International Research Journal of MMC (IRJMMC)
ISSN 2717-4999 (Online) | 2717-4980 (Print)
Volume 6, Issue 1 │ March 2025

Guidelines for Simple Linear Regression Analysis in IBM SPSS: A Step-by-Step Approach

Rabin Koirala[1]

[1]Lecturer of Mathematics, Makawanpur Multiple Campus

Corresponding Author
Rabin Koirala
Email: rabinkoirala52@gmail.com

To Cite this article: Koirala, R. (2025). Guidelines for simple linear regression analysis in IBM SPSS: A step-by-step approach. *International Research Journal of MMC*, *6*(1), 53–67. https://doi.org/10.3126/irjmmc.v6i1.77479

Submitted: 1 March 2025 **Accepted:** 10 March 2025 **Published:** 31 March 2025

Abstract

Simple linear regression is a powerful statistical tool used for scrutinizing the relationship between an independent variable, and a dependent variable. SPSS, which stands for Statistical Package for the Social Sciences, is a software application developed by International Business Machines Corporation (IBM) for a statistical analysis. This article aims to suggest some guidelines for executing a simple linear regression analysis in SPSS. The author went through books, journal articles and relevant website materials to muster required secondary qualitative data for the attainment of the objective. A comprehensive overview of literature on the regression analysis enables the author to conclude data preparation, assumption validation, data entry in SPSS, simple linear regression analysis, interpretation of the results and reporting of findings as guidelines for conducting simple linear regression analysis. This article is significant as it offers a systematic and accessible framework, aiding researchers and analysts in effectively utilizing SPSS for the accurate and efficient analysis of simple linear regression in empirical research studies.

Keywords: simple linear regression, dependent variable, independent variable, SPSS, r-square, model fit

1. Introduction

Regression is a robust statistical modus operandi used to explore the association between an outcome (dependent) variable and one or more predictor (independent) variables. It is a statistical test that examines whether an independent variable can linearly account for any of the variance in a dependent variable (Wrench et al., 2009). A simple linear regression is an influential statistical device employed to examine the connection between a dependent variable and an independent variable. Simple linear regression is a basic but powerful tool in statistics and data analysis, commonly used for tasks such as predicting outcomes, understanding relationships between variables, and making inferences about populations based on sample data. The goal is to find the best-fitting line (linear regression line) that represents the relationship between the two variables.

Regression analysis is a set of techniques used to examine the relationship between two or more variables (Field, 2018). The primary goal of regression analysis is to estimate the conditional expectation of the dependent variable given the independent variables (Hastie et al., 2009). The simplest form of regression analysis is linear regression, which models the relationship between a dependent variable and one or more independent variables using a linear equation.

Regression analysis is a widely used statistical method in various fields, including social sciences, engineering, and business. It is a predictive modeling technique that aims to establish the relationship between a dependent variable and one or more independent variables. Researchers and analysts often use statistical software like SPSS to perform regression analysis. This introduction will discuss the conducive steps of executing regression analysis with SPSS, with in-text citations from authoritative and popular internet sources. SPSS can generate descriptive statistics such as mean, median, mode, standard deviation, and range (IBM, 2021). Visual representations such as histograms and box plots can also be generated using SPSS to gain a better understanding of the data (Pallant, 2020).

Regression analysis is a fundamental tool in various fields such as economics, finance, psychology, and social sciences. It plays a significant role in the field of empirical research studies.

It helps in modeling the relationship between variables by identifying patterns and trends in the data. It allows researchers to understand how changes in one variable affect another, providing insights into the underlying dynamics of the system under study (Montgomery et al., 2012).

One of the key uses of regression analysis is to make predictions based on the relationships observed in the data. By fitting a regression model to the data, researchers can forecast future outcomes or estimate unknown values with a certain degree of confidence (Gujarati & Porter, 2009). It enables researchers to test hypotheses about the relationships between variables. By examining the significance of coefficients and overall model fit, analysts can determine whether there is a statistically significant association between the variables of interest (Kleinbaum et al., 1998). Moreover, it allows researchers to control for confounding variables that may influence the relationship between the dependent and independent variables. By including these control variables in the model, analysts can isolate the effect of specific factors on the outcome of interest (Wooldridge, 2015). Similarly, it provides valuable information for decision-making processes in various fields. Whether it is optimizing business strategies, designing public policies, or understanding consumer behavior, regression analysis helps stakeholders make informed decisions based on empirical evidence (Hair et al., 2010). Predicting economic growth based on factors like investment and inflation (Stock & Watson, 2018). Identifying risk factors for diseases (Hosmer & Lemeshow, 2000). Understanding customer behavior based on demographics and purchasing history (Hair et al., 2019).

Although there are several types of regression, this article is essentially focused on suggesting some guidelines for analyzing simple linear regression in IBM SPSS through exemplification. It has further pointed out the assumptions, and adherence to these assumptions is essential for ensuring that regression analysis provides valid and reliable insights into relationships between variables in a dataset and the robustness of the regression model and the interpretation of its results.

This article provides a comprehensive and practical guideline for conducting simple linear regression analysis using IBM SPSS, offering step-by-step instructions. It is valuable for researchers and analysts seeking a clear and structured approach to performing such analyses within the SPSS software.

2. Literature Review

2.1 Regression

Regression is a resilient statistical method exploited to investigate the relationship between one or more explanatory (independent) variables and a response (dependent) variable. It is a statistical method for investigating and modeling the relationship between two or more variables (Field, 2018). Similar idea concerning the regression analysis is expressed by Hair et al. (2019) that it is a powerful statistical technique used to investigate the relationship between a dependent variable (the variable to be predicted) and one or more independent variables (the variables to influencing the dependent variable). Regression analysis is a powerful tool for examining relationships between variables (Gujarati & Porter, 2009). Regression analysis is widely used in econometrics for modeling economic relationships and making predictions (Stock & Watson, 2019).

It helps us understand how changes in the independent variables lead to changes in the dependent variable, and quantify the strength and direction of these relationships (Montgomery & Runger, 2010). The primary goal of regression analysis is to develop a predictive model that can be used to make accurate predictions about the dependent variable based on the values of the independent variables (Hoaglin et al., 1978). Linear regression assumes that the relationship between the dependent variable and independent variables is linear (Field, 2018). Linear regression assumes a linear relationship between the independent and dependent variables. (Kleinbaum et a., 1988). Assumptions of regression include linearity, independence, homoscedasticity, and normality (Fox, 2015). Multiple regression extends linear regression to analyze the impact of several independent variables on a dependent variable (Hair et al., 2019). Simple linear regression involves one independent variable and one dependent variable (James et al., 2013), whereas multiple linear regression involves two or more independent variables and one dependent variable (Draper & Smith, 1998). Robust regression methods are employed to handle outliers and influential observations in the data (Hampel et al., 2011).

2.2 Simple Linear Regression

A simple linear regression analysis involves an independent variable, and a dependent variable.

Assumptions of Simple Linear Regression

It has several assumptions that must be met for accurate results. Violating these assumptions can lead to incorrect conclusions and inaccurate predictions. These assumptions include:

Selection of Variable: There should be an independent variable and a dependent variable. The selection of independent variables is critical for building a parsimonious and meaningful model (Hair et al., 2019). Researchers should carefully consider theoretical justification, existing literature, and potential multicollinearity (correlation among independent variables) (Menard, 2020). Techniques like correlation analysis and preliminary regression models can aid in identifying relevant and independent predictors (Field, 2013). One critical step in regression analysis is the thoughtful selection of variables. As highlighted by Johnson and Wichern (2007),

Measurement Level: Dependent and independent variables should be on continuous scale. They should be either interval or ratio. SPSS regards it as Scale. Interval scale data is a type of measurement scale in which the intervals between values are equal and meaningful, but there is no true zero point. For examples,

Temperature (in Celsius or Fahrenheit): The difference between 10°C and 20°C is the same as the difference between 20°C and 30°C. However, 0°C does not represent the absence of temperature.

IQ Scores: Intelligence Quotient (IQ) scores are measured on an interval scale. The difference between an IQ score of 90 and 100 is the same as the difference between 110 and 120. However, an IQ of 0 does not mean the absence of intelligence.

SAT Scores: SAT scores are another example of interval scale data. The difference between a score of 1200 and 1300 is the same as the difference between 1400 and 1500, but a score of 0 does not indicate a complete absence of academic ability.

Distance on a Map: When measuring distances on a map, the intervals are equal. The difference between 1 inch and 2 inches on a map is the same as the difference between 3 inches and 4 inches. However, a point with zero inches does not mean the absence of distance.

Time (in Standard Clock Time): Standard clock time is measured on an interval scale. The difference between 1:00 PM and 2:00 PM is the same as the difference between 8:00 PM and 9:00 PM. However, midnight (12:00 AM) does not represent the absence of time.
Ratio scale data is a type of measurement with a true zero point, and ratios of values are meaningful. For examples,
Age: 25 years, 5 Months, 3 weeks, 7 days
Income: Rs. 50,000, $ 500, £ 525 (Per Month, Week, Year)
Weight: 40 kg, 500 grams, 30 ounces, 10 pounds, 2 litres, etc.
Mark Score: 40, 50, 35, 20, 12.5, 12.25
Height: 4 feet, 5 feet, 7 feet

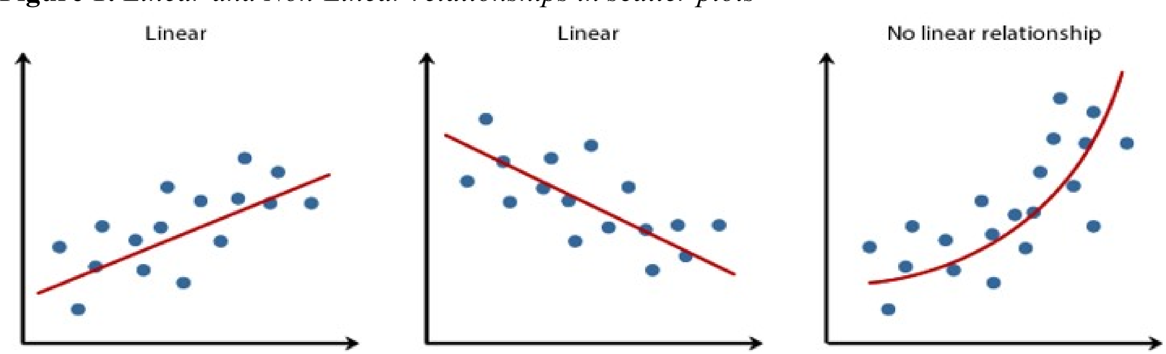
Distance: 5km, 7 Km, 10 km, 4.5 miles…
Length: 5 meters, 5.5 meters, 25 centimeters…..
Number of boys in the classroom: One, Two, Three, Four…
Number of girls in the classroom: Ten, Fifteen, Twenty…
Number of schools in the districts: 25, 55, 68. 104…
Percent: 25.00 %, 34.45%, 78.255%, 0.05% …
Mean: 4, 4.5, 5.75, 8.257…

Reading period per day: 2 hours, 4 hours and 30 minutes, 20 minutes and 30 seconds…
Let's explain: Suppose there are two sticks. One stick is 180 cm long and another stick is 90 cm long. Length is measured on a ratio scale because it has a true zero point (absolute absence of length), and the ratio of 180 cm to 90 cm is meaningful as it indicates that one stick is twice as long as another one.

Linear Relationship: There should be a linear relationship between the dependent variable and each independent variable. It means there must be
between the two variables.

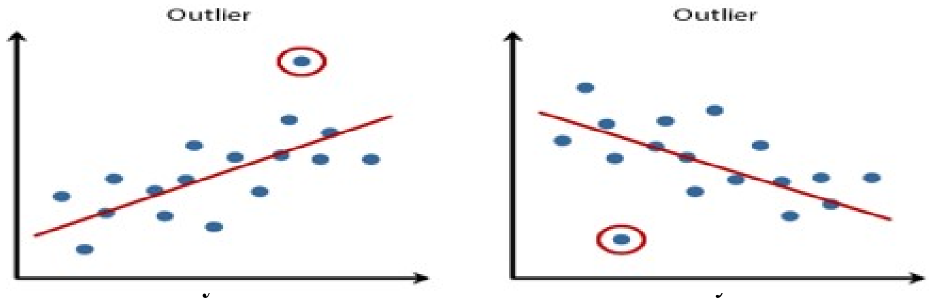
The relationship between the independent and dependent variables should be linear. This means a straight line can adequately capture the association between them (Kutner et al., 2004). The relationship between the independent and dependent variables should be linear (Hair et al., 2019). Though there are a number of ways to check whether a linear relationship exists between your two variables, we can create a scatterplot using SPSS Statistics where we can plot the dependent variable against your independent variable and then visually inspect the scatterplot to check for linearity. Our scatterplot may look something like one of the following:

Figure 1: *Linear and Non-Linear relationships in scatter plots*



No Spurious Outliers: There should be no significant outliers. An outlier is an observed data point that has a dependent variable value that is very different to the value predicted by the regression equation. As such, an outlier will be a point on a scatterplot that is (vertically) far away from the regression line indicating that it has a large residual, as highlighted below:

Figure 2: *Outliers in Linear Regression*

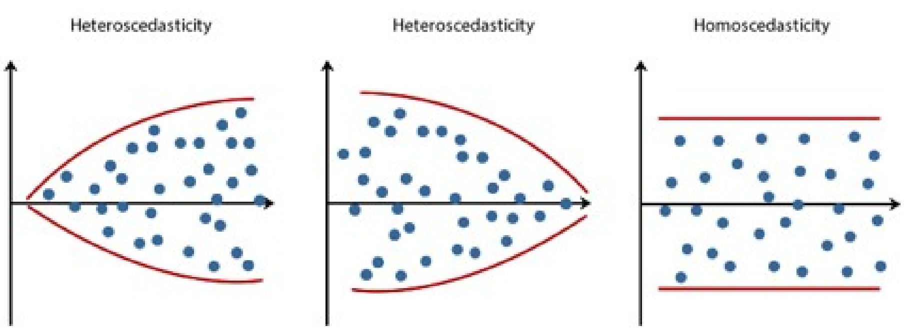


ons.
servations

umes that the variance of errors is constant across all levels of independent variables (Field, 2018). The variance of the errors should be constant across all levels of the independent variables. In simpler terms, the "spread" of the errors should be consistent regardless of the independent variable's value (Gujarati, & Porter, 2009). The variance of the residuals should be constant across all levels of the independent variable (Montgomery & Runger, 2010).

Homoscedasticity refers to the assumption in statistics that the variance of the errors (residuals) in a regression model is constant across all levels of the independent variable(s). In simpler terms, it means that the spread of the residuals is consistent across the range of predicted values. The three scatterplots below provide three simple examples: two of data that fail the assumption (called heteroscedasticity) and one of data that meets this assumption (called homoscedasticity):

Figure 3: *Homoscedasticity and Heteroscedasticity in Regression Residuals*



Normality: Normality assumes that errors are normally distributed with mean zero (Field, 2018). The residuals should be normally distributed (Kutner et al., 2005).The **residuals (errors)** of the regression line are **approximately normally distributed.** Normality of data refers to the distribution of values in a dataset and the extent to which those values follow a normal (Gaussian) distribution or bell-shaped curve. A normal distribution is characterized by certain statistical properties, such as a symmetric shape with a peak at the mean and a specific pattern of spread around the mean.
Key characteristics of a normal distribution include:

Symmetry: The distribution is symmetric, meaning that the left and right sides of the curve are mirror images of each other.

Bell-shaped Curve: The distribution has a single peak, forming a bell-shaped curve when plotted.

Mean, Median, and Mode Equality: The mean, median, and mode of the distribution are equal and located at the center of the curve.

Test: Shapiro-Wilk's test can be used to test the normality of the data. The *p- value is* greater than .05 indicates the normality. It can be executed by following steps:

SPSS: Analyze → Descriptive Statistics → Explore.
Normality of Data: Visual inspection of their histograms, normal Q-Q plots and box plot.

Figure 4: *Q-Q plots and box plot*

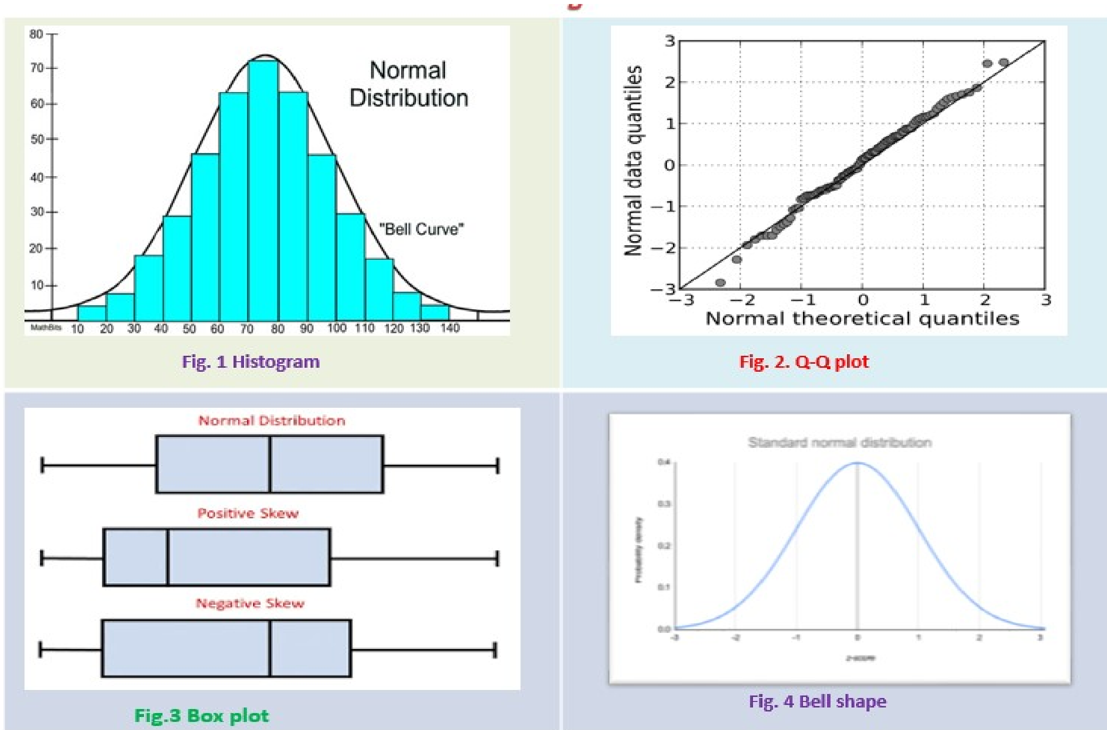


Fig. 1 Histogram
Fig. 2. Q-Q plot
Fig.3 Box plot
Fig. 4 Bell shape

No Perfect Multicollinearity: The independent variables should not be highly correlated with each other. Multicollinearity can cause unstable estimates and difficulty interpreting the model's coefficients. No multicollinearity assumes that there is no correlation between independent variables (Field, 2018). Independent variables should not be highly correlated, as it can affect the stability and interpretation of coefficients (Belsley et al., 1980). It is checked only in a linear multiple regression analysis, not in a simple linear regression. The following table can be used for interpreting the multicollinearity. **SPSS:** Analyze → Regression→Linear→Statistics → Collinearity Diagnostics

Table 1: *Variance Inflation Factor (VIF)*

Variance Inflation Factor (VIF)	Interpretation
VIF = 1	No multicollinearity between the independent variables
VIF > 1 and < 5	Moderate multicollinearity
VIF > = 5	High multicollinearity among the independent variables.
VIF > = 10	Very high multicollinearity

2.3 Steps or Guidelines for Simple Linear Regression Analysis in IBM SPSS

Data Preparation: Data preparation involves cleaning, transforming, and recoding the data to ensure its accuracy and suitability for the analysis (Field, 2018). The initial step involves ensuring data quality and preparing it for analysis (Pallant, 2020). This includes checking for missing values, outliers, and inconsistencies (Hair et al., 2019). Missing data can be handled through various methods, such as mean imputation, list wise deletion, or model-based techniques (Enders, 2010). Tabachnick and Fidell (2019) emphasize the significance of data cleaning and transformation for accurate results. SPSS offers various data preparation tools, such as computing new variables, recoding variables, and transforming variables (IBM, 2021). It is crucial to ensure that the data is free from missing values, outliers, and multicollinearity (Field, 2018). SPSS has become the preferred choice for many researchers and professionals due to its user-friendly interface, extensive functionality, and compatibility with various data formats (Tabachnick & Fidell, 2013). SPSS provides various tools to assist in data preparation, such as the Missing Values and Descriptive Statistics procedures (Tabachnick & Fidell, 2013).

- Ensure that the data are clean and organized
- Identify an independent variable a dependent variable.
- Ensure that both variables are on continuous scales.
- Check for missing values

Check the Assumptions

- Linear Relationship
- No Spurious Outliers: Outliers, however, require careful examination to determine their legitimacy and potential impact on the analysis (Field, 2013). Additionally, data transformation may be necessary to meet the assumptions of normality and linearity required for regression analysis (Tabachnick & Fidell, 2020).
- Independence of Observations**:**
- Homoscedasticity
- Normality**.**
- No Perfect Multicollinearity

Run the Simple Linear Regression Analysis:

- Navigate to Analyze > Regression > Linear.
- Enter Variables:
    - In the Dependent: box, insert the dependent variable.
    - In the Independent(s): box, insert the independent variable.

- Run the Analysis: Click OK.

Interpret the Results: The final step involves interpreting the results within the context of the research question and existing literature (Field, 2013). Researchers should present the regression coefficients, their significance levels, and confidence intervals to communicate the direction and magnitude of the relationships between variables (Arbuckle, 2023).
The output will include:

Assess Model Fit: Evaluate the overall fit of the model using R-squared, adjusted R-squared, and other goodness-of-fit measures.

Examine Coefficients: Interpret the coefficients for each predictor variable.

Regression Coefficients: Look for the coefficient of the predictor variable (slope) and the constant term (intercept).

R-squared (R^2): Indicates the proportion of variance in the response variable explained by the predictor. The **R-squared** represents the proportion of variance in the response variable explained by the predictor variables. It ranges from 0 to 1. This statistic represents the proportion of variance in the dependent variable explained by the independent variables (Cohen et al., 2013).

Adjusted R-Squared: The adjusted R-squared accounts for the number of predictors in the model. It is always lower than the R-squared. This penalizes R-squared for the number of independent variables, providing a more accurate estimate of explanatory power (Montgomery & Runger, 2010).

Residuals: These are the differences between the observed values of the dependent variable and the predicted values from the regression equation. Analyzing them helps assess model fit (Fox, 2016).

Standard Error of Estimate: Measures the average deviation of actual scores from the regression line.

Significance Levels: Check if the coefficients are statistically significant.

p-values: A p-value less than 0.05 suggests the corresponding coefficient is statistically significant, meaning the relationship between that specific independent and dependent variable is unlikely due to chance alone.

Confidence Intervals: Assess the range within which the true coefficients lie.

Report Findings:
- Present your results, including coefficients, p-values, and confidence intervals.

3. Materials and Methods

This article has been based on the secondary qualitative data for identifying the theoretical aspects of regression. To conduct regression analysis using SPSS, first, quantitative dummy data have been prepared and variables have been identified. The students' study hours per day has been considered the independent variable, and their mark scores have been taken as the dependent variable. In SPSS, navigate to "Analyze" and select "Regression." Choose the type of regression (e.g., linear) and input the dependent and independent variables. Check assumptions like linearity, normality, and homoscedasticity. Interpret the results, including coefficients and significance levels. Report findings have accurately been done. This process involves data preparation, selecting regression type, checking assumptions, interpreting results, addressing multicollinearity, and reporting outcomes effectively within the SPSS software.

4. Analysis and Interpretation of Data

Step 1: Identification of Variables

The author has produced the dummy data for conducting a simple linear regression analysis. The mark scores of 30 students in their first term examinations have been taken as a dependent variable (Y) and their daily reading hours as an independent variable (X). Both variables have been measured on the ratio scale.

Table 2: *Identification of an Independent Variable (X) and a Dependent Variable (Y)*

Code No. of Students	Reading Hours Per Day (X)	Mark Scores in English (Y)		Code No. of Students	Reading Hours Per Day (X)	Mark Scores in English (Y)
01	4	12		16	6	16
02	5	13		17	3	8
03	2	7		18	2	7
04	4	9		19	3	10
05	3	6		20	4	13
06	2	5		21	5	17
07	5	15		22	4	13
08	4	13		23	5	18
09	3	12		24	4	13
10	1	10		25	6	18
11	4	14		26	2	8
12	5	17		27	3	9
13	2	5		28	2	8
14	3	7		29	3	12
15	4	12		30	4	15

Step 2: Checking the Assumption of Simple Linear Regression

The visual examination of histograms, Q-Q plots, Box plots, and the Shapiro-Wilk's test (with $p > .05$) in SPSS indicate that the data satisfy the assumptions for the simple linear regression.

Step 3: Running the Simple Linear Regression Analysis:

- Navigate to Analyze > Regression > Linear.
- Enter Variables:

    - In the Dependent: box, insert the dependent variable (Mark Scores of Students)
    - In the Independent(s): box, insert the independent variable (Reading Hours Per Day)
- Run the Analysis: Click OK.

Step 4: Interpreting the Results

After running SPSS, we can see primarily 4 tables entitled: Variables Entered / Removed, Model Summary, ANOVA, and Coefficients. Such as:

Table 3: *Variables Entered/Removed* [a]

Model	Variables Entered	Variables Removed	Method
1	Reading Hoursb	.	Enter

a. Dependent Variable: Marks in English
b. All requested variables entered.

This table shows that the reading hours as an independent variable and marks in English as a dependent variable were entered in a single step and were given equal importance.

Table 4: *Model Summary*

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.940[a]	.883	.879	1.170

a. Predictors: (Constant), Reading Hours Per Day

Correlation coefficient, (R) = .940 indicates that there was a positive high degree of correlation between reading hours and marks secured in English. R-Square (.883) value indicates that 88.3 percent variation in the mark scores was due to the reading hours, and the remaining 11.7 percent was due to other reasons**.** Standard error of the estimate refers to how accurate the prediction around the regression line is. The value 1.170 lay between -2 and +2. The smaller the value of the standard error of the estimate, the better the fit of the regression model to the data. Adjusted R Square is used when there are two or more independent variables.

Table 5: *ANOVA*

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	288.670	1	288.670	211.057	.000[b]
	Residual	38.297	28	1.368		
	Total	326.967	29			

a. Dependent Variable: Mark Scores in English
b. Predictors: (Constant), Reading Hours Per Day

The regression sum of square (288.670) was greater than the residual sum of square (38.297). It indicates the variability in the dependent variable that was explained by the independent variable included in the regression model was effective. The total sum of squares

326.967 represents the total variability in the dependent variable. Similarly, the mean square regression (288.670) that was higher than a mean square residual / error (1.368) suggests that a larger proportion of the total variation in the dependent variable was explained by the regression model. It was further justified by the F-statistic. The observed f (1, 28) = 211.057 was higher than a critical value = 4.20. It shows that the model had a significant effect on explaining the variation in the dependent variable. Similarly, If Sig. was smaller than .05, it implies that the independent variable had a significant effect on the dependent variable after accounting for other variables in the model.

Table 6: *Coefficients* [a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	3.612	.466		7.748	.000
	Reading Hours Per Day	1.672	.115	.940	14.528	.000

a. Dependent Variable: Marks in English

Unstandardized B (Constant) tells us that the average exam score was 3.612 when the study hour of the students was equal to zero**.** Unstandardized B (reading hours) indicates that each additional hour studied was associated with an increase of 1.672 in exam score, on average. The standard error of the unstandardized Coefficients was low. It indicates more precise estimates. A standardized coefficient beta of .940 indicates that there was a positive relationship between the independent variable and the dependent variable. The t-value (14.528) was greater than the standard error (.115). It indicates that the corresponding independent variable had a stronger impact on the dependent variable. Sig .000 indicates that the reading hours had a significant impact on the mark score after accounting for other variables in the model.

Using Formula

$Y = a + bX + e$

Where,

Y = Predicted value of the dependent variable (Mark Scores)

b= Slope of the regression line = 1.672 (Coefficient of X)

X= Value of the independent variable (Reading Hour)

a = Y-intercept of the regression line (constant) = 3.612

e = error terms

Putting the value in the regression equation:

$Y = 3.612 + 1.672\,X + e$

Impact: If students increase their daily reading hours by one hour, the model suggests their mark scores are expected to increase by 1.672 units.

Some conclusions we can draw are:

Intercept (3.612): When X is zero, Y is expected to be 3.612. However, depending on the context, this value may or may not have a meaningful interpretation.

Slope (1.672): For every one-unit increase in X, Y is expected to increase by 1.672 units. The sign of the slope (positive in this case) indicates the direction of the relationship between X and Y.

If the coefficient of X is negative, it indicates that there is an inverse relationship between the independent variable (X) and the dependent variable. In other words, as the value of X increases, the predicted value of the dependent variable decreases, and vice versa.

If the intercept is negative, it simply means that the predicted value of the dependent variable is negative when all independent variables are zero.

Error Term (e): The error term represents unobserved factors or random variability in the relationship between X and Y. It is assumed to be normally distributed with a mean of zero.

Overall Model Fit: The regression equation by itself does not tell us about the overall fit of the model. To assess the model's goodness of fit, you would typically look at statistical measures like R-squared, p-values, and residuals.

Assumptions: The regression analysis assumes that there is a linear relationship between X and Y, and the residuals are normally distributed. It's essential to check these assumptions for the validity of your result.

Value: R-Square value greater than 0.5 shows that the model was effective enough to determine the relationship.

No Interpretation: Do not interpret if the Y-Intercept (a) has a value in minus**.** The mark scores do not occur in minus.

5. Conclusion

This article emphasizes the importance of following systematic and structured steps while conducting a simple linear regression analysis using IBM SPSS. It highlights the significance of understanding the assumptions of linear regression, checking for outliers, assessing multicollinearity, and interpreting the results accurately. By following these guidelines, we can assess the relationship between two continuous variables, apprehend the impact of the independent variable on the dependent variable, and evaluate the overall model fit. The R-Square value of .883 suggests that 88.3 percent of the variation in mark scores was attributed to reading hours, with the remaining 11.7 percent being influenced by other factors. A higher R-Square value, closer to 1, indicates a better fit of the model to the data, suggesting that a larger percentage of the variation in the dependent variable could be explained by the independent variable. In other words, An R-Square value exceeding 0.5 indicates the model's effectiveness in discerning the relationship between variables. According to the model, an increase of one hour in daily reading hours was associated with an expected increase of 1.672

units in mark scores of the students. The study underscores the need for researchers to carefully consider the limitations and challenges associated with linear regression analysis to ensure the validity and reliability of their findings. By providing a detailed step-by-step guide, it empowers researchers to effectively utilize IBM SPSS for a simple linear regression analysis, ultimately enhancing the quality and rigor of their research outcomes. This foundational knowledge paves the way for exploring more complex regression techniques in future analyses.

References

1. Arbuckle, J. L. (2023). *IBM SPSS statistics basics: A guide to learning and doing statistical analysis*. Sage Publications.
2. Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity.* John Wiley & Sons.
3. Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge.
4. Draper, N. R., & Smith, H. (1998). *Applied regression analysis (*3rd ed.). John Wiley & Sons.
5. Enders, C. K. (2010). *Applied missing data analysis*. Guilford Publications.
6. Field, A.P. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). Newbury Park: Sage.
7. Fox, J. (2015). *Applied regression analysis and generalized linear models*. Sage Publications.
8. Fox, J. (2016). An R companion to applied regression (3rd ed.). Sage Publications.
9. Green, R., & Jones, P. (2017). Regression analysis of climate change impact on agricultural yields. *Environmental Science and Technology, 22*(1), 45-62.
10. Gujarati, D. N., & Porter, D. C. (2009). *Basic econometrics* (5th ed.). McGraw-Hill.
11. Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.
12. Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (2011). *Robust statistics: The approach based on influence functions*. John Wiley & Sons.
13. Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (2nd ed.). Springer Series in Statistics.
14. Hoaglin, D., & Welsch, R. (1978). The hat matrix in regression and ANOVA. *American Statistician, 32*, 17–22.
15. Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). John Wiley & Sons.
16. IBM Corp. (2021). *IBM SPSS Statistics 28.0.0 documentation*. https://www.ibm.com/support/pages/ibm-spss-statistics-28-documentation
17. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning with applications*. Springer.
18. Johnson, R.A., & Wichern, D.W. (2007) *Applied multivariate statistical analysis (*6th ed.). Pearson Prentice Hall, Upper Saddle River.
19. Kleinbaum, D. G., Kupper, L. L., & Muller, K. E. (1988). *Applied regression analysis and other multivariable methods.* Duxbury Press.

20. Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2004). *Applied linear regression models* (5th ed.). McGraw-Hill.
21. Menard, S. (2020). *Applied logistic regression*. Sage Publications.
22. Montgomery, D. C., & Runger, G. C. (2010). *Applied statistics and probability for engineers and scientists* (5th ed.).
23. Montgomery, D.C., Peck, E.A., & Vining, G.G. (2012). *Introduction to linear regression analysis* (5 th ed.). John Wiley & Sons.
24. Pallant, J. (2020). *SPSS statistics for social science and research methods*. Routledge.
25. Smith, J., & Johnson, A. (2018). The impact of inflation on consumer spending: A regression analysis. *Journal of Economic Studies, 45*(2), 213-230.
26. Stock, J. H., & Watson, M. W. (2019). *Introduction to econometrics* (4th ed.). Pearson.
27. Tabachnick, B. G., & Fidell, L. S. (2013). *Practical approach to using multivariate analyses*. Pearson Education Limited.
28. Tabachnick, B. G., & Fidell, L. S. (2021). *Using multivariate statistics*. Pearson Education Limited.
29. Tabachnick, B. G., & Fidell, L. S. (2019). Using Multivariate Statistics (7th ed.). Pearson.
30. Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach* (5th ed.). South-Western Cengage Learning.
31. Wrench, J.S, Maddox, C.T., Richmond, V.P., & Mccroskey, J.C. (2009). *Quantitative research methods for communication*. New York: Oxford University Press.