

How to do a Power Analysis

By Claire Willeck

In this document, I describe how I go through a power analysis using the power calculator from EGAP that was created by Alexander Coppock: https://egap.shinyapps.io/Power_Calculator/

The calculator allows for three randomized control trial (RCT) designs research designs: 1) non-binary outcome variable 2) binary outcome variable 3) clustered RCT.

Non-Binary Outcome Variable

For a continuous outcome variable, there's two parameters that you need to approximate in the power calculation if you use the standard $\alpha=0.05$ and $\beta=0.8$. These two parameters are the treatment effect size and the standard deviation of the outcome variable. For political psychology and behavior experiments, you can use a 0.1 effect size as a substantive effect to glean from a power analysis for a normalized outcome variable (outcome variable has a range of 0 to 1). If the effect size is less than 0.1, then you'll need a larger sample size to detect it, but the treatment would have a very small effect.

Usually the more difficult parameter to approximate is the standard deviation. Small changes in the standard deviation can lead to large changes in the sample sizes needed to have sufficient power to detect the treatment effect. If you have pilot data from your experiment, you should use the standard deviation of your outcome variable from the pilot data, and you can continue to use the 0.1 treatment effect size or use the treatment effect size that you calculated from your pilot data. But often we need to do a power analysis before we can even collect pilot data. If you're in this position, then it is good to look at papers that have used a similar outcome variable as you to see the standard deviation estimated in those papers. Ultimately, it is worthwhile to use multiple standard deviation numbers to get a range of the sample size needed.

Below I compare the sample size needed to detect a treatment effect size of 0.1 with a significance level of $\alpha=0.05$ and a power target of $\beta=0.8$ if the standard deviation is 0.2 or 0.3. As you can see, the sample size with a standard deviation of 0.2 is 126 and the sample size with a standard deviation of 0.3 is 283.

Power Calculator

This calculator can help you understand the power of your experimental design to detect treatment effects. You can choose between a standard design in which individuals are randomly assigned to treatment or control and a clustered design, in which groups of individuals are assigned to treatment and control together.

- Clustered Design?
- Binary Dependent Variable?

Significance Level

Alpha = 0.05

Treatment Effect Size

.1

Standard Deviation of Outcome Variable

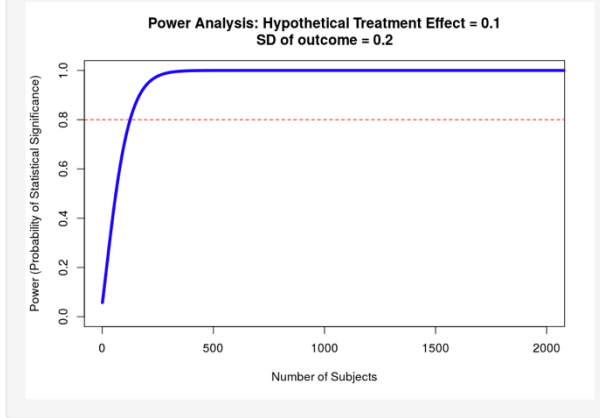
.2

Power Target

0.8

Maximum Number of Subjects

2000



In order to achieve 80% power, you'll need to use a sample size of at least 126.

Power Calculator

This calculator can help you understand the power of your experimental design to detect treatment effects. You can choose between a standard design in which individuals are randomly assigned to treatment or control and a clustered design, in which groups of individuals are assigned to treatment and control together.

- Clustered Design?
- Binary Dependent Variable?

Significance Level

Alpha = 0.05

Treatment Effect Size

.1

Standard Deviation of Outcome Variable

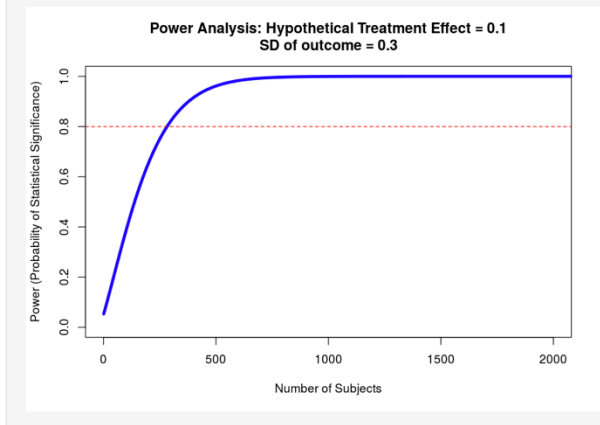
.3

Power Target

0.8

Maximum Number of Subjects

2000



In order to achieve 80% power, you'll need to use a sample size of at least 283.

