

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,

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

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

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_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 .

Tests for One-Sample Sensitivity and Specificity

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

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

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).

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 the parameters associated with this test such as the sensitivities, specificities, sample sizes, alphas, and powers.

Tests for One-Sample Sensitivity and Specificity

Solve For

Solve For

This option specifies the parameter to be solved for using the other parameters. The parameters that may be selected are *Se1*, *Sp1*, *Power*, and *Sample Size*. Under most situations, you will select either *Power* or *Sample Size*.

Since this procedure actually results in two tests, you can search for *Sample Size* based on the power of the sensitivity test or the power of the specificity test.

If you want to search for *Se1* or *Sp1*, you must indicate whether to search above or below the values of *Se0* or *Sp0*, respectively.

Test

H1 (Alternative Hypothesis)

Specify the alternative hypothesis of both the sensitivity and the specificity tests. Since the null hypothesis is equality (a difference of zero), the alternative is all that needs to be specified here.

Usually, the two-sided (“≠”) option is selected.

Note that the “<” and “>” options are one-sided tests. When you choose one of these, you must make sure that the other settings (i.e. Solve For and Effect Size) are consistent with this choice.

Power and Alpha

Power

This option specifies one or more values for the desired power. Power is the probability of rejecting a false null hypothesis, and is equal to 1 - Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is commonly used.

A single value may be entered or a range of values such as *0.8 to 0.95 by 0.05* may be entered.

Alpha

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. For this procedure, a type-I error occurs when you reject the null hypothesis of equal sensitivities when in fact they are equal.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha and this is still the most common choice today.

Note that because of the discrete nature of the binomial distribution, the alpha level rarely will be achieved exactly.

A single value may be entered here or a range of values such as *0.05 to 0.2 by 0.05* may be entered.

Sample Size

n (Sample Size)

This option specifies the total number of observations in the sample. Values must be integers greater than one.

You may enter a single value or a range of values such as *10, 50, 100* or *10 to 100 by 10*.

Tests for One-Sample Sensitivity and Specificity

P (Prevalence)

Specify one or more values for the disease prevalence: the anticipated proportion of the population of interest that has the disease. Because this is a proportion all values must be between zero and one.

You may enter a single value or a range of values such as *0.1, 0.2, 0.3*.

Effect Size

Se0 (Null Sensitivity)

This is the value of the sensitivity under the null (H0) hypothesis. The sensitivity estimated from the data is compared to this value by the statistical test. Usually, this is the base line or standard value.

Sensitivity = Pr(+Test|Disease).

Range

All sensitivities must be between zero and one.

You may enter a list of values such as *0.5, 0.6, 0.7* or *0.5 to 0.8 by 0.1*.

Se1 (Alternative Sensitivity)

This is the value of the sensitivity under the alternative (H1) hypothesis.

Sensitivity = Pr(+Test|Disease).

Detectable Difference

Se1 – Se0 is the size of the difference that the study will detect.

Range

All sensitivities must be between zero and one. Also, this value cannot be set equal to Se0.

You may enter a list of values such as *0.5, 0.6, 0.7* or *0.5 to 0.8 by 0.1*.

Sp0 (Null Specificity)

This is the value of the specificity under the null (H0) hypothesis. The specificity estimated from the data is compared to this value by the statistical test. Usually, this is the base line or standard value.

Specificity = Pr(-Test|Not Disease).

Range

All specificities must be between zero and one.

You may enter a list of values such as *0.5, 0.6, 0.7* or *0.5 to 0.8 by 0.1*.

Sp1 (Alternative Specificity)

This is the value of the specificity under the alternative (H1) hypothesis.

Specificity = Pr(+Test|Disease).

Detectable Difference

Sp1 – Sp0 is the size of the difference that the study will detect.

Range

All specificities must be between zero and one. Also, this value cannot be set equal to Se0.

You may enter a list of values such as *0.5, 0.6, 0.7* or *0.5 to 0.8 by 0.1*.

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

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for One-Sample Sensitivity and Specificity** procedure window by expanding **Proportions**, clicking on **Sensitivity and Specificity**, and then clicking on **Tests for One-Sample Sensitivity and Specificity**. 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**.

