

What is the interpretation of the regression output in R?

Authored by
stats writer

April 23, 2024

RECOMMENDED CITATION

stats writer (2024). *What is the interpretation of the regression output in R?*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=138303>

The interpretation of the regression output in R refers to the analysis and understanding of the statistical results generated from a regression model. This includes the coefficients, p-values, confidence intervals, and other relevant statistics that help to explain the relationship between the variables in the model. The output can be used to make predictions, identify significant factors, and evaluate the overall fit of the model. It is an important step in the regression analysis process and provides valuable insights into the relationship between the variables being studied.

Interpret Regression Output in R

To fit a linear regression model in R, we can use the `lm()` command.

To view the output of the regression model, we can then use the `summary()` command.

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

Example: Interpreting Regression Output in R

The following code shows how to fit a multiple linear regression model with the built-in `mtcars` dataset using `hp`, `drat`, and `wt` as predictor variables and `mpg` as the response variable:

```
#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

Here is how to interpret every value in the output:

Call

Call:

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

This section reminds us of the formula that we used in our regression model. We can see that we used mpg as the response variable and hp, drat, and wt as our predictor variables. Each variable came from the dataset called mtcars.

Residuals

Residuals:

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

This section displays a summary of the distribution of residuals from the regression model. Recall that a residual is the difference between the observed value and the predicted value from the regression model.

The minimum residual was -3.3598, the median residual was -0.5099 and the max residual was 5.7078.

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

This section displays the estimated coefficients of the regression model. We can use these coefficients to form the following estimated regression equation:

For each predictor variable, we're given the following values:

Estimate: The estimated coefficient. This tells us the average increase in the response variable associated with a one unit increase in the predictor variable, assuming all other predictor variables are held constant.

Std.Error: This is the standard error of the coefficient. This is a measure of the uncertainty in our estimate of the coefficient.

t value: This is the t-statistic for the predictor variable, calculated as (Estimate) / (Std. Error).

Pr(>|t|): This is the p-value that corresponds to the t-statistic. If this value is less than some alpha level (e.g. 0.05) than the predictor variable is said to be statistically significant.

If we used an alpha level of $\alpha = .05$ to determine which predictors were significant in this regression model, we'd say that hp and wt are statistically significant predictors while drat is not.

Assessing Model Fit

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

This last section displays various numbers that help us assess how well the regression model fits our dataset.

Residual standard error: This tells us the average distance that the observed values fall from the regression line. The smaller the value, the better the regression model is able to fit the data.

The degrees of freedom is calculated as $n-k-1$ where n = total observations and k = number of predictors. In this example, mtcars has 32 observations and we used 3 predictors in the regression model, thus the degrees of freedom is $32 - 3 - 1 = 28$.

Multiple R-Squared: This is known as the coefficient of determination. It tells us the proportion of the variance in the response variable that can be explained by the predictor variables.

This value ranges from 0 to 1. The closer it is to 1, the better the predictor variables are able to predict the value of the response variable.

Adjusted R-squared: This is a modified version of R-squared that has been adjusted for the number of predictors in the model. It is always lower than the R-squared.

The adjusted R-squared can be useful for comparing the fit of different regression models that use different numbers of predictor variables.

F-statistic: This indicates whether the regression model provides a better fit to the data than a model that contains no independent variables. In essence, it tests if the regression model as a whole is useful.

p-value: This is the p-value that corresponds to the F-statistic. If this value is less than some significance level (e.g. 0.05), then the regression model fits the data better than a model with no predictors.

When building regression models, we hope that this p-value is less than some significance level because it indicates that the predictor variables are actually useful for predicting the value of the response variable.

How to Perform Simple Linear Regression in R

How to Perform Multiple Linear Regression in R

What is a Good R-squared Value?