

How to Interpret $Pr(>|z|)$ in Logistic Regression Output in R

Authored by
stats writer

November 30, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Interpret $Pr(>|z|)$ in Logistic Regression Output in R*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=102669>

The term **Pr(>|z|)**, commonly found in the summary output of a generalized linear model (GLM) in R, is one of the most crucial yet frequently misunderstood elements of statistical modeling. In essence, **Pr(>|z|)** represents the two-sided p-value associated with the z-test statistic for each estimated regression coefficient. This value is indispensable for determining the statistical reliability and predictive power of each independent variable within the model.

When conducting a logistic regression analysis, our primary goal is to assess whether the predictors significantly influence the outcome variable. The **Pr(>|z|)** column provides the mechanism for this assessment. Specifically, it tests the null hypothesis (H_0) that the true population coefficient for that predictor is zero, meaning the variable has no effect on the log-odds of the outcome. A small p-value allows us to reject this null hypothesis, confirming the variable's significance.

Understanding the Components of R's GLM Summary

Whenever you execute a logistic regression model using the `glm()` function in R, specifying `family=binomial`, the resulting summary output presents a detailed breakdown of the model's performance. The coefficients table is the heart of this output, displaying four key columns that summarize the impact of each predictor variable on the model outcome.

These columns--Estimate, Std. Error, z value, and Pr(>|z|)--work together to provide a complete picture of hypothesis testing for individual predictors. Understanding how they interrelate is fundamental to proper statistical inference. The **Estimate** gives us the magnitude and direction of the effect, while the **Std. Error** quantifies the uncertainty around that estimate. The **z value** is the standardized test statistic, and ultimately, **Pr(>|z|)** translates this test statistic into a readily interpretable probability value.

The following structure illustrates the typical coefficient summary generated by R:

Coefficients:

```
Estimate Std. Error z value Pr(>|z|)
(Intercept) -17.638452 9.165482 -1.924 0.0543 .
disp -0.004153 0.006621 -0.627 0.5305
drat 4.879396 2.268115 2.151 0.0315 *
```

The **Pr(>|z|)** column represents the p-value derived from the corresponding **z value**. A primary focus in statistical analysis is comparing this calculated probability against a predefined significance level (alpha, often $\alpha = 0.05$). This comparison dictates whether we can conclude that the predictor variable contributes meaningfully to the prediction of the response variable in the context of the model.

The Statistical Foundation: The Z-Test Statistic

The **z value** in the R output is the actual z-test statistic used for evaluating the significance of the individual coefficient. In logistic regression, we employ a specific application of the z-test, which is calculated by taking the ratio of the estimated coefficient to its standard error. Mathematically, this is expressed as: $Z = \frac{\text{Estimate}}{\text{Standard Error}}$.

Because logistic regression models are often fit using maximum likelihood estimation, the sampling distribution of the resulting coefficients is generally assumed to approximate a Normal distribution, especially when the sample size is sufficiently large. This assumption allows us to use the standard z-test statistic, which compares the calculated Z-score to the standard normal distribution to determine the extremity of the observed estimate.

The magnitude of the **z value** reflects how many standard errors the estimated coefficient is away from zero. A larger absolute **z value** indicates a stronger deviation from zero, making it less likely that the true coefficient is zero and thus leading to a smaller p-value. The calculation of $\Pr(>|z|)$ depends directly on this z-score, representing the probability of observing a z-score as extreme or more extreme than the one calculated, assuming the null hypothesis holds true.

Determining Statistical Significance Using $\Pr(>|z|)$

The core purpose of examining $\Pr(>|z|)$ is to perform hypothesis testing. As previously mentioned, we test the null hypothesis (H_0) that the predictor variable has no linear relationship with the log-odds of the outcome (i.e., $\beta = 0$). The alternative hypothesis (H_a) is that the predictor does have an effect ($\beta \neq 0$). Since the test is concerned with effects in either the positive or negative direction, the test is inherently two-tailed, which is reflected by the use of the absolute value symbol ($|z|$) in the notation $\Pr(>|z|)$.

To establish statistical significance, we must compare the calculated $\Pr(>|z|)$ value against our predetermined significance level, denoted as α . The standard convention in most fields is to set α at 0.05. If the calculated p-value is smaller than the threshold ($\Pr(>|z|) < \alpha$), we reject the null hypothesis. This rejection signifies that the estimated coefficient is statistically significant, meaning the predictor variable is a valuable and reliable addition to the predictive model.

Conversely, if $\Pr(>|z|)$ is greater than or equal to α (e.g., $0.5305 > 0.05$), we fail to reject the null hypothesis. This does not necessarily mean the predictor has absolutely no effect, but rather that the available data does not provide sufficient evidence, at the chosen significance level, to conclude that the population coefficient is different from zero. Variables that are not statistically significant are often candidates for removal during model refinement or simplification.

Example: How to Interpret $\Pr(>|z|)$ Values in Practice

To solidify the understanding of $\Pr(>|z|)$ interpretation, let us examine a concrete example using R's built-in `mtcars` dataset. We will fit a logistic regression model predicting the type of transmission (`am`) based on engine displacement (`disp`) and rear axle ratio (`drat`).