<u>Option</u>	<u>Value</u>
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

Annotated Output

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

Numeric Results

Numeric Results for testing H0: Se = Se0 vs. H1: Se ≠ Se0 and H0: Sp = Sp0 vs. H1: Sp ≠ Sp0
 Test Statistic: Binomial Test

--- Power ---		Sample Size N1 and N	--- Sensitivity ---		--- Specificity ---		----- Alpha -----			Prevalence P
Sens.	Spec.		H0 Se0	H1 Se1	H0 Sp0	H1 Sp1	Target	Actual Sens.	Actual Spec.	
0.0874	0.9719	18 300	0.7100	0.7920	0.8200	0.9020	0.0500	0.0340	0.0360	0.0600
0.1051	0.9999	36 600	0.7100	0.7920	0.8200	0.9020	0.0500	0.0256	0.0425	0.0600
0.1813	1.0000	54 900	0.7100	0.7920	0.8200	0.9020	0.0500	0.0349	0.0488	0.0600
0.3429	1.0000	72 1200	0.7100	0.7920	0.8200	0.9020	0.0500	0.0385	0.0439	0.0600
0.3846	1.0000	90 1500	0.7100	0.7920	0.8200	0.9020	0.0500	0.0362	0.0481	0.0600
0.4183	1.0000	108 1800	0.7100	0.7920	0.8200	0.9020	0.0500	0.0333	0.0462	0.0600
.
.
.

Tests for One-Sample Sensitivity and Specificity

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.

Report Definitions

- Sens. Power is the power of the sensitivity test. It is based on the N_1 observations.
- Spec. Power is the power of the specificity test. It is based on the $N-N_1$ observations.
- N is the total sample size of the study. It is equal to $N_1 + N_2$.
- Se_0 is the sensitivity under H_0 . The sensitivity is the proportion of diseased subjects that yield a positive test result.
- Se_1 is the sensitivity under H_1 . The sensitivity is the proportion of diseased subjects that yield a positive test result.
- Sp_0 is the specificity under H_0 . The specificity is the proportion of non-diseased subjects that yield a negative test result.
- Sp_1 is the specificity under H_1 . The specificity is the proportion of non-diseased subjects that yield a negative test result.
- Target Alpha is the alpha (probability of rejecting H_0 when H_0 is true) that was desired.
- Actual Sens. Alpha is the alpha that was actually achieved by the sensitivity test, calculated from the binomial distribution.
- Actual Spec. Alpha is the alpha that was actually achieved by the specificity test, calculated from the binomial distribution.
- P is proportion of the population that actually has the condition (disease) of interest, called the prevalence.

Summary Statements

A total sample size of 300 (which includes 18 subjects with the disease) achieves 9% power to detect a change in sensitivity from 0.71 to 0.792 using a two-sided binomial test and 97% power to detect a change in specificity from 0.82 to 0.902 using a two-sided binomial test. The target significance level is 0.05. The actual significance level achieved by the sensitivity test is 0.0340 and achieved by the specificity test is 0.0360. The prevalence of the disease is 0.06.

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 N_1 observations of the diseased subjects.

Spec. Power

This is the power of the specificity test. It is calculated from the binomial distribution using the $N-N_1$ observations of the non-diseased subjects.

N

This is the total sample size of the study, n . It is equal to $N_1 + N_2$. The number of diseased subjects is $N_1 = NP$. The number of non-diseased subjects is $N_2 = N(1-P)$.

Se_0

This is the sensitivity under H_0 . The sensitivity is the proportion of diseased subjects that yield a positive test result.

Se_1

This is the sensitivity under H_1 . The difference between Se_1 and Se_0 is the difference that is detected by the study.

Tests for One-Sample Sensitivity and Specificity

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.

