Chapter 270

Tests for One-Sample Sensitivity and Specificity

Introduction

The power analysis of a diagnostic test is often based on the sensitivity and specificity of the test. In such a test, the outcome of the diagnostic screening test is compared to the gold standard. In the common case-control study, the gold standard must be known before. In a prospective study, the gold standard is determined subsequent to the study, so the case-control framework is not appropriate. In this procedure, the power analysis and sample size requirements of such a design are considered.

In a prospective study, a group of *N* subjects is obtained. Some of the subjects have the disease (or the condition of interest) and some do not. Each subject is given the diagnostic test for the disease. Subsequently, a gold standard test is used to obtain the true presence or absence of the disease. The gold standard may be a more expensive test, or it may be following the subject to determine if the disease status becomes more apparent.

The measures of diagnostic accuracy are sensitivity and specificity. *Sensitivity* (*Se*) is the probability that the diagnostic test is positive for the disease, given that the subject actually has the disease. *Specificity* (*Sp*) is the probability that the diagnostic test is negative, given that the subject does not have the disease. Mathematically,

$$\text{Sensitivity } (Se) = \Pr(+\text{Test}|\text{Disease})$$

$$\text{Specificity } (Sp) = \Pr(-\text{Test}|\text{No Disease})$$

Li and Fine (2004) present sample size methodology for testing sensitivity and specificity using a prospective design. Their methodology will be used here. Other useful references are Obuchowski and Zhou (2002), Machin, Campbell, Tan, and Tan (2009), and Zhou, Obuchowski, and McClish (2002).

Prospective Study Design

In a prospective study, a group of N subjects is split into two groups: those with the disease of interest and those without it. Suppose a particular sample has N_D with the disease and N_{ND} without the disease. A diagnostic test is administered to each subject (usually before the disease status is determined) and its output is recorded. The diagnostic test outcome is either positive or negative for the disease. Suppose that of the N_D subjects with the disease, s_+ have a positive test outcome and s_- have a negative outcome. Similarly, of the N_{ND} subjects without the disease, r_+ have positive outcomes and r_- have negative outcomes.

Sensitivity is estimated by

$$Se = \frac{s_+}{N_D},$$

and *specificity* is estimated by

$$Sp = \frac{r_-}{N_{ND}}.$$

A useful diagnostic test has high values of both *Se* and *Sp*.

Sensitivity Hypotheses

Conditional on the value of N_D, s_+ is distributed as $\text{Binomial}(N_D, Se)$. Thus, tests of the two-sided statistical hypotheses for sensitivity,

$$H_0: Se = Se_0 \quad \text{or} \quad H_0: Se - Se_0 = 0$$

$$H_1: Se \neq Se_0 \quad \text{or} \quad H_1: Se - Se_0 \neq 0$$

can be carried out using a binomial test.

The upper one-sided null and alternative hypotheses are

$$H_0: Se \leq Se_0 \quad \text{or} \quad H_0: Se - Se_0 \leq 0$$

$$H_1: Se > Se_0 \quad \text{or} \quad H_1: Se - Se_0 > 0$$

The lower one-sided null and alternative hypotheses are

$$H_0: Se \geq Se_0 \quad \text{or} \quad H_0: Se - Se_0 \geq 0$$

$$H_1: Se < Se_0 \quad \text{or} \quad H_1: Se - Se_0 < 0$$

The power analysis for the sensitivity test is based on the binomial distribution, conditional on the value of N_D.

Specificity Hypotheses

Conditional on the value of N_{ND}, r_- is distributed as $\text{Binomial}(N_{ND}, Sp)$. Thus, tests of the two-sided statistical hypotheses for specificity,

$$H_0: Sp = Sp_0 \quad \text{or} \quad H_0: Sp - Sp_0 = 0$$

$$H_1: Sp \neq Sp_0 \quad \text{or} \quad H_1: Sp - Sp_0 \neq 0$$

can be carried out using a binomial test.

The upper one-sided null and alternative hypotheses are

$$H_0: Sp \leq Sp_0 \quad \text{or} \quad H_0: Sp - Sp_0 \leq 0$$

$$H_1: Sp > Sp_0 \quad \text{or} \quad H_1: Sp - Sp_0 > 0$$

The lower one-sided null and alternative hypotheses are

$$H_0: Sp \geq Sp_0 \quad \text{or} \quad H_0: Sp - Sp_0 \geq 0$$

$$H_1: Sp < Sp_0 \quad \text{or} \quad H_1: Sp - Sp_0 < 0$$

The power analysis for the specificity test is based on the binomial distribution, conditional on the value of N_{ND}.

Power and Sample Size Calculations using the Binomial Model

A binomial variable should exhibit the following four properties:

1. The variable is binary --- it can take on one of two possible values.
2. The variable is observed a known number of times. Each observation or replication is called a Bernoulli trial. The number of replications is *n*. The number of times that the outcome of interest is observed is *r*. Thus, *r* takes on the possible values 0, 1, 2, ..., *n*.
3. The probability, π, that the outcome of interest occurs is constant for each trial.
4. The trials are independent. The outcome of one trial does not influence the outcome of the any other trial.

The binomial probability is calculated using the formula

$$b(x|n,\pi) = \binom{n}{x}\pi^x(1-\pi)^{n-x} \text{ where } \binom{n}{r} = \frac{n!}{r!\,(n-r)!}.$$

Using the binomial probability formula, the sample size necessary to meet both a significance level and a power requirement for the sensitivity test may be found by solving the following to equations simultaneously:

$$B(s_+ > s_\alpha|N_D, Se_0) = \alpha \quad \text{(Significance Level)}$$

$$B(s_+ > s_\alpha|N_D, Se_1) = 1 - \beta$$

where

$$B(s_+ > s_\alpha|N_D, \pi) = \sum_{x=s_\alpha+1}^{N_D} b(x|N_D, Se_0)$$

A similar calculation can be made for the specificity by substituting N_{ND} for N_D, r_- for s_+, r_α for s_α, Sp_0 for Se_0, and Sp_1 for Se_1. Further details of this procedure are given in the Tests for One Proportion chapter.

Note that these formulas give N_D (or N_{ND}), not N. To obtain the estimate of the total sample size, N, using N_D, we inflate N_D by the disease prevalence, P (the proportion of diseased subjects in the population), using the equation

$$N = \frac{N_D}{P},$$

This is called Method 0 in the paper by Li and Fine (2004). We can calculate N_D from N by rearranging this formula as

$$N_D = N \times P.$$

Similarly, to obtain the estimate of the total sample size, N, using N_{ND}, we inflate N_{ND} by one minus the disease prevalence, $1 - P$ (the proportion of non-diseased subjects in the population), using the equation

$$N = \frac{N_{ND}}{(1 - P)},$$

We can calculate N_{ND} from N by rearranging this formula as

$$N_{ND} = N \times (1 - P).$$

Example 1 – Finding the Power

Suppose that diagnosing a certain type of cancer has required expensive and invasive test procedure. The sensitivity of this procedure is 72% and the specificity is 82%. A new diagnostic test has been developed that is much less expensive and invasive. The developers of the test want to design a prospective study to compare the old and new tests using a two-sided binomial test with a significance level of 0.05.

They want to consider changes in sensitivity of 10%, 15%, 20%, and 25%. These changes translate to sensitivities of 78.10%, 81.65%, 85.20%, and 88.75%. The prevalence of the disease in the population of interest is 6%. The power will be determined for trials with sample sizes between 300 and 3000 incremented by 300. They want to consider a 10% increase in specificity which is 90.2%.

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	**Power**
Alternative Hypothesis for Sensitivity Test	**Two-Sided (H1: Se ≠ Se0)**
Alternative Hypothesis for Specificity Test	**Two-Sided (H1: Sp ≠ Sp0)**
Alpha	**0.05**
N (Number of Subjects)	**300 to 3000 by 300**
P (Disease Prevalence)	**0.06**
Se0 (Null Sensitivity)	**0.71**
Se1 (Actual Sensitivity)	**0.7810 0.8165 0.8520 0.8875**
Sp0 (Null Specificity)	**0.82**
Sp1 (Actual Specificity)	**0.902**

Output

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

Numeric Reports

Numeric Results

Solve For: Power
Sensitivity Hypotheses: Two-Sided (H0: Se = Se0 vs. H1: Se ≠ Se0)
Specificity Hypotheses: Two-Sided (H0: Sp = Sp0 vs. H1: Sp ≠ Sp0)
Test Statistic: Binomial Test

Power		Number of Subjects				Sensitivity		Specificity		Alpha		
											Actual	
Sensitivity	Specificity	Total N	Diseased N_D	Non-Diseased N_{ND}	Disease Prevalence P	Null Se0	Actual Se1	Null Sp0	Actual Sp1	Target	Sensitivity	Specificity
0.07259	0.97191	300	18	282	0.06	0.71	0.7810	0.82	0.902	0.05	0.03400	0.03596
0.08020	0.99992	600	36	564	0.06	0.71	0.7810	0.82	0.902	0.05	0.02558	0.04252
0.13556	1.00000	900	54	846	0.06	0.71	0.7810	0.82	0.902	0.05	0.03494	0.04883
0.26447	1.00000	1200	72	1128	0.06	0.71	0.7810	0.82	0.902	0.05	0.03848	0.04386
0.29250	1.00000	1500	90	1410	0.06	0.71	0.7810	0.82	0.902	0.05	0.03617	0.04811
0.31430	1.00000	1800	108	1692	0.06	0.71	0.7810	0.82	0.902	0.05	0.03328	0.04620
0.41442	1.00000	2100	126	1974	0.06	0.71	0.7810	0.82	0.902	0.05	0.03917	0.04632
0.42450	1.00000	2400	144	2256	0.06	0.71	0.7810	0.82	0.902	0.05	0.03422	0.04849
0.50881	1.00000	2700	162	2538	0.06	0.71	0.7810	0.82	0.902	0.05	0.04597	0.04663
0.51251	1.00000	3000	180	2820	0.06	0.71	0.7810	0.82	0.902	0.05	0.03983	0.04711
0.13169	0.97191	300	18	282	0.06	0.71	0.8165	0.82	0.902	0.05	0.03400	0.03596
0.18430	0.99992	600	36	564	0.06	0.71	0.8165	0.82	0.902	0.05	0.02558	0.04252
0.32069	1.00000	900	54	846	0.06	0.71	0.8165	0.82	0.902	0.05	0.03494	0.04883
0.54764	1.00000	1200	72	1128	0.06	0.71	0.8165	0.82	0.902	0.05	0.03848	0.04386
0.61590	1.00000	1500	90	1410	0.06	0.71	0.8165	0.82	0.902	0.05	0.03617	0.04811
0.66986	1.00000	1800	108	1692	0.06	0.71	0.8165	0.82	0.902	0.05	0.03328	0.04620
0.78448	1.00000	2100	126	1974	0.06	0.71	0.8165	0.82	0.902	0.05	0.03917	0.04632
0.81141	1.00000	2400	144	2256	0.06	0.71	0.8165	0.82	0.902	0.05	0.03422	0.04849
0.87804	1.00000	2700	162	2538	0.06	0.71	0.8165	0.82	0.902	0.05	0.04597	0.04663
0.89183	1.00000	3000	180	2820	0.06	0.71	0.8165	0.82	0.902	0.05	0.03983	0.04711
0.23102	0.97191	300	18	282	0.06	0.71	0.8520	0.82	0.902	0.05	0.03400	0.03596
0.36746	0.99992	600	36	564	0.06	0.71	0.8520	0.82	0.902	0.05	0.02558	0.04252
0.59406	1.00000	900	54	846	0.06	0.71	0.8520	0.82	0.902	0.05	0.03494	0.04883
0.82890	1.00000	1200	72	1128	0.06	0.71	0.8520	0.82	0.902	0.05	0.03848	0.04386
0.88997	1.00000	1500	90	1410	0.06	0.71	0.8520	0.82	0.902	0.05	0.03617	0.04811
0.92821	1.00000	1800	108	1692	0.06	0.71	0.8520	0.82	0.902	0.05	0.03328	0.04620
0.97130	1.00000	2100	126	1974	0.06	0.71	0.8520	0.82	0.902	0.05	0.03917	0.04632
0.98094	1.00000	2400	144	2256	0.06	0.71	0.8520	0.82	0.902	0.05	0.03422	0.04849
0.99252	1.00000	2700	162	2538	0.06	0.71	0.8520	0.82	0.902	0.05	0.04597	0.04663
0.99496	1.00000	3000	180	2820	0.06	0.71	0.8520	0.82	0.902	0.05	0.03983	0.04711
0.38294	0.97191	300	18	282	0.06	0.71	0.8875	0.82	0.902	0.05	0.03400	0.03596
0.61870	0.99992	600	36	564	0.06	0.71	0.8875	0.82	0.902	0.05	0.02558	0.04252
0.85189	1.00000	900	54	846	0.06	0.71	0.8875	0.82	0.902	0.05	0.03494	0.04883
0.97152	1.00000	1200	72	1128	0.06	0.71	0.8875	0.82	0.902	0.05	0.03848	0.04386
0.98920	1.00000	1500	90	1410	0.06	0.71	0.8875	0.82	0.902	0.05	0.03617	0.04811
0.99584	1.00000	1800	108	1692	0.06	0.71	0.8875	0.82	0.902	0.05	0.03328	0.04620
0.99927	1.00000	2100	126	1974	0.06	0.71	0.8875	0.82	0.902	0.05	0.03917	0.04632
0.99971	1.00000	2400	144	2256	0.06	0.71	0.8875	0.82	0.902	0.05	0.03422	0.04849
0.99995	1.00000	2700	162	2538	0.06	0.71	0.8875	0.82	0.902	0.05	0.04597	0.04663
0.99998	1.00000	3000	180	2820	0.06	0.71	0.8875	0.82	0.902	0.05	0.03983	0.04711

Sensitivity	The proportion of diseased subjects that yield a positive test result. Sensitivity = Pr(+Test|Disease).
Specificity	The proportion of non-diseased subjects that yield a negative test result. Specificity = Pr(-Test|No Disease).
Power	The probability of rejecting a false null hypothesis when the alternative hypothesis is true. The power for the sensitivity test is computed using the N_D diseased subjects. The power for the specificity test is computed using the N_{ND} non-diseased subjects.
N	The total number of subjects in the study. N = N_D + N_{ND}. N = N_D / P. N = N_{ND} / (1 - P).
N_D	The number of diseased subjects in the study. N_D = N × P.
N_{ND}	The number of non-diseased subjects in the study. N_{ND} = N × (1 - P).
P	Disease Prevalence. The proportion of individuals in the population that have the disease (or condition).
Se0	The sensitivity under the null hypothesis (H0).
Se1	The sensitivity under the alternative hypothesis (H1) at which power is calculated.
Sp0	The specificity under the null hypothesis (H0).
Sp1	The specificity under the alternative hypothesis (H1) at which power is calculated.
Target Alpha	The alpha (probability of rejecting H0 when H0 is true) that was desired.
Actual Alpha	The alpha that was actually achieved by the test, calculated using the binomial distribution.

Summary Statements

A prospective diagnostic test design will be used to test the sensitivity against the null value 0.71 (H0: Sensitivity = 0.71 versus H1: Sensitivity ≠ 0.71) and the specificity against null value 0.82 (H0: Specificity = 0.82 versus H1: Specificity ≠ 0.82). The sensitivity comparison will be made using a two-sided binomial test, and the specificity comparison will be made using a two-sided binomial test. Both comparisons will be made with a Type I error rate (α) of 0.05. The prevalence of the disease (or condition) in the population is assumed to be 0.06. With a total sample size of 300 (of which 18 subjects are known to have the disease (or condition) and 282 subjects are known not to have the disease (or condition)), the power to detect a sensitivity of 0.781 is 0.07259 and the power to detect a specificity of 0.902 is 0.97191.

Dropout-Inflated Sample Size

Dropout Rate	Sample Size N	Dropout-Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	300	375	75
20%	600	750	150
20%	900	1125	225
20%	1200	1500	300
20%	1500	1875	375
20%	1800	2250	450
20%	2100	2625	525
20%	2400	3000	600
20%	2700	3375	675
20%	3000	3750	750

Dropout Rate	The percentage of subjects (or items) that are expected to be lost at random during the course of the study and for whom no response data will be collected (i.e., will be treated as "missing"). Abbreviated as DR.
N	The evaluable sample size at which power is computed (as entered by the user). If N subjects are evaluated out of the N' subjects that are enrolled in the study, the design will achieve the stated power.
N'	The total number of subjects that should be enrolled in the study in order to obtain N evaluable subjects, based on the assumed dropout rate. N' is calculated by inflating N using the formula N' = N / (1 - DR), with N' always rounded up. (See Julious, S.A. (2010) pages 52-53, or Chow, S.C., Shao, J., Wang, H., and Lokhnygina, Y. (2018) pages 32-33.)
D	The expected number of dropouts. D = N' - N.

Dropout Summary Statements

Anticipating a 20% dropout rate, 375 subjects should be enrolled to obtain a final sample size of 300 subjects.

References

Obuchowski, N.A., Zhou, X.H. 2002. 'Prospective studies of diagnostic test accuracy when disease prevalence is low,' Biostatistics, Volume 3, No. 4, pages 477-492.

Li, J., Fine, J. 2004. 'On sample size for sensitivity and specificity in prospective diagnostic accuracy studies,' Statistics in Medicine, Volume 23, pages 2537-2550.

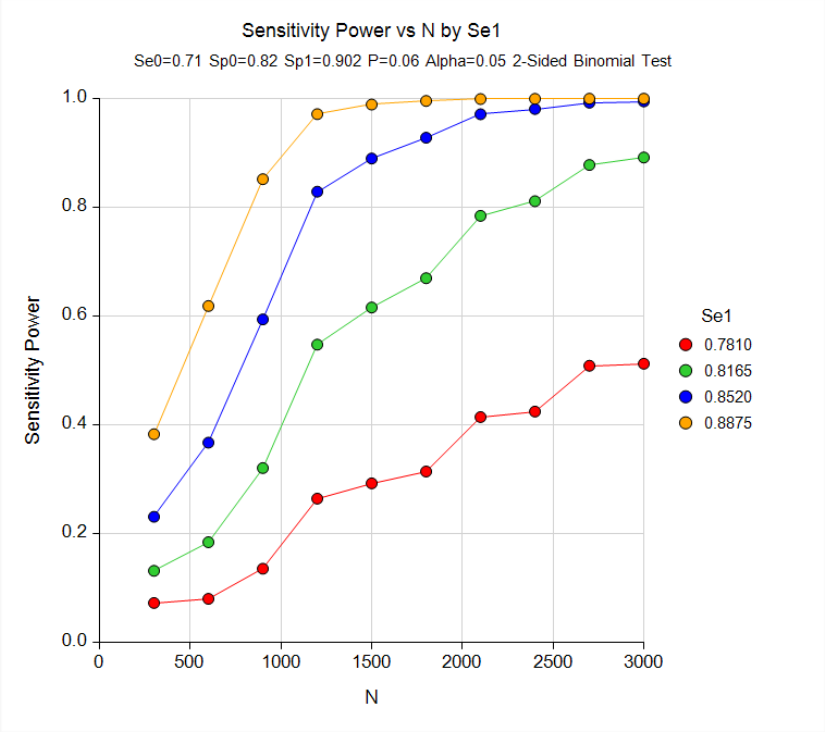
Machin, D., Campbell, M.J., Tan, S.B., Tan, S.H. 2009. Sample Size Tables for Clinical Studies, Third Edition. Wiley-Blackwell, Chichester, United Kingdom.

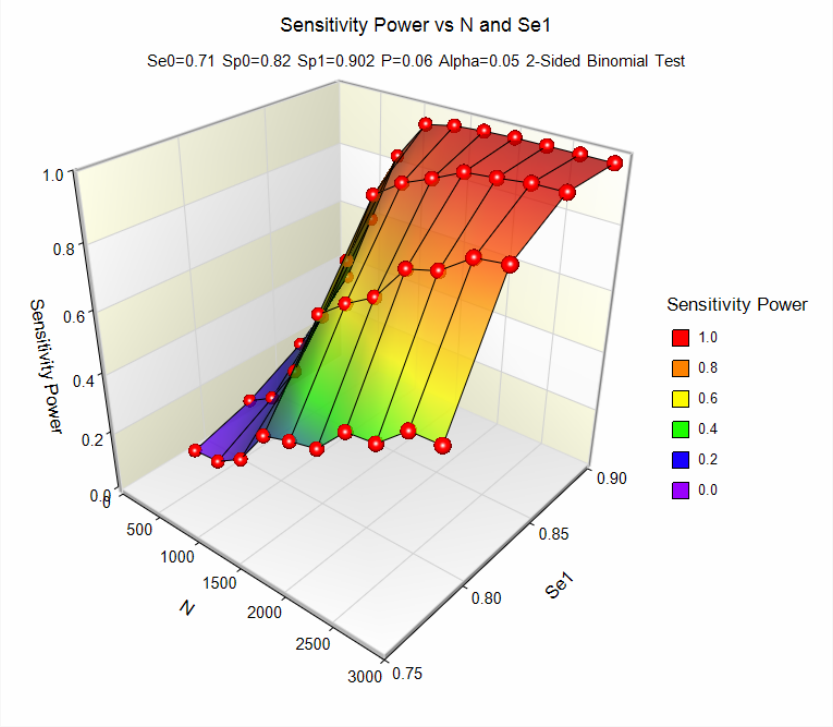
Zhou, X.H., Obuchowski, N.A., McClish, D.K. 2002. Statistical Methods in Diagnostic Medicine. Wiley-Interscience, New York.

This report shows the values of each of the parameters, one scenario per row. Because of the discrete nature of the binomial distribution, the stated (Target) alpha is usually greater than the actual alpha. Hence, we also show the Actual Alpha along with the rejection region.

Sensitivity Plots Section

Sensitivity Plots

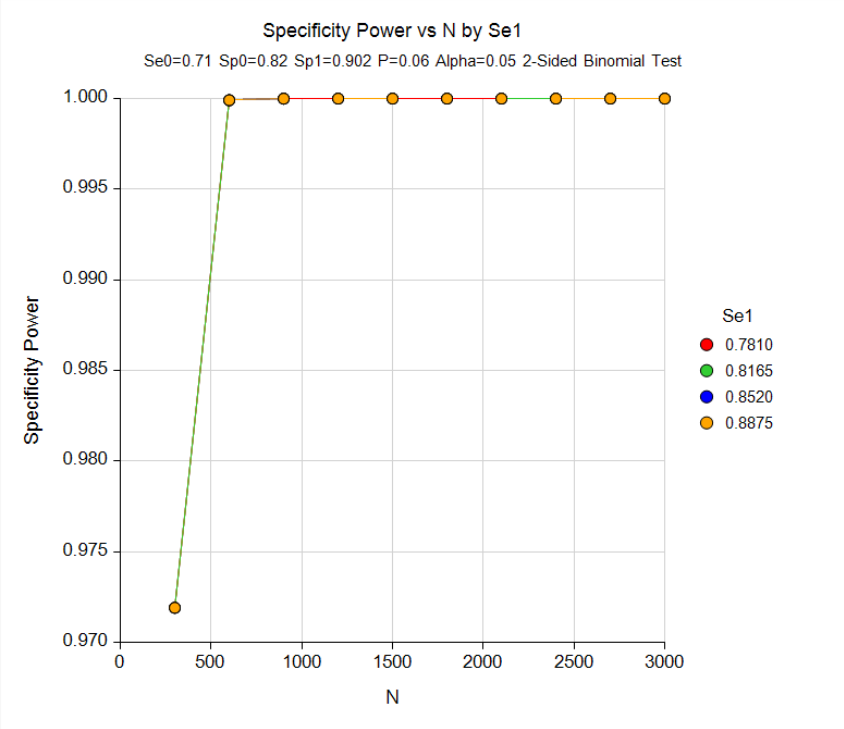


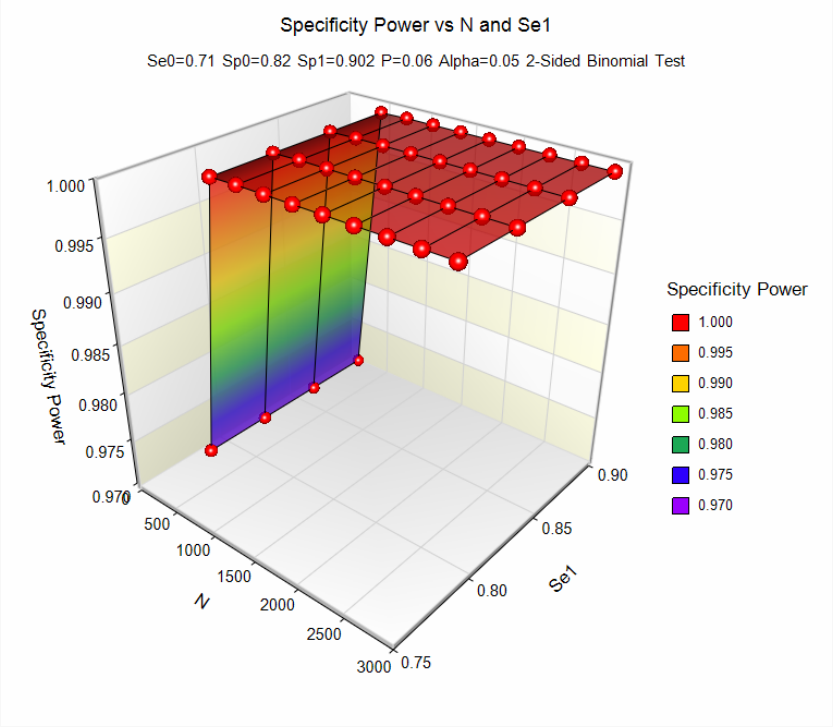


These plots show the relationship between sensitivity power, sample size, and *Se1* in this example.

Specificity Plots Section

Specificity Plots





These plots show the relationship between specificity power, sample size, and *Se1* in this example.

Example 2 – Finding the Sample Size

Continuing with Example 1, suppose you want to study the impact of various choices for Se1 on sample size. Using a significance level of 0.05 and 90% power, find the sample size when Se1 is 78.10%, 81.65%, 85.20%, and 88.75%. Assume that a two-tailed binomial test will be used.

