Power and Sample Size

John M. Boltri, M.D.
Mercer University School of Medicine
Cindy Passmore, M.S.
Waco Faculty Development Center
Robert L. Vogel, Ph.D.
Jiann-Ping Hsu School of Public Health
Georgia Southern University

Introduction to Power Analysis

- The power of a statistical test is the probability that it will yield significant results
- A priori power analysis
- A posteriori power analysis?

Introduction to Power Analysis

- How large a sample?
- Too small
  - Fail to answer question
  - Fail to detect associations
- Too large
  - More difficult
  - More costly

Introduction to Power Analysis

- α serves as a standard of proof
- Type 1 error (α) – rejecting the null hypothesis (H_0) when it is true-α set too low
- Type II error (β) – failing to reject H_0 when H_0 is false-power set too low

Introduction to Power Analysis

- Power = $1-\beta$
  = probability of correctly rejecting the null hypothesis (H_0)
- Power is a function of α, n, ES

Introduction to Power Analysis

- Power analysis is most easily discussed in terms of *statistical hypothesis testing*
- The simplest situation to consider is a single sample, one-tail test involving a proportion
- H_0: $P=P_0$ verses H_1:$P>P_0$
- H_0 is called the null hypothesis and it specifies there is no difference between the population proportion, P, and the hypothesized value P_0

Introduction to Power Analysis-2

- H_1, also written as H_A is called the alternate hypothesis or the research hypothesis. This is the hypothesis we would like to be true. The objective of the study is to show H_0 is unlikely - thus allowing us to claim H_1 is true. Power analysis helps us determine the likelihood of achieving this objective.

Factors Affecting Power

- Sample Size (n)
- Effect Size (ES)
- Variability
- Directionality of significant criterion
- α

Introduction to Power Analysis

- Four procedures to calculate sample size
  - Formulae
  - Ready made tables
  - Nomograms
  - Computer Software

Binomial Table

	P								
R	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	0.349	0.107	0.028	0.006	0.001	0.000	0.000	0.000	0.000
1	0.387	0.376	0.121	0.040	0.010	0.002	0.000	0.000	0.000
2	0.194	0.302	0.233	0.121	0.044	0.011	0.001	0.000	0.000
3	0.057	0.201	0.267	0.215	0.117	0.042	0.009	0.001	0.000
4	0.011	0.088	0.200	0.251	0.205	0.111	0.037	0.006	0.000
5	0.001	0.026	0.103	0.201	0.246	0.201	0.103	0.026	0.001
6	0.000	0.006	0.037	0.111	0.205	0.251	0.200	0.088	0.011
7	0.000	0.001	0.009	0.042	0.117	0.215	0.267	0.201	0.057
8	0.000	0.000	0.001	0.011	0.044	0.121	0.233	0.302	0.194
9	0.000	0.000	0.000	0.002	0.010	0.040	0.121	0.376	0.387
10	0.000	0.000	0.000	0.000	0.001	0.006	0.028	0.107	0.349

Binomial Table

- The column parameter P is the probability of obtaining a head on any one toss of a coin. If the coin is fair, P=0.5. If the coin is bias toward a head, P>0.5 and biased toward a tail, P<0.5. The row parameter R is the number of heads obtained in 10 tosses of the coin.
- In 10 flips of a fair coin, the probability of observing exactly 5 heads is 0.246 or 24.6%.

Interpretation of Alpha

- A type-I error occurs when an experiment results in the rejection of the null hypothesis when the null hypothesis is true.

Interpretation of Alpha - 2

- Suppose the null hypothesis is P=0.5 and the alternative hypothesis is P>0.5. Also suppose the decision rule is to reject H0 when R is 8, 9, or 10. From the table with P=0.5, the probability of obtaining 8, 9, or 10 heads is (0.044 + 0.010 + 0.001 = 0.055). That is, 5.5% of the coin tossing experiments using this rule result in a type-I error.

Interpretation of Alpha - 3

- Again - If the probability of obtaining a head on a toss of a coin is 0.5, then 5.5% of the experiments that use the rejection criterion of R=8, 9, or 10 will result in the false conclusion than P>0.5.

