

How can a t-test be performed to determine the slope of a regression line in R?

Authored by
stats writer

June 25, 2024

RECOMMENDED CITATION

stats writer (2024). *How can a t-test be performed to determine the slope of a regression line in R?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=151905>

A t-test can be performed in R to determine the slope of a regression line by using the "lm" function. This function allows the user to fit a linear regression model to the data and obtain the slope coefficient. The "summary" function can then be used to obtain the t-statistic and p-value for the slope coefficient. A low p-value (typically less than 0.05) indicates a significant relationship between the two variables, and the slope coefficient represents the change in the response variable for every one unit change in the predictor variable. This process can be repeated for different variables to determine the impact of each variable on the regression line.

Perform t-Test for Slope of Regression Line in R

Whenever we perform , we end up with the following estimated regression equation:

$$? = b_0 + b_1x$$

We typically want to know if the slope coefficient, b_1 , is statistically significant.

To determine if b_1 is statistically significant, we can perform a t-test with the following test statistic:

$$t = b_1 / se(b_1)$$

where:

$se(b_1)$ represents the standard error of b_1 .

We can then calculate the that corresponds to this test statistic with $n-2$ degrees of freedom.

If the p-value is less than some threshold (e.g. $\alpha = .05$) then we can conclude that the slope coefficient is different than zero.

In other words, there is a statistically significant relationship between the predictor variable and the in the model.

The following example shows how to perform a t-test for the slope of a regression line in R.

Example: Performing a t-Test for Slope of Regression Line in R

Suppose we have the following data frame in R that contains information about the hours studied and final exam score received by 12 students in some class:

```
#create data frame
```

```
df <- data.frame(hours=c(1, 1, 2, 2, 3, 4, 5, 5, 5, 6, 6, 8),  
score=c(65, 67, 78, 75, 73, 84, 80, 76, 89, 91, 83, 82))
```

```
#view data frame
```

```
df
```

```
hours score
```

```
1 1 65
```

2 1 67

3 2 78

4 2 75

5 3 73

6 4 84

7 5 80

8 5 76

9 5 89

10 6 91

11 6 83

12 8 82

Suppose we would like to fit a simple linear regression model to determine if there is a statistically significant relationship between hours studied and exam score.

We can use the function in R to fit this regression model:

```
#fit simple linear regression model
```

```
fit <- lm(score ~ hours, data=df)
```

```
#view model summary
```

```
summary(fit)
```

Call:

lm(formula = score ~ hours, data = df)

Residuals:

Min 1Q Median 3Q Max

-7.398 -3.926 -1.139 4.972 7.713

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 67.7685 3.3757 20.075 2.07e-09 ***

hours 2.7037 0.7456 3.626 0.00464 **

Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**

Residual standard error: 5.479 on 10 degrees of freedom

Multiple R-squared: 0.568, Adjusted R-squared: 0.5248

F-statistic: 13.15 on 1 and 10 DF, p-value: 0.004641

Exam score = 67.7685 + 2.7037(hours)

To test if the slope coefficient is statistically significant, we can calculate the t-test statistic as:

$t = b_1 / se(b_1) = 2.7037 / 0.7456 = 3.626$

The p-value that corresponds to this t-test statistic is shown in the column called $\text{Pr}(> |t|)$ in the output.

The p-value turns out to be 0.00464.

Since this p-value is less than 0.05, we conclude that the slope coefficient is statistically significant.

In other words, there is a statistically significant relationship between the number of hours studied and the final score that a student receives on the exam.