Effect Size, Power, and Sample Size

ERSH 8310

Lecture 7

September 30, 2009

Today’s Class

- Effect Size
- Power
- Sample Size

EFFECT SIZE

Descriptive Measures of Effect Size

- The report of any study should include:
  - A description of the pattern of means and their variability
    - Usually the standard deviations of the scores

- It is also useful to have an overall measure of the magnitude of the effect that incorporates all the groups at once
  - Called the effect size

Descriptive Measures of Effect Size

- An effect-size measure is a quantity that measures the size of an effect as it exists in the population

- It is independent of certain details of the experiment
  - i.e., sizes of the samples used

- Descriptive measures of effect size can generally be divided into two types
  - Those that describe differences in means relative to the study's variability
  - Those that look at how much of the variability can be attributed to the treatment conditions

Differences Relative to the Variability of the Observations

- The standardized difference between means, for example, between two groups a_1 and a_2 is:

$$d_{12} = \frac{\overline{Y}_1 - \overline{Y}_2}{s_{12}}$$

- Where:

$$s_{12} = \sqrt{\frac{SS_1 + SS_2}{df_1 + df_2}}$$

The Proportion of Variability Accounted for by an Effect

- The square of the correlation ratio is defined as

$$R^2 = \frac{SS_A}{SS_T} = \frac{SS_T - SS_{S/A}}{SS_T}$$

- An equivalent formula, based on the F statistic, is

$$R^2 = \frac{(a-1)F}{(a-1)F + a(n-1)}$$

The Proportion of Variability Accounted for by an Effect

- In a very influential book on power analysis, Cohen (1988) defined standards for interpreting effect sizes:

- A small effect with $R^2 = .01$ (or $d = 0.25$)

- A medium effect with $R^2 = .06$ (or $d = 0.5$)

- A large effect with $R^2 = .15$ (or $d = 0.8$)

Effect Sizes in the Population

- The most popular measure of treatment magnitude in the population is an index known as the ω^2
- For the single-factor design,

$$\omega^2 = \frac{\sigma^2_{\text{Total}} - \sigma^2_{\text{Error}}}{\sigma^2_{\text{Total}}} = \frac{\sigma^2_A}{\sigma^2_A + \sigma^2_{S/A}}$$

- Where:

$$\sigma^2_A = \sum_{j=1}^{a} \frac{\alpha^2_j}{a}$$

More Omega-Squared

- Equivalently, from the F statistic,

$$\hat{\omega}^2 = \frac{SS_A - (a-1)MS_{S/A}}{SS_{Total} + MS_{S/A}}$$

- It is possible to obtain a negative value of $\hat{\omega}^2$ for F < 1
- Note that $\hat{\omega}^2$ is unaffected by small sample sizes

Eta Squared Example

- Recall our vigilance task experiment

ANOVA

ERRORS

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3314.250	3	1104.750	7.343	.005
Within Groups	1805.500	12	150.458		
Total	5119.750	15			

- The effect size is:

$$\hat{\omega}^2 = \frac{SS_A - (a-1)MS_{S/A}}{SS_{Total} + MS_{S/A}} = \frac{3314.25 - (4-1)150.458}{5119.750 + 150.458} = 0.54$$

Effect Sizes for Contrasts

- Two different effect sizes are available for a contrast
  - Complete omega-squared
  - Partial omega-squared
- Recall our contrast:

$$\psi = \sum_{j=1}^{a} c_j \mu_j$$

- With:

$$\sigma_\psi^2 = \frac{\psi^2}{2\sum_{j=1}^{a} c_j^2}$$

Vigilance Contrast

- Consider a contrast in our vigilance task where we compared the mean of the 4 hour group with the mean of the 28 hour group
  - Coefficients 1, 0, 0, -1

$$F_\psi = -4.064^2 = 16.52$$

Contrast Coefficients

Contrast	HOURS			
	4	12	20	28
1	1	0	0	-1

Contrast Tests

		Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
ERRORS	Assume equal variances	1	-35.2500	8.67347	-4.064	12	.002
	Does not assume equal variances	1	-35.2500	7.77684	-4.533	4.496	.008

Complete Omega Squared

- The complete ω^2 is given by:

$$\omega_{\psi}^{2} = \frac{\sigma_{\psi}^{2}}{\sigma_{A}^{2} + \sigma_{\text{Error}}^{2}}$$

- The complete ω^2 looks at the proportion of total variance accounted for the contrast
- This can be obtained from the contrast F test:

$$\hat{\omega}_{\psi}^{2} = \frac{F_{\psi} - 1}{(a-1)(F_{A} - 1) + an}$$

Complete Omega Squared: Vigilance Style

ANOVA

ERRORS

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3314.250	3	1104.750	7.343	.005
Within Groups	1805.500	12	150.458		
Total	5119.750	15			

$$F_{\psi} = -4.064^{2} = 16.52$$

$$\hat{\omega}_{\psi}^{2} = \frac{F_{\psi} - 1}{(a-1)(F_{A} - 1) + an} = \frac{16.52 - 1}{(4-1)(7.343-1) + 4*4} = 0.443$$

The Partial Omega Squared

- The partial ω^2 is given by:

$$\omega^2_{(\psi)} = \frac{\sigma^2_\psi}{\sigma^2_\psi + \sigma^2_{\text{Error}}}$$

- It looks at the proportion of variance relative to the contrast itself and the error
- It can also be obtained by the F test for the contrast:

$$\hat{\omega}^2_{(\psi)} = \frac{F_\psi - 1}{F_\psi - 1 + 2n}$$

Partial Omega Squared: Vigilance Style

ANOVA

ERRORS

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3314.250	3	1104.750	7.343	.005
Within Groups	1805.500	12	150.458		
Total	5119.750	15			

$$F_{\psi} = -4.064^2 = 16.52$$

$$\hat{\omega}^2_{(\psi)} = \frac{F_{\psi} - 1}{F_{\psi} - 1 + 2n} = \frac{16.52 - 1}{16.52 - 1 + 2 * 4} = 0.66$$

Effect Size Recommendations

- If you wish to summarize the effects for the experiment as a whole, the authors recommend that you use $\hat{\omega}^2$
- When you are discussing the difference between two groups, you might consider reporting d
- If appropriate, use the partial omega squared

POWER AND SAMPLE SIZE

Power and Sample Size

- The power of the test is

  Power = Prob(Reject H_0 given H_0 is false) = 1 - β.

- We may use power to find the sample size in the planning stage
  - To find if the existing design would be likely to detect a particular alternative

Determinants of Power

- We can control the magnitude of Type I error through our choice of a rejection region for the F distribution (i.e., the α level)

- Type II error control and power is not simple

- Power depends on three determinants:
    1. The significance level α
    2. The size of the treatment effects ω^2
    3. The sample size n

Determinants of Power

- Controlling power is important because power reflects the degree to which we can detect the treatment differences we expect
  - …and the chances that others will be able to duplicate our findings when they attempt to repeat the experiment

- Research in the behavioral sciences is lacking power
  - Average power of about .50 for detecting medium effects

Determinants of Power

- The power of an experiment is determined by the interplay of three factors, α, ω^2, and n

- From a practical point of view, sample size is normally used to control power

Steps in Determining Sample Size

(1) Determine the experimental design

(2) Decide on the null hypothesis and α (e.g., $\alpha = .05$)

(3) Find ω^2 from the alternative and null hypotheses (e.g., $\omega^2 = .06$)

(4) Select the desired power = $1 - \beta$ (e.g., power = .80)

(5) Find the sample size from the combined information

Finding the Size of the Target Effect

- From the hypothesized results with α_j and σ_{error} (see p. 171), the effect size can be calculated

- Another way is to use a model experiment results (e.g., α, F, and n)

Finding the Sample Size

- Using α, ω^2, and $1-\beta$, as well as a, we may use the following three ways to find the sample size:

1. Table 8.1 (p. 173)
2. Appendix A.7 (pp. 590-595)
3. GPOWER (Erdfelder, Faul, & Buchner, 1996)

Table 8.1

Table 8.1: Sample size needed to achieve a power of .60, .80, and .90 in a test at $\alpha = .05$ for studies with from 2 to 8 groups and effect sizes ω^2 from .01 to .15. These values were calculated by the program GPOWER *(Erdfelder et al., 1996).*

ω^2	$a = 2$	$a = 3$	$a = 4$	$a = 5$	$a = 6$	$a = 7$	$a = 8$
				Power = .60			
.01	244	207	179	158	143	131	122
.02	122	103	89	79	72	66	61
.03	81	69	59	53	48	44	41
.04	60	51	44	39	36	33	31
.05	48	41	36	32	29	26	25
.06	40	34	30	26	24	22	20
.08	30	25	22	20	18	17	15
.10	24	20	18	16	14	13	12
.12	19	17	15	13	12	11	10
.15	15	13	12	10	10	9	8
				Power = .80			
.01	390	319	271	238	213	194	179
.02	194	159	135	118	106	97	89
.03	128	105	90	79	71	64	59
.04	96	79	67	59	53	48	44
.05	76	63	53	47	42	38	35
.06	63	52	44	39	35	32	30
.08	47	38	33	29	26	24	22
.10	37	30	26	23	21	19	18
.12	30	25	21	19	17	16	15
.15	24	20	17	15	14	12	12
				Power = .90			
.01	522	419	352	306	273	248	228
.02	259	208	175	153	136	123	113
.03	171	138	116	101	90	82	75
.04	127	103	86	75	67	61	56
.05	101	82	69	60	54	49	45
.06	84	68	57	50	44	40	37
.08	62	50	42	37	33	30	28
.10	49	39	33	29	26	24	22
.12	40	32	27	24	22	20	18
.15	31	25	22	19	17	16	14

Using the Power Charts

- Pearson and Hartley (1951, 1972) have constructed some helpful charts from which we can estimate a sample size for a particular degree of power

- Note that power, significance level, effect size, and sample size are interrelated and that fixing any two will determine the third

Non-centrality Parameter

- The power charts make use of a quantity known as the noncentrality parameter:

$$\varphi_A = \frac{\sqrt{\dfrac{\sum_{j=1}^{a}(\mu_i - \mu_T)^2}{a\sqrt{n}}}}{\sigma^2_{S/A}}$$

- With ω^2, we may derive:

