

How to Perform Multiple Linear Regression in SPSS: A Step-by-Step Guide

Authored by
stats writer

March 14, 2026

RECOMMENDED CITATION

stats writer (2026). *How to Perform Multiple Linear Regression in SPSS: A Step-by-Step Guide*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=135798>

Understanding the Foundations of Multiple Linear Regression Analysis

In the field of statistical modeling, **multiple linear regression** serves as a sophisticated analytical tool designed to explore the complex relationships between a single **dependent variable** and two or more **independent variables**. Unlike simple linear regression, which only accounts for a single predictor, this method allows researchers to isolate the specific impact of each explanatory variable while simultaneously controlling for others. By utilizing this technique, analysts can determine the extent to which several factors collectively influence an outcome, providing a more comprehensive understanding of real-world phenomena where variables rarely operate in isolation.

The primary objective of **multiple linear regression** is to find a mathematical linear equation that best predicts the value of the response variable based on the provided explanatory inputs. This process involves calculating **coefficients** that represent the strength and direction of the relationship between each predictor and the outcome. In practice, this is widely used across various disciplines, such as economics, social sciences, and medicine, to forecast trends, identify key drivers of behavior, and test theoretical hypotheses within a controlled statistical framework.

To execute this analysis effectively, researchers often rely on specialized software like **SPSS** (Statistical Package for the Social Sciences). This platform simplifies the heavy computational requirements of regression, allowing users to focus on data interpretation and model refinement. Throughout this guide, we will examine the systematic process of performing a **multiple linear regression** in **SPSS**, ensuring that every step from data entry to the interpretation of final results is handled with precision and clarity.

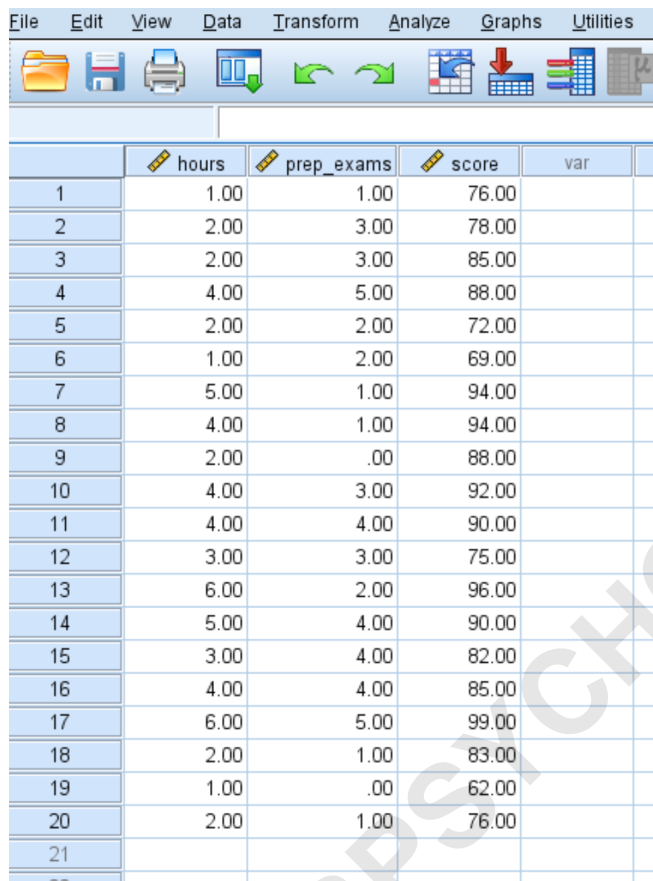
Step 1: Preparing the Dataset and Initial Data Entry

The success of any statistical analysis begins with the integrity of the data being evaluated. In this tutorial, we will utilize a practical example to illustrate the process: exploring how academic performance is influenced by study habits. Specifically, we aim to determine if the number of hours spent studying and the number of preparatory exams taken significantly affect the final exam score achieved by a group of students. In this scenario, the **exam score** acts as the response variable, while **hours studied** and **prep exams taken** serve as our primary explanatory variables.

To begin the procedure, users must launch **SPSS** and navigate to the "Data View" tab to input the raw information. Each row in the spreadsheet should represent an individual observation--in this case, a single student--while each column represents a specific variable. Proper labeling in the "Variable View" tab is essential; ensuring that variables are correctly identified as "Scale" measures will prevent errors during the calculation phase. This structured approach to data management is a prerequisite for generating reliable and valid statistical outputs.

Once the variables are defined, enter the data for all participants. For this example, we have

collected data from 20 students, tracking their study duration, the number of practice tests they completed, and their subsequent exam results. Accuracy during this phase is paramount, as outliers or data entry errors can significantly skew the **multiple linear regression** model and lead to misleading conclusions. The image below displays how the data should appear within the **SPSS** interface prior to starting the analysis:



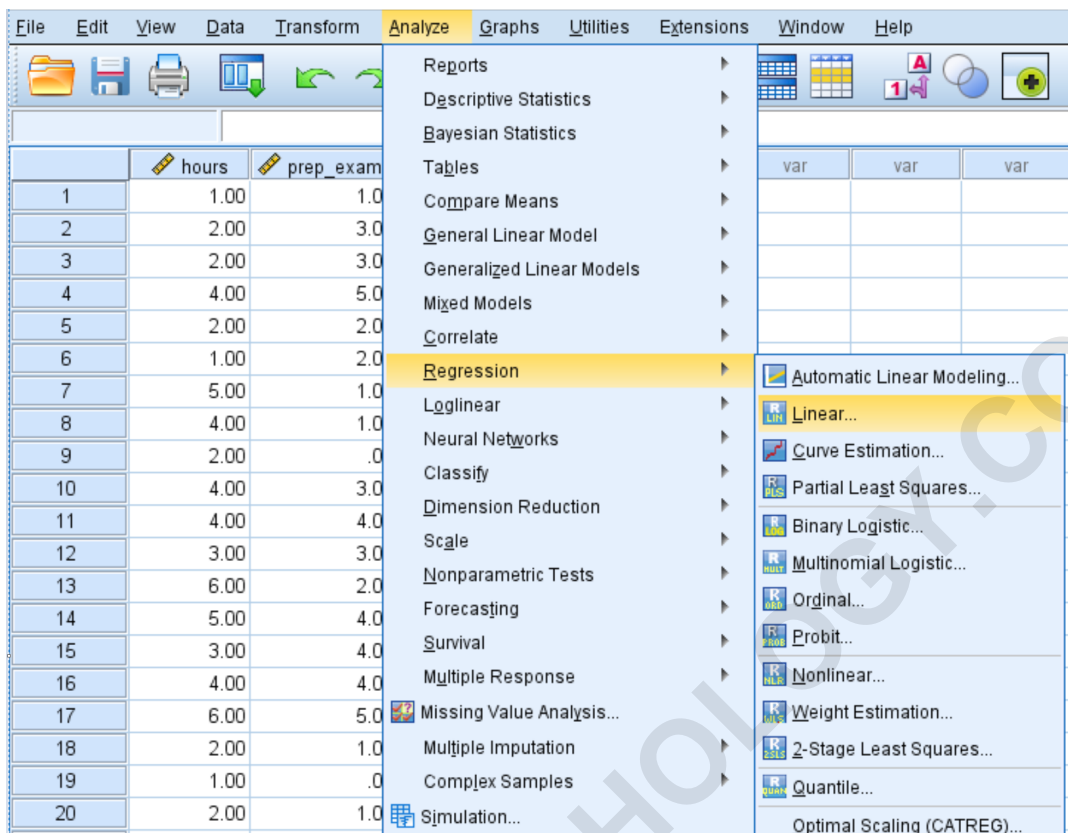
| | hours | prep_exams | score | var |
|----|-------|------------|-------|-----|
| 1 | 1.00 | 1.00 | 76.00 | |
| 2 | 2.00 | 3.00 | 78.00 | |
| 3 | 2.00 | 3.00 | 85.00 | |
| 4 | 4.00 | 5.00 | 88.00 | |
| 5 | 2.00 | 2.00 | 72.00 | |
| 6 | 1.00 | 2.00 | 69.00 | |
| 7 | 5.00 | 1.00 | 94.00 | |
| 8 | 4.00 | 1.00 | 94.00 | |
| 9 | 2.00 | .00 | 88.00 | |
| 10 | 4.00 | 3.00 | 92.00 | |
| 11 | 4.00 | 4.00 | 90.00 | |
| 12 | 3.00 | 3.00 | 75.00 | |
| 13 | 6.00 | 2.00 | 96.00 | |
| 14 | 5.00 | 4.00 | 90.00 | |
| 15 | 3.00 | 4.00 | 82.00 | |
| 16 | 4.00 | 4.00 | 85.00 | |
| 17 | 6.00 | 5.00 | 99.00 | |
| 18 | 2.00 | 1.00 | 83.00 | |
| 19 | 1.00 | .00 | 62.00 | |
| 20 | 2.00 | 1.00 | 76.00 | |
| 21 | | | | |
| 22 | | | | |

Step 2: Navigating the SPSS Interface for Regression Analysis

With the dataset correctly formatted and entered, the next phase involves selecting the appropriate analytical pathway within **SPSS**. The software provides a dedicated menu for regression, housing various models ranging from simple linear to logistic and non-linear options. For our purposes, the **Linear Regression** option is the correct choice, as we are examining a continuous dependent variable against multiple continuous or categorical independent predictors.

To initiate the analysis, click on the **Analyze** tab located in the top navigation bar. From the resulting dropdown menu, hover over **Regression** and then select **Linear**. This action will trigger a new dialog box that serves as the command center for the regression model. It is here that the user defines which variables are predictors and which is the outcome, as well as selects additional statistical tests to validate the model's assumptions. The following image illustrates the specific

menu path to follow:



After the Linear Regression dialog box appears, the user must meticulously assign the variables. Drag the **score** variable into the field labeled "Dependent." This informs the software that this is the outcome we are trying to predict. Next, drag both the **hours** and **prep_exams** variables into the "Independent(s)" box. It is important to note that **SPSS** allows for different "Methods" of variable entry, such as "Enter," "Stepwise," or "Forward"; for a standard **multiple linear regression**, the "Enter" method is typically used to include all chosen variables simultaneously.

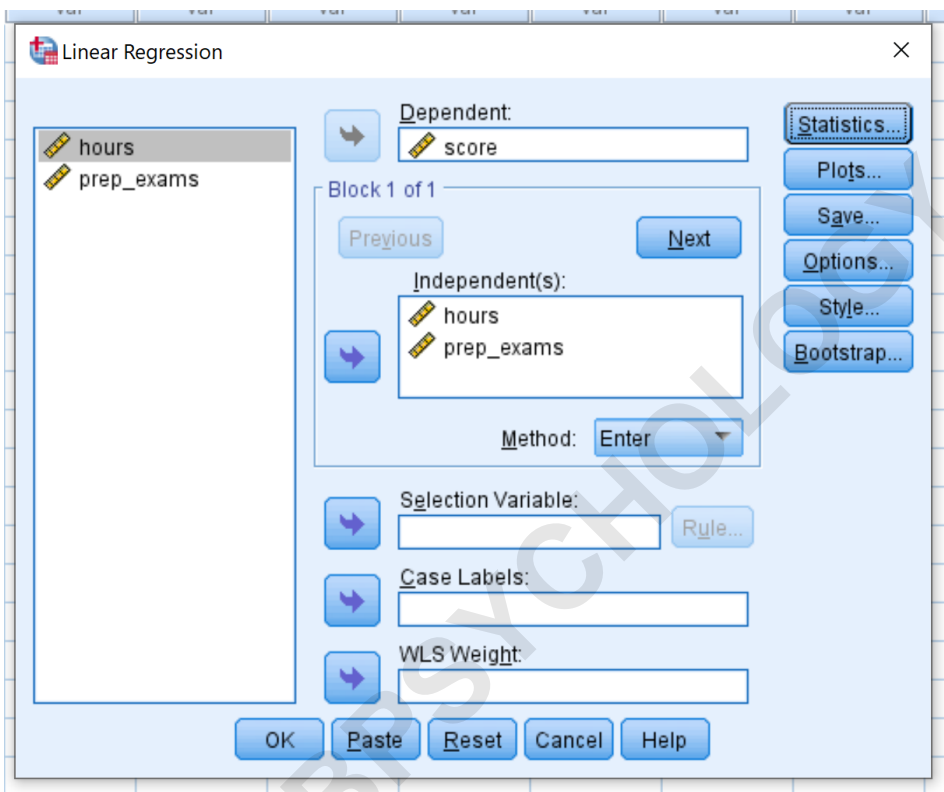
Step 3: Configuring Advanced Model Settings and Visualizations

Before finalizing the analysis by clicking "OK," it is beneficial to configure specific diagnostic settings that provide deeper insights into the model's performance. By clicking the **Statistics** button within the Linear Regression window, users can opt to include **coefficients**, model fit indicators, and descriptives. These metrics are vital for verifying the underlying assumptions of the regression, such as the absence of multicollinearity and the independence of residuals.

Furthermore, the **Plots** button allows for the generation of essential visualizations that help in assessing the model's accuracy. For instance, creating a **scatterplot** of standardized residuals against predicted values can help identify issues with heteroscedasticity. Additionally, generating

histograms of the residuals is a common practice to ensure that the error terms are normally distributed, a key requirement for valid hypothesis testing in **multiple linear regression**.

Once the variables are assigned and the desired options are selected, the setup should reflect the configuration shown in the image below. This step ensures that the output window will contain all the necessary data to perform a comprehensive evaluation of the student performance model. After verifying these settings, click **Continue** and then **OK** to execute the procedure and generate the results window.



Step 4: Interpreting the Model Summary and Goodness of Fit

The first critical output generated by **SPSS** is the **Model Summary** table. This table provides a high-level overview of how well the independent variables--hours studied and prep exams--collectively explain the **variance** in the exam scores. Understanding these metrics is the first step in determining whether the regression model is a useful tool for prediction or if it lacks explanatory power. The image below displays the Model Summary output for our example:

Model Summary

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|-------|-------------------|----------|-------------------|----------------------------|
| 1 | .857 ^a | .734 | .703 | 5.36570 |

a. Predictors: (Constant), prep_exams, hours

In this table, the most significant value to observe is the **R Square**. This statistic represents the proportion of the **variance** in the response variable that can be attributed to the linear combination of the predictor variables. In our specific case, the **R Square** value is 0.734, meaning that **73.4%** of the differences in student exam scores can be explained by the hours they studied and the number of prep exams they took. This indicates a relatively strong model with high predictive value.

Another important metric is the **Standard Error of the Estimate**. This value quantifies the average distance that the observed data points fall from the regression line. A smaller **standard error** suggests that the model's predictions are more precise. For our example, the value is **5.3657**, indicating that the actual exam scores typically deviate from the predicted scores by approximately five and a half points. This provides a clear margin of error for any predictions made using the model.

Step 5: Evaluating Statistical Significance through ANOVA

The second table produced in the output is the **ANOVA** (Analysis of Variance) table. While **ANOVA** is often used as a standalone test to compare group means, in the context of **multiple linear regression**, it serves to determine the overall **statistical significance** of the model. Essentially, it tests the null hypothesis that there is no relationship between the predictors and the outcome. The results are shown in the following image:

| ANOVA ^a | | | | | | |
|--------------------|------------|----------------|----|-------------|--------|-------------------|
| Model | | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 1350.757 | 2 | 675.378 | 23.458 | .000 ^b |
| | Residual | 489.443 | 17 | 28.791 | | |
| | Total | 1840.200 | 19 | | | |

a. Dependent Variable: score

b. Predictors: (Constant), prep_exams, hours

The key figure in this table is the **F-statistic**, which is calculated as the ratio of the Mean Square Regression to the Mean Square Residual. A higher F-value generally indicates a stronger relationship. However, the most actionable piece of information is the **Sig.** column, which provides the **p-value**. If this value is less than the standard alpha level of 0.05, the model is considered statistically significant, meaning the predictors together have a non-zero effect on the dependent variable.

In our tutorial example, the **p-value** is 0.000 (often reported as $p < .001$). Because this is significantly lower than 0.05, we can confidently reject the null hypothesis. This result confirms that the combination of study hours and preparatory exams is a statistically significant predictor of exam scores. Without a significant result in the **ANOVA** table, individual predictor analysis would be largely irrelevant, as the model as a whole would not be considered reliable.

Step 6: Analyzing Regression Coefficients and Predictor Impact

Once the overall model significance is established, we must examine the **Coefficients** table to understand the specific influence of each individual variable. This table breaks down the contribution of each independent variable, allowing us to see which factors are the most influential and whether their impact is positive or negative. It also provides the necessary data points to construct the final **multiple linear regression** equation. Observe the coefficient details below:

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|-------|------------|-----------------------------|------------|---------------------------|--------|------|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | 67.674 | 2.816 | | 24.033 | .000 |
| | hours | 5.556 | .899 | .902 | 6.179 | .000 |
| | prep_exams | -.602 | .914 | -.096 | -.658 | .519 |

a. Dependent Variable: score

The **Unstandardized B** column provides the specific weights for each variable. The **Constant** (intercept) value of **67.674** represents the predicted exam score if a student spent zero hours studying and took zero prep exams. The coefficient for **hours** is **5.556**, indicating that for every additional hour a student studies, their exam score is expected to increase by 5.556 points, provided that the number of prep exams remains constant. Conversely, the **prep_exams** coefficient is **-0.602**, suggesting a slight negative correlation, though we must check its significance before drawing conclusions.

To determine if these individual effects are meaningful, we look at the **Sig. (p-value)** for each predictor. For **hours**, the **p-value** is 0.000, which is less than 0.05, confirming that hours studied is

a significant predictor. However, for **prep_exams**, the **p-value** is 0.519. Since this is much higher than 0.05, we conclude that the number of prep exams taken does not have a statistically significant relationship with the final score in this specific model. This insight allows researchers to refine their models by potentially removing non-significant variables.

Step 7: Constructing and Applying the Predictive Regression Equation

The final stage of **multiple linear regression** is the construction of a predictive equation based on the unstandardized **coefficients**. This formula allows us to estimate the outcome for any new set of inputs. The general structure for the equation is: **Estimated Value = Constant + (B1 * Variable 1) + (B2 * Variable 2)**. Using the data from our **SPSS** output, we can define the specific equation for student exam scores as follows:

$$\text{Estimated exam score} = 67.674 + 5.556 * (\text{hours}) - 0.602 * (\text{prep_exams})$$

To illustrate the practical application of this formula, consider a student who decides to study for 3 hours and takes 2 preparatory exams. By plugging these values into our equation, we can calculate their expected performance: **67.674 + 5.556 * (3) - 0.602 * (2)**. This calculation results in an estimated exam score of **83.1**. This predictive capability is one of the most powerful aspects of regression analysis, enabling data-driven forecasting in educational, clinical, and business settings.

It is important to remember that because the **prep_exams** variable was found to be statistically non-significant ($p = 0.519$), its inclusion in the equation may add unnecessary complexity without providing real predictive value. In a professional research setting, the analyst might choose to perform a subsequent simple linear regression using only **hours studied** as the predictor. This iterative process of model building and refinement ensures that the final statistical conclusions are both robust and parsimonious, focusing only on the factors that truly drive the observed outcomes.