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 – α)% 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

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

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

A one-sided 100(1 – α)% upper confidence limit is calculated by

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

Similarly, the one-sided 100(1 – α)% 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

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

2. Directly using *s*:

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

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

Confidence Level

The confidence level, 1 – α, 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 – α.

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

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 ..... **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**

Output

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

Numeric Reports

Numeric Results

Solve For: Sample Size
Interval Type: Two-Sided

Confidence Level	Sample Size N	Distance from Slope to Limits Target	Distance from Slope to Limits Actual	Slope B	Standard Deviation of X's SX	Standard Deviation of Residuals S
0.95	293	0.5	0.49979	1.7	11.2	48.6
0.95	205	0.6	0.59903	1.7	11.2	48.6
0.95	152	0.7	0.69774	1.7	11.2	48.6
0.95	117	0.8	0.79805	1.7	11.2	48.6
0.95	93	0.9	0.89864	1.7	11.2	48.6
0.95	76	1.0	0.99838	1.7	11.2	48.6
0.95	64	1.1	1.09283	1.7	11.2	48.6
0.95	54	1.2	1.19606	1.7	11.2	48.6
0.95	47	1.3	1.28861	1.7	11.2	48.6
0.95	41	1.4	1.38777	1.7	11.2	48.6
0.95	36	1.5	1.49060	1.7	11.2	48.6
0.99	505	0.5	0.49977	1.7	11.2	48.6
0.99	352	0.6	0.59987	1.7	11.2	48.6
0.99	260	0.7	0.69970	1.7	11.2	48.6
0.99	201	0.8	0.79800	1.7	11.2	48.6
0.99	160	0.9	0.89725	1.7	11.2	48.6
0.99	130	1.0	0.99899	1.7	11.2	48.6
0.99	109	1.1	1.09504	1.7	11.2	48.6
0.99	92	1.2	1.19705	1.7	11.2	48.6
0.99	79	1.3	1.29769	1.7	11.2	48.6
0.99	69	1.4	1.39511	1.7	11.2	48.6
0.99	61	1.5	1.49112	1.7	11.2	48.6

Confidence Level	The proportion of confidence intervals (constructed with this same confidence level, sample size, etc.) that would contain the true slope.
N	The size of the sample drawn from the population.
Distance from Slope to Limits	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.
Target Distance	The value of the distance that is entered into the procedure.
Actual Distance	The value of the distance that is obtained from the procedure.
B	The simple linear regression slope.
SX	The standard deviation of the X values.
S	The standard deviation of the sample residuals.

Summary Statements

A single-group design will be used to obtain a two-sided 95% confidence interval for a single linear regression slope. The standard t-distribution-based formula will be used to calculate the confidence interval. The standard deviation of the X's is assumed to be 11.2, and the standard deviation of the residuals is assumed to be 48.6. To produce a confidence interval with a distance of no more than 0.5 from the sample slope to either limit, 293 subjects will be needed.

Dropout-Inflated Sample Size

Dropout Rate	Sample Size N	Dropout-Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	293	367	74
20%	205	257	52
20%	152	190	38
20%	117	147	30
20%	93	117	24
20%	76	95	19
20%	64	80	16
20%	54	68	14
20%	47	59	12
20%	41	52	11
20%	36	45	9
20%	505	632	127
20%	352	440	88
20%	260	325	65
20%	201	252	51
20%	160	200	40
20%	130	163	33
20%	109	137	28
20%	92	115	23
20%	79	99	20
20%	69	87	18
20%	61	77	16

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 the confidence interval is computed. If N subjects are evaluated out of the N' subjects that are enrolled in the study, the design will achieve the stated confidence interval.
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. After solving for N, 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, 367 subjects should be enrolled to obtain a final sample size of 293 subjects.

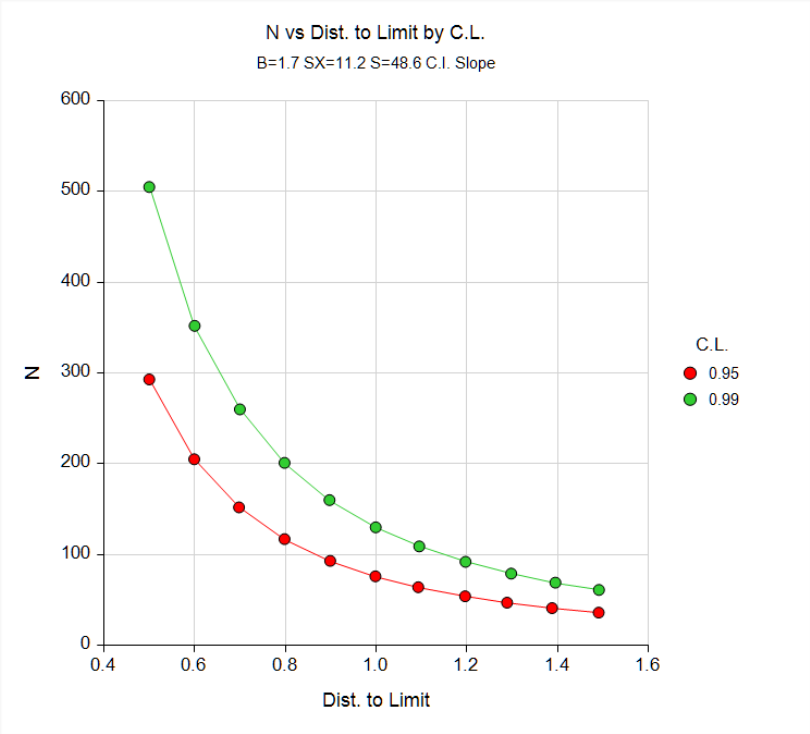
References

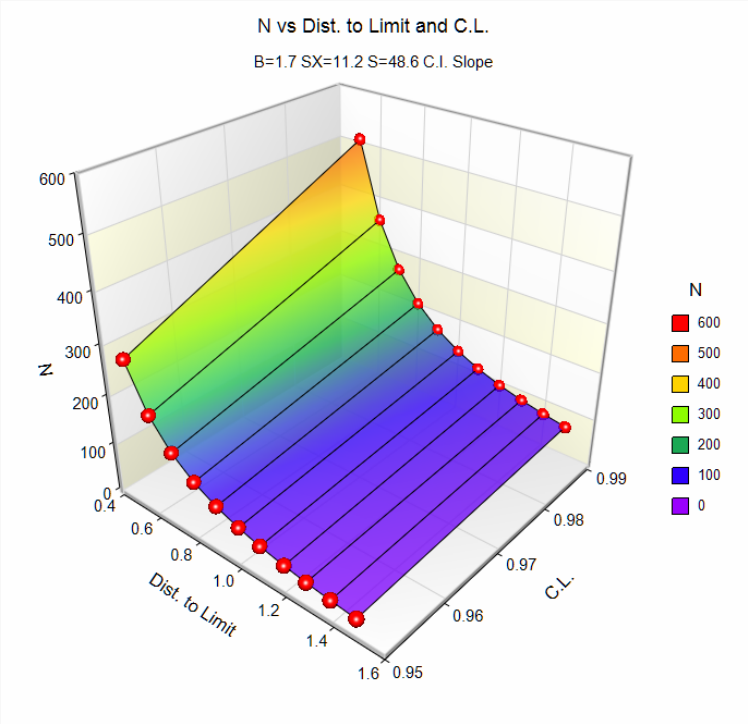
Ostle, B. and Malone, L.C. 1988. Statistics in Research. Iowa State University Press. Ames, Iowa.

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

Plots Section

Plots





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

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

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**
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

Solve For: Sample Size
Interval Type: Two-Sided

Confidence Level	Sample Size N	Distance from Slope to Limits Target	Distance from Slope to Limits Actual	Slope B	Standard Deviation of X's SX	Standard Deviation of Residuals S
0.95	13	0.87608	0.87608	7.478	3.8944	5.36979

PASS also calculated the sample size to be 13.