What is the easiest way to Perform Multiple Linear Regression in SPSS?

Authored by

stats writer

December 26, 2025

RECOMMENDED CITATION

stats writer (2025). *What is the easiest way to Perform Multiple Linear Regression in SPSS?*. PSYCHOLOGICAL SCALES. Retrieved from https://scales.arabpsychology.com/?p=108909

For researchers and analysts utilizing SPSS (Statistical Package for the Social Sciences), understanding how to efficiently execute complex statistical procedures is paramount. The most streamlined and preferred method for conducting a Multiple Linear Regression analysis involves leveraging the dedicated Regression procedure, specifically the **Linear** option, found within the Analyze menu. Although the General Linear Model (GLM) procedure can also perform this task, the standard Regression module is purpose-built for linear modeling, offering a more direct route to the desired output tables, including the comprehensive regression equation and the essential model summary statistics.

Multiple Linear Regression is a powerful statistical technique designed to explore the relationship between a single continuous outcome variable (the response variable) and two or more continuous or categorical predictor variables (the explanatory variables). This method provides crucial insights into how well the set of predictors can predict the outcome, and which individual predictors contribute most statistically significantly to the model. By following the steps outlined below, analysts can quickly generate and interpret the required output for robust decision-making and hypothesis testing within the SPSS environment.

To initiate the analysis in SPSS, navigate to the main menu bar. Select **Analyze**, then hover over **Regression**, and finally click on **Linear**. This action opens the primary dialog box where you define your dependent and independent variables, set specific options for statistics, plots, and save desired residuals, ensuring the resulting model is tailored to your research questions. This dedicated interface simplifies the process compared to the broader GLM approach, especially for those new to multivariate statistics.

Multiple linear regression serves as a fundamental analytical tool used to quantify and characterize the relationship between a response variable and a collection of two or more explanatory variables. Its primary goal is to model the linear relationship between these variables, allowing for prediction and inference about the population.

This detailed tutorial provides a step-by-step guide on the correct execution and comprehensive interpretation of a Multiple Linear Regression analysis using IBM SPSS Statistics.

Example Scenario: Factors Affecting Student Exam Performance

Consider a practical scenario in educational research where we aim to determine if two factors--the total time a student dedicates to studying and the number of preparatory exams they complete--statistically significantly influence the final score achieved on a major examination. This investigation requires the application of Multiple Linear Regression to assess the combined and individual impacts of these factors on the outcome variable. This method allows us to isolate the unique effect of each predictor while controlling for the other.

To establish this linear model, we first clearly define our variables based on their statistical role in the analysis:

Explanatory Variables (Predictors):

Hours studied (The time dedicated to preparation)
Prep exams taken (The quantity of practice tests completed)

Response Variable (Dependent):

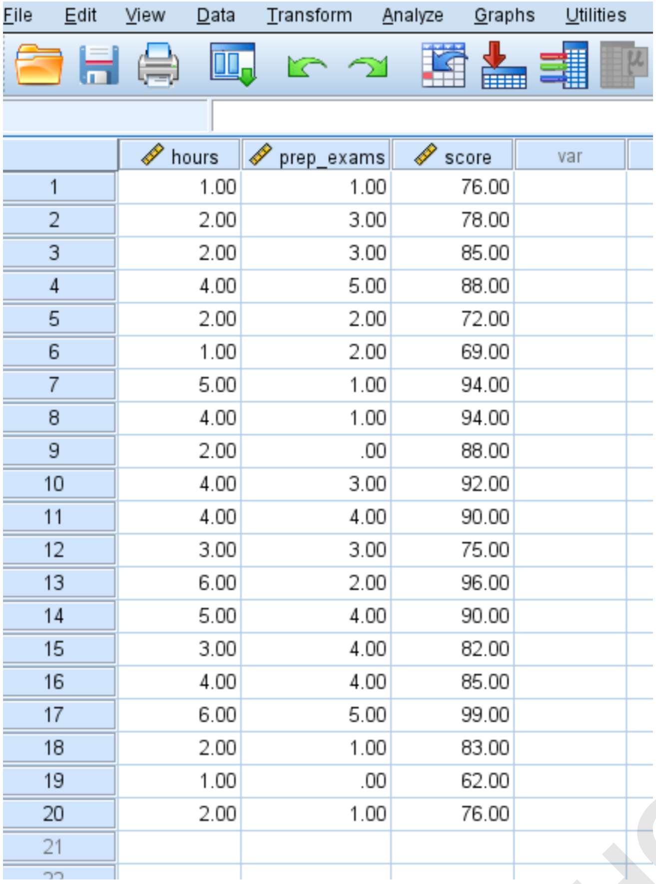
Exam score (The continuous outcome measure)

