Power and Sample Size (StatPrimer Draft)

To achieve meaningful results, statistical studies must be carefully planned and designed. Study design has many aspects. Here's just a sampling of questions you must ask yourself when planning a study:

- How will the research question be operationalized?
- How will the study outcome be measured? Will it be objective? Will it be repeatable?
- How will relations be quantified?
- What parameter will be estimated?
- Will the study be experimental or non-experimental?
- If experimental, will blinding be employed? Will randomization be employed? What is the nature of the control group?
- If non-experimental, will data be prospective or retrospective? Will the sample be cross-naturalistic, cohort, or case-control? What are you going to do about confounders?
- How large a sample will be needed?

These notes address the last question for studies that seek to infer means, mean differences, proportions, and differences in proportions. Selected estimation and testing methods are considered.

Confidence Interval for a Mean

The goal is to estimate μ with stated "plus or minus" margin of error *m*. In estimating μ with 95% confidence, $m \approx \frac{2\sigma}{\sqrt{n}}$. Therefore,

$$n = \frac{4\sigma^2}{m^2}$$

In applying this method, considerable effort should be put into getting a good estimate of σ. (via prior studies, pilot studies, or "Gestalt").

Example: To obtain a margin of error of 5 when studying a variable with a standard deviation of 15, use $n = (4)(15^2)/(5^2) = 36$. To obtain a margin of error of 2.5 when studying this variable, use $n = (4)(15^2)/(2.5^2) = 144$. Notice that requiring a smaller margin of error required a larger sample size; half the margin of error required us to quadruple the sample size.

Testing Two Means

In testing H_0: $\mu_1 = \mu_2$ with equal sized groups, use

$$n = \frac{2\sigma^2\left(z_{1-\beta} + z_{1-\frac{\alpha}{2}}\right)^2}{\Delta^2}$$

per group, where $1 - \beta$ = the desired power, α = the significance-level, σ = the within group standard deviation, and Δ = a difference worth detecting.

Example: We want to test H_0: $\mu_1 = \mu_2$ at $\alpha = 0.05$ (two-sided) with 90% power and are looking for a mean difference of 1 mmol/L. We assume a within group standard deviation of 0.5 mmol.

Therefore, $n = \dfrac{2\sigma^2\left(z_{1-\beta} + z_{1-\frac{\alpha}{2}}\right)^2}{\Delta^2} = \dfrac{2 \cdot 0.5^2(1.28 + 1.96)^2}{1^2} = 5.2$. Use 6 per group.

Notes:

- Maximum efficiency is gained when the groups have equal sizes. If n_1 is restricted, apply the formula $n_2 = nn_1 / (2n_1 - n)$ to determine the size of the group 2.
- Always round-up results to ensure adequate power.
- Use of a software utilities (e.g., WipPepi, www.OpenEpi.com) is strongly encouraged, see this:

Sample Size For Comparing Two Means

Input Data

Confidence Interval (2-sided)	95%		
Power	90%		
Ratio of sample size (Group 2/Group 1)	1		
	Group 1	**Group 2**	**Mean difference[1]**
Mean			1
Standard deviation	0.5	0.5	
Variance	0.25	0.25	
Sample size of Group 1	6		
Sample size of Group 2	6		
Total sample size	12		

[1] Mean difference= (Group 1 mean) - (Group 2 mean)

The sample size formula for can be rearranged to determine the power of a test

$$1 - \beta = \Phi\left(-z_{1-\frac{\alpha}{2}} + \frac{\Delta}{\sqrt{\frac{2\sigma^2}{n}}}\right)$$

where $\Phi(z)$ is the cumulative probability of a standard Normal random variable. Use of a software utility is strongly encouraged.

Example: What is the power of a test of H_0: $\mu_1 = \mu_2$ when $\alpha = 0.05$ (two-sided), $n = 30$ per group, $\mu_1 - \mu_2 = 0.25$, and $\sigma^2 = 0.25$? *Solution*: $1 - \beta = \Phi\left(-z_{1-\frac{\alpha}{2}} + \frac{\Delta}{\sqrt{\frac{2\sigma^2}{n}}}\right) = \Phi\left(-1.96 + \frac{0.25}{\sqrt{\frac{2(0.25)}{30}}}\right)$ $=\Phi(-0.02) =$ 0.4920 (about 50/50).

Note: You can get the same result with www.OpenEpi.com.

Estimating a Proportion

Specify your best guess for population proportion (parameter) *p*. If you have no idea, use *p* = 0.5 to derive ensure adequate precision. State your desired level of confidence (e.g., 95%) and c computations a software utility such as OpenEpi.com.

Example: How many people do I need to estimate a proportion within a margin of error no greater than ±3%? I do not have a good idea of *p*, so will assume *p* = 0.5 for the sake of calculations. OpenEpi.com derives this output:

Sample Size for Frequency in a Population

Population size(for finite population correction factor or fpc)(*N*):	1000000
Hypothesized % frequency of outcome factor in the population (*p*):	50%+/-3
Confidence limits as % of 100(absolute +/- %)(*d*):	3%
Design effect (for cluster surveys-*DEFF*):	1

Sample Size(*n*) for Various Confidence Levels

ConfidenceLevel(%)	Sample Size
95%	1066
80%	457
90%	751
97%	1307
99%	1840
99.9%	2999
99.99%	4189

suggesting the need for 1066 observations.

How many people do I need to study to the proportion within a margin of error no greater than ±5%?

Sample Size for Frequency in a Population

Population size(for finite population correction factor or fpc)(*N*):	1000000
Hypothesized % frequency of outcome factor in the population (*p*):	50%+/-5
Confidence limits as % of 100(absolute +/- %)(*d*):	5%
Design effect (for cluster surveys-*DEFF*):	1

Sample Size(*n*) for Various Confidence Levels

ConfidenceLevel(%)	Sample Size
95%	384
80%	165
90%	271
97%	471
99%	664
99.9%	1082
99.99%	1512

This larger margin of error allowed for a smaller sample size.

Testing two proportions

Sample size: In testing H_0: $p_1 = p_2$, the sample size depends on α, power (1 – β), the difference worth detecting (Δ), the relative sizes of the samples (n_2 / n_1), and the proportion in the non-exposed group (p_2). Because computations are tedious, we rely on a software utility for computations.

Example: We want to test proportions from a cohort study. We use an equal number of individuals in the two groups, let alpha = 0.05 (two-sided), and power = 80%. We wish to detect a doubling of risk and assume the risk in the non-exposed group is 10%. The results from www.OpenEpi.com are to the right. Three different methods of calculation are reported. Pick one (and report its source). For example, the Fleiss method (middle column) states that we need 199 individuals per group.

Sample Size for Cross-Sectional & Cohort Studies & Clinical Trials

Two-sided significance level(1-alpha):	95
Power(1-beta, % chance of detecting):	80
Ratio of sample size, Unexposed/Exposed:	1
Percent of Unexposed with Outcome:	10
Percent of Exposed with Outcome:	20
Odds Ratio:	2.3
Risk/Prevalence Ratio:	2
Risk/Prevalence difference:	10

	Kelsey	Fleiss	Fleiss with CC
Sample Size - Exposed	201	199	219
Sample Size-Nonexposed	201	199	219
Total sample size:	402	398	438

Power: sample size formula can be rearranged to determine the power of a test based on a given sample size. For example, we can ask what would the power of the aforementioned example had we used only 100 individuals per group. Here the www.OpenEpi.com output.

Power for Cohort Studies

	Input Data
Two-sided confidence interval (%)	95
Number of exposed	100
Risk of disease among exposed (%)	20
Number of non-exposed	100
Risk of disease among non-exposed (%)	10
Risk ratio detected	2
Power based on:	
Normal approximation	50.82%
Normal approximation with continuity correction	42.45%

This says that the power would have been only 51% (Normal approximation). That would have been inadequate.

Testing proportions from case control studies

In testing H_0: $OR = 1$ for a case-control study, specify the expected odds ratio and expected proportion of controls who are classified as exposed. Here's output for an illustration that assumes $\alpha = 0.05$, $1 - \beta = 0.8$, an equal number of cases and controls, an exposure proportion of 0.25 in controls, and an expected odds ratio of 2.

Sample Size for Unmatched Case-Control Study

For:		
	Two-sided confidence level(1-alpha)	95
	Power(% chance of detecting)	80
	Ratio of Controls to Cases	1
	Hypothetical proportion of controls with exposure	25
	Hypothetical proportion of cases with exposure:	40
	Least extreme Odds Ratio to be detected:	2.00

	Kelsey	Fleiss	Fleiss with CC
Sample Size - Cases	154	152	165
Sample Size - Controls	154	152	165
Total sample size:	308	304	330

These conditions require 152 cases and 152 controls.