

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

Authored by  
**Mohammed looti**

November 8, 2025

## RECOMMENDED CITATION

Mohammed looti (2025). *Learn How to Perform Multiple Linear Regression in SPSS: A Step-by-Step Guide*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=12914>

**Multiple linear regression** is a powerful statistical technique utilized to model the linear relationship between a continuous [response variable](#) and two or more [explanatory variables](#). This method allows researchers to determine the overall fit of the model and assess the unique contribution and statistical significance of each predictor. Understanding how to execute and interpret this analysis is fundamental for data-driven decision-making across various fields, including social science, business, and education.

This comprehensive tutorial is designed to explain the process of performing a [multiple linear regression](#) analysis using **SPSS** (Statistical Package for the Social Sciences), a widely used software platform for complex statistical computation.

## Understanding Multiple Linear Regression

Before diving into the software steps, it is essential to grasp the core purpose of this analytical method. Unlike simple linear regression, which uses only one predictor, [multiple linear regression](#) allows for the simultaneous assessment of several independent variables on a single outcome. This capability provides a more nuanced understanding of complex phenomena where outcomes are influenced by multiple interacting factors. The goal is to create a parsimonious model that maximizes the prediction of the response variable while adhering to key statistical assumptions, such as linearity, independence of errors, and homoscedasticity.

The model essentially attempts to fit a hyperplane to the multidimensional data points. The resulting equation estimates the value of the dependent variable based on the weighted contributions of the independent variables. Each weight (or coefficient) represents the expected change in the dependent variable for a one-unit change in that specific independent variable, assuming all other variables in the model are held constant (*ceteris paribus*). This control mechanism is what makes multiple regression a preferred method over running several simple regressions when dealing with correlated predictors.

To illustrate this process within **SPSS**, consider a common research question in educational psychology. We want to investigate whether a student's performance on a standardized exam (our response variable) is influenced by two distinct factors: the total number of hours they dedicate to studying, and the quantity of practice or prep exams they complete. By using multiple linear regression, we can isolate the individual impact of "Hours studied" and "Prep exams taken" on the final "Exam score."

In this specific example, we define our variables as follows:

### **Explanatory Variables (Predictors):**

Hours studied

Prep exams taken

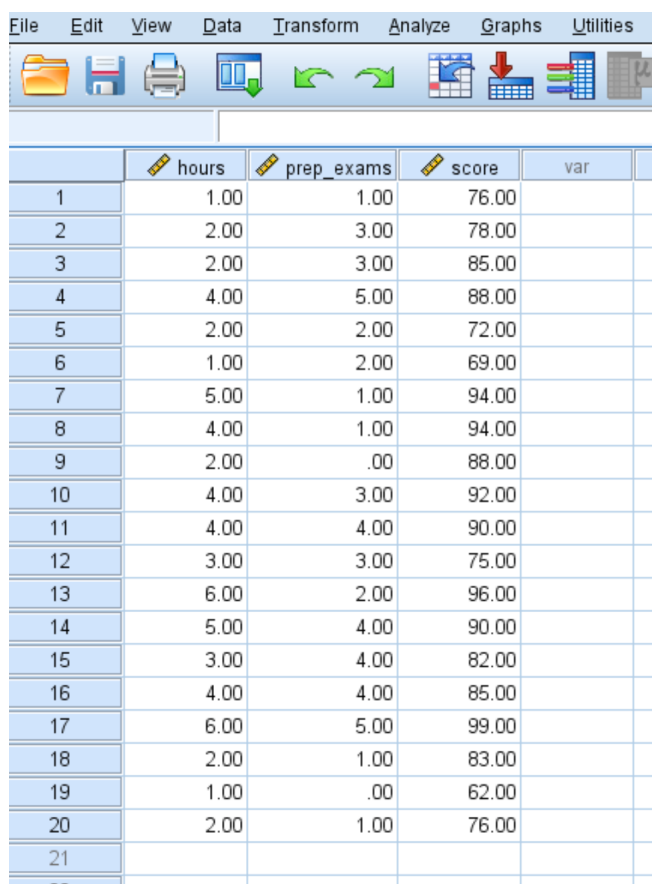
### Response Variable (Outcome):

Exam score

## Setting Up the Analysis in SPSS

The first crucial step in any statistical analysis is accurate data entry and preparation. For this demonstration, we are using a sample dataset representing 20 students, detailing their effort (hours studied and prep exams taken) and their resulting outcome (exam score). Ensure that your data is properly structured in the [SPSS](#) Data View, with each variable correctly defined in the Variable View (typically as Scale measurement for continuous variables).