The subsequent sections detail the exact sequence of steps required within the SPSS interface to process this data and generate the statistical output necessary for rigorous interpretation and hypothesis testing.

Step 1: Data Preparation and Entry in SPSS

Before any statistical analysis can commence, the data must be accurately transcribed into the SPSS Data View. Proper formatting ensures the software correctly identifies the variables and their scale of measurement. For this example, we utilize a dataset comprising observations for 20 students, tracking their hours studied, the number of prep exams completed, and their final exam score.

It is essential that each variable is assigned an appropriate variable type (numeric) and measurement level (scale) in the Variable View tab to guarantee the accuracy of the regression calculations. If the data is not correctly scaled, the assumption of linearity required for this procedure may be violated, leading to unreliable results. The data structure for the 20 hypothetical students should appear as follows in the Data View tab:



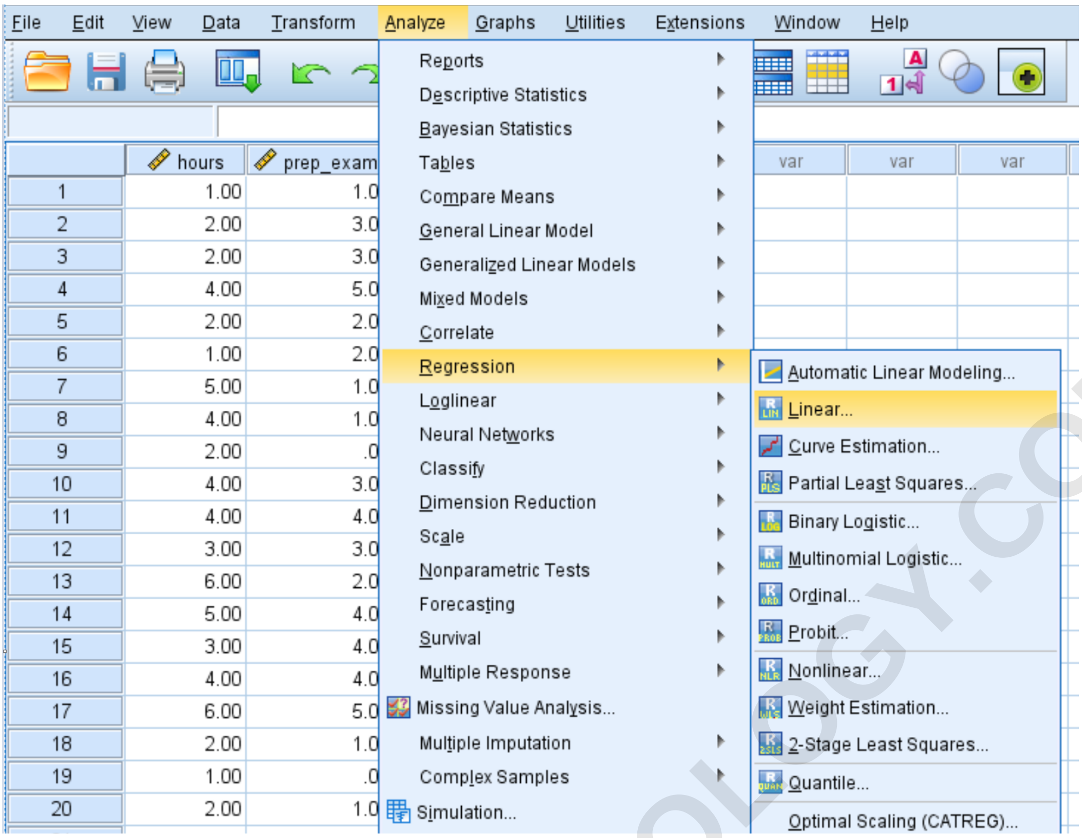
	hours	prep_exams	score	var
1	1.00	1.00	76.00	
2	2.00	3.00	78.00	
3	2.00	3.00	85.00	
4	4.00	5.00	88.00	
5	2.00	2.00	72.00	
6	1.00	2.00	69.00	
7	5.00	1.00	94.00	
8	4.00	1.00	94.00	
9	2.00	.00	88.00	
10	4.00	3.00	92.00	
11	4.00	4.00	90.00	
12	3.00	3.00	75.00	
13	6.00	2.00	96.00	
14	5.00	4.00	90.00	
15	3.00	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				

