Chapter 403

Conditional Power of Paired 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 paired *t*-test is used to test whether the mean paired difference is greater than, less than, or not equal to a specific value.

Technical Details

All details and assumptions usually made when using a paired *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 ($k = 1,\ldots, K-1$)

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 paired means with null hypothesis $H_0: \delta = \delta 0$, where $\delta 0$ is the hypothesized mean paired difference under the null hypothesis, and if $\delta 1$ is the hypothesized difference under the alternative hypothesis, these components are computed in Change (2008) page 69 as

$\theta = \delta 1 - \delta 0$ (the expected difference of mean paired differences under the alternative hypothesis)

$Z_k = (\bar{d}_k - \delta 0)\sqrt{\hat{I}_k}$ (the paired t-statistic computed from the observed data)

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

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

where

$\bar{d}_k$ is the sample mean difference, estimating δ_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

σ_d^2 is the variance of the paired differences

N is the total sample size

The value of σ_d^2 may computed directly from the differences or from the original data using $\sigma_d^2 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$. Here, σ_1^2 is the variance of the first variable from which the difference is computed

and σ_2^2 is the variance of the second variable. The correlation coefficient ρ is correlation between the first and second variables.

Computing conditional power requires you to set σ_d^2 , $\delta 0$, and $\delta 1$. 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.

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 $\delta 0$ and $\delta 1$ match your choice. For example, if you select *H1: $\delta 1 < \delta 0$*, the value $\delta 1$ should be less than $\delta 0$.

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

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 at 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 – Mean Paired Difference

δ0 (Null Difference)

Enter a value for δ_0, the mean of the paired differences under the null hypothesis. Usually, δ_1 and δ_0 are the original values upon which the study sample size was based.

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

δ0 (Actual Difference to Detect)

Enter a value for δ_1, the mean paired difference assumed by the alternative hypothesis at which conditional power is calculated. Only the difference, $\delta_1 - \delta_0$, is needed to compute the conditional power. Usually, δ_1 and δ_0 are the original values upon which the study sample size was based.

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

Effect Size – Standard Deviation

σd (Std Dev of Paired Differences)

Enter a value for σ_d, the standard deviation of the paired differences. 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 so far. This value may be positive or negative. Usually, the *t*-statistic ranges between -5 and 5.

Example 1 – Computing Conditional Power

Suppose a paired study has been planned to detect a mean change of 1 at an alpha of 0.025 using a one-sided, paired *t*-test. The sample size is 52. 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 also like to have the conditional power calculated for mean changes of 0.2, 0.4, 0.6, and 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 **Conditional Power of Paired T-Tests** procedure window by expanding **Means**, then **Paired Means**, then clicking on **Conditional Power**, and then clicking on **Conditional Power of Paired 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**.

Option   **Value**

Design Tab

Alternative Hypothesis .... **H1: δ > δ0 (One-Sided)**

Alpha **0.025**

N (Target Sample Size) **52**

nk (Sample Size at Look k) **26**

δ0 (Null Difference) **0**

δ1 (Actual Difference to Detect) **0.2 0.4 0.6 0.8 1**

σd (Std Dev of Paired Differences) **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 Paired T-Test ───
Hypotheses: H0: δ ≤ δ0   vs.   H1: δ > δ0

Cond. Power	Pred. Power	Total Sample Size N	Current Sample Size nk	Null Diff δ0	Actual Diff δ1	Diff of Diffs δ1-δ0	Std Dev σd	Test Stat Zk	Alpha	Futility
0.46603	0.85040	52	26	0.00	0.20	0.20	1.80	2.120	0.025	0.53397
0.68485	0.85040	52	26	0.00	0.40	0.40	1.80	2.120	0.025	0.31515
0.85265	0.85040	52	26	0.00	0.60	0.60	1.80	2.120	0.025	0.14735
0.94678	0.85040	52	26	0.00	0.80	0.80	1.80	2.120	0.025	0.05322
0.98541	0.85040	52	26	0.00	1.00	1.00	1.80	2.120	0.025	0.01459

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.

