

Chapter 433

Conditional Power of Two-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 two-sample *t*-test is used to test whether the means of two populations are different.

Technical Details

All details and assumptions usually made when using a two-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

Conditional Power of Two-Sample T-Tests

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

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 ($k = 1, \dots, K - 1$)

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 a test of a two means with null hypothesis $H_0: \mu_1 = \mu_2$, where μ_1 and μ_2 are the population means in groups 1 and 2, respectively, under the alternative hypothesis, these components are computed in Chang (2008) page 70 as

$\theta = \mu_2 - \mu_1$ (the expected difference under the alternative hypothesis)

$Z_k = (\bar{x}_{2k} - \bar{x}_{1k})\sqrt{I_k}$ (the t statistic computed from the observed data)

$I_k = \left(\frac{\sigma_1^2}{n_{1k}} + \frac{\sigma_2^2}{n_{2k}}\right)^{-1}$ (the interim information level)

$I_K = \left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)^{-1}$ (the final information level)

where

\bar{x}_{jk} is the sample mean for group j , estimating μ_j at stage k

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

n_{jk} is the sample size in group j at stage k

n_j is the final sample size in group j

σ_j^2 is the variance of group j

Computing conditional power requires you to set μ_1 , μ_2 , σ_1 , and σ_2 . 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.

Conditional Power of Two-Sample T-Tests

Futility Index

The *futility index* is $1 - P_k(\theta) | H_a$. The study may be stopped if this index is above 0.80 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.

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.

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 δ match your choice. For example, if you select $H_a: \delta < 0$, the value of δ should be less than zero.

Conditional Power of Two-Sample T-Tests

Alpha

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

N1 (Target Sample Size Group 1)

Enter one or more values of the target sample size of group 1, the total number of subjects planned for this group. 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*.

N2 (Target Sample Size Group 2)

Enter a value (or range of values) for the target sample size of group 2 or enter *Use R* to calculate *N2* from *N1*. You may enter a range of values such as *10 to 100 by 10*.

- **Use R**

If *Use R* is entered here, *N2* is calculated using the formula

$$N2 = [R(N1)]$$

where *R* is the Sample Allocation Ratio and the operator $[Y]$ is the first integer greater than or equal to *Y*. For example, if you want $N1 = N2$, select *Use R* and set $R = 1$.

R (Sample Allocation Ratio)

Enter a value (or range of values) for *R*, the allocation ratio between samples. This value is only used when *N2* is set to *Use R*.

When used, *N2* is calculated from *N1* using the formula: $N2 = [R(N1)]$ where $[Y]$ is the next integer greater than or equal to *Y*. Note that setting $R = 1.0$ forces $N2 = N1$.

n1k (Group 1 Sample Size at k)

Enter the group 1 sample size obtained through look k. If this value is greater than *N1*, the value of *N1* is increased to this amount.

n2k (Group 2 Sample Size at k)

Enter the group 2 sample size obtained through look k. If this value is greater than *N2*, the value of *N2* is increased to this amount.

- **n1k**

If *n1k* is entered here, *n2k* is set equal to *n1k*.

Effect Size – Means

μ_1 (Mean of Group 1)

Enter a value for μ_1 , the mean of group 1. The power and sample size of the study are based on the difference between this mean and the mean of group 2 which is called δ ($\mu_2 - \mu_1$). The null hypothesis assumes that $\mu_1 = \mu_2$.

Conditional Power of Two-Sample T-Tests

The alternative hypothesis assumes that these means are different. This difference, δ , is the minimum detectable difference.

Usually, the values μ_1 and μ_2 that were used to plan the study are used here.

You may enter a range of values such as '10 20 30' or '0 to 100 by 25'.

μ_2 (Mean of Group 2)

Enter a value for μ_2 , the mean of group 2. The power and sample size of the study are based on the difference between this mean and the mean of group 1. This difference is called δ ($\mu_2 - \mu_1$). The null hypothesis assumes that $\mu_1 = \mu_2$. The alternative hypothesis assumes that these means are different. This difference, δ , is the minimum detectable difference.

Usually, the values of μ_1 and μ_2 that were used to plan the study are used here.

You may enter a range of values such as '10 20 30' or '0 to 100 by 25'.

Effect Size – Standard Deviation

σ_1 (Std Deviation Group 1)

Enter a value for σ_1 , the standard deviation of group 1.

This value 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_1 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.

σ_2 (Std Deviation Group 2)

Enter a value for σ_2 , the standard deviation of group 2.

This value 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_1 created from the current sample.

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

Enter ' σ_1 ' to set $\sigma_2 = \sigma_1$.

Effect Size – Current Test Statistic

Z[t] (Current Test Statistic)

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

Conditional Power of Two-Sample T-Tests

Example 1 – Computing Conditional Power

Suppose a study has been planned to detect a mean change of 2 at an alpha of 0.05 using a two-sided t-test. The sample size is 60. The standard deviations are expected to be about 4. 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 a mean change of 0.5, 1, 1.5, and 2.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Conditional Power of Two-Sample T-Tests** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Conditional Power**, and then clicking on **Conditional Power of Two-Sample T-Tests**. You may make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Alternative Hypothesis	Ha: $\mu_1 \neq \mu_2$ (Two-Sided)
Alpha.....	0.05
N1 (Target Sample Size Group 1)	60
N2 (Target Sample Size Group 2)	Use R
R (Sample Allocation Ratio)	1.0
n1k (Sample Size to k Group 1)	30
n2k (Sample Size to k Group 2)	n1k
μ_1 (Mean of Group 1)	0
μ_2 (Mean of Group 2)	0.5 1 1.5 2
σ_1 (Std Deviation of Group 1).....	4
σ_2 (Std Deviation of Group 2).....	σ_1
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 the Conditional Power of the Two-Sample T-Test

Null Hypothesis: $\mu_1 = \mu_2$ Alternative Hypothesis: $\mu_1 \neq \mu_2$

Cond. Power	Pred. Power	Total Sample Sizes N1 N2	Current Sample Sizes n1k n2k	Means $\mu_1 \mu_2$	Mean Diff. $\mu_2 - \mu_1$ δ	Std Dev.s $\sigma_1 \sigma_2$	Test Stat. Zk	Alpha	Futility
0.43342	0.85040	60 60	30 30	0.0 0.5	0.5	4.0 4.0	2.120	0.05000	0.56658
0.62417	0.85040	60 60	30 30	0.0 1.0	1.0	4.0 4.0	2.120	0.05000	0.37583
0.78831	0.85040	60 60	30 30	0.0 1.5	1.5	4.0 4.0	2.120	0.05000	0.21169
0.90055	0.85040	60 60	30 30	0.0 2.0	2.0	4.0 4.0	2.120	0.05000	0.09945

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. NY, NY.
- Chang, Mark. 2008. Classical and Adaptive Clinical Trial Designs. John Wiley & Sons. Hoboken, New Jersey.

Conditional Power of Two-Sample T-Tests

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.

Predicted Power is the average conditional power, averaged over the effect size.

$N1|N2$ are the anticipated total sample sizes of groups 1 and 2.

$n1k|n2k$ are the sample sizes of groups 1 and 2 obtained through stage k.

$\mu1$ is the mean of group 1.

$\mu2$ is the mean of group 2 under the alternative hypothesis.

δ is the mean difference that is to be detected by the completed study.

$\sigma1|\sigma2$ are the standard deviations of groups 1 and 2.

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.

