Chapter 402

Conditional Power and Sample Size Reestimation of One-Sample T-Tests

Introduction

In sequential designs, one or more intermediate analyses of the emerging data are conducted to evaluate whether the experiment should be continued. This may be done to conserve resources or to allow a data monitoring board to evaluate safety and efficacy when subjects are entered in a staggered fashion over a long period of time. *Conditional power* (a frequentist concept) is the probability that the final result will be significant, given the data obtained up to the time of the interim look. *Predictive power* (a Bayesian concept) is the result of averaging the conditional power over the posterior distribution of effect size. Both of these methods fall under the heading of *stochastic curtailment* techniques. Further reading about the theory of these methods can be found in Jennison and Turnbull (2000), Chow and Chang (2007), Chang (2008), Proschan et.al (2006), and Dmitrienko et.al (2005).

This program module computes conditional and predicted power for the case when a one-sample *t*-test is used to test whether the mean of a population is greater than, less than, or not equal to a specific value. It also provides *sample size reestimation* to achieve a specified conditional power value.

Technical Details

All details and assumptions usually made when using a one-sample *t*-test continue to be in force here.

Conditional Power

The power of an experiment indicates whether a study is likely to result in useful results, given the sample size. Low power means that the study is *futile*: little chance of statistical significance even though the alternative hypothesis is true. A study that is futile should not be started. However, futility may be determined only after the study has started. When this happens, the study is *curtailed*.

The futility of a study that is underway can be determined by calculating its *conditional power*: the probability of statistical significance at the completion of the study given the data obtained so far.

It is important to note that conditional power at the beginning of the study before any data are collected is equal to the unconditional power. So, conditional power will be high even if early results are negative. Hence, conditional power will seldom result in study curtailment very early in the study.

From Jennison and Turnbull (2000) pages 205 to 208, the general upper one-sided conditional power at stage *k* for rejecting a null hypothesis about a parameter θ at the end of the study, given the observed test statistic, Z_k, is computed as

$$P_{uk}(\theta) = \Phi\left(\frac{Z_k\sqrt{I_k} - z_{1-\alpha}\sqrt{I_K} + \theta(I_K - I_k)}{\sqrt{I_K - I_k}}\right),$$

the general lower one-sided conditional power at stage *k* is computed as

$$P_{lk}(\theta) = \Phi\left(\frac{-Z_k\sqrt{I_k} - z_{1-\alpha}\sqrt{I_K} - \theta(I_K - I_k)}{\sqrt{I_K - I_k}}\right),$$

and the general two-sided conditional power at stage *k* is computed as

$$P_k(\theta) = \Phi\left(\frac{Z_k\sqrt{I_k} - z_{1-\alpha/2}\sqrt{I_K} + \theta(I_K - I_k)}{\sqrt{I_K - I_k}}\right) + \Phi\left(\frac{-Z_k\sqrt{I_k} - z_{1-\alpha/2}\sqrt{I_K} - \theta(I_K - I_k)}{\sqrt{I_K - I_k}}\right),$$

where

θ = the parameter being tested by the hypothesis

k = an interim stage at which the conditional power is computed (k = 1, …, K – 1)

K = the stage at which the study is terminated and the final test computed

Z_k = the test statistic calculated from the observed data that has been collected up to stage k

I_k = the information level at stage k

I_K = the information level at the end of the study

$z_{1-\alpha}$ = the standard normal value for the test with a type I error rate of α.

For the test of a single mean with null hypothesis H_0: $\mu_1 = \mu_0$, where μ_1 is the alternative mean at which the power is calculated and μ_0 is the hypothesized mean under the null hypothesis, these components are computed in Chang (2008) page 69 as

$\theta = \mu_1 - \mu_0$ (the expected difference under the alternative hypothesis)

$Z_k = (\bar{x}_k - \mu_0)\sqrt{\hat{I}_k}$ (the *t*-statistic computed from the observed data)

$I_k = \dfrac{n_k}{\sigma^2}$ (the interim information level)

$I_K = \dfrac{N}{\sigma^2}$ (the final information level) where

$\bar{x}_k$ is the sample mean, estimating μ_1 at stage k

$\hat{I}_k$ is the estimated information from the sample at stage k

n_k is the sample size at stage k

N is the total sample size

σ^2 is the variance of X (estimated by the sample variance).

