

Week #14: Sample Size and Power Calculation

Cheng Peng

5/4/2021

Contents

1	Introduction	1
2	Sample size estimation	1
2.1	Sample size estimations for two proportions	2
2.2	Power analysis for two proportions	3
2.3	Sample size estimations for two averages	5
2.4	Power analysis for two averages	7
3	Paired-Sample Problems	10
3.1	Sample size estimation for paired data (before and after)	10
3.2	Power analysis of paired samples (paired t test)	12
4	Power analysis with unequal sample sizes	15
5	Conclusions	17

1 Introduction

Estimating the sample size for a prospective study is one of the first most important tasks in classical statistics. In clinical studies, researchers need to determine the number of subjects to enroll in a clinical trial to guarantee the reliability of the study results. If the sample size is too small, it will not be able to achieve the power of the analysis. If the size excessively larger than needed, it will waste resources.

Estimating the optimal number of patients is necessary for developing a Statistical Analysis Plan (SAP) in all prospective study. In retrospective studies, we can determine how much power of the analysis (for example, the power of detecting a difference if one existed).

This chapter focuses on how to

- 1) estimate sample size for a prospective study (e.g., randomized controlled trial)
- 2) determine how much power we have to detect a difference if one existed (post hoc analysis)

2 Sample size estimation

Before enrolling patients in a clinical trial, we want to make sure the we have the optimal number of patients for the study. Different statistical procedures require different optimal sample sizes to attain the target power. Next we discuss sample size determination for several commonly used statistical procedures.

Before move forward, let recall two error probabilities associated with a statistical test of hypothesis H_0 :

$$\alpha = P[\text{type I error}] = P[\text{reject } H_0 \mid H_0 \text{ is true}] = \text{false positive rate}$$

$\beta = \text{P}[\text{type II error}] = \text{P}[\text{fail to reject } H_0 \mid H_0 \text{ is false}] = \text{false negative.}$

$\text{power of a test} = 1 - \beta = \text{P}[\text{reject } H_0 \mid H_0 \text{ is false}]$

Several observations about α and β :

- α and β are negatively dependent on each other (not linearly dependent).
- Estimated α and β are dependent on the sample size.

In general, if more than one test for testing a given hypothesis H_0 under the same α , the one that achieves the lowest β is the best test. This tells the relationship among estimated α , β , power, and the sample size n .

2.1 Sample size estimations for two proportions

This is a common sample size calculation when you have two groups and the outcome is dichotomous (e.g., Yes / No) Assume that we have a study that aims to estimate the efficacy of Treatment A versus Treatment B. The main end point is a dichotomous variable: "Response" or "No response."

The null hypothesis is:

H_0 : There is no difference in the proportion of responders between Treatment A and Treatment B

The alternative hypothesis is:

H_a : There is a difference in the proportion of responders between Treatment A and Treatment B

To determine the sample size needed to detect a meaningful difference in the proportion of responders between Treatment A and Treatment B, we need the following information:

- **α level** (type I error) is the threshold where we make the determination whether something is statistically significantly different. We normally use an alpha level of 0.05 (type I error).
- **Effect size** is the standardized difference one would expect to see between treatment groups. This is an estimate and something that is usually based on past studies.
- **Power** ($1 - \beta$) is the ability to correctly reject the null hypothesis if the actual effect of the population is equal to or greater than the specified effect size.

We'll set the two-tailed alpha level to 0.05 estimate size h using the following equation:

$$h = \varphi_1 - \varphi_2$$

where $\varphi_i = 2 * \arcsine(\sqrt{p_i})$ and p_i denotes the proportion in treatment i that were classified as "responders." In generally, the power is set to 80%.

With these pieces of information, we can estimate the sample size for a study with two proportions as the outcome using the following formula:

$$n_i = \frac{(Z_{\alpha/2} + Z_{1-\beta})^2 + (p_1(1 - p_1)) + (p_2(1 - p_2))}{(p_1 - p_2)^2}$$

where n_i is the sample size for one group.

Use the `pwr.2p.test()` function in `library("pwr")` to find out optimal sample size that meet the requirements on the alpha level (`alpha = 0.05`), effect size h , and power level (`power = 80%`).

Example 1: We consider Treatment A and Treatment B in a survey study. Set Treatment B as the reference (e.g., control) with 50% response rate. Assume that Treatment A is slightly better with a response rate of

60%. With these pieces of information, we can estimate the sample size required to detect a difference in “response” rate between Treatment A and Treatment B that is 10% ($p_1 - p_2$) with an alpha of 5% and power of 80%.

```
### alpha = sig.level option and is equal to 0.05
### power = 0.80
### p1 = 0.60
### p2 = 0.50
power1 <-pwr.2p.test(h = ES.h(p1 = 0.60, p2 = 0.50), sig.level = 0.05, power = .80)
power1

##
##      Difference of proportion power calculation for binomial distribution (arcsine transformation)
##
##              h = 0.2013579
##              n = 387.1677
##      sig.level = 0.05
##      power = 0.8
##      alternative = two.sided
##
## NOTE: same sample sizes
```

The effect size h is 0.201, and the sample n is 387.2 or 388 rounded to the nearest whole number. This n is only for one group. Hence, based on the parameters of our study, we need approximately 388 patients in Treatment A and 388 patients in Treatment B to detect a difference of 10% response or greater with an alpha of 0.05 and power of 80%.

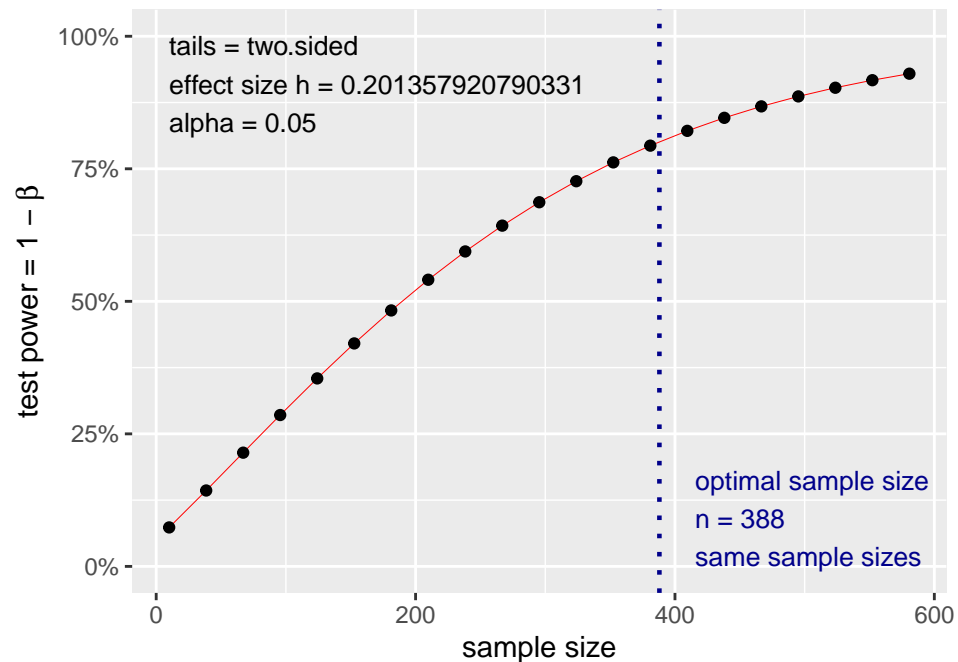
2.2 Power analysis for two proportions

Power ($1 - \beta$) is the ability to correctly reject the null hypothesis if the actual effect of the population is equal to or greater than the specified effect size. In other words, if you conclude that there is no difference in the sample, then there is no true difference in the population conditioned on the specified effect size

We can also use the plot feature to see how the power level changes with varying sample sizes. As the sample size goes up, power increases. As the sample size goes down, power decreases. This is important to understand. As we increase our sample size, we reduce the uncertainty around the estimates. By reducing this uncertainty, we gain greater precision in our estimates, which results in greater confidence in our ability to avoid making a type II error.

```
### We can plot the power relative to different levels of the sample size.
plot(power1)
```

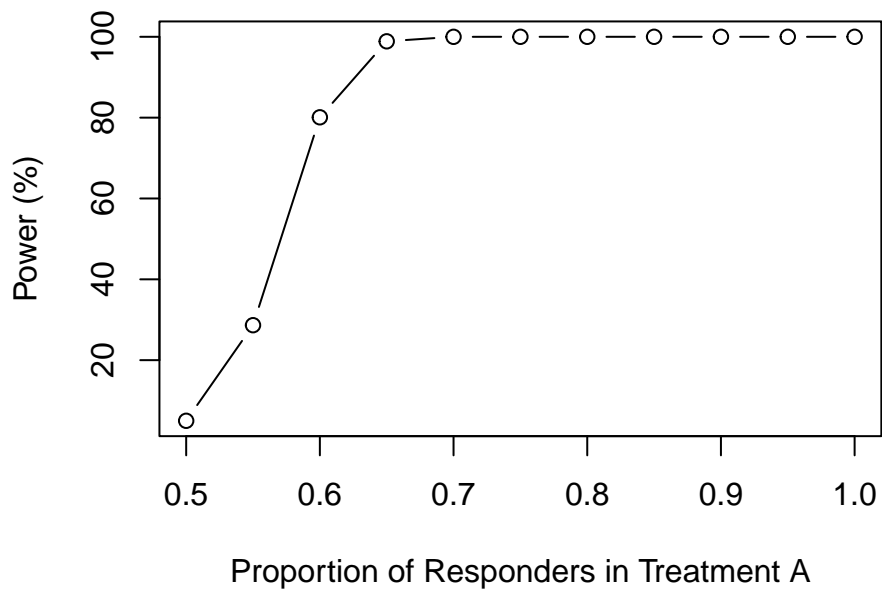
Difference of proportion power calculation for binomial distribution (arcsine transformation)



Let's change the p_i and see how the power level change; we are fixing our sample size at 388 for each group with an alpha of 0.05.

