Statistical Power Analysis
QMS Event

Xuejun Ryan Ji (MERM)
Ben Hives (KIN)
University of British Columbia, Vancouver

Mar 22, 2018

Outline

- Conceptual foundations of power analysis
  - What is the power analysis?
  - Refresh: alpha, beta, p-value, effect size
- Determining the sample size (a priori power analysis)
  - Linear bivariate regression: One group, size of slope
  - Associations among power, N, and effect size
- G* Power demonstration
- Comments on post-hoc power analysis

Statistical Power

- Statistical power is "…the probability the the null hypothesis will be rejected when it is false, that is, the probability of obtaining a statistically significant result"(Cohen, 1992)

	Truth (Unknown) in Population		
Decision on H_0	H_0 True	H_0 False	
Reject	Type I error **α**, $p_{critical}$	Power (1-β)	$p_{calculated} < p_{critical}$ or $TS_{calculated} > TS_{critical}$
Not Reject		Type II error β	$p_{calculated} > p_{critical}$ or $TS_{calculated} < TS_{critical}$

alpha and beta

- **α** (the significance level, $p_{critical}$)
  - the probability of a Type I error
  - the probability of rejecting the null hypothesis when it is true
  - not through calculation but subjective judgment
- β
  - the probability of Type II error
  - the probability of not rejecting the null hypothesis when it is false
  - through calculation

p-value vs. effect size

- p-value (the calculated p)
  - the probability of obtaining our observed results, or more extreme results, given the null hypothesis is true.
  - Caution! the p-value is the NOT the probability that the null hypothesis is true (Wasserstein & Lazar, 2016)
  - “The p value is a random variable that varies from sample to sample … Consequently, it is not appropriate to compare the p values from two distinct experiments, or from tests on two variables measured in the same experimenting, and declare that one is more important than the other. (p.100)
  - depends upon sample size

p-value vs. effect size

- Effect size
  - a statistic quantify the extent to which sample statistics diverge from the null hypothesis (Thompson, 2006)
  - the magnitude of the effect
  - independent of sample size
  - specify the ES estimation when reporting the effect size
  - Cohen's Reference Value

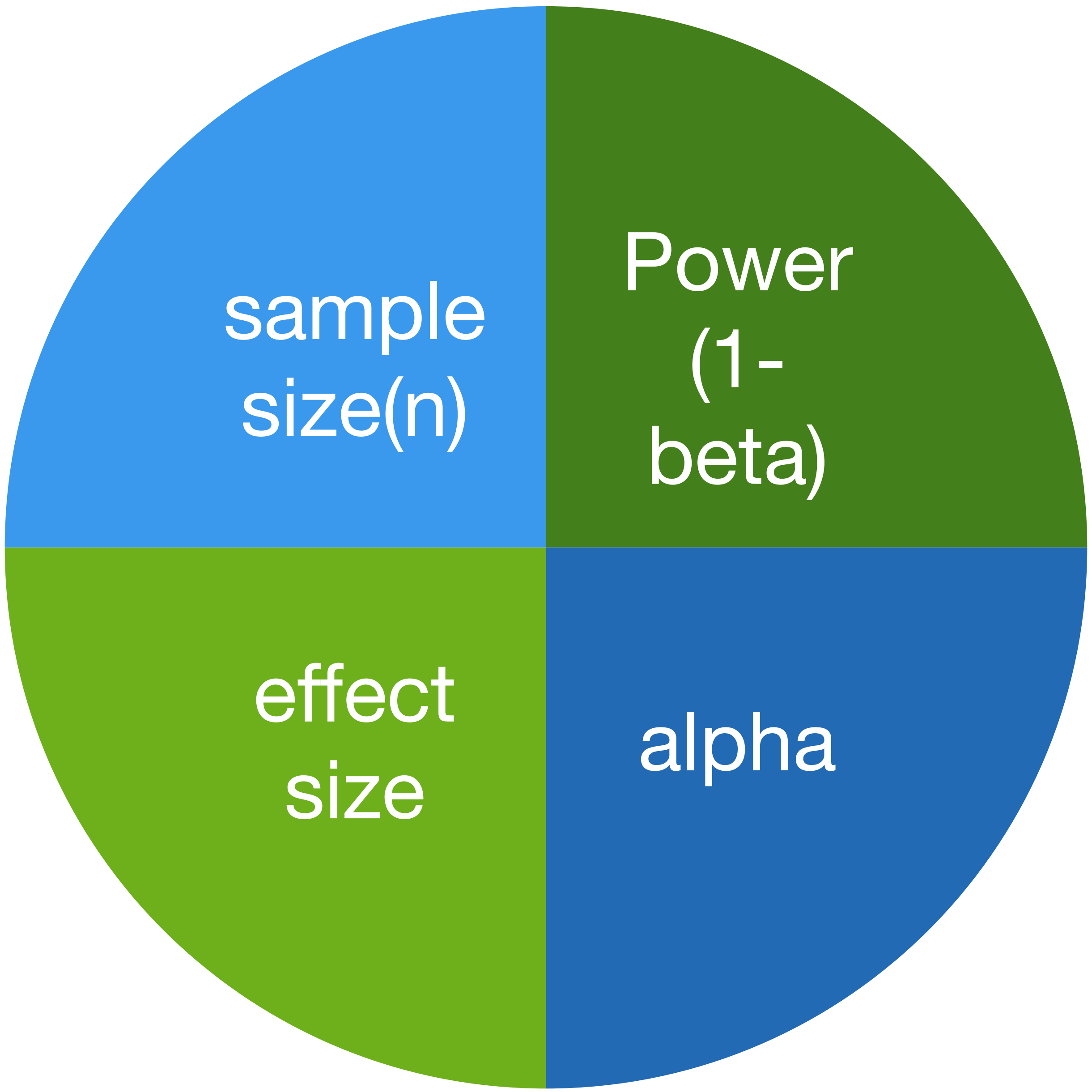
Cohen’s Reference Value

The terms "small," "medium," and "large" are relative, not only to each other, but to the area of behavioral science or even more particularly to the specific content and research method being employed in any given investigation … In the face of this relativity, there is a certain risk inherent in offering conventional operational definitions for these terms for use in power analysis in as diverse a field of inquiry as behavioral science. This risk is nevertheless accepted in the belief that more is to be gained than lost by supplying ***a common conventional frame of reference which is recommended for <u>use only when no better basis</u> for estimating the ES index is available.*** (Cohen, 1988, p25)

The table can be retrieved from the book by Cumming(2013, p39).

Cumming, G. (2013). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge.

