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 (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,

$$\text{Sensitivity} = \Pr(+\text{test} \mid \text{disease})$$

$$\text{Specificity} = \Pr(-\text{test} \mid \text{no disease})$$

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_1 with the disease and n_2 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 in the n_1 subjects with the disease, s_1 have a positive test outcome and s_2 have a negative outcome. Similarly, in the n_2 subjects without the disease r_1 have positive outcomes and r_2 have negative outcomes. Sensitivity is estimated by s_1/n_1 and specificity is estimated by r_2/n_2. A useful diagnostic test has high values of both *Se* and *Sp*.

Conditional on the values of n_1 and n_2, s_1 is $Binomial(n_1, Se)$. Thus, a one-sided test of the statistical hypothesis $H_0 : Se = Se_0$ versus $H_1 : Se = Se_1 > Se_0$ can be carried out using a binomial test. Hence, the power analysis is based on the binomial distribution conditional on the value of n_1. Similarly, the hypothesis test $H_0 : Sp = Sp_0$ versus $H_1 : Sp = Sp_1 > Sp_0$ is also based on the binomial distribution.

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, *P*, 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 may be found by solving the following to equations simultaneously:

$$\text{Significance Level } B(s_1 > s_\alpha|n_1, Se_0) = \alpha$$

$$B(s_1 > s_\alpha|n_1, Se_1) = 1 - \beta$$

where

$$B(s_1 > s_\alpha|n_1, \pi) = \sum_{x=s_\alpha+1}^{n_1} b(x|n_1, Se_0)$$

A similar calculation can be made for the specificity. Further details of this procedure are given in the Tests for One Proportion chapter.

Note that these formulas give n_1, not n. To obtain n, n_1 is inflated by the disease prevalence P to obtain $n = n_1/P$. This is called Method 0 in the paper by Li and Fine (2004).

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 71% 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 79.20%, 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**
H1 (Alternative Hypothesis) .... **H1: Se ≠ Se0, H1: Sp ≠ Sp0**
Alpha... **0.05**
n (Sample Size) .... **300 to 3000 by 300**
P (Prevalence) **0.06**
Se0 (Null Sensitivity)..................................... **0.71**
Se1 (Alternative Sensitivity) **0.7920 0.8165 0.8520 0.8875**
Sp0 (Null Specificity)..................................... **0.82**
Sp1 (Alternative Specificity) **0.902**

Plots Tab

Probabilities Decimals................................... **4**

Output

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

Numeric Reports

Numeric Results

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

Power			Sensitivity		Specificity		Alpha			
Sens.	Spec.	Sample Size N1 and N	H0 Se0	H1 Se1	H0 Sp0	H1 Sp1	Target	Sens. Actual	Spec. Actual	Prevalence P
0.0874	0.9719	18 300	0.71	0.7920	0.82	0.902	0.05	0.0340	0.0360	0.06
0.1051	0.9999	36 600	0.71	0.7920	0.82	0.902	0.05	0.0256	0.0425	0.06
0.1813	1.0000	54 900	0.71	0.7920	0.82	0.902	0.05	0.0349	0.0488	0.06
0.3429	1.0000	72 1200	0.71	0.7920	0.82	0.902	0.05	0.0385	0.0439	0.06
0.3846	1.0000	90 1500	0.71	0.7920	0.82	0.902	0.05	0.0362	0.0481	0.06
0.4183	1.0000	108 1800	0.71	0.7920	0.82	0.902	0.05	0.0333	0.0462	0.06
0.5340	1.0000	126 2100	0.71	0.7920	0.82	0.902	0.05	0.0392	0.0463	0.06
0.5526	1.0000	144 2400	0.71	0.7920	0.82	0.902	0.05	0.0342	0.0485	0.06
0.6427	1.0000	162 2700	0.71	0.7920	0.82	0.902	0.05	0.0460	0.0466	0.06
0.6530	1.0000	180 3000	0.71	0.7920	0.82	0.902	0.05	0.0398	0.0471	0.06
0.1317	0.9719	18 300	0.71	0.8165	0.82	0.902	0.05	0.0340	0.0360	0.06
0.1843	0.9999	36 600	0.71	0.8165	0.82	0.902	0.05	0.0256	0.0425	0.06
0.3207	1.0000	54 900	0.71	0.8165	0.82	0.902	0.05	0.0349	0.0488	0.06
0.5476	1.0000	72 1200	0.71	0.8165	0.82	0.902	0.05	0.0385	0.0439	0.06
0.6159	1.0000	90 1500	0.71	0.8165	0.82	0.902	0.05	0.0362	0.0481	0.06
0.6699	1.0000	108 1800	0.71	0.8165	0.82	0.902	0.05	0.0333	0.0462	0.06
0.7845	1.0000	126 2100	0.71	0.8165	0.82	0.902	0.05	0.0392	0.0463	0.06
0.8114	1.0000	144 2400	0.71	0.8165	0.82	0.902	0.05	0.0342	0.0485	0.06
0.8780	1.0000	162 2700	0.71	0.8165	0.82	0.902	0.05	0.0460	0.0466	0.06
0.8918	1.0000	180 3000	0.71	0.8165	0.82	0.902	0.05	0.0398	0.0471	0.06
0.2310	0.9719	18 300	0.71	0.8520	0.82	0.902	0.05	0.0340	0.0360	0.06
0.3675	0.9999	36 600	0.71	0.8520	0.82	0.902	0.05	0.0256	0.0425	0.06
0.5941	1.0000	54 900	0.71	0.8520	0.82	0.902	0.05	0.0349	0.0488	0.06
0.8289	1.0000	72 1200	0.71	0.8520	0.82	0.902	0.05	0.0385	0.0439	0.06
0.8900	1.0000	90 1500	0.71	0.8520	0.82	0.902	0.05	0.0362	0.0481	0.06
0.9282	1.0000	108 1800	0.71	0.8520	0.82	0.902	0.05	0.0333	0.0462	0.06
0.9713	1.0000	126 2100	0.71	0.8520	0.82	0.902	0.05	0.0392	0.0463	0.06
0.9809	1.0000	144 2400	0.71	0.8520	0.82	0.902	0.05	0.0342	0.0485	0.06
0.9925	1.0000	162 2700	0.71	0.8520	0.82	0.902	0.05	0.0460	0.0466	0.06
0.9950	1.0000	180 3000	0.71	0.8520	0.82	0.902	0.05	0.0398	0.0471	0.06
0.3829	0.9719	18 300	0.71	0.8875	0.82	0.902	0.05	0.0340	0.0360	0.06
0.6187	0.9999	36 600	0.71	0.8875	0.82	0.902	0.05	0.0256	0.0425	0.06
0.8519	1.0000	54 900	0.71	0.8875	0.82	0.902	0.05	0.0349	0.0488	0.06
0.9715	1.0000	72 1200	0.71	0.8875	0.82	0.902	0.05	0.0385	0.0439	0.06
0.9892	1.0000	90 1500	0.71	0.8875	0.82	0.902	0.05	0.0362	0.0481	0.06
0.9958	1.0000	108 1800	0.71	0.8875	0.82	0.902	0.05	0.0333	0.0462	0.06
0.9993	1.0000	126 2100	0.71	0.8875	0.82	0.902	0.05	0.0392	0.0463	0.06
0.9997	1.0000	144 2400	0.71	0.8875	0.82	0.902	0.05	0.0342	0.0485	0.06
1.0000	1.0000	162 2700	0.71	0.8875	0.82	0.902	0.05	0.0460	0.0466	0.06
1.0000	1.0000	180 3000	0.71	0.8875	0.82	0.902	0.05	0.0398	0.0471	0.06

Sensitivity Power The power of the sensitivity test. It is based on the N1 subjects that have the disease.
Specificity Power The power of the specificity test. It is based on the N2 subjects that do not have the disease.
N The total sample size of the study. It is equal to N1 + N2, where N1 = NP and N2 = N(1-P).

Se0	The sensitivity under H0. The sensitivity is the proportion of diseased subjects that yield a positive test result.
Se1	The sensitivity under H1. The sensitivity is the proportion of diseased subjects that yield a positive test result.
Sp0	The specificity under H0. The specificity is the proportion of non-diseased subjects that yield a negative test result.
Sp1	The specificity under H1. The specificity is the proportion of non-diseased subjects that yield a negative test result.
Target Alpha	The alpha (probability of rejecting H0 when H0 is true) that was desired.
Actual Sensitivity Alpha	The alpha that was actually achieved by the sensitivity test, calculated from the binomial distribution.
Actual Specificity Alpha	The alpha that was actually achieved by the specificity test, calculated from the binomial distribution.
P	The prevalence is the proportion of the population that actually has the condition (disease) of interest.

Summary Statements

A total sample size of 300 (which includes 18 subjects with the disease) achieves 9% power using a two-sided binomial test comparing sensitivities of 0.71 and 0.792 under the null and alternative hypotheses, respectively. This sample size also achieves 97% power using a two-sided binomial test comparing specificities of 0.82 and 0.902 under the null and alternative hypotheses, respectively. The target significance level is 0.05 for both tests. The actual significance level achieved by the sensitivity test is 0.034 and achieved by the specificity test is 0.036. The prevalence of the disease is 0.06.

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.

Sens. Power

This is the power of the sensitivity test. It is calculated from the binomial distribution using the N1 observations of the diseased subjects.

Spec. Power

This is the power of the specificity test. It is calculated from the binomial distribution using the N-N1 observations of the non-diseased subjects.

N

This is the total sample size of the study, n. It is equal to N1 + N2. The number of diseased subjects is N1 = NP. The number of non-diseased subjects is N2 = N(1-P).

Se0

This is the sensitivity under H0. The sensitivity is the proportion of diseased subjects that yield a positive test result.

Se1

This is the sensitivity under H1. The difference between Se1 and Se0 is the difference that is detected by the study.

Sp0

is the specificity under H0. The specificity is the proportion of non-diseased subjects that yield a negative test result.

Sp1

This is the specificity under H1. The difference between Sp1 and Sp0 is the difference that is detected by the study.

Target Alpha

This is the alpha (probability of rejecting H0 when H0 is true) that was desired. Because the binomial is a discrete distribution, the target value is seldom obtained. Rather, the actual value is lower than alpha.

Actual Sens. Alpha

This is the alpha that was actually achieved by the sensitivity test, calculated from the binomial distribution.

Actual Spec. Alpha

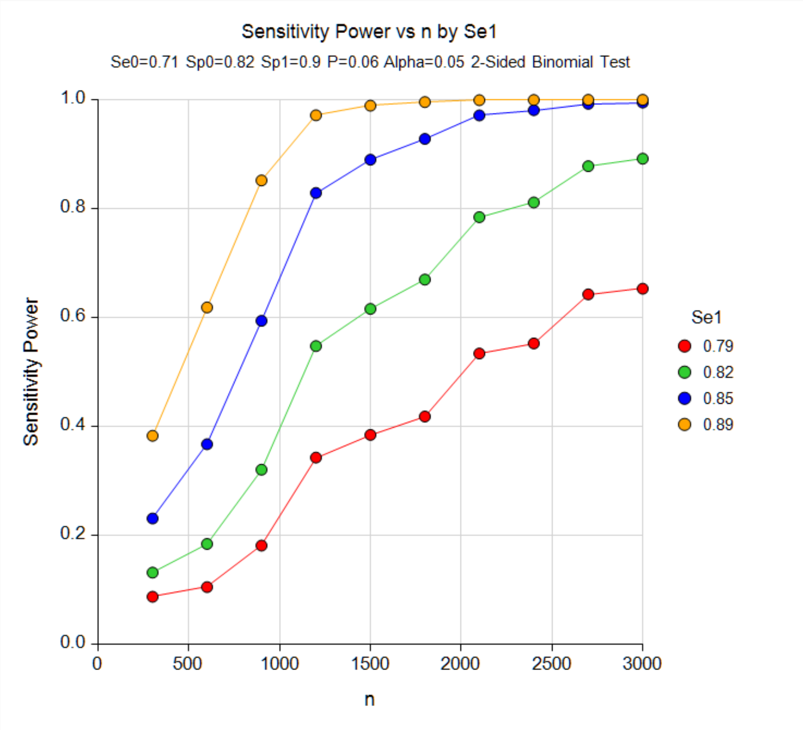
This is the alpha that was actually achieved by the specificity test, calculated from the binomial distribution.

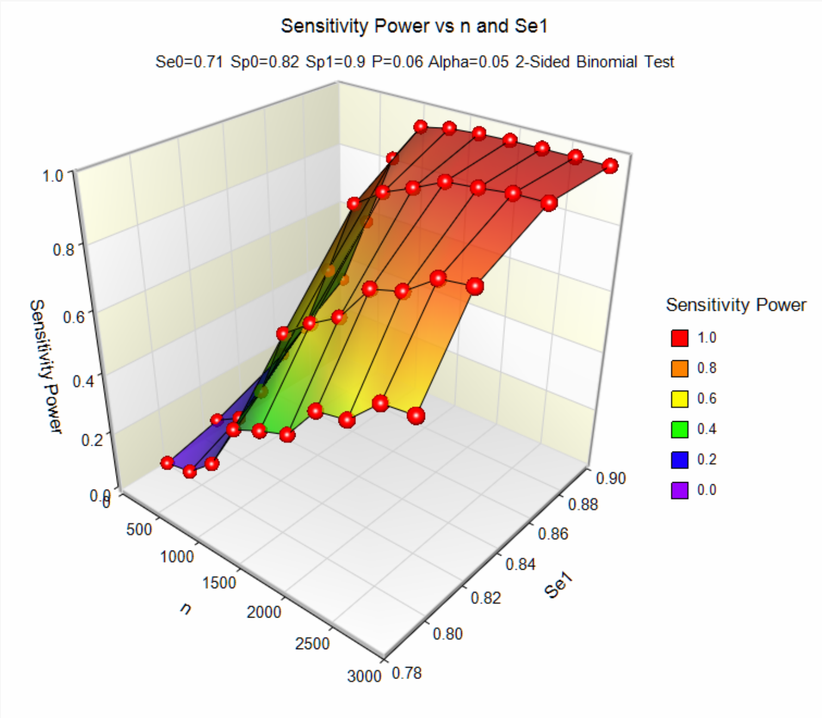
P

This is proportion of the population that actually has the condition (disease) of interest, called the prevalence.

Sensitivity Plots Section

Sensitivity Plots

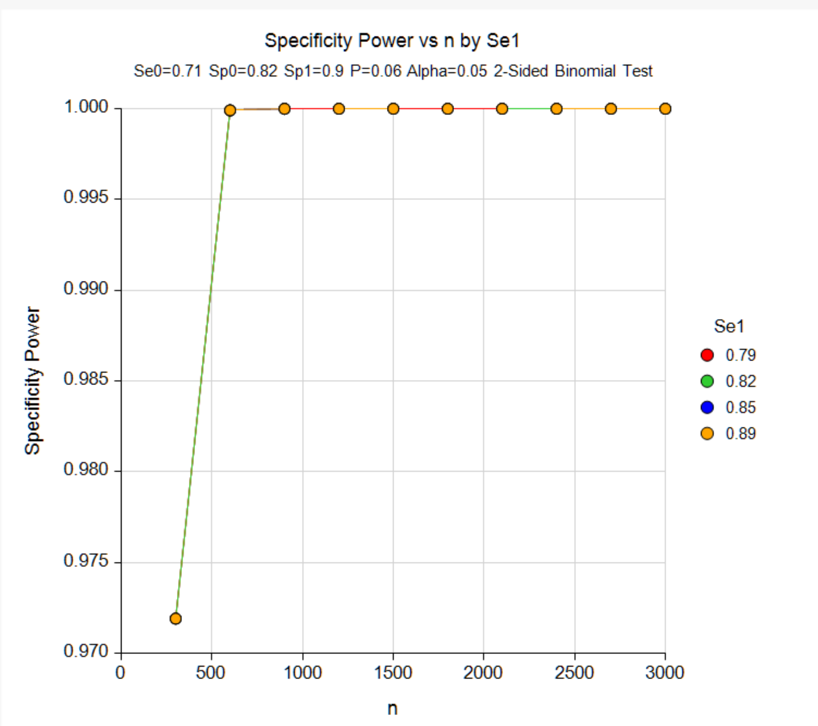


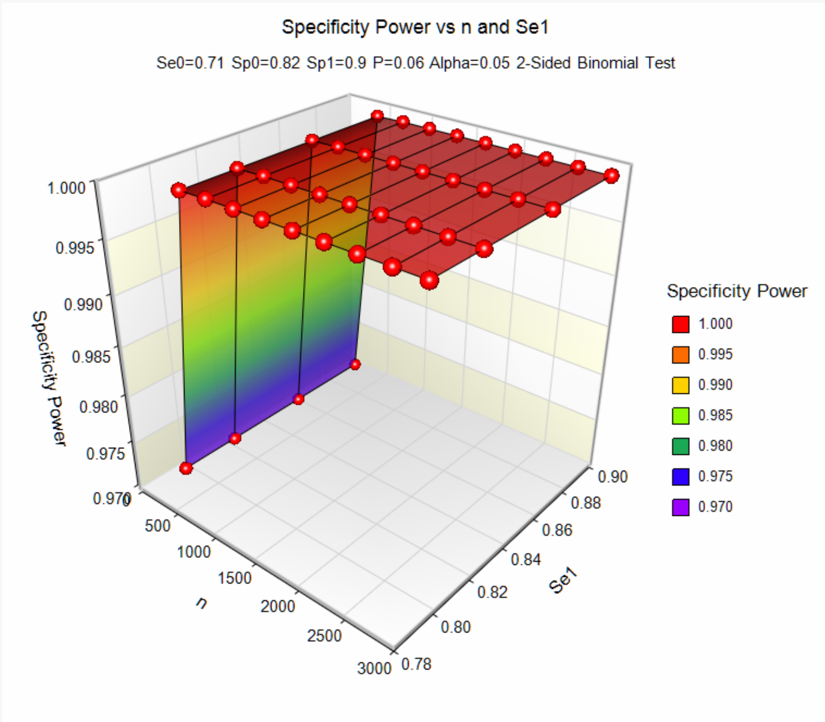


These plots show the relationship between sensitivity power, sample size, and Se1 in this example.

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 79.20%, 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)**
H1 (Alternative Hypothesis) ....**H1: Se ≠ Se0, H1: Sp ≠ Sp0**
Power...**0.90**
Alpha..**0.05**
P (Prevalence) **0.06**
Se0 (Null Sensitivity)............................**0.71**
Se1 (Alternative Sensitivity)..................**0.792 0.864 0.828 0.8875**
Sp0 (Null Specificity).............................**0.82**
Sp1 (Alternative 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: H0: Se = Se0 vs. H1: Se ≠ Se0
Specificity Hypotheses: H0: Sp = Sp0 vs. H1: Sp ≠ Sp0
Test Statistic: Binomial Test

Power			Sensitivity		Specificity		Alpha			
Sens.	Spec.	Sample Size N1 and N	H0 Se0	H1 Se1	H0 Sp0	H1 Sp1	Target	Sens. Actual	Spec. Actual	Preva-lence P
0.9060	1	299 4983	0.71	0.7920	0.82	0.902	0.05	0.0480	0.0480	0.06
0.9054	1	173 2883	0.71	0.8165	0.82	0.902	0.05	0.0441	0.0482	0.06
0.9128	1	93 1550	0.71	0.8520	0.82	0.902	0.05	0.0399	0.0479	0.06
0.9153	1	55 917	0.71	0.8875	0.82	0.902	0.05	0.0379	0.0413	0.06

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)**
H1 (Alternative Hypothesis) .... **H1: Se > Se0, H1: Sp > Sp0**
Power.. **0.90**
Alpha.. **0.05**
P (Prevalence) **0.01**
Se0 (Null Sensitivity).................................... **0.5**
Se1 (Alternative Sensitivity).......................... **0.9**
Sp0 (Null Specificity).................................... **0.5**
Sp1 (Alternative 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: H0: Se = Se0 vs. H1: Se > Se0
Specificity Hypotheses: H0: Sp = Sp0 vs. H1: Sp > Sp0
Test Statistic: Binomial Test

Power			Sensitivity		Specificity		Alpha			
Sens.	Spec.	Sample Size N1 and N	H0 Se0	H1 Se1	H0 Sp0	H1 Sp1	Target	Sens. Actual	Spec. Actual	Preva-lence P
0.9104	1	11 1100	0.5	0.9	0.5	0.9	0.05	0.0327	0.0448	0.01

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 amount of interest is N1 which is 11. If *n* is 1050, N1 = *n*P = 1050 x 0.01 = 10.5 ~ 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 0.910. This difference arises because 0.904 is the unconditional power while 0.910 is the conditional power.