

Chapter 856

Confidence Intervals for Linear Regression Slope

Introduction

This routine calculates the sample size necessary to achieve a specified distance from the slope to the confidence limit at a stated confidence level for a confidence interval about the slope in simple linear regression.

Caution: This procedure assumes that the slope and standard deviation/correlation estimates of the future sample will be the same as the slope and standard deviation/correlation estimates that are specified. If the slope and standard deviation/correlation estimates are different from those specified when running this procedure, the interval width may be narrower or wider than specified.

Technical Details

For a single slope in simple linear regression analysis, a two-sided, $100(1 - \alpha)\%$ confidence interval is calculated by

$$b_1 \pm t_{1-\alpha/2, n-2} s_{b_1}$$

where b_1 is the calculated slope and s_{b_1} is the estimated standard deviation of b_1 , or

$$s_{b_1} = \frac{s}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}}$$

where

$$s = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y})^2}{n-2}}$$

The value s^2 is often obtained from regression tables as MSE.

A one-sided $100(1 - \alpha)\%$ upper confidence limit is calculated by

$$b_1 + t_{1-\alpha, n-2} s_{b_1}$$

Confidence Intervals for Linear Regression Slope

Similarly, the one-sided $100(1 - \alpha)\%$ lower confidence limit is

$$b_1 - t_{1-\alpha, n-2} s_{b_1}$$

Each confidence interval is calculated using an estimate of the slope plus and/or minus a quantity that represents the distance from the mean to the edge of the interval. For two-sided confidence intervals, this distance is sometimes called the precision, margin of error, or half-width. We will label this distance, D .

The basic equation for determining sample size when D has been specified is

$$D = t_{1-\alpha/2, n-2} s_{b_1}$$

This equation can be solved for any of the unknown quantities in terms of the others. The value $\alpha/2$ is replaced by α when a one-sided interval is used.

In this procedure, the slope and the standard deviation of the X 's are entered as input, where

$$b_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

and

$$s_X = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}$$

One of two different additional inputs can be used to calculate s .

1. Standard deviation of the Y 's:

$$s_Y = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n-1}}$$

In this case s is generated using

$$s = \sqrt{\left(s_Y^2 - b_1^2 s_X^2\right) \frac{n-1}{n-2}}$$

2. Directly using s :

$$s = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y})^2}{n-2}}$$

which is often obtained from regression tables as the square root of MSE.

Confidence Level

The confidence level, $1 - \alpha$, has the following interpretation. If thousands of samples of n items are drawn from a population using simple random sampling and a confidence interval is calculated for each sample, the proportion of those intervals that will include the true population slope is $1 - \alpha$.

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.

One-Sided or Two-Sided Interval

Interval Type

Specify whether the interval to be used will be a two-sided confidence interval, an interval that has only an upper limit, or an interval that has only a lower limit.

Confidence

Confidence Level

The confidence level, $1 - \alpha$, has the following interpretation. If thousands of samples of n items are drawn from a population using simple random sampling and a confidence interval is calculated for each sample, the proportion of those intervals that will include the true population slope is $1 - \alpha$.

Often, the values 0.95 or 0.99 are used. You can enter single values or a range of values such as *0.90, 0.95* or *0.90 to 0.99 by 0.01*.

Sample Size

N (Sample Size)

Enter one or more values for the sample size. This is the number of individuals selected at random from the population to be in the study.

N must be greater than or equal to 3.

You can enter a single value or a range of values.

Precision

Distance from Slope to Limit(s)

This is the distance from the confidence limit(s) to the sample slope. For two-sided intervals, it is also known as the precision, half-width, or margin of error.

You can enter a single value or a list of values. The value(s) must be greater than zero.

Confidence Intervals for Linear Regression Slope

Regression – Slope

B (Slope)

Enter the magnitude of the sample slope.

In the linear equation, $Y=A+BX$, B represents the slope of the line relating the dependent variable Y to the independent variable X.

Caution: The sample size estimates for this procedure assume that the slope that is achieved when the confidence interval is produced is the same as the slope entered here. If the sample slope is different from the one specified here, the width of the interval may be narrower or wider than specified.

You can enter a single value or a range of values such as 1,2,3.

This value must be greater than zero.

Regression – Standard Deviation of X's

SX (Standard Deviation of X's)

This is the standard deviation of the X values *in the sample*.

Caution: The sample size estimates for this procedure assume that the SX that is achieved when the confidence interval is produced is the same as the SX entered here. If the sample SX is different from the one specified here, the width of the interval may be narrower or wider than specified.

It is assumed that the standard deviation formula used in calculating SX uses the divisor N-1.

The value must be greater than zero.

Regression – Residual Variance Calculation

Residual Variance Method

Two methods can be used to calculate the standard deviation of the residuals. This option specifies which method should be used.

- **SY**

This option specifies that the formula

$$S = \text{SQRT}[\text{SY}^2 - (\text{B} * \text{SX})^2]$$

should be used to calculate the standard deviation of the residuals. Notice that this formula requires only SY, SX, and B (the correlation is not used).

- **S**

This option specifies the value of S directly. The values of SY and R are not used.

SY (Standard Deviation of Y)

This is the standard deviation of the Y values *in the sample*.

This value is only used when the Residual Variance Method is set to 'SY (Std. Dev. of Y)'.

Caution: The sample size estimates for this procedure assume that the SY that is achieved when the confidence interval is produced is the same as the SY entered here. If the sample SY is different from the one specified here, the width of the interval may be narrower or wider than specified.

It is assumed that the standard deviation formula used in calculating SY uses the divisor N-1.

The value must be greater than zero.

Confidence Intervals for Linear Regression Slope

S (Standard Deviation of Residuals)

This is the standard deviation of the residuals from the regression of Y on X.

This value is only used when the Residual Variance Method is set to 'S (Std. Dev. of Residuals)'.

Caution: The sample size estimates for this procedure assume that the S that is achieved when the confidence interval is produced is the same as the S entered here. If the sample S is different from the one specified here, the width of the interval may be narrower or wider than specified.

It is assumed that the standard deviation formula used in calculating S is $\sqrt{\text{MSE}}$, which uses the divisor $N-2$.

The value (or values) must be greater than zero.

Confidence Intervals for Linear Regression Slope

Example 1 – Calculating Sample Size

Suppose a study is planned in which the researcher wishes to construct a two-sided 95% confidence interval for the slope such that the distance from the slope to the limits is no more than 1 unit. The confidence level is set at 0.95, but 0.99 is included for comparative purposes. The estimated slope is 1.7 and the standard deviation of the X's is 11.2. The standard deviation of the residuals estimate, based on the MSE from a similar study, is 48.6. Instead of examining only the interval width of 1, a series of widths from 0.5 to 1.5 will also be considered.

The goal is to determine the necessary sample size.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Confidence Intervals for Linear Regression Slope** procedure window by clicking on **Regression**, and then clicking on **Confidence Intervals for Linear Regression Slope**. 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	Sample Size
Interval Type	Two-Sided
Confidence Level (1 - Alpha)	0.95 0.99
Distance from Slope to Limit(s)	0.5 to 1.5 by 0.1
B (Slope).....	1.7
SX (Standard Deviation of X's).....	11.2
Residual Variance Method	S (Std. Dev. of Residuals)
S (Standard Deviation of Residuals)	48.6

Annotated Output

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

Numeric Results

Numeric Results for Two-Sided Confidence Intervals for the Slope in Simple Linear Regression

Confidence Level	Sample Size N	Target Distance from Slope to Limits	Actual Distance from Slope to Limits	Slope B	Standard Deviation of X SX	Standard Deviation of Residuals S
0.95000	293	0.50000	0.49979	1.70	11.20	48.60
0.95000	205	0.60000	0.59903	1.70	11.20	48.60
0.95000	152	0.70000	0.69774	1.70	11.20	48.60
0.95000	117	0.80000	0.79805	1.70	11.20	48.60
0.95000	93	0.90000	0.89864	1.70	11.20	48.60
0.95000	76	1.00000	0.99838	1.70	11.20	48.60
0.95000	64	1.10000	1.09283	1.70	11.20	48.60
0.95000	54	1.20000	1.19606	1.70	11.20	48.60
0.95000	47	1.30000	1.28861	1.70	11.20	48.60
0.95000	41	1.40000	1.38777	1.70	11.20	48.60
0.95000	36	1.50000	1.49060	1.70	11.20	48.60
0.99000	505	0.50000	0.49977	1.70	11.20	48.60
0.99000	352	0.60000	0.59987	1.70	11.20	48.60
0.99000	260	0.70000	0.69970	1.70	11.20	48.60
0.99000	201	0.80000	0.79800	1.70	11.20	48.60
0.99000	160	0.90000	0.89725	1.70	11.20	48.60
0.99000	130	1.00000	0.99899	1.70	11.20	48.60

Confidence Intervals for Linear Regression Slope

0.99000	109	1.10000	1.09504	1.70	11.20	48.60
0.99000	92	1.20000	1.19705	1.70	11.20	48.60
0.99000	79	1.30000	1.29769	1.70	11.20	48.60
0.99000	69	1.40000	1.39511	1.70	11.20	48.60
0.99000	61	1.50000	1.49112	1.70	11.20	48.60

References

Ostle, B. and Malone, L.C. 1988. Statistics in Research. Iowa State University Press. Ames, Iowa.

Report Definitions

Confidence level is the proportion of confidence intervals (constructed with this same confidence level, sample size, etc.) that would contain the true slope.

N is the size of the sample drawn from the population.

Distance from Slope to Limit is the distance from the confidence limit(s) to the sample slope. For two-sided intervals, it is also know as the precision, half-width, or margin of error.

Target Distance from Slope to Limit is the value of the distance that is entered into the procedure.

Actual Distance from Slope to Limit is the value of the distance that is obtained from the procedure.

B is the sample slope.

SX is the sample standard deviation of the X values.

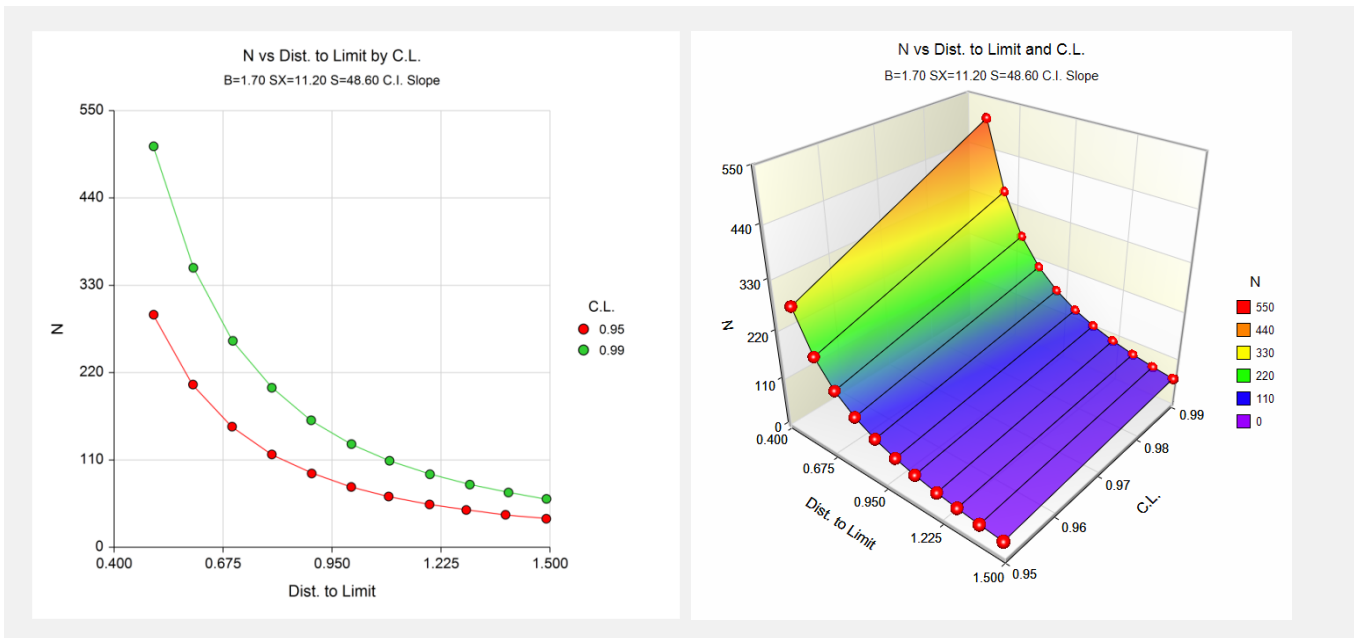
S is the standard deviation of the sample residuals.

Summary Statements

A sample size of 293 produces a two-sided 95% confidence interval with a distance from the sample slope to the limits that is equal to 0.500 when the sample slope is 1.70, the standard deviation of the X's is 11.20, and the standard deviation of the residuals is 48.60.

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

Plots Section



These plots show the sample size versus the distance from the sample slope to the limits for the two confidence levels.

Confidence Intervals for Linear Regression Slope

Example 2 – Validation using Ostle and Malone (1988)

Ostle and Malone (1988) page 234 give an example of a calculation for a confidence interval for the slope when the confidence level is 95%, the slope is 7.478, the standard deviation of the X's is 3.8944, the standard deviation of the residuals is 5.369792, and the distance from the slope to the limits is 0.87608. The sample size is 13.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Confidence Intervals for Linear Regression Slope** procedure window by clicking on **Regression**, and then clicking on **Confidence Intervals for Linear Regression Slope**. 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
Interval Type	Two-Sided
Confidence Level (1 - Alpha)	0.95
Distance from Slope to Limit(s)	0.87608
B (Slope)	7.478
SX (Standard Deviation of X's)	3.8944
Residual Variance Method	S (Std. Dev. of Residuals)
S (Standard Deviation of Residuals)	5.369792

Output

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

Numeric Results

Numeric Results for Two-Sided Confidence Intervals for the Slope in Simple Linear Regression						
Confidence Level	Sample Size N	Target Distance from Slope to Limits	Actual Distance from Slope to Limits	Slope B	Standard Deviation of X SX	Standard Deviation of Residuals S
0.95000	13	0.87608	0.87608	7.48	3.89	5.37

PASS also calculated the sample size to be 13.