

Understanding and Interpreting Linear Regression Output in R

Authored by
Mohammed loot

November 6, 2025

RECOMMENDED CITATION

Mohammed loot (2025). *Understanding and Interpreting Linear Regression Output in R*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=11620>

Mastering the interpretation of statistical output is perhaps the most critical step in applied [data analysis](#). When working within the [R](#) environment, fitting a [linear regression model](#) is straightforwardly achieved using the built-in **lm()** command.

However, the complexity arises not in running the model, but in understanding the comprehensive statistical report generated by piping the model object through the **summary()** function. This output contains all the necessary metrics--from individual coefficient significance to overall model explanatory power--required to draw robust conclusions about the relationships between variables.

This authoritative guide is designed to serve as a comprehensive reference, offering a detailed, step-by-step methodology for dissecting and interpreting every segment of the standard regression summary provided by the statistical software [R](#). Correct interpretation is fundamental for evaluating model validity and effectiveness.

Setting Up the Regression Analysis in R

To provide a clear, practical demonstration of the interpretation process, we will employ a standard dataset readily available within the [R](#) package: the **mtcars** dataset. This dataset contains information on 32 automobiles, and we will use it to construct a multiple linear regression model.

Our objective is to predict *mpg* (miles per gallon), which serves as our response variable, based on three specific predictor variables: *hp* (gross horsepower), *drat* (rear axle ratio), and *wt* (vehicle weight). This selection allows us to examine how multiple factors simultaneously influence fuel efficiency.

The following code block illustrates the exact syntax used in R to fit this specific multiple regression model and subsequently generate the detailed statistical summary that we will be analyzing:

```
#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

```

The subsequent sections will systematically dissect this output, explaining the meaning, statistical implication, and practical significance of every quantitative metric displayed.

Section 1: Confirming Model Specification (The Call)

Call:

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

The **Call** section, positioned at the very top of the output, serves an essential function: it provides a precise echo of the command executed to generate the model. This is more than just a formality; it acts as a critical validation step.

By reviewing the **Call**, the analyst confirms that the statistical software correctly processed the intended model specification. This is particularly important in complex analyses involving variable transformations or interaction terms, ensuring that the defined response variable (**mpg**) and the set of predictor variables (**hp**, **drat**, and **wt**) were correctly incorporated from the specified dataset (**mtcars**).

Section 2: Analyzing the Distribution of Residuals

Residuals:

```

Min 1Q Median 3Q Max
-3.3598 -1.8374 -0.5099 0.9681 5.7078

```

This section summarizes the distribution of the model's [residuals](#). A [residual](#) is the difference between the observed value of the response variable and the value predicted by the fitted regression line. In essence, residuals represent the error inherent in the model's predictions.

The output provides five key summary statistics: the minimum, the first quartile (1Q), the median, the third quartile (3Q), and the maximum residual. Analyzing these values offers a rapid diagnostic check of the model's underlying assumptions.

For a reliable linear model, the distribution of [residuals](#) should ideally be symmetric around zero. Consequently, the **Median** residual value should be very close to zero. Our median, **-0.5099**, is relatively small. Furthermore, we check the symmetry by comparing the absolute values of the minimum (**3.3598**) and maximum (**5.7078**). The noticeable difference here suggests a slight right skew in the error distribution, indicating that the model tends to underpredict more significantly than it overpredicts, possibly warranting further investigation using diagnostic plots like the Q-Q plot and Residuals vs. Fitted plot.

Section 3: Interpreting Individual Predictor Coefficients

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 the core of the summary, providing the estimated parameters for the regression equation and the results of individual hypothesis tests for each variable. This table allows us to quantify the relationship between each predictor and the response.

The fitted regression equation derived from these estimates is: **mpg = 29.39 - 0.032(hp) + 1.615(drat) - 3.228(wt)**.

The table provides detailed statistics for each term:

Estimate: This is the calculated regression coefficient. For the intercept, it is the predicted value of MPG when all predictors are zero. For the slope coefficients (e.g., **hp**), it signifies the expected change in MPG for a one-unit increase in HP, assuming all other predictors are held constant (*ceteris paribus*). For example, every one-unit increase in HP is associated with a decrease of 0.03223 MPG.

Std. Error: This metric measures the precision of the coefficient estimate. A smaller standard error relative to the estimate indicates higher precision and reliability of the estimated slope.

t value: The t-statistic tests the null hypothesis (H_0) that the true coefficient is zero (meaning

the predictor has no effect). It is calculated by dividing the Estimate by the Standard Error.

Pr(>|t|): This is the [p-value](#) associated with the t-statistic. It represents the probability of observing a t-value as extreme as the one calculated, assuming the null hypothesis is true.

Using a conventional significance level (α) of 0.05, we evaluate the [p-values](#). Both **hp** ($p < 0.001$) and **wt** ($p < 0.001$) have highly [statistically significant](#) relationships with MPG, as confirmed by the '***' codes. Conversely, **drat**, with a high [p-value](#) of 0.198755, fails to demonstrate [statistical significance](#) in the presence of **hp** and **wt**.

Section 4: Global Metrics for Model Fit Assessment

The final block of the regression summary provides essential global statistics necessary to evaluate the overall predictive quality and explanatory power of the entire regression model.

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

Three primary metrics define the model's fitness:

Residual Standard Error (RSE): The RSE, 2.561, is often referred to as the standard deviation of the [residuals](#). It is a measure of the average distance between the observed values and the regression line (in the units of the response variable, MPG). A lower RSE suggests that the model's predictions are, on average, closer to the actual data points, indicating a better fit. The degrees of freedom (28) are calculated as the total number of observations ($n=32$) minus the number of parameters estimated (the intercept plus three predictors, $k+1=4$).

Multiple R-squared: This is the [coefficient of determination](#). It quantifies the proportion of the total variability in the response variable (MPG) that is explained by the predictor variables included in the model. A value of **0.8369** means that approximately 83.7% of the variation in MPG can be attributed to the combined effects of horsepower, axle ratio, and weight. Higher values generally indicate a stronger explanatory model.

Adjusted R-squared: The Adjusted [R-squared](#) is a modification designed to penalize the model for including excessive or unnecessary predictor variables. Unlike the Multiple R-squared, which always increases when a new variable is added, the Adjusted [R-squared](#) will only increase if the new predictor improves the model more than would be expected by chance. This metric is crucial when comparing models with different numbers of predictors.

Section 5: Testing Overall Model Significance (The F-Test)

The final component of the summary output assesses the overall validity of the model through a

global hypothesis test using the [F-statistic](#).

The [F-statistic](#) tests the null hypothesis that all regression coefficients (except the intercept) are simultaneously equal to zero. If this null hypothesis holds, the model is essentially useless because none of the predictors, as a group, contribute significantly to explaining the response variable.

F-statistic Value: The calculated value (47.88) is compared against the F-distribution based on the model's degrees of freedom.

p-value (for F-statistic): The corresponding [p-value](#) (3.768×10^{-11}) is extremely small. Since this value is far below any conventional significance threshold (like 0.05), we decisively reject the null hypothesis.

The rejection of the null hypothesis confirms that at least one of the predictor variables in the model is significantly related to the response variable. This result validates the model as a whole, establishing that the combined set of predictors (HP, DRAT, and WT) is highly [statistically significant](#) and provides predictive power beyond simply relying on the mean of MPG.

Further Resources for Advanced R Modeling

To continue developing your expertise in statistical modeling and deepen your understanding of these complex outputs, the following related resources are recommended:

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

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

[What is a Good R-squared Value?](#)