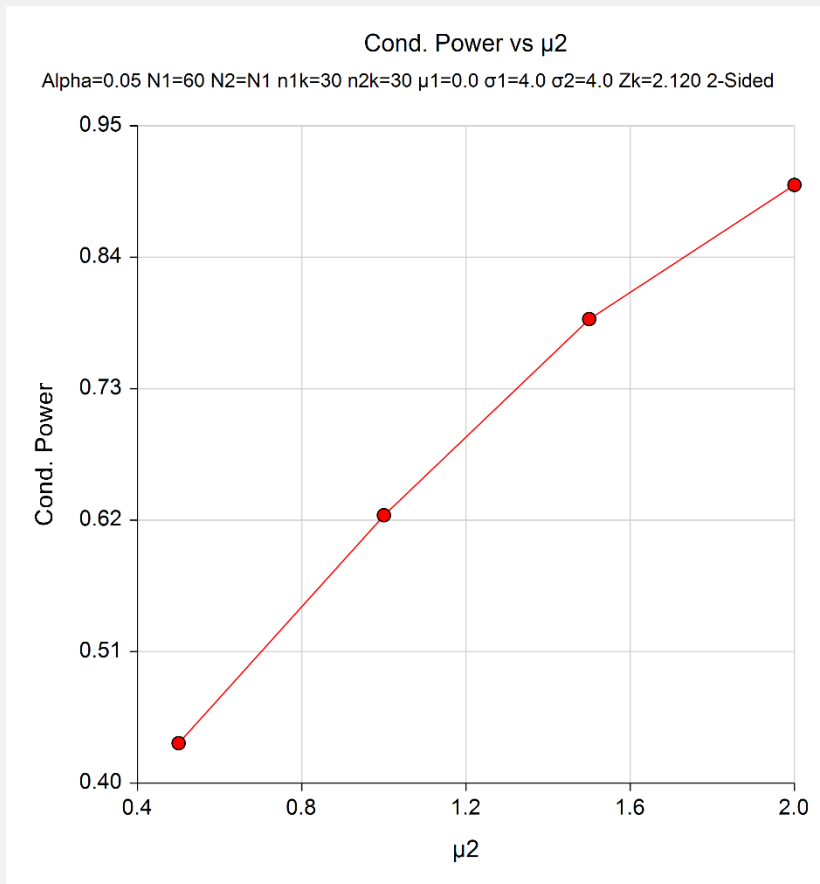
Futility is one minus the conditional power. A value greater than 0.9 or 0.8 indicates the study should be stopped because there is little chance of achieving statistical significance.

Summary Statements

The first 30 of 60 subjects in group 1 and 30 of 60 subjects in group 2 achieve 43% conditional power to detect the difference ($\mu2 - \mu1$) of 0.5 between the group 1 mean of 0.0 and the group 2 mean of 0.5 with group standard deviations of 4.0 and 4.0 and a significance level of 0.05000 using a two-sided test. The value of the test statistic, Zk , from data that have emerged so far is 2.120. The futility index is 0.56658.

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 $\mu2$.

Conditional Power of Two-Sample T-Tests

Example 2 – Validation

We could not find an example of a conditional power calculation for a two-sample t-test in the literature. Since the calculations are relatively simple, we will validate the calculation of the first scenario ($\theta = 0.5$) of Example 1 by hand.

In this case

$$\begin{aligned}
 I_k &= \left(\sigma_1^2 / n_{1k} + \sigma_2^2 / n_{2k} \right)^{-1} & I_K &= \left(\sigma_1^2 / n_1 + \sigma_2^2 / n_2 \right)^{-1} \\
 &= \left(\frac{16}{30} + \frac{16}{30} \right)^{-1} & &= \left(\frac{16}{60} + \frac{16}{60} \right)^{-1} \\
 &= 0.9375 & &= 1.875
 \end{aligned}$$

$$\begin{aligned}
 P_k(\theta) &= \Phi \left(\frac{Z_k \sqrt{I_k} - z_{1-\alpha/2} \sqrt{I_K} + \theta(I_K - I_k)}{\sqrt{I_K - I_k}} \right) + \Phi \left(\frac{-Z_k \sqrt{I_k} - z_{1-\alpha/2} \sqrt{I_K} - \theta(I_K - I_k)}{\sqrt{I_K - I_k}} \right) \\
 &= \Phi \left(\frac{2.12 \sqrt{0.9375} - (1.9599640) \sqrt{1.875} + 0.5(1.875 - 0.9375)}{\sqrt{1.875 - 0.9375}} \right) \\
 &\quad + \Phi \left(\frac{-2.12 \sqrt{0.9375} - (1.9599640) \sqrt{1.875} - 0.5(1.875 - 0.9375)}{\sqrt{1.875 - 0.9375}} \right) \\
 &= \Phi(-0.1676848) + \Phi(-5.3759305) \\
 &= 0.4334156 + 0.0000000 \\
 &= 0.4334156
 \end{aligned}$$

This value matches the first line of the report in Example 1.