Regression Slope: Confidence Interval

This lesson describes how to construct a confidence interval around the slope of a regression line. We focus on the equation for simple linear regression, which is:

$$\hat{y} = b_0 + b_1 x$$

where b_0 is a constant, b_1 is the slope (also called the regression coefficient), x is the value of the independent variable, and $\hat{y}$ is the *predicted* value of the dependent variable.

Estimation Requirements

The approach described in this lesson is valid whenever the standard requirements for simple linear regression are met.

- The dependent variable *Y* has a linear relationship to the independent variable *X*.
- For each value of X, the probability distribution of Y has the same standard deviation σ.
- For any given value of X,
  - The Y values are independent.
  - The Y values are roughly normally distributed (i.e., symmetric and unimodal). A little skewness is ok if the sample size is large.

Previously, we described how to verify that regression requirements are met.

The Variability of the Slope Estimate

To construct a confidence interval for the slope of the regression line, we need to know the standard error of the sampling distribution of the slope. Many statistical software packages and some graphing calculators provide the standard error of the slope as a regression analysis output. The table below shows hypothetical output for the following regression equation: y = 76 + 35x .

Predictor	Coef	SE Coef	T	P
Constant	76	30	2.53	0.01
X	35	20	1.75	0.04

- Notation
- AP Statistics Formulas

In the output above, the standard error of the slope (shaded in gray) is equal to 20. In this example, the standard error is referred to as "SE Coeff". However, other software packages might use a different label for the standard error. It might be "StDev", "SE", "Std Dev", or something else.

If you need to calculate the standard error of the slope (SE) by hand, use the following formula:

$$SE = s_{b_1} = \text{sqrt} [\Sigma(y_i - \hat{y}_i)^2 / (n - 2)] / \text{sqrt} [\Sigma(x_i - \bar{x})^2]$$

where y_i is the value of the dependent variable for observation *i*, $\hat{y}_i$ is estimated value of the dependent variable for observation *i*, x_i is the observed value of the independent variable for observation *i*, $\bar{x}$ is the mean of the independent variable, and n is the number of observations.

How to Find the Confidence Interval for the Slope of a Regression Line

Previously, we described how to construct confidence intervals. The confidence interval for the slope uses the same general approach. Note, however, that the critical value is based on a t score with *n* - 2 degrees of freedom.

- Identify a sample statistic. The sample statistic is the regression slope b_1 calculated from sample data. In the table above, the regression slope is 35.
- Select a confidence level. The confidence level describes the uncertainty of a sampling method. Often, researchers choose 90%, 95%, or 99% confidence levels; but any percentage can be used.
- Find the margin of error. Previously, we showed how to compute the margin of error, based on the critical value and standard error. When calculating the margin of error for a regression slope, use a t score for the critical value, with degrees of freedom (DF) equal to *n* - 2.
- Specify the confidence interval. The range of the confidence interval is defined by the *sample statistic* ± *margin of error*. And the uncertainty is denoted by the confidence level.

In the next section, we work through a problem that shows how to use this approach to construct a confidence interval for the slope of a regression line.

Test Your Understanding of This Lesson

Problem 1

The local utility company surveys 101 randomly selected customers. For each survey participant, the company collects the following: annual electric bill (in dollars) and home size (in square feet). Output from a regression analysis appears below.

Regression equation: Annual bill = 0.55 * Home size + 15

Predictor	Coef	SE Coef	T	P
Constant	15	3	5.0	0.00
Home size	0.55	0.24	2.29	0.01

What is the 99% confidence interval for the slope of the regression line?

(A) 0.25 to 0.85
(B) 0.02 to 1.08
(C) -0.08 to 1.18
(D) 0.20 to 1.30
(E) 0.30 to 1.40

Solution

The correct answer is (C). Use the following four-step approach to construct a confidence interval.

- Identify a sample statistic. Since we are trying to estimate the slope of the true regression line, we use the regression coefficient for home size (i.e., the sample estimate of slope) as the sample statistic. From the regression output, we see that the slope coefficient is 0.55.

- Select a confidence level. In this analysis, the confidence level is defined for us in the problem. We are working with a 99% confidence level.

- Find the margin of error. Elsewhere on this site, we show how to compute the margin of error. The key steps applied to this problem are shown below.

    - Find standard deviation or standard error. The standard error is given in the regression output. It is 0.24.

    - Find critical value. The critical value is a factor used to compute the margin of error. With simple linear regression, to compute a confidence interval for the slope, the critical value is a t score with degrees of freedom equal to *n* - 2. To find the critical value, we take these steps.

- Compute alpha (α): α = 1 - (confidence level / 100) = 1 - 99/100 = 0.01
- Find the critical probability (p*): p* = 1 - α/2 = 1 - 0.01/2 = 0.995
- Find the degrees of freedom (df): df = *n* - 2 = 101 - 2 = 99.
- The critical value is the t score having 99 degrees of freedom and a cumulative probability equal to 0.995. From the t Distribution Calculator, we find that the critical value is 2.63.

- Compute margin of error (ME): ME = critical value * standard error = 2.63 * 0.24 = 0.63

- Specify the confidence interval. The range of the confidence interval is defined by the *sample statistic* ± *margin of error*. And the uncertainty is denoted by the confidence level.

Therefore, the 99% confidence interval is -0.08 to 1.18. That is, we are 99% confident that the true slope of the regression line is in the range defined by 0.55 ± 0.63.

< Previous lesson | Next lesson >