Interpretation of Alpha - 4

- **The value of alpha is based on a particular value of P.** If the actual value of P is 0.4 then our calculations based on P=0.5 are incorrect.

Interpretation of Alpha - 5

- Alpha is a statement about the proportion of experiments. In our example, alpha tells us what percentage of a large number of experiments will result in 8, 9, or 10 heads. It is a statement about what to expect in future experiments. It is not a statement about the parameter P.

Interpreting Beta and Power

- Beta is the probability of a type-II error. A type-II error occurs when we fail to reject the null hypothesis when the null hypothesis is false. Power = 1-beta and is the probability of rejecting a false null hypothesis.

Interpreting Beta and Power-2

- Suppose the true value of P is 0.7 and not 0.5. Our test is H_0: P=0.5 verses H_1: P>0.5. Also suppose we will reject the null if R is 8, 9, or 10. From the probability table, the probability of seeing 8, 9, or 10 heads with P=0.7 is 0.382 (0.233 + 0.121 + 0.028).
- The power is 0.382. Beta = 1-power = 1-0.382 = 0.618. If P=0.7 then 61.8% of the coin tossing experiments will result in a type-II error.

Interpreting Beta and Power-3

- If the probability of obtaining a head on a coin toss is 0.7, then 61.8% of the experiments that use the rejection criterion of R=8, 9, or 10 will result in the false conclusion that P=0.5.
- Beta is determined by a particular value of P.
- Because the rejection region (R=8, 9, or 10) depends on alpha, ***beta also depends on alpha***.
- Type-I and type-II errors cannot occur at the same time.

Interpreting Effect Size

- The effect size is the size of the change in the parameter of interest that can be detected by an experiment.
- Effect size is often given as a percentage change and not an absolute change. For example, suppose the baseline proportion of success is 0.40 and we wish to see a 10% increase. The effect size is $|0.40-0.44|=0.04$.

Six Steps to Calculating Sample Size

1. Formulate the study
2. Specify the Parameters for the Planned Analysis
3. Specify the Effect Size
4. Compute the Sample Size or Power
5. Sensitivity Analysis
6. Choose the Sample Size

Formulate the Study

1 Detail the Study Design:

A two-group, randomized, parallel, double blind study is planned. Patients will be studied for one year; each patient will be randomly assigned to receive either vitamin C or a placebo each day for one complete year. The sample sizes in the two groups will be equal.

Formulate the Study

2 Select the Primary Endpoint:

The primary endpoint is to determine if there is a difference in the average number of colds.

3 Specify the Analysis Method:

The mean number of colds will be compared between the two groups using a two-sample t-test. The null hypothesis states the mean number of colds is the same in both groups.

Specify the Parameters

- For the two group t-test, the parameters that must be specified are:
  - The significance level for the test, α. Usually we pick $\alpha=0.05$.
  - The number of tails (one sided or two?). Because we are testing for a difference in the average number of colds, the test is two-tailed.
  - The power for the test, $1-\beta$. Usually the minimum required power is 0.80.

Effect Size

- First step in sample size determination:
- Reliably detect clinically significant magnitude.
- The smaller the effect size the larger the sample size required.
- Choosing too large ES may not detect a difference.

Specify the Effect Size

- To specify the effect size, we must specify the expected mean difference between the two groups we would like to detect and the standard deviation within each of the two groups.

Specify the Effect Size - 2

- The most important part of this step involves what we know about the outcome variable. For example, we may know people in general suffer an average of 5 colds per year.

Specify the Effect Size -2

- With a baseline of 5 colds per year, we may decide a 40% reduction in the number of colds is clinically meaningful. We would use 5 and 3 as the mean number of colds for the placebo and Vitamin C group respectively. We have also noticed, based on 37 patients last year, the number of colds per person ranged from 0 to 11. This information allows us to estimate the s.d. to be approximately 1.8.

