Interpret P-Values in Linear Regression (With Example)

Authored by

Mohammed looti

October 30, 2025

In the expansive field of statistics, **linear regression** models stand out as foundational instruments for quantifying the relationship between a **response variable** and one or more **predictor variables**. these models provide the analytical framework necessary to understand how changes in input variables are associated with changes in the outcome, serving as the cornerstone of quantitative analysis across fields like economics, biology, and data science.

When conducting a rigorous **regression analysis** using specialized statistical software, the resulting output is summarized in a comprehensive **regression table**. This table offers crucial insights into the estimated strength, direction, and reliability of the relationships identified within the dataset.

Two of the most vital components found in this output are the **regression coefficients** and their accompanying **P-values**. While the coefficients quantify the estimated magnitude of the effect (e.g., how much the response variable changes for a one-unit increase in the predictor), the P-values address the fundamental issue of **statistical significance**. They help us determine if the observed relationship is likely genuine or simply the result of random sampling variability.

A low **P-value** suggests that the relationship we observed between a predictor and the response is highly unlikely to have occurred if there were truly no relationship in the underlying population. This guide is designed to provide a clear, practical roadmap for interpreting these critical values within the context of a **linear regression** model, utilizing a detailed example to solidify your understanding.

The Role of P-Values in Statistical Hypothesis Testing

To properly interpret the numerical value of a P-value, it is essential to frame it within the mechanics of **hypothesis testing**. In regression analysis, for every predictor variable included in the model, we establish a formal **null hypothesis** (H?). This hypothesis is the default assumption of no effect; specifically, it states that the true **regression coefficient** for that predictor is zero, meaning the variable has no linear relationship with the outcome in the population.

Conversely, the **alternative hypothesis** (H?) posits that a relationship does exist--that the true coefficient is non-zero. The P-value is calculated based on the sample data and represents the probability of observing a test statistic (like the t-statistic in our regression output) as extreme as, or more extreme than, the one calculated, assuming that the **null hypothesis** is completely true.

The decision rule hinges on comparing the P-value to a predetermined **significance level**, conventionally denoted as **alpha (α)**, most commonly set at 0.05. If the P-value is less than α (e.g., $P < 0.05$), the observed data are considered too improbable under the null hypothesis, leading us to reject H? and conclude that the relationship is **statistically significant**. If the P-value is greater than α, we fail to reject the null hypothesis, meaning we lack sufficient evidence to conclude that a non-zero relationship exists.

Setting Up Our Practical Regression Example

To ground the abstract concepts of P-values in reality, let us establish a concrete scenario. We aim to construct a **linear regression** model designed to identify factors that significantly influence student performance on a standardized exam. Our primary goal is to determine if specific study habits are reliable predictors of higher scores.

We will utilize data collected on a sample of students, focusing on the following variables for our analysis:

The Predictor Variables (Inputs):

Hours studied: This **continuous predictor variable** quantifies the total time spent studying (0 to 20 hours).
Tutor: This is a **categorical predictor variable**, converted into a **dummy variable** (1 = utilized a tutor, 0 = did not utilize a tutor).

The Response Variable (Outcome):

Exam score: The final numerical score achieved by the student (0 to 100). This is the outcome we are attempting to model and predict based on the input factors.

The core inquiry of this analysis is whether the number of hours dedicated to studying and the decision to employ a tutor exert a measurable influence on a student's final score that is demonstrably greater than what could be explained by mere randomness. We must therefore assess the **P-values** associated with both **predictor variables**.

Decoding the Regression Output Table

After running the multiple **linear regression** model, we obtain the following summary table, which contains all the necessary statistics for inference.

Term	Coefficient	Standard Error	t Stat	P-value
Intercept	48.56	14.32	3.39	0.002
Hours studied	2.03	0.67	3.03	0.009
Tutor	8.34	5.68	1.47	0.138

Before focusing solely on the **P-values**, it is necessary to understand the context provided by the other columns. The **Coefficient** column details the estimated change in the response variable resulting from a one-unit change in the predictor. The **Standard Error** quantifies the precision of this coefficient estimate; a smaller standard error suggests a more reliable estimate.

The **t Stat**, or t-statistic, is derived by dividing the coefficient by its standard error. This ratio indicates how many standard errors the coefficient is away from zero. It is this t-statistic that the P-value uses to quantify the probability of observing such an extreme result if the **null hypothesis** (coefficient = 0) were true. Collectively, these statistics allow us to fully assess both the practical impact and the statistical reliability of each predictor.

Interpreting P-Values for the Intercept and Continuous Variable

First, let us examine the **Intercept**. The **intercept** (48.56) represents the predicted exam score when both predictor variables--Hours studied and Tutor status--are zero. The associated **P-value** is **0.002**. Since 0.002 is significantly less than the common **alpha (α)** level of 0.05, we conclude that the intercept is **statistically significantly different from zero**. While statistically significant, the intercept's P-value is often less crucial than the predictors' P-values, as its primary function is merely to anchor the regression line.

Next, consider **Hours studied**, our **continuous predictor variable**. The **coefficient** is **2.03**, indicating a positive relationship: every additional hour studied is associated with an expected increase of 2.03 points in the exam score, holding the tutor status constant. The corresponding **P-value** is **0.009**.

Because 0.009 is much smaller than our threshold of $\alpha = 0.05$, we have strong evidence to reject the **null hypothesis** that the true coefficient for "Hours studied" is zero. We confidently conclude that the positive relationship between study hours and exam performance is **statistically significant**. This small P-value confirms that the observed effect is highly unlikely to be the result of chance alone.

Interpreting P-Values for the Categorical Variable: Tutor

The variable **Tutor** serves as a binary, **categorical predictor variable** (1 for using a tutor, 0 otherwise). Its **coefficient** is **8.34**, suggesting that, for the same number of hours studied, a student who received tutoring is expected to score 8.34 points higher on the exam compared to a student who did not. This represents the estimated practical benefit of having a tutor in our sample.

However, we must look at the statistical reliability of this estimate. The associated **P-value** for "Tutor" is **0.138**. When compared against the standard **alpha (α)** level of 0.05, we find that 0.138 is substantially greater than 0.05. This result dictates our inferential decision.

Since the P-value exceeds the significance threshold, we must fail to reject the **null hypothesis** for the "Tutor" variable. We conclude that the observed difference of 8.34 points is not statistically significant. In practical terms, this means that the positive benefit seen in our sample could plausibly be due to random chance or sampling error, and we do not have sufficient evidence to generalize this effect to the larger population.

Critical Considerations Beyond Statistical Significance

While **P-values** are indispensable for hypothesis testing, relying solely on them can lead to flawed conclusions. It is crucial to distinguish between **statistical significance** and practical significance. An effect can be statistically significant ($P < 0.05$) if the sample size is very large, even if the actual coefficient magnitude is negligible in the real world. Conversely, a large, practically important effect might be non-significant ($P > 0.05$) if the sample size is too small or data variability is too high, as we saw with the "Tutor" variable.

The selection of **alpha (α)** (the 0.05 threshold) is inherently arbitrary. Analysts must consider the context of the study and the potential risks of Type I errors (rejecting a true null hypothesis) versus Type II errors (failing to reject a false null hypothesis). Best practice dictates reporting the exact P-value (e.g., 0.138) rather than simplifying the result to "significant" or "not significant," allowing readers maximum transparency.

Furthermore, the reliability of all P-values generated by a **linear regression** model depends on the model meeting several critical assumptions, including linearity, independence of errors, and homoscedasticity. Failing to verify these assumptions means the statistical inferences drawn from the P-values may be invalid. Thorough diagnostic checking is a non-negotiable step in responsible regression analysis.

Summary of Key Findings

Interpreting **P-values** is the fundamental mechanism for determining the statistical reliability of **predictor variables** within a regression context. By comparing the P-value against the chosen **alpha (α)** level, we transition from observing a sample effect to making an informed inference about the population.

Our student performance example clearly illustrated this distinction: "Hours studied" exhibited a statistically significant relationship with the exam score ($P = 0.009$), suggesting a genuine effect. Conversely, the "Tutor" variable showed a positive effect in the sample (Coefficient = 8.34), but its high P-value (0.138) indicated this effect was not statistically significant at the 0.05 level. A successful statistical analysis always requires synthesizing these statistical metrics with subject-matter expertise to derive meaningful conclusions.

Further Reading and Resources

To enhance your mastery of linear regression and advanced statistical concepts, we recommend exploring the following authoritative resources:

Interpreting Regression Output: P-Values, R-squared
Assumptions of Linear Regression: Explained
IBM SPSS Documentation: Interpreting Linear Regression Results