Chapter 433

Conditional Power and Sample Size
Reestimation 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 predictive power for the case when a two-sample t-test is used to test whether the means of two populations are different. It also provides *sample size reestimation* to achieve a specified conditional power value.

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 \(\theta\) at the end of the study, given the observed test statistic, \(Z_k\), is computed as

\[
P_{uk}(\theta) = \Phi\left( \frac{Z_k\sqrt{l_k} - z_{1-\alpha}\sqrt{l_k} + \theta(l_k - l_k)}{\sqrt{l_k - l_k}} \right).
\]
the general lower one-sided conditional power at stage $k$ is computed as

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

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

- $\theta$ = 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 $\alpha$.

For a test of a two means with null hypothesis $H_0: \mu_1 = \mu_2$, where $\mu_1$ and $\mu_2$ 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)

$$Z_k = (\bar{x}_{2k} - \bar{x}_{1k}) \sqrt{\frac{1}{I_k}}$$ (the z-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 $\mu_j$ 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 $j$
- $\sigma_j^2$ is the variance of group $j$

Computing conditional power requires you to set $\mu_1, \mu_2, \sigma_1, \sigma_2$, in addition to the current test statistic value, $Z_k$. 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.
Converting from a T-Statistic to a Z-Statistic

A common problem is that the procedure requires a Z-statistic, but the results from an analysis usually provide a T-statistic with a given degrees of freedom. So, the T-statistic must be transformed to a Z-statistic. One way to do this is to use the associated p-value. This is accomplished using the following steps:

**Step 1. Find the p-value associated with the t-statistic**

For example, suppose we have a t-statistic of 2.33 with 30 degrees of freedom. Using the PASS Probability Calculator, the probability for Student's T distribution is \( \text{Prob}(t \geq T) = 0.0133616 \).

**Step 2. Convert this p-value to a z-statistic**

Continuing the example, we can use the Normal distribution in the PASS Probability Calculator to determine that the z-score associated with a p-value of 0.0133616 is -2.215537403212. Since the t-statistic was positive, we use \( z = 2.215537403212 \).

**Step 3. Enter the z-statistic for \( Z_k \) (Current Test Statistic)**

Continuing the example, we would enter 2.1255374 for \( Z_k \) (Current Test Statistic).

**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

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

and 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*.
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, \( \mu_2 - \mu_1 \), of 0.5, 1, 1.5, and 2.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the Example 1 settings file. To load these settings to the procedure window, click Open Example Settings File in the Help Center or File menu.

Design Tab

- **Solve For**: Conditional Power
- **Alternative Hypothesis**: \( H_1: \mu_2 - \mu_1 \neq 0 \) (Two-Sided)
- **Alpha**: 0.05
- **N1 (Group 1 Target Sample Size)**: 60
- **N2 (Group 2 Target Sample Size)**: Use R
- **R (Sample Allocation Ratio)**: 1.0
- **n1k (Group 1 Sample Size at Look k)**: 30
- **n2k (Group 2 Sample Size at Look k)**: n1k
- **\( \mu_1 \) (Mean of Group 1)**: 0
- **\( \mu_2 \) (Mean of Group 2)**: 0.5 1 1.5 2
- **\( \sigma_1 \) (Standard Deviation of Group 1)**: 4
- **\( \sigma_2 \) (Standard Deviation of Group 2)**: \( \sigma_1 \)
- **Test Statistic Input Type**: Tk
- **Tk (Current Test Statistic)**: 2.12
### Output

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

### Numeric Reports

#### Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>Solve For: Conditional Power</th>
<th>Hypotheses: H0: μ2 - μ1 = 0 vs. H1: μ2 - μ1 ≠ 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Size</td>
<td>N1</td>
<td>N2</td>
</tr>
<tr>
<td>0.41449</td>
<td>0.83393</td>
<td>60</td>
</tr>
<tr>
<td>0.60569</td>
<td>0.83393</td>
<td>60</td>
</tr>
<tr>
<td>0.77405</td>
<td>0.83393</td>
<td>60</td>
</tr>
<tr>
<td>0.89184</td>
<td>0.83393</td>
<td>60</td>
</tr>
</tbody>
</table>

- **Conditional Power**: The probability of rejecting a false null hypothesis at the end of the study given the data that have emerged so far.
- **Predictive Power**: The result of averaging the conditional power over the posterior distribution of the effect size.
- **N1 and N2**: The target sample sizes of groups 1 and 2, respectively.
- **n1k and n2k**: The actual sample sizes of groups 1 and 2, respectively, obtained through stage k.
- **μ1 and μ2**: The means of group 1 and 2.
- **δ1**: The actual difference to detect under the alternative hypothesis at which conditional power is calculated. δ1 = μ2 - μ1.
- **σ1 and σ2**: The standard deviations of groups 1 and 2, respectively.
- **Tk**: The value of the test statistic from the observed data at stage k.
- **Alpha**: The probability of rejecting a true null hypothesis.
- **Futility**: Equal to 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 41.449% conditional power to detect a difference of δ1 = μ2 - μ1 = 0.5 - 0 = 0.5 using a two-sided T-test with group 1 standard deviation of σ1 = 4, group 2 standard deviation of σ2 = 4, and a significance level of alpha = 0.05. The value of the T-test statistic from data that have emerged so far is Tk = 2.12. The futility index is 0.58551.

#### References


This report shows the values of each of the parameters, one scenario per row.
This plot shows the relationship between conditional power and $\mu^2$. 
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 (θ = 0.5) of Example 1 by hand.

In this case

\[ I_k = \left( \frac{\sigma_1^2}{n_1 k} + \frac{\sigma_2^2}{n_2 k} \right)^{-1} \]
\[ \frac{16}{30} + \frac{16}{30} = 0.9375 \]
\[ I_K = \frac{16}{60} + \frac{16}{60} = 1.875 \]

\[ P_k(\theta) = \Phi \left( \frac{Z_k \sqrt{I_k} - z_{1-a/2} \sqrt{I_k} + \theta(I_k - I_K)}{\sqrt{I_K - I_k}} \right) + \Phi \left( \frac{-Z_k \sqrt{I_k} - z_{1-a/2} \sqrt{I_k} - \theta(I_k - I_K)}{\sqrt{I_K - I_k}} \right) \]
\[ = \Phi \left( \frac{2.12 \sqrt{0.9375} - (1.9599640) \sqrt{1.875} + 0.5(1.875 - 0.9375)}{\sqrt{1.875 - 0.9375}} \right) \]
\[ + \Phi \left( \frac{-2.12 \sqrt{0.9375} - (1.9599640) \sqrt{1.875} - 0.5(1.875 - 0.9375)}{\sqrt{1.875 - 0.9375}} \right) \]
\[ = \Phi(-0.1676848) + \Phi(-5.3759305) \]
\[ = 0.4334156 + 0.0000000 \]
\[ = 0.4334156 \]

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the Example 2 settings file. To load these settings to the procedure window, click Open Example Settings File in the Help Center or File menu.

<table>
<thead>
<tr>
<th>Design Tab</th>
<th>Conditional Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solve For</td>
<td>H1: μ2 - μ1 ≠ 0 (Two-Sided)</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N1 (Group 1 Target Sample Size)</td>
<td>60</td>
</tr>
<tr>
<td>N2 (Group 2 Target Sample Size)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Sample Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>n1k (Group 1 Sample Size at Look k)</td>
<td>30</td>
</tr>
<tr>
<td>n2k (Group 2 Sample Size at Look k)</td>
<td>n1k</td>
</tr>
<tr>
<td>μ1 (Mean of Group 1)</td>
<td>0</td>
</tr>
<tr>
<td>μ2 (Mean of Group 2)</td>
<td>0.5</td>
</tr>
</tbody>
</table>
σ1 (Standard Deviation of Group 1) .......... 4
σ2 (Standard Deviation of Group 2) .......... σ1
Test Statistic Input Type ......................... Zk
Zk (Current Test Statistic) ......................... 2.12

Output

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

<table>
<thead>
<tr>
<th>Power</th>
<th>Target</th>
<th>Current</th>
<th>Mean</th>
<th>Mean Diff</th>
<th>Std Dev</th>
<th>Test Stat</th>
<th>Alpha</th>
<th>Futility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cond</td>
<td>Pred</td>
<td>N1</td>
<td>N2</td>
<td>nk1</td>
<td>nk2</td>
<td>μ1</td>
<td>μ2</td>
<td>δ</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.43342</td>
<td>0.8504</td>
<td>60</td>
<td>60</td>
<td>30</td>
<td>30</td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
</tr>
</tbody>
</table>

The conditional power of 0.43342 matches the value calculated by hand.
Example 3 – Sample Size Reestimation Example

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 original target sample size was 60. The standard deviations were expected to be about 4. An interim analysis was run after half the data were collected. This analysis yielded a z-test value of 2.12. This value was obtained by transforming the t-test p-value using the inverse normal. The standard deviations were found to be 6.7.

The data monitoring board would like to recalculate the sample size for a mean change of 1.5, a conditional power of 0.8, and a standard deviation of 6.7.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the Example 3 settings file. To load these settings to the procedure window, click Open Example Settings File in the Help Center or File menu.

Design Tab

- Solve For: Sample Size Reestimation
- Alternative Hypothesis: H1: μ2 - μ1 ≠ 0 (Two-Sided)
- Conditional Power: 0.8
- Alpha: 0.05
- N2 (Group 2 Target Sample Size): Use R
- R (Sample Allocation Ratio): 1.0
- n1k (Group 1 Sample Size at Look k): 30
- n2k (Group 2 Sample Size at Look k): n1k
- μ1 (Mean of Group 1): 0
- μ2 (Mean of Group 2): 1.5
- σ1 (Standard Deviation of Group 1): 6.7
- σ2 (Standard Deviation of Group 2): σ1
- Test Statistic Input Type: Zk
- Zk (Current Test Statistic): 2.12
Output

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

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cond Power</td>
<td>0.8</td>
</tr>
<tr>
<td>Pred Power</td>
<td>0.93153</td>
</tr>
<tr>
<td>Target N1</td>
<td>203</td>
</tr>
<tr>
<td>Target N2</td>
<td>203</td>
</tr>
<tr>
<td>Current N1</td>
<td>30</td>
</tr>
<tr>
<td>Current N2</td>
<td>30</td>
</tr>
<tr>
<td>Mean 1 μ1</td>
<td>0</td>
</tr>
<tr>
<td>Mean 2 μ2</td>
<td>1.5</td>
</tr>
<tr>
<td>Mean Diff δ1</td>
<td>1.5</td>
</tr>
<tr>
<td>Std Dev 1 σ1</td>
<td>6.7</td>
</tr>
<tr>
<td>Std Dev 2 σ2</td>
<td>6.7</td>
</tr>
<tr>
<td>Test Stat Zk</td>
<td>2.12</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>Futility</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Notice that the target sample size has increased from 60 per group (N = 120), to 203 per group (N = 406).