

# How to Easily Read and Interpret a Regression Table

Authored by  
**stats writer**

March 2, 2026

## RECOMMENDED CITATION

stats writer (2026). *How to Easily Read and Interpret a Regression Table*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=133538>

## Understanding the Fundamentals of Regression Analysis and Its Output

In the expansive field of **statistics**, regression analysis serves as a foundational tool for uncovering the intricate relationships between various factors. At its core, this analytical method allows researchers and data scientists to determine how a **dependent variable**--often referred to as the response or outcome--is influenced by one or more **independent variables**, commonly known as predictors. The primary goal of this technique is to create a mathematical model that can predict future outcomes or explain the underlying dynamics of a dataset. When executing these complex calculations using sophisticated software such as R, SPSS, or SAS, the results are summarized in a structured format known as a regression table. This table is an essential artifact that distills a vast amount of mathematical computation into a digestible summary, providing the evidence needed to support or refute a **research hypothesis**.

To effectively read and interpret a regression table, one must approach it with a systematic mindset, recognizing that each section of the table serves a specific purpose in evaluating the model's validity. The table acts as a bridge between raw data and actionable insights, offering a clear view of the **coefficients**, **standard errors**, and **significance levels**. Without a thorough understanding of these components, the results of a regression analysis could easily be misinterpreted, leading to erroneous conclusions. Therefore, mastering the interpretation of these tables is not merely a technical skill but a critical competency for anyone engaged in data-driven decision-making. By carefully examining the layout and the values presented, analysts can discern the strength of the evidence and the reliability of their predictions.

Furthermore, it is important to recognize that a regression table does not exist in a vacuum. It is part of a broader analytical framework that includes exploratory data analysis, diagnostic checking, and validation. While the table provides a concise summary of the model's performance, it should always be interpreted alongside other diagnostic tools, such as residual plots and influence metrics. This holistic approach ensures that the relationships identified in the model are not just artifacts of the data but represent meaningful, reproducible patterns. In the following sections, we will delve deeper into a specific example of multiple linear regression, walking through each component of the output to provide a comprehensive guide on how to derive value from these statistical summaries.

### A Practical Case Study: Predicting Academic Success

To ground these theoretical concepts in reality, let us consider a specific scenario involving educational data. Suppose we are investigating the factors that contribute to a student's performance on a final examination. In this study, we have collected data from 12 individual students, tracking three key variables: the total number of hours they spent studying, the number of preparatory exams they completed, and their final exam scores. This dataset serves as the basis

for our **statistical modeling**, where we aim to understand if study habits and practice testing can reliably predict academic achievement. The collection of such data is the first step in performing a rigorous quantitative analysis, providing the empirical evidence necessary to test our assumptions about learning and performance.

	Study Hours	Prep Exams	Final Exam Score
Student 1	3	2	76
Student 2	7	6	88
Student 3	16	5	96
Student 4	14	2	90
Student 5	12	7	98
Student 6	7	4	80
Student 7	4	4	86
Student 8	19	2	89
Student 9	4	8	68
Student 10	8	4	75
Student 11	8	1	72
Student 12	3	3	76

In this case study, the **final exam score** is designated as the response variable, as it is the outcome we are attempting to explain. The predictor variables are **hours studied** and **prep exams taken**. By employing a linear regression model, we are essentially trying to fit a line (or in the case of multiple predictors, a plane) through the data points that minimizes the distance between the observed values and the values predicted by the model. This process, often referred to as **Ordinary Least Squares** (OLS), allows us to quantify the specific impact of each study hour and each practice exam on the final score, while also accounting for the simultaneous influence of both factors.

Once the regression analysis is performed using statistical software, the resulting output is presented in a multi-part table. This output is the definitive source for determining whether our predictors are statistically significant and how much of the variation in exam scores can be attributed to them. For any researcher, this is the "moment of truth" where the data reveals whether the hypothesized relationships actually exist. Below is the regression table generated from our student dataset, which we will use as a reference throughout this tutorial to explain the various metrics and their implications for the study.

<i>Regression Statistics</i>	
Multiple R	0.728550902
R Square	0.530786416
Adjusted R Square	0.426516731
Standard Error	7.326766656
Observations	12

ANOVA					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	2	546.53308	273.2665	5.090515	0.033202256
Residual	9	483.1335867	53.68151		
Total	11	1029.666667			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	66.99010508	6.211445265	10.78495	1.9E-06	52.93883968	81.04137
Study Hours	1.299900324	0.417012868	3.117171	0.012375	0.356551677	2.243249
Prep Exams	1.117275106	1.025145307	1.08987	0.30409	-1.201764693	3.4363149

## Evaluating Model Fit: Correlation and Determination

The initial section of a regression table is dedicated to assessing the "Goodness of Fit." This refers to how well the statistical model describes the observed data. If a model fits well, the differences between the observed values and the model's predicted values will be small and unbiased. The first metric we encounter is **Multiple R**, which represents the Pearson correlation coefficient between the observed values of the dependent variable and those predicted by the model. This value ranges from 0 to 1, where a value closer to 1 indicates a stronger linear relationship. In our example, the **Multiple R is 0.72855**, suggesting a robust positive correlation between the predictors (hours and prep exams) and the final score.

---

<i>Regression Statistics</i>	
Multiple R	0.728550902
R Square	0.530786416
Adjusted R Square	0.426516731
Standard Error	7.326766656
Observations	12

---

The next, and perhaps most frequently cited metric, is the **R-Squared** value, also known as the coefficient of determination. This value quantifies the proportion of the **variance** in the dependent variable that can be explained by the independent variables in the model. An R-Squared of 0.5307, as seen in our output, indicates that approximately 53.07% of the fluctuations in students' final exam scores can be explained by the hours they studied and the number of prep exams they took. While this suggests the model has significant explanatory power, it also implies that nearly 47% of the variance is due to other factors not included in our model, such as prior knowledge, intelligence, or even luck.

To provide a more conservative estimate, especially in models with multiple predictors, we look at the **Adjusted R-Squared**. Unlike the standard R-Squared, which can artificially increase as more variables are added to the model (even if they are irrelevant), the Adjusted R-Squared accounts for the number of predictors relative to the sample size. It penalizes the addition of variables that do not improve the model's fit by a sufficient amount. In our scenario, the **Adjusted R-Squared is 0.4265**. This lower value is a more realistic measure of the model's generalizability and is particularly useful when comparing different models to see which one provides the most efficient explanation of the data without **overfitting**.

## The Precision of Estimates and Observation Counts

Beyond correlation metrics, the regression table provides a measure of the model's error. The **Standard Error of the Regression** represents the average distance that the observed values fall from the regression line. It is essentially the standard deviation of the **residuals** (the errors in prediction). A smaller standard error indicates that the data points are clustered closely around the regression line, suggesting that the model's predictions are more precise. In our study, the **standard error is 7.3267**, meaning that on average, our predicted exam scores deviate from the actual scores by about 7.3 points. This value is expressed in the same units as the dependent variable, making it a highly practical measure of the model's accuracy.

Understanding the standard error is vital for constructing prediction intervals. If we were to use our

model to predict a new student's score, the standard error would tell us the range within which we could expect that score to fall. A high standard error relative to the mean of the dependent variable might suggest that the model is not very useful for individual predictions, even if the R-Squared is high. Therefore, analysts must always weigh the explanatory power of the model (R-Squared) against its predictive precision (Standard Error) to get a full picture of the model's utility.

Finally, this section includes the count of **Observations**, which is simply the size of the sample used to build the model. In our case, the **total observations is 12**. Sample size is a critical factor in statistical significance; with only 12 students, our model might be susceptible to the influence of outliers, and the results may not be as stable as those derived from a larger group. A larger sample size generally leads to more reliable estimates and narrower confidence intervals, increasing our confidence that the observed relationships are representative of the broader population of students.

### Analyzing the ANOVA Table: Variance Decomposition

The second major section of the regression output is the **ANOVA** (Analysis of Variance) table. This section breaks down the total variability in the dataset into two distinct components: the variability explained by the regression model and the remaining unexplained variability, often called the residual or error. By comparing these two types of variance, we can determine if the model as a whole is statistically significant. The table lists the **Degrees of Freedom** (df), the **Sum of Squares** (SS), and the **Mean Squares** (MS) for both the regression and the residuals. These mathematical components are the building blocks used to calculate the F-statistic, which is the primary test for the model's overall performance.

ANOVA					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	2	546.53308	273.2665	5.090515	0.033202256
Residual	9	483.1335867	53.68151		
Total	11	1029.666667			

**Degrees of freedom** are an essential concept in inferential statistics. For the regression component, the degrees of freedom are calculated as the number of predictor variables. Since we have two predictors (hours and prep exams), the regression df is 2. The total degrees of freedom are calculated as the number of observations minus one ( $12 - 1 = 11$ ). The residual degrees of freedom are the difference between the total and the regression df ( $11 - 2 = 9$ ). These values are used to normalize the Sum of Squares, resulting in the Mean Squares. Specifically, the **Regression MS** is 273.2665, and the **Residual MS** is 53.68151. These MS values represent the

average variance explained by the model versus the average variance left unexplained.

The **Sum of Squares** (SS) provides a look at the total amount of variation. The Regression SS (546.53) represents the total variation explained by the relationship between the study habits and the scores. The Residual SS (483.13) is the variation that remains, likely due to factors we did not measure. A successful model will have a much larger Regression SS than Residual SS. This decomposition of variance is what allows statisticians to partition the "signal" from the "noise," providing a clear mathematical basis for deciding whether the model's predictors are actually doing a good job of capturing the patterns in the data.

## The F-Statistic and Global Model Significance

The **F-statistic** is the ratio of the Regression Mean Square to the Residual Mean Square. It serves as a global test to determine if at least one of the predictor variables has a non-zero coefficient. In simpler terms, the F-test asks: "Is this model better than a model with no predictors at all?" An F-statistic significantly greater than 1 suggests that the model provides a better fit than a simple average of the dependent variable. In our student exam example, the **F-statistic is calculated as 5.09**. This value on its own is difficult to interpret without context, which is why we look at the corresponding p-value, labeled as **Significance F**.

The **Significance F**, or the p-value for the F-test, tells us the probability of obtaining an F-statistic as large as 5.09 by pure chance, assuming that the null hypothesis (that all coefficients are zero) is true. To determine significance, we compare this p-value to a pre-determined **alpha level**, usually set at 0.05. If the p-value is lower than the alpha, we reject the null hypothesis. In our output, the **p-value is 0.033**. Because 0.033 is less than 0.05, we conclude that our regression model is statistically significant. This gives us the "green light" to proceed with interpreting the individual variables, knowing that the overall model is not just a result of random noise.

It is important to note that a significant F-test does not mean that *all* of your predictors are significant; it only indicates that *at least one* of them is. In some cases, a model might have a very high F-statistic because one predictor is incredibly powerful, even if the others are useless. Conversely, if the Significance F is greater than 0.05, the model is generally considered unreliable, and any individual p-values for the predictors should be viewed with extreme skepticism. The F-test acts as a gatekeeper, ensuring that we only invest time in interpreting models that have demonstrated a basic level of statistical validity.

## Interpreting Regression Coefficients and the Equation

The final section of the regression table is often the most scrutinized, as it provides the specific estimates for each variable. The **Coefficients** column gives us the values needed to construct the **regression equation**. This equation allows us to calculate the predicted response for any given

set of input values. The general form of the equation for multiple regression is  **$y = \text{Intercept} + (\text{Coefficient 1} * \text{Variable 1}) + (\text{Coefficient 2} * \text{Variable 2})$** . For our study, the equation is: **Final Exam Score = 66.99 + 1.299(Study Hours) + 1.117(Prep Exams)**. This formula is the culmination of the entire analysis, providing a predictive tool for academic performance.

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	66.99010508	6.211445265	10.78495	1.9E-06	52.93883968	81.04137
Study Hours	1.299900324	0.417012868	3.117171	0.012375	0.356551677	2.243249
Prep Exams	1.117275106	1.025145307	1.08987	0.30409	-1.201764693	3.4363149

Interpreting these coefficients requires careful attention to detail. Each coefficient represents the average change in the dependent variable for every one-unit increase in the predictor, holding all other variables constant. For instance, the coefficient for **Study Hours is 1.299**. This means that for every additional hour a student studies, their final exam score is expected to increase by approximately 1.3 points, provided the number of prep exams they take remains the same. The **Intercept** (66.99) represents the predicted score when both study hours and prep exams are zero. While the intercept is mathematically necessary, it may not always be practically meaningful (e.g., if it's impossible to score a 66 without studying, or if the data doesn't include students who studied zero hours).

When analyzing coefficients, it is also vital to consider their sign. A positive coefficient indicates a positive relationship--as the predictor goes up, the outcome goes up. A negative coefficient would indicate an inverse relationship. In our example, both study hours and prep exams have positive coefficients, which aligns with the intuitive expectation that more preparation leads to better results. However, the **magnitude** of the coefficient is just as important as the direction. A small coefficient might suggest that while a relationship exists, its practical impact on the outcome is negligible. This distinction between **statistical significance** and practical significance is a hallmark of sophisticated data interpretation.

## Individual Significance: T-Stats and P-Values

While the coefficients tell us the size and direction of the effect, the **Standard Error**, **t-stat**, and **P-value** tell us how much confidence we should have in those estimates. The standard error measures the variability of the coefficient estimate across different possible samples. If we were to repeat this study with 12 different students, the coefficient for study hours would likely change; the standard error tells us how much we should expect it to fluctuate. The **t-stat** is calculated by dividing the coefficient by its standard error ( $1.299 / 0.417 = 3.117$ ). This statistic follows a Student's t-distribution and is used to determine the individual significance of each predictor.

The most critical value in this section is the **P-value** associated with each t-stat. This value tests the **null hypothesis** that a specific coefficient is equal to zero (meaning the variable has no effect). Just like the Significance F, we typically compare these p-values to an alpha of 0.05. In our example, the p-value for **Study Hours is 0.012**, which is significant. However, the p-value for **Prep Exams is 0.304**, which is well above the 0.05 threshold. This leads to a crucial insight: while study hours are a significant predictor of exam scores, the number of prep exams taken is not statistically significant in this specific model. This highlights the power of regression in isolating the true drivers of an outcome.

The lack of significance for "Prep Exams" could be due to several reasons. It might be that practice exams truly don't help, or perhaps their effect is already captured by the "Hours Studied" variable (a phenomenon known as multicollinearity). It could also be a result of our small sample size; with only 12 observations, the model may lack the **statistical power** to detect a smaller effect from the prep exams. This is why researchers must be cautious when interpreting "non-significant" results-- it doesn't necessarily mean the effect is zero, but rather that our data does not provide enough evidence to conclude it is different from zero with high confidence.

## Confidence Intervals and Final Considerations

To round out the interpretation, the regression table provides **Confidence Intervals** for each coefficient, usually at the 95% level. These intervals give us a range of plausible values for the true population coefficient. For **Study Hours**, the interval is (0.356, 2.24). Because this range does not include zero, we are 95% confident that the true effect of an hour of study is a positive increase in the score between 0.35 and 2.24 points. This adds a layer of robustness to our finding that study hours are a significant predictor. It provides a "margin of error" for our estimate, which is essential for honest reporting of statistical results.

In contrast, the 95% confidence interval for **Prep Exams** is (-1.201, 3.436). Because this interval includes zero, it reinforces our previous finding that the variable is not statistically significant. When an interval contains zero, it means that the data is consistent with a scenario where the variable has a positive effect, a negative effect, or no effect at all. This ambiguity is why we fail to reject the null hypothesis for this specific predictor. Understanding confidence intervals is key to moving beyond simple "yes/no" significance testing and toward a more nuanced understanding of the uncertainty inherent in all statistical estimations.

In conclusion, reading a regression table is a multi-step process that involves checking the overall model fit, verifying the global significance of the model, and then drilling down into the individual predictors. By combining the insights from **R-Squared**, the **F-test**, and individual **p-values**, an analyst can build a compelling story about their data. Remember that regression is a tool for simplification; it takes complex, real-world relationships and distills them into a set of numbers.

However, these numbers are only as good as the data and the assumptions behind them. Always consider the context of your research question and use these statistical findings as one piece of a larger evidentiary puzzle in your decision-making process.

ARABPSYCHOLOGY.COM