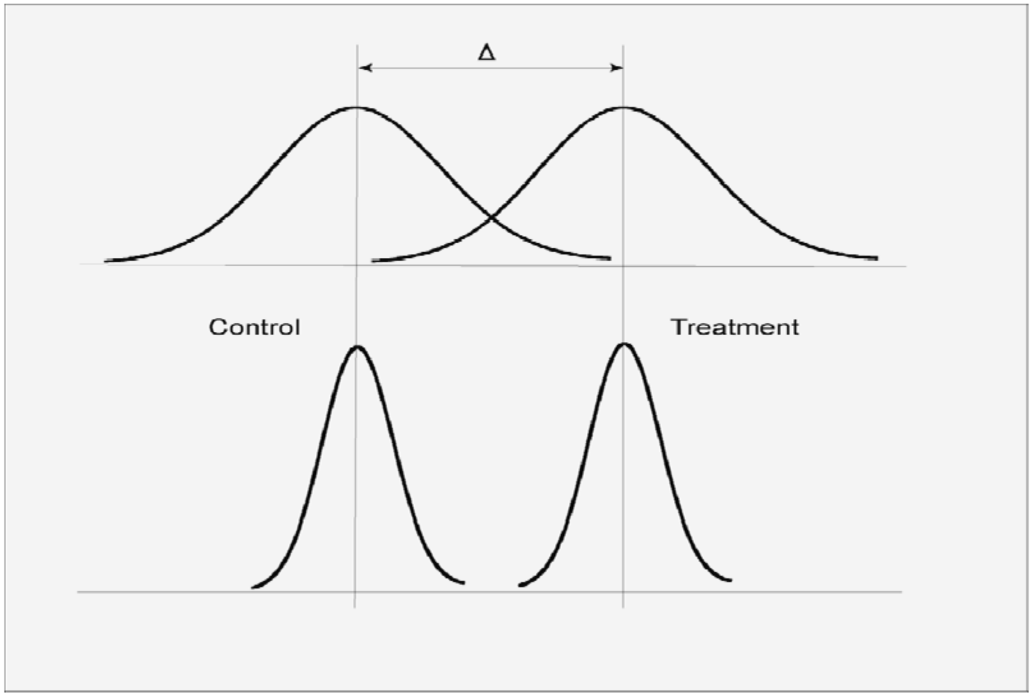
Understanding and Quantifying EFFECT SIZES



Karabi Nandy, Ph.d.

Assistant Adjunct Professor

Translational Sciences Section, School of Nursing

Department of Biostatistics, School of Public Health,

University of California Los Angeles (UCLA)

Objective

Effect size comes up in the context of

- sample size calculations for proposals
- reporting results from pilot studies.

By the end of this talk, you will be able to calculate, interpret and report effect sizes in your work.

Outline

- Why are Effect Sizes (ES) important?
- Types of Effect Sizes
- Quantifying Magnitudes of Effect Sizes
- Calculating Effect Sizes
- Use of Softwares for Calculating Effect Sizes
- Specific Guidelines for Reporting Effect Sizes

Definition

- "Effect" - A change or changed state occurring as a direct result of action by somebody or something (Encarta, 2009)
- "Size" - The degree of something in terms of how big or small it is
- 'Effect size' is simply a way of quantifying the size of the difference between two groups.

- It is particularly valuable for quantifying the effectiveness of a particular intervention, relative to some comparison. It allows us to move beyond the simplistic, 'Does it work or not?' to the far more sophisticated, 'How well does it work in a range of contexts?‘

Why is Effect Size Important

- *Knowing the magnitude of an effect allows us to ascertain the practical significance of statistical significance*
  - Can always reach statistical significance if there is a large enough sample size, unless the effect size is 0.
  - Even a large effect may not be statistically significant if the sample size is too small.

Why is Effect Size Important

- *Practical Significance*
  - Even a statistically significant treatment difference may not be practically important if the effect size is too small.
  - However, there could still be practical importance even for small effect sizes, especially in cases where cost and ease make it easy to be implemented on a large scale.

Why is Effect Size Important

- *Sample Size Calculation for Studies*

ES plays a direct role in sample size calculations for any study. It is connected to the power of a test, the level of significance α and sample size (n).

- $\uparrow ES = \uparrow power$
- $\uparrow \alpha = \uparrow power$
- $\uparrow N = \uparrow power$ or $\uparrow reliability = \uparrow power$

- Given any 3 quantities (power, ES, α, n), we can find the 4^{th}.

Why is Effect Size Important

- *Meta-Analysis*
  - pooling information from many studies to verify results of past research and inform future studies.
  - ES is computed in each study and the findings are pooled together to draw overall inferences.

Types of Effect Sizes

- Mean Differences between Groups
  - Effect Size: Cohen's d
- Correlation/Regression
  - Effect Size: Pearson's r and R^2
  - Effect Size: Cohen's f^2
- Contingency tables
  - Effect Size: Odds Ratio or Relative Risk (association between binary variables)

Types of Effect Sizes

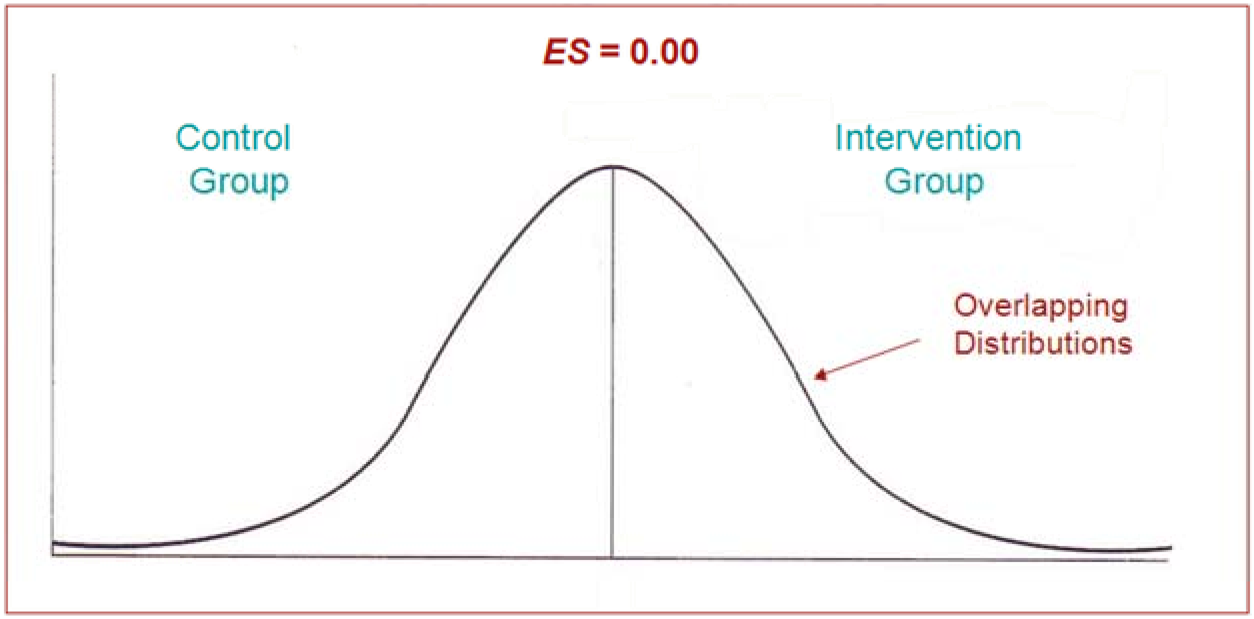
- ANOVA or GLMs
  - Effect Size: Eta-squared
  - Effect Size: Omega squared
  - Effect Size: Intraclass correlation (rater equality)
- Chi-square tests
  - Effect Size: Phi (2 binary variables)
  - Effect Size: Cramer's Phi or V (categorical variables)

Cohen's d

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s_{\text{pooled}}} \quad \text{where} \quad s_{\text{pooled}} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2}}$$

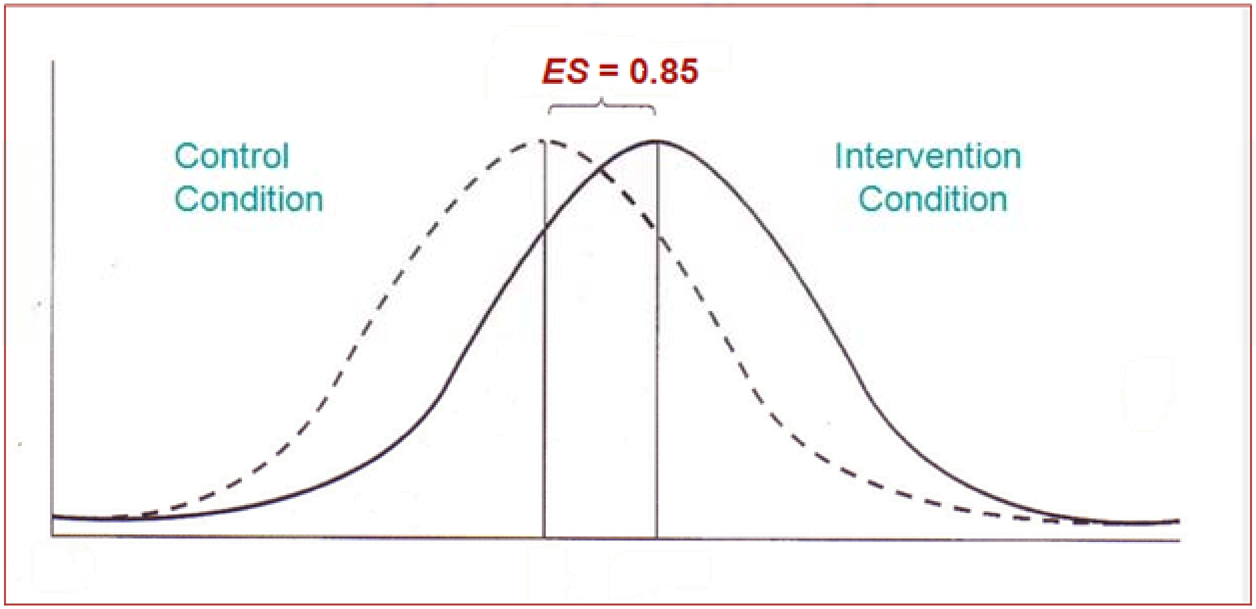
- Standardizes ES of the difference between two means
- Used for two sample independent t-tests
- d ranges from -∞ to +∞
- interpretation: the difference between the mean values is "d" standard deviations, Cohen (1988)

ES Example 1



- <u>*ES* = 0.00</u> means that the *average* treatment participant outperformed <u>50%</u> of the control participants

ES Example 2



- ES = 0.85 means that the *average* treatment participant outperformed 80% of the control participants

General Guidelines

In general, ≤ 0.20 is a small effect size, 0.50 is a moderate effect size and ≥ 0.80 is a large effect size (Cohen, 1992)

	d- standardized mean difference	Percentage of variance explained
• Small	.20	1%
• Moderate	.50	10%
• Large	.80	25%

Cohen's d

- Special Cases
  - For small sample sizes use Hedge's G

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s_{\text{pooled}}} \quad \text{where } s_{\text{pooled}} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

  - For unequal group variances, use Glass's Δ
    - uses sample sd of the control group only so that effect sizes would not differ under equal means and unequal variances

$$\Delta = \frac{\overline{x_1} - \overline{x_2}}{s_2}$$

Pearson's r

$$r_{xy} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\Sigma_{i=1}^{n}(y_i-\bar{y})^2}} \quad -1 \le r_{xy} \le 1$$

- used in the context of correlation…measuring association between 2 variables
- Interpretation: For every 1-unit standard deviation change in x, there is a "r-unit" standard deviation change in y

Pearson's R^2

- used in the context of regression…measuring how well a regression line fits to a given data
- R - linear association between 2 continuous variables
- R^2 (Coefficient of Determination) – proportion of shared variability between 2 or more variables
- Interpretation: "R^2 *100%" is percent variance of the outcome y that can be explained by the linear regression model (i.e. indicates how well the linear regression line fits the data)

Odds Ratio

- Used in the context of binary/categorical outcomes
- Odds of being in one group (eg. success) relative to the odds of being in a different group (eg. failure)
- OR ranges from 0 to ∞
- OR>1 indicates an increase in odds relative to the reference group
- OR < 1 indicates a decrease in odds relative to the reference group

Relative Risk

$$RR = \frac{a/(a+b)}{c/(c+d)} \approx \frac{ad}{bc} = OR$$

Risk	Disease status	
	Present	Absent
Smk	a	b
Non-smk	c	d

- RR measures the risk of an event relative to an independent variable
- For small probabilities, the relative risk is approximately equal to the odds ratio
- Interpretation: If RR > 1 then the risk of disease X among smokers is "RR" times the risk of disease X among non-smokers (vice versa if RR < 1)

Eta-Squared (η^2) and partial Eta-Squared ($\eta_p{}^2$)

$$\eta^2 = \frac{SS_{treatment}}{SS_{total}} \quad \eta_p^2 = \frac{SS_{treatment}}{SS_{treatment} + SS_{error}} \quad 0 \leq \eta^2 \leq 1$$

- Used with ANOVA family and GLMs
- Measures the degree of association in the sample
- Standardizes Effect Sizes of the shared variance between a continuous outcome and categorical predictors
- Partial eta-squared is the proportion of the total variability attributable to a given factor.

Eta-Squared (η^2) and partial Eta-Squared ($\eta_p^{\ 2}$)

- Interpretation: "η^2 *100%" is percent of the variance in y explained by the variance in x (similar to the R^2 interpretation for linear regression (Dattalo, 2008)).
- η^2 is biased and on average overestimates the variance explained in the population, but decreases as the sample size gets larger.
- Caution: these effect sizes depend on the number and magnitude of the other effects

Cohen's f²

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

- Used in multiple linear regression, $R^2 = \eta^2$
- Standardized effect size is the proportion of explained variance over unexplained variance
- Estimate is biased and overestimates the effect size for ANOVA (unbiased estimate is Omega-Squared)

Omega-Squared

$$\hat{\omega}^2 = \frac{SS_{treatment} - df_{treatment} * MS_{error}}{SS_{total} + MS_{error}}$$

- Estimates the proportion of variance in the *population* that is explained by the treatment
- $\hat{\omega}^2$ is always smaller than η^2 or η_p^2 since Omega measures the population variance and Eta measures the sample variance

Intraclass Correlation

- ICC is used to measure inter-rater reliability for two or more raters. It may also be used to assess test-retest reliability. ICC may be conceptualized as the ratio of between-groups variance to total variance.

$$ICC = \frac{MS_{Treatment} - MS_{Error}}{MS_{Treatment} + (n-1)MS_{Error}}$$

- Can be used in ANOVA
- Similar interpretation to Omega-Squared

Phi

$$\phi = \sqrt{\frac{\chi^2}{n}}$$

- Used for crosstabs or for chi-square tests (equality of proportions or tests of independence between 2 binary variables)… φ= 0 indicates independence
- Phi are related to correlation and Cohen's d (for 2 binary variables)
- Interpreted like Pearson's r and R^2

Cramer's Phi or V

- Cramer's Phi (Cramer's V) can be used with categorical variables with more than 2 categories ($m \geq 2$) (R x C tables)

$$\varphi_c = \sqrt{\frac{\chi^2}{N(k-1)}} \quad ; k = \min(R, C)$$

- measures the inter-correlation of the variables, but is biased since it increases with the number of cells. Increase in R and C will indicate a strong association, which is just an artifact of the type of variable used.

Magnitude of Effect Summary Table

Effect Size	Small	Medium	Large
r	0.10	0.30	0.50
r^2	0.01	0.09	0.25
η^2	0.01	0.06	0.14
R^2	0.01	0.06	0.14
Cohen's d	0.20	0.50	0.80
φ / Cramer's V	0.10	0.30	0.50
Cohen's f^2	0.02	0.15	0.35
OR	1.44	2.47	4.25

Effect Size Conversions

Effect Size	Converted to Cohen's d
Correlation	$d = 2r \Big/ \sqrt{1 - r^2}$
Chi-Square df =1 df > 1	$d = \sqrt{4\chi^2/(N - \chi^2)}$ $d = \sqrt{\dfrac{4\chi^2}{N}}$
Odds Ratio (Chinn, 2000)	$d = \dfrac{\ln(OR)}{1.81}$

Software

- Calculate effect size
  - http://www.uccs.edu/~faculty/lbecker/
  - http://faculty.vassar.edu/lowry/newcs.html
  - Statistical softwares such as SAS, R, STATA, SPSS can calculate most of the standard Effect Sizes

Calculating Cohen’s d – online calculator http://www.uccs.edu/~faculty/lbecker/

Effect Size Calculators

Calculate Cohen's *d* and the effect-size correlation, $r_{Y\lambda}$, using --

- **means and standard deviations**
- **independent groups *t* test values and *df***

For a discussion of these effect size measures see Effect Size Lecture Notes

---

Calculate *d* and *r* using means and standard deviations

Calculate the value of Cohen's *d* and the effect-size correlation, $r_{Y\lambda}$, using the means and standard deviations of two groups (treatment and control).

Cohen's $d = M_1 - M_2 / s_{pooled}$
where s_{pooled} = �[(s 1� ' s 2�) / 2]

$r_{Y\lambda}$ = d / �(d� + 4)

Note: d and $r_{Y\lambda}$ are positive if the mean difference is in the predicted direction.

Group 1		Group 2	
M_1	6.25	M_2	7.36
SD_1	3.27	SD_2	4.65

Compute Reset

Cohen's *d*	effect-size *r*

top

Calculating Cohen's d – online calculator http://www.uccs.edu/~faculty/lbecker/

Effect Size Calculators

Calculate Cohen's *d* and the effect-size correlation, $r_{Y\lambda}$, using --

- **means and standard deviations**
- **independent groups *t* test values and *df***

For a discussion of these effect size measures see Effect Size Lecture Notes

Calculate *d* and *r* using means and standard deviations

Calculate the value of Cohen's *d* and the effect-size correlation, $r_{Y\lambda}$, using the means and standard deviations of two groups (treatment and control).

Cohen's $d = M_1 - M_2 / s_{pooled}$

where s_{pooled} = �[(s 1�+ s 2�) / 2]

$r_{Y\lambda}$ = d / �(d� + 4)

Note: d and $r_{Y\lambda}$ are positive if the mean difference is in the predicted direction.

	Group 1		Group 2
M_1	6.25	M_2	7.36
SD_1	3.27	SD_2	4.65
	Compute		Reset
	Cohen's *d*		effect-size *r*
	-0.27614:		-0.13677:

top

Calculating Cramer's Phi – online calculator http://www.uccs.edu/~faculty/lbecker/

Chi-Square, Cramer's V, and Lambda

For a Rows by Columns Contingency Table

For a contingency table containing up to 5 rows and 5 columns, this unit will:

- ~ perform a chi-square analysis [the logic and computational details of chi-square tests are described in Chapter 8 of Concepts and Applications];
- ~ calculate Cramer's V, which is a measure of the strength of association among the levels of the row and column variables [for a 2x2 table, Cramer's V is equal to the absolute value of the phi coefficient];
- ~ and calculate the two asymmetrical versions of lambda, the Goodman- Kruskal index of predictive association, along with some other measures relevant to categorical prediction. [Click here for a brief explanation of lambda.]

To begin, select the number of rows and the number of columns by clicking the appropriate buttons below; then enter your data into the appropriate cells of the data-entry matrix. After all data have been entered, click the «Calculate» button.

Select the number of rows:	2	3	4	5	---
Select the number of columns:	2	3	4	5	---

Calculating Cramer's Phi – online calculator http://www.uccs.edu/~faculty/lbecker/

Chi-Square, Cramer's V, and Lambda

For a Rows by Columns Contingency Table

For a contingency table containing up to 5 rows and 5 columns, this unit will:

~ perform a chi-square analysis [the logic and computational details of chi-square tests are described in Chapter 8 of Concepts and Applications];

~ calculate Cramer's V, which is a measure of the strength of association among the levels of the row and column variables [for a 2x2 table, Cramer's V is equal to the absolute value of the phi coefficient];

~ and calculate the two asymmetrical versions of lambda, the Goodman- Kruskal index of predictive association, along with some other measures relevant to categorical prediction. [Click here for a brief explanation of lambda.]

To begin, select the number of rows and the number of columns by clicking the appropriate buttons below; then enter your data into the appropriate cells of the data-entry matrix. After all data have been entered, click the «Calculate» button.

