Sample Size Calculations and Power Analysis using Stata

Wayne Enanoria, M.P.H., Ph.D.

Biostat 212:
Statistical Computing in Clinical Research

September 8, 2015

Sample Size Calculations and Power Analysis

Sample Size Calculations: Sample size calculations tell you the number of sampling units you need to detect a given effect size for a given level of statistical significance (α) and fixed power.

Power Analysis: Power analysis tells you the statistical power you have to detect a given effect size for a fixed sample size and a given level of statistical significance (α).

Example:

A clinical dietician wants to compare two different diets, A and B, for diabetic patients. She hypothesizes that diet A (Group 1) will be better than diet B (Group 2), in terms of lower blood glucose. She plans to get a random sample of diabetic patients and randomly assign them to one of the two diets. At the end of the experiment, which lasts 6 weeks, a fasting blood glucose test will be conducted on each patient.

- She also expects that the average difference in blood glucose measure between the two groups will be about 25 mg/dl.
- Furthermore, she also assumes the standard deviation of blood glucose distribution for diet A to be 15 and the standard deviation for diet B to be 17.

The dietician wants to know the number of subjects needed in each group assuming equal sized groups.

Sample Size Calculation for a Difference of Two Means

What do we need to calculate the sample size in each group (n)?

- Mean glucose in group 1 (μ_1)
- Standard deviation of glucose in group 1 (σ_1)
- Mean glucose in group 2 (μ_2)
- Standard deviation of glucose in group 2 (σ_2)
- Level of statistical significance (α)
- Power (1 - β)

Sample Size per Group Needed for a Two-Sample t-test:

$$n = \frac{(\sigma_1^2 + \sigma_2^2)(z_{1-\frac{\alpha}{2}} + z_{1-\beta})^2}{\Delta^2}$$

where $\Delta = \mu_2 - \mu_1$.

"In words, n is the appropriate sample size in each group to have a probability of $1 - \beta$ of finding a significant difference based on a two-sided level α significance test, if the true difference in means between the two groups is $\mu_2 - \mu_1$" (Rosner 1995).

Stata Command to Compute Sample Size

Syntax

power twomeans *m1 m2* [, power(*numlist*) sd1(*sigma1*) sd2(*sigma2*) *options*]

where *m1* is the control-group mean, *m2* is the experimental-group mean, *sigma1* is the standard deviation in the control-group, and *sigma2* is the standard deviation in the experimental group.

Sample Calculations for Two-Sample Means Test

Syntax

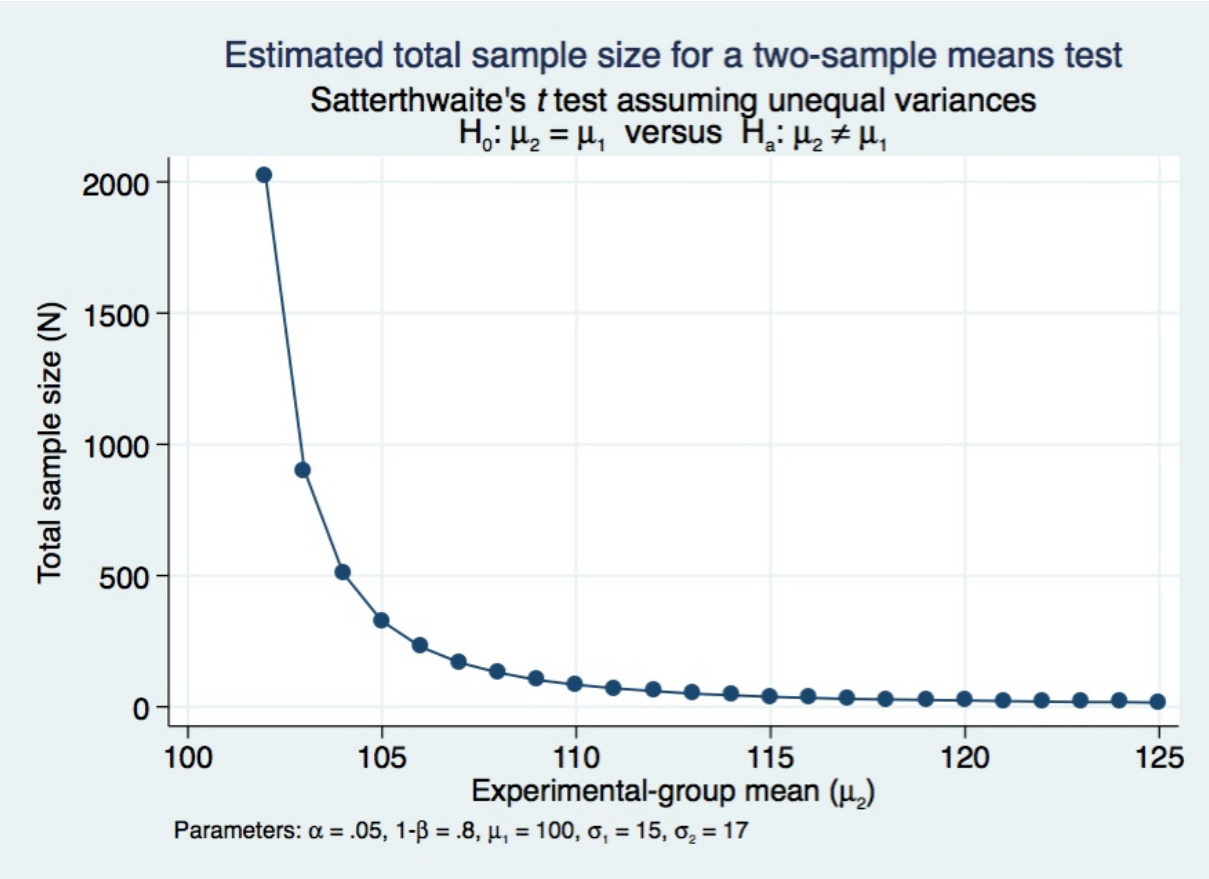
power twomeans 100 125, power(0.80) sd1(15) sd2(17)

```
. power twomeans 100 125, power(0.80) sd1(15) sd2(17)

Performing iteration ...

Estimated sample sizes for a two-sample means test
Satterthwaite's t test assuming unequal variances
Ho: m2 = m1  versus  Ha: m2 != m1

Study parameters:

        alpha =    0.0500
        power =    0.8000
        delta =   25.0000
           m1 =  100.0000
           m2 =  125.0000
          sd1 =   15.0000
          sd2 =   17.0000

Estimated sample sizes:

            N =        16
  N per group =         8
```

Sample Calculations for Two-Sample Means Test

Syntax

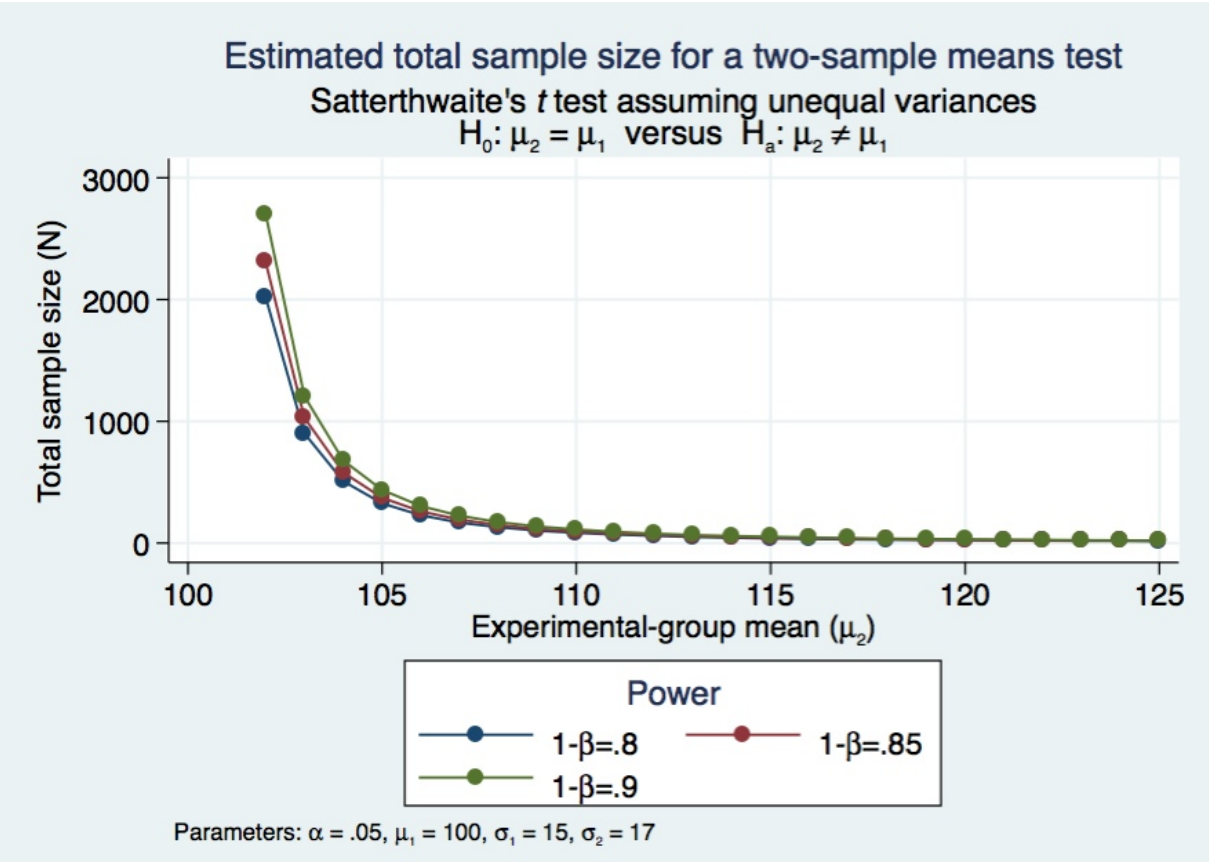
power twomeans 100 (102(1)125), power(0.80) sd1(15) sd2(17) graph



Sample Calculations for Two-Sample Means Test

Syntax

power twomeans 100 (102(1)125), power(0.80(0.05)0.9) sd1(15) sd2(17) graph graph(nosimplelabels legend(title("Power")))



Power Analysis for Two-Sample Means Test

Syntax

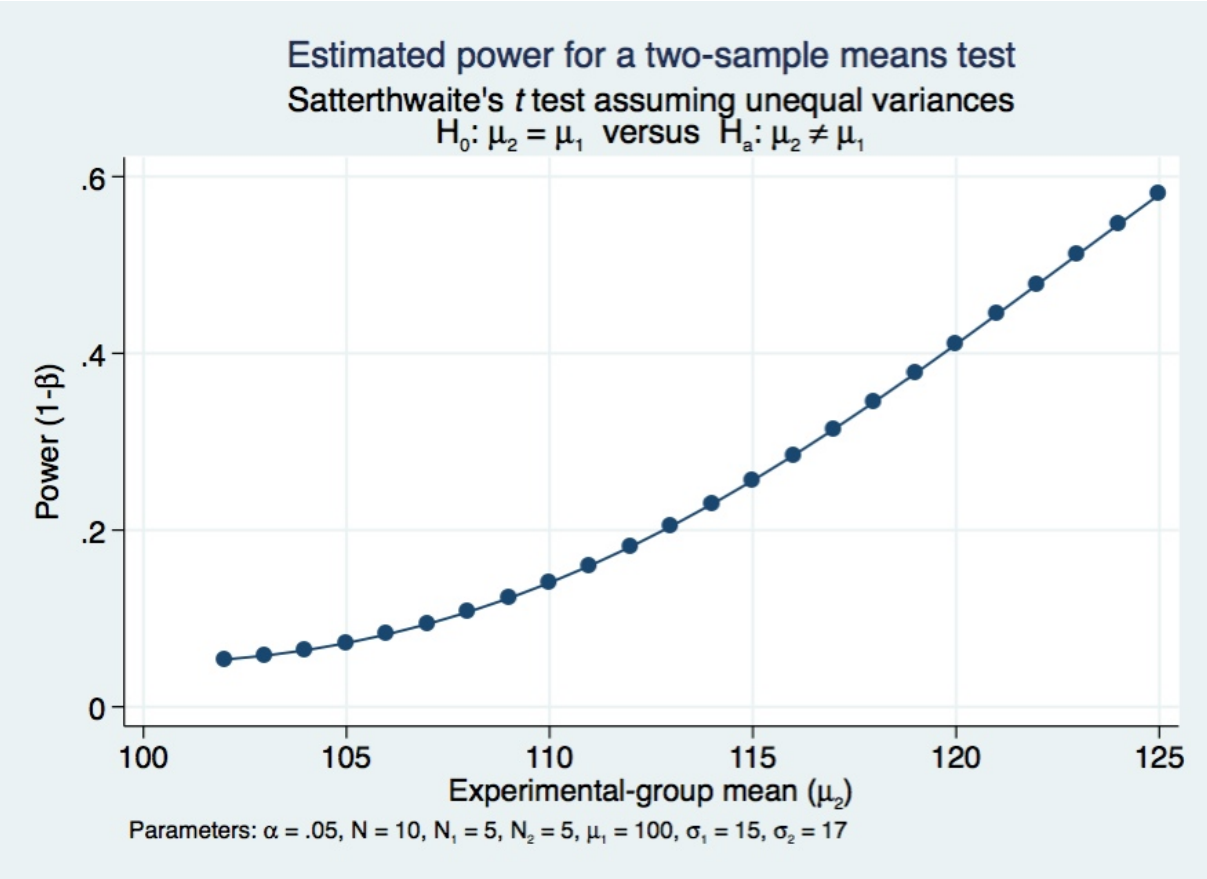
power twomeans 100 125, sd1(15) sd2(17) n(10)

```
. power twomeans 100 125, sd1(15) sd2(17) n(10)

Estimated power for a two-sample means test
Satterthwaite's t test assuming unequal variances
Ho: m2 = m1  versus  Ha: m2 != m1

Study parameters:

        alpha =    0.0500
            N =        10
  N per group =         5
        delta =   25.0000
           m1 =  100.0000
           m2 =  125.0000
          sd1 =   15.0000
          sd2 =   17.0000

Estimated power:

        power =    0.5797
```

Power Analysis for Two-Sample Means Test

Syntax

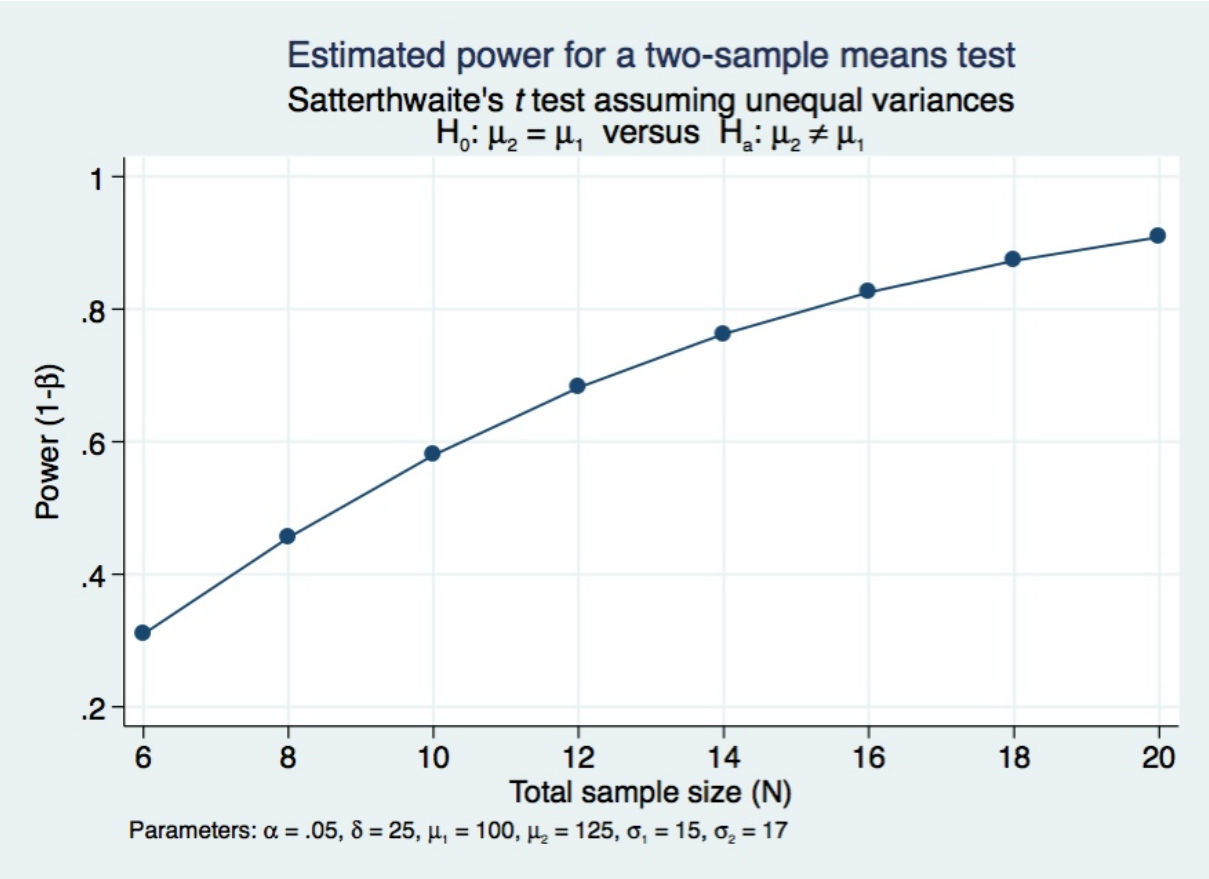
power twomeans 100 (102(1)125), sd1(15) sd2(17) n(10) graph



Power Analysis for Two-Sample Means Test

Syntax

power twomeans 100 125, sd1(15) sd2(17) n(6(2)20) graph



Other Commands

- power onemean
- power pairedmeans
- power oneproportion
- power twoproportions
- power twoway
- power mcc
- power trend
- power cox
- power logrank

For More Information:

- Stata Power and Reference Manual
  `http://www.stata.com/manuals14/pss.pdf`
- Stata Power and Sample Size Manual
  `http://www.stata-press.com/manuals/power-sample-size-reference-manual/`