Learning to Create Scatterplots with Regression Lines in SPSS

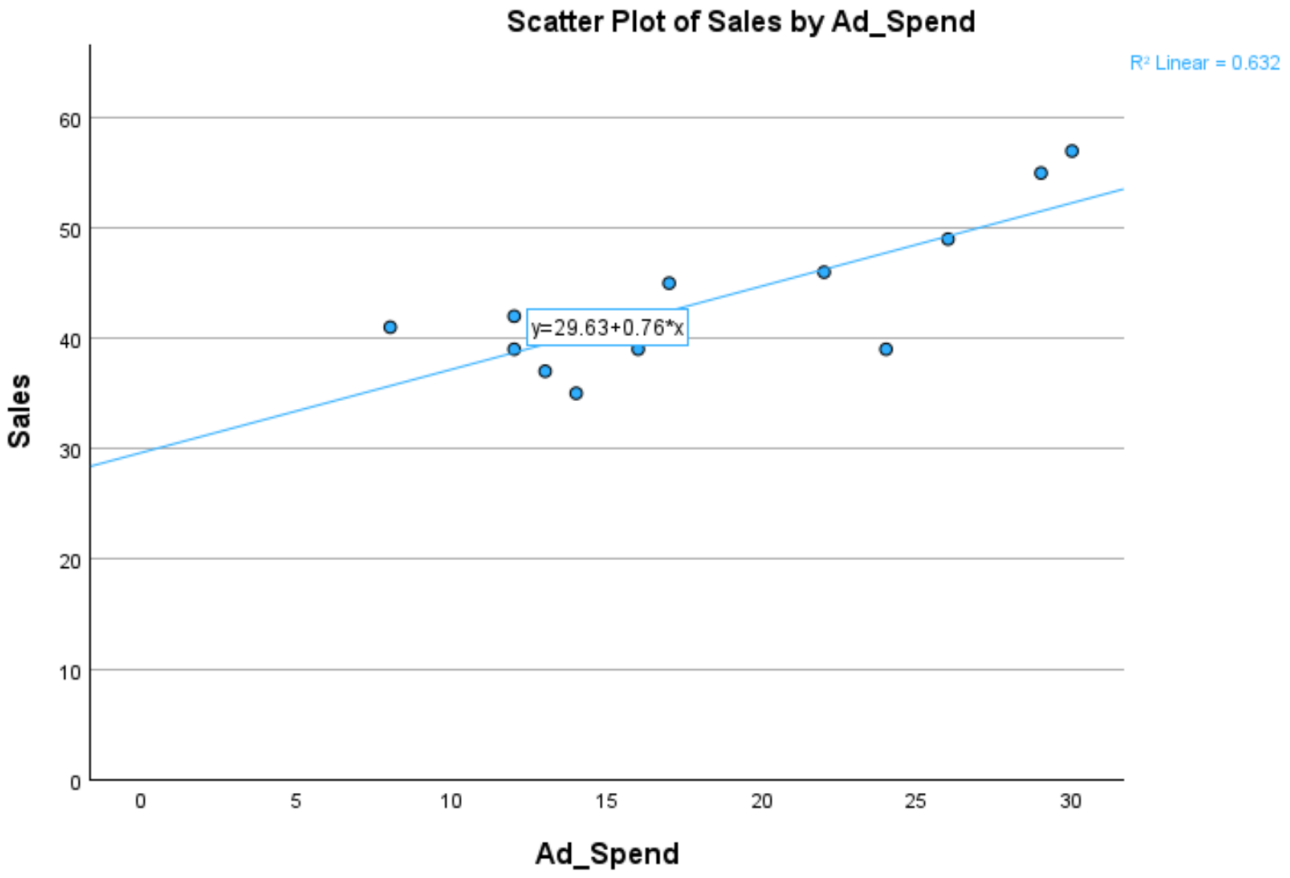
Authored by

Mohammed looti

November 12, 2025

Data visualization is an essential component of rigorous statistical analysis. One of the most insightful graphical tools used to explore the relationship between two continuous variables is the scatterplot. When analysts perform predictive modeling, such as linear regression, it is vital to overlay the calculated line of best fit directly onto the plot. This detailed guide offers a comprehensive, step-by-step tutorial on how to seamlessly generate a high-quality scatterplot complete with a fitted regression line and its corresponding equation using the powerful software, **SPSS** (Statistical Package for the Social Sciences).

The resulting visualization provides immediate and clear insight into the nature and strength of the linear relationship between your variables. Below is an example of the final output we aim to achieve, which includes the data points, the regression line, the equation, and the R^2 statistic:

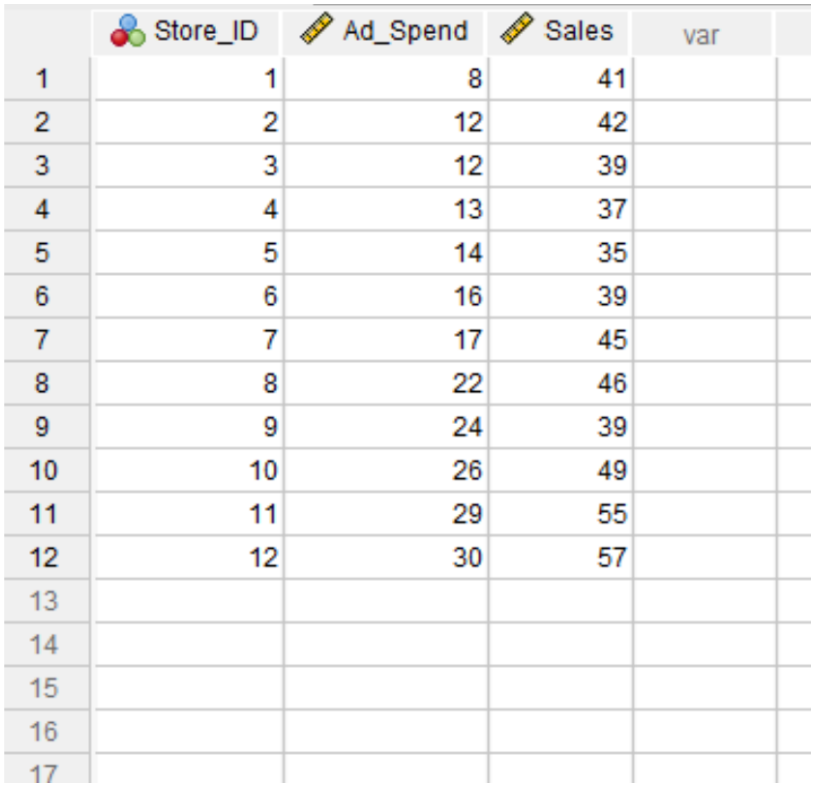


Fortunately, **SPSS** simplifies this complex graphing process through its intuitive **Chart Builder** interface. By following the subsequent steps precisely, you will learn how to generate, customize, and annotate this key statistical visualization effectively for reporting and analysis.

Step 1: Preparing and Structuring Your Data in SPSS

Before any visualization can be generated, the underlying data must be correctly structured and entered into the **Data View** within **SPSS**. For illustrative purposes, we will use a small sample dataset representing observations from twelve distinct retail stores. This dataset tracks two key continuous metrics: the total investment in advertising (labeled **Ad_Spend**) and the resulting total revenue generated (labeled **Sales**).

It is critical to ensure that your variables are correctly defined in the **Variable View**. Both variables should be set as **Numeric** and classified with the **Scale** measurement level, as they represent continuous data. The data input screen should reflect a structure similar to the image below, with each row representing a store and the columns representing the recorded metrics:



	Store_ID	Ad_Spend	Sales	var
1	1	8	41	
2	2	12	42	
3	3	12	39	
4	4	13	37	
5	5	14	35	
6	6	16	39	
7	7	17	45	
8	8	22	46	
9	9	24	39	
10	10	26	49	
11	11	29	55	
12	12	30	57	
13				
14				
15				
16				
17				

Our analytical objective is to create a scatterplot where **Ad_Spend** is designated as the **predictor variable** (or independent variable), mapped along the **x-axis**. Conversely, **Sales** will function as the **response variable** (or dependent variable), plotted along the **y-axis**. The added regression line will then visually model how fluctuations in advertising expenditure correlate with predicted changes in sales figures.

Step 2: Accessing the Chart Builder and Defining Axes

Once the dataset is prepared and verified, the next step involves navigating to the visualization tools. To begin the chart creation process, locate and click the **Graphs** tab in the main menu bar, and then select the **Chart Builder** option. This action initiates a specialized dialogue box designed for sophisticated graph generation and customization.

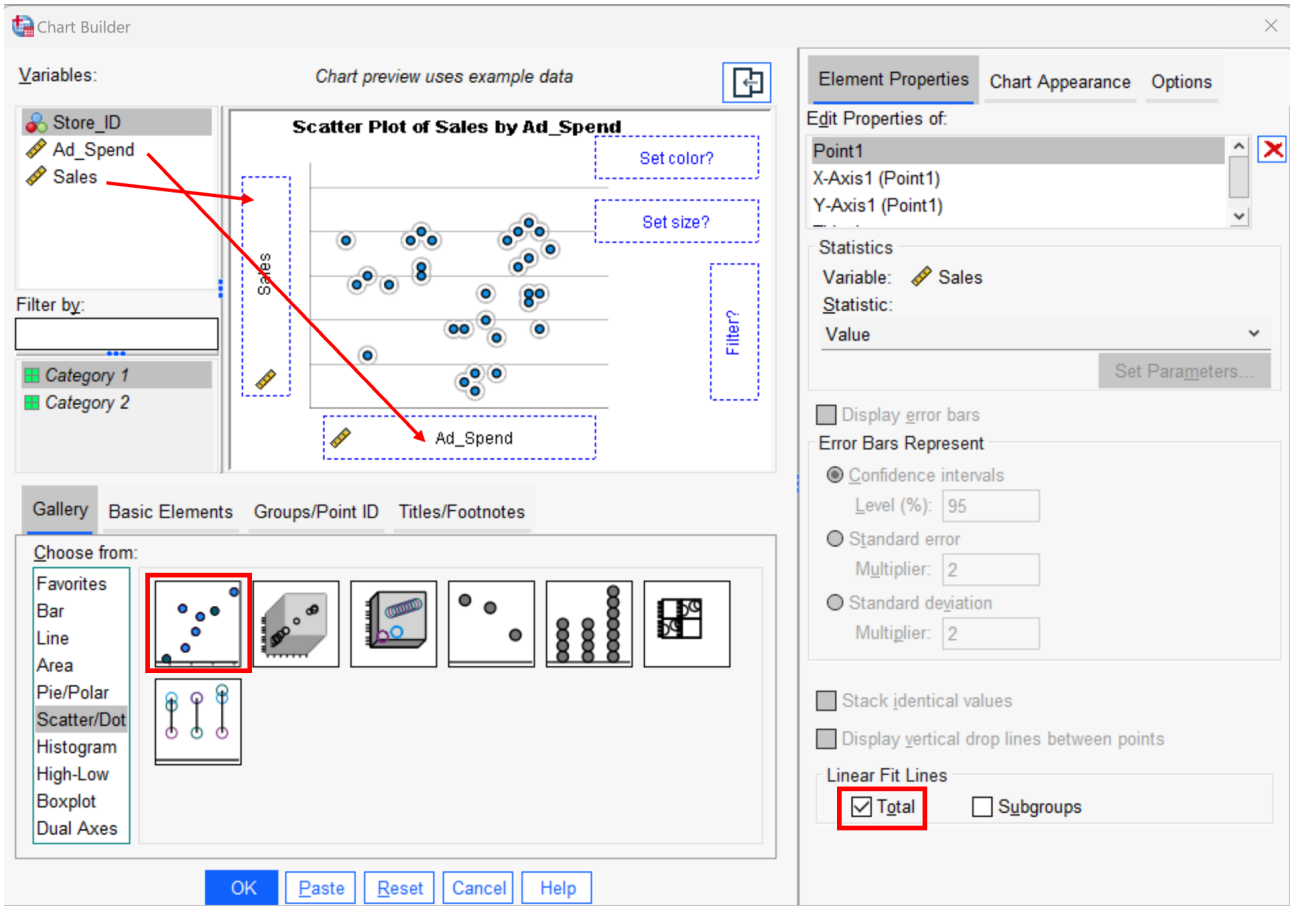
The **Chart Builder** window requires you to specify the chart type and assign your variables to the correct spatial dimensions. Follow these specific steps sequentially to set up the basic scatterplot structure:

In the bottom-left corner, adjust the **Choose from:** list to select **Scatter/Dot** as the desired chart family.
Drag the first scatter plot icon (representing the simple scatterplot) from the gallery and drop it onto the main **Chart Builder area** (the canvas).
From the list of available variables, click and drag the **Ad_Spend** variable into the designated drop zone labeled **X-Axis**.
Finally, drag the **Sales** variable into the drop zone labeled **Y-Axis**.

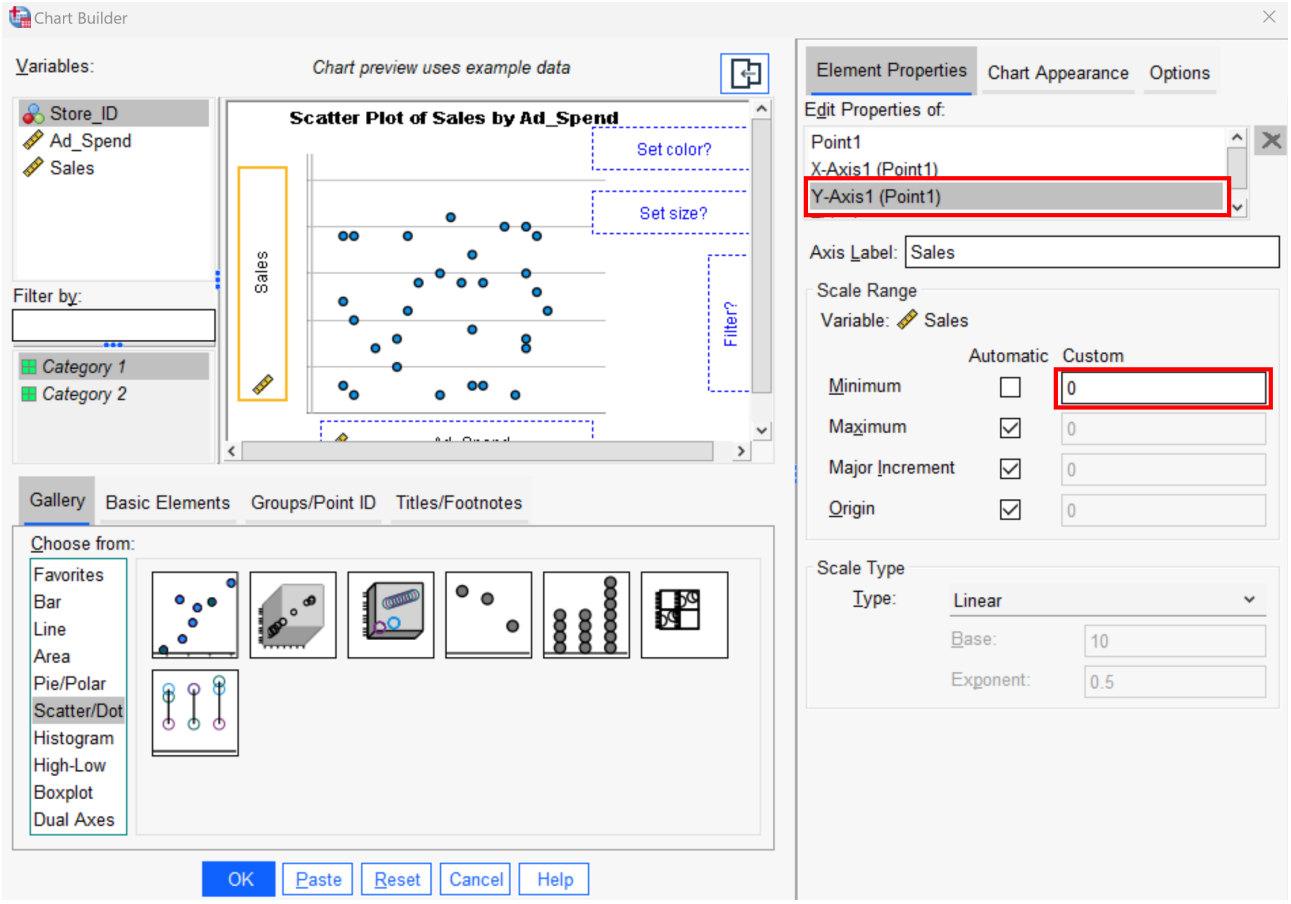
Step 3: Integrating the Linear Regression Fit Line

To transform the basic collection of points into a meaningful visualization for linear regression analysis, we must explicitly instruct **SPSS** to calculate and display the line of best fit. This critical configuration is managed within the **Element Properties** pane of the **Chart Builder**.

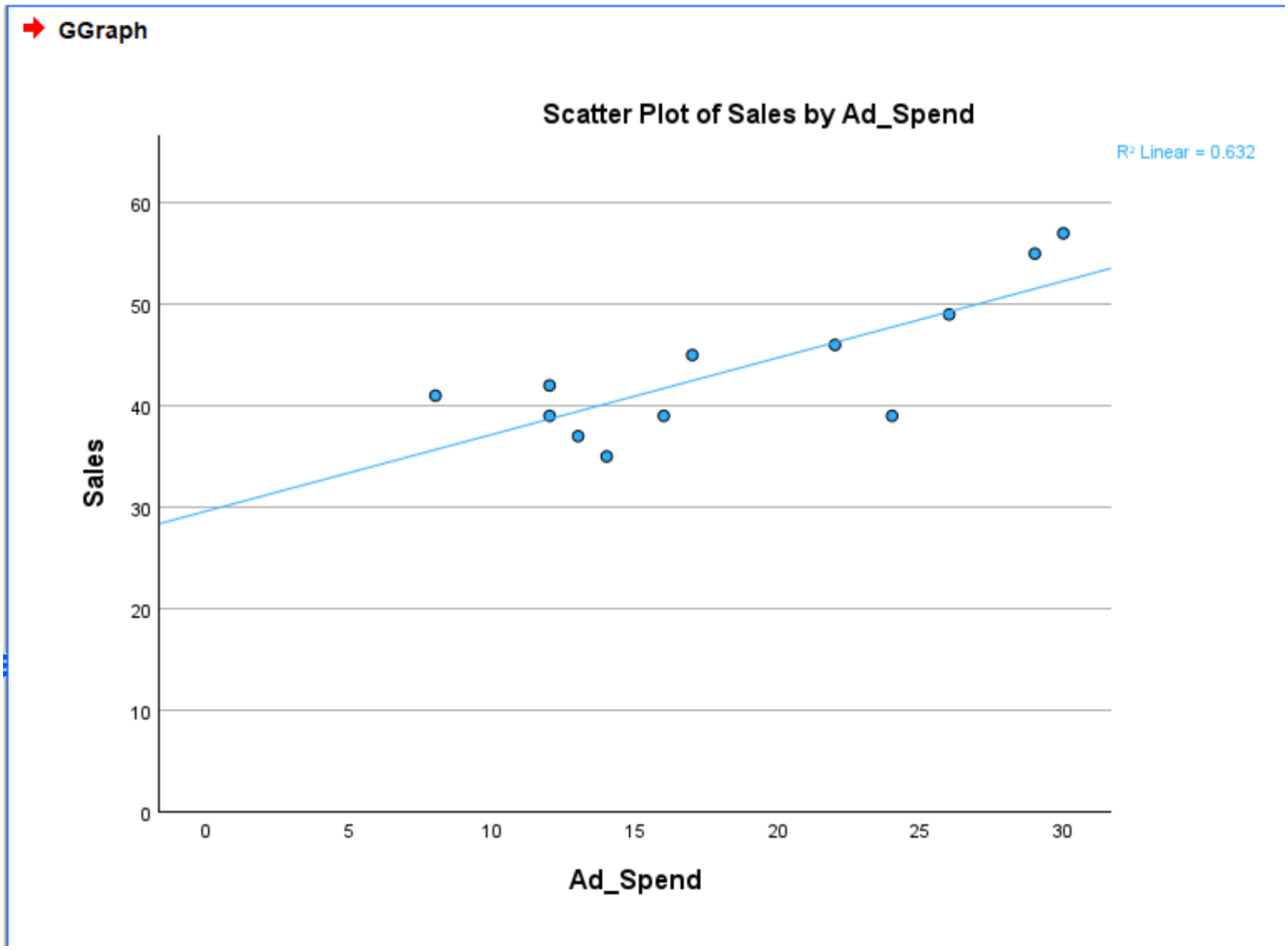
Within the properties panel, locate the section labeled **Linear Fit Lines**. It is essential to check the box associated with the **Total** option. Selecting this instructs the software to compute and draw a linear trend line--which is essentially a simple regression model--that optimally minimizes the squared distances (residuals) between the plotted data points and the line itself.