Once the data entry is verified for completeness and accuracy across all 20 cases, and all variables are correctly named ('score', 'hours', 'prep_exams'), we proceed directly to the statistical modeling phase.

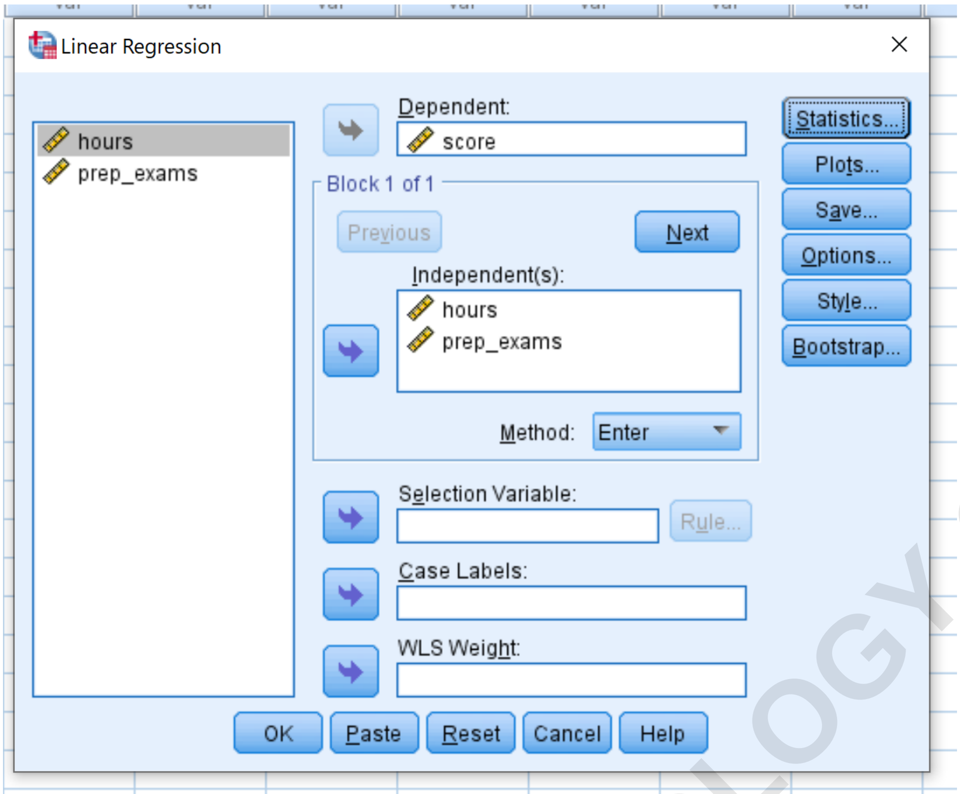
Step 2: Executing the Multiple Linear Regression Procedure

The computational execution of the regression model is initiated through the menu interface, ensuring that the appropriate variables are correctly assigned to their roles (Dependent vs. Independent) within the statistical framework. Misassignment of variables will result in a model that addresses the wrong research question.

