

PRACTICE

Power Calculator Exercise

INTRODUCTION

This exercise accompanies the [Six Rules of Thumb for Determining Sample Size and Statistical Power](#), and uses a fictional example of a housing program that provides monthly stipends that households can use to pay for rent. In the real world, this could mimic a voucher program, or a conditional cash transfer program. Imagine that the program will be evaluated using a randomized evaluation, where one group of participants will receive the treatment (i.e., the funding) and another will not (the comparison group). One outcome of interest could be the number of days a family is stably housed, or the number of days we expect a family can continue to pay their rent. All numbers used throughout this exercise are fictional and not meant to reflect the necessary targets for a true evaluation.

Note: The [answer key](#) is at the end of this document.

SET UP

Navigate to the J-PAL Power Calculator: <https://j-pal.shinyapps.io/power-app/>

To start using the power calculator, please enter the following information:

- *Minimum detectable effect (MDE):* 14 (days)
- *Standard deviation of outcome:* 30 (days)

The **MDE** is the smallest outcome that can be measured with an evaluation with the stated parameters. Practically, researchers should set an MDE that is plausible and has policy relevance. In the housing example, an MDE of 14 means that the researchers believe the cash program

could allow the family to stay housed or pay their rent for at least 14 days more than those who don't receive the program, and that if the program changes outcomes by fewer than 14 days, this evaluation would not be able to detect that smaller impact.

The **standard deviation** is a measure of variability. Researchers, for example, know that families have different levels of housing stability or number of days of ability to pay rent. The standard deviation is a measure of how much this varies across our population of interest at the outset of the evaluation. The larger the standard deviation, the larger sample you are likely to need. Researchers can measure the standard deviation of their own data or may estimate it proactively using previous literature. In this exercise, 30 days is a hypothetical example for illustrative purposes.

Check that the following information is entered as follows (these should be the default values, change to match if not):

- *Significance level:* 0.05
- *Power:* 0.8
- *Take up in treatment:* 100%
- *Take up in control:* 0%
- *Attrition:* 0%
- *Proportion in treatment:* 0.5

Note: Before each exercise, reset all values to these unless otherwise indicated.

GENERAL KNOWLEDGE CHECK

1. Based on the parameters above, what is the required *sample size* of this evaluation?

RULE OF THUMB 1

A larger sample size increases statistical power

Most studies aim for a power of 0.8. Using this calculator, you can not change the sample size. However, you can change the level of power to see how the required sample size changes. Researchers could use a tool like this to input what they know to get a sense of how large of a sample they might need for their evaluation.

1. What is the required *sample size* if you decrease *power* to 0.7?

2. What about if you increase the *power* to 0.9?

RULE OF THUMB 2

A program with a small effect size requires a larger sample size

Understanding how large or small an effect size is likely to be helps researchers understand what sample size they need to achieve a given level of power. Researchers run power calculations before beginning an intervention. Therefore, they may be able to change the intervention to increase its likely effect (for example, you could consider increasing the monthly amount of your housing program to increase its likely effect size).

1. What happens to the *sample size* if you change the *minimum detectable effect* size from 14 days to 30 days?

2. Conversely, what happens to the *sample size* if the *minimum detectable effect* is decreased to 2 days (representing a small effect size)?

Note the graph to the right of the calculator showing the inverse relationship between MDE and sample size. As your sample size increases, your MDE gets smaller, meaning you can measure an increasingly small effect. Researchers can use prior literature to get a sense of what size of an effect they need to be able to measure, and then use power tools such as this to understand how large of a sample they need.

RULE OF THUMB 3

A program with low take-up requires a larger sample size

In many research projects, not everyone who is offered a program or tool takes it up. This can affect the measured effect size on the entire population, and therefore the needed sample size.

1. What happens to the *sample size needed* if, say, only 80% of those offered this monthly payment accept or successfully enroll in the program (enter 80% into *take up treatment*)?

2. What happens to the *sample size* if even fewer, say 70% of those offered this monthly payment take up the program (enter 70% into *take up treatment*)?

RULE OF THUMB 4

A population with more variation requires a larger sample size

Understanding the natural variation in your outcome of interest of your sample or population is important. This is measured by the standard deviation. Generally, the more your sample or population varies, the larger sample you will need.

1. What happens to the *required sample size* if your *standard deviation* increases to 45 days?

2. What happens to the *required sample size* if your *standard deviation* decreases to only 7 days?

Researchers might expect a smaller variation in outcomes if the program is offered to a certain population type. The more similar a group is, the more likely their outcomes are to be similar. For example, the standard deviation might be smaller if the housing program is offered only to families exiting a rapid rehousing program, and larger if it is offered to any family who is currently homeless or at risk of homelessness.

RULE OF THUMB 5

Equal proportions in treatment and control groups usually maximizes power

Researchers can vary the proportion (or percentage) of the sample that receives the treatment and the proportion assigned to the comparison group. However, these changes do affect power.

1. If you decrease the *proportion of sample in treatment* to 0.3, what does the *power* have to be set at to keep your original sample size of 146 households? Try increasing or decreasing the *power* to get back to the original sample size. You can change numbers by 0.001.