Binary Outcome Variable

If your RCT uses a binary outcome variable, then the power calculation may be a bit easier. You need to approximate the mean of your outcome variable for the treatment group and the mean of your outcome variable for your control group. Once again, you can typically assume an effect size of 0.1 for political psychology and behavior experiments meaning the difference between the mean of your outcome for the treatment group and the mean of your outcome for the control group is 0.1. If you have pilot data, then you can use your data to estimate the mean of your outcome for the control group and either make the mean of the treatment group 0.1 greater or use the treatment effect size that you calculated in the pilot. If you do not have pilot data, then you should find published papers that use a similar outcome variable and use the mean that they calculated for the mean of your outcome variable for the control group. The closer the mean of your control group is to 0.5 and the smaller your treatment effect size is, the larger the sample size you will need.

Below I compare the sample size needed to detect a treatment effect size of 0.1 with a significance level of $\alpha=0.05$ and a power target of $\beta=0.8$ if the mean of the control group is 0.5 or 0.8. As you can see, the sample size needed when the mean of the control group is 0.5 is 770 and the sample size needed when the mean of the control group is 0.8 is 393.

Power Calculator

This calculator can help you understand the power of your experimental design to detect treatment effects. You can choose between a standard design in which individuals are randomly assigned to treatment or control and a clustered design, in which groups of individuals are assigned to treatment and control together.

- Clustered Design?
 Binary Dependent Variable?

Significance Level

Alpha = 0.05

Proportion (DV = 1) in Control Group

0.5

Proportion (DV = 1) in Treatment Group

0.6

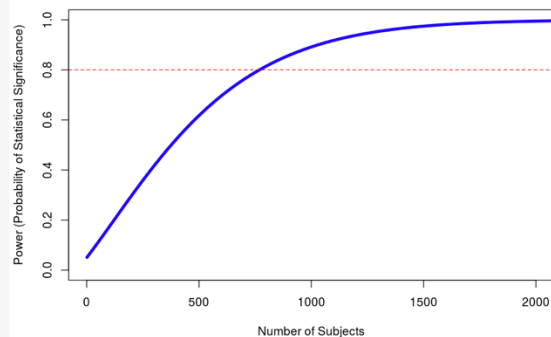
Power Target

0 0.8 1
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

Maximum Number of Subjects

2000

Power Analysis: Hypothetical Treatment Effect = 0.1 Percentage Points



In order to achieve 80% power, you'll need to use a sample size of at least 770.

Power Calculator

This calculator can help you understand the power of your experimental design to detect treatment effects. You can choose between a standard design in which individuals are randomly assigned to treatment or control and a clustered design, in which groups of individuals are assigned to treatment and control together.

- Clustered Design?
 Binary Dependent Variable?

Significance Level

Alpha = 0.05

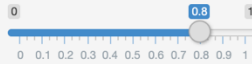
Proportion (DV = 1) in Control Group

0.8

Proportion (DV = 1) in Treatment Group

0.9

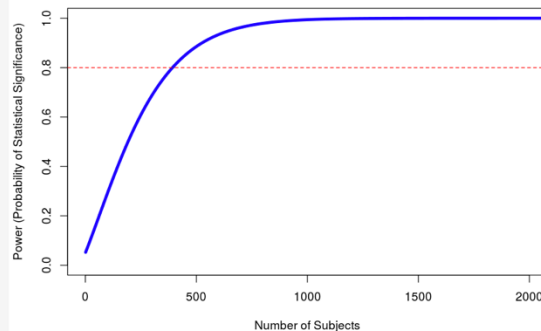
Power Target



Maximum Number of Subjects

2000

Power Analysis: Hypothetical Treatment Effect = 0.1 Percentage Points



In order to achieve 80% power, you'll need to use a sample size of at least 393.

Clustered Design

In a clustered RCT design, your randomization occurs at a higher level than your units of analysis. As an example, a clustered RCT would be when you randomly assign your treatment to classes and survey students to obtain your outcome variable where students are clustered within classes. Clustered RCT designs require a balance between the number of clusters (classes) and the number of units (students). This balance depends on how correlated units are within a cluster referred to as Intra-Cluster Correlation (ICC). If units are extremely correlated within a cluster, then the effective sample size becomes the number of clusters. If units are not very correlated within a cluster, then the effective sample size is closer to the number of units. This is because if ICC is equal to zero, then the units within clusters act no differently based on their cluster. If ICC is equal to one, then the units within clusters act the same as others in their cluster. As ICC increases, then the sample size needed to detect an effect size increases.

The ICC is difficult to estimate. If you have pilot data, then you can estimate the ICC. If you want to do this in R, you can use the `ICC()` function in the “psych” package. If you do not have pilot data, then you will need to approximate the ICC using papers that have a similar clustered design and outcome measures. Depending on your experiment, this might be challenging to find. Worst case scenario, you can calculate a range of sample sizes needed based on a range of ICC.

Once again, we can assume an $\alpha=0.05$ and $\beta=0.8$. You can use a clustered design with either a non-binary or binary outcome using the process described above. You can then set the number of clusters depending on your experiment and calculate the sample size needed based on your approximated ICC. Sometimes, if your ICC is high and number of clusters are too small, the power calculator will let you know that you need to increase the number of clusters.

