Paper 4675-2020

Sample Size Calculation Using SAS®, R, and nQuery Software

Jenna Cody, Johnson & Johnson

ABSTRACT

A prospective determination of the sample size enables researchers to conduct a study that has the statistical power needed to detect the minimum clinically important difference between treatment groups. With knowledge or assumptions about the study design, drop-out rate, variation of the outcome measure, and desired power and alpha levels, the required sample size for a study can be calculated. This paper discusses methods for calculating sample size by hand and through the use of statistical software. It walks through the method for computing sample size using the POWER procedure and the GLMPOWER procedure in SAS® and compares the commands and user interfaces of SAS with R and nQuery software for sample size calculations.

INTRODUCTION

Selecting the appropriate sample size for a study is one of the fundamental tasks required of a statistician. Whether the statistician is determining the number of patients to enroll in a clinical trial, voters to complete a political poll, or mice to include in a lab experiment, the same input factors of power, significance criteria, and effect size can be used to successfully identify the sample needed. A sample that is too small can lead to an analysis that fails to identify any trends due to inadequate power, while a sample that is too large can lead to wasted time and resources. In clinical studies, sample size determination is not only a statistical issue, but an ethical issue. Enrolling too few subjects in a clinical trial can lead to unnecessary hardship and exposure to a study agent for a study that was never capable of drawing conclusions to establish efficacy of the compound. Enrolling too many subjects can cause potentially unnecessary exposure to inferior treatments. Sample size determinations can be completed by hand or through one of the many available software packages, such as SAS, R, and nQuery.

BACKGROUND INFORMATION AND INPUTS

STATISTICAL POWER

Statistical power is defined as the probability of rejecting the null hypothesis when the alternative hypothesis is true, or, in other words, the probability of a correct rejection. Written mathematically, it can be represented as $\Pr(reject\ H_0|H_1\ is\ true)$ or as 1 − β, where β is equal to the probability of Type II error (i.e. "false negative" result). Because power is a probability, it can take on values between 0 and 1. Although this may greatly differ based on the study design and field of study, conventional thresholds for statistical power are usually around 0.8 to 0.9 (80% to 90%).

Statistical power and sample size are inextricably linked, with a positive correlation between power and sample size. That is, given equality of all other factors, a higher requirement of statistical power will yield a higher required sample size. Similarly, a higher sample size in a study will yield a higher power for that study if all other factors are held constant.

Statistical power can be used to calculate the minimum sample size required to detect a specified effect size. For example, if the aim of a study is to detect a scientifically meaningful difference in growth of two plant varieties, and the desired power and alpha

level are pre-specified, the researcher will be able to calculate exactly how many plants to include in the experiment to identify the meaningful difference in growth. Similarly, it can be used to calculate a minimum effect size likely to be detected given a specified sample size. If the same researcher only had access to a limited number of plants, she or he could identify the effect size likely to be detected at a set level of power with the available sample size.

Statistical power can be used to make comparisons between statistical tests. With all other factors equal, tests yielding higher power represent stronger evidence of the outcome identified than tests with lower power. Power analysis can reveal the statistical test likely to yield the highest level of evidence under varying sample sizes and effect sizes.

Statistical power can also play a role in determining whether studies are stopped early. In longitudinal studies with elements of adaptive design at interim time points, it is common to pre-specify stopping boundaries based on the outcome. In these types of studies, it is imperative that stopping boundaries are pre-specified. When interim stopping rules are set up correctly, data supporting a strongly positive outcome can lead to an early termination of the study for efficacy, and data supporting a non-efficacious outcome can lead to an early termination of the study for futility.

Power analysis improves the chances of conclusive results. When potential outcomes are examined prospectively and assumptions are well thought out, researchers can set up the study in a way that success is likely, and can avoid conducting studies that are likely to fail.

Type I and Type II Error

Statistics is the study of drawing inferences based on incomplete information. Therefore, there is inherent uncertainty in every statistical test completed. This uncertainty can be captured in two types of errors:

- Type I Error: the probability of rejecting the null hypothesis when the null hypothesis is true (i.e. false positive). This is represented by α and can be written mathematically as $\Pr(reject\ H_0 | H_0\ is\ true)$.
- Type II Error: the probability of accepting the null hypothesis when the alternative hypothesis is true (i.e. false negative). This is represented by β and can be written mathematically as $\Pr(accept\ H_0 | H_1\ is\ true)$.

There must be a tradeoff between these two types of error, so statisticians set up statistical tests in a way that balances these types of error, carefully mitigating risk while considering the type of task to be completed.

Table 1 depicts the types of statistical error associated with hypothesis tests and the relationships between the terms discussed. We can see that statistical power (1- β) is directly inversely proportional to Type II error (β).

	Reality (Unknown)	
Decision based on sample	***Groups are not different (H_0 true)***	***Groups are different (H_1 true)***
Groups are not different (Accept H_0)	Correct decision (1-α)	Type II error (β) *False negative*
Groups are different (Reject H_0)	Type I error (α) *False positive*	Correct decision (1-β)

Table 1. Statistical Error Associated with Hypothesis Tests

Figure 1 graphically depicts the relationship between the types of statistical error in a two sample test (Image source: Verhulst, 2016). The graph on the left-hand side displays an example of a distribution of a null and alternative hypothesis for a normal distribution, and the graph on the right-hand side displays an example of the null and alternative hypothesis of a chi square distribution. The black line indicates the critical value selected for the test, with the area shaded in red indicating Type I error and the area shaded in blue indicating Type II error. The non-shaded region represents a correct decision of, in this example, no effect to the left of the critical value and the presence of an effect to the right of the critical value.

