

Power and Effect Size

Cal State Northridge
Ψ320
Andrew Ainsworth PhD

Major Points

- ◆ Review
- ◆ What is power?
- ◆ What controls power?
- ◆ Effect size
- ◆ Power for one sample t
- ◆ Power for related-samples t
- ◆ Power for two sample t

Psy 320 - Cal State Northridge

2

Important Concepts

- ◆ Concepts critical to hypothesis testing
 - Decision
 - Type I error
 - Type II error
 - Critical values
 - One- and two-tailed tests

Psy 320 - Cal State Northridge

3

Decisions

- ◆ When we test a hypothesis we draw a conclusion; either correct or incorrect.
 - Type I error
 - Reject the null hypothesis when it is actually correct.
 - Type II error
 - Retain the null hypothesis when it is actually false.

Psy 320 - Cal State Northridge

4

Type I Errors

- ◆ Null Hypothesis really is true
- ◆ We conclude the null is false.
- ◆ This is a Type I error
 - Probability set at alpha (α)
 - α usually at .05
 - Therefore, probability of Type I error = .05

Psy 320 - Cal State Northridge

5

Type II Errors

- ◆ The Alternative Hypothesis is true
- ◆ We conclude that the null is true
- ◆ This is also an error (Type II)
 - Probability denoted beta (β)
 - We can't set beta easily.
 - We'll talk about this issue later.
- ◆ Power = $(1 - \beta)$ = probability of correctly rejecting false null hypothesis.

Psy 320 - Cal State Northridge

6

Confusion Matrix

		Reality	
		H ₀	H ₁
Your Decision	“H ₀ ”	1 - α	β
	“H ₁ ”	α	1 - β
		1.00	1.00

		Reality	
		H ₀	H ₁
Your Decision	“H ₀ ”	.95	.16
	“H ₁ ”	.05	.84
		1.00	1.00

Psy 320 - Cal State Northridge

7

Critical Values

- ◆ These represent the point at which we decide to reject null hypothesis.
- ◆ e.g. We might decide to reject null when $(p|null) \leq .05$.
 - In the null distribution there is some value with $p = .05$
 - We reject when we exceed that value.
 - That value is called the ***critical value***.

Psy 320 - Cal State Northridge

8

One- and Two-Tailed Tests

- ◆ Two-tailed test rejects null when obtained value too extreme in *either* direction
 - Decide on this before collecting data.
- ◆ One-tailed test rejects null if obtained value is too low (or too high)
 - We only set aside one direction for rejection.

Psy 320 - Cal State Northridge

9

One- & Two-Tailed Example

◆ One-tailed test

- Reject null if IQPLUS group shows an increase in IQ
 - Probably wouldn't expect a reduction and therefore no point guarding against it.

◆ Two-tailed test

- Reject null if IQPLUS group has a mean that is substantially higher or lower.

Psy 320 - Cal State Northridge

10

What Is Power?

- ◆ Probability of rejecting a false H_0
- ◆ Probability that you'll find difference that's really there
- ◆ $1 - \beta$, where β = probability of Type II error

Psy 320 - Cal State Northridge

11

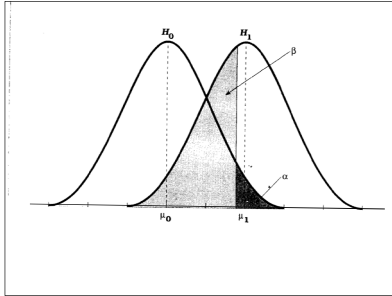
What Controls Power?

- ◆ The significance level (α)
- ◆ True difference between null and alternative hypotheses
 - $\mu_1 - \mu_2$
- ◆ Sample size
- ◆ Population variance
- ◆ The particular test being used

Psy 320 - Cal State Northridge

12

Distributions Under μ_1 and μ_0



Psy 320 - Cal State Northridge

13

Effect Size

- ◆ The degree to which the null is false
 - Depends on distance between μ_1 and μ_2
 - Also depends on standard error (of mean or of difference between means)

$$d = \frac{\mu_1 - \mu_0}{\sigma} \text{ or } \hat{d} = \frac{\bar{X}_1 - \bar{X}_0}{s} \text{ if we can assume that } \bar{X}_1, \bar{X}_0 \text{ and } s \text{ are adequate estimators or their population counterparts.}$$

Psy 320 - Cal State Northridge

14

What happened to n ?

- ◆ It doesn't relate to how different the two population means are.
- ◆ It controls power, but not effect size.
- ◆ We will add it in later.

Psy 320 - Cal State Northridge

15

Estimating Effect Size

- ◆ Judge your effect size by:
 - Past research
 - What you consider important
 - Cohen's conventions

Effect Size	d
Small	.20
Medium	.50
large	.80

Psy 320 - Cal State Northridge

16

Combining Effect Size and n

- ◆ We put them together and then evaluate power from the result.
- ◆ General formula for Delta

$$\delta = d[f(n)]$$

- where $f(n)$ is some function of n that will depend on the type of design

Psy 320 - Cal State Northridge

17

Power for One-Sample or Related samples t

- ◆ First calculate delta with:

$$\delta = \hat{d}\sqrt{n}$$

- where n = size of sample, and δ and γ as above

- ◆ Look power up in table using δ and significance level (α)

Psy 320 - Cal State Northridge

18

Power for Single Sample IQPLUS Study

- ◆ One sample z and t
 - Compared IQPLUS group with population mean = 100, sigma = 10
 - Used 25 subjects
 - We got a sample mean of 106 and $s = 7.78$

$$d = \frac{\bar{X}_1 - \mu_0}{\sigma} = \frac{106 - 100}{10} = \frac{6}{10} = 0.60 \text{ or}$$

$$\hat{d} = \frac{\bar{X}_1 - \mu_0}{s} = \frac{106 - 100}{7.78} = \frac{6}{7.78} = 0.77$$

Psy 320 - Cal State Northridge

19

IQPLUS

- ◆ Assuming we don't know sigma
 - $\hat{d} = 0.77$
 - $n = 25$
 - $\delta = .77\sqrt{25} = .77 * 5 = 3.85$
 - We are testing at $\alpha = .05$
 - Use Appendix D.5

Psy 320 - Cal State Northridge

20

Appendix D.5

δ	Alpha (α)			
	0.10	0.05	0.02	0.01
1.00	0.26	0.17	0.09	0.06
2.00	0.64	0.52	0.37	0.28
2.40	0.78	0.67	0.53	0.43
2.50	0.80	0.71	0.57	0.47
2.80	0.88	0.80	0.68	0.59
3.00	0.91	0.85	0.75	0.66
3.80	0.98	0.97	0.93	0.89
3.90	0.99	0.97	0.94	0.91
4.00	0.99	0.98	0.95	0.92

This table is severely abbreviated.

Power for $\delta = 3.85, \alpha = .05$

Psy 320 - Cal State Northridge

21

Conclusions

- ◆ If we can trust our estimates in the IQPLUS study then if this study were run repeatedly, 97% of the time the result would be significant.

How Many Subjects Do I Really Need (Single/Related Sample(s))?

- ◆ Run calculations backward
 - Start with anticipated effect size (\hat{d})
 - Determine δ required for power = .80.
 - Why .80?
 - Calculate n
- ◆ What if we wanted to rerun the IQPLUS study, and wanted power = .80?

Calculating n

- ◆ We estimated $\hat{d} = .77$
- ◆ Complete Appendix D.5 shows we need $\delta = 2.80$
- ◆ Calculations on next slide

IQPLUS n

Since $\delta = \hat{d}\sqrt{n}$, $n = \left(\frac{\delta}{\hat{d}}\right)^2$

$\frac{\delta}{\hat{d}} = \frac{\cancel{\delta}\sqrt{n}}{\cancel{\delta}}$ (divide both sides by \hat{d})

$\left(\frac{\delta}{\hat{d}}\right)^2 = (\sqrt{n})^2$ (square both sides)

$\left(\frac{\delta}{\hat{d}}\right)^2 = n = \left(\frac{2.80}{.77}\right)^2 = 13.22 \approx 14$ subjects (where we had 25)

Power for Two Independent Groups

- ◆ What changes from preceding?
 - Effect size deals with two sample means
 - Take into account both values of n
- ◆ Effect size

$$d = \frac{\mu_1 - \mu_2}{\sigma} \text{ or } \hat{d} = \frac{\bar{X}_1 - \bar{X}_2}{S_{pooled}}$$

Estimating d

- ◆ We could calculate d directly if we knew populations.
- ◆ We could estimate from previous data.
- ◆ Here we will calculate using Violent Video Games example

Example: Violent Videos Games

- ◆ Two independent randomly selected/assigned groups
 - **GTA** (violent: 8 subjects) VS. **NBA 2K7** (non-violent: 10 subjects)
 - We want to compare mean number of aggressive behaviors following game play
 - GTA had a mean of 10.25 and $s = 1.669$
 - NBA 2K7 had a mean of 8.4 and $s = 1.647$
 - $s^2_{\text{pooled}} = 2.745$, so $s_{\text{pooled}} = 1.657$

Psy 320 - Cal State Northridge

28

Two Independent Groups

- ◆ Then calculate δ from effect size \hat{d}

$$\delta = \hat{d} \sqrt{\frac{n}{2}}$$

Note: The above formula assumes that the 2 groups have equal n

Psy 320 - Cal State Northridge

29

Two Independent Groups

- ◆ Our data do not have equal n, but **let's pretend they do** for a moment (both 10)
- ◆ For our data

$$\hat{d} = \frac{10.25 - 8.4}{1.657} = \frac{1.85}{1.657} = 1.116$$
$$\delta = \hat{d} \sqrt{\frac{n}{2}} = 1.116 \sqrt{\frac{10}{2}} = 2.495 \approx 2.5$$

Psy 320 - Cal State Northridge

30

Appendix D.5

δ	Alpha (α)			
	0.10	0.05	0.02	0.01
1.00	0.26	0.17	0.09	0.06
2.00	0.64	0.52	0.37	0.28
2.40	0.78	0.67	0.53	0.43
2.50	0.80	0.71	0.57	0.47
2.80	0.88	0.80	0.68	0.59
3.00	0.91	0.85	0.75	0.66
3.80	0.98	0.97	0.93	0.89
3.90	0.99	0.97	0.94	0.91
4.00	0.99	0.98	0.95	0.92

This table is severely abbreviated.

*Power for $\delta = 2.5$, $\alpha = .05$
Estimate = .71*

Psy 320 - Cal State Northridge

31

Conclusions

- ◆ If we had equal n and we can trust our estimates in the violent video game study then if this study were run repeatedly, 71% of the time the result would be significant.

Psy 320 - Cal State Northridge

32

Unequal Sample Sizes

- ◆ With unequal samples use harmonic mean of sample sizes

$$\bar{X}_h = \frac{k}{\sum \frac{1}{n_i}}$$

- ◆ Where k is the number of groups (i.e. 2), n_i is each group size
- ◆ Standard arithmetic average will work well if n 's are close.

Psy 320 - Cal State Northridge

33

Two Independent Groups

- ◆ Our data **do not** have equal n, so we need to find the harmonic mean
- ◆ For our data

$$\bar{X}_h = \frac{k}{\sum \frac{1}{n_i}} = \frac{2}{\frac{1}{8} + \frac{1}{10}} = \frac{2}{.225} = 8.889$$

Psy 320 - Cal State Northridge

34

Two Independent Groups

- ◆ Our data **do not** have equal n, so...
- ◆ For our data

$$\hat{d} = \frac{10.25 - 8.4}{1.657} = \frac{1.85}{1.657} = 1.116$$

$$\delta = \hat{d} \sqrt{\frac{\bar{X}_h}{2}} = 1.116 \sqrt{\frac{8.889}{2}} = 2.353 \approx 2.4$$

Psy 320 - Cal State Northridge

35

Appendix D.5

δ	Alpha (α)			
	0.10	0.05	0.02	0.01
1.00	0.26	0.17	0.09	0.06
2.00	0.64	0.52	0.37	0.28
2.40	0.78	0.67	0.53	0.43
2.50	0.80	0.71	0.57	0.47
2.80	0.88	0.80	0.68	0.59
3.00	0.91	0.85	0.75	0.66
3.80	0.98	0.97	0.93	0.89
3.90	0.99	0.97	0.94	0.91
4.00	0.99	0.98	0.95	0.92

This table is severely abbreviated.

Power for $\delta = 2.4$, $\alpha = .05$
Estimate = .67

Psy 320 - Cal State Northridge

36

Conclusions

- ◆ If we can trust our estimates in the violent video game study then if this study were run repeatedly, 67% of the time the result would be significant.

Psy 320 - Cal State Northridge

37

How Many Subjects Do I Really Need (Independent Samples)?

- ◆ Run calculations backward
 - Start with anticipated effect size (\hat{d})
 - Determine δ required for power = .80.
 - Why .80?
 - Calculate n
- ◆ What if we wanted to rerun the violent video game study, and wanted power = .80?

Psy 320 - Cal State Northridge

38

Calculating n

- ◆ We estimated $\hat{d} = 1.116$
- ◆ Complete Appendix E.5 shows we need $\delta = 2.80$
- ◆ Calculations on next slide

Psy 320 - Cal State Northridge

39

Violent Video Games n

$$\text{If } \delta = \hat{d} \sqrt{\frac{n}{2}} \text{ then } n = 2 * \left(\frac{\delta}{\hat{d}} \right)^2$$

$$\frac{\delta}{\hat{d}} = \sqrt{\frac{n}{2}} \text{ (divide both sides by } \hat{d} \text{)}$$

$$\left(\frac{\delta}{\hat{d}} \right)^2 = \left(\sqrt{\frac{n}{2}} \right)^2 \text{ (square both sides)}$$

$$\left(\frac{\delta}{\hat{d}} \right)^2 = \frac{n}{2} \text{ (multiply both sides by 2)}$$

$$2 * \left(\frac{\delta}{\hat{d}} \right)^2 = n = 2 * \left(\frac{2.80}{1.116} \right)^2 12.59 \approx 13 \text{ subjects/group}$$

Psy 320 - Cal State Northridge

40
