

Understanding and Interpreting Regression Model Output in R

Authored by
Mohammed looti

November 4, 2025

RECOMMENDED CITATION

Mohammed looti (2025). *Understanding and Interpreting Regression Model Output in R*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=9743>

Mastering R's Linear Regression Model Summary

When performing rigorous [data analysis](#), especially within the powerful [R programming environment](#), fitting a [linear regression](#) model is a foundational technique. The core mechanism for this task is the [lm function](#). For any practicing data scientist or statistician, proficiency in interpreting the resulting model summary is absolutely critical. This detailed output contains specific metrics that are essential for accurately assessing the relationship between the predictor variables and the response variable.

The standard summary output presents a comprehensive coefficients table, which serves as the heart of the model diagnostics. This table provides crucial statistics for every variable incorporated into the model, allowing analysts to determine the magnitude and reliability of each relationship. Key columns range from the **Estimate** (the predicted change in the response for a one-unit change in the predictor) to the **Pr(>|t|)** (the probability value). A thorough understanding of what each of these components represents is indispensable for drawing informed and statistically sound conclusions about the underlying data structure.

Below is a typical representation of the coefficients table found in the R summary output, highlighting the critical statistics we will be focusing on:

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 |

Defining Pr(>|t|): The Probability of the T-Statistic

The column designated **Pr(>|t|)** is arguably the most vital piece of information within the regression output for determining variable importance. This numerical value is the two-sided [p-value](#) associated with the calculated **t value** for the specific predictor variable (or the intercept). It quantifies the probability of observing the data, or data more extreme, given that the underlying assumption--the null hypothesis--is true.

In the context of regression [hypothesis testing](#), the central challenge is testing the **null hypothesis** (H_0): that the true population coefficient for the predictor variable is exactly zero. If a coefficient is zero, it implies that the predictor variable has absolutely no meaningful linear relationship with the response variable, and thus, it does not contribute predictive power to the model. The **Pr(>|t|)** value provides the statistical evidence we need to either accept or reject this critical assumption.

More specifically, the [p-value](#) calculates the probability of obtaining a test statistic (the t-statistic) that is as extreme as, or more extreme than, the one derived from our sample data, assuming H_0 holds true. Consequently, a smaller [p-value](#) indicates that the observed result is unlikely to have occurred by chance if the null hypothesis were true, thereby suggesting stronger evidence against the notion of a zero coefficient.

Interpreting Significance: The Role of Alpha (α)

To effectively interpret the [p-value](#) provided by the **Pr(>|t|)** column, it must be benchmarked against a predetermined risk threshold known as the **significance level**, commonly denoted by the Greek letter α (alpha). This level represents the maximum acceptable probability of incorrectly rejecting a true null hypothesis (a Type I error). While α is most frequently set at 0.05 (or 5%), values of 0.01 or 0.10 are also utilized, depending on the field of study and the desired level of confidence required for the findings.

The decision rule for evaluating the statistical relevance of a predictor variable is remarkably straightforward: If the calculated [p-value](#) (extracted from the **Pr(>|t|)** column) is less than the predetermined significance level (α), we are statistically justified in rejecting the null hypothesis (H_0). This rejection is a powerful conclusion, as it implies that the observed linear relationship between the predictor and the response variable is considered [statistically significant](#) at the chosen confidence level.

Conversely, if the p-value is greater than or equal to the significance level (α), we lack sufficient statistical evidence to reject the null hypothesis. In this scenario, we must fail to reject H_0 , concluding that the data does not provide enough support to suggest a meaningful, non-zero linear relationship exists between the variables within the model. This framework of comparison is fundamental to drawing reliable conclusions from any regression analysis.

Practical Example: Interpreting Coefficient Results

To solidify the process of interpreting the **Pr(>|t|)** column, let us apply these principles to a concrete case. We will fit a multiple [linear regression](#) model designed to predict a single response variable, y , using two distinct predictor variables, x_1 and x_2 . This example allows us to directly observe how the p-values dictate the inclusion or exclusion of variables in our final interpretation.