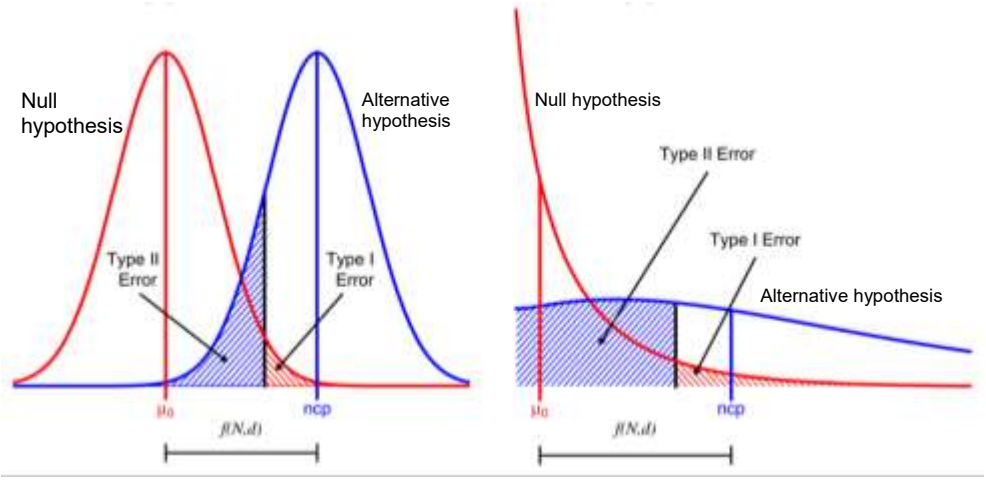


Figure 1. Graphical Depiction of Statistical Error and Power with the Normal Distribution (left) and Chi Square Distribution (right)

SIGNIFICANCE CRITERION

The next factor necessary for computing sample size in a study is the significance criterion. This is represented by α and is defined as $\Pr(reject\ H_0|H_0\ is\ true)$. It represents the probability of a "false positive" result, and has been described in the earlier section as Type I error. This value is another important assumption for calculating sample sizes. By convention, which may differ based on study design and field of study, this significance criterion is usually set at a value or 0.05 or less.

EFFECT SIZE

The next required factor for calculating sample size in a simple hypothesis test is the effect size, or the magnitude of the effect of interest in the population. The effect size encompasses both the absolute change in effect and the variability. It is important to specify an effect size that is meaningful for the question of interest. For clinical trials, effect size is quantified by a clinician and/or supported by literature outlining a clinically meaningful effect size. This could be the number of points of improvement on a test to truly make a difference in the patient's quality of life, or the improvement of a disease condition to a greater degree than existing treatments.

OTHER FACTORS THAT CAN INFLUENCE POWER

We have discussed the factors that always need to be specified in a sample size calculation. These are:

- Power (1-β): Pr(reject H_0 | H_1 true); correct rejection
- Significance criterion (α): Pr(reject H_0 | H_0 true); false positive
- Effect size: magnitude of the effect of interest in the population

Other factors that can influence power include the experimental design, precision, and expected rates of non-completion. There are many components of the experimental design that can influence the statistical power and consequently, the required sample size. Some examples of design factors that may influence statistical power include whether the number of observations in each sample group is balances or unbalanced, whether the hypothesis test is parametric or non-parametric, and whether the design of the study is crossover, parallel group, or factorial.

The next factor that can influence statistical power is the precision of the instrument used to measure the parameter of interest. For example, categorizing variables into groups, such as numeric values grouped into "low", "medium", and "high", results in reduced precision, a loss of information, and consequently a loss of power in the analysis. A reduction of measurement error improves statistical power, thus requiring a smaller sample size

Another factor influencing power is the expected rates of non-completion. In studies on human subjects, it cannot be expected that everyone enrolling in the study will complete the study. Therefore, the experiment needs to be designed to account for a reasonable amount of treatment withdrawals and protocol violations.

ADDITIONAL BACKGROUND INFORMATION FOR COMPUTING SAMPLE SIZE

The sample size for a study is typically calculated based on the primary hypothesis of interest. Because of this, secondary and exploratory analyses may be underpowered and should not be used to make claims but can influence design of future studies. This is an important distinction, because many studies seek to answer several questions. While this is permissible to include multiple endpoints, only adequately powered endpoints should be used to draw conclusions.

Generally, the sample size that is set at the beginning of the study is used as the guideline throughout the study. However, if pre-specified, sample size re-estimation can be performed while experiment is ongoing. This can be a useful technique to ensure the study is adequately powered if event rates are lower than anticipated or variability is larger than expected at the interim analysis time points (U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, & ICH, 1998).

APPROACHES FOR COMPUTING SAMPLE SIZE

COMPUTING SAMPLE SIZE BY HAND

Sample size can be calculated by hand using standard formulas when the underlying distribution is assumed to be approximately normal. Among the required inputs, z-scores for the assumed power level and significance criteria need to be included.

The z-score is derived based on the quantile of the standard normal distribution after the alpha (significant criteria) and beta (1 − power) terms are input. It equals the number of standard deviations away from the mean.

Given a quantile of a normal distribution, the z-score can be found by looking in a z-table or use the functions in SAS or in R.

The following function produces quantiles for the normal distribution under an assumed alpha level of 0.05 and beta level of 0.2:

```
DATA TEST;
  Q1=QUANTILE("Normal", 0.975);
  Q2=QUANTILE("Normal", 0.8);
```

```
RUN;
```

Output 1 shows the values of q1 and q2 assigned from the preceding data step.

q1	q2
1.95996	0.84162

Output 1. Output Quantile Assignments Using the QUANTILE Function

The following commands R equivalently compute the quantiles, and the output (in black) is immediately below the command (in blue):

```
> qnorm(0.975)
[1] 1.959964
> qnorm(0.8)
[1] 0.8416212
```

Example: 2 Sample T-Test, Equal Variances

The following formula can be used to determine the sample size required for each group in a 2 sample t-test using an approximation of the standard normal distribution.