We create a sequence of values by varying the proportion of “responders” for Treatment A. We will change these from 50% to 100% in intervals of 5%.

```
p1 <- seq(0.5, 1.0, 0.05)
power1 <- pwr.2p.test(h = ES.h(p1 = p1, p2 = 0.50),
                     n = 388,
                     sig.level = 0.05)
powerchange <- data.frame(p1, power = power1$power * 100)
plot(powerchange$p1,
     powerchange$power,
     type = "b",
     xlab = "Proportion of Responders in Treatment A",
     ylab = "Power (%)")
```



We can also write a function for this:

```
iteration <- function(p_i, P_i, i_i, p2, n, alpha) {
  p1 <- seq(p_i, P_i, i_i)
  power1 <- pwr.2p.test(h = ES.h(p1 = p1, p2 = p2),
    n = n,
    sig.level = alpha)
  powerchange <- data.frame(p1, power = power1$power * 100)
  powerchange
}

iteration(0.5, 1.0, 0.05, 0.50, 388, 0.05)
```

```
##      p1      power
## 1  0.50   5.00000
## 2  0.55  28.65038
## 3  0.60  80.08415
## 4  0.65  98.88117
## 5  0.70  99.99190
## 6  0.75 100.00000
## 7  0.80 100.00000
## 8  0.85 100.00000
## 9  0.90 100.00000
## 10 0.95 100.00000
## 11 1.00 100.00000
```

2.3 Sample size estimations for two averages

This is another common sample size estimation for two groups when the outcome is a continuous variable.

Now let's estimate the sample size for a study where we are comparing the averages between two groups.

Let's suppose that we are working on a randomized controlled trial that seeks to evaluate the difference in the average change in hemoglobin A1c (HbA1c) from baseline between Treatment A and Treatment B.

The null hypothesis is:

H_0 : There is no difference in the average change in HbA1c from baseline between Treatment A and Treatment B

The alternative hypothesis is:

H_a : There is a difference in the average change in HbA1c from baseline between Treatment A and Treatment B

To determine the sample size needed to detect a meaningful difference in the average HbA1c change from baseline between Treatment A and Treatment B, we can use the following formula:

$$n_i = \frac{2\sigma^2 * (Z_{\alpha/2} + Z_{1-\beta})^2}{(\mu_1 - \mu_2)^2}$$

where n_i is the sample size for one of the groups, μ_1 and μ_2 are the average changes in HbA1c from baseline for Treatment A and Treatment B, respectively.

As usual, we'll set the two-tailed alpha level to 0.05. The effect size, also known as Cohen's d , is estimated using the following equation:

$$d = \frac{\mu_1 - \mu_2}{\sigma_{pooled}}$$

The pooled standard deviation is estimated using the following formula:

$$\sigma_{pooled} = \sqrt{\frac{sd_1^2 + sd_2^2}{2}}$$

Once again, the R `pwr` package can make this task easy for us. However, we'll need to estimate the Cohen's d .

Example 2: Since we haven't started the study, we have to make some assumptions of about each treatment strategy's change in HbA1c. Let's assume that the expected average change in HbA1c from baseline for Treatment A was 1.5% with a standard deviation of 0.25%. Additionally, let's assume that the expected average change in HbA1c from baseline for Treatment B was 1.0% with a standard deviation of 0.20.

First, we'll calculate the pooled standard deviation (σ_{pooled}):

```
sd1 = 0.25
sd2 = 0.30
sd.pooled = sqrt((sd1^2 + sd2^2) / 2)
sd.pooled
```

```
## [1] 0.276134
```

Once we have the σ_{pooled} , we can estimate the Cohen's d :

```
mu1 = 1.5
mu2 = 1.0
d = (mu1 - mu2) / sd.pooled
d
```

```
## [1] 1.810715
```

The Cohen's d is 1.81, which is considered a large effect size.

Now, we can take advantage of the `pwr` package and estimate the sample size needed to detect a difference of 0.5% (1.5% - 1.0%) in the average HbA1c change from baseline between Treatment A and Treatment B with 80% power and a significance threshold of 5%.

```
### d = Cohen's d
### power = 0.80
### alpha = 0.05
n.i <- pwr.t.test(d = d, power = 0.80, sig.level = 0.05)
n.i
```

```
##
##      Two-sample t test power calculation
##
##              n = 5.921286
##              d = 1.810715
##      sig.level = 0.05
##      power = 0.8
##      alternative = two.sided
##
## NOTE: n is number in *each* group
```

Based on our study parameters, we need 6 patients in each group to detect a difference of 0.5% or greater with 80% power and a significance threshold of 5%.

2.4 Power analysis for two averages

Power ($1 - \beta$) is the ability to correctly reject the null hypothesis if the actual effect of the population is equal to or greater than the specified effect size.

