

How to Interpret Regression Output in R?

Authored by
stats writer

December 17, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Interpret Regression Output in R?*. PSYCHOLOGICAL SCALES.
Retrieved from <https://scales.arabpsychology.com/?p=107683>

Understanding the output generated by a statistical programming environment is the cornerstone of effective data analysis. When performing regression analysis in the R programming language, the `summary()` function provides a wealth of information necessary to judge model quality, variable relationships, and statistical significance.

A thorough interpretation of the R regression output necessitates familiarity with key statistical metrics, including the R-squared value, the estimated coefficients, the corresponding standard errors, and various test statistics. These elements collectively offer profound insights into the underlying mechanisms driving the relationships between the predictor and response variables.

This comprehensive guide details the process of fitting a linear regression model in R using the fundamental **lm()** command, followed by an in-depth, section-by-section walkthrough of the resulting output generated by the **summary()** command. Mastering these interpretations is essential for making robust, data-driven decisions based on your statistical models.

To fit a linear regression model in R, we utilize the **lm()** command.

To reveal the full statistical details of the fitted model, we then apply the **summary()** command to the model object.

This tutorial explains how to interpret every critical value provided in the R regression output summary.

Setting Up the Multiple Linear Regression Model in R

To illustrate the interpretation process, we will employ a standard multiple linear regression scenario using R's built-in **mtcars** dataset. This dataset is commonly used for demonstrating statistical methods and contains 32 observations on 11 variables related to automobile design and performance.

Our objective is to model miles per gallon (*mpg*), which serves as the response variable, based on three chosen predictor variables: gross horsepower (*hp*), rear axle ratio (*drat*), and weight (*wt*). We initialize the model using the **lm()** function and subsequently request the detailed model output using the **summary()** function.

The code block below outlines the commands necessary to fit this model and display the full statistical summary, which we will dissect in the subsequent sections. Note the use of the formula structure `mpg ~ hp + drat + wt`, which explicitly defines the relationship we are testing within the **mtcars** data frame:

```
#fit regression model using hp, drat, and wt as predictors
```

```
model <- lm(mpg ~ hp + drat + wt, data = mtcars)
```

```
#view model summary
summary(model)
```

Call:

```
lm(formula = mpg ~ hp + drat + wt, data = mtcars)
```

Residuals:

```
Min 1Q Median 3Q Max
```

```
-3.3598 -1.8374 -0.5099 0.9681 5.7078
```

Coefficients:

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

```
(Intercept) 29.394934 6.156303 4.775 5.13e-05 ***
```

```
hp -0.032230 0.008925 -3.611 0.001178 **
```

```
drat 1.615049 1.226983 1.316 0.198755
```

```
wt -3.227954 0.796398 -4.053 0.000364 ***
```

```
---
```

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

```
Residual standard error: 2.561 on 28 degrees of freedom
```

```
Multiple R-squared: 0.8369, Adjusted R-squared: 0.8194
```

```
F-statistic: 47.88 on 3 and 28 DF, p-value: 3.768e-11
```

Interpreting the Model Call Section

Call:

```
lm(formula = mpg ~ hp + drat + wt, data = mtcars)
```

The first element displayed in the R regression summary output is the **Call**. This section is essentially a verification step, reiterating the exact function and arguments used to generate the model. It confirms the structure of the model formula, which is critical for documentation and reproducibility. In our example, `lm(formula = mpg ~ hp + drat + wt, data = mtcars)` clearly states that we fitted a linear model (`lm`) predicting miles per gallon (`mpg`) based on the combined effects of horsepower (`hp`), rear axle ratio (`drat`), and weight (`wt`), all sourced from the `mtcars` data frame.

Identifying the variables used is straightforward here: `mpg` is the dependent or response variable, positioned to the left of the tilde (`~`), and `hp + drat + wt` are the independent or predictor

variables, located on the right. This clear restatement ensures that analysts immediately know the scope and composition of the tested model before delving into the statistical results, providing the foundational context for the entire summary output.

Analyzing the Distribution of Residuals

Residuals:

Min 1Q Median 3Q Max

-3.3598 -1.8374 -0.5099 0.9681 5.7078

The **Residuals** section provides a five-number summary (minimum, first quartile (1Q), median, third quartile (3Q), and maximum) of the model's residuals. A residual is defined as the difference between the actual observed value of the response variable and the value predicted by the fitted regression line. Understanding the distribution of these residuals is crucial for assessing one of the core assumptions of linear regression: that the errors are normally distributed and centered around zero.

In a well-fitting model where the assumptions hold true, we expect the median residual to be close to zero, and the first and third quartiles (1Q and 3Q) should be roughly symmetrical around that median. Our output shows a median residual of **-0.5099**, which is reasonably close to zero, suggesting the model is not systematically over- or under-predicting across the entire data range. However, we also observe that the minimum residual (**-3.3598**) and the maximum residual (**5.7078**) are not perfectly symmetrical around zero, indicating that there may be some larger positive outliers or that the distribution is slightly skewed.

Specifically, the difference between the median and the minimum is approximately 2.85, while the difference between the maximum and the median is about 6.22, confirming a positive skew. While this summary is a quick diagnostic tool, formal residual plots (such as a Q-Q plot or a histogram of residuals) should always be consulted for a definitive assessment of the error distribution and homoscedasticity assumptions.