$$n = \varphi^2 \frac{1-\omega^2}{\omega^2}$$

Sample Sizes for Contrasts

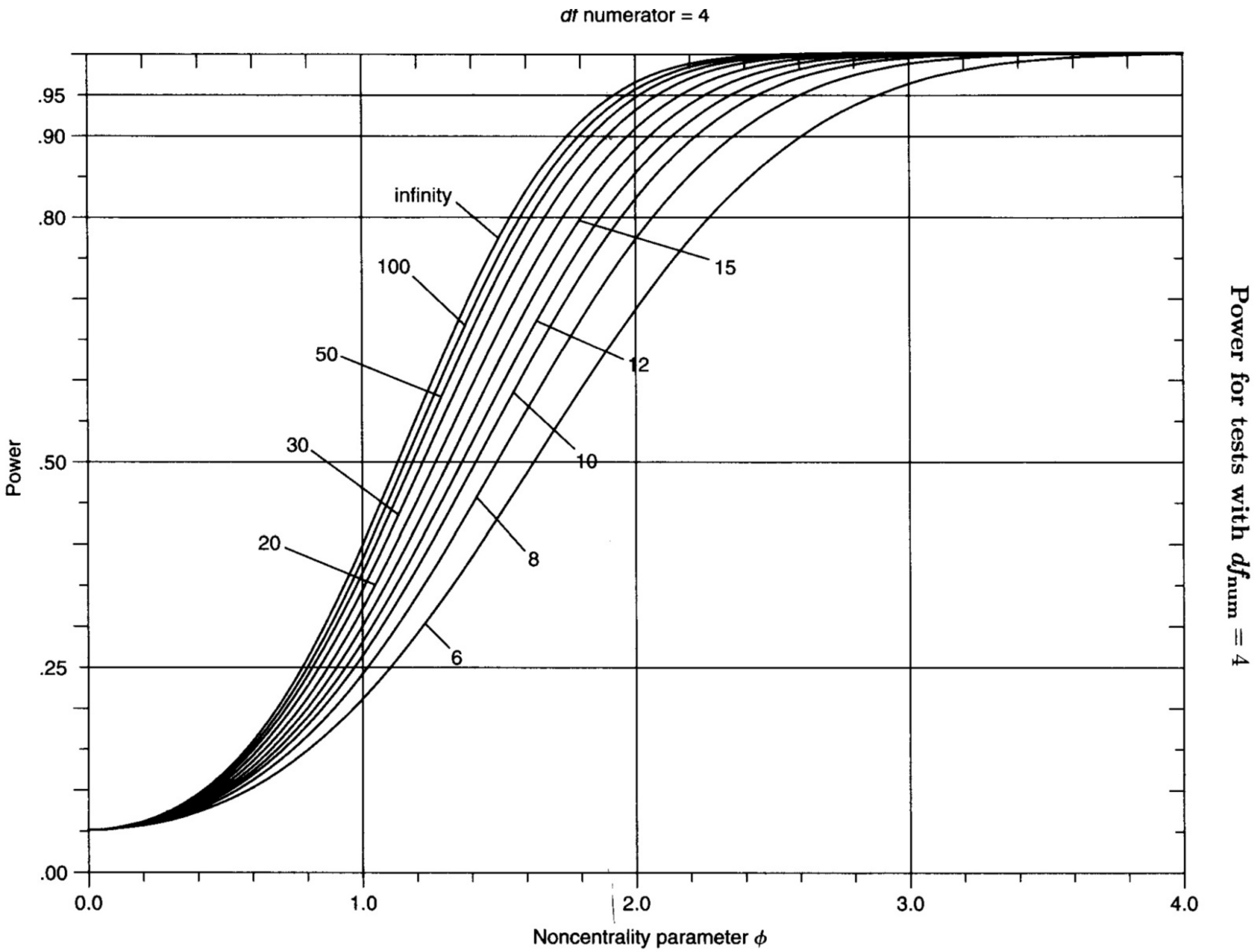
- Use $\omega^2_{(y)}$ and $df_{num} = 1$ on the power chart

Comments

- Interestingly, methodologists seem to agree that a power of about .80 is a reasonable value for researchers in the behavioral sciences

- When $\alpha = .05$, the ratio of Type II error to Type I error is 4:1.

Reading a Power Curve



Other Methods for Power

- Use a software package
  - GPOWER (free)
  - SAS
  - SPSS

- Use Monte Carlo simulations
  - In SPSS
  - In SAS
  - In Mplus

DETERMINING POWER

Calculating the Power of an Experiment

- We may obtain

$$\varphi = \sqrt{\frac{n\omega^2}{1 - \omega^2}}$$

and use power chart with df_{num} and df_{denom}

- Always use a power analysis before implementing an experiment

From our Previous Example

- Previously, we found:

$$\hat{\omega}^2 = \frac{SS_A - (a-1)MS_{S/A}}{SS_{Total} + MS_{S/A}} = \frac{3314.25 - (4-1)150.458}{5119.750 + 150.458} = 0.54$$

- Therefore:

$$\varphi = \sqrt{\frac{n\omega^2}{1 - \omega^2}} = \sqrt{\frac{4 * 0.54^2}{1 - 0.54^2}} = 1.28$$

- Finally, using the power chart on p. 592, we find power of about 0.3
  - Is that good?

Wrapping Up…

- Effect sizes are useful for determining the size of a treatment effect
  - They also play a role in power calculations

- Power is an important concept in that underpowered studies do not have a great chance of finding significance

- Use an estimate of power to determine how big of a sample is needed
  - This helps in experiment planning

Up Next…

- Now:
  - Midterm topic discussion
- In Lab:
  - How to calculate effect sizes in SPSS
  - Midterm review
- Next week
  - Midterm