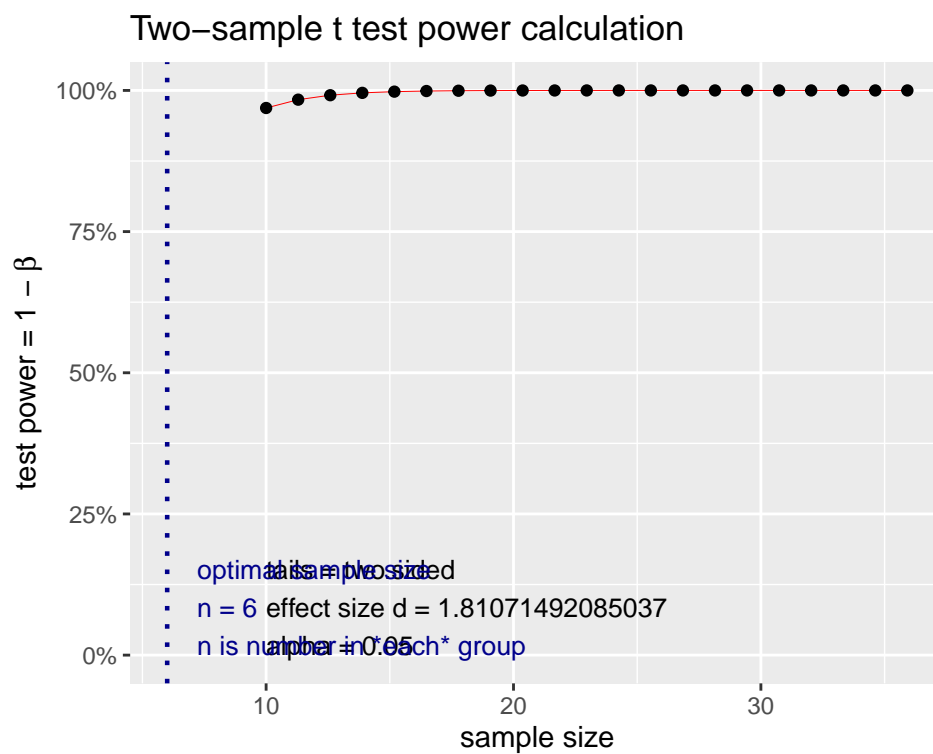
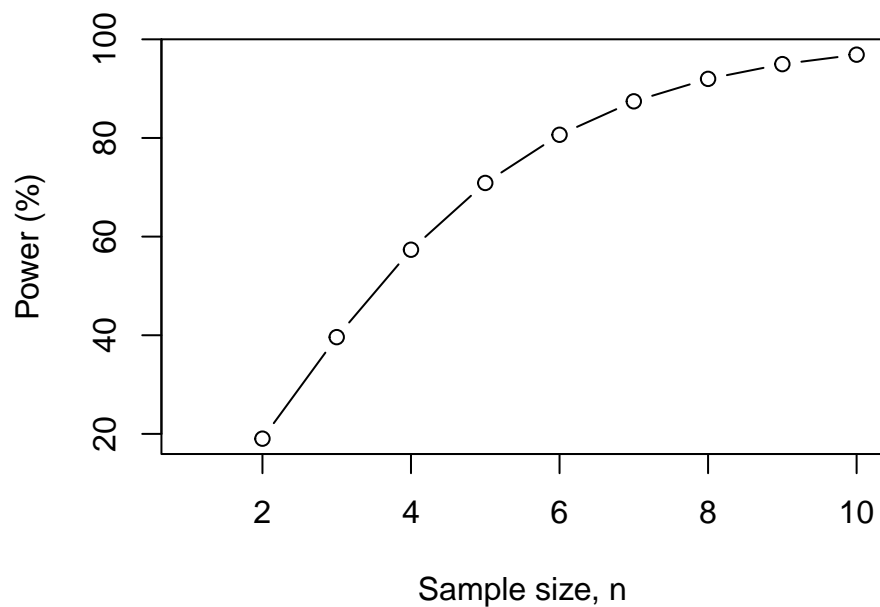
We can plot how the power will change as the sample size changes. As the sample size increases, power increases. This should make sense. Like our previous example with two proportions, as we increase our sample size, we reduce the uncertainty around the estimates. By reducing this uncertainty, we gain greater precision in our estimates, which results in greater confidence in our ability to avoid making a type II error.

```
### We can plot the power relative to different levels of the sample size.
n <- seq(1, 10, 1)
nchange <- pwr.t.test(d = d, n = n, sig.level = 0.05)
nchange.df <- data.frame(n, power = nchange$power * 100)
nchange.df
```

```
##      n      power
## 1      1      NaN
## 2      2 19.03307
## 3      3 39.61785
## 4      4 57.33850
## 5      5 70.87945
## 6      6 80.64997
## 7      7 87.42531
## 8      8 91.98145
## 9      9 94.96979
## 10    10 96.88938
```

```
plot(nchange.df$n,
     nchange.df$power,
     type = "b",
```

```
xlab = "Sample size, n",
ylab = "Power (%)")
```



As the sample size increases, our power increases. This makes sense because we have more patients to detect

differences that may be smaller. But we fixed our effect size (Cohen's d), so as we increase the sample size, our power to detect that difference ultimately increases.