Interpreting Regression Coefficients and the Model Equation

Coefficients:

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

(Intercept) 29.394934 6.156303 4.775 5.13e-05 ***

hp -0.032230 0.008925 -3.611 0.001178 **

drat 1.615049 1.226983 1.316 0.198755

wt -3.227954 0.796398 -4.053 0.000364 ***

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

The **Coefficients** table is arguably the most important part of the regression summary, providing the actual structure of the estimated model equation. The first column, **Estimate**, lists the calculated values for the intercept (often denoted as β_0) and the slope coefficients (β_1 , β_2 , dots) for each predictor variable. These estimates allow us to write the predicted model equation: $\hat{mpg} = 29.395 - 0.032 \text{ times } hp + 1.615 \text{ times } drat - 3.228 \text{ times } wt$.

Each coefficient represents the average change in the response variable associated with a one-unit increase in that specific predictor, assuming all other predictor variables in the model are held constant (*ceteris paribus*). For example, the coefficient for **hp** is **-0.032230**, indicating that for every one-unit increase in horsepower, the predicted fuel efficiency (mpg) decreases by 0.032230 units, holding *drat* and *wt* constant. Conversely, the **(Intercept)** value of **29.394934** represents the predicted value of *mpg* when all predictor variables (*hp*, *drat*, and *wt*) are zero--though this often lacks practical meaning if zero is outside the sensible range of the data.

Assessing Significance of Individual Predictors

To determine the reliability and statistical importance of these estimates, we examine the remaining columns. The **Std. Error** quantifies the uncertainty of the coefficient estimate; a smaller standard error suggests a more precise estimate. The **t value** is the test statistic, calculated by dividing the Estimate by its corresponding Standard Error. This value is used to test the null hypothesis that the true population coefficient is zero (i.e., that the predictor has no linear relationship with the response variable).

The final column, **Pr(>|t|)**, provides the p-value associated with the t-statistic. If this p-value is below a pre-determined significance level, typically $\alpha = 0.05$, we reject the null hypothesis and conclude that the predictor is statistically significant. In our example, using $\alpha = 0.05$, both **hp** (p-value: 0.001178) and **wt** (p-value: 0.000364) are highly significant predictors, indicated by the asterisks. However, **drat** (p-value: 0.198755) is not statistically significant, suggesting we lack sufficient evidence to conclude it has a meaningful effect on *mpg* when *hp* and *wt* are already included in the model.

Evaluating Model Accuracy: Residual Standard Error and R-squared

Residual standard error: 2.561 on 28 degrees of freedom

Multiple R-squared: 0.8369, Adjusted R-squared: 0.8194

The section dedicated to model fit metrics begins with the **Residual standard error (RSE)**. This

value, **2.561** in our case, estimates the standard deviation of the error term (σ). Conceptually, the RSE represents the average distance that the observed response values (*mpg*) fall from the fitted regression line. This metric is expressed in the units of the response variable; therefore, we can state that the typical prediction error of our model is approximately 2.561 miles per gallon. The associated **degrees of freedom** (28) are calculated as $n - k - 1$, where n is the total number of observations (32 in **mtcars**) and k is the number of predictor variables (3). A smaller RSE value is always desirable, as it indicates a tighter fit of the model to the observed data points.

Next, the summary provides two measures of the goodness-of-fit. The **Multiple R-squared**, or the coefficient of determination, indicates the proportion of the total variance in the response variable that is explained by the predictor variables included in the model. Our value of **0.8369** suggests that 83.69% of the variation in *mpg* can be accounted for by *hp*, *drat*, and *wt* combined. Since R-squared always increases when new variables are added, regardless of their statistical relevance, it is generally recommended to look at the adjusted version for comparison purposes.

The **Adjusted R-squared** attempts to correct this inflation by penalizing the inclusion of unnecessary predictor variables. It is always slightly lower than the Multiple R-squared, reporting **0.8194** here. The adjusted R-squared can be useful when comparing the performance of two or more regression models that are built on the same dataset but contain different numbers of predictors, serving as a more reliable indicator of model efficacy outside of the sample data.

Testing Overall Model Utility: F-Statistic

F-statistic: 47.88 on 3 and 28 DF, p-value: 3.768e-11

The final section addresses the overall utility of the entire regression equation using the **F-statistic**. This test performs a global assessment, testing the null hypothesis that all regression coefficients (except the intercept) are simultaneously equal to zero. Essentially, it determines whether the model containing the predictors provides a statistically superior fit to the data compared to a simple intercept-only model.

The calculated **F-statistic** is **47.88**, distributed on 3 and 28 degrees of freedom. The corresponding p-value is extremely small, **3.768e-11** (which is 3.768×10^{-11}). Since this p-value is far below any common significance threshold (like $\alpha = 0.05$), we decisively reject the null hypothesis. This powerful rejection indicates that at least one of the predictor variables (*hp*, *drat*, or *wt*) is significantly useful in predicting *mpg*, validating the overall strength and effectiveness of the multiple linear regression model built on the **mtcars** data.

[How to Perform Simple Linear Regression in R](#)

How to Perform Multiple Linear Regression in R

What is a Good R-squared Value?

ARABPSYCHOLOGY.COM