We begin by constructing the necessary data frame in the R environment and subsequently executing the regression analysis utilizing the [lm function](#). The code block below illustrates the setup and the detailed summary output generated by R. Pay close attention to the final section of the output, which lists the coefficients and their corresponding probabilities.

```
#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
```

Assuming the standard significance level of $\alpha = 0.05$, we proceed to evaluate the statistical relevance of predictors **x1** and **x2** by consulting the values within the **Pr(>|t|)** column. This comparison dictates whether we retain the variable in our interpretation of the model.

Analyzing the results derived from the coefficient table leads to two distinct conclusions regarding our predictors:

The p-value for predictor variable **x1** is **0.0325**. Because this value is less than our threshold ($0.0325 < 0.05$), we confidently reject the null hypothesis for **x1**. This signifies that **x1** maintains a [statistically significant](#) relationship with the response variable **y**, confirmed visually by the asterisk code (*) in the output.

The p-value for predictor variable **x2** is **0.3732**. Since this value is considerably larger than our

threshold (0.3732 ge 0.05), we must fail to reject the null hypothesis for **x2**. We conclude that **x2** does not demonstrate a [statistically significant](#) linear relationship with **y** at the 0.05 level, suggesting it could potentially be removed from a refined model.

Mathematical Derivation of $\Pr(>|t|)$

For analysts seeking a deeper understanding beyond simple interpretation, it is crucial to examine the underlying mathematical processes that generate the $\Pr(>|t|)$ value. This probability is not directly calculated from the raw data; rather, it is derived from the **t-statistic**, which itself is based on the coefficient estimate and its standard error. Understanding this two-step process--calculating the t-statistic and then finding the corresponding probability--provides complete clarity on the regression output.

The initial step involves calculating the **t value** for each coefficient. The t-statistic essentially measures how many standard deviations the estimated coefficient is away from the hypothesized value of zero (the null hypothesis). A large absolute t-value implies that the estimate is far from zero, suggesting a potentially significant effect. This statistic is computed as a simple ratio: the coefficient **Estimate** divided by its **Standard Error**.

The fundamental formula used to derive this critical test statistic is:

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

We can confirm this calculation using the figures for **x1** from our example (Estimate: 1.4758, Std. Error: 0.5029). The resulting t-value calculation confirms the figure provided in the R summary table:

```
#calculate t-value
```

```
1.4758 / .5029
```

```
2.934579
```

Mapping the T-Statistic to Probability via the T-Distribution

Once the **t value** is successfully established, the subsequent and final step is calculating the $\Pr(>|t|)$. This is achieved by reference to the [Student's t-distribution](#). The resulting probability represents the area in the tails of the distribution--the likelihood that a random value drawn from the [t-distribution](#), with the correct degrees of freedom, would be greater than the absolute magnitude of the calculated t-statistic.

It is important to remember that [hypothesis testing](#) for regression coefficients usually involves a

two-tailed test. This means we are testing for significance in both the positive and negative directions (i.e., whether the coefficient is significantly greater than zero OR significantly less than zero). Consequently, the probability calculated for one tail of the distribution must be doubled to obtain the final two-tailed p-value.

The generalized formula employed by R's statistical functions to determine the two-tailed p-value is as follows, where `residual df` refers to the residual degrees of freedom:

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

Crucially, this calculation relies on the **residual degrees of freedom** (df). Returning to our example, the summary output clearly indicates this value is 5:

Residual standard error: 1.867 on 5 degrees of freedom

Using the calculated t-value of 2.935 and 5 residual degrees of freedom, we can use the R function `pt` to confirm the final probability:

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

```
0.0324441
```

This computational verification reveals that the derived p-value, 0.0324441, precisely matches the 0.0325 reported in the model's coefficients table (allowing for rounding). This confirms the robust mathematical methodology underpinning the regression results.

Conclusion: The Definitive Metric for Significance

The **Pr(>|t|)** column serves as the definitive indicator for assessing the statistical merit of individual predictor variables within a [linear regression](#) model. By providing the associated probability (the p-value) that the true coefficient is zero, it empowers researchers to make objective decisions regarding variable inclusion.

A robust understanding of how to compare this probability against the established significance level (α) is non-negotiable for effective statistical practice. Mastery of this output column not only facilitates rigorous model validation but also ensures accurate and meaningful reporting of scientific findings, transforming raw data into reliable insights.