Chapter 477

Group-Sequential Tests for Two Means Assuming Normality (Simulation) (Legacy)

This procedure uses simulation for the calculation of the boundaries as well as for calculation of power (and sample size). Futility boundaries are limited. A variety of test statistics are available.

Introduction

This procedure can be used to determine power, sample size and/or boundaries for group sequential tests comparing the means of two groups. The common two-sample T-Test and the Mann-Whitney U test can be simulated in this procedure. For two-sided tests, significance (efficacy) boundaries can be generated. For one-sided tests, significance and futility boundaries can be produced. The spacing of the looks can be equal or custom specified. Boundaries can be computed based on popular alpha- and beta-spending functions (O'Brien-Fleming, Pocock, Hwang-Shih-DeCani Gamma family, linear) or custom spending functions. Boundaries can also be input directly to verify alpha- and/or beta-spending properties. Futility boundaries can be binding or non-binding. Maximum and average (expected) sample sizes are reported as well as the alpha and/or beta spent and incremental power at each look. Corresponding P-Value boundaries are also given for each boundary statistic. Plots of boundaries are also produced.

Technical Details

This section outlines many of the technical details of the techniques used in this procedure including the simulation summary, the test statistic details, and the use of spending functions.

An excellent text for the background and details of many group-sequential methods is Jennison and Turnbull (2000).

Simulation Procedure

In this procedure, a large number of simulations are used to calculate boundaries and power using the following steps:

1. Based on the specified distributions, random samples of size N1 and N2 are generated under the null distribution and under the alternative distribution. These are simulated samples as though the final look is reached.
2. For each sample, test statistics for each look are produced. For example, if N1 and N2 are 100 and there are 5 equally spaced looks, test statistics are generated from the random samples at N1 = N2 = 20, N1 = N2 = 40, N1 = N2 = 60, N1 = N2 = 80, and N1 = N2 = 100 for both null and alternative samples.

3. To generate the first significance boundary, the null distribution statistics of the first look (e.g., at N1 = N2 = 20) are ordered and the percent of alpha to be spent at the first look is determined (using either the alpha-spending function or the input value). The statistic for which the percent of statistics above (or below, as the case may be) that value is equal to the percent of alpha to be spent at the first look is the boundary statistic. It is seen here how important a large number of simulations is to the precision of the boundary estimates.
4. All null distribution samples that are outside the first significance boundary at the first look are removed from consideration for the second look. If binding futility boundaries are also being computed, all null distribution samples with statistics that are outside the first futility boundary are also removed from consideration for the second look. If non-binding futility boundaries are being computed, null distribution samples with statistics outside the first futility boundary are not removed.
5. To generate the second significance boundary, the remaining null distribution statistics of the second look (e.g., at N1 = N2 = 40) are ordered and the percent of alpha to be spent at the second look is determined (again, using either the alpha-spending function or the input value). The percent of alpha to be spent at the second look is multiplied by the total number of simulations to determine the number of the statistic that is to be the second boundary statistic. The statistic for which that number of statistics is above it (or below, as the case may be) is the second boundary statistic. For example, suppose there are initially 1000 simulated samples, with 10 removed at the first look (from, say, alpha spent at Look 1 equal to 0.01), leaving 990 samples considered for the second look. Suppose further that the alpha to be spent at the second look is 0.02. This is multiplied by 1000 to give 20. The 990 still-considered statistics are ordered and the 970^{th} (20 in from 990) statistic is the second boundary.
6. All null distribution samples that are outside the second significance boundary and the second futility boundary, if binding, at the second look are removed from consideration for the third look (e.g., leaving 970 statistics computed at N1 = N2 = 60 to be considered at the third look). Steps 4 and 5 are repeated until the final look is reached.

Futility boundaries are computed in a similar manner using the desired beta-spending function or custom beta-spending values and the alternative hypothesis simulated statistics at each look. For both binding and non-binding futility boundaries, samples for which alternative hypothesis statistics are outside either the significance or futility boundaries of the previous look are excluded from current and future looks.

Because the final futility and significance boundaries are required to be the same, futility boundaries are computed beginning at a small value of beta (e.g., 0.0001) and incrementing beta by that amount until the futility and significance boundaries meet.

When boundaries are entered directly, this procedure uses the null hypothesis and alternative hypothesis simulations to determine the number of test statistics that are outside the boundaries at each look. The cumulative proportion of alternative hypothesis statistics that are outside the significance boundaries is the overall power of the study.

Generating Random Distributions

Two methods are available in **PASS** to simulate random samples. The first method generates the random variates directly, one value at a time. The second method generates a large pool (over 10,000) of random values and then draws the random numbers from this pool. This second method can cut the running time of the simulation by 70%.

As mentioned above, the second method begins by generating a large pool of random numbers from the specified distributions. Each of these pools is evaluated to determine if its mean is within a small relative tolerance (0.0001) of the target mean. If the actual mean is not within the tolerance of the target mean, individual members of the population are replaced with new random numbers if the new random number moves the mean towards its target. Only a few hundred such swaps are required to bring the actual mean to within tolerance of the target mean. This population is then sampled with replacement using the uniform distribution. We have found that this method works well as long as the size of the pool is the maximum of twice the number of simulated samples desired and 10,000.

Test Statistics

This section describes the test statistics that are available in this procedure.

Two-Sample T-Test

The two-sample t-test assumes that the data are simple random samples from Normal populations with the same variance. This assumption implies that the data are continuous and their distribution is symmetric. The calculation of the t statistic is as follows

$$t_{df} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_{\bar{X}_1 - \bar{X}_2}}$$

where

$$\bar{X}_k = \frac{\sum_{i=1}^{N_k} X_{ki}}{N_k}$$

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sum_{i=1}^{N_1}(X_{1i} - \bar{X}_1)^2 + \sum_{i=1}^{N_2}(X_{2i} - \bar{X}_2)^2}{N_1 + N_2 - 2}\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}$$

$$df = N_1 + N_2 - 2$$

The significance of the test statistic is determined by computing the p-value based on the t distribution with degrees of freedom *df*. If this p-value is less than a specified level (often 0.05), the null hypothesis is rejected. Otherwise, no conclusion can be reached.

Mann-Whitney U Test

This test is the nonparametric substitute for the equal-variance t-test. Two key assumptions for this test are that the distributions are at least ordinal and that they are identical under H0. This implies that ties (repeated values) are not acceptable. When ties are present, the approximation provided can be used, but know that the theoretic results no longer hold.

The Mann-Whitney test statistic is defined as follows in Gibbons (1985).

$$z = \frac{W_1 - \frac{N_1(N_1 + N_2 + 1)}{2} + C}{s_W}$$

where

$$W_1 = \sum_{k=1}^{N_1} Rank(X_{1k})$$

The ranks are determined after combining the two samples. The standard deviation is calculated as

$$s_W = \sqrt{\frac{N_1 N_2 (N_1 + N_2 + 1)}{12} - \frac{N_1 N_2 \sum_{i=1}\left(t_i^3 - t_i\right)}{12(N_1 + N_2)(N_1 + N_2 - 1)}}$$

where t_1 is the number of observations tied at value one, t_2 is the number of observations tied at some value two, and so forth.

The correction factor, *C*, is 0.5 if the rest of the numerator of *z* is negative or -0.5 otherwise. The value of *z* is then compared to the standard normal distribution.

Spending Functions

Spending functions can be used in this procedure to specify the proportion of alpha or beta that is spent at each look without having to specify the proportion directly.

Spending functions have the characteristics that they are increasing and that

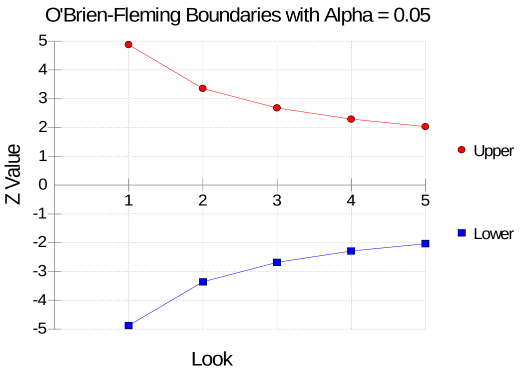
$$\alpha(0) = 0$$

$$\alpha(1) = \alpha$$

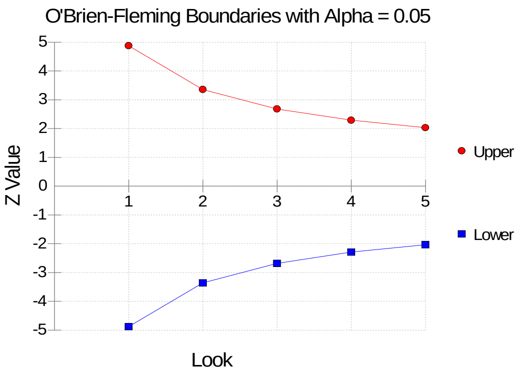
The last characteristic guarantees a fixed α level when the trial is complete. This methodology is very flexible since neither the times nor the number of analyses must be specified in advance. Only the functional form of $\alpha(\tau)$ must be specified.

PASS provides several popular spending functions plus the ability to enter and analyze your own percents of alpha or beta spent. These are calculated as follows (beta may be substituted for alpha for beta-spending functions):

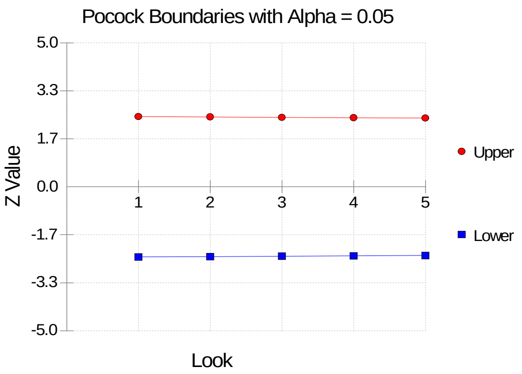
1. **Hwang-Shih-DeCani (gamma family)** $\alpha\left[\frac{1-e^{-\gamma t}}{1-e^{-\gamma}}\right], \gamma \neq 0; \quad \alpha t, \gamma = 0$



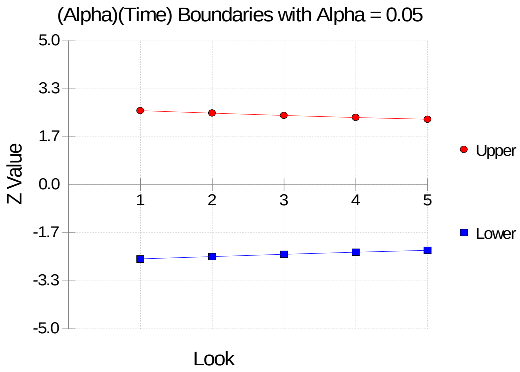
2. **O'Brien-Fleming Analog** $2 - 2\Phi\left(\frac{Z_{\alpha/2}}{\sqrt{t}}\right)$



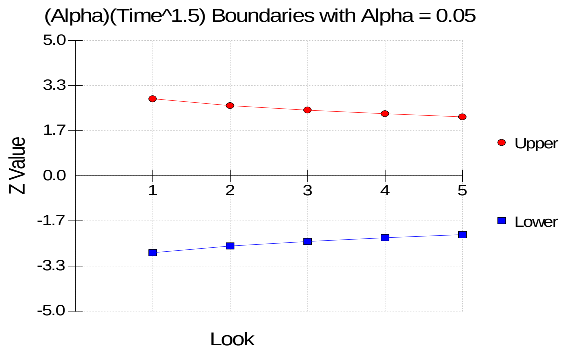
3. **Pocock Analog** $\alpha \cdot \ln(1 + (e - 1)t)$



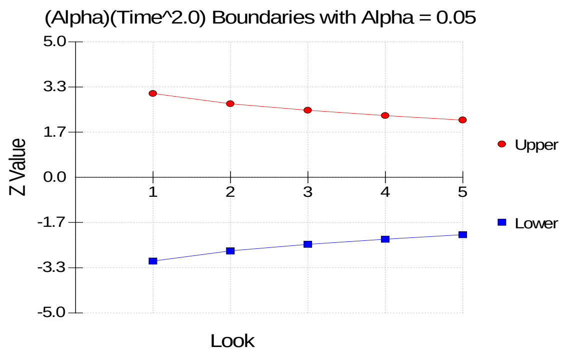
4. Alpha * time $\alpha \cdot t$



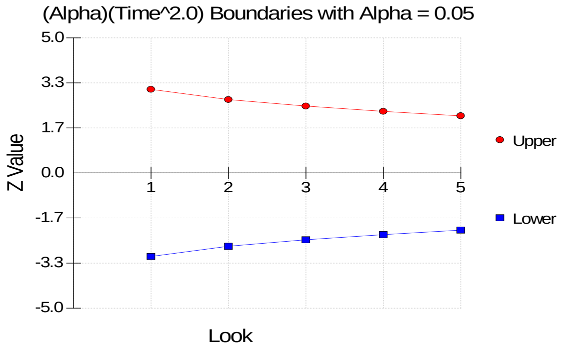
5. Alpha * time^1.5 $\alpha \cdot t^{3/2}$



6. Alpha * time^2 $\alpha \cdot t^2$



7. **Alpha * time^C** $\alpha \cdot t^C$



8. **User Supplied Percents**

   A custom set of percents of alpha to be spent at each look may be input directly.

The O'Brien-Fleming Analog spends very little alpha or beta at the beginning and much more at the final looks. The Pocock Analog and (Alpha or Beta)(Time) spending functions spend alpha or beta more evenly across the looks. The Hwang-Shih-DeCani (C) (gamma family) spending functions and (Alpha or Beta)(Time^C) spending functions are flexible spending functions that can be used to spend more alpha or beta early or late or evenly, depending on the choice of C.

Example 1 – Power and Output

A clinical trial is to be conducted over a two-year period to compare the mean response of a new treatment to that of the current treatment. The current response mean is 108. Although the researchers do not know the true mean of the new treatment, they would like to examine the power that is achieved if the mean of the new treatment is 113. The standard deviation for both groups is assumed to be 25. The sample size at the final look is to be 500 per group. Testing will be done at the 0.05 significance level. A total of five tests are going to be performed on the data as they are obtained. The O'Brien-Fleming (Analog) boundaries will be used.

Find the power and test boundaries assuming equal sample sizes per arm and two-sided hypothesis tests.

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 .... **Power**
Test Type... **T-Test**
Alternative Hypothesis .... **H1: Mean1 ≠ Mean2**
Simulations **20000**
Random Seed.. **3530283** (for Reproducibility)
Alpha... **0.05**
Group Allocation .... **Equal (N1 = N2)**
Sample Size Per Group **500**
Mean1 (Mean of Group 1, Control) **108**
Mean2 (Mean of Group 2, Treatment | H1).... **113**
Standard Deviation **25**

Looks & Boundaries Tab

Specification of Looks and Boundaries.......... **Simple**
Number of Equally Spaced Looks................. **5**
Alpha Spending Function.............................. **O'Brien-Fleming Analog**

Output

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

Scenario 1 Numeric Results for Group Sequential Testing Mean Difference = 0

Solve For:	Power
Hypotheses:	H0: Mean1 = Mean2; H1: Mean1 ≠ Mean2
Test Statistic:	T-Test
Alpha-Spending Function:	O'Brien-Fleming Analog
Beta-Spending Function:	None
Futility Boundary Type:	None
Number of Looks:	5
Simulations:	20000
Pool Size:	40000
Random Seed:	3530283 (User-Entered)

Numeric Summary for Scenario 1

Power			Alpha				
Value	95% LCL	95% UCL	Target	Actual	95% LCL	95% UCL	Beta
0.884	0.88	0.889	0.05	0.05275	0.04965	0.05585	0.116

		Average Sample Size							
		Given H0		Given H1					
N1	N2	Grp1	Grp2	Grp1	Grp2	Diff	Mean1	Mean2	Std Dev
500	500	496	496	377	377	-5	108	113	25

Power	The probability of rejecting a false null hypothesis at one of the looks. It is the total proportion of alternative hypothesis simulations that are outside the significance boundaries.
Power 95% LCL and UCL	The lower and upper confidence limits for the power estimate. The width of the interval is based on the number of simulations.
Target Alpha	The user-specified probability of rejecting a true null hypothesis. It is the total alpha spent.
Actual Alpha	The alpha level that was actually achieved by the experiment. It is the total proportion of the null hypothesis simulations that are outside the significance boundaries.
Alpha 95% LCL and UCL	The lower and upper confidence limits for the actual alpha estimate. The width of the interval is based on the number of simulations.
Beta	The probability of accepting a false null hypothesis. It is the total proportion of alternative hypothesis simulations that do not cross the significance boundaries.
N1 and N2	The sample sizes of each group if the study reaches the final look.
Average Sample Size Given H0	The average or expected sample sizes of each group if H0 is true. These are based on the proportion of null hypothesis simulations that cross the significance or futility boundaries at each look.
Average Sample Size Given H1	The average or expected sample sizes of each group if H1 is true. These are based on the proportion of alternative hypothesis simulations that cross the significance or futility boundaries at each look.
Diff	The mean difference between groups (Grp1 - Grp2) assuming the alternative hypothesis, H1.
Mean1, Mean2, and Std Dev	The parameters that were set by the user to define the null and alternative simulation distributions.

Summary Statements

A group sequential trial with group sample sizes of 500 and 500 at the final look achieves 88% power to detect a difference of -5 at the 0.05275 significance level (alpha) using a two-sided T-Test.

Accumulated Information Details for Scenario 1

Look	Accumulated Information Percent	Accumulated Sample Size Group 1	Group 2	Total
1	20	100	100	200
2	40	200	200	400
3	60	300	300	600
4	80	400	400	800
5	100	500	500	1000

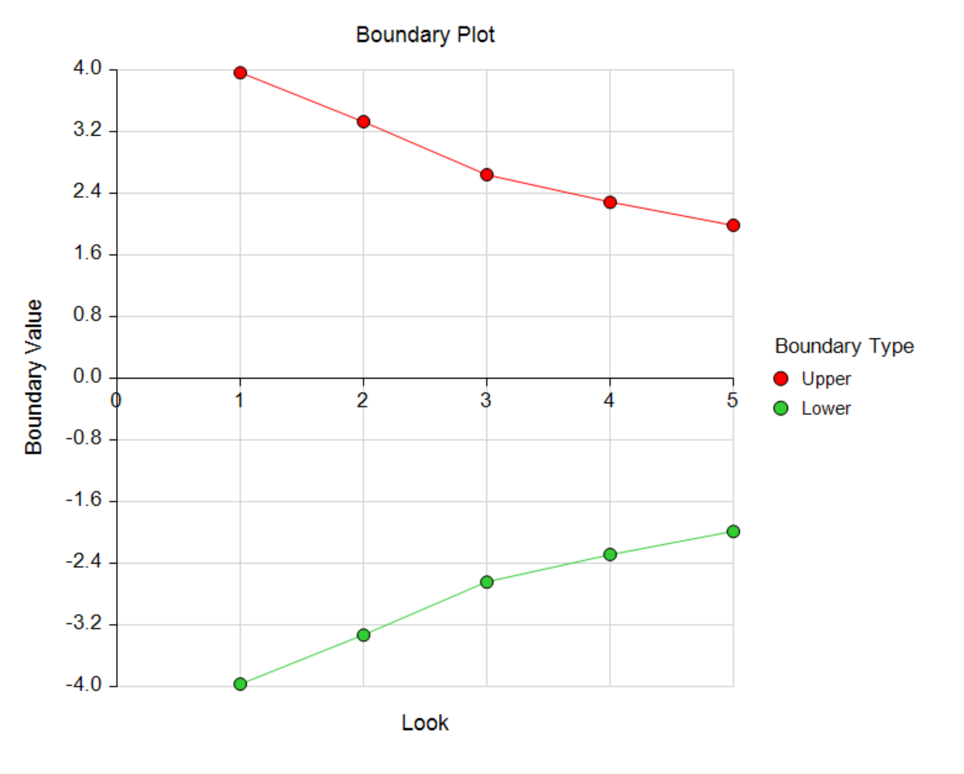
Look — The number of the look.
Accumulated Information Percent — The percent of the sample size accumulated up to the corresponding look.
Accumulated Sample Size Group 1 — The total number of individuals in group 1 at the corresponding look.
Accumulated Sample Size Group 2 — The total number of individuals in group 2 at the corresponding look.
Accumulated Sample Size Total — The total number of individuals in the study (group 1 + group 2) at the corresponding look.

Boundaries for Scenario 1

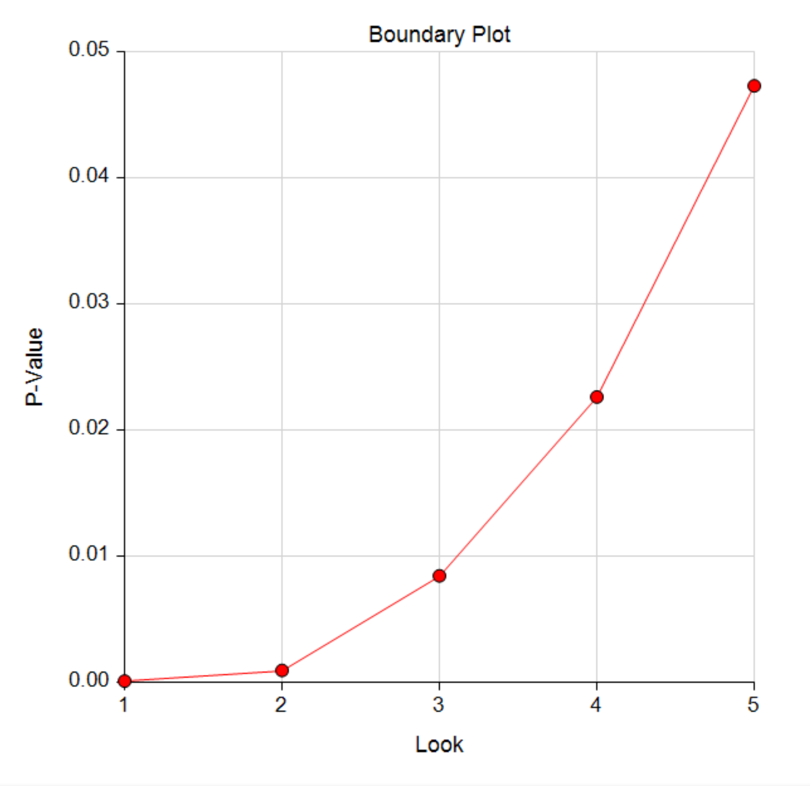
Look	Significance Boundary T-Value Scale	Significance Boundary P-Value Scale
1	+/- 3.96518	0.00010
2	+/- 3.33093	0.00095
3	+/- 2.64251	0.00844
4	+/- 2.2841	0.02263
5	+/- 1.9861	0.04730

Look — The number of the look.
Significance Boundary T-Value Scale — The value such that statistics outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. They are sometimes called efficacy boundaries.
Significance Boundary P-Value Scale — The value such that P-Values outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. This P-Value corresponds to the T-Value Boundary and is sometimes called the nominal alpha.

Boundary Plot



Boundary Plot - P-Value



Significance Boundaries with 95% Simulation Confidence Intervals for Scenario 1

	T-Value Boundary			P-Value Boundary		
Look	Value	95% LCL	95% UCL	Value	95% LCL	95% UCL
1	+/- 3.96518			0.00010		
2	+/- 3.33093	-3.57035	-3.11361	0.00095	0.00040	0.00198
3	+/- 2.64251	-2.70737	-2.57880	0.00844	0.00698	0.01015
4	+/- 2.2841	-2.35629	-2.23533	0.02263	0.01870	0.02567
5	+/- 1.9861	-2.03214	-1.95729	0.04730	0.04240	0.05059

Look The number of the look.
T-Value Boundary Value The value such that statistics outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. They are sometimes called efficacy boundaries.
P-Value Boundary Value The value such that P-Values outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. This P-Value corresponds to the T-Value Boundary and is sometimes called the nominal alpha.
95% LCL and UCL The lower and upper confidence limits for the boundary at the given look. The width of the interval is based on the number of simulations.

Alpha-Spending and Null Hypothesis Simulation Details for Scenario 1

	Signif. Boundary		Target		Actual			
Look	T-Value Scale	P-Value Scale	Spending Function Alpha	Cum. Spending Function Alpha	Alpha Spent	Cum. Alpha Spent	Proportion H1 Sims Outside Signif. Boundary	Cum. H1 Sims Outside Signif. Boundary
1	+/- 3.96518	0.00010	0.00000	0.00000	0.00010	0.00010	0.006	0.006
2	+/- 3.33093	0.00095	0.00079	0.00079	0.00130	0.00140	0.085	0.091
3	+/- 2.64251	0.00844	0.00683	0.00762	0.00800	0.00940	0.331	0.422
4	+/- 2.2841	0.02263	0.01681	0.02442	0.01620	0.02560	0.292	0.714
5	+/- 1.9861	0.04730	0.02558	0.05000	0.02715	0.05275	0.170	0.884

Look The number of the look.
Significance Boundary T-Value Scale The value such that statistics outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. They are sometimes called efficacy boundaries.
Significance Boundary P-Value Scale The value such that P-Values outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. This P-Value corresponds to the Significance T-Value Boundary and is sometimes called the nominal alpha.
Spending Function Alpha The intended portion of alpha allocated to the particular look based on the alpha-spending function.
Cumulative Spending Function Alpha The intended accumulated alpha allocated to the particular look. It is the sum of the Spending Function Alpha up to the corresponding look.
Alpha Spent The proportion of the null hypothesis simulations resulting in statistics outside the Significance Boundary at this look.
Cumulative Alpha Spent The proportion of the null hypothesis simulations resulting in Significance Boundary termination up to and including this look. It is the sum of the Alpha Spent up to the corresponding look.
Proportion H1 Sims Outside Significance Boundary The proportion of the alternative hypothesis simulations resulting in statistics outside the Significance Boundary at this look. It may be thought of as the incremental power.
Cumulative H1 Sims Outside Significance Boundary The proportion of the alternative hypothesis simulations resulting in Significance Boundary termination up to and including this look. It is the sum of the Proportion H1 Sims Outside Significance Boundary up to the corresponding look.

Numeric Results for Group Sequential Testing Mean Difference = 0.

Solve For:	Power
Hypotheses:	H0: Mean1=Mean2; H1: Mean1≠Mean2
Test Statistic:	T-Test
Alpha-Spending Function:	O'Brien-Fleming Analog
Beta-Spending Function:	None
Futility Boundary Type:	None
Number of Looks:	5
Simulations:	20000
Pool Size:	40000
Random Seed:	3530283 (User-Entered)

Numeric Summary of Scenarios

Scenario	Power	N1	N2	Alpha	Mean1	Mean2	Diff	Std Dev
1	0.884	500	500	0.05275	108	113	-5	25

Power	The probability of rejecting a false null hypothesis at one of the looks. It is the total proportion of alternative hypothesis simulations that are outside the significance boundaries.
Alpha	The alpha level that was actually achieved by the experiment. It is the total proportion of the null hypothesis simulations that are outside the significance boundaries.
N1 and N2	The sample sizes of each group if the study reaches the final look.
Mean1, Mean2, and Std Dev	The parameters that were set by the user to define the null and alternative simulation distributions.
Diff	The mean difference between groups (Grp1 - Grp2) assuming the alternative hypothesis, H1.

Power and Alpha Summary

	Power			Alpha				
Scenario	Value	95% LCL	95% UCL	Target	Actual	95% LCL	95% UCL	Beta
1	0.884	0.88	0.889	0.05	0.05275	0.04965	0.05585	0.116

Power	The probability of rejecting a false null hypothesis at one of the looks. It is the total proportion of alternative hypothesis simulations that are outside the significance boundaries.
Power 95% LCL and UCL	The lower and upper confidence limits for the power estimate. The width of the interval is based on the number of simulations.
Target Alpha	The user-specified probability of rejecting a true null hypothesis. It is the total alpha spent.
Alpha or Actual Alpha	The alpha level that was actually achieved by the experiment. It is the total proportion of the null hypothesis simulations that are outside the significance boundaries.
Alpha 95% LCL and UCL	The lower and upper confidence limits for the actual alpha estimate. The width of the interval is based on the number of simulations.
Beta	The probability of accepting a false null hypothesis. It is the total proportion of alternative hypothesis simulations that do not cross the significance boundaries.

Sample Size Summary

					Average Sample Size			
					Given H0		Given H1	
Scenario	Power	Alpha	N1	N2	Grp1	Grp2	Grp1	Grp2
1	0.884	0.05275	500	500	496	496	377	377

Power — The probability of rejecting a false null hypothesis at one of the looks. It is the total proportion of alternative hypothesis simulations that are outside the significance boundaries.
Alpha — The alpha level that was actually achieved by the experiment. It is the total proportion of the null hypothesis simulations that are outside the significance boundaries.
N1 and N2 — The sample sizes of each group if the study reaches the final look.
Average Sample Size Given H0 — The average or expected sample sizes of each group if H0 is true. These are based on the proportion of null hypothesis simulations that cross the significance or futility boundaries at each look.
Average Sample Size Given H1 — The average or expected sample sizes of each group if H1 is true. These are based on the proportion of alternative hypothesis simulations that cross the significance or futility boundaries at each look.

Run Time: 27.99 seconds.

References

Jennison, C., Turnbull, B.W. 2000. Group Sequential Methods with Applications to Clinical Trials. Chapman & Hall. Boca Raton, FL.
Devroye, Luc. 1986. Non-Uniform Random Variate Generation. Springer-Verlag. New York.
Matsumoto, M. and Nishimura,T. 1998. 'Mersenne twister: A 623-dimensionally equidistributed uniform pseudorandom number generator.' ACM Trans. On Modeling and Computer Simulations.
Zar, Jerrold H. 1984. Biostatistical Analysis (Second Edition). Prentice-Hall. Englewood Cliffs, New Jersey.

The values obtained from any given run of this example will vary slightly due to the variation in simulations.

Example 2 – Power for One-Sided Test with Futility Boundaries

Suppose researchers would like to compare two treatments with a one-sided test at each look. Further, suppose they would like to terminate the study early when it can be deemed highly unlikely that the new treatment is better than the standard. Suppose the control group mean is 108. The researchers wish to know the power of the test if the treatment group mean is 113. The sample size at the final look is to be 500 per group. Testing will be done at the 0.05 significance level. A total of five tests are going to be performed on the data as they are obtained. The O'Brien-Fleming (Analog) boundaries will be used for both significance and futility boundaries.

Find the power and test boundaries assuming equal sample sizes per arm and one-sided hypothesis tests.

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.

Design Tab

Solve For ....**Power**
Test Type..**T-Test**
Alternative Hypothesis ....**H1: Mean1 < Mean2**
Simulations **20000**
Random Seed..**3579726** (for Reproducibility)
Alpha..**0.05**
Group Allocation ....**Equal (N1 = N2)**
Sample Size Per Group **500**
Mean1 (Mean of Group 1, Control) **108**
Mean2 (Mean of Group 2, Treatment | H1)....**113**
Standard Deviation **25**

Looks & Boundaries Tab

Specification of Looks and Boundaries..........**Simple**
Number of Equally Spaced Looks.................**5**
Alpha Spending Function.............................**O'Brien-Fleming Analog**
Type of Futility Boundary ....**Non-binding**
Number of Skipped Futility Looks **0**
Beta Spending Function...............................**O'Brien-Fleming Analog**

Output

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

Scenario 1 Numeric Results for Group Sequential Testing Mean Difference = 0

Solve For:	Power
Hypotheses:	H0: Mean1 = Mean2; H1: Mean1 < Mean2
Test Statistic:	T-Test
Alpha-Spending Function:	O'Brien-Fleming Analog
Beta-Spending Function:	O'Brien-Fleming Analog
Futility Boundary Type:	Non-Binding
Number of Looks:	5
Simulations:	20000
Pool Size:	40000
Random Seed:	3579726 (User-Entered)

Numeric Summary for Scenario 1

Power			Alpha				
Value	95% LCL	95% UCL	Target	Actual	95% LCL	95% UCL	Beta
0.909	0.905	0.913	0.05	0.0454	0.04251	0.04829	0.091

		Average Sample Size							
		Given H0		Given H1					
N1	N2	Grp1	Grp2	Grp1	Grp2	Diff	Mean1	Mean2	Std Dev
500	500	304	304	335	335	-5	108	113	25

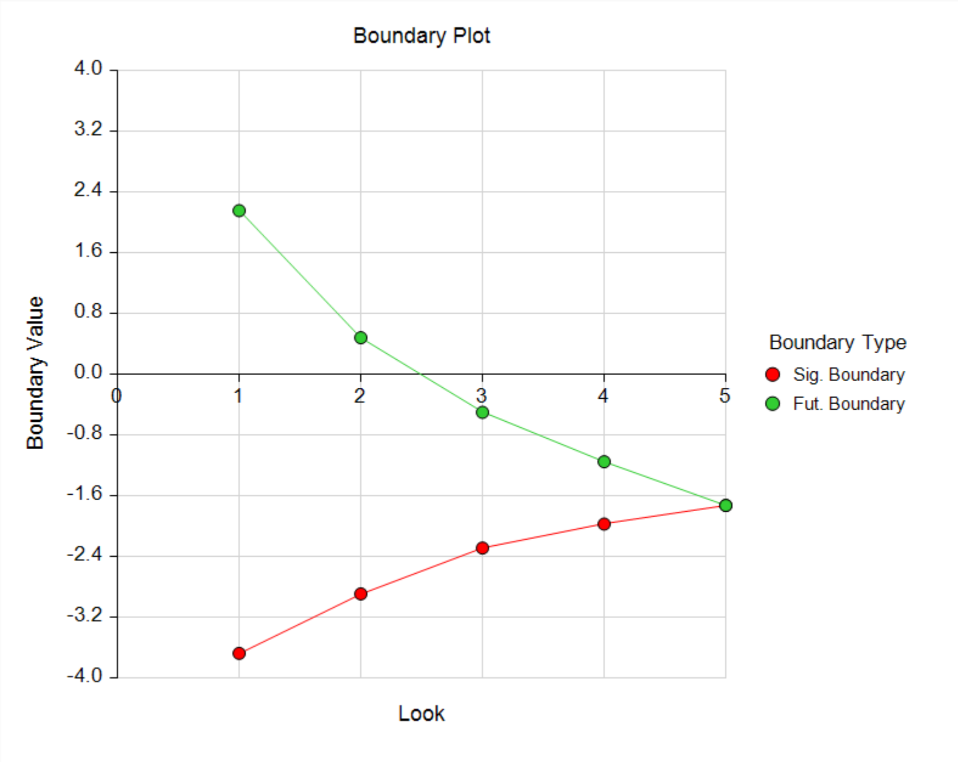
Accumulated Information Details for Scenario 1

	Accumulated	Accumulated Sample Size		
Look	Information Percent	Group 1	Group 2	Total
1	20	100	100	200
2	40	200	200	400
3	60	300	300	600
4	80	400	400	800
5	100	500	500	1000

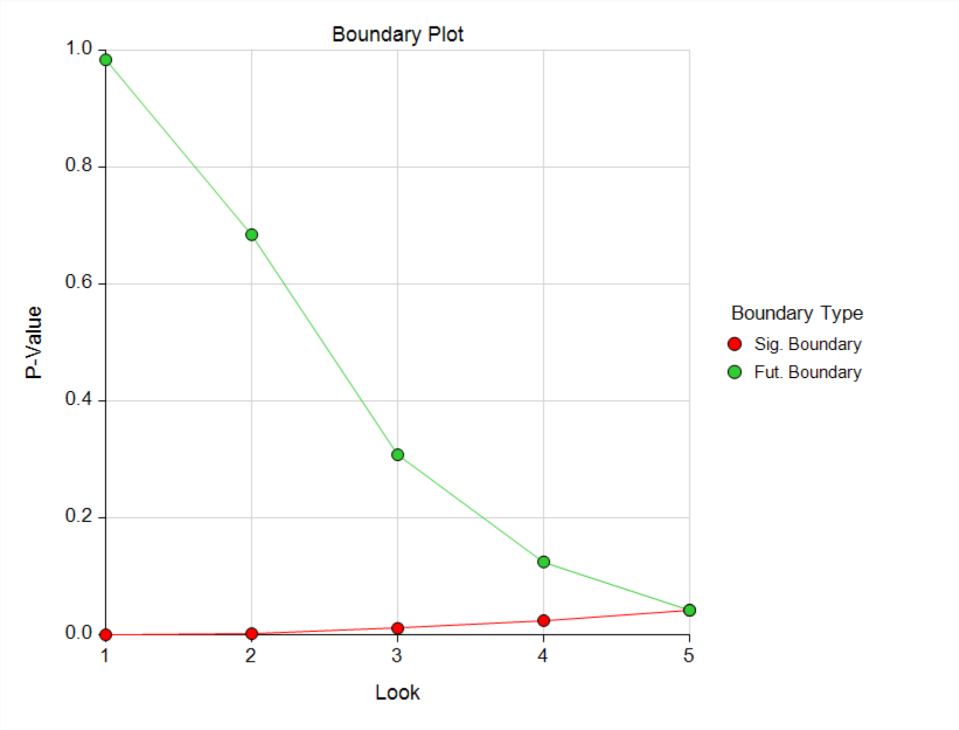
Boundaries for Scenario 1

	Significance Boundary		Futility Boundary	
Look	T-Value Scale	P-Value Scale	T-Value Scale	P-Value Scale
1	-3.67556	0.00015	2.15485	0.98381
2	-2.89209	0.00202	0.48119	0.68468
3	-2.28821	0.01124	-0.50138	0.30814
4	-1.97244	0.02445	-1.15184	0.12487
5	-1.72787	0.04216	-1.72787	0.04216

Boundary Plot



Boundary Plot - P-Value



Significance Boundaries with 95% Simulation Confidence Intervals for Scenario 1

	T-Value Boundary			P-Value Boundary		
Look	Value	95% LCL	95% UCL	Value	95% LCL	95% UCL
1	-3.67556			0.00015		
2	-2.89209	-2.99901	-2.82502	0.00202	0.00144	0.00248
3	-2.28821	-2.32608	-2.24129	0.01124	0.01017	0.01269
4	-1.97244	-1.99854	-1.93647	0.02445	0.02300	0.02658
5	-1.72787	-1.76470	-1.70138	0.04216	0.03896	0.04459

Futility Boundaries with 95% Simulation Confidence Intervals for Scenario 1

	T-Value Boundary			P-Value Boundary		
Look	Value	95% LCL	95% UCL	Value	95% LCL	95% UCL
1	2.15485	1.88581	2.90175	0.98381		
2	0.48119	0.44413	0.51648	0.68468	0.67141	0.69710
3	-0.50138	-0.53974	-0.45592	0.30814	0.29479	0.32431
4	-1.15184	-1.17897	-1.11825	0.12487	0.11938	0.13190
5	-1.72787	-1.75782	-1.70699	0.04216	0.03954	0.04407

Alpha-Spending and Null Hypothesis Simulation Details for Scenario 1

	Signif. Boundary		Target		Actual			
Look	T-Value Scale	P-Value Scale	Spending Function Alpha	Cum. Spending Function Alpha	Alpha Spent	Cum. Alpha Spent	Proportion H0 Sims Outside Futility Boundary	Cum. H0 Sims Outside Futility Boundary
1	-3.67556	0.00015	0.00001	0.00001	0.00000	0.00000	0.01560	0.01560
2	-2.89209	0.00202	0.00193	0.00194	0.00195	0.00195	0.30340	0.31900
3	-2.28821	0.01124	0.00945	0.01140	0.00945	0.01140	0.38210	0.70110
4	-1.97244	0.02445	0.01703	0.02843	0.01700	0.02840	0.18315	0.88425
5	-1.72787	0.04216	0.02157	0.05000	0.01700	0.04540	0.07030	0.95455

Beta-Spending and Alternative Hypothesis Simulation Details for Scenario 1

	Futility Boundary		Target		Actual			
Look	T-Value Scale	P-Value Scale	Spending Function Beta	Cum. Spending Function Beta	Beta Spent	Cum. Beta Spent	Proportion H1 Sims Outside Signif. Boundary	Cum. H1 Sims Outside Signif. Boundary
1	2.15485	0.98381	0.000	0.000	0.000	0.000	0.013	0.013
2	0.48119	0.68468	0.007	0.008	0.007	0.008	0.171	0.184
3	-0.50138	0.30814	0.022	0.029	0.022	0.029	0.377	0.560
4	-1.15184	0.12487	0.030	0.059	0.030	0.059	0.242	0.802
5	-1.72787	0.04216	0.032	0.091	0.032	0.091	0.107	0.909

The values obtained from any given run of this example will vary slightly due to the variation in simulations.

Example 3 – Enter Boundaries

With a set-up similar to Example 2, suppose we wish to investigate the properties of a set of significance (-3, -3, -3, -2, -1) and futility (2, 1, 0, 0, -1) boundaries.

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 .... **Alpha and Power (Enter Boundaries)**
Test Type.. **T-Test**
Alternative Hypothesis .... **H1: Mean1 < Mean2**
Simulations **20000**
Random Seed.. **3606318** (for Reproducibility)
Group Allocation .... **Equal (N1 = N2)**
Sample Size Per Group **500**
Mean1 (Mean of Group 1, Control) **108**
Mean2 (Mean of Group 2, Treatment | H1).... **113**
Standard Deviation **25**

Looks & Boundaries Tab

Number of Looks **5**
Equally Spaced... **Checked**
Types of Boundaries..................................... **Significance and Futility Boundaries**
Significance Boundary .... **-3 -3 -3 -2 -1** (for looks 1 through 5)
Futility Boundary .... **2 1 0 0 -1** (for looks 1 through 5)

Output

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

Scenario 1 Numeric Results for Group Sequential Testing Mean Difference = 0

Solve For:	Alpha and Power (Enter Boundaries)
Hypotheses:	H0: Mean1 = Mean2; H1: Mean1 < Mean2
Test Statistic:	T-Test
Type of Boundaries:	Significance and Futility Boundaries
Number of Looks:	5
Simulations:	20000
Pool Size:	40000
Random Seed:	3606318 (User-Entered)

Numeric Summary for Scenario 1

Power			Alpha			
Value	95% LCL	95% UCL	Value	95% LCL	95% UCL	Beta
0.978	0.976	0.98	0.151	0.14604	0.15596	0.022

		Average Sample Size							
		Given H0		Given H1					
N1	N2	Grp1	Grp2	Grp1	Grp2	Diff	Mean1	Mean2	Std Dev
500	500	369	369	363	363	-5	108	113	25

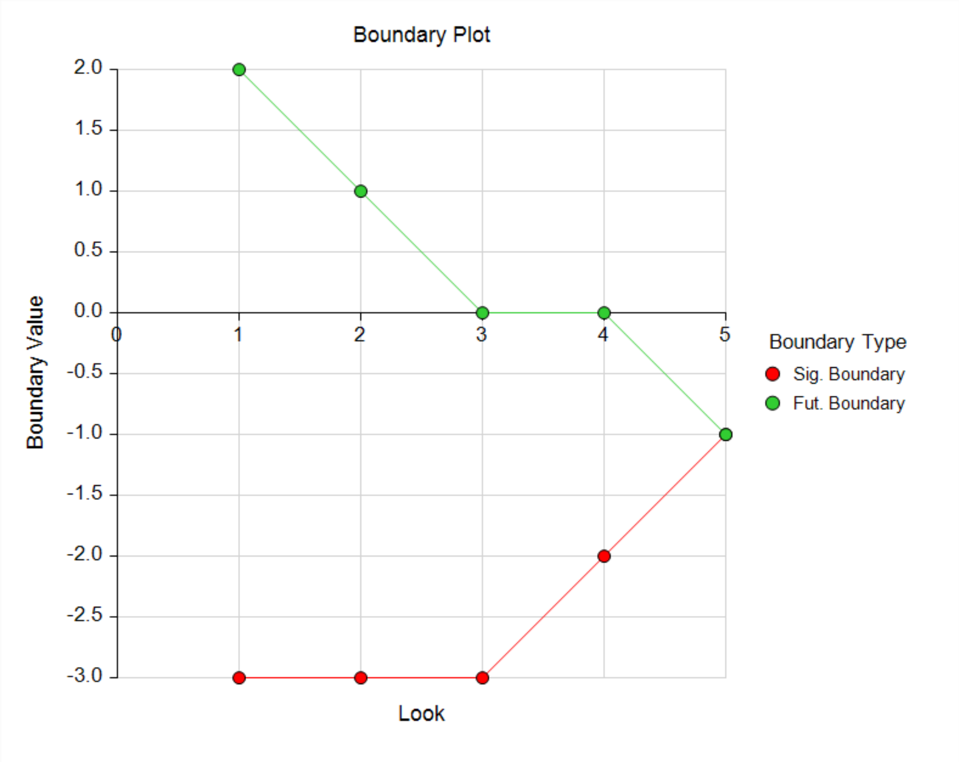
Accumulated Information Details for Scenario 1

