`pwr` cheat sheet

Clay Ford

The `pwr` package provides basic functions for power and sample size analysis as described in the book *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.) by Jacob Cohen.

```
library(pwr)
```

conventional effect sizes for certain tests

Test	`small`	`medium`	`large`
tests for proportions (`p`)	0.2	0.5	0.8
tests for means (`t`)	0.2	0.5	0.8
chi-square tests (`chisq`)	0.1	0.3	0.5
correlation test (`r`)	0.1	0.3	0.5
anova (`anov`)	0.1	0.25	0.4
general linear model (`f2`)	0.02	0.15	0.35

To get these in R: `cohen.ES(test = c("p", "t", "r", "anov", "chisq", "f2"), size = c("small", "medium", "large"))`

Example:

```
cohen.ES(test = "anov", size = "small")
```

```
## 
##      Conventional effect size from Cohen (1982) 
## 
##            test = anov
##            size = small
##     effect.size = 0.1
```

one-sample test for proportions

Say we think people place name tags on the left side of their chest 65% percent of the time versus random chance (50%). What sample size do we need to show this assuming a significance level (Type I error) of 0.05 and a desired power of 0.80?

```
pwr.p.test(h = ES.h(p1 = 0.65, p2 = 0.50), sig.level = 0.05,
           power = 0.80, alternative = "greater")
```

Say we think people place name tags on the left 75% percent of the time instead of 50%. What is the power of our test if we survey 30 people provided we accept a significance level (Type I error) of 0.05?

```
pwr.p.test(h = ES.h(p1 = 0.75, p2 = 0.50), n = 30,
           sig.level = 0.05, alternative = "greater")
```

Say we think people placing name tags on the left side of their chest is not governed by random chance (50%). What sample size do we need to detect a "small" effect of 0.2 assuming a significance level of 0.05 and a desired power of 0.80?

```
pwr.p.test(h = 0.2, sig.level = 0.05, power = 0.8, alternative = "two.sided")
```

two-sample test for proportions

I randomly sample male and female UVA undergrad students and ask them if they consume alcohol at least once a week. My null hypothesis is no difference in the proportion that answer yes. My alternative hypothesis is that there is a difference. (two-sided; one gender has higher proportion, I don't know which.) I'd like to detect a difference as small as 5%. How many students do I need to sample in each group if we want 80% power and 5% chance of Type 1 error?

```
# 55% vs. 50%
pwr.2p.test(h = ES.h(p1 = 0.55, p2 = 0.50), sig.level = 0.05, power = .80)
# 35% vs. 30%
pwr.2p.test(h = ES.h(p1 = 0.35, p2 = 0.30), sig.level = 0.05, power = .80)
# using "small" conventional effect size
pwr.2p.test(h = 0.2, sig.level = 0.05, power = .80)
```

two-sample test for proportions, unequal sample sizes

I randomly sample male and female UVA undergrad students and ask them if they consume alcohol at least once a week. My null hypothesis is no difference in the proportion that answer yes. My alternative hypothesis is that there is a difference. (two-sided; one gender has higher proportion, I don't know which.) We were able to survey 543 males and 675 females. What's the power of our test if we want to be able to detect a "small" effect (h = 0.2)?

```
pwr.2p2n.test(h = 0.2, n1 = 543, n2 = 675, sig.level = 0.05)
```

one-sample and two-sample t-tests for means

The effect size for two-sample t-tests:

$$d = \frac{m_1 - m_2}{\sigma}$$

The effect size for one-sample t-tests:

$$d = \frac{m_1 - null}{\sigma}$$

The effect size for paired t-tests, where σ_d is the standard deviation of differences between pairs:

$$d = \frac{m_1 - m_2}{\sigma_d}$$

I'm interested to know if there is a difference in the mean price of what male and female students pay at the library coffee shop. Let's say I randomly observe 30 male and 30 female students check out from the coffee shop and note their total purchase price. How powerful is this experiment if I want to detect a "small" effect in either direction?

```r
# two-sample test
pwr.t.test(n = 30, d = 0.2, sig.level = 0.05) # n is per group
```

How many do I need to observe for a test with 80% power?

```r
pwr.t.test(d = 0.2, power = 0.80, sig.level = 0.05)
# alternative: greater than 0 (positive effect)
pwr.t.test(d = 0.2, power = 0.80, sig.level = 0.05, alternative = "greater")
# alternative: less than 0 (negative effect)
pwr.t.test(d = -0.2, power = 0.80, sig.level = 0.05, alternative = "less")
```