A best practice for effective data presentation, especially when dealing with non-negative financial metrics like expenditures and sales, is to ensure the axes properly reflect the potential range of data, typically starting at zero. To implement this, click on **Y-Axis1** under the **Edit Properties of** section. In the resulting property panel, manually set the **Minimum** value for the Y-axis range to **0**. This simple adjustment prevents visual distortion or misinterpretation that can arise from a truncated y-axis.



Once these properties are configured, click **OK**. **SPSS** will then generate the initial scatterplot visualization, which now includes the critical linear fit line, displayed in the Output Viewer:



Notice that the output automatically displays the **R² value** (coefficient of determination) in the top-right corner of the chart area. The R² value is a crucial statistic, quantifying the proportion of the variance observed in the dependent variable (Sales) that can be reliably predicted or explained by the independent variable (Ad Spend).

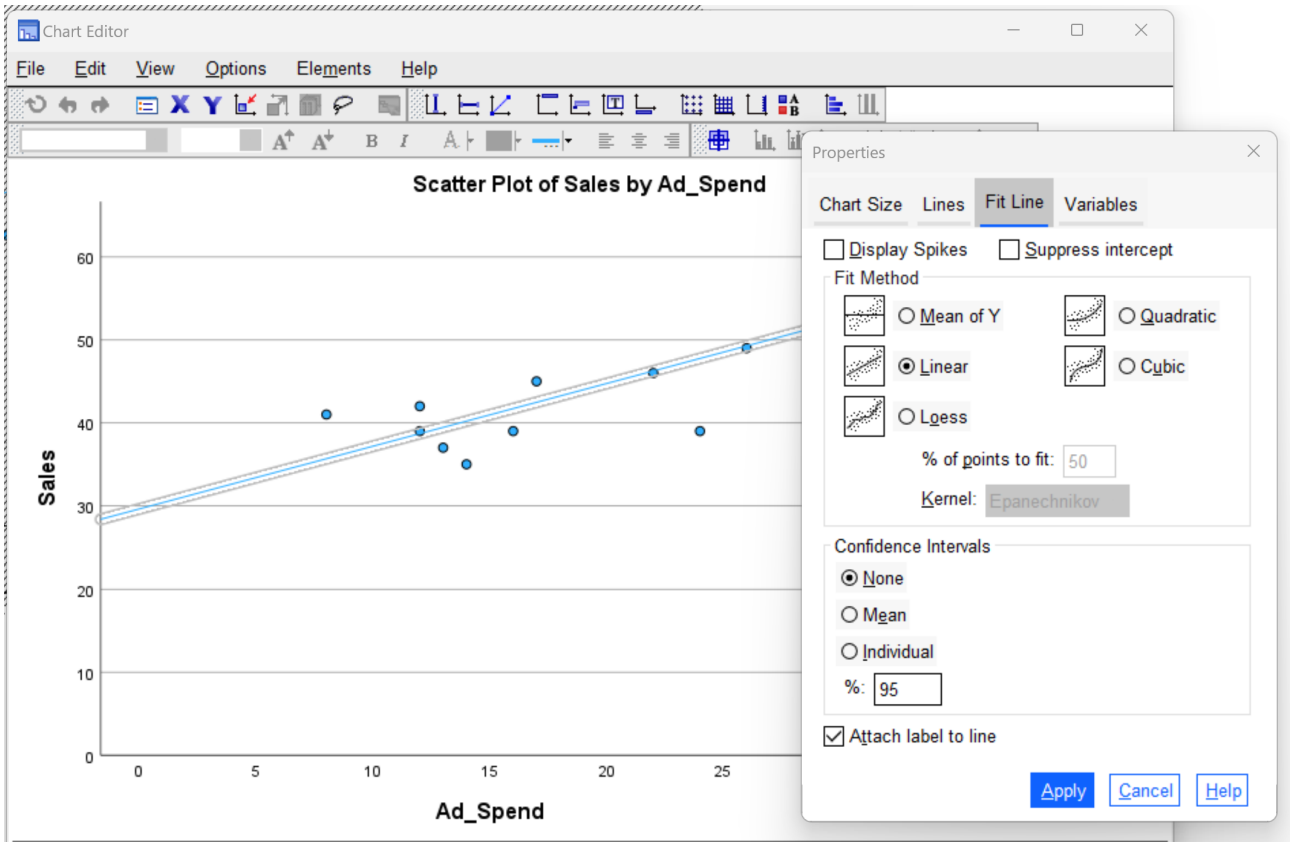
Step 4: Annotating the Plot with the Regression Equation

While the visual trend line is informative, statistical reporting standards frequently require the explicit mathematical equation of the line of best fit. To add this essential annotation, you must access the **Chart Editor**. Launch this specialized editing window by simply double-clicking anywhere on the generated plot within the **Output Viewer**.

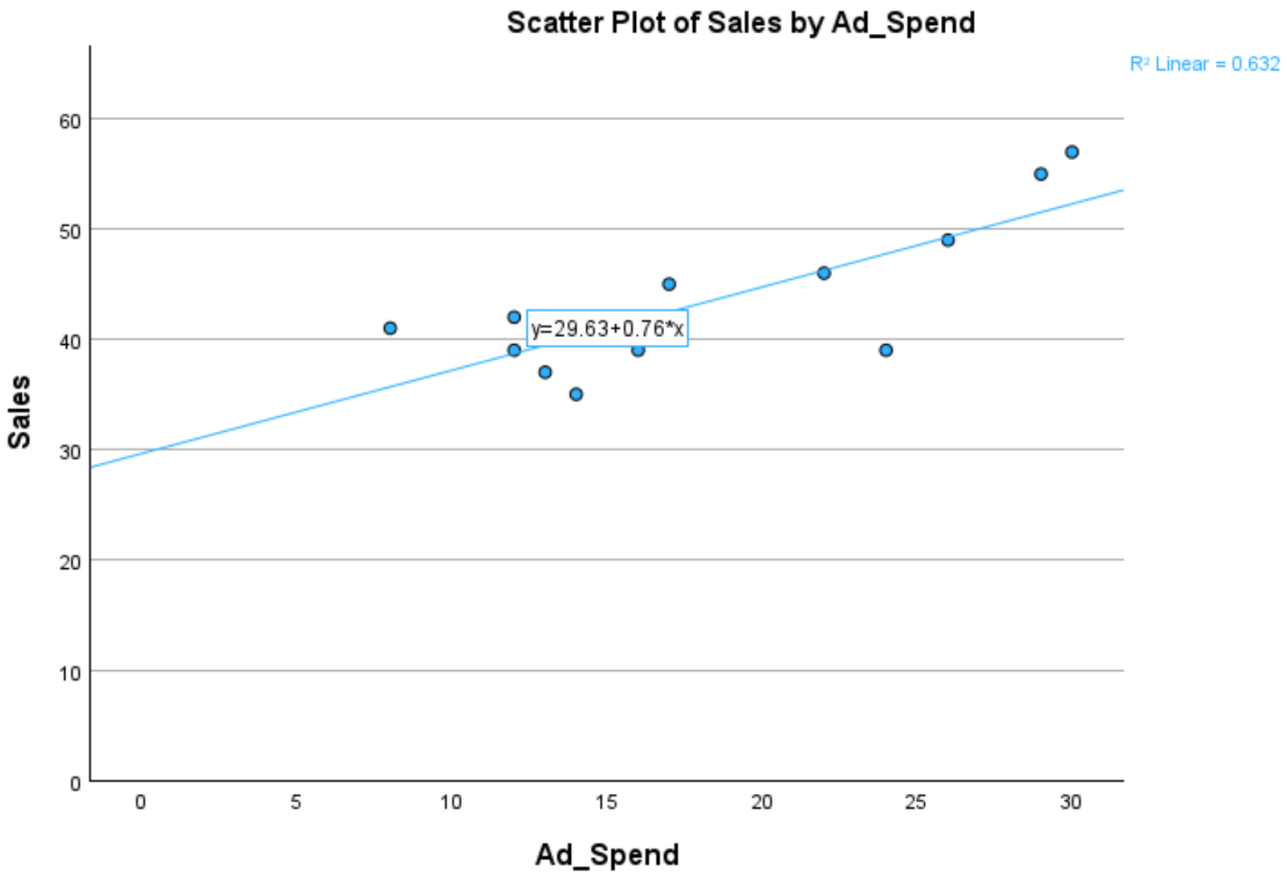
Once the **Chart Editor** is open, the next action is to specifically select the element you intend to modify--which, in this instance, is the regression line itself. Double-click directly on the colored regression line within the chart area. This precise selection will open the **Properties** dialogue box

