

Research Appreciation Week  
Presentation

Practical Strategies for Increasing Power

Christoph Maier

Coordinator, the Applied Research Lab (ARL)

April 5, 2012

# Overview

R. Baker Bausell and Yu Fang Li devote an entire chapter of their book “Power Analysis for Experimental Research: A Practical Guide for the Biological, Medical and Social Sciences to eleven strategies for increasing statistical power. The most obvious strategy is to increase the number of subjects but this is all too often not viable. This presentation will briefly discuss concepts such as power and effect size and then cover some of the most viable strategies for researchers at IUP.

# Question?

Question: Are 9-year old boys in the USA taller than 8-year old boys in the USA?

The CDC provides us with detailed stature (height) information:

Age	Mean Height (inches)
8 years	50.3
9 years	52.6

Effect size = true difference in the means

$$52.6 - 50.3 = 2.3 \text{ inches!}$$

## Information Important for Testing

CDC Tables also give percentiles. Here are the 25<sup>th</sup> and 75<sup>th</sup> percentiles, as well as the IQR.

Age	Mean Height	25th Percentile	75 <sup>th</sup> percentile	IQR
8 years	50.3	48.8	51.9	3.1
9 years	52.6	51.0	54.2	3.2

**How large is this effect?** There is very little overlap in heights of the middle 50% of boys in the two groups!  
**LARGE EFFECT!**

Within-Group variability is the yard-stick by which we measure differences between groups. IQR=measure of variability of the middle half of the data.

Symmetry (actually Normality) is also very important!

# Effect Sizes

Effect sizes are usually standardized (divided by the within-group standard deviation).

For this data, the standard deviation = 2.3 inches!

(SD  $\approx$  IQR/1.35)

Age	Mean Height (inches)	Unstandardized effect size (inches)	d = standardized effect size (inches)
96 months (8 years)	50.3	0.0	0.00
99 months	50.9	0.6	0.22 (small)
102 months (8.5 years)	51.5	1.2	0.52 (medium)
105 months	52.0	1.7	0.83 (large)
108 months (9 years)	52.6	2.3	1.00

# Effect Size Measured in Different Ways

Often determined by the statistical test

- Two-sample t-tests 
$$d = \frac{\mu_1 - \mu_2}{\sigma}$$

- One-way k-group ANOVA: 
$$f = \sqrt{\frac{\sum (\mu_j - \mu)^2}{k\sigma_e^2}}$$

Chi-square test:

$$w = \sqrt{\sum \frac{(p_{ij} - p_{0ij})^2}{p_{0ij}}}$$

Summed over  
k groups

Summed over  
all cells

# Small Medium and Large Effect Sizes

- Jacob Cohen, the power analysis guru came up with rules of thumb for small medium and large effect sizes.
- Example for  $d$ :  $d=.2$  small,  $d=.5$  medium,  $d=.8$  large
- Rules vary from discipline to discipline, and even not used in some disciplines.
- The medium effect size is often used for power analysis. Cohen: “a medium effect size is one large enough to be visible to the naked eye”.

# Other Important Parameters

## Level of significance.

$\alpha$ =the level of unacceptable risk of making a type I error (say it is different when it is not)

Usually  $\alpha = .05$  (5% level)

## Power

- The ability (think probability) to detect a specified difference (effect size).
- commonly .8 or 80% is used.

**Note:** When power is fixed, the smaller the effect size, the larger the required sample size.



# Many tables exist for determining power or sample size

Example for two-sided independent-samples t-test for finding sample size for a given power, effect size(d) , and  $\alpha = .05$  (Cohen 2.4.1)

$$\frac{\alpha_2 = .05 (\alpha_1 = .025)}{d}$$

Power	.10	.20	.30	.40	.50	.60	.70	.80	1.00	1.20	1.40
.25	332	84	38	22	14	10	8	6	5	4	3
.50	769	193	86	49	32	22	17	13	9	7	5
.60	981	246	110	62	40	28	21	16	11	8	6
2/3	1144	287	128	73	47	33	24	19	12	9	7
.70	1235	310	138	78	50	35	26	20	13	10	7
.75	1389	348	155	88	57	40	29	23	15	11	8
.80	1571	393	175	99	64	45	33	26	17	12	9
.85	1797	450	201	113	73	51	38	29	19	14	10
.90	2102	526	234	132	85	59	44	34	22	16	12
.95	2600	651	290	163	105	73	54	42	27	19	14
.99	3675	920	409	231	148	103	76	58	38	27	20

