Power Analysis

BIOE 498/598 PJ

Spring 2021

When is an effect size significant?

Linear models compute estimates of the true value of the parameters β.

The uncertainty in our estimate is quantified by the *standard error*.

Let's say we estimate β using n samples from a population with standard deviation σ. The standard error of our estimate is

$$\text{s.e.} = \sigma/\sqrt{n}$$

The 95% confidence interval for a parameter is 1.96 standard errors* on each side of the estimate:

$$\text{95\% C.I. of } \beta = [\beta - 1.96\text{s.e.}, \beta + 1.96\text{s.e.}]$$

* We often use 2 standard errors as a slightly conservative (and more convenient) estimate of the 95% confidence interval.

The 95% Confidence Interval

If a parameter estimate has a 95% C.I. that includes zero, we cannot be certain that the true value of the parameter is nonzero.

A parameter estimate is *significant* if and only if the 95% C.I. excludes zero.

The 95% C.I. depends on the number of samples (s.e. $= \sigma/\sqrt{n}$). We can narrow the 95% C.I. and improve our estimate of β by increasing n.

Example

Let's say we fit a model and found an estimate of $\beta = 3.1$ with s.e. $= 1.9$ using $n = 4$ samples. How many samples would we need before our estimate of β is significant?

Example

Let's say we fit a model and found an estimate of $\beta = 3.1$ with s.e. $= 1.9$ using $n = 4$ samples. How many samples would we need before our estimate of β is significant?

We always assume that the population standard deviation (σ) is independent of n. In our example, $\sigma = \sqrt{n}\,\text{s.e.} = 3.8$. For our estimate to be significant, the lower end of the 95% C.I. must exclude zero, so

$$\begin{aligned}
\beta - 1.96\sigma/\sqrt{n} &> 0 \\
n &> (1.96\sigma/\beta)^2 \\
n &> (1.96 \times 3.8/3.1)^2 \\
n &> 5.77
\end{aligned}$$

Example

Let's say we fit a model and found an estimate of $\beta = 3.1$ with s.e. $= 1.9$ using $n = 4$ samples. How many samples would we need before our estimate of β is significant?

We always assume that the population standard deviation (σ) is independent of n. In our example, $\sigma = \sqrt{n}\,\text{s.e.} = 3.8$. For our estimate to be significant, the lower end of the 95% C.I. must exclude zero, so

$$\begin{aligned} \beta - 1.96\sigma/\sqrt{n} &> 0 \\ n &> (1.96\sigma/\beta)^2 \\ n &> (1.96 \times 3.8/3.1)^2 \\ n &> 5.77 \end{aligned}$$

Two more samples ($n = 6$) would have been sufficient for our estimate of β to be significantly nonzero.

Power Analysis

The previous example makes two assumptions:

- The standard deviation (σ) will not change in subsequent experiments.
- The parameter estimate (β) will not change when new samples are added.

Power Analysis

The previous example makes two assumptions:

- The standard deviation (σ) will not change in subsequent experiments.
- The parameter estimate (β) will not change when new samples are added.

The first assumption is valid since σ is a property of the underlying population. Assuming our samples are drawn from the same population, they will have the same variation.

Power Analysis

The previous example makes two assumptions:

- The standard deviation (σ) will not change in subsequent experiments.
- The parameter estimate (β) will not change when new samples are added.

The first assumption is valid since σ is a property of the underlying population. Assuming our samples are drawn from the same population, they will have the same variation.

Our assumption about β is not valid. Remember that β is only an estimate of the true parameter value. If we re-sample the population we will get a new estimate. If the new estimate of β is any lower, the confidence interval in the previous example will again include zero.

Power Analysis (continued)

We need to be more conservative in our estimate of n to account for differences in the new estimates of β. Adding another 0.84 s.e. to our bound will ensure the 95% C.I. for β excludes zero **for 80% of the new estimates of** β.

Power Analysis (continued)

We need to be more conservative in our estimate of n to account for differences in the new estimates of β. Adding another 0.84 s.e. to our bound will ensure the 95% C.I. for β excludes zero **for 80% of the new estimates of β.**

Our new estimate for the sample size is

$$\beta - (1.96\text{s.e.} + 0.84\text{s.e.}) > 0$$
$$\Rightarrow n > (2.80\sigma/\beta)^2$$

Power Analysis (continued)

We need to be more conservative in our estimate of n to account for differences in the new estimates of β. Adding another 0.84 s.e. to our bound will ensure the 95% C.I. for β excludes zero **for 80% of the new estimates of** β.

Our new estimate for the sample size is

$$\beta - (1.96\text{s.e.} + 0.84\text{s.e.}) > 0$$
$$\Rightarrow n > (2.80\sigma/\beta)^2$$

Even with this conservative estimate, there is still a 20% chance that our estimate of β will not be significant, although this level of uncertainty seems to be acceptable to most experimenters.

Power calculations using a t-test

The significance of an effect size is determined by a t-test, which can differ from the normal distribution used on the previous slide.

Unfortunately, calculating n using a t-distribution is not simple. We use the R function `power.t.test` instead.

Power calculations using a *t*-test

The significance of an effect size is determined by a t-test, which can differ from the normal distribution used on the previous slide.

Unfortunately, calculating n using a t-distribution is not simple. We use the R function `power.t.test` instead.

```
power.t.test(n=NULL,
             delta=...,
             sd=...,
             power=...,
             alternative="one.sided")
```

- `delta` is the effect size (β)
- `sd` is the standard deviation
- `power` is 0.8 for an 80% chance of seeing a significant result
- `alternative="one.sided"` assumes the effect won't change signs

Back to the farm

```
##
## Call:
## lm(formula = yield ~ ammonia + phosphate, data = fert)
##
## Residuals:
##        1        2        3        4        5        6
## -0.12172  0.03404  0.12402  0.08646  0.11808 -0.24089
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.1640     0.2247  -0.730    0.518
## ammonia       0.1909     0.3152   0.606    0.588
## phosphate     0.4201     0.2273   1.848    0.162
##
## Residual standard error: 0.1922 on 3 degrees of freedom
## Multiple R-squared:  0.5327, Adjusted R-squared:  0.2212
## F-statistic:  1.71 on 2 and 3 DF,  p-value: 0.3194
```

Power analysis for phosphate

```
##               Estimate Std. Error    t value  Pr(>|t|)
## (Intercept) -0.1640188  0.2247293 -0.7298506 0.5183074
## ammonia      0.1908623  0.3151669  0.6055911 0.5875313
## phosphate    0.4200717  0.2273268  1.8478757 0.1617658
```

Power analysis for phosphate

```
##                Estimate Std. Error    t value  Pr(>|t|)
## (Intercept) -0.1640188  0.2247293 -0.7298506 0.5183074
## ammonia      0.1908623  0.3151669  0.6055911 0.5875313
## phosphate    0.4200717  0.2273268  1.8478757 0.1617658
```

```r
power.t.test(n=NULL, delta=0.4201, sd=0.2273*sqrt(3),
             power=0.8, alternative="one.sided")
```

```
##
##      Two-sample t test power calculation
##
##               n = 11.60011
##           delta = 0.4201
##              sd = 0.3936951
##       sig.level = 0.05
##           power = 0.8
##     alternative = one.sided
##
## NOTE: n is number in *each* group
```

Power analysis for phosphate

```
##               Estimate Std. Error    t value  Pr(>|t|)
## (Intercept) -0.1640188  0.2247293 -0.7298506 0.5183074
## ammonia      0.1908623  0.3151669  0.6055911 0.5875313
## phosphate    0.4200717  0.2273268  1.8478757 0.1617658
```

```
power.t.test(n=NULL, delta=0.4201, sd=0.2273*sqrt(3),
             power=0.8, alternative="one.sided")
```

```
##
##      Two-sample t test power calculation
##
##               n = 11.60011
##           delta = 0.4201
##              sd = 0.3936951
##       sig.level = 0.05
##           power = 0.8
##     alternative = one.sided
##
## NOTE: n is number in *each* group
```

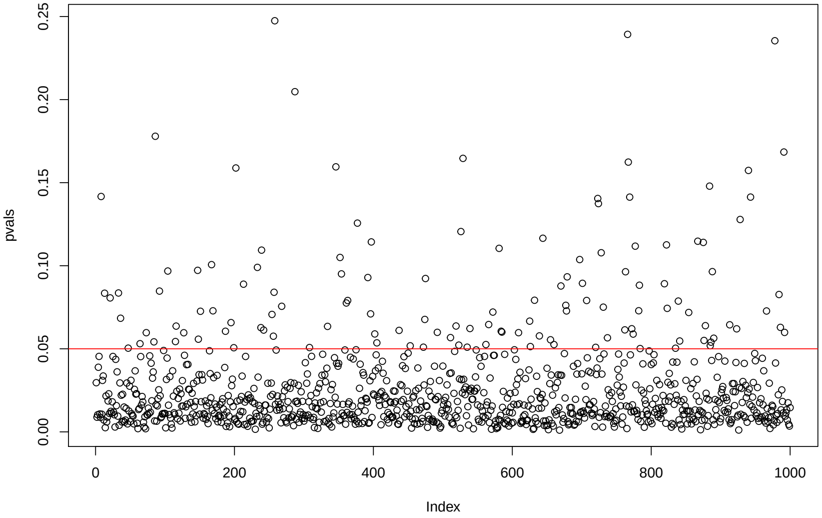
Six more runs (12 total) should be sufficient 80% of the time.

After adding six more runs. . .

```
##
## Call:
## lm(formula = yield ~ ammonia + phosphate, data = aug_fert)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.20426 -0.07663  0.05880  0.09042  0.16070
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.03192    0.08886  -0.359   0.7278
## ammonia      0.03634    0.11973   0.303   0.7684
## phosphate    0.29305    0.10419   2.813   0.0203 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
##
## Residual standard error: 0.1382 on 9 degrees of freedom
## Multiple R-squared:  0.4684, Adjusted R-squared:  0.3503
## F-statistic: 3.965 on 2 and 9 DF,  p-value: 0.05823
```

Is six more runs always enough?

Below are the *p*-values for the phosphate effect from 1,000 models fit with six additional runs.



One last note

Given enough runs, any effect size — no matter how small — will become statistically significant.

Never forget that **statistical** significance does not imply **practical** significance.

Keep your focus on the effect size, not the p-value.