

# Simple Linear Regression in SPSS: A Step-by-Step Guide

Authored by  
**Mohammed looti**

November 8, 2025

## RECOMMENDED CITATION

Mohammed looti (2025). *Simple Linear Regression in SPSS: A Step-by-Step Guide*.  
PSYCHOLOGICAL STATISTICS. Retrieved from  
<https://statistics.arabpsychology.com/?p=12921>

Simple Linear Regression is a powerful statistical method we can use to understand and model the relationship between a single predictor variable and a single response variable. This technique allows researchers to quantify the extent and nature of this relationship, ultimately enabling prediction and inference.

This comprehensive tutorial explains the step-by-step process of how to effectively perform and interpret [Simple Linear Regression](#) using the [SPSS](#) statistical software package.

## Introduction to Simple Linear Regression

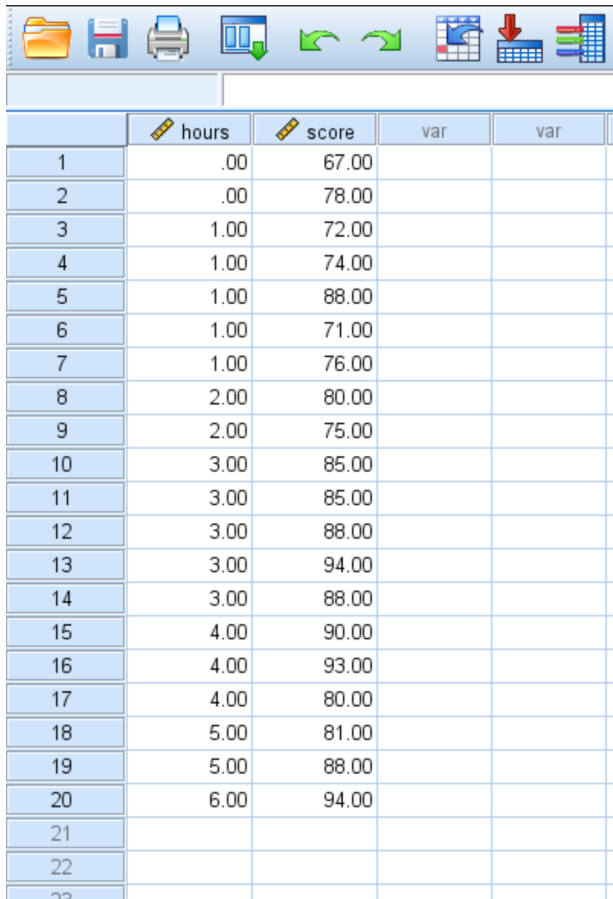
At its core, [Simple Linear Regression](#) (SLR) attempts to fit a straight line to a set of data points that best minimizes the sum of squared errors, a methodology known as the least squares method. The resulting equation is used to predict the value of the dependent variable (response) based on the value of the independent variable (predictor). It is a foundational tool in statistics, crucial for fields ranging from economics to psychology.

Before proceeding with the analysis in [SPSS](#), it is vital to ensure that the data meets several key assumptions, including linearity (the relationship must be approximated by a straight line), independence of errors, and homoscedasticity (constant variance of residuals). Failure to meet these assumptions can lead to unreliable model results and incorrect conclusions.

## Setting Up the Example Dataset in SPSS

To demonstrate the procedure, we will utilize a practical example investigating the relationship between study habits and academic performance. Suppose we have collected data from 20 students, tracking the total number of hours they studied for a particular examination and the final score they achieved. The goal is to determine if hours studied is a statistically significant predictor of exam performance.

The dataset, which must be correctly entered into [SPSS](#)'s Data View, contains two variables: **hours** (the predictor, X) and **score** (the response, Y). Ensure both variables are defined as "Scale" in the Variable View to be treated as continuous quantitative data for the regression analysis.



