What is the interpretation of Pr(>|t|) in the regression model output in R?

Authored by
stats writer

April 30, 2024

The interpretation of Pr(>|t|) in the regression model output in R refers to the probability of obtaining a t-value that is larger than the absolute value of the calculated t-statistic. This value is used to determine the significance of each independent variable in the model and is often compared to a predetermined threshold, such as 0.05. A smaller Pr(>|t|) indicates a higher level of significance and suggests that the corresponding variable has a significant effect on the dependent variable. Conversely, a larger Pr(>|t|) suggests that the relationship between the variable and the dependent variable is not statistically significant. Overall, Pr(>|t|) is an important measure in determining the validity and reliability of the regression model.

Interpret Pr(>|t|) in Regression Model Output in R

Whenever you perform linear regression in R, the output of your regression model will be displayed in the following format:

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.0035 5.9091 1.693 0.1513
x1 1.4758 0.5029 2.935 0.0325 *
x2 -0.7834 0.8014 -0.978 0.3732

The Pr(>|t|) column represents the p-value associated with the value in the t value column.

If the p-value is less than a certain significance level (e.g. α = .05) then the predictor variable is said to have a statistically significant relationship with the response variable in the model.

The following example shows how to interpret values in the Pr(>|t|) column for a given regression model.

Example: How to Interpret Pr(>|t|) Values

Suppose we would like to fit a using predictor variables x1 and x2 and a single response variable y.

The following code shows how to create a data frame and fit a regression model to the data:

```
#create data frame
df <- data.frame(x1=c(1, 3, 3, 4, 4, 5, 6, 6),
x2=c(7, 7, 5, 6, 5, 4, 5, 6),
y=c(8, 8, 9, 9, 13, 14, 17, 14))

#fit multiple linear regression model
model <- lm(y ~ x1 + x2, data=df)

#view model summary
summary(model)

Call:
lm(formula = y ~ x1 + x2, data = df)

Residuals:
1 2 3 4 5 6 7 8
```

2.0046 -0.9470 -1.5138 -2.2062 1.0104 -0.2488 2.0588 -0.1578

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.0035 5.9091 1.693 0.1513
x1 1.4758 0.5029 2.935 0.0325 *
x2 -0.7834 0.8014 -0.978 0.3732

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

Residual standard error: 1.867 on 5 degrees of freedom
Multiple R-squared: 0.7876, Adjusted R-squared: 0.7026
F-statistic: 9.268 on 2 and 5 DF, p-value: 0.0208

Here's how to interpret the values in the Pr(>|t|) column:

The p-value for the predictor variable x1 is .0325. Since this value is less than .05, it has a statistically significant relationship with the response variable in the model.The p-value for the predictor variable x2 is .3732. Since this value is not less than .05, it does not have a statistically significant relationship with the response variable in the model.

The under the coefficient table tell us that a single asterik (*) next to the p-value of .0325 means the p-value is statistically significant at $\alpha = .05$.

How is Pr(>|t|) Actually Calculated?

Here's how the value for Pr(>|t|) is actually calculated:

Step 1: Calculate the t value

First, we calculate the t value using the following formula:

t value = Estimate / Std. Error

#calculate t-value
1.4758 / .5029

2.934579

Step 2: Calculate the p-value

Next, we calculate the p-value. This represents the probability that the absolute value of the t-distribution is greater than 2.935.

We can use the following formula in R to calculate this

value:

p-value = 2 * pt(abs(t value), residual df, lower.tail = FALSE)

For example, here's how to calculate the p-value for a t-value of 2.935 with 5 residual degrees of freedom:

```
#calculate p-value
2 * pt(abs(2.935), 5, lower.tail = FALSE)

0.0324441
```

Notice that this p-value matches the p-value in the regression output from above.

Note: The value for the residual degrees of freedom can be found near the bottom of the regression output. In our example, it turned out to be 5:

```
Residual standard error: 1.867 on 5 degrees of freedom
```