	Accumulated Information	Accumulated Sample Size		
Look	Percent	Group 1	Group 2	Total
1	20	100	100	200
2	40	200	200	400
3	60	300	300	600
4	80	400	400	800
5	100	500	500	1000

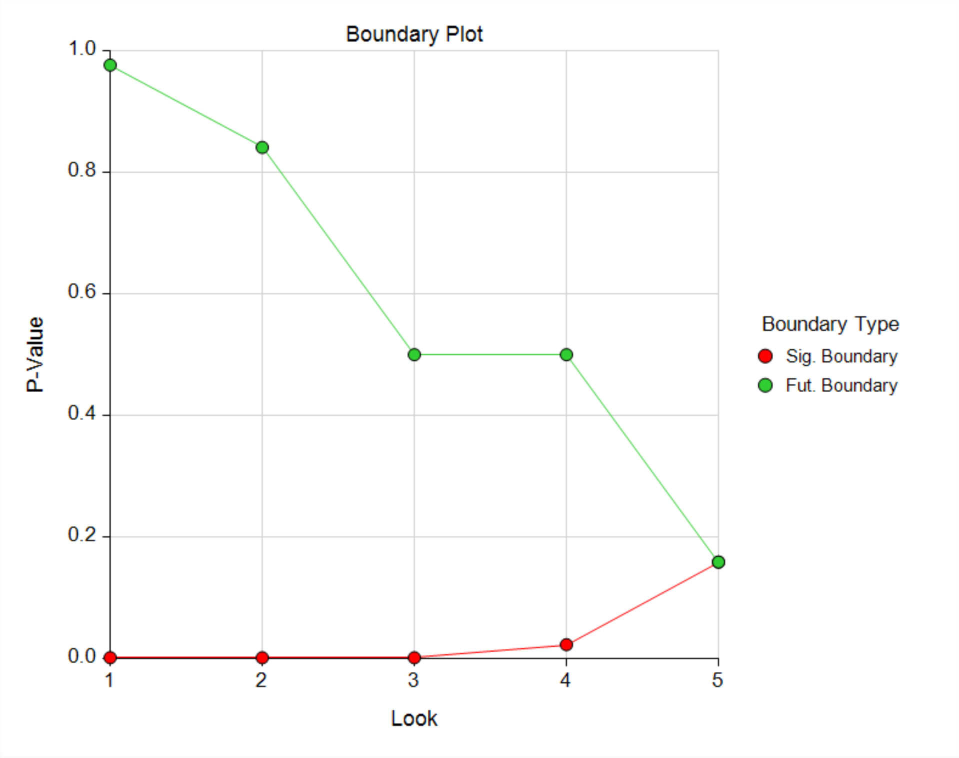
Boundaries for Scenario 1

	Significance Boundary		Futility Boundary	
Look	T-Value Scale	P-Value Scale	T-Value Scale	P-Value Scale
1	-3	0.00152	2	0.97657
2	-3	0.00143	1	0.84104
3	-3	0.00141	0	0.50000
4	-2	0.02292	0	0.50000
5	-1	0.15878	-1	0.15878

Boundary Plot



Boundary Plot - P-Value



Alpha-Spending and Null Hypothesis Simulation Details for Scenario 1

	Signif. Boundary				Proportion H0 Sims	Cum. H0 Sims
Look	T-Value Scale	P-Value Scale	Alpha Spent	Cum. Alpha Spent	Outside Futility Boundary	Outside Futility Boundary
1	-3	0.00152	0.00120	0.00120	0.02245	0.02245
2	-3	0.00143	0.00135	0.00255	0.13975	0.16220
3	-3	0.00141	0.00080	0.00335	0.34240	0.50460
4	-2	0.02292	0.02100	0.02435	0.08500	0.58960
5	-1	0.15878	0.12665	0.15100	0.25940	0.84900

Beta-Spending and Alternative Hypothesis Simulation Details for Scenario 1

	Futility Boundary				Proportion H1 Sims	Cum. H1 Sims
Look	T-Value Scale	P-Value Scale	Beta Spent	Cum. Beta Spent	Outside Signif. Boundary	Outside Signif. Boundary
1	2	0.97657	0.000	0.000	0.057	0.057
2	1	0.84104	0.002	0.002	0.117	0.174
3	0	0.50000	0.007	0.009	0.146	0.320
4	0	0.50000	0.001	0.010	0.477	0.797
5	-1	0.15878	0.012	0.022	0.181	0.978

The values obtained from any given run of this example will vary slightly due to the variation in simulations.

Example 4 – Validation Using Simulation

With a set-up similar to Example 1, we examine the power and alpha generated by the set of two-sided significance boundaries (+/- 4.026, +/- 3.396, +/- 2.682, +/- 2.302, +/- 1.990).

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 4** 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 .... **Alpha and Power (Enter Boundaries)**
Test Type.. **T-Test**
Alternative Hypothesis .... **H1: Mean1 ≠ Mean2**
Simulations **20000**
Random Seed... **3637490** (for Reproducibility)
Group Allocation .... **Equal (N1 = N2)**
Sample Size Per Group **500**
Mean1 (Mean of Group 1, Control) **108**
Mean2 (Mean of Group 2, Treatment | H1).... **113**
Standard Deviation **25**

Looks & Boundaries Tab

Number of Looks **5**
Equally Spaced... **Checked**
Significance Boundary **4.026, 3.396, 2.682, 2.302, 1.990**

Output

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

Scenario 1 Numeric Results for Group Sequential Testing Mean Difference = 0

Solve For:	Alpha and Power (Enter Boundaries)
Hypotheses:	H0: Mean1 = Mean2; H1: Mean1 ≠ Mean2
Test Statistic:	T-Test
Type of Boundaries:	Significance Boundaries Only
Number of Looks:	5
Simulations:	20000
Pool Size:	40000
Random Seed:	3637490 (User-Entered)

Numeric Summary for Scenario 1

Power			Alpha			
Value	95% LCL	95% UCL	Value	95% LCL	95% UCL	Beta
0.874	0.87	0.879	0.0524	0.04931	0.05549	0.126

		Average Sample Size							
		Given H0		Given H1					
N1	N2	Grp1	Grp2	Grp1	Grp2	Diff	Mean1	Mean2	Std Dev
500	500	497	497	382	382	-5	108	113	25

The values obtained from any given run of this example will vary slightly due to the variation in simulations. The power and alpha generated with these boundaries are very close to the values of Example 1.