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	**Sample Size (Sensitivity)**
Alternative Hypothesis for Sensitivity Test	**Two-Sided (H1: Se ≠ Se0)**
Alternative Hypothesis for Specificity Test	**Two-Sided (H1: Sp ≠ Sp0)**
Power	**0.90**
Alpha	**0.05**
P (Disease Prevalence)	**0.06**
Se0 (Null Sensitivity)	**0.71**
Se1 (Actual Sensitivity)	**0.7810 0.8165 0.8520 0.8875**
Sp0 (Null Specificity)	**0.82**
Sp1 (Actual Specificity)	**0.902**

Output

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

Numeric Results

Solve For:	Sample Size (Sensitivity)
Sensitivity Hypotheses:	Two-Sided (H0: Se = Se0 vs. H1: Se ≠ Se0)
Specificity Hypotheses:	Two-Sided (H0: Sp = Sp0 vs. H1: Sp ≠ Sp0)
Test Statistic:	Binomial Test

Power		Number of Subjects			Disease	Sensitivity		Specificity		Alpha		
											Actual	
Sensitivity	Specificity	Total N	Diseased N_D	Non-Diseased N_{ND}	Prevalence P	Null Se0	Actual Se1	Null Sp0	Actual Sp1	Target	Sensitivity	Specificity
0.90023	1	6683	401	6282	0.06	0.71	0.7810	0.82	0.902	0.05	0.04742	0.04692
0.90544	1	2883	173	2710	0.06	0.71	0.8165	0.82	0.902	0.05	0.04410	0.04822
0.91276	1	1550	93	1457	0.06	0.71	0.8520	0.82	0.902	0.05	0.03990	0.04790
0.91525	1	917	55	862	0.06	0.71	0.8875	0.82	0.902	0.05	0.03788	0.04133

This report shows the sample size needed to achieve 90% power for each value of Se1.

Example 3 – Validation using Li and Fine (2004)

Li and Fine (2004) page 2545 give the results of a power analysis indicate that if Se0 = 0.5, Se1 = 0.9, P = 0.01, alpha = 0.05 (one-sided), and power = 0.90, that *N* = 1100.

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 (Sensitivity)**
Alternative Hypothesis for Sensitivity Test	**One-Sided (H1: Se > Se0)**
Alternative Hypothesis for Specificity Test	**One-Sided (H1: Sp > Sp0)**
Power	**0.90**
Alpha	**0.05**
P (Disease Prevalence)	**0.01**
Se0 (Null Sensitivity)	**0.5**
Se1 (Actual Sensitivity)	**0.9**
Sp0 (Null Specificity)	**0.5**
Sp1 (Actual Specificity)	**0.9**

Output

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

Numeric Results

Solve For:	Sample Size (Sensitivity)
Sensitivity Hypotheses:	One-Sided (H0: Se ≤ Se0 vs. H1: Se > Se0)
Specificity Hypotheses:	One-Sided (H0: Sp ≤ Sp0 vs. H1: Sp > Sp0)
Test Statistic:	Binomial Test

Power		Number of Subjects			Disease	Sensitivity		Specificity		Alpha		
											Actual	
Sensitivity	Specificity	Total N	Diseased N_D	Non-Diseased N_{ND}	Prevalence P	Null Se0	Actual Se1	Null Sp0	Actual Sp1	Target	Sensitivity	Specificity
0.91044	1	1100	11	1089	0.01	0.5	0.9	0.5	0.9	0.05	0.03271	0.04483

PASS has also obtained an *N* of 1100.

It is interesting to note that a sample size of 1050 will also result in the identical power. This is because the sample size used in the power calculations is the number of diseased subjects, N_D, which is 11, not the total sample size, 1100. If *N* is 1050, $N_D = N \times P = 1050 \times 0.01 = 10.50$, which rounds to 11. If *N* is 1149, $N_D = N \times P = 1149 \times 0.01 = 11.49$, which also rounds to 11. Thus, all values of *N* between 1050 and 1149 will result in the same power.

Also, the value of the power in the article is 0.904, while **PASS** has obtained 0.91044. This difference arises because 0.904 is the unconditional power, while 0.91044 is the conditional power.