Chapter 125

Two-Stage Designs for Tests of One Proportion (Simon)

Introduction

This module finds two-stage designs for exact tests of a single proportion that meet the error rate (type-I and type-II) criterion and minimize the expected sample size. An algorithm, presented by Simon (1989), finds the designs with the minimum N (*minimax*) and the minimum expected N (*optimum*). Extending Simon's work, Jung, Lee, Kim, George (2004) discuss other designs which are optimum from a Bayesian point of view which they call *admissible* designs.

In a two-stage design, the subjects are divided into two groups or stages. At the completion of the first stage, an interim analysis is made to determine if the second stage should be conducted. If the number of patients responding is greater than a certain amount, the second stage is conducted. Otherwise, it is not.

Technical Details

Phase I clinical trials are designed to provide information about the maximum tolerated dose levels of a treatment. They consist of three to six patients at each dose level and provide little information about the effectiveness of the treatment.

Phase II trials obtain initial estimates of the degree of treatment activity. A patient's response may be measured by the decrease in the size of a tumor. For example, a patient may be considered to have responded to treatment if the tumor shrinks by 50% or more. There is no control group in these designs. Rather, the purpose of the trial is to determine if the treatment shows enough activity against disease to warrant a full-scale, phase III clinical trial.

Let *P0* be the largest response proportion which, if true, clearly implies that the treatment does not warrant further study. *P0* is sometimes called the response rate of a *poor* treatment. For a new anti-tumor drug, this may be set to 0.10.

Let *P1* be the smallest response proportion which, if true, clearly implies that the treatment does warrant further study. *P1* is sometimes called the response rate of a good treatment. For a new anti-tumor drug, this may be set to 0.30.

A statistical test of hypothesis may be conducted to test the null hypothesis that $P \leq P0$ versus the alternative hypothesis that $P > P0$ (P is the true proportion responding to the treatment in the population). Let α be the probability of rejecting the null hypothesis when it is true. Let β be the probability of rejecting the alternative hypothesis when $P = P1$.

A phase II design can be represented by four numbers: $N1$, $R1$, N, and R. $N1$ is the sample size in the first stage. $R1$ is the critical value in the first stage. If $R1$ or fewer responses occur in the $N1$ patients, the drug is rejected. N is the combined sample size for both the first and second stages. R is the critical value in the combined sample. If R or fewer of the N patients respond, the drug is rejected.

The expected (or average) sample size of this design is

$$E(N) = N1 + (1 - PET)(N - N1)$$

where PET is the probability of early termination of the study computed at $P = P0$.

The probability of rejecting a drug with success probability P can be found using the binomial distribution. The formulation is

$$\Pr(reject|P, N1, R1, R, N) = B(R1|P, N1) + \sum_{X=R1+1}^{\min(N1,R)} b(X|P, N1)B(R - X|P, N - N1)$$

where

$$b(X|P, N) = \frac{N!}{X!(N - X)!}P^X(1 - P)^{N-X}$$

$$B(X|P, N) = \sum_{r=0}^{X} b(R|P, N)$$

The two error rate constraints are

$$\Pr(reject|P0, N1, R1, R, N) \geq 1 - \alpha$$

and

$$\Pr(reject|P1, N1, R1, R, N) \leq \beta$$

Optimum Designs

Simon (1989) defines two optimum designs. The first is the *minimax* design which is the design with the smallest N that obeys both error rate constraints. The second is the *optimum* design. It has the minimum expected N. These designs are found through a search of all possible designs. This search may take several minutes to complete.

Occasionally, neither the minimax or the optimum design is useful. For example, N1 may be too small or too large. Jung, Lee, Kim, George (2004) developed other designs which are optimum from a Bayesian point of view which they call *admissible* designs. These are a compromise between the two Simon designs.

The optimum design minimizes the average sample size, $E(N)$, while meeting the error rate constraints.

Designs Other Than Optimal

The optimal design minimizes the average sample size. There are examples where a less-than optimal design may be more desirable. For example, suppose the optimal design were $N1 = 5$ and $N = 25$. This design is poor because only 5 patients are obtained during the first stage, but 20 are needed during the second stage. Most researchers would rather have more balance in the sample sizes of the two stages. Because of this, the actual optimal design may be rejected on other grounds. There are often designs that are near optimal, but are much more balanced.

Design Flexibility

Dealing with sequential designs is complicated. It may be difficult to achieve exactly the number of patients proscribed for each phase. However, it should be remembered that the validity of the probability statements depends on the sample size requirements being met exactly. This is because the interpretation of an error rate probability statement is for repeated studies conducted in exactly the same way. We envision that if many studies of the same drug are conducted using the specific sampling plan $N1$, $R1$, N, R when $P = P0$, a proportion α of them will be falsely terminated due to chance occurrences.

The interpretation of the error rates is for a large number of identical studies in which the sampling plan is identical and as proscribed. If the sampling plan is allowed to vary, this interpretation is invalid. Of course, the degree of possible error in interpretation depends on the degree to which the sampling plan is changed. We recognize that when dealing with human subjects, flexibility must be maintained. However, the scientist must also recognize that when the sampling plan is changed, the exact probability statements can no longer be calculated.

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.

Designs

Search Type

Specify which search algorithm to use. Both algorithms, if properly set up, should provide the same results.

The options are:

- **Automatic**

  Provides the optimum designs in about 25% of the run time.

- **Custom**

  Allows you to set custom search ranges for each of the characteristics of the design. This may result in a much longer run time.

  This option is most useful when you want to find the characteristics of a few, specific designs.

Power and Alpha

Power (Probability of Rejecting H0: P ≤ P0 | P = P1)

Power is the probability of rejecting the null hypothesis that the proportion responding to the treatment is less than or equal to $P0$ when this hypothesis is false. That is, when $P = P1$. One minus power, beta, is the probability of a Type-II error.

The common range of power is 0.6 to 0.999 (Beta = 0.001 to 0.4). Popular values for power are 0.80 and 0.90 (Beta = 0.1 and 0.2).

Alpha (Probability of Rejecting H0: P ≤ P0 | P = P0)

Alpha is the probability of rejecting the hypothesis that the proportion responding to the treatment is less than or equal to *P0* when this hypothesis is actually true. That is, when P = P0.

The range of Alpha is 0.001 to 0.25. Popular values are 0.05 and 0.10.

Effect Size

P0 (Maximum Response Rate of a Poor Treatment)

This is the largest response proportion that will still allow the treatment to be deemed "poor".

If the true proportion responding to the treatment (drug) is less than this amount, study of the treatment will not be continued.

This value must be less than *P*1.

Only one value can be entered.

The range of *P*0 is $0 < P0 < P1 < 1$.

Note that values of *P*0 over 0.5 may result in large run-times.

P1 (Minimum Response Rate of a Good Treatment)

This is the minimum response proportion of a ""good" treatment. If the true proportion responding to the treatment (drug) is greater than or equal to this amount, the study of the treatment should be continued.

This value must be greater than P0.

Only one value can be entered.

The range is defined as $0 < P0 < P1 < 1$.

This is the response proportion of a good drug. If the true proportion responding to the treatment is greater than or equal to this amount, study of the treatment should be recommended.

Search Parameters – Automatic Search

Multiplier for Upper Search Limit of N

Simon (1989) recommended that the search for the optimum N continue to an upper limit of N*, where N* is 1.25 (Ns). Here, Ns is the sample size of a single-stage study.

This option lets you increase this amount if you desire.

Range

1.1 to 3.

Recommended: 1.25.

Search Parameters – Custom Search

Mininum N (Combined Sample Size)

N is the combined sample size of the two stages of the design. This parameter sets the minimum value of N that is used during the search. The optimum value of N must be between N Min and N Max or it will not be found.