Conditional Power of Paired 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.
Predictive Power is the result of averaging the conditional power over the posterior distribution of the effect size.
N is the anticipated total sample size.
nk is the sample size obtained through stage k.
$\delta 0$ is the mean paired difference assuming the null hypothesis.
$\delta 1$ is the actual paired difference to detect under the alternative hypothesis at which conditional power is calculated.
$\delta 1 - \delta 0$ is the difference of the mean paired differences. This is the minimum value that is to be detected by the completed study.
σd is the standard deviation of the paired differences.
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 26 of the planned 52 subjects achieve 46.603% conditional power to detect a difference of $\delta 1 - \delta 0 = 0.20$ between the alternative mean difference $\delta 1 = 0.20$ and the null mean difference $\delta 0 = 0.00$ using a one-sided paired t-test with a standard deviation of paired differences of 1.80 and a significance level of 0.025. The t-value of the data that have emerged so far is 2.120. The futility index is 0.53397.

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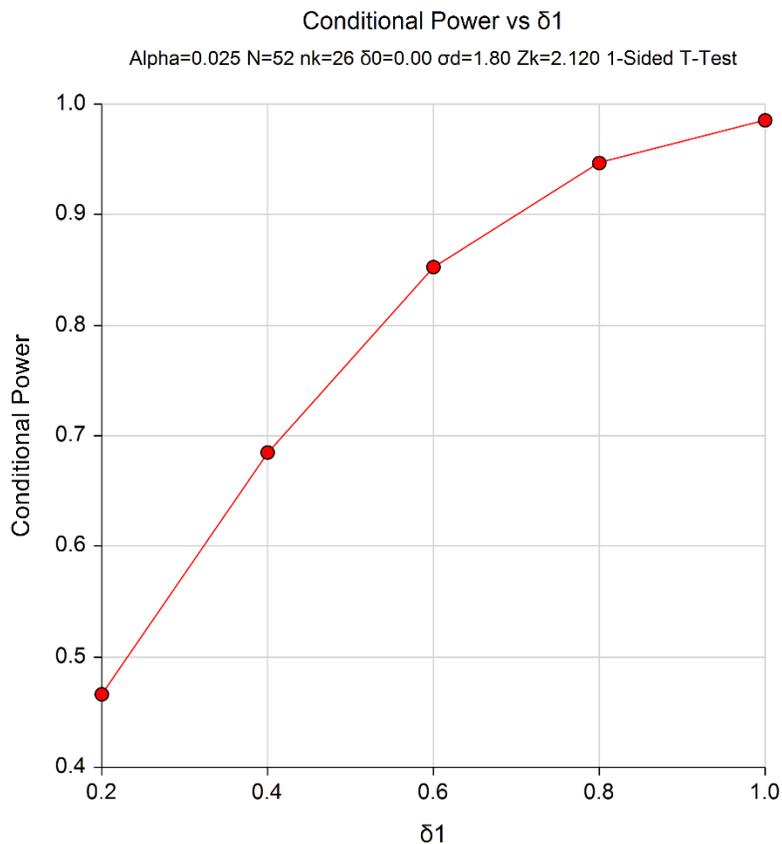
This plot shows the relationship between conditional power and $\delta 1$.

Example 2 – Validation

We could not find an example of a conditional power calculation for a paired *t*-test in the literature. Since the calculations are relatively simple, we will validate the calculation of the third scenario (δ1 = 0.6) of Example 1 by hand.

In this case

$$I_k = n_k / \sigma_d^2 \qquad\qquad I_K = N / \sigma_d^2$$

$$= \frac{26}{1.8^2} \qquad\qquad = \frac{52}{1.8^2}$$

$$= 8.0246914 \qquad\qquad = 16.0493827$$

$$
\begin{aligned}
P_{uk}(\theta) &= \Phi\left(\frac{Z_k \sqrt{I_k} - z_{1-\alpha}\sqrt{I_K} + (\delta_1 - \delta_0)(I_K - I_k)}{\sqrt{I_K - I_k}} \right) \\
&= \Phi\left(\frac{2.12\sqrt{8.0246914} - (1.959964)\sqrt{16.0493827} + (0.6 - 0)(16.0493827 - 8.0246914)}{\sqrt{16.0493827 - 8.0246914}} \right) \\
&= \Phi\left(\frac{2.9683814}{2.8327886} \right) \\
&= \Phi(1.0478655) \\
&= 0.8526497
\end{aligned}
$$

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