The screenshot shows the SPSS Data Editor window. The toolbar at the top includes icons for file operations (Save, Print, Copy, Paste), navigation (Home, Back, Forward), and data management (Delete, Insert, Sort, Filter). The data grid below has four columns: 'hours', 'score', 'var', and 'var'. The first 20 rows contain data, with the first two columns having values and the last two being empty. The data points are as follows:

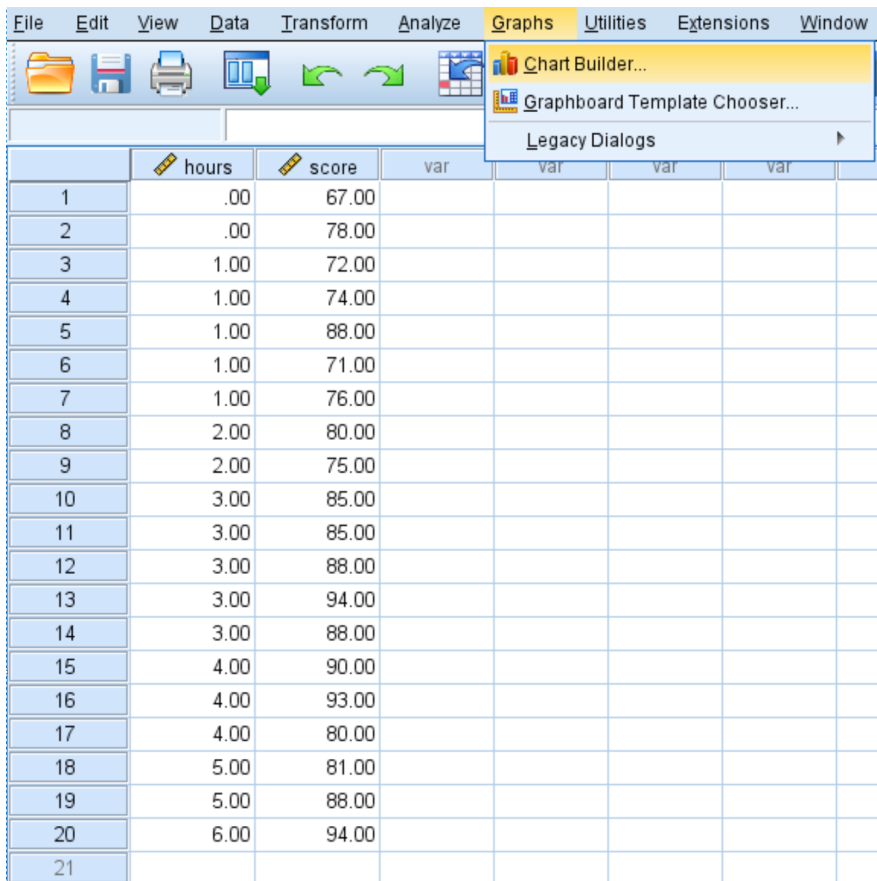
	hours	score	var	var
1	.00	67.00		
2	.00	78.00		
3	1.00	72.00		
4	1.00	74.00		
5	1.00	88.00		
6	1.00	71.00		
7	1.00	76.00		
8	2.00	80.00		
9	2.00	75.00		
10	3.00	85.00		
11	3.00	85.00		
12	3.00	88.00		
13	3.00	94.00		
14	3.00	88.00		
15	4.00	90.00		
16	4.00	93.00		
17	4.00	80.00		
18	5.00	81.00		
19	5.00	88.00		
20	6.00	94.00		
21				
22				
23				

We will now use the following structured steps to perform the [Simple Linear Regression](#) analysis on this dataset, quantifying the precise relationship between hours studied and exam score.

### Step 1: Visualizing the Data with a Scatterplot

The first and most critical step in any regression analysis is data visualization. Creating a [scatterplot](#) allows us to visually assess the linearity assumption. If the relationship between the two variables does not appear linear, then applying simple linear regression would be inappropriate, and alternative techniques (such as non-linear models) should be considered.