Let's change the μ_i and see how the power level change; we are fixing our sample size at 6 for each group with an alpha of 0.05.

We create a sequence of values by varying the average change in HbA1c from baseline for Treatment A. We will change these from 0% to 2% in intervals of 0.1%.

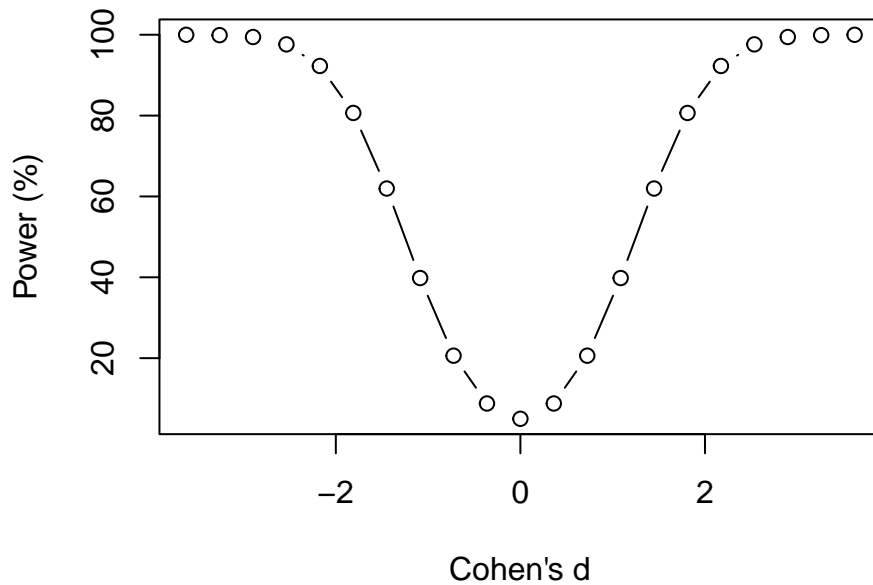
```
mu1 <- seq(0.0, 2.0, 0.1)

d <- (mu1 - mu2) / sd.pooled

power1 <- pwr.t.test(d = d, n = 6, sig.level = 0.05)
powerchange <- data.frame(d, power = power1$power * 100)
powerchange
```

```
##           d      power
## 1  -3.621430 99.986261
## 2  -3.259287 99.898978
## 3  -2.897144 99.437136
## 4  -2.535001 97.615372
## 5  -2.172858 92.271971
## 6  -1.810715 80.649971
## 7  -1.448572 61.960800
## 8  -1.086429 39.817661
## 9  -0.724286 20.610368
## 10 -0.362143  8.785855
## 11  0.000000  5.000000
## 12  0.362143  8.785855
## 13  0.724286 20.610368
## 14  1.086429 39.817661
## 15  1.448572 61.960800
## 16  1.810715 80.649971
## 17  2.172858 92.271971
## 18  2.535001 97.615372
## 19  2.897144 99.437136
## 20  3.259287 99.898978
## 21  3.621430 99.986261
```

```
plot(powerchange$d,
      powerchange$power,
      type = "b",
      xlab = "Cohen's d",
      ylab = "Power (%)")
```



This figure shows how the power changes with Cohen's d . It has a symmetrical pattern because of negative and positive range associated with Cohen's d . But the story is the same. As the effect size increases (negative and positive signs do not matter; we only care about the absolute values), the power increases. This makes sense because we only have enough power to detect large differences with the current sample size (which is fixed in this case). If the differences are small, then we do not have enough power with the current sample size of 6.

3 Paired-Sample Problems

The above section discussed two independent sample means and proportions. Next we discuss the case of two dependent samples.

3.1 Sample size estimation for paired data (before and after)

So far, we discussed how to perform sample size estimations for “between-groups” comparisons. However, many studies investigate the “within-group” changes. This are paired data, which means that the observation for one data point is dependent on another observation. A common study design where paired data is collected are longitudinal studies. These types of studies involve repeated measure. For example a pretest posttest study design will measure a data point for a patient at baseline and then repeat that measurement at another point in time. In the figure, Patient A has two measurements at t_0 and t_f . Since these measurements were made in the same person, the change is “within” the patient. Alternatively, we can think of this as a “repeated” measure since the patient had the measurement performed twice.

When you have a study where you are performing a paired t test, you can use the same `pwr.t.test()` function for a two sample test from the `pwr` package.

Let's assume that we want to conduct a prospective study measure the weight change of a cohort of patients who started a diet. You want to enroll enough patients to detect a 5 lb reduction in the weight 3 weeks after the diet started. Let's assume that at baseline the expected average weight for the cohort was 130 lbs with a standard deviation of 11. After 3 weeks of diet, the expected average weight was 125 lbs with a standard deviation of 12.



Figure 1: Figure caption: Pretest-Posttest (repeated measures) framework

We can estimate the effect size (Cohen's d_z) for a paired t test.

$$d_z = \frac{|\mu_z|}{\sigma_z} = \frac{|\mu_x - \mu_y|}{\sqrt{\sigma_x^2 + \sigma_y^2 - 2\rho_{x,y}\sigma_x\sigma_y}}$$

where x denotes “before” (or baseline), y denotes “after”, d_z denotes the Cohen's d for paired analysis ρ denotes the correlation between the measures before and after the diet. (For simplicity, I use 0.50 if I don't have prior information about this correlation.)

```
### Parameters for paired analysis or a pretest-posttest study design

mu_x <- 130    ### Average weight before the diet (baseline)
mu_y <- 125    ### Average weight after the diet

sd_x <- 11     ### Standard deviation before the diet
sd_y <- 12     ### Standard deviation after the diet

rho <- 0.5     ### Correlation between measures before and after the diet

sd_z <- sqrt(sd_x^2 + sd_y^2 - 2*rho*sd_x*sd_y)

d_z <- abs(mu_x - mu_y) / sd_z
d_z

## [1] 0.433555
```

The Cohen's d_z is 0.433. We can input this into the `pwr.t.test()` function.

```
n.paired <- pwr.t.test(d = d_z, power = 0.80, sig.level = 0.05, type = "paired")
n.paired

##
##      Paired t test power calculation
##
##              n = 43.71557
##              d = 0.433555
##      sig.level = 0.05
##      power = 0.8
```

```
##      alternative = two.sided
##
## NOTE: n is number of *pairs*
```

We need 44 patients with two measurements (before and after) they implement their diet to detect a difference of 5 lbs or greater with 80% power and a significance level of 0.05.

3.2 Power analysis of paired samples (paired t test)

We can plot how the power will change as the sample size changes for the paired t test analysis. As the sample size increases, power increases. This should make sense. Like our previous examples, as we increase our sample size, we reduce the uncertainty around the estimates. By reducing this uncertainty, we gain greater precision in our estimates, which results in greater confidence in our ability to avoid making a type II error.

```
### We can plot the power relative to different levels of the sample size for paired analysis.
```

```
n_z <- seq(1, 80, 5)
n_z.change <- pwr.t.test(d = d_z, n = n_z, sig.level = 0.05, type = "paired")
```

```
## Warning in qt(sig.level/tside, nu, lower = FALSE): NaNs produced
```

```
n_z.change.df <- data.frame(n_z, power = n_z.change$power * 100)
n_z.change.df
```

```
##      n_z      power
## 1      1      NaN
## 2      6 14.03624
## 3     11 25.58334
## 4     16 36.84309
## 5     21 47.26307
## 6     26 56.56985
## 7     31 64.66154
## 8     36 71.54769
## 9     41 77.30572
## 10    46 82.04980
## 11    51 85.90929
## 12    56 89.01478
## 13    61 91.48950
## 14    66 93.44465
## 15    71 94.97744
## 16    76 96.17076
```

```
plot(n_z.change.df$n,
     n_z.change.df$power,
     type = "b",
     xlab = "Sample size, n",
     ylab = "Power (%)")
```

As the sample size increases, we generate more power to detect a difference of 5 lbs with a significance level of 0.05 and a fixed sample size of 44 patients with two measurements (before and after) they implement their diet.

Let's see how power changes when we change the effect size. Let's change the average weight after the patients implement their diet. Instead of an average of 125 lbs, let's see how the power will change when we reduce that to 100 lbs.

```
### Vary the mu_y from 50 lbs to 130 lbs in intervals of 5 lbs.
```

```
mu_y <- seq(50, 130, 5)
```

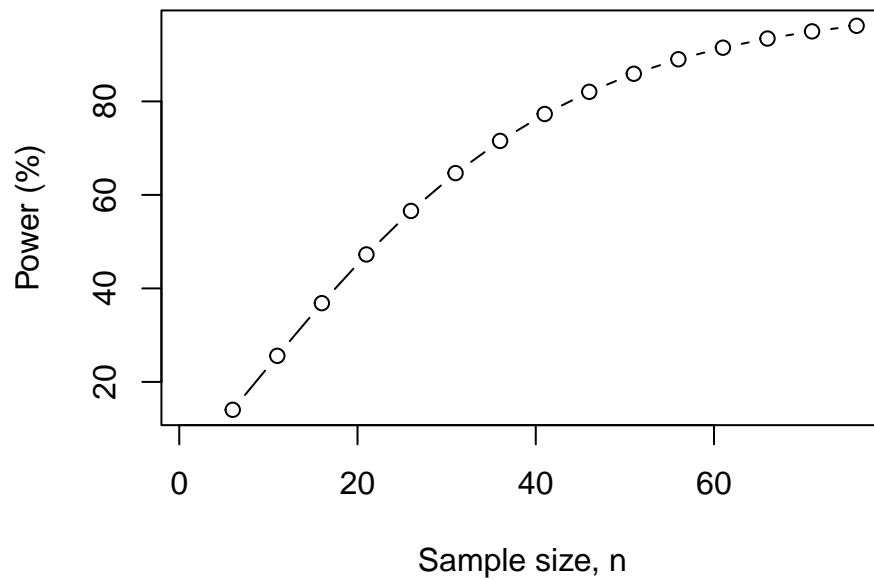


Figure 2: We increase the sample size from 1 to 80 at 5-unit intervals.

```
d_z <- abs(mu_x - mu_y) / sd_z

n_z.change <- pwr.t.test(d = d_z, n = 44, sig.level = 0.05)
n_z.change.df <- data.frame(d_z, power = n_z.change$power * 100)
n_z.change.df
```

```
##      d_z      power
## 1  6.936880 100.00000
## 2  6.503325 100.00000
## 3  6.069770 100.00000
## 4  5.636215 100.00000
## 5  5.202660 100.00000
## 6  4.769105 100.00000
## 7  4.335550 100.00000
## 8  3.901995 100.00000
## 9  3.468440 100.00000
## 10 3.034885 100.00000
## 11 2.601330 100.00000
## 12 2.167775 100.00000
## 13 1.734220 100.00000
## 14 1.300665  99.99766
## 15 0.867110  98.03639
## 16 0.433555  52.03146
## 17 0.000000   5.00000
```

```
plot(n_z.change.df$d_z,
     n_z.change.df$power,
     type = "b",
     xlab = "Cohen's d_z",
     ylab = "Power (%)",
     xlim = c(0, 2))
```

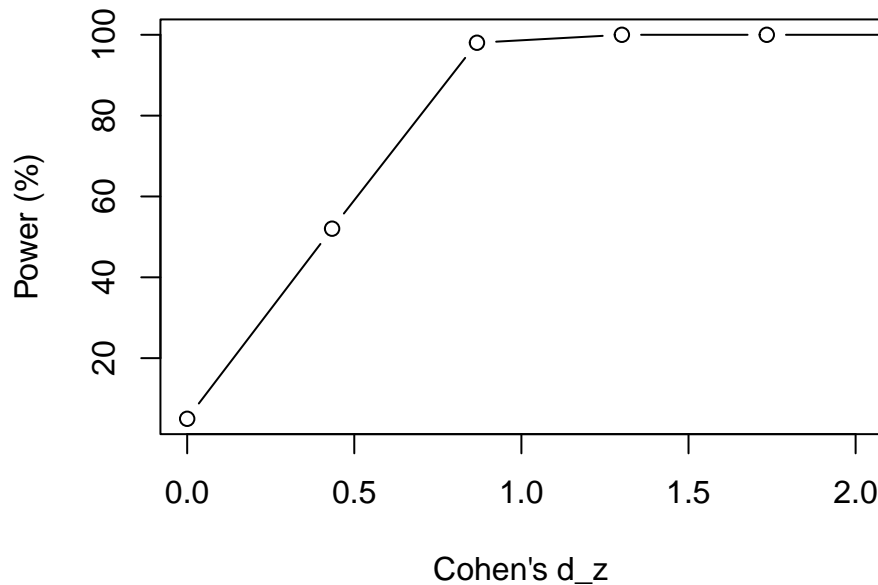


Figure 3: We vary the average weight μ_y between 50 lbs and 130 lbs in intervals of 5 lbs.

When we increase the effect size (Cohen's d_z), our power goes up; recall that the sample size is fixed at 44 and significance level is 0.05. But when the effect size gets smaller (or when the average weight loss shrinks), we lose power to detect a difference because our sample size is too small. We'll need to increase our sample size to have a reasonable power to detect small differences.

We can also see how power changes when we vary ρ . If we set $\rho = 0$, then the Cohen's $d_z = 0.307$. If we set $\rho = 1$, then the Cohen's $d_z = 5$.

```
mu_x <- 130      ### Average weight before the diet (baseline)
mu_y <- 125      ### Average weight after the diet

sd_x <- 11       ### Standard deviation before the diet
sd_y <- 12       ### Standard deviation after the diet

sd_z_1 <- sqrt(sd_x^2 + sd_y^2 - 2*1*sd_x*sd_y)
sd_z_0 <- sqrt(sd_x^2 + sd_y^2 - 2*0*sd_x*sd_y)

d_z_1 <- abs(mu_x - mu_y) / sd_z_1
d_z_0 <- abs(mu_x - mu_y) / sd_z_0
```

```
d_z_1
```

```
## [1] 5
```

```
d_z_0
```

```
## [1] 0.3071476
```

So, higher ρ results in large d_z and smaller ρ results in small d_z values.

Let's see how power changes when we change the ρ range from 0 to 1 in intervals of 0.1 units.

```
rho <- seq(0.0, 1.0, 0.1)
```

```
sd_z <- sqrt(sd_x^2 + sd_y^2 - 2*rho*sd_x*sd_y)
```

```
d_z <- abs(mu_x - mu_y) / sd_z
```

```
rho.change <- pwr.t.test(d = d_z, n = 44, sig.level = 0.05)
```

```
rho.change.df <- data.frame(d_z, power = rho.change$power * 100)
```

```
rho.change.df
```

```
##           d_z      power
## 1  0.3071476  29.65480
## 2  0.3236941  32.35129
## 3  0.3432395  35.66281
## 4  0.3668151  39.80733
## 5  0.3960280  45.10546
## 6  0.4335550  52.03146
## 7  0.4842743  61.25990
## 8  0.5583195  73.54735
## 9  0.6816774  88.52181
## 10 0.9552009  99.32401
## 11 5.0000000 100.00000
```

```
plot(rho.change.df$d_z,
     rho.change.df$power,
     type = "b",
     xlab = "Cohen's d_z",
     ylab = "Power (%)",
     xlim = c(0, 1.5))
```

As ρ increases, our power increases. This makes sense because we are nearing “perfect” correlation, which would require less sample to detect a difference if one existed. As the correlation becomes less “perfect” our power drops suggesting that we need to increase our sample size to make up for this poor correlation.

4 Power analysis with unequal sample sizes

It is common for the sample size to be different. The `pwr.t2n.test()` is a useful tool to help estimate the power given the sample sizes of the study.

It is common to perform power analysis on a study where the sample sizes between groups are different.

Suppose you have a retrospective study where the patients were prescribed Treatment A and Treatment B. There were 130 patients in Treatment A ($n_A = 130$) and 120 patients in Treatment B ($n_B = 120$). The average change in HbA1c was 1.5% with a standard deviation of 1.25% in Treatment A, and the average change in HbA1c was 1.4% with a standard deviation of 1.01% in Treatment B.

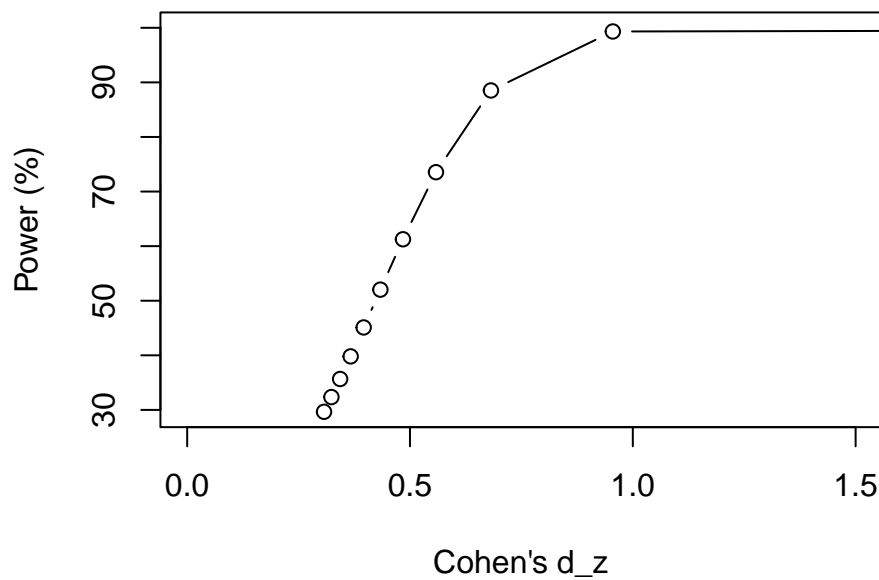


Figure 4: We vary ρ from 0 to 1 at intervals of 0.1 unit.

First, we'll calculate the pooled standard deviation (σ_{pooled}):

```
sd1 <- 1.25
sd2 <- 1.01
sd_pooled <- sqrt((sd1^2 + sd2^2) / 2)
sd_pooled
```

```
## [1] 1.136354
```

Once we have the σ_{pooled} , we can estimate the Cohen's d :

```
mu1 <- 1.5
mu2 <- 1.4
d <- (mu1 - mu2) / sd_pooled
d
```

```
## [1] 0.08800076
```

Now, we can estimate the power with the different sample sizes across the groups ($n_A = 130$, $n_B = 120$).

```
n1 <- 130
n2 <- 120

power.diff_n <- pwr.t2n.test(d = d, n1 = n1, n2 = n2, sig.level = 0.05)
power.diff_n
```

```
##
##      t test power calculation
##
##              n1 = 130
```



```
##          n2 = 120
##          d = 0.08800076
##      sig.level = 0.05
##          power = 0.1064836
##      alternative = two.sided
```

Since the average HbA1c change from baseline for Treatment A is 1.5% and 1.4% for Treatment B, the average difference in the HbA1c change from baseline is 0.1%. This is a difference (difference between the groups) of the differences (difference from baseline within the group) calculation.

You only have 11% power to detect a difference of 0.10% or greater in the HbA1c change from baseline. This means that you are underpowered to detect a difference of 0.10% or greater in the HbA1c change from baseline with $n_A = 130$, $n_B = 120$, and a significance level of 0.05. When studies are underpowered, there is a high potential for type II error. The only way to address this problem is to enroll more patients or expand the sample by relaxing inclusion criteria. But this may increase the threats to the study's internal validity.

5 Conclusions

Sample size estimations and power analysis are very useful tools to determine how many patients you need in your study and how confident you are that you didn't make a type II error. Depending on the type of study, you will need to use different functions from the **pwr** package. I highly encourage you to explore the other functions of the **pwr** package to see if those fit the study design you have planned.