

How to Run a Simple Linear Regression in SPSS: A Step-by-Step Guide

Authored by
stats writer

March 14, 2026

RECOMMENDED CITATION

stats writer (2026). *How to Run a Simple Linear Regression in SPSS: A Step-by-Step Guide*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=135787>

Understanding the Fundamentals of Simple Linear Regression

Simple linear regression is a foundational statistical method used by researchers and data analysts to model the relationship between two **continuous variables**. The primary objective of this technique is to determine how a change in one variable, known as the **independent variable** (or predictor), influences the change in another variable, known as the **dependent variable** (or response). By fitting a **mathematical model** to the observed data, analysts can quantify the strength of this association and make informed predictions about future outcomes. In the context of simple linear regression, the model assumes a linear relationship, meaning that the data can be best described by a straight line rather than a curve or more complex geometric shape.

Performing this analysis in a robust software environment like SPSS (Statistical Package for the Social Sciences) provides a systematic and efficient way to process large datasets. The software automates the complex calculations involved in **least squares estimation**, allowing the user to focus on the interpretation of the results. Beyond mere calculation, SPSS offers a variety of diagnostic tools to ensure that the assumptions of the regression model are met. These assumptions include **linearity**, **independence of observations**, **homoscedasticity**, and the **normality** of the error distribution. Understanding these theoretical underpinnings is essential before diving into the procedural steps of the software.

The practical utility of linear regression spans across numerous fields, including economics, psychology, and biology. For instance, a researcher might use it to understand how education levels impact income, or how dosage amounts affect the recovery time of a patient. In each case, the goal remains the same: to find a line of best fit that minimizes the **sum of squared residuals**. This tutorial will guide you through the process of conducting a simple linear regression using a practical example, ensuring you can navigate the software interface and interpret the subsequent statistical output with professional clarity.

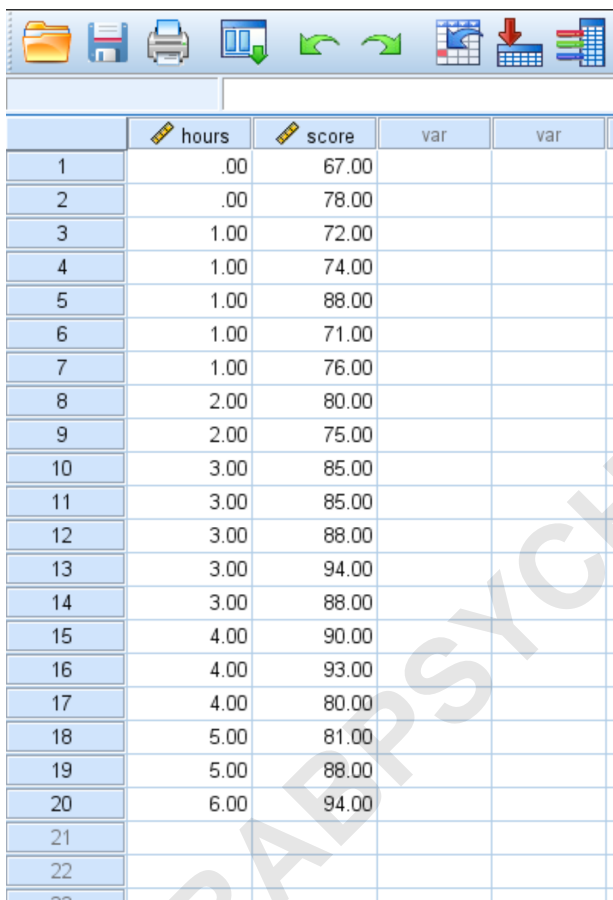
Preparing Your Dataset and Defining Research Goals

Before initiating the analysis in SPSS, it is critical to ensure that your data is properly formatted and that your research question is clearly defined. In regression analysis, variables must be measured at the **interval or ratio scale** to be considered continuous. If your data contains outliers or missing values, these must be addressed during the **data cleaning** phase, as they can significantly skew the regression line and lead to inaccurate conclusions. A well-organized dataset is the backbone of any valid statistical inference.

Consider a scenario where we want to investigate the impact of academic effort on student performance. In this specific example, we have a dataset representing 20 students. The **independent variable** is the number of "hours studied," which we hypothesize will predict the

"exam score," our **dependent variable**. By examining this relationship, we can determine not just if studying helps, but exactly how much an additional hour of study might improve a student's final grade. This quantitative insight is far more valuable than a simple qualitative observation.