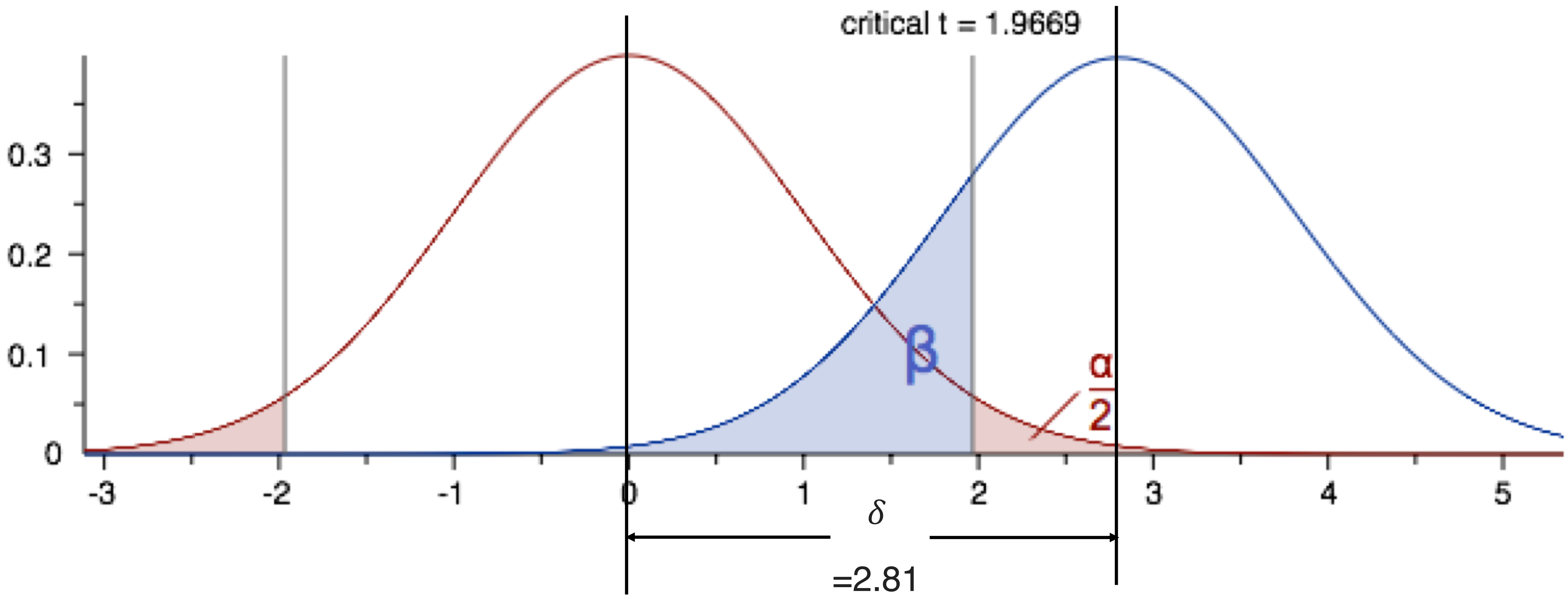
Statistical power depends upon the significance criterion (α), the sample size (*N*), and the effect size (ES)



Linear bivariate regression: One group, size of slope

```
Tail(s)                         =Two
 Slope H1                       =0.15
 α err prob                     =0.05
 Power (1-β err prob)           =0.8
 Slope H0                       =0
 Std dev σ_x                    =1
 Std dev σ_y                    =1
```

$$Y = \beta_0 + \beta_1 X + \varepsilon$$



δ is noncentrality parameter for non central *t* distribution

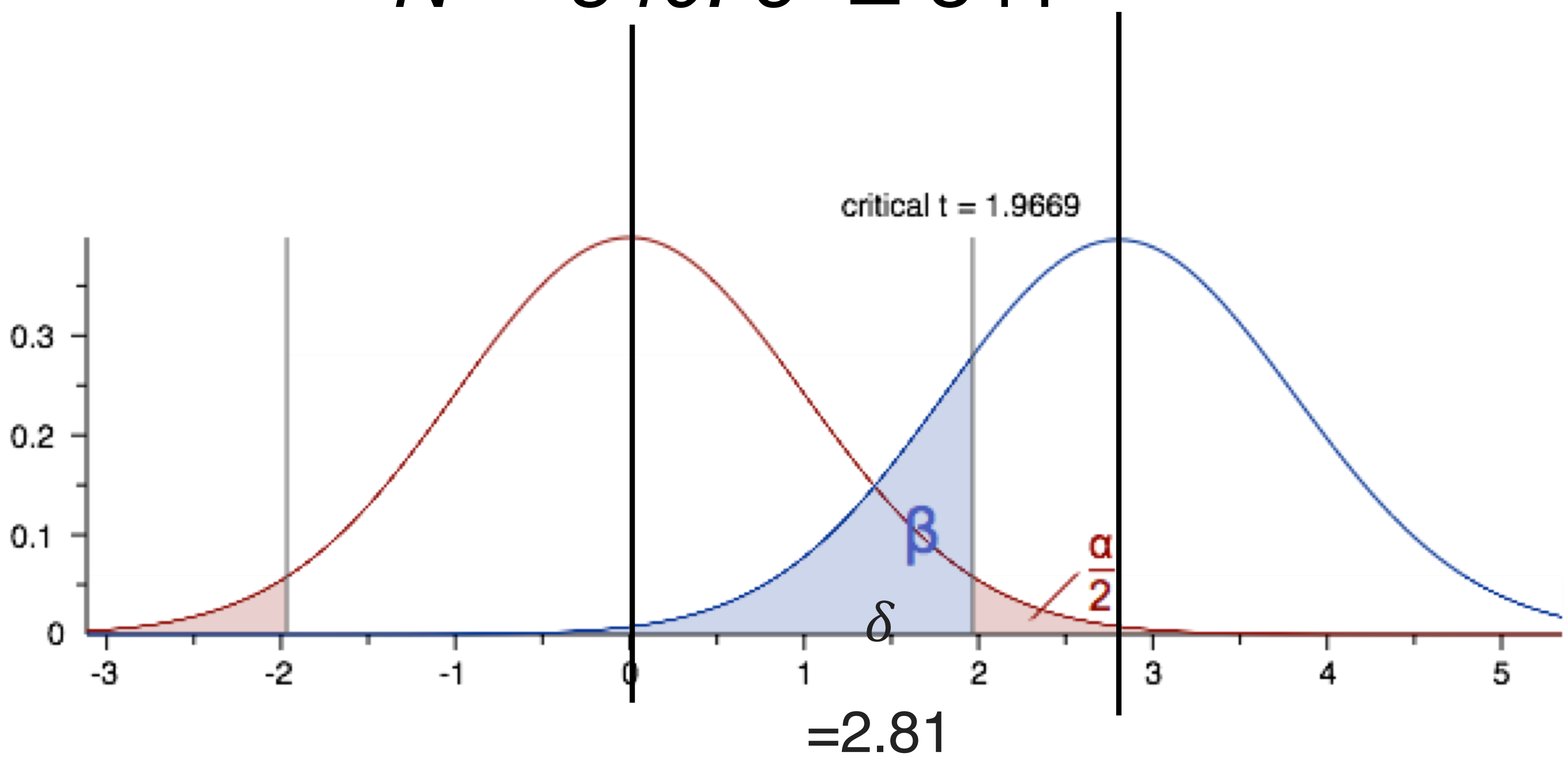
Linear bivariate regression: One group, size of slope

$$\delta = z_{\alpha/2} + (-z_{\beta}) \quad \textbf{(1)}$$

$$\delta = \frac{\beta_1\sqrt{N}}{\sqrt{\sigma_{\varepsilon}^2/\sigma_X^2}} = \frac{\beta_1\sqrt{N}}{\sqrt{(1-\beta_1^2)/\sigma_X^2}} \quad \textbf{(2)}$$

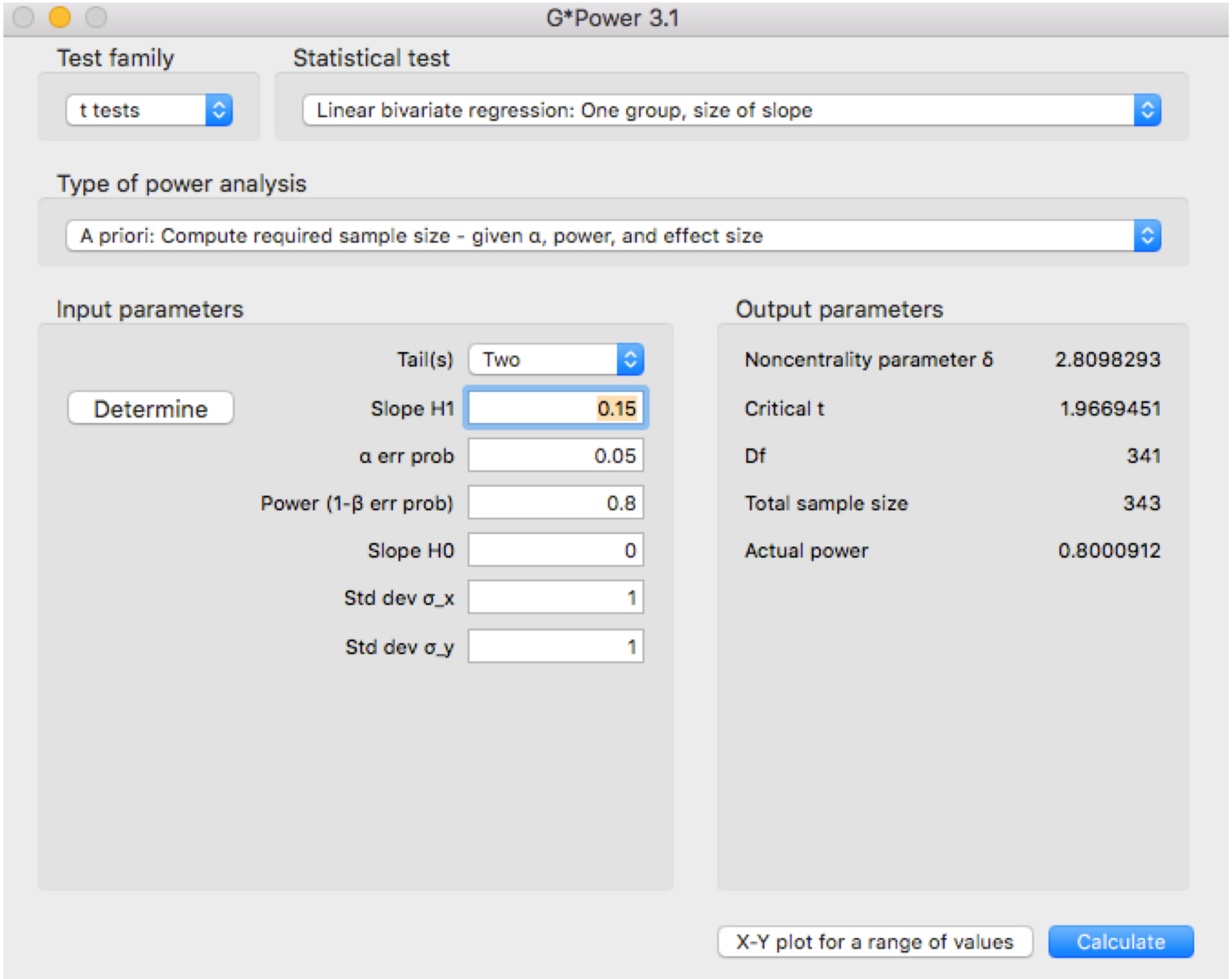
$$N = \frac{\delta^2(1-\beta_1^2)}{\beta_1^2 \cdot \sigma_X^2} = \frac{[z_{\alpha/2} + (-z_{\beta})]^2(1-\beta_1^2)}{\beta_1^2 \cdot \sigma_X^2} \quad \textbf{(3)}$$

$$N = 340.6 \cong 341$$

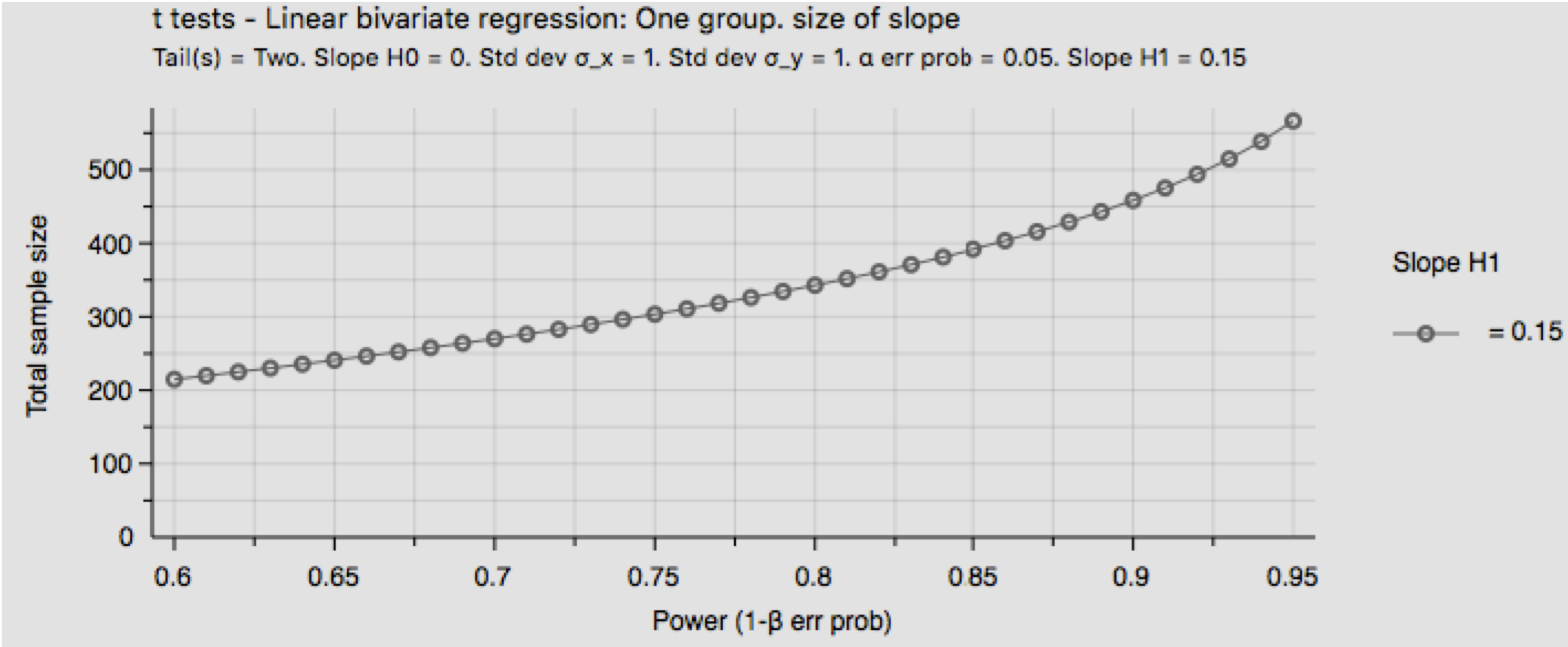


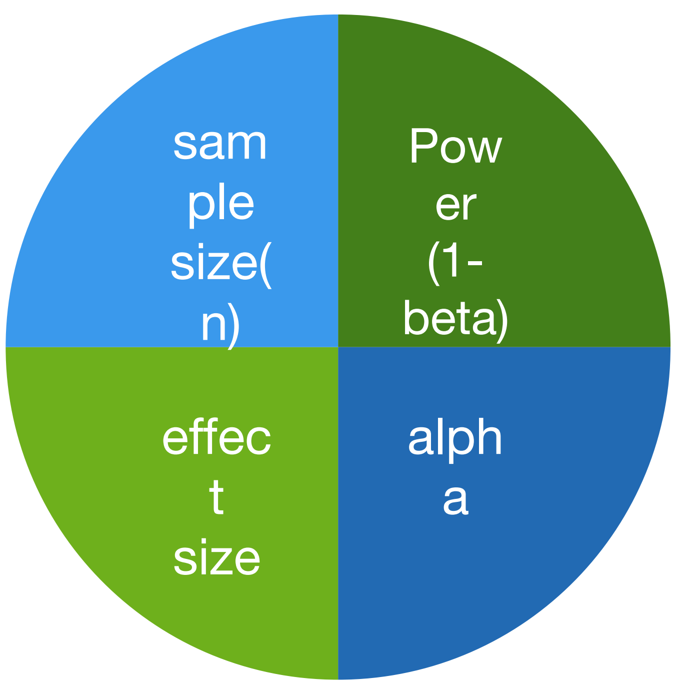
Lane, S. P., Hennes, E. P., & West, T. V. (2016, January). “I’ve got the power”: How anyone can do a power analysis of any type of data using simulation. Workshop presented at the 17th Annual Convention of the Society for Social and Personality Psychology, San Diego, CA.

Linear bivariate regression: One group, size of slope



positive association between the sample size and power

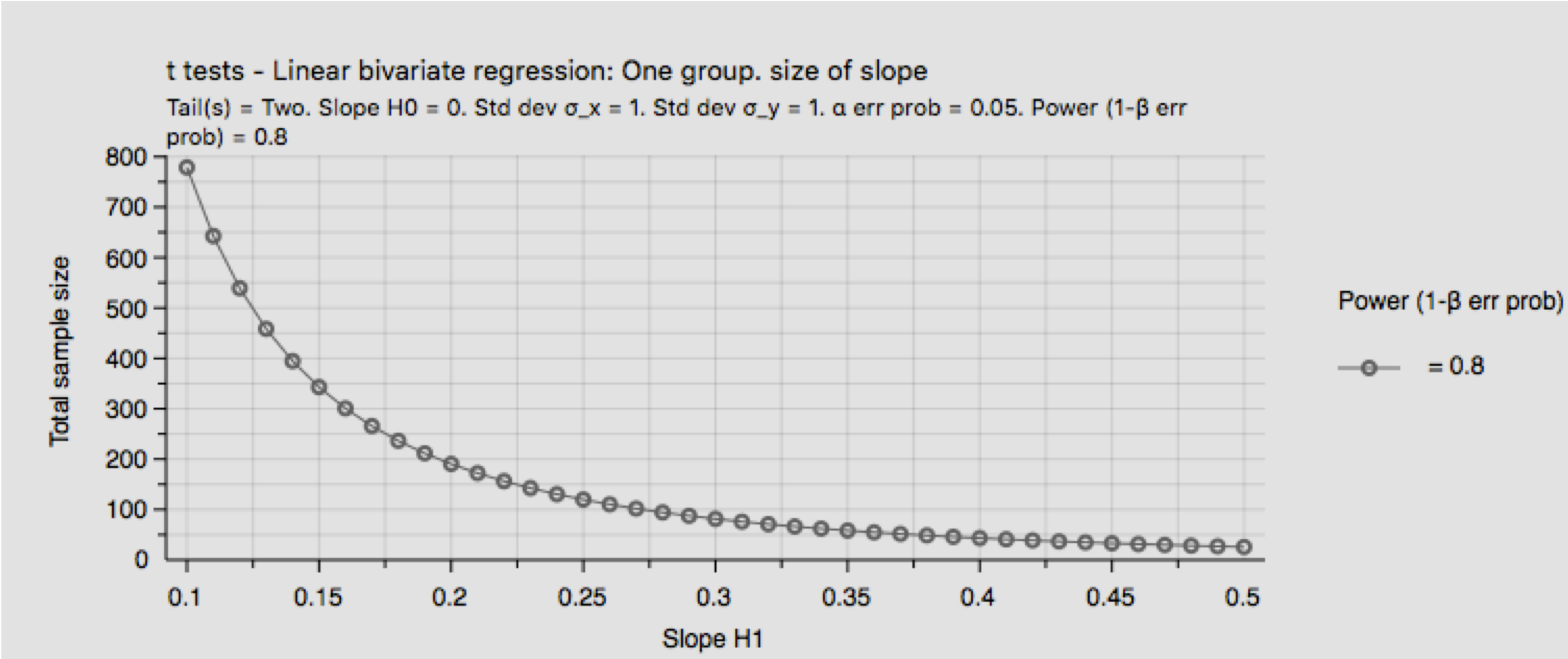


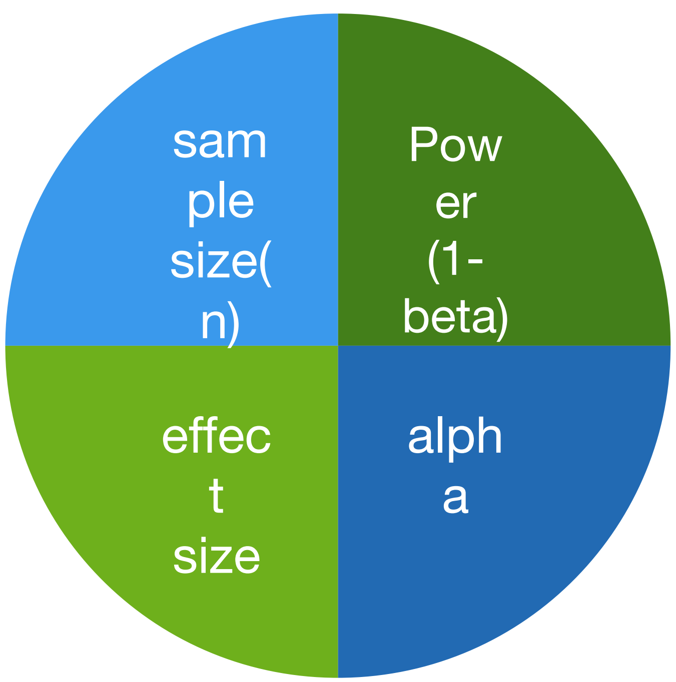


$$N = \frac{\delta^2(1-\beta_1^2)}{\beta_1^2 \cdot \sigma_X^2} = \frac{[z_{\alpha/2} + (-z_\beta)]^2(1-\beta_1^2)}{\beta_1^2 \cdot \sigma_X^2}$$

#	Power (1-β err prob)	Slope H1 = 0.15
17	0.760000	310.776340
18	0.770000	318.361804
19	0.780000	326.229663
20	0.790000	334.406019
21	0.800000	342.920717
22	0.810000	351.808103
23	0.820000	361.107987

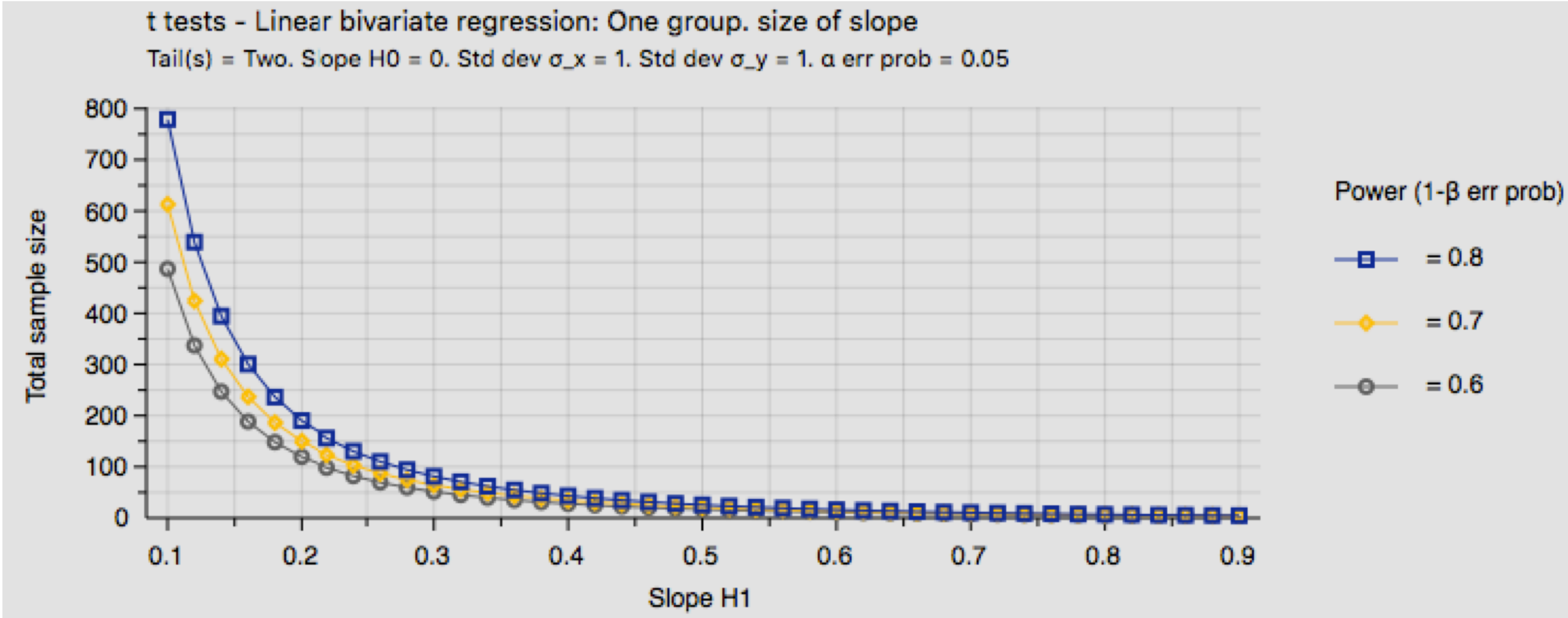
negative association between the sample size and effect size

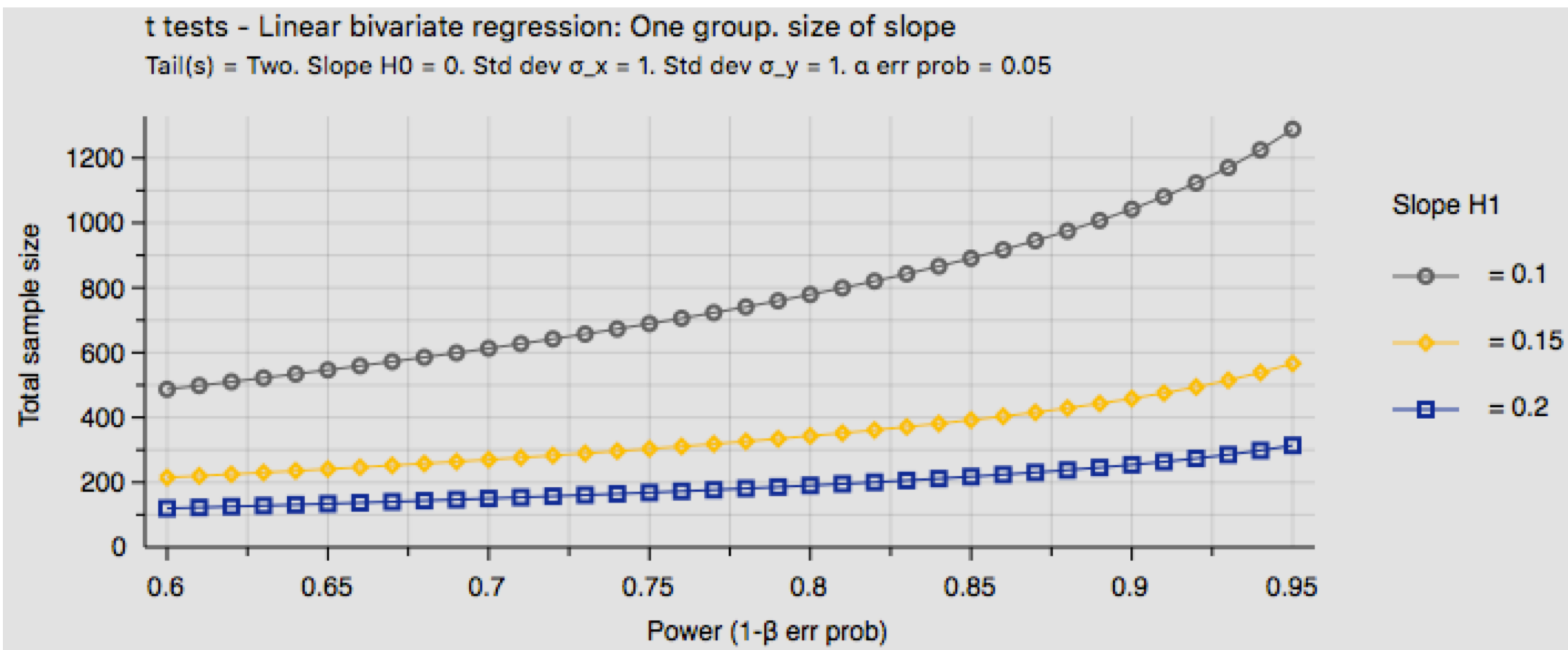


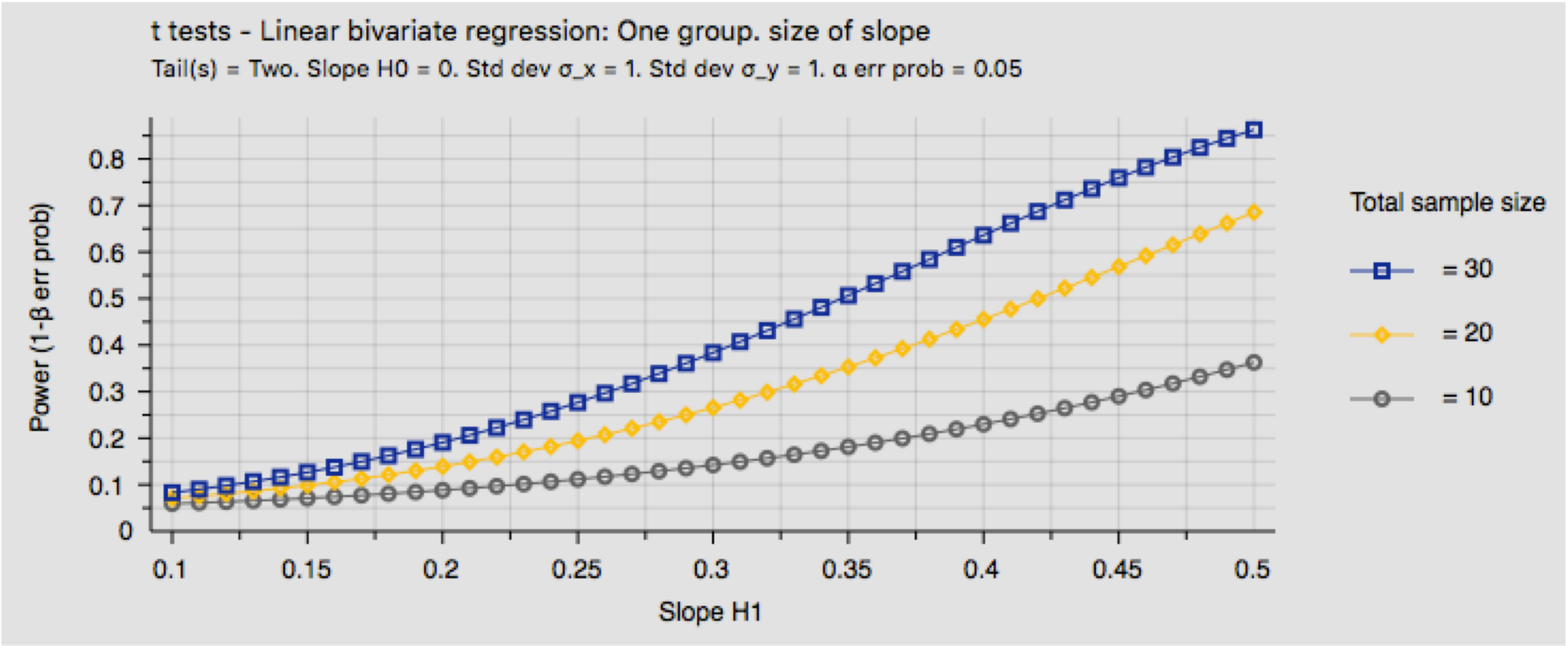


$$N = \frac{\delta^2(1-\beta_1^2)}{\beta_1^2 \cdot \sigma_X^2} = \frac{[z_{\alpha/2} + (-z_\beta)]^2(1-\beta_1^2)}{\beta_1^2 \cdot \sigma_X^2}$$

#	Slope H1	Power (1-β err prob) = 0.80
2	0.11000	642.743672
3	0.12000	539.138360
4	0.1300	458.509406
5	0.1400	394.533146
6	0.15000	342.920717
7	0.1600	300.680082


t tests - Linear bivariate regression: One group. size of slope
Tail(s) = Two. Slope H0 = 0. Std dev σ_x = 1. Std dev σ_y = 1. α err prob = 0.05
Total sample size
800
700
600
500
400
300
200
100
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Slope H1
Power (1-β err prob)
= 0.8
= 0.7
= 0.6


t tests - Linear bivariate regression: One group. size of slope
Tail(s) = Two. Slope H0 = 0. Std dev σ_x = 1. Std dev σ_y = 1. α err prob = 0.05
Total sample size
1200
1000
800
600
400
200
0
0.6
0.65
0.7
0.75
0.8
0.85
0.9
0.95
Power (1-β err prob)
Slope H1
= 0.1
= 0.15
= 0.2



t tests - Linear bivariate regression: One group. size of slope
Tail(s) = Two. Slope H0 = 0. Std dev σ_x = 1. Std dev σ_y = 1. α err prob = 0.05
Power (1-β err prob)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
Slope H1
Total sample size
= 30
= 20
= 10

Comments on the post-hoc power analysis

- ***The a priori power does not equal the post-hoc power.***
- The post-hoc power is not even an issue when we did not reject the null hypothesis.
- "…It is nonsensical to make power calculation after a study has been conducted and a statistical decision has been made."
- Researchers should report effect sizes for both significant and non-significant findings.

- Zumbo, B. D., & Hubley, A. M. (1998). A note on misconceptions concerning prospective and retrospective power. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *47*(2), 385-388

Comments on the post-hoc power analysis

The table can be retrieved from the paper by Zumbo and Hubley (1998)

- Zumbo, B. D., & Hubley, A. M. (1998). A note on misconceptions concerning prospective and retrospective power. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *47*(2), 385-388

Comments on the post-hoc power analysis

- The a priori power does not equal the post-hoc power.
- The post-hoc power is not even an issue when we did not reject the null hypothesis.
- “…It is nonsensical to make power calculation after a study has been conducted and a statistical decision has been made.”
- Researchers should report effect sizes for both significant and non-significant findings.

- Zumbo, B. D., & Hubley, A. M. (1998). A note on misconceptions concerning prospective and retrospective power. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *47*(2), 385-388

G*Power

- t-test (independent samples)
- t-test(paired samples)
- ANOVA
- Repeated-Measures ANOVA
- Multiple Linear Regression (R-squared)

a priori power analysis

Two independent groups t-test

Input:	Type of test:	t-test (means) two tails
	Alpha =	0.05
	Power (1-beta) =	0.9
	Effect size (d) =	0.5
Output:	Total sample size =	?

Two matched groups t-test

Input:	Type of test:	t-test (means) two tails
	Alpha =	0.05
	Power (1-beta) =	0.95
	ES[$dz = M_{(m1-m2)}/SD_{(m1-m2)}$]	0.4
Output:	Total sample size =	?

F-test: One way ANOVA Omnibus test

Input:	Type of test:	ANOVA (global)
	Alpha =	0.05
	Power (1-beta) =	0.90
	ES (Cohen’s f)	0.25
	# of Group	2
Output:	Total sample size =	?

Compare the results with two-independent group t -test

F-test: One way ANOVA Omnibus test

Input:	Type of test:	ANOVA (global)
	Alpha =	0.05
	Power (1-beta) =	0.90
	ES (Cohen’s f)	0.25
	# of Group	2
Output:	Total sample size =	?

Compare the results with two-independent group t -test

F-test: Multiple way ANOVA (fixed effect)

3×5 Design

Main Effect: Factor A

Input:	Type of test:	ANOVA (global)
	Alpha =	0.05
	Power (1-beta) =	0.95
	ES (Cohen's f)	0.40
	# of Group	15
	Numerator df (3-1)	2
Output:	Total sample size =	?

Main Effect: Factor B

Input:	Type of test:	ANOVA (global)
	Alpha =	0.05
	Power (1-beta) =	0.95
	ES (Cohen's f)	0.40
	# of Group	15
	Numerator df (5-1)	4
Output:	Total sample size =	?

Main Effect: Interaction effect

Input:	Type of test:	ANOVA (global)
	Alpha =	0.05
	Power (1-beta) =	0.95
	ES (Cohen’s f)	0.40
	# of Group	15
	Numerator df	8
Output:	Total sample size =	?

Multiple Way ANOVA

Effects	Groups	Numerator df	total sample size
A	15	2	101
B	15	4	122
$A \times B$	15	8	151

F-test:Repeated ANOVA

Repeated Measure ANOVA GPower Default

Input:	Type of test:	RM-within factor
	Alpha =	0.05
	Power (1-beta) =	0.95
	ES (SPSS partial eta-square)	0.71
	# of Group	1
	# of measures	2
	corr among rep measures	0.726
Output:	Total sample size =	4

Two matched groups t-test

Input:	Type of test:	t-test (means) two tails
	Alpha =	0.05
	Power (1-beta) =	0.95
	ES[$dz=M_{(m1-m2)}/SD_{(m1-m2)}$]	1.5(from 0.71)
Output:	Total sample size =	8

Repeated Measure ANOVA GPower (SPSS)

Input:	Type of test:	RM-within factor
	Alpha =	0.05
	Power (1-beta) =	0.95
	ES (SPSS partial eta-square)	0.71
	# of Group	1
	# of measures	2
	corr among rep measures	0.726
Output:	Total sample size =	9

Repeated Measure ANOVA (GPower 3.0)

Input:	Type of test:	RM-within factor
	Alpha =	0.05
	Power (1-beta) =	0.95
	ES (Gpower partial eta-square)	0.26
	# of Group	1
	# of measures	2
	corr among rep measures	0.726
Output:	Total sample size =	8

Repeated Measure ANOVA

RM-ANOVA (default)	SPSS Partial eta square = 0.71	n = 4
matched t-test	dz= 1.5	n = 8
RM-ANOVA (SPSS)	SPSS Partial eta square = 0.71	n = 9
RM-ANOVA (default)	GPower eta square = 0.26	n= 8

GPower partial eta square does not take into account the correlations among the repeated measures.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in psychology*, *4*, 863.

Multiple Linear Regression

The figure can be retrieved from the paper by Thompson (2016)

Thompson, B. (2016). The Case for Using the General Linear Model as a Unifying Conceptual Framework for Teaching Statistics and Psychometric Theory. *Journal of Methods and Measurement in the Social Sciences*, *6*(2), 30-41.

Effect Size Conversion

$$d = \frac{2r}{\sqrt{1-r^2}} \quad \text{(r to d when } n_1 = n2\text{)}$$

$$r = \frac{d}{\sqrt{d^2 + a}} \quad \text{(d to r)}$$

$$\text{(where } a = \frac{(n_1 + n_2)^2}{n_1 n_2}\text{,the correction term when } n_1 \neq n_2\text{)}$$

$$f^2 = \frac{\eta^2}{1-\eta^2} \quad (\eta^2 \text{ to Cohen's } f^2)$$

$$\eta^2 = \frac{f^2}{1+f^2} \quad (\text{Cohen's } f^2 \text{ to } \eta^2)$$

$$d = 2f \quad \text{(f to d, two independent groups)}$$

Cohen, J. (1988). Statistical power analysis for the behavioral sciences . Hilsdale. *NJ: Lawrence Earlbaum Associates*, *2*.

Two independent groups t-test

Input:	Type of test:	t-test (means) two tails
	Alpha =	0.05
	Power (1-beta) =	0.9
	Effect size (d) =	0.5
Output:	Total sample size =	172

F-test: One way ANOVA Omnibus test

Input:	Type of test:	ANOVA (global)
	Alpha =	0.05
	Power (1-beta) =	0.90
	ES (Cohen’s f)	0.25
	# of Group	2
Output:	Total sample size =	172

MLR: Fixed model R^2 deviation from zero

Input:	Type of test:	F-test
	Alpha =	0.05
	Power (1-beta) =	0.90
	ES (Cohen's f^2)	0.0625
	# of predictor	1
Output:	Total sample size =	171

MLR:single regression coefficient

Input:	Type of test:	t-test
	tails	Two
	Alpha =	0.05
	Power (1-beta) =	0.90
	ES (Cohen's f^2)	0.0625
	# of predictor	1
Output:	Total sample size =	171

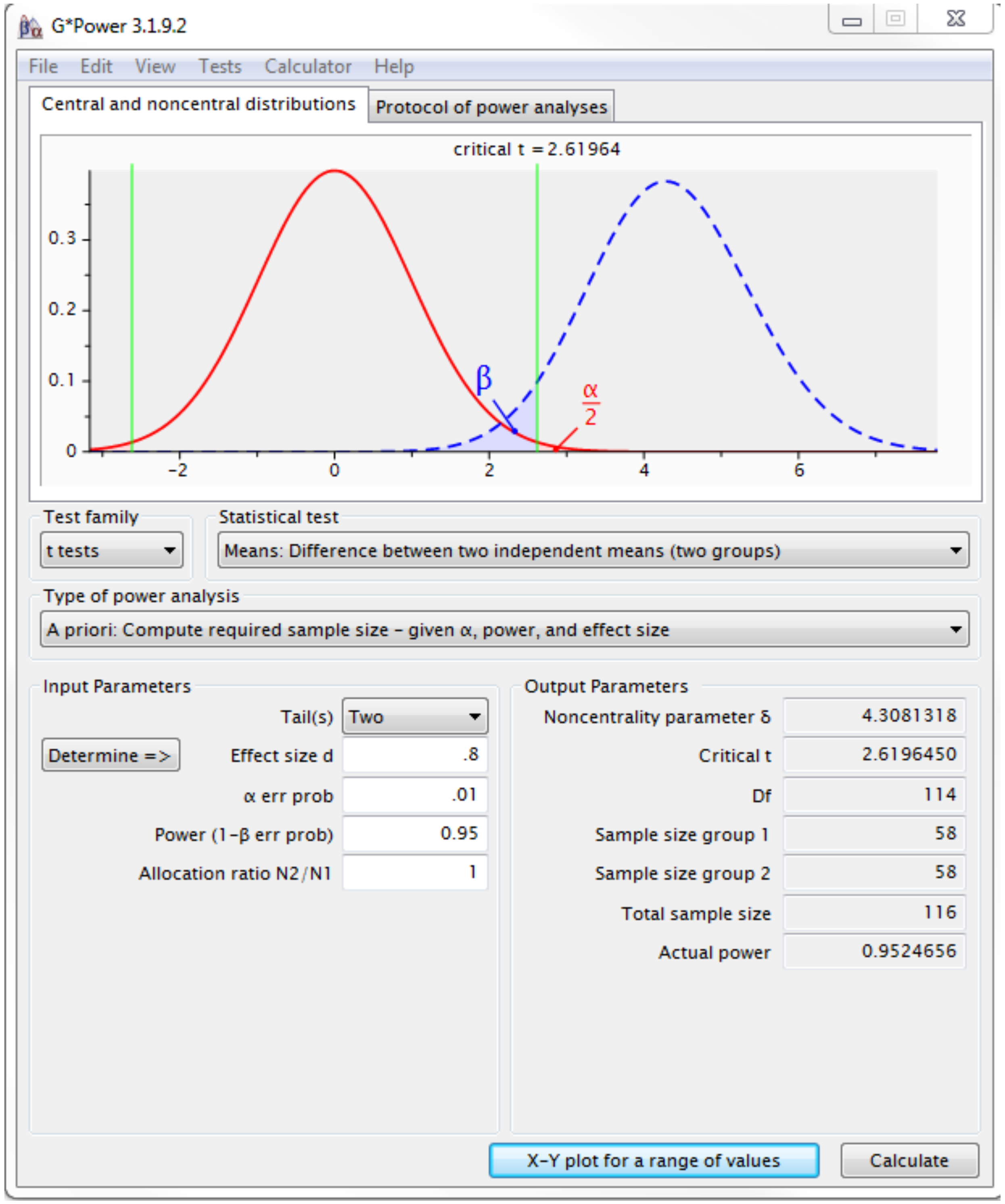
Multiple Linear Regression

t-test	d = 0.5	n = 172
ANOVA	f = 0.25	n = 172
MLR (t-test)	f^2 = 0.0625	n = 171
MLR (ANOVA)	f^2 = 0.0625	n= 171

Now you try!

- Suppose you work in a clinical lab testing new drug. Since you want to avoid any errors, you set Alpha at 0.01 and Beta at .05. Previous studies have found a large effect size, but haven't reported an exact number. You intend to randomly distribute participants into two groups and analyze using a t-test.

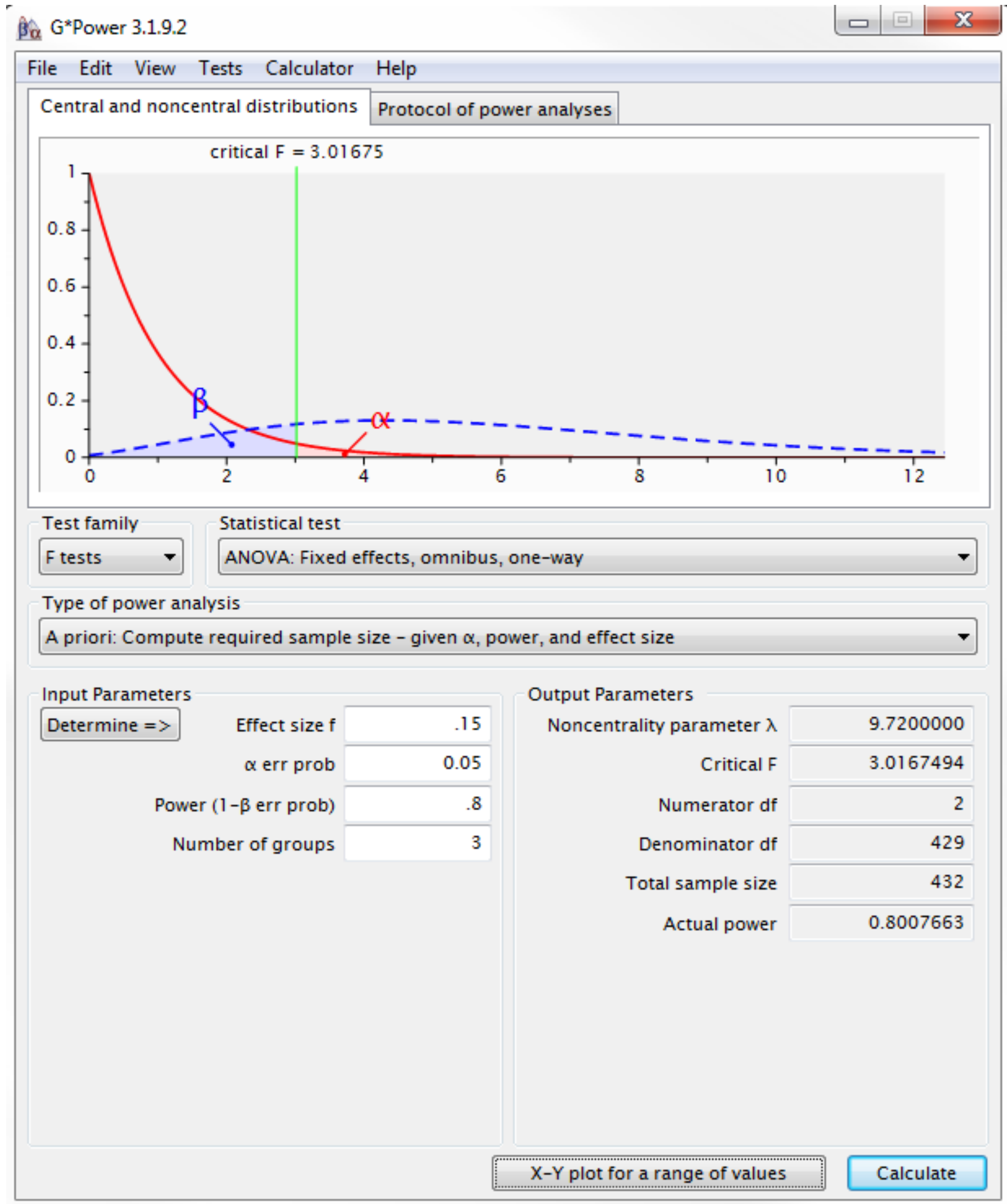
Now you try!


G*Power 3.1.9.2
File Edit View Tests Calculator Help
Central and noncentral distributions
Protocol of power analyses
critical t = 2.61964
0.3
0.2
0.1
0
−2
0
2
4
6
β
α
2
Test family
t tests
Statistical test
Means: Difference between two independent means (two groups)
Type of power analysis
A priori: Compute required sample size − given α, power, and effect size
Input Parameters
Tail(s) Two
Determine =>
Effect size d .8
α err prob .01
Power (1−β err prob) 0.95
Allocation ratio N2/N1 1
Output Parameters
Noncentrality parameter δ 4.3081318
Critical t 2.6196450
Df 114
Sample size group 1 58
Sample size group 2 58
Total sample size 116
Actual power 0.9524656
X−Y plot for a range of values
Calculate

Now you try!

- It’s spring and you want to test a new fertilizer. You have three groups of fertilizers: No Fertilizer, old and busted fertilizer, and the new hotness fertilizer. Harkening back for the days of Fisher, you decide to set alpha at 0.05 and Beta at .2, previous studies found an effect size of .15. How many daffodils should you pant to test your hypothesis?

Now you try!


G*Power 3.1.9.2
File Edit View Tests Calculator Help
Central and noncentral distributions
Protocol of power analyses
critical F = 3.01675
Test family
F tests
Statistical test
ANOVA: Fixed effects, omnibus, one-way
Type of power analysis
A priori: Compute required sample size – given α, power, and effect size
Input Parameters
Determine =>
Effect size f .15
α err prob 0.05
Power (1-β err prob) .8
Number of groups 3
Output Parameters
Noncentrality parameter λ 9.7200000
Critical F 3.0167494
Numerator df 2
Denominator df 429
Total sample size 432
Actual power 0.8007663
X-Y plot for a range of values
Calculate

Any Questions?

Thank You!

Power Analysis in R