IEHC0046 BASIC STATISTICS FOR MEDICAL SCIENCES

Analysis of Continuous Data III: Practical

08 November, 2020

In this practical we will use R to calculate statistical power and sample sizes for analyses. This practical is split into two sections.

1. Power calculations in R
2. Sample size calculations in R

Unlike previous practicals we do not require a dataset for this practical so there is no need to open the teaching dataset here. But, we will be using the pwr package. Let’s install and load this now.

install.packages("pwr")  
library(pwr)

Remember to make sure you save your code in a script!

# 1. Power calculations in R

## 1.1. Comparing Means

In this first section we are going to use R to calculate the power of analyses in various scenarios. We will firstly compare the characteristics of two random samples under different conditions – different alpha levels, sample sizes and unequal sizes of groups.

First we will compare the mean age in two samples; each with 2000 individuals, with alpha level of 0.05 and then alpha of 0.01. The mean age in sample 1 is 48.7 years (standard deviation = 17.6). The mean age in sample 2 is 47.0 years (standard deviation = 17.7).

**Q. Use the function pwr.t.test() to estimate the statistical power at both alpha levels. What can you say about these results?**

We use pwr.t.test() to estimate power when we have equal sample sizes in each group. pwr.t.test() takes two inputs: the effect size and the sample size in each group. To use the function we need to first calculate the effect size from the information above.

The effect size is the effect (i.e. 48.7 - 47 = 1.7) expressed in standard deviations. Because each group has different standard deviations (17.6 and 17.7), we need to calculate the *pooled* standard deviation. With equal sample sizes the formula for this is:

means <- c(48.7, 47)  
sds <- c(17.6, 17.7)  
n <- c(2000, 2000)  
  
pool\_sd <- sqrt(sum(sds^2)/2)  
  
d <- (means[1] - means[2])/pool\_sd

pwr.t.test(n = n, d = d)

##   
## Two-sample t test power calculation   
##   
## n = 2000, 2000  
## d = 0.09631689  
## sig.level = 0.05  
## power = 0.8610642, 0.8610642  
## alternative = two.sided  
##   
## NOTE: n is number in \*each\* group

By default pwr.t.test() assumes a significance (alpha) level of 0.05. We can see that the power of the analysis is 86.1%. This means there is 86% probability of finding a statistical difference at the 5% significance level in the two population means if a true difference exists.

Next we try altering the alpha level to 0.01:

pwr.t.test(n = n, d = d, sig.level = 0.01)

##   
## Two-sample t test power calculation   
##   
## n = 2000, 2000  
## d = 0.09631689  
## sig.level = 0.01  
## power = 0.6803632, 0.6803632  
## alternative = two.sided  
##   
## NOTE: n is number in \*each\* group

We can see that if we alter the significance level to 0.01 there is 68% power. That is, there is 68% power to detect a difference at this level if one exists. The power is lower than when the significance level was set at 0.05.

**Q. Repeat the previous exercise but with 1500 individuals in each sample and then with 2500 individuals in each at each level of alpha (0.05 and 0.01). How do the power levels change as the sample size decreases or increases?**

n <- c(1500, 1500)  
pwr.t.test(n = n, d = d)

##   
## Two-sample t test power calculation   
##   
## n = 1500, 1500  
## d = 0.09631689  
## sig.level = 0.05  
## power = 0.7507794, 0.7507794  
## alternative = two.sided  
##   
## NOTE: n is number in \*each\* group

We have 75% power to detect a difference at the 0.05 level if one exists with 1500 people in each group. This is lower power than when there were 2000 people in each group.

Next with same size groups but 0.01 alpha level:

pwr.t.test(n = n, d = d, sig.level = 0.01)

##   
## Two-sample t test power calculation   
##   
## n = 1500, 1500  
## d = 0.09631689  
## sig.level = 0.01  
## power = 0.5241047, 0.5241047  
## alternative = two.sided  
##   
## NOTE: n is number in \*each\* group

We observe something similar for alpha level of 0.01. The reduced sample size of 1500 people in each group reduces the power to 52.4%.

Next we try increasing the sample size to 2500 individuals in each group to see how this affects the power, first at alpha level of 0.05:

n <- c(2500, 2500)  
pwr.t.test(n = n, d = d)

##   
## Two-sample t test power calculation   
##   
## n = 2500, 2500  
## d = 0.09631689  
## sig.level = 0.05  
## power = 0.9257287, 0.9257287  
## alternative = two.sided  
##   
## NOTE: n is number in \*each\* group

The power is now very high at 92.6%. This means there is very high probability of detecting a difference in mean age at the 0.05 level should one exist. This has increased with a bigger sample. We also see an improvement in power at the 0.01 level when we increase the sample size:

pwr.t.test(n = n, d = d, sig.level = 0.01)

##   
## Two-sample t test power calculation   
##   
## n = 2500, 2500  
## d = 0.09631689  
## sig.level = 0.01  
## power = 0.7962658, 0.7962658  
## alternative = two.sided  
##   
## NOTE: n is number in \*each\* group

We will now try to see what happens to the power level when we change the group sizes so that they are no longer of equal size.

**Q. Calculate the study power with 1500 people in sample 1 and 2500 in sample 2; then with 500 people in sample 1 and 3500 in sample 2. Keep alpha at 0.01. It might be useful to first recalculate the power for 2000 people in each sample as you did in the first exercise for ease of reference.**

This time we have unequal sample sizes. For this we use the function pwr.t2n.test(). Again, we need to express the difference in means as an effect sizes. This requires calculating the pooled standard deviation. Because the sample sizes are unequal, the formula is a little more complicated:

We’ll need to recalculate the effect size more than once, so let’s create a new function get\_d() which calculates it for use.

get\_d <- function(n, means, sds){  
 pool\_sd <- sqrt(sum((n-1)\*(sds^2))/(sum(n)-2))   
 d <- (means[1] - means[2])/pool\_sd  
}

If you find the function difficult to read, try running the component parts (e.g. n-1\*(sds^2)) to see if it fits with your understanding.

As suggested in the question let’s first re-run the twomeans command with the original 2000 individuals in each group for reference. The get\_d() and pwr.t2n.test() functions will work if you have identical sample sizes.

n <- c(2000, 2000)  
d <- get\_d(n, means, sds)  
pwr.t2n.test(n1 = n[1], n2 = n[2], d = d, sig.level = 0.01)

##   
## t test power calculation   
##   
## n1 = 2000  
## n2 = 2000  
## d = 0.09631689  
## sig.level = 0.01  
## power = 0.6803632  
## alternative = two.sided

We now change this to 1500 in group 1 and 2500 in group 2:

n <- c(1500, 2500)  
d <- get\_d(n, means, sds)  
pwr.t2n.test(n1 = n[1], n2 = n[2], d = d, sig.level = 0.01)

##   
## t test power calculation   
##   
## n1 = 1500  
## n2 = 2500  
## d = 0.09624872  
## sig.level = 0.01  
## power = 0.6442906  
## alternative = two.sided

By changing the imbalance in groups slightly the power has reduced slightly to 64.6%. This imbalance makes it slightly less probable that we will detect a difference in mean age if it exists.

If we make the imbalance more pronounced by allocating 500 people in sample 1 and 3500 in sample 2 the power really reduces:

n <- c(500, 3500)  
d <- get\_d(n, means, sds)  
pwr.t2n.test(n1 = n[1], n2 = n[2], d = d, sig.level = 0.01)

##   
## t test power calculation   
##   
## n1 = 500  
## n2 = 3500  
## d = 0.0961128  
## sig.level = 0.01  
## power = 0.2855905  
## alternative = two.sided

In this case the power is really insufficient (28.7%) and we are far less likely to detect a difference in mean age between the two samples if one exists.

## 1.2. Comparing Proportions

We have so far been comparing means, but what if we had different starting information, such as proportions?

Suppose we have a random sample of individuals and we want to know if there is a differences in the prevalence of depression between men and women. We have the following 2x2 table from our study:

|  |  |  |  |
| --- | --- | --- | --- |
| Sex | Not Depressed | Depressed | Total |
| Male | 3,981 | 1,205 | 5,186 |
| Female | 4,288 | 1,917 | 6,205 |
| Total | 8,269 | 3,122 | 11,391 |

**Q. What is the proportion of men with depression? And that of women?**

We can calculate this by inputting the data into R and use some basic calculations.

men <- c(3981, 1205)  
women <- c(4288, 1917)  
  
men[2]/sum(men)

## [1] 0.2323563

women[2]/sum(women)

## [1] 0.3089444

23.2% of men in this sample have depression. 30.9% of women in this sample have depression. Let’s stores these results.

m\_dep <- men[2]/sum(men)  
w\_dep <- women[2]/sum(women)

We can now use these two proportions in the power calculation instead of the means we used in the previous exercises.

**Q. Perform a power test for the difference between the two proportions.**

We have unequal sample sample sizes so we need the pwr.2p2n.test() function. (You can use pwr.2p.test() if you have equal sample sizes.) We also need to calculate the effect size for the difference in proportions. We can do this using the ES.h() function which comes with the pwr packages.

pwr.2p2n.test(h = ES.h(m\_dep, w\_dep), n1 = sum(men), n2 = sum(women))

##   
## difference of proportion power calculation for binomial distribution (arcsine transformation)   
##   
## h = 0.1727681  
## n1 = 5186  
## n2 = 6205  
## sig.level = 0.05  
## power = 1  
## alternative = two.sided  
##   
## NOTE: different sample sizes

This output show that there is excellent power ~100% to detect a difference in these two proportions at 0.05 alpha if one exists.

# 2. Sample size calculations in R

## 2.1. Difference in means

In this second practical section we will work the other way around. We will instead work out the required sample sizes from a desired statistical power and other study characteristics but we don’t yet have the data, we are instead at the point of designing our study. We will again use the pwr package.

**Q. Calculate the required sample size for comparing the mean cholesterol of two groups. You have information from a pilot study that measured cholesterol in people following a special diet (mean cholesterol = 5.7, standard deviation = 1.3) and those following their normal diet (mean cholesterol = 5.9, standard deviation = 1.2). Use a power of 0.8 followed by a power of 0.9.**

Power, sample size and effect size are intimately related. Whatever we don’t provide the pwr functions will be calculated for us. If we know the sample size and effect size, we can calculate power. If we know the effect size and power, we can calculate the sample size. So now, we’ll be passing power and effect size into the pwr.t.test() function.

Firstly looking at a power of 80%, the function we use is as follows:

n <- c(1000, 1000)  
means <- c(5.7, 5.9)  
sds <- c(1.3, 1.2)  
  
d <- get\_d(n, means, sds)  
pwr.t.test(d = d, power = 0.8)

##   
## Two-sample t test power calculation   
##   
## n = 615.1352  
## d = 0.1598722  
## sig.level = 0.05  
## power = 0.8  
## alternative = two.sided  
##   
## NOTE: n is number in \*each\* group

The output shows that to have 80% power we would need to have at least 1232 people in our study, with 616 people in each diet group.

If we want greater statistical power (90%):

pwr.t.test(d = d, power = 0.9)

##   
## Two-sample t test power calculation   
##   
## n = 823.1671  
## d = 0.1598722  
## sig.level = 0.05  
## power = 0.9  
## alternative = two.sided  
##   
## NOTE: n is number in \*each\* group

**Q. Calculate the required sample size for the difference in mean BMI in people with diabetes and those without. From previous work you know that the standard deviation of BMI is 4.6. Complete the following table reporting the total number of people required and the number in each group.**

|  |  |  |  |
| --- | --- | --- | --- |
| Difference in BMI | Power = 0.8, alpha = 0.05 | Power = 0.8, alpha = 0.01 | Power = 0.9, alpha = 0.05 |
| 1 unit |  |  |  |
| 2 units |  |  |  |
| 4 units |  |  |  |

To get you started:

pwr.t.test(d = 1/4.6, power = 0.8, sig.level = 0.1)

##   
## Two-sample t test power calculation   
##   
## n = 262.3113  
## d = 0.2173913  
## sig.level = 0.1  
## power = 0.8  
## alternative = two.sided  
##   
## NOTE: n is number in \*each\* group

|  |  |  |  |
| --- | --- | --- | --- |
| Difference in BMI | Power = 0.8, alpha = 0.05 | Power = 0.8, alpha = 0.01 | Power = 0.9, alpha = 0.05 |
| 1 unit | 668 | 992 | 892 |
| 2 units | 170 | 252 | 226 |
| 4 units | 44 | 66 | 58 |

The bigger the difference in mean BMI between the two groups, the fewer the people needed in each group to detect this difference within each power and alpha level. To detect the difference in means at a lower alpha level, a bigger sample is needed. To specify a higher power level more people are also needed.

## 2.2. Difference in proportions

As with pwr.t.test(), we can use pwr.2p.test() to calculate requires sample sizes to achieve a set power.

**Q. Estimate the sample size for the difference in the proportion of men and women taking statins, from your pilot study you found that 12% of men are taking statins versus 10.5% of women. Use the power at 80% and 90% in your calculations**

pwr.2p.test(ES.h(0.105, 0.12), power = 0.8)

##   
## Difference of proportion power calculation for binomial distribution (arcsine transformation)   
##   
## h = 0.04749588  
## n = 6958.645  
## sig.level = 0.05  
## power = 0.8  
## alternative = two.sided  
##   
## NOTE: same sample sizes

To report a difference at the 5% level with power 80%, 6,959 men and women would be needed (total 13,918 people). If we specified a higher power level (90%), we would need even more people in our study (18,648, 9,324 each of men and women):

pwr.2p.test(ES.h(0.105, 0.12), power = 0.9)

##   
## Difference of proportion power calculation for binomial distribution (arcsine transformation)   
##   
## h = 0.04749588  
## n = 9315.671  
## sig.level = 0.05  
## power = 0.9  
## alternative = two.sided  
##   
## NOTE: same sample sizes

In order to detect this difference in proportions at the 5% level and with 90% power we need 1812 people in each group (total of at least 3624).

**Q. Estimate the sample size for the difference in the proportion of retired and employed men doing voluntary work. From previous studies you know that 34% of retired men are engaged in voluntary work while only 29% of employed men are engaged in voluntary work. You want to have power of your study 90%.**

pwr.2p.test(ES.h(0.34, 0.29), power = 0.9)

##   
## Difference of proportion power calculation for binomial distribution (arcsine transformation)   
##   
## h = 0.1077158  
## n = 1811.202  
## sig.level = 0.05  
## power = 0.9  
## alternative = two.sided  
##   
## NOTE: same sample sizes