Below, I use a non-binary outcome variable and an $\alpha=0.05$, $\beta=0.8$, treatment effect size of 0.1, standard deviation of 0.2, and number of clusters per arm of 40 (meaning 80 clusters total). I vary the ICC to be 0.2, 0.5, and 0.7. As you can see, the sample size needed when ICC is 0.2 is 152. The sample size needed when ICC is 0.5 is 331. When ICC is 0.7, the sample size required exceeds 10,000 which means the number of clusters per arm needs to increase.

Power Calculator

This calculator can help you understand the power of your experimental design to detect treatment effects. You can choose between a standard design in which individuals are randomly assigned to treatment or control and a clustered design, in which groups of individuals are assigned to treatment and control together.

- Clustered Design?
 Binary Dependent Variable?

Significance Level

Alpha = 0.05

Treatment Effect Size

.1

Standard Deviation of Outcome Variable

.2

Intra-cluster Correlation

0 0.2 1

Number of Clusters per Arm

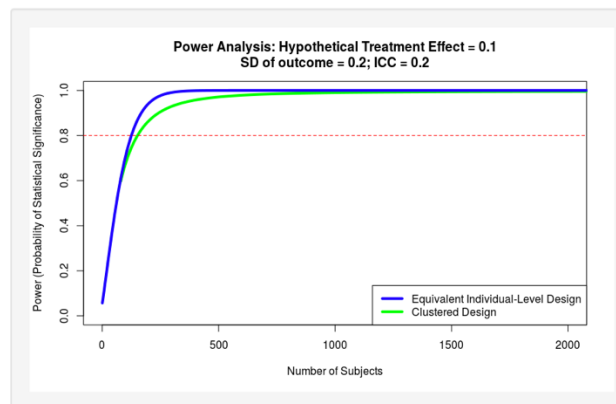
40

Power Target

0 0.8 1

Maximum Number of Subjects

2000



In order to achieve 80% power, you'll need to use a sample size of at least 152, or an average of at least 2 subjects in each of 80 clusters. Right-click to download image.

Power Calculator

This calculator can help you understand the power of your experimental design to detect treatment effects. You can choose between a standard design in which individuals are randomly assigned to treatment or control and a clustered design, in which groups of individuals are assigned to treatment and control together.

- Clustered Design?
- Binary Dependent Variable?

Significance Level
Alpha = 0.05

Treatment Effect Size
.1

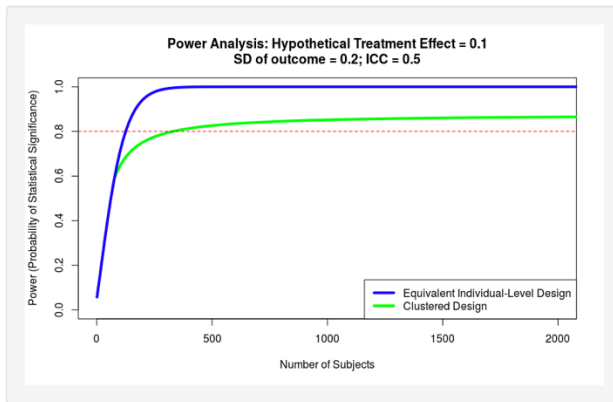
Standard Deviation of Outcome Variable
.2

Intra-cluster Correlation
0 0.5 1

Number of Clusters per Arm
40

Power Target
0 0.8 1

Maximum Number of Subjects
2000



In order to achieve 80% power, you'll need to use a sample size of at least 331, or an average of at least 4 subjects in each of 80 clusters. Right-click to download image.

Power Calculator

This calculator can help you understand the power of your experimental design to detect treatment effects. You can choose between a standard design in which individuals are randomly assigned to treatment or control and a clustered design, in which groups of individuals are assigned to treatment and control together.

Clustered Design?

Binary Dependent Variable?

Significance Level

Alpha = 0.05

Treatment Effect Size

.1

Standard Deviation of Outcome Variable

.2

Intra-cluster Correlation

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

Number of Clusters per Arm

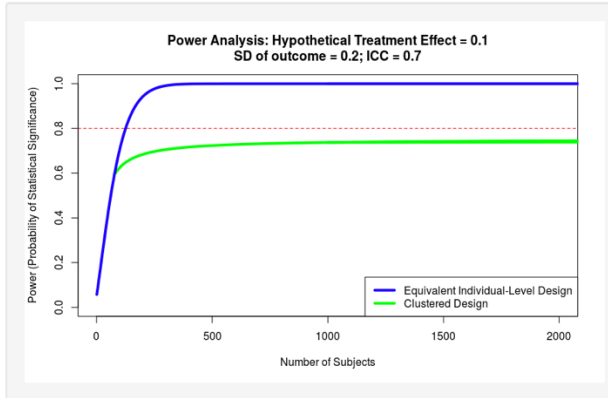
40

Power Target

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

Maximum Number of Subjects

2000



In order to achieve 80% power, you'll need to use a sample size of well more than 10,000,000. You may need to increase the number of clusters.