The following image illustrates the dataset structure within the SPSS Data View. Note how each row represents a unique observation (a student) and each column represents a variable of interest. This alignment is necessary for the software to correctly map the data points during the regression procedure.



The image shows a screenshot of the SPSS Data View window. The window title bar includes standard icons for file operations (Open, Save, Print, Copy, Paste) and data manipulation (Undo, Redo, Sort, Filter, Hide, Show). The data grid below has the following structure:

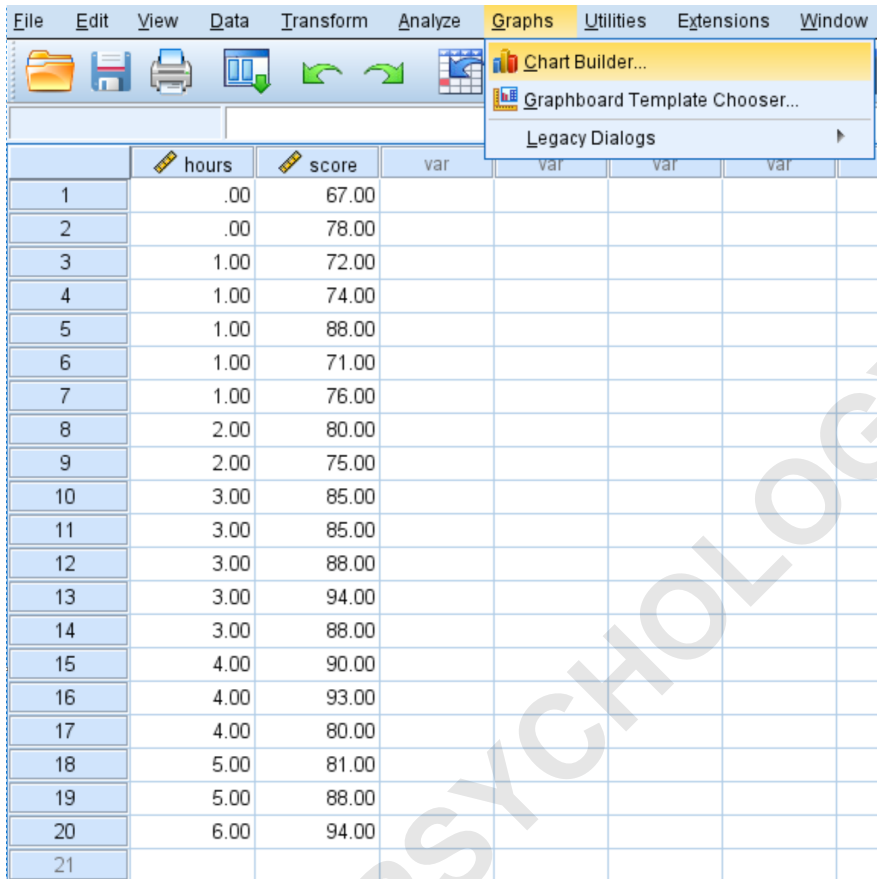
	hours	score	var	var
1	.00	67.00		
2	.00	78.00		
3	1.00	72.00		
4	1.00	74.00		
5	1.00	88.00		
6	1.00	71.00		
7	1.00	76.00		
8	2.00	80.00		
9	2.00	75.00		
10	3.00	85.00		
11	3.00	85.00		
12	3.00	88.00		
13	3.00	94.00		
14	3.00	88.00		
15	4.00	90.00		
16	4.00	93.00		
17	4.00	80.00		
18	5.00	81.00		
19	5.00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

Step 1: Visualizing the Data through Scatterplots

The first imperative step in any regression analysis is **exploratory data analysis**, specifically the creation of a scatterplot. Visualizing the data allows you to confirm that a linear relationship actually exists between your variables. If the scatterplot reveals a curved pattern, a simple linear model would be inappropriate, and you might need to consider **polynomial regression** or **data transformations**. Additionally, the plot helps in identifying **influential points** or outliers that might disproportionately affect the slope of the regression line.

To generate this visualization in SPSS, navigate to the **Graphs** menu and select **Chart Builder**.

