

How to Interpret $Pr(>|t|)$ Values in R Regression Output

Authored by
stats writer

December 5, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Interpret $Pr(>|t|)$ Values in R Regression Output*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=105495>

The Pr(>|t|) values in Regression Model Output in R represent the p-values associated with the two-sided t-test performed for each estimated regression coefficient. This calculation is vital as it quantifies the statistical significance of each predictor variable's contribution to predicting the outcome variable. Specifically, a low p-value, typically less than the chosen significance level of 0.05, serves as strong evidence to reject the null hypothesis, indicating that the coefficient is reliably non-zero and should be considered an important component of the model.

Understanding the Role of Pr(>|t|) in Regression Output

The column labeled Pr(>|t|) within the summary output of a Regression Model generated in R holds one of the most critical pieces of information for determining the quality and validity of your predictive factors. Specifically, the values in this column represent the p-values that correspond to the two-sided hypothesis test for each coefficient in the model. This test, known as a coefficient-level t-test, assesses whether the true population coefficient for that specific predictor variable is statistically different from zero. If a coefficient is not significantly different from zero, it suggests that the predictor variable contributes little to no explanatory power for the response variable, potentially making its inclusion in the model unnecessary.

Interpreting these p-values correctly is paramount in applied statistics and machine learning, as they directly inform variable selection and model refinement. The hypothesis being tested for each predictor is the null hypothesis (H_0), which states that the coefficient is zero ($\beta = 0$), versus the alternative hypothesis (H_1), which states that the coefficient is non-zero ($\beta \neq 0$). The numerical output in the Pr(>|t|) column quantifies the probability of observing a t-statistic as extreme as, or more extreme than, the one calculated, assuming the null hypothesis is true. Consequently, low p-values provide strong evidence against the null hypothesis, supporting the conclusion that the corresponding predictor has a genuine linear relationship with the outcome variable.

A common threshold used across scientific disciplines, particularly in social sciences and business analytics, is the significance level, denoted as alpha (α), typically set at 0.05. If the calculated p-value in the Pr(>|t|) column falls below this chosen alpha level (e.g., $p < 0.05$), we reject the null hypothesis. Rejecting H_0 means that we conclude the predictor variable is deemed statistically significant, meaning its estimated coefficient is reliably non-zero and should be retained in the final explanatory model. Conversely, a high p-value suggests that the observed effect could easily be due to random chance, and thus, the variable might be considered insignificant.

Deciphering the Standard R Regression Output Table

When executing a standard linear model using the `lm()` function in R and then requesting the summary, the output is structured into several sections, with the 'Coefficients' table demanding the most detailed scrutiny. This table typically features five key columns: Estimate, Std. Error, t value,

Pr(>|t|), and sometimes a column for significance markers (Signif. codes). Understanding the arrangement of these columns is fundamental to interpreting model performance beyond simply looking at R-squared values.

The first two columns, Estimate and Std. Error, provide the core statistical parameters. The **Estimate** represents the calculated regression coefficient (slope) for the corresponding predictor variable, indicating the change in the response variable for a one-unit change in the predictor, holding all other variables constant. The **Std. Error**, or standard error of the estimate, measures the precision and variability of this coefficient estimate. A smaller standard error implies greater precision in the estimate. These two values form the basis for calculating the third column, the **t value**, which serves as the test statistic for the coefficient hypothesis test.

It is the final two statistical columns, **t value** and **Pr(>|t|)**, that encapsulate the result of the hypothesis test. The **t value** measures how many standard errors the estimated coefficient is away from zero. A larger absolute t-value implies that the coefficient is far from zero, suggesting significance. The **Pr(>|t|)** column then translates this t-statistic into a probability measure--the p-value. This probability is central to determining whether the observed relationship between the predictor and the response is robust or merely a random fluctuation, linking directly back to the decision of whether to reject the null hypothesis of zero effect.

Whenever you perform linear regression in R, the crucial 'Coefficients' section of the output will be displayed in a structure similar to the following example, demonstrating how these components align:

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
```

As shown above, the **Pr(>|t|)** column is derived directly from the corresponding **t value** column. If the resulting p-value is less than a predetermined significance level (e.g., $\alpha = 0.05$), the predictor variable is understood to exhibit a statistically significant relationship with the response variable within the context of the fitted model.

The Relationship Between t-Value and Pr(>|t|)

The relationship between the calculated t-statistic and the resulting Pr(>|t|) value is inverse and probabilistic. The **t value** quantifies the distance of the estimated coefficient from the hypothesized population mean (which is zero under the null hypothesis), measured in units of standard errors. A

large absolute t-value indicates a strong deviation from zero, making the null hypothesis less plausible. The t-test assumes that the coefficients are normally distributed, allowing us to use the t-distribution curve to assess probabilities.

The calculation of the p-value ($\Pr(>|t|)$) involves determining the area under the t-distribution curve that lies beyond the calculated t-statistic in both the positive and negative directions (since this is typically a two-tailed test). The formula reflects the probability of observing an absolute t-value greater than the calculated one, hence the notation $\Pr(>|t|)$. If the calculated t-statistic falls far out into the tails of the distribution, the area remaining under the curve--the p-value--will be small. Conversely, if the t-statistic is close to zero, meaning the coefficient is statistically indistinguishable from zero, the p-value will be large.

Therefore, a large absolute **t value** always leads to a small **$\Pr(>|t|)$** value, signifying higher statistical significance. This relationship is central to interpreting model results: one must look for predictor variables that generate large t-statistics, as these are the variables whose estimated effects are robust relative to their measurement uncertainty (Standard Error). Understanding this underlying statistical theory provides confidence in model interpretation, ensuring that decisions about variable inclusion are grounded in rigorous probabilistic evidence rather than simply checking if a number is small.

Practical Example: Interpreting Regression Coefficients in R

To solidify the interpretation of the $\Pr(>|t|)$ values, let us walk through a concrete example involving the fitting of a multiple linear regression model. Suppose we are interested in predicting a response variable **y** using two potential predictor variables, **x1** and **x2**. The objective is to determine which of these predictors contributes meaningfully to explaining the variation in **y**.

The process begins with generating the data structure and fitting the model in R. This involves defining a data frame and then applying the `lm()` function, followed by calling `summary()` to obtain the detailed statistical output necessary for coefficient evaluation. The code below illustrates the necessary setup for creating the sample dataset and running the regression analysis on it:

```
#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

Upon examining the coefficient table from the output, we focus exclusively on the Pr(>|t|) column to draw conclusions about the predictive variables **x1** and **x2**. Assuming we are employing the conventional alpha level (α) of 0.05 for hypothesis testing, we compare each predictor's p-value against this benchmark. This comparison dictates whether we retain the variable as statistically influential in the model or consider removing it due to lack of significant evidence supporting its effect.

The interpretation for each coefficient proceeds as follows:

The p-value calculated for the predictor variable **x1** is **0.0325**. Since this value is less than the chosen significance level of 0.05 ($0.0325 < 0.05$), we conclude that **x1** has a statistically significant relationship with the response variable **y** in the fitted model. The single asterisk (*) next to this value, as defined by the Signif. codes, visually confirms its significance at the 0.05 level.

In contrast, the p-value for the predictor variable **x2** is **0.3732**. Because this value is substantially larger than 0.05 ($0.3732 > 0.05$), we fail to reject the null hypothesis for **x2**. Therefore, **x2** does not demonstrate a statistically significant relationship with the response variable in this specific model configuration.

Based on this analysis, variable **x1** should be retained, while variable **x2** is likely a candidate for removal if the goal is to create a parsimonious model containing only influential predictors. The

Significance codes displayed beneath the coefficient table are crucial visual aids, summarizing the strength of the evidence against the null hypothesis using symbols like asterisks, where more stars indicate a lower p-value and greater significance (e.g., three stars (***) denotes $p < 0.001$).

Deep Dive into Significance Levels (Alpha)

The significance level, denoted as α , is the probability threshold set prior to conducting the hypothesis test. It represents the maximum risk of committing a Type I error--the error of incorrectly rejecting a true null hypothesis (false positive). While $\alpha = 0.05$ is the conventional standard, it is not an immutable rule. Researchers may choose a stricter level, such as $\alpha = 0.01$, when the cost of a Type I error is very high (e.g., in medical trials), or a looser level, such as $\alpha = 0.10$, in exploratory research where the goal is to identify potential relationships for future study.

The choice of alpha directly influences the interpretation of the Pr(>|t|) values. If a researcher sets $\alpha = 0.01$, a coefficient must have a p-value less than 0.01 to be considered statistically significant. If the p-value is 0.0325 (as seen for x1 in the example), it would be significant at $\alpha = 0.05$ but not significant at the stricter $\alpha = 0.01$ level. This highlights that statistical significance is always relative to the chosen threshold. It is always better practice to report the actual p-value rather than simply stating "significant" or "not significant," allowing readers to assess the evidence based on their own preferred alpha level.

It is vital to distinguish between statistical significance and practical significance. A tiny p-value indicates that the coefficient is genuinely non-zero, but the magnitude of the coefficient (the Estimate) might be so small that it holds no real-world importance. Conversely, a large coefficient might have a p-value slightly above 0.05 due to high variability (a large Standard Error), suggesting that while the observed effect is large, the evidence supporting its non-zero nature is marginally weak. Therefore, model interpretation requires considering both the p-value and the coefficient estimate in conjunction with domain expertise.

Detailed Calculation of the Pr(>|t|) Value (Step 1: Calculating the t-value)

Understanding the calculation process behind Pr(>|t|) reinforces the statistical foundation of regression model outputs. The t-test statistic, or **t value**, is the essential first step in determining the p-value. This calculation standardizes the coefficient estimate by its variability, giving a metric of reliability.

The **t value** is calculated using a straightforward ratio. It is the quotient of the estimated coefficient value and the standard error of that estimate. This can be expressed by the following simple formula, which is applied independently to each predictor variable and the intercept:

$$\mathbf{t\ value} = \text{Estimate} / \text{Std. Error}$$

Using the example from the coefficient table for variable **x1**, where the Estimate is 1.4758 and the Standard Error is 0.5029, we can verify the calculated **t value**:

```
#calculate t-value
```

```
1.4758 / .5029
```

```
2.934579
```

As expected, the calculated t-value of approximately 2.935 matches the output provided in the regression summary. This numerical result signifies that the estimated coefficient of 1.4758 is nearly three standard deviations away from zero, suggesting a highly reliable, non-zero effect. A strong t-statistic is the prerequisite for obtaining a low p-value.

Detailed Calculation of the Pr(>|t|) Value (Step 2: Determining the P-value)

Once the **t value** is computed, the next step involves converting this test statistic into the probability measure, **Pr(>|t|)**. This is achieved by reference to the Student's t-distribution. Critically, the shape of the t-distribution curve depends entirely on the degrees of freedom associated with the error term of the model.

The p-value represents the probability that a random sample would yield a t-statistic whose absolute magnitude is greater than the one we calculated (2.935 in our case), under the assumption that the true coefficient is zero. Since we are testing for a difference in either direction (positive or negative), the resulting probability must account for both tails of the distribution. The common method utilizes the cumulative distribution function (CDF) of the t-distribution in R, specifically the `pt()` function.

The standard formula in R for calculating this two-tailed p-value is:

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

Here, `abs(t value)` ensures we use the magnitude of the t-statistic; `residual df` stands for the residual degrees of freedom, which dictates the distribution shape; and `lower.tail = FALSE` calculates the probability in the upper tail. We multiply by 2 to account for the probability in both the upper and lower tails, making it a two-sided test.

For the variable **x1**, we found the t-value to be 2.935. We must also locate the residual degrees of freedom (df) from the overall model summary. Looking at the bottom of the output provided previously:

Residual standard error: 1.867 on 5 degrees of freedom

The residual degrees of freedom is 5. Using these inputs, we calculate the p-value in R:

```
#calculate p-value
```

```
2 * pt(abs(2.935), 5, lower.tail = FALSE)
```

```
0.0324441
```

This calculated value, 0.0324441, precisely matches the **Pr(>|t|)** of 0.0325 listed in the official regression model output. This successful reproduction validates the underlying statistical process, confirming that the Pr(>|t|) column is indeed the result of applying the t-test procedure to each estimated coefficient based on the degrees of freedom available in the model.

Conclusion and Best Practices for Model Interpretation

The Pr(>|t|) column in the R regression summary is not just another number; it is the probabilistic cornerstone of coefficient interpretation. It translates the reliability of the coefficient estimate (measured by the t-statistic) into a concise probability (the p-value), which directly guides decisions on variable importance and model structure. Low p-values furnish strong evidence that a predictor variable contributes genuine explanatory power to the model, suggesting its effect is unlikely to be zero.

Best practices for reporting and interpreting these results require transparency and context. Always state the chosen significance level (α). When discussing model findings, ensure that you report both the coefficient estimate (magnitude and direction of effect) and the corresponding p-value (reliability of effect). Avoid the pitfall of relying solely on p-values; a model with high statistical significance (many small p-values) may still be a poor model if the coefficients are practically meaningless or if the model violates key assumptions of linear regression.

Ultimately, mastering the interpretation of Pr(>|t|) is essential for any analyst working with R. It transforms raw output into actionable insight, allowing for the construction of parsimonious, robust, and meaningful predictive models that accurately reflect the underlying data generating process. By understanding the link between the Estimate, Standard Error, t-value, and the final p-value, researchers can confidently select the most influential variables and communicate their findings with statistical rigor.