

Power, precision, and sample-size analysis in Stata

April 7, 2020

Power and sample-size analysis for hypothesis testing

Suppose we want to test whether students who take an SAT coaching program improve their SAT scores by 50 points, on average.

- How many students would we need to recruit to detect a 50-point difference in average SAT scores?
- What is the estimated power of the study if we can only recruit 60 students?

Precision and sample-size analysis for confidence intervals

Suppose we want to construct a confidence interval for the average change in SAT scores for students taking an SAT coaching program.

- If we only have enough resources to include 60 students in our study, what kind of precision would we expect our confidence interval to have?
- How would this precision change as we increase our sample size to 70, 80, or 90?

Overview

- Introduction to power, precision, and sample-size analysis
- Power and sample-size analysis for hypothesis testing
- Precision and sample-size analysis for confidence intervals
- Summary

Introduction to power, precision, and sample-size analysis

Power and sample-size analysis

- Hypothesis tests for inference
 - Test that a population parameter of interest is equal, or smaller, or larger than a specified value.
- Typically, we take the following steps
 1. Obtain a random sample and compute an estimate of the parameter of interest
 2. Calculate a test statistic
 3. Reject the null hypothesis or fail to reject it

Power and sample-size analysis

- Power is the probability of correctly rejecting the null hypothesis when it is false.
- Power analysis estimates the required sample size for a future study to ensure that the test will have high power (chance) to detect when the parameter estimate is different from the specified value.

Precision and sample-size analysis

- Confidence intervals for inference
 - Estimate a plausible range of values for the population parameter
- Like hypothesis tests, confidence intervals are data dependent and so their precision will vary across samples.
- Use precision and sample-size analysis to account for the variability, and plan a study with the desired precision.

Precision and sample-size analysis

- Precision analysis estimates the required sample size for a future study to ensure that the estimated interval will have the desired precision, so that it is not too wide
- Precision and sample-size analysis for confidence intervals is analogous to power and sample-size analysis for hypothesis tests

Power and sample-size analysis

Components of power and sample-size analysis

Study design

- One sample
- Two independent samples
- Paired samples
- Possibly matched on selected characteristics

Components of power and sample-size analysis

Statistical method

- One-sample t test
- Two-sample t test
- Paired t test
- Log-rank test
- And more

Components of power and sample-size analysis

Significance level

- Typically denoted α
- Upper bound for the probability of incorrectly rejecting the null hypothesis.

Power

- Typically denoted $1 - \beta$
- Probability of correctly rejecting the null hypothesis when the null hypothesis is false.

Components of power and sample-size analysis

Effect size

- Magnitude of the effect of interest
- Examples
 - Difference in proportions
 - Odds ratio
 - Log hazard-ratio

Components of power and sample-size analysis

Sample size

- Simple random sample: N = number of observations
- CRD: N = number of clusters and cluster size
- Survival-time data: N = number of events

Power and sample-size analysis

One sample

- Given a power of 90%, what is the smallest detectable difference in average SAT scores if we only recruit 60 students?

Power and sample-size analysis

One sample

- Given a power of 90%, what is the smallest detectable difference in average SAT scores if we only recruit 60 students?

Survival data

- How many subjects would we need to ensure that a hazard ratio of 0.5 can be detected with a power of 90% with a two-sided, 5%-level test?

Power and sample-size analysis

One sample

- Given a power of 90%, what is the smallest detectable difference in average SAT scores if we only recruit 60 students?

Survival data

- How many subjects would we need to ensure that a hazard ratio of 0.5 can be detected with a power of 90% with a two-sided, 5%-level test?

Matched case-control data

- What is the estimated power of detecting an odds ratio of 1.7 with a 5%-level two-sided test, if we have 80 case-control pairs?

The **power** command will help you answer these questions and more.

A first example

Power analysis for a population mean

Consider an example from Tamhane and Dunlop (2000,209).

- On average, students who don't take any coaching programs improve their SAT scores by 15 points when retaking the SAT, with a standard deviation of 40 points.

Power analysis for a population mean

Consider an example from Tamhane and Dunlop (2000,209).

- On average, students who don't take any coaching programs improve their SAT scores by 15 points when retaking the SAT, with a standard deviation of 40 points.
- Suppose a coaching program claims to improve SAT scores by 40 points.

Power analysis for a population mean

Consider an example from Tamhane and Dunlop (2000,209).

- On average, students who don't take any coaching programs improve their SAT scores by 15 points when retaking the SAT, with a standard deviation of 40 points.
- Suppose a coaching program claims to improve SAT scores by 40 points.
- How many students do we need to recruit to detect an average change in SAT scores of 40 points with 80% power using a 5%-level two-sided test?

- Summaries, tables, and tests
- Linear models and related
- Binary outcomes
- Ordinal outcomes
- Categorical outcomes
- Count outcomes
- Fractional outcomes
- Generalized linear models
- Choice 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)
- Multivariate analysis
- Survey data analysis
- Lasso
- Meta-analysis
- Multiple imputation
- Nonparametric analysis
- Exact statistics
- Resampling
- Power, precision, and sample size

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
- ▶ Hypothesis test
- ▶ Confidence interval
- ▶ Sample

Filter methods here

- Test comparing one mean to a reference value
- Test comparing one mean to a reference value in a cluster randomized design
- CI for one mean

Main Table Graph Iteration

Compute:

* Accepts numlist [\(Examples\)](#)

Sample size



Error probabilities

0.05

* Significance level

0.8

* Power



Sample size

☐ Allow fractional sample size

Effect size

Means

* Null

* Alternative



Standard deviation

1

* Standard deviation

☐ Assume a known standard deviation

* Finite population correction:

None



Sides:

Two-sided test

☐ Treat number lists in starred(*) options as parallel

Submit

Cancel

OK

Main Table Graph Iteration

Compute:

* Accepts numlist [\(Examples\)](#)

Sample size



Error probabilities

0.05

* Significance level

0.8

* Power



Sample size

☐ Allow fractional sample size

Effect size

Means

15

* Null

40

* Alternative



Standard deviation

40

* Standard deviation

☐ Assume a known standard deviation

* Finite population correction:

None



Sides:

Two-sided test

☐ Treat number lists in starred(*) options as parallel

Submit

Cancel

OK

Computing sample size

```
. power onemean 15 40, sd(40)
```

Performing iteration ...

Estimated sample size for a one-sample mean test
t test

Ho: $m = m_0$ versus Ha: $m \neq m_0$

Study parameters:

alpha =	0.0500
power =	0.8000
delta =	0.6250
m0 =	15.0000
ma =	40.0000
sd =	40.0000

Estimated sample size:

N =	23
-----	----

Computing power

Assuming our budget allows us to recruit 30 students, what is the power of our study?

Computing power

```
. power onemean 15 40, sd(40) n(30)
```

Estimated power for a one-sample mean test
t test

Ho: $m = m_0$ versus Ha: $m \neq m_0$

Study parameters:

```
alpha =    0.0500  
  N =      30  
delta =    0.6250  
  m0 =    15.0000  
  ma =    40.0000  
  sd =    40.0000
```

Estimated power:

```
power =    0.9112
```

Computing results for a range of parameters

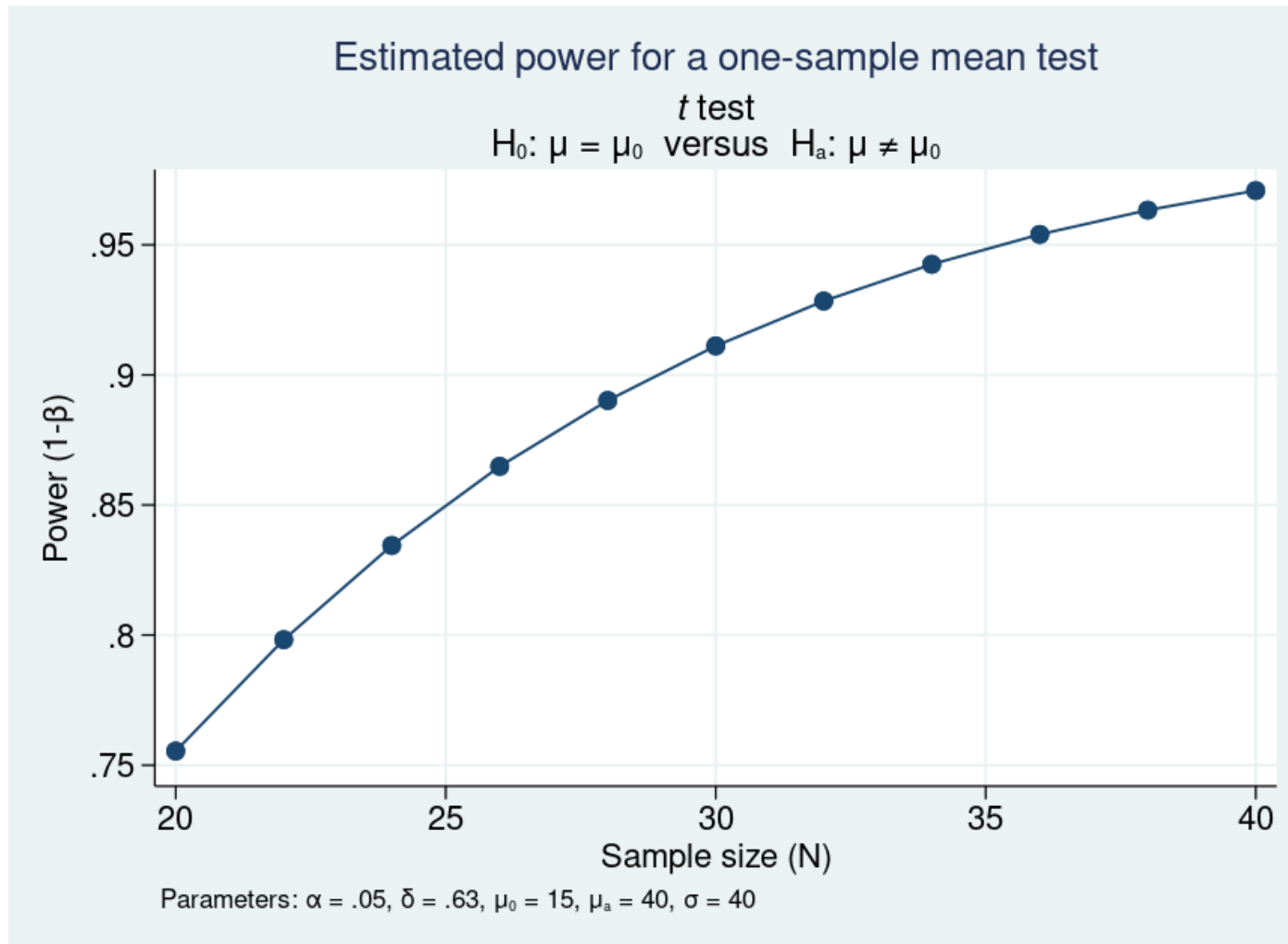
```
. power onemean 15 40, sd(40) n(20(2)40) table graph
```

Estimated power for a one-sample mean test
t test

Ho: $m = m_0$ versus Ha: $m \neq m_0$

alpha	power	N	delta	m0	ma	sd
.05	.7554	20	.625	15	40	40
.05	.7982	22	.625	15	40	40
.05	.8344	24	.625	15	40	40
.05	.8648	26	.625	15	40	40
.05	.8902	28	.625	15	40	40
.05	.9112	30	.625	15	40	40
.05	.9284	32	.625	15	40	40
.05	.9425	34	.625	15	40	40
.05	.954	36	.625	15	40	40
.05	.9634	38	.625	15	40	40
.05	.9709	40	.625	15	40	40

Power curve



See [\[PSS-2\] power, graph](#) to learn more about graphing the results from `power`.

Computing target mean

Assuming we can recruit 30 students, what is the smallest mean change in SAT scores that can be detected with a power of 80%?

Computing target mean

```
. power onemean 15, sd(40) n(30) power(0.8)
```

Performing iteration ...

Estimated target mean for a one-sample mean test
t test

Ho: $m = m_0$ versus Ha: $m \neq m_0$; $m_a > m_0$

Study parameters:

```
alpha =    0.0500
power =    0.8000
  N =         30
  m0 =   15.0000
  sd =   40.0000
```

Estimated effect size and target mean:

```
delta =    0.5292
  ma =   36.1694
```

Computing target mean

```
. power onemean 15, sd(40) n(30) power(0.8)
```

Performing iteration ...

Estimated target mean for a one-sample mean test
t test

Ho: $m = m_0$ versus $H_a: m \neq m_0; m_a > m_0$

Study parameters:

```
alpha =    0.0500
power =    0.8000
  N =         30
  m0 =   15.0000
  sd =   40.0000
```

Estimated effect size and target mean:

```
delta =    0.5292
  ma =   36.1694
```

Estimated effect size for a one-sample
mean test = $(ma - m_0)/sd$

Syntax overview for a population mean

- Compute sample size (assuming default 80% power)
 - `power onemean 15 40, sd(40)`
- Compute power
 - `power onemean 15 40, sd(40) n(30)`
- Compute effect size and target mean
 - `power onemean 15, sd(40) n(30) power(0.8)`

Also see documentation entry [\[PSS-2\] power onemean](#) for details.

General syntax for **power**

- Compute sample size
 - **power** *method* ... [, **power**(*numlist*) *power_options*]
- Compute power
 - **power** *method* ..., **n**(*numlist*) [*power_options*]
- Compute effect size and target parameter
 - **power** *method* ..., **n**(*numlist*) **power**(*numlist*) [*power_options*]

power_options:

table

graph

direction(upper | lower)

...

See [\[PSS-2\] power](#) for an overview of power and sample-size analysis in Stata.

Methods implemented with **power**

One sample

`onemean`

`oneproportion`

`onecorrelation`

`onevariance`

Two independent samples

`twomeans`

`twoproportions`

`twocorrelations`

`twovariances`

Two paired samples

`pairedmeans`

`pairedproportions`

Analysis of variance

`oneway`

`twoway`

`repeated`

Linear regression

`oneslope`

`rsquared`

`pcorr`

Contingency tables

`cmh`

`mcc`

`trend`

Survival analysis

`cox`

`exponential`

`logrank`

Cluster randomized design (CRD)

`onemean, cluster`

`oneproportion, cluster`

`twomeans, cluster`

`twoproportions, cluster`

`logrank, cluster`

User-defined methods

`usermethod`

Power and sample-size analysis with data from a cluster randomized design (CRD)

Data from a simple random sample

- A coaching program claimed to improve SAT scores by 40 points.
- We estimated the sample size required to detect an average change in SAT scores of 40 points with 80% power.
- Now we'll be sampling classes and students within classes, and test scores of students from the same class will be correlated.

Data from a cluster randomized design (CRD)

- In a CRD, the number of clusters and the cluster size play the role of a sample size. We can compute the number of clusters given the cluster size or vice versa.
- We plan on sampling 10 students from each class, and we assume the intraclass correlation is 0.3.
- How many classes do we need to sample from to detect a 25 point difference in the mean change in SAT scores, considering the cluster randomized design?

Computing the number of clusters

```
. power onemean 15 40, sd(40) rho(0.3) m(10)
```

Performing iteration ...

Estimated number of clusters for a one-sample mean test

Cluster randomized design, z test

Ho: $m = m_0$ versus Ha: $m \neq m_0$

Study parameters:

```
alpha =    0.0500
power =    0.8000
delta =    0.3249
m0 =    15.0000
ma =    40.0000
sd =    40.0000
```

Cluster design:

```
M =        10
rho =    0.3000
```

Estimated number of clusters and sample size:

```
K =         8
N =        80
```

Computing cluster size

- What if we instead had a number of classes in mind, and wanted to know how many students we should sample from each class (cluster)?
- Suppose we plan on sampling 12 classes.
- How many students would we need to sample from each class to detect a 25 point difference in the average change in SAT scores?

Computing cluster size

```
. power onemean 15 40, sd(40) rho(0.3) k(12)
```

Performing iteration ...

Estimated cluster size for a one-sample mean test

Cluster randomized design, z test

Ho: $m = m_0$ versus Ha: $m \neq m_0$

Study parameters:

```
alpha =    0.0500
power =    0.8000
delta =    0.4941
m0 =    15.0000
ma =    40.0000
sd =    40.0000
```

Cluster design:

```
K =        12
rho =    0.3000
```

Estimated cluster size and sample size:

```
M =         3
N =        36
```

Computing power

- We previously computed the number of clusters required to detect a 25 point difference in the average change in SAT scores with 80% power, assuming that we would be sampling 10 students from each class, $m(10)$.
- We also computed the cluster size assuming that we would sample from 12 classes, $k(12)$.
- Now let's compute power assuming that we'll sample 10 students per class from 12 classes, $k(12)$ and $m(10)$.

Computing power

```
. power onemean 15 40, sd(40) rho(0.3) k(12) m(10)
```

Estimated power for a one-sample mean test

Cluster randomized design, z test

Ho: $m = m_0$ versus Ha: $m \neq m_0$

Study parameters:

```
alpha =    0.0500
delta =    0.3249
m0 =    15.0000
ma =    40.0000
sd =    40.0000
```

Cluster design:

```
K =        12
M =        10
N =       120
rho =     0.3000
```

Estimated power:

```
power =    0.9451
```

Other hypothesis tests with data from a CRD

- We performed power and sample-size analysis for a population mean with data from a CRD. **power** also supports CRDs for a
 - one-sample proportion test ([\[PSS-2\]power oneproportion, cluster](#))
 - two-sample means test ([\[PSS-2\]power twomeans, cluster](#))
 - two-sample proportions test ([\[PSS-2\]power twoproportions, cluster](#))
 - log-rank test ([\[PSS-2\]power logrank, cluster](#))
- Additionally, you can compute both cluster size and the number of clusters by specifying the total sample size, and you can have varying cluster sizes if you specify the coefficient of variation.

Power and sample-size analysis for matched case-control studies

Matched case-control studies

- The goal is to examine the relationship between disease and exposure, controlling for the effect of confounding variables.

Matched case-control studies

- The goal is to examine the relationship between disease and exposure, controlling for the effect of confounding variables.
- Cases are observations that have the outcome of interest, controls are observations that do not.

Matched case-control studies

- The goal is to examine the relationship between disease and exposure, controlling for the effect of confounding variables.
- Cases are observations that have the outcome of interest, controls are observations that do not.
- Cases are matched to controls based on similar values of the confounding variables.

Matched case-control studies

- The goal is to examine the relationship between disease and exposure, controlling for the effect of confounding variables.
- Cases are observations that have the outcome of interest, controls are observations that do not.
- Cases are matched to controls based on similar values of the confounding variables.
- A study may have a 1:1 or 1: M matched design.

Matched case-control studies

- The goal is to examine the relationship between disease and exposure, controlling for the effect of confounding variables.
- Cases are observations that have the outcome of interest, controls are observations that do not.
- Cases are matched to controls based on similar values of the confounding variables.
- A study may have a 1:1 or 1: M matched design.
- Interest lies in the odds ratio of developing the disease in exposed and unexposed subjects who have equal values of matching variables.

PSS for a matched case-control study

- Consider a study comparing the proportion of smokers and nonsmokers that develop lung cancer.
- We'll take a look at the proportions that previous studies found when matching smokers and nonsmokers on the basis of age, gender, race, and alcohol consumption.

Power analysis for a 1:1 matched case-control study

Lung Cancer (Case)	No Lung Cancer (Control)		Total
	Smoker	Nonsmoker	
Smoker	0.180	0.144	0.324
Nonsmoker	0.040	0.636	0.676
Total	0.220	0.780	1

Power analysis for a 1:1 matched case-control study

Lung Cancer (Case)	No Lung Cancer (Control)		Total
	Smoker	Nonsmoker	
Smoker	0.180	0.144	0.324
Nonsmoker	0.040	0.636	0.676
Total	0.220	0.780	1

Probability of smoking among cases, p_1

Probability of smoking among controls, p_0

Probability of not smoking among controls, q_0

Probability of not smoking among cases, q_1

Power analysis for a 1:1 matched case-control study

Lung Cancer (Case)	No Lung Cancer (Control)		Total
	Smoker	Nonsmoker	
Smoker	0.180	0.144	0.324
Nonsmoker	0.040	0.636	0.676
Total	0.220	0.780	1

Probability of smoking among cases, p_1

Probability of smoking among controls, p_0

Probability of not smoking among controls, q_0

Probability of not smoking among cases, q_1

Assuming there is no correlation of exposure in matched pairs, the odds ratio is calculated as:

$$\theta = p_1 q_0 / p_0 q_1 = (0.324 \times 0.78) / (0.22 \times 0.676) = 1.7$$

Power analysis for a 1:1 matched case-control study

- We want to design a study in which we'll match one case to one control. Our goal is to determine the number of case-control pairs required to achieve 80% power to detect an odds ratio of 1.7 with a 5%-level, two-sided test.
- We're assuming that the proportions from the previous table are representative of the population probabilities, and thus the probability that the sampled control patient is a smoker is 0.22. We'll specify this probability along with the odds ratio, 1.7.

Computing sample size for a 1:1 matched case-control study

```
. power mcc .22, oratio(1.7)
```

Performing iteration ...

Estimated sample size for a matched case-control study

Asymptotic z test, 1:1 matched design

Ho: OR = 1 versus Ha: OR != 1

Study parameters:

alpha =	0.0500
power =	0.8000
delta =	1.7000
p0 =	0.2200
oratio =	1.7000
corr =	0.0000
M =	1

Estimated sample size:

N cases =	285
-----------	-----

Computing power for a one-sided test

- The two-sided test allowed for the odds ratio to be greater than or less than one.

Computing power for a one-sided test

- The two-sided test allowed for the odds ratio to be greater than or less than one.
- We expect the odds of developing lung cancer would be higher for smokers than nonsmokers.

Computing power for a one-sided test

- The two-sided test allowed for the odds ratio to be greater than or less than one.
- We expect the odds of developing lung cancer would be higher for smokers than nonsmokers.
- Say we instead want to test whether the odds ratio is greater than 1.

Computing power for a one-sided test

- The two-sided test allowed for the odds ratio to be greater than or less than one.
- We expect the odds of developing lung cancer would be higher for smokers than nonsmokers.
- Say we instead want to test whether the odds ratio is greater than 1.
- Also, suppose our budget will allow us to recruit 300 subjects.

Computing power for a one-sided test

- The two-sided test allowed for the odds ratio to be greater than or less than one.
- We expect the odds of developing lung cancer would be higher for smokers than nonsmokers.
- Say we instead want to test whether the odds ratio is greater than 1.
- Also, suppose our budget will allow us to recruit 300 subjects.
- What is the estimated power for this 5% level, one-sided test?

Computing power for a one-sided test

```
. power mcc .22, oratio(1.7) n(300) onesided
```

Estimated power for a matched case-control study

Asymptotic z test, 1:1 matched design

Ho: OR = 1 versus Ha: OR > 1

Study parameters:

alpha =	0.0500
N cases =	300
delta =	1.7000
p0 =	0.2200
oratio =	1.7000
corr =	0.0000
M =	1

Estimated power:

power =	0.8931
---------	--------

Computing sample size with correlated exposures

- Previously, we assumed that there was no correlation of exposure in matched pairs.
- Matching based on confounding variables will generally lead to correlation of the exposure between cases and controls.
- In our case, alcohol consumption is correlated with whether an individual smokes.
- If we ignore this correlation, we can underestimate the required sample size, or overestimate power.

Computing sample size with correlated exposures

Below, we use the values from the contingency table to compute the correlation of exposure:

Lung Cancer (Case)	No Lung Cancer (Control)		Total
	Smoker	Nonsmoker	
Smoker	0.180	0.144	0.324
Nonsmoker	0.040	0.636	0.676
Total	0.220	0.780	1

$$\rho = (0.180 \times 0.636 - 0.144 \times 0.04) / \sqrt{0.324 \times 0.676 \times 0.2 \times 0.78} = 0.56$$

Computing sample size with correlated exposures

```
. power mcc .22, oratio(1.7) corr(.56)
```

Performing iteration ...

Estimated sample size for a matched case-control study

Asymptotic z test, 1:1 matched design

Ho: OR = 1 versus Ha: OR != 1

Study parameters:

alpha =	0.0500
power =	0.8000
delta =	1.7000
p0 =	0.2200
oratio =	1.7000
corr =	0.5600
M =	1

Estimated sample size:

N cases =	703
-----------	-----

Sample size for a range of correlated exposures

```
. power mcc .22, oratio(1.7) corr(.4(.05).6)
```

Performing iteration ...

Estimated sample size for a matched case-control study

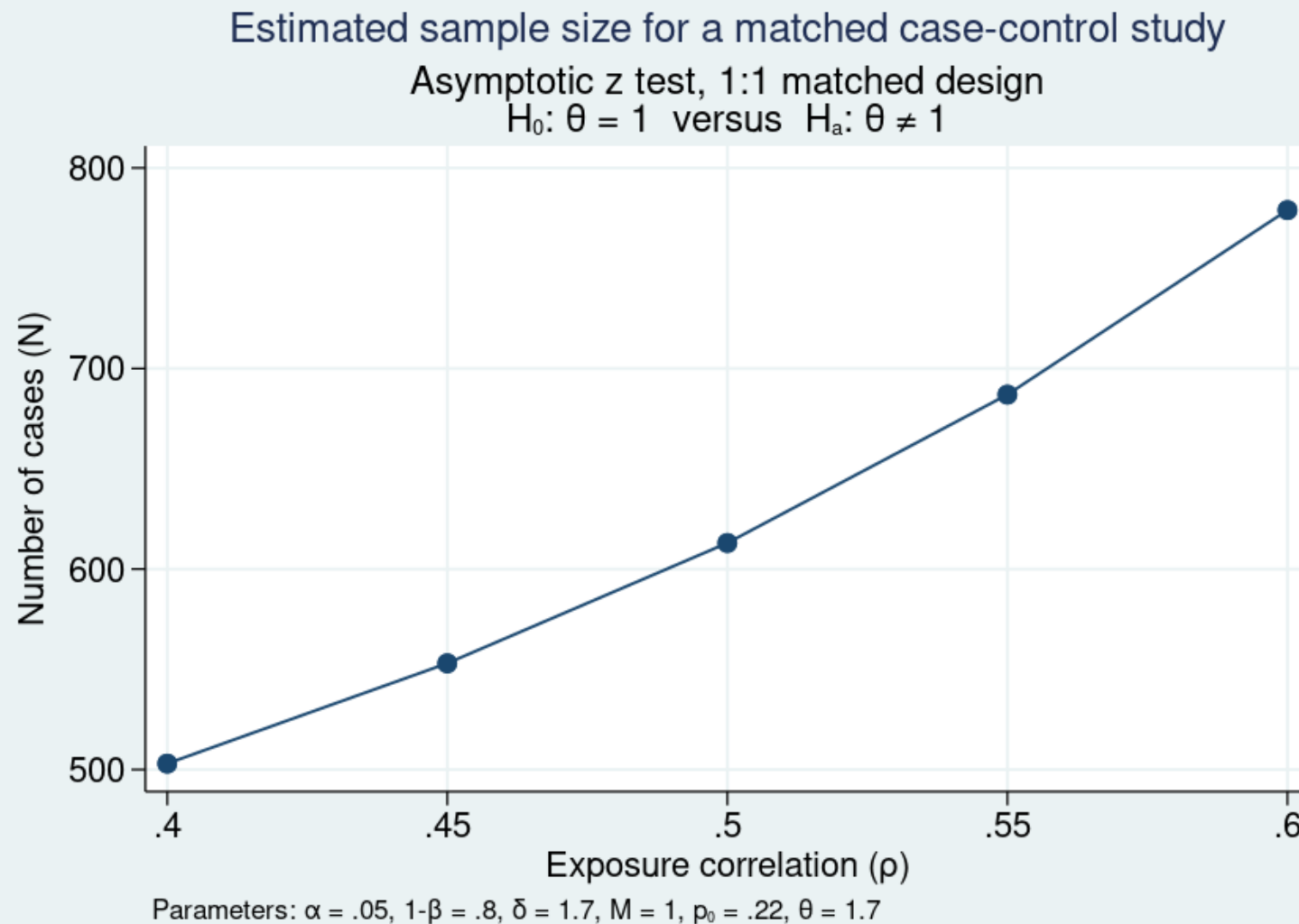
Asymptotic z test, 1:1 matched design

Ho: OR = 1 versus Ha: OR != 1

alpha	power	N	delta	M	p0	oratio	corr
.05	.8	503	1.7	1	.22	1.7	.4
.05	.8	553	1.7	1	.22	1.7	.45
.05	.8	613	1.7	1	.22	1.7	.5
.05	.8	687	1.7	1	.22	1.7	.55
.05	.8	779	1.7	1	.22	1.7	.6

Sensitivity analysis graph

```
. power mcc .22, oratio(1.7) corr(.4(.05).6) graph
```



Power analysis for contingency tables

We performed power and sample-size analysis for a 1:1 matched case-control study.

- See [\[PSS-2\] power mcc](#) for an example of power analysis for a 1: M matched case-control study.

We simply had the binary exposure variable (smoker status) and the outcome (lung cancer).

- If you have data in which a third variable confounds this relationship, take a look at [\[PSS-2\] power cmh](#).
- If your exposure variable is ordinal, and you want to test whether there is a linear trend in exposure and a binary outcome, take a look at [\[PSS-2\] power trend](#).

Power and sample-size analysis for survival-time data

Power analysis for survival-time data

- Survival-time data record the time from an origin to the time an event occurs, typically referred to as a failure.
- Some examples include the time elapsed from a medical diagnosis to death, and the time from which treatment started to the time the patient was fully recovered.
- The sample size in survival analysis is determined by the number of events (or failures).

Power analysis for survival-time data

One model that can be used with this type of data is the Cox proportional hazards model:

$$h(t) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)$$

where $h_0(t)$ is the baseline hazard

- β_1 is the log hazard-ratio associated with a one-unit increase in x_1 .
- Does x_1 have an effect on the time to failure, all else held constant?

Power analysis for survival-time data

- Several factors play a role in the power and sample size requirements for the Cox proportional hazards model.
 - Censoring – Failure may not be observed in every subject.
 - Withdrawal – Subjects may withdraw from the study before it terminates.
 - Dependence between covariates in the model.

Power analysis for survival-time data

Consider a study of multiple-myeloma patients treated with alkylating agents (Krall, Uthoff, and Harley 1975).

- 17 out of 65 patients survived until the end of the study.

Power analysis for survival-time data

Consider a study of multiple-myeloma patients treated with alkylating agents (Krall, Uthoff, and Harley 1975).

- 17 out of 65 patients survived until the end of the study.
- The variable BUN measures the log of the amount of blood urea nitrogen measured in a patient, and its standard deviation is reported to be 0.3126.

Power analysis for survival-time data

Consider a study of multiple-myeloma patients treated with alkylating agents (Krall, Uthoff, and Harley 1975).

- 17 out of 65 patients survived until the end of the study.
- The variable BUN measures the log of the amount of blood urea nitrogen measured in a patient, and its standard deviation is reported to be 0.3126.

We want to compute the sample size required for a one-sided, 5%-level test to detect a log hazard-ratio (β_1) of 1 for a unit increase in BUN with a power of 80%.

Power analysis for survival-time data

Consider a study of multiple-myeloma patients treated with alkylating agents (Krall, Uthoff, and Harley 1975).

- 17 out of 65 patients survived until the end of the study.
- The variable BUN measures the log of the amount of blood urea nitrogen measured in a patient, and its standard deviation is reported to be 0.3126.

We want to compute the sample size required for a one-sided, 5%-level test to detect a log hazard-ratio (β_1) of 1 for a unit increase in BUN with a power of 80%.

- The probability of an event occurring, death, is $48/65=0.738$.

Compute sample size for a Cox PH model

```
. power cox 1, sd(0.3126) onesided eventprob(0.738)
```

Estimated sample size for Cox PH regression

Wald test

Ho: $\beta_1 = 0$ versus Ha: $\beta_1 > 0$

Study parameters:

alpha =	0.0500	
power =	0.8000	
delta =	1.0000	(coefficient)
b1 =	1.0000	
sd =	0.3126	

Censoring:

Pr_E =	0.7380
--------	--------

Estimated number of events and sample size:

E =	64
N =	86

Compute sample size in the presence of correlation between covariates

- Now suppose that we want to adjust the effect of the covariate BUN for eight other covariates in the model. A previous study (Hsieh and Lavori 2000) found the R^2 to be 0.1837, when regressing the log of BUN on eight other covariates.

Compute sample size in the presence of correlation between covariates

```
. power cox 1, sd(0.3126) onesided eventprob(0.738) r2(0.1837)
```

Estimated sample size for Cox PH regression

Wald test

Ho: $\beta_1 = 0$ versus Ha: $\beta_1 > 0$

Study parameters:

alpha =	0.0500	
power =	0.8000	
delta =	1.0000	(coefficient)
b1 =	1.0000	
sd =	0.3126	
R2 =	0.1837	

Censoring:

Pr_E =	0.7380
--------	---------------

Estimated number of events and sample size:

E =	78
N =	106

Compute power in the presence of correlation between covariates

Suppose that our budget will only allow us to recruit 65 subjects. How will this smaller sample size affect the power of the test to detect the alternative?

Compute power in the presence of correlation between covariates

```
. power cox 1, sd(0.3126) onesided eventprob(0.738) r2(0.1837) n(65)
```

Estimated power for Cox PH regression

Wald test

Ho: $\beta_1 = 0$ versus Ha: $\beta_1 > 0$

Study parameters:

alpha =	0.0500	
N =	65	
delta =	1.0000	(coefficient)
b1 =	1.0000	
sd =	0.3126	
R2 =	0.1837	

Number of events and censoring:

E =	48
Pr_E =	0.7380

Estimated power:

power =	0.6222
---------	--------

Compute effect size in the presence of correlation between covariates

- Now we're interested in determining the smallest value of the log hazard-ratio (effect size) that can be detected with 80% power.
- The default is to compute a lower one-sided test, in which the alternative hypothesis is that the log hazard-ratio is less than zero.
- We'll specify **direction (upper)** for an alternative value of the log-hazard ratio greater than zero.

Compute effect size in the presence of correlation between covariates

```
. power cox, sd(0.3126) onesided eventprob(0.738) r2(0.1837) n(65) ///  
> power(0.8) direction(upper)
```

Estimated target coefficient for Cox PH regression

Wald test

Ho: $\beta_1 = 0$ versus Ha: $\beta_1 > 0$

Study parameters:

```
alpha = 0.0500  
power = 0.8000  
N = 65  
sd = 0.3126  
R2 = 0.1837
```

Number of events and censoring:

```
E = 48  
Pr_E = 0.7380
```

Estimated effect size and target coefficient:

```
delta = 1.2711 (coefficient)  
b1 = 1.2711
```

Power analysis for survival-time data

- For more examples of performing power analysis for the Cox proportional hazards model, see [\[PSS-2\] power cox](#).
- You can also use [\[PSS-2\] power logrank](#) for power and sample-size computations for survival analysis comparing survivor functions in two groups by using the log-rank test.
- See [\[PSS-2\] power exponential](#) for sample size and power computations for survival analysis comparing two exponential survivor functions by using parametric tests for the difference between hazards or log hazards.
 - Accommodates flexible accrual of subjects into the study and group-specific losses to follow-up.

What else is available with **power**?

- We can also perform power and sample-size analyses for
 - One-sample proportion test `-- power oneproportion`
 - One-sample correlation test `-- power onecorrelation`
 - One-sample variance test `-- power onevariance`
 - Two-sample means tests `-- power twomeans`
 - Two-sample proportions test `-- power twoproportions`
 - Two-sample correlations test `-- power twocorrelations`
 - Two-sample variances test `-- power twovariances`
 - Paired-means test `-- power pairedmeans`
 - Paired-proportions test `-- power pairedproportions`
 - One-way ANOVA `-- power oneway`
 - Two-way ANOVA `-- power twoway`
 - Repeated-measures ANOVA `-- power repeated`
 - Slope test in a simple linear regression `-- power oneslope`
 - R^2 test in a multiple linear regression `-- power rsquared`
 - Partial-correlation test in a multiple linear regression `-- power pcorr`
- And you can add your own methods to **power**. See [\[PSS-2\] power](#) for details and more examples.

Adding your own methods to **power**

Adding your own method to **power**

Adding your own methods to **power** is easy. Suppose you want to add a method called **mymethod** to **power**. Simply

1. write an r-class program called **power_cmd_mymethod** that computes power, sample size, or effect size and follows **power**'s convention for naming common options and storing results; and
2. place the program where Stata can find it.

You are done. You can now use **mymethod** within **power** like any other official **power** method.

Program for computing power for a one-sample z test

```
program power_cmd_myztest, rclass  
version 16.1
```

Program for computing power for a one-sample z test

```
program power_cmd_myztest, rclass  
version 16.1  
syntax, n(integer) STDDiff(real) Alpha(string)
```

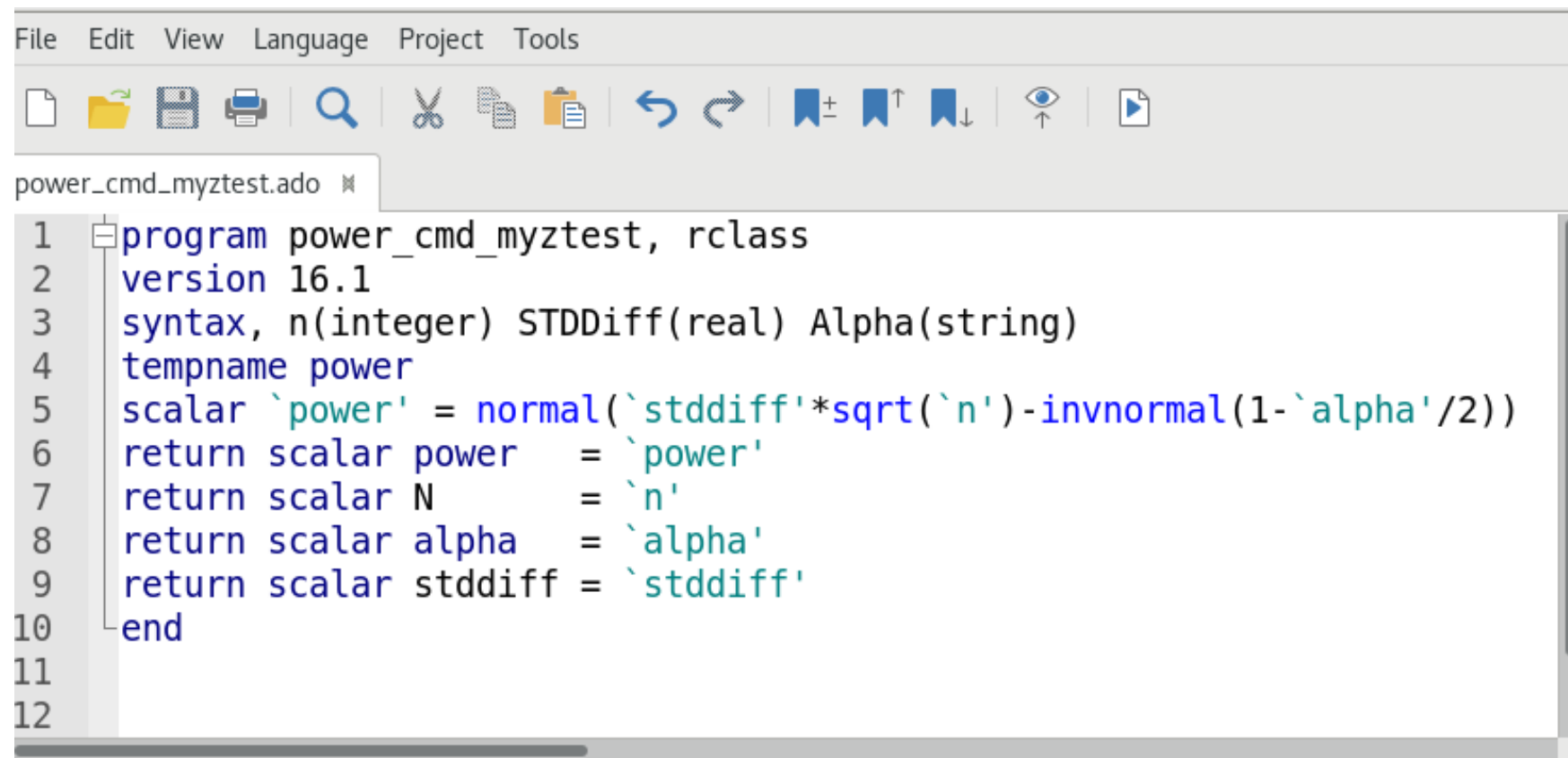
Program for computing power for a one-sample z test

```
program power_cmd_myztest, rclass
version 16.1
syntax, n(integer) STDDiff(real) Alpha(string)
tempname power
scalar `power' = normal(`stddiff'*sqrt(`n')-invnormal(1-`alpha'/2))
```

Program for computing power for a one-sample z test

```
program power_cmd_myztest, rclass
version 16.1
syntax, n(integer) STDDiff(real) Alpha(string)
tempname power
scalar `power' = normal(`stddiff'*sqrt(`n')-invnormal(1-`alpha'/2))
return scalar power = `power'
return scalar N = `n'
return scalar alpha = `alpha'
return scalar stddiff = `stddiff'
end
```

Program for computing power for a one-sample z test



The image shows a screenshot of a Stata command editor window. The title bar at the top reads "power_cmd_myztest.ado". Below the title bar is a menu bar with "File", "Edit", "View", "Language", "Project", and "Tools". Under the "Edit" menu, there is a toolbar with icons for file operations (new, open, save, print, find, copy, paste, undo, redo) and navigation (bookmarks, search, zoom). The main editing area contains the following Stata code:

```
1 program power_cmd_myztest, rclass
2 version 16.1
3 syntax, n(integer) STDDiff(real) Alpha(string)
4 tempname power
5 scalar `power' = normal(`stddiff'*sqrt(`n')-invnormal(1-`alpha'/2))
6 return scalar power = `power'
7 return scalar N = `n'
8 return scalar alpha = `alpha'
9 return scalar stddiff = `stddiff'
10 end
11
12
```

power myztest : Compute power

```
. power myztest, alpha(0.05) n(10) stddiff(0.25)
```

Estimated power

Two-sided test

alpha	power	N
.05	.1211	10

power myztest : Specifying multiple values

```
. power myztest, alpha(0.01 0.05) n(10(10)100) stddiff(0.25)
```

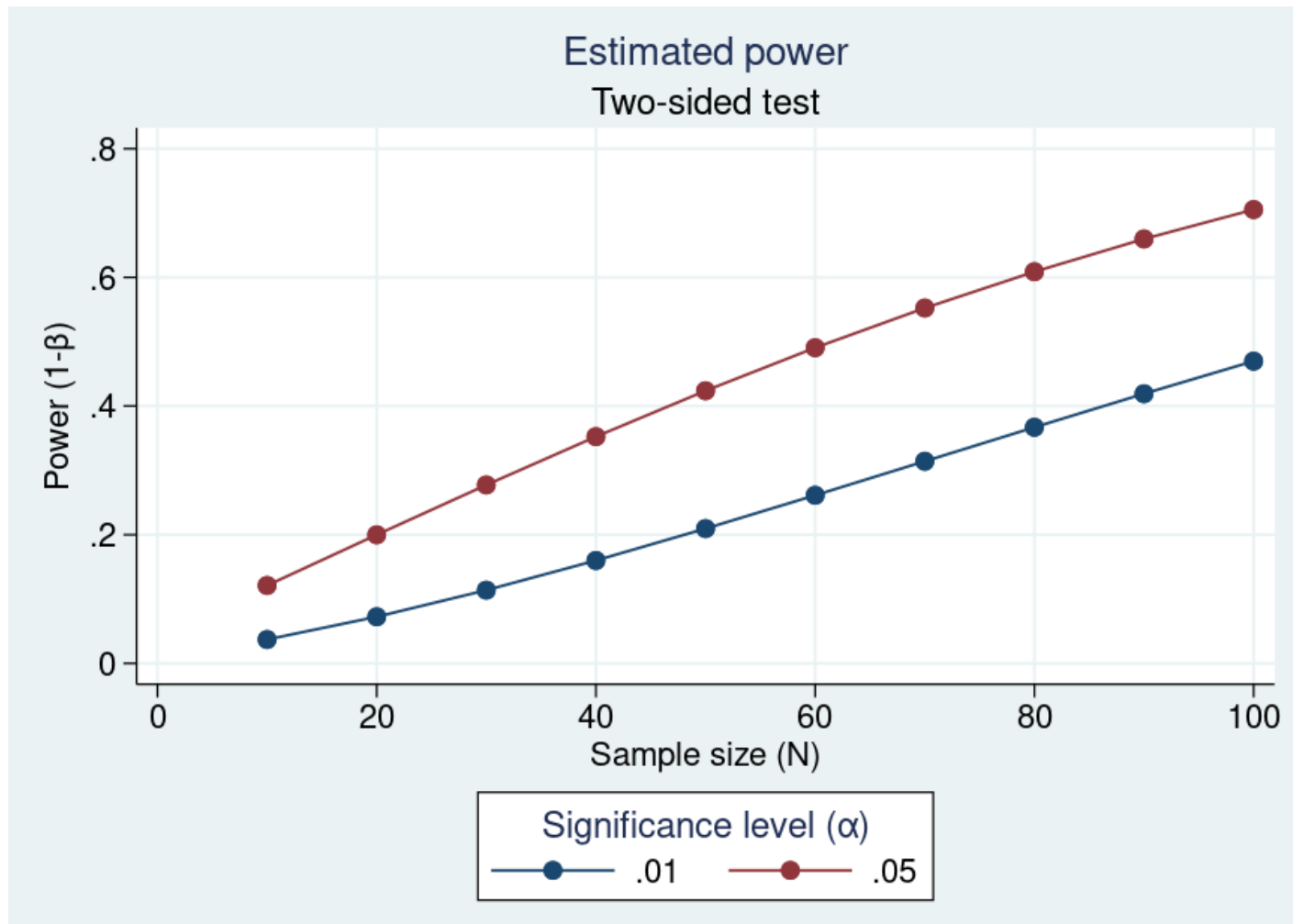
Estimated power

Two-sided test

alpha	power	N
.01	.03711	10
.01	.07245	20
.01	.1138	30
.01	.1599	40
.01	.2095	50
.01	.2613	60
.01	.3141	70
.01	.367	80
.01	.4191	90
.01	.4698	100
.05	.1211	10
.05	.1999	20
.05	.2774	30
.05	.3524	40
.05	.4238	50
.05	.4906	60
.05	.5524	70
.05	.6088	80
.05	.6597	90
.05	.7054	100

power myztest : Automatic graphs

```
. power myztest, alpha(0.01 0.05) n(10(10)100) stddiff(0.25) graph
```



Customizing your **power** command

- Add method-specific options and set them up to allow multiple values
- Tables
 - Change column labels, formats, and widths to modify the look and contents for the table created by default
- Graphs
 - Change the default column labels
 - Use different symbols to label the results
- See [\[PSS-2\] *power usemethod*](#) for more examples.

Precision and sample-size analysis

Motivating questions

Suppose you want to construct an interval for the average weight loss for a new exercise program.

- How many subjects would be required to ensure that the confidence interval for the mean is no wider than 2?
- If we only have enough resources to include 40 subjects in our study, what kind of precision would we expect our confidence interval to have?
- How would this precision change as we increase our sample size to 50, 60, and 70?

The **ciwidth** command will help you answer these questions and more.

Overview of `ciwidth` examples

- Perform precision and sample-size analysis for confidence intervals
 - Population mean
 - Comparison of means from independent samples
 - Comparison of means from paired samples
- Compute
 - Sample size
 - CI precision
 - Probability of CI precision
- Perform sensitivity analysis
 - Present results graphically or in a table
 - Customize graphs and tables

Computation of a confidence interval

Confidence intervals

- Computation
 - Confidence level
 - 95%, 97%, other
 - Sample size
 - How many participants can you afford to have in your study?
 - Standard deviation
 - How did weight loss vary in the sample?
- Precision
 - Measured by the CI width
 - Ensured by the probability of CI width

Confidence interval for a population mean

A $100 \times (1 - \alpha)\%$ CI:

$$\left[\bar{x} - \underbrace{t_{n-1, 1-\alpha/2} \left(\frac{s}{\sqrt{n}} \right)}_{1/2 \text{ width}}, \bar{x} + \underbrace{t_{n-1, 1-\alpha/2} \left(\frac{s}{\sqrt{n}} \right)}_{1/2 \text{ width}} \right]$$

α = significance level

s = sample standard deviation

n = sample size

Confidence interval for a population mean

$$\left[\bar{x} - t_{n-1, 1-\alpha/2} \left(\frac{s}{\sqrt{n}} \right), \bar{x} + t_{n-1, 1-\alpha/2} \left(\frac{s}{\sqrt{n}} \right) \right]$$

Sample mean	Width	Sample mean \pm half-width	95% Confidence interval
5	10	5-5, 5+5	0 ————— 10
5	8	5-4, 5+4	1 ————— 9
5	4	5-2, 5+2	3 ————— 7

Confidence interval for a population mean

$$\left[\bar{x} - t_{n-1, 1-\alpha/2} \left(\frac{s}{\sqrt{n}} \right), \bar{x} + t_{n-1, 1-\alpha/2} \left(\frac{s}{\sqrt{n}} \right) \right]$$

Choose the values of the **confidence level** ($100 \times (1-\alpha)$), **sample standard deviation**, and **sample size** that will provide the desired level of precision, given the resources you have

A first example

Precision analysis for a population mean

- How long do plug-in air fresheners last?



- How many air fresheners would we need to test to obtain a two-sided 95% CI for the mean scent duration with a width no larger than 4 days?

Computing sample size for a population mean

```
. ciwidth onemean, sd(6) probwidth(0.96) width(4)
```

Performing iteration ...

Estimated sample size for a one-mean CI
Student's t two-sided CI

Study parameters:

level =	95.00
Pr_width =	0.9600
width =	4.0000
sd =	6.0000

Estimated sample size:

N =	51
-----	-----------

Computing CI width for a population mean

How big of an interval width would we expect if we could only afford to sample 30 air fresheners?

Computing CI width for a population mean

```
. ciwidth onemean, sd(6) probwidth(0.96) n(30)
```

Estimated width for a one-mean CI
Student's t two-sided CI

Study parameters:

level =	95.00
N =	30
Pr_width =	0.9600
sd =	6.0000

Estimated width:

width =	5.4945
---------	--------

Computing probability of CI width for a population mean

What's the probability that we'll obtain a desired CI width of 4, if we only sample 30 air fresheners?

Computing probability of CI width for a population mean

```
. ciwidth onemean, sd(6) width(4) n(30)
```

Estimated probability of width for a one-mean CI
Student's t two-sided CI

Study parameters:

```
level =    95.00  
  N =      30  
width =    4.0000  
  sd =    6.0000
```

Estimated probability of width:

```
Pr_width =    0.2285
```

Syntax overview for a one-mean CI

- Estimate sample size
 - `ciwidth onemean, sd(6) probwidth(0.96) width(4)`
- Estimate CI width
 - `ciwidth onemean, sd(6) probwidth(0.96) n(30)`
- Estimate probability of CI width
 - `ciwidth onemean, sd(6) width(4) n(30)`
- Also see documentation entry [\[PSS-2\] ciwidth onemean](#) for details.