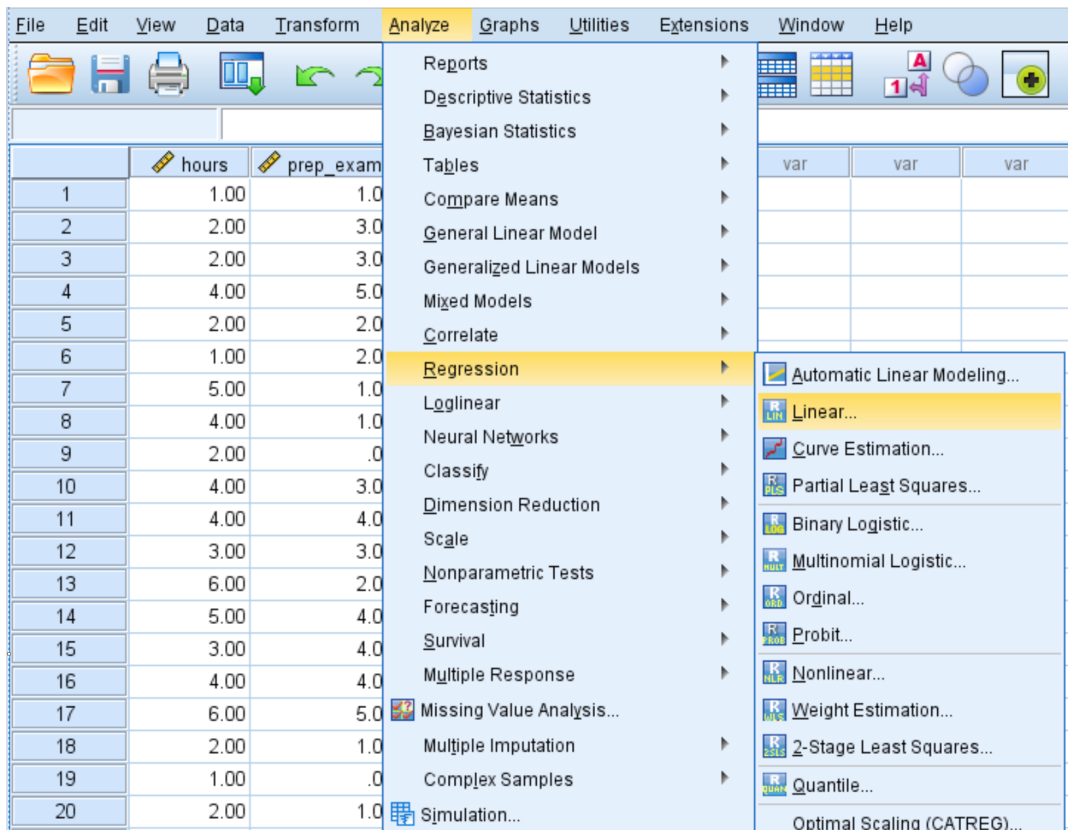
**Step 1: Enter the Data.** The following table visualizes the necessary data structure and the specific observations for our 20 participants. Careful attention to data integrity at this stage prevents errors in the final model output.



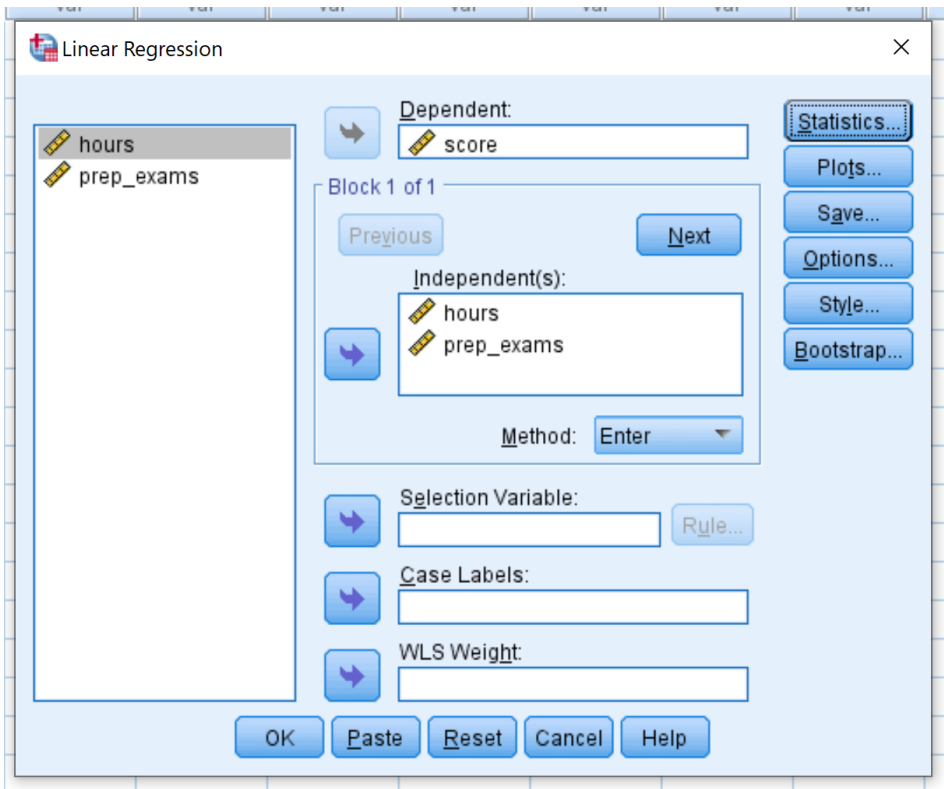
	hours	prep_exams	score	var
1	1.00	1.00	76.00	
2	2.00	3.00	78.00	
3	2.00	3.00	85.00	
4	4.00	5.00	88.00	
5	2.00	2.00	72.00	
6	1.00	2.00	69.00	
7	5.00	1.00	94.00	
8	4.00	1.00	94.00	
9	2.00	.00	88.00	
10	4.00	3.00	92.00	
11	4.00	4.00	90.00	
12	3.00	3.00	75.00	
13	6.00	2.00	96.00	
14	5.00	4.00	90.00	
15	3.00	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: Perform Multiple Linear Regression.** Once the data is entered, navigate through the [SPSS](#) menu system to initiate the regression procedure. Click the **Analyze** tab, hover over

**Regression**, and then select the **Linear** option. This sequence opens the primary dialog box where the predictor and outcome variables are assigned roles.



Within the Linear Regression dialog box, you must correctly assign the variables based on their role in the analysis. Drag the variable representing the outcome, **score**, into the box labeled **Dependent**. Subsequently, drag both predictor variables, **hours** and **prep\_exams**, into the box labeled **Independent(s)**. Ensure that the default Method (Enter) is selected, which forces all independent variables into the model simultaneously. After correctly configuring the variables, click **OK** to execute the analysis and generate the output tables.



Upon execution, the [SPSS](#) output viewer will display several tables necessary for a complete interpretation of the regression results. These tables detail the overall model fit, the statistical significance of the model, and the unique contributions of the individual predictors.

## Interpreting the Model Summary Output

The first table requiring detailed attention is the **Model Summary**. This output provides critical metrics regarding the explanatory power and overall fit of the regression model. It helps us understand how well the chosen [explanatory variables](#) collectively predict the [response variable](#).

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.857 <sup>a</sup>	.734	.703	5.36570

a. Predictors: (Constant), prep\_exams, hours

The two most relevant statistics presented here are the **R Square** and the **Std. Error of the Estimate**. These statistics quantify the quality of the fit and the precision of the predictions:

**R Square (Coefficient of Determination):** This value represents the proportion of variance in the dependent variable that is predictable from the independent variables. Essentially, it tells us how much of the variation in exam scores can be accounted for by knowing the hours studied and the number of prep exams taken. In this example, the **R Square** is **0.734**. This highly favorable result indicates that **73.4%** of the variance in student exam scores is explained by the combination of hours studied and the number of prep exams taken.

**Std. Error of the Estimate:** Often referred to as the root mean square error, this statistic measures the average distance that the observed data points fall from the calculated regression line (or hyperplane in multiple regression). It is expressed in the units of the dependent variable. In our case, the observed exam scores fall an average of **5.3657** units (points) away from the predicted score based on the regression model. A smaller standard error suggests a more accurate and precise predictive model.

While the **R Square** provides an initial indication of model strength, researchers often also look at the Adjusted R Square, particularly when comparing models with different numbers of predictors. The Adjusted R Square penalizes the model for including excessive predictors that do not contribute meaningfully, offering a more conservative estimate of population variance explained. However, the raw R Square remains the primary measure of explanatory power within the sample data.

## Analyzing the ANOVA Table for Overall Fit

The next crucial output table is the **ANOVA** (Analysis of Variance) table. The purpose of the [ANOVA](#) table in regression is to test the null hypothesis that all regression coefficients are zero--meaning that none of the [explanatory variables](#) significantly predict the response variable. This table determines if the overall regression model provides a better prediction than simply using the mean of the dependent variable.

**ANOVA<sup>a</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1350.757	2	675.378	23.458	.000 <sup>b</sup>
	Residual	489.443	17	28.791		
	Total	1840.200	19			

a. Dependent Variable: score

b. Predictors: (Constant), prep\_exams, hours

The key components of the [ANOVA](#) output are the F-statistic and its associated [p-value](#) (labeled Sig.):

**F Statistic:** This is the overall F ratio for the regression model, calculated by dividing the Mean Square of the Regression by the Mean Square of the Residual (Error). A larger F value suggests that the model explains more variance than the unexplained error. This statistic is critical for assessing the collective impact of the predictors.

**Sig. (p-value):** This is the probability associated with the calculated F statistic. It indicates whether the regression model as a whole is **statistically significant**. If the **p-value** is less than the chosen alpha level (typically 0.05), we reject the null hypothesis and conclude that the set of explanatory variables significantly predicts the response variable. In this analysis, the **p-value** is **0.000** (which SPSS displays when the value is less than 0.0005). Since 0.000 is much less than 0.05, we conclude that the variables "hours studied" and "prep exams taken" together have a **statistically significant** association with the "exam score."

The overall significance established by the **ANOVA** table confirms that the model is useful for prediction. However, it does not tell us which individual predictor variables are contributing to this significance. For that detailed information, we must proceed to the Coefficients table.

## Deciphering the Regression Coefficients

The **Coefficients** table is arguably the most important output, as it provides the specific parameter estimates (Unstandardized B values) necessary to construct the regression equation, as well as the individual significance testing for each predictor.

**Coefficients<sup>a</sup>**

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 interpretation of the coefficients is central to understanding the unique impact of each variable:

**Unstandardized B (Constant/Intercept):** This coefficient represents the predicted mean value of the **response variable** when all predictor variables in the model are equal to zero. In this context, the average exam score is predicted to be **67.674** points for a student who studied for zero hours and took zero prep exams. This value serves as the baseline for the regression equation.

**Unstandardized B (hours):** This coefficient (5.556) indicates the average change in the exam score associated with a one-unit increase in hours studied, while strictly controlling for the number of prep exams taken. Specifically, each additional hour spent studying is associated with an

increase of **5.556** points in the exam score, holding the other predictor constant.

**Unstandardized B (prep\_exams):** This coefficient (-0.602) suggests the average change in exam score associated with a one-unit increase in prep exams taken, while controlling for hours studied. Interestingly, this value is negative, meaning each additional prep exam taken is associated with a decrease of **0.602** points in the exam score, assuming hours studied remain the same.

**Sig. (hours):** This is the [p-value](#) for the predictor variable **hours**. Since this value (.000) is less than the significance level of .05, we conclude that hours studied has a [statistically significant](#) independent association with the exam score.

**Sig. (prep\_exams):** This is the [p-value](#) for the predictor variable **prep\_exams**. Since this value (.519) is substantially greater than .05, we cannot conclude that the number of prep exams taken has a [statistically significant](#) unique association with the exam score when hours studied is already included in the model. This suggests that its contribution to explaining variance in the exam score is negligible or redundant compared to the hours studied.

## Formulating and Applying the Regression Equation

The Unstandardized B values from the Coefficients table are the building blocks of the [multiple linear regression](#) equation. This equation allows for practical prediction of the dependent variable based on known values of the independent variables. The general form of the equation is:

$$\hat{Y} = B_0 + B_1X_1 + B_2X_2 + \dots + B_kX_k$$

Using the specific constants derived from our SPSS output (Constant = 67.674, B for hours = 5.556, B for prep\_exams = -0.602), we formulate the predictive model for the estimated exam score:

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

This equation can now be employed for forecasting. For instance, if a student studies for 3 hours and takes 2 prep exams, we can substitute these values into the derived equation to predict their expected performance.

$$\text{Estimated exam score} = 67.674 + 5.556 * (3) - 0.602 * (2)$$

$$\text{Estimated exam score} = 67.674 + 16.668 - 1.204 = 83.138$$

Therefore, a student exhibiting this specific combination of preparation efforts is expected to receive an exam score of approximately 83.1 points based on our regression model. This demonstrates the practical utility of the coefficients in making point estimates.

## Considerations and Model Refinement

While the overall model was deemed **statistically significant** (as determined by the **ANOVA** F-test), the individual test for the variable **prep\_exams** indicated a lack of significance ( $p = 0.519$ ). This lack of significance suggests that, in the presence of the "hours studied" variable, "prep exams taken" does not provide a unique or meaningful contribution to the prediction of the exam score.

In practical modeling, the presence of non-significant predictors often prompts model refinement. If a predictor is found not to be **statistically significant**, researchers may consider removing it to create a more parsimonious model, especially if theoretical justification for its inclusion is weak. Removing a non-significant predictor often improves the clarity and interpretability of the remaining coefficients, and can sometimes lead to a slight increase in the adjusted **R Square** for the final model.

In this scenario, a reasonable next step would be to eliminate the **prep\_exams** variable and perform a simple linear regression using only **hours studied** as the explanatory variable. This simplified model may provide a clearer and equally effective prediction, as the primary explanatory power seems to reside solely with the hours dedicated to studying. Always prioritize models that are both statistically robust and easy to interpret within the context of the underlying theory.