R Data Analysis Examples
Power Analysis for Two-group Independent sample t-test

Examples

Example 1. A clinical dietician wants to compare two different diets, A and B, for diabetic patients. She hypothesizes that diet A (Group 1) will be better than diet B (Group 2), in terms of lower blood glucose. She plans to get a random sample of diabetic patients and randomly assign them to one of the two diets. At the end of the experiment, which lasts 6 weeks, a fasting blood glucose test will be conducted on each patient. She also expects that the average difference in blood glucose measure between the two group will be about 10 mg/dl. Furthermore, she also assumes the standard deviation of blood glucose distribution for diet A to be 15 and the standard deviation for diet B to be 17. The dietician wants to know the number of subjects needed in each group assuming equal sized groups.

Example 2. An audiologist wanted to study the effect of gender on the response time to a certain sound frequency. He suspected that men were better at detecting this type of sound then were women. He took a random sample of 20 male and 20 female subjects for this experiment. Each subject was be given a button to press when he/she heard the sound. The audiologist then measured the response time - the time between the sound was emitted and the time the button was pressed. Now, he wants to know what the statistical power is based on his total of 40 subjects to detect the gender difference.

Prelude to The Power Analysis

There are two different aspects of power analysis. One is to calculate the necessary sample size for a specified power as in Example 1. The other aspect is to calculate the power when given a specific sample size as in Example 2. Technically, power is the probability of rejecting the null hypothesis when the specific alternative hypothesis is true.

For the power analyses below, we are going to focus on Example 1, calculating the sample size for a given statistical power of testing the difference in the effect of diet A and diet B. Notice the assumptions that the dietician has made in order to perform the power analysis. Here is the information we have to know or have to assume in order to perform the power analysis:

- The expected difference in the average blood glucose; in this case it is set to 10.
- The standard deviations of blood glucose for Group 1 and Group 2; in this case, they are set to 15 and 17 respectively.
- The alpha level, or the Type I error rate, which is the probability of rejecting the null hypothesis when it is actually true. A common practice is to set it at the .05 level.
- The pre-specified level of statistical power for calculating the sample size; this will be set to .8.
- The pre-specified number of subjects for calculating the statistical power; this is the situation for Example 2.

Notice that in the first example, the dietician didn't specify the mean for each group, instead she only specified the difference of the two means. This is because that she is only interested in the difference, and it does not matter what the means are as long as the difference is the same.

Power Analysis

In R, it is fairly straightforward to perform power analysis for comparing means. For example, we can use the **pwr** package in R for our calculation as shown below. We first specify the two means, the mean for Group 1 (diet A) and the mean for Group 2 (diet B). Since what really matters is the difference, instead of means for each group, we can enter a mean of zero for Group 1 and 10 for the mean of Group 2, so that the difference in means will be 10. Next, we need to specify the pooled standard deviation, which is the square root of the average of the two standard deviations. In this case, it is sqrt((15^2 + 17^2)/2) = 16.03. The default significance level (alpha level) is .05. For this example we will set the power to be at .8.

```
library(pwr)

pwr.t.test(d=(0-10)/16.03,power=.8,sig.level=.05,type="two.sample",alternative="two.sided")

     Two-sample t test power calculation

              n = 41.31968
              d = 0.6238303
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

   NOTE: n is number in *each* group
```

The calculation results indicate that we need 42 subjects for diet A and another 42 subject for diet B in our sample in order the effect. Now, let's use another pair of means with the same difference. As we have discussed earlier, the results should be the same, and they are.

```
pwr.t.test(d=(5-15)/16.03,power=.8,sig.level=.05,type="two.sample",alternative="two.sided")

     Two-sample t test power calculation

              n = 41.31968
              d = 0.6238303
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

   NOTE: n is number in *each* group
```

Now the dietician may feel that a total sample size of 84 subjects is beyond her budget. One way of reducing the sample size is to increase the Type I error rate, or the alpha level. Let's say instead of using alpha level of .05 we will use .07. Then our sample size will reduce by 4 for each group as shown below.

```
pwr.t.test(d=(5-15)/16.03,power=.8,sig.level=.07,type="two.sample",alternative="two.sided")
```

```
        Two-sample t test power calculation

              n = 37.02896
              d = 0.6238303
      sig.level = 0.07
          power = 0.8
    alternative = two.sided

   NOTE: n is number in *each* group
```