specifically tailored for the line element.

In the new **Properties** window, navigate to the **Fit Line** tab. This tab contains all the relevant options for the display and calculation of the trend line. To ensure the equation is visible, check the box labeled **Attach label to line**. Enabling this feature instructs **SPSS** to calculate and affix the standard linear equation (Y = b0 + b1*X) directly onto the visualization for easy reference.



Click **Apply**, and then close the **Chart Editor**. The finalized chart, now featuring the regression equation clearly displayed alongside the line, is ready for interpretation and integration into your statistical report. This completes the detailed visualization process:



Step 5: Interpreting the Linear Regression Model Results

The fitted regression model, now clearly annotated on the scatterplot, provides a precise mathematical expression for the relationship observed in the data. The equation displayed is:

Sales = 29.63 + 0.76(Ad Spend)

A thorough understanding of the components of this equation is essential for drawing accurate and meaningful conclusions from the analysis. The equation defines the predicted linear relationship between the predictor variable (Ad Spend) and the response variable (Sales), offering concrete figures that explain the observed trend.

Here is the interpretation of the key coefficients in this regression model:

Intercept (29.63): This value represents the expected baseline sales figure when the ad spend is exactly zero. In a practical context, 29.63 units of sales are projected to occur regardless of any

advertising investment, potentially due to other factors like brand loyalty or location.
Slope (0.76): This coefficient indicates the average change in sales for every one-unit increase in ad spend. Specifically, for each additional dollar invested in advertising, sales are predicted to increase by an average of **0.76** units. This positive slope confirms a direct and positive linear relationship between the two variables, suggesting that increased advertising investment generally leads to increased sales.

Furthermore, evaluating the **R^2 value**, which assesses the goodness of fit of the linear model, is crucial. A higher R^2 value (closer to 1.0) indicates that the regression line provides a substantially better explanation for the variance observed in the Sales figures, making the model more reliable for prediction within the observed data range.

Additional Resources for Advanced SPSS Analysis

Mastering data visualization and regression analysis using **SPSS** forms a critical foundation for quantitative research. For those seeking to deepen their understanding of statistical operations and explore other analytical procedures within the software, the following resources provide guidance on common complex procedures: