formula for a t-test (mean test) of two samples of potentially different sizes. Here the `n` is replaced by `n1` and `n2`:

```
# ex ii.
pwr.t2n.test(n1= 1000,          # number of observations in the first sample
             n2= 1000,          # number of observations in the second sample
             d = ,
             sig.level = .05,
             power = .9,
             alternative = "two.sided")
```

```
##
##      t test power calculation
##
##          n1 = 1000
##          n2 = 1000
##          d = 0.1450343
##      sig.level = 0.05
##          power = 0.9
##      alternative = two.sided
```

With a sample of 2000 (1000 in each arm) we are powered to detect a 0.14 MDE.

Question 3. Say we finished our trial, and we had some differential attrition between arms. Using `pwr.t2n.test()`, vary the sample size in group `n2` to 950, then to 800, then to 500.

# 3.

- a. What happened to the MDE at each of the three different sizes for `n2`?
- b. How concerned should we be about the power of our study based on differential attrition?
- c. What else should we be concerned about regarding differential attrition?

*example iii.*

Lets look at a trickier command for a binary outcome at the individual level: `power.prop.test()`. There are two new parameters here. Let's think of `p1` as the proportion of those with a value of 1 (our main binary outcome) in the control group, and `p2` as the proportion of those with a value of 1 in the treatment group, and consider that we are conducting a sample size (or power) calculation when we have some idea of what the outcome of the study is, or might be.

```
# ex iii.

power.prop.test(n = ,
  p1 = .15,          ## Proportion with a value of 1 in the absence of treatment
  p2 = .25,          ## Proportion with a value of 1 with the treatment
  sig.level = 0.05,
  power = .8,
  alternative = "two.sided")
```

```
##
##      Two-sample comparison of proportions power calculation
##
##              n = 249.982
##              p1 = 0.15
##              p2 = 0.25
##      sig.level = 0.05
##              power = 0.8
##      alternative = two.sided
##
## NOTE: n is number in *each* group
```

In this case, we suspect that the proportion of the control group (`p1`) with a value of 1 (in the absence of the treatment) is 15%. The treatment group (`p2`) meanwhile, will be higher at 25%. Thus, we want to know what sample size (`n`) will necessary to detect a .10 percentage point difference between the two groups. The answer seems to be 500 (250 in each group).

Question 4.1. Imagine you're a broke PhD student (big stretch, I know) and you'd like to conduct a RCT for your dissertation. You can only afford to sample about 500 households. You have some idea that the baseline prevalence for the control group ( $p_1$ ) will be around .25. The literature (and your advisors) tell you that you need to be powered to detect a 10 percentage point difference between treatment and control, i.e.  $(p_2) = .35$ . Use the `power.prop.test()` command to determine how big your sample needs to be.

# 4.1.

- a. Compare the sample size you find to the sample size from example iii. How does it differ?



Question 4.2. You've gotten too curious for your own good. You worry about your assumption regarding the baseline prevalence of the control group, but you want to make your advisors happy and be able to detect that .10 pp effect size. You wonder, what if .25 was too low? What sample size will you need if  $p_1 = .35$  and  $p_2 = .45$ ?

# 4.2.

Question 4.3. You broke the bank. But you try not to panic. What is the tradeoff between sample size and MDE anyway? You go back to your original assumption that  $p_1 = .25$  and keep your sample size constant. What value in the treatment group will you be powered to detect if you keep your sample size at 500 households?

# 4.3.

a. What did you learn about the tradeoff between sample size and MDE?

## Part II. Individual-level data vizualization

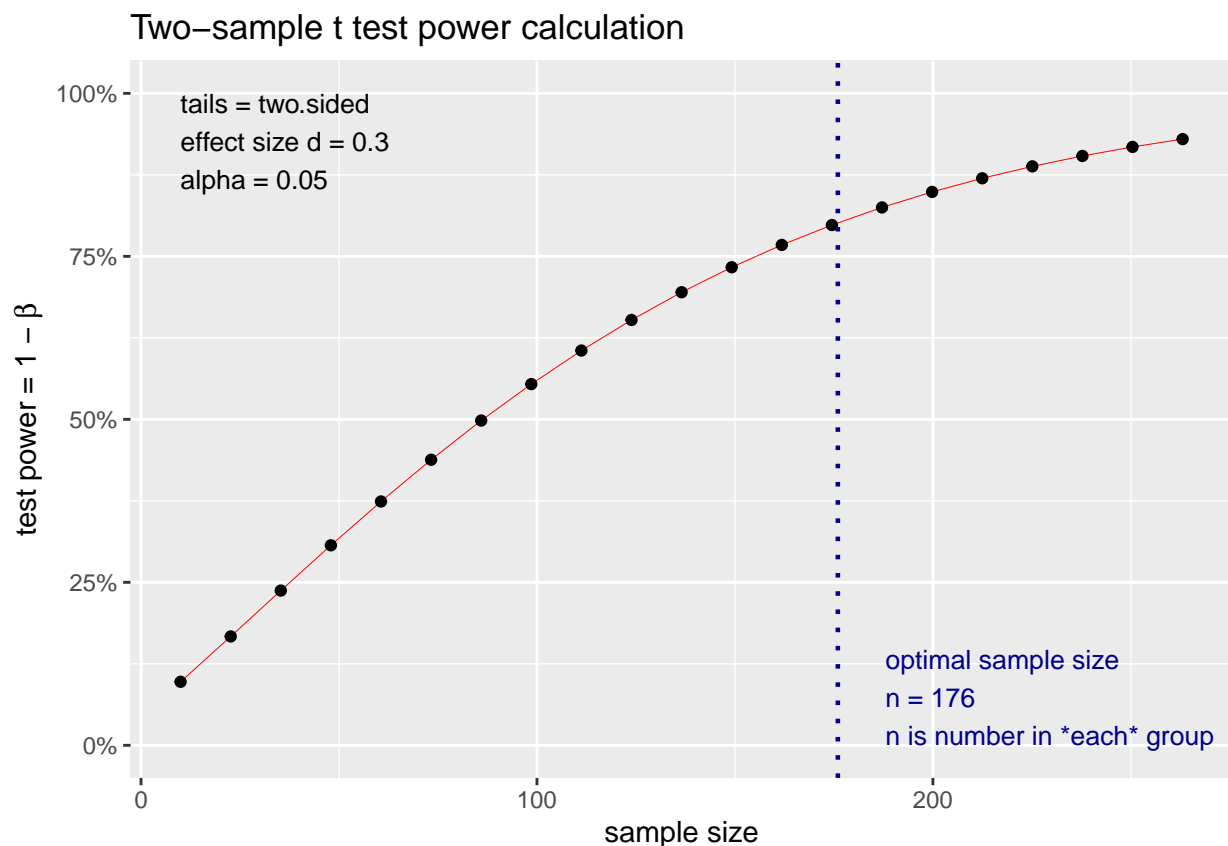
Next, lets take some of the guesswork and a lot of the ‘samba’ out of our power calculations by using vizualization tools to examine the tradeoff between sample size and power.

*example iv.*

Its easy to generate a plot of the power sample size tradeoff. First, store the results of the `pwr.t.test()` command in an object called ‘ex4’. Next, simply use the `plot.power.htest()` command to plot the output for example four.

```
# ex iv.
ex4 <- pwr.t.test(n = ,
                  d = 0.3 ,
                  sig.level = 0.05,
                  power = 0.8,
                  type = "two.sample")

plot.power.htest(ex4)
```



We can see that `plot.power` adds a vertical dotted line where the sample size line crosses the test power threshold of 80%, indicating that we would need a sample size of 176 in each arm of a 2-armed trial in order to detect an effect size of  $d = 0.3$ , two-sided with  $\alpha = 0.05$ .

Question 5. Again, use `pwr.t.test()`, but now calculate the effect size that could be detected. Use  $n = 50$  samples per arm, a significance level of 0.01, and power of 0.8.

# 5.

- a. What is the MDE?
- b. How well powered would we be ( $1 - \beta = \text{---}$ ) if we dropped the sample size from 50 to 40 samples per arm?

## Part III. Power and Sample Size for Cluster RCTs

Finally, let's have a look at cluster randomized controlled trials. Luckily, R has a prefabricated package for these power calculations. Sadly, visualization tools are a bit harder to come by. Let's explore the command for CRCTs with continuous outcomes: "crtpwr.2mean()". Note that commands are also available for "crtpwr.2prop()" (proportion) and "crtpwr.2rate()" (rate).

*example v.*

A few parameters here are worth note. (m) is the number of clusters needed. (n) is either the mean (if a number) or a vector (if we know the mean of each) of the clusters. (cv) is our coefficient of variation, (icc) is our intra-class correlation, and (varw) is the within-cluster correlation.

```
# ex v.
crtpwr.2mean(alpha = 0.05,          ## Significance level (.05 default)
              power = 0.8,          ## Power (.08 default)
              m = ,                 ## No. of clusters per condition (>1)
              n = 20,               ## Mean of cluster sizes (or vector)
              cv = 0,               ## Coefficient of variation
              d = .25,              ## Difference in condition means
              icc = .51,            ## Intra-Class Correlation
              varw = .3)            ## Within cluster variation

##          m
## 41.25688
```

In this example, we've set  $\alpha = 0.05$  and power to 0.8. We want to find the number of clusters needed (m) if we have an average of 20 samples per cluster ( $n=20$ ), a coefficient of variation of 0 (meaning all the clusters have the same size), a mean difference between treatment and control of  $d = 0.25$ , an ICC of .51, and within cluster variation of  $\text{varw} = 0.3$ .

We find that we need 42 clusters per condition (treatment vs. control), so 84 in total.

Question 6. Find the number of clusters per condition needed for a trial with  $\alpha = .05$ , power = 0.8, 5 observations per cluster, no variation in cluster size, a difference of 1 unit,  $\text{icc} = 0.1$  and a variance of five units.

# 6.

a. How many clusters per condition are required?