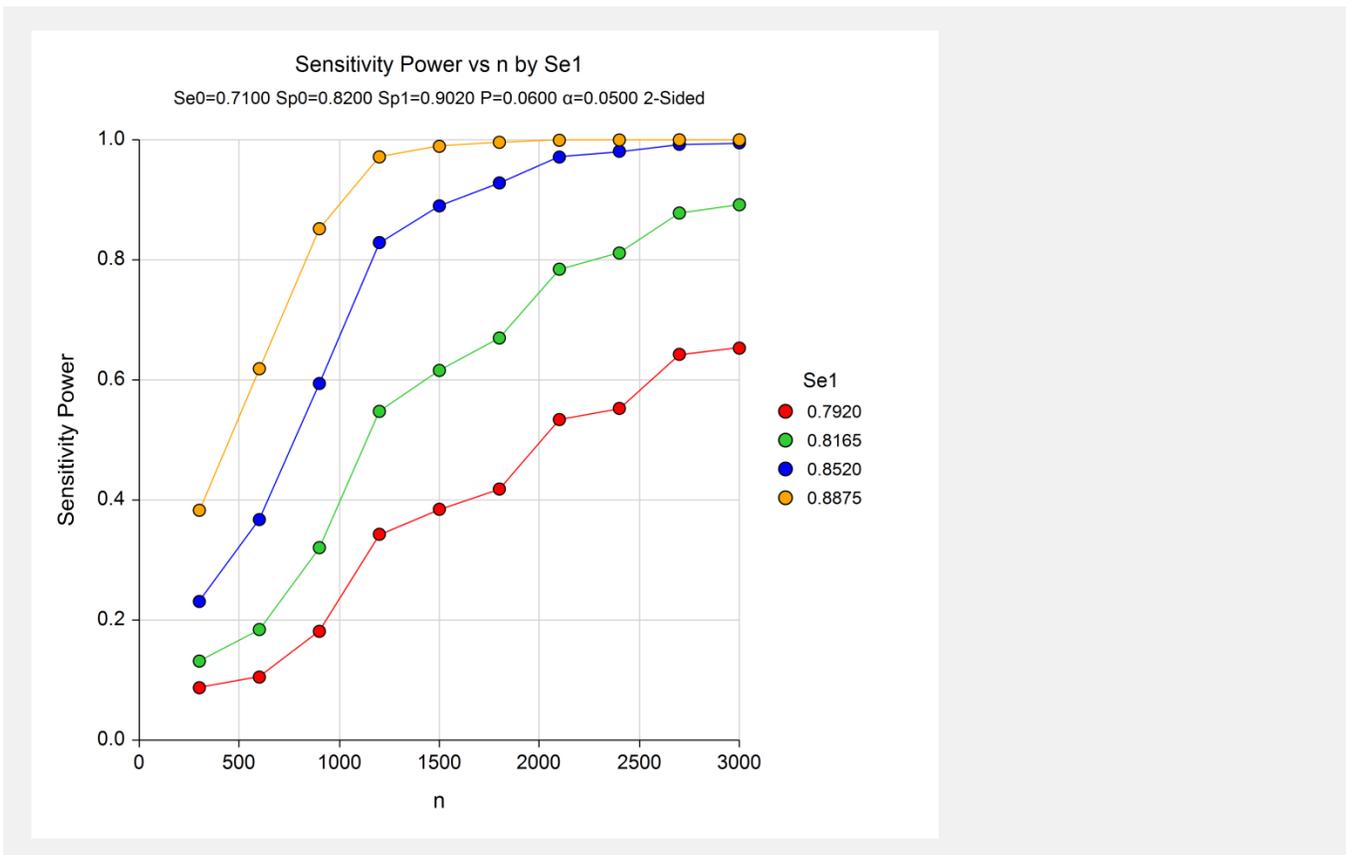
Actual Spec. Alpha

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

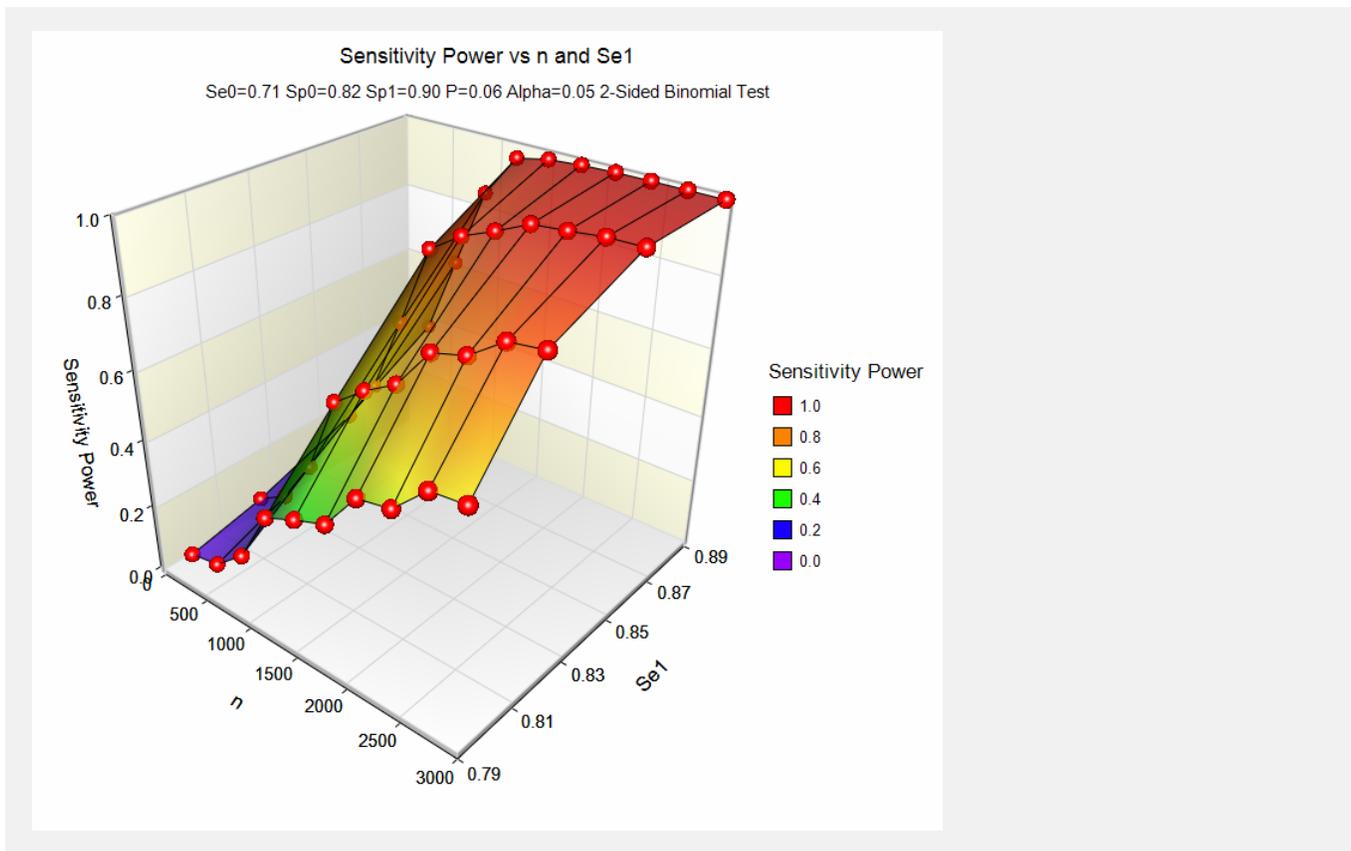
P

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

Plots Section



Tests for One-Sample Sensitivity and Specificity



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

Tests for One-Sample Sensitivity and Specificity

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

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for One-Sample Sensitivity and Specificity** procedure window by expanding **Proportions**, clicking on **Sensitivity and Specificity**, and then clicking on **Tests for One-Sample Sensitivity and Specificity**. 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**.

<u>Option</u>	<u>Value</u>
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

Annotated Output

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

Numeric Results

Numeric Results for testing H0: Se = Se0 vs. H1: Se ≠ Se0 and H0: Sp = Sp0 vs. H1: Sp ≠ Sp0										
Test Statistic: Binomial Test										
---- Power ----		Sample Size N1 and N	--- Sensitivity ---		--- Specificity ---		----- Alpha -----		Prevalence P	
Sens.	Spec.		H0 Se0	H1 Se1	H0 Sp0	H1 Sp1	Target	Sens. Actual		Spec. Actual
0.9060	1.0000	299 4983	0.7100	0.7920	0.8200	0.9020	0.0500	0.0480	0.0480	0.0600
0.9054	1.0000	173 2883	0.7100	0.8165	0.8200	0.9020	0.0500	0.0441	0.0482	0.0600
0.9128	1.0000	93 1550	0.7100	0.8520	0.8200	0.9020	0.0500	0.0399	0.0479	0.0600