Using the Range to Estimate σ

- Given the range, an estimate of σ is approximately the 50^{th} percentile of the distribution of the range. In general, the probability density function for the range is $h(w)=\int n(n-1)[F(w+z)-F(z)]^{n-2}f(z)f(w+z)dz$ where F(*) is the probability distribution function of the standard normal distribution. h(w) is integrated from $-\infty$ to x, we want to know the value of x so the integral = 0.50.

Compute Sample Size

1. The t-value for significance is given from a central t-distribution based on alpha and the number of tails (ReqT).
2. The effect size, ES, is computed as |difference in means|/(common standard deviation).
3. For equal sample sizes, $n_1=n_2=n$

Compute Sample Size-2

4. The non-centrality parameter is computed as $NCP=ES*\sqrt{2}/N$
5. Power is given by the non-central t for NCP, ReqT and degrees of freedom, df=2n-2.
6. Since we know the power, we iteratively solve step 5.

- Most computer programs are designed to find the power given a sample size. To find sample size given power, an iterative algorithm is used. Usually starting with n=2,10 and 100.

Sample Size for the Example

- For our example with α =0.05, power = $1-\beta$=0.80, the average number of colds with placebo = 5 and an expected 40% reduction with Vitamin C (3 colds) and finally an estimate of the standard deviation based on the range is 1.8; the effect size is 2/1.8=1.11 and n=14 subjects per group.

Sensitivity Analysis

- Sensitivity Analysis allows the researcher to assess variability in required sample size or in resulting power.
- When computing sample sizes, sensitivity analysis involves exploring the changes in sample size when the design effect changes.

Sensitivity Analysis - 2

- The design effect may change due to changes in the treatment mean or the standard deviation.
- Researchers may also be interested in changing the power or alpha.

Sensitivity Analysis for the Example

- By changing the number of expected colds from 3 to 2, ES= 3/1.8=1.67 and n = 7/group.
- The expected colds is increased to 4 for the Vitamin C group gives an ES=1/1.8=0.55 and n=52/group.

Choose the Sample Size

- Once the sample size is calculated based on the original parameter estimates and a sensitivity analysis is performed, the researcher needs to select the appropriate sample size based on the analysis and cost constraints.
- Write a statement about the selected sample size, alpha, power and expected effect size.

Computer Packages Freebies-nada-nothing

- Russ Lenth – www.stat.uiowa.edu/~rlenth/Power/ - provides basics with interactive internet calculator
- Members.aol.com/johnp71/javastat.html – provides a list of links – one goes to Russ Lenth

Computer Packages Freebies-nada-nothing - 2

- PS: Power and Sample size calculation - www.mc.vanderbilt.edu/twiki/bin/main/powersamplesize (google PS: Power and Sample Size Calculation) – download (basics plus, survival analysis, regression, case-control studies)

Computer Packages Freebies-nada-nothing -3

- UnifyPow is a SAS Module/Macro available from Ralph O'Brien of the Cleveland Clinic Foundation.(steep learning curve) - www.bio.ri.ccf.org/Power/

- Much of this is also found in SAS Proc Power and Proc GLMpower.

Cheap Software

- Study Size -www.studysize.com/index.htm
- Cost - $79.00
- T-test, anova (1-way), Fisher's exact test, McNemar's test, Chi-square test, Life table test, Log-rank test, Log-rank test (No of Events), Wilcoxon rank-sum test, Wilcoxon signed-rank test, Sign test, Bioequivalence, Correlation, Regression
- I do not recommend

Computer Packages

- Three popular commercial packages are:
- Power and Precision - $595.00
- nQuery version 6 - $750.00
- PASS – PASS2002 regular $450.00
  PASS2005 regular $750.00
- All three have websites and provide demos
- I prefer nQuery

SAS Procs

- If you have SAS then you have Proc Power and Proc GLMPOWER
- Proc Power provides for t-tests, equivalence tests for means, confidence intervals for means, tests of binomial proportions, multiple regression, tests of correlation and partial correlation, one-way analysis of variance, two survival curves
- Proc GLMPOWER – allows for contrasts

Computer Packages

Stat. Test	Power/Pre	NQuery	PASS	Agree
1-sample t-test	Yes (CI)	Yes (CI)	Yes	Yes
Paired t-test	Yes (CI)	Yes (CI)	Yes	Yes
2-sample t-test; $=\sigma$	Yes (CI)	Yes (CI)	YES	Yes
2-sample t-test; $\neq\sigma$	Yes	Yes	Yes	Yes
Single prop.	Yes (CI)	Yes (CI)	Yes	Yes

Computer Packages - 3

Stat. Test	Power/Pre	NQuery	PASS	Agree
2 indep. Prop.	Yes (CI)	Yes (CI)	Yes	Yes
Paired Prop.	Yes (CI) McNemar	Yes McNemar	Yes McNemar	Yes
Cross-tabs	Yes	2-by-2 only	YES	Yes
Odds Ratio	Yes	Yes	Yes **	Yes
Oneway Anova	Yes	Yes	Yes	Yes

** Indirectly via Case-Control-Matched Samples

Computer Packages - 4

Stat. Test	Power/Pre	NQuery	PASS	Agree
AOV contrast	No	Yes	Yes *	--------
Factorial AOV	Yes – 3 factors	No	Yes – 3 factors	Yes
ANCOV.	Yes	No	No	------
Repeated Measures	No	Yes**	Yes***	No
Corr.	Yes (CI)	Yes	Yes	Yes

**Has a Geisser-Greenhouse Correction and allows for a within factor contrast
***Treats the design as if it were a split plot design

Computer Packages - 5

Stat. Test	Power/Pre	NQuery	PASS	Agree
Linear Reg.	Yes	Yes	Yes	Yes
Logistic Reg.	Yes	Yes	Yes	Yes
Survival Analysis	Yes	Yes	Yes	Yes
Effect Calc.	Yes	Yes	No	----
Formulae	Yes	Yes	Yes	------

Computer Packages - 6

Stat. Test	Power/Pre	NQuery	PASS	Agree
Graphics	Yes	Yes	Yes	--------
Unequal n	Yes	Yes	Yes	-------
Cost	$595	$750	$450 $750	------
Rec.	Yes	Yes	Yes	----

Computer Packages -7

- All three packages provide information and/or references about the algorithms used to calculate power and sample size.
- All three packages provide graphical output.
- nQuery and Power and Precision provide effect calculators.

SAS Procedures POWER and GLMPOWER

- Proc Power – t-tests, equivalence, proportions, multiple regression, correlation, one-way anova and rank tests for survival.
- Proc GLMPower – specify models and contrasts using code similar to PROC GLM and MIXED.

Example 1

- We wish to compare the reduction in blood pressure resulting from use of a placebo, a standard drug and a new drug at both a low dose and a high dose. Previous studies with the standard drug suggest a standard deviation of about 6mmHg. In the past the placebo has resulted in reductions of about 5mmHg and the standard drug of about 12mmHg. The researcher guesses the low dose new drug will result in a reduction of about 10.5 mmHg and the high dose will result in a reduction of about 13.5 mmHg.

nQuery Advisor answer

Single one-way between means contrast (equal n's)

	1	2	3	4
Test significance level, α	0.013	0.013	0.013	0.013
1 or 2 sided test?	2	2	2	2
Number of groups, G	4	4	4	4
Contrast, C = $\Sigma c_i \mu_i$	-5.5	-8.5	1.5	-1.5
Scale, D = SQRT(Σc_i^2)	1.4	1.4	1.4	1.4
Common standard deviation, σ	6	6	6	6
Effect size, Δ = |C|/(σD)	0.648	1.002	0.177	0.177
Power (%)	90	90	90	90
n per group	35	16	455	458

nQuery effect size calculator

Group	Mean	Coefficient
1	5.000	1.000
2	12.000	0.000
3	10.500	0.000
4	13.500	-1.000

*Contrast, $C = \Sigma c_i \mu_i$ -8.500 $\Sigma C_i = 0.000$

*Scale, $D = SQRT(\Sigma c_i^2)$ 1.414

* Values are pasted into previous table

```
data bp;
   input treat$ reduce cellwgt @@;
   datalines;
   1placebo     5     1          2standard     12    1
   3low         10.5  1          4high         13.5  1
run;
proc glmpower data=bp data=order;
   class treat;
   model reduce=treat;
   weight cellwgt;
   contrast 'placebo vs. low'   treat 1 0 -1 0;
   contrast 'placebo vs. high'  treat 1 0 0 -1;
   contrast 'standard vs. low'  treat 0 1 -1 0;
   contrast 'standard vs. high' treat 0 1 0 -1;
   power
      stddev=6    alpha=0.0125            ntotal = .
   power=0.9;
run;
```

SAS GLMPower Results for BP

Fixed Scenario Elements	
Dependent Variable	reduce
Weight Variable	cellwgt
Alpha	0.0125
Error Standard Deviation	6
Nominal Power	0.9

SAS GLMPower Results for BP

Computed N Total						
Index	**Type**	**Source**	**Test DF**	**Error DF**	**Actual Power**	**N Total**
1	Effect	treat	3	68	0.909	72
2	Contrast	placebo vs low	1	136	0.902	140
3	Contrast	placebo vs high	1	60	0.920	64
4	Contrast	standard vs low	1	1828	0.900	1832
5	Contrast	standard vs high	1	1828	0.900	1832

Example 2: Survival Analysis

- We wish to compare survival for two groups using the logrank test. The hazard for the reference group is estimated as 0.7 and the hazard ratio is estimated at 1.2. There is expected a two-year accrual period and a two year follow-up for each patient. The requirements for power and alpha are 0.90 and 0.05 respectively.

nQuery Advisor Solution

Two group test of equal exponential survival (n large), no dropouts

	1
Test significance level, α	0.050
1 or 2 sided test?	2
Length of accrual period	2.00
Maximum length of followup	2.00
Group 1 exponential parameter, λ_1	0.8400
Group 2 exponential parameter, λ_2	0.7000
Hazard ratio, $h=\lambda_1 / \lambda_2$	1.200
Power (%)	90
n per group	1298
Total number of events required, E	1264

SAS PROC POWER – Survival Example

```
ods listing close;
ods rtf file = 'c:\bass\power2.rtf';
proc power;
twosamplesurvival test=logrank
   groupsurvexphazards = 0.84 | 0.7
   accrualtime=2
   totaltime = 4
   nsubintervals=1
   ntotal = .
   power = 0.9;
   run;
ods rtf close;
ods listing;
```

SAS Output – Example 2

Fixed Scenario Elements	
Method	Lakatos normal approximation
Form of Survival Curve 1	Exponential
Form of Survival Curve 2	Exponential
Accrual Time	2
Total Time	4
Number of Time Sub-Intervals	1
Group 1 Survival Exponential Hazard	0.84
Group 2 Survival Exponential Hazard	0.7
Nominal Power	0.9
Number of Sides	2
Group 1 Loss Exponential Hazard	0
Group 2 Loss Exponential Hazard	0
Alpha	0.05
Group 1 Weight	1
Group 2 Weight	1

Computed N Total	
Actual Power	**N Total**
0.900	1308

Bibliography

- Desu MM, Raghavarao D. Sample Size Methodology. 1990; Academic Press. New York.
- Elashoff JD. nQuery Advisor Version 4.0 User's Guide. 2000; Statistical Solutions Ltd. Cork, Ireland.
- Borenstein M, Rothstein H, and Cohen J. , et.al. Power and Precision. 2001; Biostat. Teaneck, NJ.
- Hintze JL. Pass 2000: Power Analysis and Sample Size for Windows User's Guide. 2000; NCSS. Kaysville, Utah.
- Gupta S. Quick Results with Output Delivery System. 2003. Cary, NC: SAS Institute Inc.
- SAS/STAT User's Guide version 9.2. 2005. Cary, NC: SAS Institute Inc.