Statistical Power - What is it?

Joseph J. Luczkovich, PhD

February 6, 2014

A common question before begining any study is ”How many samples or experimental units do I need?”. The short answer is: A lot. Remember the Central Limit Theorem and the coin flipping experiments you did earlier. It was unclear what the pattern was going to be with small numbers of flips, but the pattern became clear with 1000 or more flips. But time and money are limited, so what if you cannot establish 10,000 experimental plots? How many experimental units can you get away with? In part this depends on a trade-off between statistical power and effect size, and what you are willing to give up to answer your question.

Statistical Power is an estimate of the power or ability to detect a difference between means in ANOVA or other statistical procedures. More statitical power means that smaller differences in means can be detected. Generally, power increases with sample size. But what sample size is the minimum for getting a valid answer to your question or rejecting or accepting the hypothesis you have proposed? In most cases, you will need to have some preliminary data from a pilot study to complete the power calculation. You will need to know the size of the effect you wish to detect **effect size** (value δ is the difference between means) and also some estimate of variation (standard deviation σ of the means or variance σ^2).

First, you need to run a pilot study or use some previous data and know:

1. How many groups to be compared?
2. What is the between group and within group variance?
3. What probability do you want to reject the null hypothesis with (p =0.05)
4. What amount of ”power” will you accept? i.e. an 0.80 chance of rejecting the null hypothesis? Power =0.8

Lets do a simple two-sample t-test power calculation to get started. First, I import the Systat Wysocking Bay Juncus biomass data, then compute the means for 0 cm and 10 cm treatment groups. Next I compute delta δ is the difference between means.

```
> library("foreign", lib.loc="C:/Program Files/R/R-3.0.2/library")
> JB<-read.systat(file="C:/Users/Joseph/Documents/CRM7008/HW2_2014/HW2/bijr9298.SYD")
```

```
> xbar0=mean(JB$SEP93[JB$TREAT==0])
> xbar0 # mean of the 0 cm treatment in Sep 1993

[1] 43.05625

> xbar10=mean(JB$SEP93[JB$TREAT==10])
> xbar10 # mean of the 10 cm treatment in Sep 1993

[1] 9

> delta.0.10=xbar0-xbar10
> delta.0.10 # This is delta

[1] 34.05625

> power.t.test(n=4,delta=delta.0.10)

     Two-sample t test power calculation

              n = 4
          delta = 34.05625
             sd = 1
      sig.level = 0.05
          power = 1
    alternative = two.sided

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

Here the output shows that the power is very high (1.0) with just 4 replicates in each treament to detect a difference between means with effect size delta. This is because there was a huge effect size for these two treatments, i.e. delta was large, relative to the variance. We can also figure out what sample size we would need to detect a difference between 4 cm and 2 cm:

```
> xbar4=mean(JB$SEP93[JB$TREAT==4])
> xbar4 # mean of the 4 cm treatment in Sep 1993

[1] 30.56875

> xbar2=mean(JB$SEP93[JB$TREAT==2])
> xbar2 # mean of the 10 cm treatment in Sep 1993

[1] 24.575

> delta.4.2=xbar4-xbar2
> delta.4.2 # This is delta

[1] 5.99375
```

```
> power.t.test(n=4,delta=delta.4.2)

     Two-sample t test power calculation

              n = 4
          delta = 5.99375
             sd = 1
      sig.level = 0.05
          power = 0.9999994
    alternative = two.sided

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

The power to detect a difference at 0.05 between these groups is lower, and we would need more plots.

Let's look at an ANOVA example from the Wysocking Bay data we just analyzed. Lets say we did a pilot study in April 1992, before the study began.

```
> xbar0=mean(JB$APR92[JB$TREAT==0])
> xbar2=mean(JB$APR92[JB$TREAT==2])
> xbar4=mean(JB$APR92[JB$TREAT==4])
> xbar10=mean(JB$APR92[JB$TREAT==10])
> grmeans<-c(xbar0,xbar2,xbar4,xbar10)
> btvar<-var(grmeans)
> var0=var(JB$APR92[JB$TREAT==0])
> var2=var(JB$APR92[JB$TREAT==2])
> var4=var(JB$APR92[JB$TREAT==4])
> var10=var(JB$APR92[JB$TREAT==10])
> withinvar<-mean(c(var0,var2,var4,var10))
> #Write the means for each treatment
> xbar0 # the mean for the 0 cm treatments is:

[1] 23.975

> xbar2 # the mean for the 2 cm treatments is:

[1] 28.55625

> xbar4 # the mean for the 4 cm treatments is:

[1] 26.1625

> xbar10 # the mean for the 10 cm treatments is:

[1] 21.5375

> grmeans # A vector of the four treament means:

[1] 23.97500 28.55625 26.16250 21.53750
```

```
> btvar # the between treatment group varience
[1] 9.008161
> var0 # the variance for the 0 cm treatments is:
[1] 85.66333
> var2 # the variance for the 2 cm treatments is:
[1] 216.5786
> var4 # the variance for the 4 cm treatments is:
[1] 123.9012
> var10 # the variance for the 10 cm treatments is:
[1] 91.85183
> withinvar # the average of the variance in the four treatments
[1] 129.4987
> power.anova.test(groups=4, n=NULL, between.var=btvar, within.var=withinvar,power=0.9)

     Balanced one-way analysis of variance power calculation

         groups = 4
              n = 68.89388
    between.var = 9.008161
     within.var = 129.4987
      sig.level = 0.05
          power = 0.9

NOTE: n is number in each group
```