2. How about if you increase the *proportion of sample in treatment* to 0.7?

RULE OF THUMB 6

Randomizing at the cluster level reduces power, and power is largely driven by the number of clusters

A clustered design means you are randomizing at a different unit than you are analyzing the data. In all the previous examples, we were both assigning and measuring outcomes at the household level (non-clustered). A clustered design means we would need to:

- Randomize at a different level—maybe a neighborhood or zip-code level—to measure household-level outcomes, or
- Continue to randomize at the household level, but get outcome measures for each individual in a household/family.

For this part, we are not using the previous housing program example, as to keep the focus only on the change in direction and size of the parameters, and not complicate comparisons to the previous examples.

Please reset the power calculator by refreshing the page entirely so that *MDE* reads 0.25, *standard deviation* is 0.4, and *proportion in treatment* is 0.5.

1. What is the *required sample size*?

Now select the box that says *clustered design*.

2. What is the *required sample size* now?

The large increase in required sample size shows that at a given level of power, the required sample size is higher in a clustered design. Instead of only needing 82 units, we need 400 units keeping all other parameters the same, if we are grouping those units in larger clusters. Which means, for a given sample size, power will be reduced using a clustered design.

Set the number of *units within a cluster* to 17. Leave all other selections.

3. What is the *sample size* and *number of clusters*?

4. What is the *sample size* and *number of clusters* if you:
 - a. Increase the *number of units* to 20

- b. Increase the *number of units* to 25

- c. Decrease the *number of units* to 15

Change the cluster input to Number of Clusters.

5. What is the *total sample size* and *units within a cluster*?

6. Try increasing *the number of clusters* to 25. What are the new *cluster size* and *sample size*?

7. Try decreasing *the number of clusters* to 17. What are the new *cluster size* and *sample size*?

Changing the number of units in a cluster took a relatively long time (in either direction) to change the sample size needed. And within those movements, the number of clusters changed even less. Remember from Rule of Thumb 1, sample size and power move in the same direction. So a large change (increase or decrease) in the sample size correlates with a change (increase or decrease in the same direction) of power.

However, moving the number of clusters has a much more dramatic effect on the sample size (in both directions). In this case, increasing the number of clusters from 20 to 25 actually reduced the total sample from 340 to 200, and decreasing the number of clusters quickly raised your sample size dramatically. This highlights the Rule of Thumb 6, changing the number of clusters will usually more quickly affect power than changing the number of units in a cluster.

ANSWER KEY

General Knowledge Check

1. Based on the parameters above, what is the required *sample size* of this evaluation? (The required sample size is 146 households)

Rule of Thumb 1 - A larger sample size increases statistical power

1. What is the required *sample size* if you decrease *power* to 0.7? (The required sample size is 114 households)
2. What about if you increase the *power* to 0.9? (The required sample size is 194 households)

Rule of Thumb 2 - A program with a small effect size requires a larger sample size

1. What happens to the *sample size* if you change the *minimum detectable effect* size from 14 days to 30 days? (The sample size drops to 32 households)
2. Conversely, what happens to the *sample size* if the *minimum detectable effect* is decreased to 2 days (representing a small effect size)? (The sample size increases to 7,064 households)

Rule of Thumb 3 - A program with low take-up requires a larger sample size

1. What happens to the *sample size needed* if say only 80% of those offered this monthly payment accept or successfully enroll in the program (enter 80% into *take up treatment*)? (The sample size increases to 226 households)
2. What happens to the *sample size* if even fewer, say 70% of those offered this monthly payment take up the program (enter 70% into *take up treatment*)? (The sample size increases to 296 households)

Rule of Thumb 4 - A population with more variation requires a larger sample size

1. What happens to the *required sample size* if your *standard deviation* increases to 45 days? (The required sample size increases to 326 households)
2. What happens to the *required sample size* if your *standard deviation* decreases to only 7 days? (The required sample size decreases to 8 households)

Rule of Thumb 5 - Equal proportions in treatment and control usually maximizes power

1. If you decrease the *proportion of sample in treatment* to 0.3, what does the *power* have to be set at to keep your original sample size of 146 households? Try increasing or decreasing the *power* to get back to the original sample size. You can change numbers by 0.001. (The power has to be 0.73)
2. How about if you increase the *proportion of sample in treatment* to 0.7? (Same answer. The power has to be 0.73)

Rule of Thumb 6 - Randomizing at the cluster level reduces power, and power is largely driven by the number of clusters

1. What is the *required sample size*? (The required sample size is 82 units)
2. What is the *required sample size* now? (The required sample size is 400 units)
3. What is the *sample size* and *number of clusters*? (The sample size is 340 units using 20 clusters)
4. What is the *sample size* and *number of clusters* if you:
 - a. Increase the *number of units* to 20 (The sample size is 400 units using 20 clusters)
 - b. Increase the *number of units* to 25 (The sample size is 500 units using 20 clusters)
 - c. Decrease the *number of units* to 15 (The sample size is 330 units using 22 clusters)
5. What is the *total sample size* and *units within a cluster*? (The total sample size is 340 with 17 units in a cluster)
6. Try increasing the *number of clusters* to 25. What are the new *cluster size* and *sample size*? (The sample size is 200 using 25 clusters, with a cluster size of 8)
7. Try decreasing the *number of clusters* to 17. What are the new *cluster size* and *sample size*? (The sample size is 1,190 using 17 clusters, with a cluster size of 70)