

Understanding $Pr(>|z|)$ Values in Logistic Regression Output Using R

Authored by
Mohammed looti

October 31, 2025

RECOMMENDED CITATION

Mohammed looti (2025). *Understanding $Pr(>|z|)$ Values in Logistic Regression Output Using R*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=6971>

When performing [logistic regression](#) analysis, particularly within the powerful statistical environment of [R](#), the ability to accurately interpret the generated output is essential for deriving meaningful and actionable conclusions. Unlike its linear counterpart, [logistic regression](#) is specifically designed to model binary or categorical outcomes, estimating the probability of a specific event occurring. The summary output generated by R's generalized linear model function (`glm`) provides a rigorous statistical overview of the modeled relationships between the [predictor variables](#) and the binary [response variable](#).

Within this comprehensive summary, the coefficients table contains a critical metric often labeled as [Pr\(>|z|\)](#). This specific column serves as the primary indicator of statistical significance for each independent variable. For researchers and data analysts, deciphering this value determines which predictors possess a robust and non-random association with the outcome being modeled. A precise understanding of [Pr\(>|z|\)](#) is therefore fundamental to ensuring the integrity and reliability of the reported results from any [logistic regression](#) study.

This expert guide offers a deep dive into the definition, calculation, and practical interpretation of [Pr\(>|z|\)](#) within the [R](#) environment. We will break down the structure of the coefficient table and use clear examples to illuminate how this critical p-value is utilized in statistical inference. The following display shows the typical format for the coefficients section in a standard R regression model summary:

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 *
```

Dissecting the Logistic Regression Output in R

The "Coefficients" table, a central feature of the [R](#) regression summary, systematically organizes the impact of every [predictor variable](#) included in the model. Each row is dedicated to either the intercept or a specific predictor, and the columns provide the metrics necessary to evaluate both the direction and the statistical reliability of that variable's effect. A thorough interpretation requires a clear grasp of the function of each column.

The first key column, **Estimate**, provides the coefficient for the predictor. In the context of [logistic regression](#), this value represents the change in the [log-odds](#) of the outcome occurring for a one-unit increase in the predictor, assuming all other variables in the model remain constant. The **Std. Error (Standard Error)** is critical for assessing the stability and precision of the Estimate; a smaller [Standard Error](#) implies greater certainty regarding the true population parameter. These two

components are combined to produce the **z value**, which serves as the test statistic for the individual coefficient.

The **z value** is calculated simply by dividing the Estimate by the **Standard Error** ($z = \text{Estimate} / \text{Std. Error}$). This statistic measures how many standard errors the coefficient is away from zero. This brings us to the final, and perhaps most crucial, column: **Pr(>|z|)**. This value is the corresponding **p-value** derived from the calculated **z value**. The **p-value** represents the probability of observing a test statistic as extreme as, or more extreme than, the z-statistic calculated, assuming the **null hypothesis** is true. Because the test is two-tailed (considering effects both positive and negative), the notation $>|z|$ signifies that we are looking at the probability in both tails of the distribution.

Pr(>|z|) as the Cornerstone of Statistical Inference

The core purpose of the **Pr(>|z|)** column is to facilitate **statistical significance** testing for each coefficient in the **logistic regression** model. This process is rooted in the framework of hypothesis testing, where two competing statements are evaluated: the **null hypothesis** (H_0) and the **alternative hypothesis** (H_A). In regression, H_0 posits that the true coefficient for the **predictor variable** is exactly zero, implying no relationship exists between the predictor and the **response variable**. H_A , conversely, suggests that the true coefficient is non-zero, indicating a meaningful association.

To make a decision regarding H_0 , the calculated **p-value** ($\text{Pr}(>|z|)$) is compared against a predefined threshold known as the **significance level**, typically denoted by the Greek letter alpha (α). This α value represents the maximum risk a researcher is willing to take of committing a Type I error--that is, incorrectly rejecting a true **null hypothesis**. While $\alpha = 0.05$ is the academic standard, choices like 0.01 (for higher rigor) or 0.10 (for exploratory analysis) are also utilized depending on the field of study and the consequences of error.

The decision rule is clear and absolute: if the **p-value** is strictly less than the chosen **significance level** ($\text{Pr}(>|z|) < \alpha$), we possess sufficient evidence to reject the **null hypothesis**. This acceptance of H_A signifies that the **predictor variable** is deemed **statistically significant** in influencing the **response variable**. Conversely, if the **p-value** is greater than or equal to α , we "fail to reject" H_0 , concluding that the data does not provide adequate evidence to support a statistically reliable relationship.

Practical Example: Interpreting `Pr(>|z|)` with `mtcars` Dataset

To solidify the interpretation of **Pr(>|z|)**, we turn to a concrete example using the widely recognized built-in **mtcars** dataset in **R**. Our objective is to construct a **logistic regression model** aimed at predicting the type of transmission (0 for automatic, 1 for manual, stored in the 'am' variable). We

will use two continuous factors as **predictor variables**: engine displacement ('disp') and rear axle ratio ('drat').

The following **R** code executes the model fit using the `glm()` function, specifying `family=binomial` to correctly implement **logistic regression**. The subsequent summary output allows us to directly examine the coefficient table, paying close attention to the **Pr(>|z|)** column for each predictor.

#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

Assuming a standard **significance level** of $\alpha = 0.05$, we analyze the **Pr(>|z|)** results for our two **predictor variables** using the decision rule established earlier:

For the **predictor variable** "**disp**" (engine displacement), the calculated **p-value** is **0.5305**. Since this value is considerably greater than our threshold of 0.05, we must fail to reject the **null**

hypothesis. We conclude that 'disp' does not demonstrate a **statistically significant** relationship with the likelihood of having a manual transmission within this specific model.

For the **predictor variable** "drat" (rear axle ratio), the **p-value** is **0.0315**. Because 0.0315 is smaller than 0.05, we successfully reject the **null hypothesis**. This confirms that 'drat' is a **statistically significant** predictor of transmission type. The visual indicator of a single asterisk (*) next to the p-value confirms its significance at this level, as defined by the "Signif. codes" legend below the table.

The "Signif. codes" provided in R's summary output serve as a convenient, albeit simplified, method for quickly gauging the strength of the evidence against the **null hypothesis**. These codes map different **p-value** ranges to symbols (e.g., '***' for $p < 0.001$, '**' for $p < 0.05$), allowing analysts to rapidly identify the most impactful **predictor variables** in the model.

The Underlying Calculation: How $\text{Pr}(>|z|)$ is Derived

A deeper comprehension of $\text{Pr}(>|z|)$ requires understanding the mathematical steps R uses to compute it. The process is a standard statistical procedure involving two stages: calculating the **z value** test statistic, and subsequently mapping that **z value** onto the **standard normal distribution** to obtain the probability (p-value).

The first step involves standardizing the coefficient estimate. The **z value** for any given **predictor variable** is computed by taking its **Estimate** and dividing it by its associated **Standard Error**. This calculation transforms the coefficient into a standard score that allows comparison across different variables and models. It essentially measures the distance, in units of standard errors, from the hypothesized population mean of zero (the **null hypothesis**).

Formula for z value = Estimate / Std. Error

Referring back to our "drat" example (Estimate: 4.879396; **Standard Error**: 2.268115), we can verify the reported **z value**:

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

