REGRESI LINIER

Aryan Eka Prastya Nugraha

2018

Regression and correlation are closely related. Both techniques involve the relationship between two variables, and they both use the same set of paired scores taken from the same subjects

There are many reasons why researchers want to predict one variable from another. For example, knowing a person’s IQ, what can we say about this person’s prospects of successfully completing a university course?

Knowing a person’s prior voting record, can we make any informed guesses concerning his vote in the coming election? Knowing his mathematics aptitude score, can we estimate the quality of his performance in a course in statistics?

These questions involve predictions from one variable to another, and psychologists, educators, biologists, sociologists, and economists are constantly being called upon to perform this function.

Requirements

- For each subject in the study, there must be *related pairs of scores*. That is, if a subject has a score on variable *X*, then the same subject must also have a score on variable *Y*.

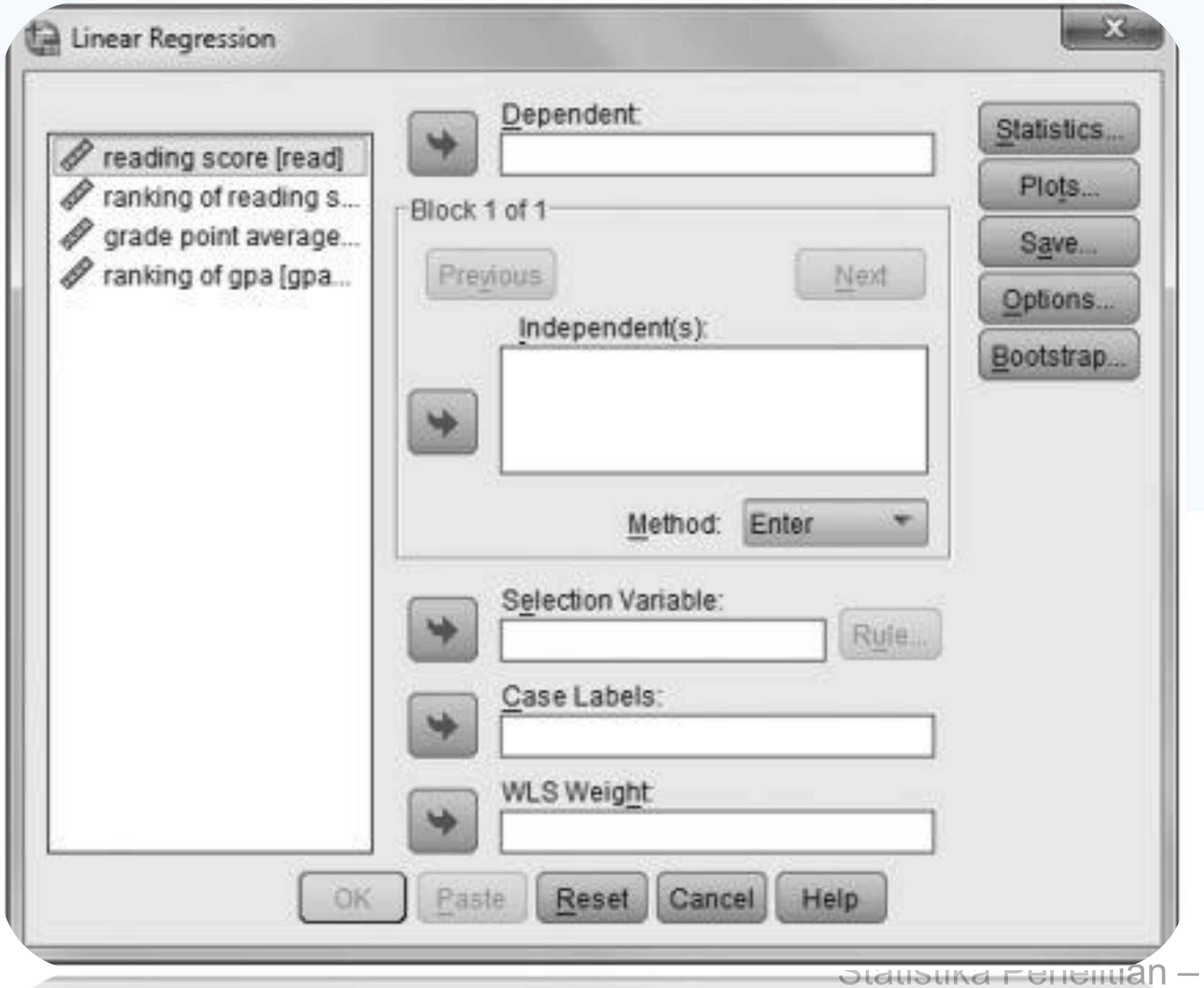
- The variables should be measured at least at the *ordinal level*.

Assumptions

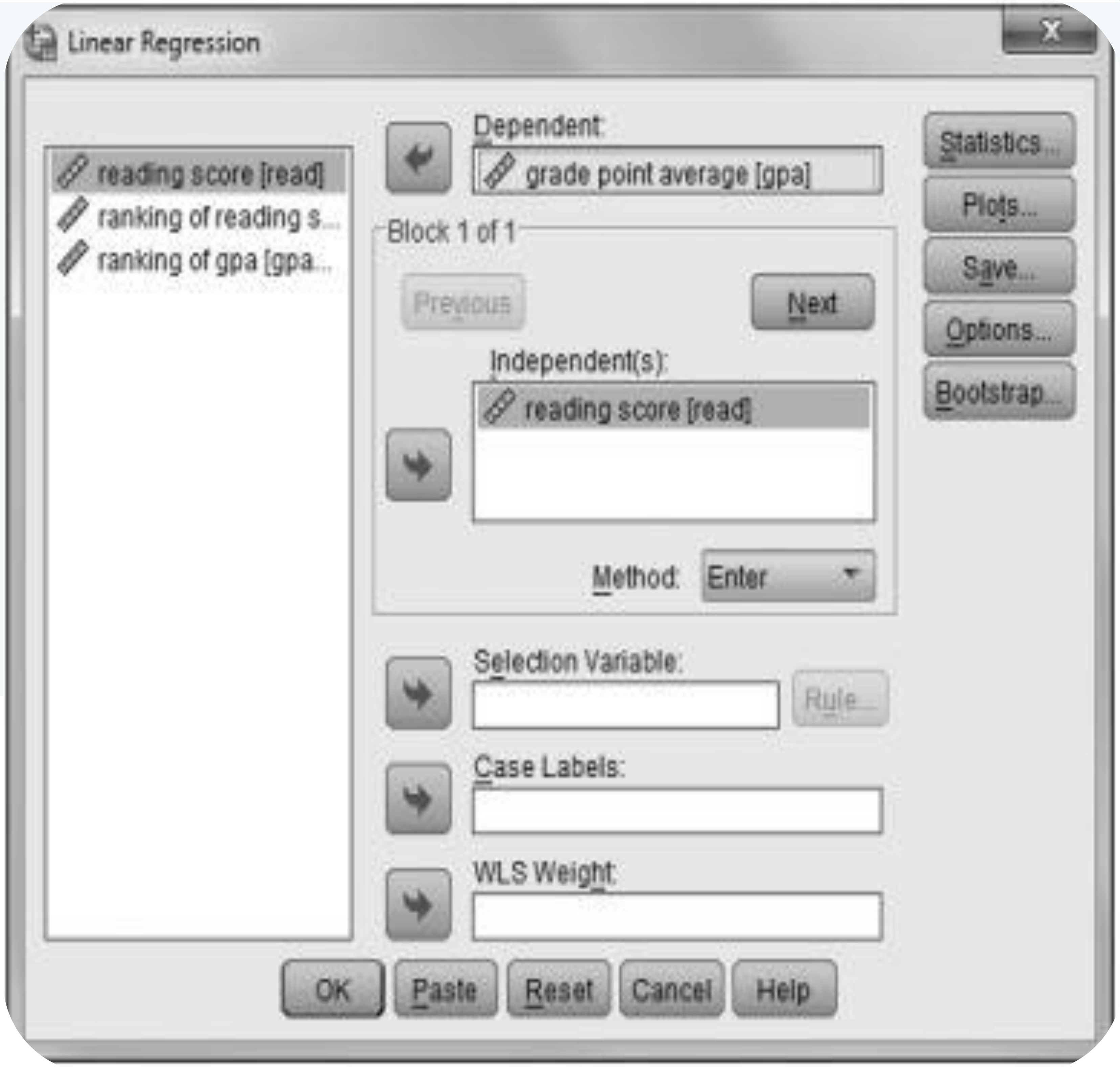
- **Linearity**—The relationship between the two variables must be *linear*, that is, the relationship can be more accurately represented by a straight line.
- **Homoscedasticity**—The variability of scores on the *Y* variable should remain constant at all values of the *X* variable.

This example employs the data set **CORR.SAV**
In this example, we wish to (1) fnd the prediction equation that allows us to best predict students’ grade point average scores (**GPA**) from their reading scores (**READ**), (2) determine the strength of this relationship, and (3) test the null hypothesis that **READ** and **GPA** scores are unrelated.

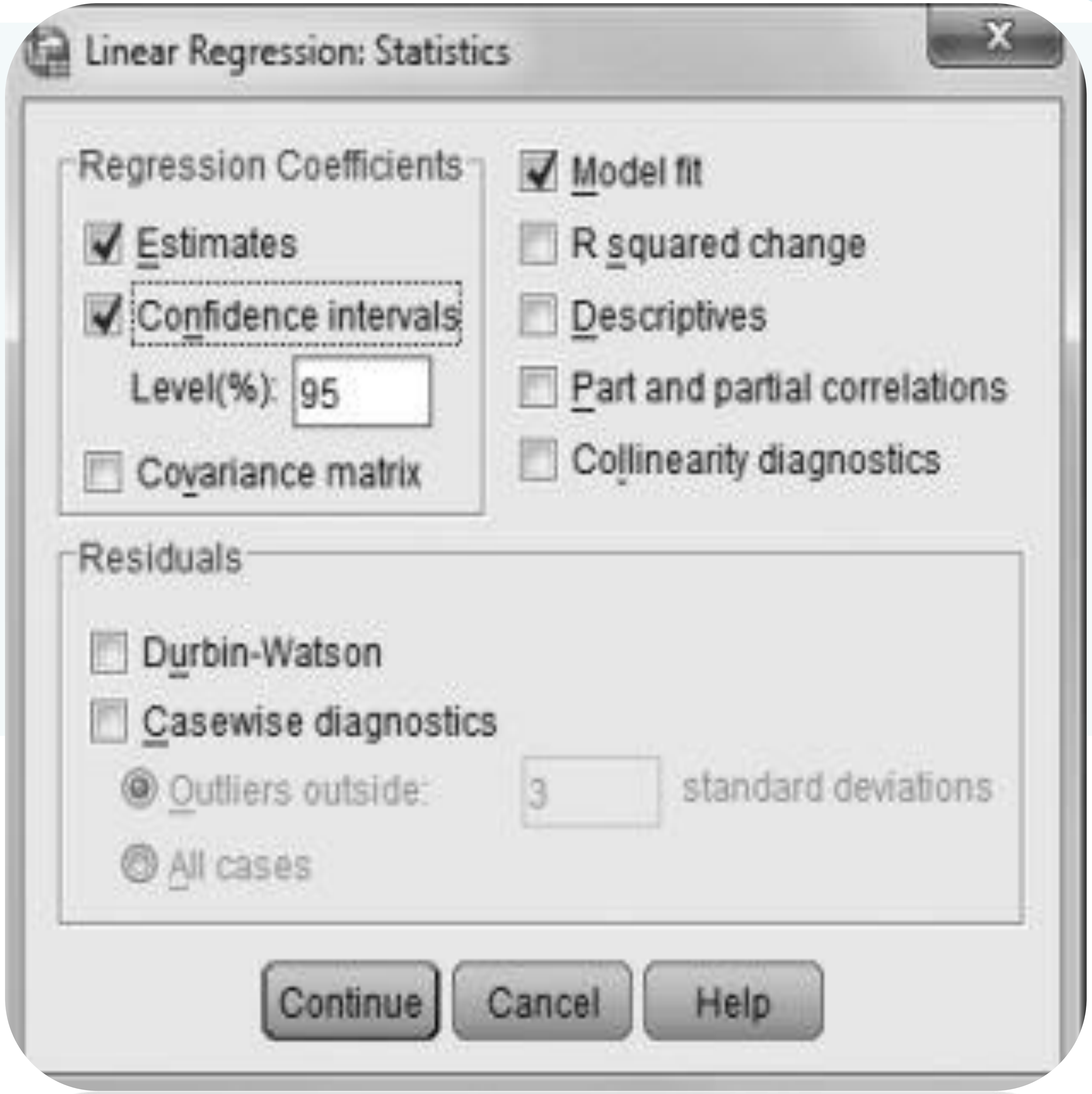
From the menu bar, click **Analyze**, then **Regression**, and then **Linear**. The following **Linear Regression** window will open.



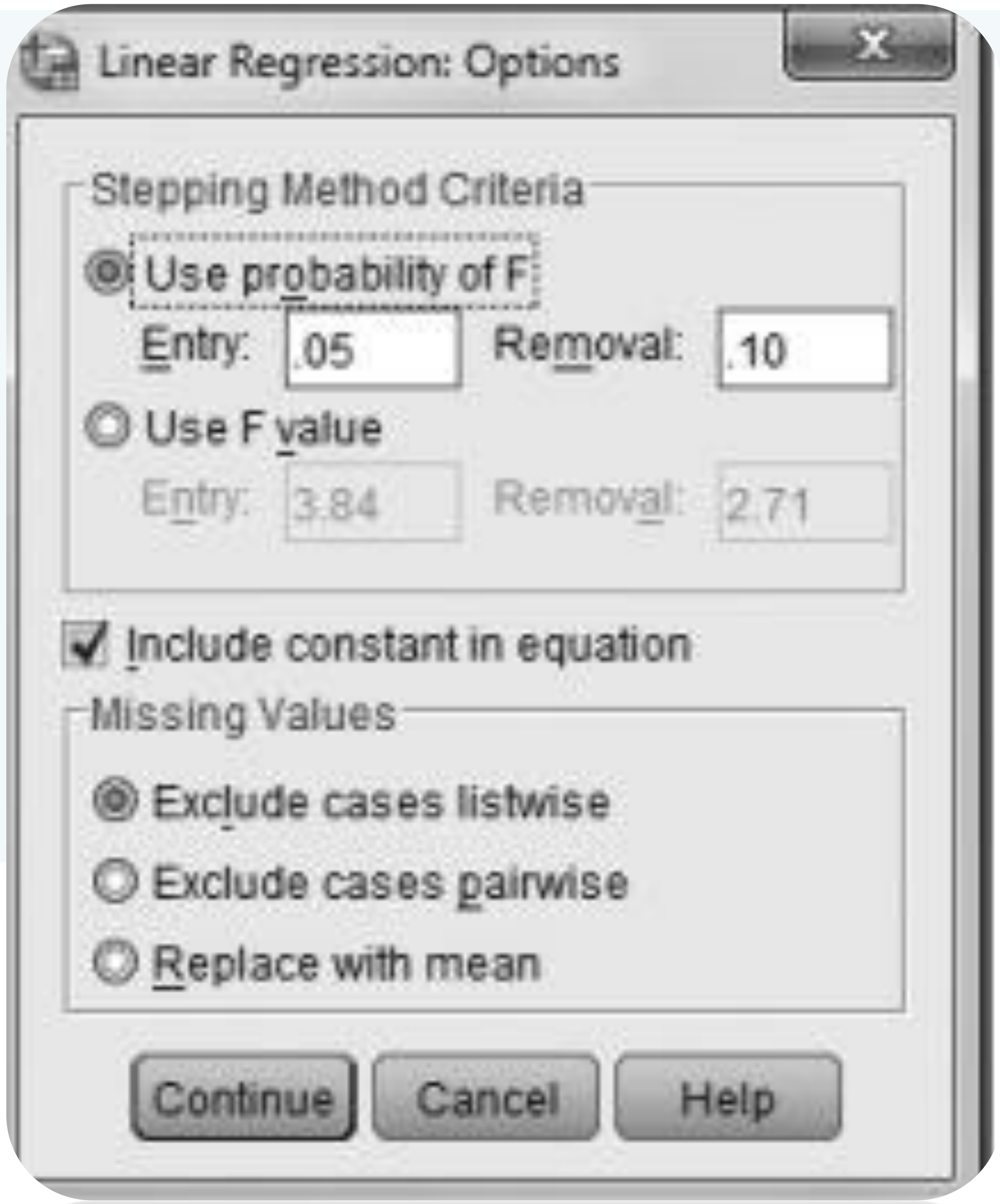
Click (highlight) the **GPA** variable and then click to transfer this variable to the **Dependent:** feld. Next, click (highlight) the **READ** variable and then click to transfer this variable to the **Independent(s):** feld. In the **Method:** feld, select **Enter** from the drop-down list as the method of entry for the independent (predictor) variable into the prediction equation.



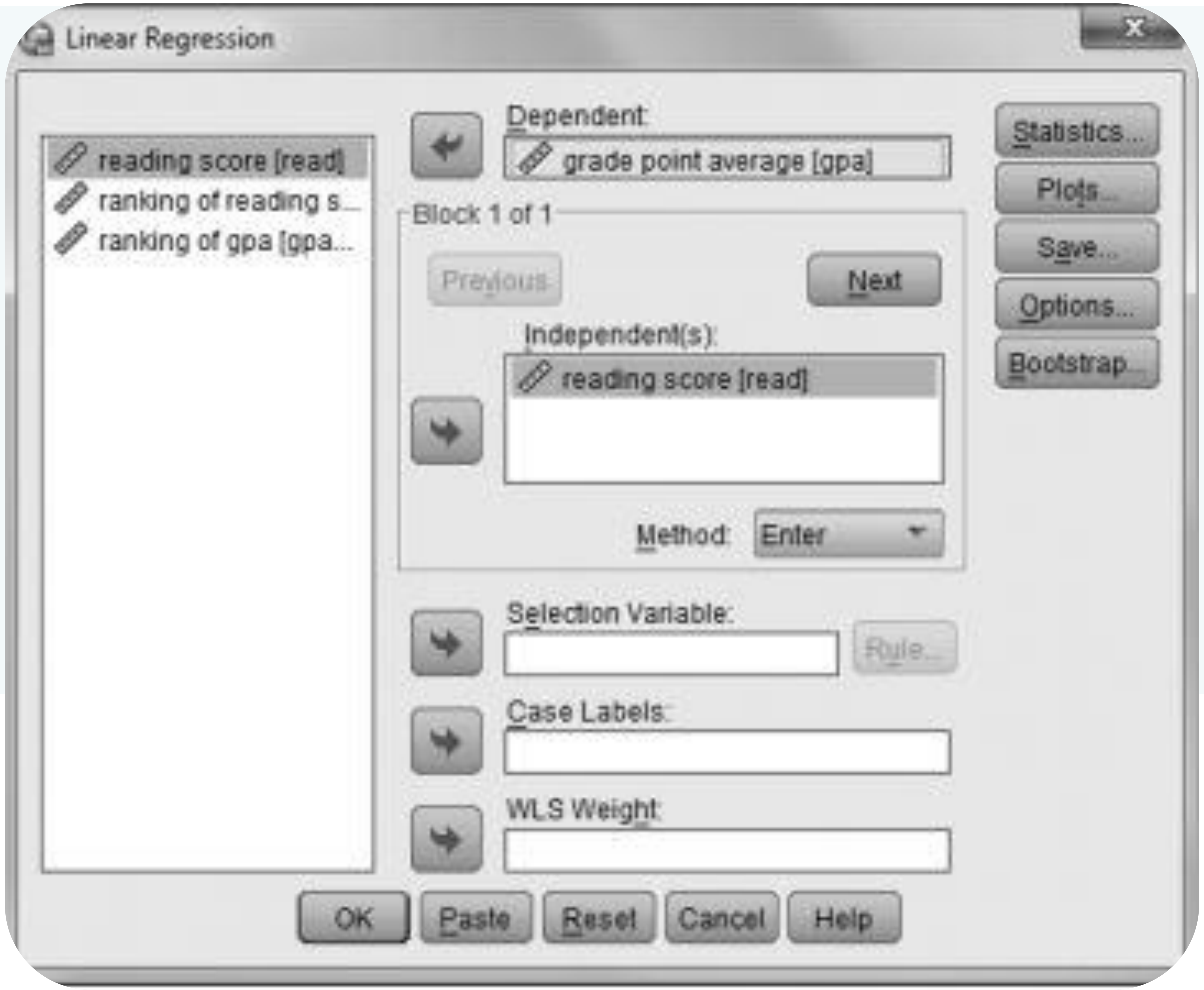
Click to open the **Linear Regression: Statistics** window. Check the felds to obtain the statistics required. For this case, check the felds for **Estimates, Confdence intervals,** and **Model ft**. Click when finished.



When the **Linear Regression** window opens, click to open the **Linear Regression: Options** window below. Ensure that both the **Use probability of F** and the **Include constant in equation** felds are checked. Click to return to the **Linear Regression** window



When the **Linear Regression** window opens, click to complete the analysis.



SPSS Output

Linear Regression Analysis Output Regression

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	Reading score[a]	.	Enter

[a] All requested variables entered.
[b] Dependent Variable: grade point average.

Model Summary

Model	*R*	*R* Square	Adjusted *R* Square	Std. Error of the Estimate
1	.867[a]	.752	.733	.32848

[a] Predictors: (Constant), reading score.

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4.253	1	4.253	.39.418	.000[a]
	Residual	1.403	13	.108		
	Total	5.656	14			

[a] Predictors: (Constant), reading score.
[b] Dependent Variable: grade point average.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			95% Confidence Interval for B	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	–.111	.446		–.248	.808	–1.075	.853
	reading score	.061	.010	.867	6.278	.000	.040	.082

[a] Dependent Variable: grade point average.

Results and Interpretation

The prediction equation is $Y' = A + BX$ where Y' is the predicted dependent variable, A is the constant, B is the unstandardized regression coeffcient, and X is the value of the predictor variable.

The relevant information for constructing a least-squares regression (prediction) equation is presented in the **Coeffcients** table

In order to predict the students’ grade point average scores (**GPA**) from their reading scores (**READ**), use the values presented in the **Unstandardized Coeffcients** column.

Using the **Constant** and **B** (unstandardized coeffcient) values, the prediction equation would be:
Predicted GPA = -0.111 + (.061 × READ)

Thus, for a student who has a reading score of 56, his/her predicted GPA score will be:

Predicted GPA = -0.111 + (.061 × 56) = **3.31**

However, in the **Model Summary** table, the **Standard Error of the Estimate** is 0.32848. This means that at the 95% confdence interval, the predicted GPA score of 3.31 lies between the scores of **2.66** (3.31 - (1.96 × 0.32848)) and **3.95** (3.31 + (1.96 × 0.32848))

Evaluating the Strength of the Prediction Equation

A measure of the strength of the computed equation is ***R*-square**, sometimes called the **coeffcient of determination**. *R*-square is simply the square of the multiple correlation coeffcient listed under ***R*** in the **Model Summary** table, and represents the proportion of variance accounted for in the dependent variable (GPA) by the predictor variable (READ)

For this example, the multiple correlation coeffcient is 0.867, and the *R*-square is 0.752. Therefore, for this sample, the predictor variable of READ has explained 75.2% of the variance in the dependent variable of GPA.

The **ANOVA** table presents results from the test of the null hypothesis that *R*-square is zero. An *R*-square of zero indicates no linear relationship between the predictor and the dependent variable.

The **ANOVA** table shows that the computed *F* statistic is 39.42, with an observed signifcance level of less than 0.001. Thus, the hypothesis that there is no linear relationship between the predictor and the dependent variable is rejected.

Identifying an Independent Relationship

The **Coeffcients** table presents the **standardized Beta** coeffcient between the predictor variable READ and the dependent variable GPA.

The Beta coeffcient is shown to be positive and statistically signifcant at the 0.001 level.

Thus, the higher the students' reading scores, the higher their GPA scores, (Beta = 0.87), $t = 6.28$, $p < .001$. Note that the standardized Beta coeffcient of 0.87 is identical to the multiple R coeffcient. This is because there is only one predictor variable.

REFLEKSI

1. **Informasi penting hari ini**

2. **Manfaat penting dari informasi penting hari ini**

3. **Tindak lanjut yang dapat saudara lakukan**

THANK YOU!

ANY QUESTIONS?