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

Where:

- n is the sample size required for each group
- z_x is the critical value at the point on the standard normal distribution corresponding with the quantile in subscript
- σ is the standard deviation of the population
- Δ is the standardized difference between the 2 groups

This approximation is generally acceptable to use over the t distribution when the sample size is large (~>100). The values can be input into this formula and algebraically computed to obtain the sample size required for each group under the pre-specified conditions.

Example: 2 Sample Test of Proportions

The following formula can be used to determine the sample size required for each group in a 2 sample test of proportions.

$$n = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2[p_1(1-p_1) + p_2(1-p_2)]}{(p_1 - p_2)^2}$$

Where:

- n is the sample size required for each group
- z_x is the critical value at the point on the standard normal distribution corresponding with the quantile in subscript
- p_1 is the proportion of events expected to occur in group 1
- p_2 is the proportion of events expected to occur in group 2

The denominator, $(p_1\text{-}p_2)^2$, is the minimum meaningful difference or effect size.

COMPUTING SAMPLE SIZE USING SAS

There are two procedures available to compute sample size in SAS: PROC POWER and PROC GLMPOWER. The procedures are included in the SAS/STAT package, and have different capabilities that will be outlined in this section. Both procedures perform prospective power and sample size analyses. A prospective analysis is conducted when planning for a future study. Retrospective analysis, or power analysis of a study that has already taken place, is not supported by these procedures.

PROC POWER is used for sample size calculations for tests such as:

- t tests, equivalence tests, and confidence intervals for means,
- tests, equivalence tests, and confidence intervals for binomial proportions,
- multiple regression,
- tests of correlation and partial correlation,
- one-way analysis of variance,
- rank tests for comparing two survival curves,
- logistic regression with binary response, and
- Wilcoxon-Mann-Whitney (rank-sum) test (SAS, 2010).

PROC GLMPOWER is used for sample size calculations for more complex linear models, and cover Type III tests and contrasts of fixed effects in univariate linear models with or without covariates. (SAS, 2011).

Inputs: Comparison of PROC POWER and PROC GLMPOWER

Table 2Table 1 compares required inputs for PROC POWER and PROC GLMPOWER (SAS, 2010; SAS, 2011).

PROC POWER	PROC GLMPOWER
Design	Design (including subject profiles and their allocation weights)
Statistical model and test	Statistical model and contrasts of class effects
Significance level (alpha)	Significance level (alpha)
Surmised effects and variability	Surmised response means for subject profiles (i.e. "cell means") and variability
Power	Power
Sample size	Sample size

Table 2. Comparison of Inputs for Power Procedures in SAS

Not all of the inputs need to be filled out. Users should leave the result parameter (in this case, sample size) missing by designating it with a missing value in the input. If users are seeking to compute power with a predetermined sample size, the power field could be left missing if the sample size field is populated.

The POWER Procedure

The basic syntax of the POWER procedure is as follows:

```
PROC POWER <options> ;
  LOGISTIC <options> ;
  MULTREG <options> ;
  ONECORR <options> ;
  ONESAMPLEFREQ <options> ;
  ONESAMPLEMEANS <options> ;
  ONEWAYANOVA <options> ;
  PAIREDFREQ <options> ;
  PAIREDMEANS <options> ;
  PLOT <plot-options> </ graph-options> ;
  TWOSAMPLEFREQ <options> ;
  TWOSAMPLEMEANS <options> ;
  TWOSAMPLESURVIVAL <options> ;
  TWOSAMPLEWILCOXON <options> ;
RUN;
```

When using this procedure, users should specify at least one analysis statement and optionally, one or more PLOT statements. The analysis statements are all of the other statements in the procedure besides the PLOT statement. Within each analysis statement, there are different keywords used to specify the inputs. These keywords can be found in the SAS documentation and in the following examples. Each PLOT statement refers to the previous analysis statement and generates a separate graph or set of graphs.

Example: 2 Sample T-Test for Difference in Means

A two-sample t test assuming equal variances uses the following syntax:

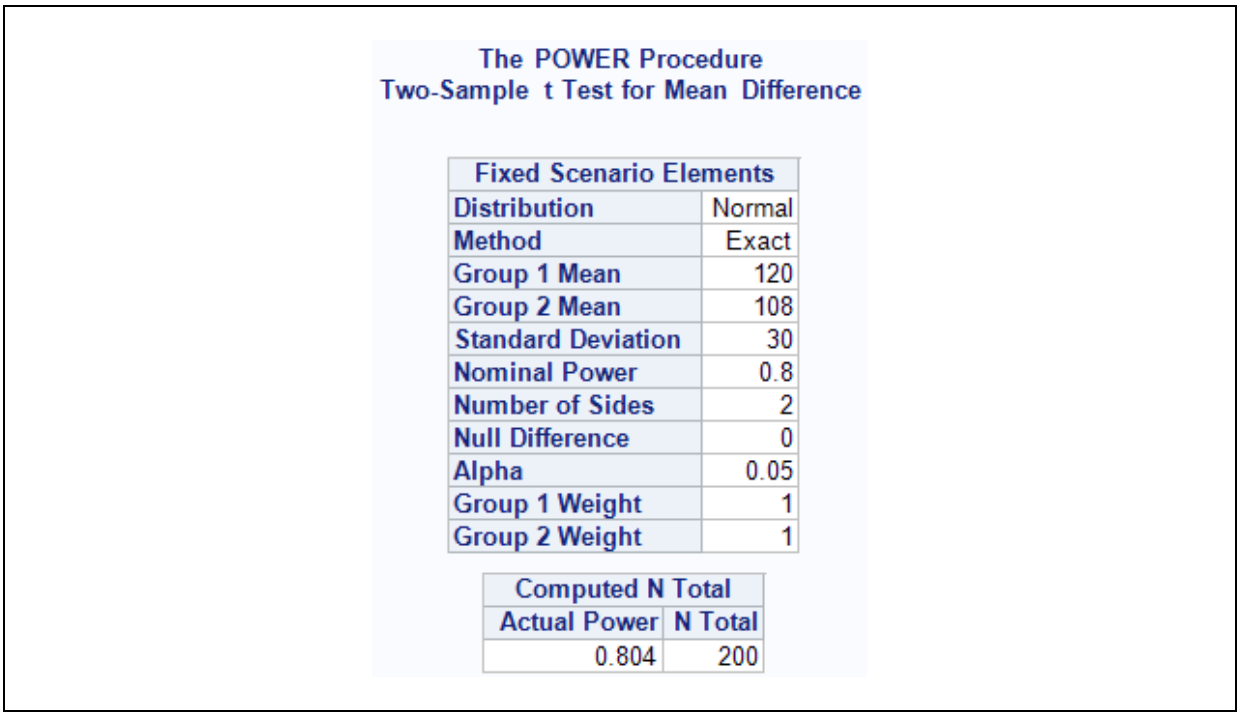
```
PROC POWER;
  TWOSAMPLEMEANS TEST=DIFF
  GROUPMEANS = mean1 | .
  STDDEV = .
  NTOTAL = .
  POWER = .
;
RUN;
```

Users can solve for any of the factors indicated as missing with a "." but all remaining factors need to be filled in. To calculate sample size, the NTOTAL field should be left missing with the other fields populated based on the underlying assumptions. Sample values have been input for illustrative purposes:

```
PROC POWER;
  TWOSAMPLEMEANS TEST=DIFF
  GROUPMEANS = 120 | 108
  STDDEV = 30
  NTOTAL = .
```

```
  POWER = 0.8
;
RUN;
```

Output 2 shows the SAS output from the PROC POWER statement with these sample values. We can see the informative display of each of the parameters as well as the computed N Total value of 200. This has been rounded up to the next highest integer, as the sample size needs to be a whole number.



The POWER Procedure
Two-Sample t Test for Mean Difference

Fixed Scenario Elements	
Distribution	Normal
Method	Exact
Group 1 Mean	120
Group 2 Mean	108
Standard Deviation	30
Nominal Power	0.8
Number of Sides	2
Null Difference	0
Alpha	0.05
Group 1 Weight	1
Group 2 Weight	1

Computed N Total	
Actual Power	N Total
0.804	200

Output 2. Output from the POWER Procedure Using an Example of a Two-Sample t Test for Mean Difference

When planning a study with limited resources, it is often advantageous to examine the effect of varying sample sizes on the statistical power. A useful plot can be produced by adding the following statement to the PROC POWER statement:

```
PLOT X=POWER MIN=0.8 MAX=0.95;
```

Figure 2 displays the output of this command, showing the total sample size required to attain to achieve a range of power values.

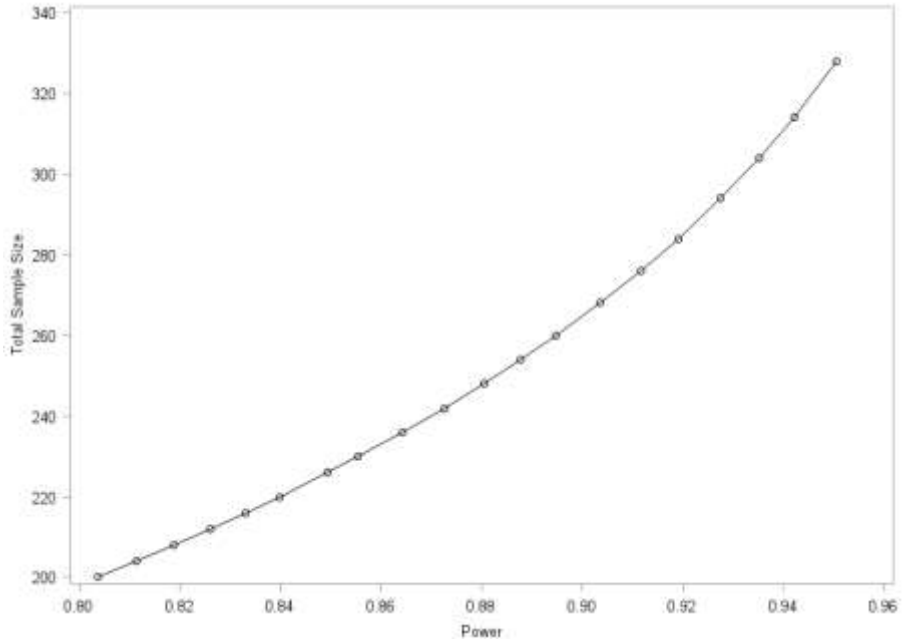


Figure 2. Required Sample Size for a Range of Power for a Two Sample t-Test Using PROC POWER

Example: Chi-Square Test for Difference in Proportions

The following code shows an example of how the POWER procedure can be used to compute a chi-square test for difference in proportions:

```
PROC POWER;
  TWOSAMPLEFREQ TEST=PCHI
  GROUPPROPORTIONS = 0.8 | 0.5
  POWER = 0.8
  NTOTAL = .
;
RUN;
```

Output 3 shows the SAS output from the PROC POWER statement for a Chi-Square Test for Difference in Proportions. The total sample size required for this example has been calculated to be 78.

The POWER Procedure
Pearson Chi-square Test for Proportion Difference

Fixed Scenario Elements	
Distribution	Asymptotic normal
Method	Normal approximation
Group 1 Proportion	0.8
Group 2 Proportion	0.5
Nominal Power	0.8
Number of Sides	2
Null Proportion Difference	0
Alpha	0.05
Group 1 Weight	1
Group 2 Weight	1

Computed N Total	
Actual Power	N Total
0.805	78

Output 3. Output from the POWER Procedure Using an Example of a Chi-Square Test for Proportion Difference

The GLMPOWER Procedure

The basic syntax of the GLMPOWER procedure is as follows:

```
PROC GLMPOWER <options> ;
  BY variables ;
  CLASS variables ;
  CONTRAST 'label' effect values <...effect values> </ options> ;
  MODEL dependents = independents ;
  PLOT <plot-options> </ graph-options> ;
  POWER <options> ;
  WEIGHT variable ;
RUN;
```

This procedure supports more complicated study designs, with additional options allowing for specification of contrasts and addition of covariates. To add a covariate, specify the ncovariates=1 option in the POWER statement and specify CORRXY, or the correlation between the covariate and response variable. Another useful customization in this procedure is the WEIGHT statement, which can be used for studies with an unbalanced design.

Example: Two-Way ANOVA Test

Sample size can be computed for a 2-way ANOVA using PROC GLMPOWER with the following syntax:

```
PROC GLMPOWER DATA = dataset;
  CLASS expvar1 expvar2;
  MODEL responsevar = expvar1 | expvar2;
  POWER
```

```
    STDDEV = .
    NTOTAL = .
    POWER = .
;
RUN;
```

The procedure allows users to solve for any of the factors indicated as missing with a ".", but all remaining factors need to be filled in. To calculate sample size, NTOTAL should be left missing.

The first step is to create an exemplary data set with expected population means. In this example, these are lab values at each level of treatment and dose:

```
DATA Exemplary;
  DO trt = 1 to 2;
    DO dose = 1 to 3;
      INPUT lab @@;
      OUTPUT;
    END;
  END;
  DATALINES;
    14 16 21
    10 15 16
  ;
RUN;
```

Output 4 shows the SAS output from the DATA step creating an Exemplary dataset for lab values at several treatment group and dose values.

trt	dose	lab
1	1	14
1	2	16
1	3	21
2	1	10
2	2	15
2	3	16

Output 4. Exemplary Dataset Created to Input in the GLMPOWER Procedure

The next step is to input the exemplary dataset into the GLMPOWER procedure. In this example, standard deviation is assumed to be common in both groups:

```
PROC GLMPOWER DATA = Exemplary;
  CLASS trt dose;
  MODEL lab = trt | dose;
  POWER
    STDDEV = 5
    NTOTAL = .
```

```
    POWER = 0.8
RUN;
```

Output 5 shows the SAS output from the PROC GLMPOWER statement for a Two-Way ANOVA Test.

Fixed Scenario Elements	
Dependent Variable	lab
Error Standard Deviation	5
Nominal Power	0.8
Alpha	0.05

Computed N Total					
Index	Source	Test DF	Error DF	Actual Power	N Total
1	trt	1	72	0.828	78
2	dose	2	36	0.849	42
3	trt*dose	2	336	0.807	342

Output 5. Output from the GLMPOWER Procedure Using an Example of a Two-Way ANOVA Test

Similar to the feature in PROC POWER, users can add the following statement to the PROC GLMPOWER command to produce a plot:

```
PLOT X=POWER MIN=0.1 MAX=0.9;
```

Figure 3 identifies the total sample size to attain to achieve a range of power for the two-way ANOVA test.

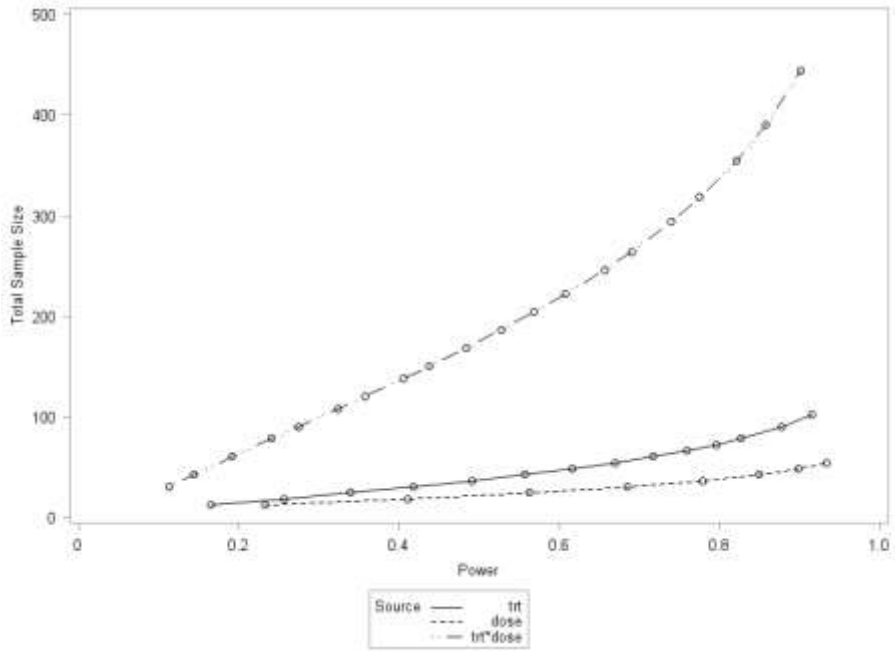


Figure 3. Required Sample Size for a Range of Power for a Two-Way ANOVA Test Using PROC GLMPOWER

COMPUTING SAMPLE SIZE USING R

R has several functions available for computing sample sizes. These functions are contained in the "pwr" package, which needs to be downloaded before attempting to run any of these functions.