```
2.151
```

The second step involves calculating the **two-tailed p-value**, denoted by $\text{Pr}(>|z|)$. This probability measures the combined area under the **standard normal distribution** curve that is more extreme than the calculated **z value** in either the positive or negative direction. By calculating the probability of observing a result this extreme, we quantify the likelihood of our data occurring if the **null hypothesis** of zero effect were actually true.

Formula for p-value = $2 * (1 - \text{pnorm}(\text{abs}(z \text{ value})))$

Using the [R](#) function `pnorm()`, which provides the cumulative distribution function, we confirm the [p-value](#) for "drat":

```
#calculate p-value
```

```
2*(1-pnorm(2.151))
```

```
0.0314762
```

This result confirms that the reported [Pr\(>|z|\)](#) value in the [logistic regression](#) summary is derived directly from the standard two-tailed z-test statistic based on the estimated coefficient and its associated uncertainty.

Important Considerations and Best Practices

While [p-values](#) are indispensable for determining [statistical significance](#), they should never be interpreted in isolation. A finding that is statistically significant ($p < \alpha$) merely indicates that the observed effect is unlikely to be due to random chance, assuming the [null hypothesis](#) is true. It does not inherently guarantee [practical importance](#) or a large effect size. Conversely, a nonsignificant [p-value](#) ($p \geq \alpha$) does not prove the absence of an effect; it only suggests that the current data set lacks sufficient power or evidence to detect one at the chosen [significance level](#).

Expert analysis mandates a holistic review of the model output beyond just [Pr\(>|z|\)](#). For [logistic regression](#), analysts must examine the magnitude and direction of the **Estimate** (coefficient) alongside its [Standard Error](#). Furthermore, evaluating overall model fit using metrics such as [AIC](#) (Akaike Information Criterion), [Residual Deviance](#), and various forms of [Pseudo R-squared](#) provides a comprehensive understanding of the model's explanatory power. Calculating and interpreting [confidence intervals](#) for the coefficients is also crucial, as they provide a range of plausible values for the true population effect, offering a much richer context than a point estimate and a p-value alone.

Analysts must also be acutely aware of how sample size influences [p-values](#). In extremely large data sets, even trivial effects can yield statistically significant results ($p < 0.05$). Conversely, small samples may mask genuinely strong relationships. Finally, ethical practice requires avoiding manipulation techniques, such as [p-hacking](#), which undermines the scientific rigor of the analysis. Robust statistical conclusions integrate quantitative evidence from [R](#) with theoretical knowledge and domain expertise.

Conclusion and Further Resources

Mastering the interpretation of $\Pr(>|z|)$ is indispensable for anyone conducting or reviewing [logistic regression](#) in [R](#). This specific p-value serves as the gatekeeper for statistical inference, allowing researchers to confidently identify which [predictor variables](#) have a statistically verifiable impact on the probability of the outcome. By understanding both the decision rule (comparison against α) and the underlying mathematical derivation involving the [z value](#), analysts can move beyond simple observation to profound statistical reporting.

We strongly recommend adopting a comprehensive approach to model evaluation. Relying solely on a low [p-value](#) risks misinterpreting results; instead, integrate evidence from coefficient estimates, [confidence intervals](#), and overall model fit statistics to ensure your statistical conclusions are not only correct but also practically relevant and robust for decision-making.

For those seeking to further refine their skills, the following resources and tutorials provide detailed guidance on fitting and interpreting advanced regression models within the [R](#) environment and exploring related statistical concepts: