

Learning Guide: Calculating Confidence Intervals for Regression Slopes

Authored by
Mohammed Iooti

October 29, 2025

RECOMMENDED CITATION

Mohammed Iooti (2025). *Learning Guide: Calculating Confidence Intervals for Regression Slopes*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=5160>

The Foundation of Simple Linear Regression

Simple linear regression (SLR) stands as a cornerstone statistical methodology used to rigorously model and quantify the linear association between two continuous variables. This technique is invaluable for analysts seeking to understand how variation in one factor, designated as the **predictor variable** (or independent variable), reliably translates into changes in a second factor, the **response variable** (or dependent variable). Through SLR, researchers gain the capacity to explain, predict, and ultimately interpret real-world phenomena across diverse academic and professional fields.

The primary goal of conducting a simple linear regression analysis is to determine the optimal straight line--often referred to as the "line of best fit"--that accurately represents the underlying trend within the observed data points. This optimized line is mathematically derived by minimizing the sum of the squared residuals, which are the vertical distances between the actual response values and the values predicted by the model. The foundational algebraic representation of this fitted relationship is given by the formula:

$$y = b_0 + b_1x$$

Each element within this equation plays a critical and distinct role in defining the linear model:

y: Represents the **estimated response value**. This is the predicted outcome for the dependent variable corresponding to a specific input value of x .

b₀: Denotes the **y-intercept** of the regression line. Statistically, it is interpreted as the expected mean value of the response variable when the predictor variable x is exactly zero.

b₁: Signifies the **slope coefficient** of the regression line. This is arguably the most critical component, as it quantifies the average magnitude and direction of change in the estimated response variable for every single-unit increase in the predictor variable x .

x: Is the observed **value of the predictor variable**, serving as the independent factor driving the analysis.

Focusing on the Regression Slope (b₁): The Measure of Relationship

While the intercept (b₀) provides a baseline, the **slope** coefficient, **b₁**, is typically the parameter of greatest interest in regression studies. This coefficient provides the direct, actionable insight into the strength and nature of the linear connection between the predictor and response variables. Specifically, **b₁** tells us, on average, how much we anticipate the response variable to shift when the predictor variable increases by one unit, assuming all other conditions remain constant.

It is crucial to recognize that the calculated slope (**b₁**) derived from a finite sample is merely an estimation of the true, underlying **population slope (β₁)**. Because we base our calculations on a

limited subset of the entire population, our estimate of β is inherently subject to sampling error and uncertainty. Therefore, relying solely on the point estimate b_1 is insufficient for robust statistical inference.

To address this inherent variability and provide a more reliable measure of the relationship, statisticians construct a [confidence interval](#) for β . This interval provides a range of plausible values for the true population slope, established with a specific, high degree of certainty (e.g., 90%, 95%, or 99%). The confidence interval is significantly more informative than the single point estimate b_1 , as it clearly communicates the precision of our estimation and forms the basis for making sound conclusions about the population.

Deriving the Confidence Interval Formula for the Population Slope (β)

The process for calculating a confidence interval for the true population slope β involves combining the sample estimate, the precision of that estimate, and a critical value determined by the desired confidence level. This methodology results in a margin of error that is applied symmetrically around the sample slope. The universally accepted formula for calculating this interval is as follows:

Confidence Interval for β : $b_1 \pm t_{1-\alpha/2, n-2} * se(b_1)$

Understanding the contribution of each element is essential for accurate interpretation of the result:

b_1 : The **sample slope coefficient**. This value is the point estimate calculated directly from the [simple linear regression](#) output and serves as the center of our confidence interval.

$t_{1-\alpha/2, n-2}$: The **t critical value**. This value is sourced from the [t-distribution](#) and is necessary because the population standard deviation is unknown (a common scenario in regression). It depends on two factors: the specified [confidence level](#) ($1-\alpha$) and the [degrees of freedom](#) ($n-2$), where n is the total number of observations and α (alpha) is the significance level.

$se(b_1)$: The [standard error of \$b_1\$](#) . This measure, also provided by the regression output, quantifies the expected variability of the sample slope estimate b_1 across multiple samples. A smaller standard error signifies a more reliable and precise estimate of the true slope β .

The product of the t critical value and the standard error constitutes the margin of error, which defines the uncertainty surrounding our sample estimate and ultimately establishes the width of the interval.

Practical Application: Setting Up the Calculation

To effectively demonstrate how to construct a confidence interval for the regression slope, let us analyze a typical educational scenario. Imagine a study designed to investigate the linear

relationship between the time students dedicate to studying and their resulting exam scores. We gather data from 15 students ($n=15$), using "hours studied" as the predictor variable (x) and "exam score" as the response variable (y).

The raw dataset, outlining the paired observations for study hours and corresponding exam performance, is presented in the image below:

Hours Studied	Exam Score
1	64
2	66
4	76
5	73
5	74
6	81
6	83
7	82
8	80
10	88
11	84
11	82
12	91
12	93
14	89

Following the data collection, we execute a [simple linear regression analysis](#) using appropriate statistical software. The software generates a comprehensive output table that contains all the necessary coefficients and measures required for inference. For the purpose of determining the [confidence interval](#) for the slope, our focus must be placed squarely on the sample slope coefficient (b_1) and its associated [standard error of the slope](#) ($se(b_1)$).

The resulting output from the regression analysis run on our sample data is shown here:

SUMMARY OUTPUT						
<i>Regression Statistics</i>						
Multiple R	0.9115809					
R Square	0.8309798					
Adjusted R Square	0.8179782					
Standard Error	3.6409321					
Observations	15					
<i>ANOVA</i>						
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>	
Regression	1	847.2669759	847.267	63.91387	2.2527E-06	
Residual	13	172.3330241	13.25639			
Total	14	1019.6				
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	65.333952	2.105989497	31.02292	1.41E-13	60.7842381	69.8836655
hours	1.9823748	0.247963753	7.994615	2.25E-06	1.44668165	2.51806789

From this detailed output, we can immediately extract the essential statistical values needed to proceed with our calculation. The coefficient listed for the 'Hours Studied' variable is our estimated [regression slope](#), b_1 , and the value in the adjacent 'Std. Error' column is the corresponding standard error, $se(b_1)$.

Step-by-Step Derivation of the Confidence Interval

Based on the regression output presented above, we have established the estimated [regression slope](#) (b_1) as **1.982**, and the [standard error of the slope](#) ($se(b_1)$) as **0.248**. This slope value suggests that for every additional hour a student studies, their exam score is expected, on average, to increase by 1.982 points. Our objective now is to construct a 95% [confidence interval](#) for the true population slope β .

We first compile the required input parameters for the formula:

$b_1 = 1.982$ (The sample point estimate)

$se(b_1) = 0.248$ (The standard error of the estimate)

Confidence Level = 95%. This implies the significance level $\alpha = 0.05$. Therefore, we look for the t-value corresponding to $1 - 0.05/2 = 0.975$.

Degrees of Freedom (df) = $n - 2 = 15 - 2 = 13$.

The next crucial step involves identifying the correct [t critical value](#). By consulting a [t-distribution table](#) or utilizing statistical calculation tools, we find the two-tailed t-value for a 95% confidence

level with 13 **degrees of freedom**. This value, **t.975, 13**, is determined to be approximately **2.1604**.

Finally, we substitute these values into the confidence interval formula to determine the lower and upper bounds:

95% C.I. for β_1 : $b_1 \pm t_{1-\alpha/2, n-2} * se(b_1)$

95% C.I. for β_1 : $1.982 \pm t.975, 15-2 * 0.248$

95% C.I. for β_1 : $1.982 \pm 2.1604 * 0.248$

95% C.I. for β_1 : 1.982 ± 0.5357792 (This is the Margin of Error)

95% C.I. for β_1 :

95% C.I. for β_1 : (Rounded for practical use)

The resulting 95% confidence interval for the true population regression slope is .

Interpreting the Statistical Significance of the Interval

The derived 95% **confidence interval**, spanning from **1.446 to 2.518**, provides a precise range of highly plausible values for the true population **regression slope** ($\beta?$). The interpretation is straightforward: we are 95% confident that the genuine average increase in exam score associated with each additional hour of study lies somewhere between 1.446 and 2.518 points for the population of students under consideration.

The most critical step in interpreting this interval is examining its relationship with the value zero. If the interval were to contain zero (i.e., if the lower bound were negative and the upper bound were positive), it would imply that a zero slope is a plausible value for the population. A population slope of zero indicates that there is no true linear relationship between the predictor variable (hours studied) and the response variable (exam scores).

In our specific case, the interval is entirely positive and therefore emphatically excludes zero. This exclusion is a powerful statistical statement, leading us to confidently conclude that there is a **statistically significant** positive association between study time and exam performance. This confirms the hypothesis that increased study effort is indeed linked to higher academic outcomes within this population.

Further Resources for Advanced Statistical Understanding

To enhance your knowledge of linear regression diagnostics and related statistical concepts, the following supplementary tutorials and articles offer valuable depth and insight:

[How to Interpret Regression Output](#)

[What is R-Squared?](#)

[What is Adjusted R-Squared?](#)