Another input parameter that needs to be calculated ahead of time before using the functions is d, which is defined as follows:

$$d = \frac{|\mu_1 - \mu_2|}{\sigma}$$

Where μ_1 = mean of group 1,

μ_1 = mean of group 2, and

σ^2 = common error variance.

Once these steps are completed, users can proceed with inputting values into the R function associated with the appropriate statistical test.

Example: 2 Sample T-Test for Difference in Means

The syntax of a sample size calculation for a 2 sample t-test in R is:

```
pwr.t.test(n = , d = , sig.level = , power = , type = c("two.sample",
"one.sample", "paired"))
```

This example inputs the same values as in the previous example where we used PROC POWER in SAS to conduct the sample size for a 2 sample t-test. Similar to SAS, we can leave the field we want to calculate as blank. In this case, we leave "n=" as blank to compute the sample size:

```
pwr.t.test(n =, d = 0.4, sig.level = 0.05, power = 0.8, type="two.sample")
```

Output 6Output 5 shows the R output from the pwr.t.test function for a 2 sample t-test. Note that the output in R is slightly different, as the output has not been rounded up to the nearest whole number. The output also displays the required number of subjects for each group rather than overall, as it did in the SAS output.

```
     Two-sample t test power calculation

              n = 99.08032
              d = 0.4
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: n is number in *each* group
```

Output 6. Output from the pwr.t.test Function in R Using an Example of 2 Sample T-Test

We can produce a plot in R similar to the plot produced in SAS. This can be done by submitting the following statements, where we assign the power output to an object in R and plot the object:

```
> x <- pwr.t.test(n = , d = 0.4, sig.level = 0.05, power = 0.8,
type="two.sample")
> plot(x)
```

Figure 4 identifies the total sample size to attain to achieve a range of power for the example of the 2 sample t-test.

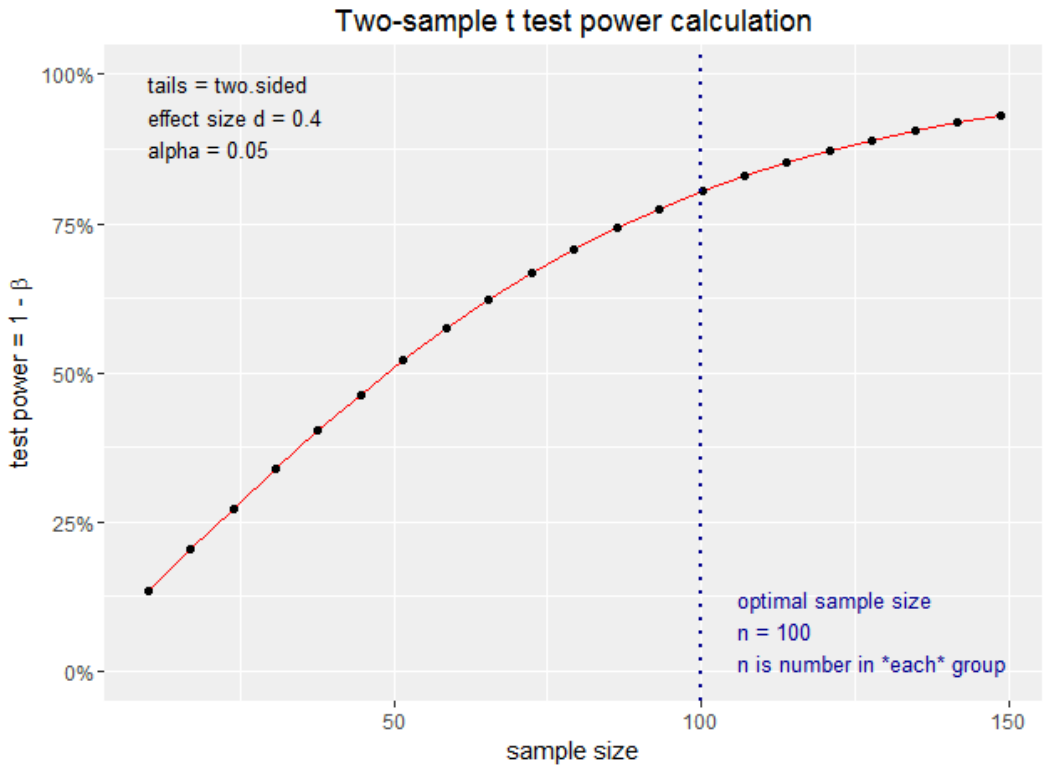


Figure 4. Required Sample Size for a Range of Power for a Two Sample t-Test Using R

Syntax of Functions for Other Tests

Sample sizes can be computed for other statistical tests using different functions. Some examples are listed below:

- t test with unequal sample sizes:

  `pwr.t2n.test(n1 =, n2 = , d = , sig.level = , power = )`

- One-way ANOVA:

  `pwr.anova.test(k = , n =, f = , sig.level = , power = )`

$$f = \sqrt{\frac{\sum_{i=1}^{k} p_i * (\mu_i - \mu)^2}{\sigma^2}}$$

Where $p_i = n_i/N$,

n_i = number of observations in group I,

N = total number of observations,

μ_i = mean of group i,

μ = grand mean, and

σ^2 = error variance within groups.

- Chi-square test:

  `pwr.chisq.test(w = , N = , df = , sig.level = , power = )`

$$w = \sqrt{\sum_{i=1}^{m} \frac{(p0_i - p1_i)^2}{p0_i}}$$

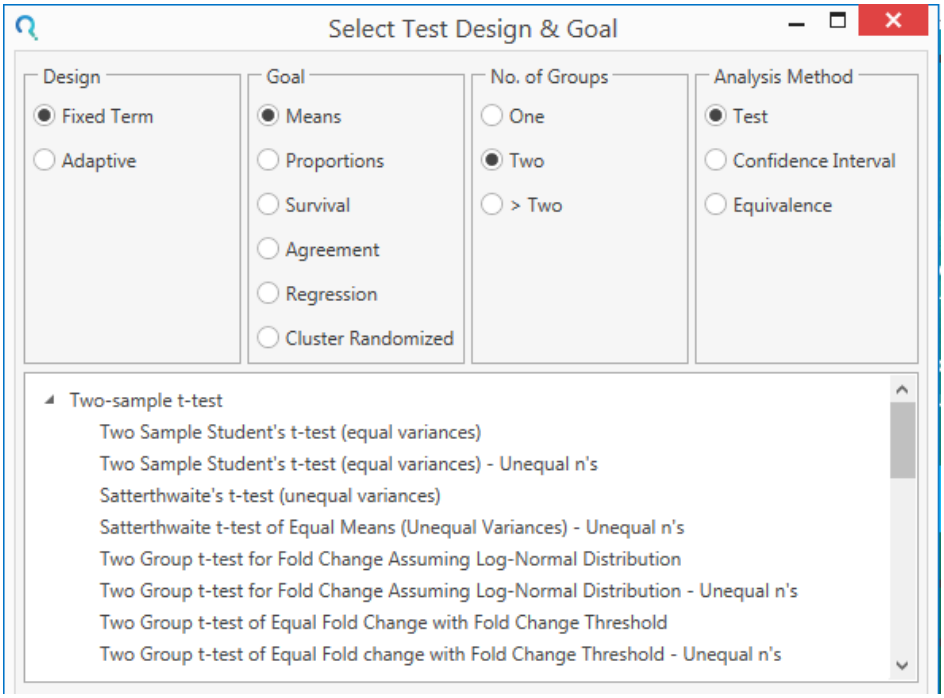
Where $p0_i$ = cell probability in the i^{th} cell under H_0, and

$p1_i$ = cell probability in the i^{th} cell under H_1.

- Other designs include linear models (`pwr.f2.test`), correlations (`pwr.r.test`), test of proportions (`pwr.2p.test`/ `pwr.2p2n.test`/ `pwr.p.test`)

COMPUTING SAMPLE SIZE USING NQUERY

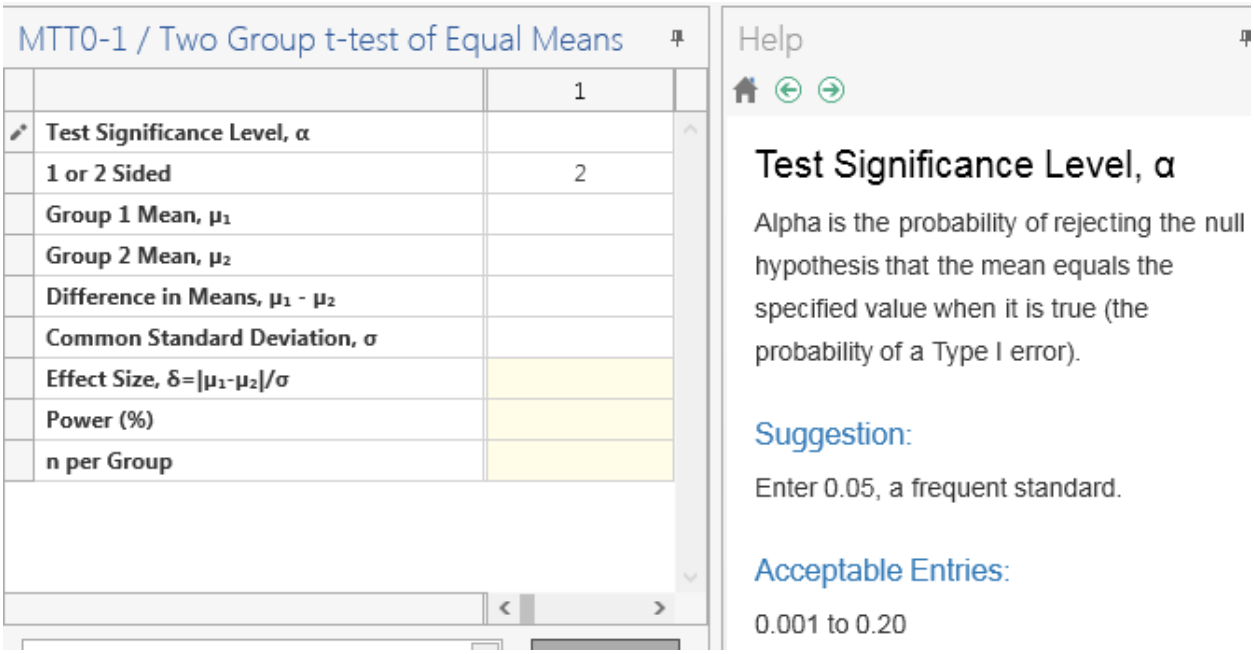
Lastly, we'll demonstrate how to compute sample size using nQuery. Display 1 shows the nQuery Wizard interface, which recommends a statistical test based on the study design and goals.



Display 1. nQuery Wizard Interface

Example: 2 Sample T-Test for Difference in Means

Proceeding with our example computing the sample size for a two sample t-test, we can see in Display 2 the user interface for nQuery. In this second step, the user fills in known information and the software defines and suggests values, as in the example below.



Display 2. nQuery Wizard Interface: Enter Background Information

A useful feature of nQuery is that it automatically fills in fields once enough information is entered. For example, the Difference in means value is populated after Group 1 and Group 2 are filled out, and the Effect size is populated after Difference in means and σ are filled out.

As we've seen in the other software demonstrations, the field of interest should be left blank. Once enough information is filled out in the other fields, the result for the blank field will be shown in the output, which is automatically generated.