Computing conditional power requires you to set σ^2, μ_0, and μ_1. Their values can come from the values used during the planning of the study, from similar studies, or from estimates made from the data that has emerged.

Futility Index

The *futility index* is $1 - P_k(\theta)|H_1$. The study may be stopped if this index is above 0.8 or 0.9 (that is, if conditional power falls below 0.2 or 0.1).

Predictive Power

Predictive power (a Bayesian concept) is the result of averaging the conditional power over the posterior distribution of effect size. From Jennison and Turnbull (2000) pages 210 to 213, the general upper one-sided predictive power at stage k is given by

$$P_{uk} = \Phi\left(\frac{Z_k\sqrt{I_K} - z_{1-\alpha}\sqrt{I_k}}{\sqrt{I_K - I_k}}\right).$$

the general lower one-sided predictive power at stage k is given by

$$P_{lk} = \Phi\left(\frac{-Z_k\sqrt{I_K} - z_{1-\alpha}\sqrt{I_k}}{\sqrt{I_K - I_k}}\right),$$

the general two-sided predictive power at stage k is given by

$$P_k = \Phi\left(\frac{|Z_k|\sqrt{I_K} - z_{1-\alpha/2}\sqrt{I_k}}{\sqrt{I_K - I_k}}\right) + \Phi\left(\frac{-|Z_k|\sqrt{I_K} - z_{1-\alpha/2}\sqrt{I_k}}{\sqrt{I_K - I_k}}\right),$$

with all terms defined as in the equations for conditional power.

Sample Size Reestimation

As Chang (2014) points out, after an interim analysis, it is often desirable to recalculate the target sample size using updated values for various nuisance parameters such as the variance. This process is known as *sample size reestimation*.

One method of calculating an adjusted sample size estimate is to search for the sample size that results in a predetermined value of conditional power. PASS conducts a binary search using the conditional power as the criterion. The result is called the *target sample size*.

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Design tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Design Tab

The Design tab contains most of the parameters and options that you will be concerned with.

Solve For

Solve For

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are *Conditional Power* or *Sample Size Reestimation*.

Note that the value selected here always appears as the vertical axis on the charts.

The program is set up to calculate conditional power directly. To find appropriate values of sample size, a binary search is made using an iterative procedure until an appropriate value is found.

Test

Alternative Hypothesis

Specify the alternative hypothesis of the test. Since the null hypothesis is the opposite, specifying the alternative is all that is needed.

When you choose a one-sided test option, you must be sure that the value(s) of μ1 and μ0 match your choice. For example, if you select H1: $\mu < \mu0$, the values of μ1 should be less than μ0.

Conditional Power and Alpha

Conditional Power

Enter one or more values for the conditional power that is to be attained by the sample size search.

Conditional power is the probability of rejecting a false null hypothesis given the data that has been obtained so far. It is a function of the effect size and the data that have been collected.

The valid range is between 0 to 1. Typically, values over 0.70 are considered high enough to continue a study, while values below 0.10 considered grounds for curtailing the study.

You can enter a single value such as 0.80 or a series of values such as 0.80 0.85 0.90 0.95 or 0.7 to 0.97 by 0.05.

Alpha

This option specifies one or more values for the probability of a type-I error at the end of the study. A type-I error occurs when a true null hypothesis is rejected.

Values must be between zero and one. Historically, the value of 0.05 has been used for two-sided tests and 0.025 for one-sided tests.

You may enter a range of values such as *0.01 0.05 0.10* or *0.01 to 0.10 by 0.01*.