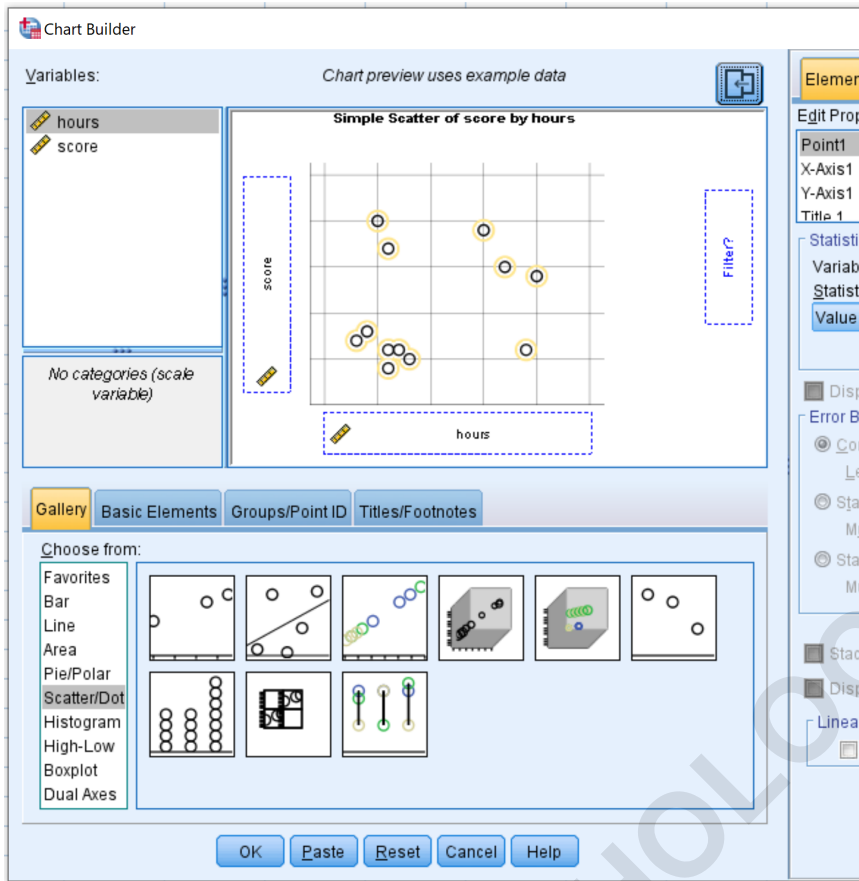
This tool provides a highly customizable interface for generating professional-grade graphics. Within the Chart Builder, you will select the "Scatter/Dot" option from the gallery and drag the "Simple Scatter" icon into the preview window. The intuitive drag-and-drop system ensures that even complex visualizations are accessible to users of all skill levels.



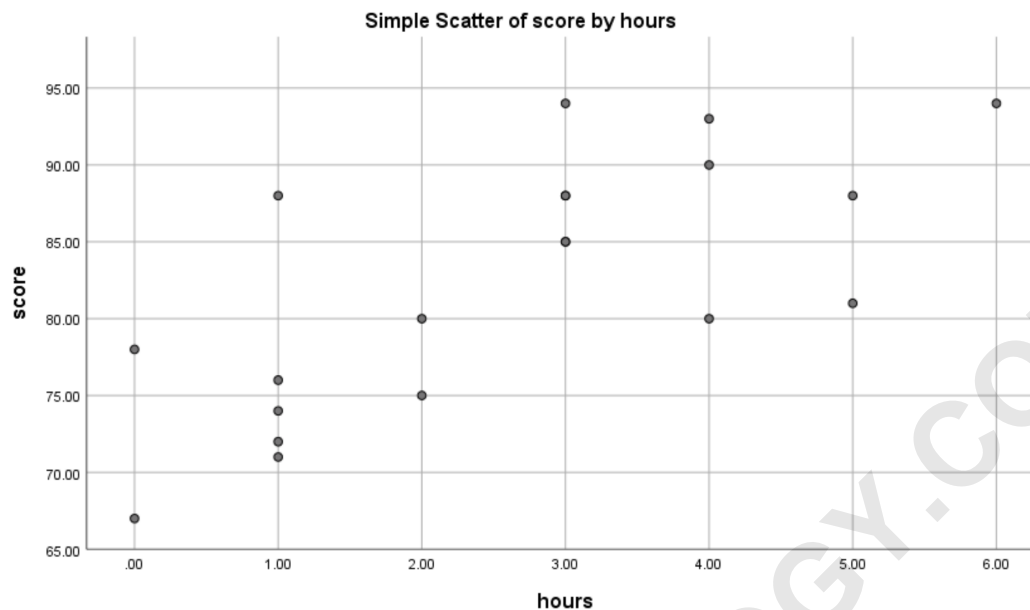
The screenshot shows the SPSS software interface. The 'Graphs' menu is open, displaying options: 'Chart Builder...', 'Graphboard Template Chooser...', and 'Legacy Dialogs'. Below the menu, a data table is visible with columns 'hours' and 'score'. The data points are as follows:

	hours	score	var	var	var	var
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						

Once the basic chart structure is selected, you must assign your variables to the correct axes. In standard statistical practice, the **independent variable** (hours) is placed on the **x-axis** (horizontal), while the **dependent variable** (score) is placed on the **y-axis** (vertical). This layout allows the viewer to see how the score changes as hours increase. After configuring these settings, click **OK** to generate the output.

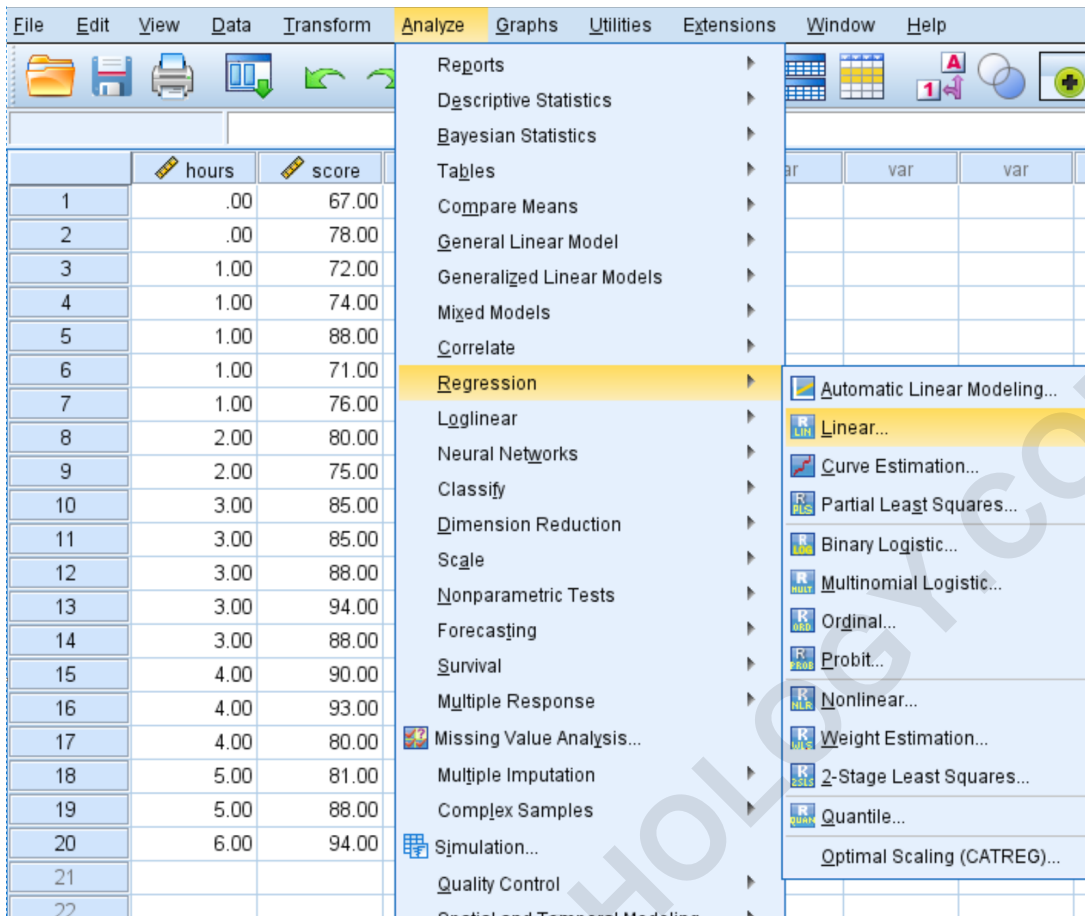


The resulting scatterplot provides an immediate visual confirmation of our hypothesis. As seen in the generated chart below, the data points follow a clear upward trend. This indicates a **positive correlation**: as students invest more time in studying, their exam scores generally increase. Because the points cluster around an imaginary straight line, we can confidently proceed with the **linear regression model**.

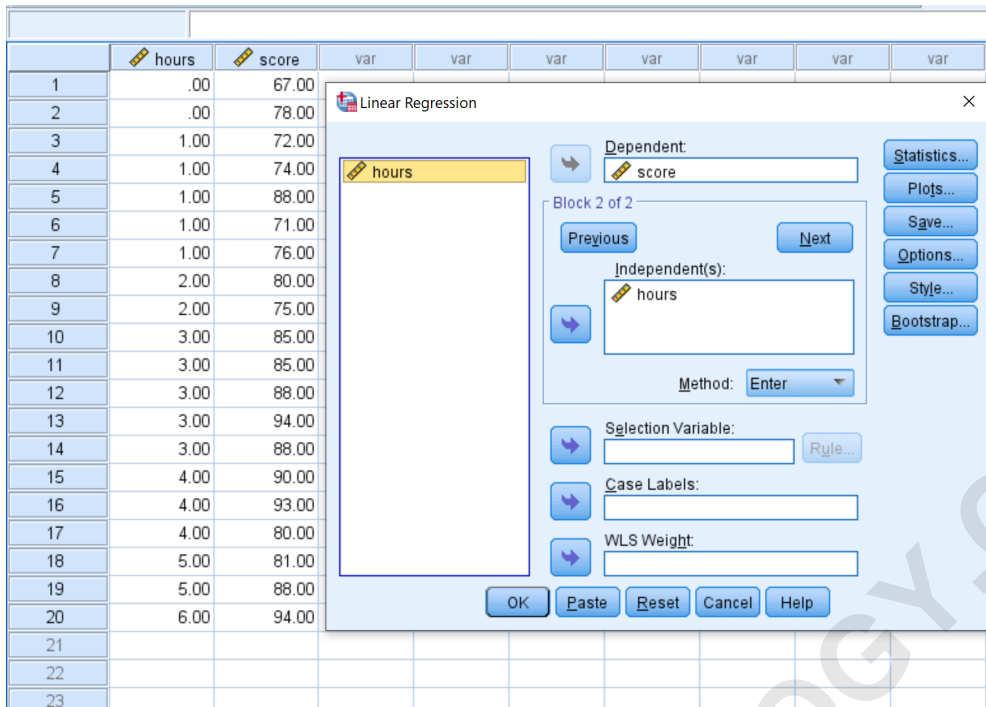
→ GGraph

Step 2: Executing the Linear Regression Procedure

With the visual confirmation complete, the next phase involves the actual execution of the regression algorithm. In SPSS, this is handled through the **Analyze** menu. By selecting **Regression** and then **Linear**, you open the primary configuration dialog for the model. This window allows you to define the mathematical structure of your test and request additional statistics, such as **confidence intervals** or **residual diagnostics**, which are vital for a comprehensive analysis.



In the "Linear Regression" dialog box, you must specify the roles of your variables. Move the "score" variable into the **Dependent** box and the "hours" variable into the **Independent(s)** box. While this tutorial focuses on **simple linear regression** (one predictor), this same interface is used for **multiple linear regression**, where several predictors are analyzed simultaneously. For now, we will maintain the simple model to isolate the effect of study hours.



Before finalizing the procedure, you might explore the **Statistics** and **Plots** sub-menus to request further data. For example, requesting a **Durbin-Watson test** can help check for **autocorrelation**, while plotting **standardized residuals** can help verify the assumption of **homoscedasticity**. For this basic tutorial, clicking **OK** will suffice to generate the standard regression output tables in the SPSS Viewer window.

Step 3: Evaluating Model Fit and the R-Squared Value

The first critical table in the SPSS output is the **Model Summary**. This table provides a high-level overview of how well the regression model fits the observed data. The most prominent metric here is the **R Square** (or coefficient of determination), which ranges from 0 to 1. An R-squared value of 1 indicates a perfect fit, while a value of 0 indicates that the model explains none of the variability in the dependent variable.

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.712 ^a	.506	.479	5.86100

a. Predictors: (Constant), hours

b. Dependent Variable: score

Interpreting the **Model Summary** involves looking at two key statistics:

R Square: In our exam score example, the value is **0.506**. This means that **50.6%** of the variance in exam scores can be explained by the number of hours studied. This is a substantial percentage in the social sciences, suggesting that study time is a major factor in performance, though other factors (like prior knowledge or sleep) likely account for the remaining 49.4%.

Std. Error of the Estimate: This metric represents the average deviation of the data points from the regression line. Here, the value is **5.861**, indicating that the model's predictions are, on average, within approximately 5.86 points of the actual exam scores. A smaller standard error generally indicates a more precise model.