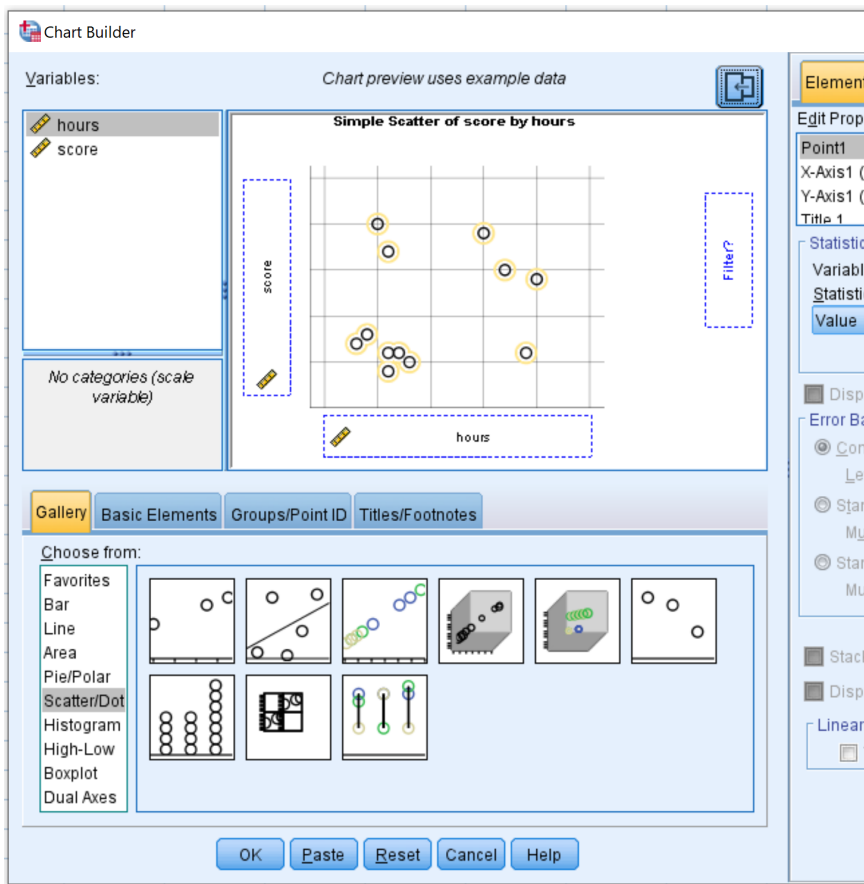
To generate the visualization in [SPSS](#), navigate to the main menu bar and click the **Graphs** tab, followed by **Chart Builder**. This tool provides a flexible interface for generating various types of plots.



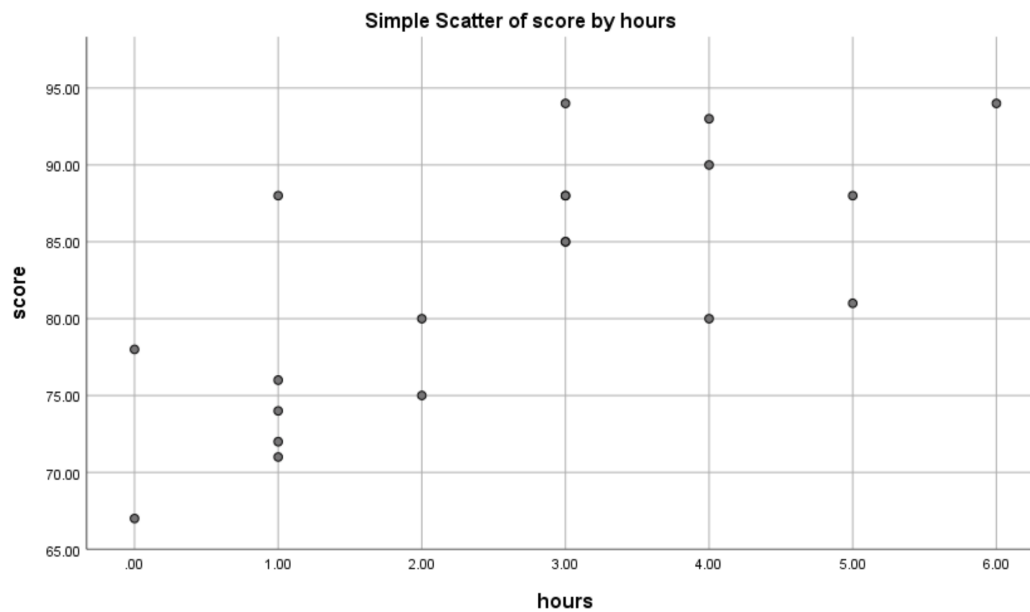
The screenshot shows the SPSS software interface. The 'Graphs' menu is open, displaying options: 'Chart Builder...', 'Graphboard Template Chooser...', and 'Legacy Dialogs'. Below the menu, a data table is visible with columns 'hours' and 'score'. The data points are as follows:

	hours	score	var	var	var	var	v
1	.00	67.00					
2	.00	78.00					
3	1.00	72.00					
4	1.00	74.00					
5	1.00	88.00					
6	1.00	71.00					
7	1.00	76.00					
8	2.00	80.00					
9	2.00	75.00					
10	3.00	85.00					
11	3.00	85.00					
12	3.00	88.00					
13	3.00	94.00					
14	3.00	88.00					
15	4.00	90.00					
16	4.00	93.00					
17	4.00	80.00					
18	5.00	81.00					
19	5.00	88.00					
20	6.00	94.00					
21							

In the Chart Builder interface, select **Scatter/Dot** from the **Choose from** menu at the bottom left, and drag the icon into the main editing window. Next, assign your variables: drag the **hours** variable onto the x-axis (Independent) and the **score** variable onto the y-axis (Dependent). Once the variables are assigned, click **OK** to generate the chart.

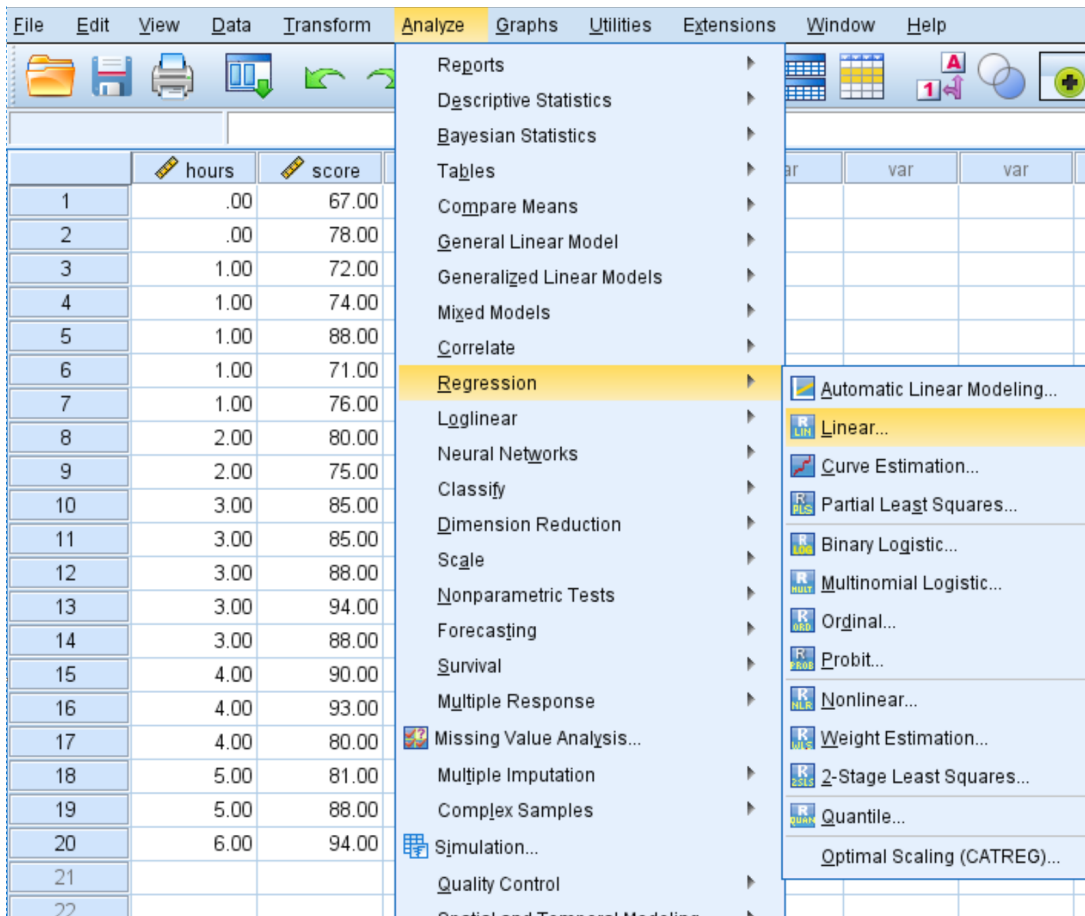


The resulting [scatterplot](#) clearly reveals a **positive linear relationship** between the two variables. We can observe that as the number of hours studied increases, the corresponding exam score generally rises. Since this visualization confirms the linearity assumption, we can confidently proceed with fitting the simple linear regression model to the dataset.

**→ GGraph**

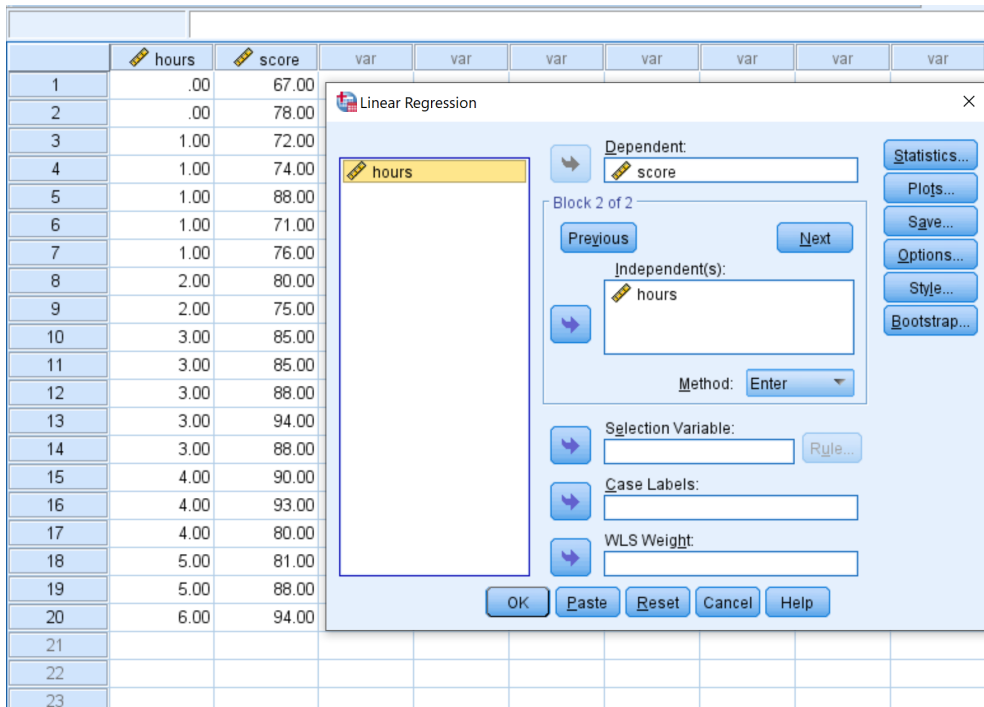
## Step 2: Fitting the Simple Linear Regression Model

With the linearity assumption confirmed, we can now instruct [SPSS](#) to calculate the regression line parameters. To initiate the regression procedure, navigate to **Analyze**, select **Regression**, and then choose **Linear**. This opens the main dialog box for specifying the model.



In the new dialog box that appears, accurately designate the dependent and independent variables. Drag the **score** variable (the outcome we are trying to predict) into the box labeled **Dependent**, and drag the **hours** variable (the predictor) into the box labeled **Independent(s)**. For simple linear regression, only one independent variable is used.

While advanced options exist for selecting statistics or plotting residuals, for a basic analysis, simply ensuring the variables are correctly placed is sufficient. Click **OK** to execute the command and generate the statistical output, which will appear in the SPSS Output Viewer window.



### Step 3: Detailed Interpretation of SPSS Output

The results output generated by [SPSS](#) provides several tables detailing the model's performance and the specific coefficients of the regression line. We must analyze two key tables: the Model Summary and the Coefficients table.

#### Model Summary Interpretation

The first critical table to examine is the **Model Summary**, which provides metrics regarding the overall fit and strength of the relationship established by the regression equation.

**Model Summary<sup>b</sup>**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.712 <sup>a</sup>	.506	.479	5.86100

a. Predictors: (Constant), hours

b. Dependent Variable: score

**R Square:** This value represents the [Coefficient of Determination](#). It is the proportion of the variance in the response variable (exam score) that is predictable from the explanatory variable (hours studied). In this instance, the R Square value of 0.506 indicates that **50.6%** of the total

variation observed in the exam scores can be explained by the variation in the number of hours studied. The remaining variance is attributed to other factors not included in the model.

**Std. Error of the Estimate:** This metric quantifies the average magnitude of the errors (residuals). It is the average distance that the observed score values fall from the calculated regression line. A lower value indicates a more precise model. In this example, the observed values fall an average of **5.861** units (score points) from the regression line, which serves as a measure of the model's accuracy in prediction.

### Coefficients Table Interpretation

The **Coefficients** table is arguably the most important, as it provides the necessary values (the intercept and the slope) required to construct the final regression equation and interpret the specific impact of the predictor variable.

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	73.662	2.410		30.571	.000
	hours	3.342	.778	.712	4.297	.000

a. Dependent Variable: score

**Unstandardized B (Constant):** This coefficient represents the Y-intercept of the regression line. It tells us the predicted average value of the response variable (score) when the predictor variable (hours studied) is equal to zero. Here, the constant of **73.662** implies that a student who studies for zero hours is predicted to achieve an average exam score of 73.662 points. Caution should always be exercised when interpreting the intercept if zero is outside the range of observed data.

**Unstandardized B (hours):** This is the regression coefficient, or slope. It indicates the average change in the response variable associated with a one-unit increase in the predictor variable. The value of **3.342** means that for each additional hour studied, the exam score is expected to increase by 3.342 points, on average. This is the key quantification of the relationship.

**Sig (hours):** This column contains the [P-value](#) associated with the t-test for the predictor's coefficient. This test assesses whether the slope is significantly different from zero. Since the observed [P-value](#) (reported as 0.000, meaning less than 0.0005) is far below the conventional significance level of 0.05, we reject the null hypothesis. We conclude that the predictor variable **hours** is **statistically significant** in predicting exam scores.

### Step 4: Constructing and Applying the Regression Equation

The ultimate practical output of a simple linear regression analysis is the predictive equation, formed by utilizing the unstandardized coefficients (Constant and Slope) derived from the Coefficients table. The general form of the equation is: Estimated Y = Constant + (Slope \* X).

Substituting our specific results into this linear model yields the following regression equation:

$$\text{Estimated exam score} = 73.662 + 3.342 * (\text{hours})$$

This equation can now be used for prediction within the range of the observed data. For instance, we can calculate the expected score for a student who reports studying for exactly 3 hours.

By plugging the value X=3 into the model, we calculate the estimated score:

$$\text{Estimated exam score} = 73.662 + 3.342 * (3) = 73.662 + 10.026 = 83.688$$

Therefore, a student who studies for 3 hours is expected to receive an exam score of 83.688, based on the statistical relationship modeled by our dataset.

## Step 5: Reporting the Results

The final stage of the analysis involves summarizing the findings in a clear, standardized format, typically adhering to academic or professional reporting guidelines. This summary should include the descriptive statistics, the overall model fit (R-square), and the specific parameters of the regression line, along with the test statistics.

When reporting the results, it is essential to cite the degrees of freedom and the relevant test statistics (t-value and p-value) to allow readers to evaluate the statistical significance of the findings independently. Here is an example of how the results of this [Simple Linear Regression](#) could be formally presented:

A simple linear regression was performed to quantify the relationship between hours studied and exam score received, utilizing a sample of 20 students.

Results showed that there was a statistically significant positive relationship between hours studied and exam score ( $t = 4.297$ ,  $p < 0.001$ ). Hours studied accounted for 50.6% ( $R^2 = 0.506$ ) of the explained variability in exam score.

The final regression equation was determined to be:

$$\text{Estimated exam score} = 73.662 + 3.342 * (\text{hours})$$

This indicates that each additional hour studied is associated with an increase of **3.342** points in the predicted exam score, on average.

## **Additional Resources for SPSS Analysis**

Mastering statistical analysis in [SPSS](#) often requires exploring different methods and advanced options.

The following tutorials explain how to perform other common statistical tasks in SPSS, expanding upon the foundational knowledge gained from this regression guide: