

## Chapter 741

# Non-Zero Null Tests for Simple Linear Regression

---

### Introduction

Simple linear regression is a commonly used procedure in statistical analysis to model a linear relationship between a dependent variable  $Y$  and an independent variable  $X$ . One of the main objectives in simple linear regression analysis is to test hypotheses about the slope (sometimes called the regression coefficient) of the regression equation. This module calculates power and sample size for testing whether the slope is a value other than the value specified by the null hypothesis. The conditional power calculation method is used.

---

### Difference between Simple Linear Regression and Correlation

The correlation coefficient is used when  $X$  and  $Y$  are from a bivariate normal distribution. That is,  $X$  is assumed to be a random variable whose distribution is normal. The values of  $X$  will not be known until the study is completed. In the simple linear regression context, no statement is made about the distribution of  $X$ . In fact,  $X$  does not have to be a random variable. In this procedure the distribution of  $Y$  is conditioned on  $X$ .

---

### Fixed or Random $X$

Gatsonis and Sampson (1989) present power analysis results for two approaches: *unconditional* and *conditional*. This procedure provides a calculation for the *conditional* (fixed  $X$ ) approach.

The *unconditional* approach assumes that  $X$  is normally distributed and is based on the correlation coefficient. The normality assumption might occasionally be met, but not frequently. Our impression is that usually, the values of  $X$  will not be known at the planning stage and they will not follow (even approximately) the normal distribution. Hence, the only option available is to proceed with the sample size calculation using the *conditional* approach to power calculation and estimate the standard deviation of the  $X$ 's as best you can.

## Technical Details

Suppose that the dependence of a variable  $Y$  on another variable  $X$  can be modeled using the simple linear equation

$$Y = A + BX$$

In this equation,  $A$  is the  $Y$ -intercept,  $B$  is the slope,  $Y$  is the dependent variable, and  $X$  is the independent variable.

The nature of the relationship between  $Y$  and  $X$  is studied using a sample of  $N$  observations. Each observation consists of a data pair: the  $X$  value and the  $Y$  value. The values of  $A$  and  $B$  are estimated from these observations. Since the linear equation will not fit the observations exactly, estimated values of  $A$  and  $B$  must be used. These estimates are found using the method of least squares. Using these estimated values, each data pair may be modeled using the equation

$$Y_i = a + bX_i + e_i$$

Note that  $a$  and  $b$  are the estimates of the population parameters  $A$  and  $B$ . The  $e$  values represent the discrepancies between the estimated values ( $a + bX$ ) and the actual values  $Y$ . They are called the errors or residuals.

If it is assumed that these  $e$  values are normally distributed, tests of hypotheses about  $A$  and  $B$  can be constructed. Specifically, we can employ a T-test to test the null hypothesis that the  $B$  is  $B_0$  versus the alternative hypothesis that the slope is something else.

## Linear Regression Slope T-Test Statistic

It is anticipated that a  $t$ -test of a regression coefficient will be used to conduct the test. Hence, the formula of the test statistic is

$$t_{N-2} = \frac{b - B_0}{s_b}$$

where  $N$  is the sample size,  $b$  is the estimate of  $B$ , and  $s_b$  is the standard error of  $b$ .

## Power Calculation of the Test of the Regression Coefficient, $B$

The following presentation is based on the standard results for a  $t$ -test as shown by Neter, Wasserman, and Kutner (1983) pages 71 and 72.

The power is calculated as follows for a directional alternative (one-tailed test) in which  $H_1: B > B_0$ .

1. Find  $t_{1-\alpha}$  such that  $T_{df}(t_{1-\alpha}) = 1 - \alpha$ , where  $T_{df}(x)$  is the area under a central- $t$  curve to the left of  $x$  and  $df = N - 2$ .
2. Calculate:  $X_0 = (t_{1-\alpha})\sigma_e/\sqrt{N}$ .
3. Calculate the noncentrality parameter:  $\lambda = \sqrt{N}(B_1 - B_0)\sigma_X/\sigma_e$ , where  $\sigma_X$  is the standard deviation of the  $X$  values in the regression and  $B_1$  is the slope at which the power is to be calculated.
4. Calculate:  $t_1 = (X_0 - (B_1 - B_0)\sigma_X\sqrt{N}/\sigma_e) + \lambda$ .
5. Power =  $1 - T'_{df,\lambda}(t_1)$ , where  $T'_{df,\lambda}(x)$  is the area to the left of  $x$  under a noncentral- $t$  curve with degrees of freedom  $df$  and noncentrality parameter  $\lambda$ .

The sample size can be easily found using a binary search with this power formula.

## Non-Zero Null Tests for Simple Linear Regression

### Calculation of $\sigma_x$

The above calculation requires the value of  $\sigma_x$ , the (population) standard deviation of the X values in the regression analysis. Except for the occasional experimental design that includes a specification of the X values (e.g., doses), the specific X values are unknown in the planning phase. Hence, a reasonable estimate must be found. PASS includes a special tool called the *Standard Deviation Estimator* that will aid in your search for accurate estimates of this parameter.

The following table provides examples of typical data configurations and their corresponding standard deviations.

$\sigma_x$	X Values	$\sigma_x$	X Values	$\sigma_x$	X Values	$\sigma_x$	X Values
0.500	1, 2	0.816	1, 2, 3	1.118	1, 2, 3, 4	1.414	1, 2, 3, 4, 5
1.000	1, 3	1.633	1, 3, 5	2.236	1, 3, 5, 7	2.828	1, 3, 5, 7, 9
1.500	1, 4	2.449	1, 4, 7	3.354	1, 4, 7, 10	4.243	1, 4, 7, 10, 13
2.000	1, 5	3.266	1, 5, 9	4.472	1, 5, 9, 13	5.657	1, 5, 9, 13, 17
4.000	1, 9	6.532	1, 9, 17	8.944	1, 9, 17, 25	11.314	1, 9, 17, 25, 33

Because of the direct impact on the power and sample size, it will be important to spend some time determining appropriate values for this parameter.

One final note: when a basic pattern is repeated, its population standard deviation remains the same. For example, the standard deviation of the values 1, 2, 1, 2, 1, 2, 1, 2 is 0.5. This is also the standard deviation of 1, 2 or 1, 2, 1, 2.

### 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.

#### Solve For

##### Solve For

This option specifies the parameter to be solved for from the other parameters. Under most situations, you will select either *Power* for a power analysis or *Sample Size* for sample size determination.

Select *Sample Size* when you want to calculate the sample size needed to achieve a given power and alpha level.

Select *Power* when you want to calculate the power of an experiment.

## Non-Zero Null Tests for Simple Linear Regression

---

### Test

#### Alternative Hypothesis

Specify the alternative hypothesis assessed by the test.

Let  $B$  represent the slope variable and  $B_0$  represent the specific value of the slope assumed by the null hypothesis. Possible choices are:

- **H1:  $B \neq B_0$**

This is a two-sided hypothesis. The hypothesis set is  $H_0: B = B_0$  vs.  $H_1: B \neq B_0$ .

- **H1:  $B < B_0$**

This is a lower, one-sided hypothesis. The hypothesis set is  $H_0: B \geq B_0$  vs.  $H_1: B < B_0$ .

- **H1:  $B > B_0$**

This is an upper, one-sided hypothesis. The hypothesis set is  $H_0: B \leq B_0$  vs.  $H_1: B > B_0$ .

Note that this parameter impacts the value of alpha. The value of alpha is used directly for one-sided tests. For two-sided tests, alpha is replaced by  $\alpha/2$ .

---

### Power and Alpha

#### Power

This option specifies one or more values for power. Power is the probability of rejecting a false null hypothesis and is equal to one minus 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 here 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 (alpha). A type-I error occurs when you reject the null hypothesis when in fact it is true.

Values of alpha must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as *0.01 0.05 0.10* or *0.01 to 0.10 by 0.01*.

---

### Sample Size

#### N (Sample Size)

Enter one or more values for the number of observations (e.g., subjects) in the study.

## Non-Zero Null Tests for Simple Linear Regression

---

### Effect Size – Slope

#### B0 (Slope|H0)

Enter one or more values of the slope under the null hypothesis,  $B_0$ .

In the linear equation,  $Y = A + BX$ ,  $B$  represents the slope of the line relating the dependent variable  $Y$  to the independent variable  $X$ . A  $t$ -test can be used to test whether  $B = B_0$  versus  $B \neq B_0$ .

Any value is valid as long as  $B_0 \neq B_1$ . This value must be in the same scale as  $\sigma_x$ .

#### B1 (Slope|H1)

Enter one or more values of the slope assumed by the alternative hypothesis,  $B_1$ . This represents the actual value of  $B$  at which the power is computed.

You can enter a single value such as '1' or a series of values such as '1 2 3' or '1 to 10 by 1'.

The only restriction is that  $B_1 \neq B_0$ .

---

### Effect Size – Standard Deviation of X

#### $\sigma_x$ Input Type

Select the method you want to use to enter the value(s) of  $\sigma_x$ . Your choices are

- **$\sigma_x$  (Std Dev of X)**  
Enter one or more values for  $\sigma_x$  directly.
- **List of X Values**  
Enter a list of two or more numbers from which the standard deviation is to be calculated.

#### $\sigma_x$ (Standard Deviation of X)

Enter one or more values for  $\sigma_x$ , the *population* standard deviation of the  $X$  values that will occur in a sample.

Usually, the actual  $X$  values are not known at the planning stage. When they are not known, you will have to estimate this value. You can press the *Standard Deviation Estimator* button at the right to obtain help in determining appropriate values for this parameter. Just be sure to use the *population*, not the *sample*, formula. That is, divide the sum of squares by  $N$ , not  $N-1$ .

The individual numbers can be any numeric value: positive, negative, or zero.

#### Fixed Xs

Determine the standard deviation of a typical set of fixed  $X$ s. For example, suppose the  $X$  values will be five -1's and five 1's. The population standard deviation of these values (dividing by  $N$ , not  $N - 1$ ) is 1.0. This is the value of  $\sigma_x$ . Note that '1 2' will result in the same  $\sigma_x$  as '1 2 1 2', '1 1 1 2 2 2', '-1 -2', or '11 12'.

#### Random Xs

Estimate one or more values of  $\sigma_x$ , the standard deviation of  $X$ , from your knowledge of  $X$ . If nothing else is available, you can use the likely range divided by 4, 5, or 6.

#### List of X Values

Enter a list of values from which the value of  $\sigma_x$  will be calculated. For example, entering "1, 3" results in  $\sigma_x = 1.0$ . Note that this calculation assumes that the  $N$  observations are allocated equally among the  $X$ 's.

## Non-Zero Null Tests for Simple Linear Regression

---

**Effect Size –  $\sigma_e$  (Standard Deviation of Residuals)** **$\sigma_e$  Input Type**

Select the method to use to enter  $\sigma_e$  (the standard deviation of the residuals).

- **$\sigma_e$  (Std Dev of Residuals)**  
Specify  $\sigma_e$  directly.
- **$\sigma_Y$  (Std Dev of Y)**  
Specify  $\sigma_Y$ . Calculate:  $\sigma_e^2 = \sigma_Y^2 - B1^2 (\sigma_X^2)$

 **$\sigma_e$  (Std Dev of Residuals)**

Enter one or more values of the standard deviation of the residuals from the regression of Y on X. The possible range is  $0 < \sigma_e$ .

 **$\sigma_Y$  (Std Dev of Y)**

Enter one or more values for the standard deviation of Y, ignoring the independent variable X. The value of  $\sigma_Y$  is converted to  $\sigma_e$  using the formula:  $\sigma_e^2 = \sigma_Y^2 - B1^2 (\sigma_X^2)$ . The allowable range is  $0 < |B1(\sigma_X)| < \sigma_Y$ .

## Non-Zero Null Tests for Simple Linear Regression

**Example 1 – Calculating the Power**

Suppose a power analysis is required for a simple linear regression study that will test the relationship between two variables,  $Y$  and  $X$ . The analysis will look at the power of several sample sizes between 20 and 180. A one-sided test will be used with a significance level of 0.025. Based on previous studies,  $\sigma_e$  will be assumed to be 0.6.  $\sigma_x$  will assume that  $X$  is binary with equally-likely values of -1 and 1. The experimenter decides that unless the slope is at least 0.5, the relationship between  $X$  and  $Y$  is too weak to be considered. The power will be computed at  $B1 = 0.6, 0.7, \text{ and } 0.8$ .

**Setup**

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Non-Zero Null Tests for Simple Linear Regression** 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**.