Select the number of rows:	2	3	4	5	2
Select the number of columns:	2	3	4	5	2

Calculating Cramer's Phi – online calculator http://www.uccs.edu/~faculty/lbecker/

Data Entry

	B_1	B_2	B_3	B_4	B_5	Totals
A_1	12	7	-----	-----	-----	19
A_2	8	4	-----	-----	-----	12
A_3	-----	-----	-----	-----	-----	-----
A_4	-----	-----	-----	-----	-----	-----
A_5	-----	-----	-----	-----	-----	-----
Totals	20	11	-----	-----	-----	31

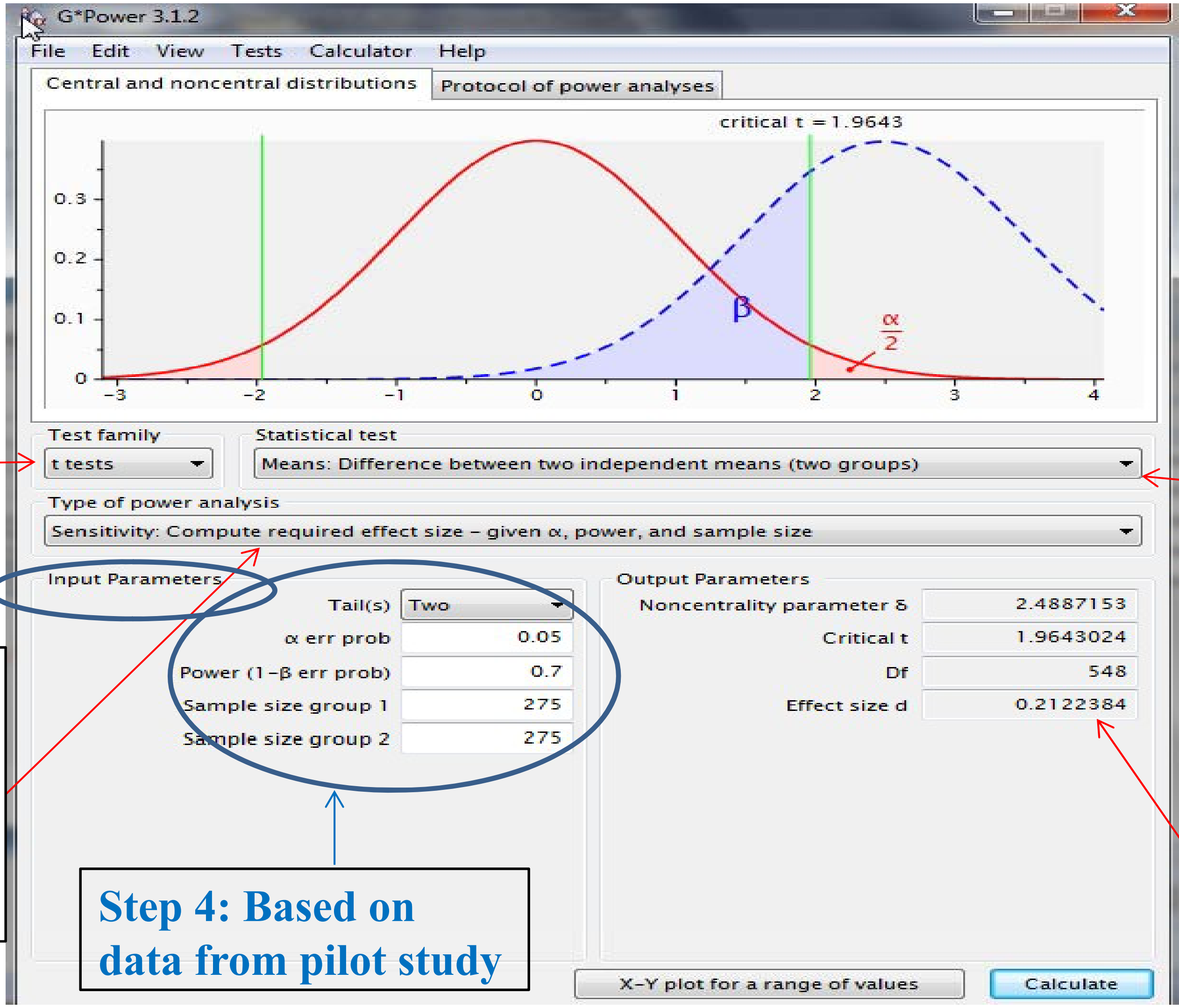
Reset Calculate

Chi-Square	df	P
0.03	1	0.8625

Cramer's V = 0.0359

Note that one of your expected cell frequencies is smaller than 5. For a rows by columns chi-square test, at least 80% of the cells must have an expected frequency of 5 or greater, and no cell may have an expected frequency smaller than 1.0. For a 2x2 table, the chi-square test is valid only if all expected cell frequencies are equal to or greater than 5. If this requirement is not met for a 2x2 table, use instead the Fisher Exact Probability Test. The Fisher Exact Test is also available for 2x3, 2x4, and 3x3.