Let's say we want to be able to detect a difference of at least 75 cents in the mean purchase price. How can we convert that to an effect size? First we need to make a guess at the population standard deviation. If we have absolutely no idea, one rule of thumb is to take the difference between the maximum and minimum values and divide by 4 (or 6). Let's say max is 10 and min is 1. So our guess at a standard deviation is $9/4 = 2.25$.

```r
# one-sample test
d <- 0.75/2.25 # 0.333
pwr.t.test(d = d, power = 0.80, sig.level = 0.05, type = "one.sample")
```

I think the average purchase price at the Library coffee shop is over \$3 per student. My null is \$3 or less; my alternative is greater than \$3. If the true average purchase price is \$3.50, I would like to have 90% power to declare my estimated average purchase price is greater than \$3. How many transactions do I need to observe assuming a significance level of 0.05? Let's say max purchase price is \$10 and min is \$1. So our guess at a standard deviation is $9/4 = 2.25$. Therefore d is...

```r
d <- 0.50/2.25
pwr.t.test(d = d, sig.level = 0.05, power = 0.90, alternative = "greater",
           type = "one.sample")

# or with power.t.test:
power.t.test(delta = 0.50, sd = 2.25, power = 0.90, sig.level = 0.05,
             alternative = "one.sided", type = "one.sample")
```

24 high school boys are put on a ultraheavy rope-jumping program. Does this increase their 40-yard dash time? We'll measure their 40 time before the program and after. We'll use a paired t-test to see if the difference in times is greater than 0. Assume the standard deviation of the differences will be about 0.25. How powerful is the test to detect a difference of about positive 0.08 with 0.05 significance?

```r
# paired t-test
pwr.t.test(n = 24, d = 0.08 / 0.25,
           type = "paired", alternative = "greater")

# or
power.t.test(n = 24, delta = 0.08, sd = 0.25,
             type = "paired", alternative = "one.sided")
```

two sample t test for means, unequal sample sizes

Find power for a t test with 28 in one group and 35 in the other group and a medium effect size. (sig.level defaults to 0.05.)

```r
pwr.t2n.test(n1 = 28, n2 = 35, d = 0.5)
```

Find n1 sample size when other group has 35, desired power is 0.80, effect size is 0.5 and significance level is

0.05:

```
pwr.t2n.test(n2 = 35, d = 0.5, power = 0.8)
```

chi-squared tests

(From Cohen, example 7.1) A market researcher is seeking to determine preference among 4 package designs. He arranges to have a panel of 100 consumers pick their favorite design. He wants to perform a chi-square goodness of fit test against the null of equal preference (25% for each design) with a significance level of 0.05. What's the power of the test if 3/8 of the population actually prefers one of the designs and the remaining 5/8 are split over the other 3 designs?

```
# Goodness of Fit
# We need to create vectors of null and alternative proportions:
P0 <- rep(0.25, 4)
P1 <- c(3/8, rep((5/8)/3, 3))

# To calculate power, specify effect size, N, and degrees of freedom (4-1).
pwr.chisq.test(w=ES.w1(P0,P1), N=100, df=(4-1), sig.level=0.05)

# How many subjects do we need to achieve 80% power?
pwr.chisq.test(w=ES.w1(P0,P1), df=(4-1), power=0.8, sig.level = 0.05)
```

I want to see if there's an association between gender and flossing teeth among UVA students. I randomly sample 100 students (male and female) and ask whether or not they floss daily. I want to carry out a chi-square test of association to determine if there's an association between these two variables. As usual I set my significance level to 0.05. To determine effect size I need to propose an alternative hypothesis, which in this case is a table of proportions.

```
# test of association
# using hypothesized proportions
prob <- matrix(c(0.10,0.20,0.40,0.30), ncol=2,
               dimnames = list(c("M","F"),c("Floss","No Floss")))
pwr.chisq.test(w = ES.w2(prob), N = 100, df = 1, sig.level = 0.05)

# How many students should I survey if I wish to achieve 90% power?
pwr.chisq.test(w = ES.w2(prob), power = 0.9, df = 1, sig.level = 0.05)

# using "small" conventional effect
pwr.chisq.test(w = 0.1, N = 100, df = 1, sig.level = 0.05)
pwr.chisq.test(w = 0.1, power = 0.9, df = 1, sig.level = 0.05)
```