<u>Option</u>	<u>Value</u>
<b>Design Tab</b>	
Solve For .....	<b>Power</b>
Alternative Hypothesis .....	<b>One-Sided (H1: B &gt; B0)</b>
Alpha .....	<b>0.025</b>
N (Sample Size).....	<b>20 60 100 140 180</b>
B0 (Slope H0) .....	<b>0.5</b>
B1 (Slope H1) .....	<b>0.6 0.7 0.8</b>
$\sigma_x$ Input Type.....	<b>List of X Values</b>
List of X Values.....	<b>-1 1</b>
$\sigma_e$ Input Type.....	<b><math>\sigma_e</math> (Std Dev of Residuals)</b>
$\sigma_e$ (Std Dev of Residuals) .....	<b>0.6</b>

**Annotated Output**

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

<b>Numeric Results</b>									
Hypotheses: H0: B ≤ B0 vs. H1: B > B0									
Power	Sample Size N	H0 Slope B0	Actual Slope B1	Std Dev of X $\sigma_x$	Std Dev of Y $\sigma_y$	Std Dev of Resids $\sigma_e$	R <sup>2</sup>	Alpha	
0.1050	20	0.500	0.600	1.000	0.849	0.600	0.500	0.025	
0.2917	20	0.500	0.700	1.000	0.922	0.600	0.576	0.025	
0.5620	20	0.500	0.800	1.000	1.000	0.600	0.640	0.025	
0.2450	60	0.500	0.600	1.000	0.849	0.600	0.500	0.025	
0.7187	60	0.500	0.700	1.000	0.922	0.600	0.576	0.025	
0.9677	60	0.500	0.800	1.000	1.000	0.600	0.640	0.025	
0.3784	100	0.500	0.600	1.000	0.849	0.600	0.500	0.025	
0.9100	100	0.500	0.700	1.000	0.922	0.600	0.576	0.025	
0.9986	100	0.500	0.800	1.000	1.000	0.600	0.640	0.025	
0.4993	140	0.500	0.600	1.000	0.849	0.600	0.500	0.025	
0.9748	140	0.500	0.700	1.000	0.922	0.600	0.576	0.025	
1.0000	140	0.500	0.800	1.000	1.000	0.600	0.640	0.025	
0.6041	180	0.500	0.600	1.000	0.849	0.600	0.500	0.025	
0.9936	180	0.500	0.700	1.000	0.922	0.600	0.576	0.025	
1.0000	180	0.500	0.800	1.000	1.000	0.600	0.640	0.025	

## Non-Zero Null Tests for Simple Linear Regression

### References

Dupont, W.D. and Plummer, W.D. Jr. 1998. 'Power and Sample Size Calculations for Studies Involving Linear Regression'. *Controlled Clinical Trials*, Vol. 19, Pages 589-601.  
 Sampson, Allan R. 1974. 'A Tale of Two Regressions'. *JASA*, Vol. 69, No. 347, Pages 682-689.  
 Neter, J., Wasserman, W., and Kutner, M. 1983. *Applied Linear Regression Models*. Richard D. Irwin, Inc. Chicago, Illinois.

### Report Definitions

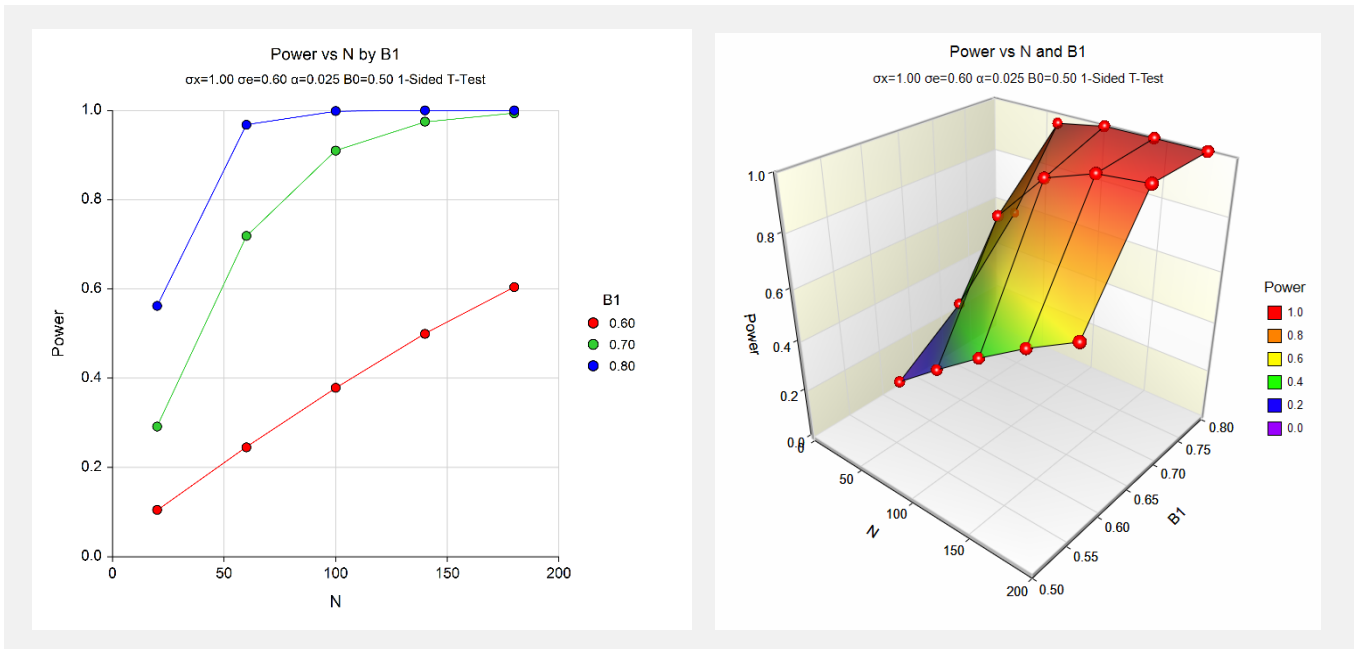
Power is the probability of rejecting a false null hypothesis. It should be close to one.  
 N is the size of the sample drawn from the population. To conserve resources, it should be small.  
 B0 is the slope under the null hypothesis.  
 B1 is the slope at which the power is calculated.  
 $\sigma_x$  is the standard deviation of the X values.  
 $\sigma_y$  is the standard deviation of Y (ignoring X).  
 $\sigma_e$  is the standard deviation of the residuals.  
 $R^2$  is the R-squared when Y is regressed on X.  
 Alpha is the probability of rejecting a true null hypothesis.

### Summary Statements

A sample size of 20 achieves 10% power to detect a change in slope from 0.500 under the null hypothesis to 0.600 under the alternative hypothesis when the statistical hypothesis is one-sided, the significance level is 0.025, the standard deviation of X is 1.000, the standard deviation of Y is 0.849, the standard deviation of residuals is 0.600, and  $R^2$  is 0.500.

This report shows the calculated sample size for each of the scenarios.

## Plots Section



These plots show the power versus the sample size for the three values of B1.



## Non-Zero Null Tests for Simple Linear Regression

## Example 2 – Validation using Neter, Wasserman, and Kutner (1983)

Neter, Wasserman, and Kutner (1983) pages 71 and 72 present a power analysis when

$$N = 10, B_0 = 0, B_1 = 0.25, \alpha = 0.05, \sigma_X = \sqrt{3400/10} = 18.439, \quad \sigma_e = \sqrt{10} = 3.16228.$$

They found the power to be approximately 0.97.

### Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Non-Zero Null Tests for Simple Linear Regression** procedure window. 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>
<b>Design Tab</b>	
Solve For .....	<b>Power</b>
Alternative Hypothesis .....	<b>Two-Sided (H1: B ≠ B0)</b>
Alpha .....	<b>0.05</b>
N (Sample Size) .....	<b>10</b>
B0 (Slope H0) .....	<b>0.0</b>
B1 (Slope H1) .....	<b>0.25</b>
$\sigma_X$ Input Type .....	<b><math>\sigma_X</math> (Std Dev of X)</b>
$\sigma_X$ (Std Dev of X) .....	<b>18.439</b>
$\sigma_e$ Input Type .....	<b><math>\sigma_e</math> (Std Dev of Residuals)</b>
$\sigma_e$ (Std Dev of Residuals) .....	<b>3.16228</b>

### Output

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

<b>Numeric Results</b>									
Hypotheses: H0: B = B0 vs. H1: B ≠ B0									
	Sample Size	H0 Slope B0	Actual Slope B1	Std Dev of X $\sigma_X$	Std Dev of Y $\sigma_Y$	Std Dev of Resids $\sigma_e$	R <sup>2</sup>	Alpha	
<b>Power</b>	<b>N</b>								
<b>0.9797</b>	<b>10</b>	<b>0.000</b>	<b>0.250</b>	<b>18.439</b>	<b>5.590</b>	<b>3.162</b>	<b>0.680</b>	<b>0.050</b>	

The power of 0.9797 matches their approximate result to two decimals. Note that they used interpolation from a table to obtain their answer.

## Non-Zero Null Tests for Simple Linear Regression

## Example 3 – Observational Study given in Dupont and Plummer (1998)

Dupont and Plummer (1998) page 593 present a power analysis example for an *observational* study in which the values of the X variable is not fixed. In this example,

$$N = 100, B_0 = 0, B_1 = -0.0667, \alpha = 0.05, \sigma_X = 7.5, \quad \sigma_Y = 4.$$

They found the power to be approximately 0.24.

### Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Non-Zero Null Tests for Simple Linear Regression** procedure window. 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>
<b>Design Tab</b>	
Solve For .....	<b>Power</b>
Alternative Hypothesis .....	<b>Two-Sided (H1: B ≠ B0)</b>
Alpha.....	<b>0.05</b>
N (Sample Size).....	<b>100</b>
B0 (Slope H0) .....	<b>0.0</b>
B1 (Slope H1) .....	<b>-0.0667</b>
$\sigma_X$ Input Type.....	<b><math>\sigma_X</math> (Std Dev of X)</b>
$\sigma_X$ (Std Dev of X) .....	<b>7.5</b>
$\sigma_Y$ Input Type.....	<b><math>\sigma_Y</math> (Std Dev of Y)</b>
$\sigma_Y$ (Std Dev of Y) .....	<b>4</b>

### Output

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

<b>Numeric Results</b>									
Hypotheses: H0: B = B0 vs. H1: B ≠ B0									
	Sample Size	H0 Slope B0	Actual Slope B1	Std Dev of X $\sigma_X$	Std Dev of Y $\sigma_Y$	Std Dev of Resids $\sigma_e$	R <sup>2</sup>	Alpha	
<b>Power</b>	<b>N</b>								
0.2390	100	0.000	-0.067	7.500	4.000	3.969	0.016	0.050	

The power of 0.2390 matches their result of 0.24 to two decimals.

## Non-Zero Null Tests for Simple Linear Regression

## Example 4 – Fixed-X Sample Size Study given in Dupont and Plummer (1998)

Dupont and Plummer (1998) pages 593-594 present a sample size example for a *fixed-X* study in which the values of the X variable are fixed at 10, 30, and 50. In this example,

$$\text{Power} = 0.90, B1 = 0.01, \alpha = 0.05, r = 0.4. \text{ So, using the relationship } r = \frac{B\sigma_X}{\sigma_Y},$$

we obtain  $\sigma_Y = \frac{0.01(16.330)}{0.4} = 0.4082$ . They found the sample size to be 57.

### Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Non-Zero Null Tests for Simple Linear Regression** procedure window. You may then make the appropriate entries as listed below, or open **Example 4** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
<b>Design Tab</b>	
Solve For .....	<b>Sample Size</b>
Alternative Hypothesis .....	<b>Two-Sided (H1: B ≠ B0)</b>
Power .....	<b>0.9</b>
Alpha .....	<b>0.05</b>
B0 (Slope H0) .....	<b>0.0</b>
B1 (Slope H1) .....	<b>0.01</b>
σ <sub>x</sub> Input Type .....	<b>List of X Values</b>
List of X Values .....	<b>10 30 50</b>
σ <sub>e</sub> Input Type .....	<b>σ<sub>Y</sub> (Std Dev of Y)</b>
σ <sub>Y</sub> (Std Dev of Y) .....	<b>0.4082</b>

### Output

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

<b>Numeric Results</b>									
Hypotheses: H0: B = B0 vs. H1: B ≠ B0									
	Sample Size	H0 Slope B0	Actual Slope B1	Std Dev of X σ <sub>x</sub>	Std Dev of Y σ <sub>Y</sub>	Std Dev of Resids σ <sub>e</sub>	R <sup>2</sup>	Alpha	
Power	N								
0.9044	58	0.000	0.010	16.330	0.408	0.374	0.160	0.050	

PASS has computed the sample size as 58, one more than the 57 Dupont and Plummer found. We assume that this is a rounding problem. PASS computed the power for an N of 57 to be 0.8993 which does round to 0.90 but is actually slightly less than the desired 0.90.