Display 3 shows the menu bar in nQuery, where the users can click on the graph icon on the right-hand side of the menu to produce a plot that identifies the total sample size to attain to achieve a range of power.

Display 3. nQuery Menu Bar, Indicating the Location of the Graph Icon

Figure 5 displays the result of the graph command in nQuery, producing a plot of sample size against power.

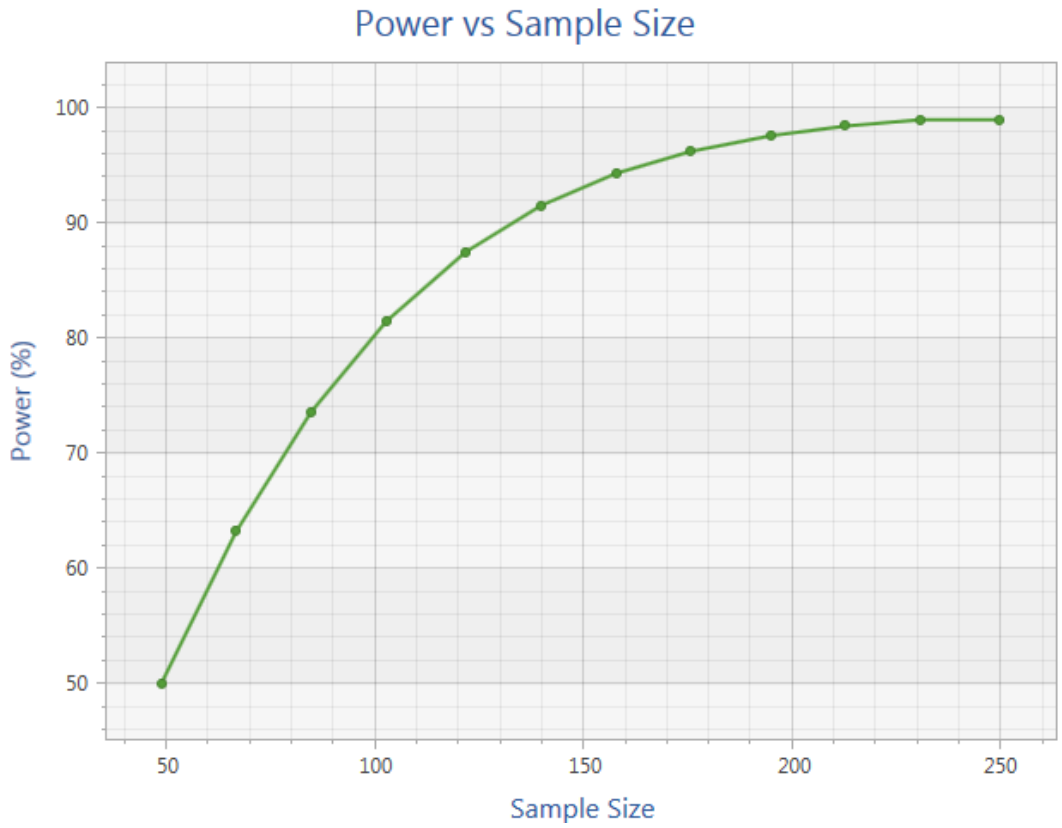


Figure 5. Required Sample Size for a Range of Power for a Two Sample t-Test Using nQuery

CONCLUSION

All of the software packages described in this paper are useful for computing sample sizes, and can accomplish many of the same tasks. Users should determine the software package based on familiarity with the software and their anticipated needs. The advantages of each software package individually as well as the shared advantages are outlined below.

SAS:

- Ability to calculate sample size for complex linear models and contrasts
- Blend of user friendly features and advanced options
- Requires SAS/STAT package

SAS and R

- Can quickly and easily test a range of values
- Plots are easily customizable
- Sample size can be computed in a program so it is easily replicable and "macrotized"

R

- Requires more extensive computations by the user for input parameters
- Plots are most informative

- Requires "pwr" package
- Free and open source

R and nQuery

- Limited in their ability to compute sample size for very complicated models
- Choose test first and then enter inputs, rather than customizing inputs to influence test

nQuery

- Wizard interface, so no programming required
- Explanations of each input parameter and plain text description of output
- Great for non-programmers

nQuery and SAS

- No extensive computations required by the user
- User-friendly
- Capabilities for many tests
- Not free, but documentation is comprehensive

REFERENCES

Kabacoff, Robert (2017). Power Analysis. *Quick R: by DataCamp.* Retrieved from https://www.statmethods.net/stats/power.html

SAS (2010). Overview: POWER Procedure. *SAS/STAT® 9.22 User's Guide.* Retrieved from http://support.sas.com/documentation/cdl/en/statug/63347/HTML/default/viewer.htm#statug_power_a0000000967.htm

SAS (2011). The GLMPOWER Procedure. *SAS/STAT® 9.3 User's Guide*. Retrieved from https://support.sas.com/documentation/cdl/en/statug/63962/HTML/default/viewer.htm#statug_glmpower_a0000000158.htm

UCLA: Statistical Consulting Group. *Introduction to SAS Power and Sample Size Analysis.* Retrieved from https://stats.idre.ucla.edu/sas/seminars/proc-power/

U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, & ICH. (1998). E9 Statistical Principles for Clinical Trials. Retrieved from https://www.fda.gov/regulatory-information/search-fda-guidance-documents/e9-statistical-principles-clinical-trials

Verhulst, B. (2016). A Power Calculator for the Classical Twin Design. *Behavior Genetics*, *47*(2), 255–261. doi: 10.1007/s10519-016-9828-9

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Jenna Cody
Johnson & Johnson
JCody1@its.jnj.com
Linkedin.com/in/JennaCody