# Many software packages and website calculators (applets) are available as well

- **Websites:**

Russ Lenth: Java applets for power and sample size:

<http://www.stat.uiowa.edu/~rlenth/Power/>

Daniel Soper's Free Statistics Calculators website.

<http://www.danielsoper.com/statcalc3/default.aspx>

- **PASS:** Power Analysis and Sample Size Software for Windows.  
(available at the ARL)

- The “pwr” package in r (language and environment for statistical computing and graphics). Can be downloaded for free.

See <http://www.statmethods.net/stats/power.html>

## Power and Sample Size in r

Function:

```
pwr.t.test(n = , d = , sig.level = 0.05, power = ,  
           type = c("two.sample", "one.sample", "paired"),  
           alternative = c("two.sided", "less", "greater"))
```

where

n=sample size

d=effect size

sig.level=significance level

power =power(as decimal, e.g. .8)

Leave out the term for which you are solving (usually power or n).

# Example 1: Finding sample size

Example to find sample size for two-sample two-sided t-test at the 5% level with 80% power.

```
pwr.t.test(d = .5 , sig.level = 0.05, power = .8, type = c("two.sample"),  
alternative = c("two.sided"))
```

Two-sample t test power calculation

n = 63.8 ← 64 in each group

d = 0.5

sig.level = 0.05

power = 0.8

alternative = two.sided

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

## Example 2: Finding Power

Example to find power for two-sample two-sided t-test at the 5% level with  $n=50$  in each group:

```
pwr.t.test(n=50, d = .5 , sig.level = 0.05, type = c("two.sample"), alternative = c("two.sided"))
```

Two-sample t test power calculation

```
      n = 50  
      d = 0.5  
sig.level = 0.05  
power = 0.6969 ← .70  
alternative = two.sided
```

NOTE: n is number in *each* group

# Strategies for Increasing Power

- 1) Increase the sample size in the control group.
- 2) Collapse categories into fewer groups.
- 3) Adding a Covariate.
- 4) Use a Blocking Variable.
- 5) Use a repeated measures design.
- 6) Employing reliable measures.
- 7) Employing measures that are sensitive to change.
- 8) Using direct, rather than indirect dependent measures.

## Strategy 1

### Increase the sample size in the control group

- Consider experiments involving a control and an intervention
  - control is inexpensive and easily obtained
  - intervention is expensive or time-consuming.
- Tables give the sample size required for each group.

## Strategy 1 Continued

- Suppose that
  - my intervention sample is limited to  $N_{\text{exp}}$  individuals.
  - The tables indicate that I need a sample size of  $N_{\text{table}}$  individuals in each group.
- How many individuals ( $N_{\text{control}}$ ) do I need for the control group?

$$N_{\text{Table}} = \frac{2N_{\text{control}}N_{\text{exp}}}{N_{\text{control}} + N_{\text{exp}}} \Rightarrow N_{\text{Control}} = \frac{N_{\text{Table}}N_{\text{exp}}}{2N_{\text{exp}} - N_{\text{Table}}}$$

**Harmonic Mean**



## Strategy 1 Continued

For an independent-sample two-sided t-test, we need 64 individuals in each group to have the ability to detect a difference of  $d = .5$  (medium) at the 5% level of significance with 80% power.

Here is the relationship between  $N_{\text{exp}}$  and  $N_{\text{control}}$

$N_{\text{exp}}$	64	60	56	52	48	44	40	36
$N_{\text{control}}$	64	69	75	83	96	117	160	288
Total n	128	129	132	135	144	161	200	324

Small increases in n

Rapid acceleration in total n

## Strategy 2 Collapse into fewer Groups

Ask for an individual's race as (6 categories):

One-race

White

Black/African American,

Asian

Some Other Race

Two or More Races

But collapse to three categories for the analysis

One-race White

One-race Black/African American

Other"

Ask for agreement as (4 categories)

"Strongly Disagree", "Somewhat Disagree", "Agree", "Strongly Agree",

But collapse to two categories for the analysis

"Disagree" and "Agree"

## Example for Strategy 2

RQ: Is opinion (SD, D, A, SA) on some issue independent of race(One-race White, One-race Black, One-race Asian, One-race Other, Two more races):

**Test:** Chi-Square Test for Independence

**Level of Significance:**  $\alpha = .05$

**Power:** 80%

**Effect size:**  $w=.3$  (medium)

**Df :**  $(\text{number columns}-1)(\text{number rows}-1)=(4-1)(5-1)=12$

R function:

`pwr.chisq.test(w=.3,df=12,power=.8,sig.level=.05)`

## Strategy 2

Required Sample Size Using all categories

```
pwr.chisq.test(w=.3,df=12,power=.8,sig.level=.05)
```

Chi squared power calculation

$w = 0.3$

**$N = 192.6216$**

**193 total  
individuals**

$df = 12$

$sig.level = 0.05$

$power = 0.8$

NOTE: N is the number of observations

## Required Sample Sizes Collapsing Categories

Collapsing race to three categories, agreement uncollapsed

Using all  $4 * 3 = 12$  combinations,

$Df = (\# \text{ of cols}) * (\# \text{ of rows}) = (4 - 1) * (3 - 1) = 6$

Tells R to display  
Sample size only.

`pwr.chisq.test(w=.3,df=6,power=.8,sig.level=.05)$N`

`[1] 151.381` total individuals reduced from 196 to 152

Collapsing race to three categories and agreement to two categories

Using all  $2 * 3 = 6$  combinations,

$Df = (\# \text{ of cols}) * (\# \text{ of rows}) = (2 - 1) * (3 - 1) = 2$

`round(pwr.chisq.test(w=.3,df=2,power=.8,sig.level=.05)$N,1)`

Tells R to round N to 1 digit  
beyond the decimal point.

`[1] 107.1` total individuals  
reduced to 108

## Strategy 3 Add a Covariate

- Adding a variable that is correlated with the dependent sample increases the effect size

$$f' = \frac{f}{\sqrt{1-r^2}}$$

where  $f$ =effect size for the ANOVA and  $r$  is the correlation between the covariate and the dependent variable.

- The higher the correlation, the more the effect size increases.
- Remember that larger effect sizes are easier to detect, thus require smaller sample sizes.

# How much does it help?

$$\frac{f'}{f} = \frac{1}{\sqrt{1-r^2}}$$

r	0	.1	.2	.3	.4	.5	.6	.7	.8	.9
f'/f	1.00	1.01	1.02	1.05	1.09	1.15	1.25	1.40	1.67	2.29
n per group (k=2)	64	64	62	59	54	49	42	34	24	14
n per group (k=3)	53	52	51	48	45	40	34	28	20	11
n per group (k=4)	45	45	43	41	38	34	29	24	17	10

K=number of groups

## Strategy 3 Example

(alternative to the small-medium-large approach)

- Wish to look at whether there is a difference in the prices that used-car dealers offer male and female customers for cars, here restricted to Honda Civics.
- How many men and women should they use in their experiment?

**Test:** independent-samples t-test

**Two-tailed test** (although more logically one-tailed)

**Level of significance:**  $\alpha = .05$

**Power:** 80%



## Questions that the Researcher needs to answer

- 1) What constitutes an important difference between offers made to men and women in terms of average car price?

\$500 dollars

- 2) What is the within-group standard deviation?

From pilot study, previous studies, or if all else fails, by the following conservative estimate:

Range of reasonable prices/4:

$$(\$17,000 - \$1,000) / 4 = \$4000$$

- 3) Compute  $d = \text{important difference} / \text{SD}$ :

$$500 / 4000 = 1/8 = .125$$

- 4) `pwr.anova.test(k = 2 , f = .125, sig.level = .05, power = .80)`

result for t-test: n= 1006 male and 1006 female customers.

Say that conservatively that  $r = -.6$  between  
mileage and price

$$f' = \frac{1}{\sqrt{1-r^2}} f = \frac{1}{\sqrt{1-.6^2}} f = 1.25 f = 1.25(1/16)$$

Can also get  
this from table  
on slide 23

`pwr.anova.test(k = 2 , f = 1.25*1/16,  
sig.level = .05, power = .8)`

Result: 644 male and 644 female customers.

## Strategy 4 Use a Blocking Variable

- Similar to covariate, but blocking uses a categorical variable. Individuals in a study are usually assigned to blocks before the study begins.
- **Example:** Look at the difference in learning for two learning strategies.
- Students' typical level of conscientiousness (effort) is a very important predictor of learning. Estimated as one of three levels (low, medium, high). Students in each "effort" block are randomized to the treatments.
- Goal: to reduce the within-strategy variability.

# Strategy 5 Using a repeated-measures design

- designs in which subjects receive all treatments (think repeated measures).
- As the correlation between the dependent observations increases, the power increases.
- Not always possible to use this design.
- Even more effective than using a covariate.
- Example: Each subject gets both treatments.  
(within-subjects). For an effect size of  $d=.5$  (medium effect size), 5% level of significance, and power=.8

r	0	.2	.4	.5	.6	.7	.8	.9
Adjusted effect size	0.50	0.56	0.65	0.71	0.79	0.91	1.12	1.58
n per group (k=2)	64	52	39	33	27	21	14	8
Using <u>covariate</u> : n per group (k=2)	64	62	54	49	42	34	24	14

# Other Strategies

## Strategy 6 Employing reliable measures.

- the greater the internal consistency of the items in the scale, the less within-group variability, which means more power to detect differences of the effects.

## Strategy 7 Employing measures that are sensitive to change.

### Example:

The Multidimensional Body-Self Relations Questionnaire (MBSRQ)

- It is a well-validated self-report inventory for the assessment of body image.
- It consists of many subscales including appearance evaluation, appearance orientation, fitness evaluation, fitness orientation, health orientation, and illness orientation.
- If your intervention deals only with health orientation, administer the items of the health subscale only, rather than the entire MBSRQ. Otherwise the effect you are trying to measure will be diluted by unrelated information.

# Other Strategies Continued

Strategy 8 Using direct, rather than indirect dependent measures.

Example Measuring the effect of an intervention to reduce strokes

The intervention consists of educating the subjects about healthy diets and good exercises to reduce the likelihood of suffering a stroke.

- 1) Education might have an effect on their knowledge about how they should eat and how they should exercise.
- 2) Which might have an effect on their behavior.
- 3) Which might affect their blood cholesterol levels.
- 4) Which might lead to a reduction in strokes.

BUT, as you move further down the chain of effects (further away from the intervention) , the ability to detect the effect decreases.

# Installing r and the Cohen Sample Size Package

To install r:

<http://www.r-project.org/>

To install the “pwr” package:

```
install.packages(c('pwr'))
```

Choose CRAN mirror site: USA(PA1) and click “OK” button.

```
Type: library(pwr)
```

## Reference Books

R Barker Bausell and Yu-Fang Li. Power Analysis for Experimental Research A practical Guide for the Biological, Medical and Social Sciences. Cambridge University Press, 2002.

Jacob Cohen. Statistical Power for the Behavioral Sciences, Second Edition. Lawrence Erlbaum Associates, 1988.

Robert J. Grissom and John Kim. Effect Sizes for Research: Univariate and Multivariate Applications, Second Edition. Routledge Academic, 2011.