

How to Easily Interpret Regression Output in Excel

Authored by
stats writer

December 4, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Easily Interpret Regression Output in Excel*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=105171>

Mastering the interpretation of regression analysis output, particularly when generated using tools like Microsoft Excel, is a fundamental skill for data professionals. The core task involves meticulously examining key statistical indicators such as the R-squared value and the coefficients. These metrics provide critical insights into the underlying relationships within the data, quantifying both the strength and the nature of the association between the variables under study.

The coefficients serve as the bedrock of the model, quantifying the magnitude and directional impact--positive or negative--that the independent variables exert on the dependent variable. Concurrently, the R-squared value offers a measure of model fit, specifying the precise proportion of the total variation observed in the dependent variable that the chosen independent variables collectively account for. Furthermore, the inclusion of the p-value is essential for assessing the reliability of individual variable estimates, allowing us to determine if an observed relationship is truly statistically significant or likely due to random chance.

In the expansive realm of quantitative methods, multiple linear regression stands out as one of the most versatile and frequently utilized statistical techniques. It allows researchers to model the relationship between a single dependent variable and multiple independent (or predictor) variables simultaneously. This detailed guide is designed to systematically walk you through the process of interpreting every crucial value presented in the Microsoft Excel regression output, providing a robust framework for drawing accurate conclusions from your data.

Case Study: Defining Variables for Multiple Regression

To illustrate the interpretation process clearly, we will analyze a practical example concerning student performance. Our primary objective is to investigate whether a student's commitment to studying and their preparation strategy--specifically, the number of preparatory exams taken--have a quantifiable effect on their final score in a crucial college entrance examination. This scenario requires a model capable of handling more than one predictor, making multiple linear regression the appropriate methodology.

In this specific statistical model, we designate the student's observed final **exam score** as the **response variable** (or dependent variable), which is the outcome we are attempting to predict or explain. Conversely, the measurable factors believed to influence this score--namely, the **hours studied** and the **number of prep exams taken**--are classified as the **predictor variables** (or independent variables).

The output presented below, generated by running the Regression tool in Excel's Analysis ToolPak on a dataset, summarizes the findings from this specific model. A thorough understanding of each section of this output is essential for deriving meaningful conclusions about the factors influencing exam scores.

D	E	F	G	H	I	J	K
SUMMARY OUTPUT							
<i>Regression Statistics</i>							
Multiple R	0.857						
R Square	0.734						
Adjusted R Square	0.703						
Standard Error	5.366						
Observations	20						
ANOVA							
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>		
Regression	2	1350.76	675.38	23.46	0.00		
Residual	17	489.44	28.79				
Total	19	1840.20					
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	
Intercept	67.67	2.82	24.03	0.00	61.73	73.61	
hours	5.56	0.90	6.18	0.00	3.66	7.45	
prep_exams	-0.60	0.91	-0.66	0.52	-2.53	1.33	

Interpreting the Regression Statistics Block

The top section of the Excel output, labeled "Regression Statistics," provides key metrics that summarize the overall performance and fit of the model. These statistics are crucial for determining how well the set of predictor variables explains the variation in the response variable.

We begin with the **Multiple R** value, which is calculated as **0.857** in this case. This statistic represents the multiple correlation coefficient--a measure of the linear relationship between the response variable (Exam Score) and the combined set of all predictor variables (Hours Studied and Prep Exams Taken). Since the value is close to 1, it suggests a strong positive correlation, indicating that the predictors are effective at estimating the outcome.

Next, the **R Square** value, standing at **0.734**, is perhaps the most critical indicator of model fit. Known formally as the coefficient of determination, this metric tells us the proportion of the variation in the dependent variable that is predictable from the independent variables. Specifically, in this example, 73.4% of the variability observed in the student exam scores can be successfully explained by the combined influence of the number of hours studied and the number of prep exams taken. This is a robust finding, suggesting the model has substantial explanatory power.

Following this is the **Adjusted R Square**, which is **0.703**. While similar to R Square, the Adjusted R

Square incorporates a correction factor based on the number of predictors included in the model and the sample size. It is a more conservative estimate of model fit, particularly useful when comparing models with varying numbers of independent variables. It serves to penalize models that include unnecessary or redundant predictors, thereby providing a truer assessment of the model's predictive capability in the population. Note that it will always be equal to or less than the standard R Square value.

The **Standard Error**, recorded as **5.366**, quantifies the average magnitude of the error in the prediction--that is, the average distance that the observed data points fall from the calculated regression line. This value is expressed in the units of the dependent variable (exam scores). A standard error of 5.366 means that, on average, the model's predictions for the exam score will deviate from the actual observed scores by approximately 5.366 points. A lower standard error generally signifies a more precise model. Finally, the **Observations** count of **20** simply confirms the total sample size used to estimate the model parameters.

Evaluating Overall Model Fit: The ANOVA Summary

The Analysis of Variance (ANOVA) table within the regression output is designed to test the overall usefulness and statistical significance of the entire linear model. It compares the variability explained by the model (Regression Sum of Squares) against the unexplained variability (Residual Sum of Squares).

The most important value derived from the ANOVA table is the **F statistic**, which is calculated here as **23.46**. The F statistic tests the null hypothesis that all of the true regression coefficients (excluding the intercept) are simultaneously equal to zero. If the calculated F value is sufficiently large, we reject this null hypothesis, concluding that at least one predictor variable significantly contributes to the model.

Adjacent to the F statistic is the **Significance F** value, which is the corresponding p-value for the overall F-test. In this analysis, the value is listed as **0.0000** (which implies a value extremely close to zero). This small p-value is crucial: because it is substantially less than the conventional alpha level of 0.05, we confidently reject the null hypothesis. This rejection indicates that the regression model as a collective whole is statistically significant. In practical terms, the predictor variables--hours studied and prep exams taken--combined have a reliable association with the exam score.

Interpreting Individual Predictors: The Coefficient Table

While the ANOVA table confirms that the model is significant overall, the Coefficient table dives deeper, assessing the unique contribution of each independent variable. This section provides the actual estimated coefficients (the core of the regression equation) and accompanying statistics necessary for making specific inferences about each predictor.

The **Coefficients** column lists the estimated regression weights. The fundamental interpretation of a regression coefficient is the expected average change in the dependent variable for a one-unit increase in that specific independent variable, assuming all other predictor variables in the model are held constant (the "ceteris paribus" condition). These values are the building blocks of the Estimated Regression Equation.

Let's analyze the specific results from our example:

Intercept (67.67): This is the baseline prediction. It represents the estimated exam score when all predictor variables are zero. In this context, it is the expected exam score for a hypothetical student who studies zero hours and takes zero prep exams. The predicted score is **67.67**.

Hours Studied (5.56): This positive coefficient indicates that for every additional hour a student spends studying, their expected exam score increases by **5.56** points, assuming the number of prep exams taken remains unchanged.

Prep Exams Taken (-0.60): This negative coefficient suggests that for every additional prep exam a student takes, their expected score decreases by **0.60** points, holding study hours constant. We must, however, check the significance of this finding before drawing strong conclusions.

The **P-values** associated with each predictor are vital for determining whether that variable's relationship with the outcome is statistically significant on its own. We examine the results against a standard significance level, typically $\alpha = 0.05$.

A closer look reveals distinct differences between our predictors:

Hours Studied: The p-value here is **0.00** (or very close to zero). Since $0.00 < 0.05$, the relationship between hours studied and exam score is highly statistically significant. We can be confident that the positive impact (5.56 points per hour) is a reliable finding.

Prep Exams Taken: The corresponding p-value is **0.52**. Since $0.52 > 0.05$, the relationship is **not statistically significant** at the 5% level. This implies that the observed negative relationship (-0.60) could easily be due to random chance, and we do not have sufficient evidence to conclude that the number of prep exams, when controlling for hours studied, truly affects the score.

Formulating and Using the Estimated Regression Equation

The primary practical application of the regression output is the construction of the Estimated Regression Equation. This equation formalizes the relationship observed in the sample data, allowing us to generate predictions for the dependent variable based on specific values of the independent variables. We use the estimated coefficients directly from the Excel output table to define this model.

Based on our coefficients (Intercept: 67.67; Hours Studied: 5.56; Prep Exams Taken: -0.60), the

estimated linear model is written as:

$$\text{Exam score} = 67.67 + 5.56 * (\text{Hours Studied}) - 0.60 * (\text{Prep Exams Taken})$$

This formula allows for direct application, providing an expected score for any combination of study hours and preparatory exams taken within the domain of the data used to train the model. It translates complex statistical relationships into a simple, actionable predictive tool.

Applying the Model: Calculating Predicted Outcomes

To demonstrate the practical use of the Estimated Regression Equation, consider a scenario involving a specific student. Suppose this student committed to studying for three hours and took exactly one preparatory examination before the official college entrance test. We can substitute these values into our derived equation to calculate their expected score:

The calculation proceeds as follows:

$$\text{Exam score} = 67.67 + 5.56 * (3) - 0.60 * (1)$$

Which simplifies to:

$$\text{Exam score} = 67.67 + 16.68 - 0.60 = 83.75$$

Therefore, based on the established statistical model, a student exhibiting these specific preparation habits is predicted to achieve an exam score of **83.75**. This provides a clear, quantitative prediction derived directly from the underlying data relationships.

Refining the Model: Dealing with Insignificant Predictors

A crucial step in generating robust and parsimonious statistical models involves reviewing predictors that do not demonstrate sufficient statistical reliability. As observed in the Coefficient table, the variable **prep exams taken** yielded a high p-value of 0.52. Since this value significantly exceeds the standard threshold of 0.05, we conclude that this variable does not contribute uniquely or reliably to the prediction of the exam score once the effect of study hours is already accounted for.

When a predictor variable lacks p-value significance, standard statistical practice often dictates its removal from the model. Retaining insignificant variables unnecessarily complicates the interpretation and may introduce noise without genuinely improving the predictive accuracy, especially considering the constraints placed by the Adjusted R Square metric.

In this particular instance, the logical step for model refinement would be to transition from multiple

linear regression to **simple linear regression**. This revised model would include only **hours studied** as the sole explanatory variable, as it has been proven to be the most influential and statistically reliable predictor of the exam score. This iterative process ensures that the final predictive model is both effective and efficient.

ARABPSYCHOLOGY.COM