Refresher on Power and Sample Size

Dawn Peters, Ph.D.

October 7, 2009

Outline

- Example
- Comparison of Means
- Definition of Power and Power Properties
- Power calculations using web-based tool
- Comparison of Proportions
- Other considerations for planning an experiment and analyzing data
- Confounding

Research Aim

- Compare two weight reduction diets (A, B)
  - Which is better?

Better?

- Compare weight reductions on both diets
- Compare sustained weight loss for both diets
- Compare times to achieve certain weight loss
- Compare compliance
- Compare reduction in LDL
- Compare success proportions

Which one results in the greatest weight reduction?

Hypothesize:

New weight reduction diet results in greater weight reduction after 1 year on the average than standard weight reduction diet

Experiment

Take 20 people—randomly assign 10 to standard program, 10 to new program

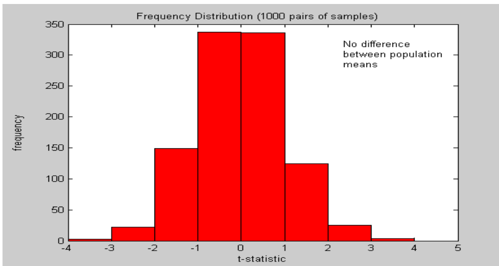
Compare mean weights after 1 year for each program?

Compare mean reductions after 1 year for each program? (Other options as well.)

T-statistic (assumptions required)

$$t = \frac{\bar{Y}_1 - \bar{Y}_2}{SE}$$

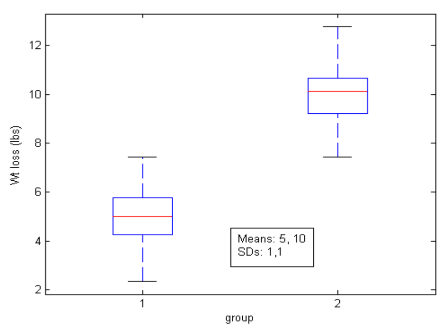
Sample sizes of 10 for each group
No difference between the population means

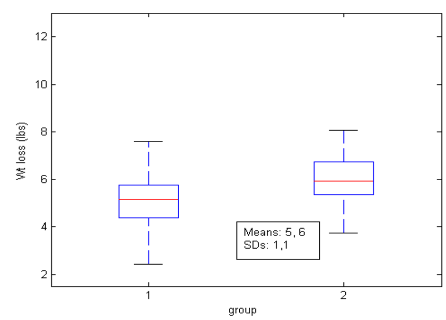


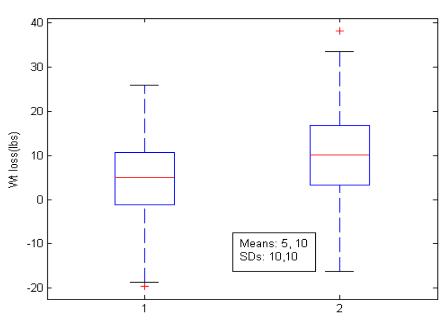
- Conclude there is a difference in mean weight reductions if the t-statistic is large in magnitude which means the p-value is small.
- For power considerations, pick a significance level, e.g .05, then p-value < .05 if

$$t > 2.101 \text{ or } t < -2.101$$

- We set up the test so that the proportion of times we incorrectly conclude there is a difference between the population means is small (Type I error rate: $\alpha = .05$ here)







- Power for testing a hypothesis about the difference in means: chance of **concluding** there is a difference between the means when there really is a difference.
- Want power to be high when there is a meaningful difference between the means

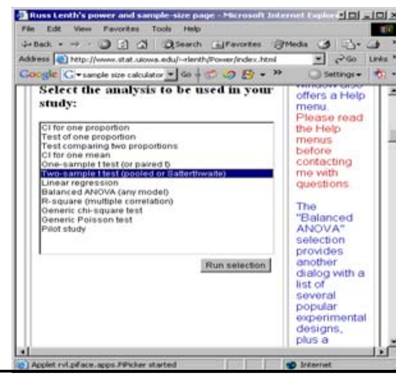
- Power for comparing means depends on
  - True difference in the means
  - SD’s within each group
  - Sample size
  - Significance level
  - One vs Two-sided Test

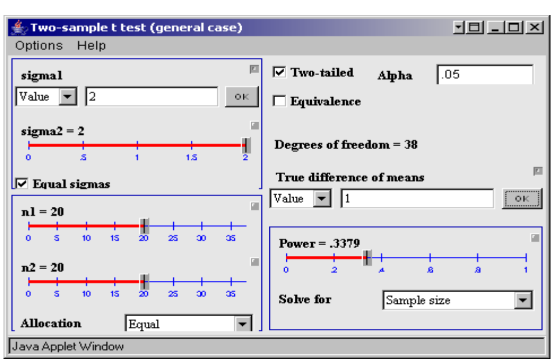
- Examples:

• True means: 5, 7	True means: 5, 6
• SD’s: 2,2	SD’s: 2, 2
• Sample sizes: 20,20	Sample sizes: 20, 20
• Two-sided test	Two-sided test
• $\alpha = .05$	$\alpha = .05$
Power = 87%	Power=34%

http://www.stat.uiowa.edu/~rlenth/Power/index.html

Two-Sample t-test

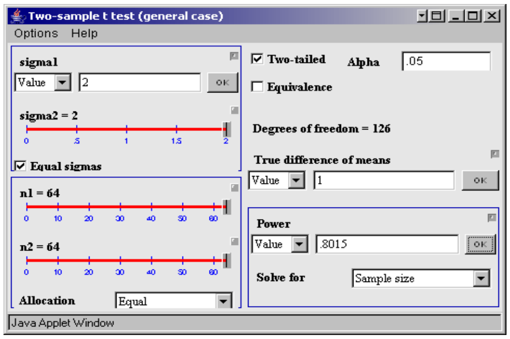




- When planning a study to compare means via a two-sample t-test, typically specify
  - Power
  - Significance level
  - Mean difference want to detect
  - SD’s
  - One or two-sided test

  Then

  Find Needed Sample Sizes



For comparing means:

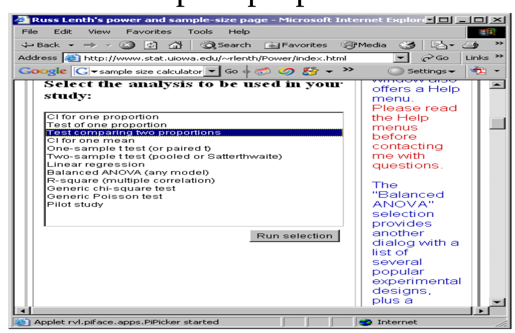
- Power increases as difference between means increases
- Power increases as SD(s) decreases
- Power increases as sample size increases
- Power increases as significance level α increases

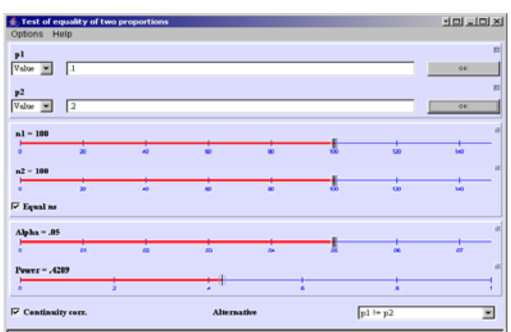
How do you pick some of these quantities?

- Difference you want to be able to detect
  - Want to detect difference in mean weight loss of 100 lbs between Atkins and Low-Fat diet after 1 year?
  - Difference of .25 lbs?

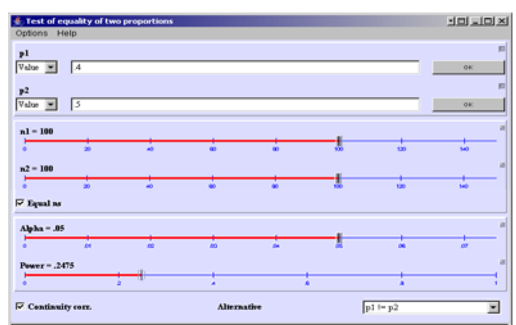
  What is the smallest difference that is meaningful?
- SD's
  - Pilot data, other studies

Compare proportions





Power =.4289



Power=.2475

Note some differences in different programs ar based on options—such as continuity corr, her

- Note that power depends on actual proportions not just the difference—in the example, power greater for comparing .1 to .2 (power = .43) than for comparing .4 to .5 (power = .25).

Additional Issues in Sample Size Selection

- Adjustment for other factors/design issues in determining sample sizes—SD's after adjusting for confounders, general regression analyses, repeated measures analyses
- Multiple comparison issues
- Interim analyses

Planning the experiment

- Determine measurements of interest, hypotheses and methods for analysis
- Determine sample sizes needed so that power is as desired
  - Clinical vs statistical significance
- Different objectives/analyses require different power calculations
- Other approaches besides power calculations

Alternative approach to sample size selection

- Precision of estimators

  Want estimate of population mean, i.e. the sample mean, to be within

  d units of population mean with high probability, say .95

  equivalently

  Want 95% confidence interval to have width 2d.

  What sample size is needed?

More about Covariates

- Men and Women Combined? Ignore gender in design and analysis?
- Age?
- Initial Weight?
- Exercise level?

- What if everyone in new program exercised a lot and everyone in standard program exercised very little?
- Difference due to diet? Exercise?
- Great variability due to exercise?

- Important to consider potential confounding variables (such as exercise level)
  - At design stage
  - In analysis

- Study from the United Kingdom recruiting subjects in 1972-1974 and obtaining smoking history (and other information) and following them up 20 years later.
- Look at data for 609 women aged 45-74 at start of study.

	Smoker	
	Yes	No
Dead	107	153
Alive	174	175
Total	281	328
%Dead	38.1%	46.6%

OR= $\frac{\left(\frac{107/281}{174/281}\right)}{\left(\frac{153/329}{175/329}\right)} = .703$ p-value=.033, Risk ratio=.82

	Age: 45-54		Age: 55-64		Age: 65-74	
	Smoker		Smoker		Smoker	
	Yes	No	Yes	No	Yes	No
Dead	27	12	51	40	29	101
Alive	103	66	64	81	7	28
%Dead	20.8%	15.4%	44.3%	33.1%	80.6%	78.3%

Example of Interaction

- Sample mean reduction for women in new program = 15 lbs (n=10)
- Sample mean reduction for men in new program =5 lbs (n=10)
- Combining men and women (equal sample sizes) mean reduction for new program (ignoring gender) is 10 pounds
- Sample mean reduction for standard program is 10 pounds for both genders
- Ignoring gender, will not see a difference between programs.

Equivalence Study

Aim to show that two treatments are equivalent.

Need to specify what is meant by equivalence

For example for comparing two diets you might say the two diets are equivalent if the mean difference in weight loss after 1 year is less than 3 lbs.

Equivalence, continued

To obtain sample sizes need:

Criteria for equivalence
Hypothesized means
Standard deviations
Significance level
Power