First, access the appropriate statistical module by clicking the **Analyze** tab, then selecting **Regression**, followed by **Linear**. This sequence opens the core dialog box for defining the model parameters, as illustrated below, initiating the request for the regression output:



Within this dialog box, the user must define the structural components of the model. The variable **score**, which represents the outcome we are attempting to predict, must be transferred into the box labeled **Dependent**. Subsequently, the predictor variables, **hours** (hours studied) and **prep_exams** (prep exams taken), are collectively moved into the box labeled **Independent(s)**. It is also advisable to check the **Statistics** dialog box to ensure options like **Descriptives** and **Model Fit** are selected. After confirming the variable assignments, click **OK** to execute the analysis and generate the comprehensive output tables in the SPSS Output Viewer.



Upon clicking **OK**, SPSS processes the data and generates a series of statistical tables. The interpretation of these tables is essential for drawing accurate conclusions about the hypothesis being tested regarding the relationship between study habits and scores.

Step 3: Interpreting the Model Summary Output

The first crucial output table generated by the regression analysis is the **Model Summary**. This table provides an essential overview of how well the combination of explanatory variables manages to account for the variability observed in the dependent variable. It serves as the initial assessment of the overall model's predictive power and goodness of fit.

The **Model Summary** table for our example is displayed below, showing key metrics such as R, R Square, and the Standard Error of the Estimate:

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

Interpretation of the key metrics within the Model Summary is critical for assessing model efficacy:

R (Multiple R): This value represents the multiple correlation coefficient between the observed values of the response variable and the predicted values generated by the model. A value close to 1.0 indicates a very strong linear relationship between the predictor variables and the outcome.
R Square: Also known as the Coefficient of Determination, the R Square value quantifies the proportion of the total variance in the dependent variable that is statistically explained by the independent variables included in the model. In this specific educational example, the R Square value of **0.734** indicates that **73.4%** of the variability observed in the student exam scores can be accounted for by the combined influence of hours studied and the number of preparatory exams taken. This suggests the model has substantial explanatory power in predicting student success.
Adjusted R Square: This statistic is generally preferred, especially in multivariate models, as it accounts for the number of independent variables in the model and the sample size. It provides a more conservative and often more accurate estimate of the population R-squared value, penalizing the inclusion of predictors that do not improve the model fit proportionally to the degrees of freedom used.
Std. Error of the Estimate: This metric represents the standard deviation of the residuals, which are the differences between the observed and predicted values. It can be conceptualized as the average distance that the observed data points fall from the computed regression line. In this analysis, the observed scores fall an average distance of **5.3657** units (points) from the regression line. This value is used to gauge the precision of the predictions; smaller values indicate greater accuracy in the prediction of the exam score.

Step 4: Analyzing the ANOVA Table for Overall Model Significance

The ANOVA (Analysis of Variance) table provides the necessary statistics to test the null hypothesis that all regression coefficients are simultaneously equal to zero (i.e., that the predictors, as a group, have no effect). This determines if the regression model, as a whole, is statistically significant in predicting the outcome variable.

The relevant ANOVA output is presented below, detailing the sums of squares and the resultant F-statistic:

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

Interpretation focuses primarily on the F-statistic and its associated significance level (p-value):

F Statistic: This is the overall F ratio for the regression model. It is mathematically calculated by dividing the Mean Square Regression by the Mean Square Residual (or Mean Square Error). This ratio tests whether the variance explained by the model (Regression Sum of Squares) is significantly greater than the unexplained variance (Residual Sum of Squares). A larger F value suggests that the predictors explain a substantial amount of variance relative to the error.

Sig. (p-value): This crucial value is the probability associated with the overall F statistic. It indicates the likelihood of observing the current data if the null hypothesis (that the model has no predictive power, $R^2 = 0$) were true. In this example, the p-value is reported as **0.000** (which conventionally means $p < 0.001$). Since this value is considerably lower than the standard alpha level of 0.05, we confidently reject the null hypothesis. This conclusion implies that the combined set of explanatory variables (hours studied and prep exams taken) has a statistically significant association with the response variable, exam score, confirming that the model is useful.