It is important to note that while a high R-squared value is desirable, it does not necessarily mean the model is "good" or that the relationship is **causal**. Correlation does not imply causation, and the R-squared must be interpreted within the context of the specific field of study. Furthermore, the **Adjusted R-Square** is often preferred in multiple regression as it penalizes the addition of unnecessary variables that do not contribute to the model's explanatory power.

Step 4: Analyzing Regression Coefficients and Statistical Significance

After assessing the overall fit, we move to the **Coefficients** table. This section provides the specific values needed to construct our **regression equation** and evaluates the **statistical significance** of the predictor. The unstandardized coefficients represent the raw change in the dependent variable for every one-unit increase in the independent variable. This table is the "heart" of the analysis, as it reveals the nature and reliability of the relationship.

Coefficients^a

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

To interpret the **Coefficients** table correctly, you must examine the following components:

Unstandardized B (Constant): Often called the **y-intercept**, this represents the predicted value of the dependent variable when the independent variable is zero. In this case, if a student studies for zero hours, their predicted exam score is **73.662**. This provides a baseline for the analysis.

Unstandardized B (hours): This is the **slope** of the regression line. It indicates that for every additional hour a student studies, their exam score is expected to increase by **3.342** points. This

positive value confirms the upward trend observed in our scatterplot.

Sig. (hours): This is the p-value associated with the **null hypothesis** that the slope is equal to zero (meaning study hours have no effect). Since the p-value is **< 0.05**, we reject the null hypothesis and conclude that "hours studied" is a **statistically significant** predictor of "exam score."

Furthermore, the **t-statistic** (found in the "t" column) is calculated by dividing the B coefficient by its standard error. A larger absolute t-value corresponds to a smaller p-value. In professional reporting, it is common to include both the t-statistic and the p-value to provide a complete picture of the statistical significance of the findings.

Step 5: Formulating and Applying the Regression Equation

The final practical output of a **simple linear regression** is the **regression equation**. This mathematical formula allows you to predict the value of the dependent variable for any given value of the independent variable. The general form of the equation is $Y = a + bX$, where Y is the predicted score, 'a' is the constant (intercept), 'b' is the slope coefficient, and X is the value of the predictor variable. Using the values from our SPSS output, we can construct the following specific model:

Estimated exam score = 73.662 + 3.342 * (hours)

This equation is not merely theoretical; it has significant predictive power. For example, if a teacher wants to estimate the score for a student who plans to study for 3 hours, they can simply plug that value into the formula. By performing the calculation $73.662 + 3.342 * (3)$, we arrive at an estimated score of **83.688**. This ability to forecast outcomes based on historical data is what makes linear regression such a potent tool in decision-making and strategic planning.

However, users should be cautious about **extrapolation**--predicting values outside the range of the original data. If our dataset only included students who studied between 0 and 10 hours, using the equation to predict the score of someone studying for 50 hours would likely result in an invalid or impossible score (e.g., over 100%). Always ensure that your applications of the model remain within the logical bounds of the **sampled population** and the observed data range.

Best Practices for Reporting Regression Results

When presenting your findings in a research paper, thesis, or business report, it is essential to summarize the results in a clear and standardized format. A good report should mention the **sample size**, the **statistical significance** of the model, the **effect size** (R-squared), and the **regression equation** itself. This ensures that your audience can verify your claims and understand the practical implications of your analysis. Following formal APA style or similar professional

guidelines is highly recommended.

A comprehensive summary of our study might look like the following:

A **simple linear regression** was conducted to determine the relationship between the number of hours studied and the resulting exam scores among a sample of 20 students. The analysis revealed a **statistically significant** positive relationship between the two variables (**t = 4.297, p < 0.001**). The predictor variable, hours studied, accounted for **50.6%** of the total variance in exam scores, indicating a moderate to strong explanatory effect.

The derived regression equation was: **Estimated exam score = 73.662 + 3.342 * (hours)**. On average, for every additional hour a student spent studying, their exam score increased by **3.342** points. These findings suggest that academic preparation is a vital component of student success in this specific context, providing a clear quantitative basis for academic counseling and intervention strategies.

By following these systematic steps--from initial visualization to final reporting--you can leverage SPSS to transform raw data into actionable insights. Whether you are a student, a researcher, or a professional analyst, mastering **simple linear regression** is a critical milestone in your journey toward statistical literacy and data-driven excellence.