The keyword MIN indicates that the value used is the minimum of the smallest sample size from a single stage design and *MIN*2 where *MIN*2 is calculated using

$$MIN2 = \frac{p_0 + p_1}{2}\left(1 - \frac{p_0 + p_1}{2}\right)\left[\frac{z_{1-\alpha} + z_{1-\beta}}{p_1 - p_0}\right]^2$$

Since it is unlikely that the two stage sample size will be less MIN, this provides a reasonable starting point for a search for N. **However, experience has shown that you should use a small number such as 2 to insure that you obtain the optimum.**

You can also enter a value like MIN-x where x is a positive integer. This will cause the search to begin x units below the MIN.

The problem here is that this procedure may take a long time to run. Specifying a good starting value significantly reduces the running time.

Examples of valid entries are

2, 10, 20, MIN, MIN-1, MIN-15.

Maximum N(Combined Sample Size)

N is the combined sample size of the two stages of the design. This parameter sets the maximum value of N used during the search. The optimum value of N should be between N Min and N Max or it will not be found.

The keyword BEST+X indicates that the search should try at least X units above the latest optimum value of N. For example, suppose the N Min is set at 10. The search algorithm begins at 10, and then continues by examining 11, 12, and so. Suppose that the search finds a candidate optimum at $N = 13$. To make sure that 13 is the optimum, the search continues on from 13 to 13+X (if, for example, $X = 5$, this value is 18). If no new optimum designs are found, the design at $N = 13$ is selected.

When using this option, X should be set large enough to guarantee that the true optimum can be found, but small enough so that the search does not take hours to complete. Our experience is that X should be greater than or equal to 8.

Examples of valid entries are for this parameter are:

20

30

BEST+8

BEST 8 (the plus sign is optional)

BEST 3

Best 4 (capitalization is not necessary)

Minimum R (Rejection Number)

R is the treatment rejection number for the combined samples. If the total number of patients responding to the treatment is not greater than R, the treatment is deemed unworthy of further study. R Min sets the lower boundary for R during the search for the optimum design. The optimum design must have an R value between R Min and R Max.

The recommended value for this parameter is zero. Its range is from zero to N.

Maximum (Rejection Number)

R is the rejection number for the combined samples. If the total number of patients responding to the treatment is not greater than R, the treatment is deemed unworthy of further study.

R Max sets the upper boundary for R during the search for the optimum design. The optimum design must have an R value between R Min and R Max.

Since the upper value is N and N is also a varying parameter, you can set this parameter to *MAX* or *MAX-X* (replacing X with an appropriate integer like 1, 2, or 3). This causes the maximum value of R to be set to the current value of *N-X* during each iteration of the search.

Minimum N1 (First Stage Sample Size)

$N1$ is the sample size of the first stage. This value sets the minimum value of $N1$ that is used during the search. The optimum value must be between N1 Min and N1 Max or it will not be found.

Although, in theory, the sample first stage design may have only 1, 2, or 3 patients, you may want to ignore such designs from consideration by setting this value to 4 or 5.

The actual range of this parameter is from 1 to N.

Maximum N1 (First Stage Sample Size)

$N1$ is the sample size of the first stage of the design. This parameter sets the maximum value of $N1$ used during the search. The optimum value of $N1$ should be between N Min and N Max or it will not be found. Although, in theory, the sample first stage design may have N-3, N-2, or N-1 patients, you may want to ignore such designs from consideration by setting this value to a smaller number.

Since the upper value is N-1 and N is also a varying parameter, you can set this parameter to MAX or MAX-X (replacing X with an appropriate value like 1 or 2). This causes the maximum value of $N1$ to be set to the current value of *N-X*.

Examples:

    10

    20

    MAX

    Max-2

    Max-4

    Step

This parameter sets the step size in the search for $N1$. Usually, you would enter a *1* here. Occasionally, you may want to increase the step size during the initial part of your search to speed up convergence. Once you have determined a likely range, you can tighten up the search boundaries and reset this value to 1.

Minimum R1 (First Stage Rejection Number)

$R1$ is the drug rejection number for the first stage. If the number of patients responding to the treatment in the first stage is not greater than $R1$, the treatment is deemed unworthy of further study. This parameter sets the lower boundary for $R1$ during the search for the optimum design. The optimum design must have an $R1$ value between R Min and R Max.

The recommended value for this parameter is zero. Its range is from zero to $N1$.

Maximum R1 (First Stage Rejection Number)

$R1$ is the rejection number for the first stage. If the number of patients responding to the treatment in the first stage is not greater than $R1$, the treatment is deemed unworthy of further study. This parameter sets the upper boundary for $R1$ in the search for the optimum design.

Since the upper value is $N1$ and $N1$ is a varying parameter, you can set this parameter to *MAX* or MAX-X (replacing *X* with an integer like 1,2, or 3). This causes the maximum value of $R1$ to be set to the current value of $N1-X$.

The valid range of $R1$ is between zero and $N1$.

Reports Tab

The Reports tab lets you specify which reports you want to view.

Which Designs Listed on Report

The number of designs displayed on the output report is controlled by this parameter. Since many thousands (often millions) of designs are considered, it is important to limit the display to those designs that are of interest.

- **Optimum only**

  This option displays the optimum, minimax, admissible, and one-stage designs. Although they are generated during the search for the optimum, all other designs are omitted from the output report.

- **Only those meeting error requirements**

  This option displays all designs that meet the Type-I (alpha) and Type-II (1 - power) requirements. Since there may be thousands of designs that meet these requirements, you should only select this option when you are sure that you want to see all the detail.

- **All designs**

  This option displays all designs considered. The output report may be very large, so this option should only be used when you are sure you want to see all the details.

Maximum Lines on Report

This amount limits the number of designs listed on the report. It does not limit the which designs are considered during the search.

About 50 designs are shown per page, so 1000 lines will result in a 50 page report.

Range

10 to 5000.

Example 1 – Finding the Optimal Designs

Suppose a design is wanted for the case alpha = 0.05, power = 0.80, *P0* = 0.1, and *P1* = 0.25. This would be set up as follows.

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 then 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	
Search Type ....**Automatic**	
Power (Probability of Rejecting H0 P ≤ P0 | P = P1)............**0.80**	
Alpha (Probability of Rejecting H0 P ≤ P0 | P = P0).............**0.05**	
P0 (Maximum Response Rate of a Poor Treatment) **0.10**	
P1 (Minimum Response Rate of a Good Treatment)...........**0.25**	
Multiplier for Upper Search Limit of N..................................**1.25**	

Annotated Output

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

Numeric Results

Designs ——
Proportions: P0 = 0.1000, P1 = 0.2500
Targets: Alpha = 0.0500, Power = 0.8000
N Search Range: Min of 39 to Max of 50

Design Type	Total Sample Size N	Stage 1 Sample Size N1	Stage 2 Sample Size N2	Stage 1 Critical Value R1	Stage 2 Critical Value R	Expected Sample Size E(N)	Prob Stop Early PET	Actual Alpha	Actual Power
Single Stage	40	40		7				0.0419	0.8180
Min N	40	22	18	2	7	28.84	0.6200	0.0398	0.8032
Admissible	41	15	26	1	7	26.72	0.5490	0.0430	0.8029
Admissible	42	14	28	1	7	25.63	0.5846	0.0464	0.8042
Min E(N)	43	18	25	2	7	24.66	0.7338	0.0480	0.8003

References

Simon, Richard. 'Optimal Two-Stage Designs for Phase II Clinical Trials', Controlled Clinical Trials, 1989, Volume 10, pages 1-10.

Jung, S.H., Lee, T., Kim, K.M., and George, S.L. 'Admissible two-stage designs for phase II cancer clinical trials', Statistics in Medicine, 2004, Volume 23, pages 561-569.

Report Definitions

Single Stage is regular, no-interim look design. It is provided for reference.

Min N is the design with the smallest N. Simon called this the 'Minimax' design.

Min E(N) is the design with the smallest E(N). Simon called this the 'Optimum' design.

Admissible is a compromise design developed using a Bayesian argument by Jung (2004). It has properties between Simon's Minimax and Optimum designs.