Sample Size

N (Target Sample Size)

This option specifies one or more values of the target sample size, the total number of subjects planned for the study. This value must be an integer greater than one. Note that you may enter a list of values using the syntax *50,100,150,200,250* or *50 to 250 by 50*.

nk (Sample Size to Look k)

Enter the sample size obtained through look k. If this value is greater than N, the value of N is increased to this amount.

Effect Size – Means

μ0 (Null or Baseline Mean)

Enter a value for μ0, the mean assumed by the null hypothesis. Note that the difference (δ1 = μ1 - μ0) between this and the alternative mean, μ1, is the value that is actually tested. Usually, this is the mean difference upon which the total sample size of the study was based.

You may enter a range of values such as *0 0.5 1* or *0 to 3 by 0.5*.

μ1 (Actual Mean to Detect)

Enter a value for μ1, the mean assumed by the alternative hypothesis at which conditional power is calculated. Note that the difference (δ1 = μ1 - μ0) between this and the null mean, μ0, is the amount that is actually tested.

You may enter a range of values such as *0 0.5 1* or *0 to 3 by 0.5*.

Effect Size – Standard Deviation

σ (Standard Deviation)

Enter a value for σ, the standard deviation. The source of the value you enter is controversial. Some think you should enter the value used in planning the study. Others think you should use the value estimated from the data obtained so far. Still others think you should use a confidence limit for σ created from the current sample.

You can enter a range of values such as *1 2 3* or *1 to 5 by 1*.

Press the σ button to the right to load the Standard Deviation Estimator window.

Effect Size – Current Test Statistic

Zk (Current Test Statistic)

Enter the value of the *t*-statistic calculated from the data obtained through stage *k*. This value may be positive or negative. Usually, the *t*-statistic ranges between -5 and 5.

Example 1 – Computing Conditional Power

Suppose a study has been planned to detect a mean change of 1 at an alpha of 0.05 using a one-sided *t*-test. The sample size is 50. The standard deviation is expected to be about 1.8. An interim analysis is run after half the data have been collected. This analysis yields a *t*-test value of 2.12. The data monitoring board would like to have the conditional power calculated for actual mean values of 0, 0.5, 1, and 1.5.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the procedure window. You may make the appropriate entries as listed below or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Design Tab	
Solve For	**Conditional Power**
Alternative Hypothesis	**H1: μ > μ0 (One-Sided)**
Alpha	**0.05**
N (Target Sample Size)	**50**
nk (Sample Size at Look k)	**25**
μ0 (Null or Baseline Mean)	**0**
μ1 (Actual Mean to Detect)	**0 0.5 1 1.5**
σ (Standard Deviation)	**1.8**
Zk (Current Test Statistic)	**2.12**

Annotated Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results for One-Sample T-Test
Hypotheses: H0: μ ≤ μ0 vs. H1: μ > μ0

Cond. Power	Pred. Power	Target Sample Size N	Current Sample Size nk	Null Mean μ0	Actual Mean μ1	Diff δ1	Std Dev σ	Test Statistic Zk	Alpha	Futility
0.41833	0.91202	50	25	0.00	0.00	0.00	1.80	2.120	0.050	0.58167
0.88154	0.91202	50	25	0.00	0.50	0.50	1.80	2.120	0.050	0.11846
0.99494	0.91202	50	25	0.00	1.00	1.00	1.80	2.120	0.050	0.00506
0.99996	0.91202	50	25	0.00	1.50	1.50	1.80	2.120	0.050	0.00004

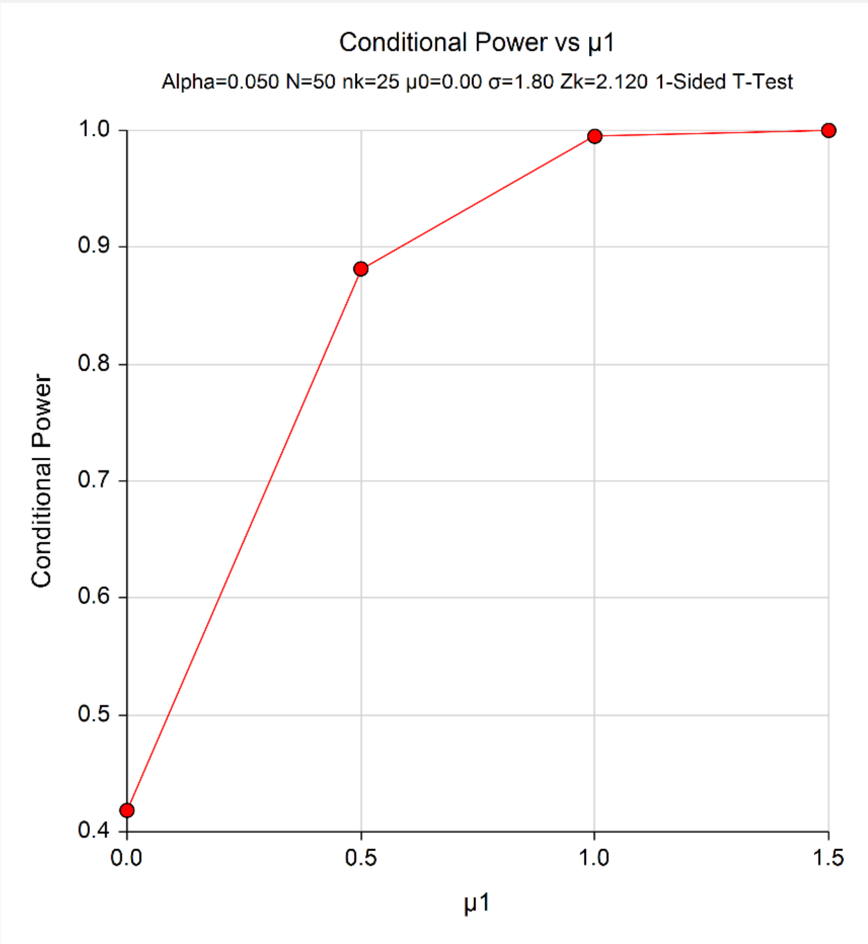
References
Jennison, C., and Turnbull, B.W. 2000. Group Sequential Methods with Applications to Clinical Trials. Chapman & Hall/CRC. New York.
Proschan, M., Lan, K.K.G., Wittes, J.T. 2006. Statistical Monitoring of Clinical Trials. Springer. New York.
Chang, Mark. 2008. Classical and Adaptive Clinical Trial Designs. John Wiley & Sons. Hoboken, New Jersey.
Chang, Mark. 2014. Adaptive Design Theory and Implementation Using SAS and R. CRC Press. New York.

Report Definitions
Conditional Power is the probability of rejecting a false null hypothesis at the end of the study given the data that have emerged so far.
Predictive Power is the result of averaging the conditional power over the posterior distribution of the effect size.
N is the target sample size.
nk is the sample size obtained through stage k.
μ0 is the mean assuming the null hypothesis.
μ1 is the actual mean to detect under the alternative hypothesis at which conditional power is calculated.
δ1 = μ1 - μ0 is the mean difference that is to be detected by the completed study.
σ is the standard deviation of the response.
Zk is the value of the test statistic from the observed data at stage k.
Alpha is the probability of rejecting a true null hypothesis.
Futility is one minus the conditional power. A value greater than 0.9 or 0.8 indicates the study should be stopped because there is little chance of achieving statistical significance.

Summary Statements
The first 25 of the planned 50 subjects achieve 41.833% conditional power to detect a difference of δ1 = μ1 - μ0 = 0.00 between the alternative mean 0.00 and the null mean 0.00 using a one-sided one-sample t-test with a standard deviation of 1.80 and a significance level of 0.050. The t-value of the data that have emerged so far is 2.120. The futility index is 0.58167.

This report shows the values of each of the parameters, one scenario per row. The definitions of each column are given in the Report Definitions section.

Plots Section



This plot shows the relationship between conditional power and $\mu 1$.

Example 2 – Validation

We could not find an example of a conditional power calculation for a one-sample *t*-test in the literature. Since the calculations are relatively simple, we will validate the calculation of the third scenario of Example 1 by hand.

In this case

$$I_k = n_k / \sigma^2 = \frac{25}{1.8^2} = 7.7160494 \qquad I_K = N / \sigma^2 = \frac{50}{1.8^2} = 15.432099$$

$$\begin{aligned}
P_{uk}(\theta) &= \Phi\left(\frac{Z_k\sqrt{I_k} - z_{1-\alpha}\sqrt{I_K} + (\mu_1 - \mu_0)(I_K - I_k)}{\sqrt{I_K - I_k}}\right) \\
&= \Phi\left(\frac{2.12\sqrt{7.7160494} - (1.6448536)\sqrt{15.432099} + (1-0)(15.432099 - 7.7160494)}{\sqrt{15.432099 - 7.7160494}}\right) \\
&= \Phi\left(\frac{7.1433431}{2.7777777}\right) \\
&= \Phi(2.5716035) \\
&= 0.9949386
\end{aligned}$$

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the procedure window. You may make the appropriate entries as listed below or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Design Tab	
Solve For	**Conditional Power**
Alternative Hypothesis	**H1: μ > μ0 (One-Sided)**
Alpha	**0.05**
N (Target Sample Size)	**50**
nk (Sample Size at Look k)	**25**
μ0 (Null or Baseline Mean)	**0**
μ1 (Actual Mean to Detect)	**1**
σ (Standard Deviation)	**1.8**
Zk (Current Test Statistic)	**2.12**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results for One-Sample T-Test
Hypotheses: H0: μ ≤ μ0 vs. H1: μ > μ0

Cond. Power	Pred. Power	Target Sample Size N	Current Sample Size nk	Null Mean μ0	Actual Mean μ1	Diff δ1	Std Dev σ	Test Statistic Zk	Alpha	Futility
0.99494	0.91202	50	25	0.00	1.00	1.00	1.80	2.120	0.050	0.00506

The conditional power of 0.99494 matches the value calculated by hand.

Example 3 – Sample Size Reestimation

Suppose a study has started. It is being conducted to detect a mean change of 1.0 at an alpha of 0.025 using a one-sided *t*-test. The target sample size is 50. The standard deviation is expected to be about 1.8.

An interim analysis was run after half the data were collected. This analysis yields a z-test value of 2.12. This test value was the result of transforming the p-value of the t-test with the inverse, standard normal distribution. The sample standard deviation was 4.28.

The data monitoring board would like to have sample size reestimated to obtain a conditional power of 0.8.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the procedure window. You may make the appropriate entries as listed below or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Design Tab	
Solve For	**Sample Size Reestimation**
Alternative Hypothesis	**H1: μ > μ0 (One-Sided)**
Conditional Power	**0.8**
Alpha	**0.05**
nk (Sample Size at Look k)	**25**
μ0 (Null or Baseline Mean)	**0**
μ1 (Actual Mean to Detect)	**1**
σ (Standard Deviation)	**4.28**
Zk (Current Test Statistic)	**2.12**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results for One-Sample T-Test
Hypotheses: H0: μ ≤ μ0 vs. H1: μ > μ0

Cond. Power	Pred. Power	Target Sample Size N	Current Sample Size nk	Null Mean μ0	Actual Mean μ1	Diff δ1	Std Dev σ	Test Statistic Zk	Alpha	Futility
0.80107	0.89583	85	25	0.00	1.00	1.00	4.28	2.120	0.025	0.19893

Notice that the target sample size has increased from 50 to 85.