0.9153	1.0000	55 917	0.7100	0.8875	0.8200	0.9020	0.0500	0.0379	0.0413	0.0600

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

Tests for One-Sample Sensitivity and Specificity

Example 3 – Validation using Li and Fine

Li and Fine (2004) page 2545 give the results of a power analysis indicate that if $Se_0 = 0.5$, $Se_1 = 0.9$, $P = 0.01$, $\alpha = 0.05$ (one-sided), and power = 0.90, that $n = 1100$.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for One-Sample Sensitivity and Specificity** procedure window by expanding **Proportions**, clicking on **Sensitivity and Specificity**, and then clicking on **Tests for One-Sample Sensitivity and Specificity**. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
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

Annotated Output

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

Numeric Results

Numeric Results for testing H0: Se = Se0 vs. H1: Se > Se0 and H0: Sp = Sp0 vs. H1: Sp > Sp0											
Test Statistic: Binomial Test											
---- Power ----		Sample Size	--- Sensitivity ---		--- Specificity ---		----- Alpha -----		Prevalence		
Sens.	Spec.	N1 and N	H0 Se0	H1 Se1	H0 Sp0	H1 Sp1	Target	Actual	Actual	P	
0.9104	1.0000	11 1100	0.5000	0.9000	0.5000	0.9000	0.0500	0.0327	0.0448	0.0100	

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, $N1 = nP = 1050 \times 0.01 = 10.5 \sim 11$. 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.