A (not-so-great) Alternative: Calculating Cohen's d using GPower



Step 1: Choose "t-test" from drop down menu

Step 2: Choose "means" from drop down menu

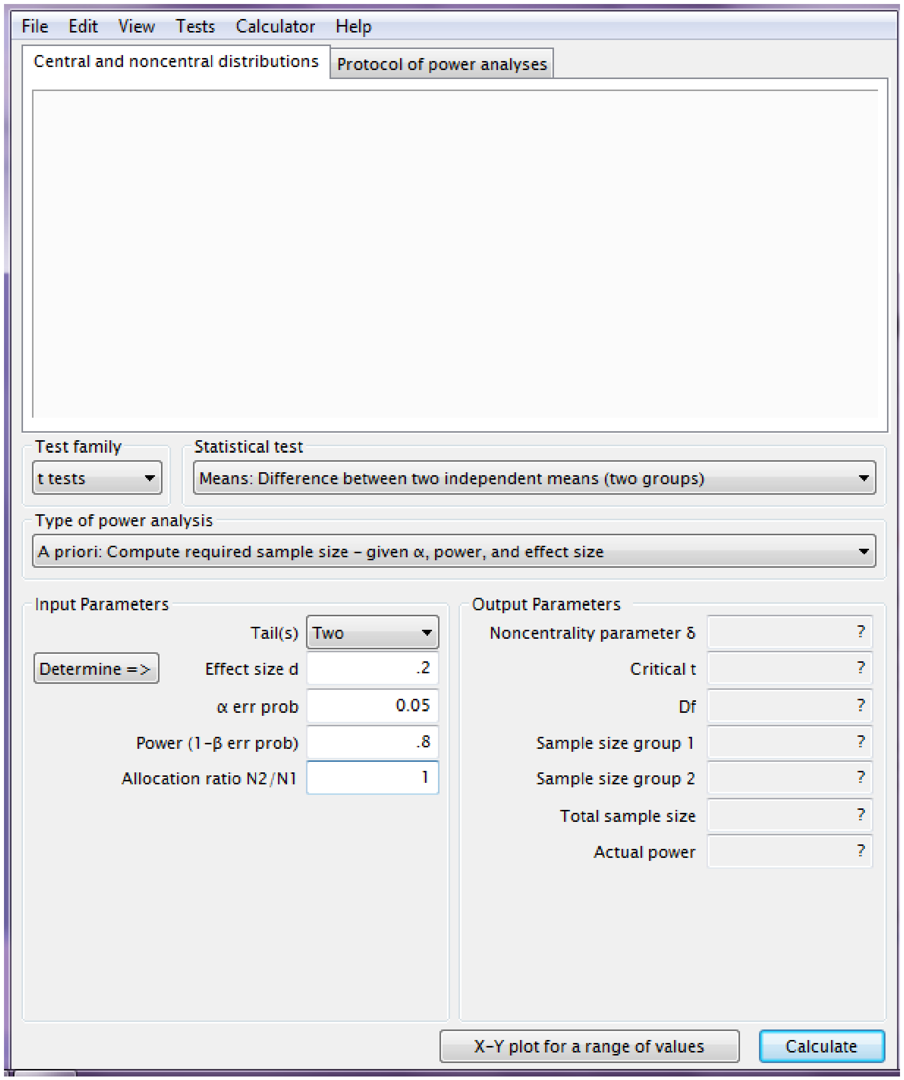
Step 3: Choose "Sensitivity" from the drop down menu

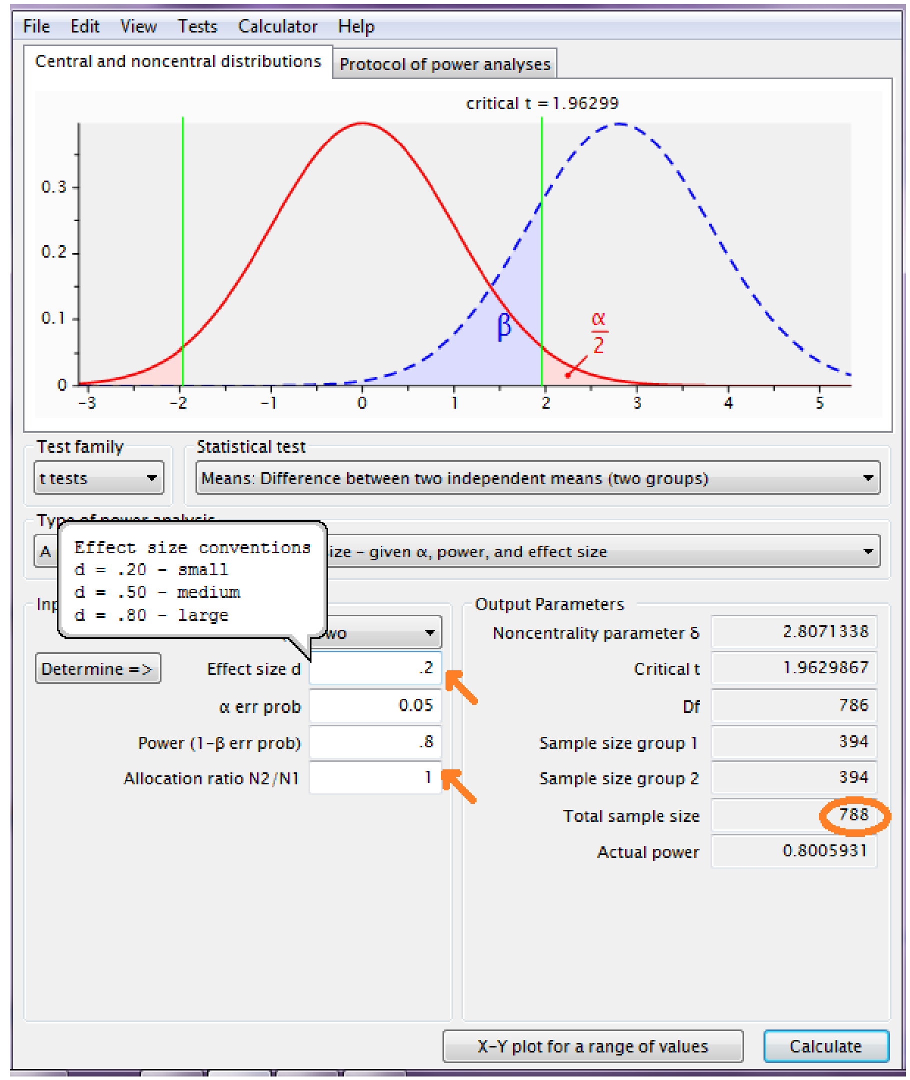
Step 4: Based on data from pilot study

Solution: Calculated effect size

Calculating SS given Cohen's d - GPower

A better use of Gpower is to calculate sample sizes.





Reporting Guidelines and Trends

- Reporting effect sizes has three important benefits (APA, 1999):
  - Meta-analysis
  - Informing subsequent research
  - Interpretation and evaluation of results within the context of related literature

Reporting Guidelines and Trends

- What to report (APA, 2010):
  - Type of effect size
  - Value of the effect size (in original units, such as lbs., or mean differences on a scale, and/or the effect size statistic)
  - Interpretation of the effect size
  - Practical significance of the effect size

References

- Chinn, S.(2000). A simple method for converting an odds ratio to effect size for use in meta-analysis. Statistics in Medicine, 19, 3127-3131.
- Cohen, J.(1992). A power primer. Psychological Bulletin, 112(1), 155-159.
- Cohen, J. (1988) “Statistical power analysis for the behavioral sciences”. New Jersey: Lawrenced Erlbaum Associates, Inc. Publishers. pp 283-286.

References

- Dunst, Carl J. et al. (2004) Guidelines for Calculating Effect Sizes for Practice-Based Research Syntheses. Centerscope 3(1) http://courseweb.unt.edu/gknezek/06spring/5610/centerscopevol3no1.pdf
- A presentation on effect size: http://www.family.umaryland.edu/ryc_research_and_evaluation/publication_product_files/selected_presentations/presentation_files/pdfs/effect%20size%20and%20intervention%20research.pdf
- Lee B.R., Bright C.L., Svobo da, D., Fakunmoju, S. and Barth R.P. Outcomes of group care for youth: A review of comparative literature. Research on Social Work Practice