I. Introduction

The importance of sample size and power considerations in planning and evaluating research studies should not be underestimated. If the sample size is inadequate, one can be faced with the problem of large (and substantively important) sample differences not being statistically significant. Similarly, failing to reject the null hypothesis in studies based on small samples, showing only trivial differences in sample estimates, can be very risky without some estimate of the probability of detecting a relevant difference (i.e. the power). In view of this, Lachin states three questions one must consider in planning and evaluating research studies:

Q1. What sample size is required to ensure adequate power in detecting a relevant difference?

Q2. What is the power for detecting a relevant difference when the sample sizes are fixed?

Q3. What is the minimum relative difference detectable with fixed sample size and specified power?

To answer these questions, it can often be necessary to perform several sample size (power) calculations in planning studies with multiple endpoints and/or studies where a wide range of alternative hypotheses (i.e. relevant differences) are to be evaluated. To facilitate these calculations, SAS macros have been written to evaluate Q1 and Q2, above, for comparing means and proportions in the one- and two-sample cases. In addition, sample size and power calculations are presented for the two-sample censored survival data problem.

It should be noted that the answers to Q1 and Q2 are generally easy to compute when the sample means are assumed to be normally distributed. However, Q3 requires iterative methods when the variance of the sample mean depends on the true mean (e.g. proportions). The macros written allow the user to evaluate Q1 and Q2 over a range of relevant differences, hence, usually allowing evaluation of Q3.

II. Equations Used in Macros

Since the development of the formulas to be presented can be found in several texts, only a minimal amount of background information (allowing one to understand the notation) is presented. The following notation will be used throughout this section:

- $H_0$ and $H_A$ represent the null and alternative hypotheses to be tested.
- $a$, Probability of Type I error = $Pr(\text{reject } H_0 | H_0 \text{ is true})$.
- $\beta$, Probability of Type II error = $Pr(\text{fail to reject } H_0 | H_A \text{ is true})$.
- $z$, $(z, Pr(Z > z) = \alpha)$ where $Z$ is a normally distributed random variable with mean zero and variance one, i.e. $Z \sim N(0,1)$.
- $z$, $(z, Pr(Z > z) = \beta)$.
- $N_i$, integer part of $N$.

All formulas are presented for one-sided alternative hypotheses. If a two-tailed test is desired, one should replace $\alpha$ with $\alpha/2$. Sample size values are rounded up to the next largest integer before being printed. The equation for $z$ is presented rather than the power directly. The macro output prints the power directly where the power = $(1 - Pr(Z > z))$.

A. One-sample test for mean: Suppose we wish to study the random variable $X$ in a population where $X \sim N(\mu, \sigma^2)$. The aim of the study is to choose between $H_0: \mu = \mu_0$ and $H_A: \mu = \mu_1$ $(\mu_1 > \mu_0)$ on the basis of a sample of size $N_1$. Note that if $X$ represents the difference in measurements of paired data and $\mu_1 = \mu_0$, then we have the paired $t$-test.

1. The sample size required in this situation with $\sigma^2$ known is:

$$N_1 = \frac{\sigma^2(z_{\alpha/2} + z_{\beta})^2}{(\mu_1 - \mu_0)^2}$$

using standard normal theory. If $\sigma^2$ is unknown (as is almost always the case), one would perform a $t$-test and would need a slightly larger sample size than that given by (1) to preserve the desired $\alpha$ and $\beta$. In view of this, a simple method suggested in Snedecor and Cochran to increase the sample size appropriately is to use a sample of size $N_1^*$, where $N_1^* = N_1 \cdot f$, and $N_1$ is calculated from (1) using $f = (df+3)/(df+1)$ and $df = (N_1^* - 1)$.  

2. The power for detecting a difference of $|\mu_1 - \mu_0|$ with a sample size $N_1$ is calculated by solving (1) for $z_B$:

$$z_B = \frac{\sqrt{N_1} |\mu_1 - \mu_0|}{\sigma} - z_\alpha$$

if $\sigma^2$ is known.

If $\sigma^2$ is unknown, the following expression is used to estimate the power:

$$z_B = \frac{\sqrt{N_1} |\mu_1 - \mu_0|}{\hat{\sigma}} - z_\alpha$$

633
where $N^* = \left[ \frac{N_1}{f} \right]$. 

$f = \frac{(df+3)}{(df+1)}$ and $df = \frac{N_1 - 1}{2}$.

Note that the sample size used in calculating $N^*$ is decreased to reflect the slight loss of power when using the t-test.

B. Two-sample test comparing means: Given the random variable $X \sim N(\mu_1, \sigma^2)$ in Population 1 and $X \sim N(\mu_2, \sigma^2)$ in Population 2, one wishes to choose between $H_0: \mu_1 = \mu_2$ and $H_A: \mu_2 > \mu_1$.

1. The sample size from Population 1 is denoted by $N_1$ and that from Population 2 by $N_2 = N_1 + \xi$. (While letting $R=1$ provides the most efficient use of one's resources, there may be occasions when values of $R$ are desirable. Hence, the macros written allow one to design studies with unequal sample sizes.) For $\sigma^2$ known, $N_1$ is calculated as follows:

$$N_1 = \frac{(z_\alpha^2 + z_{1-\alpha}^2) (1 + \frac{1}{R})}{\sigma^2 (1/N_1 + 1/N_2)}$$

For $\sigma^2$ unknown, the two-sample t-test is used again using the Snedecor and Cochran adjustment the sample size is:

$$N_1^* = N_1 f, \quad \text{where } f = \frac{(df+3)}{(df+1)}, \quad \text{df} = \frac{N_1 + 1}{2} (1+R)-2.$$

2. The power of the t-two-sample test (given samples of size $N_1$ and $N_2$) is calculated using:

$$\frac{\sigma}{\sigma_1} - \frac{\sigma}{\sigma_2}$$

where $\sigma_1 = \left[ \frac{N_1}{f} \right]$, $\sigma_2 = \left[ \frac{N_2}{f} \right]$, $f = \frac{(df+3)}{(df+1)}$ and, $\text{df} = \frac{N_1 + 1}{2} (1+R)-2.$

C. One-sample test for a proportion: Let $X$ be the Bernoulli random variable of interest with $Pr(X=1) = \pi$ and $Pr(X=0) = 1-\pi$. The aim of the study is to choose between $H_0: \pi = \pi_1$ and $H_A: \pi_2 > \pi_1$ (say), on the basis of a sample of size $N_1$.

1. If we assume that $\hat{\pi} = \frac{1}{N_1} \sum Y_i \sim N(\pi, \frac{\pi(1-\pi)}{N_1})$, then the sample size required is:

$$N_1 = \frac{\left( z_{\alpha/2}^2 (1-\pi_1) + z_{1-\alpha/2}^2 (1-\pi_2) \right)}{\left( \pi_2 - \pi_1 \right)^2}.$$  

Transforming the proportions using $A(\pi) = \arcsin(\sqrt{\pi})$ before doing the normal test has been advocated by several authors. Discussion concerning the desirability of this transformation in estimating sample size can be found in papers by Lemeshow and Haseman. In view of this, the sample size calculated using the arcsin transformation and the normal test (where $A(\pi) \leq N(A(\pi), 1/N_1)$) is also presented and is:

$$N_1 = \frac{\left( z_{\alpha}^2 + z_{1-\alpha}^2 \right)}{\left( A(\pi_2) - A(\pi_1) \right)^2}.$$  

2. The estimated power given $N_1$ is calculated using:

$$z_\alpha = \sqrt{N_1} \left[ \frac{1}{\sqrt{\sigma_1^2 + 1/\sqrt{\sigma_2^2}}} - 1 \right]$$

for the normal approximation, and

$$z_\alpha = \sqrt{N_1} \left[ \frac{\pi_2 - \pi_1}{\sqrt{\pi_2(1-\pi_2)} + \sqrt{\pi_1(1-\pi_1)}} \right]$$

for the approximate normal test based on the arcsin transformation.

U. Two-sample test comparing proportions: Assume that $X$ is distributed Bernoulli ($\pi_1$) in Population 1 and $X$ is distributed Bernoulli ($\pi_2$) in Population 2. It is desired to test $H_0: \pi_1 = \pi_2$ versus $H_A: \pi_2 > \pi_1$, by taking samples of size $N_1$ and $N_2$ from each population.

1. If it is assumed that

$$\hat{\pi}_1 = \frac{\sum Y_1}{N_1}, \quad \hat{\pi}_2 = \frac{\sum Y_2}{N_2},$$

$$\hat{\pi}_1 \sim N(\pi_1, \frac{\pi_1(1-\pi_1)}{N_1}), \quad \hat{\pi}_2 \sim N(\pi_2, \frac{\pi_2(1-\pi_2)}{N_2})$$

then the sample size $(N_1, N_2)$ required is:

$$N_1 = \frac{\left( z_{\alpha}^2 (1-\pi_1) (1+\pi_2/R) + \pi_2 \left( (1-\pi_2) \sqrt{1+\frac{R}{\pi_2(1-\pi_2)}} \right) \right)}{\left( \pi_2 - \pi_1 \right)^2},$$

$$N_2 = \frac{\left( z_{\alpha}^2 (1-\pi_2) (1+\pi_1/R) + \pi_1 \left( (1-\pi_1) \sqrt{1+\frac{R}{\pi_1(1-\pi_1)}} \right) \right)}{\left( \pi_2 - \pi_1 \right)^2},$$

where $\pi = \frac{\pi_1 + \pi_2}{2}$.

If the arcsin transformation is used, assuming that

$$A(\hat{\pi}_1) \leq N(A(\pi_1), 1/N_1)$$

where

$$A(\hat{\pi}_1) \leq N(A(\pi_1), 1/N_1)$$

then

$$N_1 = \frac{\left( z_{\alpha}^2 + z_{1-\alpha}^2 \right)}{\left( A(\pi_2) - A(\pi_1) \right)^2}.$$
E. Comparison of two median survivals in a clinical trial

If one is planning a study to compare the median survival time to some event between two populations (e.g. one given the standard treatment for a specific type of cancer, the other given a new treatment), and wishes to estimate the necessary sample size in each group or the power for a fixed sample size, some assumptions must be made about the distribution of survival time and the recruitment of subjects. The assumptions and equations which follow are essentially as presented by Lachin. Another excellent reference on the topic is the text by Gross and Clark.

If we let \( P_1(t) \) be probability of surviving to time \( t \) for subjects in Population 1 and assume exponential survival with hazard rate \( \lambda_1 \), then \( P_1(t) = e^{-\lambda_1 t} \) and the median survival time is \( M_1 = \frac{\ln(2)}{\lambda_1} \). If we further assume uniform accrual (i.e. uniform censoring) to time \( T_0 \) from the start of the study and termination of follow-up at time \( T(T_0) \) after the start of the study, then we have the following:

\[
\hat{M}_1 = \frac{N_i}{\sum \phi(\hat{\lambda}_1) / \hat{N}_i},
\]

where \( \hat{M}_1 \) is the estimated median survival time in Population 1.

Hence, the sample size \( N_1, N_2 \) required for testing \( H_0 : M_1 = M_2 \) versus \( H_1 : M_1 > M_2 \) using a normal approximation is:

\[
N_1 = \frac{\left( z_{\alpha/2} \phi(\hat{\lambda}_1) / \hat{N}_1 \right)^2 + \phi(\hat{\lambda}_2) / \hat{N}_2}{\phi(\hat{\lambda}_1)^2}
\]

where \( \lambda = \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2} \)

and \( \lambda_1 = \frac{\ln(2)}{N_1} \) and \( \lambda_2 = \frac{\ln(2)}{N_2} \).

2. The estimated power given calculation as follows for the normal approximation (chi-square test):

\[
z_\beta = \frac{[\phi(\hat{\lambda}_1)]^2}{\phi(\hat{\lambda}_1) / \hat{N}_1 + \phi(\hat{\lambda}_2) / \hat{N}_2}
\]

Using the normal test based on the arcsin transformation, the following expression for power is used:

\[
z_\beta = \frac{[\phi(\hat{\lambda}_1) - \phi(\hat{\lambda}_2)]}{\sqrt{\phi(\hat{\lambda}_1) / \hat{N}_1 + \phi(\hat{\lambda}_2) / \hat{N}_2}}
\]

Note that the macros as written require \( M_1 \) and \( M_2 \) as input for (17) and (18). If the investigator cannot provide \( M_1 \) and \( M_2 \), but instead estimates the 10-year survival in Population 1 as \( P_1(t_0) \) and that of Population 2 as \( P_2(t_0) \), then \( M_1 \) and \( M_2 \) can be calculated as

\[
M_i = \frac{\ln[P_i(t_0)]}{tn(2)}
\]

If the logrank test is to be used to compare the survival, Rubinstein, Gail and Satter have done simulation studies (for estimating trial lengths) which suggest that sample size and power estimates derived using the above methods will be very accurate when the true survivals are exponential and give reasonable approximations if the true survivals are Weibull.

III. Macro Use and Description

A. General instructions for use: Because the input required to evaluate sample size (power) changes depending on the type of comparison, 13 separate SAS macros (sample size and power for each of the 5 comparisons in II.A-II.E above) were written. Figure 1 provides a listing of the name of the macro (column 2) to be called for execution of the equations given in II. This will be referred to as the "execution macro". Prior to calling the specific execution macro, one must define its "variable definition macro" (column 3) which is called by the execution macro. Column 4 of Figure 1 lists the name of the variables to be defined for each variable definition macro.

(If there is any doubt as to the meaning of the variables, one should refer to the equations.) Note that one can input the desired power directly (e.g. POWER=.90) without determining the value of \( z_\beta \). The range of values for the parameter of interest under the alternative hypothesis (i.e. \( M_2 \) or \( M_2 \), \( T_2 \) or \( T_2 \), \( M_2 \) or \( M_2 \)) is determined by using the variables whose first 4 letters start with \( M_1 \) (for the minimum value under \( H_1 \)), \( \text{MAX} \) (for the maximum value under \( H_2 \)), and \( \text{INC} \) (for the increments). The \( \text{MAX} \) and \( \text{INC} \) variables are needed only if more than one alternative is to be evaluated. Prior to calling the desired execution macro, one must have defined the variable definition macro and two additional macros specifying the \( \alpha \) level and whether a one- or two-sided test is desired. These macro names are...
Examples: Two typical examples will now be presented, both using \( n = 0.05 \) and a one-sided test. Example 1 is to calculate the power for comparing two proportions where \( m = 10 \) and \( n \) is allowed to vary from \( 12 \) to \( 30 \) in increments of \( 0.02 \), and \( N = 50 \) and \( N_2 = 100 \). The execution macro is \( T2PR \) and the variable definition macro is \( T2PR \).

Example 2 demonstrates how to estimate the sample size necessary (with power \( = 0.90 \) and \( \alpha = 0.05 \)) to detect an increase in median survival from \( M_1 = 2 \) to \( M_2 = 3 \) years. It is assumed that patients will be accrued for two years (\( T = 2 \)) and that follow-up will be terminated after five years (\( T = 5 \)). The code for both examples is listed below and the output in Figure 2.

```
MACRO ALPHA ALPHA=.05;
MACRO SIDES SIDES=1;

/* EXAMPLE1: EQUATIONS (15) & (16) */
MACRO ZPR PR P11=.10; MIN_P12=12; MAX_P12=.30; INC_P12=.02; N1=50; N2=100; Z2SV SS

/* EXAMPLE2: EQUATION (17) */
MACRO 2SV SS; MIN_MED2=3;

MACRO 25V S5; MED1=2; MIN_MED2=3;

/* EXAMPLE2: EQUATION (17) */
MACRO 2PR PR P11=.10; MIN_P12=12; MAX_P12=.30; INC_P12=.02; N1=50; N2=100; Z2SV SS
```

C. SAS code: Figure 3 is a listing of the SAS code for a typical macro, \( T2PR \) (written as \( T2PR \)). Calculations over the specified range of alternative hypotheses are carried out by repeated execution of the equations using the DO loop. \( T2PR \) is used to allow plotting of the power (sample size) versus alternative hypotheses conditional on the \( N \) variable being present. Copies of the other nine macros, which are similar in structure to \( T2PR \), are available by writing the author at Medical Research Statistics, Mayo Clinic, Rochester, MN 55903.

Acknowledgement

I would like to thank Ms. Marilyn Ness for skillfully typing this paper.

References


Figure 1. Listing of the Macro Names for Each Type of Sample Size or Power Calculation and the Macro Name Needed for Definition of Input Variables

<table>
<thead>
<tr>
<th>Type of Comparison</th>
<th>Macro Name for Execution Equations</th>
<th>Macro Name for Variable Definition</th>
<th>Required Variables for the Variable Definition Macro</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Size for 1 Mean vs Constant (3)</td>
<td>T2PR</td>
<td>T2PR</td>
<td>Z2PR SS, MIN MED1, MIN MED2, MAX MED1, MAX MED2, POWER, ( r )</td>
</tr>
<tr>
<td>Sample Size for 1 Mean vs Constant (4)</td>
<td>T2PR</td>
<td>T2PR</td>
<td>Z2PR SS, MIN MED1, MIN MED2, MAX MED1, MAX MED2, POWER, ( r )</td>
</tr>
<tr>
<td>Sample Size for 2 Means (x)</td>
<td>T2PR</td>
<td>T2PR</td>
<td>Z2PR SS, MIN MED1, MIN MED2, MAX MED1, MAX MED2, POWER, ( r )</td>
</tr>
<tr>
<td>Sample Size for 1 Proportion vs Constant (9, 10)</td>
<td>T2PR</td>
<td>T2PR</td>
<td>Z2PR SS, MIN P12, MAX P12, POWER</td>
</tr>
<tr>
<td>Sample Size for 1 Proportion vs Constant (11, 12)</td>
<td>T2PR</td>
<td>T2PR</td>
<td>Z2PR SS, MIN P12, MAX P12, POWER</td>
</tr>
<tr>
<td>Sample Size for 2 Proportions (13, 14)</td>
<td>T2PR</td>
<td>T2PR</td>
<td>Z2PR SS, MIN P12, MAX P12, POWER</td>
</tr>
<tr>
<td>Sample Size for 2 Proportions (15, 16)</td>
<td>T2PR</td>
<td>T2PR</td>
<td>Z2PR SS, MIN P12, MAX P12, POWER</td>
</tr>
<tr>
<td>Sample Size for 1 Median Survival (17)</td>
<td>T2PR</td>
<td>T2PR</td>
<td>MED1, MIN MED2, MAX MED2, POWER, ( T )</td>
</tr>
<tr>
<td>Sample Size for 2 Median Survival (18)</td>
<td>T2PR</td>
<td>T2PR</td>
<td>MED1, MIN MED2, MAX MED2, POWER, ( T )</td>
</tr>
</tbody>
</table>

If a single alternative is desired, the 'MAX' & 'INC' variables are not needed.
POWER CALCULATIONS FOR COMPARISON OF TWO PROPORTIONS

Both the normal (PR_NML) & arcsin (PR_AR S) with normal approx. are presented

<table>
<thead>
<tr>
<th>ALPHA</th>
<th>SIDES</th>
<th>N1</th>
<th>N2</th>
<th>PI1</th>
<th>PI2</th>
<th>PR_NML</th>
<th>PR_AR S</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>1</td>
<td>50</td>
<td>100</td>
<td>0.1</td>
<td></td>
<td>0.094140</td>
<td>0.101075</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.156886</td>
<td>0.175703</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.244165</td>
<td>0.271487</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.346148</td>
<td>0.381784</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.497632</td>
<td>0.497460</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.570582</td>
<td>0.609187</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.676042</td>
<td>0.709403</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.767771</td>
<td>0.793423</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.842161</td>
<td>0.859592</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.898500</td>
<td>0.908733</td>
</tr>
</tbody>
</table>

PLOT OF PR_NML*PI2 SYMBOL USED IS N

PLOT OF PR_AR S*PI2 SYMBOL USED IS A

SAMPLE SIZE REQUIREMENTS FOR COMPARISON OF TWO MEDIAN SURVIVALS

Assuming uniform accrual for T years and analysis at T2 years
Survival is assumed to be exponential & a normal approx. used

<table>
<thead>
<tr>
<th>ALPHA</th>
<th>SIDES</th>
<th>POWER</th>
<th>T TO T_MED1</th>
<th>T_MED2</th>
<th>N1</th>
<th>N2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>1</td>
<td>0.9</td>
<td>1 2 5 2</td>
<td>3 159 159</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 3: Power Macro for Comparing Two Proportions

PROC PRINT; TITLE6 POWER CALCULATIONS FOR COMPARISON OF TWO PROPORTIONS;
TITLE7 BOTH THE NORMAL(PR_NML) & ARCSIN(PR_ARC) WITH NORMAL APPROX. ARE
PRESENTED;
%INCLUDE PLOT;

MACRO _2PR_PR_
DATA T1;
_ALPHA _SIDES _2PP_PP
  ZALPHA=(PROBIT(1-ALPHA))/%SIDES=1 + (PROBIT(1-ALPHA/2))/%SIDES=2;
  IF MAX_PI2 NE . THEN DO; FILE PLOT; PUT
    'PROC PLOT; PLOT PR_NML*PI2="N" PR_ARC*PI2="A" / ';
    'OVERLAY HPOS=601';
  END;
  IF MAX_PI2=. THEN DO:
    MAX_PI2=MIN_PI2+1; INC_PI2=MIN_PI2+2; *NEED 1 EXEC OF DO;
    FILE PLOT; PUT
      'TPI1=PI1; TN1-N1; TN2=n2;
      DO PI2=MIN_PI2 TO MAX_PI2 BY INC_PI2;
      TPI2=PI2 TO MAX_PI2 BY INC_PI2;
      PBAR=(TN1+TPI1+TN2+PI2)/(TN1+TN2);
      *NORMAL APPROX. TO BINOMIAL:
      ZB_NML=(ABS(TPI1-PI2)-ZALPHA*SQRT(PBAR*(1-PBAR)*1/TN1+1/TN2))/$SORT(TPI1=1-TPI1/1+PI2/1-TN2);
      PR_NML=PROBNORM(ZB_NML);
      *NORMAL APPROX. FOLLOWING ZARCSIN(SORT(PI)) TRANSFORMATION:
      AS_PI1=ZARCSIN(SORT(TPI1));
      AS_PI2=ZARCSIN(SORT(PI2));
      ZD_ARC=ABS(AS_PI1-AS_PI2)/$SORT(1/TN1+1/TN2)-ZALPHA;
      PR_ARC=PROBNORM(ZD_ARC);
      IF PI2=MIN_PI2 THEN DO:
        ALPHA=. SIDES=. N1=. N2=. PI1=.;
      END;
      OUTPUT;
    END;
  END;
END;
PROC PPINT; TITLE6 POWER CALCULATIONS FOR COMPARISON OF TWO PROPORTIONS;
TITLE7 BOTH THE NORMAL(PR_NML) & ARCSIN(PR_ARC) WITH NORMAL APPROX. ARE
PRESENTED;
%INCLUDE PLOT;

638