The following R code snippet demonstrates how to fit the model and generate the comprehensive summary output:

```
#fit logistic regression model
model <- glm(am ~ disp + drat, data=mtcars, family=binomial)
```

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

Call:

```
glm(formula = am ~ disp + drat, family = binomial, data = mtcars)
```

Deviance Residuals:

Min 1Q Median 3Q Max

```
-1.5773 -0.2273 -0.1155 0.5196 1.8957
```

Coefficients:

Estimate Std. Error z value $\Pr(>|z|)$

```
(Intercept) -17.638452 9.165482 -1.924 0.0543 .
```

```
disp -0.004153 0.006621 -0.627 0.5305
```

```
drat 4.879396 2.268115 2.151 0.0315 *
```

```
---
```

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 43.230 on 31 degrees of freedom

Residual deviance: 21.268 on 29 degrees of freedom

AIC: 27.268

Number of Fisher Scoring iterations: 6

By focusing specifically on the results for the predictor variables `disp` and `drat`, we can interpret their statistical relevance based on their respective $\Pr(>|z|)$ values, assuming a standard significance level (α) of 0.05:

The p-value for the predictor variable "disp" is **0.5305**. Since this value is significantly greater than the 0.05 threshold ($0.5305 > 0.05$), we fail to reject the null hypothesis. This indicates that `disp` does not have a statistically significant relationship with the probability of having a manual transmission in this model.

The p-value for the predictor variable "drat" is **0.0315**. Since this value is less than the 0.05 threshold ($0.0315 < 0.05$), we reject the null hypothesis. This result confirms that `drat` has a statistically significant relationship with the response variable, suggesting it is a reliable predictor in the model.

The Importance of Significance Codes

The Significance Codes under the coefficient table provide an immediate visual cue regarding the strength of the evidence against the null hypothesis. These codes map p-values to symbols: $0 < p < 0.0001$ '***', $0.0001 < p < 0.001$ '**', $0.001 < p < 0.01$ '*', $0.01 < p < 0.05$ '.', $0.05 < p < 0.1$ ' ', $p > 0.1$ '1'. This system allows the analyst to immediately gauge the degree of statistical significance achieved by each predictor.

For the `drat` variable in the example above, the single asterisk (*) next to the p-value of 0.0315 means the p-value is statistically significant at the $\alpha = 0.05$ level. Had the p-value been, say, 0.0005, it would have received three asterisks (***), indicating significance at a much stricter threshold ($\alpha = 0.001$), thereby showing stronger evidence against the null hypothesis.

It is critical, however, to interpret these codes alongside the actual $\Pr(>|z|)$ value and the estimated coefficient. While a small p-value indicates statistical significance, it does not inherently guarantee practical or clinical significance. A variable might be statistically significant ($p < 0.05$) yet have a negligible effect size (a very small Estimate), particularly in large datasets. Therefore, a comprehensive analysis requires considering the Estimate and its Standard Error along with the calculated $\Pr(>|z|)$.

A Detailed Look at the Calculation of $\Pr(>|z|)$

Understanding the underlying calculation reinforces the theoretical basis of the z-test statistic and the resulting p-value. The process involves two primary steps: calculating the standardized test statistic (the z value) and then converting that statistic into a two-tailed probability (the $\Pr(>|z|)$ value).

Step 1: Calculating the Z Value

First, we calculate the **z value** using the following relationship, which serves as the measure of evidence against the null hypothesis:

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

For example, here's how to calculate the z value for the predictor variable "drat":

```
#calculate z-value  
4.879396 / 2.268115
```

```
2.151
```

The resulting z value is 2.151, which matches the value reported in the model summary output for `drat`.

Step 2: Calculating the Two-Tailed P-Value

Next, we calculate the two-tailed p-value. This represents the probability that the absolute value of a random variable drawn from the standard Normal distribution is greater than 2.151 or less than -2.151. This area represents the probability of observing a result as extreme as ours, assuming the true effect is zero.

We use the following formula in R, utilizing the `pnorm()` function, which calculates the cumulative probability up to the specified Z-score:

```
p-value = 2 * (1-pnorm(z value))
```

For example, here's how to calculate the two-tailed p-value for a z-value of 2.151:

```
#calculate p-value  
2*(1-pnorm(2.151))
```

```
0.0314762
```

Notice that this p-value of approximately 0.0315 precisely matches the **Pr(>|z|)** reported in the regression output from above, confirming the statistical mechanism behind the output.

Conclusion: Synthesizing the Results

The **Pr(>|z|)** column is the definitive metric for assessing individual predictor significance in logistic regression models generated in R. It summarizes the findings of the hypothesis test by providing a clear, interpretable probability that can be directly compared against the established significance level.

Mastering the interpretation of **Pr(>|z|)** is essential for any analyst working with GLMs. A value below 0.05 indicates statistical significance, suggesting strong evidence that the predictor variable contributes meaningfully to the model's ability to predict the binary outcome. Conversely, a high p-

value warrants cautious interpretation of that variable's effect. Always remember that statistical significance must be evaluated alongside the practical implications of the coefficient magnitude (the Estimate) to ensure the variable is not only statistically reliable but also relevant to the domain context.

ARABPSYCHOLOGY.COM