ADS2 Practical 2.4 - Power analysis and sample size

Rob Young (based on Xianghua Li and Chaochen Wang)

23/02/2024

Work through this guide alone or in groups. Facilitators are here to help. The time it takes to complete this practical vary between individuals - this is OK. Do not worry if you do not finish within the sesstion.

Learning objectives

After completing this practical you will be able to:

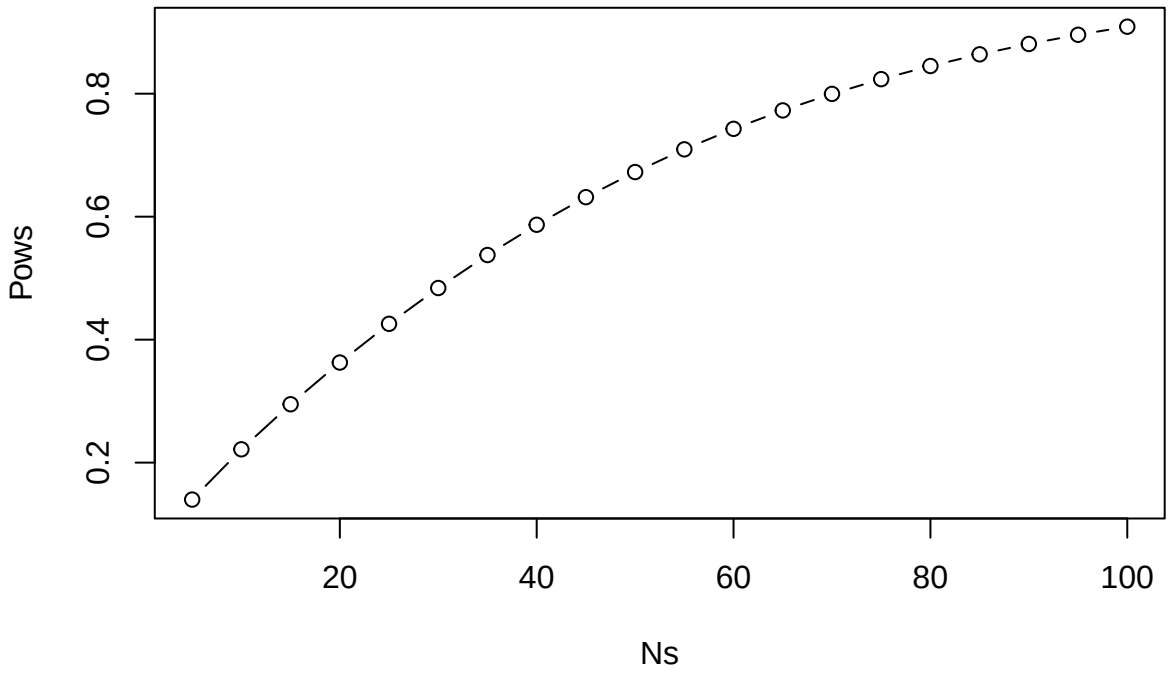
- Understand the intuituion behind power calculations
- Reveal the relationship among significance leve, power, effect size and sample size
- Know how to perform power/sample size analysis using simulations or with formulas in R

Task 1. Calculate a one-sample test power with similation in R.

Let us assume that the heights of the adult men in a country follow a normal distribution with average 175cm and standard derivation 10cm. A college is providing a free bottle of milk per day to the students. Based on the history of a nearby country, giving a bottle of milk to the college students could improve the average height from 175cm to 178cm. Now, you go to the college and measure the heights of 10 male students, and perform a t-test to see if the students are indeed higher than the average of the country. Simulate the data you get from the students, and record the p- value in the t.test, which is the number?

Now repeat the simulation for 10^5 times and look at all the p-values.

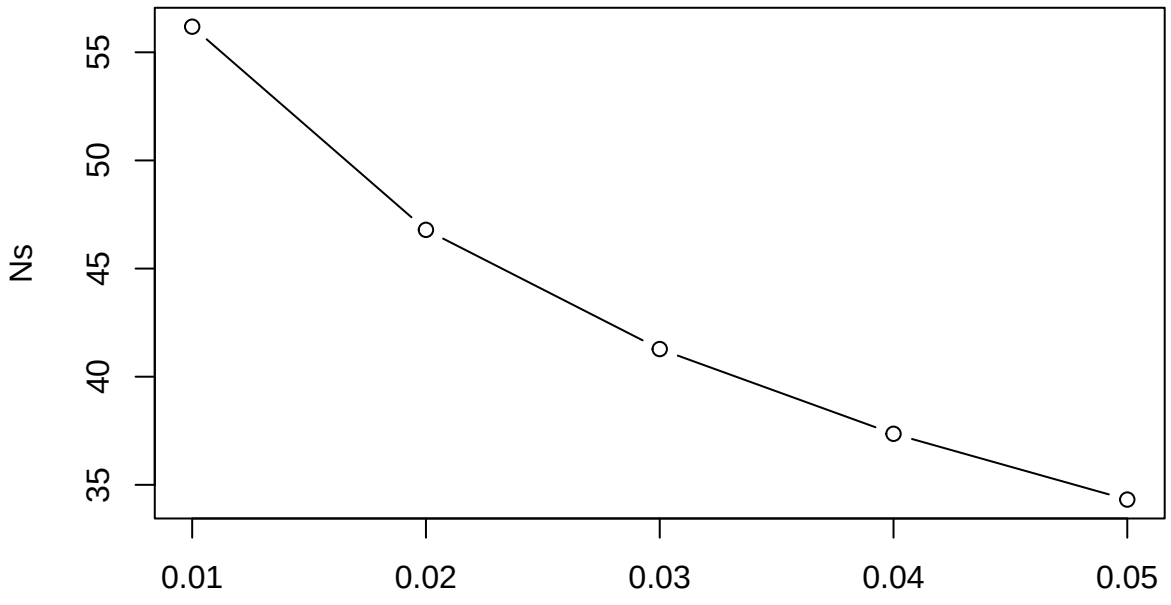
1. If you set a cutoff of 0.05, what is the percentage of simulations that you get the pvalue larger than 0.05? What does this number mean? What is the power?
2. If you measure 50 students instead of 10 students, then does the power change?
3. The cheating way is to use the command `power.t.test`. Try it out to calculate the power number with different Ns (5, 10, 15, 20. . . .100). Try to plot the power versus N. You should see a plot like this:

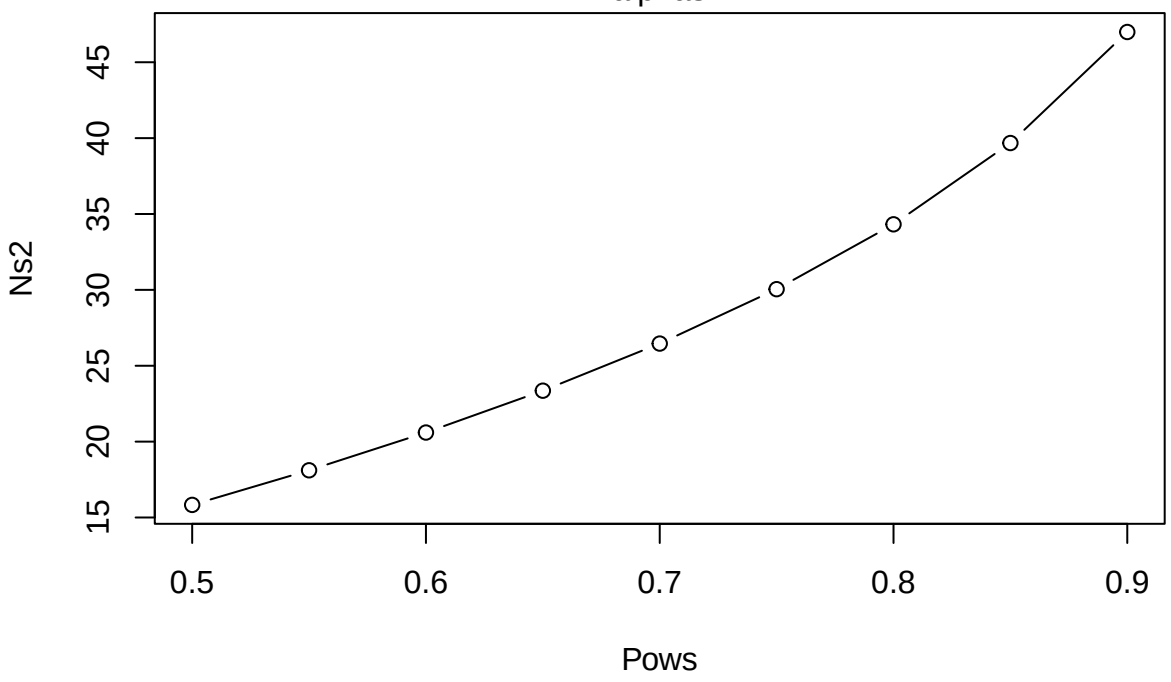


Task 2. Choose the right sample size

A company is developing a diet pill, which may help people lose weight. In the animal model, the drug could lead to 10% of weight loss. Now the company recruits 20 volunteers and separate them into two groups: placebo group and drug group to perform a trial. Let us assume that the weight in the normal population is 130 pound with standard derivation of 30.

1. If the drug is indeed effective as showed on the animal model, then what is the probability that they do not see a significant effect of the drug (p-value cutoff = 0.05)? Please perform a simulation as you did before to give an answer.
2. Do you think the power is good enough? If the company truly believes the effect of the drug and want to be sure that they will not largely miss the effect in the trial (type II error rate < 0.2), then how many volunteers do they need to recruit?
3. If the company changes their strategy, asking all the volunteers to take the pills and measuring their weights before and afterward, how many volunteers do they need?
4. Manipulate the p-value cutoff or power to see the change of the needed sample size. You could also plot them out. The output should be similar to these.


Ns
55
50
45
40
35
0.01
0.02
0.03
0.04
0.05
alphas


Ns2
45
40
35
30
25
20
15
0.5
0.6
0.7
0.8
0.9
Pows