α, β met is a design that meets both the Type-I and Type-II error requirements, but is not optimal or admissible.

N is the combined sample size of both stages.

N1 is the sample size in the first stage.

N2 is the sample size in the second stage.

R1 is the critical value of the first stage. Stop trial if response ≤ R1. Continue trial if response is > R1.

R is the critical value after both stages. Reject H0: P ≤ P0 if response ≤ R.

E(N) is the expected (average) sample size if this design were repeated many times. It is calculated assuming P = P0.

PET is the probability of early termination of the study. It is the probability that the design will be stopped after
    stage 1. It is calculated assuming P = P0.
Actual Alpha is the probability of rejecting H0: P ≤ P0 when this is true.
Actual Power is the probability of rejecting H0: P ≤ P0 when P = P1 (H0 is false).
P0 is the maximum response proportion that will be considered as POOR (Not worth pursuing).
P1 is the minmum response proportion that will still be considered GOOD (Worth pursuing).

Summary Statements ───
The two-stage design with the smallest expected sample size to test the null hypothesis that P ≤ 0.1000 versus
the alternative that P > 0.1000 has an expected sample size of 24.66 and a probability of early termination of
0.7338. The power is computed at P = 0.2500 to be 0.8003. If the treatment is actually not effective, there is a
0.0480 probability of concluding that it is (the target for this value was 0.0500). If the treatment is actually
effective, there is a 0.8003 probability of concluding that it is (the target for this value was 0.8000). After testing
the treatment on 18 patients in the first stage, the trial will be terminated if 2 or fewer respond. If the trial goes on
to the second stage, a total of 43 patients will be studied. If the total number responding is less than or equal to 7,
the treatment is rejected.

This report shows three optimal designs after showing the single stage design. The first is the Min N (minimax) solution. This is the design with the smallest total sample size (*N*). The last is the Min E(N) (optimum) design—the one that minimizes the average sample size. Two admissible designs are also displayed.

Example 2 – Validation using Simon (1989)

Simon (1989) page 4 in his Table 1 presents designs for several scenarios. The first row of the table sets *P0* to 0.05, *P1* to 0.25, alpha to 0.10, and power to 0.90.

The minimax, Min N, design is *N1* = 13, *R1* = 0, *N* = 20, and *R* = 2.

The optimal, Min E(N), design is *N1* = 9, *R1* = 0, *N* = 24, and *R* = 2.

We will now run this example through **PASS**.

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. You may then 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	
Search Type ....**Automatic**	
Power (Probability of Rejecting H0 P ≤ P0 | P = P1)............**0.90**	
Alpha (Probability of Rejecting H0 P ≤ P0 | P = P0).............**0.1**	
P0 (Maximum Response Rate of a Poor Treatment) **0.05**	
P1 (Minimum Response Rate of a Good Treatment)...........**0.25**	
Multiplier for Upper Search Limit of N...................................**1.25**	

Output

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

Numeric Results

Designs ——
Proportions: P0 = 0.0500, P1 = 0.2500
Targets: Alpha = 0.1000, Power = 0.9000
N Search Range: Min of 20 to Max of 25

Design Type	Total Sample Size N	Stage 1 Sample Size N1	Stage 2 Sample Size N2	Stage 1 Critical Value R1	Stage 2 Critical Value R	Expected Sample Size E(N)	Prob Stop Early PET	Actual Alpha	Actual Power
Single Stage	20	20		2				0.0755	0.9087
Min N	20	13	7	0	2	16.41	0.5133	0.0736	0.9030
Admissible	21	11	10	0	2	15.31	0.5688	0.0784	0.9054
Admissible	22	10	12	0	2	14.82	0.5987	0.0831	0.9050
Min E(N)	24	9	15	0	2	14.55	0.6302	0.0931	0.9028

PASS calculates the same Min N and Min E(N) designs.

Example 3 – Using Custom Search

This example will show the use of the custom search option. In this example, we will find that the two optimal Simon designs are not of practical use. We will use the custom search option to find a more appealing design.

Suppose a design is wanted for the case alpha = 0.05, power = 0.80, $P0 = 0.7$, and $P1 = 0.9$. This would be set up as follows.

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 then make the appropriate entries as listed below, or open **Example 3a** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Design Tab	
Search Type	**Automatic**
Power (Probability of Rejecting H0 P ≤ P0 | P = P1)	**0.8**
Alpha (Probability of Rejecting H0 P ≤ P0 | P = P0)	**0.050**
P0 (Maximum Response Rate of a Poor Treatment)	**0.7**
P1 (Minimum Response Rate of a Good Treatment)	**0.9**
Multiplier for Upper Search Limit of N	**1.25**

Output

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

Designs ——
Proportions: P0 = 0.7000, P1 = 0.9000
Targets: Alpha = 0.0500, Power = 0.8000
N Search Range: Min of 24 to Max of 35

Design Type	Total Sample Size N	Stage 1 Sample Size N1	Stage 2 Sample Size N2	Stage 1 Critical Value R1	Stage 2 Critical Value R	Expected Sample Size E(N)	Prob Stop Early PET	Actual Alpha	Actual Power
Single Stage	28	28		23				0.0474	0.8579
Min N	26	23	3	19	21	23.16	0.9462	0.0453	0.8010
Min E(N)	27	6	21	4	22	14.82	0.5798	0.0492	0.8042

Notice that neither of the two optimal designs is of practical use. The Min N (or minimax) design sets N1 at 23 and N at 26. Thus, this design does not call for the interim analysis until 88% of the subjects are observed. Not many researchers would adopt such a design.

The Min E(N) (or optimum) design is also disappointing for the opposite reason. In this design, N1 is 6 and N is 27. Thus, only 22% of the subjects are observed before the interim analysis is planned. Not many researchers would adopt this design either.

And so, we are stuck with two designs that both have obvious flaws.

We will now conduct a search of near optimal designs that have more favorable properties. This is done as follows.

Suppose a design is wanted for the case alpha = 0.05, power = 0.80, $P0 = 0.7$, and $P1 = 0.9$. This would be set up as follows.

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 then make the appropriate entries as listed below, or open **Example 3b** by going to the **File** menu and choosing **Open Example Template**.

Option   **Value**

Design Tab

Search Type ....**Custom**

Power (Probability of Rejecting H0 P ≤ P0 | P = P1)............**0.8**

Alpha (Probability of Rejecting H0 P ≤ P0 | P = P0).............**0.050**

P0 (Maximum Response Rate of a Poor Treatment) **0.7**

P1 (Minimum Response Rate of a Good Treatment) **0.9**

N (Combined Sample Size) Min **26**

N (Combined Sample Size) Max **27**

R (Rejection Number) Min **4**

R (Rejection Number) Max **27**

N1 (First Stage Sample Size) Min **12**

N1 (First Stage Sample Size) Max **15**

R1 (First Stage Rejection Number) Min...............................**2**

R1 (First Stage Rejection Number) Max..............................**14**

Output

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

Designs ──
Proportions: P0 = 0.7000, P1 = 0.9000
Targets: Alpha = 0.0500, Power = 0.8000
N Search Range: Min of 26 to Max of 27

Design Type	Total Sample Size N	Stage 1 Sample Size N1	Stage 2 Sample Size N2	Stage 1 Critical Value R1	Stage 2 Critical Value R	Expected Sample Size E(N)	Prob Stop Early PET	Actual Alpha	Actual Power
Single Stage	28	28		23				0.0474	0.8579
Min N & E(N)	27	12	15	9	22	15.79	0.7472	0.0495	0.8223
α, β met	27	13	14	10	22	15.83	0.7975	0.0472	0.8088

Now we have two designs to choose from. Both designs have similar total sample sizes of 27. The Min E(N) design had an E(N) of 14.82. These designs only increase the expected sample size by one. However, now N1 is 12 or 13 which translates to an interim analysis at about 50% of the total sample. We would recommend the second since it is nearly balanced.