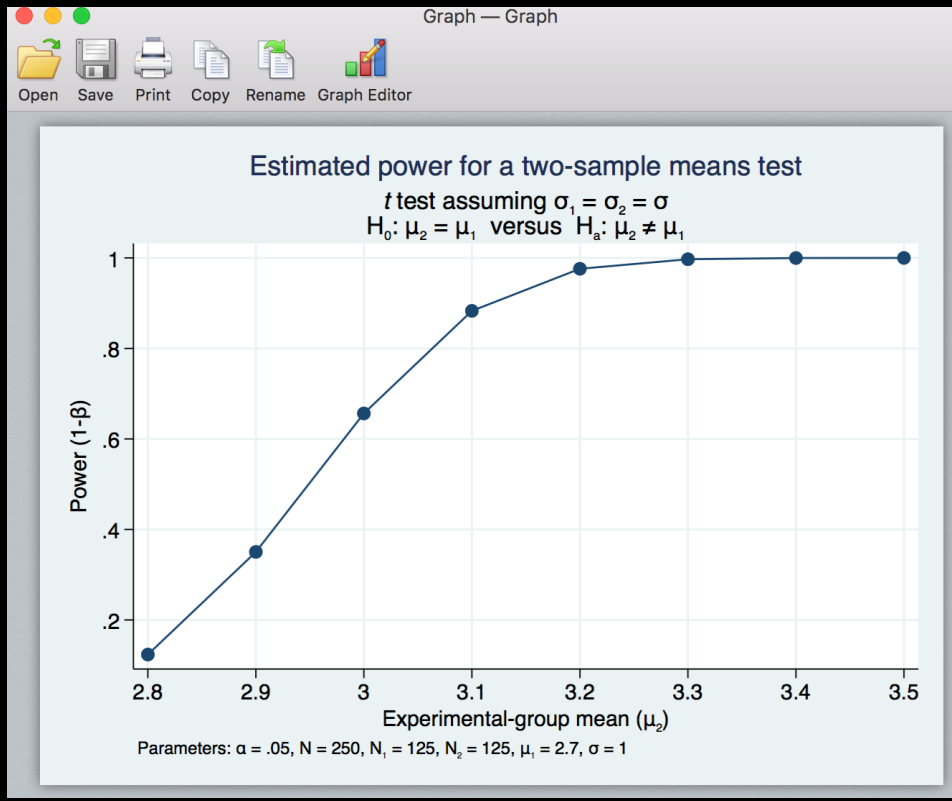
Biostatistics 212

Power and Sample Size Calculations

August 22, 2017



Phillips Perera, MD
Clinical Associate Professor
Department of Emergency Medicine
Stanford University Medical Center
Master's Degree Candidate
UCSF TICR Program 6/2018

Lecture Goals:

- Review basic approach to determining sample size for a given study
- Discuss basic concepts (alpha, beta, power)
- Learn how STATA can calculate these numbers for different study types

Sample Size Calculation:

* Critical to the success of a given study
- *Decide how many subjects to include in your study to achieve 'statistical significance'*
- Calculate too few subjects-may fail to answer the research question
- Calculate too many subjects-makes study more difficult and potentially costly
- *Calculate correct sample size-first step to optimally planning your study (although ballpark estimate-dropouts?)*

Hurley SB, Cummings SR, Browner WS, Grady DG, Newman TB. Designing Clinical Research. 2013; 4th Edition. Walters, Kluwer, Lippincott.

Null vs Alternative Hypothesis:

* State your research question as a hypothesis that summarizes main elements of your study
* Restate your study hypothesis to one that proposes no difference between the study groups

- *This null hypothesis* will become the formal basis for statistical testing on data analysis
- *Statistical tests will then help to estimate the probability that an observed association in your study is due to chance*
- *As opposed to the alternative hypothesis*-that an association in your study is actually present

Hurley SB, Cummings SR, Browner WS, Grady DG, Newman TB. Designing Clinical Research. 2013; 4th Edition. Walters, Kluwer, Lippincott.

Underlying Statistical Principles:

* Type 1 Error:

- *Rejection error*-failure to reject the null hypothesis when it is actually true (conclude that an association is present when one truly does not exist)
- Convict an innocent man, conclude there is an association between CT scans and cancer-when there truly is not one *(false positive error)*

* Type 2 Error:

- *Acceptance error*-failure to reject the null hypothesis when actually false (conclude that there is no association when one truly exists)
- Acquit a murderer, conclude that there is no association between CT scans and cancer-when there is truly one present *(false negative error)*

Hurley SB, Cummings SR, Browner WS, Grady DG, Newman TB. Designing Clinical Research. 2013; 4th Edition. Walters, Kluwer, Lippincott.

Underlying Statistical Principles:

* Alpha:

- *Level of statistical significance*
- Maximum probability of committing a type I error, or level of reasonable doubt to accept result, using statistical tests
- *Typically set at 0.05* (5% chance of incorrectly rejecting the null hypothesis when it is true)

* Beta:

- *Level at which one is willing to accept the chance of missing an association when actually present*
- *Typically set at 0.20* (20% chance of failure to recognize an association)

Hurley SB, Cummings SR, Browner WS, Grady DG, Newman TB. Designing Clinical Research. 2013; 4th Edition. Walters, Kluwer, Lippincott.

Underlying Statistical Principles:

* Power:

- *Calculated as 1-beta* (beta 0.2 = power 0.8)

* *80% chance of finding an association between study groups of that size, or greater*

- Optimally alpha and beta would be 0.00, and power would be 100%

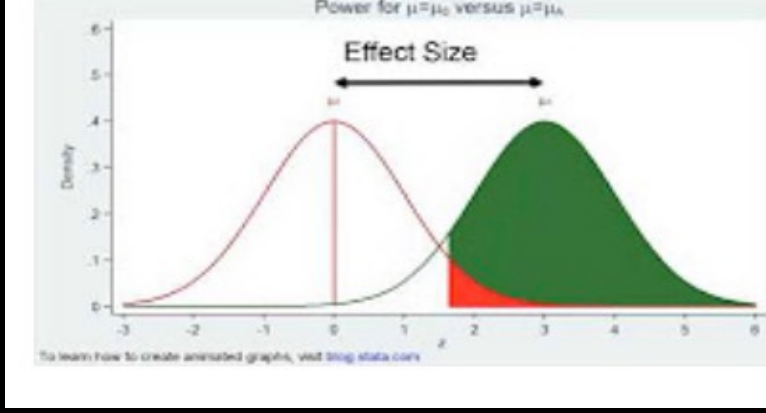
- However, that requires increasing the sample size (making the study potentially more difficult)

* Power Analysis:

- *Evaluates statistical power at a given effect size for a fixed sample size and level of significance*

Hurley SB, Cummings SR, Browner WS, Grady DG, Newman TB. Designing Clinical Research. 2013; 4th Edition. Walters, Kluwer, Lippincott.

Underlying Statistical Principles:



* Calculate effect size

- Estimation of the size of the association in the study sample that one looks to detect
- *Difference in mean value of continuous outcome between 2 study groups*
- *Difference in proportion of subjects with dichotomous outcome between 2 study groups*

* Calculate study variability:

- *Standard deviation of outcome variable*

* Calculate standardized effect size (effect size / SD of outcome)

- *Increasing standardized effect size —> decreased sample size*

Hurley SB, Cummings SR, Browner WS, Grady DG, Newman TB. Designing Clinical Research. 2013; 4th Edition. Walters, Kluwer, Lippincott.

UCSF Clinical & Translational Science Institute

Explore the *Training in Clinical Research Program* at UCSF

Sample Size Calculators

SAMPLE SIZE CALCULATORS FOR DESIGNING CLINICAL RESEARCH

Sample Size Calculators

If you are a clinical researcher trying to determine how many subjects to include in your study or you have another question related to sample size or power calculations, we developed this website for you. Our approach is based on Chapters 5 and 6 in the 4th edition of Designing Clinical Research (DCR-4), but the material and calculators provided here go well beyond an introductory textbook on clinical research methods.

DCR-4 Chapter 6 "Estimating Sample Size and Power" Appendixes A-F provide instructions and tables for calculating sample sizes. However, the tables provided for two-group studies assume that the groups are of equal size. If you have unequal group sizes (as occurs in most observational studies) or a study parameter that does not correspond to one of the table headings, then you may have to use one of these online calculators.

Online Calculators Corresponding to Tables in DCR-4 Chapter 6

Two-group analytic study comparing

(A) Means
- (A1) sample size given effect size
- (A2) effect size given sample size

(A') Means with clustered sampling
- (A'1) sample size given effect size (clustered sampling)
- (A'2) effect size given sample size (clustered sampling)

(B) Proportions
- (B1) sample size given both proportions or baseline proportion and relative risk or odds ratio
- (B2) second proportion (and relative risk and odds ratio) given baseline proportion and sample sizes

One-group analytic study comparing

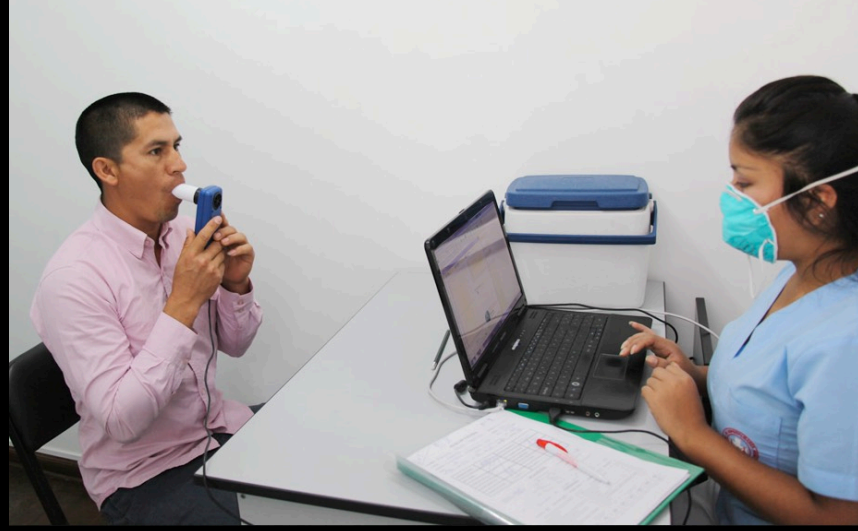
Study Example:

- Let's think of a study looking at the effects of smoking on the lung function of subjects

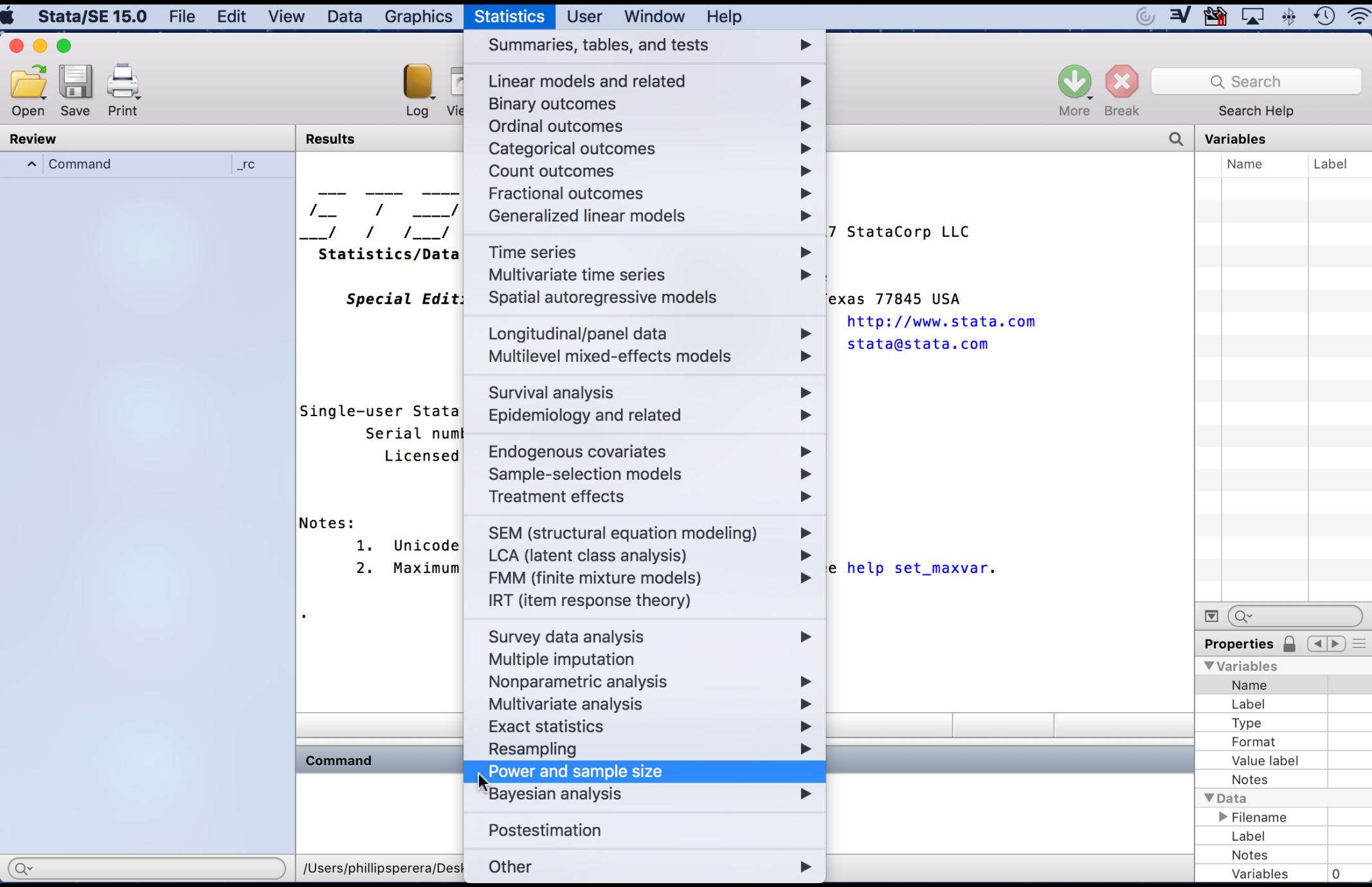
- *We will use a two-sided alternative hypothesis that specifies that there may be an association between smoking and FEV that could be either smaller or larger (smoking—> either decreased or increased FEV)*

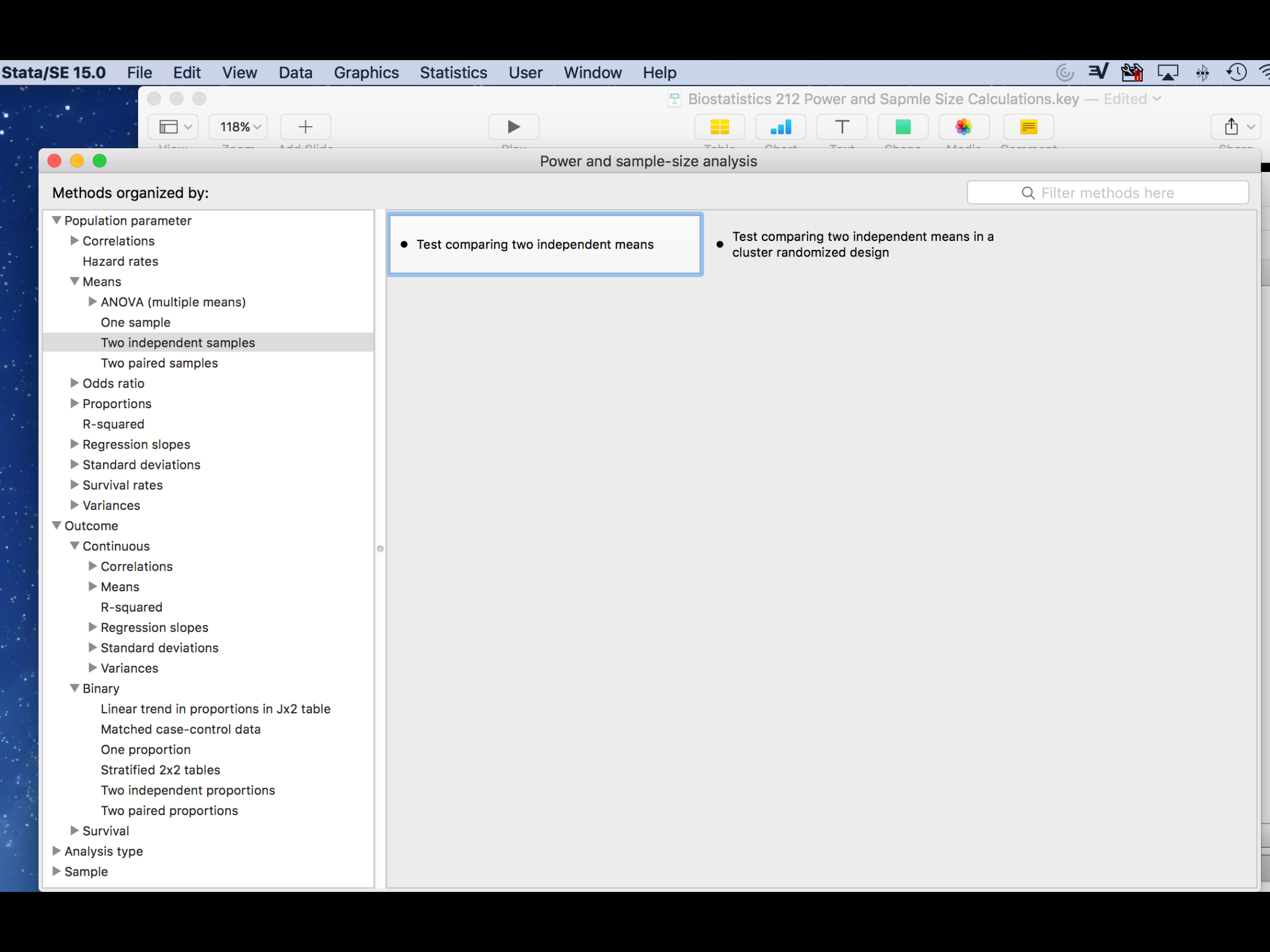
- Typically avoid a *one-side alternative hypothesis* that captures only one direction of a potential association between variable and outcome

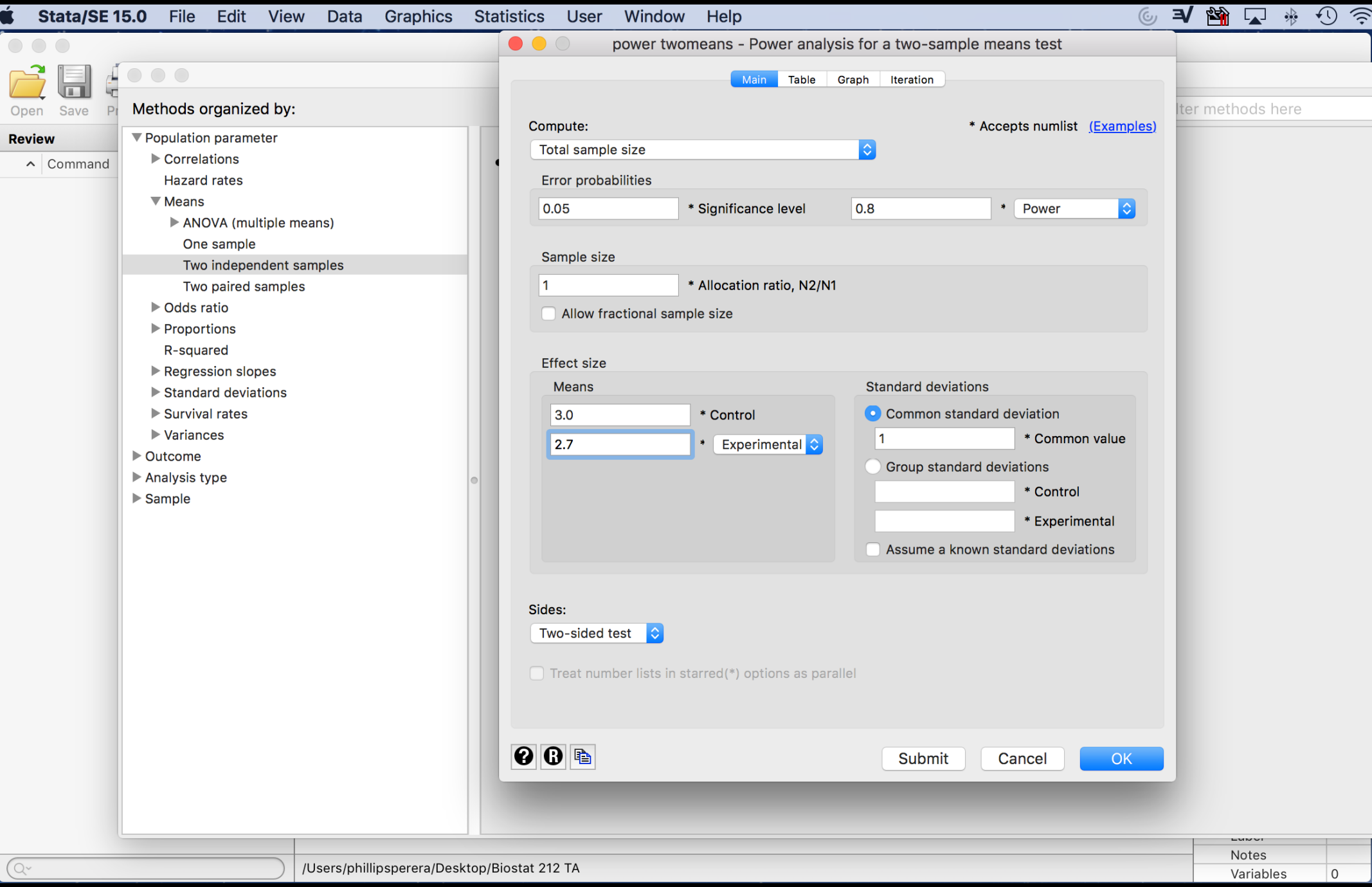
Study Example:



- Mean FEV non-smokers 3.0 liters, smokers 2.7 liters
- What test should we use?
- T-test to determine if the mean value of a continuous outcome in the non-smoker group (3.0 L) significantly differs from that in the smoker group (2.7 L)
- *T-test traditionally used in analysis of continuous variables and outcomes*
- Will use a standard deviation of 1


Stata/SE 15.0
File
Edit
View
Data
Graphics
Statistics
User
Window
Help
Summaries, tables, and tests
Linear models and related
Binary outcomes
Ordinal outcomes
Categorical outcomes
Count outcomes
Fractional outcomes
Generalized linear models
Time series
Multivariate time series
Spatial autoregressive models
Longitudinal/panel data
Multilevel mixed-effects models
Survival analysis
Epidemiology and related
Endogenous covariates
Sample-selection models
Treatment effects
SEM (structural equation modeling)
LCA (latent class analysis)
FMM (finite mixture models)
IRT (item response theory)
Survey data analysis
Multiple imputation
Nonparametric analysis
Multivariate analysis
Exact statistics
Resampling
Power and sample size
Bayesian analysis
Postestimation
Other
Open
Save
Print
Log
More
Break
Search
Search Help
Review
Command
_rc
Results
Statistics/Data
Special Edit
7 StataCorp LLC
exas 77845 USA
http://www.stata.com
stata@stata.com
Single-user Stata
Serial numb
Licensed
Notes:
1. Unicode
2. Maximum
e help set_maxvar.
Command
/Users/phillipsperera/Desk
Variables
Name
Label
Properties
Variables
Name
Label
Type
Format
Value label
Notes
Data
Filename
Label
Notes
Variables
0


Stata/SE 15.0 File Edit View Data Graphics Statistics User Window Help
Biostatistics 212 Power and Sapmle Size Calculations.key — Edited
118%
Power and sample-size analysis
Methods organized by:
Filter methods here
Population parameter
Correlations
Hazard rates
Means
ANOVA (multiple means)
One sample
Two independent samples
Two paired samples
Odds ratio
Proportions
R-squared
Regression slopes
Standard deviations
Survival rates
Variances
Outcome
Continuous
Correlations
Means
R-squared
Regression slopes
Standard deviations
Variances
Binary
Linear trend in proportions in Jx2 table
Matched case-control data
One proportion
Stratified 2x2 tables
Two independent proportions
Two paired proportions
Survival
Analysis type
Sample
Test comparing two independent means
Test comparing two independent means in a cluster randomized design


Stata/SE 15.0 File Edit View Data Graphics Statistics User Window Help
power twomeans - Power analysis for a two-sample means test
Main Table Graph Iteration
Compute:
* Accepts numlist (Examples)
Total sample size
Error probabilities
0.05 * Significance level
0.8 * Power
Sample size
1 * Allocation ratio, N2/N1
Allow fractional sample size
Effect size
Means
3.0 * Control
2.7 * Experimental
Standard deviations
Common standard deviation
1 * Common value
Group standard deviations
* Control
* Experimental
Assume a known standard deviations
Sides:
Two-sided test
Treat number lists in starred(*) options as parallel
Submit
Cancel
OK
Methods organized by:
Population parameter
Correlations
Hazard rates
Means
ANOVA (multiple means)
One sample
Two independent samples
Two paired samples
Odds ratio
Proportions
R-squared
Regression slopes
Standard deviations
Survival rates
Variances
Outcome
Analysis type
Sample
Review
Command
/Users/phillipsperera/Desktop/Biostat 212 TA
Notes
Variables 0

Stata/SE 15.0

Open Save Print Log Viewer Graph Do-file Editor Data Editor Data Browser

Review

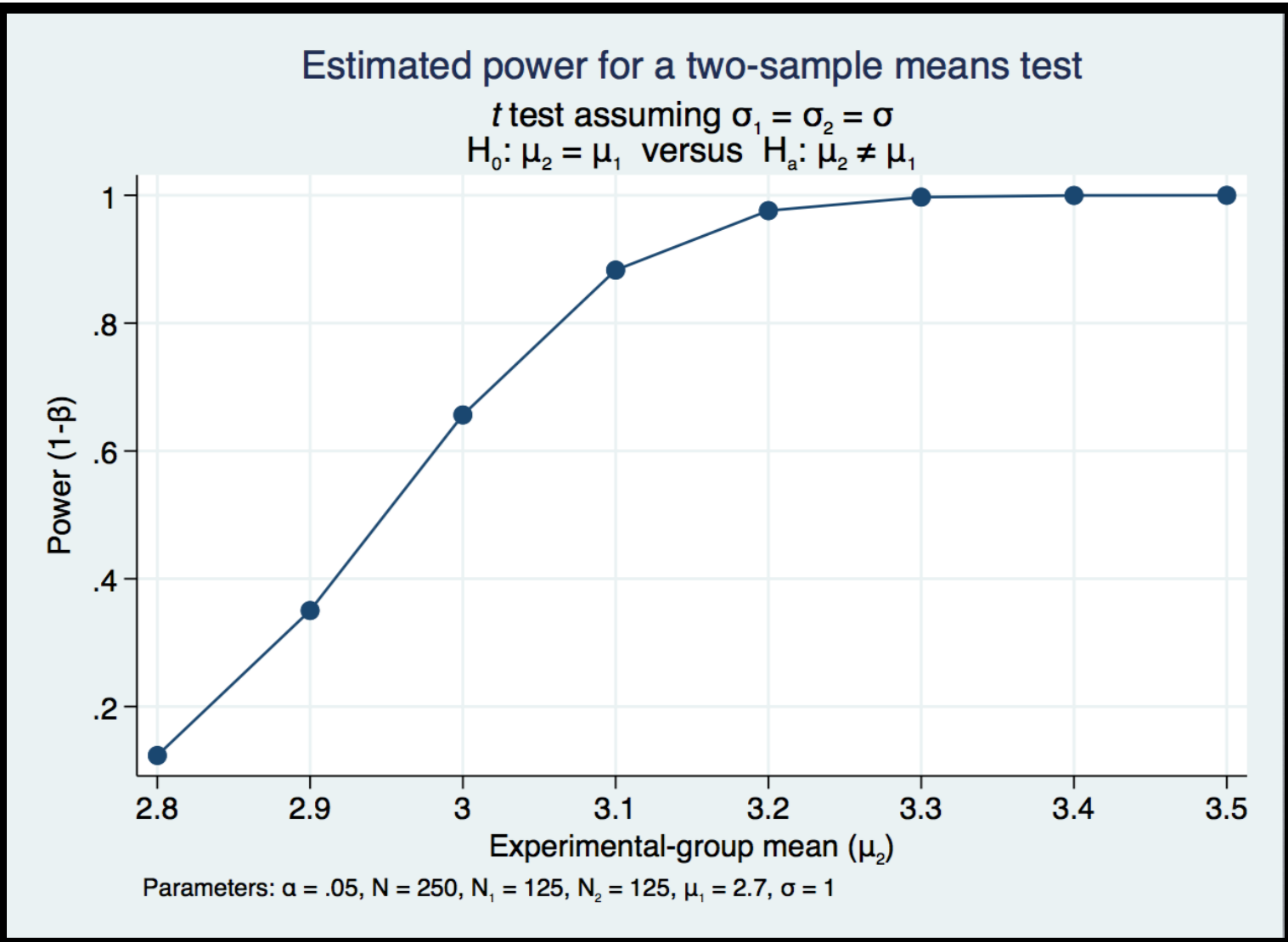
	Command	_rc
1	power twomeans 3.0 2.7	

Results

```
.
. power twomeans 3.0 2.7

Performing iteration ...

Estimated sample sizes for a two-sample means test
t test assuming sd1 = sd2 = sd
Ho: m2 = m1  versus  Ha: m2 != m1

Study parameters:

        alpha =    0.0500
        power =    0.8000
        delta =   -0.3000
           m1 =    3.0000
           m2 =    2.7000
           sd =    1.0000

Estimated sample sizes:

            N =       352
  N per group =       176
```

```
. power twomeans 2.7 (2.8(0.1)3.5), n(250) graph
```

Power Curve:



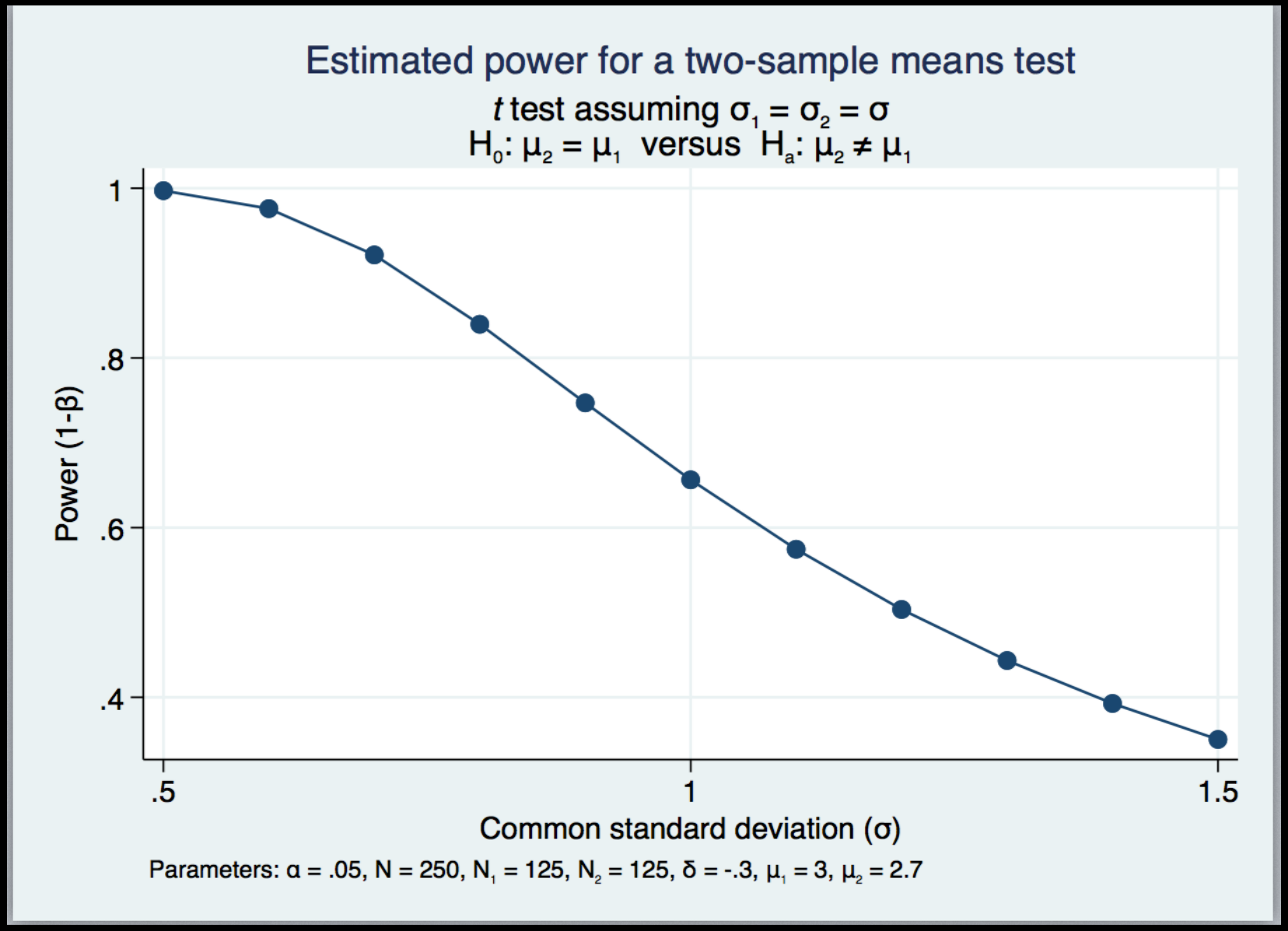
- Increasing mean difference between study groups increases power

```
. power twomeans 3 2.7, sd(0.5(0.1)1.5) n(250)

Estimated power for a two-sample means test
t test assuming sd1 = sd2 = sd
Ho: m2 = m1  versus  Ha: m2 != m1
```

alpha	power	N	N1	N2	delta	m1	m2	sd
.05	.9972	250	125	125	-.3	3	2.7	.5
.05	.976	250	125	125	-.3	3	2.7	.6
.05	.9215	250	125	125	-.3	3	2.7	.7
.05	.8397	250	125	125	-.3	3	2.7	.8
.05	.747	250	125	125	-.3	3	2.7	.9
.05	.6564	250	125	125	-.3	3	2.7	1
.05	.5745	250	125	125	-.3	3	2.7	1.1
.05	.5036	250	125	125	-.3	3	2.7	1.2
.05	.4434	250	125	125	-.3	3	2.7	1.3
.05	.3928	250	125	125	-.3	3	2.7	1.4
.05	.3503	250	125	125	-.3	3	2.7	1.5

- Increasing standard deviation of measurements between study groups decreases power by increasing overlap between groups

```
. power twomeans 3 2.7, sd(0.5(0.1)1.5) n(250) graph
```



Power Calculation:

```
. power twomeans 3.0 2.7, n(250)

Estimated power for a two-sample means test
t test assuming sd1 = sd2 = sd
Ho: m2 = m1  versus  Ha: m2 != m1

Study parameters:

        alpha =    0.0500
            N =       250
  N per group =       125
        delta =   -0.3000
           m1 =    3.0000
           m2 =    2.7000
           sd =    1.0000

Estimated power:

        power =    0.6564
```

Power Calculation:

```
. power twomeans 3.0 2.7, n(1000)

Estimated power for a two-sample means test
t test assuming sd1 = sd2 = sd
Ho: m2 = m1  versus  Ha: m2 != m1

Study parameters:

        alpha =    0.0500
            N =     1,000
  N per group =       500
        delta =   -0.3000
           m1 =    3.0000
           m2 =    2.7000
           sd =    1.0000

Estimated power:

        power =    0.9973
```