Now suppose the dietician can only collect data on 60 subjects with 30 in each group. What will the statistical power for her t-test be with respect to alpha level of .05?

pwr.t.test(d=(5-15)/16.03,n=30,sig.level=.05,type="two.sample",alternative="two.sided")

```
        Two-sample t test power calculation

              n = 30
              d = 0.6238303
      sig.level = 0.05
          power = 0.6612888
    alternative = two.sided

   NOTE: n is number in *each* group
```

As we have discussed before, what really matters in the calculation of power or sample size is the difference of the means over the pooled standard deviation. This is a measure of effect size. Let's now look at how the effect size affects the sample size assuming a given sample power. We can simply assume the difference in means and set the standard deviation to be 1 and create a table with effect size, d, varying from .2 to 1.2.

```
ptab<-cbind(NULL, NULL)        # initalize ptab

for (i in c(.2, .3, .4, .5, .6, .7, .8, .9, 1, 1.1, 1.2)){
   pwrt<-pwr.t.test(d=i,power=.8,sig.level=.05,type="two.sample",alternative="two.sided")
   ptab<-rbind(ptab, cbind(pwrt$d, pwrt$n))
}

ptab
```
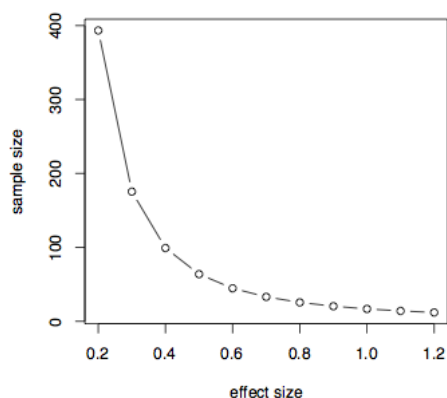
```
        [,1]       [,2]
 [1,]  0.2 393.40570
 [2,]  0.3 175.38467
 [3,]  0.4  99.08032
 [4,]  0.5  63.76561
 [5,]  0.6  44.58579
 [6,]  0.7  33.02458
 [7,]  0.8  25.52457
 [8,]  0.9  20.38633
 [9,]  1.0  16.71473
[10,]  1.1  14.00190
[11,]  1.2  11.94226
```

We can also easily display this information in a plot.

plot(ptab[,1],ptab[,2],type="b",xlab="effect size",ylab="sample size")

It shows that if the effect size is small, such .2 then we need a very large sample size and that sample size drops as effect size increases. We can also easily plot power versus sample size for a given effects size, say, d = 0.7

pwrt<-pwr.t.test(d=.7,n=c(10,20,30,40,50,60,70,80,90,100),sig.level=.05,type="two.sample",a

plot(pwrt$n,pwrt$power,type="b",xlab="sample size",ylab="power")
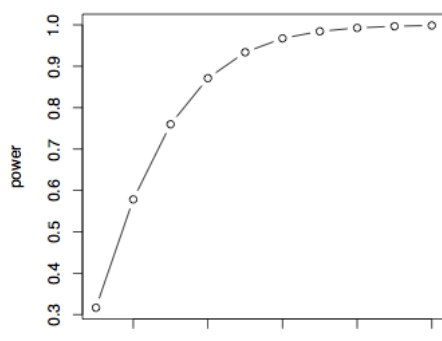
Discussion

An important technical assumption is the normality assumption. If the distribution is skewed, then a small sample size may not have the power shown in the results, because the value in the results is calculated using the method based on the normality assumption. We have seen that in order to compute the power or the sample size, we have to make a number of assumptions. These assumptions are used not only for the purpose of calculation, but are also used in the actual t-test itself. So one important side benefit of performing power analysis is to help us to better understand our designs and our hypotheses.

We have seen in the power calculation process that what matters in the two-independent sample t-test is the difference in the means and the standard deviations for the two groups. This leads to the concept of effect size. In this case, the effect size will be the difference in means over the pooled standard deviation. The larger the effect size, the larger the power for a given sample size. Or, the larger the effect size, the smaller sample size needed to achieve the same power. So, a good estimate of effect size is the key to a good power analysis. But it is not always an easy task to determine the effect size. Good estimates of effect size come from the existing literature or from pilot studies.

See Also

- References

D. Moore and G. McCabe, Introduction to the Practice of Statistics, Third Edition, Section 6.4

How to cite this page   Report an error on this page or leave a comment

The content of this web site should not be construed as an endorsement of any particular web site, book, or software product by the University of California.