Power Calculations

March 3, 2016

Joseph Massaro
Department of Biostatistics
jmm@bu.edu

Introduction

- A statistician's role in a study begins well before data are ready for analysis
- Statistician assists in the "design" of the study prior to study start
  - Major design component: Determining appropriate sample size for the study
    - i.e., determine sample size large enough to represent population with reasonable confidence;
    - while not having a sample size be cumbersome from cost and/or time perspective.

Superiority

- Suppose have two treatment groups that we wish to compare on a dichotomous outcome in a clinical trial:

  H_o: $p_E \geq p_C$ ("Null" hypothesis)
  H_1: $p_E < p_C$ ("Alternative" hypothesis – what we hope to prove)

  where p_E and p_C are the TRUE (but unknown) event rates of a negative outcome (e.g., Major Adverse Event) for experimental and control groups IN THE POPULATION.

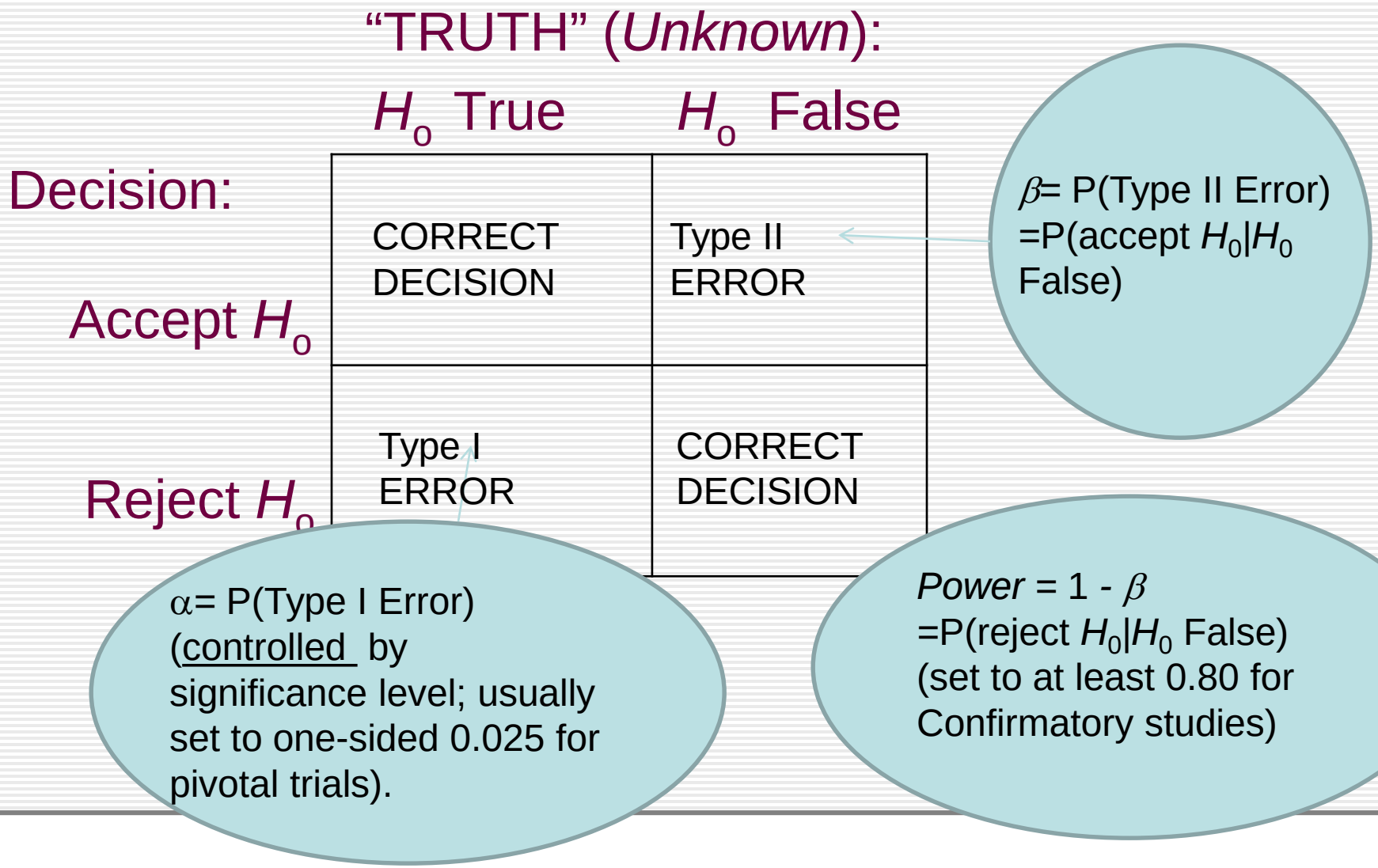
Introduction

- We hope the null hypothesis is false in the population
- We base our decision about whether the null hypothesis is false (whether to reject H_0) on a *sample* from the population.
- We want a sample large enough to represent the population with reasonable confidence (there's always a chance we'll obtain a sample not representative of the population, but we want to minimize this chance).

Introduction

Decision:	"TRUTH" (*Unknown*): H_o True	H_o False
Accept H_o	CORRECT DECISION	Type II ERROR
Reject H_o	Type I ERROR	CORRECT DECISION

Introduction



β and Power

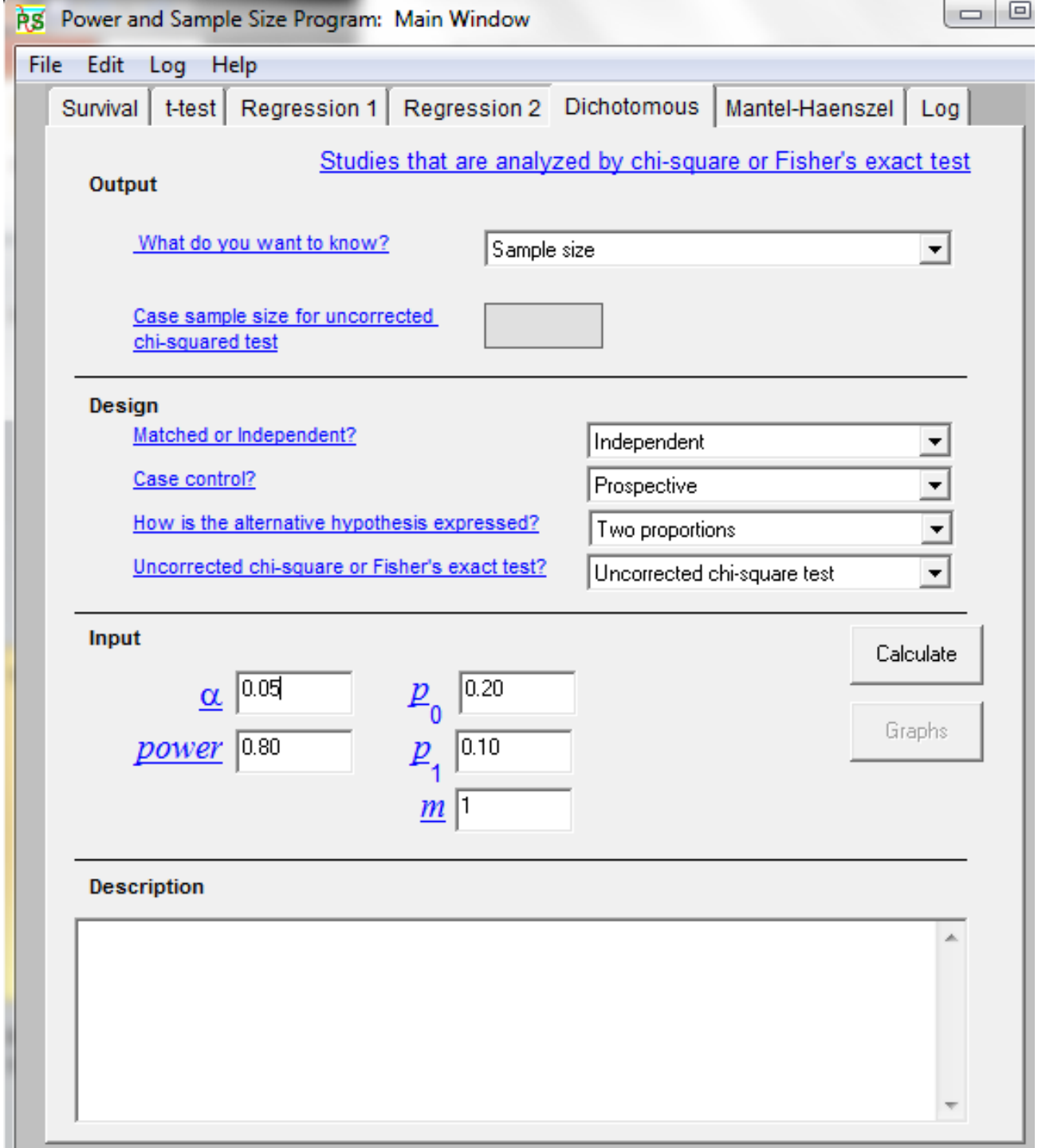
- As α decreases, β increases (and power decreases)
- As sample size (***n)*** per group increases, power increases (if the null hypothesis is truly false)
  - For a confirmatory study: choose desired power; determine *n* per group
- Power also depends on the true treatment effect

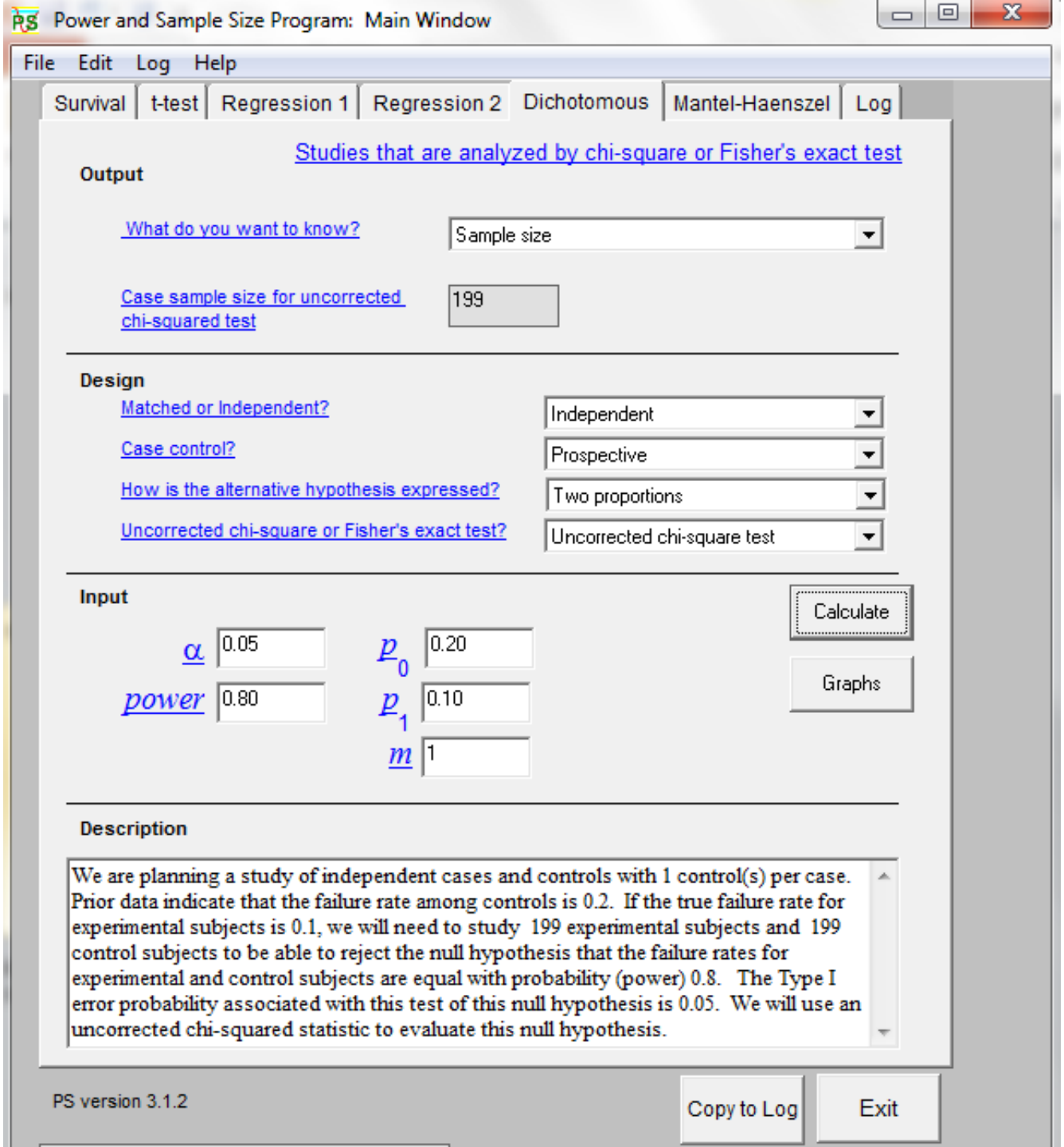
Example

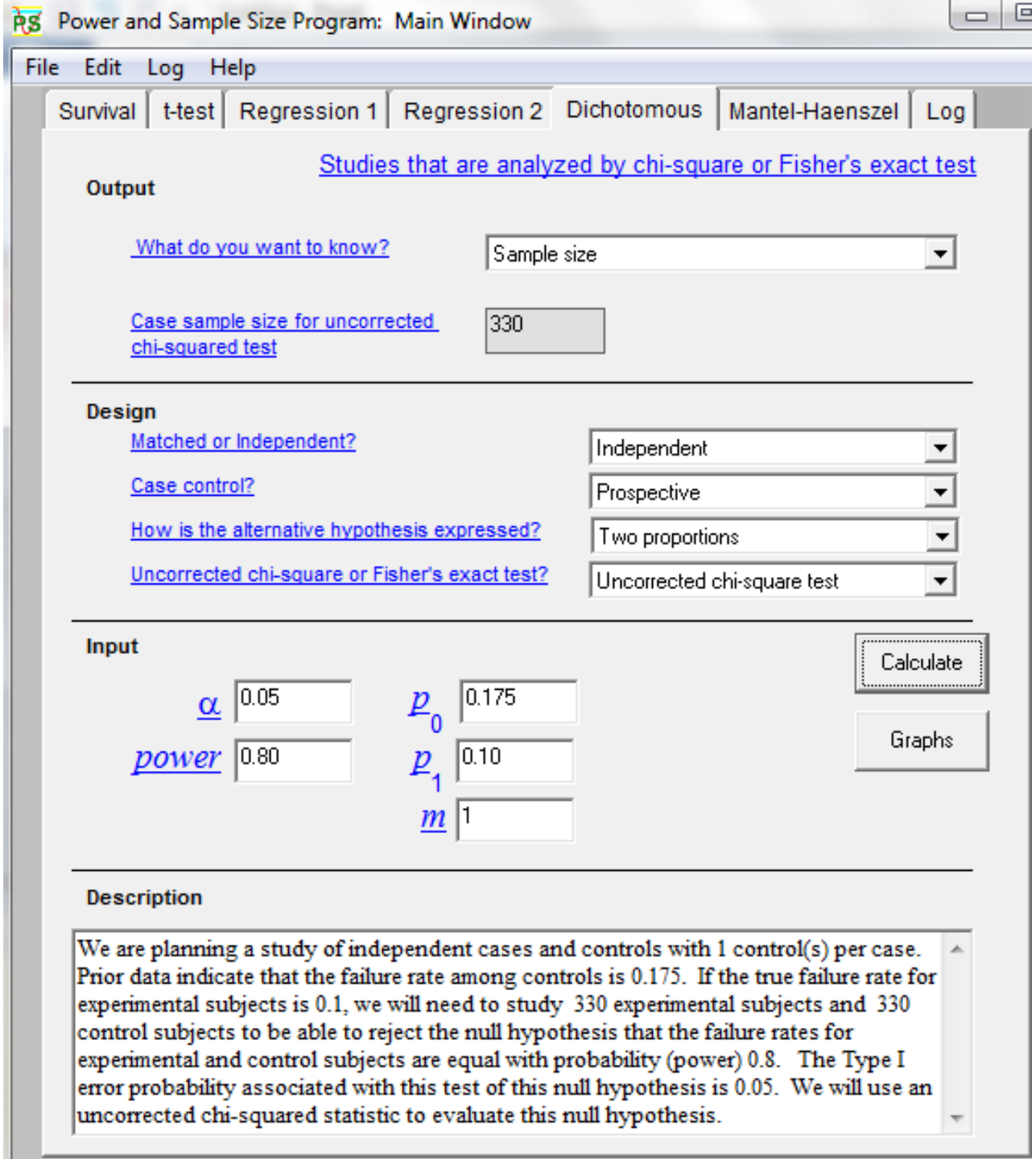
H_o: $p_E \geq p_C$ ("Null" hypothesis)
H_1: $p_E < p_C$ ("Alternative" hypothesis – what we hope to prove)

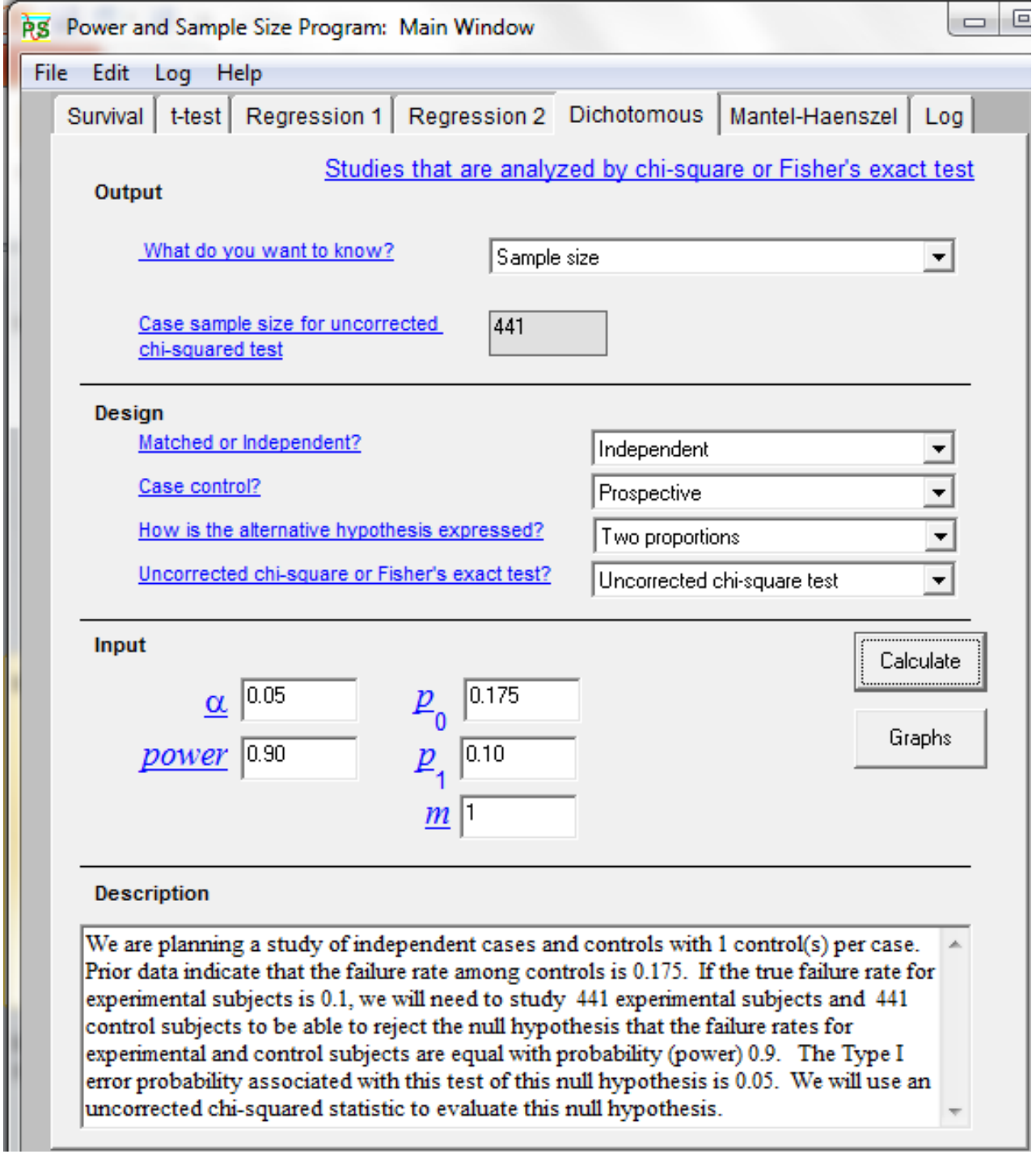
- Free sample size calculation software (for basic but commonly required calculations) at:

http://biostat.mc.vanderbilt.edu/wiki/Main/PowerSampleSize#Downloading_and_Installing_the_PS_Software


Power and Sample Size Program: Main Window
File Edit Log Help
Survival t-test Regression 1 Regression 2 Dichotomous Mantel-Haenszel Log
Studies that are analyzed by chi-square or Fisher's exact test
Output
What do you want to know?
Sample size
Case sample size for uncorrected chi-squared test
Design
Matched or Independent?
Independent
Case control?
Prospective
How is the alternative hypothesis expressed?
Two proportions
Uncorrected chi-square or Fisher's exact test?
Uncorrected chi-square test
Input
α 0.05
power 0.80
p0 0.20
p1 0.10
m 1
Calculate
Graphs
Description


Power and Sample Size Program: Main Window
File Edit Log Help
Survival t-test Regression 1 Regression 2 Dichotomous Mantel-Haenszel Log
Studies that are analyzed by chi-square or Fisher's exact test
Output
What do you want to know?
Sample size
Case sample size for uncorrected chi-squared test
199
Design
Matched or Independent?
Independent
Case control?
Prospective
How is the alternative hypothesis expressed?
Two proportions
Uncorrected chi-square or Fisher's exact test?
Uncorrected chi-square test
Input
Calculate
α 0.05
p0 0.20
power 0.80
p1 0.10
m 1
Graphs
Description
We are planning a study of independent cases and controls with 1 control(s) per case. Prior data indicate that the failure rate among controls is 0.2. If the true failure rate for experimental subjects is 0.1, we will need to study 199 experimental subjects and 199 control subjects to be able to reject the null hypothesis that the failure rates for experimental and control subjects are equal with probability (power) 0.8. The Type I error probability associated with this test of this null hypothesis is 0.05. We will use an uncorrected chi-squared statistic to evaluate this null hypothesis.
PS version 3.1.2
Copy to Log
Exit


Power and Sample Size Program: Main Window
File Edit Log Help
Survival t-test Regression 1 Regression 2 Dichotomous Mantel-Haenszel Log
Studies that are analyzed by chi-square or Fisher's exact test
Output
What do you want to know?
Sample size
Case sample size for uncorrected chi-squared test
330
Design
Matched or Independent?
Independent
Case control?
Prospective
How is the alternative hypothesis expressed?
Two proportions
Uncorrected chi-square or Fisher's exact test?
Uncorrected chi-square test
Input
Calculate
α 0.05
p0 0.175
power 0.80
p1 0.10
m 1
Graphs
Description
We are planning a study of independent cases and controls with 1 control(s) per case. Prior data indicate that the failure rate among controls is 0.175. If the true failure rate for experimental subjects is 0.1, we will need to study 330 experimental subjects and 330 control subjects to be able to reject the null hypothesis that the failure rates for experimental and control subjects are equal with probability (power) 0.8. The Type I error probability associated with this test of this null hypothesis is 0.05. We will use an uncorrected chi-squared statistic to evaluate this null hypothesis.


Power and Sample Size Program: Main Window
File Edit Log Help
Survival t-test Regression 1 Regression 2 Dichotomous Mantel-Haenszel Log
Studies that are analyzed by chi-square or Fisher's exact test
Output
What do you want to know?
Sample size
Case sample size for uncorrected chi-squared test
441
Design
Matched or Independent?
Independent
Case control?
Prospective
How is the alternative hypothesis expressed?
Two proportions
Uncorrected chi-square or Fisher's exact test?
Uncorrected chi-square test
Input
Calculate
α 0.05
p0 0.175
power 0.90
p1 0.10
m 1
Graphs
Description
We are planning a study of independent cases and controls with 1 control(s) per case. Prior data indicate that the failure rate among controls is 0.175. If the true failure rate for experimental subjects is 0.1, we will need to study 441 experimental subjects and 441 control subjects to be able to reject the null hypothesis that the failure rates for experimental and control subjects are equal with probability (power) 0.9. The Type I error probability associated with this test of this null hypothesis is 0.05. We will use an uncorrected chi-squared statistic to evaluate this null hypothesis.

Sample Size

- How do we come up with our assumptions of the "true" mean effect (i.e., true mean difference)?
  - Past Studies
  - Literature review (maybe someone else did similar work and published it)
  - Best guess?
  - What we "hope" is the effect (even though we may not have scientific evidence yet)
- Assumption should be "clinically" meaningful

Underpowering

- We do not want to "underpower" our study
  - I.e., we don't want a sample size that is too small.

- This can happen if, when powering the study, we assume the effect size (e.g., treatment difference) is larger than it truly is

Underpowering

- E.g., if we powered previous example assuming an experimental treatment 10% MAE rate and a control treatment 20% MAE rate, we need 199 per treatment group

- If, in reality (and unbeknownst to us) the true MAE rates are 10% and 17.5%, then in reality 199 per group yields only 58% power.
  - But at least we still have some power

Evaluable vs. All Randomized

- The above sample size calculations yield number of *evaluable* subjects
  - i.e., subjects who will not prematurely withdraw and hence be available for follow-up
  - Add a certain % of patients to number of evaluable to obtain number of patients who should be enrolled
  - e.g., if 10% dropout rate is anticipated, enroll 199/0.90 = 222 patients/group.

Other Sample Size Software (not free)

- SAS (PROC POWER)
- PASS ("Power and Sample Size")
  - http://www.ncss.com/software/pass/
- nQuery
  - http://www.statsols.com/products/nquery-advisor-nterim/
- Power and Precision
  - http://www.power-analysis.com/home.htm

PASS Home

Home View Procedures Tools Window Help

Options Web

Select a Procedure

Category

Favorites Recent Show All

- Cluster-Randomized
- Conditional Power
- Confidence Intervals
- Correlation
- Design of Experiments
- Equivalence
- Group-Sequential
- Means
- Microarray
- Non-Inferiority
- Nonparametric
- Normality
- Proportions
- Quality Control
- Regression
- ROC
- Simulation
- Superiority by a Margin
- Survival
- Variances
- Tools

Recent Procedures (8)

Two-Sample T-Tests Assuming Equal Variance (Enter Means)

Mann-Whitney-Wilcoxon Tests (Simulation)

Tests for Proportion

Non-Inferiority Tests for Two Proportions using Differences

Two-Sample T-Tests Allowing Unequal Variance (Enter Means)

Tests for

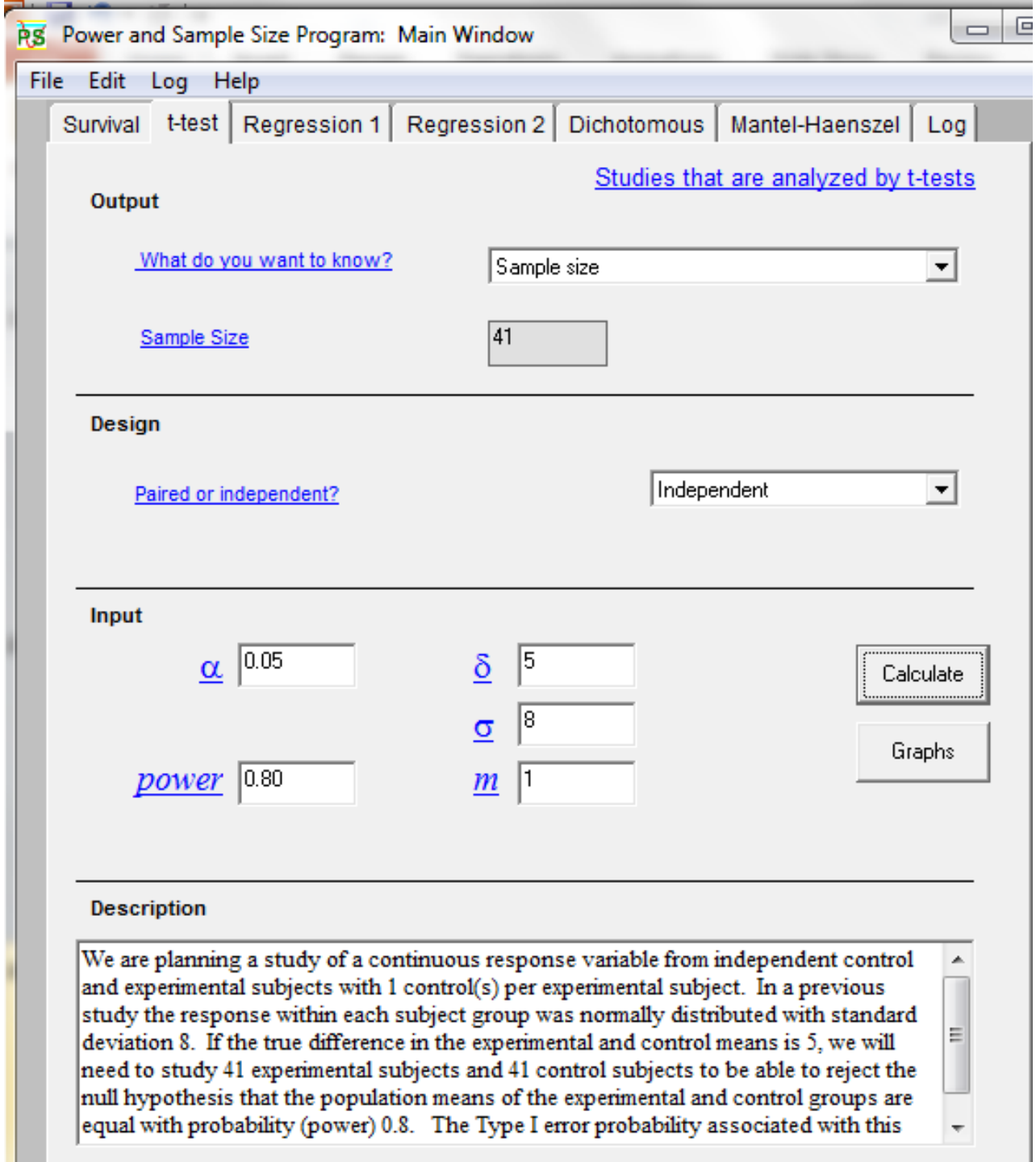
Continuous Outcome

- A study is conducted to assess the effect of a new stem cell treatment on LVEF improvement at 4 months following acute MI
- Control group is a placebo (sham control procedure)
- $H_0: \mu_T \leq \mu_{PL}$ vs. $H_1: \mu_T > \mu_{PL}$

  where μ_T and μ_{PL} are mean improvement in LVEF % from baseline to 4 months for experimental treatment and placebo

Continuous Outcome

- Assumptions made in powering study:
  - True $\mu_T - \mu_{PL} = 5\%$
  - Standard deviation of change in LVEF = 8% in each group
  - Desire power of 80%
  - One-sided significance level of 0.025
- Note: We don't need assumption of the true means; just the difference in means
- Note: We need an assumption of the variability (the more variability, the larger the sample size needed)


Power and Sample Size Program: Main Window
File Edit Log Help
Survival t-test Regression 1 Regression 2 Dichotomous Mantel-Haenszel Log
Studies that are analyzed by t-tests
Output
What do you want to know?
Sample size
Sample Size
41
Design
Paired or independent?
Independent
Input
α 0.05
δ 5
σ 8
power 0.80
m 1
Calculate
Graphs
Description
We are planning a study of a continuous response variable from independent control and experimental subjects with 1 control(s) per experimental subject. In a previous study the response within each subject group was normally distributed with standard deviation 8. If the true difference in the experimental and control means is 5, we will need to study 41 experimental subjects and 41 control subjects to be able to reject the null hypothesis that the population means of the experimental and control groups are equal with probability (power) 0.8. The Type I error probability associated with this

Assumptions Needed for Common Statistical Tests

Objective	Methodology	Assumptions Needed for Power
Show superiority of one treatment over another on binary outcome rate	Chi-square test (or Fisher's Exact test)	Event rate in each treatment
Show superiority of one treatment over another on mean of outcome	Two-sample t-test	Difference between means; sd per treatment
Show that increase in a variable X increases/decreases the odds of the outcome event	Univariate Logistic Regression	Odds Ratio, binary outcome rate when X=0 (if X is binary) or at mean of X (if X is continuous)
Show that increase in a variable X increases/decreases the odds of the outcome event, after adjusting or other coariates	Adjusted Logistic Regression	Odds Ratio, binary outcome rate when X=0 (if X is binary) or at mean of X (if X is continuous), R-squared of X with other covariates
Show that increase in a variable X increases/decreases the average outcome of Y	Simple Linear Regression	Slope (average change in Y per 1 unit increase of X), sd of X, sd of Y or correlation of X and Y
Show that increase in a variable X increases/decreases the average outcome of Y, after adjusting for other covariates	Multiple Linear Regression	Slope (average change in Y per 1 unit increase of X), sd of X, sd of Y or correlation of X and Y, R-squared of X with other covariates

Multiple Primary Analyses

- Suppose there are two primary hypotheses; e.g.:

  $H_0: p_{E,SAF} \geq p_{C,SAF}$ $\quad$ $H_0: p_{E,EFF} \leq p_{C,EFF}$

  $H_1: p_{e,SAF} < p_{C,SAF}$ $\quad$ $H_1: p_{E,EFF} > p_{C,EFF}$

  (SAFETY) $\quad$ (EFFICACY)

- Suppose we need to reject BOTH nulls for the study to be considered successful
  - Each can be tested at a 0.05 level of significance
  - Power each endpoint at 80% could yield only 64% power to reject BOTH null hypotheses
  - We need to power each null with at least 90% power to have an 80% chance of rejecting BOTH null hypotheses

Multiple Primary Analyses

- Suppose there are two primary hypotheses; e.g.:

  $H_0: p_{E,SAF} \geq p_{C,SAF}$ $H_0: p_{E,EFF} \leq p_{C,EFF}$

  $H_1: p_{e,SAF} < p_{C,SAF}$ $H_1: p_{E,EFF} > p_{C,EFF}$

  (SAFETY) (EFFICACY)

- Suppose study is a success if we reject AT LEAST ONE null hypothesis
  - Powering each hypothesis at 80% at alpha = 0.025 (=0.05/2) yields 80% power to reject at least one null hypothesis