Step 5: Examining Individual Predictor Contributions (Coefficients)

While the ANOVA table confirms that the model as a whole is useful, the **Coefficients** table is necessary to understand the unique contribution and direction of the relationship for each independent variable. This table also provides the necessary figures to formulate the prediction equation.

The Coefficients output table is shown here, displaying the unstandardized (B) and standardized (Beta) coefficients, along with their respective t-statistics and significance levels:

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
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

Detailed analysis of the Unstandardized Coefficients (B column) and their significance (Sig. column) reveals the specific role and unique predictive power of each variable:

Unstandardized B (Constant/Intercept): The Constant (**67.674**) represents the predicted average value of the response variable (Exam Score) when all predictor variables in the model are set to zero. Substantively, this indicates that a student who studies for zero hours and takes zero prep exams is predicted to achieve an average baseline score of 67.674 points.

Unstandardized B (hours): The coefficient for 'hours studied' is **5.556**. This partial regression coefficient reflects the expected change in the exam score associated with a one-unit increase in hours studied, assuming the other variable (number of prep exams) is held constant (controlled for). The positive sign indicates that more study time leads to higher scores.

Sig. (hours): The p-value for 'hours studied' is **0.000**. Since this value is far less than the conventional significance threshold of 0.05, we conclude that hours studied has a statistically significant unique association with the exam score, making it a critical predictor in the model.

Unstandardized B (prep_exams): The coefficient for 'prep exams taken' is **-0.602**. This indicates that holding hours studied constant, each additional prep exam taken is associated with an average decrease of 0.602 points in the exam score. This unexpected negative relationship, though small, warrants careful consideration of multicollinearity or other confounding factors.

Sig. (prep_exams): The p-value for 'prep_exams' is **0.519**. Because this value is much greater than the 0.05 threshold, we must conclude that the number of prep exams taken does **not** have a statistically significant unique association with the exam score when study hours are already included in the model.

Step 6: Formulating and Applying the Regression Equation

The Unstandardized Coefficients derived in the previous step are used directly to construct the formal Multiple Linear Regression prediction equation. This equation allows researchers to estimate the expected value of the dependent variable based on specific input values of the independent variables.

Using the coefficients from the SPSS output (B? = Constant, B? = hours, B? = prep_exams), our prediction equation is:

Estimated Exam Score = 67.674 + 5.556 * (Hours Studied) - 0.602 * (Prep Exams Taken)

We can use this formula for predictive purposes. For example, consider a hypothetical student who studies for 3 hours and completes 2 prep exams. We can estimate their expected score:

Estimated exam score = 67.674 + 5.556*(3) - 0.602*(2) = 83.1

Therefore, based on our model, this student is expected to receive an exam score of approximately **83.1** points.

Step 7: Considerations for Model Refinement and Parsimony

The thorough interpretation of the Coefficients table revealed a critical statistical finding: the variable 'prep exams taken' was not found to be statistically significant (p = 0.519) in predicting the exam score, once the significant effect of study hours was already accounted for. This is a common occurrence in multiple regression where variables may be highly correlated or where one variable simply dominates the prediction.

In rigorous statistical modeling, this finding suggests that the predictor variable 'prep exams taken' may not be necessary for the most efficient or parsimonious model. While the overall model is statistically significant (as shown by the ANOVA F-test), the goal of good modeling is often to use the fewest predictors necessary to achieve the best fit (highest R Square or Adjusted R Square).

Therefore, a statistician might recommend removing the non-significant variable ('prep exams taken') and re-running the analysis using only 'hours studied' as the explanatory variable. This step simplifies the model to a Simple Linear Regression and ensures that every predictor retained offers a unique, demonstrable, and statistically significant contribution to the prediction of the outcome, aligning with the principle of parsimony.