General syntax for `ciwidth`

- Estimate sample size
 - `ciwidth method ..., probwidth() width() ...`
- Estimate CI width
 - `ciwidth method ..., probwidth() n()...`
- Estimate probability of CI width
 - `ciwidth method ..., width() n()...`

method:

`onemean`

`onevariance`

`twomeans`

`pairedmeans`

`usermethod`

General syntax for `ciwidth`

- Estimate sample size
 - `ciwidth method ..., probwidth(numlist) width(numlist) [options]`
- Estimate CI width
 - `ciwidth method ..., probwidth(numlist) n(numlist) [options]`
- Estimate probability of CI width
 - `ciwidth method ..., width(numlist) n(numlist) [options]`

method:

`onemean`
`onevariance`
`twomeans`
`pairedmeans`
`usermethod`

options:

`table`
`graph`
`lower`
`upper`
`...`

- See [\[PSS-3\] ciwidth](#) for an overview of precision and sample-size analysis in Stata.

One-sided confidence intervals

One-sided confidence intervals

- You may want one-sided confidence intervals if
 - You are certain of the direction of an effect.
 - Examples include toxicity studies and analysis of occurrences of adverse drug reaction data (Chow, Shao, Wang, & Lokhnygina, 2017).
 - You are mainly interested in either the lower or upper limit.
 - An example would be product quality and reliability (Meeker, Hahn, & Escobar, 2017).
- In general, use caution with one-sided confidence intervals.
 - There have been cases where the placebo effect is superior to the actual drug effect (Meeker, Hahn, & Escobar, 2017).

Lower 95% confidence interval

```
. ciwidth onemean, sd(6) probwidth(0.96) n(30) lower
```

Estimated width for a one-mean CI
Student's t lower CI

Study parameters:

level =	95.00
N =	30
Pr_width =	0.9600
sd =	6.0000

Estimated width:

width =	2.2823
---------	--------

Sensitivity analysis

Sensitivity analysis

- Precision and sample-size analysis is centered on a prospective study, and we are obtaining estimates based on values that are truly unknown.
 - Estimates may come from historical data or pilot studies
- Sensitivity analysis is used to investigate the effect of varying study parameters on CI precision, probability of CI width, sample size, and other components of a study.
 - In Stata, you can perform sensitivity analysis with tables and graphs.

Computing probability of CI width for a population mean

```
. ciwidth onemean, sd(6) width(4) n(30)
```

Estimated probability of width for a one-mean CI
Student's t two-sided CI

Study parameters:

level =	95.00
N =	30
width =	4.0000
sd =	6.0000

Estimated probability of width:

Pr_width =	0.2285
------------	--------

Probability of CI width for a range of standard deviations

```
. ciwidth onemean, sd(6 7 8 9) width(4) n(45)
```

Estimated probability of width for a one-mean CI
Student's t two-sided CI

level	N	Pr_width	width	sd
95	45	.8599	4	6
95	45	.3476	4	7
95	45	.06042	4	8
95	45	.00627	4	9

Probability of CI width for ranges of multiple parameters

```
. ciwidth onemean, sd(6 7 8 9) width(4) n(45(10)65)
```

Estimated probability of width for a one-mean CI
Student's t two-sided CI

level	N	Pr_width	width	sd
95	45	.8599	4	6
95	45	.3476	4	7
95	45	.06042	4	8
95	45	.00627	4	9
95	55	.9918	4	6
95	55	.7419	4	7
95	55	.2336	4	8
95	55	.03245	4	9
95	65	.9999	4	6
95	65	.9599	4	7
95	65	.5635	4	8
95	65	.1295	4	9

Tables for sensitivity analysis

```
. ciwidth onemean, sd(6 7 8 9) width(4) n(45(10)65) table(, separator(4))
```

Estimated probability of width for a one-mean CI
Student's t two-sided CI

level	N	Pr_width	width	sd
95	45	.8599	4	6
95	45	.3476	4	7
95	45	.06042	4	8
95	45	.00627	4	9
95	55	.9918	4	6
95	55	.7419	4	7
95	55	.2336	4	8
95	55	.03245	4	9
95	65	.9999	4	6
95	65	.9599	4	7
95	65	.5635	4	8
95	65	.1295	4	9

Reordering table columns

```
. ciwidth onemean, sd(6 7 8 9) width(4) n(45(10)65) table(N sd Pr_width width, separator(4))
```

Estimated probability of width for a one-mean CI
Student's t two-sided CI

N	sd	Pr_width	width
45	6	.8599	4
45	7	.3476	4
45	8	.06042	4
45	9	.00627	4
55	6	.9918	4
55	7	.7419	4
55	8	.2336	4
55	9	.03245	4
65	6	.9999	4
65	7	.9599	4
65	8	.5635	4
65	9	.1295	4

See [\[PSS-3\] ciwidth, table](#) to learn more about producing tables with results from `ciwidth`.

Paired samples

Paired-means for bad (LDL) cholesterol levels

Cholesterol _{Before} (mg/dL)	Cholesterol _{After} (mg/dL)	Difference
167	140	-27
149	138	-11
192	194	2
200	180	-20
162	162	0
52	41	67.25
...

Mean	129.58	111.96	-17.62
S.D.	52.27	42.15	6.47

CI for a paired-means difference

```
. ciwidth pairedmeans, width(6) probwidth(0.98) sddiff(18)
```

CI for a paired-means difference

```
. ciwidth pairedmeans, width(6) probwidth(0.98) sddiff(18)
```

Performing iteration ...

Estimated sample size for a paired-means-difference CI
Student's t two-sided CI

Study parameters:

```
level =    95.0000  
Pr_width =   0.9800  
width =     6.0000  
sd_d =    18.0000
```

Estimated sample size:

```
N =      174
```

Sample-size determination using the correlation

```
. ciwidth pairedmeans, width(6) probwidth(0.98) sd1(16) sd2(23) corr(0.78)
```

Performing iteration ...

Estimated sample size for a paired-means-difference CI
Student's t two-sided CI

Study parameters:

level =	95.0000	sd1 =	16.0000
Pr_width =	0.9800	sd2 =	23.0000
width =	6.0000	corr =	0.7800
sd_d =	14.5231		

Estimated sample size:

N = **119**

See [\[PSS-3\] ciwidth pairedmeans](#) to learn more about precision and sample-size analysis for a paired-means-difference confidence interval.

CI for a difference of two means

- Let's consider the cholesterol example, but instead randomly sample individuals who take cholesterol medication and individuals who don't
 - Two independent samples
- Group sizes may differ

CI for a difference of two means

```
. ciwidth twomeans, width(6) probwidth(0.99) sd(12)
```

Performing iteration ...

Estimated sample sizes for a two-means-difference CI
Student's t two-sided CI assuming $sd_1 = sd_2 = sd$

Study parameters:

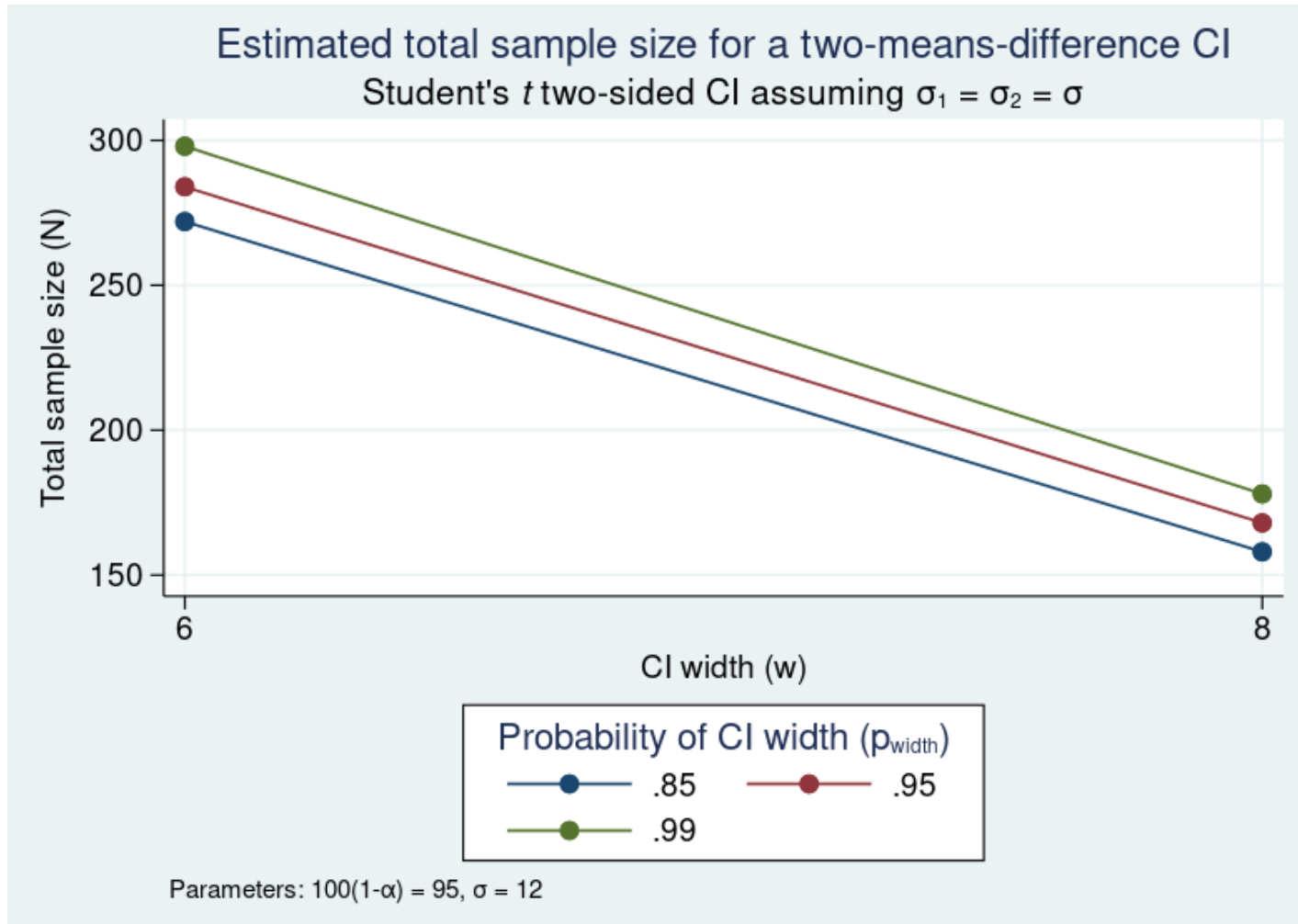
```
level =      95.00  
Pr_width =    0.9900  
width =      6.0000  
sd =      12.0000
```

Estimated sample sizes:

```
      N =      298  
N per group =    149
```

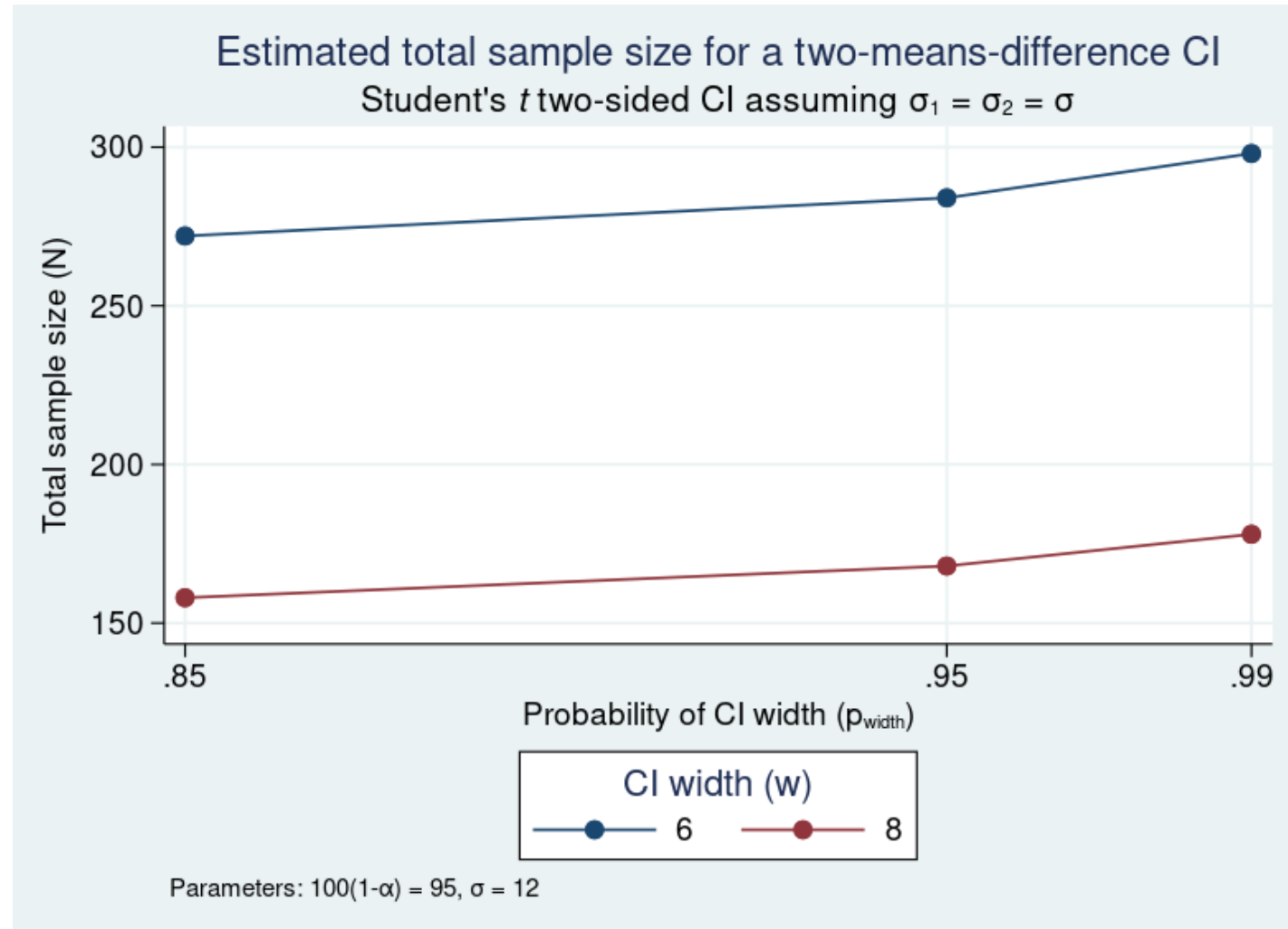
Graphs for sensitivity analysis

```
. ciwidth twomeans, width(6 8) probwidth(0.85 0.95 0.99) sd(12) graph
```



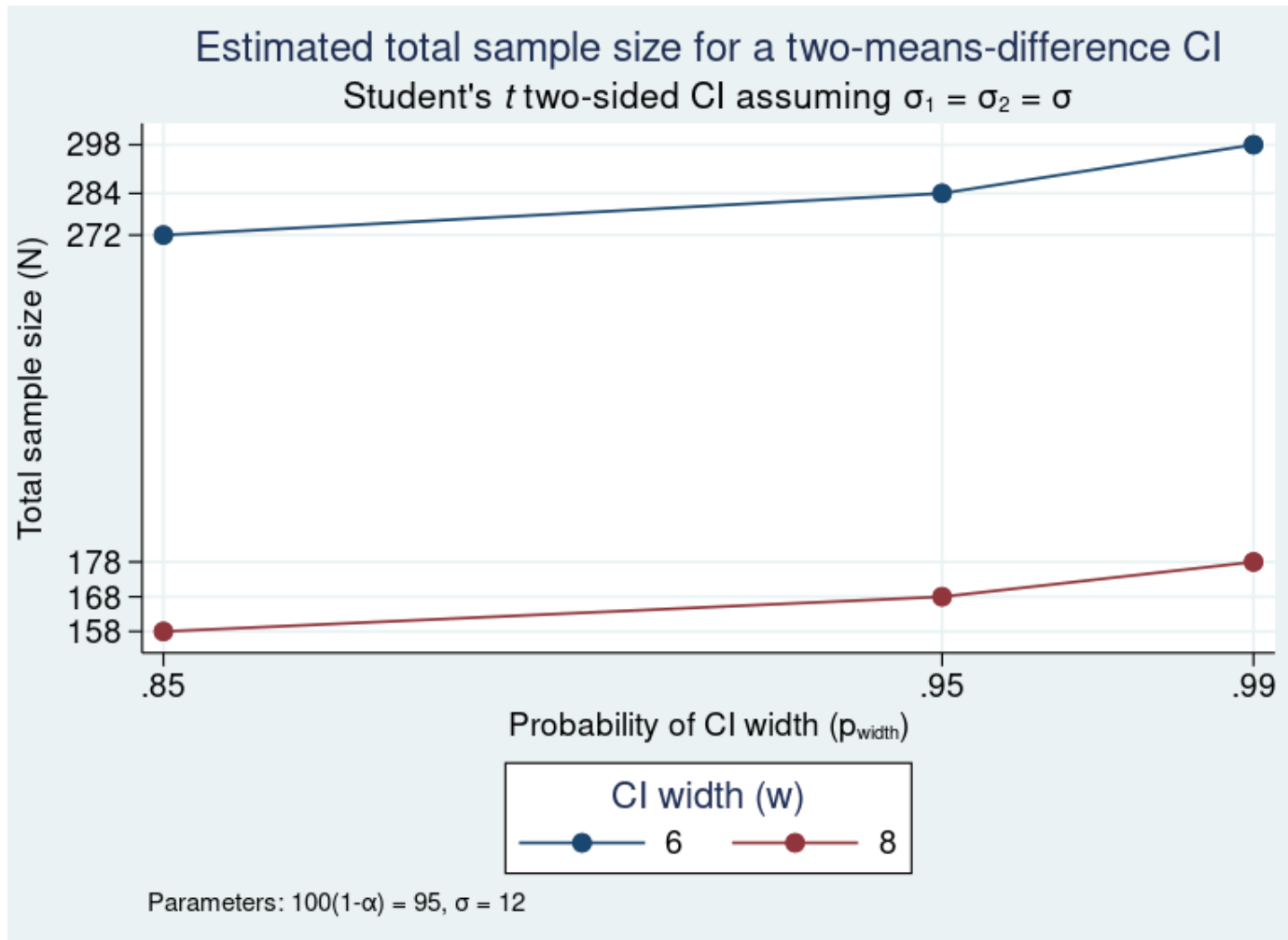
Define the x axis

```
. ciwidth twomeans, width(6 8) probwidth(0.85 0.95 0.99) sd(12)  
> graph(xdimension(Pr_width))
```



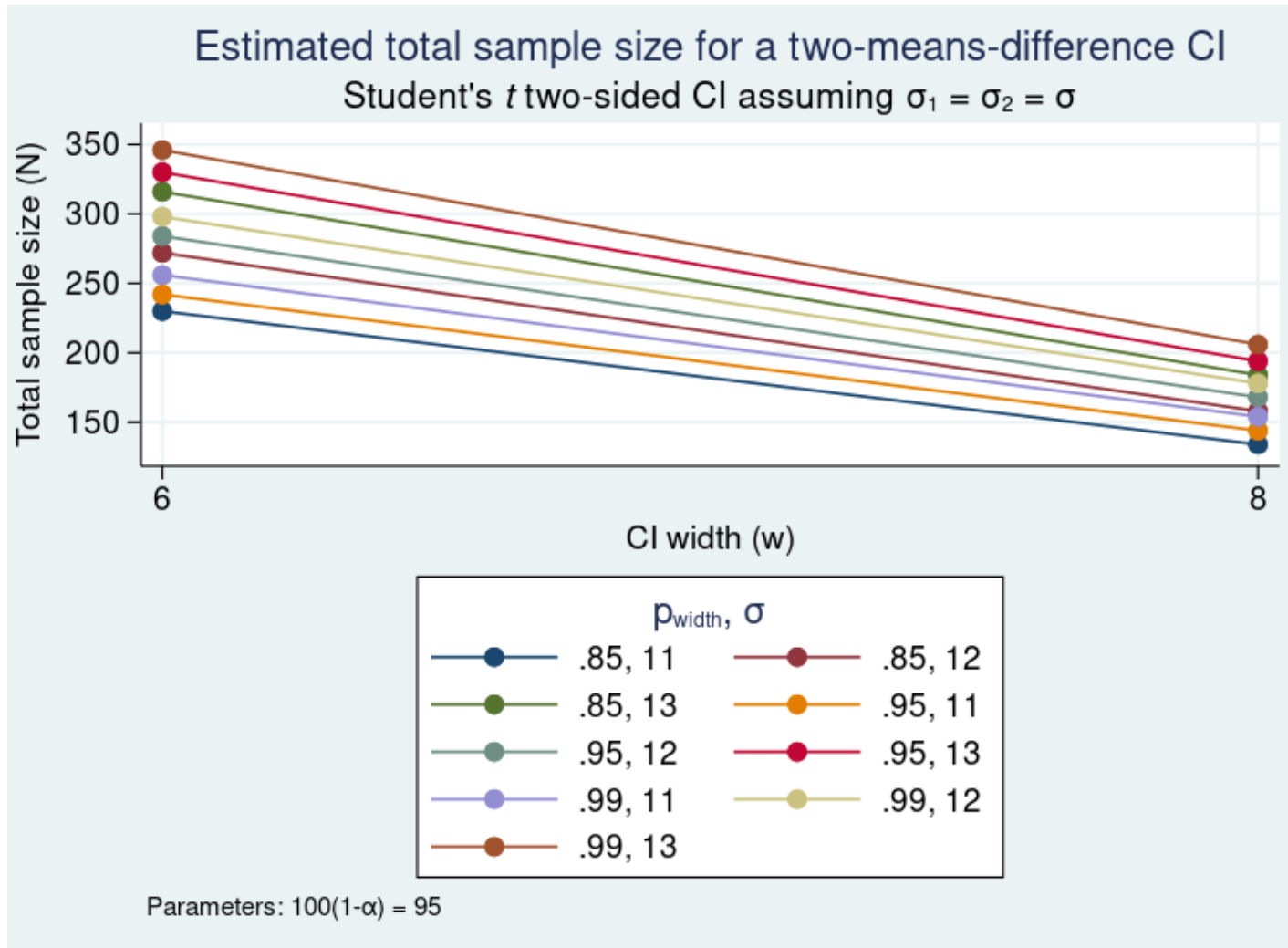
Label distinct values on the y axis

```
. ciwidth twomeans, width(6 8) probwidth(0.85 0.95 0.99) sd(12)  
> graph(xdimension(Pr_width) yvalues)
```



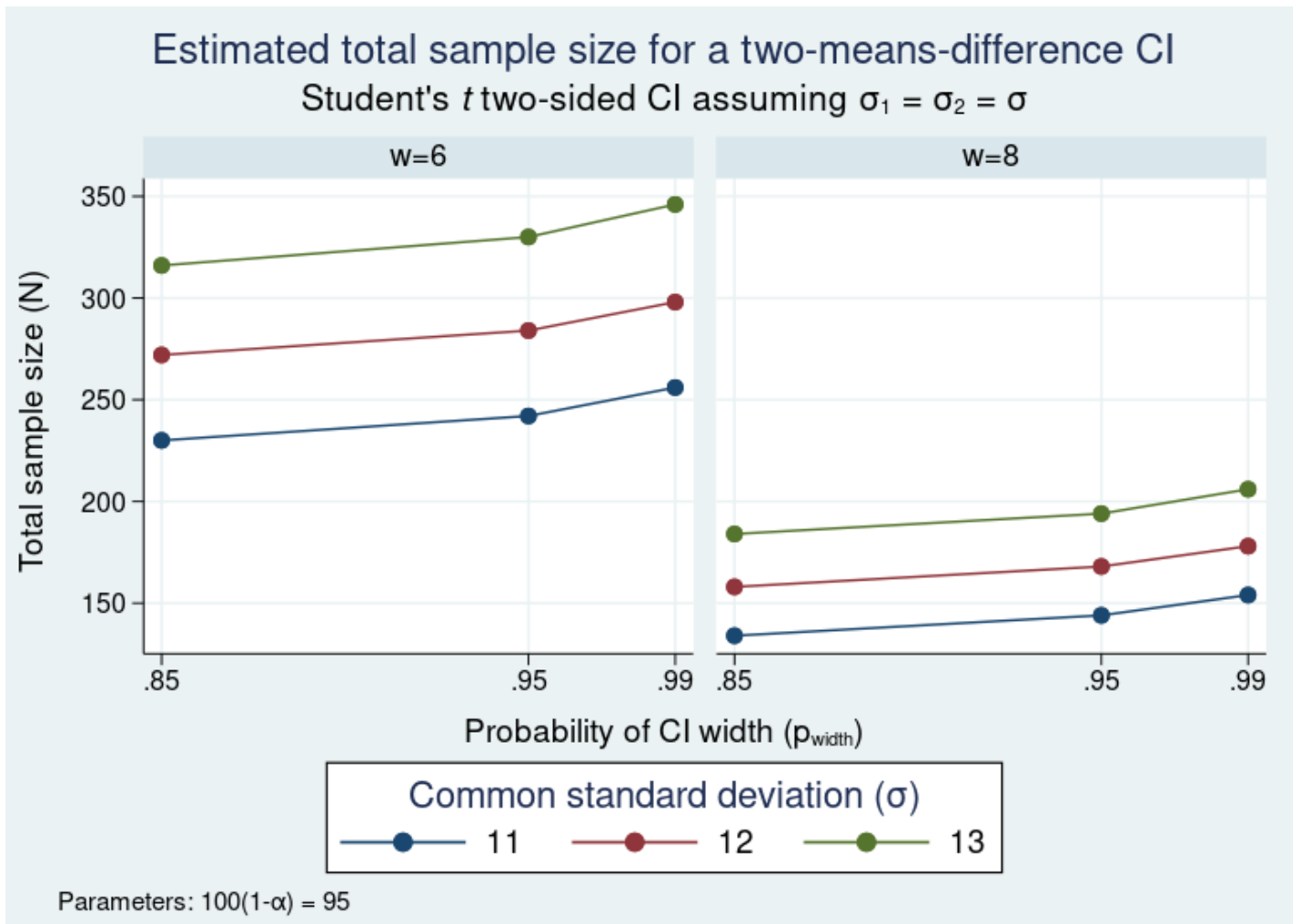
Specifying multiple values for multiple parameters

```
. ciwidth twomeans, width(6 8) probwidth(0.85 0.95 0.99)  
> sd(11 12 13) graph
```



Create subgraphs for values of width

```
. ciwidth twomeans, width(6 8) probwidth(0.85 0.95 0.99)  
> sd(11 12 13) graph(bydimension(width) legend(rows(1)))
```

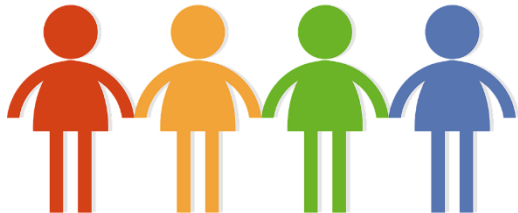


Where to learn more

- See [\[PSS-3\] ciwidth twomeans](#) to learn more about precision and sample-size analysis for a confidence interval for a difference between two means from independent samples.
- See [\[PSS-3\] ciwidth, graph](#) to learn more about graphing results from the `ciwidth` command.

Unbalanced sample sizes

Experimental group
(Medication)



Control group
(No medication)



Specifying the sample-size ratio

```
. ciwidth twomeans, width(6) sd(12) probwidth(0.99) nratio(0.5)
```

Performing iteration ...

Estimated sample sizes for a two-means-difference CI
Student's t two-sided CI assuming $sd1 = sd2 = sd$

Study parameters:

level =	95.00
Pr_width =	0.9900
width =	6.0000
sd =	12.0000
N2/N1 =	0.5000

Estimated sample sizes:

N =	333
N1 =	222
N2 =	111

Compute one sample size given another

```
. ciwidth twomeans, width(6) sd(12) probwidth(0.99) n2(100) compute(n1)
```

Performing iteration ...

Estimated sample sizes for a two-means-difference CI
Student's t two-sided CI assuming $sd1 = sd2 = sd$

Study parameters:

```
level =      95.00  
Pr_width =    0.9900  
width =      6.0000  
sd =      12.0000  
N2 =        100
```

Estimated sample sizes:

```
N =        370  
N1 =       270
```

Known standard deviations

```
. ciwidth twomeans, width(6) sd(12) n2(100) compute(n1) knownsds
```

Estimated sample sizes for a two-means-difference CI
Normal two-sided CI assuming $sd1 = sd2 = sd$

Study parameters:

```
level =    95.00  
width =    6.0000  
sd =    12.0000  
N2 =     100
```

Estimated sample sizes:

```
N =     260  
N1 =    160
```

More on unbalanced designs

- The **power** and **ciwidth** suites support unbalanced designs for two-sample studies.
 - Compute the total sample size given the allocation ratio
 - Compute one group size given the other
 - Specify the total sample size and the allocation ratio
 - Specify one group size and the allocation ratio
 - ...
- The syntax is consistent across **power** and **ciwidth** commands.
- See [\[PSS-4\] unbalanced designs](#) to learn more.

Do you have another method in mind?

Adding your own method to **ciwidth**

Similar to **power**, you can add your own method to **ciwidth** by following the same steps.

Simply write an r-class program to perform the computations, making sure to follow **ciwidth**'s convention for naming common options and storing results, and place the program where Stata can find it.

You can then create tables and graphs as you would with any official **ciwidth** method.

To learn more, see [\[PSS-3\] *ciwidth usermethod*](#).

Summary

- We performed power and sample-size analysis for
 - A population mean
 - A population mean, when using a sample from a cluster randomized design
 - An odds ratio for matched case-control data
 - A log hazard-ratio from a Cox proportional hazards model
- We performed precision and sample-size analysis for
 - A population mean
 - A difference between two independent means
 - A difference between paired means
- We were able to perform sensitivity analysis graphically and with a table.
- We implemented our own method, and easily created graphs and tables as if we were using an official **power** command.
- See the Stata Power, Precision, and Sample-Size Reference Manual ([\[PSS\]](#)) for more examples.
- See stata.com/features/power-precision-and-sample-size for a full list of features and overview examples.

References

- Dixon, W. J., and F. J. Massey, Jr. 1983. *Introduction to Statistical Analysis*. 4th ed. New York: McGraw–Hill.
- Chow, Shein-Chung, J. Shao, H. Wang, and Y. Lokhnygina. 2017. *Sample Size Calculations in Clinical Research*. 3rd ed. Boca Raton: Taylor & Francis.
- Meeker, W. Q., G. J. Hahn, and L. A. Escobar. 2017. *Statistical Intervals: A Guide for Practitioners and Researchers*. 2nd ed. Hoboken, NJ: Wiley.
- Tamhane, A. C., and D. D. Dunlop. 2000. *Statistics and Data Analysis: From Elementary to Intermediate*. Upper Saddle River, NJ: Prentice Hall.

Thank you !!