Wow! To detect a difference between means between the treament plots prior to adding the dredge spoil, we would need a lot of replicates, 69 or so in each treatment group, to detect a difference. There is no way we could affor this. Let's see what power we actually have with the replicate plots available (four):

```
> power.anova.test(groups=4, n=4, between.var=btvar, within.var=withinvar,power=NULL)

     Balanced one-way analysis of variance power calculation

         groups = 4
              n = 4
    between.var = 9.008161
     within.var = 129.4987
      sig.level = 0.05
          power = 0.08889619

NOTE: n is number in each group
```

In this power computation, we entered n=4 as the number of plots per treatment and left the power argument "NULL". The power, or ability to detect a difference between the plots is computed. It is low, abou 0.08 or an 8 percent chance of detecting a difference in group means. This is not a very powerful experimental design. **But this is before any treament was applied!** We are not really interested in that...The test we want to run power computation on is when we apply the dredge spoil. We don't want to detect differences of the unmanipulated plots, perhaps due to some other factor, but due to a dredging spoil application.

Let's do this for September 1993, one year after dredge spoil was applied. What are the power computations for smaple size needed based on application of the actual treaments we are tryoing to detect?

```
> xbar0=mean(JB$SEP93[JB$TREAT==0])
> xbar2=mean(JB$SEP93[JB$TREAT==2])
> xbar4=mean(JB$SEP93[JB$TREAT==4])
> xbar10=mean(JB$SEP93[JB$TREAT==10])
> grmeans<-c(xbar0,xbar2,xbar4,xbar10)
> btvar<-var(grmeans)
> var0=var(JB$SEP93[JB$TREAT==0])
> var2=var(JB$SEP93[JB$TREAT==2])
> var4=var(JB$SEP93[JB$TREAT==4])
> var10=var(JB$SEP93[JB$TREAT==10])
> withinvar<-mean(c(var0,var2,var4,var10))
> #Write the means for each treatment
> xbar0 # the mean for the 0 cm treatments is:

[1] 43.05625

> xbar2 # the mean for the 2 cm treatments is:

[1] 24.575

> xbar4 # the mean for the 4 cm treatments is:

[1] 30.56875

> xbar10 # the mean for the 10 cm treatments is:

[1] 9

> grmeans # A vector of the four treament means:

[1] 43.05625 24.57500 30.56875  9.00000

> btvar # the between treatment group varience

[1] 200.0866

> var0 # the variance for the 0 cm treatments is:
```

```
[1] 594.0093

> var2 # the variance for the 2 cm treatments is:

[1] 241.362

> var4 # the variance for the 4 cm treatments is:

[1] 426.4743

> var10 # the variance for the 10 cm treatments is:

[1] 188.7347

> withinvar # the average of the variance in the four treatments

[1] 362.6451

> #The number of plots needed to detect a change, given our observed variation, is:
> power.anova.test(groups=4, n=NULL, between.var=btvar, within.var=withinvar,power=0.9)

     Balanced one-way analysis of variance power calculation

         groups = 4
              n = 9.602795
    between.var = 200.0866
     within.var = 362.6451
      sig.level = 0.05
          power = 0.9

NOTE: n is number in each group

> #The power to detect a difference in means, given our plot replication and observed variat
> power.anova.test(groups=4, n=4, between.var=btvar, within.var=withinvar,power=NULL)

     Balanced one-way analysis of variance power calculation

         groups = 4
              n = 4
    between.var = 200.0866
     within.var = 362.6451
      sig.level = 0.05
          power = 0.4259762

NOTE: n is number in each group
```