correlation test

Let's say I'm a web developer and I want to conduct an experiment with one of my sites. I want to randomly select a group of people, ranging in age from 18 - 65, and time them how long it takes them to complete a task, say locate some piece of information. I suspect there may be a "small" positive linear relationship between time it takes to complete the task and age. How many subjects do I need to detect this relationship with 80% power and the usual 0.05 significance level?

```
# "small" effect is 0.1
pwr.r.test(r = 0.1, sig.level = 0.05, power = 0.8, alternative = "greater")
# detect small effect in either direction
pwr.r.test(r = 0.1, sig.level = 0.05, power = 0.8, alternative = "two.sided")
```

balanced one-way analysis of variance tests

Let's say I'm a web developer and I'm interested in 3 web site designs for a client. I'd like to know which design(s) help users find information fastest, or which design requires the most time. I design an experiment where I have 3 groups of randomly selected people use one of the designs to find some piece of information and I record how long it takes. (All groups look for the same information.) How many people do I need in each group to detect a "medium" effect if I desire power and significance levels of 0.8 and 0.05?

```
pwr.anova.test(k = 3, f = 0.25, sig.level = 0.05, power = 0.8)
```

What's the power of my test to detect a medium effect if I get 50 for each group?

```
pwr.anova.test(k = 3, f = 0.25, sig.level = 0.05, n = 50)
```

How many people do I need in each group if I believe two of the designs will take 30 seconds and one will take 25 seconds? Assume population standard deviation is 5 and that I desire power and significance levels of 0.8 and 0.05. Using `power.anova.test`:

```
power.anova.test(groups = 3, between.var = var(c(30, 30, 25)),
                 within.var = 5^2, power = 0.8)
```

test for the general linear model

Effect size for testing if set of predictors explain any variance:

$$f^2 = \frac{R^2}{1 - R^2}$$

Effect size for test if set of predictors explain variance above and beyond a second set of predictors:

$$f^2 = \frac{R^2_{AB} - R^2_A}{1 - R^2_{AB}}$$

R^2 is percent variance explained. R^2_{AB} is percent variance explained by variable sets A and B. R^2_A is percent variance explained by variable set A.

Let's say I'm hired to survey a company's workforce about job satisfaction. I ask employees to rate their satisfaction on a scale from 1 (hating life) to 10 (loving life). I know there will be variability in the answers, but I think two variables that will explain this variability are salary and gender. In fact I think it will explain at least 30% ($R^2 = 0.30$) of the variance. How powerful is my "experiment" if I randomly recruit 40 employees and accept a 0.05 significance level?

Two predictors, so u = 3. 40 employees, so v = 40 - 3 - 1. $R^2 = .30$, so effect size is $f2 = 0.3/(1 - 0.3)$

```
pwr.f2.test(u = 3, v = 40 - 3 - 1, f2 = 0.3/(1 - 0.3), sig.level = 0.05)
```

How many employees do I need to survey if I want to be able to detect at least 30% explained variance ($R^2 = 0.30$) with 80% power and the usual 0.05 significance level? We have to find v and then derive n.

```
pwr.f2.test(u = 2, f2 = 0.3/(1 - 0.3), sig.level = 0.05, power = 0.8)
# n = round(v) + u + 1
```

Continuing with previous example, it would be of interest if having your own office accounted for at least 5% beyond the variance explained by the model with salary and gender. We could think of this as a 0/1 indicator in the model that takes the value 1 if an employee has his/her own office, and 0 otherwise. In this case our effect size is

$$f^2 = \frac{0.35 - 0.30}{1 - 0.35}$$

How many employees would we need to survey to identify the office contribution to variance explained with 90% power and a significance level of 0.05?

```
pwr.f2.test(u = 2, f2 = (0.35 - 0.30) /(1 - 0.35), sig.level = 0.05, power = 0.9)
# n = round(v) + number of variables in A & B + 1
# n = 165 + 3 + 1 = 169
```

References

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (2nd ed.)*. LEA.

Dalgaard, P. (2002). *Introductory Statistics with R*. Springer. (Ch. 2)

Hogg, R and Tanis, E. (2006). *Probability and Statistical Inference (7th ed.)*. Pearson. (Ch. 9)

Kabacoff, R. (2011). *R in Action*. Manning. (Ch. 10)

Ryan, T. (2013). *Sample Size Determination and Power*. Wiley.