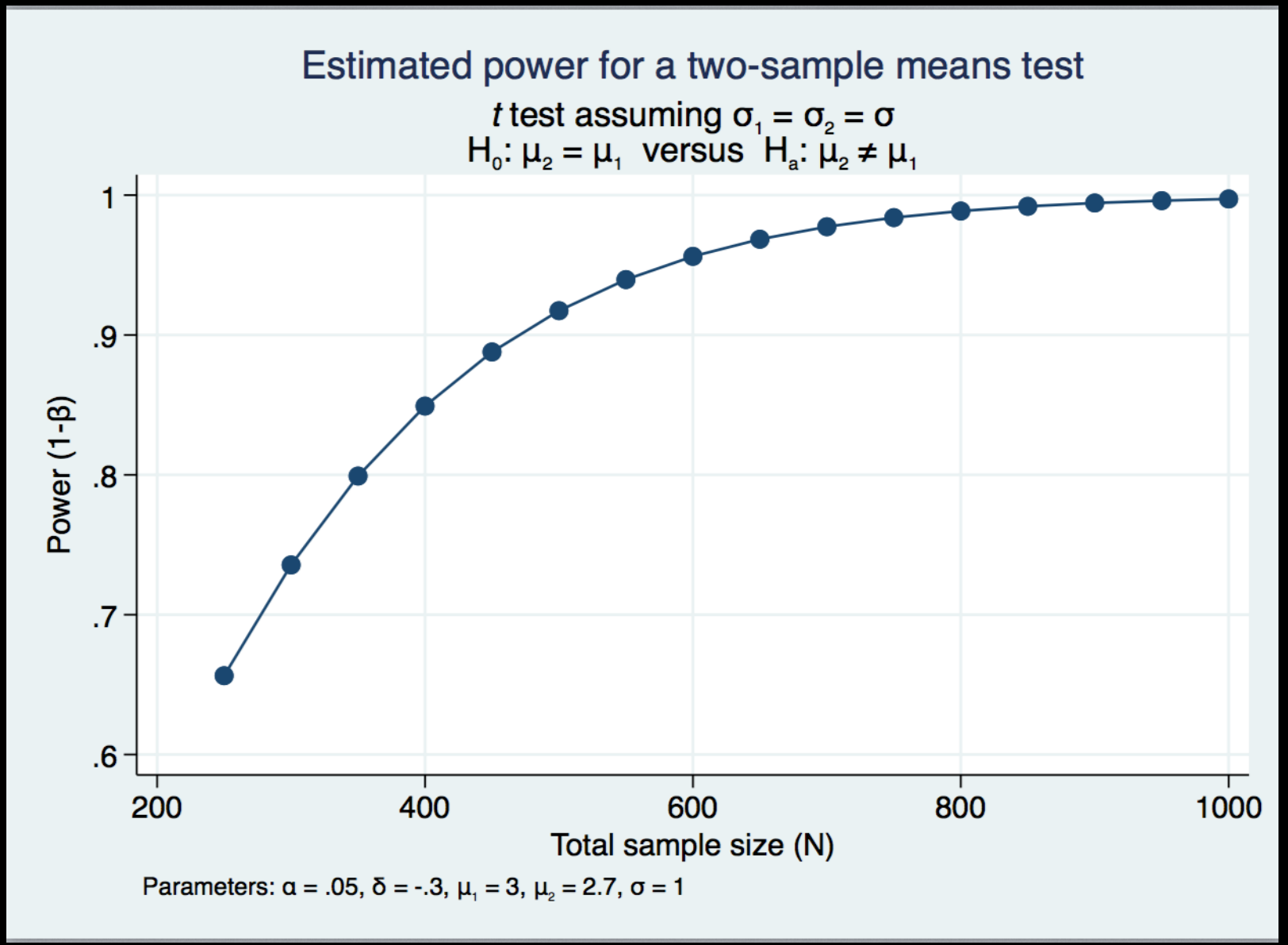
```
. power twomeans 3.0 2.7, n(250(50)1000)
```

Estimated power for a two-sample means test
t test assuming sd1 = sd2 = sd
Ho: m2 = m1 versus Ha: m2 != m1

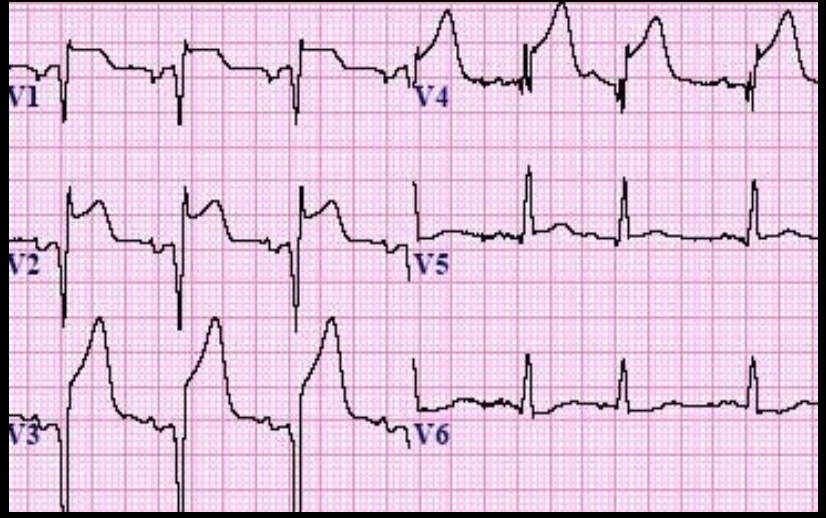
alpha	power	N	N1	N2	delta	m1	m2	sd
.05	.6564	250	125	125	-.3	3	2.7	1
.05	.7356	300	150	150	-.3	3	2.7	1
.05	.7991	350	175	175	-.3	3	2.7	1
.05	.8491	400	200	200	-.3	3	2.7	1
.05	.8879	450	225	225	-.3	3	2.7	1
.05	.9174	500	250	250	-.3	3	2.7	1
.05	.9396	550	275	275	-.3	3	2.7	1
.05	.9562	600	300	300	-.3	3	2.7	1
.05	.9685	650	325	325	-.3	3	2.7	1
.05	.9774	700	350	350	-.3	3	2.7	1
.05	.9839	750	375	375	-.3	3	2.7	1
.05	.9886	800	400	400	-.3	3	2.7	1
.05	.992	850	425	425	-.3	3	2.7	1
.05	.9944	900	450	450	-.3	3	2.7	1
.05	.9961	950	475	475	-.3	3	2.7	1
.05	.9973	1,000	500	500	-.3	3	2.7	1

```
. power twomeans 3.0 2.7, n(250(50)1000) graph
```

Sample Size Curve:



Next Example:



- Let's think of a study looking at the effects of aspirin on the rate of myocardial infarction (MI)
- Estimated proportion of MI in non-users of aspirin (control group) 189/11034 = 0.0171 vs the users of aspirin (experimental group) 104/11037 =0.094
- What test should we use?
- Chi-squared test
- *Used to compare the proportion of subjects in each study group who have a dichotomous outcome (categorical variable)*

```
. power twoproportions 0.0171 .0094

Performing iteration ...

Estimated sample sizes for a two-sample proportions test
Pearson's chi-squared test
Ho: p2 = p1  versus  Ha: p2 != p1

Study parameters:

        alpha =    0.0500
        power =    0.8000
        delta =   -0.0077  (difference)
           p1 =    0.0171
           p2 =    0.0094

Estimated sample sizes:

            N =     6,922
  N per group =     3,461

.
```

Next Example:

- Let's imagine a study looking at the effects of concussion in those who try a new 'super-thin and comfortable' bicycle helmet and fall with blunt head trauma
- Proportion of subjects with concussion in control group (no helmet) 47%
- Proportion of subjects with concussion in experimental group (wearing helmet) 59%
- 32 patient enrolled in control group
- 27 patients enrolled in experimental group
- What is estimated power of study?

```
. power twoproportions 0.47 0.59, n1(32) n2(27)

Estimated power for a two-sample proportions test
Pearson's chi-squared test
Ho: p2 = p1  versus  Ha: p2 != p1

Study parameters:

        alpha =    0.0500
            N =        59
           N1 =        32
           N2 =        27
        N2/N1 =    0.8438
        delta =    0.1200  (difference)
           p1 =    0.4700
           p2 =    0.5900

Estimated power:

        power =    0.1489
```

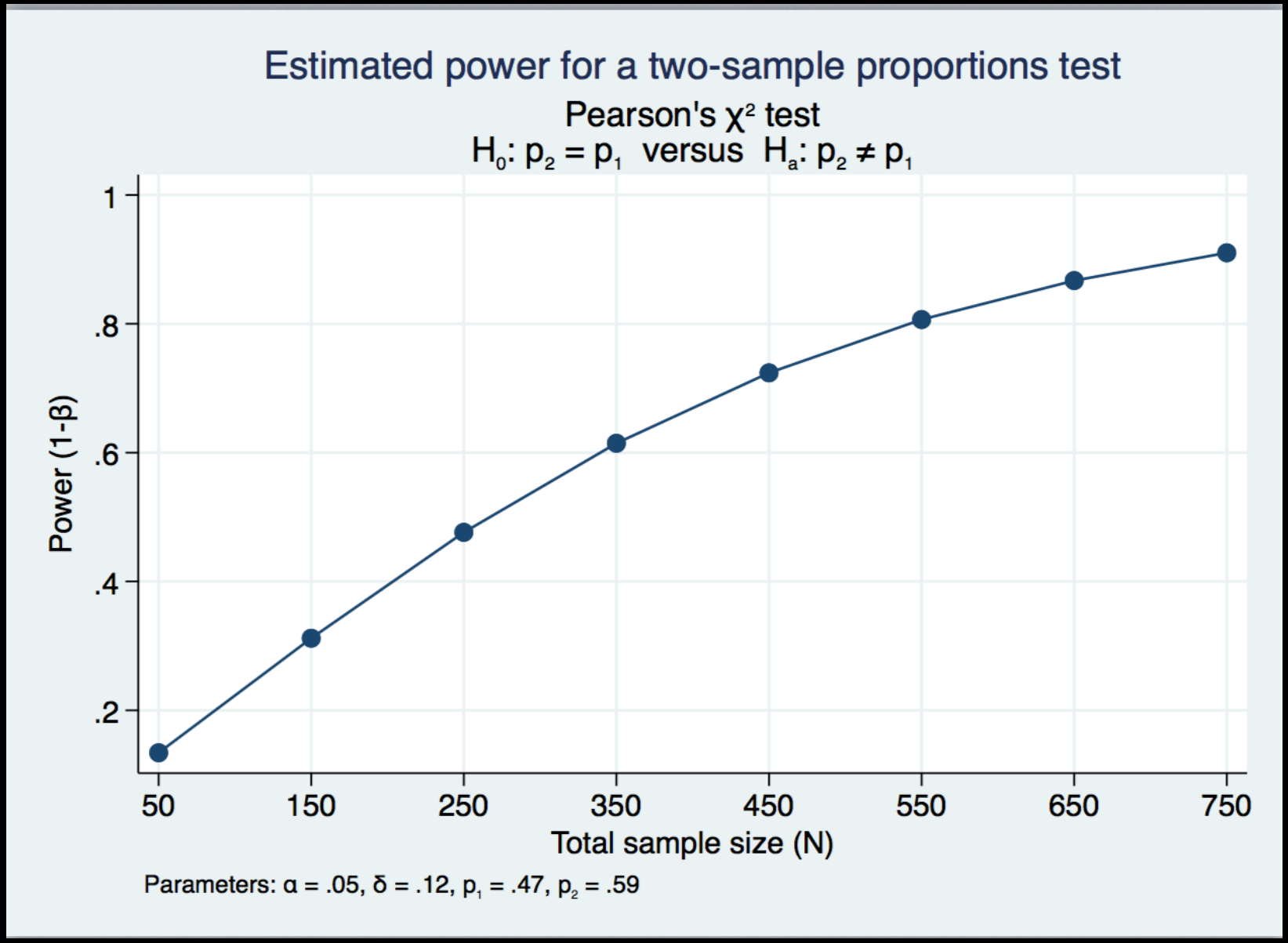
Sample Size Calculation:

```
. power twoproportions 0.47 0.59, n (50(100)750)

Estimated power for a two-sample proportions test
Pearson's chi-squared test
Ho: p2 = p1  versus  Ha: p2 != p1
```

alpha	power	N	N1	N2	delta	p1	p2
.05	.1341	50	25	25	.12	.47	.59
.05	.3119	150	75	75	.12	.47	.59
.05	.4763	250	125	125	.12	.47	.59
.05	.6146	350	175	175	.12	.47	.59
.05	.7239	450	225	225	.12	.47	.59
.05	.8067	550	275	275	.12	.47	.59
.05	.8672	650	325	325	.12	.47	.59
.05	.9102	750	375	375	.12	.47	.59

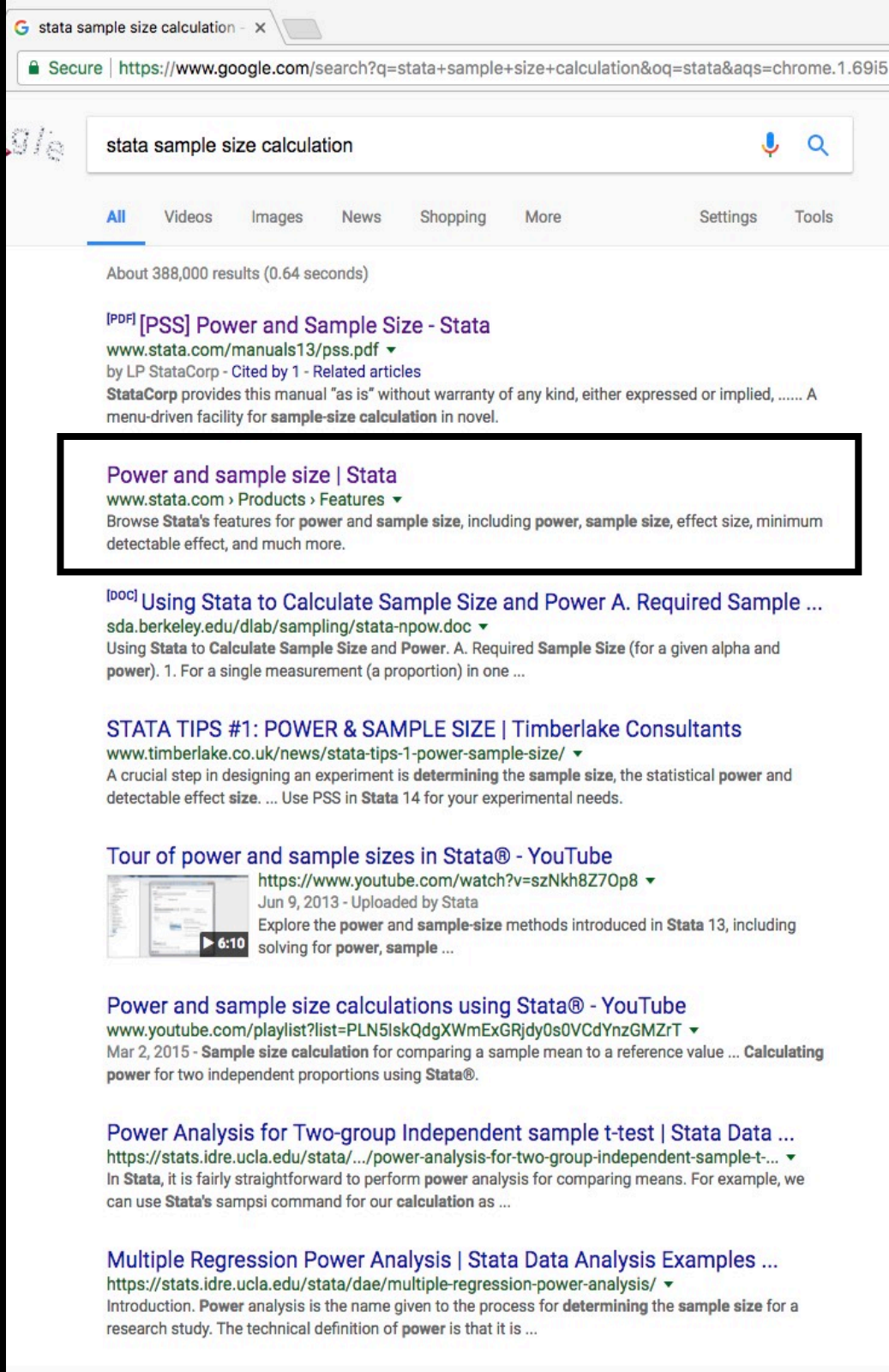
```
. power twoproportions 0.47 0.59, n (50(100)750) graph
```

Sample Size Curve:



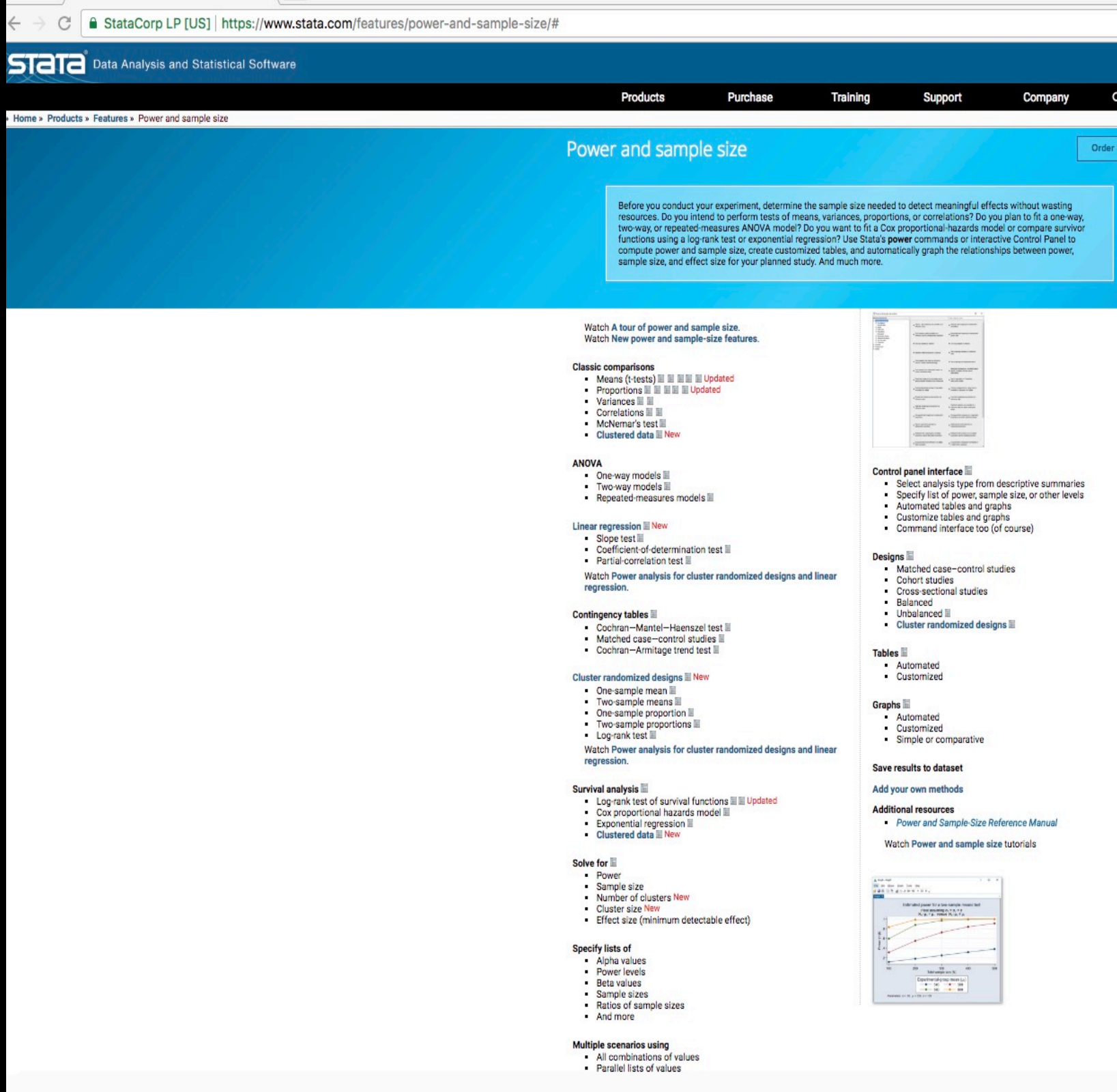
Other Commands to Know:

- Power onemeans
- Power twomeans
- Power pairedmeans
- Power oneproportion
- Power twoproportions
- Power oneway (ANOVA)
- Power mcc
- Power trend
- Power cox (proportional hazards)
- Power log rank

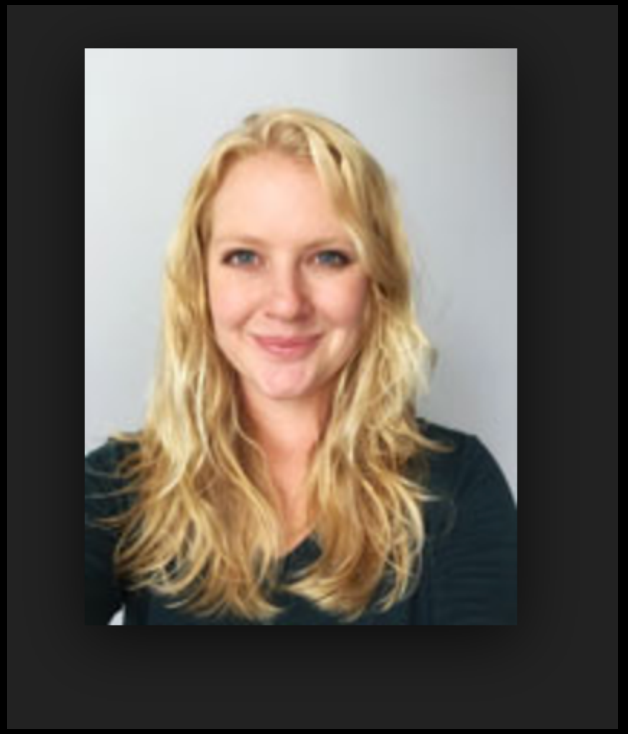


Resources:

- Older **sampsi** command no longer in version 15



Thank You!



* Kristen Aiemjoy, PHD in progress
- A truly awesome teacher!

* Biftu Mengesha, MD
- Gave excellent lecture last year